The Soft Neighborhood Model: A Dynamic Enrollment-Balancing Framework

Brooke Cowan
Matt Marjanović
{brooke,maddog}@softneighborhoodmodel.org
http://www.softneighborhoodmodel.org/

October 5, 2015, v1.0

...Before I built a wall I'd ask to know
What I was walling in or walling out,
And to whom I was like to give offense.
Something there is that doesn't love a wall,
That wants it down...

The Mending Wall, Robert Frost

Contents

1 The Problem With Boundaries

1.1 Boundaries are Like Fences

The Soft Neighborhood model is a novel framework for enrollment balancing that is designed to place the right number of students at each school while keeping families in schools close to their homes. The elimination of boundaries is a critical component of the model. By its very nature, a system of hard boundaries drives our school district into a state of enrollment imbalance and reinforces segregation and inequity.

You can think about a boundary like a fence. In the physical world, the lines on the district's boundary maps form fences. We all know where these fences are because they determine to which school community we belong. Redrawing the boundaries means digging up the fence posts and moving the fences. Boundary change is a painful process because these fences, once we erect them, become a strong part of our identity, and we have a lot at stake in them. We resort to boundary change because there are pressing problems that require immediate attention, and a boundary change appears unavoidable. Following a boundary change, we may get some short-term relief; but there has never been a boundary change that has solved enrollment problems in the long-term. This is because:

1. Families move. PPS can decide where to put its fences, but it can't decide where to put families.
2. Families with more resources and more privilege have more freedom to move when and where they want.
3. Moving is not just a matter of economics but also of culture, so some families are less empowered to move than others due to racism and classism.

In short, a district that maintains fences promotes a system of imbalance, choice, and privilege. No matter where you move the fences today to even out enrollment and inequity, people will continue to move around and the inequities and imbalances between one side of a fence and the other will return. That's just what any system of fences always does.

The Soft Neighborhood model starts by acknowledging that boundaries are a social construct that drive the district toward imbalance. Removing boundaries facilitates enrollment balancing by allowing the district to share capacity across nearby schools. It encourages the mixing rather than the segregation of populations, and it curtails the ability of families with more resources to buy into a public school of their choosing. Lastly, eliminating boundaries eliminates the need for boundary change. There are no boundary change events in the Soft Neighborhood model because there are no boundaries.

1.2 A Boundary-Free Solution

The enrollment balancing solution that the district is currently developing involves regular and frequent boundary change: whenever boundary adjustments need to occur, the district wants to make them in a way that is consistent with its equity goals. The trouble is, boundaries are inherently incompatible with equity and balance. PPS boundaries turn school assignment into a commodity: they allow people with means to buy assignment to the public school of their choice when they purchase a house (Figure 1). Boundaries partition Portland geographically and intensify stratification. Aligning school boundaries with any equity policy is deeply contradictory, and any framework that claims it can accomplish this should be viewed with skepticism.

The Soft Neighborhood model is an alternative to a boundary-based solution. The framework is designed to dynamically adapt to shifts in population. When there are more kids for a few years in one area of the city, nearby schools will accommodate them. If there are fewer kids in some area, nearby kids will fill in the gap. The model's ability to fully balance the enrollments at PPS schools is not unlimited. It is constrained by proximity, and so it cannot assign kids to schools clear across town. In other words, the Soft Neighborhood model makes a necessary compromise between proximity and enrollment balancing. However, the model is designed to be as elastic and fluid as possible to accommodate population shifts, which are both inevitable and hard to predict. Also, the model is intended to provide an assignment mechanism for new assignees — students who need a new public school assignment, for example because they have moved or because they are newly entering the system. This implies that the model is not at liberty to re-assign students for the sake of balancing enrollments. Because of these restrictions, there may still be local enrollment imbalances from time to time, but the Soft Neighborhood model should reduce the incidence of enrollment crises significantly in the long run.

The Soft Neighborhood model was designed with increased equitable access to our public schools in mind. Boundaries extend a guarantee and a special privilege to people with more resources. The Soft Neighborhood model dilutes that privilege. Boundary adjustment — even if done frequently — does not. Frequent boundary change will engender zones of predictability (close to schools) and zones of uncertainty (in the margins between schools). Those who have the means will avoid the zones of uncertainty. The Soft Neighborhood model has no such zones. It creates a system of overlapping neighborhoods in which public schools are viewed as communities of families who all live nearby.

In a Hard Boundary model, the geographic space is partitioned into non-overlapping regions, and those regions are inextricably tied to a school community. In the Soft Neighborhood model, school communities overlap geographically. This change provides the district with a very flexible means for achieving enrollment stability across schools from year to year. This is because the model distributes the students in a geographic region over more than just a single school. In the current system, students who live across the street from one another, but on opposite sides of a cluster boundary, will almost never be assigned to the same school (Figure 2, top). This school assignment model makes it possible to say, "*These* kids will never go to school with *those* kids." In contrast, in the Soft neighborhood model, any two children who live close to each other might be assigned to the same school, no matter what sides of the street they live on (Figure 2, bottom). Schools are populated with a mix of nearby students.

1.3 Values in the Soft Neighborhood Model

Central to the Soft Neighborhood model are the notions of *neighborhood*, *family*, *equity*, *stability*, and *consistency*:

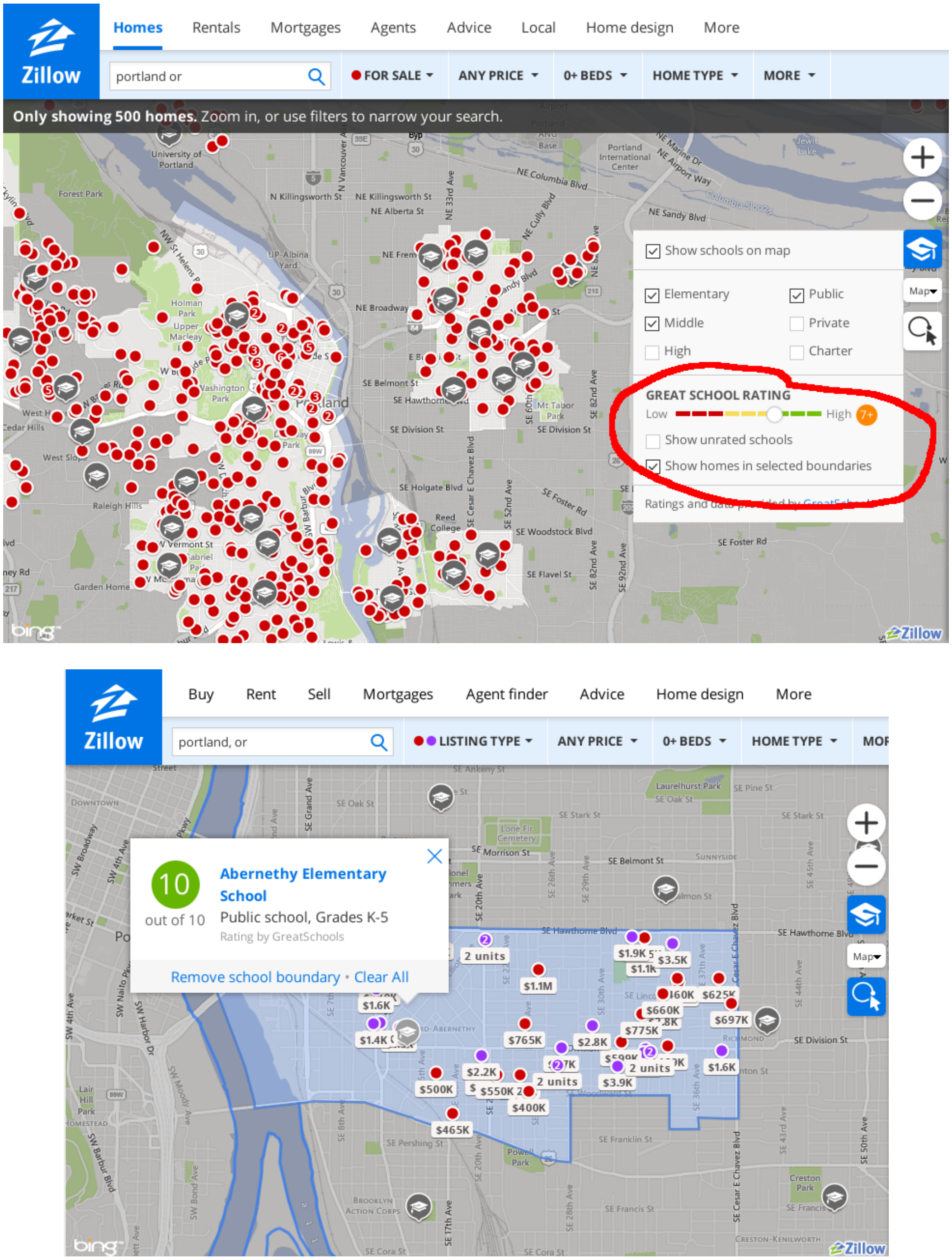


Figure 1: School assignment is a commodity in Portland. People with means can buy guaranteed assignment to the school of their choice when they purchase a house. This is a deeply-rooted source of inequity in our city.

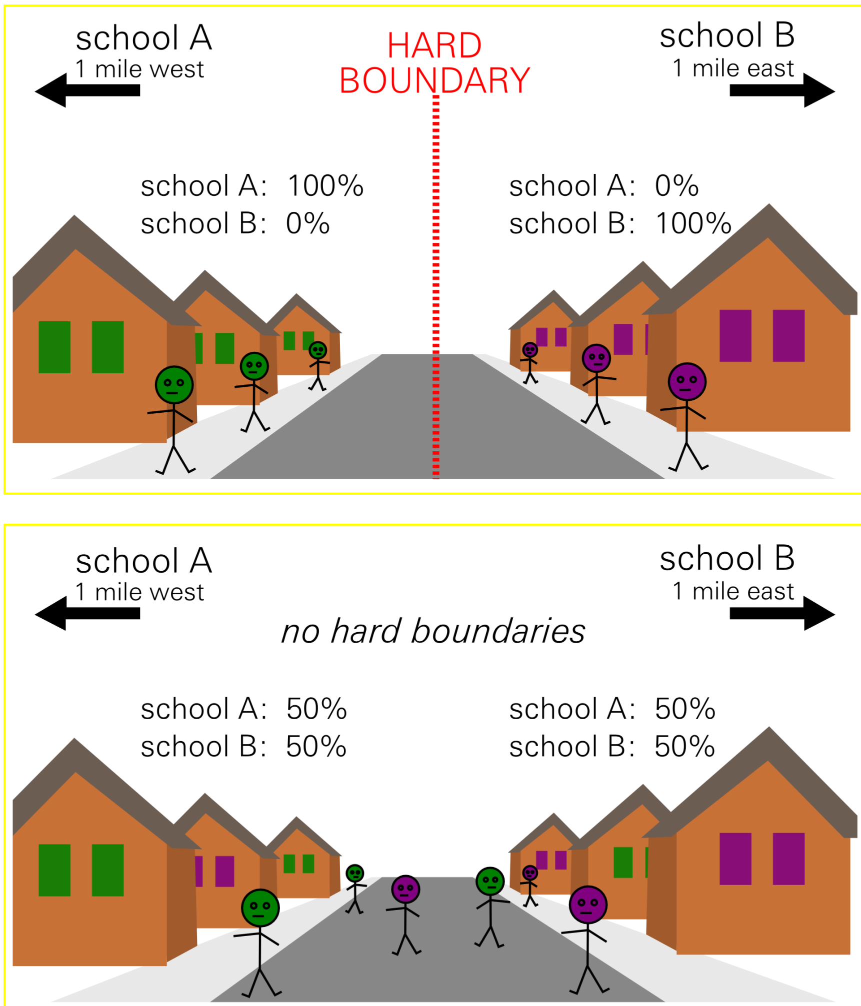


Figure 2: (top) A hard boundary will segregate students on one side of the street from those on the other. (bottom) In the Soft Neighborhood model, hard boundaries don't exist, and schools are populated with a mix of nearby students.

Neighborhood In the Soft Neighborhood model, what is close to your home is your neighborhood, and children are assigned to nearby schools. The neighborhood does not change suddenly by simply crossing the street; families that live close to each other are basically in the same neighborhood. This definition is consistent with an intuitive sense of "neighborhood" and "proximity".

Family Schools are treated as communities of families in the Soft Neighborhood model. The option for families to co-enroll younger siblings with older ones, and the emphasis on proximity between home and school are two concrete ways in which the model expresses this value. Proximity of home and school is critical because it makes it easier for families to participate in school life.

Equity It seems unlikely that there exists a solution to the enrollment balancing problem that can also fix all of the equity problems in our school district. Where possible, the Soft Neighborhood model aims to improve equity of access to and equity of opportunity within our public schools. By removing boundaries, the model reduces the ability of people with more resources to buy into a public school. By addressing enrollment balancing directly, the model avoids chronic under- and over-enrollment, which can become an equity problem. By encouraging mixing from overlapping school communities, the model mitigates the social and economic stratification that can exacerbate equity problems.

Stability The Soft Neighborhood model enables long-term enrollment stability for all schools. By sharing capacity between nearby schools, the model grants the district far more flexibility than it has today to meet capacity-based enrollment targets in the face of unknown population changes. This property guards against both under- and over-enrollment. The underlying assumption is that greater stability in school enrollments leads to better educational environments for everyone.

Consistency Long-term consistency and predictability is a driving feature of the Soft Neighborhood model. The model provides a consistent and stable set of expectations to parents around assignment. In contrast, the current PPS assignment model might appear predictable, except whenever the district needs to make changes in response to crises. These cataclysmic events have unpredictable outcomes (sometimes even forcing families to change schools mid-course, or to decide whether to send siblings to different schools or move an older sibling to a new school) and breed feelings of fear and insecurity. In the Soft Neighborhood framework, the intent is to provide an enrollment balancing mechanism that is robust enough to enable the district to guarantee that once a family is enrolled at a school, their children can remain at that school until they finish.

1.4 Overview of this Document

This document serves two related purposes: it is a proof-of-concept for the Soft Neighborhood model, and it is a proof-of-concept for the development process of any solution to the

PPS enrollment balancing problem. As a proof-of-concept for the model, it addresses three important questions about the model:

1. How far do students have to travel?
2. How well is the model able to stabilize enrollments?
3. To what degree does it result in the mixing of populations?

This document provides answers to these questions that strongly position the Soft Neighborhood model as a viable and robust enrollment balancing solution for PPS.

As a proof-of-concept for a rational process for developing solutions to the enrollment balancing problem, this document defines objective metrics and criteria that can be used to test any proposal, and to evaluate its performance once it's in place. Any proposed solution that is under serious consideration — whether it comes from the community, from the district, or from DBRAC — should be thoroughly evaluated before it is implemented, and should continue to be evaluated after it is implemented. For example, we would like to see the same questions we've asked of the Soft Neighborhood model applied to the Frequent-Boundary-Change solution that the district is promoting. If this solution were implemented on the same data set supplied to us by PPS, how stable would enrollments be? How far would students have to travel? And how diverse would school assignments be? What would the boundaries look like in that seven-year time frame? How much work goes into making those boundary adjustments? Answering these questions objectively, and making predictions and projections ahead of time of what the community can expect to see under some proposed model, are critical steps in the development process.

The remainder of this document is structured as follows: in Section 2, we give a qualitative overview of the Soft Neighborhood model. Section 3 uses a synthetic data set to demonstrate how the model works in action. In Section 4, we run the model on a data set containing seven years of historical PPS data and make three key empirical observations:

1. The Soft Neighborhood model does a far better job of controlling over- and under-enrollment at the kindergarten level.
2. Both models — Soft Neighborhood and historical — assign most students to a school within a reasonable distance of their home.
3. Soft Neighborhood assignments exhibit a higher degree of school assignment diversity compared with historical assignments.

Section 5 explains how the Soft Neighborhood model fits into the larger body of school assignment literature, and in particular how it differs from School Choice models. Section 6 answers a number of frequently-asked questions about the model. Section 7 concludes the work and outlines what we see as the next steps.

2 How Does the Soft Neighborhood Model Work?

The Soft Neighborhood model is a framework for assigning students to schools. The model makes assignments using an algorithm that satisfies schools' capacity constraints, and otherwise is more likely to place students in schools closer to their homes. This algorithm is used to place *new assignees*: students who need a new assignment — for example, because they have moved or because they are entering the system for the first time. School assignment in the Soft Neighborhood model is not something that is meant to be done to every student every year for the purposes of keeping enrollments balanced. Students matriculating up from an earlier grade are pre-assigned by the model, and are not placed using the assignment algorithm. Similarly, when a younger sibling enters the system, families can choose to co-enroll at the same school as the older sibling(s).

In any given year, the Soft Neighborhood model will satisfy the constraints imposed by target capacities at each school to the extent possible. Before assignments are made, the model uses a *proximity function* to compute a set of "nearby" schools for each student based on his or her home address. This set contains the schools to which he or she can be assigned. Since assignment can only be made to one of the schools in this set, the Soft Neighborhood model's ability to fully satisfy capacity constraints is not unlimited; however, in practice, it does a very good job of balancing enrollments even when subject to these restrictions (see Section 4).

One of the open questions that requires further investigation on an improved PPS data set is how to best define the proximity function. For the purposes of this proof-of-concept, we use a *3-School* rule: for each student, assignment can be made to any school within

- 110% of the distance to the 3rd closest school, or
- 1.25 miles,

whichever is larger. This definition has the desirable property that it scales itself to the school-density near each student. Students living in areas of the city with more schools nearby (i.e., more densely-populated regions) will be assigned to closer schools than students living in less densely-populated areas, but every student will have at least three available schools.

2.1 Examples: The Soft Neighborhood Model on the East and West Sides

Let's consider a couple of examples to help illustrate how student assignment actually works under this model. It's instructive to consider one example on the east side of the river, and one on the west side, since the densities in the two regions are so different, and what defines a "nearby" school depends on density.

We'll start with an incoming kindergartner who lives at NE 22nd Ave and Mason, pretty much equidistant from Sabin and Alameda, who would be assigned to Sabin under the current system. Under the Soft Neighborhood model, that student will be assigned to one of several nearby schools. Using the 3-School rule defined above, she can be assigned to Sabin (0.3 mi), Alameda (0.4 mi), Vernon (0.7 mi), King (0.95 mi), Irvington (1.0 mi), or

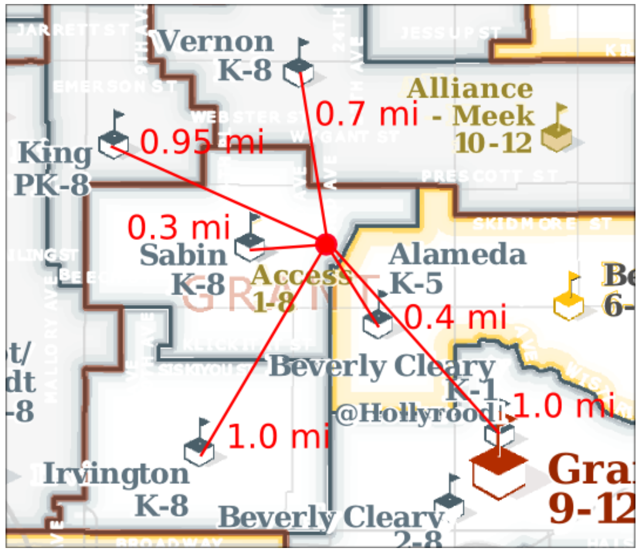


Figure 3: Possible school assignments for a student living at NE 22nd and Mason, with Cartesian distances to each school.

Beverly Cleary Hollyrood Campus (1.0 mi) (Figure 3).[1] Other kids on the same block may be assigned to different schools, but every child will have classmates that live nearby, no matter what school she is assigned to. Families are guaranteed placement at one of these nearby schools, but not a particular school.

Now let's see what happens in an example on the west side of the river. Let's consider an incoming student who lives near SW 57th and Taylor near East Sylvan School. His family currently resides within in the Chapman boundary. In this case, the closest three schools are Bridlemile (1.6 mi), Ainsworth (1.8 mi), and Chapman (2.1 mi).[2] Because the west side is more spread out, the closest three schools are farther away than on the east side. None is within 1.25 miles, and so this student has just those three schools available for assignment. The Soft Neighborhood model will assign the child to one of these three schools, and will optimize the overall assignment process so as to avoid overcrowding or underpopulating any one school.

2.2 The Soft Neighborhood Model in Four Steps

We've seen that the Soft Neighborhood model assigns students to one school in a set of nearby schools. Now let's take a closer look at how it actually picks one of those schools. For most kids (all except for those who are pre-assigned, like siblings and students moving up from the preceding grade), assignment is based on two factors: how far they'd have to travel to get to school from their house, and target enrollments at the schools near their

[1]These are Cartesian distances. Walking distances according to Google maps are 0.4 mi (Sabin), 0.4 mi (Alameda), 0.8 mi (Vernon), 1.3 mi (King), 1.4 mi (Irvington), and 1.4 mi (Beverly Cleary at Hollyrood). The proximity function and the metric used to estimate distance both have a strong impact on how the model actually works when implemented. What works best for Portland needs to be researched and validated against a proper data set.

[2]Driving distances are 3.2 mi (Bridlemile), 3.4 mi (Ainsworth), and 3.0 mi (Chapman).

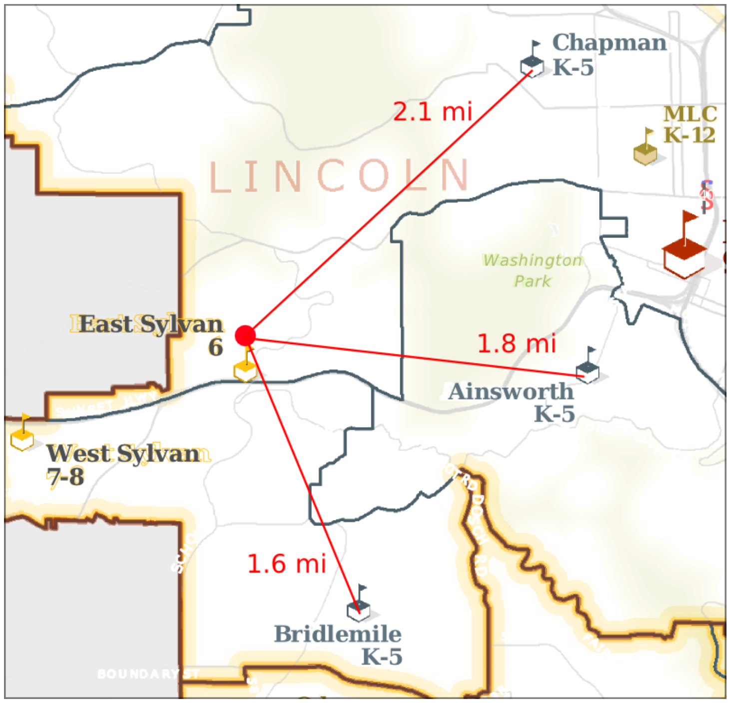


Figure 4: Possible school assignments for a student living near SW 57th and Taylor, with Cartesian distances to each school.

house. The assignment process is an algorithm in which each student is assigned to a nearby school, and the assignment is made by rolling a weighted die. Each side of the die represents one nearby school, and the probability of rolling a school (i.e., being assigned to a school) depends on both proximity and capacity. To the extent possible, the assignment algorithm ensures that once the assignment process is complete and all students have been assigned a school, schools are neither under-enrolled nor over-enrolled. The probabilities will tend to assign kids to schools that are closer to their houses. Living closer to a particular school makes it more likely to be assigned to that school but does not guarantee assignment there, and each school is populated by students who all live nearby.

We can think about the Soft Neighborhood model as a sequence of four steps:

1. **Set Capacity Constraints**: Establish school capacities by grade level,[3] after assigning children needing pre-assignment.

2. **Seed Probabilities by Proximity**: For each new assignee, calculate the proximity-based probability of attending each nearby school.

3. **Balance Probabilities with Capacity Constraints**: Modify the seed probabilities from Step 2 so that the expected number of students at each school equals its target capacity from Step 1.

4. **Assign Students**: Assign children to schools using an algorithm that enforces each

[3]The *capacities* or *target enrollments* are the number of open seats in each grade level at each school.

school's capacity constraint, but otherwise is more likely to place them closer to their home.

Note that it is expected that enrollments will shift over time as families move, etc., and that these shifts will introduce imbalances. The framework keeps the system balanced to the extent possible given the actual students entering the system each year, where they live, and how much space there is at each school.

Now we will look at each step of the model in more detail.

Step 1: Set Capacity Constraints For each school, the district must set per-grade target enrollments, or capacity constraints. For example, the district might decide it wants 3 sections of kindergarten with 25 students each at Duniway Elementary; then the capacity constraint for kindergarten at that school would be 75. Certain students, for example those continuing on from the previous grade, and co-enrolled siblings entering the system for the first time,[4] are pre-assigned to their respective schools, so targets need only account for brand new enrollments.

Step 2: Seed Probabilities by Proximity Given the set of students needing to be assigned to a school, the next step is to seed, or initialize, the probability that each student is assigned to a certain school. It is during this step that we formalize the notion that families should be assigned to schools that are closer to their homes whenever possible. Let's revisit the example from Section 2.1 in which an incoming kindergartner living near NE 22nd and Mason needs to be assigned to one of Sabin, Alameda, Vernon, King, Irvington, or Beverly Cleary. Since Sabin is the closest school (0.3 mi), it will have the highest initial probability. Alameda, which is slightly farther away (0.4 mi), will have a smaller initial probability. Irvington and Beverly Cleary, which are farthest away (1.0 mi), will have the smallest initial probabilities.

Step 3: Balance Probabilities with Capacity Constraints The key idea here is to modify the probabilities from Step 2 to account for the capacity constraints from Step 1. At the end of Step 3, we will be able to run the assignment algorithm in Step 4 and avoid over-enrollment at any particular school, while still preserving the proximity-based preferences from the preceding step as much as possible.

A simple example illustrates why we need this step and how it works. Figure 5 depicts a city with two schools, Green School and Purple School, and 200 kindergarten students waiting to be enrolled. Green School has room for 80 kindergartners; Purple School has space for 120. The students all live in one of two large apartment buildings. 100 students live in a building right in the middle between the two schools. The other 100 students live in the second building located east of Purple School.

If we use the proximity-based probabilities alone to do assignments, then we are likely to overenroll Green School and underenroll Purple School. The kids who live in the middle

[4] Families should be allowed to decline the automatic assignment of co-enrolled siblings and allow a younger sibling to be randomly assigned if they want to for some reason; no harm in allowing that. Note, however, that the intent is not to allow older siblings to join younger siblings at a newly-assigned school. That would give an unfair statistical advantage to families with more children.

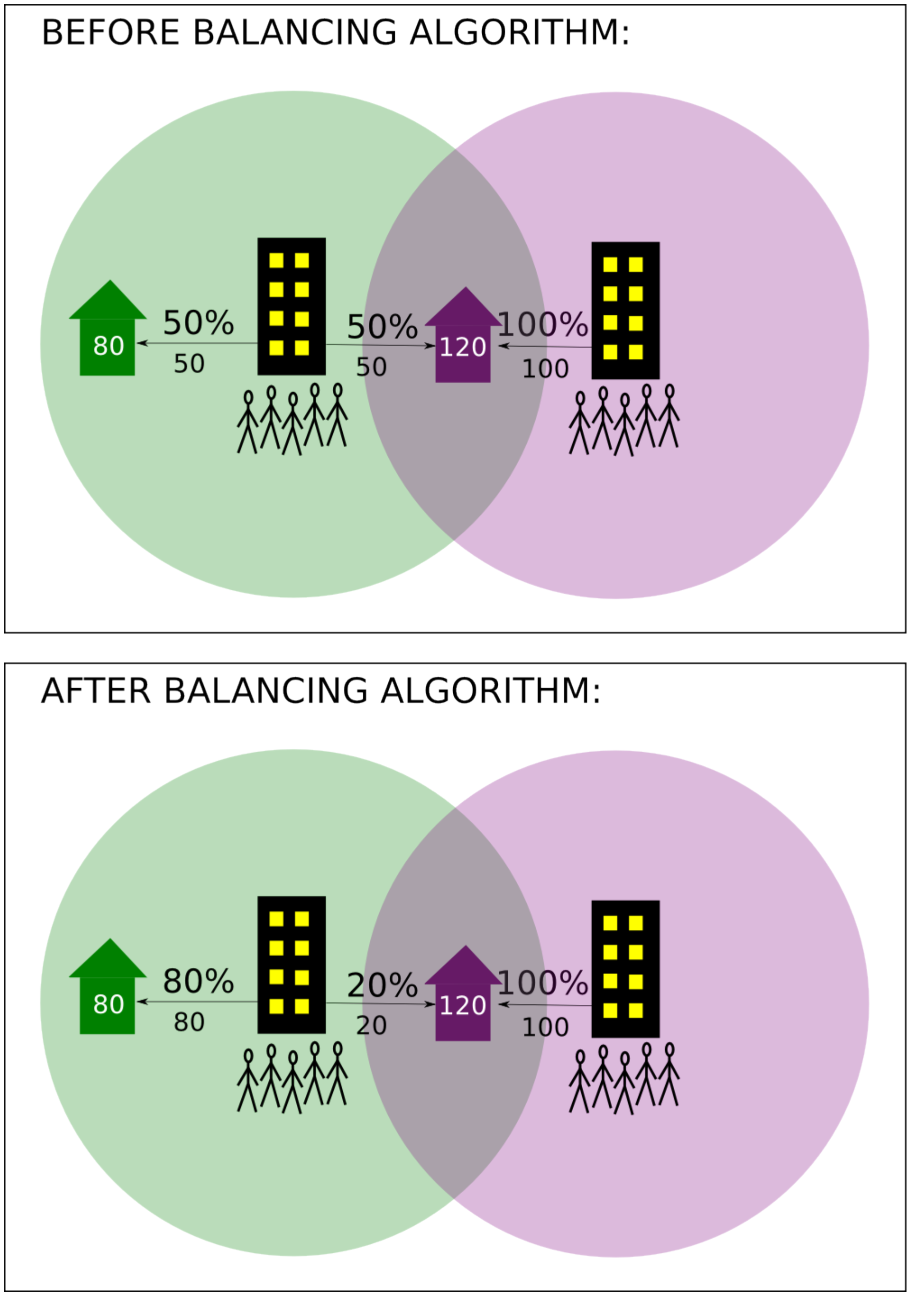


Figure 5: (top) Using raw proximity-based probabilities, the expected number of students at each school does not match target capacities. (bottom) After running the balancing step, expectations match capacities.

are equally likely to attend either school, so we expect 50 to be assigned to Green and 50 to Purple. The kids living east of Purple School are too far away from Green School, thus they all must attend Purple School. Purple School is likely to be assigned 150 students, Green only 50 (Figure 5 top). This is precisely the kind of imbalance we are trying to avoid. The balancing step nudges the probabilities in the right direction, so that we're likely to get enrollments that are aligned with capacity-based targets.

After running the balancing step, we get the modified distribution at the bottom of Figure 5 (the algorithm used to do this is described in detail in Appendix A). Now the kids who live between Green and Purple have an 80% chance of going to Green, and a 20% chance of going to Purple. This means that the expected number of kids at Green and Purple matches the target enrollments of 80 and 120, respectively.

Step 4: Assign Students The final step is to assign students to schools in a way that respects the capacity constraints as much as possible. The balanced probabilities from Step 3 are used to make the assignments. Though the idea behind the assignment algorithm is fairly simple, the algorithm itself is actually fairly complicated. We direct the interested reader to the detailed description in Appendix A.

3 Simulation: An Illustrative Example

To better understand how Soft Neighborhood assignment operates and what it accomplishes, it helps to step through an example. In this section, we demonstrate how the model behaves in an illustrative scenario. The following simulation shows the assignment of kindergarten students in a four-school district over two years. While the population shifts from Year 1 to Year 2, the capacities of the schools are assumed to be constant. In a hard boundary framework, these shifts in population put stress on the district, causing some schools to be over-enrolled, and others to be under-enrolled. In contrast, the Soft Neighborhood framework is a dynamic system that automatically assigns students to schools in a way that maintains appropriate sizing relative to target enrollments. Note that while the simulation shows a sample kindergarten assignment, the intent of the framework is to handle all new assignees in grades K–8[5].

Recall that the Soft Neighborhood model can be described as a sequence of four steps. The pictures in Figures 6–9 show what happens at each step in this four-stage process as kindergartners enter the fictional Colorville Public School (CPS) system over two years. There are four schools in CPS: PinkA, OrangeB, GreenC, and BlueD. BlueD has room for 2 kindergarten classes (50 students);[6] PinkA and GreenC have room for 3 (75 students each school); and OrangeB has room for 4 (100 students). For the purposes of comparing assignments in Year 1 and Year 2, both the total capacity (300) and the number of students entering the system (290) have been held constant over the two years. This allows us to focus on how the system adapts to changes in population density.

We'll now walk through how the Soft Neighborhood model behaves at each step in the process. Figures 7–9 show how each stage in the framework contributes to the final outcome, where students are assigned to schools subject to capacity constraints imposed by the target enrollments, and are more likely to be assigned to schools closer to their homes.

3.1 Step 1: Set Capacity Constraints

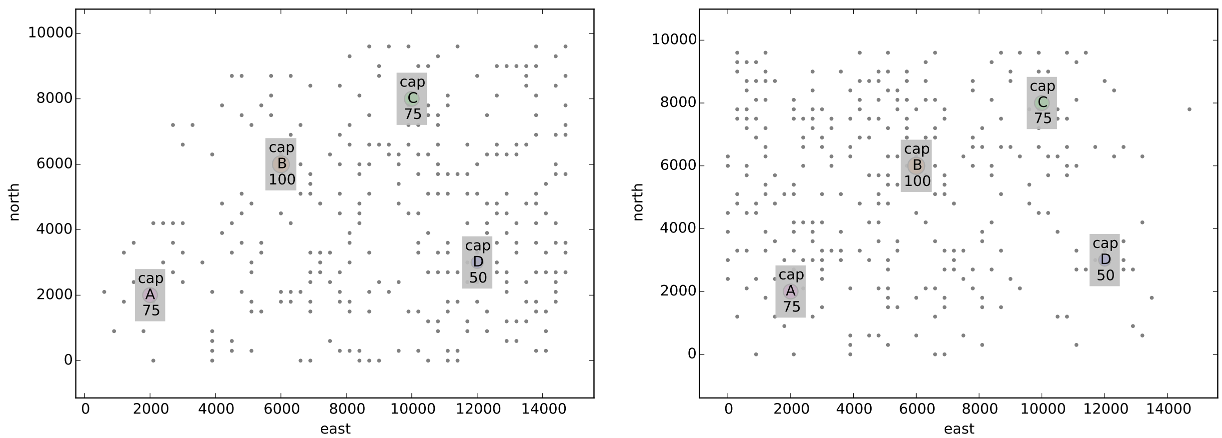
In Year 1, students are more concentrated in the southeast quadrant of the district, while in Year 2, the population shifts toward the northwest. Figure 6 shows how the population of incoming kindergartners is distributed for each year. This represents the state of the system at Step 1 when we set capacity constraints for a set of incoming new assignees.

3.2 Step 2: Seed Probabilities by Proximity

Figure 7 illustrates what the system does during Step 2 in Year 1 when it seeds the probabilities by proximity. The proximity-based probabilities are represented in the picture using colors corresponding to the name of each school. Each colored square in the picture corresponds to one student (i.e., one dot). The intensity of a particular color in a student's square corresponds to the probability of attending the school with that name. Figures 7(a)

[5]Recall that *new assignees* are students who need to be placed at a PPS school, for example, because they moved or because they are entering the system for the first time.

[6]Throughout this work, we use a target of 25 students per kindergarten section — a reasonable target for kindergarten based on conversations with PPS staff.



(a) Year 1: Population skewed to SE.

(b) Year 2: Population skewed to NW.

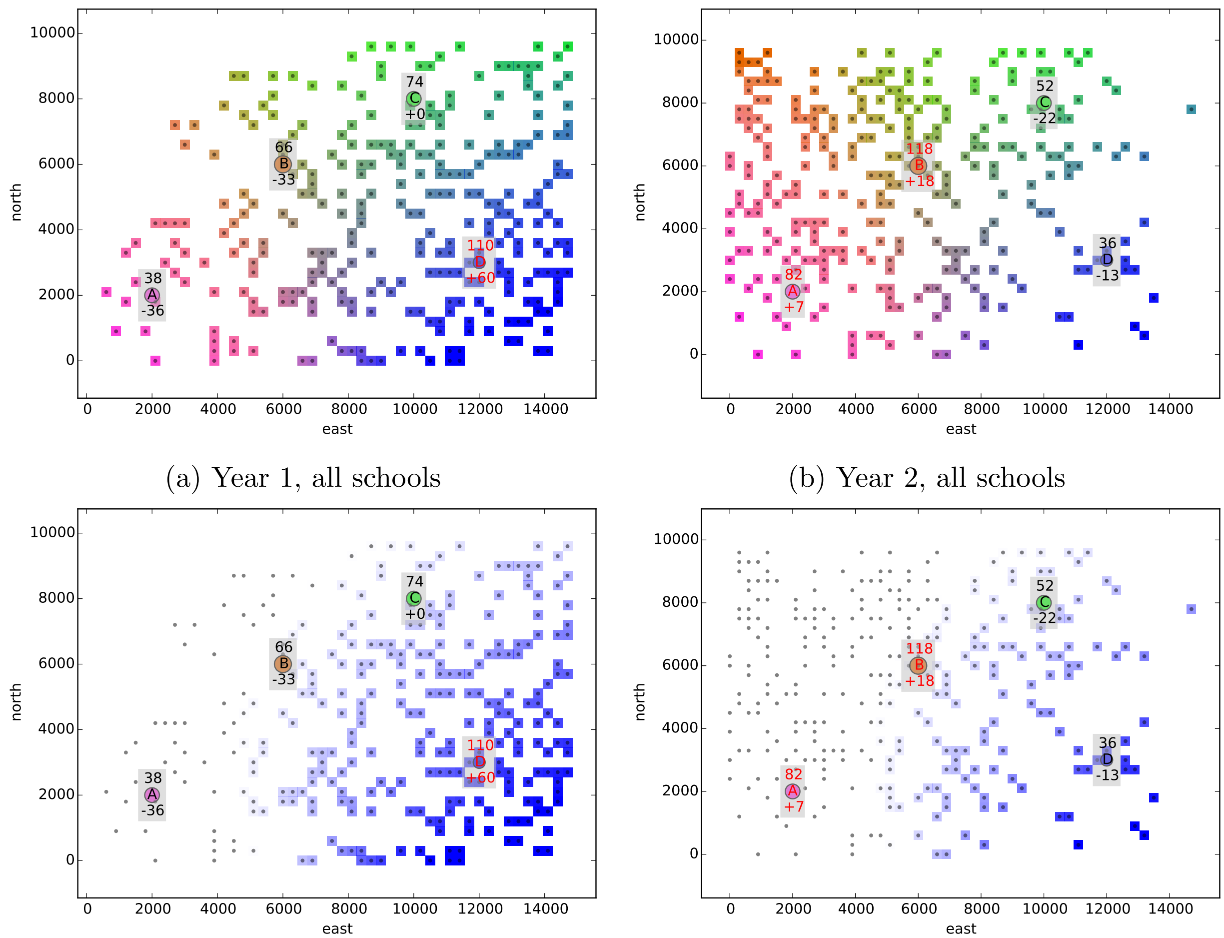
Figure 6: *Set Capacity Constraints.* This depicts the state of the Soft Neighborhood system in the first step of the model. The gray rectangles indicate the location of the four schools, annotated with their target kindergarten capacities. PinkA and GreenC have seats for 75 kids; OrangeB has room for 100 kids; BlueD has room for 50 kids. The total capacity of this district is thus 300 students.

Each gray dot represents a student needing to be assigned to a school, and both Year 1 and Year 2 have 290 incoming students to place. However, the students are not uniformly distributed, and the distribution changes from one year to the next. In Year 1, the students are concentrated toward the SE; in Year 2, they are concentrated toward the NW.

and (b) show the blended probabilities for all four schools at once. Squares that are more monochromatic (blue, pink, orange, or green) signify that there is one school with very high probability in that student's neighborhood. Squares that are combinations of colors signify that there are two or more schools with significant probability. Based on distance alone, a student living between OrangeB, GreenC, and BlueD is equally likely to attend any of those three schools, independent of their capacity or of the distribution of the population.

For clarity, in Figures 7(b) and (c) we show only the probability of attending BlueD. The squares of the students in the district's southeast quadrant, closest to BlueD, are more intensely blue. As we move toward the center of the district, the blue fades. This gradient signifies the likelihood of attending BlueD School, according to the proximity-based probabilities: students who live close to BlueD are more likely to attend that school.

The numbers in the gray rectangles in Figure 7 tell us the number of students expected to attend each school according to the seed probabilities (top), and how over- or under-enrolled each school is expected to be (bottom). Note that in Year 1, because there are more students living near BlueD, we expect BlueD to be overcrowded according to the proximity probabilities. Likewise, because there are fewer students living near BlueD in Year 2, we expect it to be under-enrolled. Under- and over-enrollments for both years will be corrected in the next step.



(a) Year 1, all schools

(b) Year 2, all schools

(c) Year 1, BlueD only

(d) Year 2, BlueD only

Figure 7: *Seed Probabilities by Proximity.* This is the result of Step 2. In (a) and (b), each colored square represents one student in Year 1, and the color is a blend of the schools' colors, weighted by the probabilities. Students in the SE corner of the district (intensely blue) are highly likely be assigned to BlueD, whereas those living between schools (blended colors) are equally likely to be assigned to one of several schools. In (c) and (d), for clarity we show only the probabilities of attending BlueD.

The schools are now annotated with two numbers. The top number is how many students we expect to assign to the school given the proximity-based probabilities. The bottom number is the expected over- or under-enrollment relative to the original target capacity. In Year 1, BlueD is expected to receive 110 students, when it only has room for 50. PinkA and OrangeB are expected to be severely under-enrolled. In Year 2, BlueD and GreenC are expected to be under-capacity, and PinkA and OrangeB are expected to be over-capacity. This will be corrected for both years in the next step.

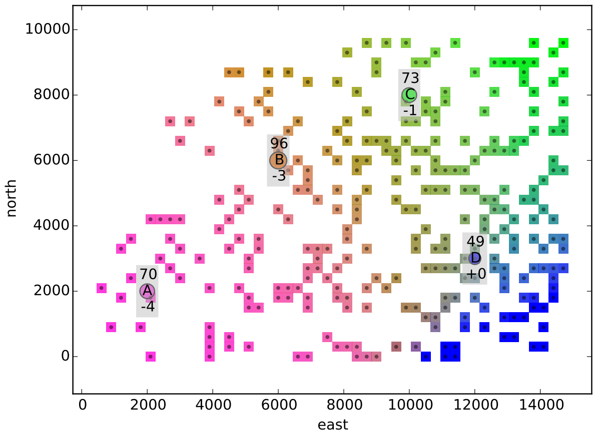
3.3 Step 3: Balance Probabilities with Capacity

In Step 3, we bring the probabilities into alignment with target capacities. Figure 8 depicts the changes to the student probabilities that result from executing Step 3. Comparing

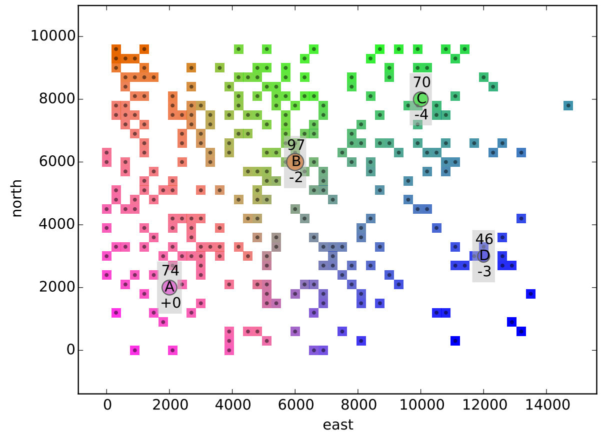
Figures 8(a) and (b) with Figures 7(a) and (b), we see that the schools that were over- or under-filled in the previous step now have expected population sizes that are well-aligned with their target capacities. In Figure 8(c), we see that in Year 1, BlueD is at capacity with an expectation of 49. It is now much less likely (though still possible) for students who live between OrangeB, GreenC, and BlueD to be assigned to BlueD. In contrast, in Year 2 (Figure 8(d)), we see that the balancing step has slightly increased the likelihood of attendance at BlueD for students living near the center of the district. This slight increase in probability brings the expected enrollment at BlueD into alignment with its target capacity.

3.4 Step 4: Assign Students

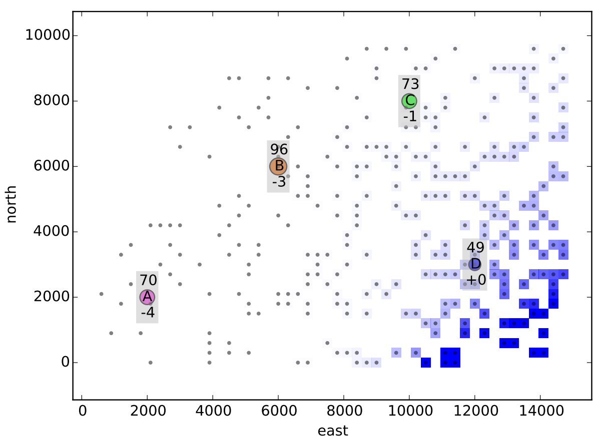
Finally, we see what happens after Step 4, when we make assignments. In Figure 9, each square is now a single, solid color corresponding to the school to which that student has been assigned. The numbers in the gray rectangles now denote the size of the kindergarten class at each school (top) and the number of unfilled seats (bottom). On the left-hand side of the figure, we see assignments over all schools. On the right-hand side of the figure, we see how Blue's student population is distributed. Year 1 results are on the top, and Year 2 results are on the bottom. Crucially, no school is over-enrolled, and empty seats are distributed over all schools. Furthermore, the student populations are blended: there are no dividing lines that separate Blue students from Orange students or Pink students from Green students.



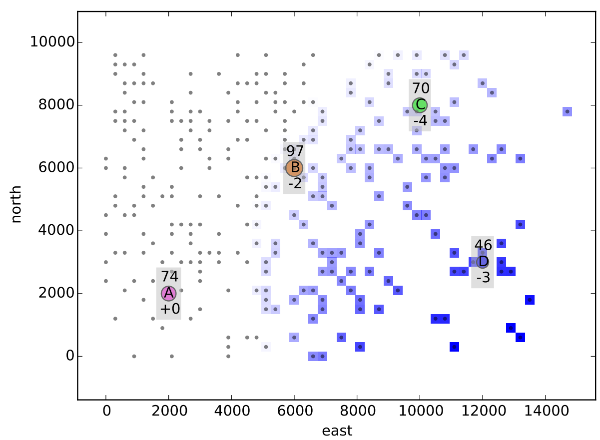
(a) Year 1, all schools



(b) Year 2, all schools

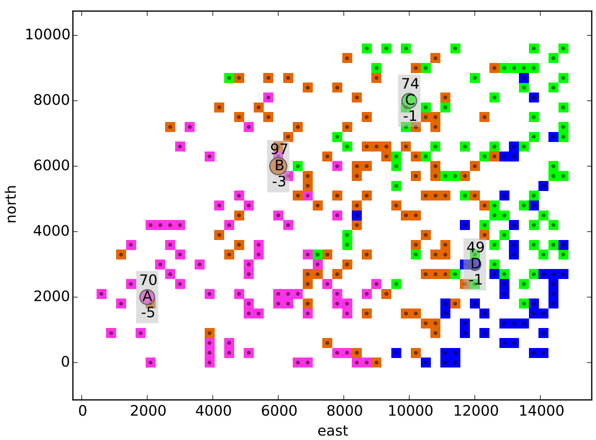


(c) Year 1, BlueD only

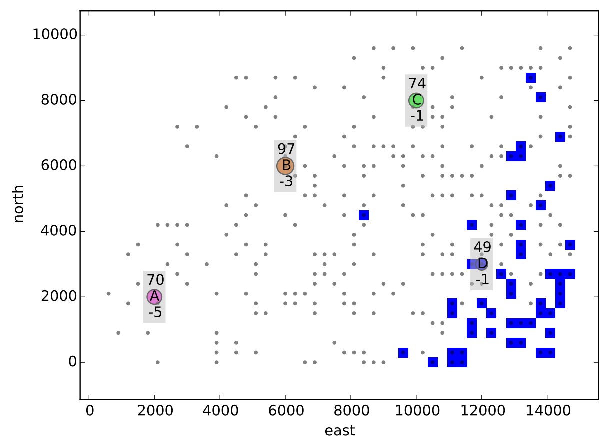


(d) Year 2, BlueD only

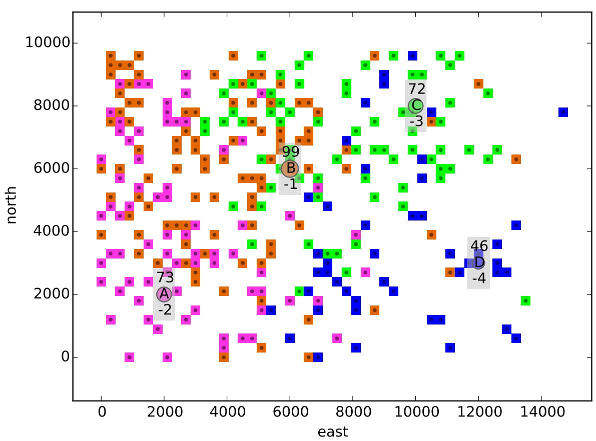
Figure 8: *Balance Probabilities With Capacity.* This is the result of the capacity balancing step. As in Figure 7, the colors reflect each student's probability of being assigned to the four schools, and (c) and (d) show only BlueD, for clarity. The balancing step shifts the probabilities around, minimally, until no school is expected to be over-capacity. Compared to Figure 7(c), students in the far southeast corner are still very likely to be assigned to BlueD, but students living between BlueD, OrangeB, and GreenC are now less likely to be sent to BlueD. The numbers on the schools reflect these shifts. BlueD is now expected to be at-capacity, and the other three schools are expected to be very slightly under-capacity — which is reasonable since there are only 290 students to fill 300 seats. The changes in Year 2 (c) are less dramatic, but the balancing step has increased the overall probability of attendance at BlueD enough to bring the expected number of students into alignment with its capacity.



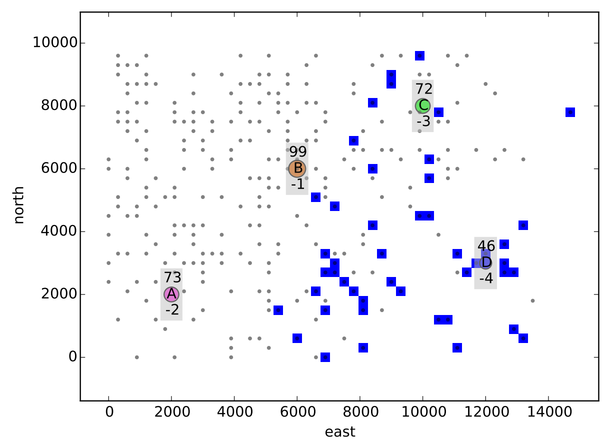
(a) Year 1, all schools



(b) Year 1, BlueD only



(c) Year 2, all schools



(d) Year 2, BlueD only

Figure 9: *Assignment.* This is the result of the final assignment step. The color of each square indicates the school to which that student has been assigned. In the final tally, no school is over-capacity; every school is very slightly under-capacity, consistent with having fewer students than seats. Students tend to live near other students assigned to the same school. Also, note the high degree of mixing: students from any given area are generally assigned to a mix of different schools, with no clear boundary isolating them from one another.

4 Results on Real Data

In the previous section, we demonstrated how the Soft Neighborhood model adapts to population shifts in a synthetic data set. In this section, we apply the Soft Neighborhood model to real PPS data provided by the district. While the data set contains flaws that limit our ability to fully validate the Soft Neighborhood model, we've used the data we have to run simulations and make some empirical observations of how the model operates on kindergarten enrollment.

In this section, we define metrics for evaluating the Soft Neighborhood model along three dimensions: enrollment balancing, travel distance, and school assignment diversity. Using these metrics to compare the Soft Neighborhood model to the current Hard Boundary model on historical data, we report three major findings:

1. Enrollment Balancing: The Soft Neighborhood model does a far better job of controlling over- and under-enrollment at the kindergarten level.

2. Travel Distance: Both models — Soft Neighborhood and historical — assign most students to a school within a reasonable distance of their home.

3. Assignment Diversity: Soft Neighborhood assignments exhibit a higher degree of school assignment diversity compared with historical assignments.

These results demonstrate that the Soft Neighborhood model is an effective means for dynamically balancing enrollments that requires reasonable travel distances and results in increased assignment diversity.

In interpreting these results, it is important to bear in mind two points. First, which we alluded to above, is that the data set we are using contains defects which make it impossible for us to fully validate the model. Appendix B contains detailed information about the data, including its flaws and how the district can fix them without too much effort. Second, how the Soft Neighborhood model is configured strongly influences its behavior. The model has a configuration parameter — the *proximity function* — that is used to calculate the set of schools to which a student may be assigned, and to set their initial proximity-based probabilities. The function we use for these simulations (defined formally in Appendix A) strongly influences the outcomes we observe in our experiments. These outcomes and our interpretations of them are best viewed as an empirical demonstration of the model's potential as a robust and equitable solution to the enrollment balancing problem in PPS. In short, there is follow-up work that needs to be done to fully validate the model, and more in-depth tuning (via the proximity function) that needs to be explored in order to develop a version of the model that works optimally for Portland.

The analysis that we undertake in this section serves two purposes. One, as we mentioned, is to demonstrate the Soft Neighborhood model's potential. The other is to establish an experimental framework for the evaluation of any solution to the enrollment balancing problem. This section defines objective metrics and criteria that can be used to test any proposal, and to evaluate its performance once it's in place. Any proposed solution that is under serious consideration — whether it comes from the community, from the district, or from DBRAC — should be thoroughly evaluated before it is implemented, and should continue to be evaluated after it is implemented. The availability of a corrected data set in the public domain would be a huge step toward realizing this goal. Anyone who has a solution to enrollment balancing could implement and test it using a common set of metrics and a common data set. Anyone who wanted to evaluate a proposal could do the same. A methodology that compares different systems on a common data set and using a common set of metrics generates comparable and meaningful results.

4.1 An Evaluation Framework

The simulations in this section were run using data provided by PPS in two data sets, the Students data set and the Schools data sets (see Appendix B for details). The most valuable of the two is the Students data set, which contains student data from seven years of PPS history (2008–15). In this section, we establish a framework for running and evaluating simulations of the Soft Neighborhood model against the Students data set. This framework also establishes criteria for comparing the results of the simulations to the current PPS assignment framework as evidenced by the historical data.

4.1.1 Methodology

There are a few methodological decisions we've made that should be taken into consideration when interpreting the results below:

Kindergarten only The simulations in this section have been run on kindergarten data only. Given the limitations in the Students data set (see Appendix B), there is no way to run meaningful simulations for any grades beyond kindergarten.

Neighborhood programs only In our historical analysis, we consider every student attending a neighborhood school to be enrolled in a neighborhood program. The data set contains anonymized information about all students who attend each neighborhood school, but it does not provide any way to distinguish between those who attend neighborhood programs and those who attend co-located focus option programs. Also, the data set does not contain any information about focus option schools. This means that the simulations are unable to account for the students who attend these schools.

No siblings Because sibling relationships are not available in the data set (see Appendix B), every kindergarten student to be enrolled is considered a non-sibling in our simulations.

No transfers All students are assigned to a nearby school in our simulations of the Soft Neighborhood model. In other words, we do not maintain historical transfers, assigning all kindergartners based on their home address.

Distances The distances computed between students' home addresses and schools are "as-the-crow-flies" Cartesian distances. In general, the use of Cartesian distances means that students appear to travel smaller distances under both the current system and the Soft

Neighborhood system. However, since both systems are evaluated using this definition, we can draw meaningful conclusions about how the two compare to one another according to the travel-distance metric defined in Section 4.1.3.

We acknowledge that using Cartesian distances is not the best approximation for real travel distances, and that a metric based on minimum driving distances or even travel time would be better. Note that for the Soft Neighborhood model, using a more realistic distance metric would result in different assignment outcomes under the definition of the *proximity function* as discussed below. We used Cartesian distance because it is simple and quick to implement. We intend to use a more realistic map-based distance metric in future analysis, and we expect that PPS would do the same with the analysis and implementation of this or any other assignment model.

Setting target thresholds A crucial question in implementing the simulations was how to set target enrollments. The method used is as follows: for each year, and at each school, we look at the total number of kindergarten students enrolled. This enrollment is considered a proxy for how many kindergarten students can comfortably fit at a school.[7] The number of desired sections is then derived by dividing the total enrollment by 25 (a reasonable target number of kindergarten students per section based on conversations with PPS staff), and rounding the result to the nearest integer. For example, a school with 38 kindergartners will have a target of 2 sections since $38/25 = 1.52$.

4.1.2 Configuring the Soft Neighborhood Model

As mentioned previously, the Soft Neighborhood model has a required function that needs to be evaluated for each student: the proximity function. For the purposes of these simulations, the proximity function uses the *3-School* rule defined in Section 2: for each student, assignment can be made to any school within (1) 110% of the distance to the third closest school, or (2) 1.25 miles, whichever is larger. This definition of the proximity function has the desirable property that it scales itself to the school-density near each student. Students living in areas of the city with more schools nearby (i.e., more densely-populated regions) will be assigned to closer schools than students living in less densely-populated areas, but every student will have at least three possible schools. The method used for seeding the initial proximity-based probabilities of these schools is defined in Appendix A.

4.1.3 Metrics

An important aspect of evaluating the Soft Neighborhood model is to define a set of objective metrics with which to compare the results to the current assignment framework. In this document, we define three metrics and use them to report some initial results:

Enrollment Balancing This measures how well a system hits enrollment targets. The Enrollment metric is implemented by calculating the difference between the actual enrollment

[7] Of course, this is a major oversimplification since the historical numbers may not reflect the actual capacity at each building. In practice, this means that the targets don't necessarily reflect the degree of suffering at particular schools.

and the target enrollment of 25 students in each kindergarten section. For example, a K section with 28 students is +3 over-enrolled, and a K section with 20 students is −5 under-enrolled. This number is considered to be that section's *over/underenrollment*. Results are reported by calculating the cumulative mass function (CMF) over the set of ⟨*section, over/underenrollment*⟩ pairs. Table 1 compares the Soft Neighborhood model to the current system in terms of the Enrollment Balancing metric.

Travel Distance This measures how far students have to travel from their home to get to school. In this document, the Travel Distance metric has been implemented using Cartesian distance, as discussed above. To obtain results for the Soft Neighborhood model using this metric, we calculate the CMF of expected travel distances as follows: first, we look at the probabilities after the balancing step (*Balance Probabilities with Capacities*) and collect ⟨*assignment-probability, distance*⟩ pairs over all students, where *assignment-probability* is the probability of a student being assigned to a particular school. Results are then obtained by calculating the CMF over the resulting set of pairs by year, where the mass of each pair is weighted by the probability. Table 2 compares the Soft Neighborhood model to the current system in terms of the Travel Distance metric.

Assignment Diversity This measures how many schools are represented within 1000 ft of each student's home. In other words, how diverse are the school assignments near any student? To implement this metric, we calculate the ${}^{1}D$ diversity[8] (exponentiated Shannon entropy) of the set of schools assigned to all students living within a circle of radius 1000 ft of each student. Results are reported by calculating the CMF over the set of ⟨*student, diversity*⟩ pairs. Since this needs to be performed on actual assignments and not expectations, we extract pairs over four simulated assignments and aggregate the results.

4.2 Results and Discussion

Tables 1, 2, and 3 show results comparing the Soft Neighborhood model to historical assignments. The Soft Neighborhood model has been configured as described above, and there are two historical models for the purposes of comparison: one that includes inter-neighborhood transfers (*historical*), and one that does not (*no transfers*). In the "no-transfer" version, all transfer students are reassigned to their original neighborhood schools. The no-transfer model better reflects the intended future state of PPS's Hard Boundary assignments, since the transfer lottery was eliminated beginning with the 2015–16 school year. Note that the district continues to grant transfers via petition; our historical dataset does not distinguish between lottery and petition transfers. Each table shows results for two years independently (2008–09 and 2014–15) as well as aggregated results over all seven years (2008–15).

Table 1 shows results for the Enrollment Balancing metric. Looking at all seven years (third section in the table), we see that the Soft Neighborhood model is much better at hitting enrollment targets than either historical model. It hits the 25-student enrollment target about 42% of the time, and is +/−2 (23–27) students 96% of the time. In contrast, even with transfers, historical assignments are either extremely under (<23 students) or

[8]See https://en.wikipedia.org/wiki/Diversity_index.

over (> 27 students) target around 42% of the time. Without transfers, under- and over-enrollment is much worse (61%). Similar observations hold in the two independent years shown (first two sections). Note that 2014–15 appears to have been a more difficult year for enrollment balancing: historically, even with transfers, only around 7% of sections were at the 25-student target (c.f. 40% for the Soft Neighborhood model). However, even in a year when balancing is difficult, the Soft Neighborhood model avoids extreme under- and over-enrollment, with 98% of sections within the range of 23–27 students.

Year(s)	**Model**	**Kindergarten Section Sizes**				
		<23	23–24	25	26–27	>27
2008-09	Hard Boundary, historical	21.7%	28.7%	14.0%	16.8%	18.9%
	Hard Boundary, no transfers	37.8%	12.6%	4.9%	7.7%	37.1%
	Soft Neighborhood	1.4%	24.5%	55.9%	17.5%	0.7%
2014-15	Hard Boundary, historical	28.4%	26.4%	6.8%	17.6%	20.9%
	Hard Boundary, no transfers	35.1%	19.6%	4.7%	12.2%	28.4%
	Soft Neighborhood	2.0%	42.6%	39.9%	15.5%	0.0%
2008-15	Hard Boundary, historical	22.4%	27.9%	11.0%	19.0%	19.7%
	Hard Boundary, no transfers	31.5%	16.7%	6.3%	15.6%	30.0%
	Soft Neighborhood	2.3%	32.8%	41.9%	21.3%	1.7%

Table 1: Under- and over-enrollment of kindergarten sections for two years (2008-09 and 2014-15), and over all 7 years, for the current system and the Soft Neighborhood model. The Soft Neighborhood model is much better at hitting enrollment targets than the current hard-boundary framework. If we look at the data over all seven years (third section of the table), the Soft Neighborhood model comes within $+/-2$ of the 25-student target (23–27 students) 96% of the time. The historical models fall within this range only 58% of the time (with transfers) and 39% (without transfers).

Table 2 shows results for the Travel Distance metric. Here we include not just statistics for the district as a whole (top), but also for the east side of the river (middle) and west side of the river (bottom) independently. Overall, the Soft Neighborhood model assigns students to a school within a reasonable distance from home. If we look at the district as a whole, over all seven years, we see that under the Soft Neighborhood model, around 33% of students travel less than 0.5 mi, about 81% of students travel less than 1.0 mi, and about 96% travel less than 1.5 mi. Recall that in our simulations, the Soft Neighborhood model does not maintain historical transfers, so it's most instructive to compare it against the historical no-transfer model. Here, we see that around 55% of students travel less than 0.5 mi, around 89% travel less than 1 mi, and around 97% travel less than 1.0 mi. More students travel very short distances under the current system, but both assignment models keep most students within reasonable distances of their homes.

Year(s)	Model, whole district	Distance from Home to School			
		<0.5mi	<1.0mi	<1.5mi	>1.5mi
2008-09	Hard Boundary, historical	47.7%	78.1%	87.6%	12.4%
	Hard Boundary, no transfers	55.9%	89.3%	97.2%	2.8%
	Soft Neighborhood	32.7%	80.6%	95.8%	4.2%
2014-15	Hard Boundary, historical	46.3%	78.5%	88.5%	11.5%
	Hard Boundary, no transfers	53.3%	87.8%	96.7%	3.3%
	Soft Neighborhood	31.2%	79.0%	95.7%	4.3%
2008-15	Hard Boundary, historical	47.1%	78.2%	88.4%	11.6%
	Hard Boundary, no transfers	55.4%	88.8%	97.3%	2.7%
	Soft Neighborhood	32.5%	81.0%	96.2%	3.8%
Year(s)	**Model, east of river**	**Distance from Home to School**			
		<0.5mi	<1.0mi	<1.5mi	>1.5mi
2008-09	Hard Boundary, historical	50.8%	80.7%	88.8%	11.2%
	Hard Boundary, no transfers	60.3%	93.0%	99.2%	0.8%
	Soft Neighborhood	33.8%	84.3%	98.8%	1.2%
2014-15	Hard Boundary, historical	49.2%	81.4%	89.8%	10.2%
	Hard Boundary, no transfers	57.6%	91.8%	98.6%	1.4%
	Soft Neighborhood	32.3%	82.3%	98.0%	2.0%
2008-15	Hard Boundary, historical	49.9%	80.4%	89.3%	10.7%
	Hard Boundary, no transfers	59.4%	92.2%	98.8%	1.2%
	Soft Neighborhood	33.6%	84.5%	98.7%	1.3%
Year(s)	**Model, west of river**	**Distance from Home to School**			
		<0.5mi	<1.0mi	<1.5mi	>1.5mi
2008-09	Hard Boundary, historical	35.4%	68.0%	83.2%	16.8%
	Hard Boundary, no transfers	38.8%	74.7%	89.5%	10.5%
	Soft Neighborhood	28.5%	66.5%	84.0%	15.9%
2014-15	Hard Boundary, historical	35.7%	68.0%	83.7%	16.3%
	Hard Boundary, no transfers	37.7%	73.5%	90.0%	10.0%
	Soft Neighborhood	27.2%	67.0%	87.5%	12.4%
2008-15	Hard Boundary, historical	35.6%	68.9%	84.6%	15.4%
	Hard Boundary, no transfers	38.6%	74.8%	90.7%	9.3%
	Soft Neighborhood	27.8%	66.2%	85.7%	14.3%

Table 2: Travel distance from home to school for two years (2008-09 and 2014-15), and over all 7 years, for the current system and the Soft Neighborhood model. Overall, more students travel very short distances under the current system, but both assignment models keep most students within a reasonable distance of their home.

Table 3 shows results for the Assignment Diversity metric. Again, we see the whole district as well as the breakdown for the east and west sides. The Soft Neighborhood model displays greater assignment diversity in all cases: the fraction of students who have only one school represented among all students within 1000 ft is only 10.8% for the district as a whole. The fraction with two or more schools well-represented is 64.2%, and the fraction with three or more schools is 24.3%. The distributions for the historical models without transfers are highly skewed toward only 1 school. This makes sense because in most regions, students are assigned to the same school. Only at boundaries will there be regions with even the semblance of mixing. For the historical model with transfers, the distributions show greater diversity, though not as much as the Soft Neighborhood model. Note that we achieve this diversity in the Soft Neighborhood model without any transfers.

Year(s)	Model, whole district	Assignment Diversity		
		≤1.0	≥2.0	≥3.0
2008-09	Hard Boundary, historical	22.7%	43.5%	15.3%
	Hard Boundary, no transfers	57.6%	7.6%	0.3%
	Soft Neighborhood	10.8%	64.2%	24.3%
2014-15	Hard Boundary, historical	26.3%	38.7%	12.3%
	Hard Boundary, no transfers	59.0%	7.8%	0.1%
	Soft Neighborhood	13.6%	61.4%	23.1%
2008-15	Hard Boundary, historical	23.7%	42.1%	16.1%
	Hard Boundary, no transfers	58.3%	7.6%	0.4%
	Soft Neighborhood	9.8%	64.4%	25.1%
Year(s)	**Model, east of river**	**Assignment Diversity**		
		≤1.0	≥2.0	≥3.0
2008-09	Hard Boundary, historical	14.4%	49.7%	18.5%
	Hard Boundary, no transfers	52.1%	8.5%	0.4%
	Soft Neighborhood	4.7%	72.1%	29.8%
2014-15	Hard Boundary, historical	18.9%	44.4%	15.1%
	Hard Boundary, no transfers	53.7%	9.0%	0.1%
	Soft Neighborhood	6.6%	70.7%	28.0%
2008-15	Hard Boundary, historical	16.4%	48.1%	19.4%
	Hard Boundary, no transfers	53.0%	8.4%	0.5%
	Soft Neighborhood	4.6%	72.3%	30.3%
Year(s)	**Model, west of river**	**Assignment Diversity**		
		≤1.0	≥2.0	≥3.0
2008-09	Hard Boundary, historical	54.7%	19.5%	2.6%
	Hard Boundary, no transfers	79.2%	4.0%	0.0%
	Soft Neighborhood	34.3%	33.3%	2.9%
2014-15	Hard Boundary, historical	52.7%	18.5%	2.0%
	Hard Boundary, no transfers	78.0%	3.3%	0.0%
	Soft Neighborhood	38.6%	27.9%	5.4%
2008-15	Hard Boundary, historical	54.7%	17.2%	2.5%
	Hard Boundary, no transfers	80.7%	4.1%	0.0%
	Soft Neighborhood	31.7%	31.0%	3.2%

Table 3: Assignment diversity for two years (2008-09 and 2014-15), and over all 7 years, for the current system and the Soft Neighborhood model. Overall, the Soft Neighborhood model shows greater assignment diversity in all scenarios.

5 Comparison with Other Models

Many US school districts have stopped using the kind of deterministic neighborhood-based framework that Portland uses today (e.g., [10, 8, 9, 11]). Proponents of alternative frameworks such as *School Choice* often cite equity as a primary reason for rejecting hard-boundary neighborhood assignment:

> ...neighborhood-based assignment eventually leads to socioeconomically segregated neighborhoods, as wealthy parents move to the neighborhoods of their school of choice. Parents without such means have to continue to send their children to their neighborhood schools, regardless of the quality or appropriateness of those schools for their children. [2]

The School Choice movement is an alternative that attempts to counteract the socio-economic and racial segregation that has resulted from discriminative social, educational, and housing policies [7, 6, 4]. Like School Choice, the Soft Neighborhood framework represents an alternative to deterministic neighborhood models, though it is not a School Choice model. Understanding a little about School Choice will help to clarify the differences between the two approaches.

In this section, we will describe several models: the hard-boundary neighborhood model, a pure lottery model, and several variants of School Choice models. We will also consider how two hypothetical students would be assigned to kindergarten in each case. The two students in our example live across the street from one another, one at 4541 NE 22nd Ave. and the other at 4540 NE 22nd Ave. These addresses lie on opposite sides of a current PPS cluster boundary (Grant and Jefferson/Madison dual assignment).

In the Soft Neighborhood model, the set of nearby schools for these two children children is the same because they live so close to each other (Figure 10). If we populate this set using the 3-School rule from Section 2, then the candidate schools are Sabin, Vernon, Alameda, King, and Irvington. Since they live so close to each other, the two students have nearly the same odds of attending each of these schools. That is, the likelihood of the model assigning student A to Sabin is the same as the likelihood of assigning student B to Sabin. Because the set of nearby schools is similar for all children living near each other, there will always be children living nearby who are assigned to the same school.

5.1 Hard-Boundary Neighborhood Model

In a Hard Boundary model like the one PPS currently uses, boundaries are used to designate neighborhood attendance areas. For a given grade range, everyone residing within a certain boundary pretty much goes to the same school.[9] The two students living across the street from one another in Figure 11 will never be assigned to the same schools because they span a cluster boundary. Another way to think about it is that the probability that they go to the same school is zero. For the most part, everyone who lives on the same side of the boundary goes to the same schools with 100% probability.

[9] There are exceptions, but this is the norm.

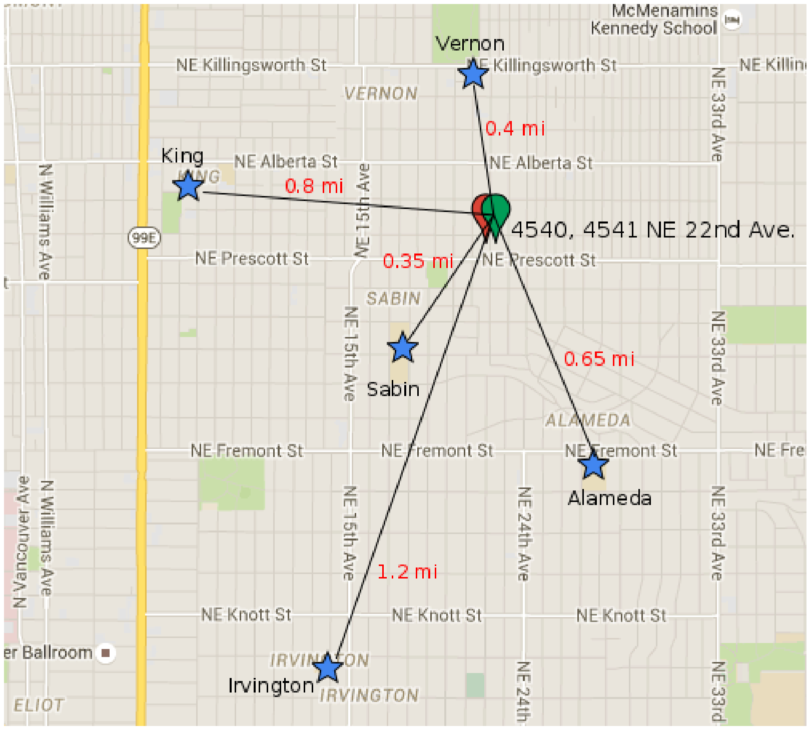


Figure 10: In the Soft Neighborhood model, the two students living across the street from each other will both be assigned to one of five nearby schools.

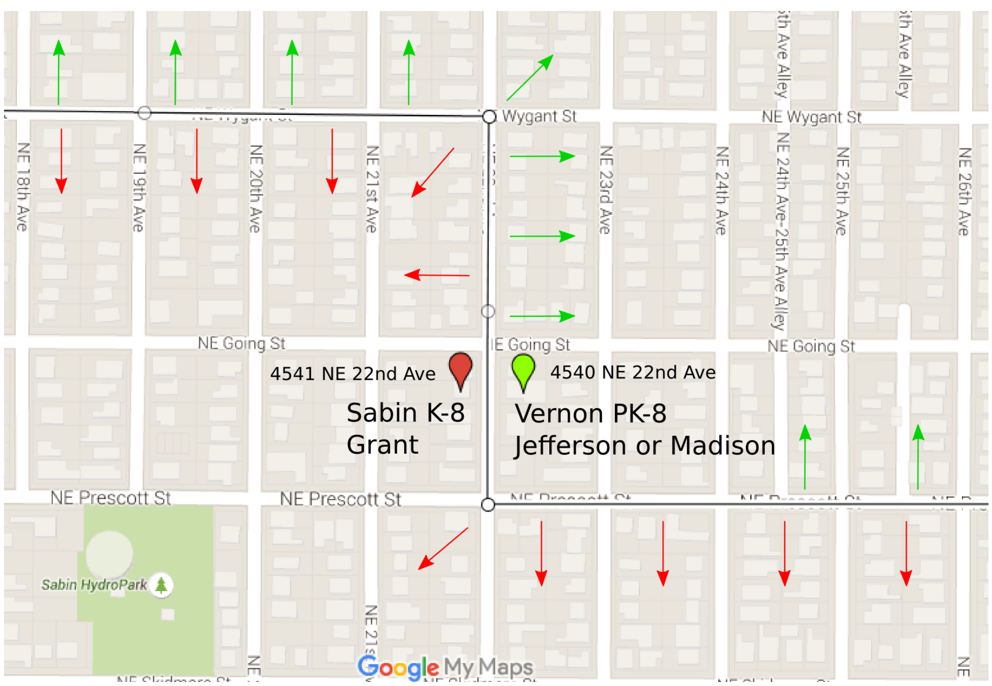


Figure 11: In a Hard Boundary model, the two students are always assigned to different schools because they live across a boundary. Only students on the same side of the boundary can go to the same school.

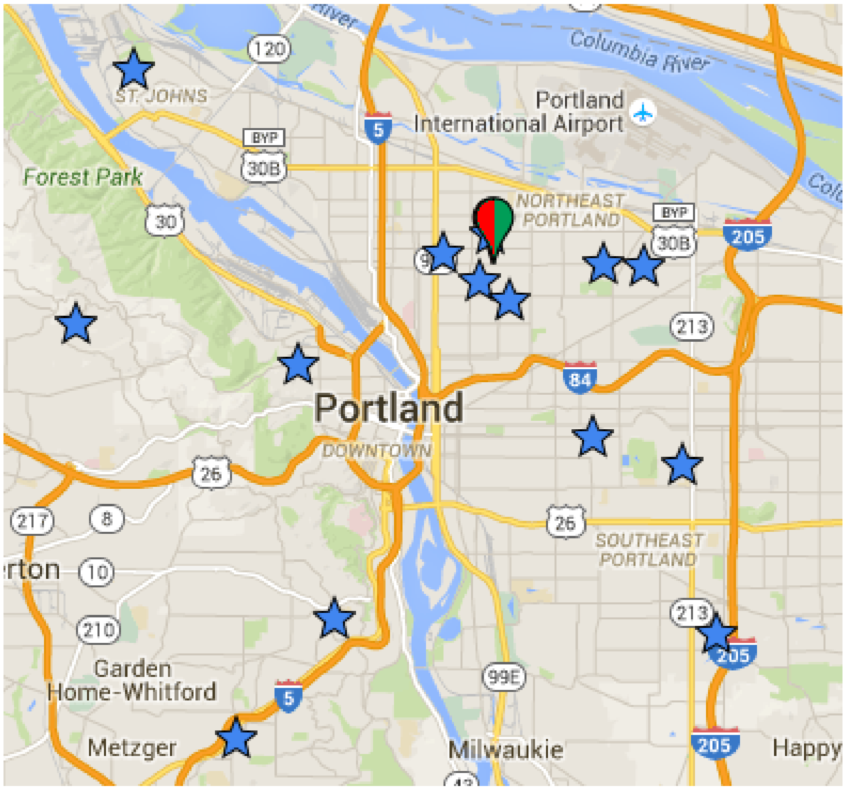


Figure 12: In a pure lottery model, students may be placed anywhere in the district. Some (but not all) district schools are shown in the picture.

5.2 Pure Lottery Model

A Pure Lottery model is geographically unconstrained and may assign children to any school in the district (Figure 12). The two children in our example have some chance of assignment to the same school; however, because of the number of possible schools, the likelihood of this happening is very small. The same holds for the children nearby: everyone will tend to go to a different school. Also, the students may have to travel very far away from where they live to get to school. A Pure Lottery model eliminates hard boundaries and can allow districts to fill schools with populations that match the district-wide demographic averages. It also offers the district a high degree of flexibility for balancing enrollments, since a child can be placed at any school. However, these advantages come at the expense of neighborhood and proximity. Travel times make a Pure Lottery model impractical in Portland. However, we include it in the discussion as a point of contrast with PPS's Hard Boundary model. The two models together (Pure Lottery and Hard Boundary) represent two extremes of the assignment spectrum.

5.3 School Choice

The School Choice movement gained momentum as an alternative to both Hard Boundary models and desegregation plans (see [5] for an overview). There are many variants of School Choice, but in the canonical version, families rank order their preferred schools, and the district places students in schools, taking into account these preferences as well as the district's priorities.

There exists a large body of academic literature, particularly in the field of economics, that has emerged from and bolstered the School Choice movement. The academic literature tends to focus on "mechanism design" — the algorithms used for making assignments — and the properties of proposed mechanisms. Researchers usually use three criteria to evaluate assignment mechanisms:

- Pareto efficiency: A Pareto efficient mechanism results in assignments that cannot be improved without making at least one student worse off.

- Stability: In a stable mechanism, there are no assignment outcomes with any *blocking pairs*. A blocking pair is a ⟨Student A, School A⟩ pair where Student A prefers School A to her assigned school, and Student B, who has lower priority than Student A at School A, has been assigned to School A.

- Strategyproofness: Strategyproof mechanisms encourage truthful rankings from all students.

Pathak [5] summarizes very well the three primary mechanisms associated with School Choice: the student-proposing deferred acceptance mechanism [3], the top trading cycles mechanism [1, 12], and the Boston mechanism [1]. It's important to understand that each assignment mechanism is different: student-proposing deferred acceptance is stable and strategyproof; top trading cycles is Pareto efficient and strategyproof; and the Boston mechanism is Pareto efficient but neither stable nor strategyproof. Choice-based programs have been characterized as conferring strategic advantages to better-educated and wealthier families, or even to resegregating school districts [13], but each individual program's chosen assignment mechanism and implementation is instrumental in how it functions.

The Soft Neighborhood model differs from School Choice models along several dimensions. The most obvious difference is that in the Soft Neighborhood model, families do not rank preferences. In the Soft Neighborhood model, proximity is used to seed an initial probability distribution, and then that distribution is modified so that the expected number of students at each school equals its capacity. Other differences are that the probabilities used in the Soft Neighborhood model are real numbers, whereas in School Choice models, usually only ordinal ranking are considered. Another major difference is in the math underpinning the two models. Choice models use an optimal-matching algorithm like the ones we described above. Soft Neighborhood assignment uses probabilities in order to eliminate determinism between an address and a particular school, but does so in a way that maintains a strong notion of neighborhood.

6 Frequently-Asked Questions

Are you suggesting that kids get reassigned to different schools every year in order to balance enrollments? No, we are definitely not saying that. School assignment in the Soft Neighborhood model is not something that is meant to be done to every student every year for the purposes of keeping enrollments balanced. Assignments are only made when students need to be placed at a PPS school — for example, because they moved or because they are entering the system for the first time.

We believe strongly in stability and consistency. Involuntary school reassignment creates a lot of stress on families and school communities, and should be avoided whenever possible. That said, there are "pivot points" in the typical K–12 grade sequence — at 6th and 9th grades when students enter middle school and high school, respectively. At these pivot points, it might make sense to allow students to apply for reassignment by the Soft Neighborhood framework if they want.

What happens if a family moves to some other location within Portland? A family moving to a different location might trigger reassignment depending on the location of the new residence. If the original school is still one of the "nearby" schools according to the model, then the family should stay at the original school, and no reassignment occurs. If the original school is no longer a "nearby" school, then reassignment might make the most sense. The district could consider allowing students to stay at the original school even after a move — perhaps via petition, and without a transportation guarantee.

How does middle school and high school assignment work in this framework? The Soft Neighborhood model provides a framework for assigning students to schools in a way that satisfies capacity constraints and otherwise is more likely to place students in schools closer to their homes. Keeping schools close to students is most important in the lower grades, when students are younger and less independent and require more day-to-day involvement from their families. There are several ways Soft Neighborhood assignment could be implemented across grade-levels within PPS. We like the idea of allowing students to re-apply for Soft Neighborhood assignment at critical pivot points, e.g., at 6th and 9th grades. In the case that a student didn't want to apply for re-assignment, he or she would continue on with the regular "feeder pattern" sequences for K–5 into 6–8, and from 8th grade into high school. Students in K–8 could also apply for reassignment at 6th grade if so desired.

This idea is compelling for a number of reasons. For one, not everyone wants to stay with their cohort all the way from K–12, and this would allow these students a chance to go to a different school. At the same time, students who would like to remain with the same cohort would be able to. Second, it provides the system an opportunity to restabilize and correct for enrollment "drift" — where enrollments shift away from a balanced state as students move, leave the system, etc. The Soft Neighborhood model will have its best chance at balancing enrollments at the kindergarten level when most new students enter the system, and it can be used to assign new assignees at any grade level. But allowing students to be voluntarily reassigned at regular intervals gives the system a better chance to maintain a balanced state.

So you mean schools will get filled with students who live closest first? No... in the Soft Neighborhood model, proximity makes it more likely that a student will get assigned to one school over another, but it doesn't make any guarantees. It figures out how to roll the dice to ensure that schools are filled to capacity with nearby students, and then it rolls the dice for everybody. Simply filling schools with the closest students first would solve the enrollment balancing problem, but with respect to equity and socioeconomic stratification, it would be just as bad (or worse) than classic hard boundaries.

Would this mean kids have to travel a lot farther to get to school? In our simulations on PPS data, both assignment models (historical and Soft Neighborhood) place students in schools within a reasonable distance from their homes. Under the Soft Neighborhood model, overall around 33% of students travel less than 0.5 mile. Around 81% travel less than 1.0 mile, and around 97% travel less than 1.5 miles. The full results are shown in Section 4, Table 2.

Why are you using Cartesian (straight-line) distances and not driving distances? We agree that using Cartesian distances is not the best approximation for real travel distances, and that a metric based on minimum driving distances or even travel time would be better. We used Cartesian distance because it is simple and quick to implement. We intend to use a more realistic map-based distance metric in future analysis, and we expect that PPS would do the same with the analysis and implementation of this or any other assignment model.

Would this even work on the west side of the city? The Soft Neighborhood assignment algorithm is applied the same way to the west side of the city as the east. On the west side, schools are more spread out and so the proximity function will generate schools that are farther away on average than on the east side. This is not surprising since students on the west side of the river already travel farther to get to school, on average, than students on the east side. In Section 4, Tables 2 and 3, we show simulation results specific to the west and east sides separately: travel distances are greater on the west side, and assignment diversity is lower. This is true no matter what model you look at.

Could kids be assigned to schools across the river in this framework? There are no boundaries in the Soft Neighborhood model, and a boundary running down the Willamette River is the only thing preventing cross-river assignment in the district today. The Soft Neighborhood model allows cross-river assignment; however, with a realistic (travel-based) distance metric, no one would be assigned to a school across the river unless it were truly as nearby as schools on the same side. That said, the district could still impose a hard boundary down the middle of the river and forcefully segregate the community for other reasons independent of balancing, proximity, and equity, but the district and the community should be honest about what those reasons are.

Hmmm....this sounds interesting, but I want my kids to go to school with the kids next door. Well, you're in luck — they just might! They might also get to go to school with the kids across the street, or in the house behind you, or one block over. That's a feature of the Soft Neighborhood model: if kids live close together, they might go to school together. There is no guarantee, though, but it's likely that some kid on your block will go to the same school. Under the current hard boundary system, if you live near an existing school boundary you don't have any guarantees either, since the district can move that boundary in unpredictable ways whenever it sees the need. Either way, if your kids like to play with the kids next door, that probably won't change just because of the schools they go to.

Isn't this kind of like the old transfer lottery? Nope, the Soft Neighborhood model is a neighborhood model, where assignment is based only on the home address, and it must provide an assignment to everyone who applies to it. The old transfer lottery was built on top of the PPS Hard Boundary neighborhood model and was under no obligation to assign all applicants.

I've heard of this before in ⟨Cambridge, San Francisco, etc.⟩ — isn't it called Controlled Choice or something like that? The Soft Neighborhood model is not a School Choice model (Controlled Choice is a variant of School Choice). School Choice systems involve eliciting ranked preferences over schools from families, and then making assignments based on those preferences. Choice-based programs may confer strategic advantages to better-educated and wealthier families and have been blamed for intensifying segregation in some cases [13]. We wanted to develop a solution to enrollment balancing in PPS that would counteract the tendency towards segregation. The Soft Neighborhood model is tailored to achieve enrollment balancing (capacity) and retain a sense of neighborhood (proximity), while inducing a gentle mixing of student populations (equity).

Shouldn't PPS ensure a baseline of equitable academic program offerings at every school first, before considering something like the Soft Neighborhood model? We agree that PPS should ensure a baseline of equitable academic program offerings at every school, but we think the community shouldn't have to wait for this to happen before implementing a robust enrollment balancing solution. In fact, having predictable and optimal school enrollment will help the district distribute educational resources to the schools in an equitable way. We are skeptical that a solution involving boundary change, even frequent boundary change, is sufficiently robust to address enrollment balancing. Moreover, we see boundaries as a historical artifact that have coevolved with and reinforced the racial and socio-economic inequities the community is trying to counteract. In short, our school district needs a functional solution to enrollment balancing that doesn't directly conflict with our equity goals.

7 Conclusions: Where Do We Go From Here?

PPS is currently investing a substantial amount of time and energy trying to fix its enrollment balancing problem. It has convened an advisory committee (District-Wide Boundary Review Advisory Committee, or DBRAC) to assess the problem, and DBRAC has recommended a framework for regular boundary adjustments. As a means for solving enrollment balancing, we believe that boundary adjustment is inadequate. Any hard-boundary system will result in imbalanced enrollments and segregated populations. It will also confer an advantage to families with more resources, who can buy assignment to the public school of their choice.

We have developed a boundary-free solution to the enrollment balancing problem that focuses directly on satisfying school-specific target enrollments. Moreover, our model reinforces the district's stated equity goals by eliminating any determinism between home address and school assignment, and by encouraging mixing between populations. We designed the Soft Neighborhood model to be a robust and equitable solution to enrollment balancing that retains the community ideal of neighborhood. Like any traditional neighborhood school model, the Soft Neighborhood model aims to place students in schools close to their homes, and the only input into the system from a family is its home address.

We have implemented and tested the Soft Neighborhood model against seven years of historical PPS enrollment data. We have defined three metrics for evaluating our model on the basis of balanced enrollments, travel distance, and assignment diversity. Our experimental results suggest that the Soft Neighborhood model outperforms the historical Hard Boundary system with respect to enrollment balancing; that it improves assignment diversity; and that students travel comparable distances whether using Soft Neighborhood assignments or historical assignments. These empirical results demonstrate the potential of the Soft Neighborhood model as a robust and equitable solution to the enrollment balancing problem in PPS. Moving forward, the model needs to be fully tuned and evaluated against a proper data set.

We encourage the PPS Board of Directors and the larger community to seriously consider the Soft Neighborhood model for implementation in PPS. Moreover, we would like to see compelling empirical evidence that the DBRAC proposal can actually achieve its stated balancing and equity goals, and we request an objective evaluation of the district's proposed solution in terms of the metrics we have defined in this document. Any such evaluation must be implemented with transparency to the public — which means it should use a data set that has been properly anonymized and released to the public, and that it should make the entirety of its evaluation methodology (e.g., the source code, too) available to the public for independent verification and validation.

A Appendix: Soft Neighborhood Algorithms

This appendix delves into the algorithmic details of the Soft Neighborhood assignment model. As with any mathematical algorithms, one could carry out a school assignment using just pencil and paper, or a pen-knife and bark chips, but it would certainly be faster if executed by modern computing hardware.

The basic inputs to the system are:

1. A set of schools $S = \{s_i\}$ with known geographic locations (address).
2. A set of students $C = \{c_j\}$ with known geographic locations (home address).

A.1 Step 1: Set Capacity Constraints

The first step is to establish the enrollment targets t_i for each school s_i, also referred to as the capacity constraints. For the given grade level, the district must decide how many spots are available for new assignees. The general formula looks like this:

$$t_i = ((\textit{\# of sections in } s_i) \times (\textit{ideal section size})) - (\textit{\# of pre-assignees to } s_i)$$

The *# of sections* and the *ideal section size* are things for the district to decide, based on whatever factors already go into filling a school with students. The pre-assignees are all the students already assigned to the school and not subject to the random assignment process, including:

- students continuing on from an earlier grade;
- co-enrolling younger siblings of older students already assigned.

The pre-assignees are removed from the corpus of students C, as they have already been assigned to a school.

The capacity constraints can be expressed as a vector $\{t_i\}$ of targets, one per school s_i.

A.2 Step 2: Seed Probabilities by Proximity

In Step 2, we establish a probability matrix $P = (p_{ij})$ in which each element p_{ij} represents the probability with which student c_j should be assigned to school s_i. This matrix is initialized (seeded) using weights from a proximity function that favors schools that are closer to a student over schools that are farther away.

The precise form of this seed weighting function is important to the performance of the model overall, as it sets the limits on how far students can live from candidate schools.

The function used in our simulations is:

$$W(s,c) = \begin{cases} D(c) - d(s,c) & \text{if } d(s,c) <= D(c) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where

$$D(c) = \text{maximum of } \begin{cases} 1.1 \times \text{distance to third closest school to } c \\ 1.25 \text{ miles} \end{cases} \tag{2}$$

Intuitively, this means: "For a student c_j, the candidate schools are: the three closest schools, any school within 110% of the third closest school, and any schools within 1.25mi distance."

In our simulations so far, we have used a simple Cartesian metric ("crow's flight") to measure distance $d(s, c)$. However, a proper distance metric should reflect the true commute distance from a student's home to a school (i.e., walking or driving). The district already has methodologies for estimating such distances.

The seed weights are plugged into the probability matrix — though, they must be normalized for each student (over all schools) before they become actual probabilities:

$$p_{ij} = \frac{W(s_i, c_j)}{\sum_i W(s_i, c_j)}$$

At this point, the matrix P contains the probabilities p_{ij} that a student c_j should be assigned to school s_i based solely on her proximity to that school. For any given student, most p_{ij} will be zero, and only the schools with non-zero probability are candidates for that student.

A.3 Step 3: Balance Probabilities with Capacity Constraints

The Balancing Algorithm is a kind of iterative *relaxation algorithm* that repeatedly alternates two steps until it converges to a solution.

In a nutshell, the algorithm works like this:

1. For each school, sum the probabilities applying to that school to find the expected number of children assigned to that school, which may be more or less than the target capacity. Go back and multiply each of those probabilities by a correction factor *target-capacity/expected-assignments*.

2. For each child, renormalize his/her assignment probabilities: divide each probability by their sum, so that they add up to 1.

3. Repeat 1 and 2 until the expected assignments to each school converge to their target capacities, quitting if the convergence completes or stalls.

If there *is* a solution that exactly satisfies the target capacities, this algorithm will find it — but there must be a solution. A necessary condition for this is that the capacity of the district (total of target capacities of all schools) is equal to the number of children being assigned. Even if there is no exact solution, the algorithm will find something close. If the algorithm stalls leaving one or more schools in an large over- or under-capacity state, that indicates a grave capacity problem in some part of the district, e.g., there simply aren't enough nearby schools to handle the local student population.

A.4 Step 4: Assign Students

The complete assignment step will be described in a future revision of this document, but the gist of it is:

1. Partition the probability matrix and the sets of students into groups which have non-zero probabilities for the same subset of schools.

2. Nudge the probability submatrix for each of these groups to ensure an integer expectation for the number of students assigned to each of the candidate schools in the group.

3. (Optionally) propagate probability mass along the network formed by the groups to further balance the integer expectations with the capacity constraints.

4. For each group, divide its students among its schools according to the integer expectations in a way that honors the probability submatrix, using an approximation algorithm similar to the balancing algorithm.

B Appendix: Detailed Description of the Data Sets

Portland Public Schools provided two data sets for the purposes of running the simulations described in this document: the STUDENTS data set and the SCHOOLS data set. The STUDENTS data set contains historical and current information from the last seven school years. The SCHOOLS data set contains information about K–8, K–5, and 6–8 schools in PPS in the year 2014-15.

B.1 The Students Data Set

The STUDENTS data set consists of just over 200K records for all students in kindergarten through eighth grade over the seven-year period from 2008–2015. Table 4 shows the number of student records (K–8 and K only) each year.

Year	**K–8**	**K**
2008-09	28,760	3,601
2009-10	29,077	3,726
2010-11	29,211	3,657
2011-12	29,875	3,715
2012-13	30,056	3,884
2013-14	30,671	3,843
2014-15	30,744	3,681
2008-15	208,395	26,107

Table 4: The number of student records (K–8 and K only) per year in the STUDENTS database.

Each record in the data set represents a student and contains the following information:

- **School year**: The school year for which this record is valid (2008-09 through 2014-15).
- **Grade**: Student's grade level (K–8) during that school year.
- **Campus enrolled**: Student's assigned school during that school year. Schools included in the data set are K–8, K–5, and 6–8 schools that have a neighborhood program on their campus. District-only focus options, alternative, and charter schools are not included.
- **Campus code**: A numeric ID associated with the assigned school.
- **Neighborhood**: The neighborhood in which the student lives. Values for this attribute can be a PPS neighborhood for in-district students (205,113 records)), *out of district* (3,264 total records), or *not yet assigned* (18 total records).
- **Capture?**: TRUE if the enrolled campus is the same as the neighborhood in which the student lives, and FALSE otherwise (e.g., transfer students).

- **Home Coordinates**: The X, Y coordinates of the student's home address, randomly offset by the district for the purposes of anonymity. Addresses are represented in Oregon State Plane coordinates, North Zone, in feet.

B.2 The Schools Data Set

The Schools data set contains information about K–8, K–5, and 6–8 schools in PPS during the year 2014–15. Each record in the data set represents a school (or campus for schools with more than one building) and includes the following information:

- **Name**: The name of the school/campus
- **Configuration**: K–8, K–5, or 6–8
- **Address**: street, city, state, zip of the building
- **School Coordinates**: The X, Y coordinates of the schools address represented in Oregon State Plane coordinates, North Zone, in feet.
- **Kindergarten Homeroom Count**: Number of kindergarten homeroom classrooms for 2014-15.
- **Demographics**: total number enrolled, number free-or-reduced lunch status (by direct certification) enrolled, and a breakdown by racial designation, at each school

B.3 Flaws in the Data Sets

The data sets provided by the district are flawed in ways that limit the extent to which we have been able to validate the model via the simulations. It is important to understand these flaws. Fixing them is a prerequisite for full validation of the Soft Neighborhood framework.

1. **No sibling information**: Sibling relationships are not available in the data set, making it impossible to evaluate the effect of guaranteed placement of co-enrolled siblings at the same school. Basically, in this data set, there are no siblings.
2. **Lack of information about new vs continuing students**: The address information in the Students database has been perturbed so as to preserve student anonymity. This is a good thing. However, there is no address continuity within the data set. In other words, it is impossible to tell whether or not a 1st-grader has continued on from kindergarten or is newly-enrolled at a school. This makes it impossible to assess how well the Soft Neighborhood model is able to maintain balanced enrollments as students move up from kindergarten through the primary and middle grades.
3. **Lack of sufficient information about focus options**: The data set lacks sufficient detail about focus option programs co-located at neighborhood schools, and focus option schools. There is no information at all about the latter (focus option schools). At schools with a neighborhood program and co-located focus option program(s), there is no way to distinguish which students are in which program. Basically, in this data

set, focus option schools and the students who attend them don't exist, and everyone at a neighborhood school is enrolled in a neighborhood program. This makes it impossible to estimate the effect of program placement.

There are also some limiting factors in the data sets that cannot be fixed but that should be acknowledged:

- **No student-specific socio-economic and racial/ethnic information**: This information cannot be released, and therefore any evaluation of the effect of the Soft Neighborhood model on mixing and diversity is limited.

- **Lack of real information regarding target capacities**: In the Soft Neighborhood framework, it is important for the district to set school-specific target capacities based on the preferred configuration of the actual space available at each building. We don't have these target capacities. The Schools data set does report the number of kindergarten classrooms at each school in 2014-15. However, this information does not generalize well to earlier years, and it does not capture whether those numbers are an artifact of overcrowding or underenrollment (e.g., district was forced to convert some space into an extra classroom). See Section 4.1.1 for more details regarding how target enrollments were set in the simulations used to generate the results in this document.

B.4 How to Fix the Data Sets

The following corrections to the Students data set would allow for complete validation of the model without compromising student anonymity:

1. **Data for all programs, and complete program identification**: The data set should include data for the student population of all programs, e.g., focus option schools/programs as well as neighborhood schools. Furthermore, student data should identify the specific program in which a student is enrolled (e.g., *Mandarin Immersion* or *Odyssey*), not just the campus. This would allow one to tease apart the populations attending such programs, and also allow for analyzing how focus-option populations interact with the distribution of neighborhood school populations.

2. **Consistently anonymized addresses**: Ideally, when a given address (in this case, a coordinate pair of (feet-east, feet-north)) is randomly perturbed to anonymize it, the same perturbed result should be used for every instance of it. For instance, if the address (84757.1E, 33364.2N) is tweaked to (84790.2E, 33298.4N) the first time it is encountered, then it should be tweaked to that same value everywhere in the data set. This would allow one to make educated guesses as to which students are newly-enrolled at a school, and which students are siblings.

3. **Consistent student identifier**: Additionally, the data would be even more clear if each student were tagged with a unique-ID that was consistent from one year to the next. The actual PPS student-ID should certainly remain private, but a one-way hash of that ID could keep it private and serve the same purpose in the data set. This would make it very clear which students were continuing on from one year to the next.

4. **Explicit sibling references**: Consistent student ID's would also allow for an explicit reference to the next older sibling and/or same-year sibling (if any) — which would clarify which students should be handled as co-enrolled siblings and how to handle them.

References

[1] Atila Abdulkakiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *The American Economic Reivew*, 93(3):729–747, 2003.

[2] Lars Ehlers, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648–683, 2014.

[3] David Gale and Lloyd Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), 1962.

[4] Ethan Johnson and Felicia Williams. Desegregation and multiculturalism in the Portland public schools. *Oregon Historical Quarterly*, 3(1), 2010.

[5] Parag A. Pathak. The mechanism design approach to student assignment. *Annual Review of Economics*, 3(1):513–536, 2011.

[6] Richard Rothstein. The racial achievement gap, segregated schools, and segregated neighborhoods — a constitutional insult. *Race and Social Problems*, 6(4), Dec 2014.

[7] Richard Rothstein. Should we force integration on those who don't want it?, and other commonplace questions about race relations. Economic Policy Institute, Working Economics Blog, March 2015.

[8] Boston Public Schools. Student assignment policy. `http://www.bostonpublicschools.org/assignment`. Accessed: 2015-09-15.

[9] Cambridge Public Schools. About controlled choice. `http://www.cpsd.us/departments/frc/making_your_choices/about_controlled_choice`. Accessed: 2015-09-15.

[10] Jefferson County Public Schools. Student assignment. `http://www.jefferson.k12.ky.us/departments/StudentAssignment/Index.html`. Accessed: 2015-09-15.

[11] San Francisco Public Schools. SFUSD: The assignment process. `http://www.sfusd.edu/en/enroll-in-sfusd-schools/how-student-assignment-works/the-assignment-process.html`. Accessed: 2015-09-15.

[12] Lloyd Shapley and Herbert Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, pages 23–37, 1974.

[13] Jeremy Adam Smith. As parents get more choice, S.F. schools resegregate. San Francisco Public Press, Feb 2015.