

# Estimating spatial availability/mismatch using singly constrained accessibility measures

Author One<sup>a</sup>, Author Two<sup>a</sup>

<sup>a</sup> *Address*

## **Abstract**

This is the abstract.

It consists of two paragraphs.

---

\*Corresponding Author

*Email addresses:* `author.1@example.com` (Author One), `author.2@example.com` (Author Two)

## Introduction

Accessibility analysis is widely employed in transportation, geography, public health, and many other areas. It is a relatively simple concept whose appeal derives from combining the spatial distribution of opportunities and the cost of reaching them. Recent works have aimed at making accessibility easier to interpret. This vignette introduces the concept of spatial *availability*, a form of singly-constrained accessibility that preserves the number of opportunities available, and therefore leads to more interpretable results.

## Background: Accessibility Analysis

Most accessibility measures (excluding utility-based measures) are derived from the gravity model, and are known as *gravity-based* accessibility. Briefly, consider the following conventional accessibility measure  $A_i$  :

$$A_i = \sum_{j=1}^J W_j f(c_{ij})$$

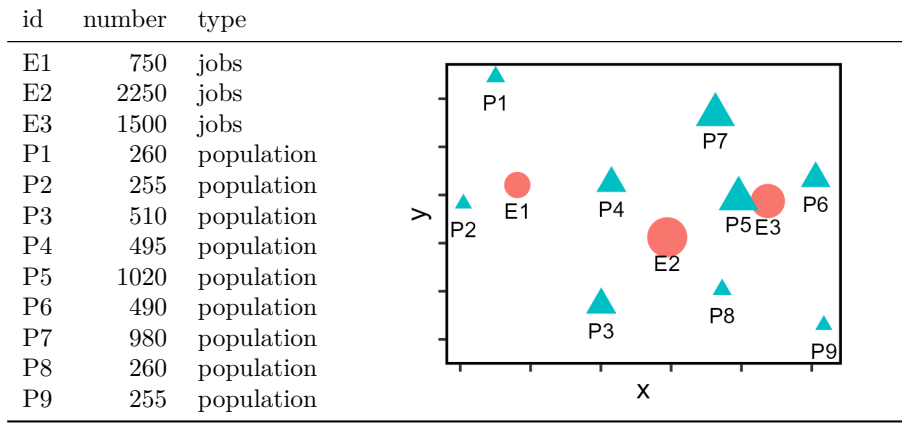
where:

- $i$  is a set of origin locations.
- $j$  is a set of destination locations.
- $W_j$  is the number of opportunities at location  $j$ . These are opportunities for activity and add some sort of *supply* to the area;
- $c_{ij}$  is a measure of the cost of moving between  $i$  and  $j$
- $f(\cdot)$  is an impedance (or distance-decay) function (a monotonically non-increasing or decreasing function of  $c_{ij}$ ).

The accessibility value  $A_i$ , it can be seen, is the weighted sum of opportunities that can be reached from location  $i$ , given the cost of travel  $c_{ij}$  and an impedance function. Summing the opportunities in the neighborhood of  $i$  (the neighborhood is defined by the impedance function) estimates of the total number of opportunities that can be reached from  $i$  at a certain cost. Depending on the impedance function, the measure could be cumulative opportunities (if  $f(\cdot)$  is a binary or indicator function) or a more traditional gravity measure, for instance with a Gaussian impedance function or an inverse cost impedance function.

We use a simple numerical example to introduce the key concepts, and we will use the usual accessibility measure for comparison. In this way, we aim to show the differences between accessibility and spatial availability, which helps to explain how spatial availability can improve interpretability in the analysis of spatially dispersed opportunities.

Table 1: Numerical example



### Numerical Example

In this section we present a simple numerical example. The setup for the example is a simple system with three employment centers and nine population centers, as seen in Table 1.

The accessibility to employment of each of the population centers can be calculated using the expression above for  $A_i$ . As noted, this yields the number of jobs (opportunities) that are accessible (i.e., can be reached) from each population center, given the cost.

In this example we use the straight line distance between the population and jobs for  $c_{ij}$ , and a negative exponential function with  $\beta = 0.0015$ .

Plot 1 shows the three employment centers (black circles), with the size of the symbol in proportion to the number of jobs there. We also see nine population centers (triangles), where the size of the symbol is proportional to the accessibility to jobs. At a quick glance:

- Population centers (triangles) in the middle of the map are relatively close to all three employment centers and thus have the highest levels of job accessibility. Population center P5 is relatively central and close to all employment centers, and it is the closest population to the second largest employment center in the region. Unsurprisingly, this population center has the highest accessibility 680.64);
- Population centers (triangles) near the left edge of the map (only in proximity to the small employment center) have the lowest levels of job accessibility. Population center P1 is quite peripheral and the closest employment center is also the smallest one. Consequently, it has the lowest accessibility with  $A_i = 17.12$ );

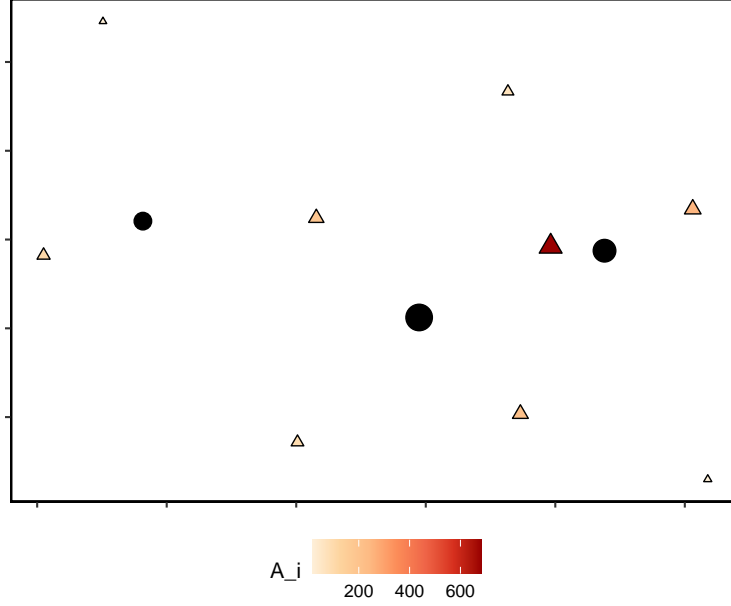


Figure 1: Accessibility results using simple numerical example

#### *What are the Issues?*

Accessibility measures such as that illustrated above are excellent indicators of the intersection between urban structure and transportation infrastructure. However, beyond enabling comparisons of cardinality (e.g., P5 has higher accessibility than P1, which although low is still better than zero) they are not highly interpretable on their own.

First, the “total accessibility” in the region is not a meaningful quantity since it is not constrained to resemble, let alone match the number of opportunities available in a region. Further, this quantity, if calculated, depends on the number of origins: since every additional origin in the analysis makes at least one and possibly more opportunities at destinations to be multiple-counted, the “total accessibility” is vulnerable to the modifiable areal unit problem.

Secondly, it is unclear what the highest and lowest levels of accessibility in the region actually represent. For example, an individual at P5 in the example has accessibility to 680.6373657 jobs. But since this is also a large population center, there is potentially large competition for those accessible jobs. In other words, the value of  $A_i$  is not sensitive to the size of population at the origin seeking the opportunity (in this case jobs), let alone the population at other locations. This unfortunately limits the interpretability of the measure. Floating catchment areas purport to account for competition/congestion effects, but as discussed by Paez, Higgins, and Vivona (2019), they are vulnerable to inflation/deflation

issues, which makes them prone to bias unless corrected.

As an alternative, we propose a measure of *spatial availability/mismatch*.

Spatial availability is an appropriate framework for *indivisible* opportunities. These are opportunities that, once they have been taken by someone, are no longer available to other members of the population. Examples of indivisible opportunities include jobs (when a person takes up a job, the same job cannot be taken by someone else) and seats at schools (once a student takes a seat at a school, that particular opportunity is no longer there for another student). From a different perspective, employers may see workers as opportunities, so when a worker takes a job, this particular individual is no longer in the available pool of candidates for hiring. As we will show, spatial availability is a singly-constrained measure of accessibility.

In the following sections we will describe and illustrate this new measure using simple numerical examples. First, we will describe the measure. Second, we will calculate the spatial availability of jobs from the perspective of the population. Then, we will calculate the spatial availability of workers from the perspective of employers. Next, we will illustrate how to calculate the spatial availability with catchment restrictions. Finally, we will compare conventional accessibility estimates to the proposed measure of spatial availability.

## Spatial Availability/Mismatch

### *Definitions*

As recent research on accessibility shows (see for instance Paez, Higgins, Vivona, 2019 and Allen and Farber (2020)), accounting for competitive access in a meaningful way requires the proportional allocations of quantities in the accessibility calculations, or their normalization. At issue is the fact that that multiple-counting is commonplace when calculating conventional accessibility  $A_i$  for  $i = 1, \dots, n$ . This is easy to see once we realize that every opportunity enters the weighted sum once for every origin  $i$  that can reach it. This has the unfortunate effect of obscuring the interpretability of  $A_i$  and fails to answer for a individual at a specific population center the question: “many jobs are accessible, but the same jobs are also accessible to my (possibly) numerous neighbors... what does high accessibility actually mean to me?”

In the spatial availability framework proposed, and in line with the gravity tradition, we distinguish between opportunities at a destination and demand for opportunities at the origin. To explain the analytical framework, the example of accessibility to employment is illustrative, with “population” in the role of demand and “jobs” in the role of opportunities.

We begin with allocation based on demand (the number of individuals in the population in the labor market who “demand” employment). Consider an employment center  $j$  with  $W_j^r$  jobs of type  $r$ . In the general case where there are  $K$  population centers in the region, the following factor can be defined to allocate

the jobs proportionally based on the size of the population at each center:

$$f_{ij}^p = \frac{P_{i \in r}^\alpha}{\sum_{k=1}^K P_{k \in r}^\alpha}$$

where  $f_{ij}^p$  is a factor that proportionally allocates a share of the jobs at  $j$  to population center  $i$ . On the right hand side of the equation,  $P_{k \in r}$  is the population at location  $k$  that is eligible for jobs of type  $r$  (maybe those with a certain level of training, or in a designated age group). Here we add an empirical parameter  $\alpha$  that can be used to modulate the effect of size in the calculations (i.e., if  $\alpha < 1$  employment centers draw more rapidly from small centers than from large centers, and when  $\alpha > 1$  employment centers draw more rapidly from large population centers). The summation in the bottom is over  $k = 1, \dots, K$ , the number of population centers in the region. The factors  $f_{kj}^p$  satisfy the property that  $\sum_k^K f_{kj}^p = 1$ .

The share of jobs at  $j$  allocated to (i.e., available to) each population center is:

$$V_{kj} = W_j f_{kj}^p$$

and since  $\sum_k^K f_{kj}^p = 1$  it follows that:

$$\sum_{k=1}^K V_{kj} = W_j$$

In other words, the number of jobs is preserved.

As an example, consider an employment center  $j$  in a region with two population centers (say  $i$  and  $k$ ). For simplicity, assume that the all the population in the region is eligible for these jobs, that is, that the entirety of the population is included in the set  $r$ . The allocation factors for the jobs at  $j$  would be:

$$\begin{aligned} f_{ij}^p &= \frac{P_i^\alpha}{P_i^\alpha + P_k^\alpha} \\ f_{kj}^p &= \frac{P_k^\alpha}{P_i^\alpha + P_k^\alpha} \end{aligned}$$

Suppose that there are three hundred jobs in the employment center ( $W_j = 300$ ), and that the populations are  $P_j = 240$  and  $P_k = 120$ . The jobs are allocated as follows (assuming that  $\alpha = 1$ ):

$$\begin{aligned} V_{ij} &= W_j \frac{P_i^\alpha}{P_i^\alpha + P_k^\alpha} = 300 \frac{240}{240+120} = 300 \frac{240}{360} = 200 \\ V_{kj} &= W_j \frac{P_k^\alpha}{P_i^\alpha + P_k^\alpha} = 300 \frac{120}{240+120} = 300 \frac{120}{360} = 100 \end{aligned}$$

It can be seen that proportionally more jobs are allocated to the bigger center and also that the total number of jobs is preserved ( $\sum_{k=1}^K W_{kj} = W_j$ ).

The factors above account for the total number of opportunities at the destination (i.e., the number of jobs at the employment center), but they do not account for their location relative to the population centers. The proportional

allocation procedure above is insensitive to how far population center  $i$  is from employment center  $j$ . To account for this effect we define a second set of allocation factors based on distance to the employment centers. These are defined as:

$$f_{ij}^c = \frac{f(c_{ij})}{\sum_{k=1}^K f(c_{kj})}$$

where  $c_{ij}$  is the cost (e.g., the distance, travel time, etc.) to employment center  $j$  from population center  $i$ , and  $f(\cdot)$  is an impedance function, that is a monotonically decreasing function of  $c_{kj}$ . The idea is that proportionally more jobs are allocated to closer locations. Assume that the impedance function is a negative exponential function as follows, and assume that  $\beta$  (which modulates the steepness of the impedance effect and is an empirical parameter) is one:

$$f(c_{ij}) = \exp(-\beta c_{ij})$$

Continuing the example, suppose that the distance from population center  $i$  to employment center  $j$  is 0.6 km, and the distance from population center  $k$  to employment center  $j$  is 0.3 km. Being closer, we would expect more jobs to be allocated to the population of  $i$ . The jobs would be sorted as follows:

$$\begin{aligned} f_{ij}^c &= \frac{\exp(-\beta D_{ij})}{\exp(-\beta D_{ij}) + \exp(-\beta D_{kj})} \\ f_{kj}^c &= \frac{\exp(-\beta D_{kj})}{\exp(-\beta D_{ij}) + \exp(-\beta D_{kj})} \end{aligned}$$

Numerically, the jobs allocation is:

$$\begin{aligned} V_{ij}^d &= W_j \frac{\exp(-D_{ij})}{\exp(-D_{ij}) + \exp(-D_{kj})} = 300 \frac{\exp(-0.6)}{\exp(-0.6) + \exp(-0.3)} = 3 \times 0.426 = 127.8 \\ V_{kj}^d &= W_j \frac{\exp(-D_{kj})}{\exp(-D_{ij}) + \exp(-D_{kj})} = 300 \frac{\exp(-0.3)}{\exp(-0.6) + \exp(-0.3)} = 3 \times 0.574 = 172.2 \end{aligned}$$

A larger share of jobs is allocated to the population center that is closest. As before, the sum of jobs allocated to the population centers matches the total number of jobs available.

We can combine the proportional allocation factors by population and distance as follows:

$$V_{ij} = W_i \frac{f_{ij}^p \cdot f_{ij}^c}{\sum_{k=1}^K f_{kj}^p \cdot f_{kj}^c}$$

In the example:

$$\begin{aligned} V_{ij} &= W_j \cdot \frac{f_{ij}^o \cdot f_{ij}^c}{f_{ij}^o \cdot f_{ij}^c + f_{kj}^o \cdot f_{kj}^c} = 300 \frac{\left(\frac{2}{3}\right) \left(\frac{0.426}{0.475}\right)}{\left(\frac{2}{3}\right) \left(\frac{0.426}{0.475}\right) + \left(\frac{1}{3}\right) \left(\frac{0.574}{0.475}\right)} = (300) \left(\frac{0.284}{0.475}\right) = 179.4 \\ V_{kj} &= W_j \cdot \frac{f_{kj}^o \cdot f_{kj}^c}{f_{ij}^o \cdot f_{ij}^c + f_{kj}^o \cdot f_{kj}^c} = 300 \frac{\left(\frac{1}{3}\right) \left(\frac{0.574}{0.475}\right)}{\left(\frac{2}{3}\right) \left(\frac{0.426}{0.475}\right) + \left(\frac{1}{3}\right) \left(\frac{0.574}{0.475}\right)} = (300) \left(\frac{0.191}{0.475}\right) = 120.6 \end{aligned}$$

Notice how fewer jobs are allocated to population center  $i$  compared to the allocation by population only, to account for the higher cost of reaching the employment center. On the other hand, distance alone allocated more jobs to

the closest population center (i.e.,  $k$ ), but since it is smaller, it also get a smaller share of the jobs overall. Again, the sum of jobs at employment center  $j$  that are allocated to population centers  $i$  and  $k$  simultaneously based on *population*- and *distance*- based allocation is preserved (i.e.,  $W_{ij} + W_{kj} = W_j$ ).

Availability is simply the sum of the above by origin:

$$V_i = \sum_{j=1}^J V_{ij}$$

This quantity represents opportunities (e.g., jobs) that can be reached from  $i$  (i.e., they are accessible), and that are *not* allocated to a competitor: therefore the weighted sum of available opportunities. Compare  $V_i$  to the singly-constrained gravity model (see Wilson (1971)). In essence,  $V_i$  is the result of constraining  $A_i$  to match one of the marginals in the origin-destination table, the known total of opportunities.

Since the sum of opportunities is preserved in the procedures above, it is possible to calculate a highly interpretable measure of spatial availability per capita (call it lower-case  $v_i$ ) as follows:

$$v_i = \frac{V_i}{P_i}$$

In the example above:

$$\begin{aligned} v_{ij} &= \frac{V_{ij}}{P_i} = \frac{179.4}{240} \\ v_{kj} &= \frac{V_{kj}}{P_k} = \frac{120.6}{120} \end{aligned}$$

Less competition ( $P_k$  is the smallest population center in the region) and being closer to the jobs clearly works in favor of individuals at  $k$ . Where the overall ratio of jobs to population in the region is  $300/(240 + 120) = 0.83$ , the spatially available jobs per capita at  $k$  is closer to unity.

In the following sections we use the same numerical example presented above to illustrate how availability  $V_i$  is calculated. We conclude by contrasting the two measures in the final section.

#### *1st Use Case: Available Jobs for The Working Population*

In the following examples we use the same impedance function that was used to illustrate the accessibility calculations in the numerical example. The spatial availability calculations are implemented in the function `sp_avail`. The inputs are an Origin-Destination table with labels for the origins (`o_id`), labels for the destinations (`d_id`), the population (`pop`) and number of opportunities (`opp`), an indicator for catchments or other eligibility constraints (`r`), and a pre-calculated impedance function (`f`). For this example, we assume that there are no catchment restrictions by setting `r` to 1.

The value of the function (its output) is a vector with  $V_{ij}$  given the inputs, that is, the opportunities available to  $i$  from  $j$ :



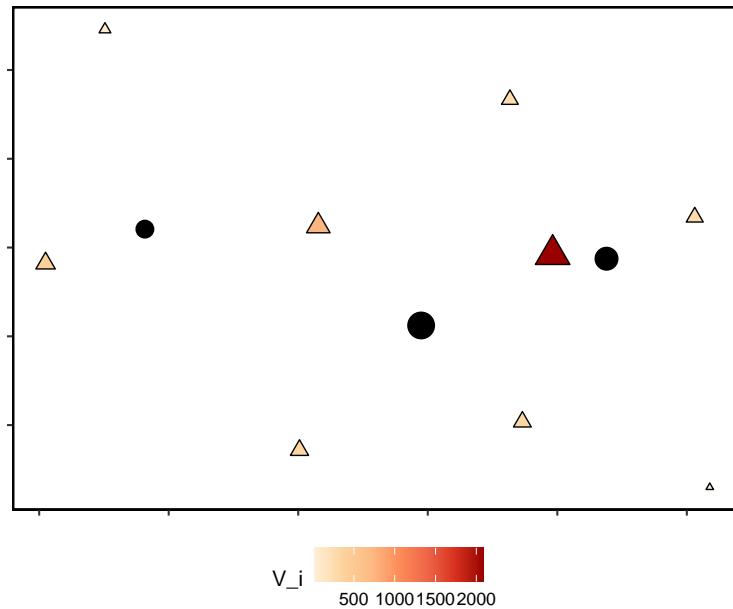


Figure 2: Spatial availability of jobs

```
[1] 4500
```

```
[1] 4500
```

```
# A tibble: 9 x 2
```

```
  Origin      V_i
  <fct>    <dbl>
1 Population 1    67.0
2 Population 2   414.
3 Population 3   336.
4 Population 4   745.
5 Population 5  2081.
6 Population 6   270.
7 Population 7   256.
8 Population 8   316.
9 Population 9    14.9
```

The following plot shows the estimates of spatial availability:

We see that population center 5 has the highest level of spatial availability, due to being a large population center that is moreover relatively close to jobs. To improve the interpretability of this measure, we first note that the regional measure of jobs per capita is 0.994. We then calculate the spatially available jobs per person at each population center:

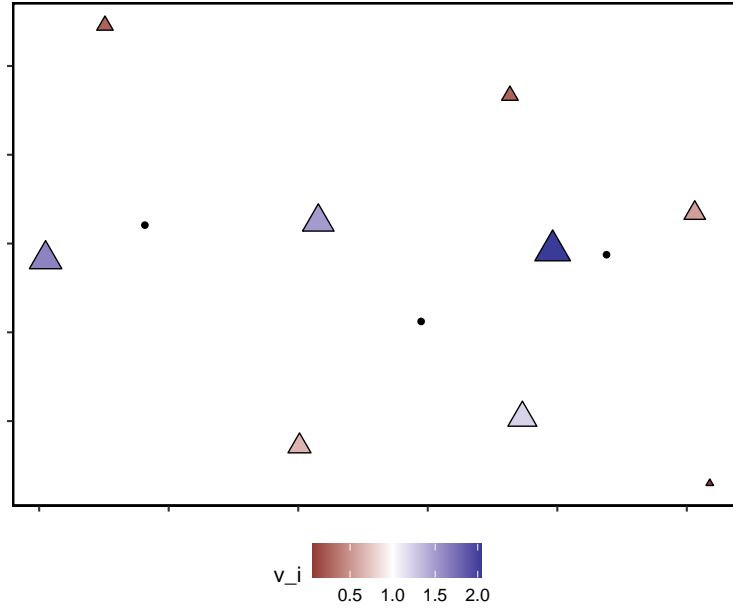


Figure 3: Spatial availability of jobs per capita

	id	v_i
1	Population 1	0.25787932
2	Population 2	1.62445872
3	Population 3	0.65947335
4	Population 4	1.50495734
5	Population 5	2.03989862
6	Population 6	0.55082935
7	Population 7	0.26164373
8	Population 8	1.21352615
9	Population 9	0.05842544

Plot the spatial availability per person:

Some population centers have almost two jobs available per person (compared to the overall regional value of approximately one job per capita), while others have less than one job available per person. This does not mean that people are not taking some of the jobs. It means that controlling for the cost of reaching jobs, they are worse off than those with more jobs spatially available.

#### *2nd Use Case: Available Workers for Employment Centers*

We can also examine the pool of workers available to each employment center by considering the workers as the opportunities and the jobs as the population.

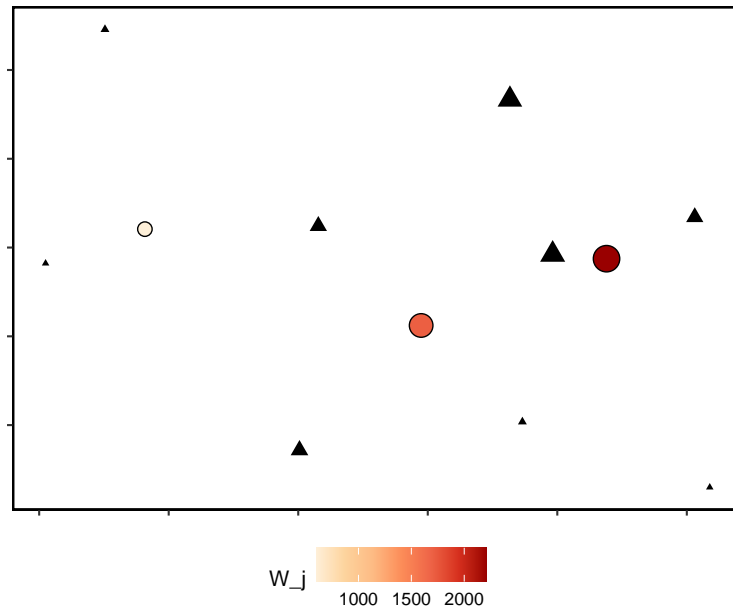


Figure 4: Spatial availability of workers

Calculate the proportional allocation of jobs to population (referred to as  $W_{ji}$ ):

```
[1] 4525
```

```
[1] 4525
```

```
# A tibble: 3 x 2
  Destination      W_j
  <fct>          <dbl>
1 Employment Center 1 610.
2 Employment Center 2 1710.
3 Employment Center 3 2205.
```

Plot the availability estimates:

```
      id      w_j
1 Employment Center 1 0.8137264
2 Employment Center 2 0.7597858
3 Employment Center 3 1.4701247
```

Plot the spatial availability of workers per job:

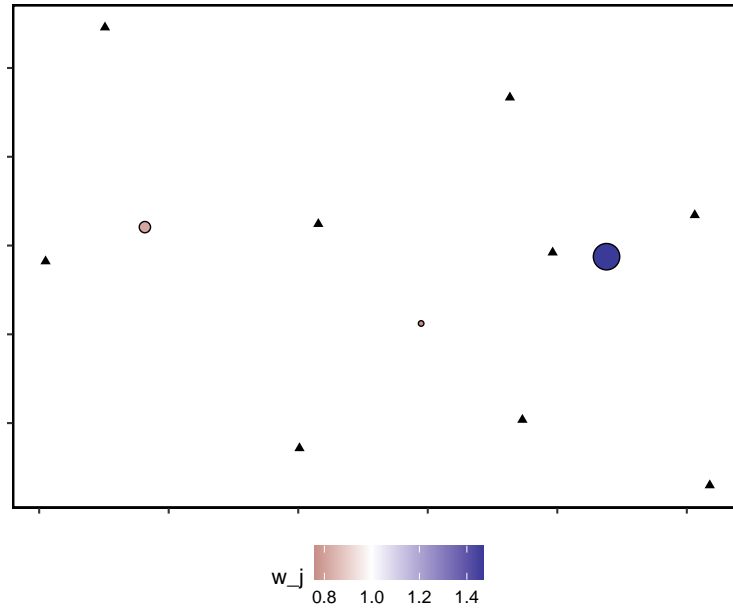


Figure 5: Spatial availability of workers per job

### 3rd Use Case: Available Jobs for Specialized Working Populations

In this section we introduce catchment/eligibility constraints. Due to differences in educational achievement among the population, the jobs in Employment Center 1 can only be taken by individuals in population centers 1 and 2. Jobs in Employment Center 2 can be taken by individuals in population centers 3, 4, 5, 7, and 8. Lastly, jobs in Employment Center 3 require qualifications available only among individuals in population centers 5, 6, 8, and 9. See figure below.

```
[1] 4500
```

```
[1] 4500
```

```
# A tibble: 9 x 2
  Origin      V_i_r
  <fct>      <dbl>
1 Population 1  104.
2 Population 2  646.
3 Population 3  303.
4 Population 4  527.
5 Population 5 2173.
6 Population 6  257.
7 Population 7  161.
```

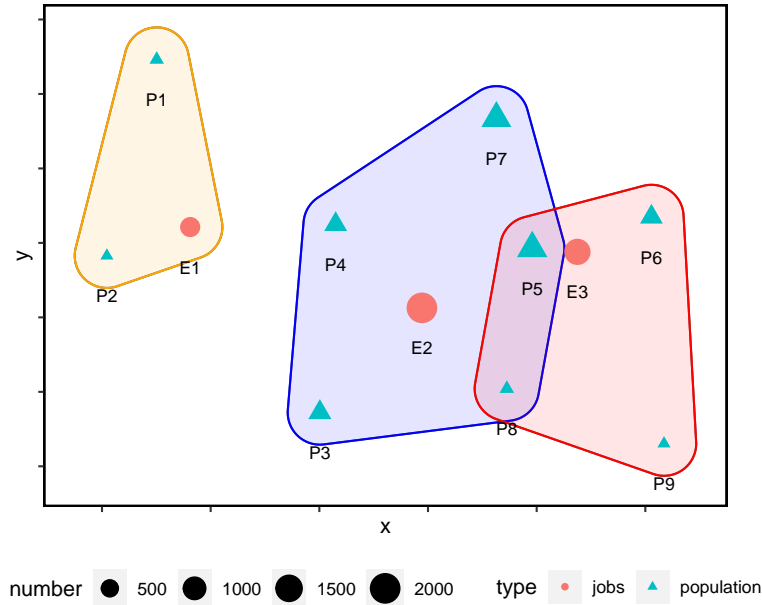


Figure 6: Catchments areas in numeric example

```
8 Population 8 322.
9 Population 9 7.52
```

Plot the availability estimates:

Available jobs per person with catchment/eligibility conditions:

The plot in Fig. 8 shows the availability per person without and with catchment restrictions.

We can see that when there are catchment restrictions population center 2, despite being relatively peripheral, has higher spatial availability due to spatial specialization. With catchments, the spatial availability of jobs declines from the perspective of population center 4: the population here has skills required for jobs at a small employment center, where they face substantial competition from other population centers.

### Empirical Example

Words go here.

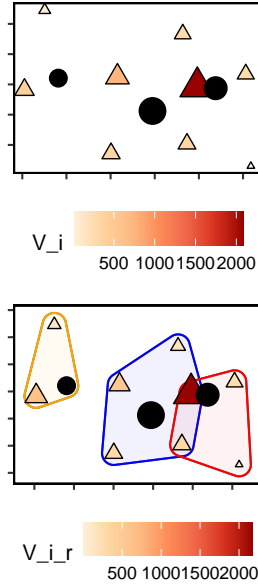


Figure 7: Spatial availability of jobs with catchment restrictions

## Discussion

### *Comparing Accessibility and Spatial Availability*

Below we compare accessibility and the first case of spatial availability illustrated above. To facilitate the comparison, we index the measures so that the lowest value corresponds to an index of 100:

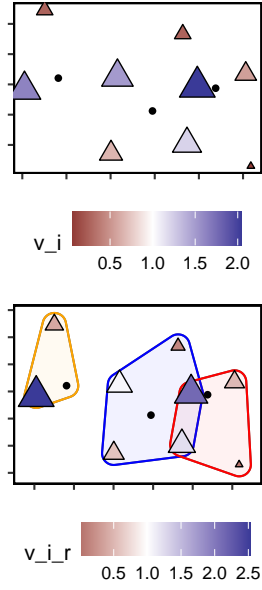
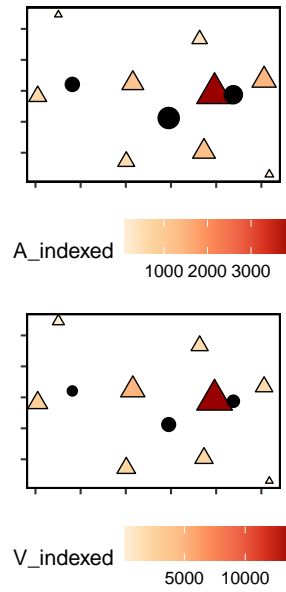


Figure 8: Spatial availability of jobs per capita with and without catchment restrictions



The figure suggests that high accessibility does not necessarily mean high avail-

ability.

*How are these measures different?*

The sum of all spatial availability values is equal to the total number of opportunities within the region. As such, the value for each population center reflects how many opportunities are *available* accounting for the indivisibility of the opportunities. This provides greater interpretability of the results. Comparatively, accessibility values associated with each population centers reflects how many jobs can *potentially* be reached by each population center; these values are not adjusted proportionally to the number of population (i.e. workers ).

Secondly, spatial availability is not robust to the modifiable unit problem, since the estimates account for the population at the centers. If more population centers are added, the availability adjusts accordingly by allocating opportunities proportionally.

Further, the measure of spatial availability can be a useful way to distinguish between high accessibility/high population centers (which potentially can result in lower availability due to competition), and contrariwise, low accessibility/low population centers, which may enjoy higher availability than the accessibility calculations may suggest. Remote, smaller population centers can be sufficiently accessible and in close proximity to the smaller employment centers; however this “sufficiency” is obscured by accessibility measure by over-inflating the accessibility of population centers which are more central to more (and larger) employment centers. Conventional accessible does not shed light on how sufficiently accessible opportunities are available to the population.

As a final point, we note that the measure proposed, by producing a concrete and interpretable number of opportunities available, can be meaningfully compared to the total number of opportunities in the region. Likewise for opportunities available per capita. This is an important topic in equity analysis. For instance, considering the two remote small population centers in the top left corner it is evident that for individuals in population center 2, despite facing low nominal accessibility, have a relatively high number of available jobs per capita.

## Conclusion

Words go here.

## References