

$${}_xP_y^* = \sum_{i=1}^n \left( \frac{x_i}{X} \right) \left( \frac{y_i}{t_i} \right) \quad (1.2.14)$$

$${}_xP_x^* = \sum_{i=1}^n \left( \frac{x_i}{X} \right) \left( \frac{x_i}{t_i} \right) \quad (1.2.15)$$

where all variables are defined as in Eqs. 1.2.4 and 1.2.5.

The core ideas behind these measures have already been translated into the activity-space context in a number of studies. Schnell and Yoav (2001) propose a set of isolation indexes that rely on the proportions of same-group members that an individual encounters in the course of various social and geographic settings, weighted by the duration of time in each setting and the salience attached to it by the individual. Wong and Shaw (2011) rely on the proportion of same-group members whose activity spaces overlap a given individual’s activity space irrespective of timing. Toomet et al. (2012) use the concept of co-presence—the number of occasions in which two individuals are proximate in time and space and thus could potentially interact—to construct a measure of homophily defined as the proportion of same-group member co-presence events in an individual’s total set of co-presence events.

Each of these indexes is tailored to the particular data that was available for the study in which it was employed: Schnell and Yoav (2001) relied on surveys that asked relatively small numbers of respondents about the people they encountered in various settings and the salience they attached to these settings; Wong and Shaw (2011) relied on travel diaries; and Toomet et al. (2012) relied on a mobile network provider’s massive database of call detail records (CDRs), which identify the cell tower from which each mobile phone call is made or received, along with a unique identifier and limited demographic information about the phone number’s owner.

For a general index of activity-space interaction or isolation, ideally we would combine Schnell and Yoav (2001)’s duration weights with the spatial detail of Wong and Shaw

(2011)’s travel diaries and the huge sample size of Toomet et al. (2012)’s CDRs.<sup>5</sup> The necessary data for such an index could be obtained from CDRs augmented by inferences about trajectories and durations between call locations drawn from models generated from active mobile positioning methods or detailed travel surveys. Given such data, I propose a duration-weighted extension of Toomet et al. (2012)’s co-presence approach.

The extension is relatively straightforward, although I define both an isolation and an interaction version, following the same logic as is used in the residential segregation indexes. I define activity-space isolation,  ${}_xP_x^a$  (analogous to Toomet et al. (2012)’s homophily), as:

$${}_xP_x^a = \sum_{i=1}^n \frac{x_i}{b_i n} \quad (1.2.16)$$

where  $x_i$  is the number of G1 person-hours that fall within G1 individual  $i$ ’s personal interaction space, defined as a cylinder with radius  $r$  along the spatial axes and height  $h$  along the time axis,  $b_i$  is the number of person-hours of both groups that fall within G1 individual  $i$ ’s personal interaction space, and  $n$  is the total G1 members.

Activity-space interaction,  ${}_xP_y^a$ , follows as:

$${}_xP_y^a = \sum_{i=1}^n \frac{y_i}{b_i n} \quad (1.2.17)$$

where  $y_i$  is the number of G2 person-hours that fall within G1 individual  $i$ ’s personal interaction space (defined in Eq. 1.2.16),  $b_i$  is the number of person-hours of both groups that fall within G1 individual  $i$ ’s personal interaction space, and  $n$  is the total number of G1 members.

Note that, like the activity-space extension of White’s spatial proximity index discussed above, these indexes do not rely on areal unit measures. This is advantageous in that it (1) allows for better-tailored distances and (2) avoids the well-known modifiable areal unit

---

<sup>5</sup>I do not propose combining Schnell and Yoav (2001)’s salience weights here because (1) the necessary data would be difficult to gather on a large scale, and (2) salience adds a level of specificity to the index that makes it less useful for broad inter-study comparisons.

problem (Kwan, 2009). An additional benefit is that this allows altitude to be incorporated into the indexes to the extent that it can be captured in the data. Although altitude might seem like a strange concern, any index constructed in a city with tall buildings will need to confront it. People who overlap in two-dimensional space may be on different floors of the same building and never come into contact. Moreover, the environments to which they are exposed on their different floors may also be very different. Just consider the situation of a CEO in a sky-scraper’s top-floor office compared to the doorman working in the lobby or the janitor working in the basement.

### **1.2.3 Segregation as a question of movement**

Activity-space segregation also involves systematic differences in movement. The distance a person travels each day for work, for example, is an important activity-space property, irrespective of the implications this has for contact with places or people (even though these implications may be significant as well), as it affects the amount of time that may be devoted to work and leisure activities and a variety of other important factors related to health and quality of life. Whether a person commutes by car or by foot has health implications, some of which will depend on places through which the commuting path passes, but some of which will be entirely independent of these places. In theory, one can imagine a city in which the place- and people-based measures of segregation indicate only small differences among social groups, but where one group commutes a long distance to work while another stays close to home, or where one group travels everywhere by car while another walks. These are important differences to capture, and while they never appear in the residential segregation context, they are readily apparent in the activity-space context.

The recognition that there are systematic differences in different groups’ movement properties has long been part of the activity-space literature, particularly in its focus on gender, and there are already a number of methods for measuring these differences. In fact, the basic measures and definitions of activity space itself are largely drawn to facilitate compar-

isons between groups. Most of these measures focus on the size or spatial extent of activity spaces and the technical details of how they do this are based on different assumptions about where people are located in the moments for which we lack data.

The most conservative approach to measuring activity space considers only the places in which people are actually observed in the data. This is sometimes referred to as the point-based method, although it may be relaxed slightly for some purposes to include not just observed points, but also the areal units into which they fall (Wong & Shaw, 2011). The shortest-path distances (either the direct geodesic distance or the distance along a constraining road network) between sequential points can be summed to produce a measure of total travel distance, and a buffer may be added around the paths to produce a measure of activity-space area (Schönfelder & Axhausen, 2003). Alternatively, the maximum distance between any two points in a given day may be used as an estimate of the spatial extent of a person’s activity space, what Isaacman et al. (2010) refers to as daily range.

Kernel density estimates provide a somewhat less conservative approach to measuring activity space. Here, the goal is to estimate a function that captures the probability of an individual being found at any given location in space based on the observed data. Different kernel functions and bandwidths may be used, but the overall effect is to smooth the discontinuous location data into a probability density surface over a given geographic area (Schönfelder & Axhausen, 2003; Kwan, 2000).

Another measure of activity space can be achieved by drawing ellipses around the person’s observed locations. This is usually done using only one ellipse, which is centered on the arithmetic mean of the locations and drawn with its long axis oriented in the direction of maximum dispersion and its short axis oriented in the direction of minimum dispersion (Wong & Shaw, 2011). The axis lengths are based on the standard deviations of the observed distances from the mean, so these are sometimes called confidence ellipses, implying that they express some level of statistical confidence about the person’s expected location. Somewhat more realistic approaches use the person’s home as the center of the ellipse, or rely on

two ellipses, one centered on the home and one centered on the workplace (Schönfelder & Axhausen, 2003).

Finally, activity spaces may be measured using minimum convex polygons. This is simply the area of the smallest convex polygon that can be drawn around all or some fraction of a person's locations (Buliung & Kanaroglou, 2006; Jones & Pebley, in press). For instance, Figure 1.6 shows the minimum convex polygon around the central 95% of the observed locations of one participant in the Human Mobility Project (Palmer et al., 2013).

The choice of a method for measuring activity-space size is often linked to the available data. The more comprehensive the data, the better suited are the point-based or kernel density approaches, as these are more likely to capture all of the spaces actually used and avoid the biases of ellipses and polygons, which are more sensitive to activity-space shape and sampling error (Burgman & Fox, 2003). Even without comprehensive data, the point-based daily range approach used in (Isaacman et al., 2010) may be best for facilitating comparisons between cities and across studies because it is simple to implement and interpret, and it is not dependent on many assumptions. Chapters 2 and 3 compare the bias and uncertainty of daily range and minimum convex polygon estimates using empirical data from a mobile phone study and two population-level simulations.

Whatever the method, a basic index for comparing two groups' activity-space sizes can be calculated as a difference in means. Where data is available for each individual over the course of multiple days, then the individual's activity-space size may be first aggregated as the maximum or median of that individual's observed activity spaces (Isaacman et al., 2010).

### **1.2.4 Additional aspects**

Although places, people, and movement are the three forms of separation on which this dissertation focuses in constructing indexes by which activity-space segregation may be measured,

a number of additional aspects of segregation in space and time should be mentioned.

First, just as residential segregation indexes capture separation in purely spatial terms, one could think of separation in purely temporal terms. Doing this would require some delimitation of spatial boundaries within which this temporal segregation is measured. This could be done functionally, which would directly parallel residential segregation. For instance, one could measure temporal segregation in people's presence at their homes, capturing the extent to which different groups tend to be at home at different times of day. The delimitation could also be explicitly spatial, for instance, measuring temporal segregation in different groups' presence in a specific neighborhood or census tract.

Second, one could focus more on the "activity" in activity-space segregation, which is has been largely ignored here. Instead of focusing only on the places and times in which people carry out their daily activities or the characteristics of the movements employed in doing this, one could focus on what the activities actually are, and the extent to which different groups perform similar activities at different times of day or in different spaces within a city.

Finally, there is important overlap and interaction in all of the aspects of activity-space segregation discussed so far. Presence in a given space at a given time, and proximity to the people within that space may have important social consequences that go beyond the mere sum of the consequences discussed so far with respect to places and people. These consequences may be further compounded by the specific activities performed.

A fascinating example of this last point is provided by Sewell (2011), who describes women's use of public space in San Francisco at the turn of the Twentieth Century. Sewell stresses the complex inter-relationships between the landscape as it is imagined, the landscape as it is actually experienced, and the landscape as a built environment. She shows how San Francisco's built landscape during this period encouraged women to use public spaces in ways that challenged popular images of gender segregation. This, in turn, had important consequences in shaping views about women's roles in society. For example, during the

women’s suffrage campaign of 1911,

suffragists utilized spaces they had made their own through everyday use, as consumers and workers. Only because these spaces had first become part of women’s everyday lives, and people were used to seeing them there, were suffragists successful in their efforts to refashion them as political spaces (Sewell, 2011, p. 166).

### 1.3 Demonstration with simulated data

To demonstrate some of the proposed indexes, I have applied them to data generated from a large computer simulation of human movement set in Buffalo, New York. The simulation started with the full 2000 census populations of non-Hispanic whites ( $n = 151,488$ ) and non-Hispanic blacks ( $n = 107,103$ ) in the City of Buffalo, placed randomly on streets within their census tracts of residence. Each simulated person “walked” along the Buffalo City and Erie County road network at 4 km per hour for 8 hours of simulation time. As explained in more detail in Chapter 3, the walks were programmed to take the form of network-constrained, truncated Lévy walks: After choosing a random direction along the road, each person then chose a random distance drawn from a truncated power law distribution. The minimum and maximum of this distribution were set to 100 m and 100 km, and after each step a new direction and distance were randomly drawn. Road and direction were also randomly selected at each intersection, and when people reached the end of a road or the county boundary they simply turned around and continued walking.

The simulation is obviously a far stretch from reality, but there are reasons to think Lévy walks provide a good statistical approximation to actual movement patterns, at least at large scales (see Chapter 3), and it provides a rich set of data with which activity-space indexes may be tested and demonstrated. The simulation also serves as a baseline against which to compare empirical data and helps in tackling the question of how activity-space

segregation is influenced by place of residence. We expect people to be well mixed when moving along random paths, but this mixing may take some time when their starting points and ending points each day are highly segregated, as is the case in Buffalo. The simulation helps us to illustrate heuristically how quickly residentially-induced segregation might drop during the day if there were no additional factors causing segregation outside the home.

As a simple visualization of the changes in census tract composition that occurred during the simulation, Figure 1.7 in the electronic version of this dissertation shows an animation of the Buffalo City black-white tract ratios at each half-hour time slice. Buffalo's high level of residential segregation is immediately apparent in the first frame, with over four times as many blacks as whites in the central census tracts, and less blacks than whites in the surrounding ones. As the simulation progresses, segregation clearly decreases, with no tracts containing more than four times as many blacks as whites by the end. Even at the end of the simulation, however, notable differences remain in the composition of the central tracts and their surroundings. The simulation covers only 8 hours, but one can imagine these people following the same paths back to their homes, and with the level of segregation consequently increasing to the residential starting point by the end of the day.

Results of the basic activity-space segregation index calculations are shown in Figure 1.8. (Note that the scales of these plots differ so that as much detail from each one can be shown.) The averaging and person hours indexes are indicated with the red and green horizontal lines. For comparison, the corresponding traditional residential segregation indexes (calculated at time zero, when all simulated people are within their actual census tracts of residence) are indicated with the blue horizontal lines, and the traditional indexes calculated at each half-hour increment are indicated with the black curves. The averaging and person-hours approaches yield nearly identical results and in all cases the activity-space measures show lower segregation than the residential measures. (The two measures are not expected to be identical in this simulation because the simulated people were allowed to leave Buffalo City as long as they remained within Erie county.) More interestingly, moderate levels of



segregation remain even after 8 hours of random mixing.

According to the activity-space extensions of the dissimilarity index, over the course of the 8-hour period on average 55% of blacks would have needed to move to different census tracts in order to achieve an even distribution relative to whites, and 45% of blacks would have needed to do so in order to achieve an even distribution relative to tracts. Similarly, 55% of black person hours would have needed to be moved to different space-time units in order to achieve an even distribution relative to white person-hours, and 44% would have needed to be moved in order to achieve an even distribution relative to tracts.

In addition, blacks were closer to other blacks and whites were closer to other whites throughout the simulation than they were to members of the other group, giving an average spatial proximity score of 1.25 when the index is calculated with census tracts. Likewise, black person-hours were closer in time and space to other black person hours and white person-hours were closer in time and space to other white person hours than either of these were to person-hours of the other group. Finally, on average blacks were located in census tracts with contemporaneous populations that were 36% white and black person-hours were located in tracts with contemporaneous person-hour populations that were 36% white.

Figure 1.9 shows the individual distance version of the spatial proximity index ( $SP^a$ ) and the personal exposure index ( ${}_xP_y^a$ ). These indexes are calculated using simple random samples of 2000 individuals from the full census population in order to reduce computational requirements. As an estimate of uncertainty, the indexes are calculated with each of 480 samples and the values within which 90% of the results lie are plotted. The horizontal red bands in the figure indicate the results of the indexes calculated over the 8 hour period; the grey bands indicate results for each half-hour time slice and the black curve indicates the mean of these results.

These two individual-level measures of activity-space segregation correspond closely with the areal unit extensions of the traditional residential segregation measures. On one hand, this suggests that the individual level measures may not add much to what we can

*Review Letters*, 88(9).

- Bartumeus, F., da Luz, M. G. E., Viswanathan, G., & Catalan, J. (2005). Animal search strategies: A quantitative random-walk analysis. *Ecology*, 86(11), 3078–3087.
- Bartumeus, F., & Levin, S. A. (2008). Fractal reorientation clocks: Linking animal behavior to statistical patterns of search. *Proceedings of the National Academy of Sciences*, 105(49), 19072–19077.
- Basta, L. A., Richmond, T. S., & Wiebe, D. J. (2010). Neighborhoods, daily activities, and measuring health risks experienced in urban environments. *Social Science & Medicine*, 71(11), 1943–1950.
- Blair, S., & Lichter, D. (1991). Measuring the division of household labor gender segregation of housework among american couples. *Journal of Family Issues*, 12(1), 91–113.
- Breckenridge, K. (2005). Verwoerd’s bureau of proof: Total information in the making of apartheid. *History Workshop Journal*, 59(1), 83–108.
- Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075), 462–465.
- Buliung, R. N., & Kanaroglou, P. S. (2006). Urban form and household activity-travel behavior. *Growth and Change*, 37(2), 172–199.
- Burgman, M., & Fox, J. (2003). Bias in species range estimates from minimum convex polygons: Implications for conservation and options for improved planning. *Animal Conservation*, 6(01), 19–28.
- Cassels, S., Curran, S. R., & Kramer, R. (2005). Do migrants degrade coastal environments? migration, natural resource extraction and poverty in North Sulawesi, Indonesia. *Human Ecology*, 33(3), 329–363.
- Chaix, B., Merlo, J., Evans, D., Leal, C., & Havard, S. (2009). Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. a response to Riva, Gavin, Apparicio and Brodeur. *Social Science & Medicine*, 69(9), 1306–1310.
- Chaskin, R. J. (1997). Perspectives on neighborhood and community: A review of the