Catchment zones of co-location: A statistical analysis of spatial proximities among various types of stores on a street network

Wataru Morioka [1*], Mei-Po Kwan [2], Atsuyuki Okabe [3], Sara L McLafferty [1]

1. Department of Geography & GIScience, University of Illinois at Urbana-Champaign; 2. The Chinese University of Hong Kong; 3. Aoyama Gakuin University

INTRODUCTION

In many central districts in cities across the world, different types of stores form clusters resulting from the benefits of spatial agglomeration. To precisely analyze the co-location relationships among multi-types of stores in a micro-scale space, this study develops GIS-based spatial statistical methods, the original and incremental network dual K function methods, by addressing the limitations of the ordinary cross K function method.

Objectives

- To formulate an exact statistical method for analyzing co-location in a fairly small area with a street network.
- To specify the attractiveness (catchment zone) of each store type by extend the above method.
- To demonstrate the usefulness of the methods through an empirical analysis.

METHOD

Formularization

a network, $N = (V, L)$, consisting of
a set of nodes $V = \{v_1, \ldots, v_{n_V}\}$; a set of links $L = \{l_1, \ldots, l_{n_L}\}$.
On the network, $N = (V, L)$, there are
a set of type A points $P_A = \{p_{A1}, \ldots, p_{An_A}\}$;
a set of type B points $P_B = \{p_{B1}, \ldots, p_{Bn_B}\}$.

$\tilde{L}$: the total length of the set of links.

$\tilde{L}(x_{i-1}, x_i | P_B)$: the total length of the buffer-ring network of type B points between the distance band x_{i-1} and x_i.

$K'_{A\to B(x_{i-1},x_i)}$: the number of type A points on $\tilde{L}(x_{i-1}, x_i | P_B)$.

Under homogeneous binomial point process or Complete Spatial Randomness (CSR) hypothesis, the probability that k points of type A are on the specific distance band $N(x_{i-1}, x_i | P_B)$ is mathematically given by,

$$\Pr\left[K'_{A\to B(x_{i-1},x_i)} = k\right] = {}_{n_{P_A}}C_k \left(\frac{\tilde{L}(x_{i-1}, x_i|P_B)}{\tilde{L}}\right)^k \left(1 - \left(\frac{\tilde{L}(x_{i-1}, x_i|P_B)}{\tilde{L}}\right)\right)^{n_{P_A}-k}. \qquad (1)$$

, where n_{P_A} is the number of type A points. Similarly,

$$\Pr\left[K'_{B\to A(x_{i-1},x_i)} = k\right] = {}_{n_{P_B}}C_k \left(\frac{\tilde{L}(x_{i-1}, x_i|P_A)}{\tilde{L}}\right)^k \left(1 - \left(\frac{\tilde{L}(x_{i-1}, x_i|P_A)}{\tilde{L}}\right)\right)^{n_{P_B}-k}. \qquad (2)$$

The overall effect of co-location across a certain range is computed by the following equations,

$$\overline{p^+}_{A\to B(0,x_i)} = \frac{1}{m}\sum_{i=1}^{m} \Pr\left[K'_{A\to B(0,x_i)} \geq k\right], \qquad (3)$$

$$\overline{p^+}_{B\to A(0,x_i)} = \frac{1}{m}\sum_{i=1}^{m} \Pr\left[K'_{B\to A(0,x_i)} \geq k\right]. \qquad (4)$$

We will judge that type of points are co-located to another type points, if $\overline{p^+}_{A\to B\,(x_0,x_i)} < 0.05$. To specify the catchment zone of co-location, we counts type A points separately in each band, i.e., the incremental network dual K function, $K'_{A\to B(x_{i-1},x_i)}$, as well as the original one, $K'_{A\to B(0,x_i)}$.

Illustrations of our methods' advantages, compared to the ordinary cross K function

- **Measuring the distance correctly by assuming the shortest-path distance on a network-constrained space** (e.g., red squared part in Figure 1)
- **Considering a dual relationship** (e.g., A to B, B to A)

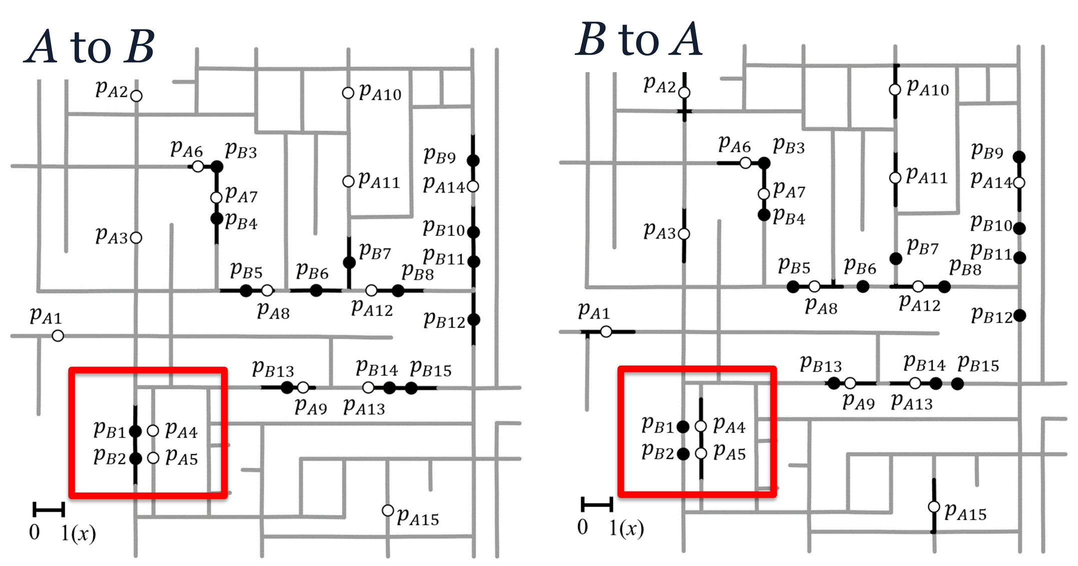

Figure 1. (a)The buffer network of $\tilde{L}(0,1|P_B)$. (b)The buffer network of $\tilde{L}(0,1|P_A)$. The bold lines indicate the buffer network of type B or A points with $x = 1$.

- **By focusing on the nearest type B points relative to each type A point, the number of type A points within x from type B points follows the simple binomial distribution. This characteristic enables us to analyze co-location with exact statistical formula.**

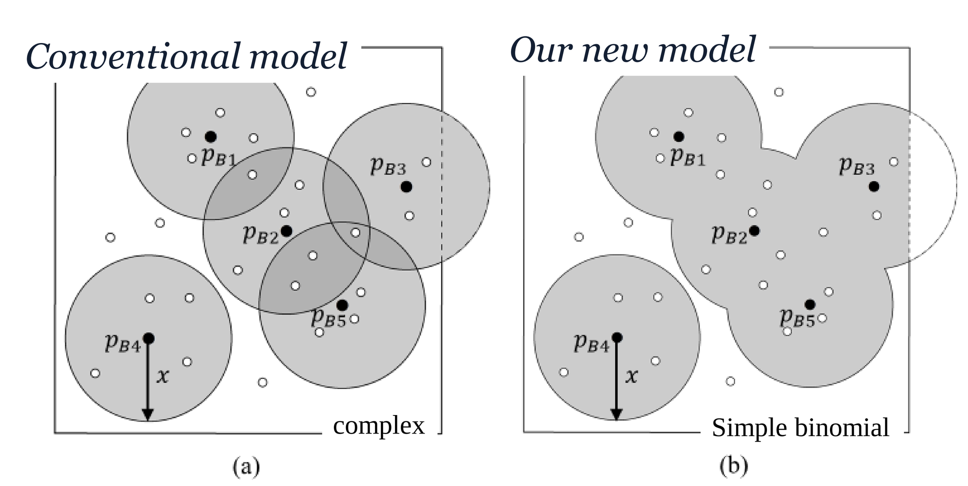

Figure 2. Buffer zones of type B points. (a) In the ordinary cross K function, the buffer zone consists of overlapping circles, some of which are outside the square (study area), (b) In our model, the buffer zone is the dissolved circles within the square.

- **Besides counting type A points cumulatively, we count them for each interval, specifying catchment zones.**

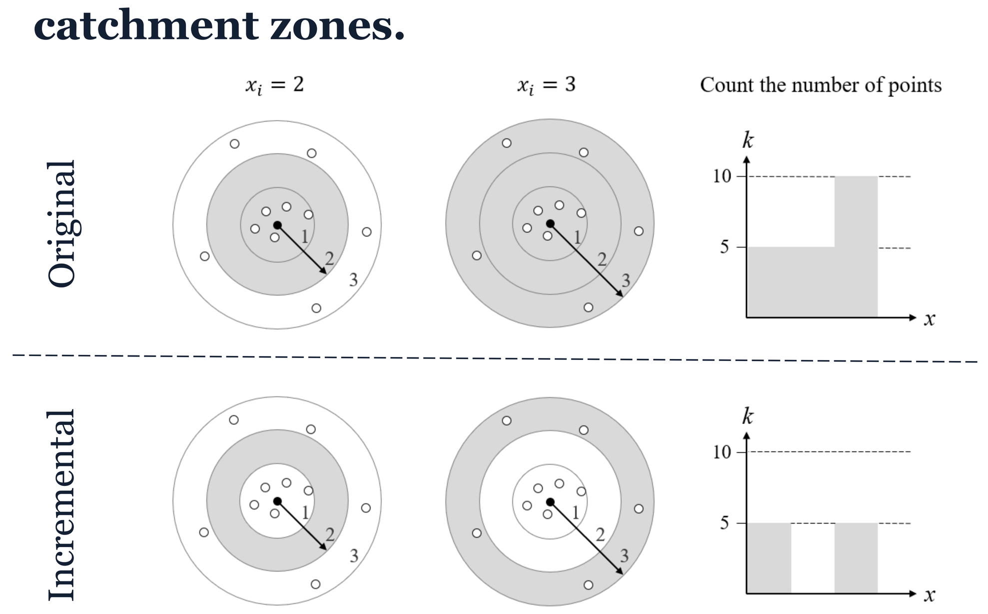

Figure 3. (upper) original cross K, (lower) incremental cross K

EMPIRICAL ANALYSIS

Study Area & Data

The study area is Shibuya and Minato wards, one of the trendiest districts in Tokyo. Store locations are obtained by TownPage telephone book (NTT TownPage Corporation) in July, 2019. We choose top 30 types of stores (listed below) and set x_i ranging from 20 meters to 400 m at intervals of 20 m.

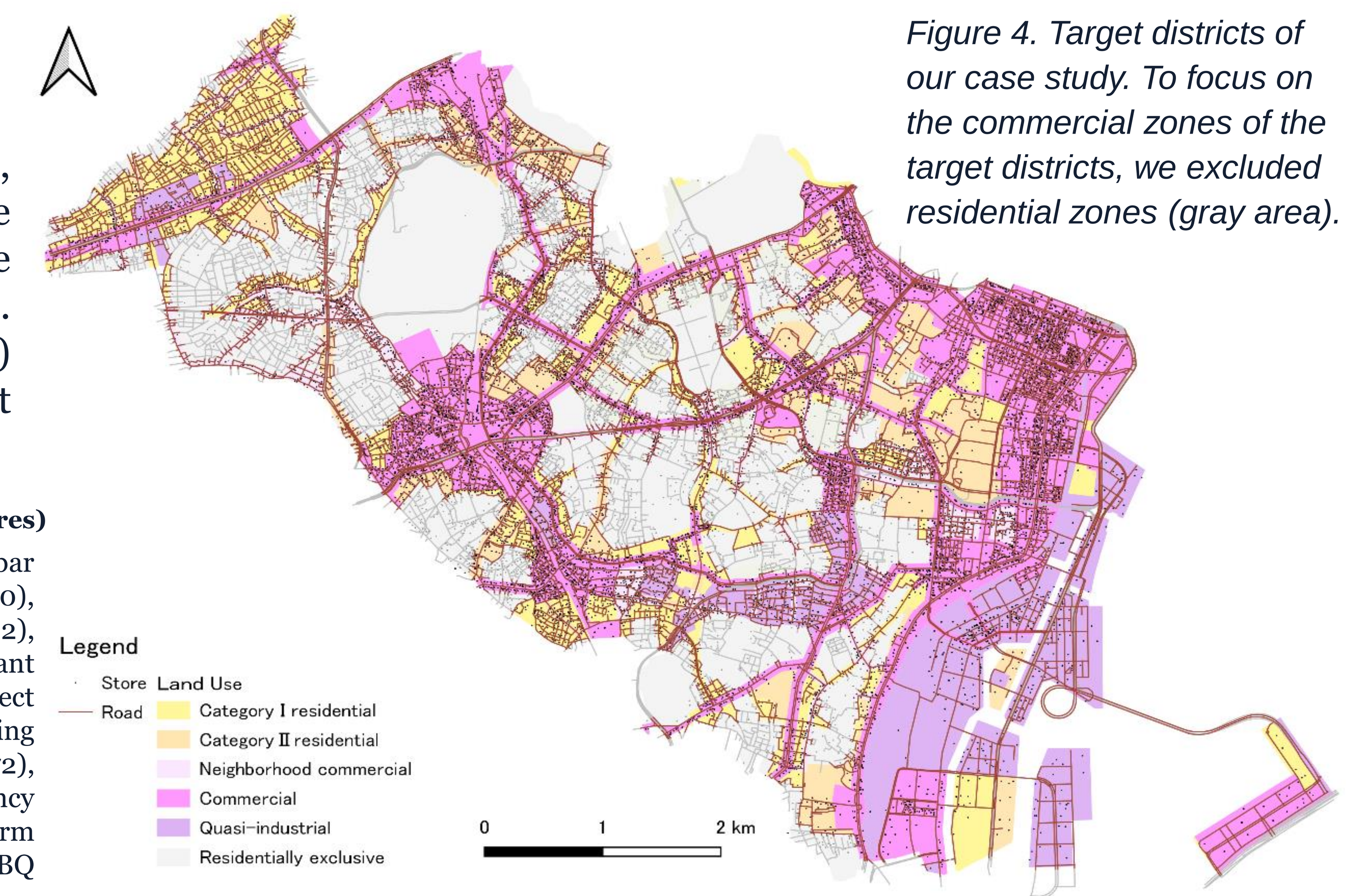

Figure 4. Target districts of our case study. To focus on the commercial zones of the target districts, we excluded residential zones (gray area).

Top 30 types of stores (order by the number of stores)

real estate agency (1605), Japanese restaurant (1103), Japanese bar (1049), hair salon (943), bar & club (843), dental clinic (690), advertising agency (663), software service (587), law office (562), convenience store (546), accounting firm (513), western restaurant (506), rental office (436), building management (426), architect design (414), internet service (359), beauty salon (334), consulting firm (327), sushi restaurant (228), café (278), pharmacy (272), Italian restaurant (265), entertainment agency (253), travel agency (243), printing service (242), yakitori stand (242), construction firm (239), graphic design (236), insurance company (234), BBQ restaurant (233)

Results

Figure 5. The Incremental network dual K function method to the top 30 types of stores, grouped by categories. X-axis is the parametric distance band (x_i) and Y-axis is the upper probability of $p^+_{A\to B(x_{i-1},x_i)}$. Each line depicts the median value, summarized by each store type.

When we applied the original network dual K function method (cumulative one) to the top 30 store types in the target area, 830 out of 870 cases are judged as co-location, which means many store types are gathered in a fairly small area. To understand the catchment zone (distance effects of attractiveness) of each store type, we also applied the incremental network dual K function method. The results are shown in Figure 5. If the value of the Y-axis (observed upper p-value) is less than 0.05, i.e., within the red-colored part, the type of store is likely to attract other types of stores until the parametric distance. In the group of light meal/alcohol, for example, it seems that while Japanese bars attract others until 100 m; cafes attract others a little bit longer than the Japanese bars, i.e., until 160 m. It can be said from the line graphs that each store type has a different catchment zone. Note that the figure shows the median case; so, to deeply understand the attractive distance, it is necessary to look at the case of each pair.

ACKNOWLEDGMENTS

This study is supported by the Joint Research Program No.943 at the Center for Spatial Information Science (CSIS) at the University of Tokyo, and NTT TownPage Corporation.

Contact Info (Wataru Morioka)
Email: watarum2@illinois.edu **Twitter:** @ruwata_mori