Exact Statistical Method for Analyzing Co-location on a Street Network and Its Computational Implementation

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

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

INTRODUCTION

In many central districts in cities across the world, different types of stores form clusters based on 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 a novel GIS-based spatial statistical method by addressing the limitations of the ordinary cross K function method.

Objectives

- To formulate an exact statistical method for analyzing co-location and repulsive-location in a fairly small area with a street network.
- To implement the statistical method through a computational method using free GIS software.
- To demonstrate the usefulness of the methods through an empirical analysis.

METHOD

Statistical method

Comparing to the conventional cross K function, the main advantages of our method are:

- To assume a network-constrained space instead of a Euclidean space.
- By focusing on the nearest type i points relative to each type j point, the number of type j points within x from type i points follows the simple binomial distribution.

Formularization

A network, $N = (V, L)$,
a set of nodes $V = \{v_1, \ldots, v_{n_V}\}$; a set of links $L = \{l_1, \ldots, l_{n_L}\}$.

On $N = (V, L)$,
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.
$\tilde{L}(x|P_B)$: the total length of the buffer network of type B points of width x.
$K'_{A\to B(x)}$: the number of type A points on $\tilde{L}(x|P_B)$.

Under the Complete Spatial Randomness (CSR) hypothesis,

$$\Pr\left[K'_{A\to B(x)} = k\right] = {}_{n_{P_A}}C_k \left(\frac{\tilde{L}(x|P_B)}{\tilde{L}}\right)^k \left(1 - \left(\frac{\tilde{L}(x|P_B)}{\tilde{L}}\right)\right)^{n_{P_A}-k} \quad (1)$$

$$\Pr\left[K'_{B\to A(x)} = k\right] = {}_{n_{P_B}}C_k \left(\frac{\tilde{L}(x|P_A)}{\tilde{L}}\right)^k \left(1 - \left(\frac{\tilde{L}(x|P_A)}{\tilde{L}}\right)\right)^{n_{P_B}-k} \quad (2)$$

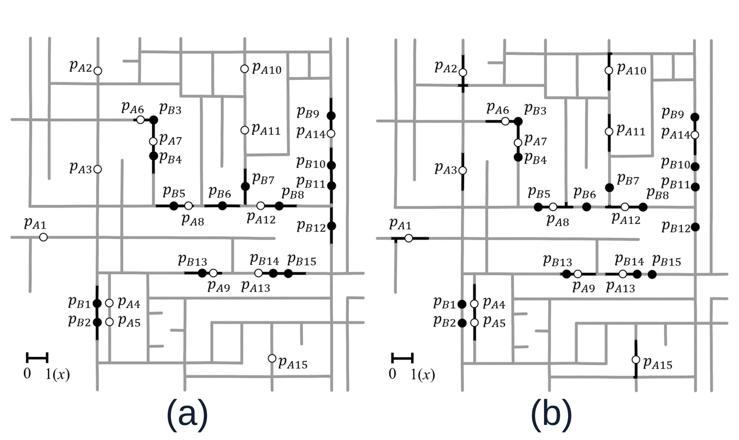


Figure 1. (a)The buffer network of $\tilde{L}(1|P_B)$. (b)The buffer network of $\tilde{L}(1|P_A)$. The bold lines indicate the buffer network of type B or A points with $x = 1$.

We depict the results of observed p-values with respect to parametric distance x, $\Pr\left[K'_{i\to j(x)} = k\right]$, for examining the degree of co-location more directly.

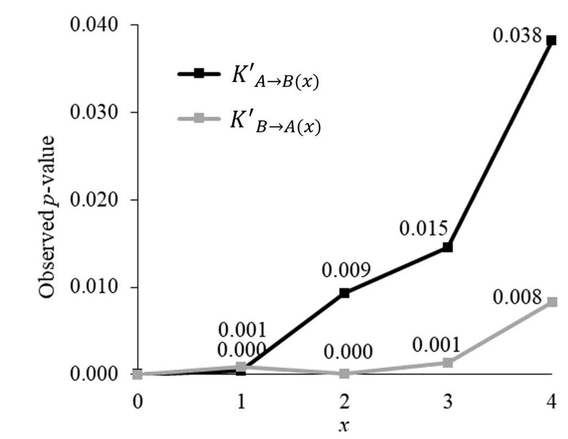


Figure 2. The observed p-values, $\Pr\left[K'_{A\to B(x)} \geq k\right]$ and $\Pr\left[K'_{B\to A(x)} \geq k\right]$ with respect to parameter distance in the case of Figure1. $\overline{p^+}_{A\to B(x)} = .016$; $\overline{p^+}_{B\to A(x)} = .003$. So, it is a bilateral co-location.

The average effect of co- and repulsive-location across a certain range is computed by the equations below.

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

$$\overline{p^+}_{B\to A(x)} = \frac{1}{m}\sum_{i=1}^{m} \Pr\left[K'_{B\to A(x_m)} \geq k\right] \quad (4)$$

$$\overline{p^-}_{A\to B(x)} = \frac{1}{m}\sum_{i=1}^{m} \Pr\left[K'_{A\to B(x_m)} \leq k\right] \quad (5)$$

$$\overline{p^-}_{B\to A(x)} = \frac{1}{m}\sum_{i=1}^{m} \Pr\left[K'_{B\to A(x_m)} \leq k\right] \quad (6)$$

The spatial relationships are classified into six groups:

a. **Bilateral co-location**: $\overline{p^+}_{A\to B(x)} < .05$ & $\overline{p^+}_{B\to A(x)} < .05$ (Fig. 3a)
b. **Unilateral co-location**: $\overline{p^+}_{A\to B(x)} < .05$ & $\overline{p^+}_{B\to A(x)} \geq .05$ & $\overline{p^-}_{B\to A(x)} \geq .05$. Or, $\overline{p^+}_{B\to A(x)} < .05$ & $\overline{p^+}_{A\to B(x)} \geq .05$ & $\overline{p^-}_{A\to B(x)} \geq .05$ (Fig. 3b)
c. **Bilateral repulsive-location**: $\overline{p^-}_{A\to B(x)} < .05$ & $\overline{p^-}_{B\to A(x)} < .05$ (Fig. 3c)
d. **Unilateral repulsive-location**: $\overline{p^-}_{A\to B(x)} < .05$ & $\overline{p^+}_{B\to A(x)} \geq .05$ & $\overline{p^-}_{B\to A(x)} \geq .05$. Or, $\overline{p^-}_{B\to A(x)} < .05$ & $\overline{p^+}_{A\to B(x)} \geq .05$ & $\overline{p^-}_{A\to B(x)} \geq .05$ (Fig. 3d)
e. **Co-/repulsive-location**: $\overline{p^+}_{A\to B(x)} < .05$ & $\overline{p^-}_{B\to A(x)} < .05$. Or, $\overline{p^+}_{B\to A(x)} < .05$ & $\overline{p^-}_{A\to B(x)} < .05$ (Fig. 3e)
f. **Neither co-locative nor repulsive-location**: $\overline{p^+}_{A\to B(x)} \geq .05$ & $\overline{p^-}_{A\to B(x)} \geq .05$ & $\overline{p^+}_{B\to A(x)} \geq .05$ & $\overline{p^-}_{B\to A(x)} \geq .05$ (Fig. 3f)

To deal with more than two types of stores, we use the signed digraph (Fig. 3) and the degree of co-location among n types is measured by *[the number of positive links]* / $n(n-1)$.

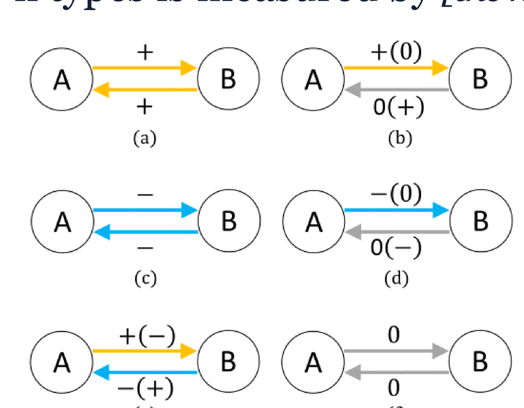


Figure 3. The signed digraphs representations of the six groups of spatial relationships. The alphabets correspond to the above classifications (e.g., a is bilateral co-location).

Geometric computational method

Two key operations are:

- To assign store points on a network (i.e., finding the nearest point on a network, called an access point).
- To generate a buffer network (i.e., solving the single-source shortest path problem).

Note that these steps can be performed by the [v.net.iso] & [NNJOIN] tools in QGIS with GRASS, or the [Voronoi diagram] & [Voronoi cross K function] tools in SANET.

EMPIRICAL ANALYSIS

Study area & Data

We target the Aoyama district, one of the trendiest districts in Tokyo (Fig. 4a). Store locations are obtained by TownPage telephone book (NTT TownPage Corporation) in July, 2019. We choose top eight types of stores as follows and set x ranging from 10 m to 200 m at intervals of 10 m.

Table 1. Top eight types of stores in the Aoyama district

Rank	Name	Rank	Name
1	Hair salon [458]	4	Dental clinic [141]
2	Real estate agency [361]	6	Aesthetic salon [129]
3	Japanese bar [164]	7	Advertising production [114]
4	Japanese restaurant [141]	8	Accounting firm [107]

Figure 4. (a) The study area, Aoyama district, formed by a 500 meters buffer (Euclidean distance) zone whose major streets are National Route 246 and Omote Sando (bold lines). (b) The buffer network of Japanese Restaurants with x = 200 meters.

Results

Table 2. Average co-location p-values (<.05 are light-yellow shade) and an average repulsive-location p-value (light-blue shade).

↓ is co-located to →	Hair salon	Real estate agency	Japanese bar	Japanese restaurant	Dental clinic	Aesthetic salon	Advertising production	Accounting firm
Hair salon	---	0.0017	0.0052	0.0000	0.0000	0.0000	0.0000	0.5198
Real estate agency	0.0000	---	0.0000	0.0000	0.0000	0.0003	0.0051	0.0000
Japanese bar	0.0004	0.0000	---	0.0000	0.0000	0.0000	0.9875 (0.0272)	0.4640
Japanese restaurant	0.0000	0.0000	0.0000	---	0.0000	0.0000	0.6638	0.3755
Dental clinic	0.0005	0.0000	0.0000	0.0000	---	0.0002	0.0766	0.0000
Aesthetic salon	0.0000	0.0026	0.0000	0.0001	0.0001	---	0.1108	0.1211
Advertising production	0.0797	0.0339	0.7280	0.5424	0.1588	0.1956	---	0.0002
Accounting firm	0.2267	0.0000	0.2400	0.1954	0.0000	0.5587	0.0002	---

- **Bilateral Co-location**
  19 pairs
- **Unilateral Co-location**
  1 pair (Hair → Ad)
- **Unilateral Repulsive-location**
  1 pair (Jpn bar→ Ad bar)
- **Degree of co-location**
  39 / 56 = 0.696

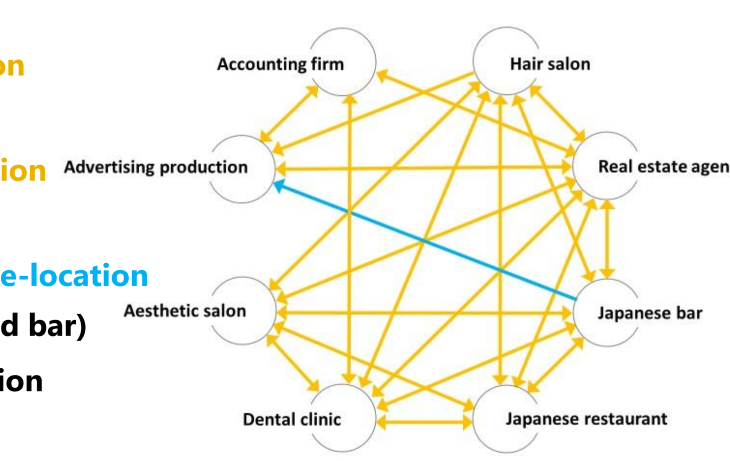
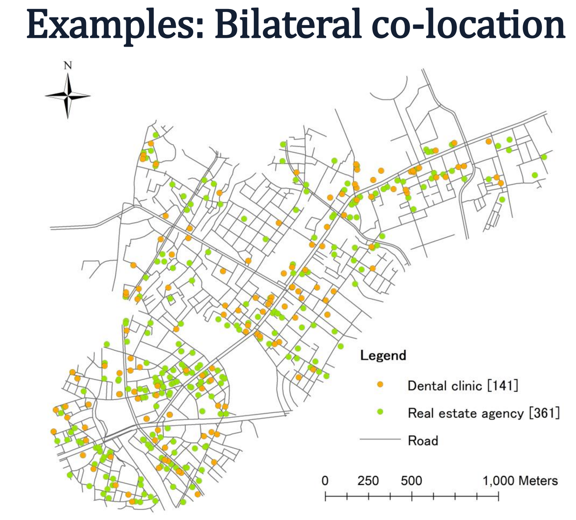


Figure 5. The graph-theoretic representation of spatial relationships among the eight types of stores. The bilateral relation is represented by a double-headed arrow. The graph has some subgraphs called cliques, in which all nodes are connected to each other by double-headed positive arrows, in this study. These combinations are likely to have a strong positive influence on urban agglomeration.

Examples: Bilateral co-location



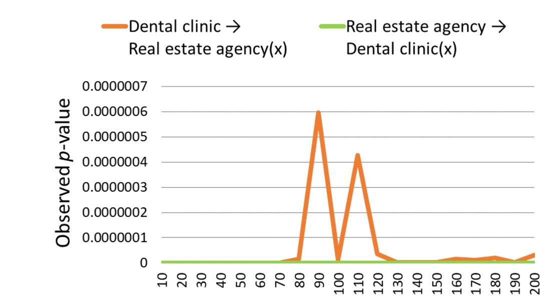
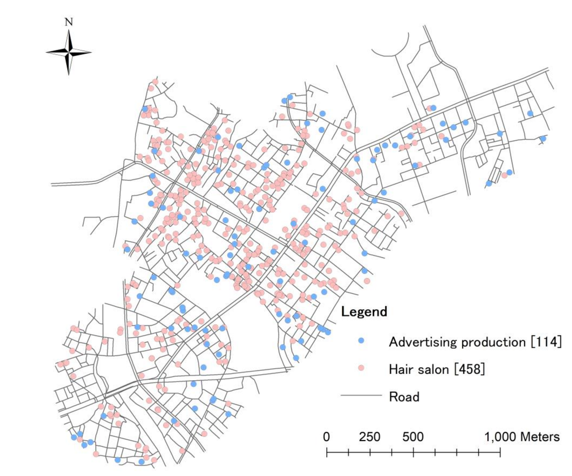


Figure 6. Spatial distributions & observed co-location p-values between real estate agencies and dental clinics.

Unilateral co-location



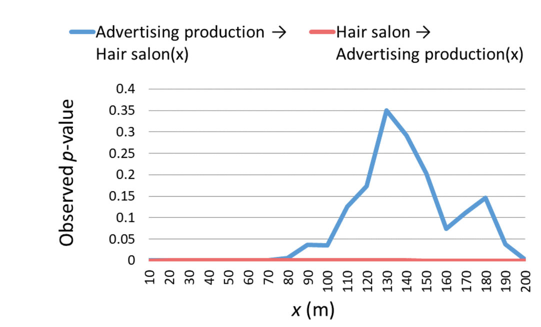
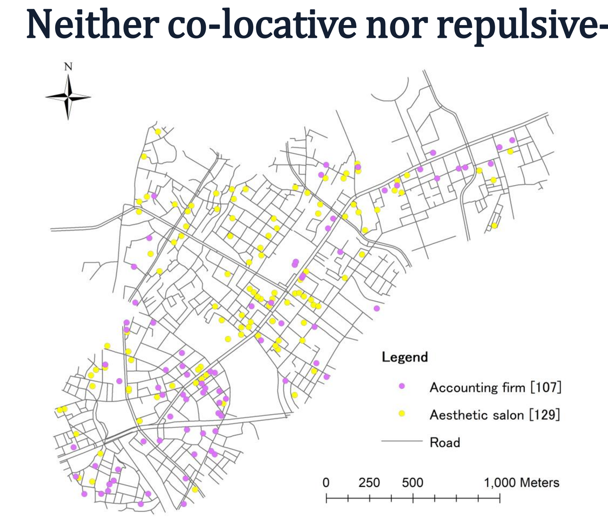


Figure 7. Spatial distributions & observed co-location p-values between hair salons and advertising productions.

Neither co-locative nor repulsive-location



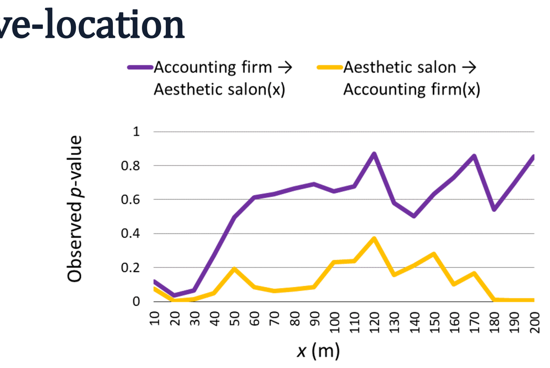
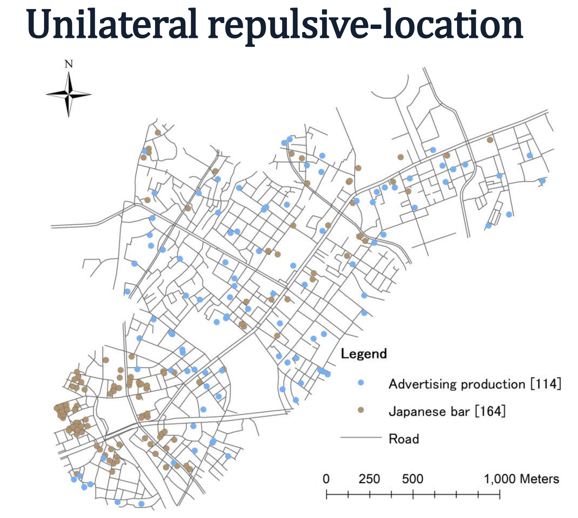


Figure 8. Spatial distributions & observed co-location p-values between aesthetic salons and accounting firms.

Unilateral repulsive-location



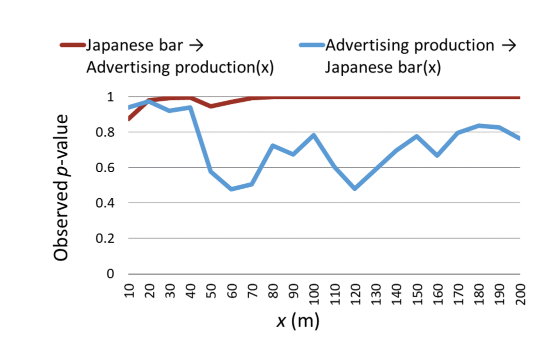


Figure 9. Spatial distributions & observed co-location p-values between Japanese bars and advertising productions.

CONCLUSION

We have developed a set of methods for analyzing the degree and types of co-location and repulsive-location among retail and service activities along street networks.

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.

Email: watarum2@illinois.edu **Twitter:** @ruwata_mori