Understanding Business Location Choice Pattern: A Co-Location Analysis on Urban POI Data

Jeffrey Chiu, Amin Vahedian Khezerlou, and Xun Zhou

The University of Iowa

{jeffrey-chiu, amin-vahediankhezerlou,xun-zhou}@uiowa.edu

Abstract

The co-localization of businesses has concerned researchers for a long time. With the advances in technology, for the first time we have access to accurate and up-to-date location information of businesses in form of public digital maps. This creates an opportunity to analyze the co-location patterns of the businesses with a data-driven approach to obtain an objective and realistic view of such patterns. In this study, we analyze the clustering tendencies and the co-location patterns of the businesses in the three largest cities of the United States. We obtain the dataset using the Google Maps Places API. We first obtain top co-locating patterns using co-location pattern mining techniques. Then we test the significance of the patterns using statistical tests and Monte-Carlo simulation. We find interesting co-location and clustering tendencies among brand names within and across industries as well as clustering tendencies between businesses of certain industries.

1 Introduction

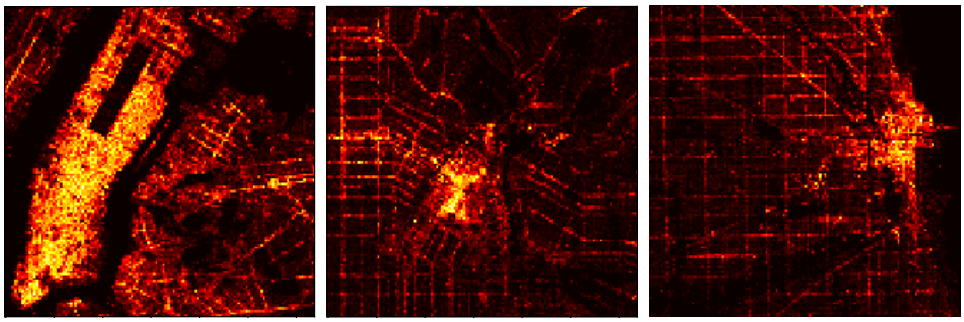
One of the most important aspects of a successful business is its location, whether they be clustered near competitors to compete for clients or far away from them in order to establish a customer base. Studying the location patterns of businesses relative to each other can reveal interesting insight to their relationship. Co-location pattern mining [7] is a set of techniques developed to discover such patterns.

In this study, we analyze four categories of co-location patterns: patterns among specific brand-names within each industry, specific brand-names across industries, co-location patterns of businesses from different industries, and clustering tendencies of the same industries. The discovered co-location patterns are tested for statistical significance using a Monte-Carlo simulation [11]. Through this study, we aim to discover previously unknown patterns that potentially influence the extent and nature of interactions between businesses as well as industries. For example, two brands offering the same service or product that choose locations far from each other for their stores are likely trying to avoid competition. While brands with significantly co-located stores are likely to be involved in a close competition. On the other hand, co-location of stores that offer different products will draw different conclusions. For example, co-location of a certain type of business with restaurants likely means that employees of such businesses visit restaurants mid-day.

Economic approach to business co-localization is theoretical. Meaning that the theories are formed to produce predictions of how the co-location patterns must be. A different approach to study the co-location patterns, is the data-driven approach, in which the objective reality of the existing co-locations is discovered and emphasized. Discovering such

co-location patterns in turn help us obtain previously unknown insight that will shape our understanding of the dynamics of interactions among businesses that serve urban dwellers. To the best of our knowledge, this is the first time business locations are studied in the context of co-location pattern mining for the largest metropolitan areas of the United States, that can help form theories that led to the existence of such patterns.

We use Google Maps Places API [6] to obtain our dataset of Point of Interests (POI). We build datasets for three large cities in the United States: New York City, Los Angeles and Chicago. We consider equally-sized regions of 12.8 km by 12.8 km for each city, fitting the entire downtown area. Figure 1 to Figure 3 show visualizations of the POIs on the map. The brighter areas are denser in terms of number of POIs. These areas include 183604 POIs in NYC, 86425 in LA and 85302 in Chicago. These numbers are consistent with the visualizations, as well as the metro populations of the three cities.



(a) New York City. (b) Los Angeles. (c) Chicago.

Figure 1: Heatmap of POIs in three studied metropolitan areas.

We used a measure developed in the context of Spatial Association Rule Mining, called participation index [7] to build a list of co-location candidates. We then analyzed the co-location patterns of those candidates using the Cross-K function and Ripley's K function [2]. A major challenge of this study is to determine whether discovered patterns are significant. Comparing the discovered patterns to a simple Poisson Complete Spatial Randomness (CSR) can produce invalid results, because all businesses are restricted to certain regions, therefore they are naturally clustered and co-located. However, the reason for this co-location is not their specific relationships, but as mentioned earlier, the reason is because they are restricted to locate only within commercial regions. Therefore, even if two businesses spread their locations independently of each other, they are co-located when considering the entire metropolis. To address this challenge, we designed a Monte-Carlo simulation based on shuffling existing locations. This way we are able to ensure the discovered patterns were statistically significant.

We find that all businesses in the urban area are extremely clustered. We found interesting co-locating brands within industries and across industries. We also found interesting de-clustering patterns among specific brands. Moreover, we analyzed the co-location patterns of POIs from different industries and found strong clustering tendencies between certain industries.

2 Related Work

In spatial statistics, the $K(d)$ function is used to describe the characteristics of point processes for given distances (d) [2]. Ripley's K function [13], is a version of the $K(d)$ function developed to measure the clustering tendency of points in space. The Cross-K function is

a generalization of the Ripley's K to two point processes [2], which measures the clustering tendencies of two types of events in space. These measures were developed without considering the computational cost of their application on large datasets. For instance, in the case of the current study, calculating the Cross-K function for all the pairs of POIs and all the pairs of location types and testing their significance using Monte-Carlo simulation is computationally infeasible, because there are hundreds of thousands of POIs and more than a hundred POI types.

Another approach to study spatial clustering is using spatial data mining techniques. Spatial data mining techniques aim at finding interesting but previously unknown patterns in spatial data [14, 15]. Spatial clustering techniques are part of spatial data mining and fall into two broad categories: clustering-based map overlay approaches and association rule-based approaches [7]. In the first approach, clusters are mined by placing the spatial attributes in map layers and producing candidates for spatial association [4, 3]. In association rule-based methods, the spatial co-locations are mined using versions of the Apriori Algorithm [1] adapted to spatial point processes [8, 12, 7]. The data mining techniques are computationally more efficient than the statistical measures discussed above. However, lack of significance testing in the such approaches is their main and important disadvantage.

In this paper, we use a data mining measure, Participation Index (PI) [7] in our analysis. However, due to above disadvantage - lack of statistical significance test - the discovered top measures are not necessarily significant patterns. We apply the statistical tests of the Ripley's K and Cross-K functions to the top co-location patterns identified by PI. This way we take advantage of the statistical measures of clustering to obtain valid results, while avoiding their prohibitive computational cost by using computer-efficient data mining approaches.

Economists have concerned themselves with the location patterns of businesses from very early days [9]. This issue has been studied throughout the 20^{th} century [10]. However, many of these studies were based on economic theories rather than observation. Some considered the benefit of clustering self-evident that argued the discussion was of little value [5]. Most explanations of the *cluster* were in the context of the balanced forces of two kinds: centripetal forces and centrifugal forces, the former pulling the businesses together and the latter pushing them apart [10]. In this paper, we take a different approach to understand such clusterings. We undertake a data-driven study of business locations in three largest metropolitan areas of the United States. The goal of this analysis is to produce an objective picture of the reality of the co-localization and clustering of businesses in the leading commercial settings in the world, which will provide valuable insight into formation of important economic theories.

3 Methods and Analysis Techniques

In this section we present the three components of our analysis. First, we present the co-location pattern mining technique that we use to identify top co-location patterns. Then we present the statistical tests that discover the significant co-location and clustering patterns. Figure 2 shows the flow of the analysis in this paper.

3.1 Co-Location Pattern Mining

Our first step is to identify top co-locating POIs. We use the concept of participation index of co-location patterns proposed in [7]. Participation index (PI) is a measure of how

frequent a co-location pattern is. For a set of POIs, the more frequently they are co-located in space, the higher their PI will be. First we define a neighbor relationship in space.

Definition 1 *Points l_i and l_j are neighbors if they are within a distance d from each other.*

The concept of neighbor relationship is fundamental to co-location patterns. Many different types of distance for the neighbor relationship can be defined, such as time distance or network distance, etc. In this paper we use distance on Earth. Based the neighborhood definition, we define the co-location pattern.

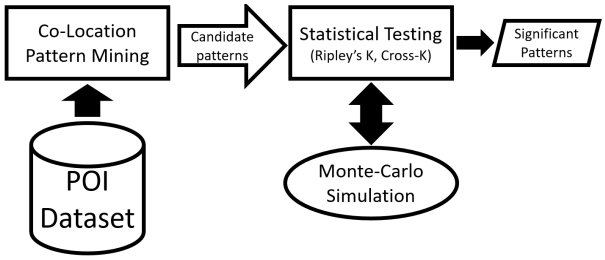


Figure 2: The flow of analyses in this paper.

Definition 2 *Set of point types $C = \{T^1, \ldots, T^k\}$ is a co-location pattern, if there is at least one instance of points $L = \{l_1, \ldots, l_k\}$ where l_i is type T^i and all members of L are neighbors. L is a* ***table instance*** *of C.*

Co-location patterns are interesting if there are many table instances of them. Moreover, they are interesting if instances of the members types are less frequent outside the pattern. In other words, if a POI type is always co-located with a set of other POI types, then its participation in the pattern is interesting. To quantify this concept, the participation ratio is defined.

Definition 3 *Given a co-location pattern C, participation ratio of type $T^i \in C$ is given by the following equation:*

$$pr(T^i; C) = \frac{|\text{Table Instances of } C|}{|\text{Table Instances of } T^i|} \quad (1)$$

This measure is defined for an individual type in the pattern. This measure is 1 if the instances of the type stay exclusive to the pattern. On the other hand, the measure is lower if the type has many instances located out of the pattern. The participation index is defined based on the participation ratio of the member types.

Definition 4 *The participation index of pattern C denoted as $pi(C)$ is defined as the minimum participation ratio of its member types. Formally:*

$$pi(C) = \min_{T^i \in C} \{pr(T^i; C)\} \quad (2)$$

The participation index is designed to only reward the participation of every member in the co-location pattern. If a co-location pattern has a high participation index, it means that the instances of the member types tend to co-locate with each other in space.

We use participation index of patterns of different size to find interesting co-location patterns. Given a set of point types, all the table instances, spatial neighborhood measure and a PI threshold, the algorithm proposed by Huang et al. [7] can efficiently find all the co-location patterns that have a PI above the threshold. However, despite computational efficiency, participation index tells us an incomplete story, i.e. we will still be able to discover a ranking of most interesting co-location patterns even if the businesses were randomly scattered in space. To address this issue, we use two statistical measures of clustering called Ripley's K function and the Cross-K function.

3.2 Statistical Tests of Co-Location and Clustering Patterns

Ripley's K function [2] is defined to measure deviation from spatial homogeneity. Specifically, we would like to test the following hypotheses:

H_0: The points in spaces are scattered randomly.

H_1: The points in space tend to cluster together.

Ripley's K function, defined as the expected number of neighbors for a randomly chosen point, is used as a test statistic of the above test. Considering Definition 1, if points are scattered randomly across the space, then the expected number of neighbors for a point with neighborhood radius of d should be equal to d^2 (under H_0). If points tend to cluster, the expected number of neighbors should be higher (H_1). A common way to estimate Ripley's K function is given by the following equation.

$$\widehat{K}(d) = \quad {}^{1}\sum_{i \neq j} \frac{I(d_{ij} \quad d)}{n} \tag{3}$$

Where n is the total number of points in space, d_{ij} is the distance between point i and j, I is an indicator function ($I(true) = 1$ and $I(false) = 0$) and is the density of points in space. In this paper, we use the above test to determine whether a set of businesses within an industry (e.g. gas stations) are clustered, de-clustered or randomly placed.

Although Ripley's K function gives us a measure of how clustered the POIs are in space, it does not take into account the type of POIs. For instance, Ripley's K function will not give us any information on how much restaurants and banks tend to co-locate. We use the following test to determine if two types of POIs are clustered together (i.e. co-located).

H_0: Locations of points type i and type j are independent.

H_1: Points type i and type j are clustered together.

The Cross-K function [2] is used as the test statistic of the above test. The Cross-K function of types i and j is the expected number of neighbors type j a type i point has. A common way to estimate the Cross-K function is given by the following equation:

$$\widehat{K_{ij}}(d) = \quad {}_{j}{}^{1}\sum_{i \neq j} \frac{I(d_{ij} \quad d)}{n_i} \tag{4}$$

Where $_j$ is the density of points type j, d_{ij} is the distance between point i and j and n_i is the total number of points type i. Under H_0, $K_{ij}(d)$ should be equal to d^2. Higher values mean that i and j tend to cluster and lower values mean that they tend to de-cluster, i.e. avoid each other. In this paper, we find that cliques of larger than size 2 are very rare in each of our study regions. Therefore, we use the above test to determine the significance of the co-locating pairs of POIs.

Although the value d^2 is commonly used to represent Complete Spatial Randomness (CSR) in the Ripley's K and Cross-K functions, it will produce misleading results in a restricted setting such as an urban area, such that we will find all the POIs to be extremely clustered. The reason for such an outcome is that the POIs that we are studying are restricted to commercial areas of the cities. In other words, their clustering together is not a result of a business decision, but a result of an external force. Therefore, we need to obtain a different CSR for our tests. We design Monte-Carlo simulations to obtain the CSR.

Table 1: List of top 5 categories in the dataset.

Category	Count
Store	54020
Health	49291
Food	29795
Doctor	25355
Restaurant	16836

3.3 Monte-Carlo Simulations

We perform Monte-Carlo simulations to determine whether specific Ripley's or Cross-K functions are statistically significant. As mentioned in Section 3.2, the conventional CSR value (d^2) is not a suitable baseline to test significance. This is because the businesses are not free to locate anywhere in the region, except some *allowed* regions. We call the set of such locations as the Location Domain. For example, when considering brand name "A" of type "T", the location domain is the set of all locations of type "T", i.e. brand "A" is only allowed to locate on the location domain of type "T". Thus, every business type has a location domain. Accordingly, we define the set of all the locations of the businesses in the dataset as the Global Location Domain.

To obtain the CSR values for the test statistics, we simply shuffle the businesses in their location domain and calculate the test statistic. For example, to test the significance of co-locations of brands "A" and "B" of type "T", we shuffle locations of POIs of both brands in the location domain of type "T" and calculate the resulting Cross-K function. We repeat this process 999 times and sort the resulting values. The 50th Cross-K value represents the Cross-K function with p-value$= 0.05$ for this test.

4 Results and Discussion

4.1 Pre-Processing and Settings

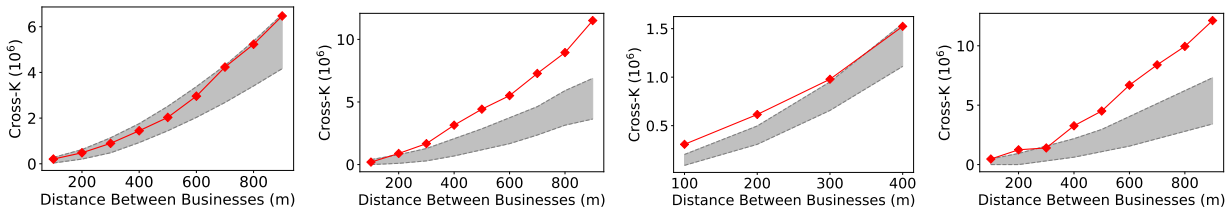
We obtain our dataset from Google Maps Places API [6]. This API returns a list of places that are registered on Google's popular and reliable map service. Each place corresponds to a POI. We focus on POIs that are associated with a business, e.g. a chain restaurant or branches of a bank. The dataset contains names, geographic coordinates and a list of categories the place belongs to. The API categorizes the registered POIs into 131 categories. Table 1 shows the list of top categories and how many times they appear, note that one POI may have multiple types.

4.2 Top Co-Location Patterns

First, we analyze all POIs and consider brand names as types and find the top co-location patterns of size 2 and more calculate the participation index for each pattern. We used PI threshold of 0.3 and distance threshold of 804 in Chicago and LA and 402 for NYC to determine the neighbor relationship. We found that patterns of size 3 and above are very rare across the three cities. Therefore, here we report top patterns of size 2. Table 2 shows the results of the top co-located brand names in the three cities. The list is consisted of POIs from the food and bank category.

Table 2: POI types with top participation index in the three cities.

New York City		Los Angeles		Chicago	
Brand 1	Brand 2	Brand 1	Brand 2	Brand 1	Brand 2
Citibank	TD Bank	Starbucks	Subway	Dunkin' Donuts	Subway
Chase Bank	Duane Reade	7Eleven	Subway	Jimmy Johns	Potbelly
CVS	TD Bank	Jack in the Box	McDonalds	7Eleven	Chase Bank
Chase Bank	Subway	7Eleven	McDonalds	Chipotle	UPS
Chipotle	HSBC Bank	76	McDonalds	7Eleven	Potbelly



(a) McDonald's and BaskinRobbins (b) Chipotle and Le Pain Quotidien (c) Dunkin' Donuts and Subway (d) Juice Press and Le Pain Quotidien

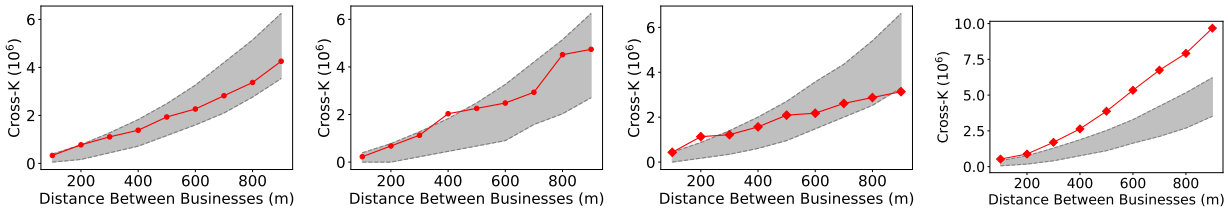
Figure 3: Cross-K functions of type *food* for New York City.

4.3 Intra-Industry Patterns

In this section, we study the significance of the co-location between the selected brands in two different industries. First, we analyze the brands in the food industry that appeared in the top participation index list and have the most locations in the area. Figure 3 shows the values of Cross-K function based on distance threshold for New York City. The shaded area shows the CSR region of the curve, with p *value* = 0.05. If the Cross-K function curve sits above this area, it means that the two POI types are clustered with p *value* of 0.05. If the curve falls below this region, it means the types are de-clustered. To obtain the CSR in this study, we used the locations of all the POIs of type *food* as the location domain and performed 999 Monte-Carlo simulations.

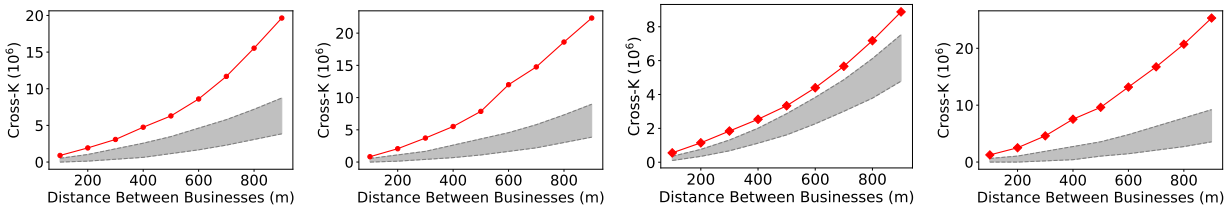
Figure 3 (a) shows the Cross-K function of McDonald's and BaskinRobbins. McDonald's is a fast food restaurant and BaskinRobbins is an ice cream shop. Although these two businesses had high participation index, this analysis shows that they are not significantly clustered. This is true for all distance thresholds. Figure 3 (b) shows the Cross-K function of Chipotle Mexican Grill (CMG) and Le Pain Quotidien (LPQ). CMG is a casual dining restaurant and LPQ is a bakery/restaurant. The Cross-K function is well above the CSR region, which means that CMG and LPQ are significantly clustered in space. Figure 3 (c) also shows significant clustering for a donut/coffee chain and a fast food chain. However, the significance is less for higher distance thresholds. Figure 3 (d) shows significant clustering for a snack shop and LPQ. Both businesses emphasize offering a variety of vegan options.

In Los Angeles, although 711 (a convenience store plus fast food) and Subway restaurant had a high participation index, their clustering is not significant as figure 4 (a) demonstrates. McDonald's has the same story in LA as in NYC, they are not clustered with any competitors (figure 4 (b)-(c)). Figure 4 (d) shows that Subway is again clustered with a coffee shop, this



(a) 711 and Subway Restaurants (b) Jack in the Box and McDonalds (c) McDonalds and Subway (d) Starbucks and Subway

Figure 4: Cross-K functions of type *food* for Los Angeles.



(a) 711 and Jimmy Johns (b) 711 and Potbelly Sandwich Shops (c) Dunkin' Donuts and Subway Restaurants (d) Jimmy Johns and Potbelly Sandwich Shops

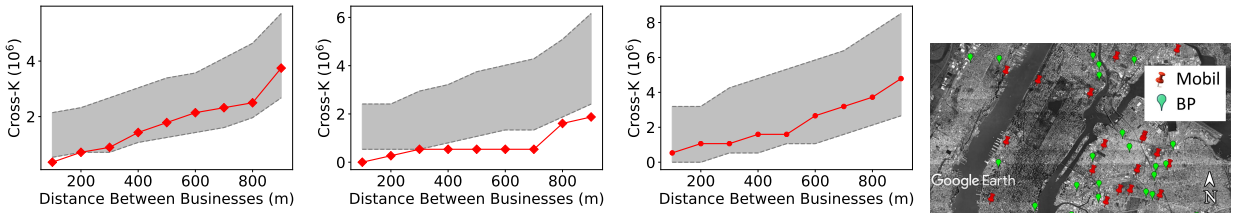
Figure 5: Cross-K functions of type *food* for Chicago.

time with Starbucks.

Figures 5 (a) and (b), unlike LA, show a significant clustering of 711 with two direct competitors Potbelly and Jimmy Johns in Chicago. Subway have the same story in Chicago as NYC and LA, while in Chicago they are again clustered with Dunkin' Donuts, similar to NYC (figure 4 (c)). Figure 4 (d) shows direct fast food competitors, Jimmy Johns and Potbelly are significantly clustered.

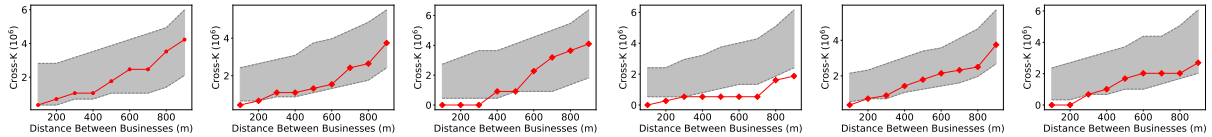
While clustering and co-location patterns can be interesting, the de-clustering patterns and places that avoid each other can be considered interesting patterns, too. Here, we present our results of analyzing the patterns of gas station locations that show some de-clustering tendencies. Figures 6, 7, and 8 show the Cross-K function for the pairs of the top 3 gas stations in the three cities. All figures show de-clustering patterns and some are significant. Figure 6 (d) shows the locations of BP and Mobil stations in NYC. It can be clearly seen that the stations are placed far from each other.

This observation suggests that gas stations tend to stay away from each other, trying to cover separated areas. This makes sense, considering that customers do not have a strong preference in buying gas for their vehicles and having any gas station serve an area is usually



(a) BP and Shell (b) BP and Mobil (c) Mobil and Shell (d) Mobil and BP on the map of NYC

Figure 6: Cross-K functions of type *gas* for New York City.

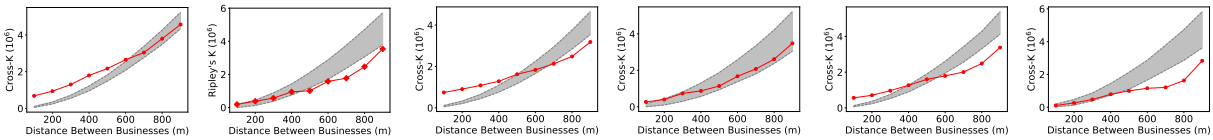


(a) 76 and Mobil (b) 76 and Shell (c) Mobil and Shell

Figure 7: Cross-K functions of type *gas* for Los Angeles.

(a) BP and Mobil (b) BP and Shell (c) Mobil and Shell

Figure 8: Cross-K functions of type *gas* for Chicago.



(a) All Gas (b) Major Gas

Figure 9: Ripley's K functions of gas stations for New York City.

(a) All Gas (b) Major Gas

Figure 10: Ripley's K functions of gas stations for Los Angeles.

(a) All Gas (b) Major Gas

Figure 11: Ripley's K functions of gas stations for Chicago.

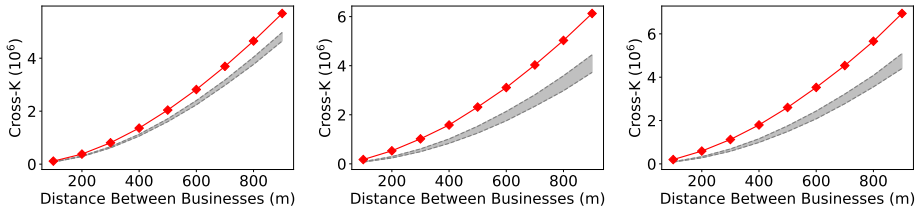
acceptable for customers, provided that the supply is sufficient. This is in contrast to the restaurant business, where customers have strong preferences and clustered restaurants are desirable in a sense that they provide options for the customer. To examine this theory, we studied the clustering tendency of gas stations as an industry. That is to say, we considered all the gas station locations as the location domain and tested the Ripley's K function. Figures 9, 10, 11 show the Ripley's K function for the gas stations in the three cities. Sub-figures (a) show clustering tendencies among all gas station brands and sub-figures (b) show clustering tendencies among the major brands. One can see that when considering all gas stations, there is no de-clustering tendencies. However, when considering only the major gas stations, we observe the same de-clustering trend. An important consideration about all the gas stations is that, most of them are not affiliated with major brands and only have one station. These gas stations seem to be locally owned and it is very unlikely that the owners had as wide option as the major brands when choosing their location. Therefore, we argue that their location in relation to other gas stations can not provide reliable insight. When disregarding the non-major gas station brands, we see the same trend as previous analysis.

4.4 Inter-Industry Co-Location Patterns

Next we examine the co-location and clustering pattern of two selected industries with most number of places in the dataset: *food* and *bank*. The *bank* category refers to all branch stores of a bank. Figure 12 shows that these two categories have strong clustering tendencies across all three cities. This finding is interesting because such a relationship between two industries seems non-trivial to guess and shows the value of data-driven approaches.

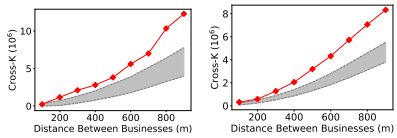
4.5 Cross-Industry Co-Location Patterns

Next, we study the co-location and clustering patterns of specific brands across the industries. Figures 13, 14, 15 show two pairs of brands from each city. Except McDonald's, all the other pairs show significant clustering and are across the two closely clustered categories of *food*



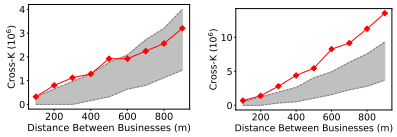
(a) New York City. (b) Los Angeles. (c) Chicago.

Figure 12: Cross-K functions of types *food* and *bank*.



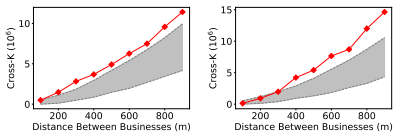
(a) Chipotle and HSBC Bank (b) Duane Reade and Chase Bank

Figure 13: Cross-K function across industries in New York City.



(a) McDonalds and 76 (b) Starbucks and Wells Fargo

Figure 14: Cross-K function across industries in Los Angeles.



(a) Chase Bank and Jimmy Johns (b) Chase Bank and Potbelly

Figure 15: Cross-K function across industries in Chicago.

and *bank*. The results of this section are consistent with the results from Section 4.4 that showed businesses of these categories to be co-located and significantly clustered.

5 Conclusions

In this paper, we took advantage of the advances in technology that for the first time gave us access to accurate and up-to-date location information on businesses in form of public digital maps to analyze the co-location patterns of the businesses with a data-driven approach to obtain an objective and realistic view of such patterns as opposed to the theoretical approaches used by the economists for a long time [10]. In this study, we analyzed the clustering tendencies and the co-location patterns of the businesses in the three largest cities of the United States. We obtained the dataset using the Google Maps Places API [6]. We first obtained top co-locating patterns using co-location pattern mining techniques. Then we tested the significance of the patterns using Statistical tests and Monte-Carlo simulation. We found interesting co-location and clustering tendencies among brand names within and across industries as well as clustering tendencies between businesses of certain industries.

This study is limited by the accuracy and completeness of the dataset used. Moreover, the real location domain of the POI types is not available. We use all the locations of instances of a specific POI type as its location domain. This means that we have assumed that the entire location domain of a specific type is covered by its instances. This can potentially lead to over-estimation of the significance of the discovered patterns. In the future, we will use real data of the location domains, i.e. official zoning data that shows where businesses are allowed to operate. Moreover, in the future, we plan to develop a unified framework to precisely identify candidate patterns by co-location pattern mining based on clear and specific criteria to ensure the discovery of every important pattern.

References

[1] Agrawal, R., Srikant, R., et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (1994), vol. 1215, pp. 487–499.

[2] Dixon, P. M. Ripley's k function. *Wiley StatsRef: Statistics Reference Online* (2014).

[3] Estivill-Castro, V., and Lee, I. Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data. In *Proc. of the 6th International Conference on Geocomputation* (2001), pp. 24–26.

[4] Estivill-Castrol, V., and Murray, A. T. Discovering associations in spatial data - an e□cient medoid based approach. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (1998), Springer, pp. 110–121.

[5] Feser, E. J. Old and new theories of industry clusters. *Clusters and regional specialisation 16* (1998).

[6] Google. Overview: Google Places API . `https://developers.google.com/places/web-service/intro`, 2018.

[7] Huang, Y., Shekhar, S., and Xiong, H. Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and Data Engineering 16*, 12 (2004), 1472–1485.

[8] Koperski, K., and Han, J. Discovery of spatial association rules in geographic information databases. In *International Symposium on Spatial Databases* (1995), Springer, pp. 47–66.

[9] Marshall, A. *Principles of economics. Vol. 1.* Macmillan And Co., Limited; London, 1898.

[10] Maskell, P. Towards a knowledge-based theory of the geographical cluster. *Industrial and corporate change 10*, 4 (2001), 921–943.

[11] Metropolis, N., and Ulam, S. The monte carlo method. *Journal of the American Statistical Association 44*, 247 (1949), 335–341.

[12] Morimoto, Y. Mining frequent neighboring class sets in spatial databases. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (2001), ACM, pp. 353–358.

[13] Ripley, B. D. The second-order analysis of stationary point processes. *Journal of applied probability 13*, 2 (1976), 255–266.

[14] Shekhar, S., Evans, M. R., Kang, J. M., and Mohan, P. Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1*, 3 (2011), 193–214.

[15] Shekhar, S., Jiang, Z., Ali, R. Y., Eftelioglu, E., Tang, X., Gunturi, V., and Zhou, X. Spatiotemporal data mining: a computational perspective. *ISPRS International Journal of Geo-Information 4*, 4 (2015), 2306–2338.