Diversity of diversities: a comparative study on the patterns of neighborhood diversities between Toronto and Manhattan, New York City

Weiyu Ran

May 19, 2020

1 Introduction

1.1 Spatial patterns of diversity in the major cities

It has been widely known that the ever-growing degree of economic interdependence between different parts of the world has brought the scale of personnel exchange to a degree that has never been seen. Such a new phenomenon is extremely observable in the major cities, which enjoy their advantages as international hubs for transportation, economy, and culture. This process brings new immigrants to these cities and makes their neighborhoods more diversified, which could be easily traced even in the briefest demographic reports.

At the same time, it is not as easy to investigate more complicated patterns beneath the simple proportional data on various ethnic groups. Among the numerous aspects related to the characters of diversity, the spatial pattern of diversity in the major cities is an extremely important one, since it indicates the potential relationship between the behaviors of people of different ethnic groups and the urban environment as a landscape. Hence, a more comprehensive understanding on the spatial pattern of diversity would not only help a company make a wise short-term commercial decision, such as “How to choose a location of a shop to involve more potential customers”, but also develop an insightful, long-term strategy based on the tendency of interest transition implied by the spatial data on the people with different cultural backgrounds. This is the focus of this study.

1.2 The changing tendency of urban spatial patterns

The traditional pattern of cities in North America is featured with a downtown filled with skyscrapers and only a few residential buildings. This core area was surrounded by suburban neighborhoods, and among them, there are neighborhoods as residents for different ethnic groups, such as “Chinatown”, ”Little Italy” and “Little Tokyo”. A classic example is the city of Chicago, in which the central business buildings and public structures are clustered in the downtown area called “the Loop”, as well as “the Magnificent Mile” in the north and “the Museum Campus” in the south. While the suburban residential neighborhoods, including the “Chinatown” for Chinese immigrants and Pilsen for Mexican immigrants, located around the business center. Figure 1 illustrates these patterns clearly.

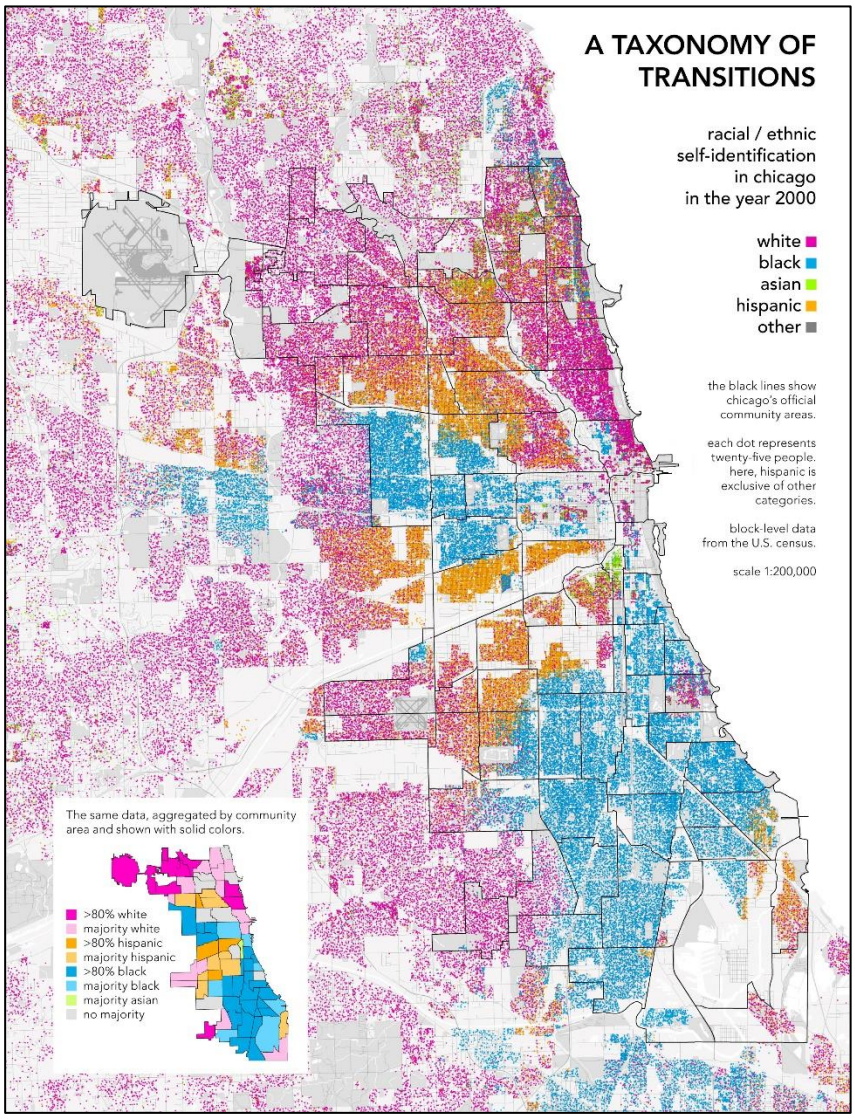

Figure 1. Racial/ethnic self-identification in Chicago in the year 2000.
(Source: http://www.radicalcartography.net/index.html?chicagodots)

In recent years, this old picture of the city landscape has been reshaped by the force of

urbanization. In Chicago, two tendencies can be traced within the map of 2010 (Figure 2): first, the boundaries between different groups are less clear than those of 2000, with many neighborhoods more diversified with more residents from different ethnic groups. Another tendency is the residents grew rapidly in the central business center, especially in the Loop and Near North neighborhoods. It is important to mention that these neighborhoods are naturally more diversified, as clearly illustrated in Figure 2. Both tendencies show clear signals indicating a new spatial pattern of diversity in the city of Chicago, with different ethnic groups do not tend to live in their "traditional" neighborhoods and form homogeneous communities. Instead, more neighborhoods start to become diversified ones whose residents are from different cultural backgrounds. With a similar overall degree of diversity, the spatial pattern of diversity has changed in Chicago.

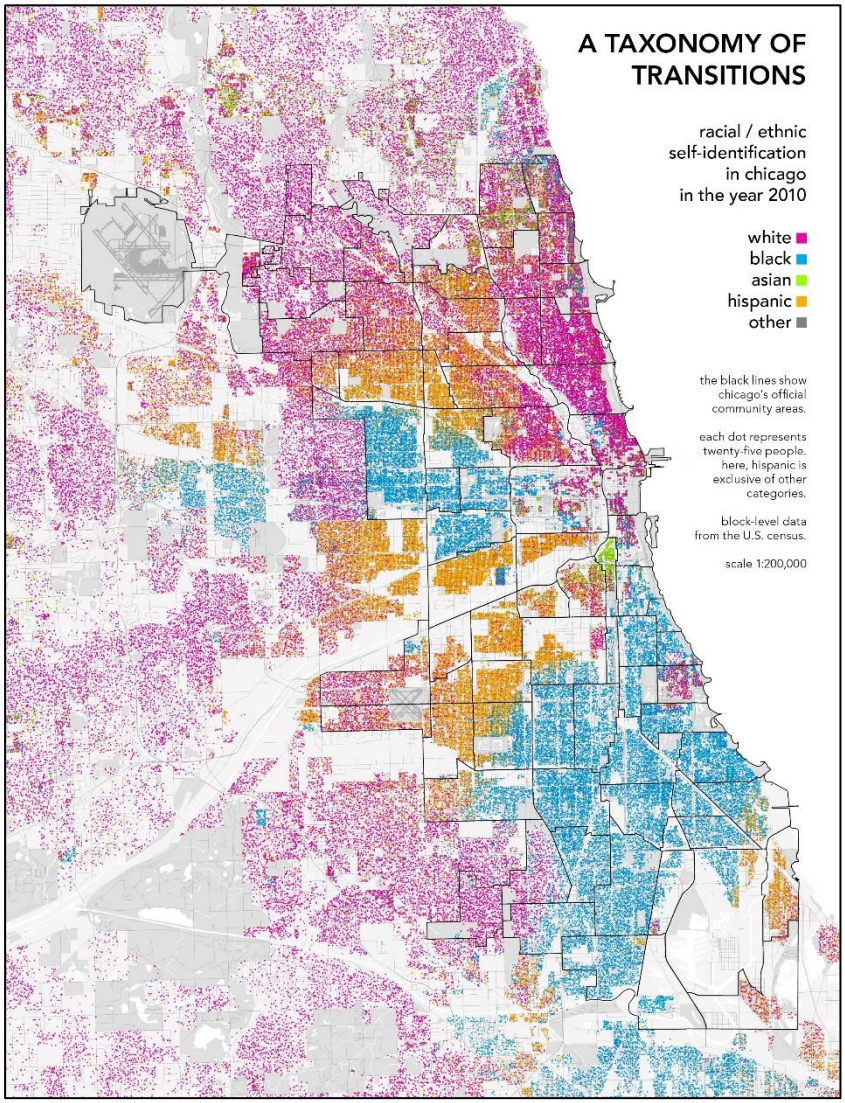

Figure 2. Racial/ethnic self-identification in Chicago in the year 2010

(Source: http://www.radicalcartography.net/index.html?chicagodots)

But how about the other major cities in North America?

This study was designed to find out the spatial patterns of diversity for two other North American cities, namely New York City and Toronto. Same to the other cities in the US and Canada, both cities have relatively short histories. In addition, they are the largest cities, as well as the financial centers and economic engines in their own country, and they are both international metropolitan and are homes of the people with wildly different races, religions, and cultures. Even though, these two cities have a different size in terms of population, and many differences in cultures and laws bring dissimilarities between them. Hence, these two research objects provide chances to understand the spatial patterns of diversity in a more comprehensive way. The questions this study seeks to answer are:

1. Are mixed spatial patterns on the neighborhood level similar to that of Chicago traceable in both cities?
2. What are the similarities and dissimilarities between the spatial patterns of diversity of the two cities?

The answers to these two questions will help us understand how did and will demographic diversity influence the development of the city, and they also have critical potential values in helping decision-makers to adjust business strategies in order to gain benefits of the changing tendency.

2 Data and data collection

With the research questions in mind, it is then important to discuss the appropriate type of data to answer them and methodologies of data collection.

2.1 Data types

There are abundant data sources that cover the topic of urban demography, among them the most important one is the census. These sources, however, have two disadvantages:

first, they do not provide up-to-date data, which weakens the ability of the results to reflect the latest tendency and predict the future direction. Second, the data collection of these sources usually focused on the race of people which doesn't necessarily reflect the cultural influence on the urban residents and their behavior under this influence, which is of critical importance to the change of diversity pattern. Hence, to answer the research questions listed above, it is important to collect recently updated data that reflect the behavior of people of different cultural backgrounds.

With the requirement of data in mind, it is suggested that applying restaurants as indicators of activities of the daily life of residents is a satisfactory and applicable choice. Since the cultural influence would affect the choice and tastes of food, a presumption of this study is that with all other being equal, the restaurants with styles related to certain cultures tend to locate in the area where corresponding cultural group activates more. Hence, the restaurants could be used as indicators of social activities, and more restaurants of certain styles existed in one area means related groups conduct more daily activities in the same area.

Since both Toronto and New York Cities are home for people belongs to too many ethnic groups, it is important to choose the typical cultural genes of restaurants to reflect the pattern of diversity as persuasive as possible. The demographic data of the two cities have revealed important ethnic groups live in them, which provide clues for choosing meaningful genes for the restaurants[1][2]. Based on them, the total number of 16 restaurant genes were selected to evaluate the cultural diversity and listed in Table 1. In the practice, the original data were collected through the Foursquare API, and category IDs for all the genes were also presented in Table 1. It is worthy to mention that although most genes were delineated based on countries, considering some genes have no category ID in the list provided by Foursquare, and some restaurants were originally classified as more general genes (such as

[1] https://en.wikipedia.org/wiki/Demographics_of_New_York_City#cite_note-nyc_popacs-34

[2] https://www12.statcan.gc.ca/census-recensement/2006/dp-pd/prof/92-591/details/page.cfm?Lang=E&Geo1=CSD&Code1=3520005&Geo2=PR&Code2=35&Data=Count&SearchText=Toronto&SearchType=Begins&SearchPR=01&B1=All&Custom=

"Caribbean Food"). Some genes cover regional food styles were included in the list instead of listing genes of their belonging counties separately.

Food Gene	Categorical ID
English	52e81612bcbc57f1066b7a05
Caribbean	4bf58dd8d48988d144941735
Chinese	4bf58dd8d48988d145941735
Filipino	4eb1bd1c3b7b55596b4a748f
French	4bf58dd8d48988d10c941735
German	4bf58dd8d48988d10d941735
Indian	4bf58dd8d48988d10f941735
Irish	52e81612bcbc57f1066b7a06
Italian	4bf58dd8d48988d110941735
Jewish	52e81612bcbc57f1066b79fd
Korean	4bf58dd8d48988d113941735
Mexican	4bf58dd8d48988d1c1941735
Polish	52e81612bcbc57f1066b7a04
Portuguese	4def73e84765ae376e57713a
Russian	5293a7563cf9994f4e043a44
South American	4bf58dd8d48988d1cd941735

Table 1. The restaurant genes involved in data analysis and related Foursquare categorical IDs

2.2 Data collection

To reflect the distribution of restaurants in different parts of New York City and Toronto with a satisfying resolution an appropriate spatial unit should be selected and applied as the basic research unit to breakdown the area of the two cities. Although "neighborhood" is a commonly used term to name and define a certain subpart in both of the two cities, applying it as a basic spatial unit for this analysis has problems, since neighborhoods are sometimes not stable administrative units, and it is especially the case for many neighborhoods in Toronto. Without clear definitions and boundaries of neighborhoods, it is hard to make accurate statistics for the basic spatial units and further affect the credibility of the research. Alternatively, the postal zones widely used in the two cities as well as others. Since postal zones are usually stable after being delineated, and cover the whole urban area with clear boundaries between each other, applying postal zones with unique postal codes for each of them would provide accurate statistical data without any omissions or duplications. Furthermore, the areas and populations postal zones in the two cities cover are

comparable and are appropriate for the proposes of this research. Hence, the postal zones were used as basic spatial units for both cities, and since the as a whole New York City is much larger than the Toronto, here only the Manhattan was included in the research to stand for the whole city. Eventually, 47 postal zones in Manhattan and 97 postal zones in Toronto are involved in the data analysis.

Due to the limitation of the free Foursquare account, the maximum results returned to any query call is only 50. Considering the heavy density of restaurants in both cities, it is necessary to make calls for each gene within each postal zone to collect complete data without serious omission. Practically, the coordinates (latitudes and longitudes) of central points of all postal zones in both cities were collected. Then calls were made to return restaurants of each gene within 1000 meters surrounding the central point of each postal zone. Since this type of queries might cover restaurants outside the targeted postal zone, further verifications were made to check the postal codes of each resulting restaurant. Finally, the count of restaurants each gene for each postal code was calculated and stored in a dataframe for data analyses.

3 Methodology for data analyses

The data analysis of this study consists of two phases:

The first phase was to illustrate the spatial distribution of data and gain a direct understanding of the distribution of restaurants with each gene. Specifically, one map for each city was drawn, with each postal zone marked with indexed color to indicate the restaurant gene with the most count. These two maps are helpful to roughly reveal the distribution patterns of different genes of restaurants in the two cities.

The second step was mainly about labeling the postal zones of the two cities. Since the research task requires unsupervised learning, the hierarchical clustering approach was applied for the analysis. This approach presents the whole process of clustering and does not require a preset number of clusters, which are advantages that match the needs of the

data analysis. Before conducting the predictive modeling, data preparation was made to provide data with the appropriate format. For that the focus of this research is on the proportional constitution of different restaurant genes for each of the postal zones, it is more reasonable to calculate the proportion of each gene to the total number of each row (postal zone), rather than applying the usual approach of standardizing the data by each column (variable). The details about the data preparation were presented in the attached notebook file.

The modified data were then used for the hierarchical clustering. As the hierarchical clustering approach doesn't require a preset number of clusters, it is important during the analysis to observe the resulting dendrogram of the analysis and try different numbers of clusters to find the optimal one.

Finally, two new maps were drawn with indexed colors marking the clusters for all postal zones of the two cities. These maps were used to illustrate the scenario of diversity patterns of the two cities and answer the research questions.

4 Results and Discussions

4.1 Results on illustrative maps

Results of the illustrative maps present the dominant restaurant gene in each of the postal zones of the two cities, which provides an insightful picture of the activity hot spots of each ethnic group. It is easy to tell the difference between the distributional patterns between the two maps (**notebook out 17 & 18**). For the map of Toronto, clusters of postal zones with the same dominant restaurant gene can be observed in the city of Toronto, especially the Chinese one in the northeastern part and the Indian one in the eastern part of the city. Also, some other genes, although they don't form clear clusters, each has its area of relatively dominant distribution, such as South American in the western and Mexican in the southwestern part of the city. Comparable clusters were almost completely absent in Manhattan, with only one exception, a cluster in the north end of Manhattan dominated by

Caribbean restaurants. Comparing the two maps, it seems that Manhattan has a more well-mixed spatial diversity pattern on the neighborhood level than that of Toronto, and this hypothesis will be tested in the following hierarchical clustering analysis.

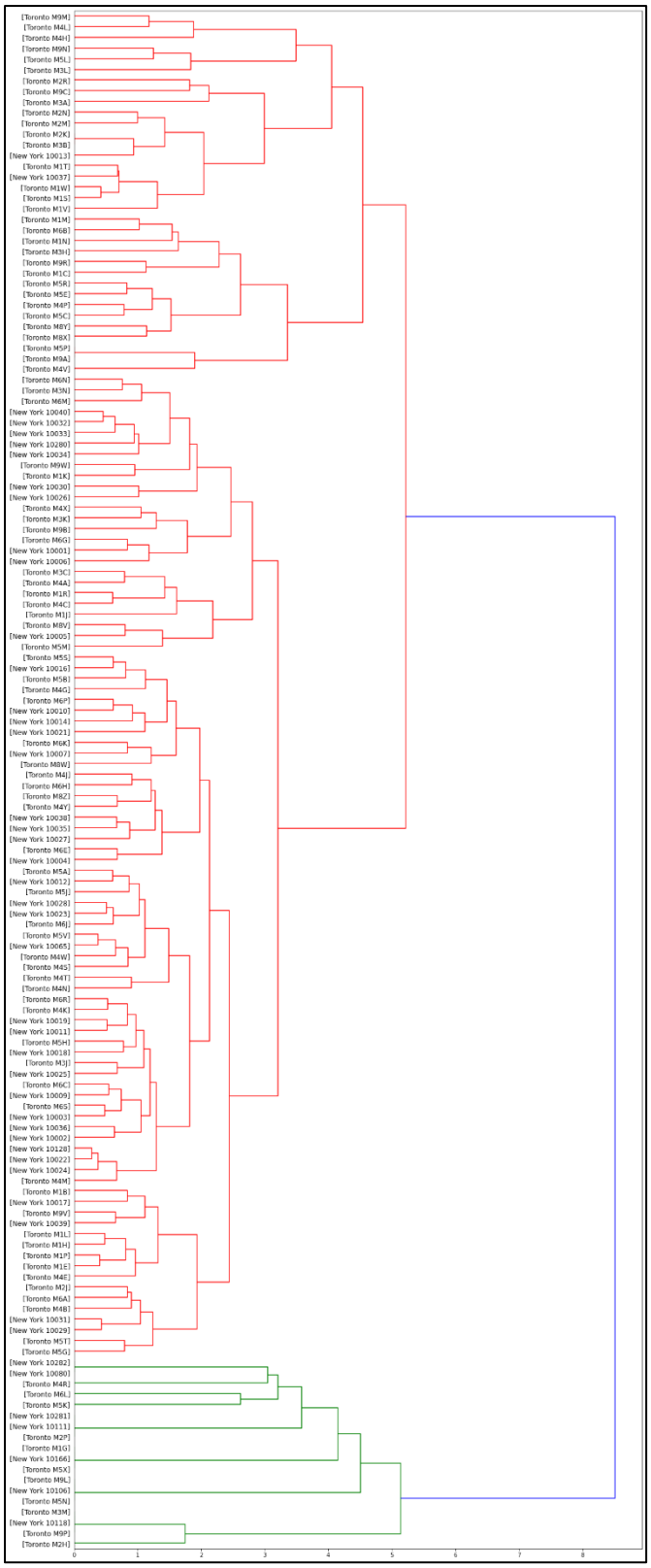

Figure 3. The dendrogram of the hierarchical clustering analysis

4.2 Results on hierarchical clustering analysis

Selecting the optimal number of clusters is critical to the hierarchical clustering analysis since it decided whether the analysis could present a clear and meaningful pattern. A trade-off must be made during the process between the details of each gene and overall tendency. To complete this mission, a pilot analysis was made to calculate the Euclidean distance between each pair of cases (postal zones) and conduct a hierarchical clustering analysis with the approach of complete linkage. The resulting dendrogram (Figure 3) provides details about the relative distances between the clusters that help choose the right number of clusters. The dendrogram indicates a range between 5-12 may be the optimal number of clusters. After trying different numbers, it turned out 10 clusters could reveal enough details without making clusters too shuttered, and was hence applied.

Cluster	English	Caribbean	Chinese	Filipino	French	German	Indian	Irish	Italian	Jewish	Korean	Mexican	Polish	Portuguese	Russian	South American
0	0.44%	11.84%	10.92%	4.43%	8.93%	0.97%	14.42%	1.83%	19.74%	1.77%	3.02%	11.87%	0.42%	2.80%	0.24%	6.36%
1	0.15%	13.69%	22.34%	1.87%	2.15%	1.59%	12.18%	1.38%	9.55%	2.30%	9.71%	15.74%	0.48%	0.87%	0.19%	5.80%
2	0.00%	0.00%	10.83%	0.00%	5.00%	0.00%	1.25%	0.00%	69.17%	0.00%	10.00%	0.00%	0.00%	0.00%	0.00%	3.75%
3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
4	0.00%	10.31%	0.42%	0.00%	0.42%	0.00%	73.33%	0.00%	4.48%	0.00%	4.17%	3.96%	0.00%	0.42%	0.00%	2.50%
5	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
6	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
8	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
9	0.00%	4.06%	73.32%	0.86%	0.00%	0.00%	7.15%	0.00%	4.58%	0.00%	5.75%	3.25%	0.00%	0.00%	0.00%	1.03%

Table 2. The proportions of 16 restaurants in each of the clusters.

To figure out what kinds of restaurants are included in each cluster, the average proportion of each gene was calculated for each cluster, and the result was presented in Table 2. From this table, it is clear that postal zones belong to clusters 0 and 1 represent have constitutions with relatively even portions of different restaurant genes, and none of the genes has a share higher than 25%. These clusters stand for those postal zones with well-mixed daily activities of different ethnic groups. There is still a difference between clusters 0 and 1, for that in cluster 0 the restaurant gene with the highest share is Italian, while postal zones labeled as cluster 1 have relatively higher proportions of Chinese restaurants. This

reveals there are still slightly different focuses between these well-mixed postal zones. In contrast, each of the other 8 clusters has clear stigmas of one or two specific restaurant gene. They are signals of the postal zones that are activity hot spots for certain ethnic groups.

Cluster	Manhattan	Toronto
0	22	40
1	16	30
2	1	9
3	2	0
4	2	6
5	0	1
6	2	1
7	0	1
8	0	1
9	2	8
Total	47	97

Table 3. Counts of postal zones of each cluster in Manhattan and Toronto.

Table 3 presents the counts of postal zones labeled for each cluster in both Manhattan and Toronto. It is clear to see in both cities most of the postal zones are of cluster 0 and 1, which suggests the well-mixed postal zones have gained dominant advantages in both cities and supports the hypothesis above. Comparing the data of the two cities, Manhattan has a higher proportion of cluster 0 and 1 postal zones than Toronto, which also matches the expectation made above that New York City has a spatial pattern of diversity that is more mixed on the neighborhood level.

Two maps were then drawn to illustrate the distributions of the postal zones belong to different clusters in each city (**notebook out 28 & 29**). It further confirms the results observed above and reveals the details of the two cities. Specifically, in Manhattan, most of the postal zones are clusters 0 and 1, and the only postal zones belong to other clusters were sparsely distributed in different parts of the city, which further proves the expectation that Manhattan has a more mixed pattern on the postal zone level.

Although postal zones in the downtown and adjacent parts of Toronto have similar patterns as those of Manhattan, in the suburban area the concentrations of postal zones belong to

other clusters can still be observed, such as Chinese one in the northeastern part and Italian one in the western. Clearly, Toronto does not have a pattern as New York City. Rather, it has a pattern with a well-mixed core area, but clusters of ethnic groups can still be found in the suburban area. It is more like a mixture between the traditional spatial pattern and the one of Manhattan.

5 Conclusion

By conducting this comparative analysis, this study provides answers to the two research questions. It is argued here that the well-mixed spatial pattern of diversity can be observed in both New York City and Toronto, which is consistent with tendencies observed in Chicago. In addition, in the core areas of the two cities host the most business and public activities we have seen a very well-mixed pattern indicated by the proportion of different restaurant genes roughly on the same level. The main difference between the two cities exists in the non-core areas: some postal zones or even clusters of postal zones with a single dominant restaurant gene could be seen in the suburban area of Toronto, while the same type of postal zones in Manhattan is almost as well-mixed as those in the core areas. A potential explanation of this difference is that Manhattan is an area with an extremely high degree of urbanization, which makes daily interaction in almost all neighborhood denser and increase the personnel mobility between the neighbor and stimulate the mixing of different cultures. The suburban areas of Toronto, on the other hand, have a significantly lower degree of urbanization compared to its downtown area and most neighborhoods of Manhattan, and the relatively lower population density and less public transportation also lead to lower personnel mobility, which keeps the shape of the already established ethnic communities.

The conclusion of this study provides insightful information for business decision-makers. For example, an owner of a restaurant featured with fusion cuisine may find a recently mixed neighborhood would be an ideal location for the business. And the rule implied by the conclusion of this study between the diversity of neighborhoods and their degree of

urbanization would be valuable to the long-term plan for the decision-maker to grasp the opportunity provided by the next wave of urbanization.