

An Approach for Measuring Spatial Similarity Among COVID-19 Epicenters

Neda Kaffash Charandabi (✉ neda.kaffash@gmail.com)

University of Tabriz <https://orcid.org/0000-0002-6281-9009>

Abolghasem Sadeghi-Niaraki

Sejong University

Soo-Mi Choi

Sejong University

Research

Keywords: COVID-19, GIS, Spatial Similarity, Environmental factors, Demographic factors

Posted Date: August 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-49216/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An approach for measuring spatial similarity among COVID-19 epicenters

Neda Kaffash Charandabi^{1*}, Abolghasem Sadeghi-Niaraki^{2,3†}, Soo-Mi Choi³

1. Geomatic Department, Marand Technical Faculty, University of Tabriz, Tabriz, Iran, n_kaffash@tabrizu.ac.ir
2. Geoinformation Tech. Center of Excellence, Faculty of Geodesy & Geomatics Engineering, K.N.Toosi University of Technology, Tehran, Iran, a.sadeghi.ni@gmail.com
3. Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea, smchoi@sejong.ac.kr

* Correspondence: neda.kaffash@gmail.com

† These authors contributed equally to this work.

Abstract

The world has been embroiled in a new epidemic known as COVID-19 since the beginning of 2020. Most countries and territories around the world are affected by the disease, and some cities have become known as epicenters due to high outbreak. The similarity of these cities can be examined within the Geographic Information System (GIS), based on various criteria. This study investigated the similarities between the eight cities of Wuhan, Tehran, Bergamo, Madrid, Paris, Daegu, New York, and Berlin in terms of the COVID-19 situation (target) in those locations based on socio-economic factors, weather, and demographic criteria. First, the factor and target layers were prepared in ArcGIS[®]10.6 software. For socio-economic data (such as: supermarkets, hospitals, metro stations etc.), the distance maps were classified with a fuzzy membership function. Weather and demographic criteria were also stored in the tables after normalization in the range of zero to one. In next step, the similar cities were identified using the similarity model and different distance metrics (Manhattan, Euclidean, Minkowski, Mahalanobis, Chebyshev, and Correlation). The results were aggregated using the Copeland method, due to the different outcomes of each metric. The most similar city was identified for each selected city and its similarity level was determined based on the criteria. Results indicated that pairs of similar cities are: Wuhan-Berlin, Tehran-Berlin, Daegu-Wuhan, Bergamo-Madrid, Paris-Mardid, and New York-Paris with a minimum and maximum similarity rate of 82.85% and 92.36%. For similar cities, the most similar factors among the socio-economic criteria are the distance from bus and metro stations; among weather, criteria are humidity and pressure; and among demographic criteria are male and female population ratios, literacy ratio, death ratio from asthma and cancer with a minimum and maximum difference of 0% and 64.94%.

Keywords: COVID-19, GIS, Spatial Similarity, Environmental factors, Demographic factors

1. Introduction

* Correspondence: n_kaffash@tabrizu.ac.ir

The Coronavirus disease 2019 (COVID-19) is a world health concern due to the rapid worldwide spread of it. More than 3750000 confirmed cases and 259000 deaths have been recorded until 8 May 2020 (WHO, 2020a). COVID-19 was first discovered at the end of 2019 in Wuhan, China, and now affects more than 210 countries (WHO, 2020b, 2020a). The outbreak of this disease has caused many economic, political, and social problems in different countries. Economic, cultural, social, and human activities have been disrupted. Therefore, it is very important to study the prevalence of this disease, the rate of spread and predictability, to identify high-risk areas and affective factors (Mollalo et al., 2020).

The GIS is an essential computer based tool for studying the spatial distribution, effective factors and spatio-temporal distribution of infectious diseases (Mollalo et al., 2020; Razavi Termeh and Sadeghi Niaraki., 2015). Despite the prevalence of COVID-19 in many countries around the world, there has been a notable increase in the number of cases and deaths in some countries such as the United States, Spain, and Italy (WHO, 2020a). Hence, it is necessary to investigate the factors that influence such a rapid increase and to discover similarities between the epicenter cities of COVID-19. One of the most important theoretical issues of the GIS is spatial similarity and it is appropriate to classify objects and solve problems. Spatial problems can be demonstrated by similar geospatial features, and the knowledge derived from the cases can help to solve other similar problems. By identifying similar cities in terms of COVID-19 prevalence, the common effective factors can be detected and the experiences of similar successful cities can be used to better manage the disease. Spatial similarity can be examined based on location, time, semantics, attributes, topology, and conditions (Liao et al., 2019).

Numerous studies have been done in the field of the COVID-19. Guan et al. (2020) studied data from 1099 patients of 552 hospitals in 30 different provinces of China. Based on their findings over the first 2 months, COVID-19 has spread around the world with different conditions and symptoms. Lai et al. (2020) studied COVID-19 patient data in countries around the world until February 11 and used graphs and maps to examine their patient numbers and specific symptoms. Recent worldwide studies have shown that the severity and speed of the COVID-19 spread can be affected by multiple factors such as air pollution (Wu et al., 2020), high temperature and humidity (Wang et al., 2020; Ma et al., 2020), smoking (Taghizadeh-Hesary and Akbari, 2020). Mollalo et al. (2020) used geographically weighted regression (GWR) and multiscale GWR (MGWR) models to investigate spatial dependence for COVID-19 cases associated with 35 environmental, socioeconomic, topographic, and demographic factors. Zhao et al. (2020) investigated the association between national train transportation and the COVID-19 outbreak in China. Their study has demonstrated a close correlation between the number of COVID-19 cases and the number of train journeys. Zhou et al. (2020) studied new challenges of the GIS with COVID-19 big data that hug data acquisition and integration, fast construction of information systems based on big data, spatial and spatiotemporal tracking, risk assessment and prediction are some examples.

Previous studies have not paid attention to data such as death rates from cardiovascular diseases, asthma, and pneumonia. Furthermore, socio-economic data such as distance from metro stations, bus stations, supermarkets, etc. have not been investigated, where the possibilities of touching and transmitting are increasing at such locations. In addition, the study of spatial similarity in cities with a high number of confirmed COVID-19 cases and death leads to the discovery of common effective factors.

In many studies, the similarity of the trajectories has been considered by the Manhattan and Euclidean distance (Furtado et al., 2016; Lehmann et al., 2019; Wan et al., 2017). In addition, the time sequence similarity of the spatial objects was investigated in different research (Yi et al., 1998). However, some research has been done on the similarities between different places and cities. Dobesova (2019) considered the similarity of European cities based on image processing with neural networks. He found similar cities based on the green areas patterns, the patterns of urban fabrics or special areas forms. The study did not pay attention to urban facilities and variable data such as climate. Preotiuc-Pietro et al. (2013) explored the venue-based similarity of cities through social network data. They applied three vector-based representation methods for foursquare venue data from 17 cities across the United States. Liao et al. (2019) proposed an approach to spatial data attribute similarity measurement based on the granular computing method. Their proposed model measures the similarity for the spatial entity in the form of a comprehensive index, based on distance and degree of membership. In their study, the similarity between different land-use was calculated based on the Manhattan distance, and no comparison was done between other distance metrics. The previous studies did not pay attention to measuring the socio-economic, statistical, and temporal similarity of cities, especially the study of such similarities in crises such as COVID-19.

This study investigated the spatial similarity of Wuhan, Tehran, Bergamo, Madrid, Paris, Daegu, New York, and Berlin-based on socio-economic, weather, demographic, and COVID-19 data. These data include the location of supermarkets, bus stations, metro stations, bakeries, butchers, hospitals, pharmacies, gas stations, and banks; population, male population ratio, female population ratio, median age, old age dependency percentage, literacy rate, and the death rate from cancer, air pollution, cardiovascular diseases, asthma, pneumonia, smoking; the maximum, minimum and average temperature, humidity, pressure and wind speed. The input data were prepared with the fuzzy membership function due to their uncertainty. The similarities between cities were calculated with a similarity model based on six distance metrics (Manhattan, Euclidean, Minkowski, Mahalanobis, Chebyshev, and Correlation). The results were aggregated with the Copeland method, as the results of each metric could be different. In terms of the COVID-19 situation, the most similar city was selected for each city, and the similarity between selected cities was calculated based on socio-economic, weather, and demographic criteria and was compared based on Ripley's K-function.

2. Materials and methods

2.1. Data collection

In this research, eight cities including Wuhan (1), Tehran (2), Bergamo (3), Madrid (4), Paris (5), Daegu (6), New York (7) and Berlin (8) were selected which have high levels of COVID-19 confirmed cases. In the first step, the required data was gathered from suitable sources in the city scale. Socio-economic data for different cities has been downloaded from OSM (<https://www.openstreetmap.org/>). These data include the location of supermarkets, bus stations, metro stations, bakeries, butchers, hospitals, pharmacies, gas stations, and banks. These criteria were selected due to the possibility of people in these places being close to each other and violating social distancing and the probability of COVID-19 transmission increase.

Demographic data were collected due to the impact of gender, age, and medical history on the probability of COVID-19 transmission (UNHCR, 2020). For example, since the COVID-19 pandemic began many people who need treatment for diseases such as cancer, cardiovascular disease and diabetes have not received appropriate health services and are at higher risk for COVID-19. Demographic data such as population, male population ratio, female population ratio, median age, old age dependency percentage, literacy rate, and the death rate from cancer, air pollution, cardiovascular diseases, asthma, pneumonia, and smoking were obtained for study areas from our world data until 2017 (<https://ourworldindata.org/>).

Because of severity and speed of the COVID-19 emergence can be affected by multiple factors such as air pressure, high temperature and humidity, weather data were considered among the epicenter cities (Wang et al., 2020; Ma et al., 2020). For each city, the maximum, minimum, and average temperature, humidity, pressure, and wind speed were gathered as weather data for March 2020 (<http://timeanddate.com/>). The COVID-19 data such as the number of confirmed cases, total deaths, total recovered cases, and transition speed were used based on WHO reports until April 15, 2020 (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>). The COVID-19 data that used as target data were shown in Table 1.

Table 1. The COVID-19 data for study areas

	Confirmed cases until 15 April (C)	Deaths until 15 April	Recovered until 15 April	Number of days from first cases until 15 April (N)	C/N	Confirmed cases in 21 days period	Population
Wuhan (1)	82341	3343	77816	96	857.71875	278	800500
Tehran (2)	74877	4683	48129	57	1313.6316	2282	8737510
Bergamo (3)	162488	21067	37130	77	2110.2338	14649	119834
Madrid (4)	174060	18255	67504	76	2290.2632	4165	3174300
Paris (5)	143303	15729	28805	83	1726.5422	6798	2190327
Daegu (6)	10618	228	7616	87	122.04598	7136	2449700
New York (7)	614606	26081	38820	86	7146.5814	30841	8537673
Berlin (8)	132210	3495	72600	70	1888.7143	1425	3520000

In this study, 49 factors were obtained according to nine socio-economic, four weather, and twelve demographic criteria chosen in different classes. The collected data were stored in a table where the table rows are the names of the cities and the columns are the names of the factors. For each city, the table value is the number of pixels in each factor. According to county-level and city-level data of novel COVID-19, 7 criteria were determined. Then, according to the 8 selected cities, a 49*8 table of the factors and a 7*8 table of Target (COVID-19 related data) were prepared.

2.2. MS small fuzzy membership

The fuzzy set's membership function is a generalization of the classical sets indicator function. It reflects the degree of truth, in fuzzy logic, as an extension of valuation. Zadeh presented membership functions in the first paper on fuzzy sets (1965). He proposed using a membership function on the domain of all possible values with a range (0,1). Several fuzzy membership functions have been introduced and the MS small function has been used in this study.

The fuzzy small membership function is utilized when it is more likely that the smaller input values are a member of the set. The identified midpoint determines the crossover point. Values greater than the midpoint having a lower probability of being a member of the set as well as taking a higher probability of membership for the values under the midpoint. The form and character of the transition zone shall be determined by the spread parameter. The role of fuzzy MS small membership is similar to the fuzzy small. In fuzzy MS small the mean and standard deviation are defined instead of midpoint and spread. The fuzzy MS small function may be more applicable than fuzzy small if a member of the set is more likely to have very small values (Ziani, 2017).

2.3. Similarity metrics

Spatial similarity is the similarity between objects on maps or in geospatial space that can show location, attribute, direction, distance, geometric, topology, and so on. The more factors that measure similarity, the more complex it becomes (Ding, 2004). Granular computing-based methods have been used to address this complexity. Granular computing is a modern type of conceptual method for the problem-solving process to solve hierarchical issues. Granular computing requires knowledge to be granulated first, and there are several methods to do so. In this paper, the atomic formula is used to granulate the data, and hence the similarity is calculated according to Eq. (1), (2) and (3) (Liao et al., 2019; Patra et al., 2011).

$$\mu_p^j(\varphi_i) = \frac{|m_p^j(\varphi_i) \cap p|}{|p|}; (i = 1,2,3, \dots, r) (j = 1,2,3, \dots, h) \quad (1)$$

$$d(p, q) = (\sum_{j=1}^h \sum_{i=1}^r |\mu_p^j(\varphi_i) - \mu_q^j(\varphi_i)|^z)^{\frac{1}{z}} \quad (2)$$

$$n(p, q) = 1 - \frac{1}{r * h} d(p, q) \quad (3)$$

Where p is the vector of the input pixels for the p city. $|p|$ is the total number of pixels in the p city, and the φ_i is an i th attribute which is examined in this research. Different attributes have many variable values, but in this research are investigated at three-level based on fuzzy logic exports. $m_p^j(\varphi_i)$ is the total number of pixels with attribute φ_i at level j , such as far, medium, and near levels of the distance from supermarkets. It is clear, $0 \leq \mu_p^j(\varphi_i) \leq 1$ and $\sum_{j=1}^3 \mu_p^j(\varphi_i) = 1$. $d(p, q)$ and $n(p, q)$ are the distance and the similarity between p and q cities, respectively. Also, r is the total number of pixels for each city, and h is the total number of levels at each attribute (Liao et al., 2019; Patra et al., 2011).

In Eq. (2), the value of z determined the distance type. In this research, the Manhattan, Euclidean, Minkowski, Mahalanobis, Chebyshev, and Correlation distance were investigated. The Manhattan, Euclidean and Minkowski distance were calculated with $z=1,2$ and 3 in Eq. (2). The Mahalanobis distance was calculated according to Eq. (4) and ρ is the correlation between p and q cities (De Maesschalck et al., 2000). The Correlation distance was calculated by $1 - \rho$ and the Chebyshev distance was obtained through $\max_{i,j=1}^{r,h} (\mu_p^j(\varphi_i) - \mu_q^j(\varphi_i))$ (Abello et al., 2002).

$$d(p, q) = (\sum_{j=1}^h \sum_{i=1}^r |\mu_p^j(\varphi_i) - \mu_q^j(\varphi_i)|^T \rho |\mu_p^j(\varphi_i) - \mu_q^j(\varphi_i)|)^{\frac{1}{2}} \quad (4)$$

2.4. Ripley's K-function

Ripley's K-function has been used to compare most similar criteria distributions among similar cities. The K-function is a valuable tool for identifying point patterns such as dispersed, clustered, or distributed at the study area. This function is similar to the Moran I function but is capable of describing point patterns at multiple scales defined by the user. The K-function is given as Eq.5 (Mitchell, 2009).

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, i \neq j}^n k_{ij}}{\pi n(n-1)}} \quad (5)$$

Where d is the distance, n is the total numbers of points, A is the total area, and k_{ij} represents a weight. No edge correction occurs, $k_{ij}=1$ if the distance between i and j is less than d , and otherwise $k_{ij}=0$ (Mitchell, 2009).

3. Results

3.1. Fuzzy modeling

Euclidean distance has been used to create distance layers from selected socio-economic data. Due to the uncertainty in the calculation of distances, fuzzy logic was used to prepare them. The MS small membership function was selected for fuzzification, since the probability of COVID-19 transmission increases with proximity to these places. Therefore, the prepared criteria were classified into three classes: far, medium, and near. For example, the fuzzified distance layers from supermarkets are shown in Fig. 1 that were demonstrated the number and distribution of supermarkets in eight study areas. All fuzzified factor layers are shown in Fig. 2. The columns and rows demonstrated cities and factor layers, respectively.

3.2. Factor and target-based similarity

The prepared tables were entered in the similarity model and the results of Table 2 were obtained with six distance metrics. Factor and target-based similarity were shown as F and T in Table 2's columns, respectively. According to the results obtained for each city, the most similar city has been identified in terms of factors and targets. For instance, for Wuhan with the Manhattan metric, the most similar cities in terms of factors were Bergamo and Daegu, and in terms of Target, Berlin, and Daegu. While for this city, according to the Mahalanobis metric, the most similar cities in terms of factors and targets were New York, Paris, Daegu, and Tehran, respectively.

The results of the Manhattan metric were shown in Fig. 3. The similarity range is 0.7-1 that one similarity measure, for each city, is the city itself. For each city, the largest number, less than one represents the city with the most similarity.

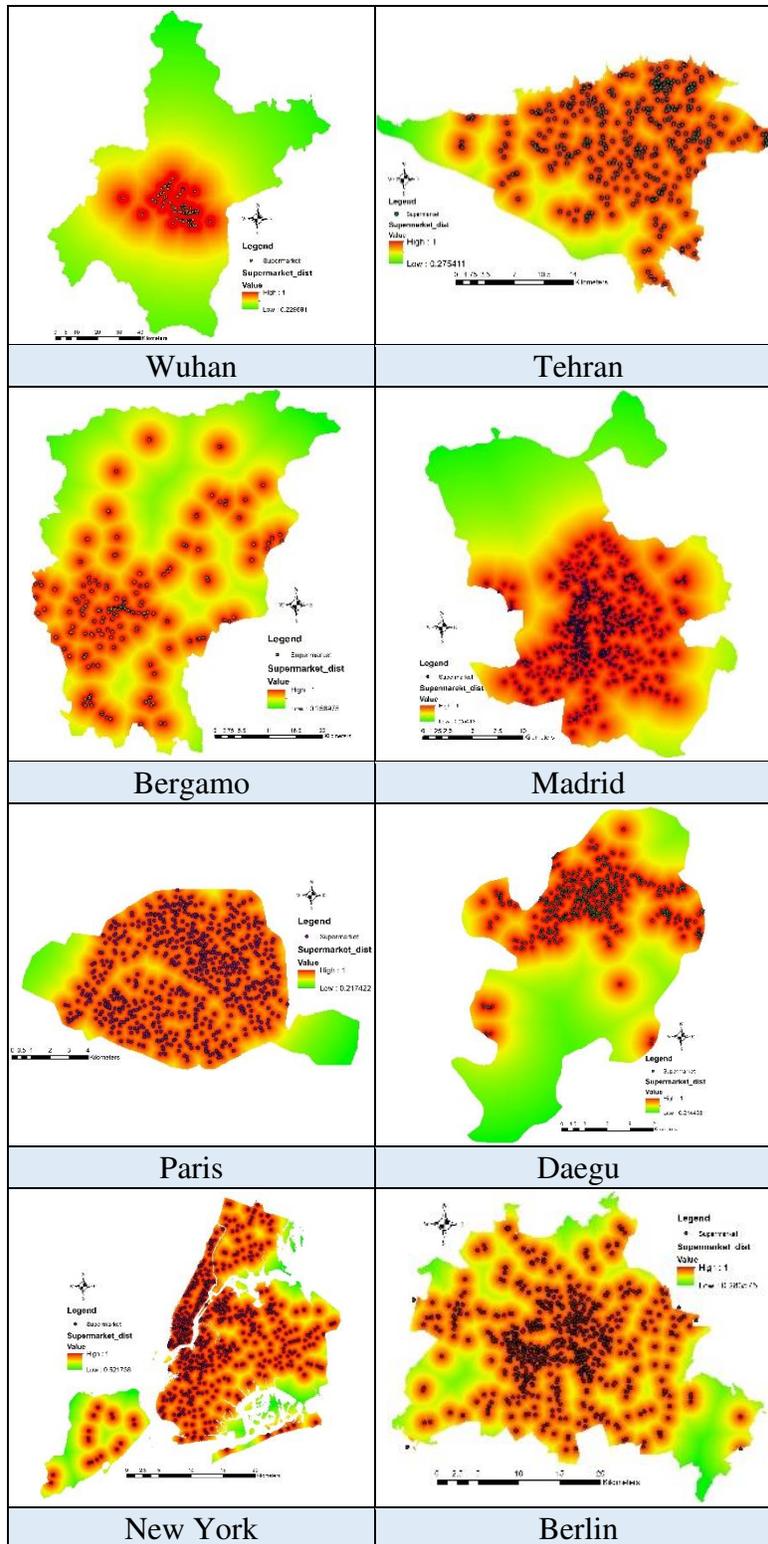


Fig. 1. The fuzzified distance layers from supermarkets in eight study areas

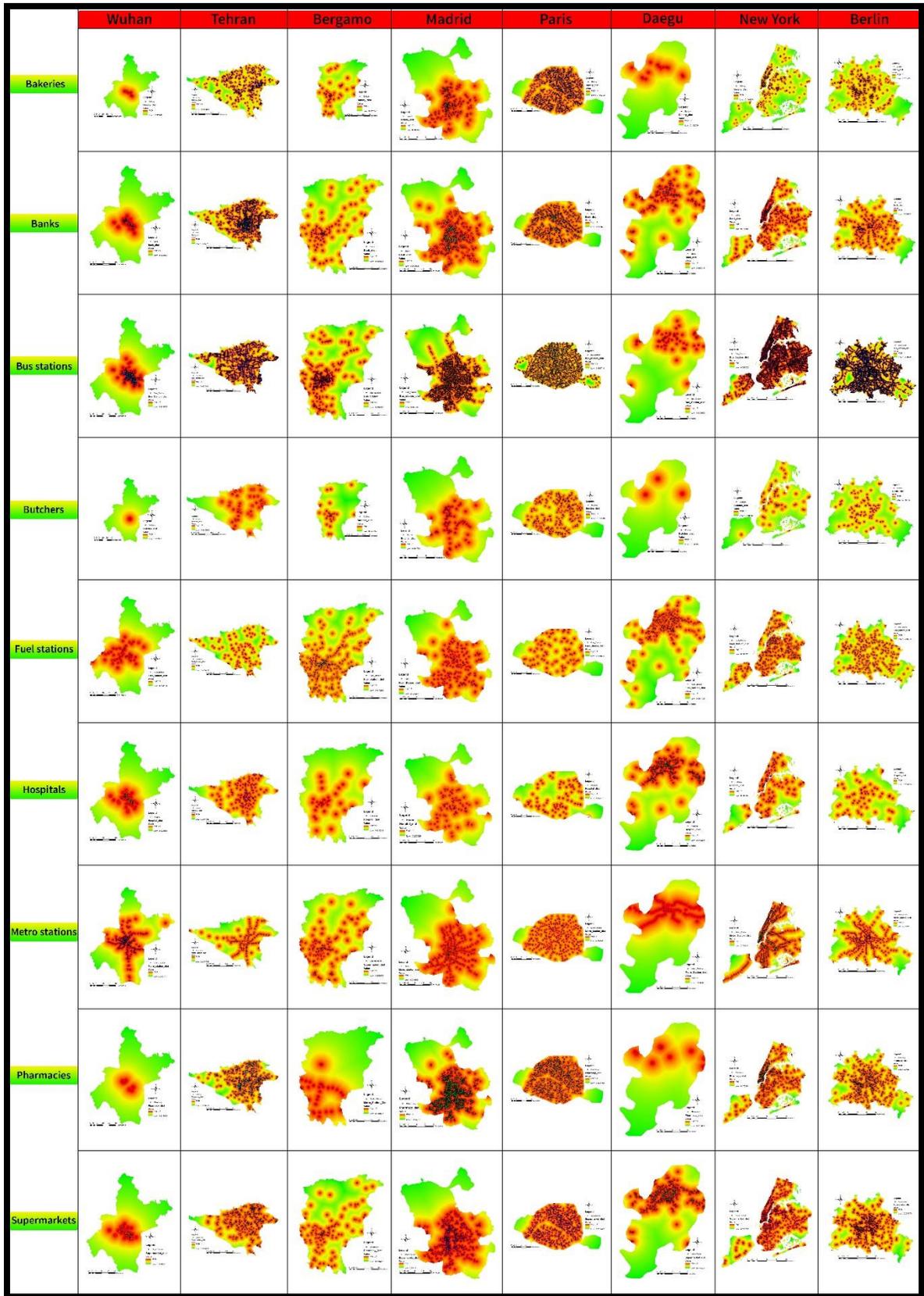


Fig. 2. The fuzzified factor layers in eight study areas

Table 2. The results of the similarity model for study areas based on six distance metrics

Cities	Manhattan		Euclidean		Minkowski		Mahalanobis		Chebyshev		Correlation	
	F	T	F	T	F	T	F	T	F	T	F	T
Wuhan (1)	3,6	8,6	6,3	8,4	7,5	8,5	7,5	6,2	3,6	8,5	6,3	8,4
Tehran (2)	5,7	8,1	5,8	8,4	8,7	4,8	6,5	1,4	5,8	8,4	5,7	8,6
Bergamo (3)	4,6	5,4	4,6	5,4	4,6	5,2	4,6	4,6	4,6	5,4	4,6	5,7
Madrid (4)	3,8	5,8	3,6	5,8	3,8	3,2	3,6	3,2	3,6	5,3	8,7	1,5
Paris (5)	2,8	4,3	2,4	4,3	8,2	4,3	6,1	6,4	2,4	4,3	2,4	3,4
Daegu (6)	4,3	5,1	3,4	5,1	4,2	1,8	2,5	1,5	3,4	5,8	1,3	2,1
New York (7)	8,2	3,4	8,2	3,4	1,8	3,4	1,8	5,3	8,4	3,5	8,2	3,5
Berlin (8)	4,3	1,4	4,3	1,4	7,2	5,4	4,3	6,4	7,4	1,5	7,4	1,2

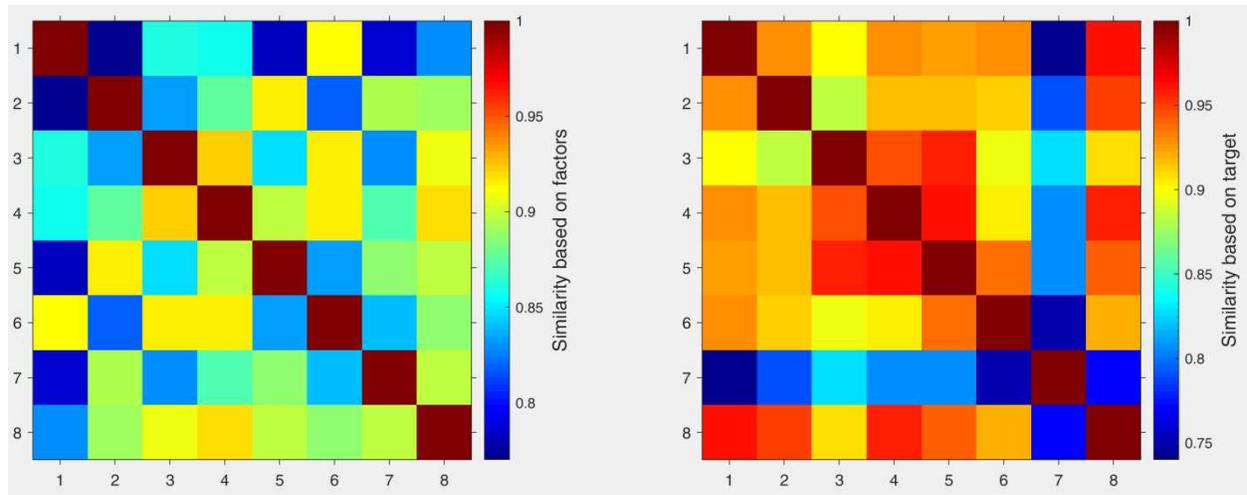


Fig. 3. The results of the similarity model based on factors and target by Manhattan distance metric

3.3. Aggregate results based on the Copeland method

Due to the different results obtained from different distance metrics of similarity, the aggregation method must be used to achieve the optimal unique answer. In the Copeland pairwise aggregation process, candidates are categorized by the number of pairwise wins, minus the number of pairwise defeats. Based on the Copeland method, different results of Table 2 were scored and the final similar cities were selected.

In this way, for each city, similar cities were obtained by different methods, based on factors. The score for each city was calculated based on Copeland, and the most similar city with the highest score was selected. This procedure has also been carried out for targets. The results are shown in Table 3.

For example, in terms of COVID-19 based targets for Wuhan, Berlin is the most similar city. The similarity rate between Wuhan and Berlin is 82.88 % based on the factors selected. The same for Bergamo is Madrid with 92.36% similarity. Therefore, the most similar city was chosen for all of the studied cities in terms of Corona based statistics, and the similarity percentage of environmental, climate, and demographic factors were calculated. The selected factors seem appropriate because the highest and lowest factor similarity percentages were 92.36 and 82.85 percent, respectively.

For all similar cities, the criteria have been examined and the most similar and different criteria have been presented in Table 4 with their percentage difference. The 49*8 table of the factors were investigated and the columns with the lowest and highest differences were selected for each similar

city. For example, between Wuhan and Berlin, difference among maximum humidity is 0.06%, while the difference between the numbers of pixels in areas far from bus stations is 64.94%. So, the maximum humidity and areas far from bus stations were selected as the most similar and most difference factors, respectively. For other similar cities, the same was determined and the results were presented in Table 4.

Table 3. The Copeland aggregation results of similarity model for study areas

Cities	Similar cities based on factors			Similar cities based on COVID-19 data			Similarity % of cities based on factors	Most similar city
	City	Copeland score	Similar city	City	Copeland score	Similarity % from Manhattan dist		
Wuhan (1)	6	2	6,3	8	4	96.74	82.88	Berlin (8)
	3	2		6	-2	93.95	91.35	
	7	-2		4	-2	93.69	85.93	
	5	-2		2	-4	93.94	77.01	
Tehran (2)	5	2	5	8	4	95.48	88.96	Berlin (8)
	8	0		4	2	92.87	87.63	
	7	0		1	-2	93.94	77.01	
	6	-4		6	-4	92.48	81.77	
Bergamo (3)	4	6	4,6	5	4	96.11	84.86	Madrid (4)
				4	2	95.13	92.36	
	6	6		2	-4	89.74	83.26	
				6	-4	90.86	91.40	
				7	-4	84.87	82.85	
Madrid (4)	3	4	3	5	2	96.66	89.70	Bergamo (3)
	6	0		3	0	95.13	92.36	
	8	0		8	-2	96.41	91.84	
	7	-4		2	-2	92.87	87.63	
				1	-4	93.69	85.93	
Paris (5)	2	4	2	4	6	96.66	89.70	Madrid (4)
	4	0		3	4	96.11	84.86	
	8	-2		6	-4	94.62	83.31	
	6	-4						
Daegu (6)	4	2	4,3	1	4	93.95	91.35	Wuhan (1)
	3	2		5	2	94.62	83.31	
	2	-2		8	-2	93.21	88.84	
	5	-4		2	-4	92.48	81.77	
	1	-4						
New York (7)	8	6	8	3	6	84.87	82.85	Bergamo (3)
	2	0		4	0	83.17	87.32	
	1	-2		5	0	82.98	88.83	
	4	-4						
Berlin (8)	4	4	4	1	2	96.74	82.88	Wuhan (1)
	7	0		4	2	96.41	91.84	
	3	0		5	-2	94.88	89.74	
	2	-4		6	-4	93.21	88.84	
				2	-4	95.48	88.96	

Table 4. The most similar and difference factors between similar cities

Cities		Most similar factors (difference %)	Most difference factors (difference %)
Wuhan (1)	Berlin (8)	Maximum humidity (0.06%) Areas far from hospitals (0.02%)	Areas near to pharmacies (64.50%) Areas far from bus stations (64.94%)
Tehran (2)	Berlin (8)	Areas near to bus stations (0.13%) Female population ratio (0.17%)	Areas near to bakeries (39.75%) Areas near to butchers (43.17%)
Bergamo (3)	Madrid (4)	Maximum humidity (0%) Female population ratio (0.08%)	Areas medium from bus stations (20.56%) Areas far from pharmacies (20.8%)
Paris (5)	Madrid (4)	Maximum humidity (0%) Female population ratio (0.11%)	Areas near to bus stations (36.14%) Areas far from bus stations (44.00%)
Daegu (6)	Wuhan (1)	Pressure (0%) Areas near to bus stations (0.13%)	Areas near to banks (32.77%) Area (50.39%)
New York (7)	Bergamo (3)	Literacy rate (0%) Male population ratio (0.22%)	Areas far from banks (52.62 %) Areas far from bus stations (58.83%)

The results of the K-function for most similar criteria among Wuhan-Berlin, Tehran-Berlin, and Daegu-Wuhan were shown in Fig. 4. For example, Fig. 4-a and 4-b, demonstrated K-function value for hospital locations in Wuhan and Berlin, respectively. Red, blue and dash lines show observed spatial patterns, expected spatial patterns, lower and higher confidence envelopes, respectively. The observed spatial pattern higher than the expected spatial pattern indicates the existence of statistically significant clusters at different distances. The slope of the red line, the higher value of the red line relative to the blue line, and the distance from it, indicate the similarity between (4-a and 4-b), (4-c and 4-d) and (4-e and 4-f) plots in terms of the distribution of hospitals and bus stations, which were selected as the most similar criteria. Since the similar pair plot properties demonstrate the similarity of the pair of cities. For 86 Wuhan and 99 Berlin hospitals, the K-function ranges were 2000-12000 and 4000-13000 respectively, which showed a similar pattern of spatial distribution. Details of the area, number of points, distances and K-function values of other similar cities are shown in Figure 4.

4. Discussion

The COVID-19 pandemic is one of the most important human problems of 2020. Socio-economic, temporal, and statistical analysis of this disease were investigated using GIS capabilities in various cities. In this research, based on socio-economic, weather, and demographic data, similarities of high-risk cities were examined. For this purpose, a combined model of fuzzy logic, similarity index, and Copeland method was used. The results showed that there is a high similarity between the COVID-19-implies cities. Based on the result, for example, Berlin is the most similar city to Wuhan and Tehran, considering the COVID-19 situation. Due to the 7 * 8 table target and the similarity index, the difference between these cities was the lowest and they were selected as the most similar cities. Based on the 49*8 table of the factors and similarity metric, the socio-economic, weather, and demographic similarities between Wuhan and Berlin are 82.88% and 88.96% between Tehran and Berlin.

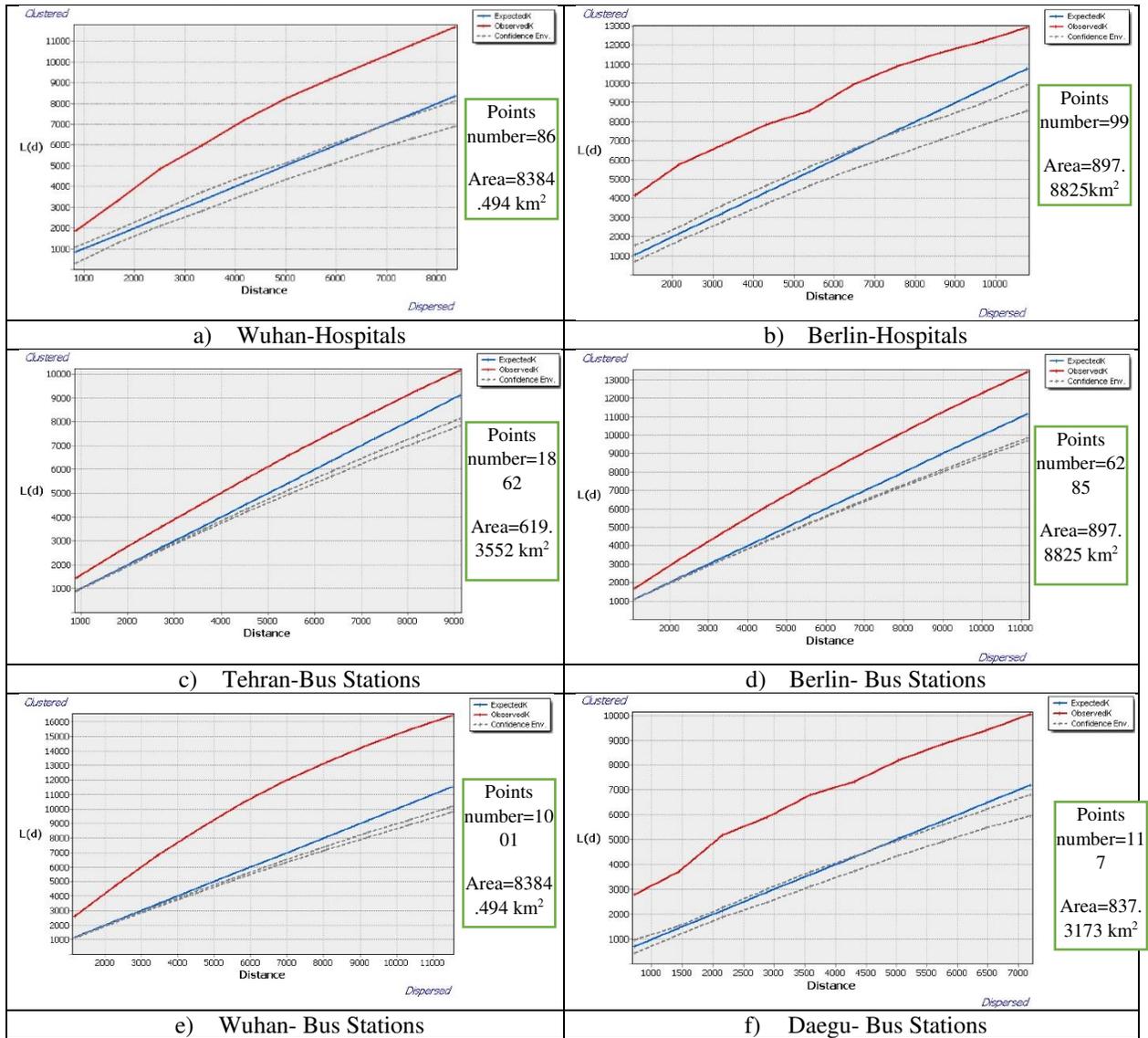


Fig. 4. The results of the K-function for most similar criteria of similar cities

Likewise, the most similar cities to Madrid are Bergamo and Paris with similarities rates of 92.36% and 89.70%. For Daegu and New York, Wuhan and Bergamo are the most similar cities, with similarities of 91.35% and 82.85%, respectively. In terms of selected factors, the most similar city was often different from the most similar city in terms of the COVID-19 situation (target). Except for Bergamo and Madrid, which were the most similar in terms of both factors and targets.

The most difference and most similar criteria between similar cities were extracted based on 49 factors and similarity metrics (Eq. 2). For similar cities, the most similar factors among the socio-economic criteria are the distance from bus and metro stations and hospital; among climatic criteria are humidity and pressure; and among demographic criteria are male and female population ratios, literacy ratio, death ratio from asthma and cancer. These criteria differed very little in two similar cities. For example, the difference percentage between the number and distribution of metro and bus stations by area was about 4% for Tehran and Berlin. For these two cities, the most similar

criteria were areas near bus stations and the most different criteria were the areas close to the butchers with a difference of 0.13% and 43.17%, respectively. While the most similar and different criteria for Madrid and Bergamo are maximum humidity and areas away from pharmacies with a difference of 0% and 20.8%.

One of the limitations of this study was the availability of data for all cities around the world. Due to the time-consuming collection of environmental and demographic data for all cities, eight cities were examined in this study. Further research would complete the required data for all countries and classified cities around the world on the similarities of the COVID-19 situation. Another limitation was the lack of cultural and political data. Furthermore, by adding cultural and political criteria such as how to implement quarantine law, its time, how to implement social distancing, and fines, similarities would be examined more accurately.

5. Conclusions

In this study, the similarity indexes between the eight cities of Wuhan, Tehran, Bergamo, Madrid, Paris, Daegu, New York and Berlin were calculated in terms of the COVID-19 situation. Similarity results were verified with six different distance metrics. Then, for a pair of similar cities, the similarity indexes were calculated based on socio-economic, weather and demographic factors. In general, Wuhan, Berlin, Tehran, and Daegu, with an average factor based similarity of 86.51%, and Madrid, Bergamo, New York, and Paris, with an average factor based similarity of 89.32%, were the most similar cities in terms of COVID-19.

These similarities help to learn and understand issues better. The similarity in GIS can be investigated in the context of socio-economic, temporal, demographic, and statistical data. Identifying cities with similar conditions in terms of COVID-19 will help to manage the disease well. Furthermore, determining the effective factors among similar cities is a way of identifying effective factors in the transmission of COVID-19. However, studies on COVID-19 and its effective factors are ongoing and the idea of this study can be regarded as a basis for future researches in this field.

Acknowledgment (Funding):

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00312) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation)

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the co-authors consent the publication of this work.

Competing interests

There are no competing interests.

CRediT authorship contribution statement

Neda Kaffash Charandabi: Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Abolghasem Sadeghi Niaraki:** Conceptualization, Writing -review & editing. **Soo-Mi Choi:** review & editing.

Data Availability

Data from this work are publicly archived online:

[\(https://ourworldindata.org/\)](https://ourworldindata.org/)

[\(http://timeanddate.com/\)](http://timeanddate.com/)

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

[\(https://www.openstreetmap.org/\)](https://www.openstreetmap.org/)

References

- Abello, J.M., Pardalos, P. M., Resende M. G., 2002. Handbook of Massive Data Sets. Springer. ISBN 1-4020-0489-3.
- De Maesschalck, R., Jouan-Rimbaud, D., Massart, D. L., 2000. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50 (1), 1–18. [https://doi.org/10.1016/s0169-7439\(99\)00047-7](https://doi.org/10.1016/s0169-7439(99)00047-7).
- Ding, H., 2004. A study on Spatial Similarity Theory and Calculation Model. Ph.D. Thesis, Wuhan University, Wuhan, China.
- Dobesova, Z., 2019. The Similarity of European Cities Based on Image Analysis, *Intelligent Systems Applications in Software Engineering*, 341-348. https://doi.org/10.1007/978-3-030-30329-7_31.
- Furtado, A. S., Kopanaki, D., Alvares, L. O., Bogorny, V., 2016. Multidimensional similarity measuring for semantic trajectories, *Transactions in GIS*, 20 (2), 280-298. <https://doi.org/10.1111/tgis.12156>.
- Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D, S, C., Du, B, Li, L., Zeng, G., Yuen, K, Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, J, L., Liang, Z., Peng, Y., Wei, L., Liu, Y., Hu, Y, H., Peng, P., Wang, J, M., Liu, J., Chen, Z., Li, G., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S., Zhong, N., 2020. Clinical Characteristics of Coronavirus Disease 2019 in China, *The new England journal of medicine*. <https://doi.org/10.1056/NEJMoa2002032>.
- Lai, Ch, Ch., Shih, T, P., Ko, W, Ch., Tang, H, J., Hsueh, P, R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, 55, 105924. <https://doi.org/10.1016/j.ijantimicag.2020.105924>.
- Lehmann, A. L., Alvares, L. O., Bogorny, V., 2019. SMSM: a similarity measure for trajectory stops and moves, *International Journal of Geographical Information Science*, 1-26. <https://doi.org/10.1080/13658816.2019.1605074>.
- Liao, W., Hou, D., Jiang, W., 2019. An approach for a spatial data attribute similarity measure based on granular computing closeness. *Applied Sciences*, 9, 2628. <https://doi.org/10.3390/app9132628>.
- Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., Luo, B., 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Science of the Total Environment*, 724, 138226. <https://doi.org/10.1016/j.scitotenv.2020.138226>.
- Mitchell, A. 2009. The ESRI Guide to GIS Analysis. Volume 2: Spatial Measurements and Statistics. Redlands, CA: Esri Press.
- Mollalo, A., Vahedi, B., Rivera, K. M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the Total Environment*, 728. <https://doi.org/10.1016/j.scitotenv.2020.138884>.
- Patra, B.K., Nandi, S., Viswanath, P., 2011. A distance based clustering method for arbitrary shaped clusters in large datasets. *Pattern Recognit.* 44, 2862–2870. <https://doi.org/10.1016/j.patcog.2011.04.027>.
- Preotiuc-Pietro, D., Cranshaw, J., Yano, T., 2013. Exploring venue-based city-to-city similarity measures, *UrbComp 13: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 16, 1-4. <https://doi.org/10.1145/2505821.2505832>.

Razavi Termeh, V., Sadeghi Niaraki, A., 2015. Design and implementation of ubiquitous health system (u-health) using smart-watch sensors. *International archives of the photogrammetry, remote sensing and spatial information sciences- ISPRS archives*, 40(1W5), 607-612. <https://doi:10.5194/isprsarchives-XL-1-W5-607-2015>.

Taghizadeh-Hesary, F., Akbari, H., 2020. The powerful immune system against powerful COVID-19: a hypothesis. *Preprints 2020*, 2020040101. <https://doi.org/10.20944/preprints202004.0101.v1>.

United Nations High Commissioner for Refugees (UNHCR), 2020. AGD CONSIDERATIONS – COVID-19. Retrieved from <https://data2.unhcr.org/en/documents/download/75295>.

Wan, Y., Zhou, C., Pei, T., 2017. Semantic-geographic trajectory pattern mining based on a new similarity measurement, *ISPRS International Journal of Geo-Information*, 6(7), 212. <https://doi.org/10.3390/ijgi6070212>.

Wang, Jingyuan., Tang, Ke., Feng, Kai., Lv, Weifeng., 2020. High temperature and high humidity reduce the transmission of COVID-19.

World Health Organization (WHO), 2020a. Coronavirus Disease 2019 (COVID-19) Situation Report – 109. Retrieved from. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200508covid-19-sitrep-109.pdf?sfvrsn=68f2c632_6.

World Health Organization (WHO), 2020b. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Retrieved from <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>.

Wu, X., Nethery, R.C., Sabath, B.M., Braun, D., Dominici, F., 2020. Exposure to Air Pollution and COVID-19 Mortality in the United States. COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv, <https://doi.org/10.1101/2020.04.05.20054502>.

Yi, B, K., Jagadish, H., Faloutsos, C., 1998. Efficient retrieval of similar time sequences under time warping, in *Proceedings 14th International Conference on Data Engineering: IEEE*, 201-208.

Zadeh L.A., 1965. Fuzzy sets. *Information and Control*, 8, 338–353.

Ziani, S., 2017. Time-varying fuzzy sets based on Gaussian membership functions for developing fuzzy controller. *Iranian Journal of Fuzzy Systems*, 14(3), 15-39.

Zhao, S., Zhuang, Z. , Ran, J. , Lin, J. , Yang, G. , Yang, L. , He, D., 2020. The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: A data-driven correlational report. *Travel Med. Infect. Dis.* 33, 101568. <https://doi.org/10.1016/j.tmaid.2020.101568>.

Zhou, Ch., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Zh., Wang, J., Yuan, W., Zhu, Y., Song, C., Chen, J., Xu, J., Li, F., Ma, T., Jiang, L., Yan, F., Yi, J., Hu, Y., Liao, Y., Xiao, H. 2020. COVID-19: Challenges to GIS with Big Data. *Geography and Sustainability*, 1, 77-87. <https://doi.org/10.1016/j.geosus.2020.03.005>.

Figures

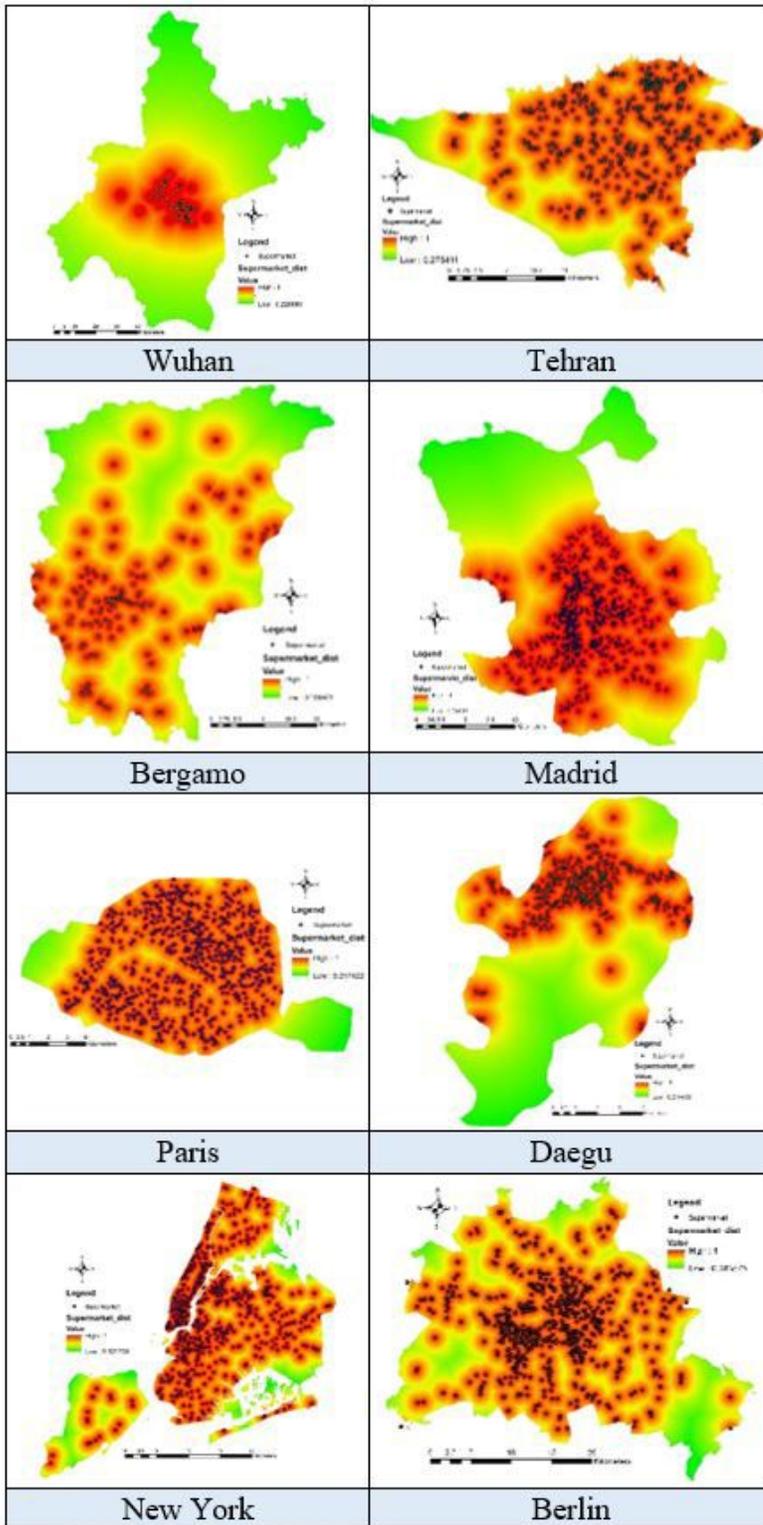


Figure 1

The fuzzified distance layers from supermarkets in eight study areas. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its

authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

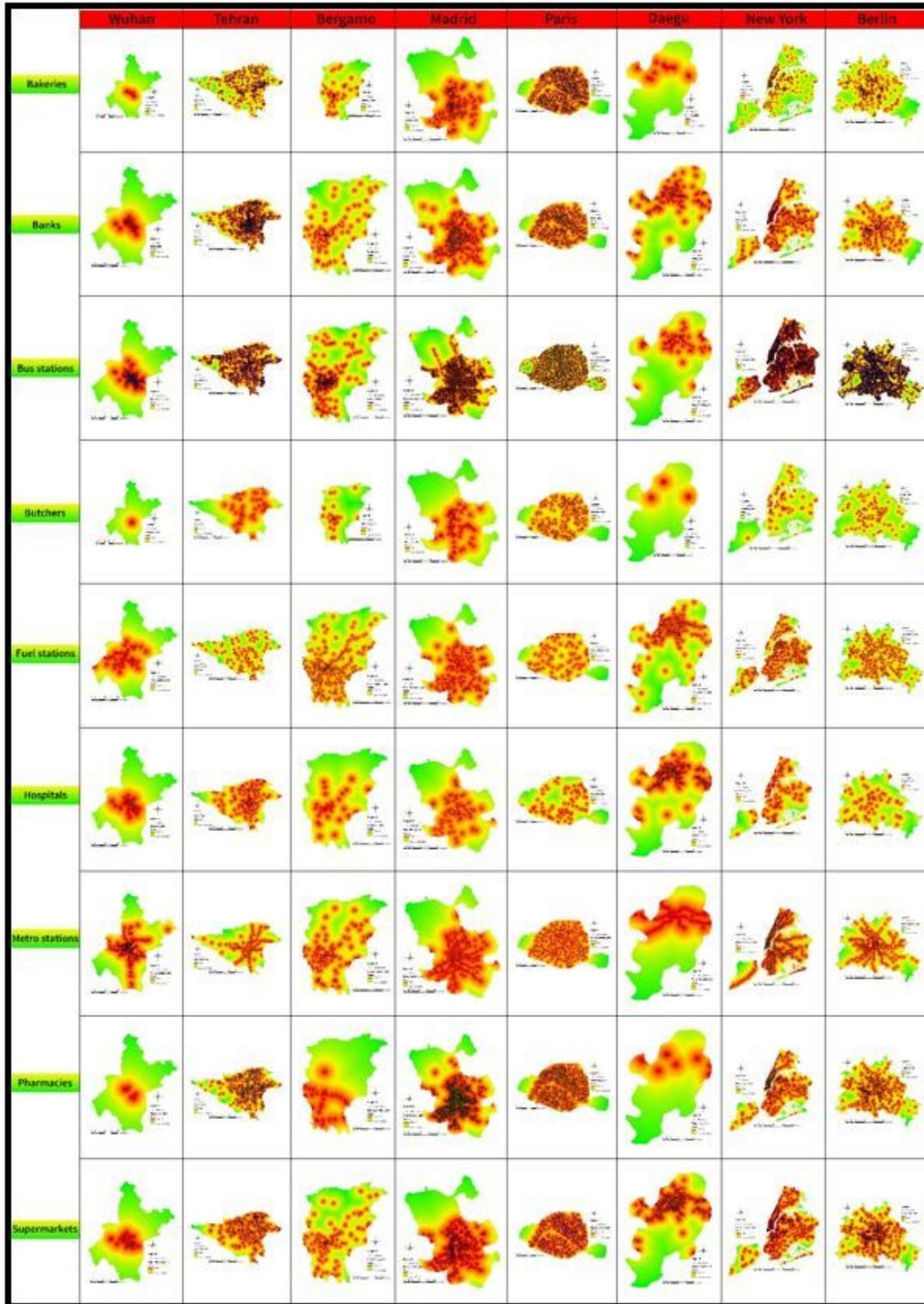


Figure 2

The fuzzified factor layers in eight study areas

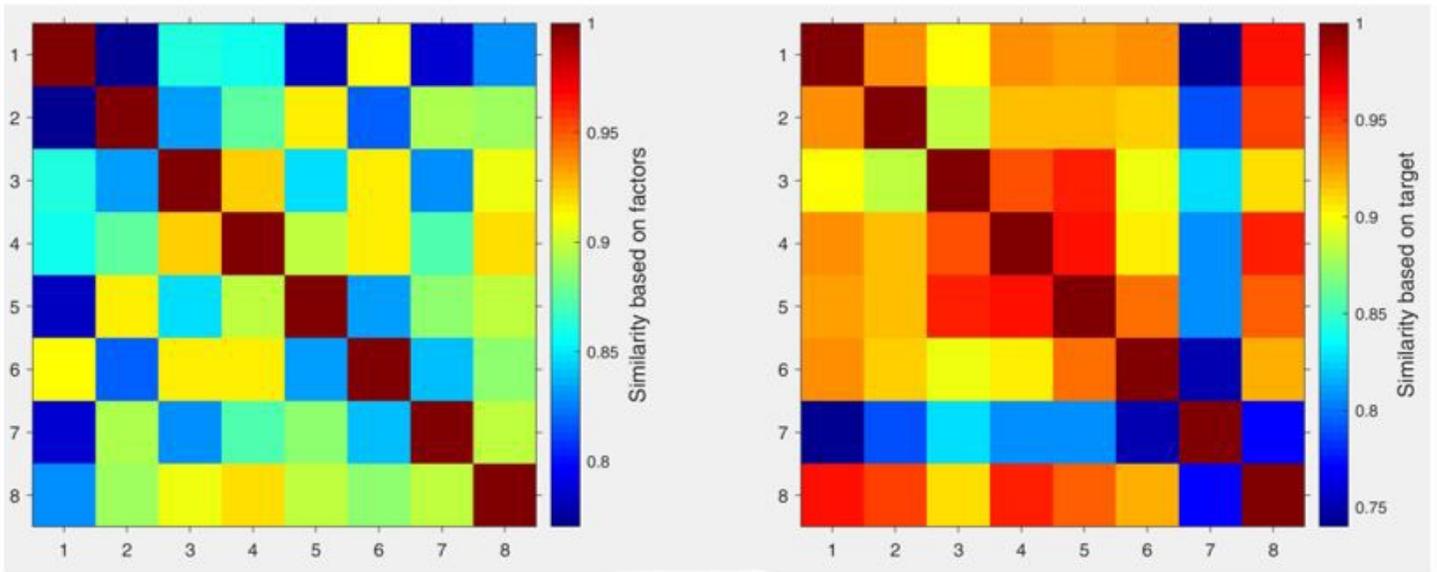


Figure 3

The results of the similarity model based on factors and target by Manhattan distance metric

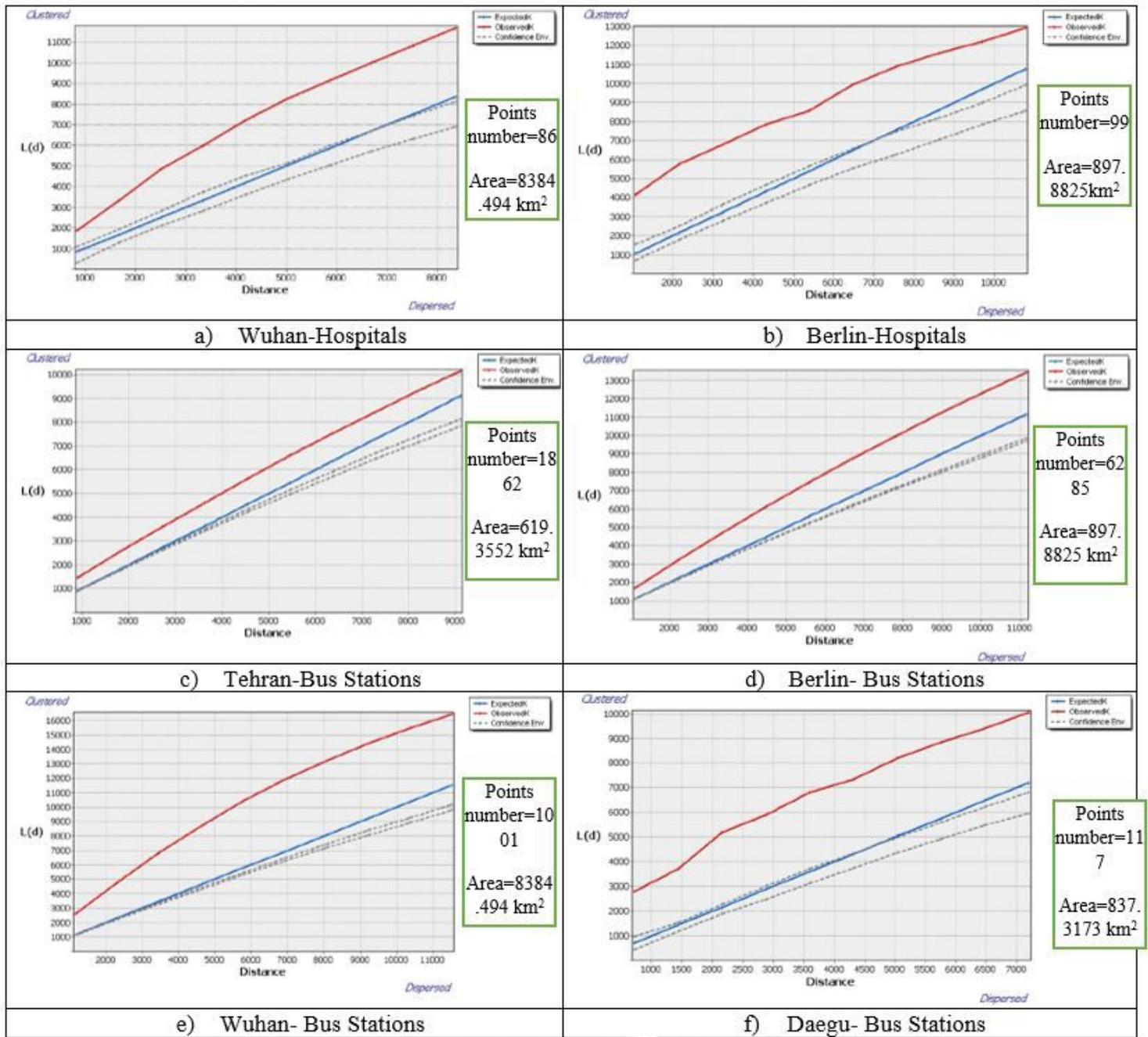


Figure 4

The results of the K-function for most similar criteria of similar cities