

Article

# Ranking the City: The Role of Location-Based Social Media Check-Ins in Collective Human Mobility Prediction

Omid Reza Abbasi, Ali Asghar Alesheikh \* and Mohammad Sharif

GIS Engineering Department, K. N. Toosi University of Technology, Tehran, Iran;  
oabbasi@mail.kntu.ac.ir (O.R.A.); msharif@mail.kntu.ac.ir (M.S.)

\* Correspondence: alesheikh@kntu.ac.ir; Tel.: +98-912-159-7191

Academic Editors: Georg Gartner, Haosheng Huang and Wolfgang Kainz

Received: 13 January 2017; Accepted: 25 April 2017; Published: 28 April 2017

**Abstract:** Technological advances have led to an increasing development of data sources. Since the introduction of social networks, numerous studies on the relationships between users and their behaviors have been conducted. In this context, trip behavior is an interesting topic that can be explored via Location-Based Social Networks (LBSN). Due to the wide availability of various spatial data sources, the long-standing field of collective human mobility prediction has been revived and new models have been introduced. Recently, a parameterized model of predicting human mobility in cities, known as rank-based model, has been introduced. The model predicts the flow from an origin toward a destination using “rank” concept. However, the notion of rank has not yet been well explored. In this study, we investigate the potential of LBSN data alongside the rank concept in predicting human mobility patterns in Manhattan, New York City. For this purpose, we propose three scenarios, including: rank-distance, the number of venues between origin and destination, and a check-in weighted venue schema to compute the ranks. When trip distribution patterns are considered as a whole, applying a check-in weighting schema results in patterns that are approximately 10 percent more similar to the ground truth data. From the accuracy perspective, as the predicted numbers of trips are closer to real number of trips, the trip distribution is also enhanced by about 50 percent.

**Keywords:** human mobility; rank-based model; location-based social networks

## 1. Introduction

Development of new data acquisition techniques has facilitated the study of human mobility patterns. Taking advantages of Global Positioning System (GPS) devices embedded in smartphones, location-based social networks (LBSN) have provided the possibility of studying the relationships between users and places. A LBSN is a special type of social network allowing the users to share their personal location. Researchers can, then, benefit from this data source. These kind of data have opened the doors onto novel research about activities within the cities [1].

Up until now, numerous models of predicting collective human mobility have been developed. Interdisciplinary in nature, human mobility prediction has a broad range of applications, from epidemic control [2,3], spatial economy [4,5], energy management [6] and urban planning [7,8], to location-based services [9] and tourism management [10,11]. There are two major, but different, assumptions in modeling the collective human mobility patterns. Some models (e.g., the gravity model) assume that a trip is directly related to the distance between an origin and a destination. In other words, the greater the distance between an origin and a destination, the lower the probability of traveling from

the origin to that destination will be [12]. Equation (1) presents the doubly-constrained gravity model of spatial interaction.

$$T_{ij} = A_i T_i B_j T_j f(r_{ij}) \quad (1)$$

where  $T_i$  and  $T_j$  are the number of trips departing from  $i$  and ending in  $j$  respectively.  $f(r_{ij})$ , a termed deterrence function, is a function in terms of the cost of making the trip; usually the distance between the two locations. It is commonly considered as a power law or exponential function of the distance.  $A_i$  and  $B_j$  are balancing factors and their values are determined through an iterative process.

In contrast to the gravity framework, some models explain human mobility using an “opportunities” concept. These models assume that trips are not directly related to the distance, but induced by opportunities provided at the destination. The most known model that employs opportunities concept is the Intervening Opportunities (I.O.) model. Equation (2) describes how trips distribute over an area according to the I.O. model [13].

$$T_{ij} = T_i \frac{e^{-\alpha(S_{ij}-m_j)} - e^{-\alpha S_{ij}}}{1 - e^{-\alpha M}} \quad (2)$$

where  $S_{ij}$  is the number of opportunities located between an origin and a destination,  $m_j$  is the population of the destination  $j$ , and  $M$  is the total population of the study area. The parameter  $\alpha$  is adjustable and its value should be determined through a calibration procedure. The I.O. model assumes that opportunities have a specific effect on the probability of making a trip toward a region.

Some researchers have taken the advantages of LBSN data in context of human mobility prediction. Liu and his colleagues have used social media check-ins to study the inter-urban trip patterns at a collective level. They have employed the gravity model to study the collective movements in China [14]. Yuan and Medel have also studied international travel behaviour using geotagged photos from Flickr. They constructed a gravity model and investigated how the popularity of a given destination affects the travel choice [15]. Due to the advent of novel positioning technologies, researchers have developed various models based on the above assumptions. Among them is the Rank-based model [16] that is a probabilistic model of human mobility prediction. It uses a rank concept to predict the probability of going from an origin toward a destination. In fact, each destination has a rank, with respect to the origin, that expresses the probability of going from a region to another. However, the question that “*how the rank should be computed?*” has not been answered well, so far. In this paper, we consider different versions of rank concept and evaluate each method to see which version yields more accurate results. First, we use the common approach, that is, ranks are computed using the distance between origins and destinations. Second, we propose a venue-based method in which the number of venues located within a circle between origin and destination is used. Finally, we apply a check-in weighting schema to the venues of the second scenario.

Similar to the research conducted by Noulas and his colleagues [16], who have also presented the rank-based model, Yan et al. [17] employed the rank-based model to compare its results with that of their own proposed model, called the Population Weighted Opportunities (PWO) model. They have used the rank-distance between origins and destinations to compute the model. On the other hand, the leading parameter of the PWO model, as its name implies, is the population between origin and destination. Since the population is, to some extent, more dynamic than pure physical distance, the comparison of results obtained from the PWO model and rank-based model does not seem to be fair.

Liang and his colleagues [18] have overcome the above issue in a sense, by using the population located inside a circle, centered at the origin, with a radius equal to the travel distance. Moreover, they have presented an alternative version of a rank-based model in which the adjustable parameter has been eliminated. Although there should be a relatively large number of locations across the area for their model to be formed, this is not necessarily the case, particularly in intra-urban scenarios.

Chen et al. [19] have utilized the rank-based model to investigate the urban mobility patterns and understand the impact of spatial distribution of places of interest on them. In fact, they have simulated

the urban mobility in some synthetic cities by rank-based model. Their approach of using Places Of Interest (POIs) to compute the rank is replicated here, and the results are compared with that of our proposed method.

Relating the human mobility to places of interest also has the advantage of conducting analyses on land parcels within the city. Lu et al. [20], for instance, have developed a framework to characterize the life cycle of POIs in Manhattan using human mobility patterns. A thorough understanding of the POI life cycle may help in urban planning [21,22], site selection procedure [23], and real estate evaluation [24].

Making use of the relationships between human mobility and activity records such as check-in data extracted from social media, Long et al. [25], have proposed a methodology to evaluate the effectiveness of urban growth in Beijing.

The structure of this paper is as follows: Section 2 explains the mechanism of a rank-based model and the way it is calibrated and balanced. Section 3 introduces the study area and the datasets used. Section 4 provides the methods of evaluating the model and presents the results. Finally, Section 5 concludes the key remarks of the study.

## 2. Methodology

According to the rank-based model, given a set of zones  $u \in U$  in a city, the probability of moving from zone  $u \in U$  to a zone  $v \in U$  is defined as [16]

$$P[u \rightarrow v] \propto \frac{1}{\text{rank}_u(v)^\gamma} \quad (3)$$

where  $\text{rank}_u(v)$  is the rank of zone  $v$  relative to zone  $u$  and  $\gamma$  is an adjustable parameter. Assuming that the total number of trips generated in each zone  $T_u$  is known, the trip distribution matrix,  $T_{uv}$  can be computed as [17]

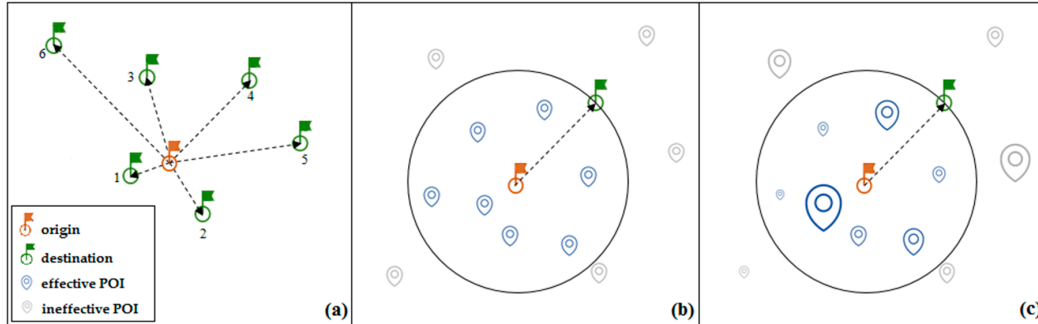
$$T_{uv} = T_u \frac{\text{rank}_u(v)^{-\gamma}}{\sum_{k \neq u}^N \text{rank}_u(k)^{-\gamma}} \quad (4)$$

where  $N$  is the total number of zones in the city.

A rank-based model is more similar to the well-known gravity framework than an Intervening Opportunities model, as they commonly use distance to compute the rank. However, in contrast to the gravity model, the cost variable is considered as rank-distance rather than spatial distance. Since people and their behaviour, as dynamic components in mobility, are neglected in this approach (i.e., using distance alone to rank the zones), the resulting mobility patterns always remain unchanged. To tackle this issue, our paper proposes three methods to compute the ranks in the city using rank-distance and LBSN data. These methods are as follows:

1. First, we implemented the model using the common method of rank computation (i.e., computing the rank as rank-distance). This is more or less similar to the assumptions made in the gravity model. In this approach, the resultant distribution is not subject to change, because the distances between regions are constant. Figure 1a represents the schematic illustration of this scenario.
2. Second, we employed the number of venues located in a circle centered at the origin with a radius equal to the distance between the origin and destination to compute the ranks. This way of computing the ranks is similar to that of an intervening opportunities model. With such an approach, resulting patterns are more dynamic. This is because new venues are created, some venues may shut down, or may be transferred to other regions. This scenario can be seen in Figure 1b.
3. The third approach is to use check-in data for weighting the venues. Thus, humans have participated in the modeling, because venues are not the same in terms of mobility. For instance, cinemas attract much more people, in a certain period of time, than hotels. Therefore, since the amount of importance of venues, in terms of mobility are not the same, the relative importance

should be taken into account. Therefore, we weighted each venue using check-ins occurred at that venue, so that more dynamicity can be reflected in the resultant patterns. The schematic view of this scenario is shown in Figure 1c.



**Figure 1.** The schematic view of the proposed methods of computing the rank. (a) rank-distance method computes the ranks based on mere physical distances between origin and destination; (b) POI-derived rank computation in which the number of POIs between origin and destination is utilized; and (c) check-in-weighted POI-derived rank computation method in which each POI is weighted based on check-ins.

Since the rank-based model is parameterized, a calibration procedure is needed to determine the adjustable parameter. This parameter serves, for geographers and planners, as a context and explanatory power in the model, so that different conditions of a study area can be taken into account. In this paper, due to its higher efficiency and simple implementation, the method introduced by Hyman was employed to determine the parameter [26]. The Hyman method tries to minimize the difference between the real average travel distance and modeled average travel distance in a repetitive manner [27]. In fact, the following equation should be minimized.

$$E(\gamma) = |\bar{r}(\gamma) - \bar{r}| = \left| \frac{\sum_i \sum_j T_{ij}(\gamma) r_{ij}}{\sum_i \sum_j T_{ij}(\gamma)} - \frac{\sum_i \sum_j T_{ij} r_{ij}}{\sum_i \sum_j T_{ij}} \right| \quad (5)$$

where  $\bar{r}$  is the average distance of observed trips and  $\bar{r}(\gamma)$  is the average distance of predicted trips using the parameter  $\gamma$ . Since obtaining a closed-form solution for Equation (5) is not straightforward, it should be minimized in a repetitive manner. One way of solving the above equation is to use the Secant method. Algorithm 1 provides the pseudocode of the Secant method for solving the above equation.

---

**Algorithm 1** Hyman method: pseudocode

---

- 1: Let  $\gamma_0$  (the initial value of the parameter) equal to  $1/\bar{r}$ .
  - 2: Compute the trip distribution matrix using  $\gamma_0$ .
  - 3: Calculate the average distance of predicted trips.
  - 4: Consider the next approximation for the parameter as  $\gamma_1 = \gamma_0 \bar{r}(\gamma_0) / \bar{r}$ .
  - 5: **while true do**
  - 6: Compute the trip distribution matrix using  $\gamma_i$ .
  - 7: Calculate the average distance of predicted trips.
  - 8: **if**  $|\bar{r}(\gamma_i) - \bar{r}| < \varepsilon$  :
  - 9: **break** while loop
  - 10: **else:**
  - 11:  $\gamma_{i+1} = \frac{(\bar{r} - \bar{r}(\gamma_{i-1}))\beta_i - (\bar{r} - \bar{r}(\gamma_i))\gamma_{i-1}}{\bar{r}(\gamma_i) - \bar{r}(\gamma_{i-1})}$
  - 12: **end if**
  - 13: **end while**
-

In addition, since the rank-based model presented in Equation (4) does not guarantee the equality of the real and predicted number of attracted trips, a balancing procedure called Furness is applied on the matrix. Algorithm 2 shows the pseudocode of the Furness algorithm that aims at minimizing the difference between the total number of real and predicted trips.

---

**Algorithm 2** Furness method: pseudocode
 

---

- 1: Compute the trip distribution matrix.
  - 2: Let  $T_i$  and  $T'_j$  be sum of elements in row  $i$  and column  $j$ , respectively.
  - 3: **while**  $|T_i - T'_i| > \varepsilon$  **and**  $|T_j - T'_j| > \varepsilon$  **do**:
  - 4: Multiply each column by  $N_j = T_j/T'_j$ .
  - 5: **if**  $|T_j - T'_j| < \varepsilon$  **and**  $|T_i - T'_i| < \varepsilon$ :
  - 6: **break** while loop
  - 7: **end if**
  - 8: Multiply each row by  $N_i = T_i/T'_i$ .
  - 9: **if**  $|T_j - T'_j| < \varepsilon$  **and**  $|T_i - T'_i| < \varepsilon$ :
  - 10: **break** while loop
  - 11: **end if**
  - 12: **end while**
- 

Figure 2 illustrates the steps required to apply the model. First, it is required for the study area to be partitioned into smaller parts, called zones. In addition, in order to compute the proposed ranking schema, information about location of venues and check-ins are required. Using these information, the zones are ranked. Then, the raw trip distribution matrix is obtained according to Equation (4). To find the best value of the adjustable parameter, the Hyman method is applied on the matrix. Also, the Furness method is applied on the matrix to balance it. Finally, the model is evaluated based on various criteria. If the results of evaluation are acceptable, the model along with the determined parameter can be employed to compute the distribution of mobility flow within the study area.

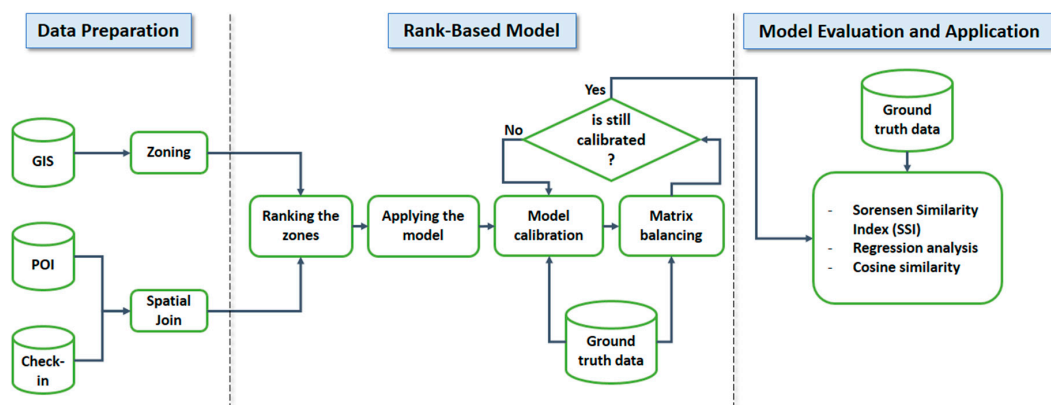
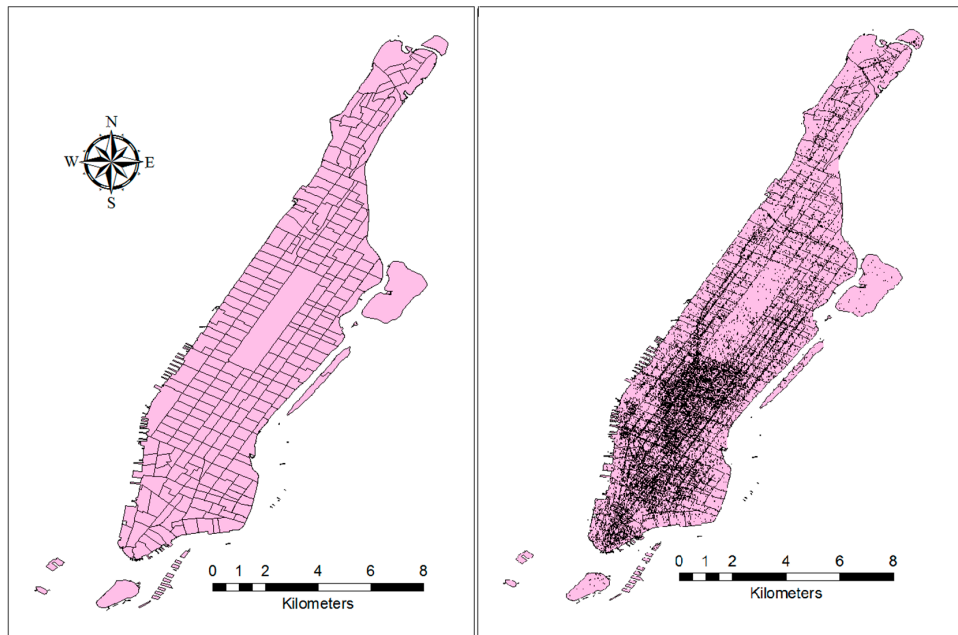


Figure 2. A diagrammatic view of the steps of the methodology.

### 3. Study Area and Datasets Used

In this paper, we considered a census tract reference map of Manhattan as trip production and attraction zones. Manhattan is one of the most densely populated areas in the United States and the rate of mobility is, then, very high. Being one of the world's major commercial and financial centers, Manhattan encounters a significant number of daily travels toward itself. In addition, thanks to the networks of tunnels, bridges, railroad lines, and subways that link Manhattan to the surroundings, there is an enormous influx of daily commuters from neighboring boroughs, including The Bronx, Brooklyn, Queens, and Staten Island, and even from New Jersey, Connecticut, and NYC suburbs such

as White Plains and Long Island who are surely making various trips within Manhattan throughout the day [28]. Thus, modeling the mobility pattern within intra-urban areas such as Manhattan is among the challenges that the spatial interaction models are facing [8,29]. There are 288 zones in the 2010 census tract reference map of Manhattan. In addition to the GIS map of the study area, we used a dataset which includes long-term (about 18 months from April 2012 to September 2013) global-scale check-in data collected from Foursquare through its Twitter Application Programming Interface [30]. It contains about 33 million check-ins on about three million venues worldwide. The location of each venue within Manhattan was extracted from this dataset and mapped on the reference map. The distribution of venues over census tracts are shown in Figure 3.



**Figure 3.** The census tracts of Manhattan (left), and the distribution of POIs over it (right).

In order to evaluate the results obtained from models and to compare the rank concepts, GPS traces of taxi vehicles over Manhattan were used. The datasets were collected and provided to the New York City (NYC) Taxi and Limousine Commission (TLC) under the Taxicab and Livery Passenger Enhancement Programs (TPEP/LPEP) (refer to Taxi and Limousine Commission website at <http://www.nyc.gov/tlc>). The yellow and green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, passenger counts, and some other information about payment and rate types. Although taxi data are not representative of the whole mobility within a city, they can present it to some extent [17]. Usually, finer resolution datasets such as taxi data suffer from having many zero-counts. In order to improve the completeness of evaluation data, we combined yellow and green taxi traces. Since taxis are allowed to pick up passengers from other boroughs of NYC, only trips originating from (and ending in) Manhattan were considered. Figure 4 visualizes the sparsity pattern of the ground truth matrix of the trip distribution.



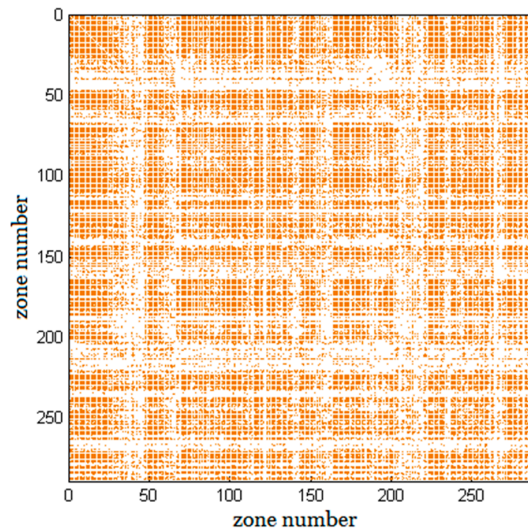


Figure 4. The sparsity pattern of GPS traces of taxi vehicles over Manhattan.

#### 4. Results and Discussion

Following the research conducted by [31], the Sorensen Similarity Index (SSI), also known as Sorensen-Dice Index, was used as a measure of similarity between real and predicted trip distribution matrices. This index ranges from 0 to 1, where numbers closer to one indicate more similarity between two matrices. Figure 5 presents a comparison among the performances of different rank concepts in the model based on SSI. The Sorensen Similarity Index is shown in the following equation:

$$SSI = \frac{2 \sum_i^N \sum_j^N \min(T_{ij}, T'_{ij})}{\sum_i^N \sum_j^N T_{ij} + \sum_i^N \sum_j^N T'_{ij}} \tag{6}$$

where  $T_{ij}$  and  $T'_{ij}$  are  $i$ th and  $j$ th element of real and predicted trip distribution matrices and  $N$  is the total number of zones in the study area.

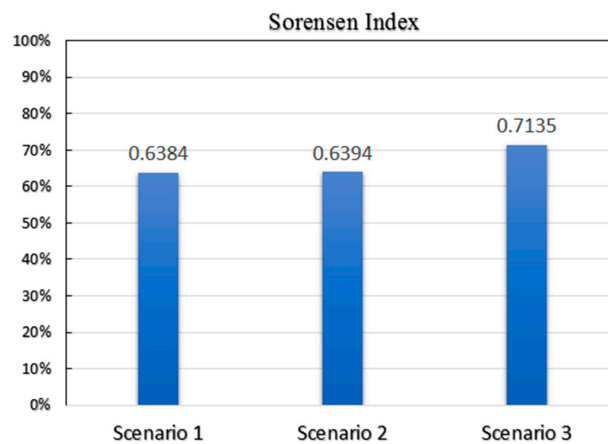
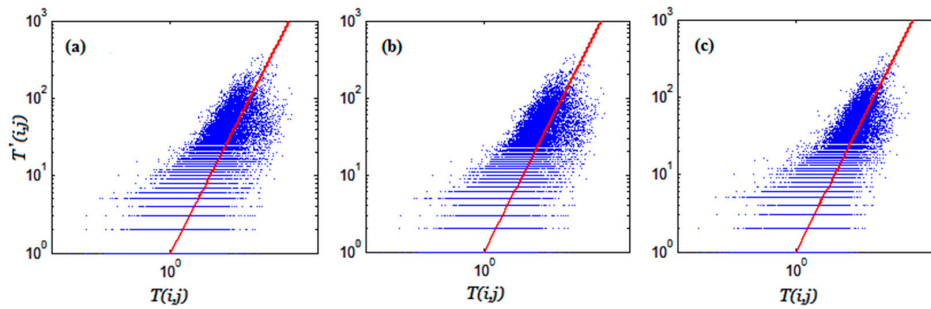


Figure 5. The performance of each ranking method in the model in terms of SSI.

Figure 5 indicates that employing opportunities (venues) rather than distance in the rank-based model is not effective. However, according to the SSI, a check-in-weighted ranking schema will result in predictions that are more similar to reality.

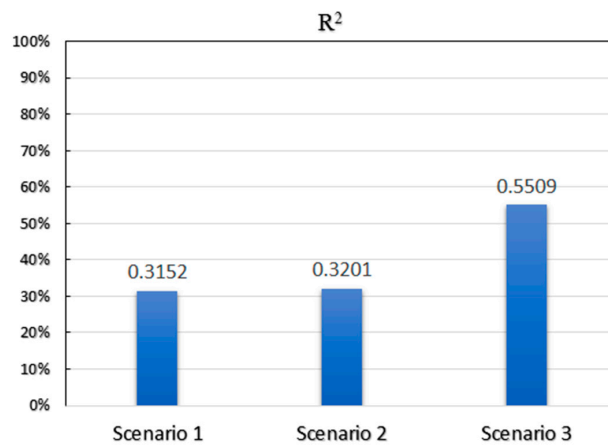
The scatter plot of predicted trips against observed trips for each scenario is shown in Figure 6. This plot shows the relationship between two matrices. The points in the plot correspond to individual

elements of trip distribution matrices. The values on the vertical and horizontal axes are from real and predicted trip distribution matrices, respectively, and the plot is on a logarithmic scale. The red line is the identity line ( $y = x$ ) where predicted number of trips are equal to the real number of trips. Obviously, the more the two matrices agree, the more the scatters tend to concentrate in the vicinity of the identity line.



**Figure 6.** The scatter plot of predicted trips against observed trips for (a) using rank-distance, (b) using POI-derived ranks, and (c) using check-in-weighted POI-derived ranks.

As can be seen from the above results, the difference between the two first scenarios is quite negligible. However, the scatter on the left is much narrower than the others and concentrated about the red line. To have a quantitative measure of how close the predicted numbers of trips are to the ground truth data, we determined  $R^2$  values from the regression analysis applied on scatter plots of Figure 6. Figure 7 compares the results in terms of  $R^2$ .



**Figure 7.** The performance of each ranking method in the model in terms of R-squared obtained from regression analysis of Figure 6.

Again, the difference between using rank-distance and number of venues is not remarkable. However, as Figure 7 illustrates, the value of  $R^2$  for the model, along with a check-in-weighted rank concept, that has been dramatically increased.

In order to analyze the results in a more detailed manner, we computed the cosine similarity between matrices at the zone level, rather than the city level. In fact, we partitioned the matrices row- and column-wise. Then, the cosine similarities between corresponding rows and columns were computed. In other words, each row (column) is considered to be a vector in a 288-dimensional space (i.e., the dimension of space is equal to the number of zones). Now, if the angle between this vector and the corresponding vector of the ground truth matrix in the space is equal to  $0^\circ$ , there will be a complete (1) similarity. On the other hand, if two vectors are in opposite orientations, the value of index will be



–1. Since the trip distribution matrix is a non-negative matrix, the index practically ranges from 0 to 1, where 0 and 1 refer to perpendicular and parallel vectors, respectively. Figure 8 shows the frequency histograms of cosine similarities for rows and columns.

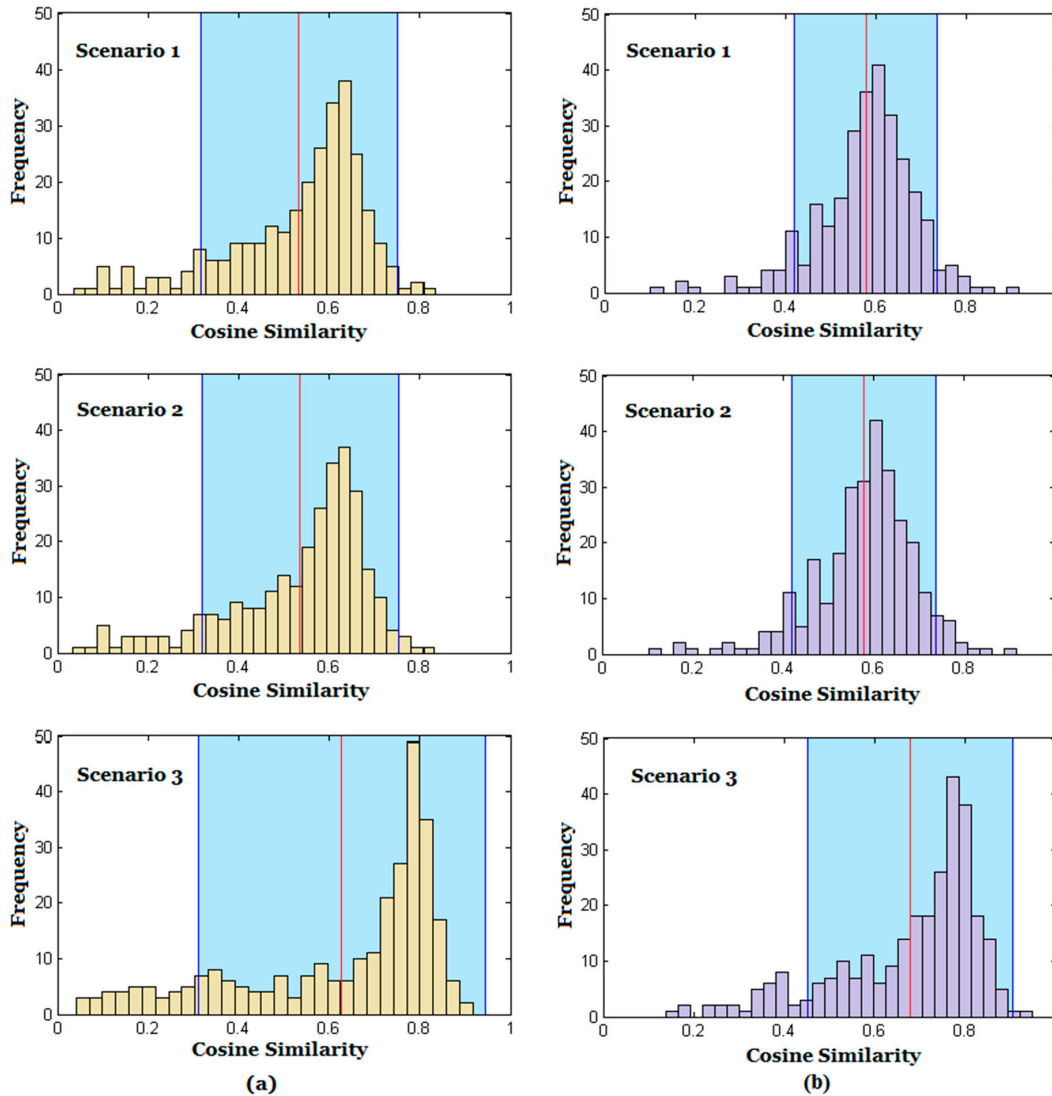


Figure 8. Frequency histograms of cosine similarities for (a) rows and (b) columns

The red line refers to the mean value of the histogram ( $\mu$ ) and the blue bounds show the interval  $\mu - \sqrt{2}\sigma$  and  $\mu + \sqrt{2}\sigma$ , where  $\sigma$  is the standard deviation. According to Chebyshev’s inequality, at least 50 percent of values lie within the blue area. The mean value of the histogram in the case of Scenario 3 is, again, higher than that of the others. In addition, the blue area has been forwarded toward higher similarities.

The reason for obtaining similar results for the two first scenarios can be attributed to the fact that as the radius of the circle between origin and destination is increased, the number of venues within this circle is also increased. As the rank-based model does not take into consideration the magnitude of differences between ranks, the resultant trip distributions will be similar for similar rank distributions. For example, in scenario 2, the number of venues located within each circle is counted. Then, these values are sorted out in a matrix and ranks are assigned to each element of the matrix. Clearly, the differences between ranks are always constant and equal to 1. However, the differences between the numbers of venues in the circles vary remarkably. This is a major disadvantage of the model. Although

applying a weighting schema on the venues using check-ins changes the distribution of ranks within the matrix.

Figure 9 shows flow maps illustrating the interactions between Central Park of Manhattan, designated by a green pin, and other zones. Apparent in the upper maps is the fact that the trips in the two first scenarios are regularly distributed. However, the ground truth map indicates irregularities in the distribution which are too complicated to be modeled using a mere physical parameter such as distance. In fact, large flows, in these scenarios, are directed towards closer zones; while the lower left map conducts trips towards southern zones at which commercial and activity centers are located.

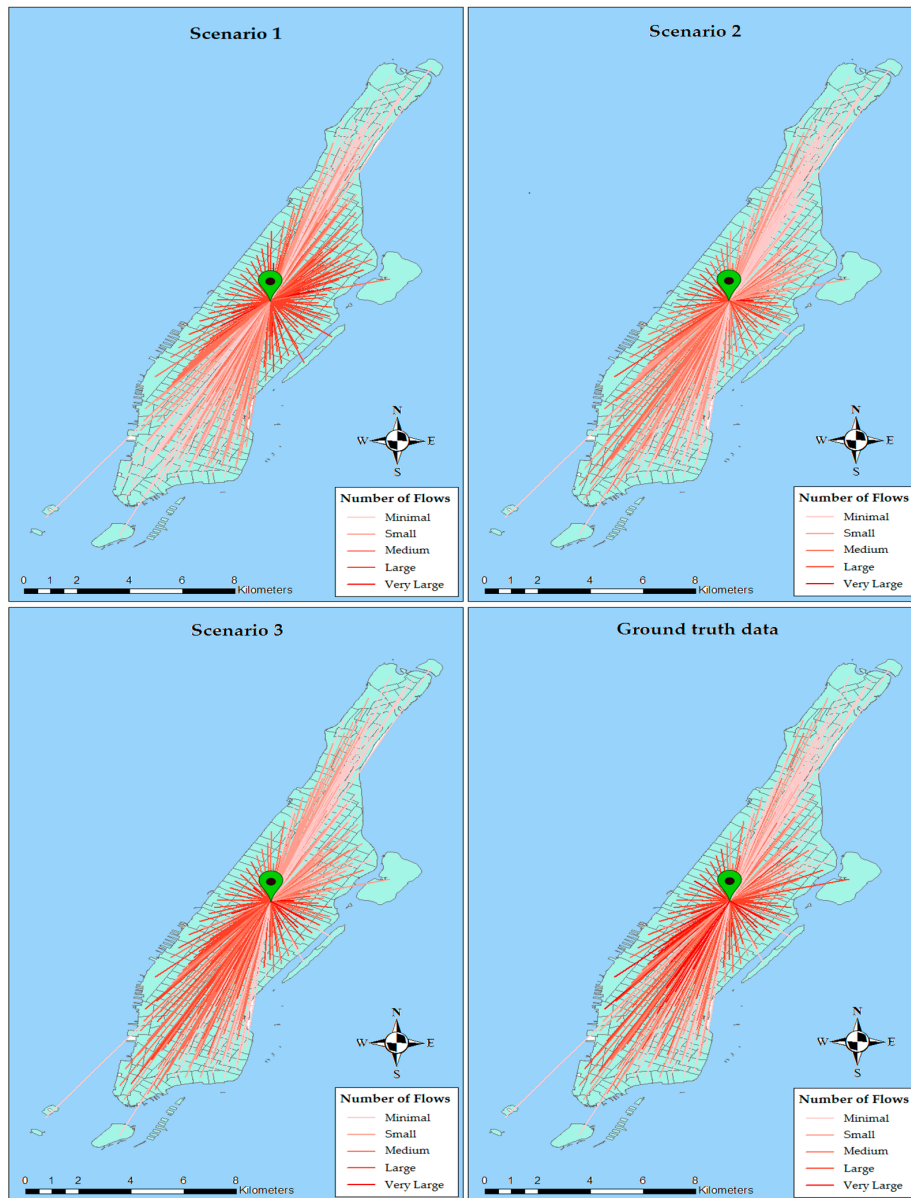


Figure 9. Flow maps showing the distribution of trips originating from Central Park.

### 5. Conclusions

In this paper, we conducted research to study the concept of ranking, in a model of predicting collective human mobility, called a rank-based model. For this purpose, we proposed three scenarios including rank-distance, the number of venues in the region, and a check-in weighted venue schema to compute the ranks in the model.

Results show that the two first scenarios will result in very similar patterns. The reason is that as the distance between origin and destination increases, the number of venues located between them also increases. However, results of applying the third scenario show a remarkable increase in the accuracy of predictions. Therefore, check-ins play a significant role in improving the predictability of mobility patterns.

Considering the rank as distance and number of places of interest (POIs) is, to some extent, objective in the sense that they do not represent real specific conditions of a city. In other words, the concepts of distance and number of POIs are the same for all cities in the world. However, using check-in weights, has some aspects of reality added to the model. Surely, the role of a crowded park in human mobility is not the same as of a hotel, for instance. Thus, applying check-ins occurred at each POI will result in closer predictions to reality.

Moreover, using a check-in weighting schema, the dynamicity of human mobility could be accounted for. As check-in data are dynamic, they can consider the variations in people's interests and behaviors. By using check-ins, any change in land use of POIs is also accountable. In addition, since ongoing events are reflected in check-in data, but not in distance and number of venues in the city, check-in data are particularly useful when events are in progress in the region. The use of check-in data to weigh the POIs implies no particular assumption or restriction on the study area. Since the assumptions made about the ranking method are independent of the study area and its conditions, the conclusions can be potentially generalized to any city. Although employing check-in data in the rank-based model resulted in more promising predictions, the differences between rank numbers is still fixed, and should be addressed in future works. In addition, the temporal dimension of human mobility has not been discussed in this paper. The applicability of the proposed method in spatiotemporal analysis of human mobility can be elaborated on in the future.

In a complex system, such as human mobility, the interactions between its many constituent parts (i.e., people, venues, distances, etc.) determine the properties of the system's behavior (i.e., urban environments). The study of the relationships between the venues, people, and distance deepens our understanding of the urban properties. The usability of these relationships in real world applications, however, needs an accurate model of human mobility prediction. This study better our understanding of human mobility by including dynamic aspects of destination selection in the rank-based model. Moreover, the increase in the accuracy of the resulting patterns using the proposed method also increases the applicability of the model in real world applications.

**Author Contributions:** Omid Reza Abbasi conceived, designed, and performed the experiments; Ali Asghar Alesheikh and Mohammad Sharif revised the methodology and analyzed the results; Omid Reza Abbasi and Mohammad Sharif wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Agryzkov, T.; Martí, P.; Tortosa, L.; Vicent, J.F. Measuring urban activities using foursquare data and network analysis: A case study of Murcia (Spain). *Int. J. Geogr. Inf. Sci.* **2016**, *31*, 100–121. [[CrossRef](#)]
2. Prothero, R.M. Disease and mobility: A neglected factor in epidemiology. *Int. J. Epidemiol.* **1977**, *6*, 259–267. [[CrossRef](#)] [[PubMed](#)]
3. Tizzoni, M.; Bajardi, P.; Decuyper, A.; King, G.K.K.; Schneider, C.M.; Blondel, V.; Smoreda, Z.; González, M.C.; Colizza, V. On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **2014**, *10*, e1003716. [[CrossRef](#)] [[PubMed](#)]
4. Matyas, L. The gravity model: Some econometric considerations. *World Econ.* **1998**, *21*, 397–401. [[CrossRef](#)]
5. Huo, J.; Wang, X.-M.; Hao, R.; Wang, P. Statistical dynamics of regional populations and economies. *arXiv* **2016**, arXiv:1609.00876.
6. Mohammadi, N.; Taylor, J.E. Urban energy flux: Human mobility as a predictor for spatial changes. *arXiv* **2016**, arXiv:1609.01239.

7. Camagni, R.; Gibelli, M.C.; Rigamonti, P. Urban mobility and urban form: The social and environmental costs of different patterns of urban expansion. *Ecol. Econ.* **2002**, *40*, 199–216. [[CrossRef](#)]
8. Kang, C.; Ma, X.; Tong, D.; Liu, Y. Intra-urban human mobility patterns: An urban morphology perspective. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 1702–1717. [[CrossRef](#)]
9. Noulas, A.; Scellato, S.; Lathia, N.; Mascolo, C. Mining user mobility features for next place prediction in location-based services. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012.
10. Buhalis, D.; Amaranggana, A. Smart tourism destinations. In *Information and Communication Technologies in Tourism 2014*; Springer: Berlin, Germany, 2013; pp. 553–564.
11. Zheng, W.; Huang, X.; Li, Y. Understanding the tourist mobility using gps: Where is the next place? *Tour. Manag.* **2017**, *59*, 267–280. [[CrossRef](#)]
12. Zipf, G.K. The  $p \propto 1/d$  hypothesis: On the intercity movement of persons. *Am. Sociol. Rev.* **1946**, *11*, 677–686. [[CrossRef](#)]
13. Stouffer, S.A. Intervening opportunities: A theory relating mobility and distance. *Am. Sociol. Rev.* **1940**, *5*, 845–867. [[CrossRef](#)]
14. Liu, Y.; Sui, Z.; Kang, C.; Gao, Y. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE* **2014**, *9*, e86026. [[CrossRef](#)] [[PubMed](#)]
15. Yuan, Y.; Medel, M. Characterizing international travel behavior from geotagged photos: A case study of flickr. *PLoS ONE* **2016**, *11*, e0154885. [[CrossRef](#)] [[PubMed](#)]
16. Noulas, A.; Scellato, S.; Lambiotte, R.; Pontil, M.; Mascolo, C. A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE* **2012**, *7*, e37027. [[CrossRef](#)]
17. Yan, X.-Y.; Zhao, C.; Fan, Y.; Di, Z.; Wang, W.-X. Universal predictability of mobility patterns in cities. *J. R. Soc. Interface* **2014**, *11*, 20140834. [[CrossRef](#)] [[PubMed](#)]
18. Liang, X.; Zhao, J.; Xu, K. A general law of human mobility. *Science* **2015**, *10*, 1–14. [[CrossRef](#)]
19. Chen, W.; Gao, Q.; Xiong, H.-G. Uncovering urban mobility patterns and impact of spatial distribution of places on movements. *Int. J. Mod. Phys. C* **2016**. [[CrossRef](#)]
20. Lu, X.; Yu, Z.; Sun, L.; Liu, C.; Xiong, H.; Guan, C. Characterizing the life cycle of point of interests using human mobility patterns. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016.
21. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and pois. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012.
22. Fan, Z.; Song, X.; Shibasaki, R. Cityspectrum: A non-negative tensor factorization approach. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014.
23. Niu, H.; Liu, J.; Fu, Y.; Liu, Y.; Lang, B. Exploiting human mobility patterns for gas station site selection. In Proceedings of the International Conference on Database Systems for Advanced Applications, Dallas, TX, USA, 16–19 April 2016.
24. Fu, Y.; Liu, G.; Papadimitriou, S.; Xiong, H.; Ge, Y.; Zhu, H.; Zhu, C. Real estate ranking via mixed land-use latent models. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015.
25. Long, Y.; Han, H.; Tu, Y.; Shu, X. Evaluating the effectiveness of urban growth boundaries using human mobility and activity records. *Cities* **2015**, *46*, 76–84. [[CrossRef](#)]
26. Celik, H.M. Sample size needed for calibrating trip distribution and behavior of the gravity model. *J. Transp. Geogr.* **2010**, *18*, 183–190. [[CrossRef](#)]
27. Hyman, G. The calibration of trip distribution models. *Environ. Plan.* **1969**, *1*, 105–112. [[CrossRef](#)]
28. Moss, M.L.; Qing, C. The Dynamic Population of Manhattan. New York. Available online: [http://wagner.nyu.edu/rudincenter/publications/dynamic\\_pop\\_manhattan.pdf](http://wagner.nyu.edu/rudincenter/publications/dynamic_pop_manhattan.pdf) (accessed on 20 August 2012).
29. Liang, X.; Zhao, J.; Dong, L.; Xu, K. Unraveling the origin of exponential law in intra-urban human mobility. *Sci. Rep.* **2013**. [[CrossRef](#)] [[PubMed](#)]

30. Yang, D.; Zhang, D.; Chen, L.; Qu, B. Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *J. Netw. Comput. Appl.* **2015**, *55*, 170–180. [[CrossRef](#)]
31. Kang, C.; Liu, Y.; Guo, D.; Qin, K. A generalized radiation model for human mobility: Spatial scale, searching direction and trip constraint. *PLoS ONE* **2015**. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).