Classification of typical Bluetooth OD matrices based on structural similarity of travel patterns- Case study on Brisbane city

Krishna N. S. Behara, Corresponding Author
Smart Transport Research Center
Queensland University of Technology
S Block #720-02, 2 George Street, Brisbane, QLD 4000
Tel: +61-426047487; Email: krishnanikhilsumanth.behara@hdr.qut.edu.au

Ashish Bhaskar
Smart Transport Research Center
Queensland University of Technology
S Block #723, 2 George Street, Brisbane, QLD 4000
Tel: +61 7 3138 9985; Fax: +61 7 3138 1170; Email: ashish.bhaskar@qut.edu.au

Edward Chung
Smart Transport Research Center
Queensland University of Technology
S Block #712, 2 George Street, Brisbane, QLD 4000
Tel: +61 7 3138 1143; Fax: +61 7 3138 1170; Email: edward.chung@qut.edu.au

Word count: 5750 words text + 7 tables/figures x 250 words (each) = 7500 words

Submission Date: 1st August, 2017

ABSTRACT

The structure of a daily Origin-Destination (OD) matrix represents the distribution of travel patterns in terms of number of trips ending into different destinations within a region. However, the daily travel patterns could be significantly different due to different characteristics such as regular working days, weekends, long weekends, public holidays, school holidays and special event days etc. Most of the travel patterns are recurrent in nature and they can be classified into different clusters of typical travel patterns represented by their corresponding typical OD matrices. Among many statistical measures, Structural SIMilarity (SSIM) index is identified as an approporiate statistical measure to classify the typical daily OD matrices based on the similarity of travel patterns. The paper discusses the strengths and practical limitations of state-of-the-art application of SSIM for structural comparison of OD matrices for large scale networks and proposes a new practical approach based on geographical window for using SSIM in transport applications. The SSIM is then used as a proximity measure for clustering that provides basis for the identification of typical daily OD matrices. The proposed approach is tested by a case study on real Bluetooth based proxy OD matrices from Brisbane city, Australia.

Keywords: Bluetooth OD matrix, Stuctural Similarity (SSIM) index, Geographical window, SSIM as Proximity measure, Clustering OD matrices, Travel Patterns analysis, Typical OD matrices, DBSCAN

INTRODUCTION

Origin-Destination (OD) matrix is a way of representing travel demand distributed between different origin and destination pairs across the network over a period of time. While OD matrix plays a significant role in transport planning, it is a complex task to estimate the demand matrix for a large scale network. The actual OD demand is not a direct observable measurement. The traditional way of estimating OD matrix is by updating *a priori* matrix (obtained from scaled-up travel surveys or four-step models) until the deviation between observed and estimated traffic counts is minimum. Travel surveys are time consuming, exhaustive and very expensive and are conducted once in every 5-10 years and hence often outdated for planning application after the survey period.

On the other hand, with the availability of seamless traffic data from advanced data sources such as Bluetooth, Cell phone etc., it is possible to measure travel more directly as compared to traditional survey based approaches. Although these emerging sources do not provide a detailed demographic and contextual information about the commuter trips, they do have high spatial and temporal resolution as compared to travel surveys (*1*). In cities like Brisbane (with over 845 Bluetooth Scanners), Bluetooth data sets are currently used for travel time and speed analysis (*2*). With a good penetration rate and detection layout, Bluetooth detections can also be used to build OD demand matrices (*3-6*).

The underlying structure of an OD matrix travel pattern information is in the form of travel demand distributed to different destinations. The knowledge of structural information of OD matrices is helpful in identifying the differences between travel patterns over different times of the day, or between different daily demands or recurrence of demand patterns over a period of time (*7*). The difference in the travel patterns are attributed to different characteristics of activities distributed spatio-temporally within a region. Generally, characteristics of a typical weekend are different from that of a weekday because non-work oriented trips such as shopping and entertainment occur mostly during the weekends. Also some studies identified that proportion of non-work trips to total trips is much higher in the sub-centers compared to the CBD region (*8*). Not limiting the classification to just weekdays and weekends, travel patterns are also observed to be different during long weekends, public holidays, school holidays and during special events days (such as Ekka, The Royal Queensland Show held for 10 full days). Many questions in regards to travel patterns are intriguing: What are the other significant travel patterns observed besides a typical weekday and weekend? How is a Saturday travel pattern different from that of a Sunday? How different are the travel patterns during major festival days? How close the long weekend's patterns are to a Sunday pattern?

To compare and analyse the travel patterns of different days and answer the above questions, there is a high need for potential statistical measures that can compare OD matrices by accounting the underlying structural information. Among many statistical tools, Structural SIMilarity (SSIM) index has recently gained recognition in computing the structural similarity between OD matrices (*9*). Although SSIM has attained immense popularity in the field of image processing, its practical applications in transportation are yet to be fully explored (*10*). The initial objective of this study is to give a physical meaning to local SSIM values so that the approach is more practical in transport applications. To achieve this, the concept of geographical window is introduced (see page 8). The second objective is to further extend the SSIM application as a proximity measure for clustering OD matrices that provides basis for the identification of typical daily OD matrices. To achieve this, mean SSIM values between typical daily OD matrices are converted into distance values for clustering using DBSCAN algorithm.

This research study is based on zonal Bluetooth based origin-destination (bOD) matrices constructed by spatially aggregating Bluetooth data with Statistical Areas-3 (SA3s) of large scale Brisbane network. Generally, Transport Analysis Zones (TAZ) are aggregations of Statistical Areas (*11*) and any reference to OD matrix should be considered as bOD matrix. The bOD matrices and SA3s are used as proxies for the actual OD matrices and TAZs and by assuming so, it will not have any impact on the findings from the current research. The study is based on realistic assumption that traffic patterns are recurrent in nature (*12*). This assumption helps to classify the corresponding OD matrices into different clusters and identify the typical OD matrices representing each classified pattern. The knowledge attained by this classification is useful for strategic modeling and policy development.

The rest of the paper is organised as follows. The paper introduces the study site and the Bluetooth data used for travel patterns analysis; Traditional statistical measures and Structural SIMilarity (SSIM) index are then introduced, followed by a discussion on the strengths of SSIM as compared to other traditional measures for computing structural similarity of OD matrices. The paper then discusses the practical limitations of existing SSIM approach and introduces the first contribution of the paper: a new practical approach for computing SSIM based on geographical window. Next, the paper discusses its second major contribution: SSIM as a proximity measure for clustering and classifying typical daily Bluetooth OD matrices based on the similarity of travel patterns followed by a brief discussion of results and then conclusion.

STUDY SITE AND DATA

Brisbane City Council (BCC) region is chosen as the study site. Raw Bluetooth data, representing temporal detections of MAC IDs (*2*), is collected by BCC from over 845 Bluetooth MAC Scanners installed along many key corridors and intersections within the BCC region (FIGURE 1). Based on population distribution, BCC region is divided into four Statistical Areas namely SA4, SA3, SA2 and SA1 (order from higher to lower) respectively. Trips identified from Bluetooth detections (*13*) are critical construct for the bOD matrices (of size 845 x 845) at scanner level. The dimensions of bOD matrices are then reduced to 20 x 20 by geographically integrating Bluetooth detections with Statistical Area-3 (SA3) obtained from Australian Bureau of Statistics (*14*). The SA4 zonal information is used for splitting SA3 OD matrices into 25 local geographical windows. In this study, Brisbane East refers to a portion of entire Brisbane East that is equipped with Bluetooth scanners. The data used in this study are from the months of January, February, March, May and August of the year 2016 and December, 2015. December, January and March are chosen to account for School Holidays and Long weekends; February and May are chosen due to continuous Regular Working days and month of August to analyse the travel patterns due to special events (Queensland largest annual event, The Royal Queensland Show (Ekka) for 10 consecutive days during August).

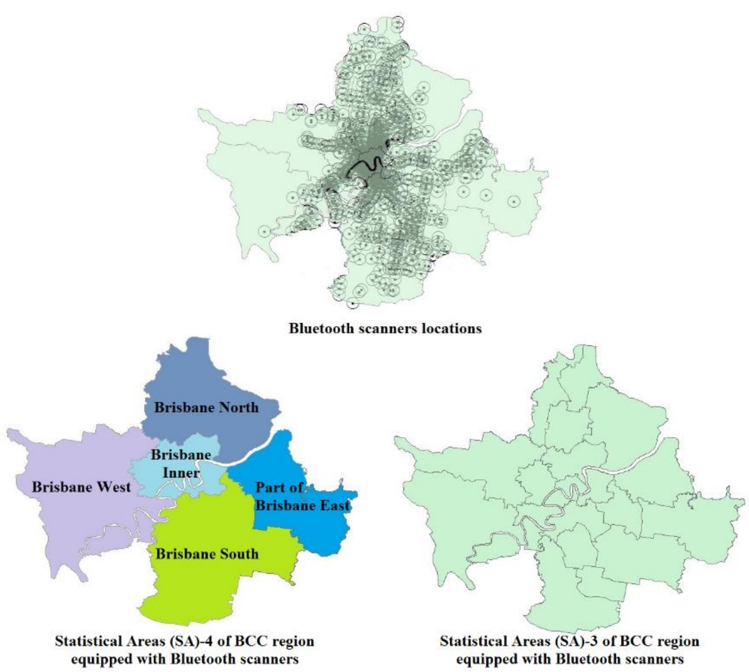


FIGURE 1 Location of Bluetooth MAC Scanners (top) and Statistical Areas-SA4 (bottom left) and SA3 (bottom right) of Brisbane City Council.

STRUCTURAL SIMILARITY (SSIM) INDEX

Several researchers have used traditional measures such as Mean Square Error (MSE), Root Mean Square Error (RMSE) and Maximum Absolute Error (MAE) etc., (*15*) for quantifying distances (differences) between OD matrices. They are widely used because of their simplicity, statistical significance and ease in the optimization process.

The formulation for a few traditional measures, expressed as deviations of OD demands, are as follows:

$$\text{MSE}= \frac{1}{W}\sum_{w\in W}(x_w - \hat{x}_w)^2 \quad (1)$$

$$\text{RMSE}=\sqrt{\frac{1}{W}\sum_{w\in W}(x_w - \hat{x}_w)^2}; \quad (2)$$

$$\text{MAE} = \frac{\sum |x_w - \hat{x}_w|}{W} \quad (3)$$

Where x_w and $\hat{x}_w$represent the values from the w^{th} cell of estimated OD and target/true OD, respectively.

Although most of the traditional measures compute statistical deviations but they lack the ability to capture structural similarity between OD matrices. For example, it is hard to interpret the structural difference in travel patterns from the above equations. In this light, Structural SIMilarity (SSIM) index is identified as an appropriate tool to compute the structural similarity between OD matrices (*9*). This concept had its early appearance in image processing, as a tool to compare two greyscale natural images (*16*). It is computed as the mean of several comparisons of local image patches in both the images. Local image patch from the query image is compared only with its corresponding local image patch from the reference image. Local sliding window is proposed to allow the metric's adaptability to compute local statistical characteristics so that local image distortions were accounted better (*17*). For the first time in transportation, Djukic et al. (*9*) applied Structural SIMilarity (SSIM) as a fitness function within the dynamic OD matrix estimation process and as a performance measure to benchmark various dynamic OD estimation methods. They proposed to re-order the OD matrix and use a sliding window of fixed size (explained with an example in FIGURE 2) or to compute SSIM on the entire OD matrix without any window. The underlying concept of matrix reordering is to deploy the similar rationale - "*neighbourhood pixels are correlated in natural images*" within the context of OD matrix. Generally, the correlations between OD pairs are possible due to sharing similar activities, trip attractions, trip productions, distances, travel cost or similar geographical locations etc. According to Djukic et al. (*9*), correlations between OD pairs are reflected in their demand volumes (especially if volumes are high) and by matrix reordering (i.e. sorting each row of the OD matrix in the order of OD pair volumes), correlated OD pairs lie in the same neighbourhood.

In the existing SSIM application for OD matrices comparison, the local window slides cell by cell over entire OD matrix. For example, consider a 4x4 OD matrix as shown in FIGURE 2. Here, the first and second column represents two OD matrices (OD-1 and OD-2). These two OD matrices need to be compared using SSIM. The local sliding window for computing SSIM is a 2x2 sub-matrix and is represented in coloured cells. Using sliding window of 2x2 sub-matrix there are 9 matrix pairs to be compared in order to achieve overall comparison of OD1 and OD2. These pairs are illustrated from *a* to *l* in FIGURE 2. The local SSIM computes the structural similarity between the sub-matrices corresponding to the windows from both the OD matrices. The final SSIM value, represented as Mean SSIM (MSSIM), is computed by averaging all local SSIM values computed for all the sliding windows. The sliding window is generally a square box of size N x N (where N x N << size of OD matrix).

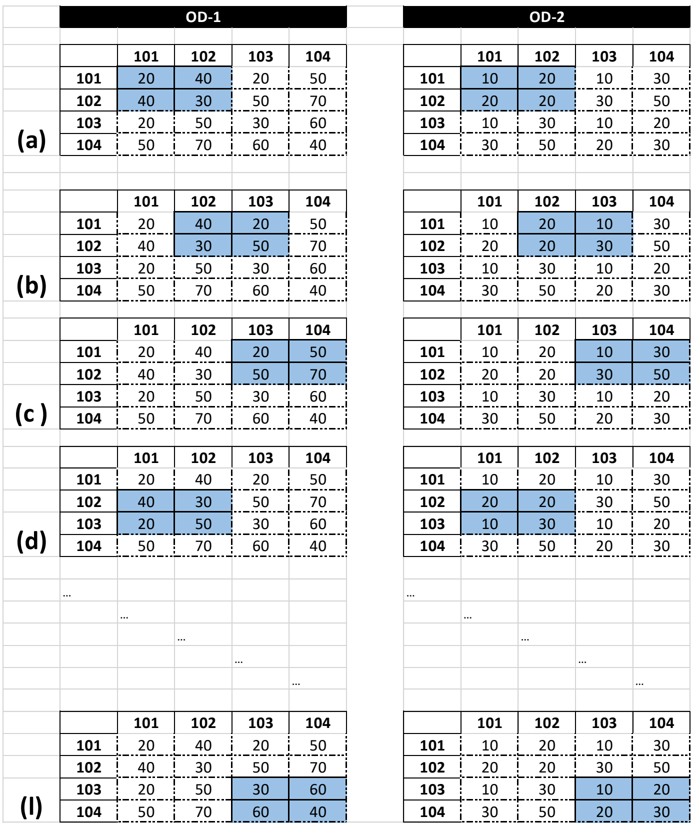


FIGURE 2 An example to demonstrate the sliding window for SSIM calculation.

The formulation for Structural Similarity Index (SSIM) is explained below

$$\mathrm{l}(\mathbf{x},\hat{\mathbf{x}}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)} \tag{4a}$$

$$\mathrm{c}(\mathbf{x},\hat{\mathbf{x}}) = \frac{(2\sigma_x\sigma_{\hat{x}} + C_2)}{(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)} \tag{4b}$$

$$\mathrm{s}(\mathbf{x},\hat{\mathbf{x}}) = \frac{(2\sigma_{x\hat{x}} + C_3)}{(\sigma_x\sigma_{\hat{x}} + C_3)} \tag{4c}$$

$$\mathrm{SSIM}(\mathbf{x},\hat{\mathbf{x}}) = [\mathrm{l}(\mathbf{x},\hat{\mathbf{x}})^{\alpha}]\big[\mathrm{c}(\mathbf{x},\hat{\mathbf{x}})^{\beta}\big][\mathrm{s}(\mathbf{x},\hat{\mathbf{x}})^{\Upsilon}];\ \alpha > 0, \beta > 0 \text{ and } \Upsilon > 0;$$

$$\mathrm{SSIM}(\mathbf{x},\hat{\mathbf{x}}) = \frac{(2\mu_x\mu_{\hat{x}}+C_1)(2\sigma_{x\hat{x}}+C_2)}{(\mu_x^2+\mu_{\hat{x}}^2+C_1)(\sigma_x^2+\sigma_{\hat{x}}^2+C_2)};\ [-1<=\mathrm{SSIM}<=1] \tag{4d}$$

$$\mathrm{MSSIM}(\mathbf{X},\hat{\mathbf{X}}) = \frac{1}{M}\sum_{m=1}^{M}\mathrm{SSIM}(\mathbf{x},\hat{\mathbf{x}})\ ;\ [-1<=\mathrm{MSSIM}<=1] \tag{4e}$$

Where $\mathbf{X}$ and $\hat{\mathbf{X}}$ represent OD matrices OD-1 and OD-2, respectively; $\mathbf{x}$ and $\hat{\mathbf{x}}$ represent the group of OD pairs within local windows in both the matrices.

$\mathrm{l}(\mathbf{x},\hat{\mathbf{x}})$: compares the mean values (μ_x and $\mu_{\hat{x}}$) of group of OD pairs in both matrices

$\mathrm{c}(\mathbf{x},\hat{\mathbf{x}})$: compares the standard deviations (σ_x and $\sigma_{\hat{x}}$) of group of OD pairs in both matrices

$\mathrm{s}(\mathbf{x},\hat{\mathbf{x}})$: compares the structure by computing correlation between normalised group of OD pairs in both matrices. Normalized $\mathbf{x}$ and $\hat{\mathbf{x}}$ with unit standard deviation and zero mean are equal to $\frac{\mathbf{x}-\mu_x}{\sigma_x}$ and $\frac{\hat{\mathbf{x}}-\mu_{\hat{x}}}{\sigma_{\hat{x}}}$, respectively.

C_1, C_2 and C_3: Constants to stabilise the result when either mean or standard deviation is close to zero.

Generally, C_3 is assumed to be $C_2/2$. Previous studies suggest values of 10^{-10} and 10^{-2} for C_1 and C_2 respectively (*10*). However, in this study, they are assumed to be zero as the results are stable.
α, β and Υ : These parameters are used to adjust relative importance of mean, standard deviation and structural components respectively. Generally, they are assumed to be equal to 1.
$\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}})$: Structural Similarity of the group of OD pairs from both matrices. The size of the window over which local SSIM is computed is N x N.
$\mathrm{MSSIM}(\mathbf{X}, \hat{\mathbf{X}})$: Overall similarity of OD matrices, OD1 and OD2, computed by taking average of local SSIM values of M number of windows. For instance, in FIGURE 2, M = 9.
The range of values for SSIM or MSSIM is between -1 and 1. While the OD matrices are exactly the same if the value is 1 and they are extremely dissimilar if the value is -1.

To briefly summarise the differences between SSIM and traditional measures;

1. Traditional measures are expressed as deviations of OD demands (see Equations (1-3)). SSIM is based on three components independent of each other – mean, standard deviation and structural comparisons (see Equations (4a-4e)) to compute overall structural similarity of OD matrices.
2. Traditional indicators, compute statistics on all OD pairs of the OD matrix at a time. SSIM computes on local windows consisting group of OD pairs and considers the average value of local SSIM values as mean SSIM (MSSIM) value.

NEW PRACTICAL APPROACH FOR COMPARING OD MATRICES USING SSIM

Practical limitations of state-of-the art SSIM

Although the concept of SSIM was originally developed in the context of images comparison, the physical meaning of it should be understood more clearly before implementing it into transport applications because the correlation properties of pixels in natural images are different from that of OD pairs within OD matrices. The state-of-the-art application of SSIM in transportation is still theoretical in nature and needs further exploration of its potential in more realistic settings by emphasising on the physical meaning of it, so that it can be applied best in practice (*10*). Also there is no clarity on the acceptable values of SSIM because SSIM is sensitive to two important parameters namely, OD matrix arrangement and window size, as discussed below.

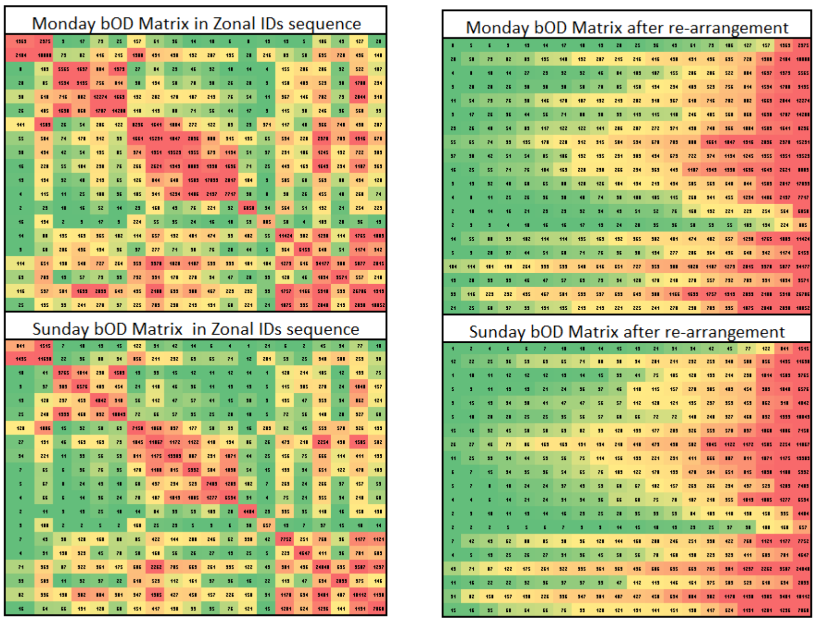
OD matrix arrangement

SSIM is sensitive to the arrangement of OD matrix. For example, MSSIM value is 0.7675 for Monday and Sunday OD matrices arranged in a sequential order of the SA3 zonal ID numbers (see FIGURE 3 (a)); the MSSIM value is 0.6858 if the rows of OD matrix are sorted in ascending order (from left to right) of OD pair demand volumes (see FIGURE 3 (b)).

Djukic et al. proposed matrix re-ordering if sliding window has to be used. However, as discussed earlier, although the concept is borrowed from a different field, its applicability for OD matrices comparison needs to be checked. If two daily OD matrices for a large scale network are individually row sorted based on OD demand volumes then the order of the OD pairs in both the matrices may be different. For example, consider two daily OD matrices from Sunday and Monday with siginificantly different travel patterns (i.e. trips ending into different destinations). FIGURE 3 (c) demonstrates this, with destinations arranged in descending order of the number of trip attractions. The destinations order between **6** and **15** during Monday, is entirely different as compared to that of Sunday. The SA3s “Nundah” and “Sunnybank” are 7th and 14th most attractive destinations on Monday, while they are in 15th and 10th postions during Sunday. Thus matrix re-ordering may not always result in a fair comparison of OD matrices.

Window size

To demonstrate the sensitivity of SSIM with the window size, consider mean SSIM (MSSIM) values computed using different window sizes (3x3 to 20x20) for Sunday-Monday (Blue line) and Tuesday-Monday (Red line) OD matrix pairs as shown in FIGURE 3(d). The example here is demonstrated for a general OD matrix representation i.e. sequential arrangement of zonal ID numbers. It is observed in this study that, larger the size of sliding window, lesser is the sensitivity of SSIM towards fine correlation distortions within the OD matrix. In FIGURE 3(d), x-axis represents the size of the local window and y-axis is the MSSIM value. The MSSIM values increase as the sliding window size increases. The rate of increment of MSSIM values is less for Tuesday-Monday pair as compared to Sunday-Monday pair. This attributes to similar travel patterns between Tuesday-Monday (both of them being working days) as compared to Sunday-Monday pair. Thus, if a sliding window is used, then there is no clear consensus on the level of acceptability of the window size and its corresponding SSIM values.

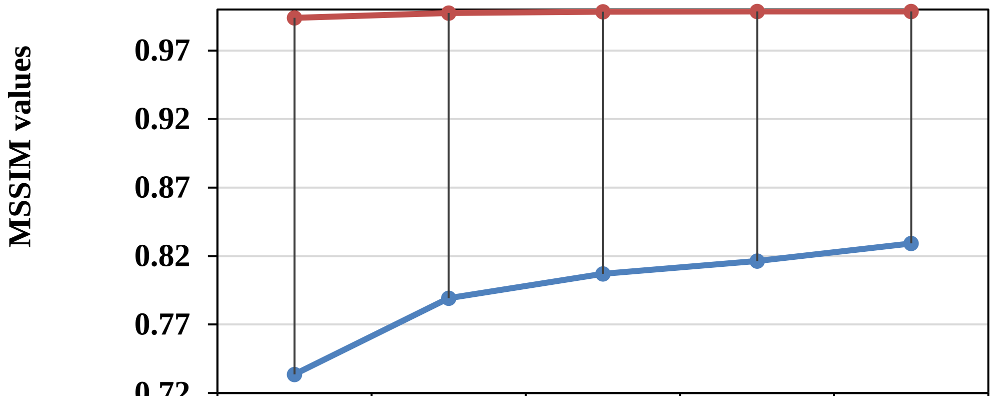


(a) MSSIM=0.7675 (b) MSSIM=0.6858

	Monday Destinations(SA3)	Dest_SA3_ID
1	'Brisbane Inner'	30501
2	'Brisbane Inner - North'	30503
3	'Holland Park - Yeronga'	30302
4	'Mt Gravatt'	30303
5	'Wynnum - Manly'	30103
6	'Rocklea - Acacia Ridge'	30305
7	'Nundah'	30203
8	'Sandgate'	30204
9	'Sherwood - Indooroopilly'	30403
10	'Brisbane Inner - West'	30504
11	'Nathan'	30304
12	'Chermside'	30202
13	'Carindale'	30301
14	'Sunnybank'	30306
15	'The Gap - Enoggera'	30404
16	'Bald Hills - Everton Park'	30201
17	'Brisbane Inner - East'	30502
18	'Centenary'	30401
19	'Capalaba'	30101
20	'Kenmore - Brookfield - Moggill'	30402

	Sunday Destinations(SA3)	Dest_SA3_ID
1	'Brisbane Inner'	30501
2	'Brisbane Inner - North'	30503
3	'Holland Park - Yeronga'	30302
4	'Mt Gravatt'	30303
5	'Wynnum - Manly'	30103
6	'Sandgate'	30204
7	'Sherwood - Indooroopilly'	30403
8	'Brisbane Inner - West'	30504
9	'Carindale'	30301
10	'Sunnybank'	30306
11	'Nathan'	30304
12	'Rocklea - Acacia Ridge'	30305
13	'Chermside'	30202
14	'The Gap - Enoggera'	30404
15	'Nundah'	30203
16	'Bald Hills - Everton Park'	30201
17	'Brisbane Inner - East'	30502
18	'Centenary'	30401
19	'Capalaba'	30101
20	'Kenmore - Brookfield - Moggill'	30402

(c)



	(N=3X3)	(N=6X6)	(N=8X8)	(N=15X15)	(N=20X20)
Sun-Mon	0.7337	0.7892	0.807	0.8164	0.8292
Tue-Mon	0.9939	0.9975	0.9985	0.9986	0.9986

Size of the sliding window

(d)

FIGURE 3 (a) Monday and Sunday OD matrices arranged in sequential order of zonal IDs; (b) Monday and Sunday OD matrices individually row sorted in ascending order (left to right) of OD volumes; (c) Destinations preference for Monday and Sunday; (d) Increase in MSSIM values for Sun-Mon and Tue-Mon OD matrices, with increase in sliding window size.

To address the aforementioned limitations a new approach is proposed to compute SSIM based on geographical window (see below). Firstly by comparing OD pairs from the same geographical window, it is ensured that the arrangement of OD pairs is not disturbed and secondly, there is no question of SSIM sensitivity for different sizes of the window, because the size and shape of the window are defined by the geographical boundaries of higher zonal level (SA4) OD pairs. The following section introduces geographical window based SSIM for comparing OD matrices.

Proposed practical approach- Geographical window based SSIM

The paper proposes a new concept of considering geographical windows before computing SSIM for OD matrices comparison. The geographical window proposed in this study consists of lower zonal level (i.e. SA3) OD pairs belonging to the same higher zonal level (SA4) OD pair ensuring geographical correlation between OD pairs within the window. Since the highest zonal level for BCC region is Statistical Area-4 (SA4), the window boundaries represent geographical boundaries of SA4 OD pair, thus adding physical siginificance to the local window.

The SA4 zones for BCC region are Brisbane East, Brisbane North, Brisbane South, Brisbane West and Brisbane Inner (see FIGURE 1). FIGURE 4 demonstrates the application of SA4 based geographical windows for comparing SA3 (20 x 20) OD matrices of Monday (see FIGURE 4 (a)) and Sunday (see FIGURE 4 (b)), respectively. For example consider a geographical window of SA4 OD pair “Brisbane East” and “Brisbane North”. It consists of SA3 OD pairs i.e. *30101-30201, 30101-30202, 30101-30203, 30101-30204, 30103-30201, 30103-30202, 30103-30203,* and *30103-30204*. These OD pairs are geographically correlated because they have same origin i.e. “Brisbane East” and destination i.e. “Brisbane North” of higher zonal level. Since “Brisbane East” and “Brisbane North” consist of 2 and 4 lower level (SA3) zones respectively, the size of the local geographical window is 2 x 4. It is to be noted that, the geographical window neither has a fixed size nor a fixed shape as it is constrained by the size of the higher level zones. The local SSIM values are then calculated for all geographical windows exclusively and the overall MSSIM is the average of all local SSIM values. For example, MSSIM for Sunday-Monday matrices pair, computed based on geographical window is 0.7231 (see TABLE 1).

By averaging, it implies that, the overall SSIM value is obtained by smoothing over all local values. Although mean SSIM values are used in this study, the local SSIM values based on geographical windows have physical siginificance in their own respects. For example, local SSIM computed for any local window provides valuable insights towards local travel patterns between different suburbs of the region. If the purpose is to compute the similarity of Sunday and Monday travel patterns between major suburbs, then the concept of fixed size sliding window will not work. From FIGURE 4 (c), it can be observed that Sunday travel patterns between the major suburbs "Brisbane South" and "Brisbane North" are less similar to their corresponding patterns on Monday with a local SSIM value of 0.4653 (bold in the TABLE 1). On the other hand, for another major suburb pair- "Brisbane South to Brisbane West", Sunday travel patterns are similar to that of Monday with an overwhelming SSIM value of 0.8037. SSIM values are also justified from visual perception of travel patterns in FIGURE 4 (c), where the first column is the comparison between Brisbane South and Brisbane North pair for Monday and Sunday and the second column is the comparision between Brisbane South and Brisbane West for Monday and Sunday. These insights into different distributions of travel patterns is not possible if a sliding window without any physical meaning are chosen for SSIM computations.

MONDAY

Origin \ Dest		Brisbane East		Brisbane North				Brisbane South						Brisbane West				Brisbane Inner			
		30101	30103	30201	30202	30203	30204	30301	30302	30303	30304	30305	30306	30401	30402	30403	30404	30501	30502	30503	30504
Brisbane East	30101	1369	2375	9	17	79	25	157	61	36	14	18	6	0	19	13	5	106	49	127	20
	30103	2184	18080	79	82	416	215	1308	491	438	192	207	135	20	216	89	58	635	720	496	140
Brisbane North	30201	8	103	5565	1637	804	1973	27	84	29	46	32	10	14	4	155	286	206	32	522	107
	30202	20	85	1534	9195	756	814	38	194	58	78	38	26	20	3	158	483	523	38	1708	234
	30203	90	618	716	882	12274	1669	192	282	170	187	213	76	54	11	367	146	782	73	2044	310
	30204	26	405	1690	868	1787	14200	118	113	80	71	56	44	17	3	115	90	246	36	568	99
Brisbane South	30301	141	1589	26	54	206	122	8296	1641	1004	272	122	89	23	371	117	48	966	740	430	207
	30302	55	504	74	178	312	93	1661	15291	1847	2096	808	315	135	65	594	228	2978	789	1916	678
	30303	38	494	42	54	195	85	974	1951	19523	1355	679	1134	51	37	231	106	1245	192	722	309
	30304	16	220	55	104	238	76	266	2621	1343	8889	1398	1636	71	25	443	163	1643	294	1187	363
	30305	19	194	32	40	219	65	126	844	648	1509	17893	2017	184	9	505	60	563	80	434	120
	30306	4	115	11	25	100	36	105	341	1234	1486	2137	7717	38	8	90	26	455	40	260	74
Brisbane West	30401	2	29	10	16	52	14	23	160	49	76	221	32	6050	34	564	51	192	21	254	223
	30402	16	134	2	3	17	3	224	55	35	24	16	10	53	805	50	4	103	28	36	19
	30403	14	88	135	163	365	102	114	657	192	401	474	99	482	55	11424	382	1230	114	1765	1809
	30404	9	68	286	496	134	96	**37**	277	71	98	76	20	44	5	364	6153	640	51	1174	942
Brisbane Inner	30501	114	651	198	540	727	264	959	3370	1020	1107	539	399	181	104	1273	616	34177	980	5077	2015
	30502	63	783	19	57	73	39	732	931	178	278	94	47	20	39	120	46	1094	3571	557	210
	30503	116	597	501	1693	2099	643	435	2180	633	908	467	223	232	39	1757	1166	5310	593	26786	1919
	30504	25	135	99	241	270	97	225	783	290	219	131	68	221	21	1875	995	2040	219	2090	10052

(a)

SUNDAY

Origin \ Dest		Brisbane East		Brisbane North				Brisbane South						Brisbane West				Brisbane Inner			
		30101	30103	30201	30202	30203	30204	30301	30302	30303	30304	30305	30306	30401	30402	30403	30404	30501	30502	30503	30504
Brisbane East	30101	841	1515	7	10	19	15	122	31	42	14	6	4	1	21	6	2	45	34	77	10
	30103	1495	11690	22	36	88	94	856	211	232	63	65	71	12	201	59	25	340	508	253	90
Brisbane North	30201	10	41	3765	1014	290	1509	13	39	15	12	11	12	14	1	120	214	105	12	199	75
	30202	9	37	989	6576	403	454	21	110	46	36	11	13	13	5	115	305	278	24	1040	157
	30203	19	120	297	453	4042	918	56	112	47	57	41	15	38	3	135	47	353	34	862	121
	30204	25	248	1393	468	892	10043	72	66	57	35	25	20	10	5	72	56	140	20	327	60
Brisbane South	30301	128	1086	15	32	50	63	7150	1068	837	177	58	93	16	289	82	45	553	578	326	133
	30302	27	191	46	163	163	79	1045	11067	1172	1122	410	194	86	26	473	218	2254	498	1505	502
	30303	34	221	11	33	56	53	811	1175	13909	807	291	1071	44	25	156	75	666	114	411	193
	30304	7	65	6	36	76	35	178	1188	815	5392	504	1090	54	15	193	34	651	122	470	109
	30305	5	67	8	24	43	18	60	437	294	529	7403	1289	102	7	263	24	266	37	157	53
	30306	4	66	6	14	36	24	78	187	1013	1085	1277	6534	31	4	75	21	355	34	210	68
Brisbane West	30401	2	11	9	13	25	10	14	84	39	53	103	28	4484	23	395	35	118	16	150	130
	30402	9	108	2	2	5	2	168	25	23	5	9	6	38	657	19	7	37	15	18	14
	30403	7	49	90	120	168	80	85	422	144	200	246	62	330	42	7752	251	768	96	1177	1121
	30404	4	31	190	323	45	78	**50**	160	56	26	27	19	25	5	223	4647	411	36	701	609
Brisbane Inner	30501	71	363	87	322	361	175	606	2262	705	669	261	335	122	49	901	436	24048	635	3507	1297
	30502	39	503	11	32	37	22	618	523	112	161	37	36	16	22	113	47	694	2899	375	146
	30503	82	336	190	982	884	381	347	1385	427	458	157	226	150	31	1178	694	3481	407	18112	1190
	30504	16	64	66	131	120	60	151	417	198	99	35	76	121	15	1201	624	1296	141	1191	7060

(b)

MONDAY					
Origin \ Dest		Brisbane North			
		30201	30202	30203	30204
Brisbane South	30301	26	54	206	122
	30302	74	178	312	93
	30303	42	54	195	85
	30304	55	104	238	76
	30305	32	40	219	65
	30306	11	25	100	36
SUNDAY					
Origin \ Dest		Brisbane North			
		30201	30202	30203	30204
Brisbane South	30301	15	32	50	63
	30302	46	163	163	79
	30303	11	33	56	53
	30304	6	36	76	35
	30305	8	24	43	18
	30306	6	14	36	24
Local SSIM=0.4653					

MONDAY					
Origin \ Dest		Brisbane West			
		30401	30402	30403	30404
Brisbane South	30301	23	371	117	48
	30302	135	65	594	228
	30303	51	37	231	106
	30304	71	25	443	163
	30305	184	9	505	60
	30306	38	8	90	26
SUNDAY					
Origin \ Dest		Brisbane West			
		30401	30402	30403	30404
Brisbane South	30301	16	289	82	45
	30302	86	26	473	218
	30303	44	25	156	75
	30304	54	15	193	34
	30305	102	7	263	24
	30306	31	4	75	21
Local SSIM=0.8037					

(c)

FIGURE 4 Splitting (a) Monday and (b) Sunday SA3 OD matrices into SA4 based geographical windows; (c) Visual representation of difference in local SSIM values.

TABLE 1 Local SSIM values based on geographical windows computed for BCC region

	Brisbane East	Brisbane North	Brisbane South	Brisbane West	Brisbane Inner
Brisbane East	0.8319	0.2437	0.7650	0.9517	0.7755
Brisbane North	0.3311	0.7353	0.4034	0.7378	0.6299
Brisbane South	0.7771	**0.4653**	0.8062	**0.8037**	0.8117
Brisbane West	0.8340	0.7754	0.7562	0.8884	0.8165
Brisbane Inner	0.7716	0.6265	0.8257	0.8385	0.8750
Mean SSIM (MSSIM)					0.7231

TYPICAL BLUETOOTH OD MATRICES CLASSIFICATION

Clustering daily OD matrices using SSIM as proximity measure

In this era of big traffic data, there are many practical applications of data mining and clustering such as clustering trajectories to identify major traffic flow groups in a network level (*18*); clustering transit riders based on travel regularity to enable transit operators in targetting different transit user segments (*19*); clustering historical traffic data to classify traffic profiles for real time traffic management (*20*); and analysing transit riders travel patterns (*21*) etc. Proximity is a general term used to measure the closeness in terms of dissimilarity, distance or similarity between two variables and the threshold proximity value is the key for clustering (*22*).

The study perfomed clustering analysis for 163 days by exploring the inherent potential of SSIM as a proximity measure. Mean SSIM values for 163x163 OD matrix pairs are computed based on the proposed geographical window approach. Before clustering, SSIM values are converted into distance values (ε) using equation (5). Density Based Scanning (DBSCAN) algorithm (*23*) is deployed for identifying different clusters of OD matrices. The two important parameters in DBSCAN algorithm that define the number of clusters and their corresponding sizes are - distance threshold value (ε_T) and minimum number of OD matrices (*Minpts*) in the ε_T neighbourhood of each OD matrix. Based on these two parameters, 163 daily matrices are segmented into three types – core matrices, border matrices and noise. Criteria for OD matrix “q” to become a core matrix is that, the number of its neighbourhood matrices within the threshold value (ε_T) should be atleast equal to ‘*Minpts*’. If the number is less than ‘*Minpts*’, but lies in ε_T neighbourhood of any core matrix, it is called a border matrix. The remaining matrices are categorised as noise. A combination of core matrices within threshold reach (ε_T) forms a cluster.

$$\varepsilon = 1000(1\text{-MSSIM}) \tag{5}$$

The parameters used for clustering analysis are:

1. ε_T: The study initially considers a range of distance threshold values (ε_T) i.e. 40, 35, 30, 25 and 20. Threshold values above 40 and below 20 are not considered because the clusters are not prominent. Before clustering, it is assumed that clusters have few expected characteristics such as regular weekdays, weekends, special-festival days, long weekends, school holidays and public holidays. Among the range of threshold values, ε_T of 20 is recommened in this study (explained in results and discussion section below).
2. $Minpts$: The minimum number of daily matrices to form a cluster is considered to be 2.

Typical daily OD matrices

The purpose of generating typical OD matrices is to identify typical daily travel patterns within the region. A typical OD matrix represents a typical daily travel pattern. One of the naïve ways to infer a typical OD matrix is by considering arithmetic average of all OD matrices within the cluster. For ε_T of 20, seven typical daily OD matrices are representing seven typical travel patterns for BCC region.

RESULTS AND DISCUSSION

TABLE 2 shows different types of clusters for different types of threshold values. Threshold values of 40, 35 and 30 formed less number (3, 5 and 5) of expected clusters as compared to those corresponding to threshold values of 25 and 20 (7 clusters each). Regular Sundays and Saturdays are in one cluster for ε_T value of 25 (see cluster 5). However, they are clearly distinguished as two separate clusters for ε_T value of 20. Thus clusters corresponding to ε_T value of 20 are recommended to classify typical travel patterns within BCC region.

The following are the inferences made from clustering analysis for ε_T value of 20 (bold in TABLE 2).

1. It is interesting to note that, Public Holidays (NewYear , Ekka and Labor Day), Easter and Christmas Long Weekends, School Holidays before-after Christmas and Ekka Sundays have similar travel patterns and form one single cluster (see cluster 1). This cluster has a strategical importance associated with it. For instance, if public holidays are shifted towards weekends, they can form more number of long weekends. This encourages public to enjoy more and spend more via excursions, short-stay holiday trips etc., boosting the nation's economy. For example in Australia, Queen's Birthday (Public Holiday) is always on Monday. In Japan, Public Holidays have already been shifted into Long Weekends as a strategic move to improve nation's ailing economy (*24*).
2. There is no typical weekend travel pattern because Sundays (cluster 2) and Saturdays (cluster 5) are in two separate clusters. Also Sundays and Saturdays amidst of school holidays are not much different from regular Sundays and Saturdays. Australia Day is similar to regular Sunday and different from other Public Holidays.
3. School holidays during normal weekdays (i.e. excluding those before-after Christmas) are not much different from regular weekdays (see cluster 4).
4. Last three school holidays of the year end (i.e. 29th, 30th and 31st of December in cluster 7) are different from the other school holidays as it is a peak holiday time. These days follow long weekend of Christmas and end into a public holiday for the New Year (1st Jan, 2017).
5. Ekka festival, that attracts half a million visitors every August, have entirely different travel patterns. Saturdays (Cluster 3) and weekdays before-after Ekka (Cluster 6) have entirely different travel patterns as compared to regular Saturdays and regular weekdays respectively.

From MSSIM matrix in TABLE 3 (left) the following inferences are made with respect to similarity of travel patterns between typical daily OD matrices:

1. The similarity of travel patterns between Sundays (*Typical OD2*) and Saturdays (*Typical OD5*) is 0.9323;
2. Travel patterns during last three days (weekday school holidays) of December (*Typical OD7*) are similar by 0.9668 to that of regular Saturdays (*Typical OD5*);
3. Regular Sundays (*Typical OD2*) are similar by 0.9735 as compared to the cluster of Public Holidays, Long Weekends, Ekka Sundays and School Holidays before-after Christmas (*Typical OD1*);
4. Ekka Saturdays (*Typical OD3*) are similar by 0.9622 to regular Saturdays (*Typical OD5*);
5. Before-after Ekka weekdays (*Typical OD6*) is close to regular weekdays (*Typical OD4*) with MSSIM value of 0.9601.

The significant differences between typical OD matrices is also validated from the fact that, no distance values in the distance matrix shown in TABLE 3 (right), is less than or equal to the distance threshold value (ε_T) of 20 recommended in this study.

CONCLUSION

This paper proposes a new practical approach to compute SSIM based on geographical window and then explores SSIM as a proximity measure for classifying typical daily Bluetooth OD matrices. SSIM based on geographical window rather than fixed size sliding window is practical oriented. The size and shape of the geographical window are defined by the geographical boundaries adding physical siginificance to local SSIM values. Moreover, it also facilitates indepth investigation of local travel patterns comparison within the region. From the correlation perspective, it accounts for the geographical correlation of OD pairs within the window by ensuring all lower zonal level OD pairs belong to the same higher zonal level OD pair. The zonal consistency is also guaranteed as the matrix is not re-arranged.

SSIM is proposed, for the first time as a proximity measure for clustering and classifying typical daily Bluetooth OD matrices for Brisbane City Council region. Seven different types of typical daily travel patterns are identified and their corresponding daily Bluetooth OD matrices are computed. The study concludes that, there are other significant travel patterns besides typical weekdays and weekends and recognizing them can be strategically important in transport planning.

As a part of future work different methods to classify typical daily OD matrices shall be explored and structural similarity of Bluetooth OD matrices as compared to Houeshold Travel Surveys and Journey to Work shall be evaluated to confirm the acceptability of Bluetooth based OD matrices.

ACKNOWLEDGEMENT

The Bluetooth data used in this study was provided by Brisbane City Council to Smart Transport Research Center, Queensland University of Technology, Brisbane. The conclusions of this paper reflect the understandings of the authors, who are responsible for the accuracy of the data.

TABLE 2 Clusters for different distance threshold values

Cluster No	Clusters for ε_T=40	Clusters for ε_T=35	Clusters for ε_T=30	Clusters for ε_T=25	Clusters for ε_T=20
Cluster 1	Public Holidays (New Year , Ekka and Christmas)	Public Holidays (New Year , Ekka and Christmas)	Public Holidays (New Year , Ekka, Christmas, Good Friday & Easter Sunday) + School Holiday before Christmas	Public Holidays (New Year , Ekka, Christmas, Good Friday & Easter Sunday) + School Holiday before-after Christmas	**Public Holidays (New Year , Ekka, Labor Day) + Long Weekends (Easter, Christmas) + School Holiday before-after Christmas + Ekka Sundays**
Cluster 2	Ekka Sundays	Ekka Sundays	Ekka Sundays	Ekka Sundays	**Regular Sundays + School Holiday Sundays+ Public Holiday (Australia Day)**
Cluster 3	Rest of the days	Public Holidays (Good Friday & Easter Sunday)	Ekka Saturdays	Ekka Saturdays	**Ekka Saturdays**
Cluster 4	NA	Regular Weekdays+ School Holiday Weekdays	Regular Weekdays+ School Holiday Weekdays	Regular Weekdays+ School Holiday Weekdays	**Regular Weekdays+ School Holiday Weekdays**
Cluster 5	NA	School Holiday Sundays and Saturdays + Regular Saturdays and Sundays + Public Holidays (Australia Day, Labor Day, Day After Good Friday & Easter Monday)	School Holiday Sundays and Saturdays + Regular Saturdays and Sundays + Public Holidays (Australia Day, Labor Day, Day After Good Friday & Easter Monday)	School Holiday Sundays and Saturdays +Regular Saturdays and Sundays + Public Holidays (Australia Day, Labor Day, Day After Good Friday & Easter Monday)	**Regular Saturdays + School Holiday Saturdays**
Cluster 6	NA	NA	NA	Before-after Ekka Weekdays	**Before-after Ekka Weekdays**
Cluster 7	NA	NA	NA	Last 3 days of December (School Holidays)	**Last 3 days of December (School Holidays)**

TABLE 3 MSSIM matrix (left) and Distance Matrix (right) for seven typical OD matrices

MSSIM matrix	*Typical OD1*	*Typical OD2*	*Typical OD3*	*Typical OD4*	*Typical OD5*	*Typical OD6*	*Typical OD7*
Typical OD1	1.0000	0.9735	0.6043	0.8771	0.8641	0.6598	0.9181
Typical OD2	0.9735	1.0000	0.6876	0.9189	0.9323	0.7269	0.9501
Typical OD3	0.6043	0.6876	1.0000	0.7942	0.8506	0.9601	0.8019
Typical OD4	0.8771	0.9189	0.7942	1.0000	0.9622	0.8667	0.9512
Typical OD5	0.8641	0.9323	0.8506	0.9622	1.0000	0.8647	0.9668
Typical OD6	0.6598	0.7269	0.9601	0.8667	0.8647	1.0000	0.8359
Typical OD7	0.9181	0.9501	0.8019	0.9512	0.9668	0.8359	1.0000

Distance matrix	*Typical OD1*	*Typical OD2*	*Typical OD3*	*Typical OD4*	*Typical OD5*	*Typical OD6*	*Typical OD7*
Typical OD1	0	26	396	123	136	340	82
Typical OD2	26	0	312	81	68	273	50
Typical OD3	396	312	0	206	149	40	198
Typical OD4	123	81	206	0	38	133	49
Typical OD5	136	68	149	38	0	135	33
Typical OD6	340	273	40	133	135	0	164
Typical OD7	82	50	198	49	33	164	0

REFERENCES

1. Toole, J. L., S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González. The path most traveled: Travel demand estimation using big data resources. *Transportation research part C: emerging technologies,* Vol. 58, 2015, pp. 162-177.
2. Bhaskar, A., and E. Chung. Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. *Transportation research part C: emerging technologies,* Vol. 37, 2013, pp. 42-72.
3. Carpenter, C., M. Fowler, and T. Adler. Generating route-specific origin-destination tables using Bluetooth technology. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2308, 2012, pp. 96-102.
4. Laharotte, P.-A., R. Billot, E. Come, L. Oukhellou, A. Nantes, and N.-E. El Faouzi. Spatiotemporal analysis of Bluetooth data: Application to a large urban network. *IEEE Transactions on intelligent transportation systems,* Vol. 16, No. 3, 2015, pp. 1439-1448.
5. Michau, G., N. Pustelnik, P. Borgnat, P. Abry, A. Nantes, A. Bhaskar, and E. Chung. A primal-dual algorithm for link dependent origin destination matrix estimation. *IEEE Transactions on Signal and Information Processing over Networks,* Vol. 3, No. 1, 2017, pp. 104-113.
6. Barceló, J., F. Gilliéron, M. Linares, O. Serch, and L. Montero. Exploring link covering and node covering formulations of detection layout problem. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2308, 2012, pp. 17-26.
7. Kumar, A., and D. M. Levinson. Temporal variations on the allocation of time. 1995.
8. Gordon, P., A. Kumar, and H. W. Richardson. Beyond the journey to work. *Transportation Research Part A: General,* Vol. 22, No. 6, 1988, pp. 419-426.
9. Djukic, T., S. Hoogendoorn, and H. Van Lint. Reliability assessment of dynamic OD estimation methods based on structural similarity index.In *Transportation Research Board 92nd Annual Meeting*, 2013.
10. Pollard, T., N. Taylor, T. van Vuren, and M. MacDonald. Comparing the Quality of OD Matrices in Time and Between Data Sources.In *Proceedings of the European Transport Conference*, 2013.
11. Australian Transport Assesment and Planning Guidelines. 2016.
12. Ul, S., I. Bajwa, and M. Kuwahara. A travel time prediction method based on pattern matching technique. *Publication of: ARRB Transport Research, Limited*, 2003.
13. Michau, G., A. Nantes, E. Chung, P. Abry, and P. Borgnat. Retrieving trip information from a discrete detectors network: The case of Brisbane Bluetooth detectors. 2014.
14. ABS. *Australian Standard Geographical Classification (ASGC)*. http://www.abs.gov.au/ausstats/abs@.nsf/latestproducts/671F2CC1E1381235CA2574CF00138E27?opendocument. Accessed 30 Nov, 2016.
15. Bera, S., and K. Rao. Estimation of origin-destination matrix from traffic counts: the state of the art. *European Transport,* Vol. 49, 2011, pp. 2-23.
16. Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on image processing,* Vol. 13, No. 4, 2004, pp. 600-612.
17. Brooks, A. C., X. Zhao, and T. N. Pappas. Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions. *IEEE Transactions on image processing,* Vol. 17, No. 8, 2008, pp. 1261-1273.
18. Kim, J., and H. S. Mahmassani. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Procedia,* Vol. 9, 2015, pp. 164-184.
19. Kieu, L. M., A. Bhaskar, and E. Chung. Transit passenger segmentation using travel regularity mined from Smart Card transactions data.In *Transportation Research Board 93rd Annual Meeting*, Washington, D.C., 2014.
20. Barceló, J., L. Montero, M. Bullejos, and M. Linares. A practical proposal for using origin-destination matrices in the analysis, modeling and simulation for traffic management.In *93rd TRB*

annual meeting compendium of papers, 2014. pp. 14-3793.

21. Ma, X., Y.-J. Wu, Y. Wang, F. Chen, and J. Liu. Mining smart card data for transit riders' travel patterns. *Transportation research part C: emerging technologies,* Vol. 36, 2013, pp. 1-12.
22. Everitt, B. S., S. Landau, and M. Leese. *Cluster Analysis*. Arnold Publishers, 2001.
23. Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise.In *Kdd, No. 96*, 1996. pp. 226-231.
24. Chung, E. Classification of traffic pattern.In *Proc. of the 11th World Congress on ITS*, 2003. pp. 687-694.