Imperial College London

BALANCED K-MEANS CLUSTERING ALGORITHM FOR A MULTI-OBJECTIVE PARTITIONING IN THE BIKE REPOSITIONING PROBLEM

T2

Javier Monroig Vives

Department of Civil and Environmental Engineering, South Kensington campus, Imperial College London.

1. INTRODUCTION

The Bike Sharing Schemes (BSS) are inhomogeneous and anisotropic due to the concentration of activities in the different areas of the city, the uneven altitude of the stations and the connection with other means of transport. These asymmetric flows cause the filling and emptying of some stations of the network, which provokes users abortion. This problem in the Santander Cycle Hire (SCH) will be solved with a repositioning fleet of 21 trucks that will redistribute the bikes of the BSS. To simplify the problem, the 400 stations of the SCH will be allocated in clusters, so every repositioning truck will handle the stations of one cluster.

2. DATA

RAW DATA: information about the 3,321,781 trips (November 1st 2010 – May 31st 2011) [Start date; Start day; Start time; End date; End day; End time; Start Station ID; End Station ID; Travel time].

FINDINGS: The flow trends of a random date just depend on whether it is weekday or weekend (Figure 1). Four peak periods were defined:
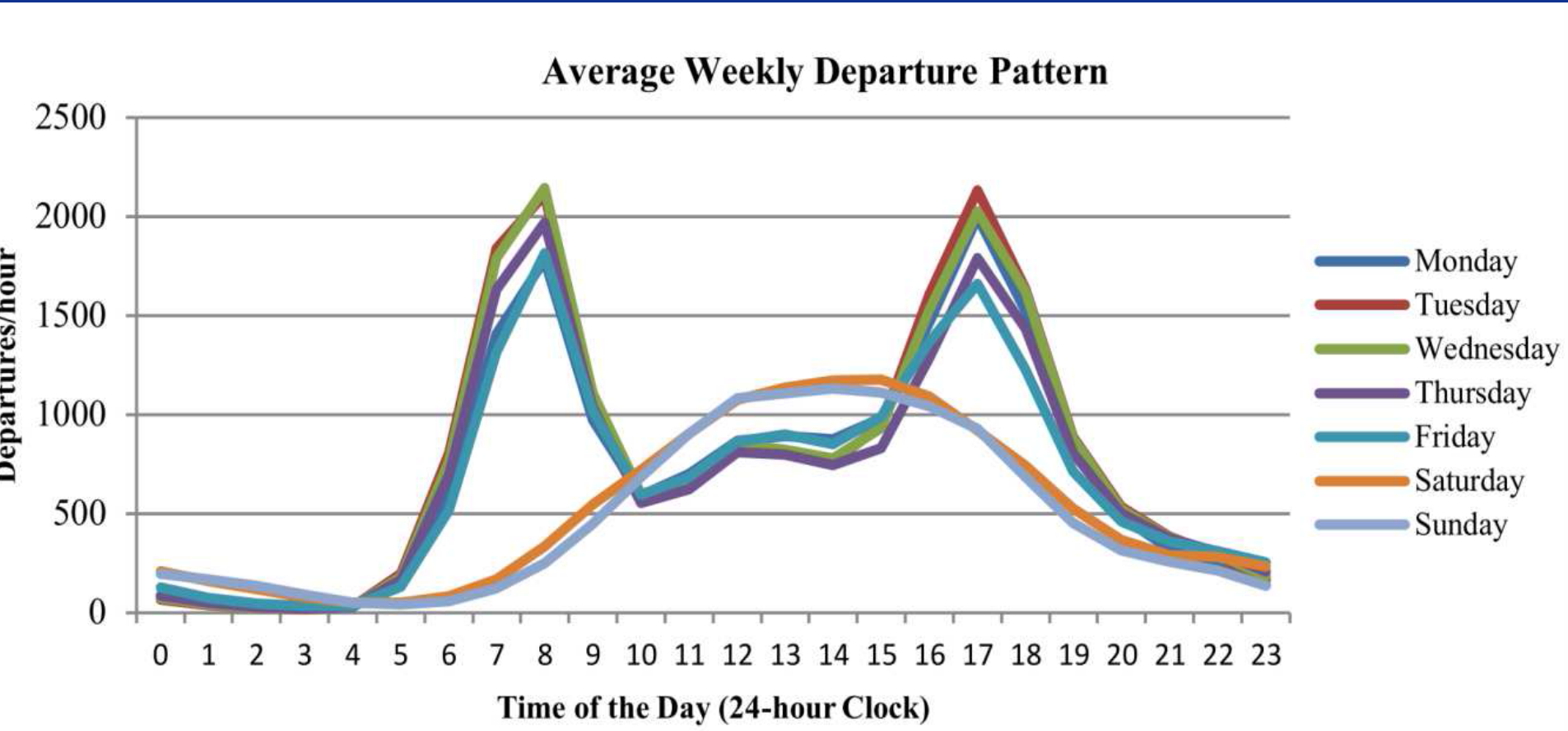
- Morning peak, Weekday (AMWD): 7am – 10am
- Afternoon peak, Weekday (PMWD): 4pm – 7pm
- Morning peak, Weekend (AMWE): 10am – 1pm
- Afternoon peak, Weekend (PMWE): 1pm – 4pm

A table with the net flow of each station for every peak period was created (Table 1).

FIGURE 1. Distribution of trips per day

TABLE 1. Sample of output information

ID	Latitude	Longitude	AMWD flow	PMWD flow	AMWE flow	PMWE flow
1	51.52916	-0.10997	-2.45	0.27	-0.97	-0.97
2	51.49961	-0.19757	-0.44	0.35	-0.56	-0.56
⋮						
411	51.49337	-0.09968	-2.39	2.47	-0.8	0.17
412	51.48811	-0.11012	-2.26	2.37	-0.52	1.37

3. METHODOLOGY: K-MEANS + BORDER ADJUSTMENT

1.K-MEANS CLUSTERING: Minimize distance between all stations and centroids of clusters.

2.BORDER ADJUSTMENT: Balance ClusterSize (number of stations of a cluster) and ClusterFlow (summation of net flow of stations composing a cluster for a peak period).

3.IMPROVEMENT: Creation of 'bigcluster' (cluster composed of the most unbalancing stations of all clusters) and 'delete_st' (most balanced stations are removed from the repositioning).

4 Scenarios considered depending on distance constraint (maxdistclusters):

$$S1: \text{maxdistclusters} = (\text{max longitude} - \text{min longitude})/2 \qquad (1)$$
$$S2: \text{maxdistclusters} = (\text{max longitude} - \text{min longitude})/3 \qquad (2)$$
$$S3: \text{maxdistclusters} = (\text{max longitude} - \text{min longitude})/4 \qquad (3)$$
$$S4: \text{maxdistclusters} = (\text{max longitude} - \text{min longitude})/5 \qquad (4)$$

4. TESTING AND RESULTS (MicroDelos)

TABLE 2. Batches tested

|	P1: AMWD	P2: PMWD	P3: AMWE	P4: PMWE
Batch 1	1-Cluster	1-Cluster	1-Cluster	1-Cluster
Batch 2	K-means	K-means	K-means	K-means
Batch 3	Scenario 1	Scenario 1	Scenario 1	Scenario 1
Batch 4	Scenario 2	Scenario 2	Scenario 2	Scenario 2
Batch 5	Scenario 3	Scenario 3	Scenario 3	Scenario 3
Batch 6	Scenario 4	Scenario 4	Scenario 4	Scenario 4
Batch 7	Genetic alg.	Genetic alg.	Genetic alg.	Genetic alg.
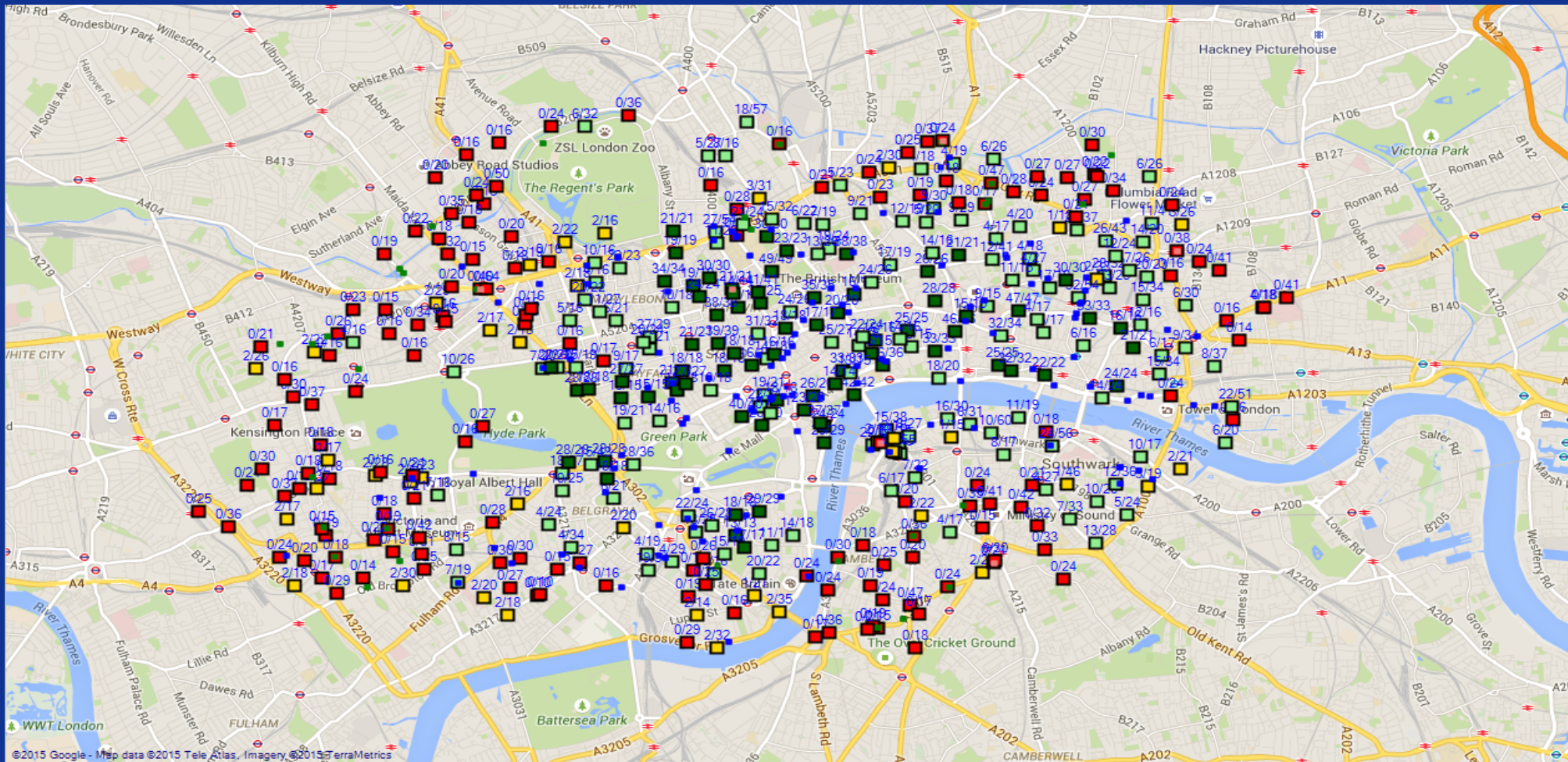
FIGURE 2. Simulator. State of the scheme

The best Scenarios for each period regarding the users aborted and the projected loss are:

AMWD: Batch 4 (Scenario 2)   PMWD: Batch 5 (Scenario 3)
AMWE: Batch 5 (Scenario 3)   PMWE: Batch 5 (Scenario 3)

FIGURE 3.
Users Aborted:
AMWD;   PMWD;
AMWE;   PMWE;
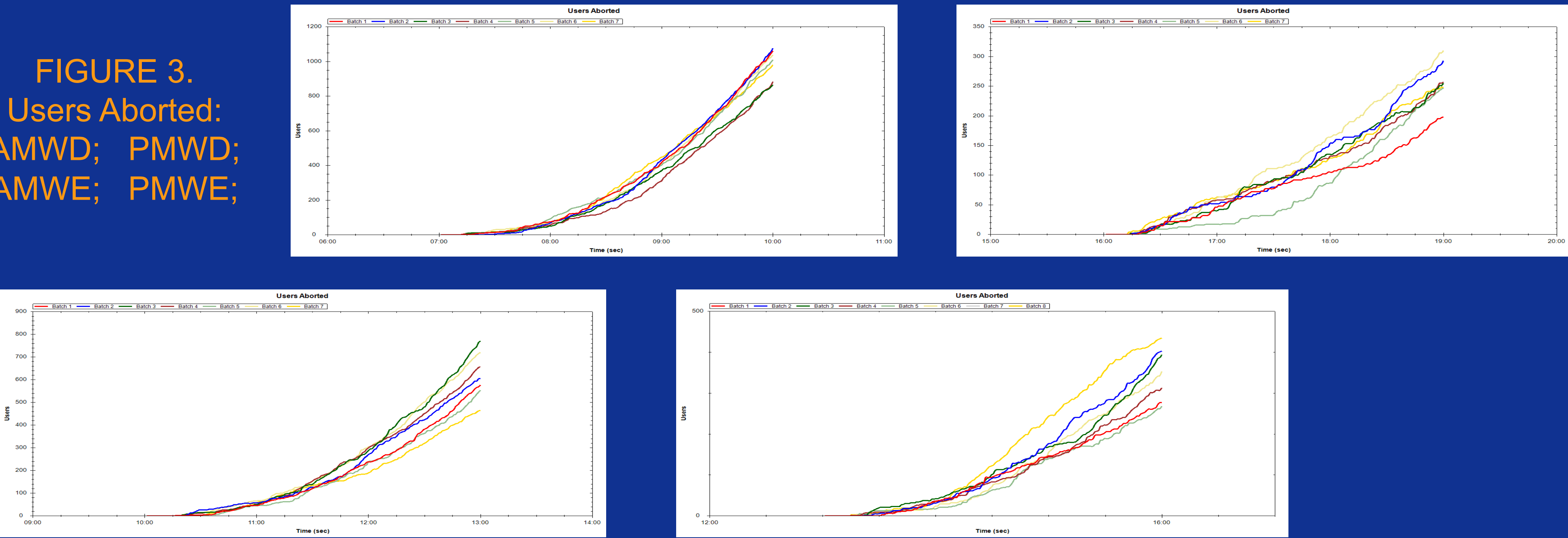
5. CONCLUSIONS

The Border adjustment method improves the K-means clustering algorithm by reallocating stations in other clusters and balancing their size and flow. The simulator Delos.BS provides the evidence of a reduction in that the *Border adjustment* reduced the projected loss and users aborted and improved the repositioning operations over simpler methods like *K-means* algorithm or the *1-cluster* arrangement. The clustering of the *Genetic algorithm* was not satisfactory because the groups of stations were not compact enough. The best scenario is the 3rd in all peak periods but the AMWD, which is the scenario 2.

ACKNOWLEDGEMENTS

REFERENCES

Angeloudis, P., Hu, J. & Bell, M. G. H. (2014) *A strategic repositioning algorithm for bicycle-sharing schemes.*