
电子科技大学
University of Electronic Science and Technology of China

# Trajectory Clustering: A Partition-and-Group Framework

**Wu Rui-zhi**


Data Mining Lab, Big Data Research Center, UESTC
Email：junmshao@uestc.edu.cn
http://staff.uestc.edu.cn/shaojunming

# Outline



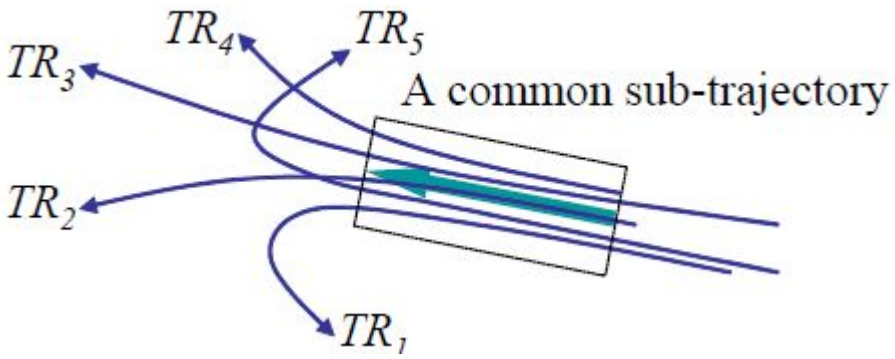
- Why do this things?
- How do this things?

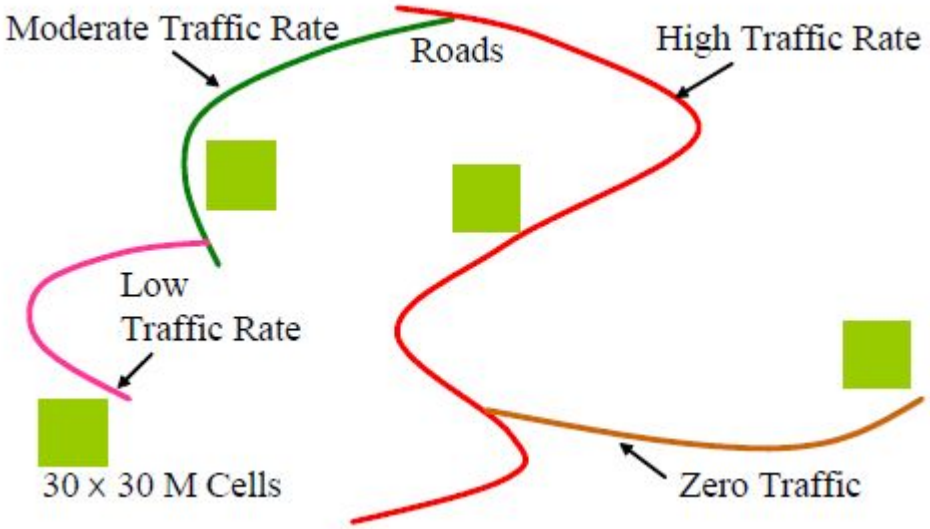
  1、a partition-and-group framework
  2、a formal trajectory partition algorithm using the MDL
  3、a density-based clustering algorithm for line segments(DBSCAN)
  4、discover the representative trajectory

- Experimental evaluation
- Discussion and Conclusions

•Why do this things?



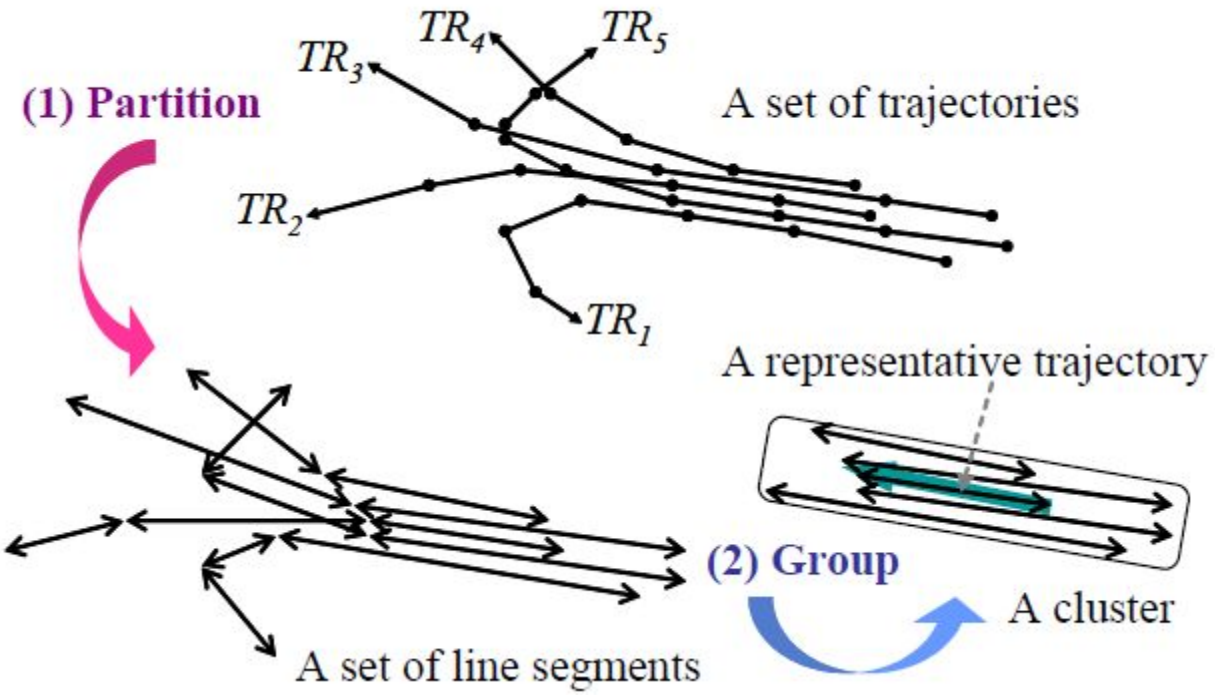


# •How do this things?

•a partition-and-group framework- Basic idea

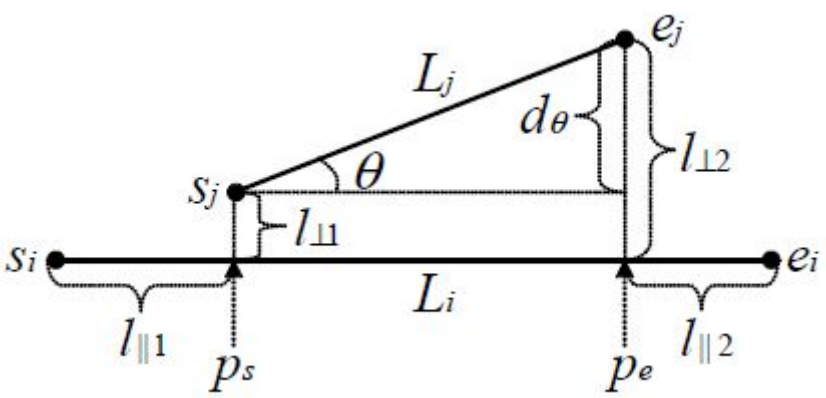




# •How do this things?

•a partition-and-group framework-distance function

数据挖掘实验室
Data Mining Lab




$$d_\perp(L_i, L_j) = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}}$$

$$d_\parallel(L_i, L_j) = \mathrm{MIN}(l_{\parallel 1}, l_{\parallel 2})$$

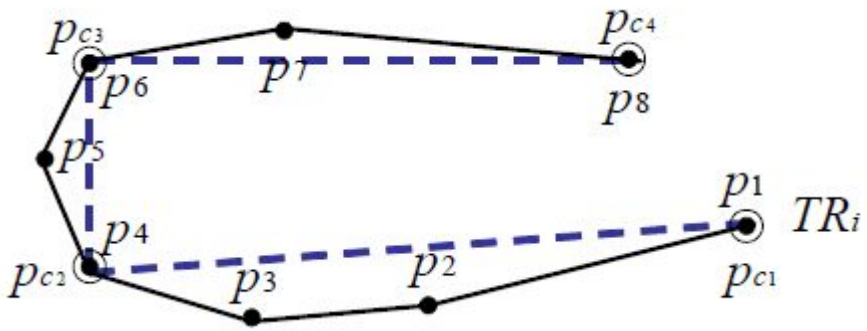
$$d_\theta = \|L_j\| \times \sin(\theta)$$

$$d_\theta(L_i, L_j) = \begin{cases} \|L_j\| \times sin(\theta), & \text{if } 0^\circ \le \theta < 90^\circ \\ \|L_j\|, & \text{if } 90^\circ \le \theta \le 180^\circ \end{cases}$$

# •How do this things?

•a formal trajectory partition algorithm using the MDL -Partition





# •How do this things?

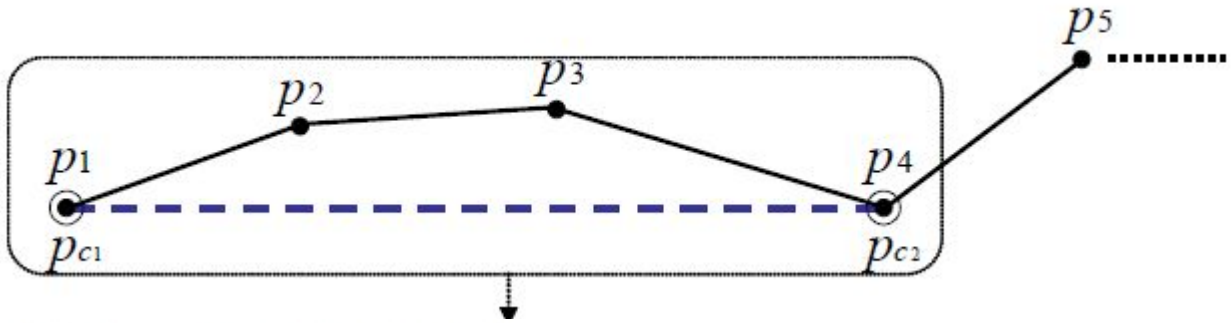
•a formal trajectory partition algorithm using the MDL

数据挖掘实验室 Data Mining Lab

MDL(minimum description length)

$$L(H) = \sum_{j=1}^{par_i - 1} \log_2(len(p_{c_j} p_{c_{j+1}}))$$

$$L(D|H) = \sum_{j=1}^{par_i - 1} \sum_{k=c_j}^{c_{j+1}-1} \{ \log_2(d_\perp(p_{c_j} p_{c_{j+1}}, p_k p_{k+1})) + \log_2(d_\theta(p_{c_j} p_{c_{j+1}}, p_k p_{k+1})) \}$$




$$L(H) = \log_2(len(p_1 p_4))$$

$$L(D|H) = \log_2(d_\perp(p_1 p_4, p_1 p_2) + d_\perp(p_1 p_4, p_2 p_3) + d_\perp(p_1 p_4, p_3 p_4)) + \log_2(d_\theta(p_1 p_4, p_1 p_2) + d_\theta(p_1 p_4, p_2 p_3) + d_\theta(p_1 p_4, p_3 p_4))$$

# •How do this things?

•a formal trajectory partition algorithm using the MDL



MDL(minimum description length)

Definition

$$MDL_{par}(p_i, p_j) = \mathrm{MDLcost}(= L(H) + L(D/H))$$

$$MDL_{nopar}(p_i, p_j) = MDL\cos t(= L(H) + L(D/H)) \qquad L(D/H) = 0$$

$$i \le k \le j$$

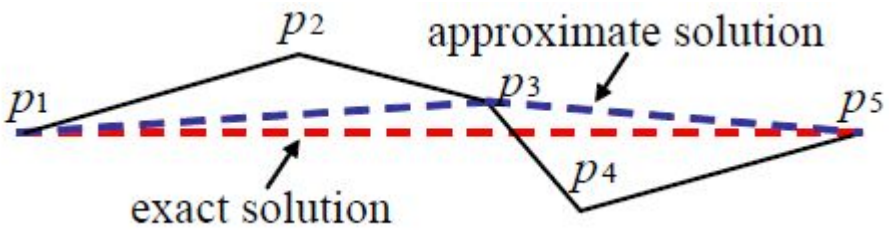
$$MDL_{par}(p_i, p_k) \le MDL_{nopar}(p_i, p_k)$$

$$MDL_{par}(p_i, p_k) \ge MDL_{nopar}(p_i, p_k)$$

# •How do this things?

•a formal trajectory partition algorithm using the MDL



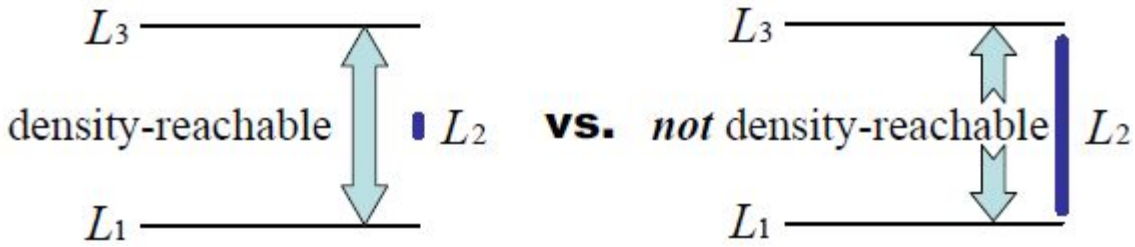



$$MDL_{par}(p_1, p_4) > MDL_{nopar}(p_1, p_4)$$
$$MDL_{par}(p_1, p_5) < MDL_{nopar}(p_1, p_5)$$

# •How do this things?

•a density-based clustering algorithm for line segments(DBSCAN)






$dist(L_1, L_2) = dist(L_2, L_3) < \varepsilon$

$\because d_\theta(L_1, L_2) = d_\theta(L_2, L_3) = \|L_2\| \approx 0$

(a) $L_2$ is very short.

$dist(L_1, L_2) = dist(L_2, L_3) > \varepsilon$

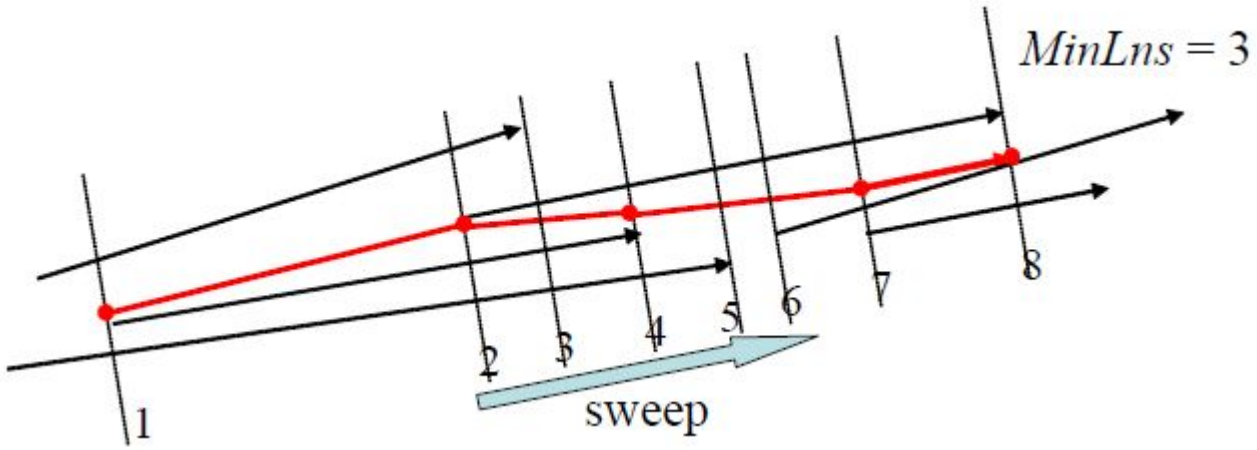
$\because d_\theta(L_1, L_2) = d_\theta(L_2, L_3) = \|L_2\| > \varepsilon$

(b) $L_2$ is not short.

# •How do this things?

•discover the representative trajectory



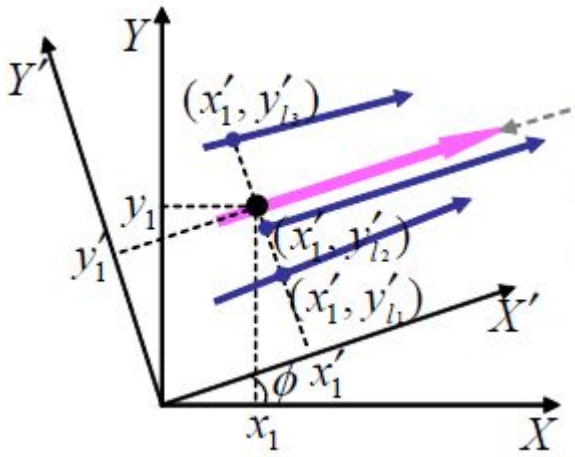


# •How do this things?

•discover the representative trajectory



$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$




average direction vector

average coordinate
in the $X'Y'$ coordinate system

$$(x'_1, y'_1) = (x'_1, \frac{y'_{l_1} + y'_{l_2} + y'_{l_3}}{3})$$

•Experimental evaluation



$$H(X) = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i),$$

$$\text{where } p(x_i) = \frac{|N_\varepsilon(x_i)|}{\sum_{j=1}^{n} |N_\varepsilon(x_j)|} \text{ and } n = num_{ln}$$

$$\begin{aligned} QMeasure &= Total\ SSE + Noise\ Penalty \qquad (11) \\ &= \sum_{i=1}^{num_{clus}} \left( \frac{1}{2|C_i|} \sum_{x \in C_i} \sum_{y \in C_i} dist(x,y)^2 \right) + \\ &\quad \frac{1}{2|\mathcal{N}|} \sum_{w \in \mathcal{N}} \sum_{z \in \mathcal{N}} dist(w,z)^2 \end{aligned}$$

# •Experimental evaluation

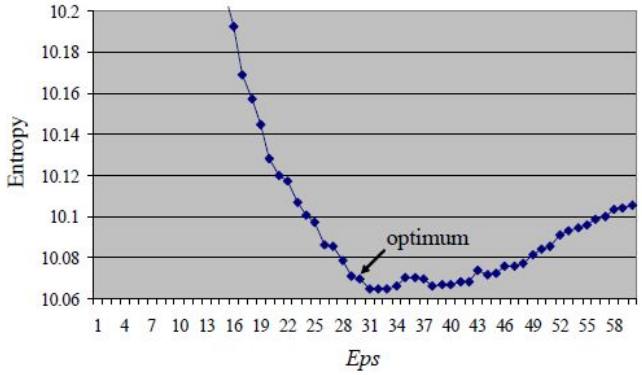



Figure 16: Entropy for the hurricane data.

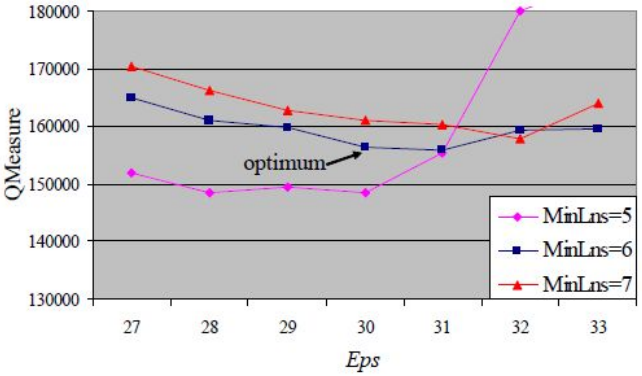



Figure 17: Quality measure for the hurricane data.

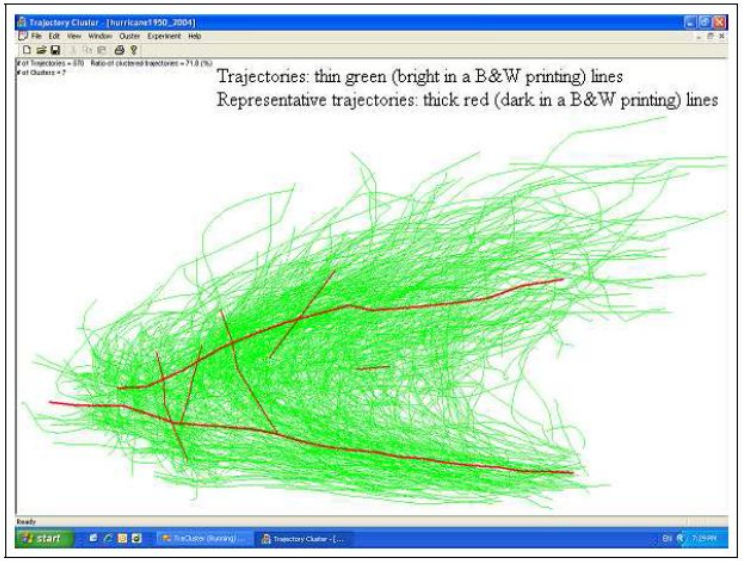



Figure 18: Clustering result for the hurricane data.

- Experimental evaluation

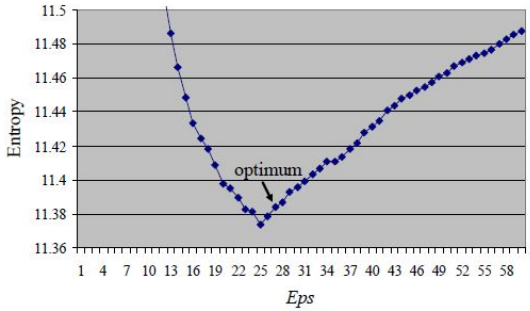



Figure 19: Entropy for the Elk1993 data.

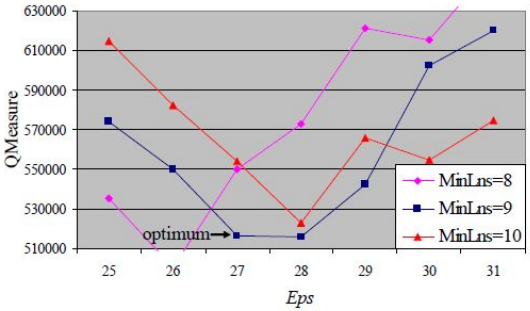



Figure 20: Quality measure for the Elk1993 data.

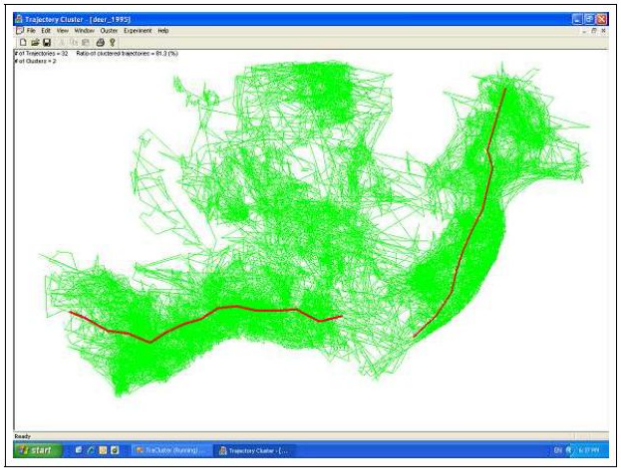


Figure 22: Clustering result for the Deer1995 data.

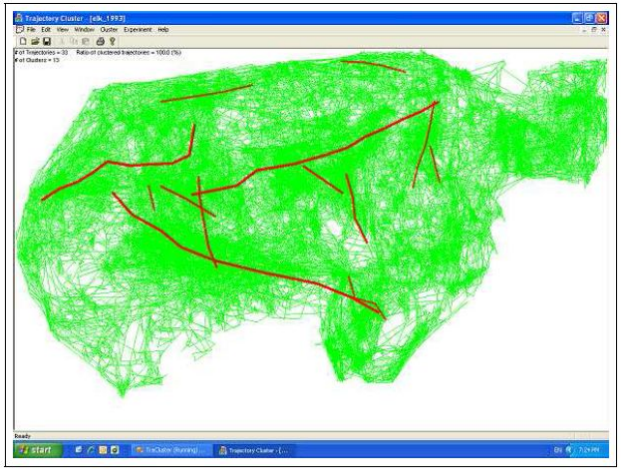


Figure 21: Clustering result for the Elk1993 data.

# Discussion and Conclusions



- partition-and-group framework
- have performed extensive experiments using two real data sets

- support undirected or weighted trajectories.
- extend our algorithm to support various types of movement patterns, especially circular motion.
- extend our algorithm to take account of temporal information during clustering.

# Thank you

# Q and A