Data Mining Lab

Moving Together Pattern

-- an overview

PRESENTED BY ZHAOYI, UESTC

xiaoFine@live.com

Outline

1

Moving Together

- Categories
- Application
- Earlier Models

2

Gathering Pattern

- Improvement
- Implementation

3

Extension Models

- Urban black holes

Part 1 Introduction

Group Moving Patterns

- Company Patterns
- Aggregation Patterns
- Divergence Patterns
- Leadership Patterns
- Popular Patterns
- Mutant Patterns

So……where're Moving Together Patterns?

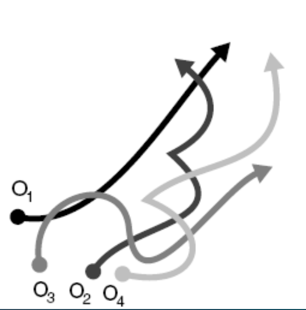
Part 1 Relative Motion Patterns

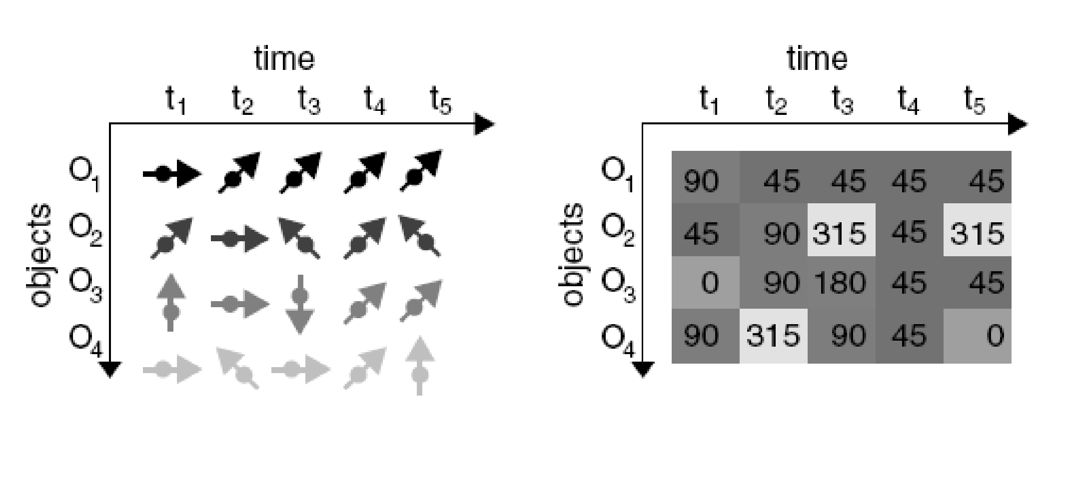
Relative Motion Patterns

- To identify similar movements in a collection of MOPs(moving point objects)
- REMO analysis
  - A transformation of lifeline data to a REMO matrix featuring motion attributes(i.e. speed, acceleration or motion azimuth)
  - Match of formalized patterns on the matrix

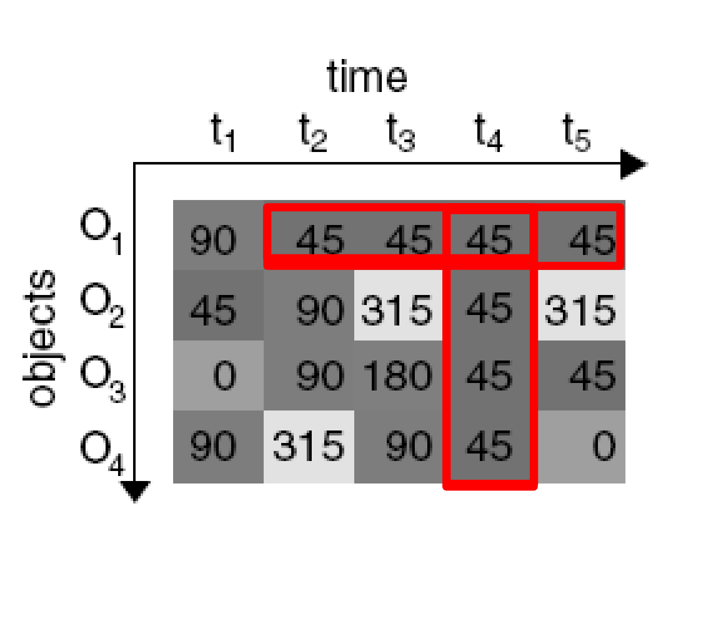
Part 1 Relative Motion Patterns

An example:





Part 1 Relative Motion Patterns



Basic Motion :

Constance: sequence of equal motion attributes for *r* consecutive timestamps

Concurrence: incident of *n* MPOs showing the same motion attributes at time *t*

Trend-setter: one trend-setting MPO anticipates the motion of *n* others

Part 1 Relative Motion Patterns

Spatial Motion Patterns

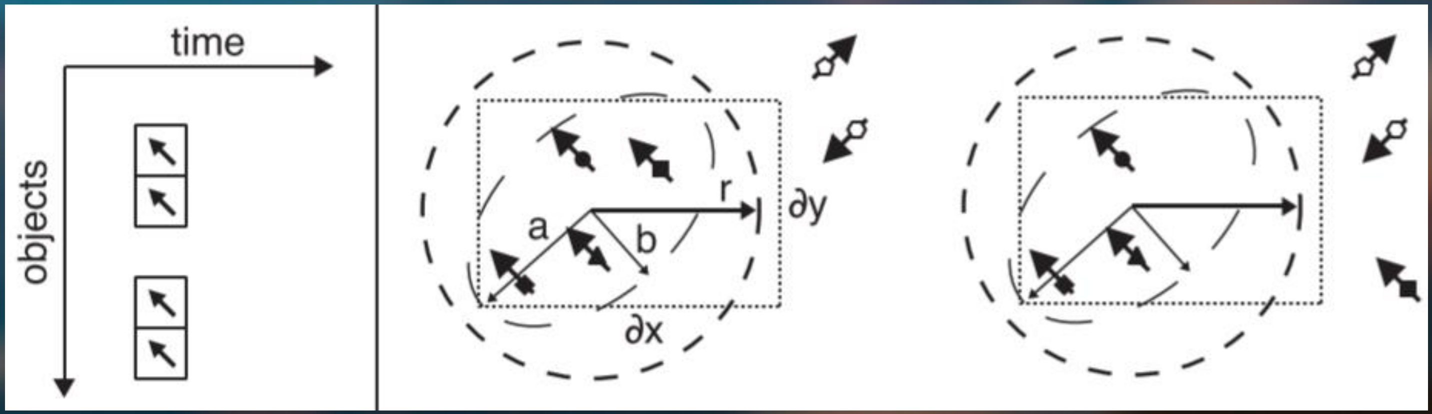
Basic Motion + Spatial Constraints(proximity measure)

- The maximal length of the cumulated distances to the mean or median center
- The average length of the Delaunay edges of the group
- MBB(i.e. a ellipse)
- The indication of a maximal border length of the convex hull

Part 1 Relative Motion Patterns

Flock:

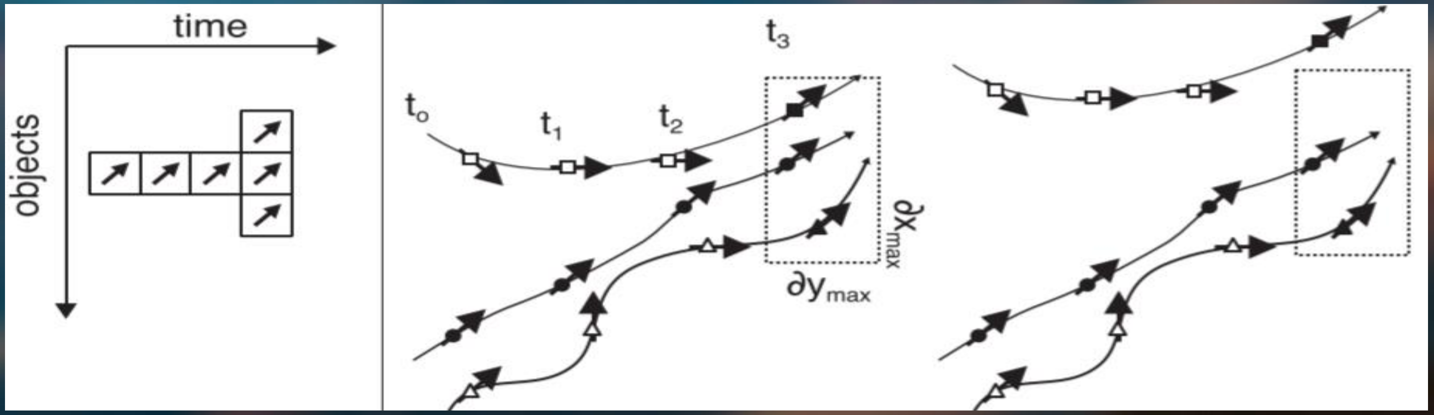
Concurrence + Spatial constraints



Part 1 Relative Motion Patterns

Leadership:

Trend-setter + Spatial constraints



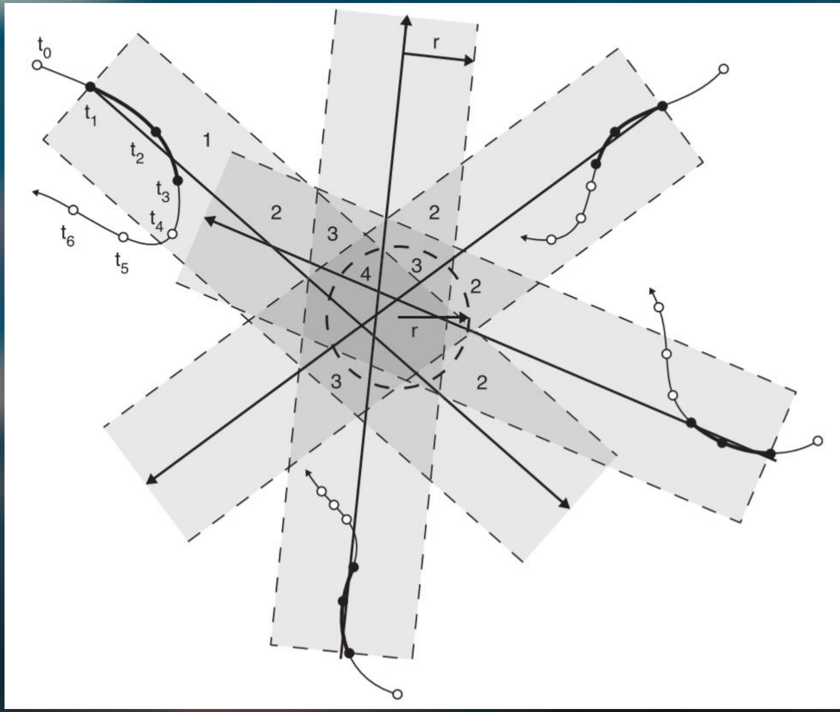
Part 1 Relative Motion Patterns

Aggregation/Disaggregation Motion Patterns

- *Convergence*: Set of m MPOs at interval *i* with motion azimuth vectors intersecting within a range *R* of radius *r*
- *Encounter*: Actually meeting within *R* extrapolating the current motion
- *Divergence*: The opposite of the Convergence
- *Breakup*: The opposite of the *Encounter*

Part 1 Relative Motion Patterns

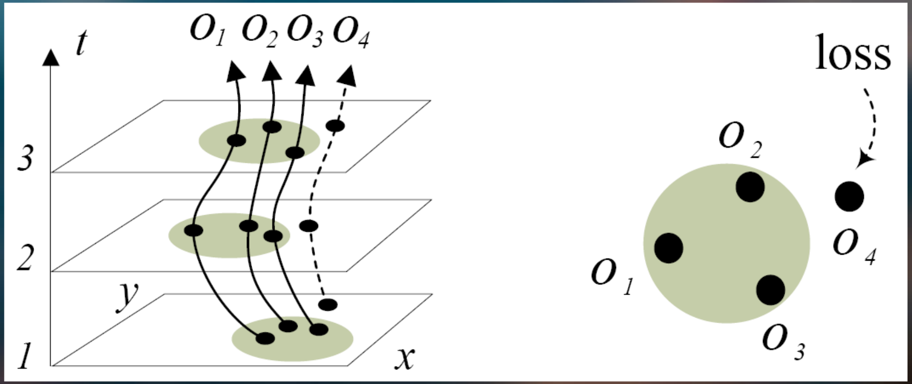
An example:
Convergence
without cluster



Part 1 Relative Motion Patterns

Drawbacks:

- Hard to define an absolute distance between two objects
- Hard to define *r* (i.e. Lossy-flock problem)
- A single *r* is unrealistic



Density-Based Motion Patterns

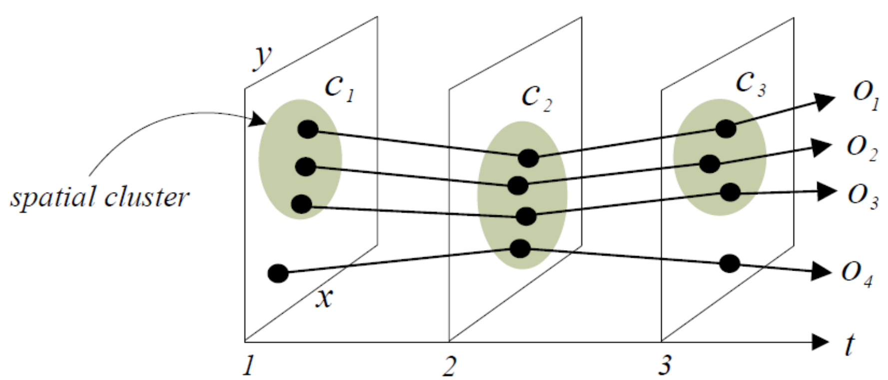
Allow the capture of trajectories of arbitrary shape

- *Convoy*: Density-Based *Flock*
- *Swarm*: Time-Relaxed *Convoy*
- Moving Cluster: A sequence of spatial cluster

Part 1 Density-Based Motion Patterns

Moving Cluster:

A set of objects that move close to each other for a time duration



$$\frac{|c_t \cap c_{t+1}|}{|c_t \cup c_{t+1}|} \geq \theta$$

Part 1 Density-Based Motion Patterns

Flock:

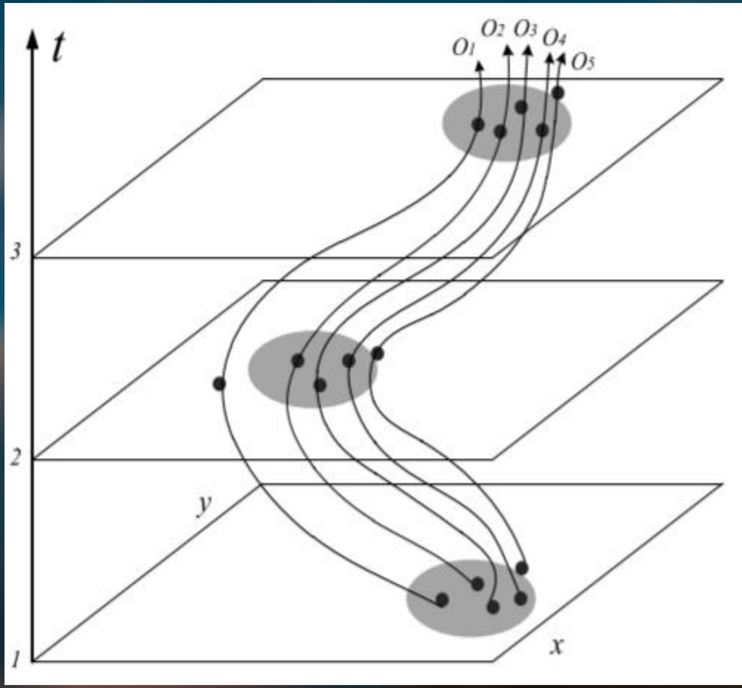
- A disc of rigid size
- K consecutive timestamps

Convoy:

- Dense-based clustering

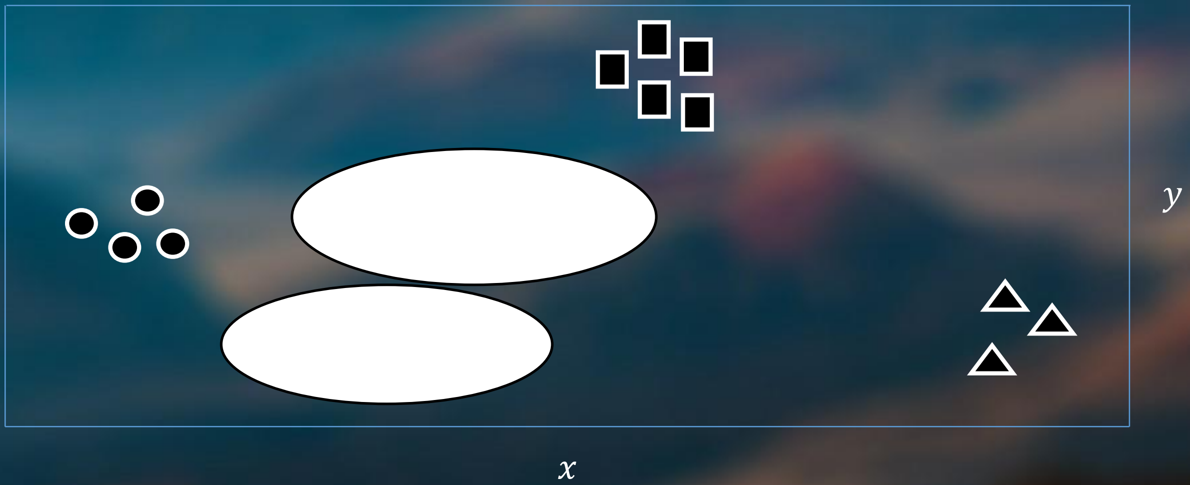
Swarm:

- K (non-consecutive) timestamps



Part 1 Density-Based Motion Patterns

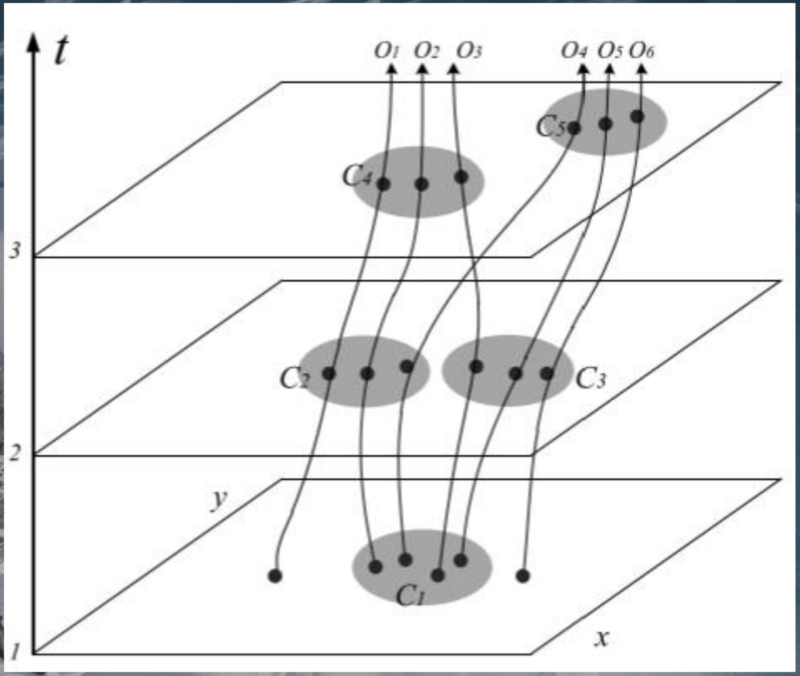
Dense Area Detection: Drawbacks



Part 2 Gathering Patterns

Gathering Patterns

- Key Attributes
- Definitions
- How does it work



Part 2 Key Attributes

- Scale: A gathering typically involves a relatively large number of individuals
- Density: Those individuals forms a dense group
- Durability: It should last for a certain time period continuously
- Stationariness: The geometric properties of the group is relatively stable
- Commitment: At any time of the gathering, there exist several dedicated members who stick to the group for a certain time(possibly non-consecutive)

Part 2 Definitions

- The trajectory of a moving object

$$o = < (p_1, t_1), (p_2, t_2), \dots, (p_n, t_n) >$$

$$where\ p_i \in \Re^2\ is\ the\ geo - spatial\ position$$

$$sample\ at\ t_i \in \mathcal{T}_{DB}$$

- Directly density-reachable

A point *p* is *directly density reachable* from a point *q* w.r.t a given distance threshold ϵ and a integer *m,* if

$$p \in N_\epsilon(q) \text{ and} |N_\epsilon(q)| \geq m$$

$$where\ N_\epsilon(p) = \{q \in S | D(p, q) < \epsilon\}$$

Part 2 Definitions

- Snapshot cluster

  The *snapshot cluster* c_t is

  - ◆ a non-empty subset of objects $\mathcal{O} \in \mathcal{O}_{DB}$
  - ◆ $\forall o_p, o_q \in \mathcal{O}$, $o_p(\mathrm{t})$is density-connected to $o_q(t)$
  - ◆ $\mathcal{O}$ is maximal

Part 2 Definitions

- Crowd

  A *crowd* C_r is

  ◆ A sequence of snapshot cluster at consecutive timestamps

  ◆ The lifetime of C_r is no less than k_c

  ◆ There should be at least m_c objects at any time

  ◆ The distance between any consecutive pair of clusters is not greater than δ

Part 2 Definitions

- Gathering

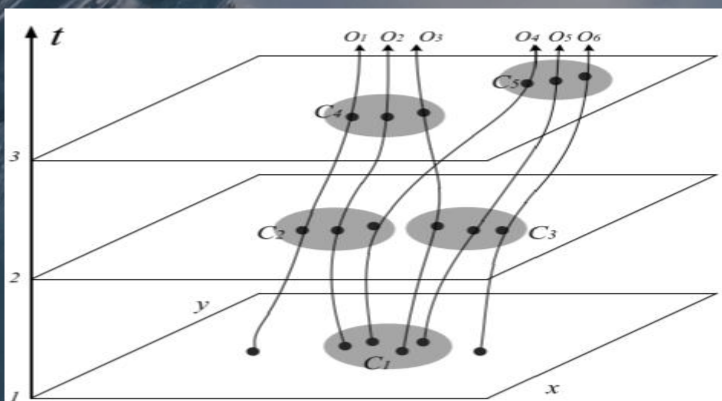
  A *crowd* C_r is called a *gathering* iff there exists at least m_p participators in each snapshot cluster of C_r

- Participator

  An object o is called a participator iff it appears in at least k_p snapshot cluster

Part 2 Definitions

object	c_1	c_2	c_4	#	object	c_1	c_3	c_4	#
$\boldsymbol{o}_1$		–	–	2	o_1			–	1
$\boldsymbol{o}_2$	–	–	–	3	$\boldsymbol{o}_2$	–		–	2
$\boldsymbol{o}_3$	–		–	2	$\boldsymbol{o}_3$	–	–	–	3
$\boldsymbol{o}_4$	–	–		2	o_4	–			1
o_5	–			1	$\boldsymbol{o}_5$	–	–		2
o_6				0	o_6		–		1
# Par.	3	3	3		# Par.	3	2	2	



$$k_p = 2, m_p = 3$$

Part 2 How does it work

How does it work

1. Snapshot cluster
2. Crowd discovery

   Indexing clusters with R-tree/grid
3. Gathering detection

   TAD

   Updating

Part 2

How does it work

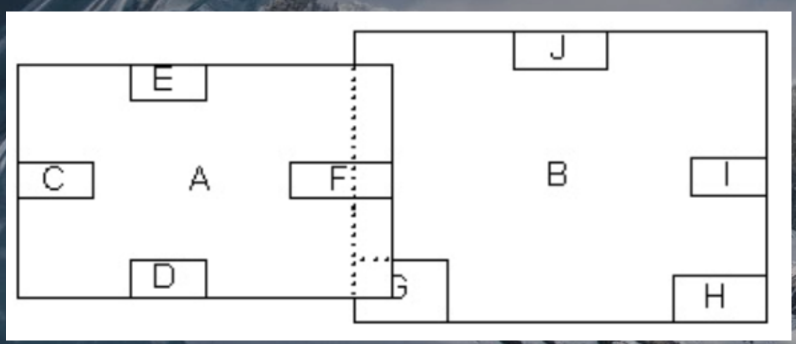
Crowd Discovery:

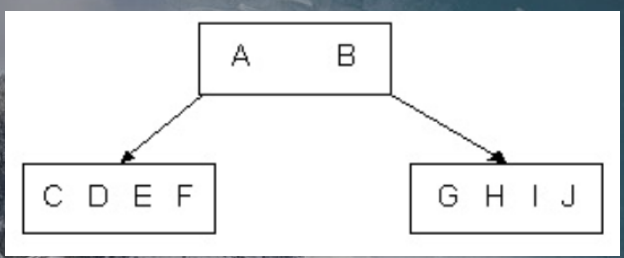
- C_r is said to be closed if it has no super-crowd
- Longer gathering can exist in super-crowd if the crowd is not closed
- Computing Hausdorff distance is high-cost!

Part 2 Crowd Discovery

Indexing cluster with R-tree:

- $d_{min}(M(c_i), M(c_j) \leq d_H(c_i, c_j)$
- Index the MBRs of the cluster in *C* by a R-tree





$\mathcal{O}(\mathrm{MN}) \dashrightarrow \mathcal{O}(\log_M N)$

Part 2 Crowd Discovery

Indexing cluster with R-tree: Drawbacks

- R-tree still costs a lot in construction and maintain
- MBRs may not capture the distribution of clusters

Part 2 Crowd Discovery

Indexing cluster with Grid:

- Partition the space into by a grid
- The side length of each cell equals to $\frac{1}{\sqrt{2}}\delta$
- Maintain a cell list for each cluster and a inverted list for each cell
- *Affect Region*: A cell g_{ab}'s AF is the set of cells whose minimum distance with g_{ab} less than δ

Part 2 How does it work

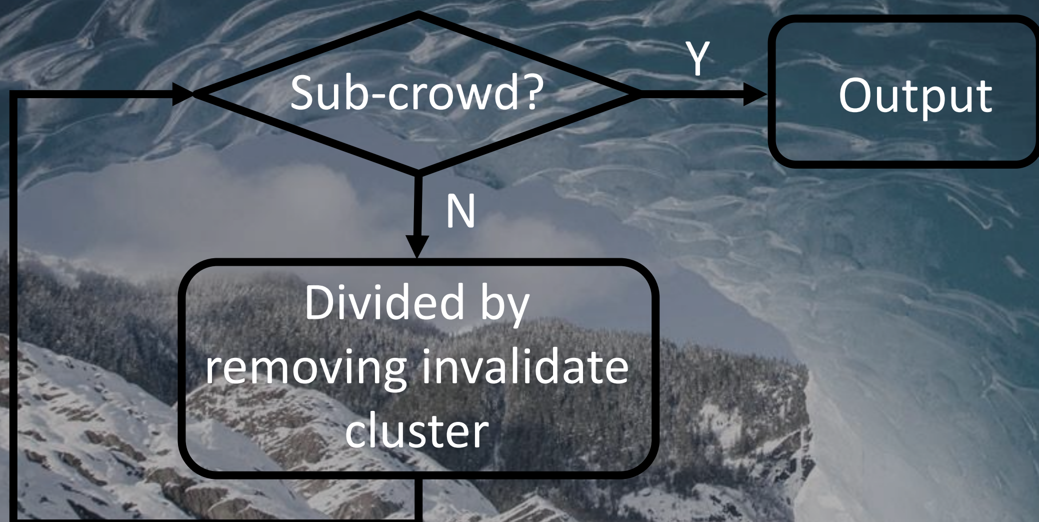
Gathering Detection

The downward closure property doesn't hold anymore

- TAD
- BVS
- Discovering gathering incrementally

Part 2 Gathering Detection

TAD(Test-and-Divide)



- The gathering output by TAD are closed

Part 2 Gathering Detection

TAD(Test-and-Divide)

$k_p = k_c = 3, \ m_p = m_c = 3$

c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
	o_1	o_1		o_1	o_1		
o_2	o_2	o_2	o_2			o_2	o_2
o_3	o_3		o_3		o_3	o_3	o_3
o_4		o_4	o_4	o_4	o_4	o_4	o_4
	o_5	o_5	o_5				
				o_6	o_6		

Part 2 How does it work

BVS(Bit Vector Signature)

$B(o_1)$	0 1 1 0 1 1 0 0
$B(o_2)$	1 1 1 1 0 0 1 1
$B(o_3)$	1 1 0 1 0 1 1 1
$B(o_4)$	1 0 1 1 1 1 1 1
$B(o_5)$	0 1 1 1 0 0 0 0
$B(o_6)$	0 0 0 0 1 1 0 0

Part 2

How does it work

TAD & BVS

- Test Step
  - Count the 1 bits in *B(o)* with bit operation

1) Let $m1 = 01010101$,
   $x = (x \& m1) + ((x \gg 1) \& m1) = 01011000$
2) Let $m2 = 00110011$,
   $x = (x \& m2) + ((x \gg 1) \& m2) = 00100010$
3) Let $m4 = 00001111$,
   $x = (x \& m4) + ((x \gg 1) \& m4) = 00000100$

m1, m2 and m3 are called *masks*

Part 2 How does it work

TAD & BVS

- Divide Step
  - No need to process BVSs of non-participators
  - Extract clusters by AND operation and *masks for clusters* i.e. 11110000

Part 2

How does it work

Discovering gathering incrementally

- New database

$$\mathcal{O}'_{DB} = \mathcal{O}_{DB} \cup \mathcal{O}_{new}$$

- New time domain

$$\mathcal{T}'_{DB} = \mathcal{T}_{DB} \cup \mathcal{T}_{new}$$

Part 2 How does it work

Discovering gathering incrementally

- Crowd Extension:

Given a closed crowd $C_r = \{c_i, \ldots, c_j\}$ in $\mathcal{O}_{DB}$, if its last cluster in not at the most recent time point of $\mathcal{T}_{DB}$, then C_r cannot be extended into $\mathcal{O}'_{DB}$

Part 2

How does it work

Discovering gathering incrementally

- Crowd Extension:

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	$\boldsymbol{t_9}$	$\boldsymbol{t_{10}}$	$\boldsymbol{t_{11}}$	$\boldsymbol{t_{12}}$
					c_6^1				c_{10}^1		
		c_3^1	c_4^1	c_5^1						c_{11}^1	c_{12}^1
c_1^1	c_2^1			c_5^2				c_9^1	c_{10}^2		
	c_2^2	c_3^2		c_5^3							
					c_6^2	c_7^1	c_8^1	c_9^2			
					c_6^3						

Part 2 How does it work

Discovering gathering incrementally

- Gathering Update:
  - $IC(C_{r_{new}}) \cap C_{r_{old}} \subseteq IC(C_{r_{old}})$

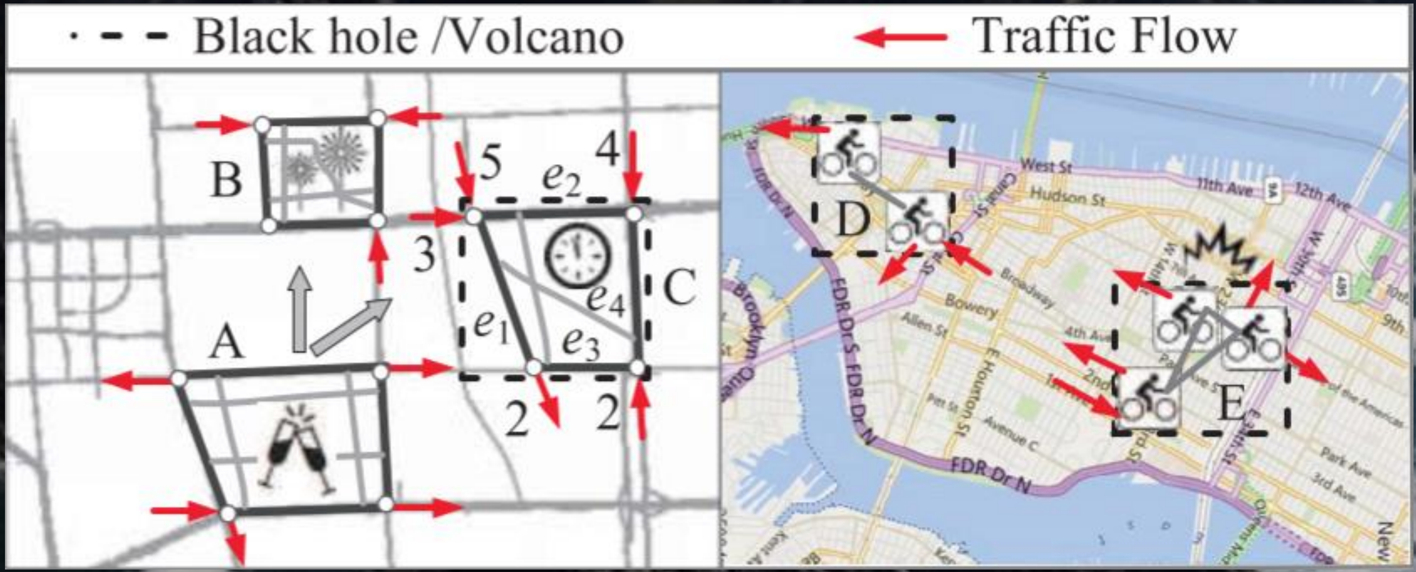
    Invalid cluster $C_{r_{old}}$ can be valid in $C_{r_{new}}$
  - Given an invalid cluster $c_j \in IC(C_{r_{new}})$

  with $j \leq n+1$, then any closed gathering $G_r \subset < c_i, \ldots, c_{j-1} >$ remains closed in $C_{r_{new}}$

    Closed gathering remain closed

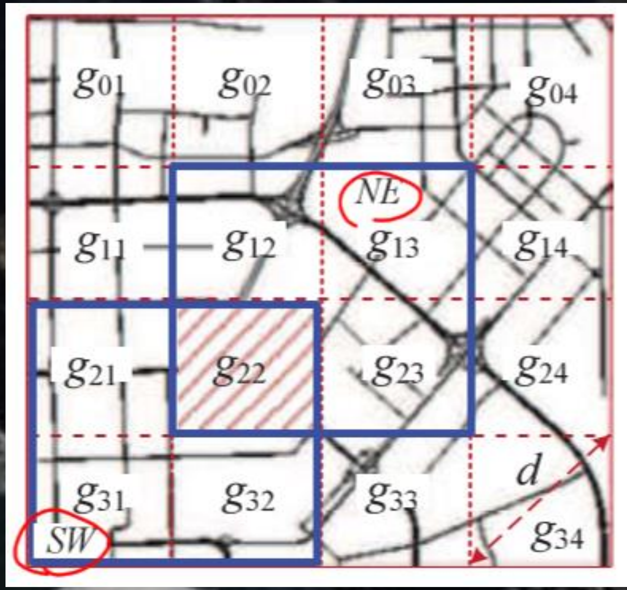
Part 3 Extension Models

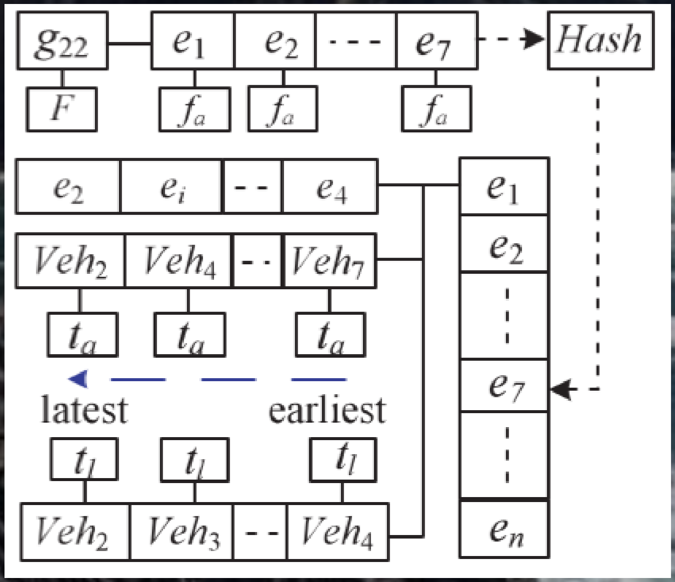
Urban Black Holes: STG(spatial-temporal Graph)



Part 3 Extension Models

Urban Black Holes: STG(spatial-temporal Graph)





Data Mining Lab

THANK YOU

PRESENTED BY ZHAOYI