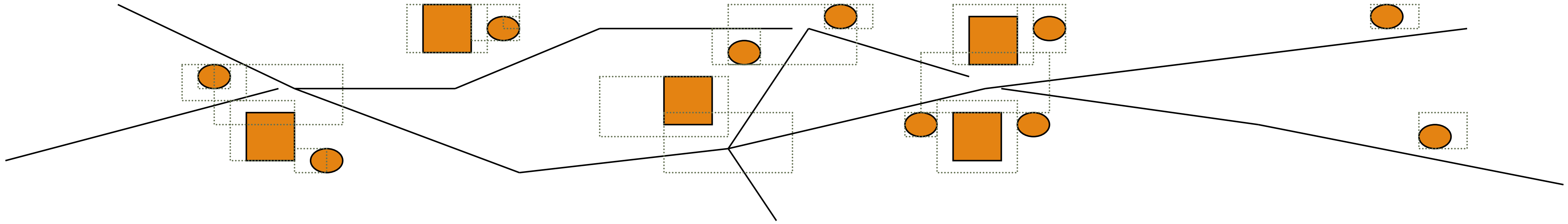
Mining Spatial Associations

Spatial Frequent Patterns and Associations

- Spatial frequent patterns and association rule: $A \Rightarrow B$ [*s%, c%*]
  - A and B are sets of spatial or non-spatial predicates, e.g.,
    - Topological relations: *intersects, overlaps, disjoint,* etc.
    - Spatial orientations: *left_of, west_of, under,* etc.
    - Distance information: *close_to, within_distance,* etc.
  - Measures: *s%*: support, and *c%*: confidence of the rule
- Example: Rules likely to be found
  - *is_a(x, large_town) ^ intersect(x, highway) → adjacent_to(x, water) [7%, 85%]*
- Explore *spatial autocorrelation*: Spatial data tends to be highly self-correlated (*nearby things are more related than distant ones)*
  - E.g., neighborhood, temperature

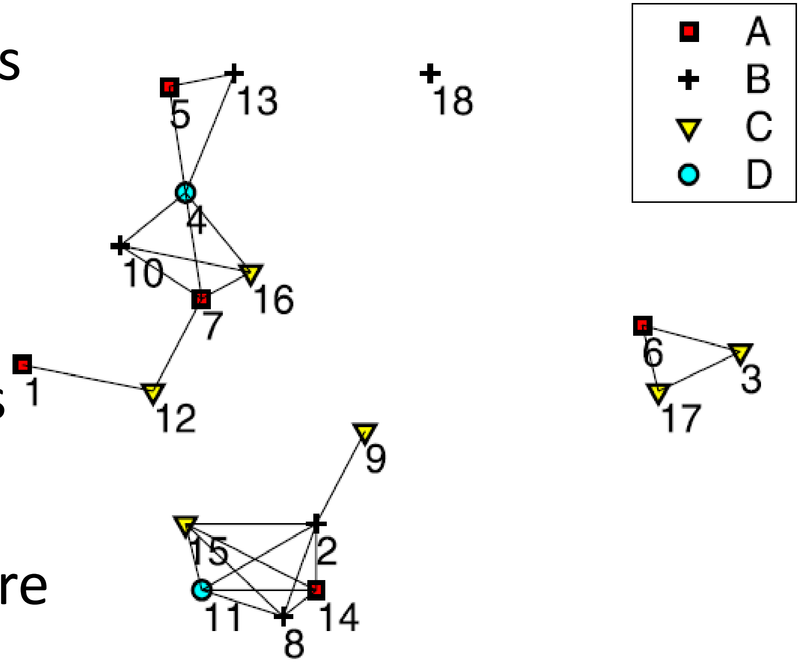
Mining Spatial Associations: Progressive Refinement

- Hierarchy of spatial relationship:
  - *close_to* is a generation of *near_by*, *touch*, *intersect*, *contain*, ...
  - Progressive refinement: First search for rough relationship and then refine it
- Two-step mining of spatial association:
  - Step 1: Rough spatial computation (as a filter)
    - Using MBR (Minimum Bounding Rectangle) or R-tree for rough estimation
  - Step2: Detailed spatial algorithm (as refinement)
    - Apply only to those objects which have passed the rough spatial association test (no less than *min_support*)



Mining Spatial Colocation Patterns

Spatial Colocation Patterns



- *Colocation pattern:* A group of spatial features or events that are frequently co-located in the same region
  - Ex. West Nile Virus often occur in regions with poor mosquito control and the presence of birds
- Figure: Neighborhood instances are connected by edges
  - Ex. {3, 6, 17}, {4, 7, 10, 16}, {2, 8, 11, 14, 15}, {2, 9}, …
- Rowset(C) if every feature in patter C appears as a feature of an instance in the neighbor-set L, e.g.,
  - rowset({A, B, C, D}) = {{4, 7, 10, 16}, {2, 11, 14, 15}, {8, 11, 14, 15}}
  - rowset({A, B}) = {{5, 13}, {7, 10}, {2, 14}, {8, 14}}
- A colocation rule R: A→ B, conditional probability cp(R) is defined as

$$\frac{|\{L \in rowset(A) | \exists L' \text{ s.t. } (L \subseteq L') \wedge (L' \in rowset(A \cup B))\}|}{|rowset(A)|}$$

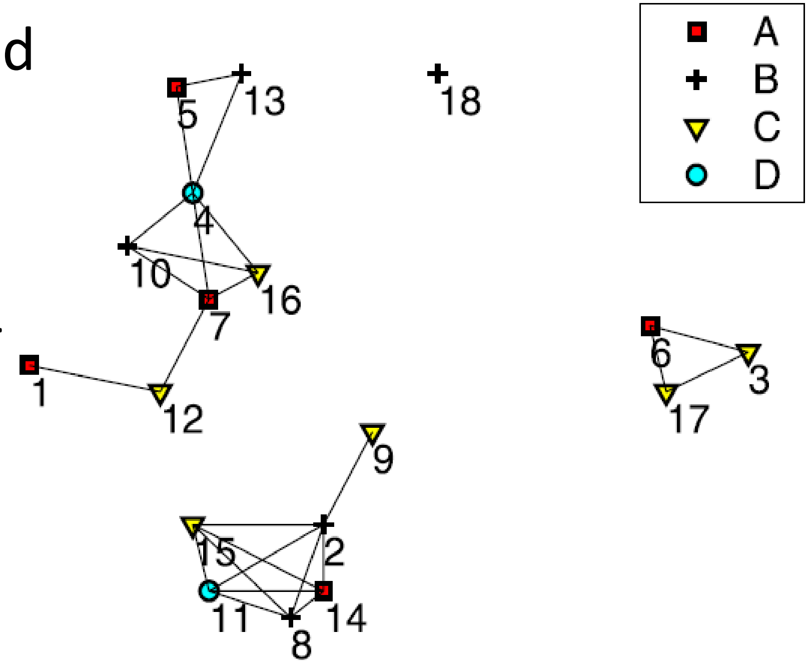
  - cp({A, B} → {C, D}) = |rowset({A, B, C, D})|/|rowset({A,B})| = ¾ = 75%

Mining Spatial Colocation Patterns

- *Participation ratio pr(C, f): probability that* C is observed in a neighbor-set wherever feature *f* is observed

$$pr(C, f) = \frac{|\{r \mid (r \in S) \wedge (r.f = f) \wedge (r \text{ is in a row instance of } C)\}|}{\{r \mid (r \in S) \wedge (r.f = f)\}|}$$

  - Ex. *pr*({A,B,C,D}, A) = 2/5, ..., *pr*({A,B,C,D}, D) = 2/2 = 1
- Monotonicity of *participation ratio*
  - Let *C, C′* be two co-location patterns such that *C′* ⊂ *C*
    - Then, for each feature *f* ∈ *C′*, *pr(C′, f)* ≥ *pr(C, f)*

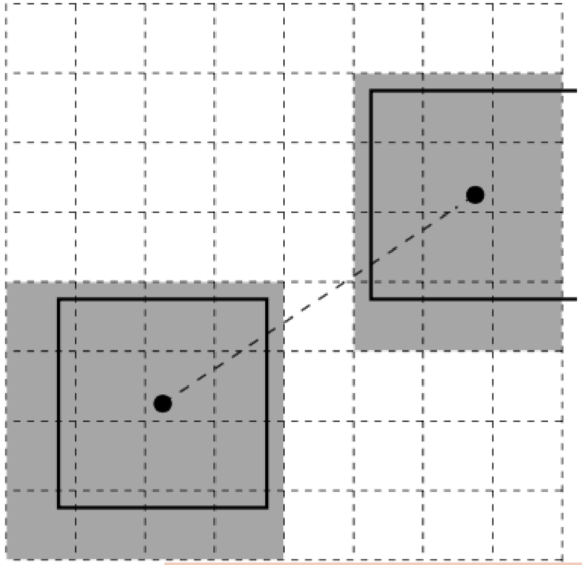


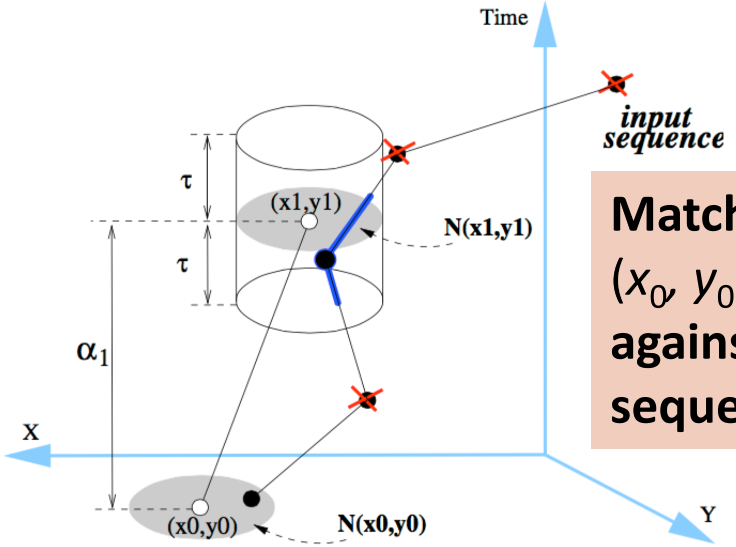
- An Apriori-like algorithm can be derived for efficient mining colocation patterns
  - Ex: Let min-feature-support = σ, min-*pr* = ρ
    - Start with a set of single feature pattern $\{p_1\}$ with support ≥ σ
      - Grow to size *k,* in Apriori way (i.e., stop growing if the pattern is infrequent)
    - For each such p, mine its super-pattern P, s.t., pr(P, p) ≥ ρ, in Apriori way

Mining and Aggregating Patterns over Multiple Trajectories

Partition-Based Trajectory Pattern Mining

- **Partition-Based Trajectory Pattern Mining (e.g., Mining T-Patterns)** [1]:
  - First partition the space into equal-width grids and obtain Regions-of-Interests (RoIs)
  - Then transform each input trajectory into a time-annotated symbolic sequence
  - Use constraint-based sequential pattern mining to find trajectory patterns





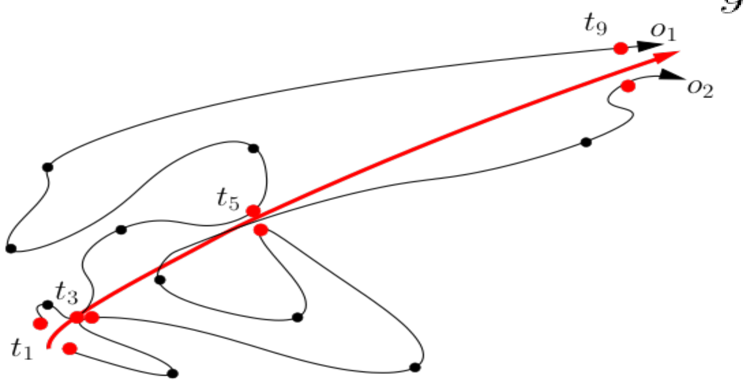
Matching T-pattern $(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$ **against an input ST-sequence**

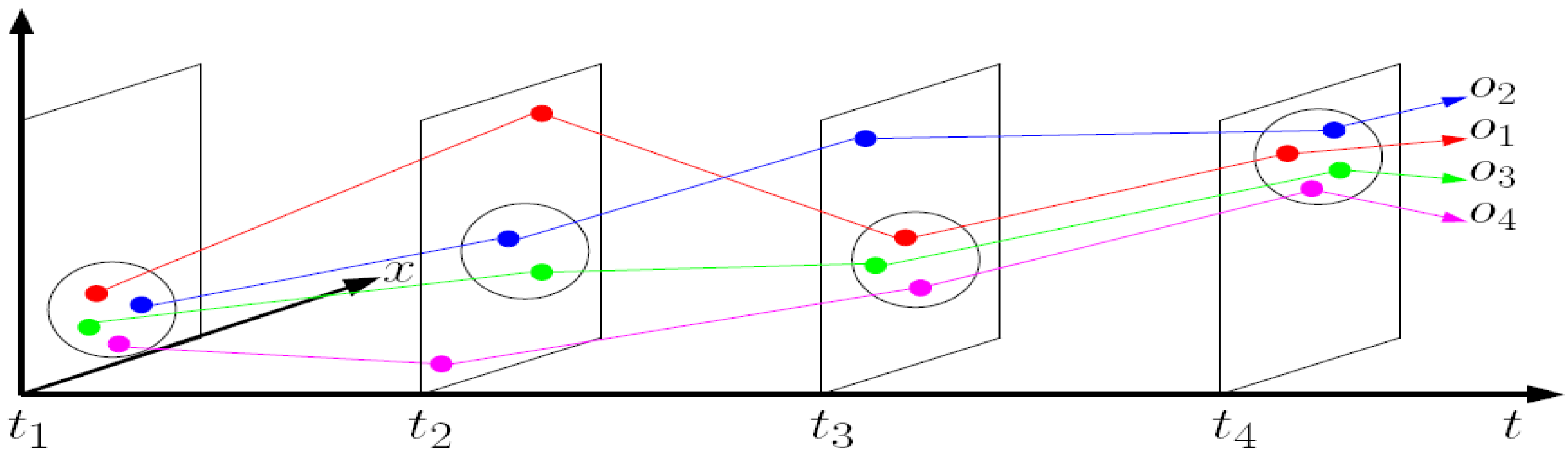
Railway Station —15*min*→ Castle Square —2*h*15min→ Museum
Railway Station —10*min*→ Middle Bridge —10*min* → Campus

[1] F. Giannotti, M. Nanni, F. Pinelli, D. Pedreschi, Trajectory Pattern Mining, KDD'07

Detecting Moving Object Clusters

- **Flock and convoy**: Both require *k* consecutive time stamps
  - **Flock:** At least *m* entities are within a *circular* region of ***radius r*** and move in the same direction
  - **Convoy:** *Density-based clustering* at each timestamp; no need to be a rigid circle
- **Swarm:** Moving objects may not be close to each other for all the consecutive time stamps
  - Efficient pattern mining algorithm can be derived for mining such swarm patterns

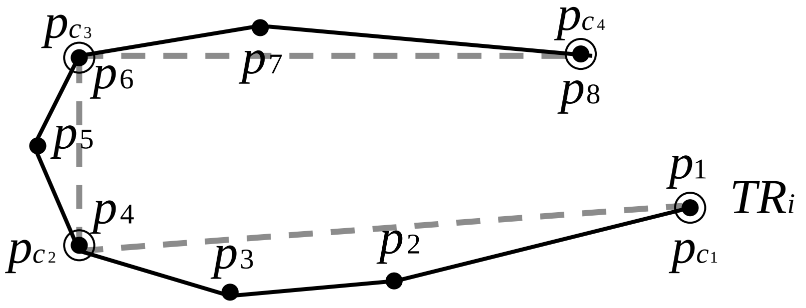




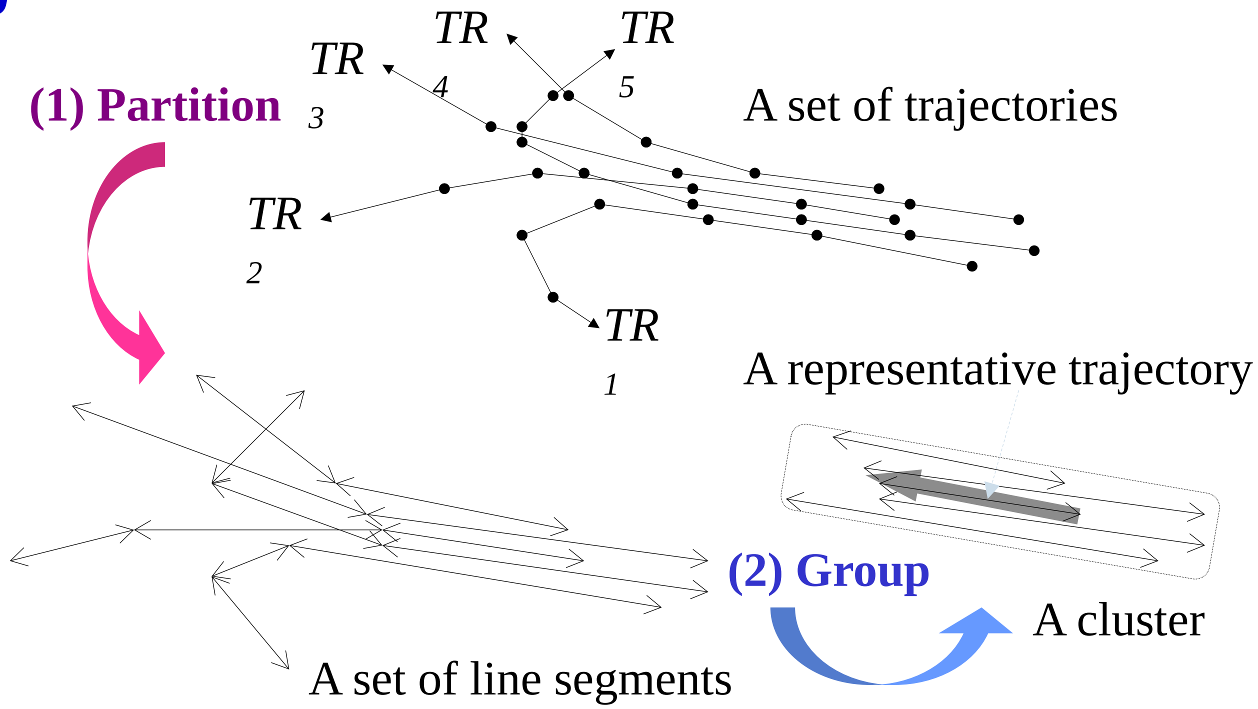
Z. Li, et al.: Swarm: Mining Relaxed Temporal Moving Object Clusters. VLDB'10

Trajectory Clustering: A Partition-and-Group Framework

- Grouping trajectories ***as a whole*** ⇒ cannot find ***similar portions*** of trajectories
- ***Solution:*** discovers common ***sub***-trajectories, e.g., *forecast hurricane landfall*
- Two phases: ***partitioning*** and ***grouping***
- Identify the points where the behavior of a trajectory changes rapidly ⇒ *characteristic points*
  - Based on the minimum description length (MDL) principle



◉: characteristic point - - - : trajectory partition

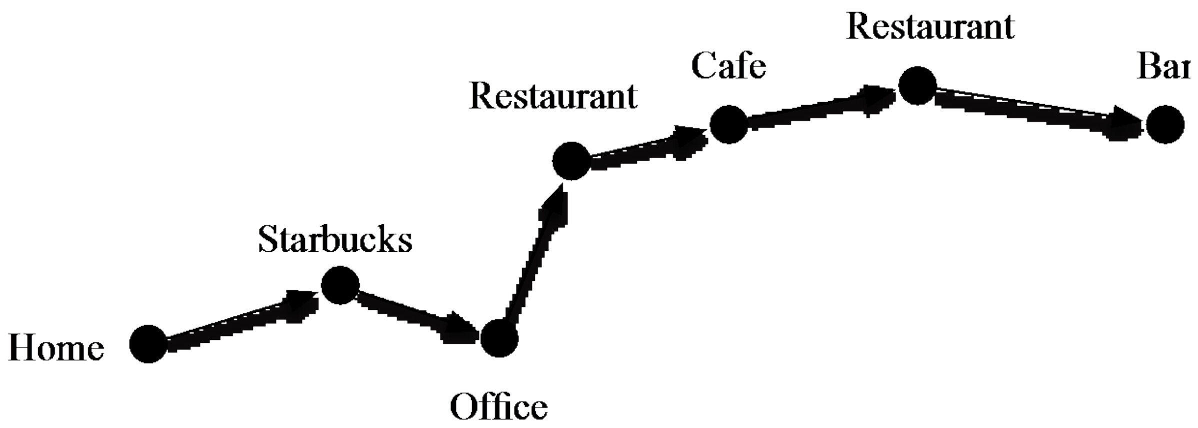


J.-G. Lee, et al., "Trajectory Clustering: A Partition-and-Group Framework", SIGMOD'07

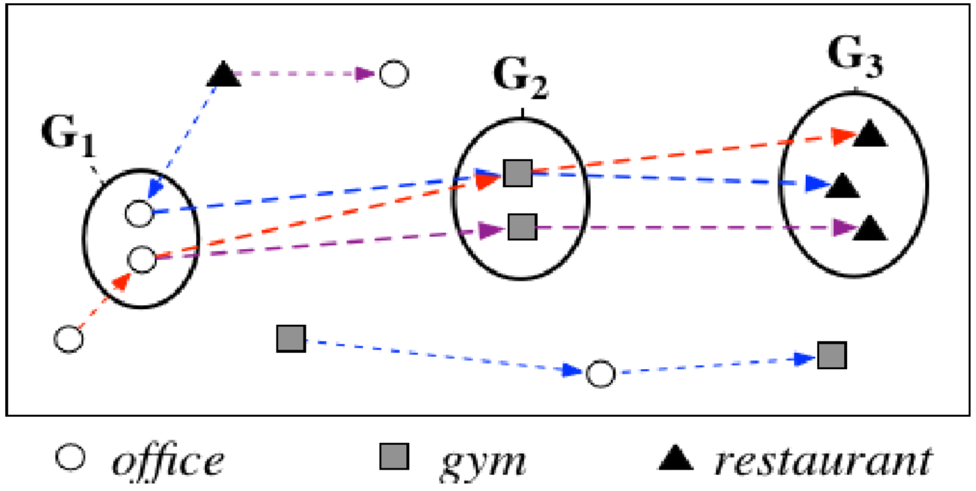
Mining Semantics-Rich Movement Patterns

Mining Frequent Movement Patterns

- **Frequent Movement Pattern:** A movement sequence that frequently appears in the input trajectory database
- **Frequent Movement Pattern vs. Frequent Sequential Pattern**
  - Both aim at finding frequent subsequences from the input sequence database
  - For mining frequent movement patterns, similar places may need to be grouped to collectively form frequent subsequences



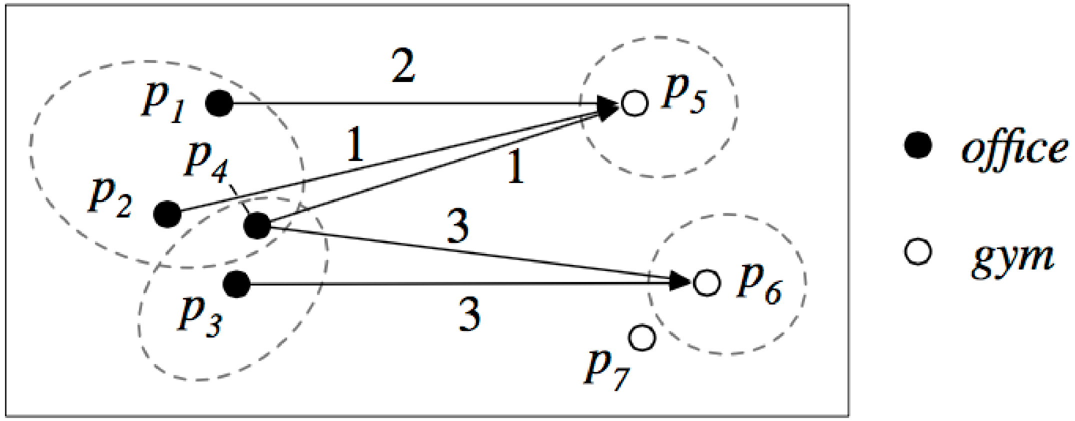
An example trajectory

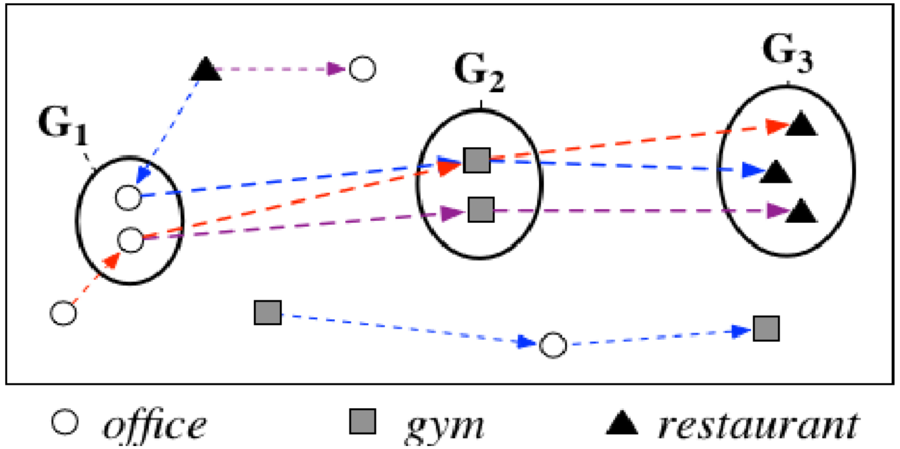


An example movement pattern

Mining Semantics-Rich Movement Patterns

- **Semantics-rich movement pattern**: In addition to knowing how people move from one region to another, we also want to understand the functions of the regions
- **A two-step top-down mining approach:**
  - Step 1: Find a set of coarse patterns that reflect people's semantics-level transitions (e.g., office → restaurant, home → gym)
  - Step 2: Split each coarse pattern into several fine-grained ones by grouping similar movement snippets



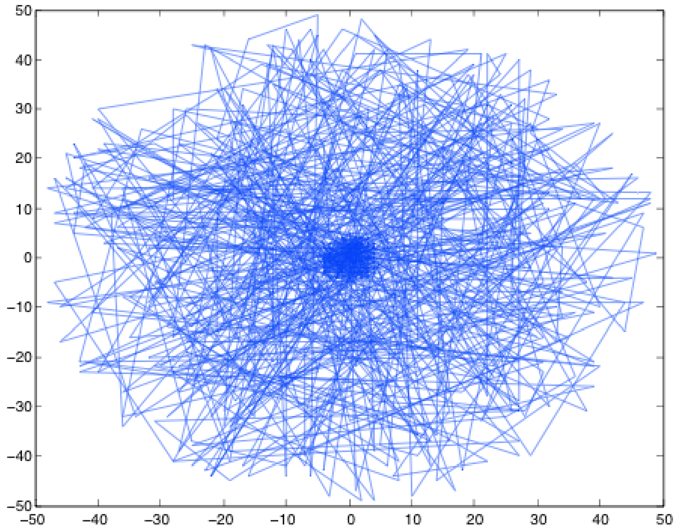


C. Zhang et al., Splitter: Mining Fine-Grained Sequential Patterns in Semantic Trajectories, VLDB 2014

Mining Periodic Movement Patterns

Pattern Discovery in Sparse Movement Data: Finding Good Reference Points

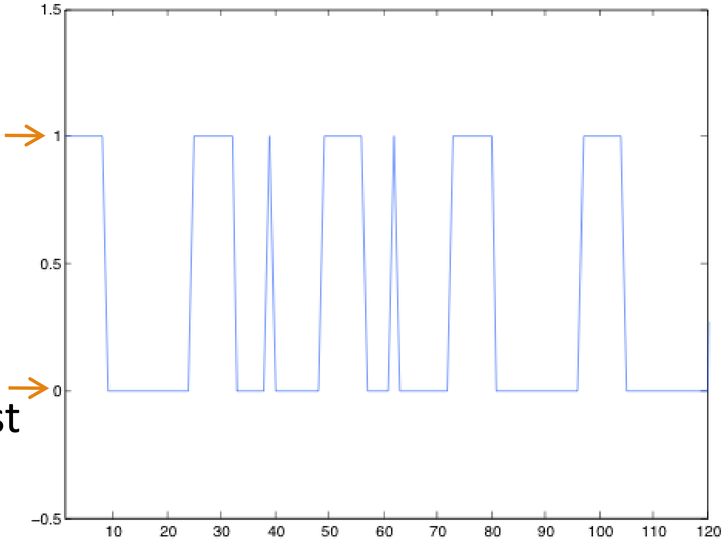
- **Pattern discovery in sparse data** (e.g., find bird flying patterns)
  - Periodicity shows up in some reference “spots” (or “cluster centers”)
  - Reference spots can be detected using **density-based method**
  - Periods are detected for each reference spot using **Fourier Transform** and **auto-correlation**



Finding bird flying patterns?
Bird Nest is a good reference point

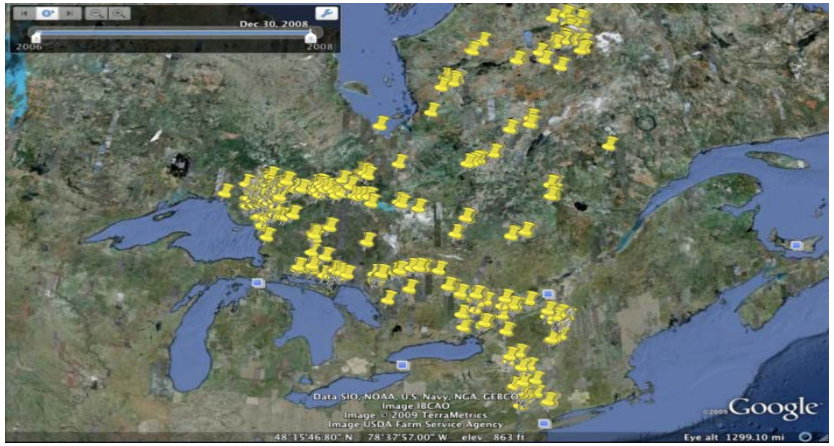
in the nest

not in the nest

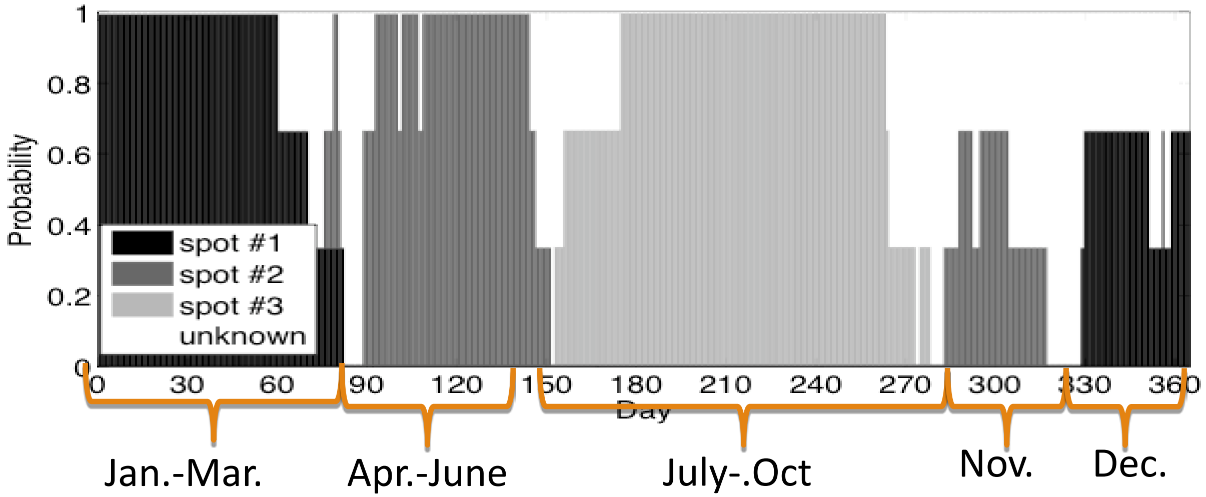


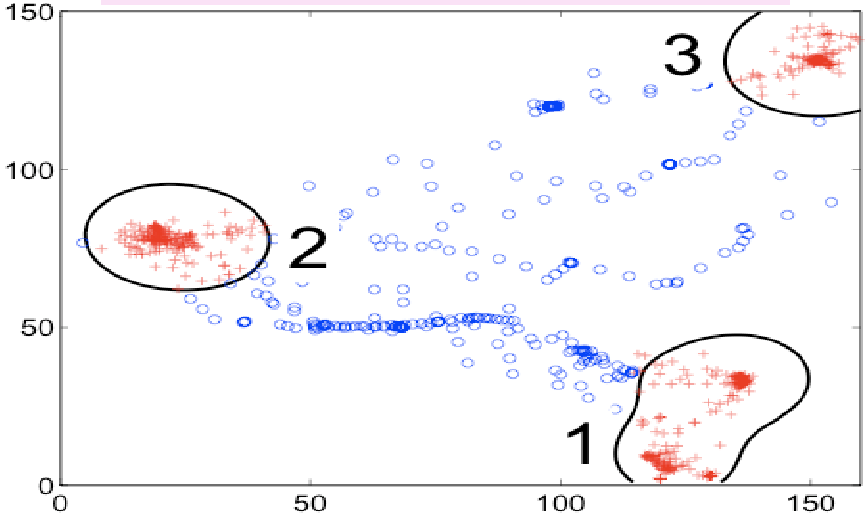
Period is more obvious in this binary sequence!

Example: Mining Periodic Patterns with Sparse Data



3-yr Bird migration data: very sparse



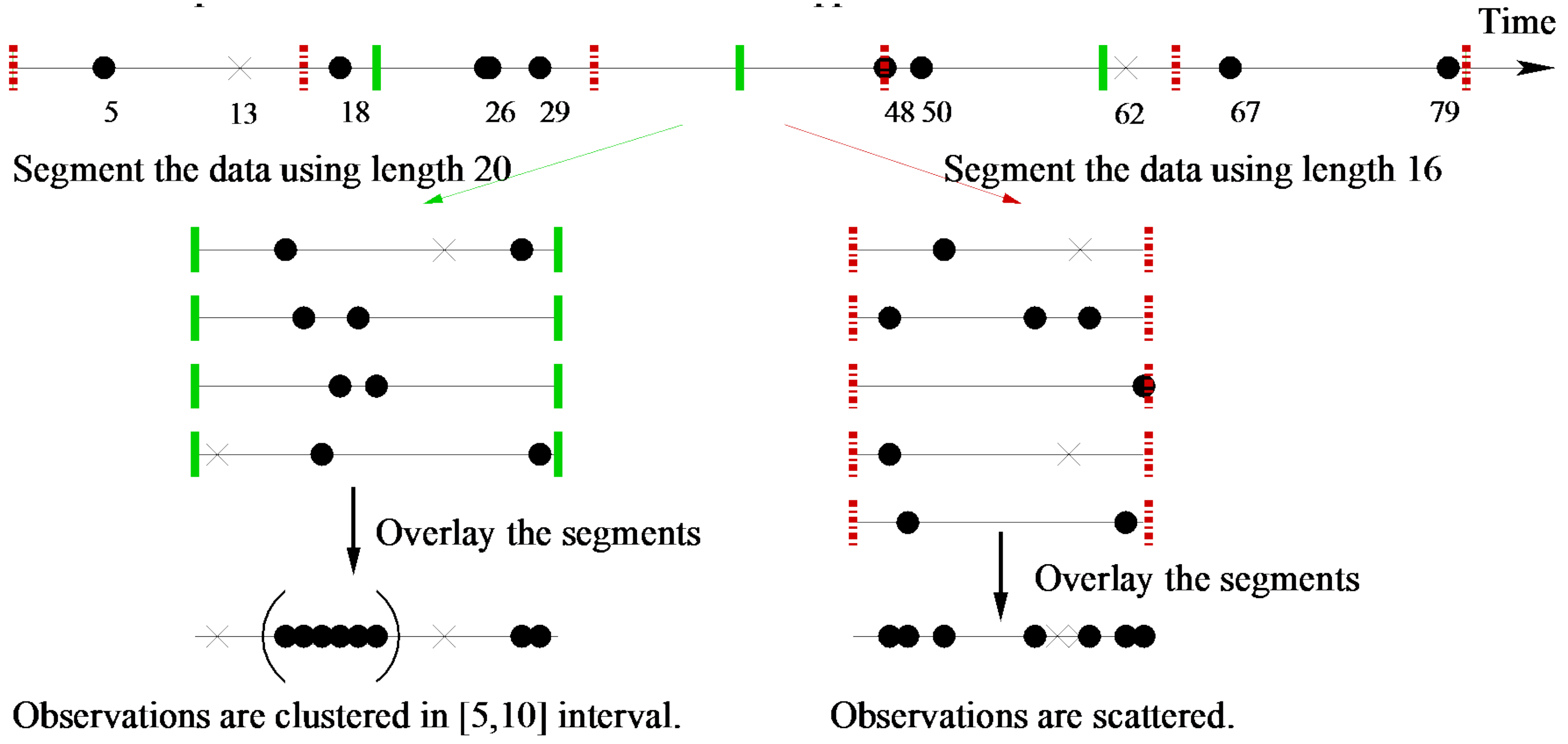


- **Detecting periods**: Cluster data to find reference "points" and then detect multiple interleaved periods by Fourier Transform and auto-correlation
- **Summarizing periodic patterns:** By clustering and pattern discovery

Z. Li, et al.: Mining Periodic Behaviors for Moving Objects. KDD'10

Periodicity Detection in Sparse Data

- Time-related data can be scattered and sparse, e.g., phone calls at a location



- Projecting on the true period, it shows a highly skewed (clustered) distribution
- Effective method can be developed based on this observation

Z. Li, et al., ePeriodicity: Mining Event Periodicity from Incomplete Observations. IEEE TKDE, 2015

Summary

Summary: Mining Spatiotemporal and Trajectory Patterns

- Mining Spatial Associations
- Mining Spatial Colocation Patterns
- Mining and Aggregating Patterns over Multiple Trajectories
- Mining Semantics-Rich Movement Patterns
- Mining Periodic Movement Patterns

Recommended Readings

- F. Giannotti, M. Nanni, F. Pinelli, D. Pedreschi: Trajectory Pattern Mining. KDD'07
- Y. Huang, S. Shekhar, H. Xiong: Discovering colocation patterns from spatial data sets: A general approach. IEEE Trans. on Knowledge & Data Eng., 16(12), 2004
- Y. Huang, J. Pei, H. Xiong: Mining Co-Location Patterns with Rare Events from Spatial Data Sets. GeoInformatica 10(3): 239-260, 2006
- K. Koperski, J. Han: Discovery of Spatial Association Rules in Geographic Information Databases. SSD'95
- J.-G. Lee, J. Han, and K.-Y. Whang: Trajectory Clustering: A Partition-and-Group Framework, SIGMOD'07
- Z. Li, B. Ding, J. Han, R. Kays: Swarm: Mining Relaxed Temporal Moving Object Clusters. VLDB'10
- Z. Li, B. Ding, J. Han, R. Kays, P. Nye: Mining Periodic Behaviors for Moving Objects. KDD'10
- Z. Li, J. Wang and J. Han, ePeriodicity: Mining Event Periodicity from Incomplete Observations. IEEE TKDE, 27(5): 1219-1232, 2015
- C. Zhang, J. Han, L. Shou, J. Lu, and T. La Porta: Splitter: Mining Fine Grained Sequential Patterns in Semantic Trajectories. VLDB'14