Second Progress Presentation

Alla Petrakova, Steve Mussmann, Mahdi Kalayeh

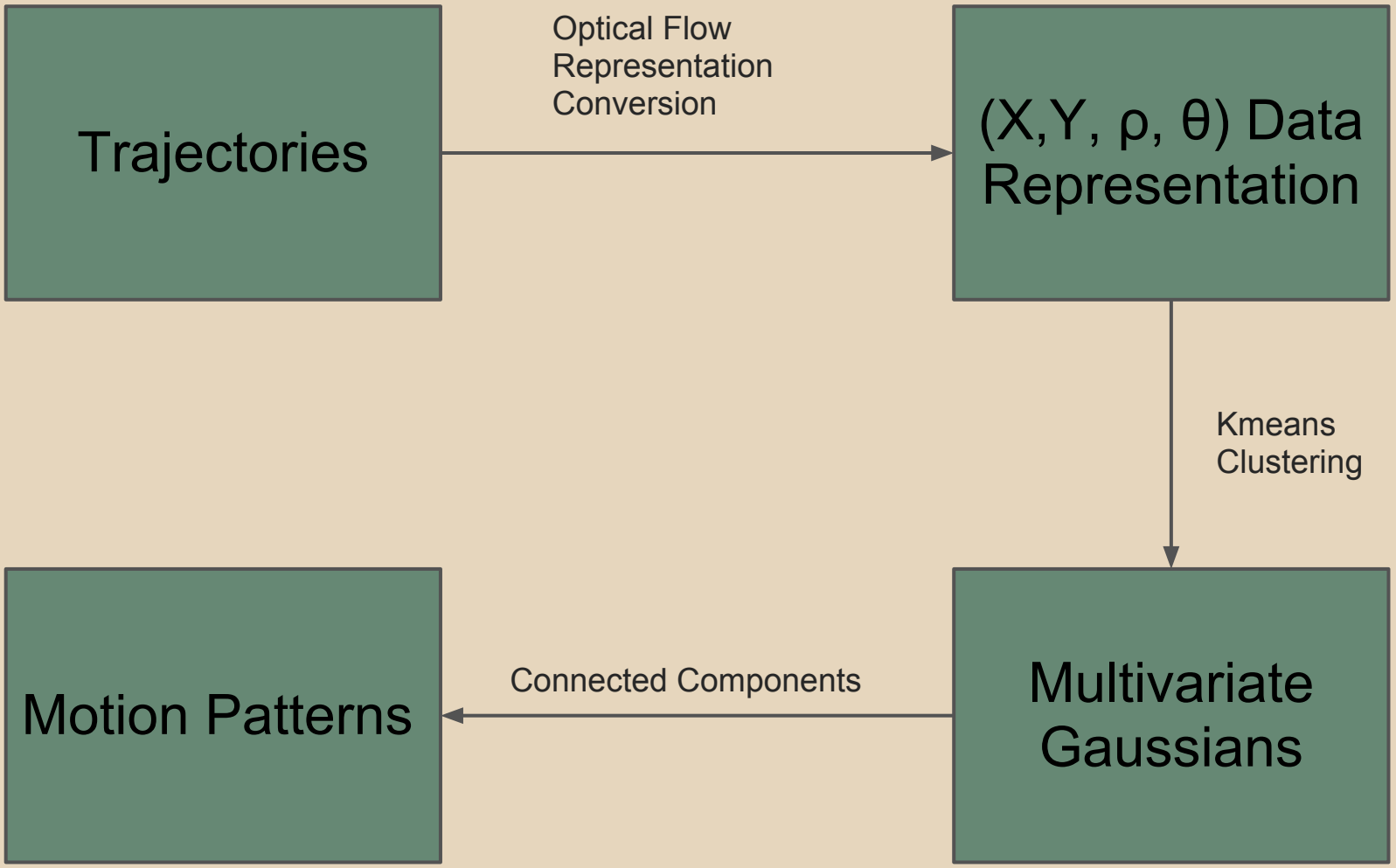
Overall Direction

-There exist two effective motion pattern mining algorithms developed in CRCV.

-Effective and efficient trajectory clustering is a contemporary problem in the data mining field.

-Modification of motion pattern algorithms can provide a solution to the data mining problem.

Motion Models Algorithm



Trajectory Clustering Papers

-Gathered 23 contemporary trajectory clustering papers.

-Identified commonly used and easily accessible trajectory datasets.

-Acquiring implementations of trajectory clustering algorithms for comparison

List of Papers

Title	Journal	Date	Cite	Data	Code
Vector Field k-Means: Clustering Trajectories by Fitting Multiple Vector Fields	arxiv	2012	0	GeoLife GPS Trajectories, [1] HURDAT, Call Detail Records (CDR)	Coming in 2 weeks from 6/4
Profiling Moving Objects by Dividing and Clustering Trajectories Spatiotemporally	IEEE	2012	0	[28] [29] typhoon, taxi & bus in Taipai	asked via email 6/6
Trajectory Analysis Using Spectral Clustering and Sequence Pattern Mining	JCIS	2012	1	urban intersection (not sourced)	
Finding homogeneous groups in trajectory streams	ACM	2012	0	(4.1) bus & truck in greece, Oregon wildlife	asked via email 6/6
Robust online trajectory clustering without computing trajectory distances	ICPR	2012	0	surveillence (not sourced)	
Using Relative Distance and Hausdorff Distance to Mine Trajectory Clusters	iaes	2012	0	Hurricane Data	asked via email 6/6
Learning trajectory patterns by clustering: Experimental studies and comparative evaluation	IEEE	2009	44	intersection surveillance (not sourced)	
Trajectory pattern mining	ACM	2007	344	trucks in greece	Code at kdd.isti.cnr.it/~nanni/
Mining Frequent Trajectory Patterns from GPS Tracks	CISE	2010	1	taxis in Shenzhen (not sourced)	asked via email 6/6
Learning Pedestrian Trajectories with Kernels	ICPR	2010	6	surveillence (not sourced)	
A Complete Framework for Clustering Trajectories	IEEE	2009	2	bus & truck in greece, Oregon wildlife	asked via email 6/6
A Trajectory Clustering Algorithm Based on Symmetric Neighborhood	World Cgrss	2009	2	Oregon wildlife	asked via email 6/6
An extended k-means technique for clustering moving objects	Egypt Info.	2011	2	[16] objects in London	
MoveMine: Mining Moving Object Databases	ACM	2010	29	movebank.org; weather.unisys.com/hurricane/at	Code at web.engr.illinois.edu/~klei2/r
Swarm: mining relaxed temporal moving object clusters.	VLDB	2010	41	Movebank.org	Code at faculty.ist.psu.edu/jessieli/S
Detection of risk factors using trajectory mining	JIIS	2011	15	ASL	
Incremental Clustering for Trajectories	DASFAA	2010	31	SF Taxi GPS	
Mining Spatio-Temporal Patterns in Trajectory Data	JIPS	2010	6	Synthetic data using spatio-temporal data generator G-TERD	
TMC-pattern: holistic trajectory extraction, modeling and mining	SIGSPATIAL	2012	0	GeoLife	
An Incremental DPMM-Based Method for Trajectory Clustering, Modeling, and Retrieval	IEEE	2013	0	ASL, Synthetic, Vehicle motion	
Clustering and aggregating clues of trajectories for mining trajectory patterns and routes	VLDB	2011	0	CarWeb	
Mining Trajectories for Spatio-temporal Analytics	ICDMW (IBM)	2012	0	mentions testing with large scale real world data; no source provided	
Efficient anomaly monitoring over moving object trajectory streams	SIGKDD	2009	32	www.ics.uci.edu/~yingyib/reposit	Code at www.ics.uci.edu/~yingyib/re

Trajectory Clustering: Partition-and-Group Framework

- Written by Lee, Han, Whang in 2007
- Serves as foundation for MoveMine set of works
- 357 citations

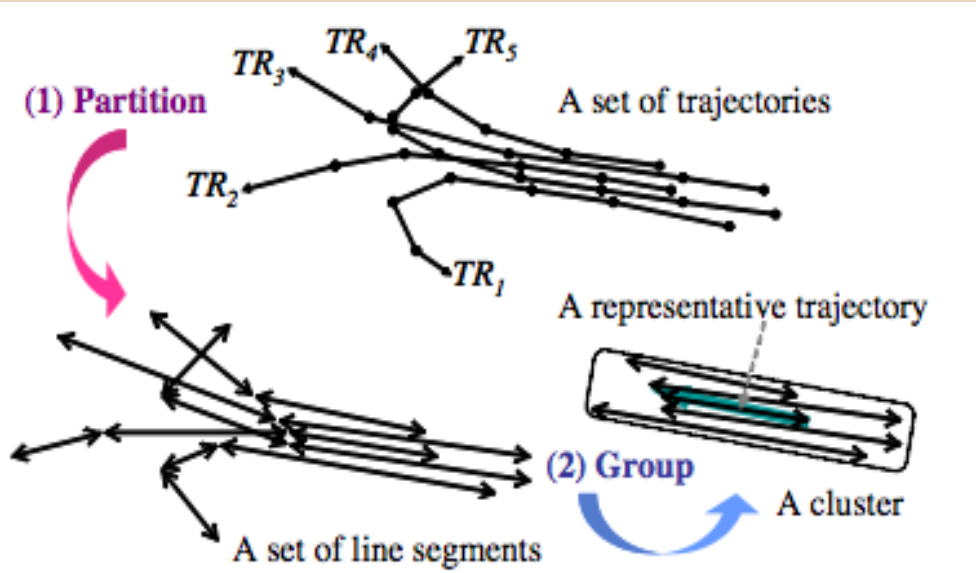
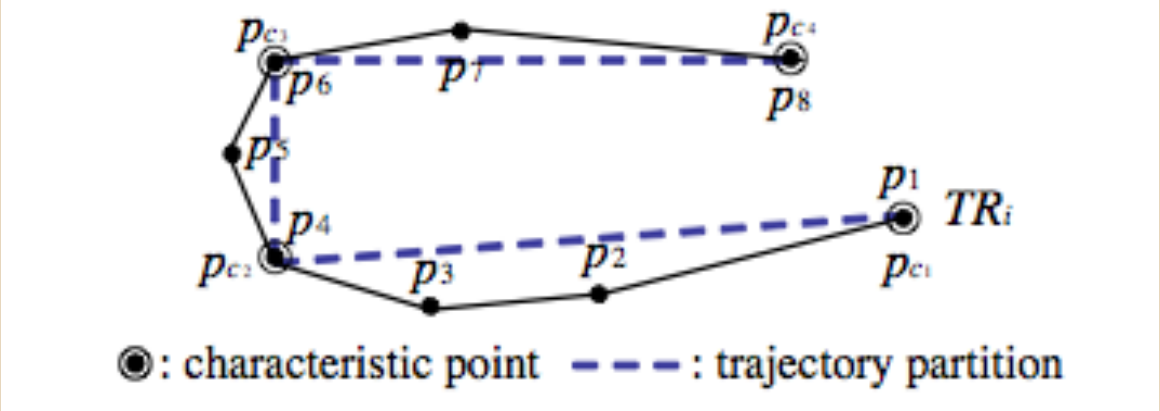


Figure 3: An example of trajectory clustering in the partition-and-group framework.

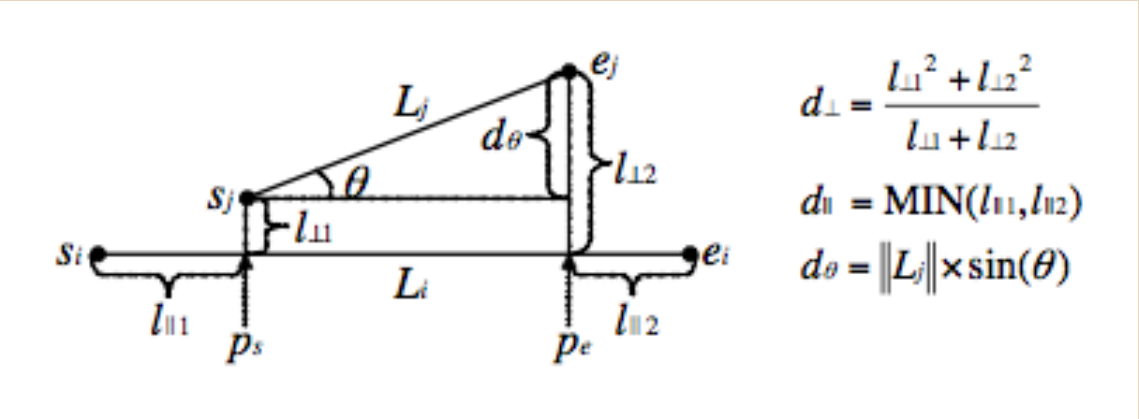
TRACLUS - Partitioning

- Preciseness vs Conciseness



- Characteristic points – points where the behavior of trajectory changes rapidly
- MDL (Minimum Description Length) principle
- L(H) conciseness (hypothesis)
- L(D|H) preciseness

Traclus Distance Formula



$$dist(L_i,L_j) = w_\perp \cdot d_\perp(L_i,L_j) + w_{//} \cdot d_{//}(L_i,L_j) + w_\theta \cdot d_\theta(L_i,L_j)$$

–The optimal partitioning of a trajectory should possess two desirable properties: preciseness and conciseness. Preciseness means that the difference between a trajectory and a set of its trajectory partitions should be as small as possible.

MDL Principle

MDL cost = L(H) + L(D|H)

–L(H) represents the sum of the length of all trajectory partitions (conciseness)

–L(D|H) represents the number of segments that deviate from actual trajectory (preciseness)

–We need to find the optimal partitioning that minimizes L(H) + L(D|H), achieving the optimal tradeoff between preciseness and conciseness.

Clustering

- Based on DBSCAN
- Parameters common to TRACLUS and DBSCAN
  - ε – the maximum distance
  - MinLns – minimum number of line segments in a cluster
- Parameter unique to TRACLUS:
  - Trajectory cardinality of a cluster:

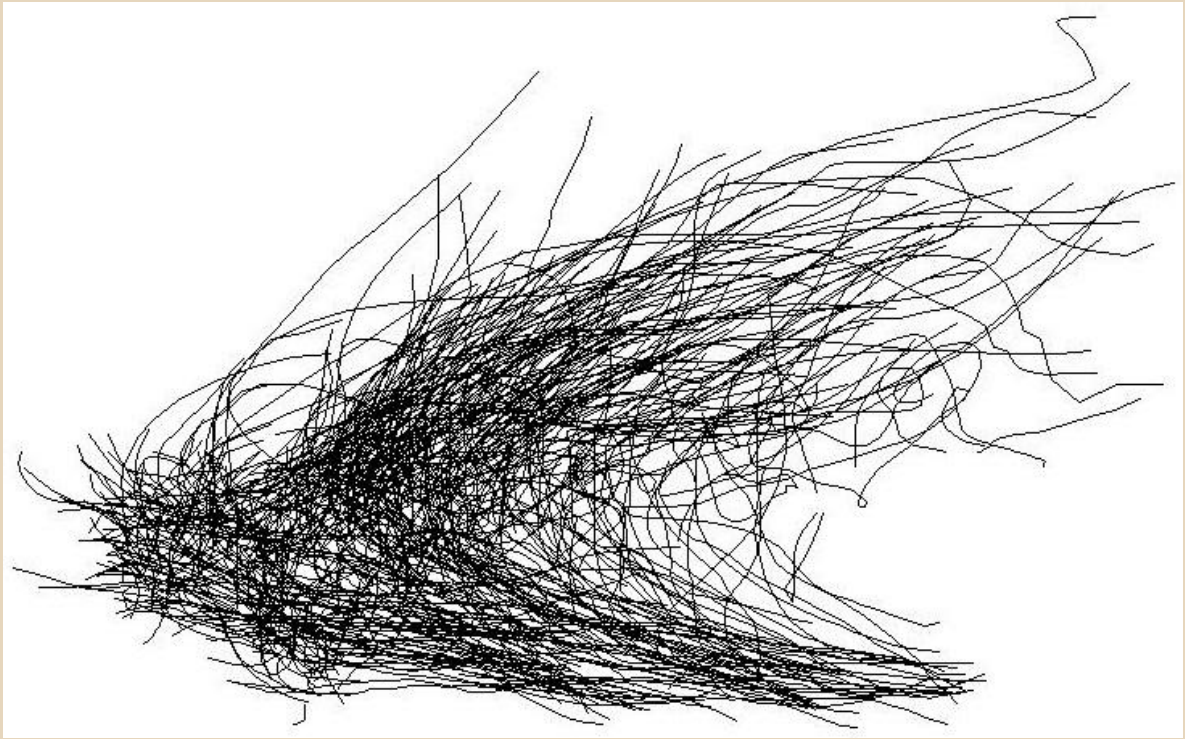
$$PTR(C_i) = \{TR(L_j) \mid \forall L_j \in C_i\}$$

TRACLUS

- Parameter selection
  - ε - simulated annealing
  - MinLns – average number of lines at an optimal ε
- Complexity –
  - $O(n^2)$
  - Depending on organization and indexing of data (line segments), complexity can be reduced to O(n log n)

Hurricane Data

1476 Atlantic hurricane trajectories recorded from 1851 to 2011.



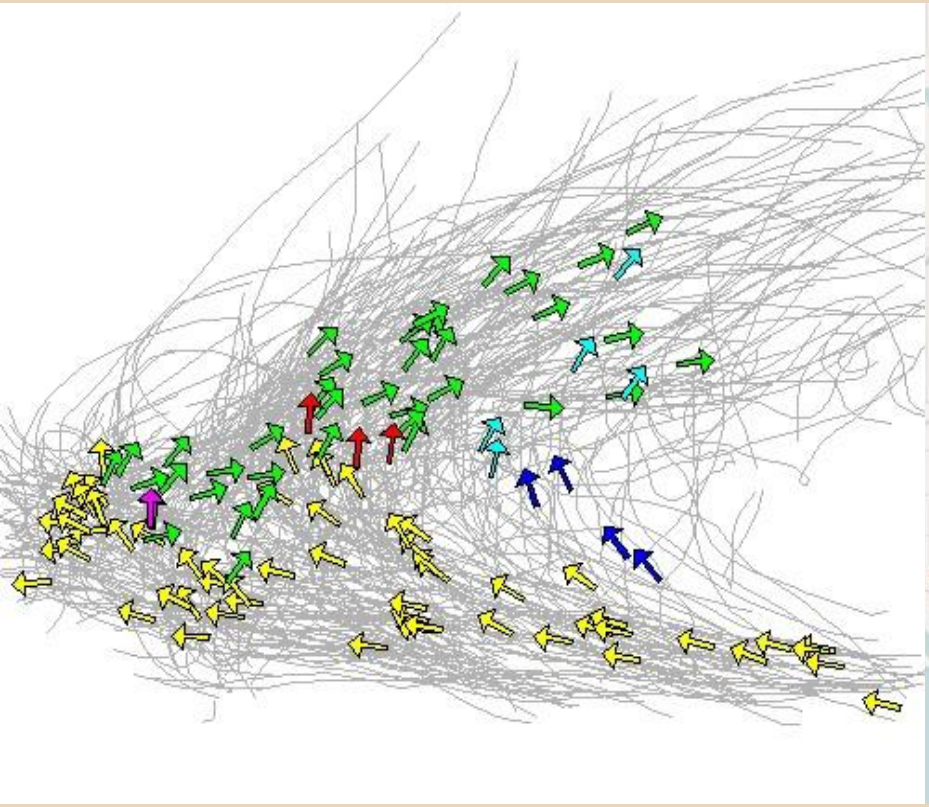
Li, Z., Ji, M., Lee, J. G., Tang, L. A., Yu, Y., Han, J., & Kays, R. (2010, June). MoveMine: mining moving object databases. In *Proceedings of the 2010 international conference on Management of data* (pp. 1203-1206). ACM. [Cited by 29]

Guan, B., Liu, L., & Chen, J. (2013). Using Relative Distance and Hausdorff Distance to Mine Trajectory Clusters. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *11*(1), 115-122.

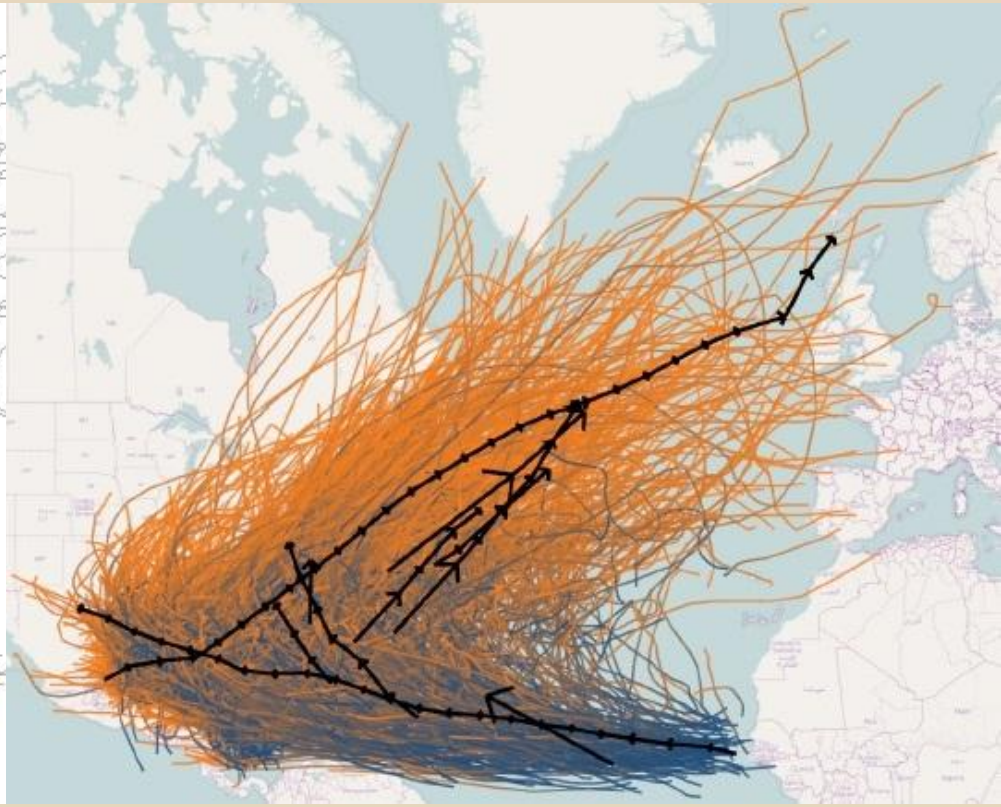
Ferreira, N., Klosowski, J. T., Scheidegger, C., & Silva, C. (2012). Vector Field k-Means: Clustering Trajectories by Fitting Multiple Vector Fields. *arXiv preprint arXiv:1208.5801*.

Algorithm Comparison

UCF

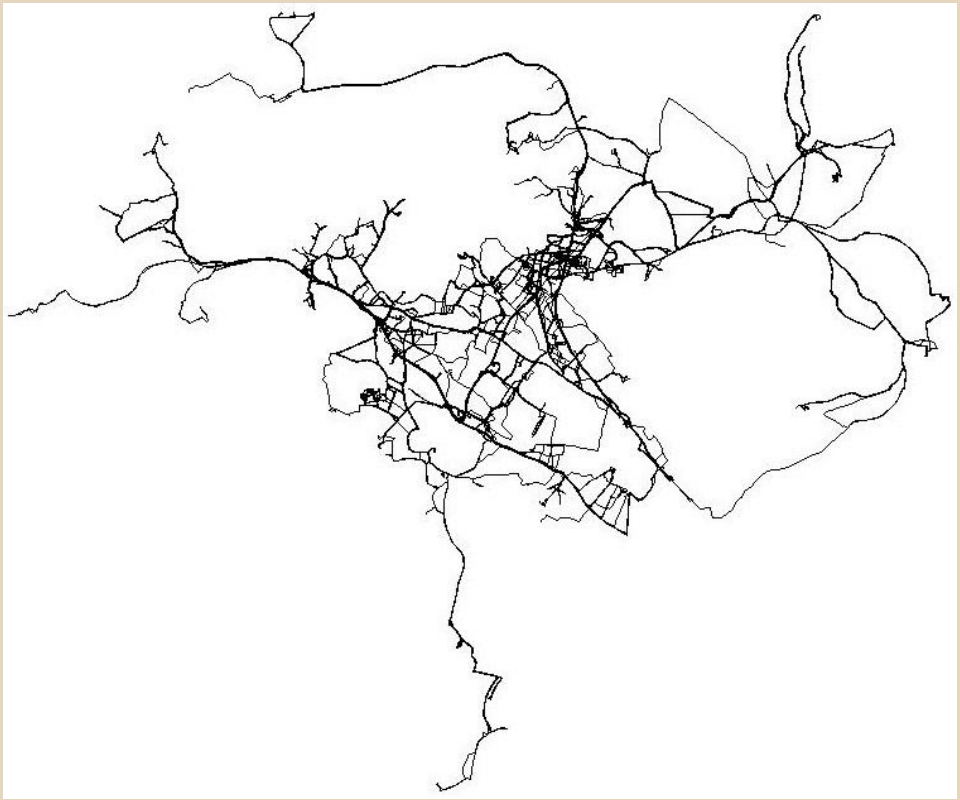


Vector Field Kmeans



Truck Data

GPS tracks from concrete trucks around Athens from 2002. 1100 trajectories.



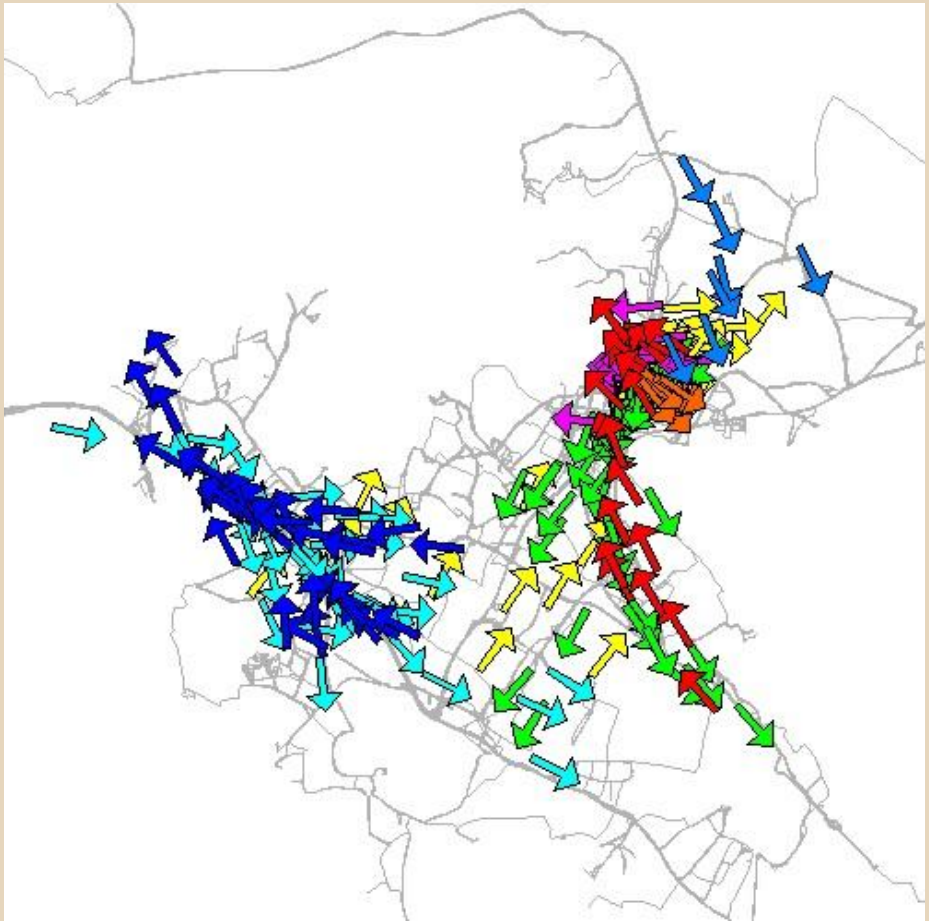
Giannotti, F., Nanni, M., Pinelli, F., & Pedreschi, D. (2007, August). Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 330-339). ACM. [Cited by 344]

Masciari, E. (2009, November). A complete framework for clustering trajectories. In *Tools with Artificial Intelligence, 2009. ICTAI'09. 21st International Conference on* (pp. 9-16). IEEE. [Cited by 2]

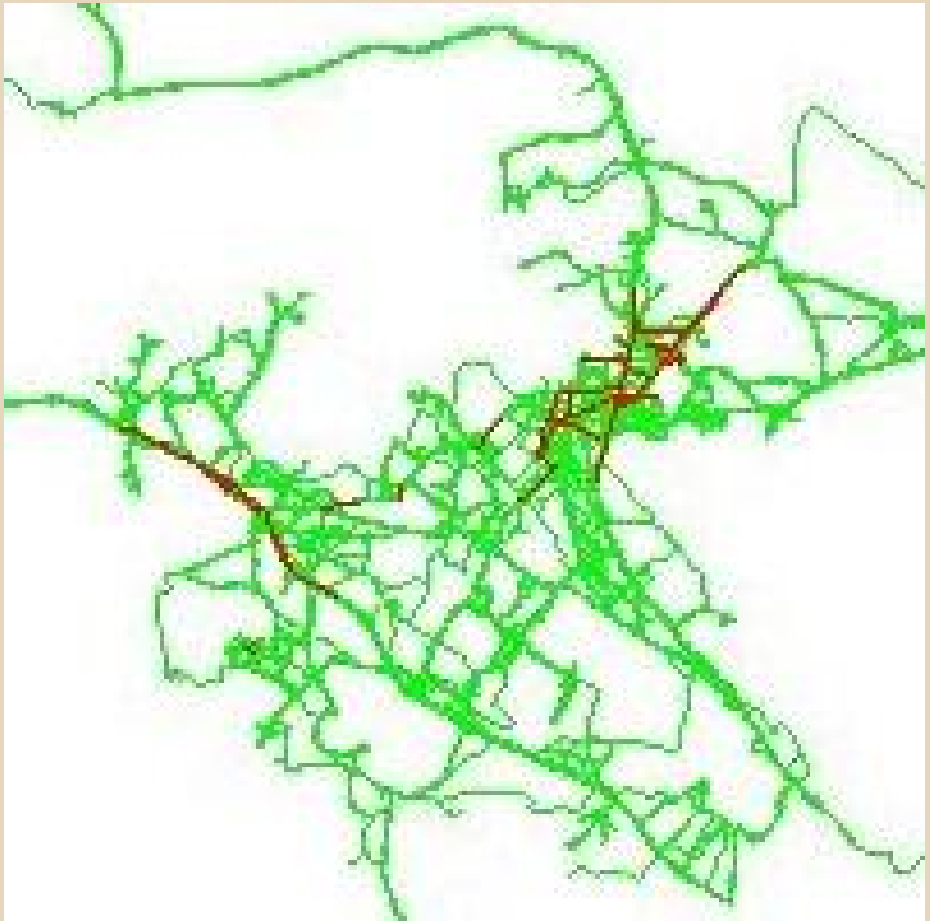
Masciari, E. (2012, November). Finding homogeneous groups in trajectory streams. In *Proceedings of the Third ACM SIGSPATIAL International Workshop on GeoStreaming* (pp. 11-18). ACM.

Algorithm Comparison

UCF



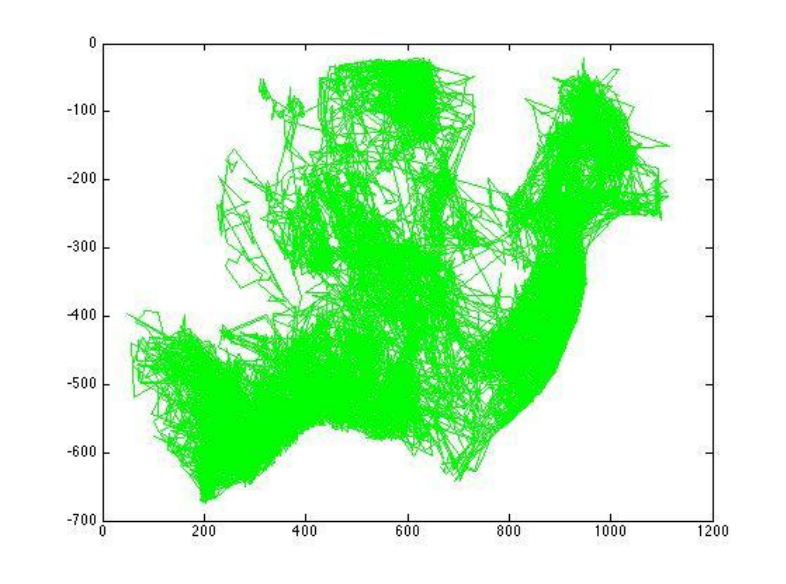
TraClus



Oregon Wildlife

Deer 1995 dataset

- 32 trajectories
- 20,065 data points



Used in the following papers:

•J. gil Lee and J. Han. Trajectory clustering: A partition-and-group framework. In Proceedings of the ACM International Conference on Management of Data (SIGMOD), [cited by 357]

•Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. 2008. *TraClass*: trajectory classification using hierarchical region-based and trajectory-based clustering. *Proc. VLDB Endow.* [cited by 83]

•Zhenhui Li, Jae-Gil Lee, Xiaolei Li, and Jiawei Han. 2010. Incremental clustering for trajectories. In Proceedings of the 15th international conference on Database Systems for Advanced Applications - Volume Part II (DASFAA'10), [cited by 31]

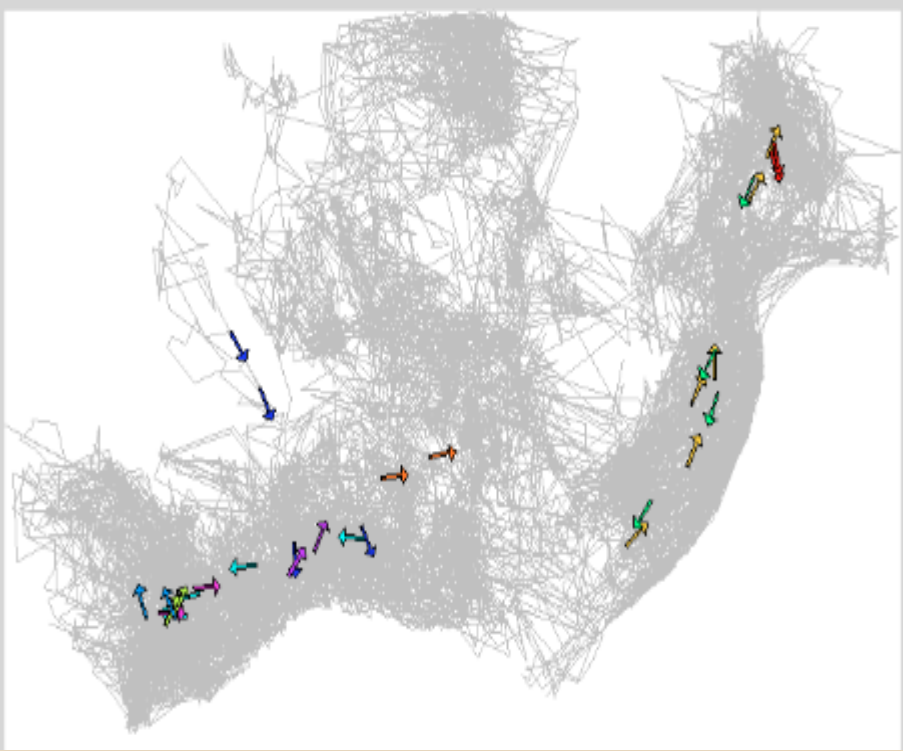
•Elio Masciari. 2009. A Complete Framework for Clustering Trajectories. In Proceedings of the 2009 21st IEEE International Conference on Tools with Artificial Intelligence (ICTAI '09). [cited by 2]

•Yu Zhang and Dechang Pi. 2009. A Trajectory Clustering Algorithm Based on Symmetric Neighborhood. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 03 (CSIE '09), [cited by 2[

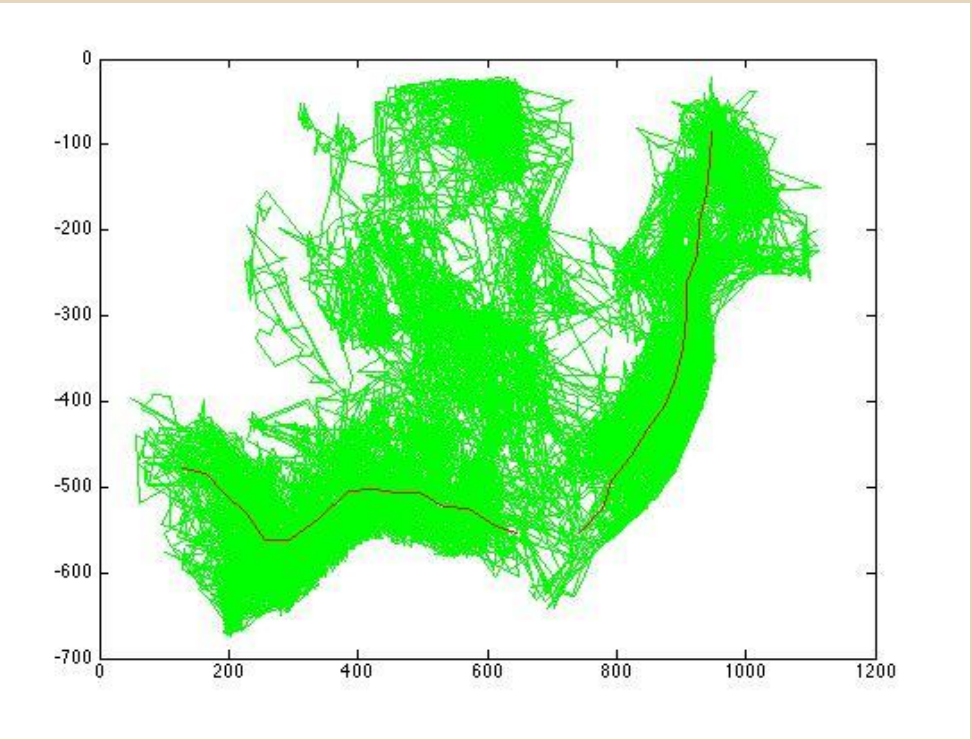
•Jae-Gil Lee, Jiawei Han, and Xiaolei Li. 2008. Trajectory Outlier Detection: A Partition-and-Detect Framework. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE '08).[cited by 112]

Performance Comparison

UCF



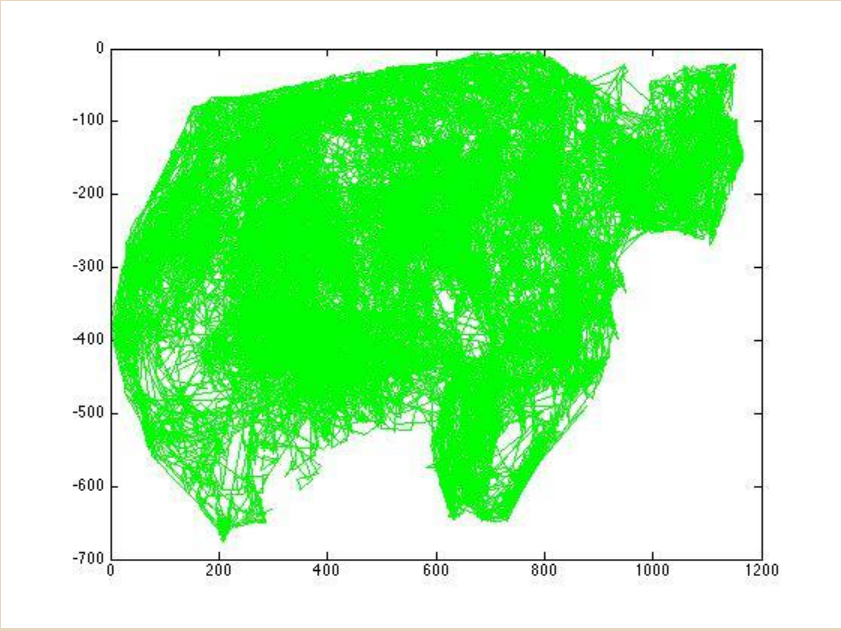
TRACLUS



Oregon Wildlife

Elk 1993 dataset

- 43 trajectories
- 47,204 points



Used in the following papers:

•J. gil Lee and J. Han. Trajectory clustering: A partition-and-group framework. In Proceedings of the ACM International Conference on Management of Data (SIGMOD),[cited by 357]

•Zhenhui Li, Jae-Gil Lee, Xiaolei Li, and Jiawei Han. 2010. Incremental clustering for trajectories. In Proceedings of the 15th international conference on Database Systems for Advanced Applications - Volume Part II (DASFAA'10), [cited by 31]

•Elio Masciari. 2009. A Complete Framework for Clustering Trajectories. In Proceedings of the 2009 21st IEEE International Conference on Tools with Artificial Intelligence (ICTAI '09). [cited by 2]

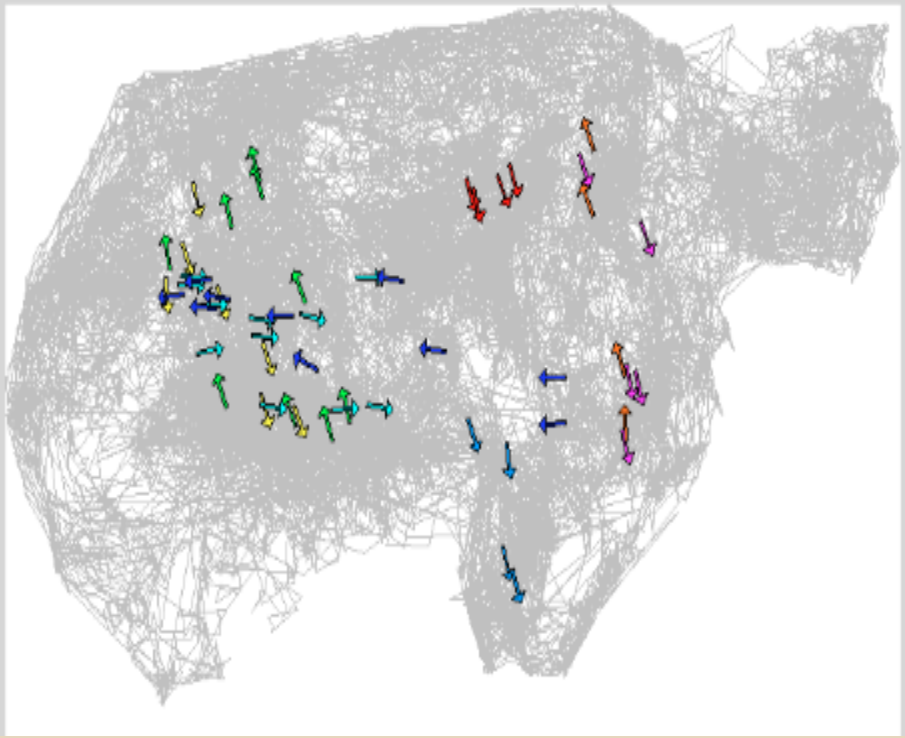
•Yu Zhang and Dechang Pi. 2009. A Trajectory Clustering Algorithm Based on Symmetric Neighborhood. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 03 (CSIE '09), [cited by 2]

•Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. 2008. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proc. VLDB Endow. [cited by 83]

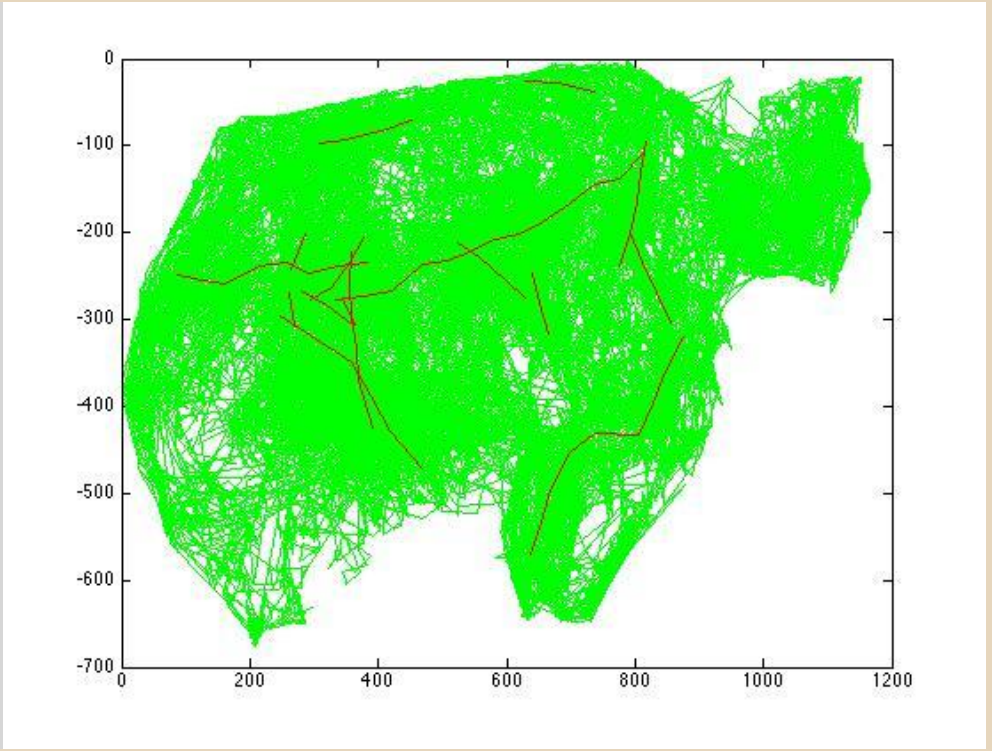
•Jae-Gil Lee, Jiawei Han, and Xiaolei Li. 2008. Trajectory Outlier Detection: A Partition-and-Detect Framework. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE '08). [cited by 112]

Performance Comparison

UCF



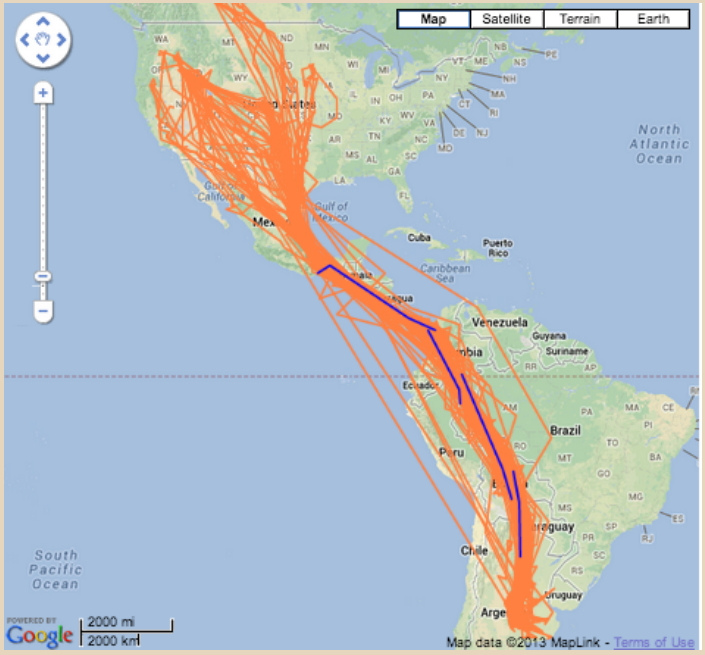
TRACLUS



MoveBank

Swainson's Hawks

- 43 trajectories
- 4,514 points

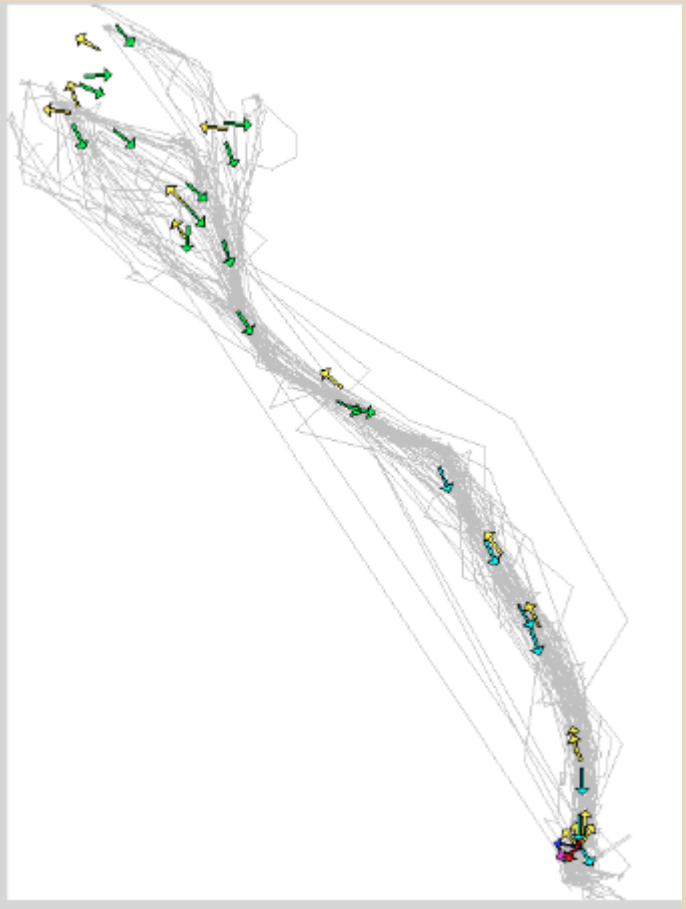


"Swainson's Hawks converged in eastern Mexico on the Gulf of Mexico coast. Southward, these hawks followed a narrow, well-defined path through Central America, across the Andes Mountains in Columbia, and east of the Andes to central Argentina where they all spent the austral summer. Swainson's Hawks northward migration largely retraced their southward route."

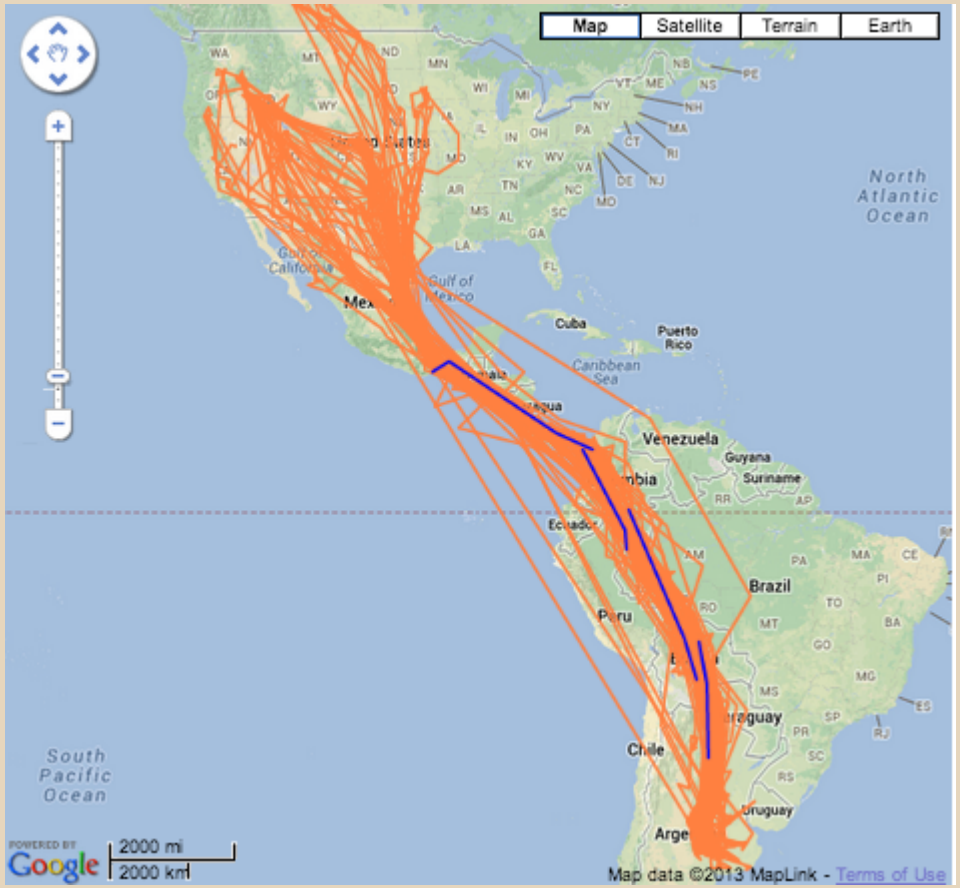
Fuller, M.R., Seegar, W.S., Schueck, L.S., 1998. Routes and Travel Rates of Migrating Peregrine Falcons Falco peregrinus and Swainson's Hawks Buteo swainsoni in the Western Hemisphere. Journal of Avian Biology 29:433-440.

Performance Comparison

UCF



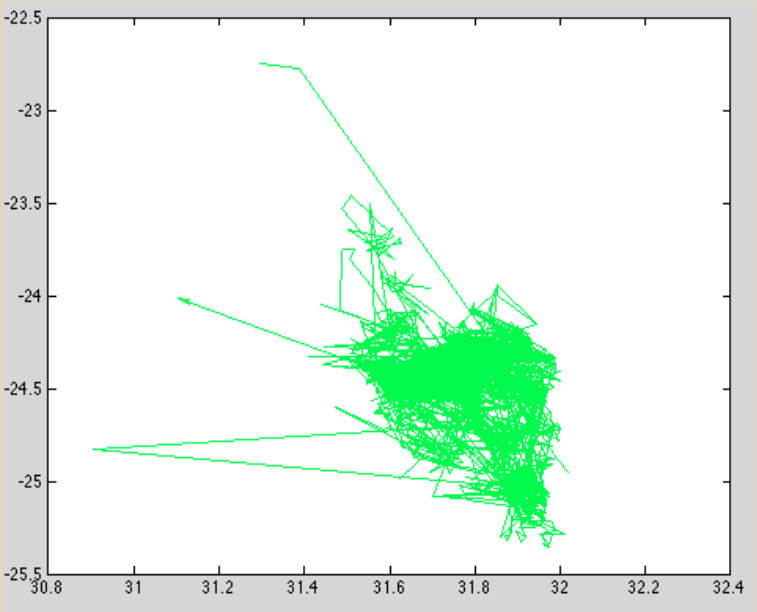
TRACLUS



Movebank

African Buffalo dataset

- 165 trajectories
- 26,609 points



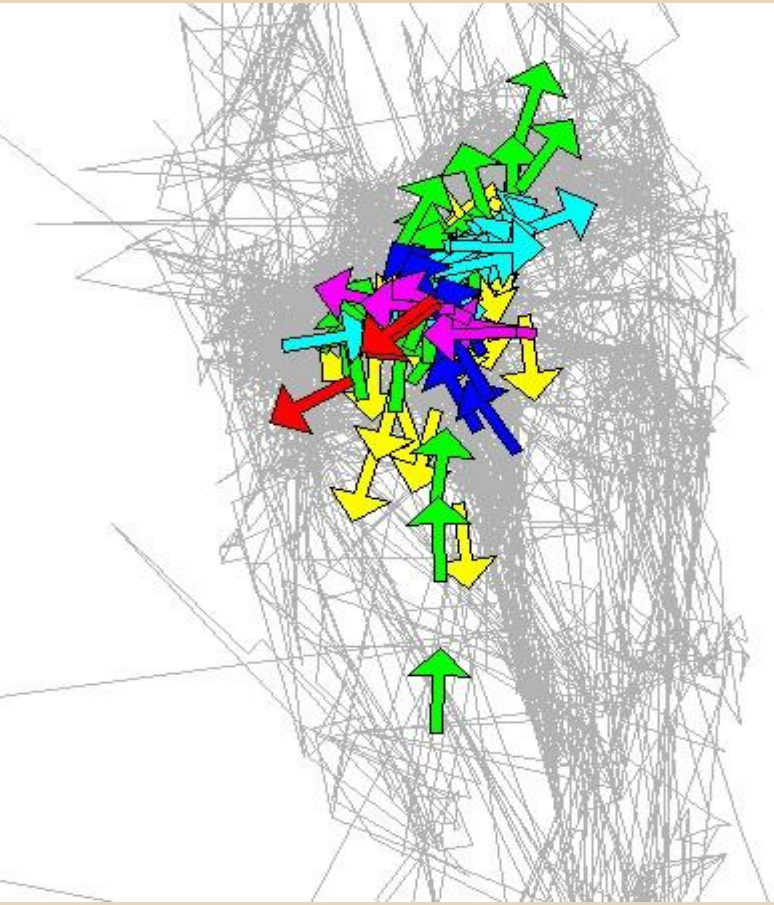
Used in the following works:

- Zhenhui Li, Ming Ji, Jae-Gil Lee, Lu-An Tang, Yintao Yu, Jiawei Han, and Roland Kays. 2010. MoveMine: mining moving object databases. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (SIGMOD '10). [cited by 29]

- Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays. 2010. Swarm: mining relaxed temporal moving object clusters. In *Procedings of VLDB Endowment.* [Cited by 41]

Performance Comparison

UCF



TRACLUS

