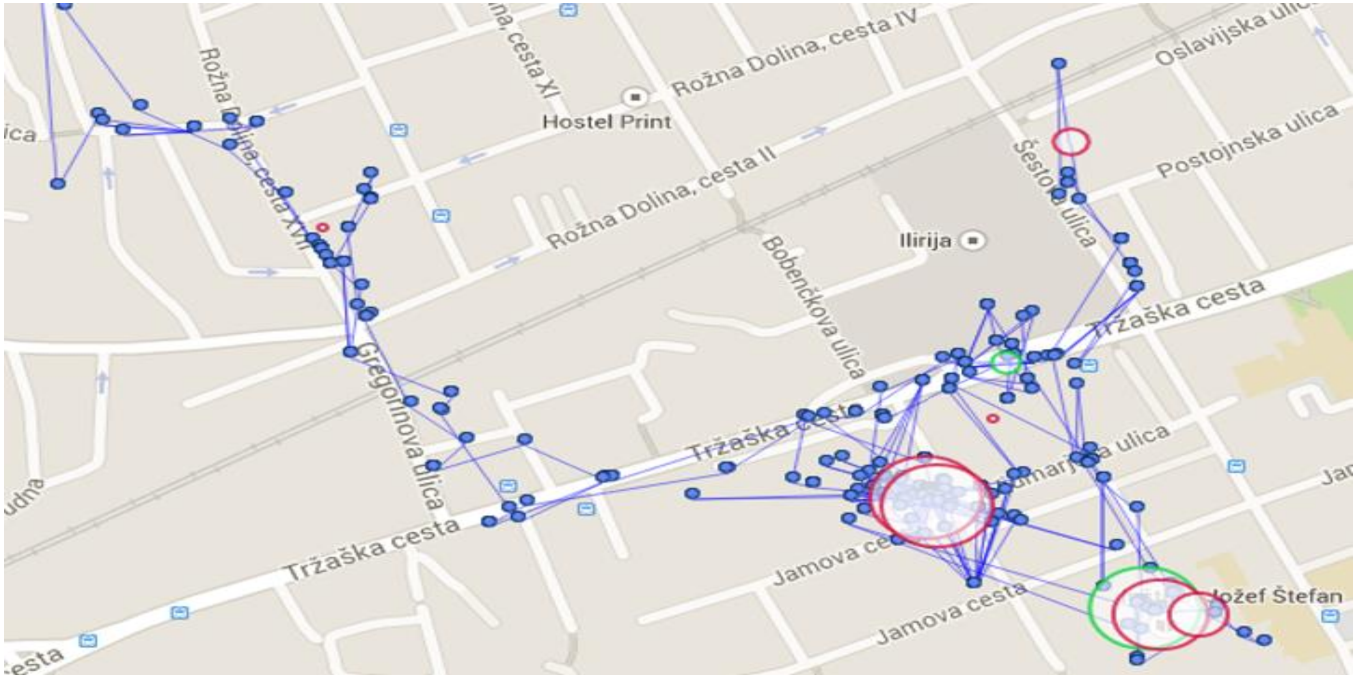
Spatio-temporal clustering methods

AUTHORS: **MATEJ SENOŽETNIK**, LUKA BRADEŠKO, BLAŽ KAŽIČ, DUNJA MLADENIĆ, TINE ŠUBIC

LJUBLJANA, 10/10/2016

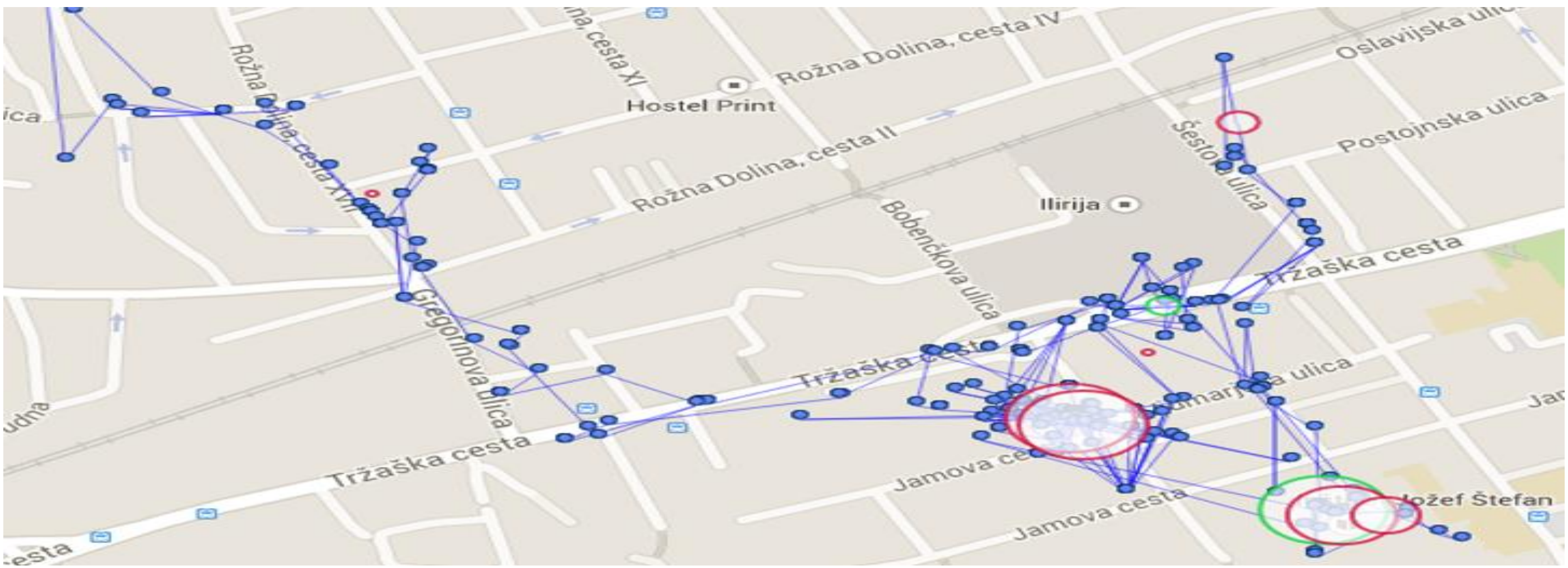
Motivation

- Analyzing user paths or behavior, measuring traffic congestion
- Application which can track coordinates from GPS
- Raw data doesn’t provide any useful information
- Stay points



Motivation (2)

- Process is called clustering
- Detects: Frequently visited locations, next place prediction



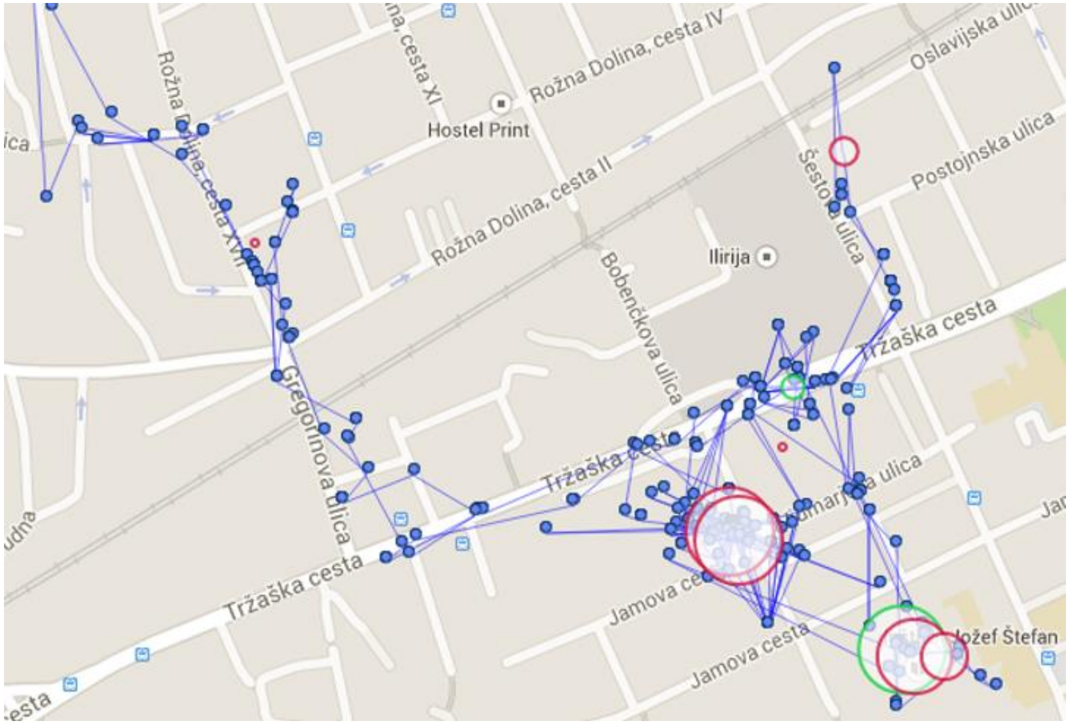
Outline

- Type of clustering data
  - Appropriate for clustering spatio-temporal data?
- Comparison of algorithms
- Find the most appropriate algorithm for clustering spatio-temporal data

Type of clustering data

Clustering methods are, in general, separated into the following categories:

- Partitioning methods
- Hierarchical methods
- **Density-based methods**



Algorithms: A comparison

	Spatio-temporal	Noise sensitivity	Returning stay points/ paths
DBSCAN [2]	No	No	No
ST-DBSCAN [5]	Yes	No	No
SMoT [1]	Yes	Yes	Yes
CB-SMoT [6]	Yes	No	Yes
SPD [7]	Yes	Yes	Yes
OPTICS [8]	Yes	No	No

Density Based Spatial Clustering of Application with Noise (DBSCAN)

Density-based clustering algorithm which identifies arbitrary-shaped objects and detects noise in a dataset

Advantages:

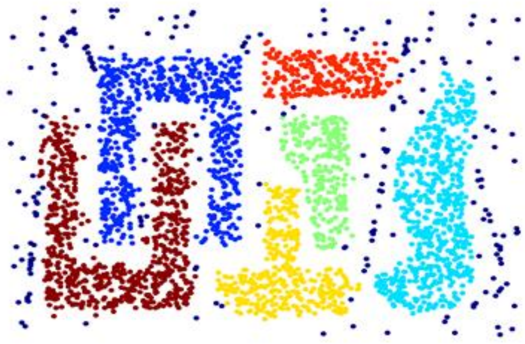
- Robustly detects outliers
- Appropriate for large databases

Drawbacks:

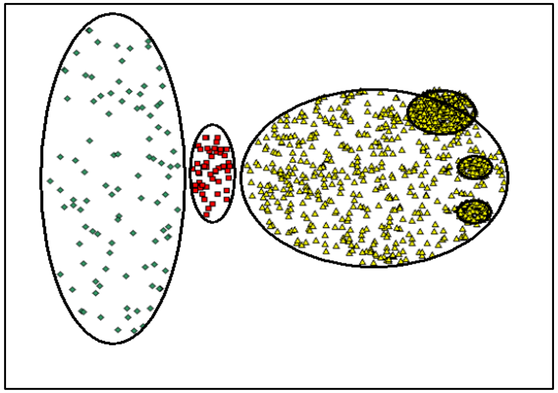
- Works on spatial elements
- Doesn‘t work on different densities.

Density Based Spatial Clustering of Application with Noise (DBSCAN (2))

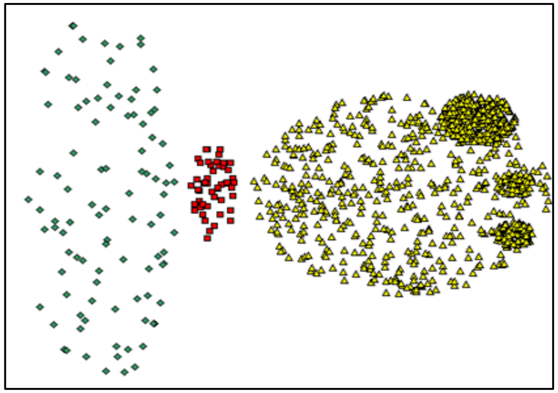
Original Points [3]



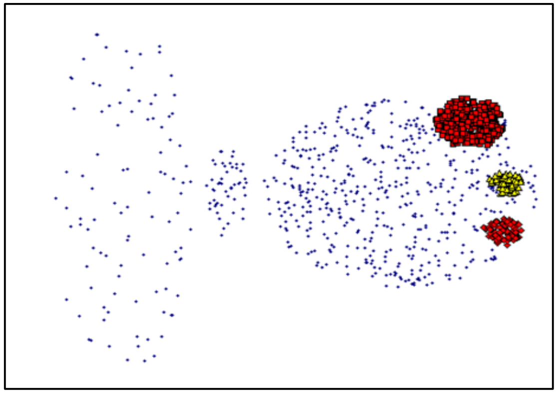
Clusters [3]



Original Points



(MinPts=4, Eps=9.92)



(MinPts=4, Eps=9.75)

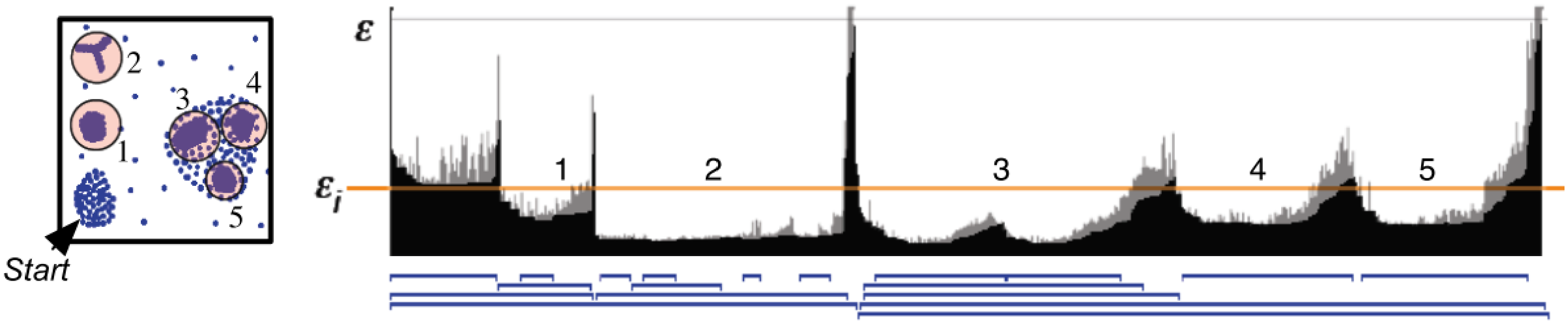
Spatio-Temporal Density Based Spatial Clustering of Application with Noise (ST-DBSCAN)

Improves DBSCAN in three ways:

- Takes into account temporal data
- Identifies noisy objects if there are various densities of the input data
- More accurately differentiates adjacent clusters

Ordering Points to Identify the Clustering Structure (OPTICS)

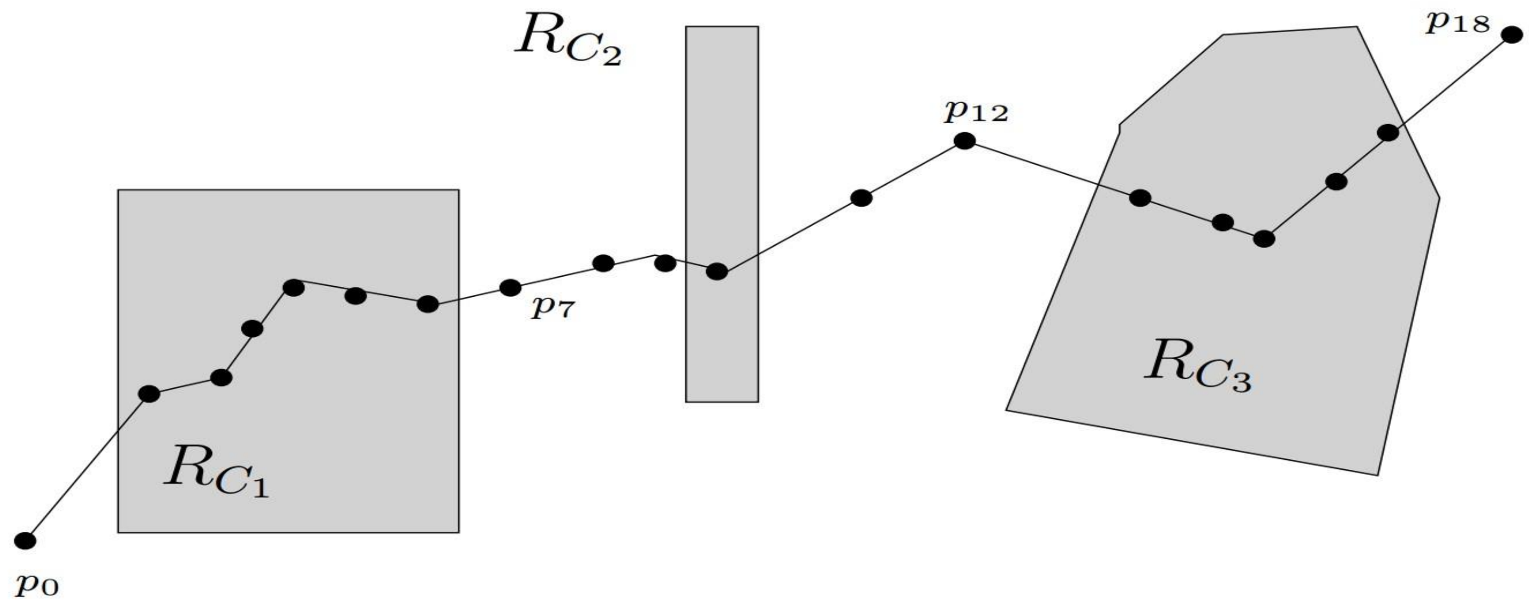
- Finding density-based clusters in spatio-temporal data
- Improves on DBSCAN's biggest weakness, the failure to detect clusters when density of the data varies



[3]

Stops and Moves of Trajectory (SMoT)

- Stop candidates R_{C1}, R_{C2}, R_{C3}



[4]

Clustering-Based Stops and Moves of Trajectory (CB-SMoT)

- SMoT drawback is incapable of detecting stay points that are not predefined by user
- Idea behind of this method:
  - We move slower than when we are traveling from one place to another
- Less incorrect stop compared to SMoT algorithm

Stay Point Detection (SPD)

Based on time and distance thresholds.

Advantages:

- Not need of predefined structures
- Computationally inexpensive

Drawbacks:

- Sensitive to noise (can be partially reduced by adjusting the parameters)

Discussion

- Algorithm needs to have the following properties:
  - Is able to cluster spatio-temporal data
  - Is noise insensitive
  - Returns stay points and a paths

	Spatio-temporal	Noise sensitivity	Returning stay points/ paths
DBSCAN	No	No	No
ST-DBSCAN	Yes	No	No
SMoT	**Yes**	**Yes**	**Yes**
CB-SMoT	**Yes**	**No**	**Yes**
SPD	**Yes**	**Yes**	**Yes**
OPTICS	Yes	No	No

SMoT vs. CB-SMoT vs. SPD

- SMoT:
  - Uses predefined regions
- CB-SMoT:
  - Predefined ratio of stay points and paths
- SPD:
  - Sensitive to noise
  - False stay points/paths

Conclusion

- SPD algorithm
  - Suffers from detecting false stay points and paths
  - This problem can be alleviated by running multiple iterations of the algorithm on resulting dataset.
- Insensitive to noise (Wi-Fi, activity recognition)

Thank you for your attention!

References

- [1] Luis Otavio Alvares, Vania Bogorny, Bart Kuijpers, Jose Antonio Fernandes de Macedo, Bart Moelans, and Alejandro Vaisman. A model for enriching trajectories with semantic geographical information. In Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, GIS ’07, pages 22:1–22:8, New York, NY, USA, 2007. ACM.
- [2] Martin Ester, Hans-Peter Kriegel, Jorg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.
- [3]http://www.hypertextbookshop.com/dataminingbook/public_version/contents/chapters/chapter004/section004/blue/page003.html
- [4] Chris Mueller. Data Preparation. 2005.
- [5] Derya Birant and Alp Kut. ST-DBSCAN: An algorithm for clustering spatial{temporal data. Data & Knowledge Engineering, 60(1):208{221, 2007.
- [6] Andrey Tietbohl Palma, Vania Bogorny, Bart Kuijpers, and Luis Otavio Alvares. A clustering-based approach for discovering interesting places in trajectories. In Proceedings of the 2008 ACM Symposium on Applied Computing, SAC '08, pages 863{868, New York, NY, USA, 2008. ACM.
- [7] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. Mining user similarity based on location history. In Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in eographic Information Systems, GIS '08,pages 34:1{34:10, New York, NY, USA, 2008. ACM.
- [8] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and J�org Sander. Optics: Ordering points to identify the clustering structure. ACM Sigmod Record, pages 49{60, 1999.