



# WEKA extension for contiguity-constrained clustering method

Carlos Mello

Geraldo Zimbrão

Jano Souza

{carlosmello, zimbrao, jano}@cos.ufrj.br

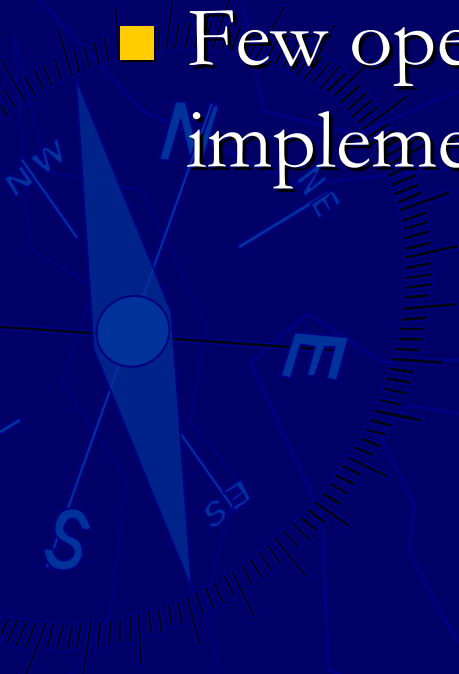


# Outline

- Introduction and motivation
- Goal
- WEKA
- Spatial clustering methods
- Contiguity-constrained clustering methods
- WEKA Extension
- Conclusion and future works

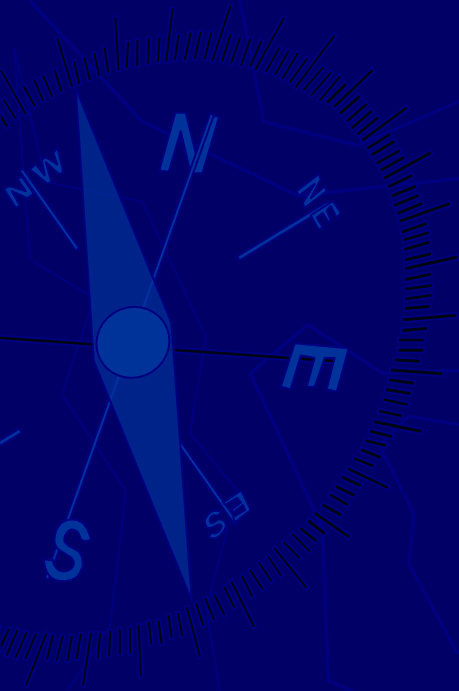
# Introduction and motivation

- Exponential growth of available huge amounts of spatial data
- Few open-source data mining tools which implement spatial data mining methods



# Goal

- Implementing a WEKA extension for spatial contiguity-constrained clustering methods

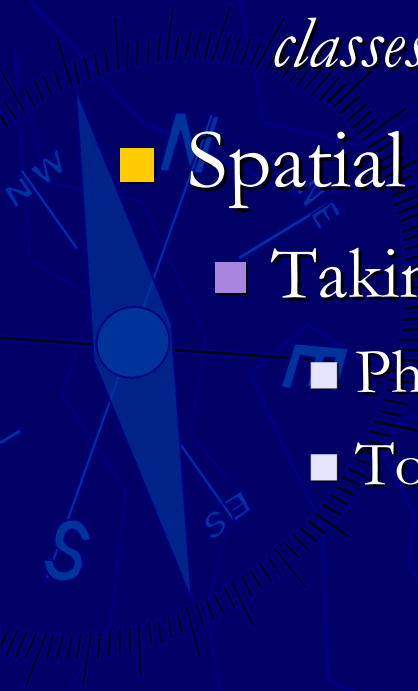


# WEKA

- Open-source
- More than 20 different conventional data mining methods have been implemented
- Some initiatives of spatial methods implementation have been developed
  - Borgony in (Borgony, 2006) presents an WEKA extension for spatial association rules method

# Spatial clustering methods

- Clustering methods
  - *“The process of grouping physical or abstract objects into classes of similar objects.” (Ng and Han, 1994)*
- Spatial clustering methods
  - Taking into account spatial information
    - Physical distance
    - Topology



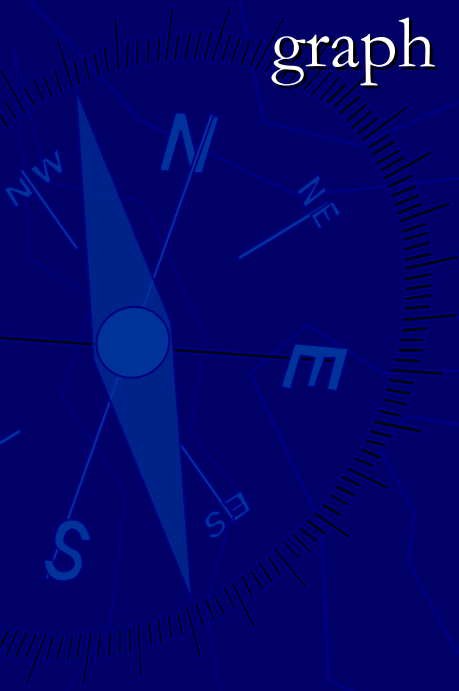
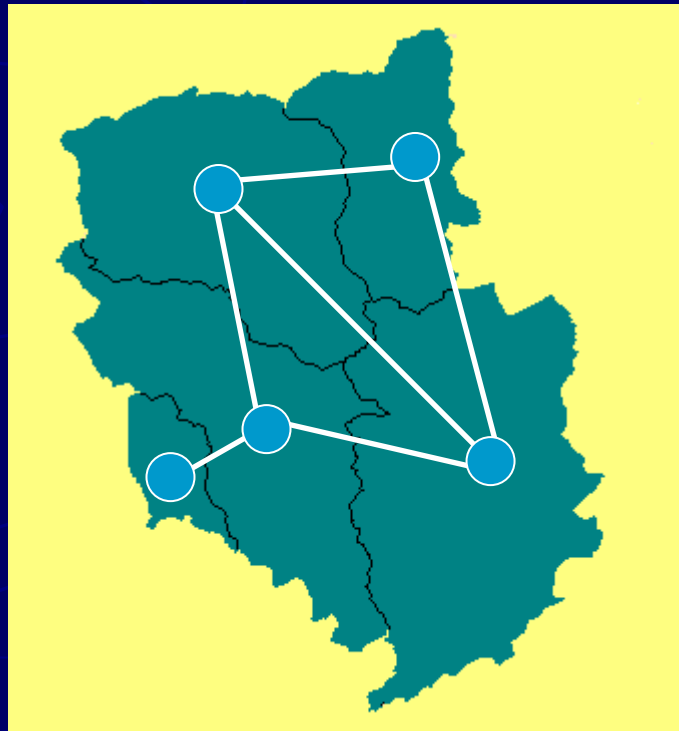
# Contiguity-constrained clustering methods

- Contiguity-constrained clustering using minimum spanning tree algorithm
  - Achieving the neighbourhood information in a matrix or graph
  - Using a minimum spanning tree (MST) to summarize the neighbourhood structure



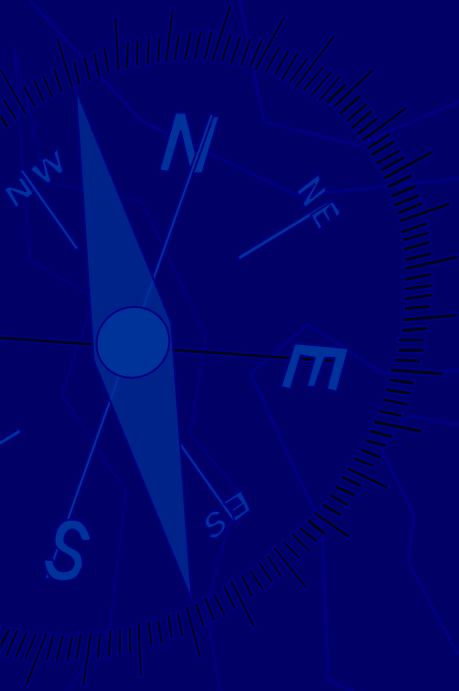
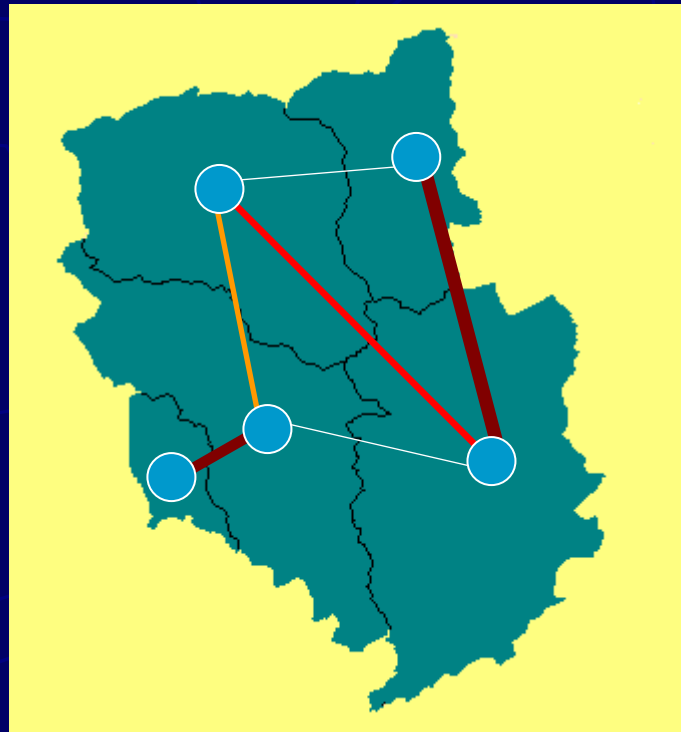
# Contiguity-constrained clustering methods

- Steps:
  - Reading and building the neighbourhood matrix or graph



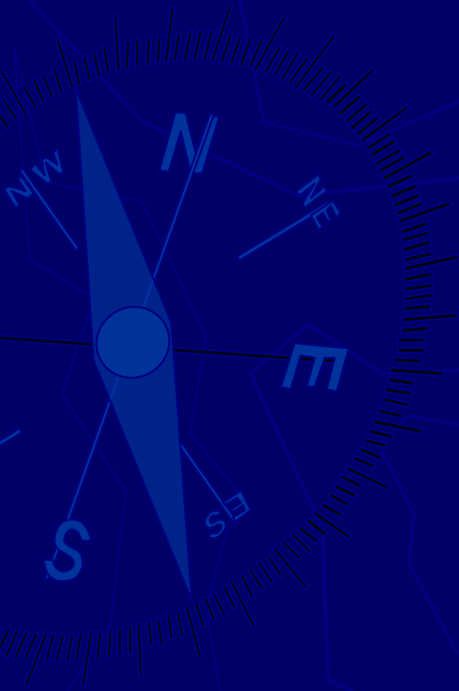
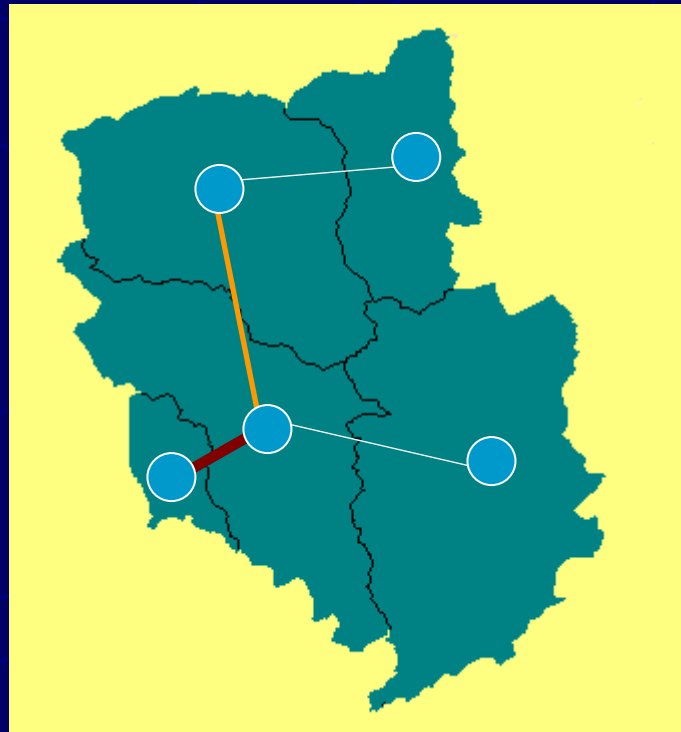
# Contiguity-constrained clustering methods

- Calculating the weight of edge which holds on two nodes



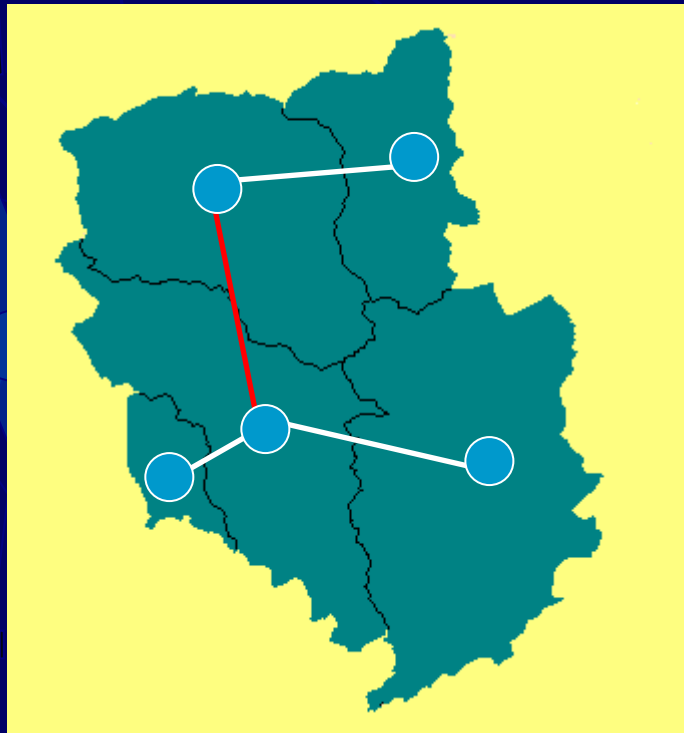
# Contiguity-constrained clustering methods

- Building the MST



# Contiguity-constrained clustering methods

- Pruning the MST



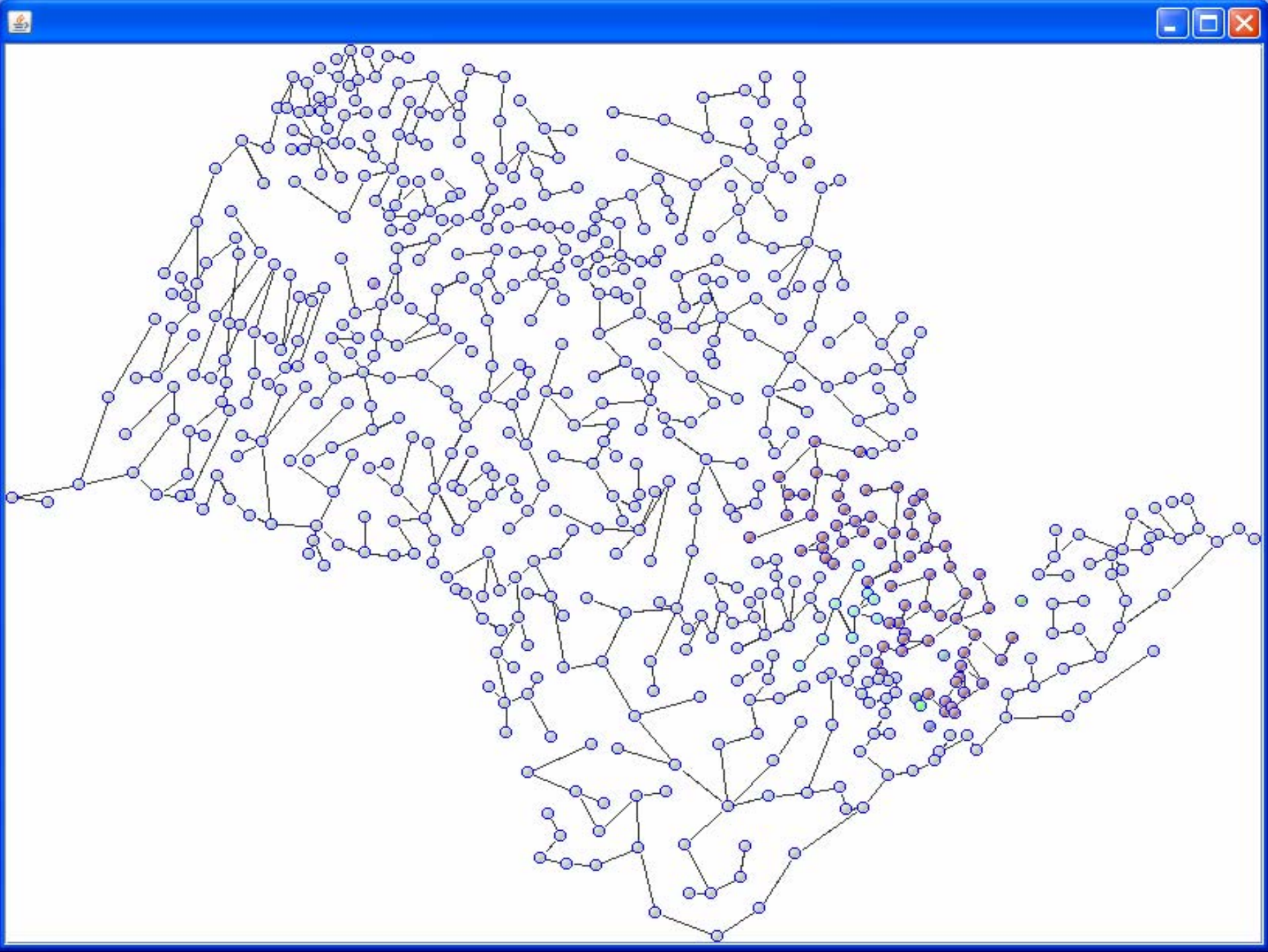
# WEKA Extension

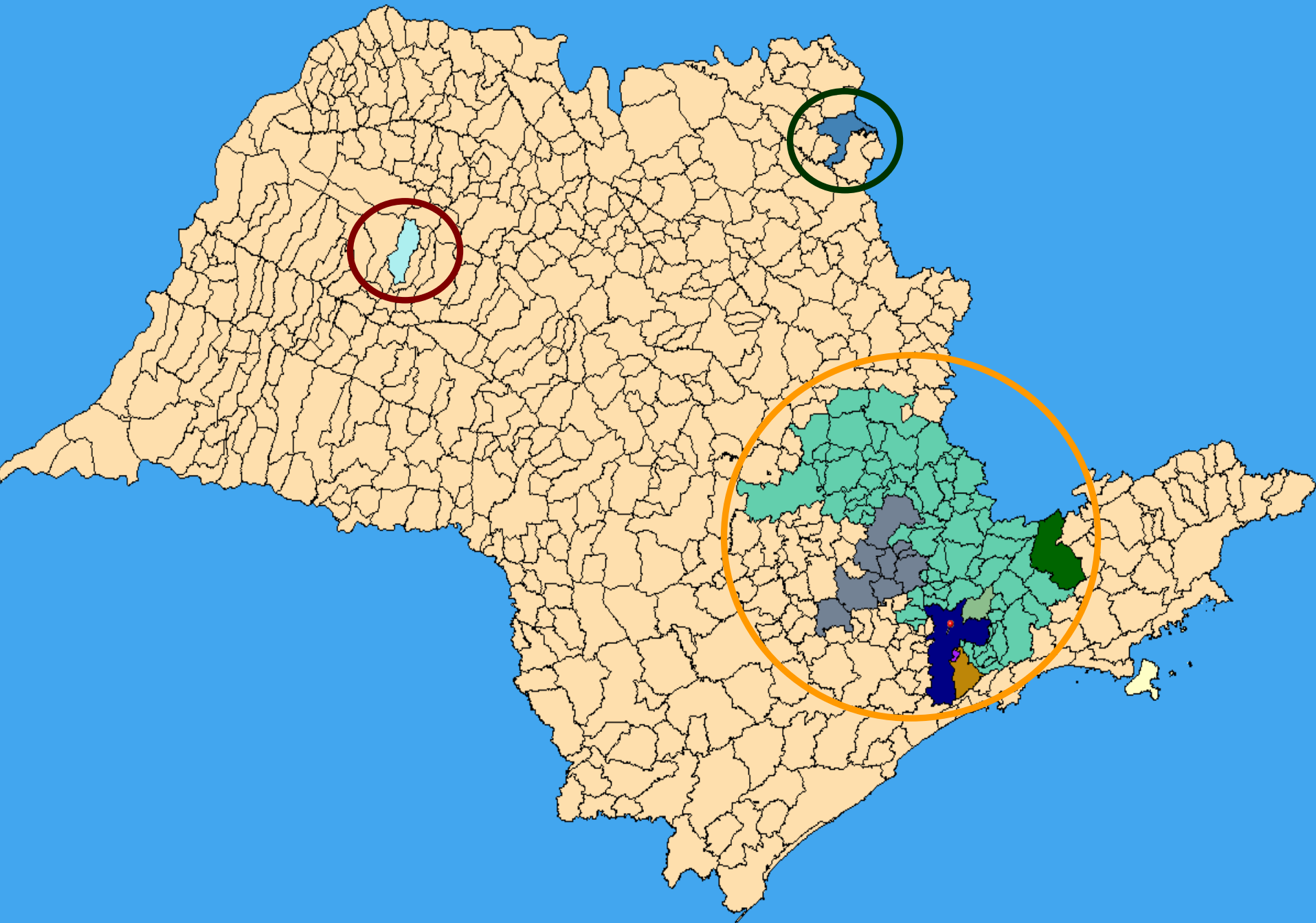
- PostGis integration
- Building the neighbourhood graph from a database table or view
- Load convencional data from a database table or view
- Visualization of neighbourhood graph, MST and Clusters
- Storing the results in a database table

# WEKA Extension

- Application example
  - Clustering the São Paulo state's cities from the number of companies in each economic activities of the industrial sector







# Conclusion and future works

## ■ Conclusion

- Implementing an WEKA extension for spatial clustering methods facilitates the implementation of new spatial clustering methods

## ■ Future work

- Implementing new spatial clustering methods based on this WEKA extension



# WEKA extension for contiguity-constrained clustering method

Carlos Mello

[carlosmello@cos.ufrj.br](mailto:carlosmello@cos.ufrj.br)

Thank you very much !!!