Universitat Pompeu Fabra Barcelona

A Heuristic Framework to Reduce Aggregation Errors in Large Classical Location Models

Daniel Serra, Carolina Castañeda P.

Department of Economics and Business
Universitat Pompeu Fabra

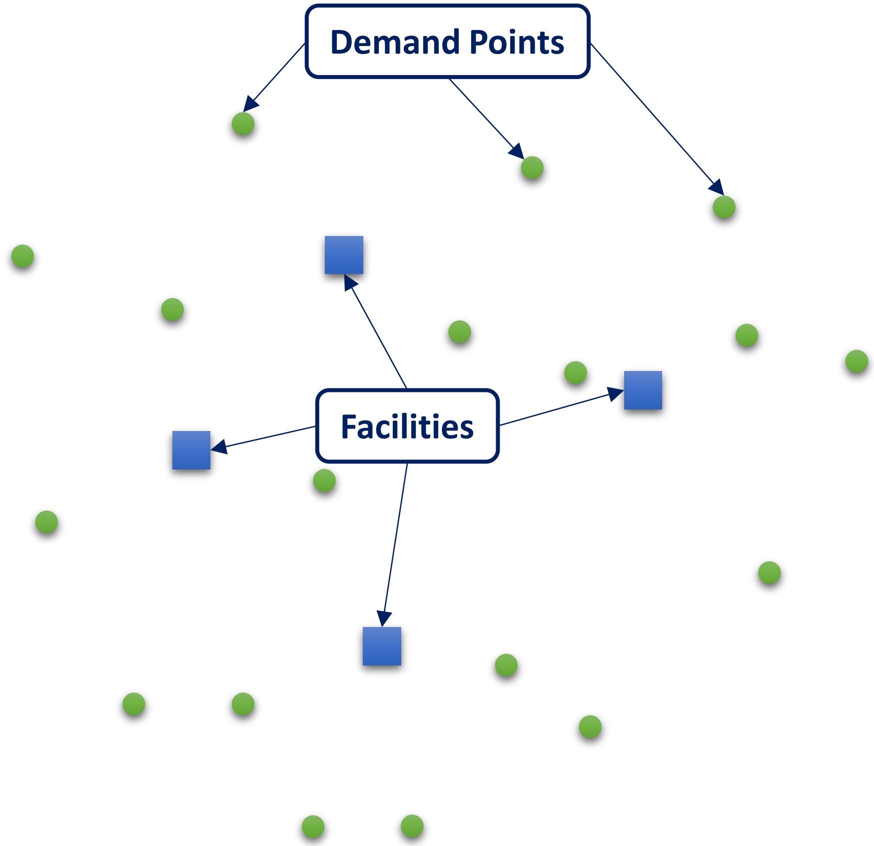
IX International Workshop on Locational Analysis and Related Problems

Cádiz, January 30 – February 1, 2019

Outline

1. Motivation
2. Aggregation Error
3. Heuristic Framework
4. Conclusions and Future Work

Motivation



Caracteristics:

- Georreferenced information
  Distance metrics
- Potential Facility Locations
- Demand Aggregation
  e.g. Postal Code

Motivation

- Location problems involve the decision of locating a set of facilities to satisfy one or more objective functions and constraints, regarding the demand for the service provided from the facilities.

- Finding good solutions to location problems helps to make better decisions in public (Marianov & Serra, 2004) and private contexts (Church & Murray, 2008).

- Applications of location models are diverse.

Motivation

Schools and Postal Offices
(Marianov & Serra et al., 2004)

Bank Offices
(Miliotis et al., 2002)

Logistics and Supply Chain
Management (Melo et al., 2009)

Retail Outlets
(Mendes & Themido, 2004)

Emergency Services
(Marianov, 2017; Marianov & Serra et al., 2004)

Telecomunications
(Gollowitzer & Ljubić I, 2011)

Routing and Transportation
(Nickel et al., 2001; Drexl & Schneider, 2015)

Motivation

- Location problems have recognized potential of interaction with other disciplines such as mathematics, engineering, economics, geography, regional science and logistics (Laporte & Nickel, 2015) (Murray, 2010) (Melo et al., 2009).
- Demand is composed by a very large number of customers
- Problems become intractable or unrealistic in terms of using traditional solution methods such as linear mixed binary programming.
- It is usual to aggregate demand points with the purpose of obtaining tractable and smaller models.

Error is induced!

Aggregation Error

- Crucial decisions in aggregation
  - Number of aggregated demand points
  - Location of aggregated points
  - How to assign the disaggregated demand points to the aggregated points
- The number of aggregated demand points determine the magnitude of error
- If possible, aggregation should be avoided

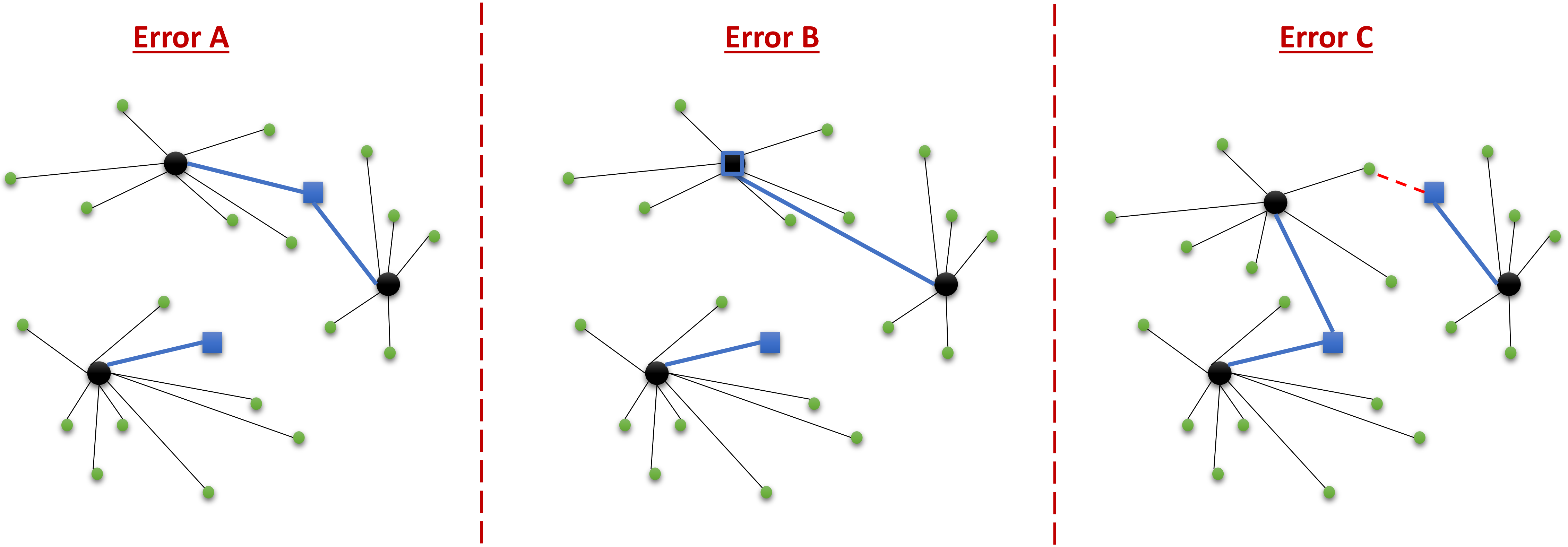
Aggregation Error

- Aggregation error was first formally defined by (Hillsman & Rhoda, 1978).
- The main classification of aggregation errors ABC type errors.
- Other error measures are based on ABC classification.
- There is no general agreement on how to measure error and depends on the location model
- Error measures are influenced by the distance metrics

Aggregation Error

- Errors type ABC (Hillsman & Rhoda, 1978)

Error A | **Error B** | **Error C**



Aggregation Error

- Frameworks to integrate aggregation algorithms and large location models
  - Francis et al. (2004) - framework for aggregating demand in a continuous space for covering and center models.
  - Plastria et al. (2006) - aggregation method that reduces the number of variables and constraints of the location model to decrease errors and avoid the optimality loss in competitive location problems.
  - Avella et al. (2012) - heuristic for large-scale *p-median* problem instances based on Lagrangean relaxation.
  - Jang & Lee (2015) - method to obtain near zero aggregation errors in covering problems avoiding the binary definition of coverage and using a random aggregation method.
  - Irawan & Salhi (2015) - multi-stage hybridization of a clustering-based technique and of a Variable Neighborhood Search (VNS) to solve large-scale *p-median* problems
  - Cebecauer & Buzna (2017) - adaptive aggregation framework for the facility location problems that keeps the problem size in reasonable limits.

Aggregation Error

- Research has been focused on developing better ways to measure the error caused by aggregation mostly in *p-median* problems
- There has been less interest in development of algorithms for making better aggregations and measure the impact in location problems solutions and aggregation errors.
- Few recent works have addressed aggregation for location models considering network distances.

Heuristic Framework

Composed by 4 stages
Use network distances

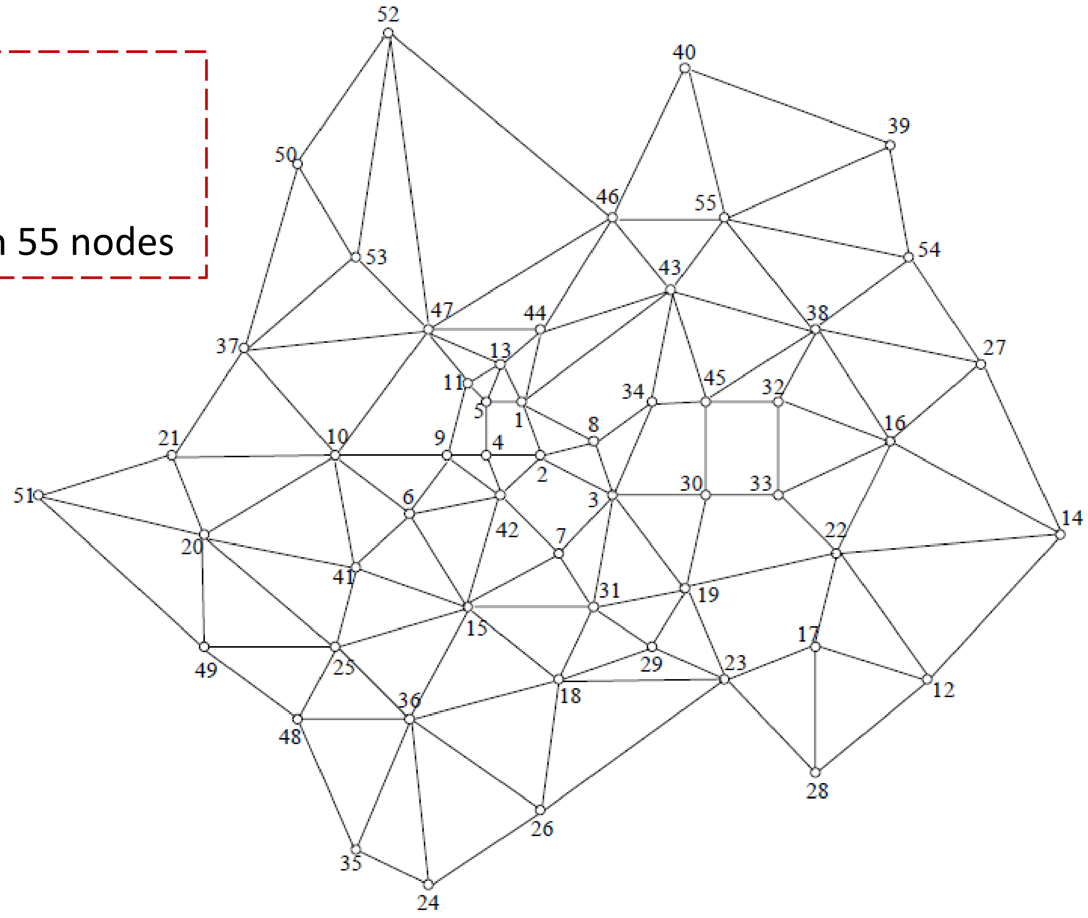
- Selection bias of points
- Groups based on distance (Super Nodes)

Stage 1
Aggregation

Heuristic Framework

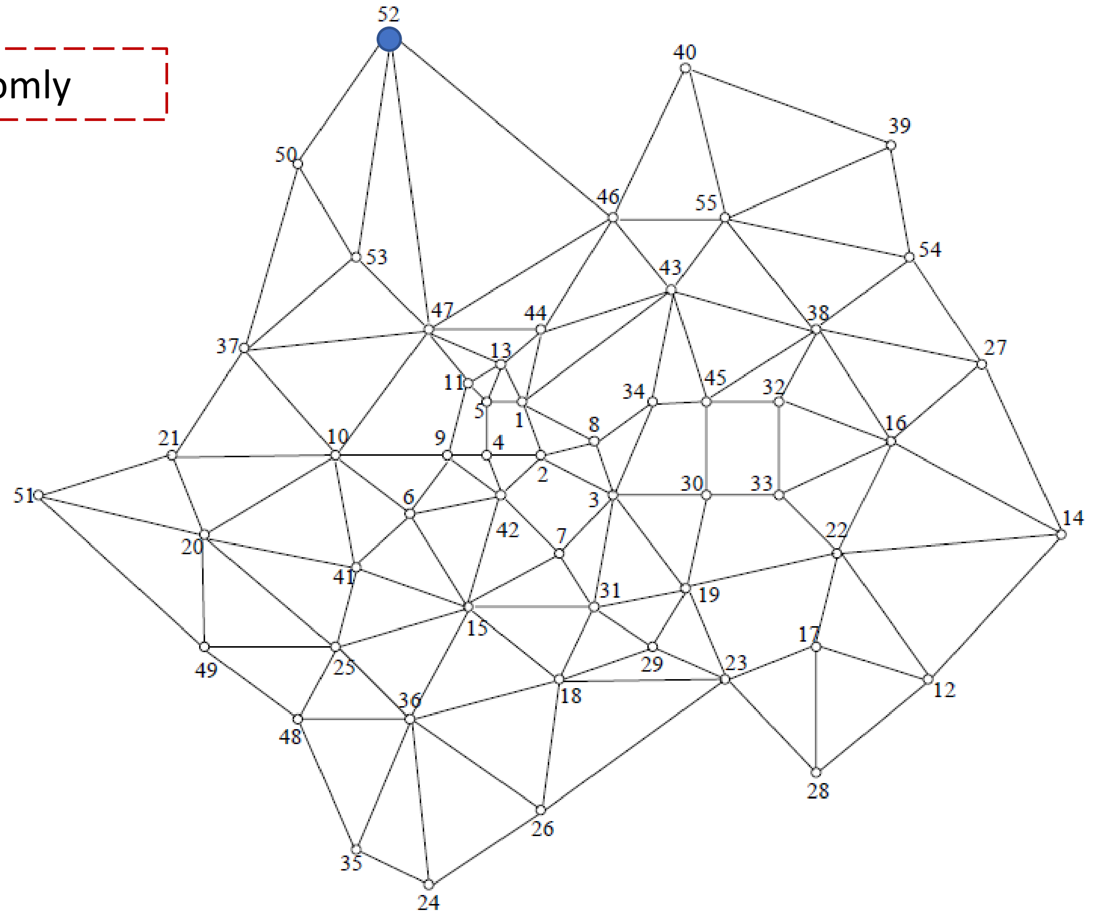
- Demand on nodes
- Distance on edges

- Example: Network with 55 nodes



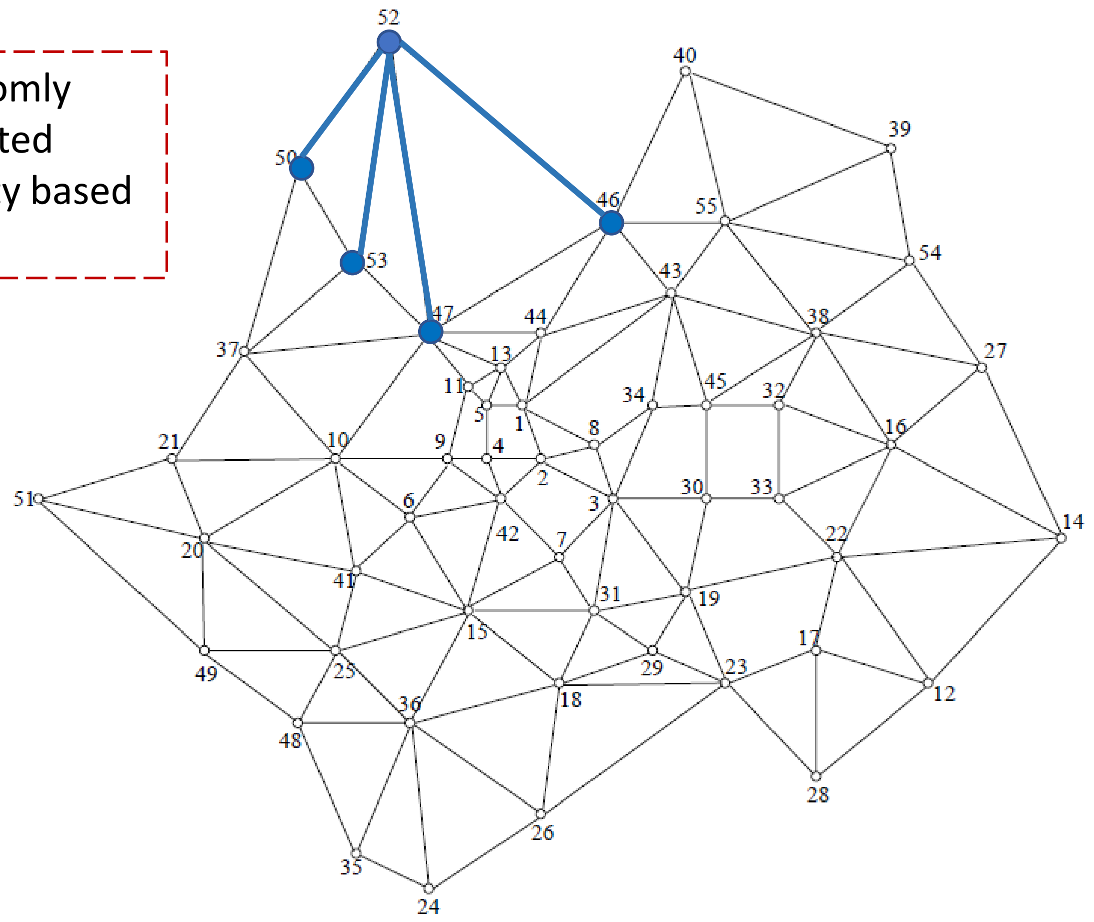
Heuristic Framework

1. Select one node randomly



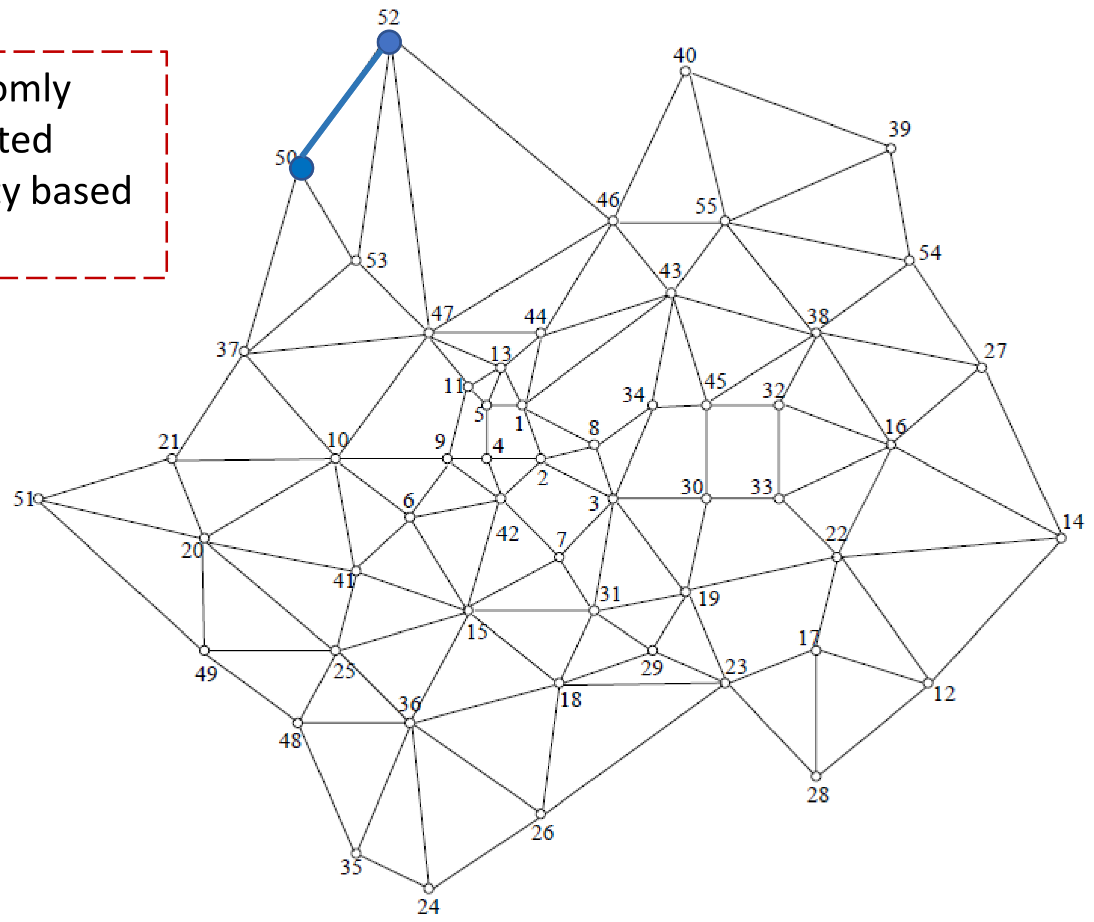
Heuristic Framework

1. Select one node randomly
2. Select a direct connected node with a probability based on distance



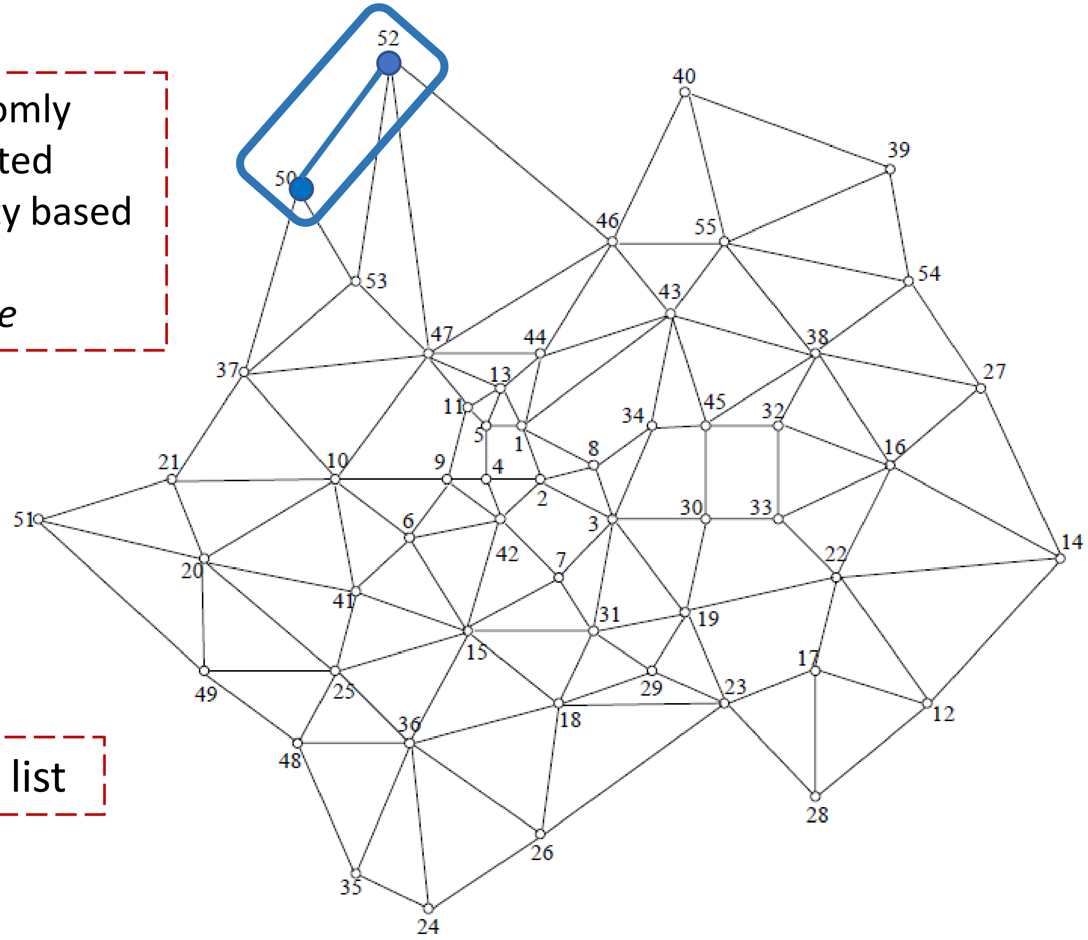
Heuristic Framework

1. Select one node randomly
2. Select a direct connected node with a probability based on distance



Heuristic Framework

1. Select one node randomly
2. Select a direct connected node with a probability based on distance
3. Create one *Super Node*

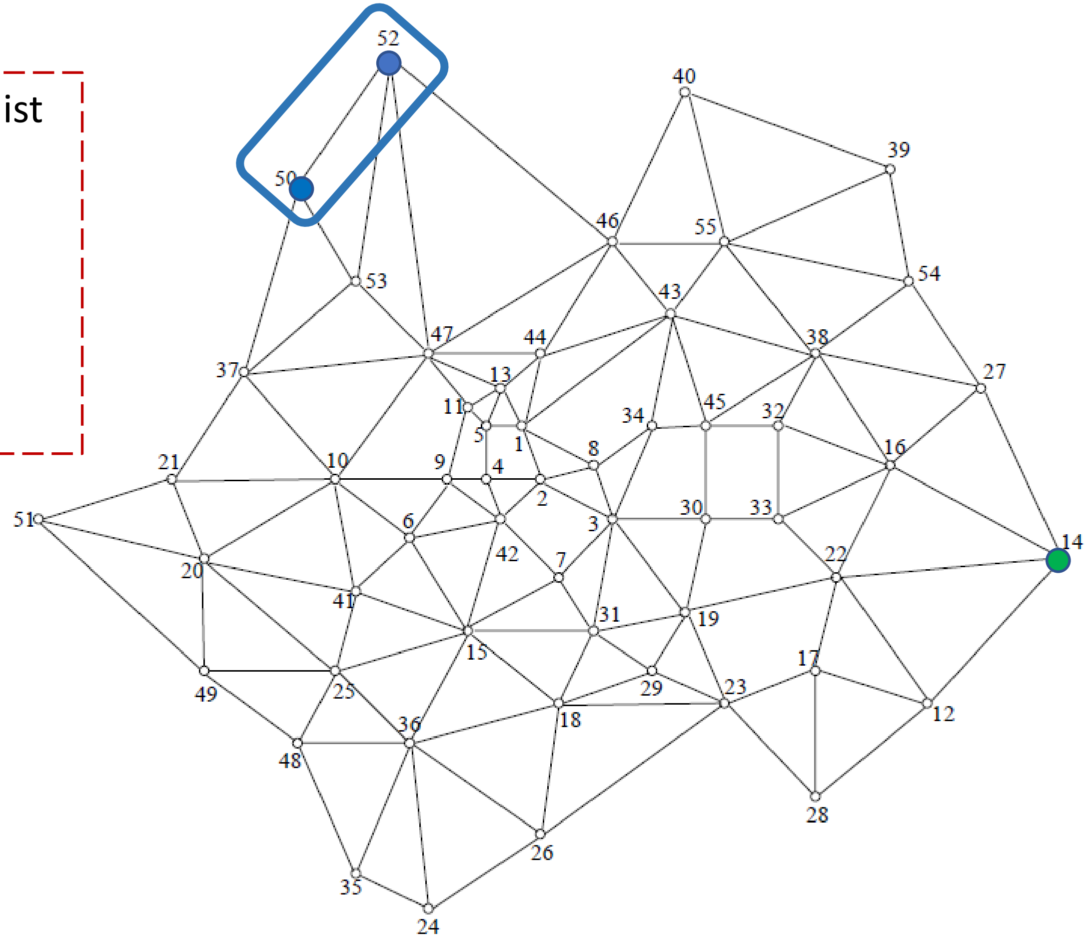


54 nodes in the eligible list

Heuristic Framework

Repeat the steps until the eligible list is empty

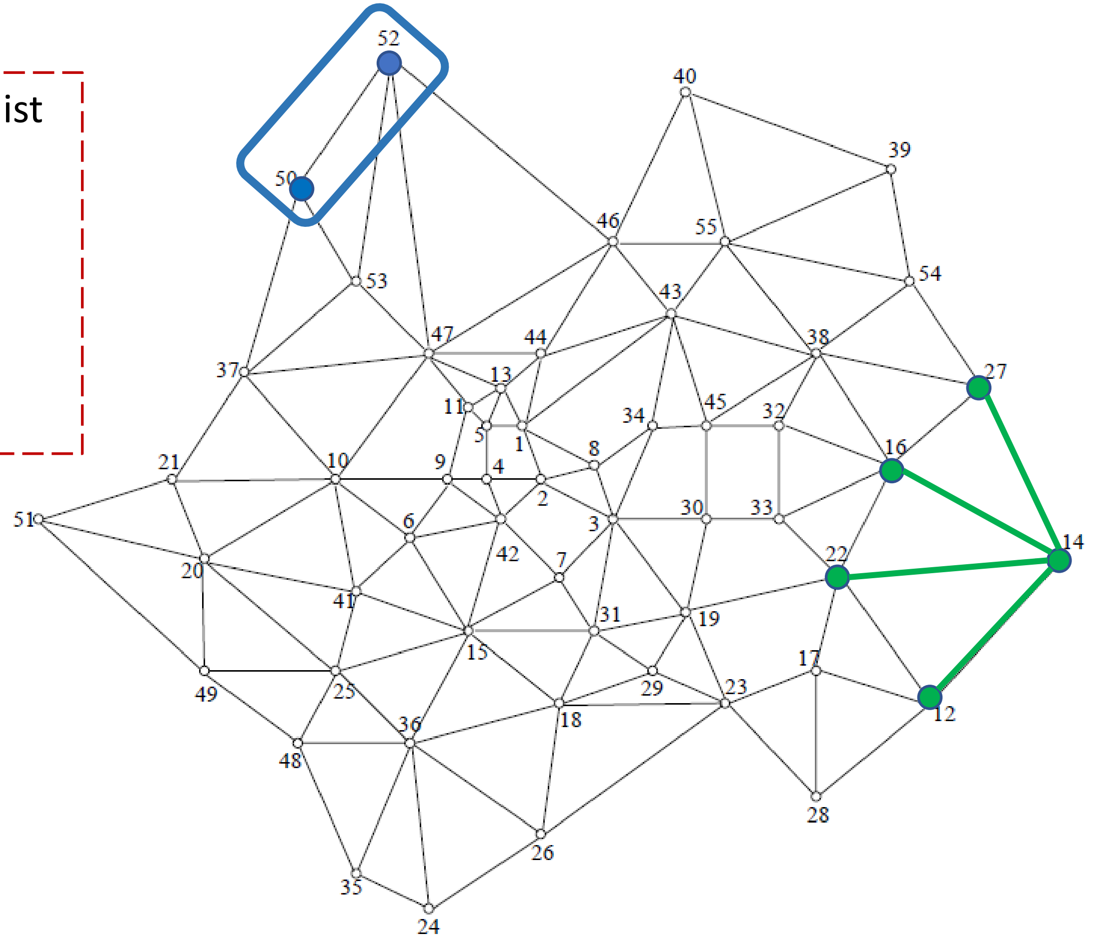
1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*



Heuristic Framework

Repeat the steps until the eligible list is empty

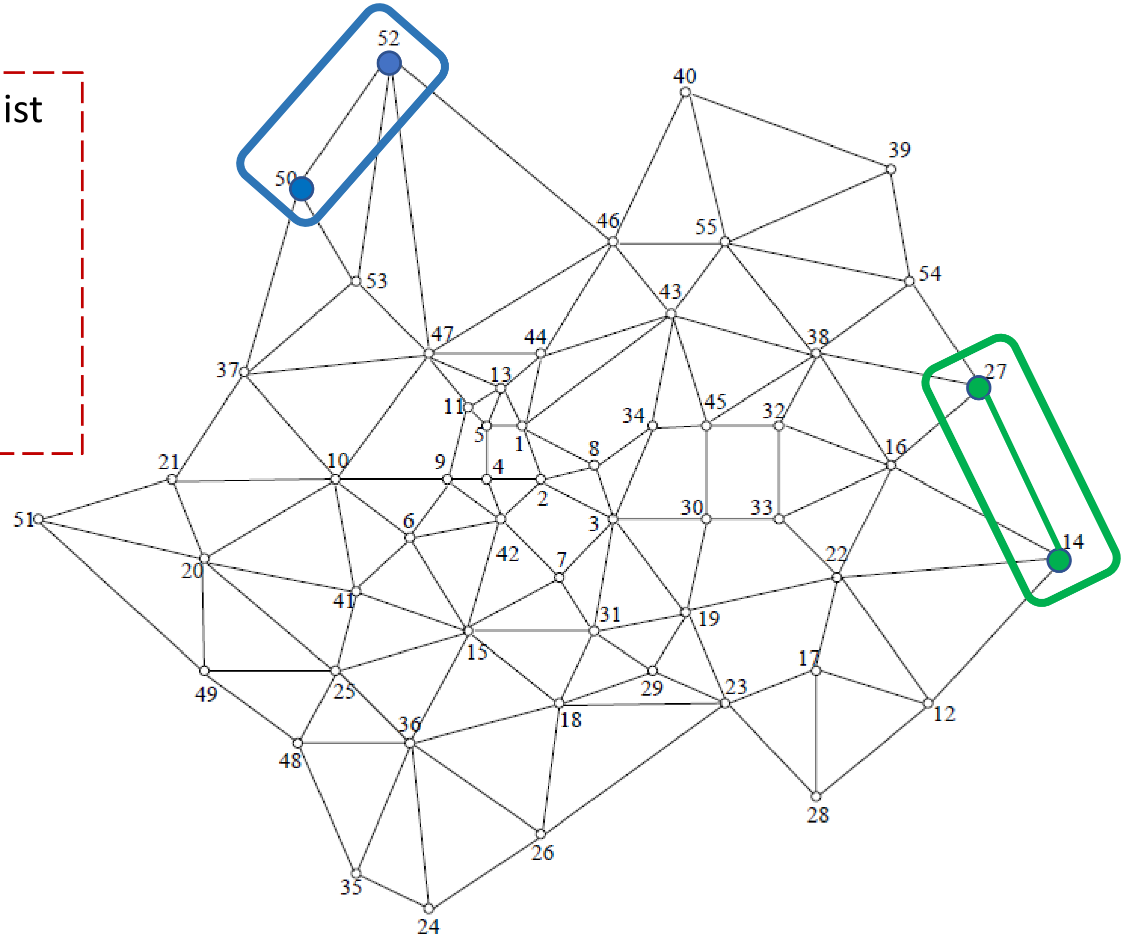
1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*



Heuristic Framework

Repeat the steps until the eligible list is empty

1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*

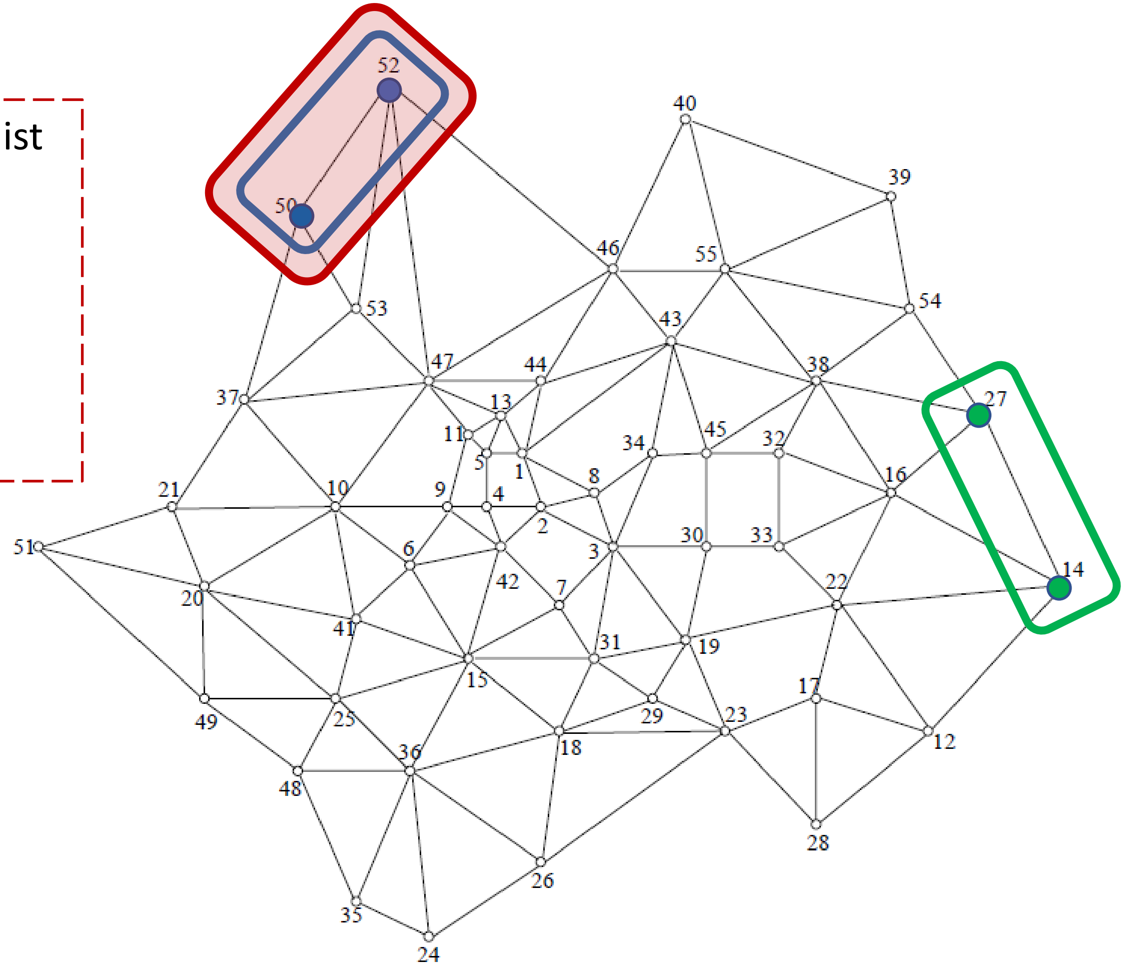


53 nodes in the eligible list
2 *Super Nodes*

Heuristic Framework

Repeat the steps until the eligible list is empty

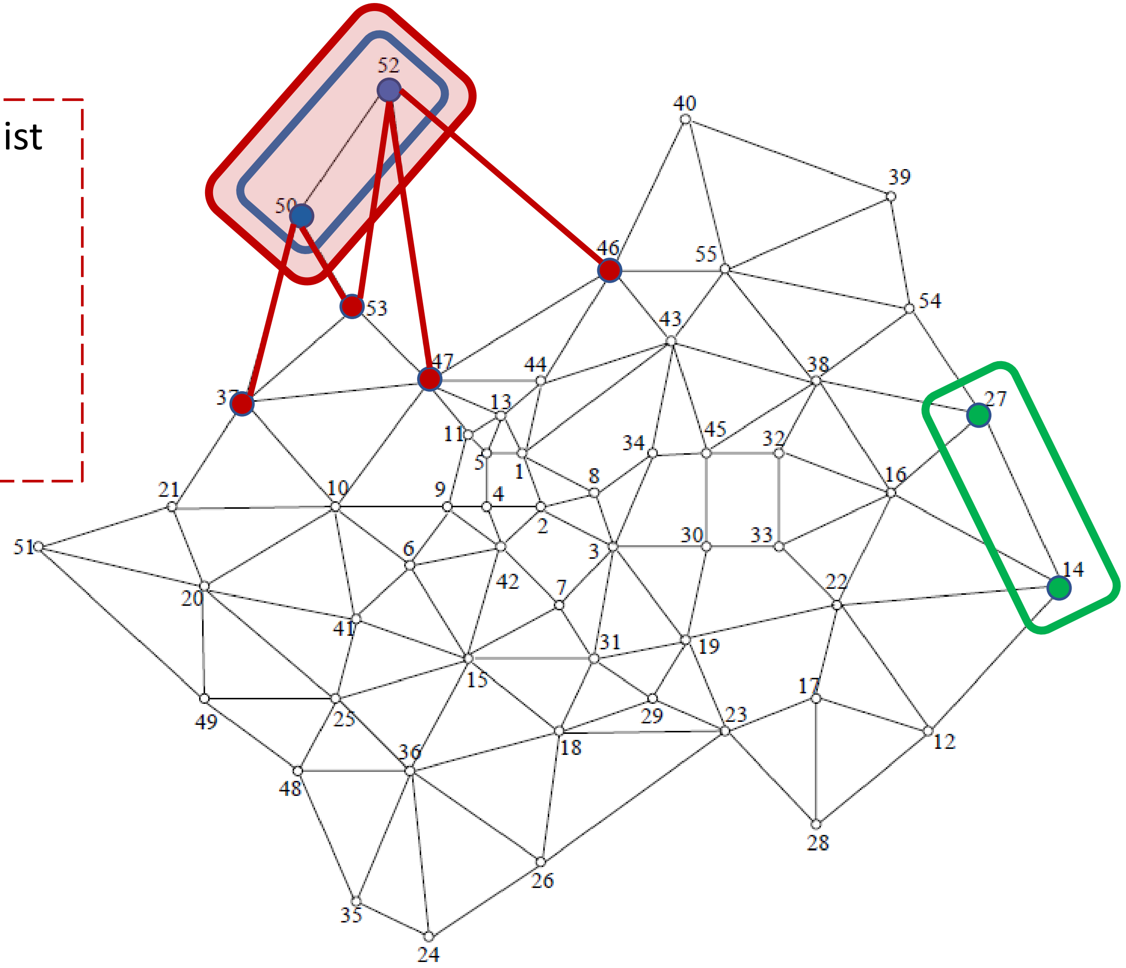
1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*



Heuristic Framework

Repeat the steps until the eligible list is empty

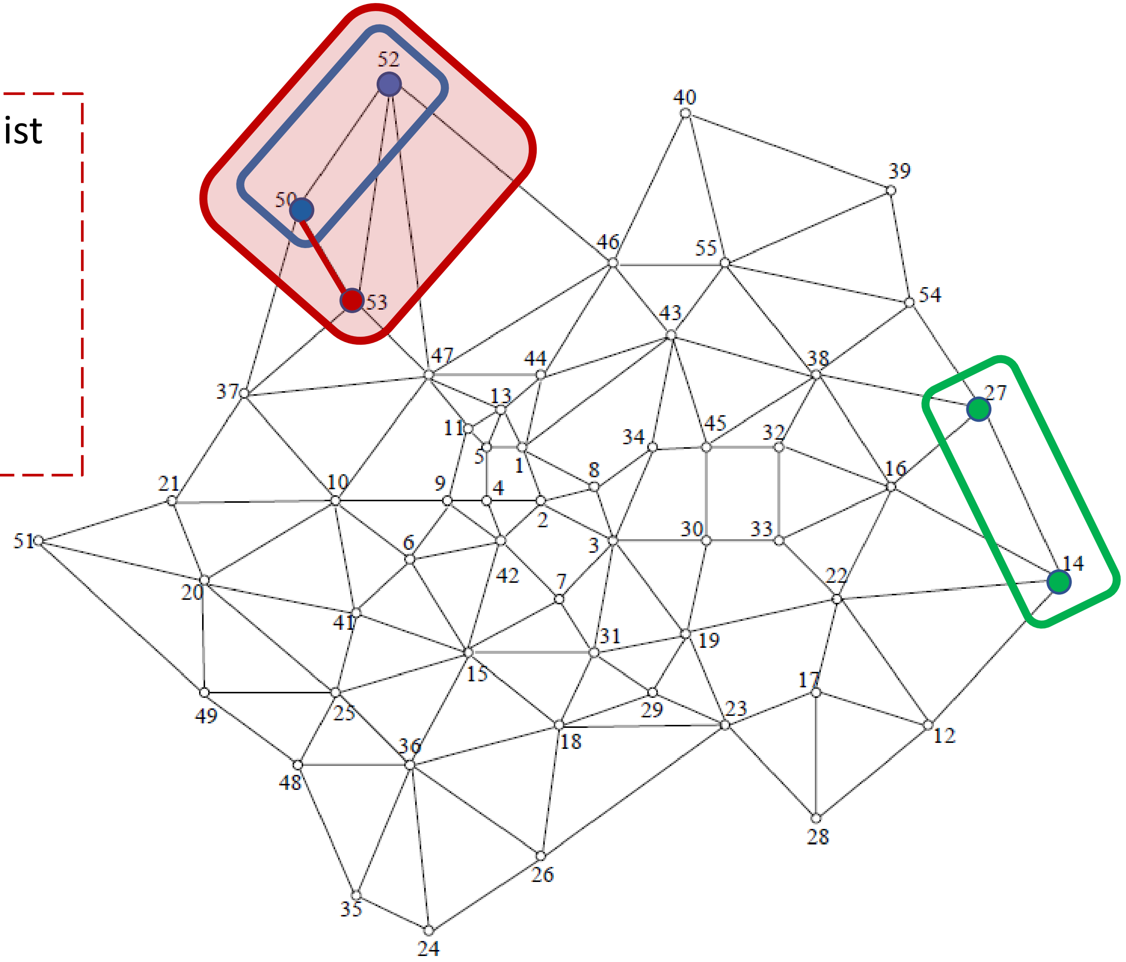
1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*



Heuristic Framework

Repeat the steps until the eligible list is empty

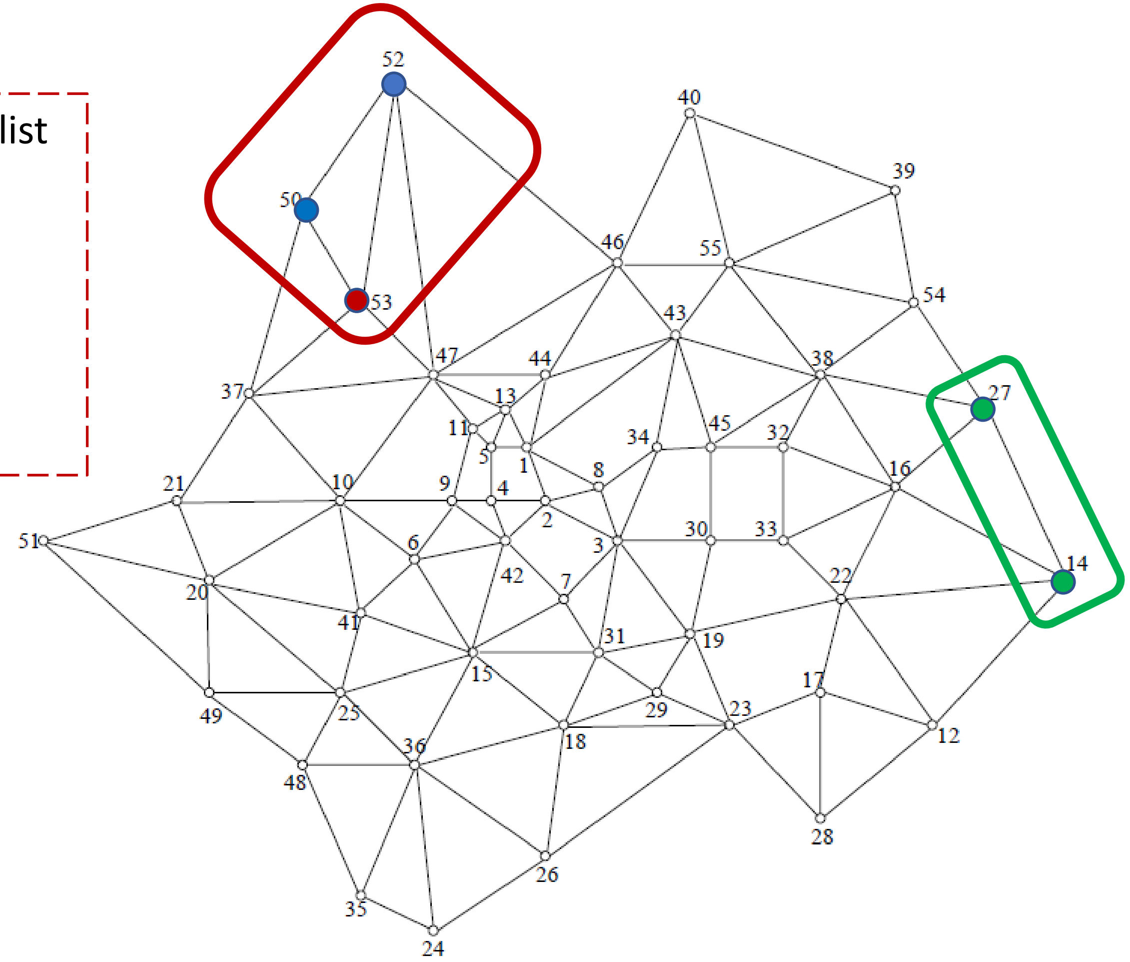
1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*



Heuristic Framework

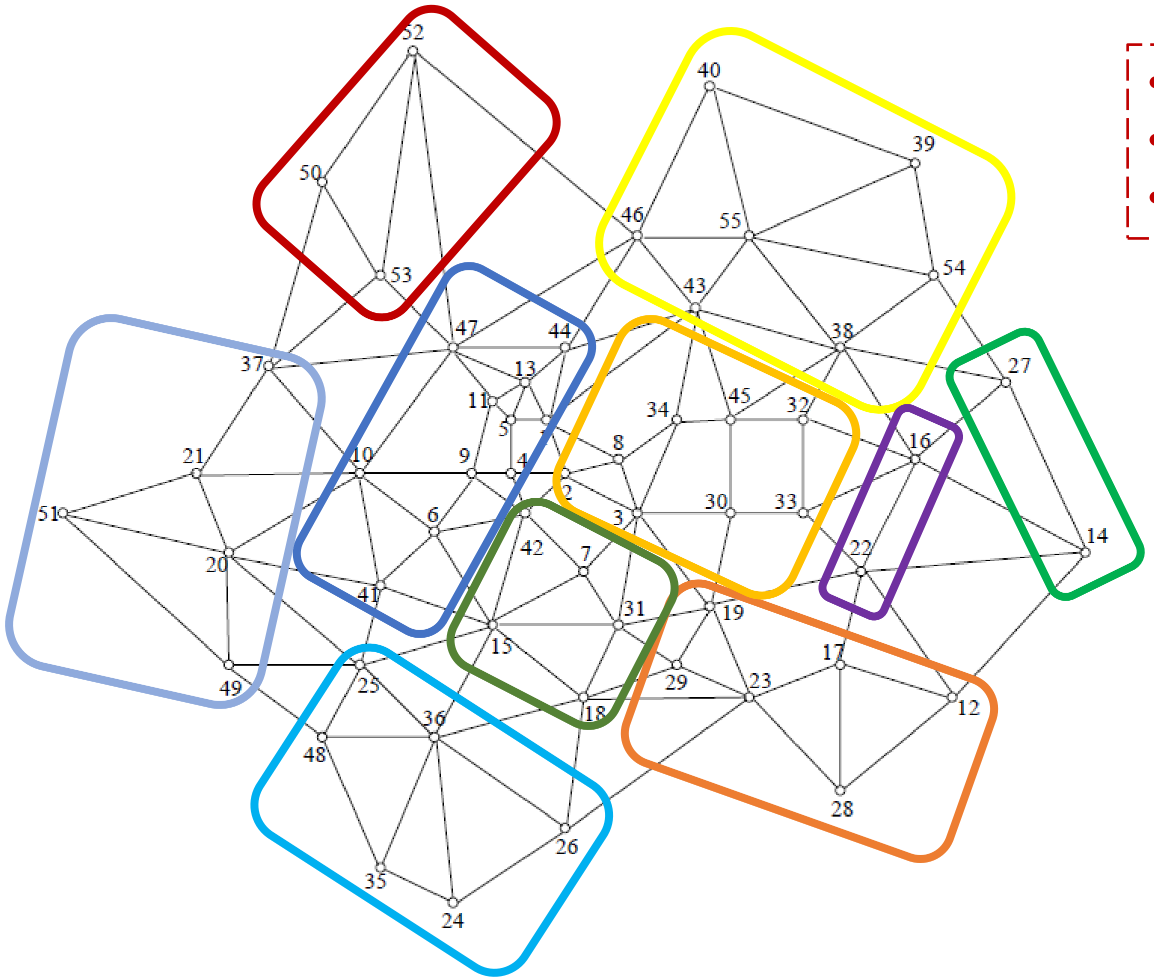
Repeat the steps until the eligible list is empty

1. Select one node randomly
2. Selection biased of a direct connected node
3. Create one *Super Node*



52 nodes in the eligible list
2 *Super Nodes*

Heuristic Framework



- Eligible list is empty
- 10 Groups
- Groups are not homogeneous

Heuristic Framework

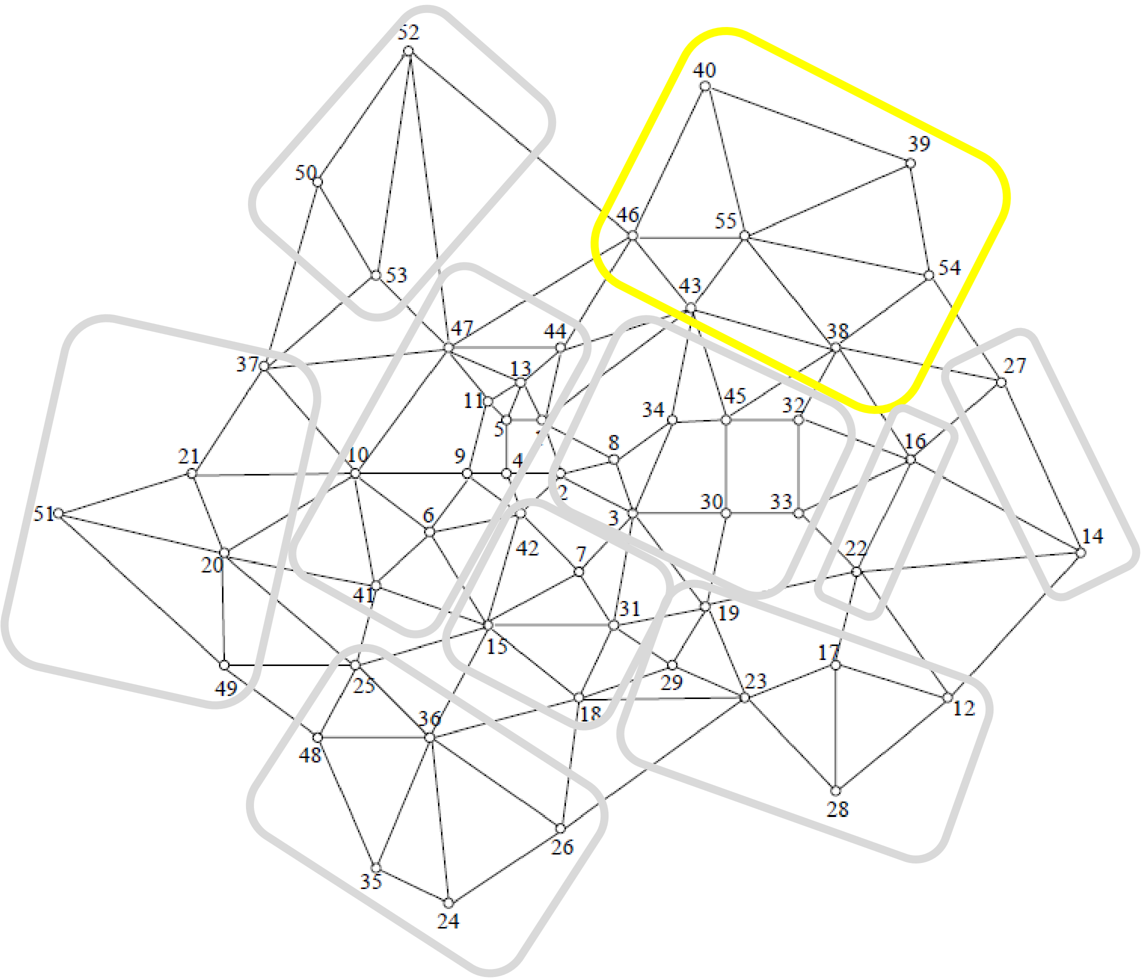
- Random selection of points
- Groups based on distance (Super Nodes)

Stage 1
Aggregation

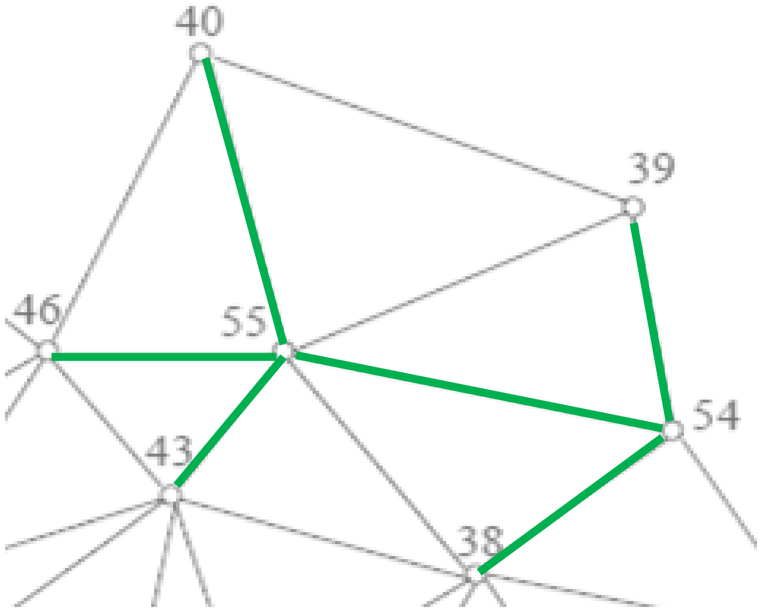
Stage 2
Centroids

- Minimum spanning tree
- Centroid is one of the nodes in the tree

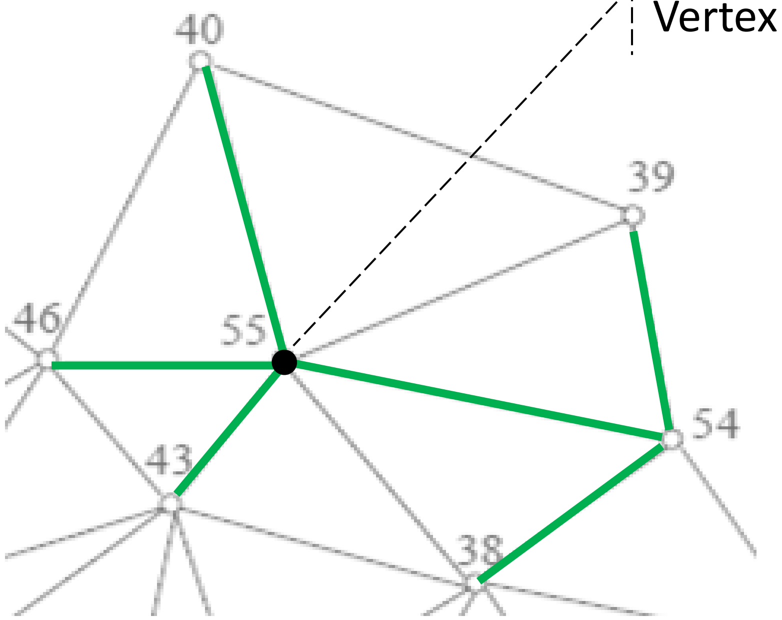
Heuristic Framework



Heuristic Framework

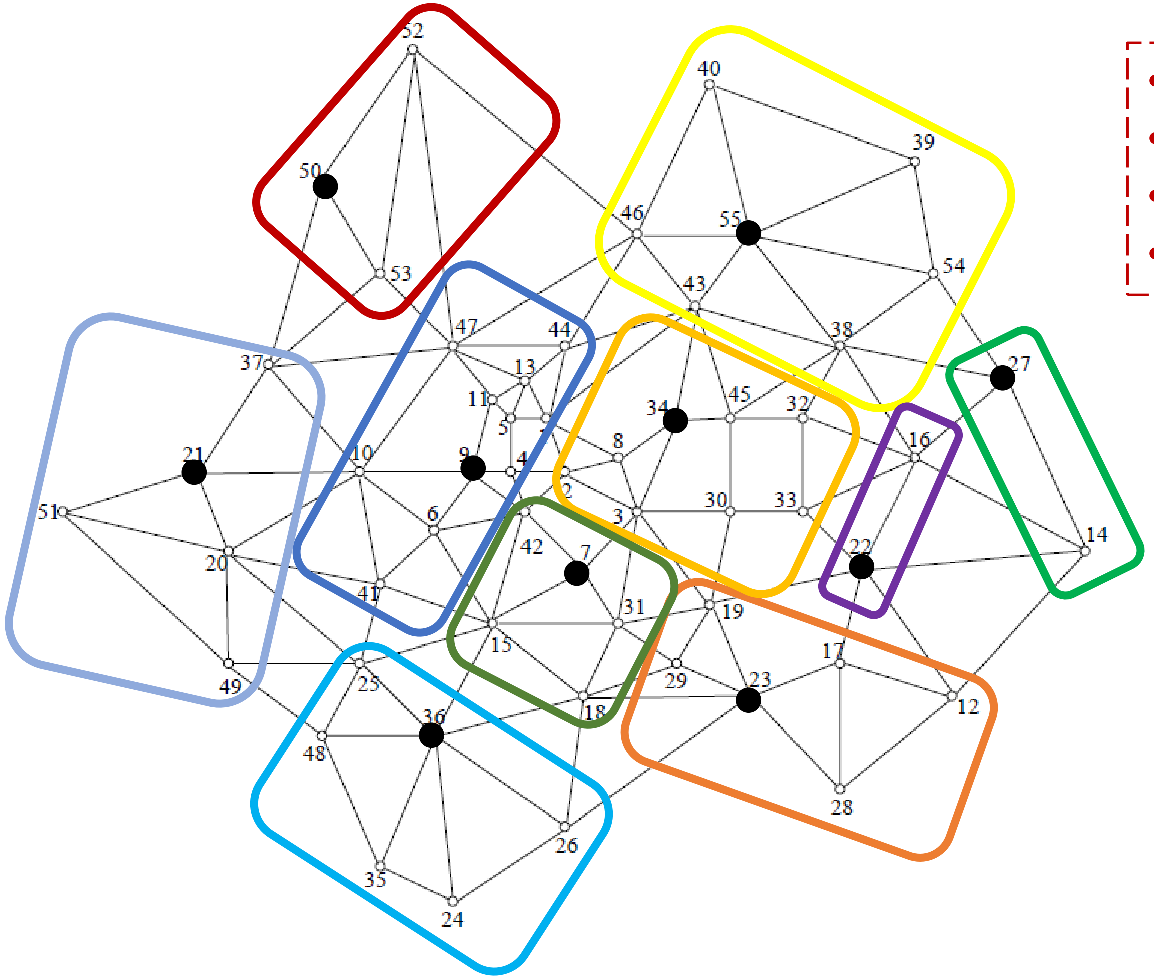


Heuristic Framework



Centroid of the minimum expansion tree
Vertex 1-center on a tree

Heuristic Framework



- Eligible list is empty
- 10 Groups
- Groups are not homogeneous
- 10 centroids

Heuristic Framework

Stage 1
Aggregation

- Random selection of points
- Groups based on distance (Super Nodes)

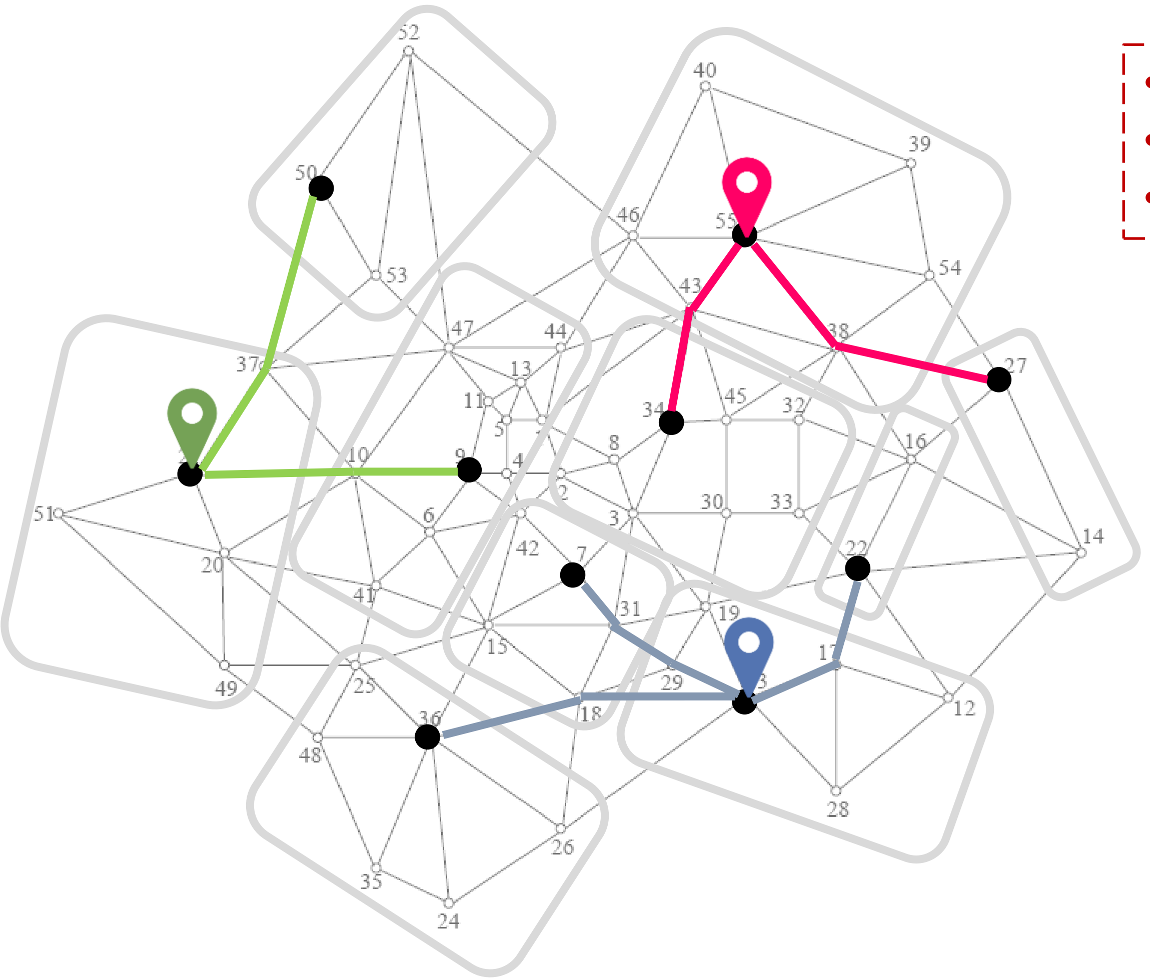
Stage 2
Centroids

- Minimum spanning tree
- Centroid is one of the nodes in the tree

Stage 3
Location Model

- p-median model
- Exact solution

Heuristic Framework



- 10 Groups
- 10 centroids
- *p-median* (locating 3 facilities)

Heuristic Framework

Stage 1
Aggregation

- Random selection of points
- Groups based on distance (Super Nodes)

Stage 2
Centroids

- Minimum spanning tree
- Centroid is one of the nodes in the tree

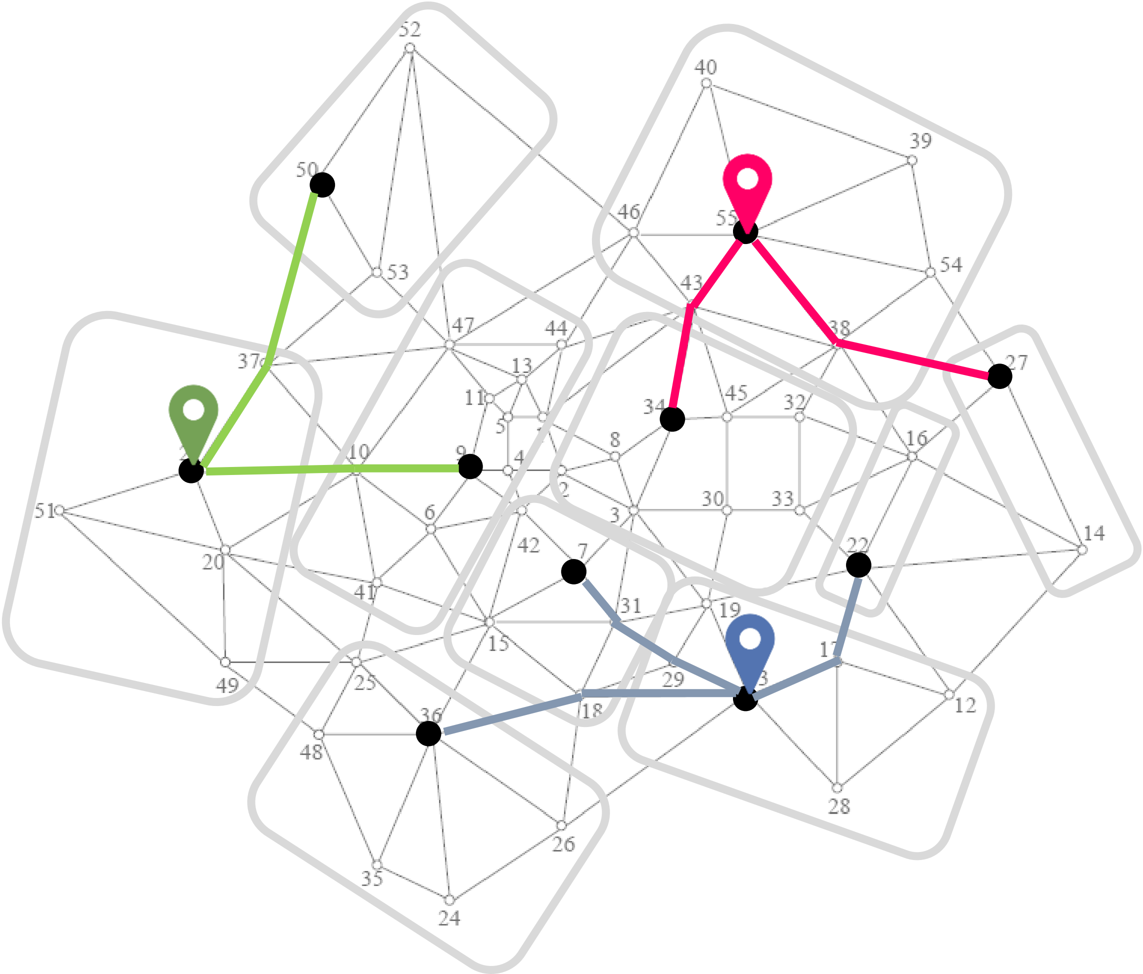
Stage 3
Location Model

- p-median model
- Exact solution

Stage 4
Repair Solution

- Evaluation of the objective function value
- Relationship between number of groups and real demand points
- Type A, B and C errors

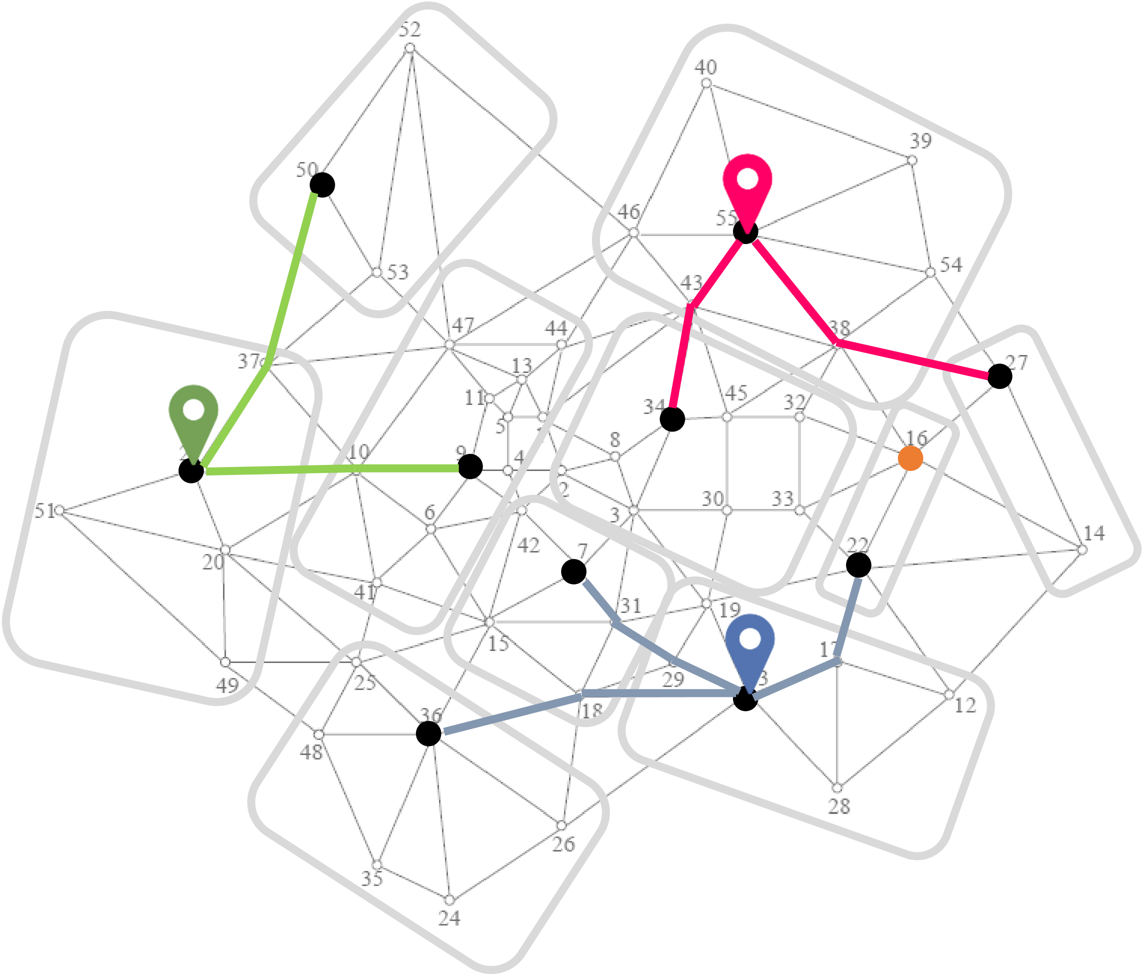
Heuristic Framework



- 10 Groups
- 10 centroids
- *p-median* (locating 3 facilities)

- Number of groups / Number of demand points
- Avoiding error type B
- Error type C – Relocate demand nodes
- Error type A is not addressed

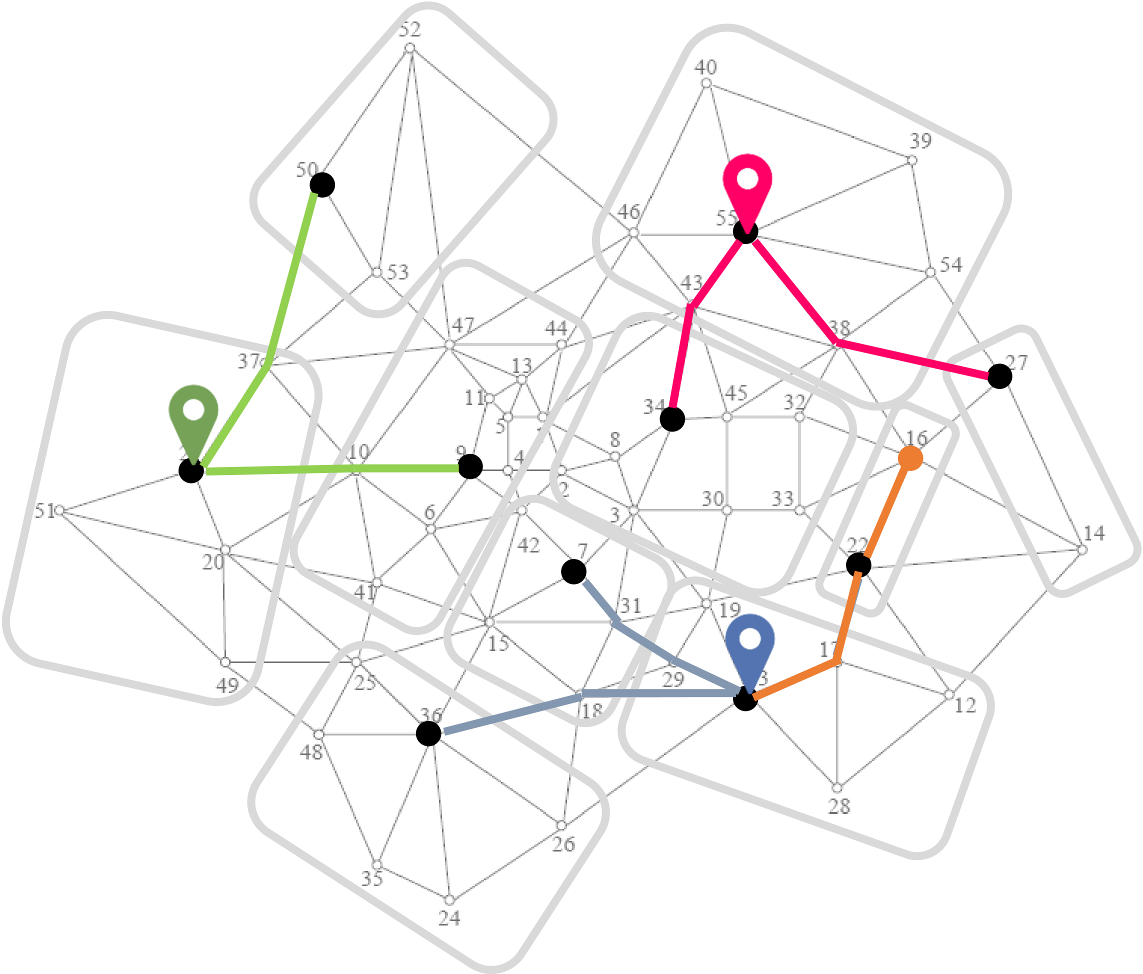
Heuristic Framework



- 10 Groups
- 10 centroids
- *p-median* (locating 3 facilities)

- Number of groups / Number of demand points
- Avoiding error type B
- **Error type C – Relocate demand nodes**
- Error type A is not addressed

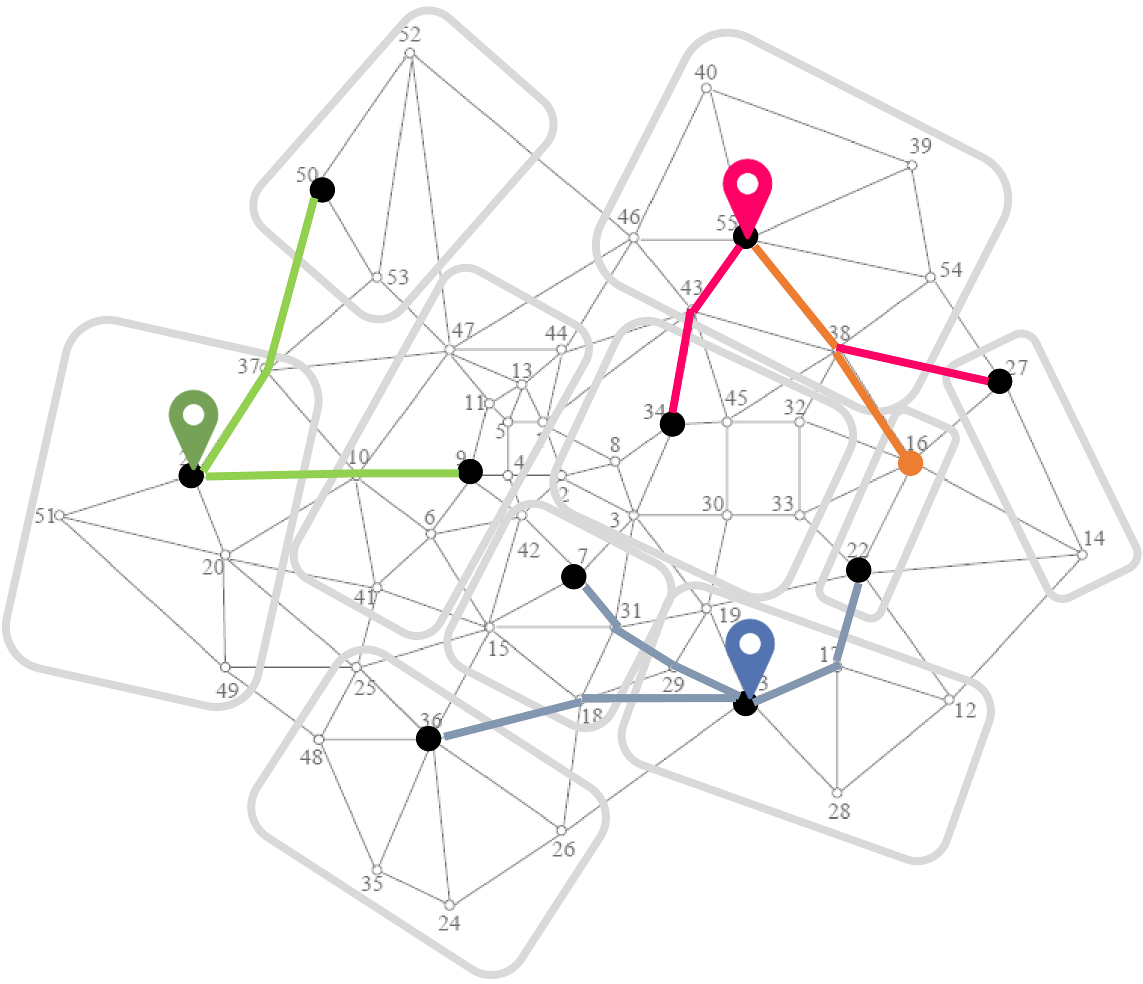
Heuristic Framework



- 10 Groups
- 10 centroids
- *p-median* (locating 3 facilities)

- Number of groups / Number of demand points
- Avoiding error type B
- **Error type C – Relocate demand nodes**
- Error type A is not addressed

Heuristic Framework



- 10 Groups
- 10 centroids
- *p-median* (locating 3 facilities)

- Number of groups / Number of demand points
- Avoiding error type B
- **Error type C – Relocate demand nodes**
- Error type A is not addressed

Heuristic Framework

Stop criterion
- Objective function
- Rate of aggregated and disaggregated demand nodes

First iteration with no fixed number of groups

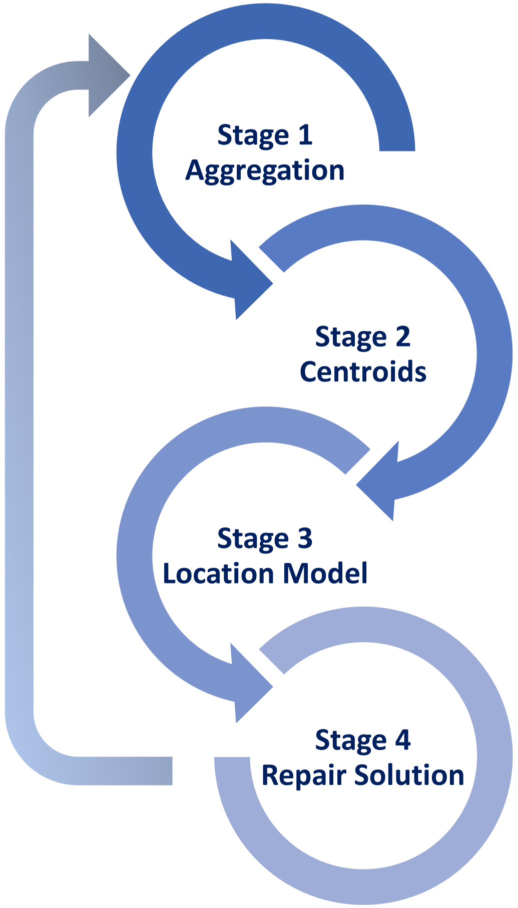
Stage 1 Aggregation
- Random selection of points
- Groups based on distance (Super Nodes)

Stage 2 Centroids
- Minimum spanning tree
- Centroid is one of the nodes in the tree

Stage 3 Location Model
- p-median model
- Exact solution

Stage 4 Repair Solution
- Evaluation of the objective function value
- Relationship between number of groups and real demand points
- Type A, B and C errors

Heuristic Framework



- **Challenges:**
  - Definition of a measure of dispersion of the demand points respect to the centroid of the group based on spatial analysis
  - Include a GRASP heuristic to repair the solution due to the multi-start characteristics of the framework
  - Perturbation of the aggregation
  - Include metaheuristic methods to solve the location problems
  - Test cluster and districting algorithms

Conclusions and Future Work

According with the literature review:

- There are more works for median models than for center and covering models
- There is little average-case analysis of aggregation errors
- There are no analytical models for addressing the trade-off of doing aggregation
- The use of centroids as aggregated demand points is limited to median models
- Data for testing algorithms is often computer-generated instead of being real data
- The development of frameworks to integrate location models with aggregation algorithms is an active research field.

Conclusions and Future Work

Extensive survey and analysis about the error induced by aggregation in location models:

Francis, R. L., Lowe, T. J., Rayco, M. B., & Tamir, A. (2009). Aggregation error for location models: survey and analysis. *Annals of Operations Research*, *167*(1), 171–208.

- Aggregation error measures vary greatly depending on the problem
- There is no agreement on how to measure errors correctly

Conclusions and Future Work

- Future work: Make a deeper analysis of some properties found in the literature
  - Self cancelling error in models with additive structure
  - Presence of diminishing returns in error measure
  - Use of centroids not only in median models but also in center and covering problems
  - Develop aggregation algorithms based on clustering and spatial analysis

References

Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, *27*(2), 93–115.

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. *Geocomputation, a Primer*. New York: Wiley.

Anselin, L. (2005). Exploring spatial data with GeoDa: A workbook. Urbana-Champaign: Center for Spatially Integrated Social Science.

Church, R. L. (2002). Geographical information systems and location science. *Computers and Operations Research*, *29*(6), 541–562.

Church, R. L., & Murray, A. T. (2008). *Business Site Selection, Location Analysis and GIS*. John Wiley & Sons.

Daskin, M. S., Haghani, A. E., Khanal, M., & Malandraki, C. (1989). Aggregation effects in maximum covering models. *Annals of Operations Research*, *18*(1), 113–139.

Duque, J. C., Anselin, L., & Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science*, *52*(3), 397–419.

Duque, J., Ramos, R., & Suriñach, J. (2007). Supervised Regionalization Methods: A Survey. *International Regional Science Review*, *30*(3), 195–220.

Francis, R. ., Lowe, T. ., Tamir, A., & Emir-Farinas, H. (2004). A framework for demand point and solution space aggregation analysis for location models. *European Journal of Operational Research*, *159*(3), 574–585.

Francis, R. L., & Lowe, T. J. (1992). On worst-case aggregation analysis for network location problems. *Annals of Operations Research*, *40*(1), 229–246.

Francis, R. L., & Lowe, T. J. (2015). Demand Point Aggregation for Some Basic Location Models. In *Location Science*.

Francis, R. L., Lowe, T. J., Rayco, M. B., & Tamir, A. (2009). Aggregation error for location models: survey and analysis. *Annals of Operations Research*, *167*(1), 171–208.

Thank you!

Questions?