Final Project for the **Data Analytics** course at the MSc in CS at **University of Milano-Bicocca**

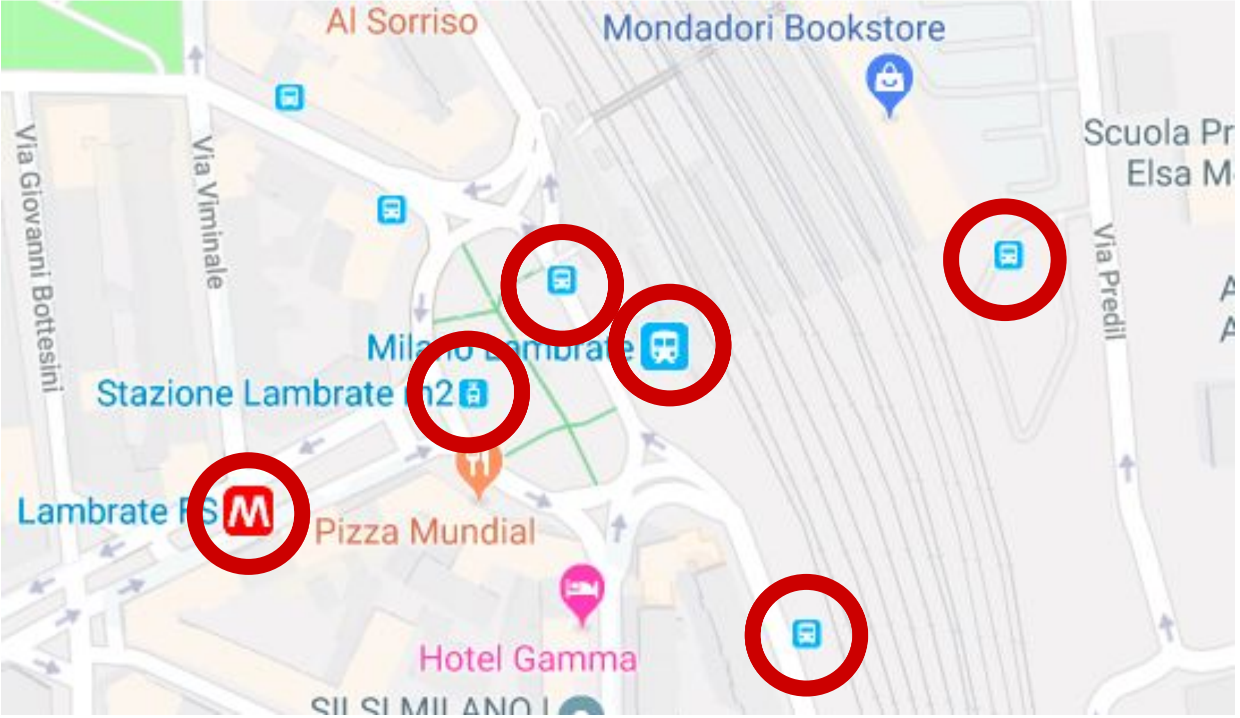
AVERAGED GRAPH ANALYSIS AND FAILURE RESILIENCE OF A PUBLIC TRANSPORT NETWORK

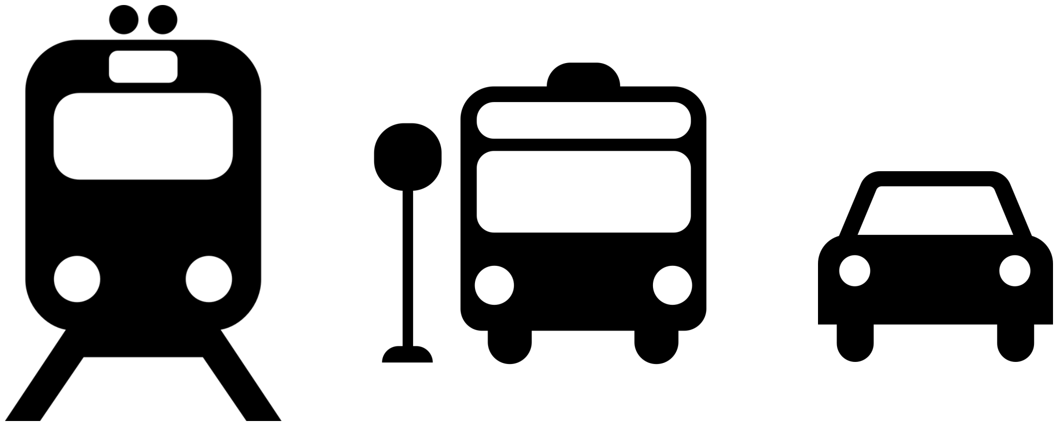
Authors:
Nassim Habbash
Ricardo Anibal Aragon Matamoros

Problems

- Transportation networks can be **disjoint**
- Transportation datasets might **not** be **standardized**
- How to correctly **represent** a network of a set of different means of transportation, possibly disjoint and not conforming to a standard?

All the circled stops service the zone of ***Lambrate****, but each one is used by* ***different lines*** *and has* ***different coordinates.*** *This can pose some issues in the zone analysis.*





Different means of transportation, possibly coming from ***different*** *datasets, might not be standardized nor coherent with each other*

Dataset

The **dataset** used in this project is Milan's AMAT GTFS feed, consisting of the feed of every ATM service in March 2019.

The data was obtained from **TransitFeeds**.

A **Docker container** has been setup for ease of reuse and multiple operability, to extract from a PSQL **database** a graph of **bus lines**, **tram lines** and **metro lines** of Milan.

What are average graphs?

Average graphs represent an averaged union of all the distinct networks considered through an operation defined as **spatial clustering**.

This operation allows us to generate graphs that represent the **spatial connectivity** of different regions (clusters) of the analyzed macro-area, based on the location of **every stop** of **every graph**.

Demo*: Average Graphs section,*
Neighbor Graph section

Given a set of seed points (the actual stops of every network), we group them together in **regions** if their **haversine distance** is minor or equal of a parameter **γ.**

This parameter represents the **clustering radius** of a region.

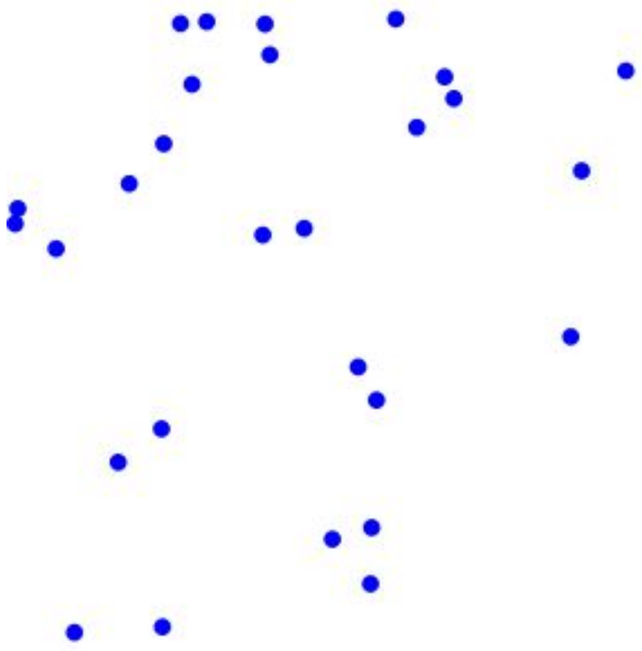
The **centroid** of each region is iteratively computed as the **mean** of its stop's coordinates.

Demo*: Average Graphs section,*
Neighbor Graph section

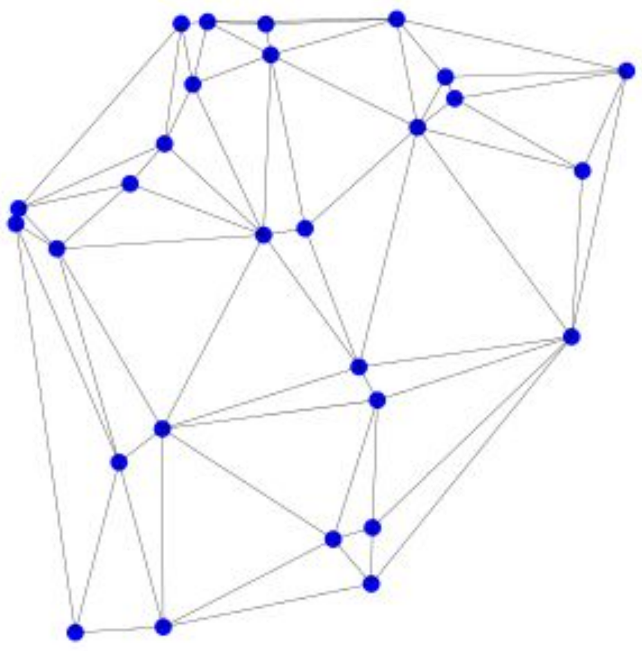
From the resulting regions, we are able to compute the **Voronoi Diagram,** representing the **spatial relation between regions**.

From the same regions we can obtain a complementary graph, called the **Delaunay Triangulation** graph, which we'll call **Neighbor Graph**, by introducing a parameter, **ρ**, that represents the **maximum distance** allowed **between two neighboring regions**.

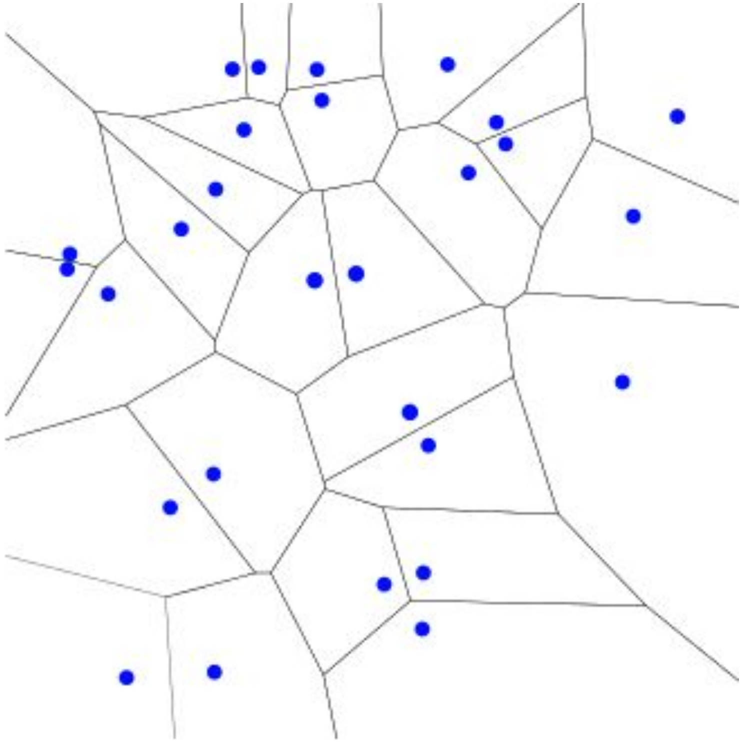
Demo: *Average Graphs section,*
Neighbor Graph section



Centroid of every region generated by the spatial clustering algorithm



Delaunay triangulation, representing the ***neighboring relationship between regions***



Voronoi diagram, representing the ***spatial subdivision of each region’s influence area***

Demo*: Neighbor Graph section*

- Spatial clustering -

From the **neighbor graph,** we go on to compute the **cell-to-cell flow graph**.

We define cell-to-cell flow as a flow indicating the presence of a **direct link** between **stops** in two **neighboring** regions.

The cell-to-cell flow graph and the neighbor graph, **share** the same **nodes**, but **do not share** the same **edges**.

The cell-to-cell graph represents **how** a public transportation **serves** a macro-area on a **topological level.**

Demo: *Cell-to-cell flow graph section*

Comparison

A **comparison** between the neighbor graph and the cell-to-cell flow graph allows us to analyze **how well** the public services **serve** regions of given parameters **γ** and **ρ**.

Some questions that we could ask are:

- How is the periphery of the city served?
- How well interconnected are different regions of the city?
- Are there many regions that act as "gateways" - and could hence be more prone to intense traffic?

Demo: *Comparison section*

Comparison considerations:

- High closeness, high degree: important central nodes with a lesser dense neighborhood
- High closeness, low betweenness: there are a lot of alternative routes in the network
- High betweenness, low closeness: there are ego elements that polarize the network
- High closeness, betweenness and closeness: relevant central, broker node

Network resilience to attacks

We analyzed how well could the cell-to-cell graph **withstand** different **attack strategies**.

The strategy are measured against the graph's Giant Component size, and test **how long** it takes to reach its **percolation threshold**.

This allows us to observe:

- How area-of-effect attacks can damage the network.
- How the original networks are affected.
- How different micro and macroregions (by varying **γ**) are affected by different strategies.

Demo: *Attacks section*

Conclusions

- Average graphs allow us to analyze network data under different **standards** and **not completely jointed**
- The average graphs generated in this project pertain to the **topology** of the network, but **temporal** average graphs can also be generated
- Milan's transport network offers good connectivity for **central** and **north-western** regions of the city, but still does not effectively connect the **periphery**
- An effective attack scenario needs in average **15 attacks** to render the network useless, and has to be **smart**

Thanks for your attention

Bibliography:

Adrienko, and Adrienko. 2011. "Spatial Generalization and Aggregation of Massive Movement Data."

Kim, Jiwon, Kai Zheng, Sanghyung Ahn, Marty Papamanolis, and Pingfu Chao. 2016. "Graph-Based Analysis of City-Wide Traffic Dynamics Using Time-Evolving Graphs of Trajectory Data." In.

Yildirimoglu, Mehmet, and Jiwon Kim. 2018. "Identification of Communities in Urban Mobility Networks Using Multi-Layer Graphs of Network Traffic.