

Algorithmic complexity and graphs: compatibility graphs

14 septembre 2024

Compatibility graphs

- ▶ Yesterday we processed graphs describing **relationship between data**
- ▶ If two nodes were related, they were linked by an edge in the graph.
- ▶ Today we are interested in **building** such graphs directly from the data, we call them **compatibility graphs**, answering the following question : Given two nodes in a graph, should there be an edge between them ?

Example applications

- ▶ Social networks management
- ▶ Recommendations

Euclidian distance and compatibility in 2D

Consider the following data :

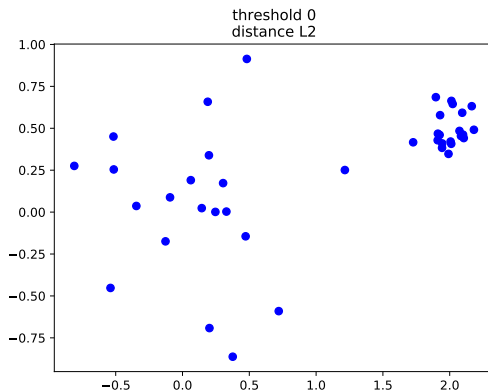


Figure – Data : we would like to define **edge** between some of them

Is this set of edges a good solution ?

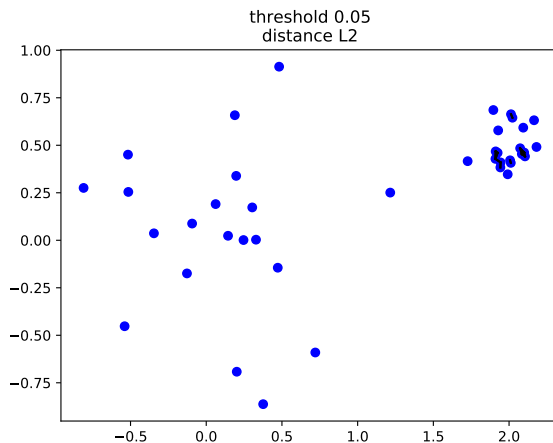


Figure – Some definition of edges

Is this set of edges a good solution ?

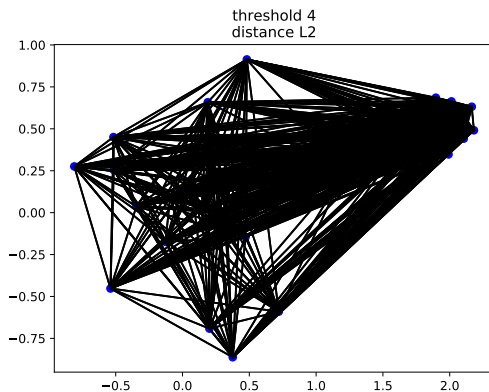


Figure – Some definition of edges

This seems to make more sense

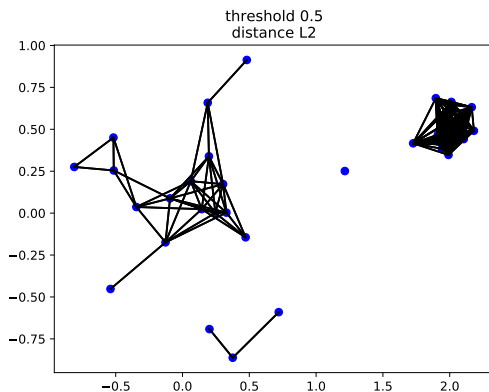


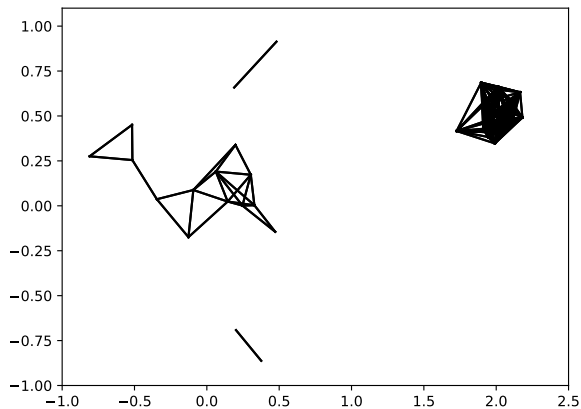
Figure – A proposition of edges

Distances

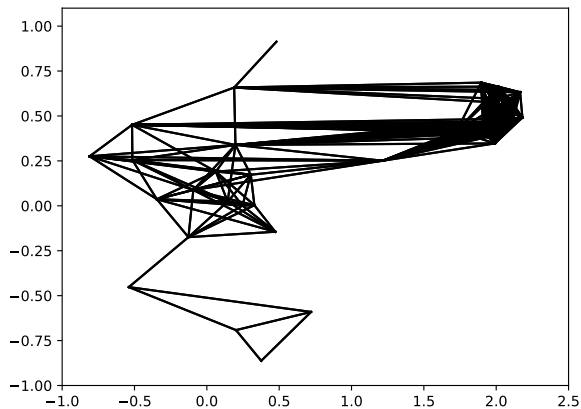
Choosing a **metric** means choosing a method to compute a distance between two objects. Often, there are many possible ways to compute this distances. Some examples are provided in **documents/Math memo.pdf** in the repo.

Exercise 1 : Experiment with the method used to compute the distance, and with the threshold, in order to obtain compatibility graphs similar to the next examples.

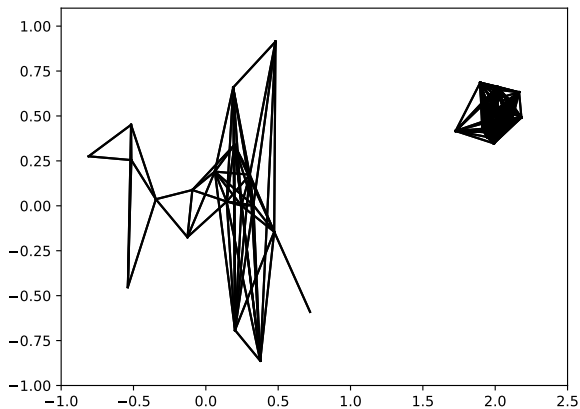
└ Simple geometrical data



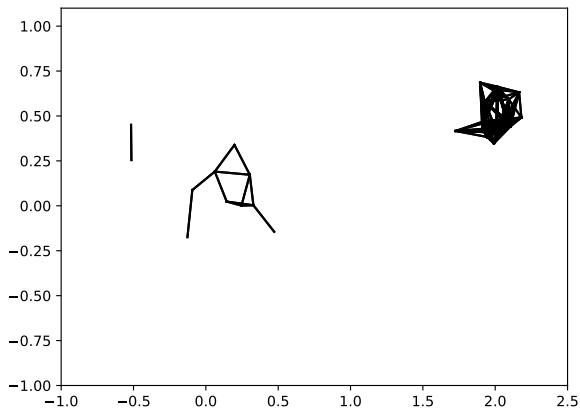
└ Simple geometrical data

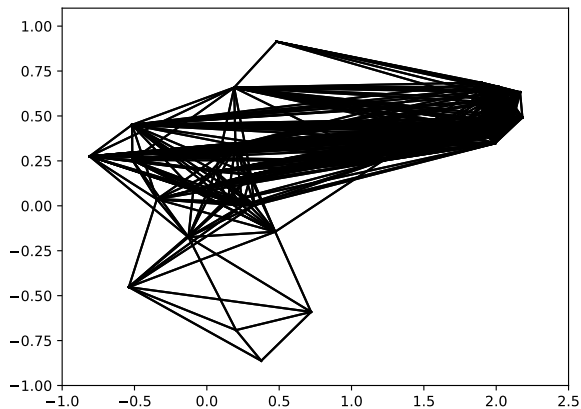


└ Simple geometrical data

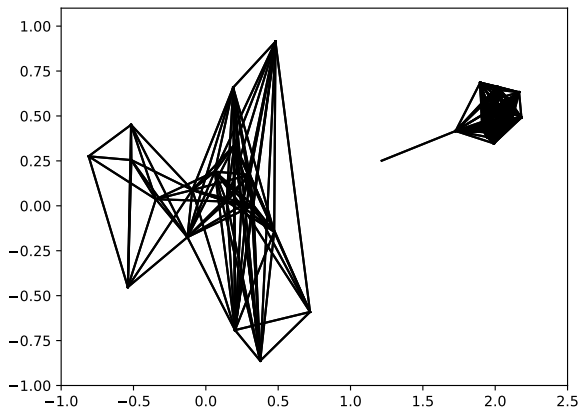


└ Simple geometrical data

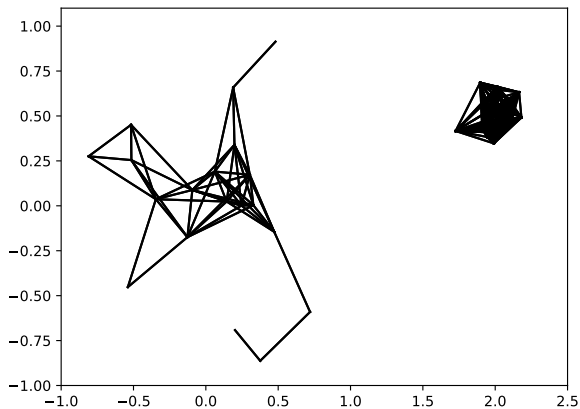




└ Simple geometrical data



└ Simple geometrical data



Building compatibility graphs for non geometrical data

- ▶ Some data are not geometric : not numbers, strings, objects or categories (categorical data)
- ▶ in **hybrid_data/** you can find a notebook contains example metrics for these hybrid data, using pandas.

