Mapping Mappers

Davide Gurnari, Paweł Dłotko

October 2020

Introduction

The goal of this project is to develop software and visualization tools that allows user to compare different mapper graphs representing the same collection of objects.

Mathematical Background

Let us consider a collection of distinct objects P and suppose we have two different ways of assigning features of the elements of P; namely $f_1 : P \to \mathbb{R}^n$ and $f_2 : P \to \mathbb{R}^m$. We can then represent the objects P as two different point clouds $X_1 = f_1(P)$ and $X_2 = f_2(P)$. mMreover we assume that f_1 and f_2 are bijections and that objects in P are distinct. We can then obtain a map between X_1 and X_2.

$$\begin{tikzcd}
P \arrow[r, "f_1"] \arrow[d, "f_2"'] & X_1 \arrow[dl, "f_2 \circ f_1^{-1}"] \\
X_2 &
\end{tikzcd}$$

Since $|X_1| = |X_2| = |P|$, we can assign an unique index to each point of P that allows us to identify $P[i] \sim X_1[i] \sim X_2[i]$. In more practical terms we can think of X_1 (respectively X_2) as a data matrix with $n = |P|$ rows and $|F_1|$ ($|F_2|$) columns. The mapping between X_1 and X_2 is given by a mapping on the row index. Moreover this map is the identity (up to an initial permutation of the indices).

Comparing Mapper Graphs

Let us consider the mapper graphs G_1, G_2 obtained from the point clouds X_1 and X_2 respectively. We can associate to each vertex $v \in V(G_1)$ (respectively $u \in V(G_2)$) the subset of points $v(X_1) \subseteq X_1$ ($u(X_2) \subseteq X_2$) that are part of the vertex's cluster. This generalizes immediately from a single vertex to a set of vertices by taking the union of the sets. Given a set of vertices $V' \in V(G_1)$, we can then define the *Coverage Function* $C_{V'}$ that, for each vertex $u \in V(G_2)$, returns the fraction of points in u that are covered by vertices in V'.

$$C_{V'}(u) = \frac{|\{p \in u(X_2) \mid id(p) \in id(V')\}|}{|u(X_2)|}\,, \tag{1}$$

where - with a little abuse of notation - we indicate with $id(V')$ the set of indices of the points covered by the vertices in V'.

By selecting a subset of vertices in G_1 we can then color G_2 by means of the coverage function, this allows us to have a visual representation on how the points clustered in G_1 are distributed in the clusters of G_2.

Equation 1 provides an online method to compute the coverage of a vertex $u \in V(G_2)$ with respect to a subset of vertices $V' \in V(G_1)$. However we could also pre-compute a *Coverage Matrix* C that contains at position (v, u) the coverage C_{vu} of $u \in V(G_2)$ with respect to a single vertex $v \in V(G_1)$. The coverage $C_{V'}(u)$ can be than obtained by means of a weighted average

$$C_{V'}(u) = \frac{\sum_{v \in V'} |v| C_{vu}}{\sum_{v \in V'} |v|} \tag{2}$$

where each vertex weight is given by the cardinality of its corresponding cluster $|v| \coloneqq v(X_1)$.

Software

The presented software can be found at `https://github.com/dgurnari/mapper_GUI`, both in a standalone version and as a jupyter notebook. It requires NetworkX 2.5 (`https://networkx.org`) to handle the graphs and Bokeh 2.2.2 (`https://bokeh.org`) as plotting library.

A short video demonstration can be found at `https://youtu.be/umSQt03kpPA`.

Using the software the user can import and visualize two mapper graphs. Moreover, vertices of the graph on the left hand side can be selected (using the mouse or the input box on top) and the graph on the right will be colored according to the coverage value of each vertex.

In order to be read and displayed, each graph must be represented by two text files. The first one is the graph's adjacency list, each row contains space separated pair of vertices. The second file contains the list of points covered by each vertex, the i-th row contains a space separated list of indices of all the points covered by vertex i. We assume that vertices are indexed by means of the natural numbers (starting from 1).

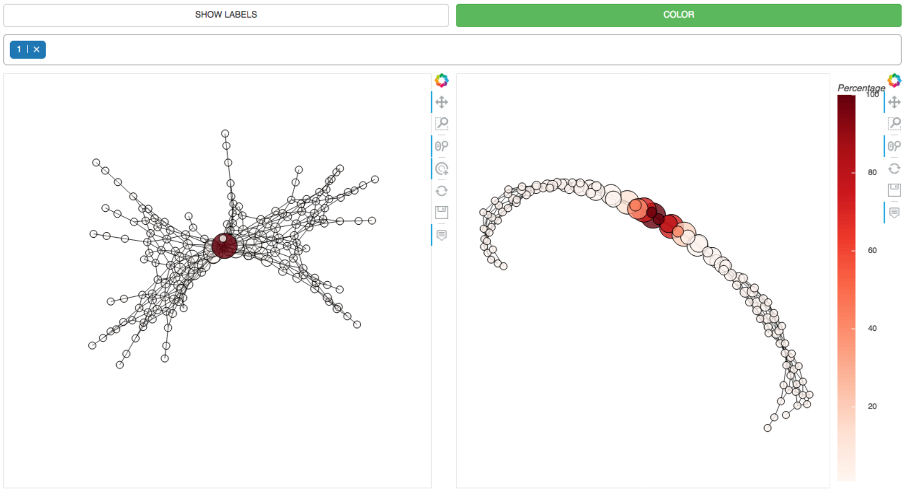

Figure 1: Only one selected vertex, vertex size is proportional to the size of the corresponding cluster

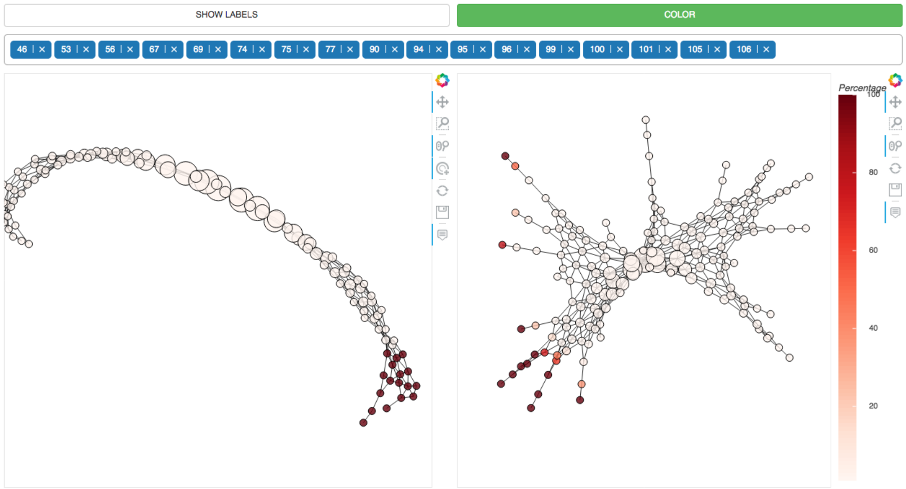

Figure 2: Multiple vertices can be selected.

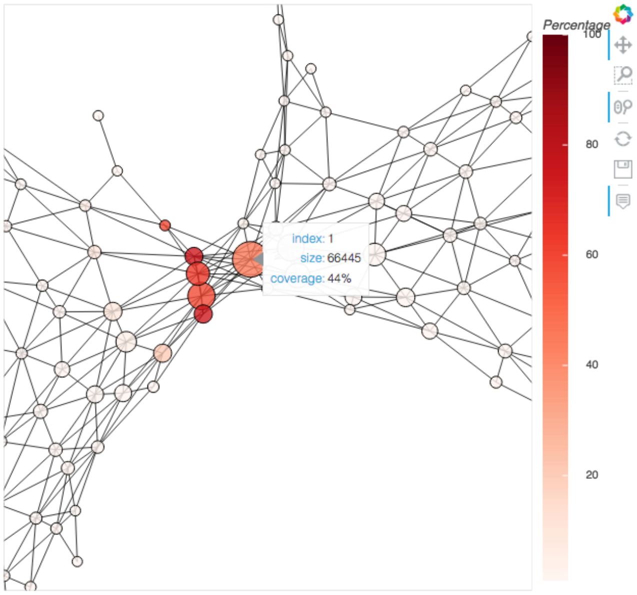

Figure 3: Hoovering over a vertexwill show its size and current coverage.