The University of Manchester

Topological Data Analysis for Economics and Finance.

Pawel Dlotko[1], Wanling Qiu[2] and Simon Rudkin[3].

[1] Dioscuri Centre in Topological Data Analysis, [2] Swansea University, [3] University of Manchester.

Topological Data Analysis Ball Mapper Algorithm

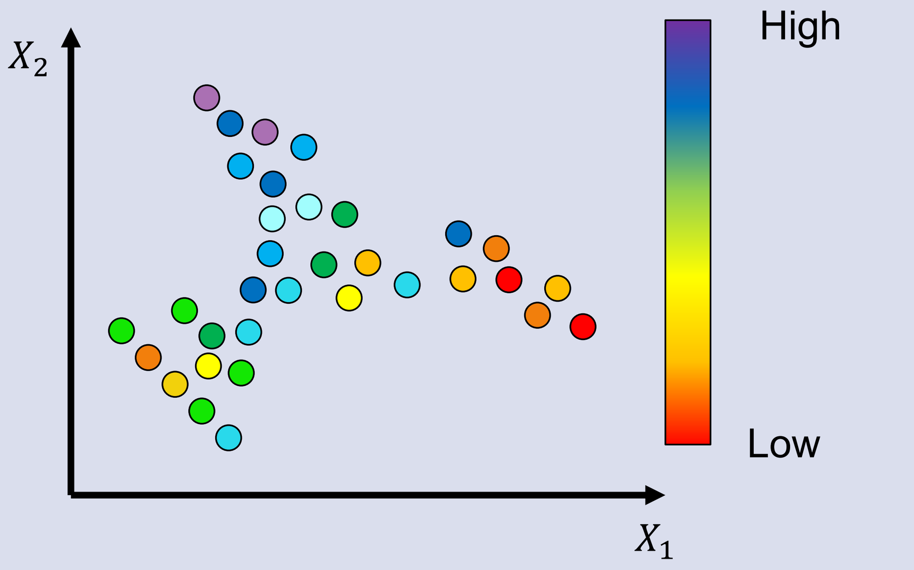

Figure 1 Two-dimensional dataset with colour representing outcome variable.

Based on two variables, X_1 and X_2, Figure 1 shows a scatter plot with points coloured according to a third variable Y. Our goal is to create an abstract visualisation of this data using BM.

Step 1: Begin with an empty cover
Step 2: Select a point at random from uncovered set
Step 3: Construct a ball of fixed radius, ε, around the point points in the constructed ball are then "covered"
Step 4: Repeat Steps 1 and 2 until all points are covered

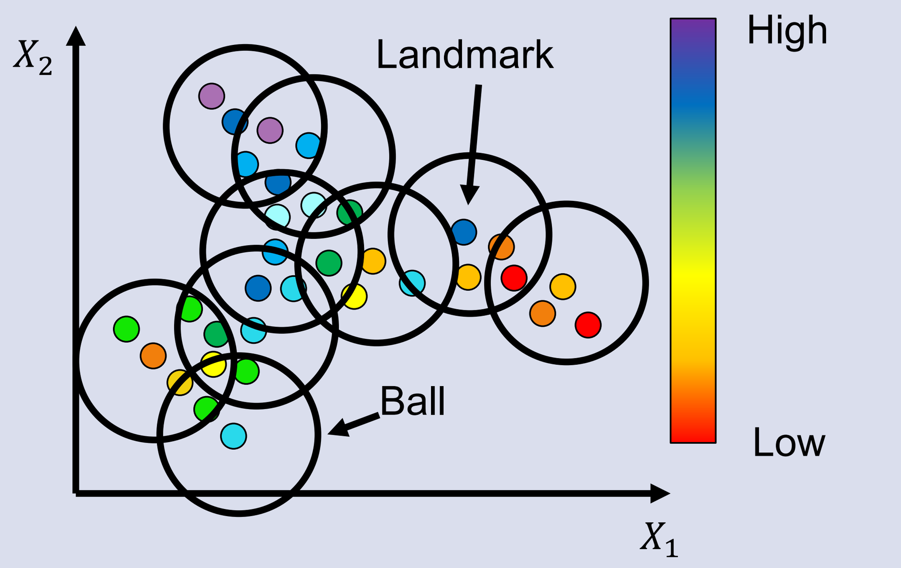

Figure 2 A cover with balls centred on landmarks.

Step 5: Balls correspond to vertices. They are coloured according to the average value of a function on a point of the ball
Step 6: Add edges between landmarks whose balls have a non-empty intersection
Step 7: In visualisation, the size of a ball corresponds to the number of points in the ball

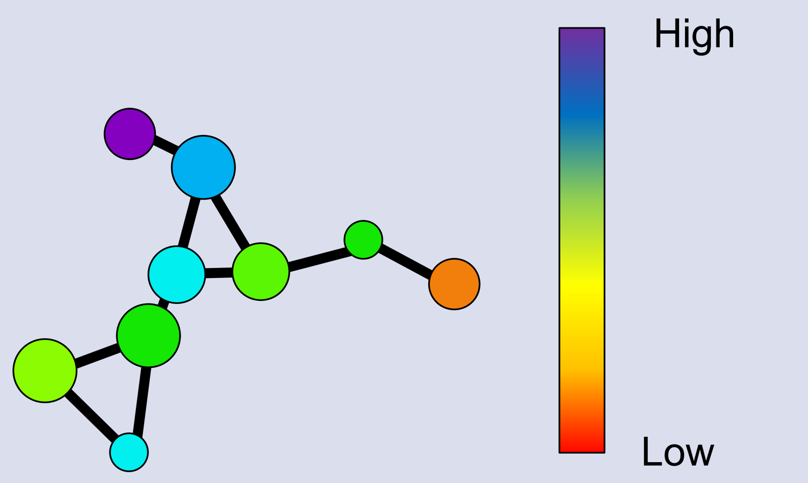

Figure 3 A BM plot of the Figure 1 data.

Figure 3 uses the average colouration in step 5. The plot may be further transformed. Provided point sizes, colourations and edges connecting the balls are not altered, the BM plot provides a topologically faithful overview of the data.

BM has a single parameter, ε, and the researcher is encouraged to consider multiple realisations for their dataset before finalising their inference on the BM plots.

BM as implemented in the R package BallMapper (Dlotko, 2019b) takes the first uncovered point in the dataset as the next landmark in step 4. Repeating the algorithm multiple times at the same ε will confirm that the messages derived from the plots are consistent. As the dataset gets larger so the BMs become more consistent.

BM plots provide information across the joint distribution of the X variables on:

Functions on the ball members
Density
Connectivity

Topological data analysis (TDA) ball mapper (BM) (Dlotko, 2019a) is an algorithm for the visualisation of multi-dimensional data sets which is advanced for applications in Economics and Finance through the work of the authors. We highlight how BM overcomes the challenges of seeing ordinal data in multiple dimensions to illuminate messages that are otherwise hidden in plain sight.

Artificial Example

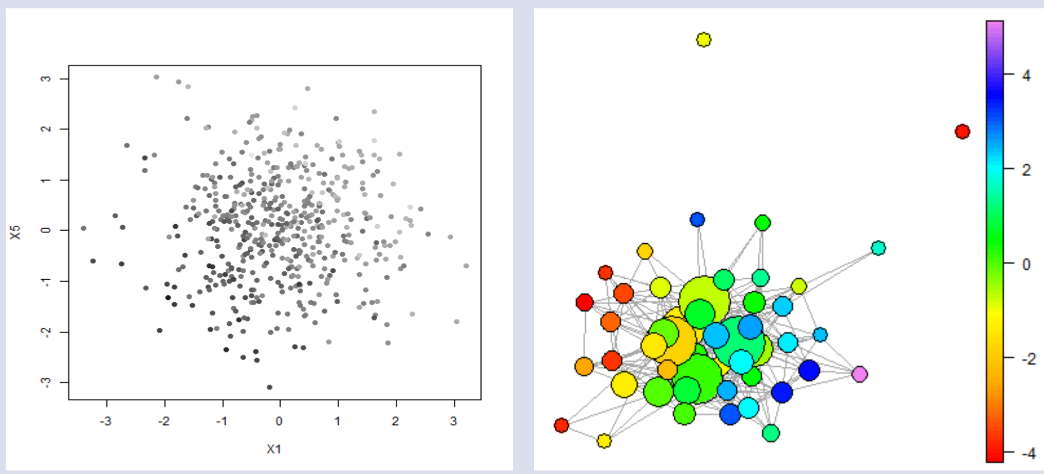

Figure 4 Scatter plot of two axes from five axis point cloud (left) with example BM plot (right).

Data across multiple axes is a point cloud. Here the point cloud is created with five axes drawn from a standard normal distribution, $X_i \sim N(0,1)$ with $i = 1,2,3,4,5$. Outcome for point j, Y_j, is the sum of the values of x_{ij}.

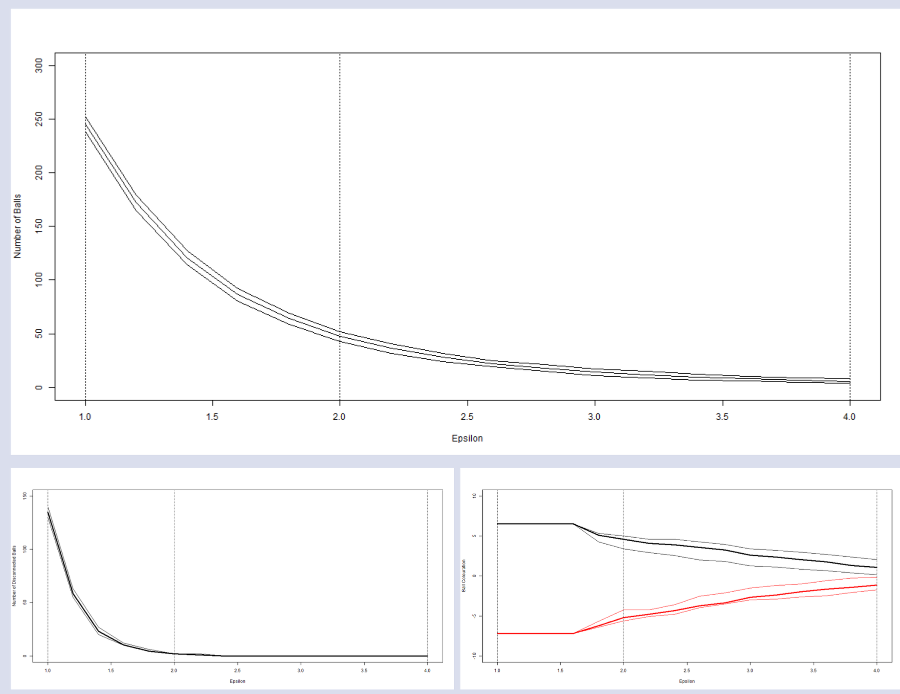

Figure 5 Number of balls (top), number of disconnected balls (bottom left), and minimum and maximum colouration of balls (bottom right).

Numbers of balls fall as radius increases. Figure 5 has a vertical line at $\varepsilon = 2$ to correspond with the radius used in the BM plot of Figure 4. The rapid reduction in the number of disconnected balls is clear. Highest and lowest values of the outcome are in the the least dense parts of the cloud and therefore connect only when the radius is sufficiently large. Detail about the colouration across the joint distribution becomes lost as the radius increases. Choice of radius is thus important to the messages from the data.

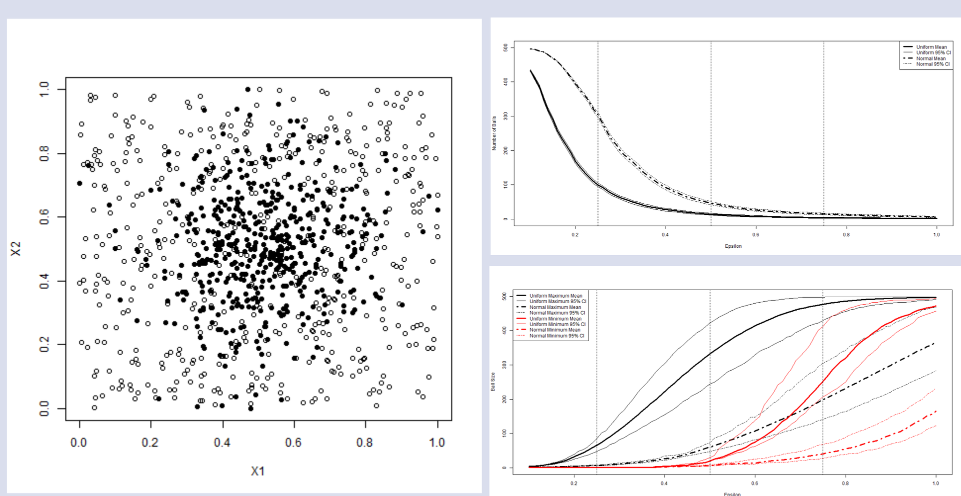

Figure 6 Comparison of normal and uniform distributions. Two-axis scatter (left), number of balls (top right) and number of points in ball (bottom right).

Normal point clouds (black on scatter and solid on graph) have denser centres such that the largest ball grows quickly and outliers only connect later. Uniform point clouds (hollow on scatter and dot-dash on graph) take longer to connect and have a smaller density range.

These plots, together with other artificial examples, may be found in Dlotko et al. (2022).

Email simon.rudkin@manchester.ac.uk to discuss more about development opportunities

Credit Scoring

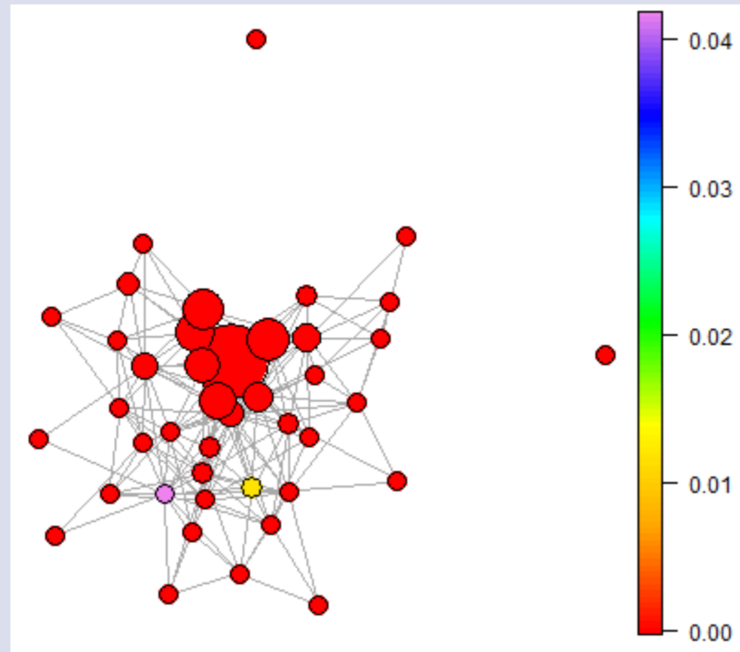

Figure 7 Proportion of firms failing in subsequent year in financial ratio space (Qiu et al., 2020).

In Qiu et al. (2020) we consider the five financial ratios used in the Altman (1968) Z-score model. Altman (1968) argues for the existence of a danger zone defined by linear discriminant analysis. Figure 5 shows no such delineation between acceptable and unacceptable credit risk exists. Consistently BM shows failures restricted to a small subset of balls in the financial ratio space.

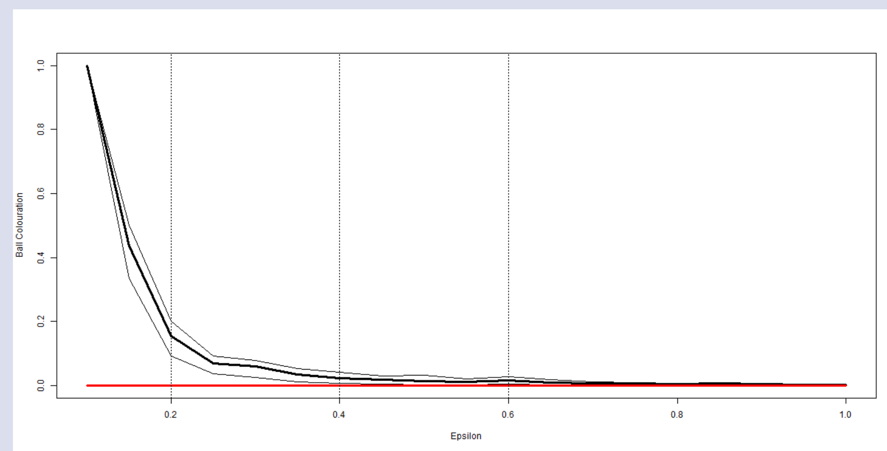

Figure 8 Minimum and maximum proportion of firms failing within a ball (Qiu et al., 2020).

Digital Economy

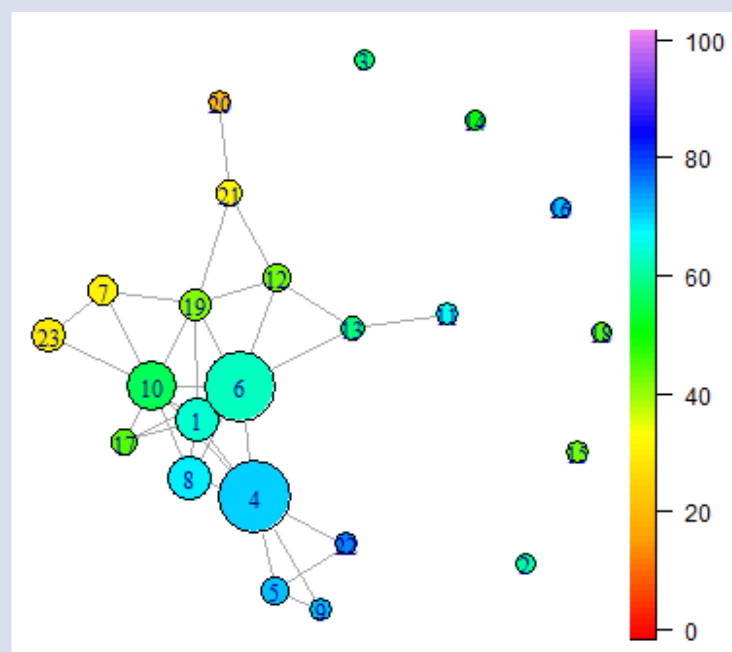

Figure 9 Minimum and maximum proportion of firms failing within a ball (Qiu et al., 2020).

Using Eurostat data, with axes from regional economic characteristics, we see strong association with buying goods and services online. Interactions are important as colour does not follow a single direction across the joint distribution. Numbering balls allows discussion of ball membership, e.g. Southern Italy in ball 20 being very different from Eastern Europe in balls 7 and 23.

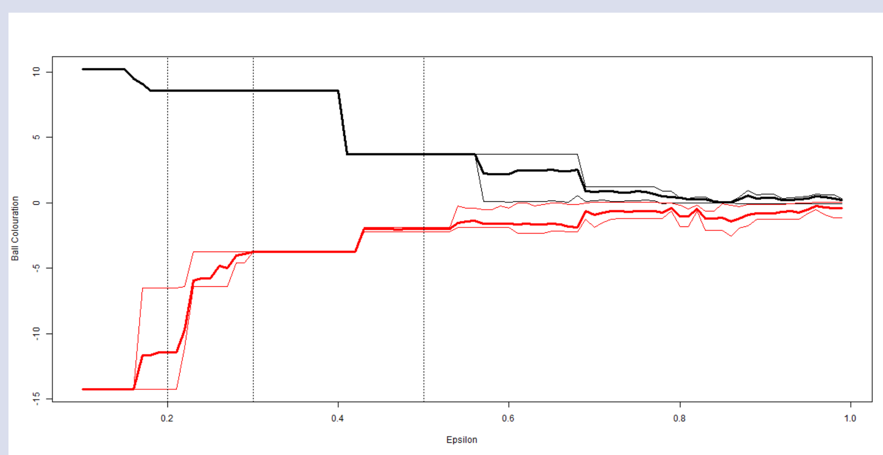

Figure 10 Minimum and maximum proportion of households buying goods and services online

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589-609.

Dlotko, P. (2019a). Ball mapper: A shape summary for topological data analysis. *arXiv preprint arXiv:1901.07410.*

Dlotko, P. (2019b). BallMapper: Create a Ball Mapper graph of the input data. *R package version 0.1. 0.*

Dlotko, P., Qiu, W., & Rudkin, S. (2022). Topological Data Analysis Ball Mapper for Finance. *arXiv preprint arXiv:2206.03622.*

Qiu, W., Rudkin, S., & Dlotko, P. (2020). Refining understanding of corporate failure through a topological data analysis mapping of Altman's Z-score model. *Expert Systems with Applications*, *156*, 113475.