Jongho Jung
Nandita Jha

1	2	Σ

Assignment 6

(Due: 12.12.2022)

Task 1 Toroidal Network Embeddings

(a) Shortest path will be: (Start 7 4 5 2 3 8 End).
Friends (neighbouring nodes) of the start node in Fig 1a are 7, 5, 2, 3, 8.
There are 8 nodes in Figure 1b: Start, 1, 2, 4, 5, 7, 8, end

(b) Figure 1 (Page 3)

(c) Benefit of Toroidal network: It lowers the stress by offering greater relaxation in representing node positions across the plane.
Limitation of toroidal Network: If interactive mode is not enabled, then there is visual cut as it is difficult to trace links wrapped across the boundaries.

(d) Cylindrical layout will have only horizontal wrapping.

Task 2 Multivariate Data

(a) To construct a parallel coordinate plot for 'n' dimensions, we will first draw 'n' equidistant vertical axes. Each axis which corresponds to one attribute will be scaled to minimum and maximum range of variables. Then for each single data point, one polyline is drawn which intersects axes at points corresponding to the attribute value of that particular axis.

(b) Type of correlation between attributes 1 and 2 in A: positive correlation.
Type of correlation between attributes 4 and 5 in D: negative correlation.

(c) STRENGTHS:
1. Scatterplot matrices: Co-relation between two attributes can be easily identified.
2. Radar charts: Visual encoding is compact.
3. Parallel coordinate plots: Required space is proportional to number of attributes

WEAKNESSES:
1. Scatterplot matrices: Co-relation or comparison between more than two variables are difficult to understand without interaction.
2. Radar charts: Representation of number of associated attributes is limited.
3. Parallel coordinate plots: It is prone to visual clutter and ordering of axes can impact results.

Preferred technique: If the data set is large then, then parallel coordinate plots should be preferred.
REASON: Aggregation or clustering by grouping similar elements in a specific color is an effective way of dimensionality reduction without losing data points.

(d) Techniques to improve parallel coordinate plots:
(i) Clustering and aggregation (adding colors) of similar elements to identify corelation between variables can help to tackle issue of massive overdrawing of lines when the data set

is large. Further mean of the cluster can be represented by a bold line in the color of the aggregation.
(ii) Dimension reordering can be done by performing nearest neighbour clustering of adjacent value pairs and exhaustive search of permutations to find minimum S clutter. This will minimize issue in axis sorting which may reveal or hide feature of data set.

Figure 1: Task 1 (b)