Clustering models for high dimensional, temporal, and dissimilarity data

Maurizio Vichi
University “La Sapienza” Rome
maurizio.vichi@uniroma1.it

1. Introduction

Often data present a multiway structure, and they can be arranged into a Three-way Data Set $X$, i.e., a set $X$ of $n \times K \times T$ values related to: $K$ variables measured (observed, estimated) on $n$ objects (individuals, products) at $T$ occasions (assessors, times, locations, etc.). Let $X_1, X_2, \ldots, X_K$ be $K$ quantitative variables observed on $n$ units (objects) at $T$ consecutive time points (Figure 1).

The observed data can be arranged into a three-way longitudinal data set

$$Y \equiv \{y_{ijt} = (x_{ij1}, x_{ij2}, \ldots, x_{ijT}) : i \in I, t = 1, \ldots, T\}$$

where $x_{ijt}$ is the value of the $j$-th variable collected on the $i$-th object at time $t$; $I = \{1, \ldots, n\}$, $J = \{1, \ldots, k\}$ and $U = \{1, \ldots, T\}$ are the set of indices pertaining to objects, variables and time points, respectively.

For each object $i$, $Y(i) = \{y_{ijt} : t = 1, \ldots, T\}$ describes a time trajectory of the $i$-th object according to the $k$ examined variables. The trajectory $Y(i)$ is geometrically represented by $T-1$ segments connecting $T$ points $y_{ijt}$ of $M^{k+1}$. Two time trajectories in $M^3$