1 Time and Space Complexity

We now study the time and space complexity of our stepwise method.

Let D a dataset composed by m time series and let L the number of features choose by PFA among the n extracted. A dataset D is evaluated by FeatTS in time $O(L(m^2) + m^2 + k(m-k)^2 + n \cdot t_f)$ and in space $O(n + L(m+E) + m^2)$.

Time Complexity. Starting from the Community Detection, the Greedy Modularity algorithm requires $O(m^2)$ time to extract the community from a graph with m nodes.

FeatTS transforms each time series in a node, thus in a dataset of m time series, one community is detected in time $O(m^2)$. However, this cost is related to a single feature. Indeed, FeatTSFeatTS creates as many graphs as the number of features chosen by PFA. Hence, assuming PFA chooses L features, the total cost will be $O(L(m^2))$.

The time complexity required by the Co-Occurrence Matrix depends strongly on the computation of Equation (1). Indeed, since we have to estimate the number of times in which two time series fall within the same community, we can easily deduce that create the Co-Occurrence Matrix will take time $O(m^2)$.

Finally, the time requested by the kMedoid to extract the cluster in the Co-Occurrence matrix is equal to $O(k(m-k)^2)$, where k is the number of requested clusters.

A further component used in the pipeline is the features extraction. TSFresh does not provide the time complexity of the method proposed but the time complexity required will be equal to the time needed by each feature to be extracted. Therefore, fixing as t_f the average extraction time of a single feature, the time required to extract the features is given by $n \cdot t_f$, where n being all the extracted features.

Space Complexity. We evaluate the space required by each component of the pipeline. In the first step, TSFresh allows us to extract a plenty of features, where each feature occupies constant memory. Therefore, suppose n are the features extracted by TSFresh, the features extraction component requires $O(n)$ space in memory.

In the Community Detection step, each community is represented by a graph. Normally, a graph in memory occupies a space equal to $O(V + E)$, where V are the vertices and E the edges. Therefore, suppose that m are all the time series, it is required $O(m+E)$ space in memory for each feature chosen by the PFA. Thus, suppose that L are all the features chosen by PFA, this Community Detection step occupies $O(L(m + E))$ space in memory.

Lastly, the space occupied by the Co-Occurrence Matrix is equal to $O(m^2)$. The space required by the Cluster is constant thus can be omitted.