Visibility Graphs-based Time Series Segmentation with Louvain Community Detection

Ivana Kesić[1], Carolina Fortuna[1], Mihael Mohorčič[1], Blaž Bertalanič[1]

[1]*Inštitut Jožef Stefan*
E-pošta: ik9401@student.uni-lj.si

Time series segmentation is an area of research in time series analysis, revealing meaningful patterns within the data in telecommunication systems. By analyzing these segments, operators can identify usage patterns and peak periods in the system. In this paper, we present a time series segmentation method that combines visibility graphs and the Louvain community detection algorithm. By transforming time series data into Natural Visibility Graphs (NVG) or Dual Natural Visibility Graphs (DNVG), we capture different structural aspects of the data. We then apply the Louvain method to detect communities within these graphs, which correspond to distinct segments of the time series. Our approach was demonstrated on six randomly selected datasets from the TSSB collection with the average percentage error of 11.23.

1 Introduction

In the era of big data, time series analysis, especially time series segmentation, has become very important across various fields [1], including telecommunications. The widespread use of information telecommunication systems is leading to a significant increase in network traffic, which must be processed and analyzed [2]. Accurate change point detection is essential for maintaining network performance, and therefore ensuring quality of service. By their detection we can divide network traffic into segments, which makes it easier to spot unusual patterns, enabling quicker responses to potential issues. Also, by analyzing these segments, operators can identify usage patterns and peak periods, ensuring efficient bandwidth usage. However, current segmentation methods usually struggle to handle the data complexity and diversity. Majority of them require predefined parameters, such as segment length, number of segments or window size [3]. Their sensitivity to noise and outliers of the real-world data makes them inefficient for large and dynamic datasets, typical in telecommunications. Our research addresses these challenges, by introducing a TS segmentation method that combines visibility graphs (VG) with the Louvain community detection algorithm [4]. We demonstrate our approach on six TS segmentation benchmark datasets from different domains.

2 Related work

Our research builds upon existing work in community detection, visibility graphs, and time series segmentation. Recent studies have demonstrated the effectiveness of combining these approaches for time series analysis. He et al. [1] introduced an adaptive time series segmentation algorithm, that integrates visibility graphs with community detection. This method transforms time series into complex networks using visibility transformation, and then for community detection it uses an adaptive Particle Swarm Optimization (PSO) algorithm. The adaptive nature of the PSO allows it to dynamically adjust the parameters based on the data characteristics. Similarly, Zheng et al. [5] proposed the Weighted Dual-Perspective Visibility Graph (WDPVG), to analyze unevenly sampled biological time series data, employing Shortest Path-based community detection to identify significant temporal patterns. Chen and Zhang [6] developed a graph-based change-point detection method, which successfully handles non-Euclidean data using general graph transformations, like Minimum Spanning Tree, based on similarity measures. Ferreira and Zhao [7] demonstrated the efficacy of transforming time series into networks, also through similarity and distance measures, for clustering purposes. Clusters are detected via several community detection algorithms like Walktrap and Infomap. Setiadi et al. [8] also applied community detection, but for network segmentation purposes. Specifically, they applied it to library book-borrower networks, finding the Louvain method most effective for network segmentation.

In comparison to mentioned researches, despite notable similarities, key differences set our work apart. It stands out by implementing two types of time series transformations, Natural and Dual Visibility Graphs, followed by applying the Louvain method for community detection. Our research is focused on the segmentation problem, which works within a TS, while many researches focus on clustering of time series traces. The segmentation problem can be used as a helpful step in the clustering process, as it simplifies the data representation. Additionally, we demonstrate the Louvain method's efficiency across various application domains, showing it is domain-agnostic.

3 Methodology

In this section, we present the proposed methodology employed in our study. It consists of two techniques, with the integration of which we get an efficient segmentation of time series.

3.1 Visibility Graphs and Transformation

Visibility graph transformation, proposed by Lacasa et al. [9], maps time series data into complex networks. In this paper we utilise two variations of Visibility graphs, namely Natural Visibility Graph (NVG) and Dual Natural Visibility Graph (DNVG), as they are able to capture more complex relationships within the data, leading to richer network structures.

3.1.1 Natural Visibility Graph

The NVG is one of the methods, that converts time series into graph structures. Each timestamp is seen as a vertical bar with a height equal to its numerical value. If the tops of the bars are mutually visible, the corresponding data points, represented as nodes in the graph structure, are connected by a visibility line. This means there is a direct, natural 'line of sight' between the bars without any interception. In mathematical terms, if any two points (t_i, Y_i) and (t_j, Y_j) have visibility in the time series $Y(t)$, the following condition holds for any point (t_k, Y_k) between them:

$$Y_k < Y_i + \frac{(Y_j - Y_i)(t_k - t_i)}{(t_j - t_i)}, \quad i < k < j \tag{1}$$

where: t_i and t_j are the time coordinates of the points (t_i, Y_i) and (t_j, Y_j), t_k is the time coordinate of the point (t_k, Y_k) that lies between t_i and t_j. Y_i and Y_j are the values of the time series at times t_i and t_j, and Y_k is the value of the time series at time t_k.

The NVG algorithm is straightforward to implement, and its computational complexity is $O(T^2)$ [9], where T is the length of the time series, making it slow for very long time series.

3.1.2 Dual Natural Visibility Graph

The DNVG extends the NVG concept by adding both upward and downward visibility conditions, providing an in-depth representation of the time series. This involves constructing two separate graphs—one for upward and one for downward visibility, and combining them. Nodes v_i and v_j are connected if any (t_k, Y_k) with $t_i < t_k < t_j$ satisfies:

$$Y_i + \frac{(Y_j - Y_i)(t_k - t_i)}{(t_j - t_i)} \leq Y_k \leq Y_i + \frac{(Y_j - Y_i)(t_k - t_i)}{(t_j - t_i)} \tag{2}$$

3.2 Community Detection via the Louvain Method

To determine the TS segments within the constructed visibility graphs we use the Louvain method [4] to detect the communities within the graphs. Communities within the graph represent the segments within the TS. Louvain method is a hierarchical clustering algorithm, meaning it organizes data points into nested clusters. It maximizes modularity, which is a metric used to evaluate the strength of the division of a network into communities. The modularity Q is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{3}$$

where: A_{ij} is the adjacency matrix of the graph, indicating the presence of an edge between nodes i and j. k_i and k_j are the degrees of nodes i and j (i.e., the number of edges connected to these nodes). m is the total number of edges in the graph. γ is the resolution parameter that influences the size and number of communities. $\delta(c_i, c_j)$ is a Kronecker delta function that takes the value 1 if nodes i and j are in the same community, and 0 otherwise.

The Louvain algorithm is divided into two main phases. Phase one focuses on local modularity optimization. Initially, each node is its own community, and then the algorithm shifts each node to a neighboring community to evaluate the increase in modularity. The community that provides the greatest modularity gain receives the nodes. This process is iterative and continues until no more modularity improvements are possible. The second phase is aggregation into super-nodes. The communities detected in the previous phase are combined into super-nodes, preserving edges between original communities. Phase one is then reapplied to this new network of super-nodes to further maximize the modularity. The process repeats until a stable modularity structure is achieved.

3.3 Louvain Method Parameter Selection

Choosing parameters for the Louvain method involves several considerations, primarily focusing on the resolution parameter and the modularity gain threshold. The parameters were chosen empirically, based on the specific characteristics of the datasets and the desired segmentation accuracy.

Resolution Parameter: The resolution parameter (γ), which appears in the previosuly defined modularity formula (3), affects the size and number of communities. A smaller γ results in fewer, larger communities, while a larger γ yields more, smaller communities. The challenge lies in selecting an appropriate γ that accurately reflects the data's natural groupings. For TS segmentation the resolution parameter is adjusted to ensure that the detected communities align with known segments of the TS.

Modularity Gain Threshold: The modularity gain threshold (ΔQ) determines when the algorithm stops optimizing modularity. If the modularity gain from moving a node between communities falls below this threshold, the algorithm stops making further adjustments. Selecting an appropriate threshold is essential to balance computational efficiency and detection accuracy. A high threshold may stop the algorithm too early, while a low threshold could lead to overfitting. For TS segmentation the modularity gain threshold is set to satisfy the segmentation precision.

In Figure 1, we can see an example of the segmentation process. At the top of the figure, the unaltered time series is shown. The first step taken in the analysis of the shown time series was applying the NVG transformation. The right side of the figure shows the resulting graph. Following this, community detection is applied to the graph, identifying different communities within the data. After applying the Louvain method each node is assigned to a specific community, and each community represent a continuous segment of the time series. The start and end points of each community are identified as the boundaries of the segments. These points are marked as potential change points, signifying the transitions between different segments of the time series. The bottom part of the figure illustrates the time series segmentation, where different segments are color-coded based on the detected communities.

3.4 Evaluation metrics

To evaluate the accuracy of the segmentation, we compare the detected change points with the true change points provided in the dataset. The segmentation accuracy is quantified by calculating two metrics: the average percentage error (APE) and the mean absolute error (MAE). These metrics provide an indication of how closely the detected segments align with the actual structure of the time series. The APE measures the average error between detected and true change points as a percentage of the true CPs :

$$APE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{CP_{\text{detected}}[i] - CP_{\text{true}}[i]}{CP_{\text{true}}[i]}\right| \quad (4)$$

where $\mathrm{CP}_{\text{detected}}$ is the list of detected change points, $\mathrm{CP}_{\text{true}}$ is the list of true change points, and n is the number of change points. The MAE provides an absolute measure of the average magnitude of the errors:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|CP_{\text{detected}}[i] - CP_{\text{true}}[i]| \quad (5)$$

4 Results

In this section, we present the results of our proposed TS segmentation approach. The experimental methodology employed to obtain these results is detailed in Section 3. For the performance evaluation we used six randomly selected datasets from the Time Series Segmentation Benchmark (TSSB) [10]. The results are presented in Table 1, where the first column represents the utilised dataset name, the second column shows the true change points (CPs) provided in each dataset, the third column contains the change points detected by our method, the fourth and the fifth columns show the calculated average percentage error and mean absolute error for each dataset.

The best performance was observed with the CBF dataset, where the APE was the lowest at 3.36%. It demonstrates the high precision of our method in detecting changes within this dataset, closely aligning with the true change points. The MAE for this dataset was 14.5, meaning that on average, the detected CPs were 14.5 time points away from the true CPs. In contrast, the Computers dataset presented the greatest challenge, with an average percentage error of 23.61%, and the mean absolute error of 1328.0. This higher error rate is likely due to the complexity and length of the data. The performance on other datasets showed acceptable APE, reflecting the method's robustness. For instance, the APE for the Adiac dataset was 9.05%, with an MAE of 88.0, whereas the errors for the Car, Coffee and OSULeaf datasets were: APE = 19.12%, MAE = 150.0, APE = 6.60%, MAE = 33.0, and APE = 5.66%, MAE = 76.0, respectively.

5 Conclusions

As stated in Section 1, time series segmentation is a problem which appears across various fields, including telecommunications. In this paper, our time series segmentation approach integrates visibility graphs with the Louvain community detection algorithm. We evaluated the method's performance on six randomly selected datasets from the TSSB segmentation benchmark repository from different domains. The experimental results showed that our approach achieves high accuracy in detecting change points, with the best performance observed in the CBF dataset, where the APE was the lowest at 3.36%, and worst performance on the Computers dataset with an APE of 23.61%. By demonstrating the methods performance on 6 datasets from different domains, we also demonstrate the generality of the approach. The Computers dataset tracks energy consumption of household devices, like desktops and laptops. Because of the fact that the Information and Communication Technology sector is responsible for 3 to 4% of global CO2 emissions, applying our segmentation approach to the mentioned dataset helps in understanding device's energy patterns, which leads to improved network energy efficiency. In the future work, the method will be applied to the specific task of TS segmentation in telecommunication domain for traffic segmentation in next generation wireless networks.

Acknowledgments

This work was supported by the Slovenian Research Agency (P2-0016).

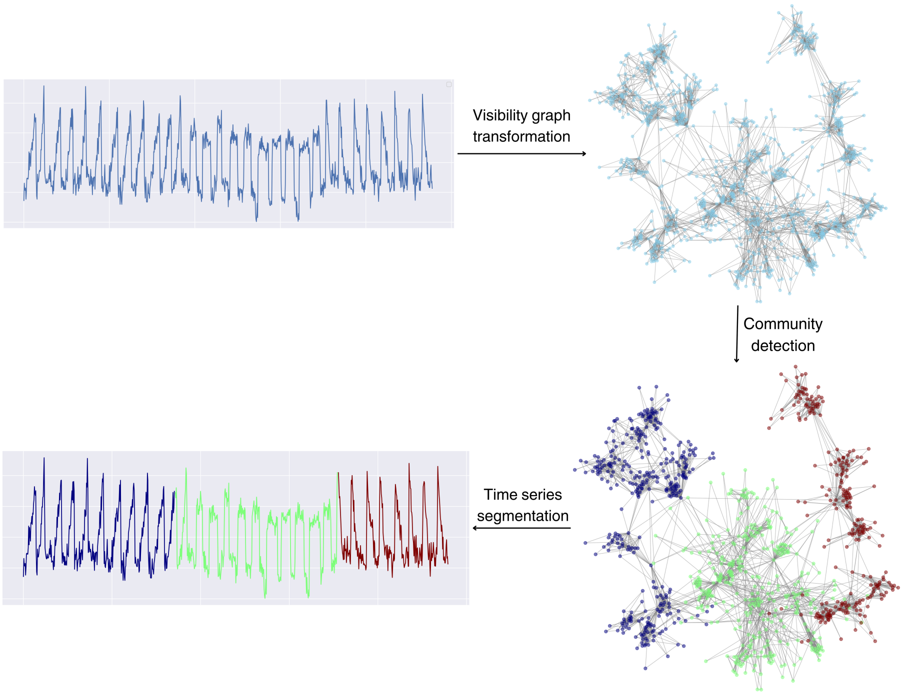

Figure 1: Visualization of time series segmentation using visibility graphs and community detection on the CBF dataset.

Dataset	True CPs	Detected CPs	Average Percentage Error	Mean Absolute Error
Adiac	[572, 1012, 1232]	[517, 1045, 1408]	9.05	88.0
Car	[577, 1154, 1550]	[323, 1235, 1675]	19.12	150.0
CBF	[384, 704]	[362, 711]	3.36	14.5
Coffee	[500]	[467]	6.60	33.0
Computers	[5625]	[6953]	23.61	1328.0
OSULeaf	[907, 1680]	[862, 1573]	5.66	76.0

Table 1: Comparison of True Change Points (CPs), Detected CPs, Average Percentage Error and Mean Absolute Error across different datasets.

References

[1] Z. He, S. Zhang, J. Hu, and F. Dai, "An adaptive time series segmentation algorithm based on visibility graph and particle swarm optimization," *Physica A: Statistical Mechanics and its Applications*, vol. 636, p. 129563, 2024.

[2] I. Lebedev and B. Rzayev, "Segmentation of data when analyzing the state of telecommunication systems," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 3, pp. 1473–1479, 2023.

[3] M. Lovrić, M. Milanović, and M. Stamenković, "Algoritmic methods for segmentation of time series: An overview," *Journal of Contemporary Economic and Business Issues*, vol. 1, no. 1, pp. 31–53, 2014.

[4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[5] M. Zheng, S. Domanskyi, C. Piermarocchi, and G. I. Mias, "Visibility graph based temporal community detection with applications in biological time series," *Scientific reports*, vol. 11, no. 1, p. 5623, 2021.

[6] H. Chen and N. Zhang, "Graph-based change-point detection," 2015.

[7] L. N. Ferreira and L. Zhao, "Time series clustering via community detection in networks," *Information Sciences*, vol. 326, pp. 227–242, 2016.

[8] T. Setiadi, R. Ratih, N. Arkiyah, and G. Prestisia, "Detection of book and borrower communities based on book borrowing records in the library using complex network analysis," in *Proceeding of International Conference on Information Science and Technology Innovation (ICoSTEC)*, vol. 2, no. 1, 2023, pp. 171–176.

[9] V. F. Silva, M. E. Silva, P. Ribeiro, and F. Silva, "Time series analysis via network science: Concepts and algorithms," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 3, p. e1404, 2021.

[10] A. Ermshaus, P. Schäfer, and U. Leser, "Clasp: parameter-free time series segmentation," *Data Mining and Knowledge Discovery*, 2023.