LVI SBPO SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
DE 04 ATÉ 07 NOVEMBRO 2024
FORTALEZA | CEARÁ

Topological Analysis of Graphs for Comparing Sets of Time Series: A Case Study with Hospital Data

Emerson Yoshiaki Okano
Instituto de Ciencia e Tecnologia (ICT), Universidade Federal de São Paulo(UNIFESP)
Av. Cesare M. G. Lattes, 1201 - Eugenio de Melo, São José dos Campos - SP
emerson.okano@unifesp.br

Daniel Aloise
École Polytechnique de Montréal
Montreal, H3C 3A7, Canada
daniel.aloise@polymtl.ca

Mariá C. V. Nascimento
Instituto Tecnológico de Aeronáutica (ITA)
Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos - SP, CEP: 12228-900, Brasil
mariah@ita.br

RESUMO

Analisar múltiplas séries temporais e suas correlações é extremamente desafiador devido à complexidade dos dados, variabilidade temporal e necessidade de técnicas avançadas para identificar padrões e prever comportamentos futuros. Este trabalho apresenta uma metodologia para esse fim ao analisar dados referentes à utilização de recursos hospitalares do Hospital Municipal de São José dos Campos em períodos que antecederam e em períodos concomitantes com a pandemia COVID-19. Nós extraímos *meta-features* de séries temporais para descrever as séries temporais de cada recurso e criamos grafos de dados de cada ano estudado baseados na distância Euclideana entre as *features*. Ao examinar a estrutura dos grafos, identificamos mudanças significativas no padrão de utilização de recursos durante a pandemia. Nossos resultados mostram que a dissimilaridade do perfil de consumo entre os itens aumentou e a conectividade dos grafos diminuiu, destacando o impacto da pandemia nas operações hospitalares.

PALAVRAS CHAVE. Análise em Grafos, Gestão em Saúde, COVID-19.

ABSTRACT

Analyzing multiple time series and their correlations is exceptionally challenging due to the complexity of the data, temporal variability, and the need for advanced techniques to identify patterns and predict future behaviors. This work presents a methodology for this purpose by analyzing data on the use of hospital resources at the Municipal Hospital of São José dos Campos before and during the COVID-19 pandemic. We extracted time-series meta-features to describe the time series of each resource and created graphs for each year based on the Euclidean distance between the features. By examining the structure of these graphs, we identified significant changes in the pattern of resource utilization during the pandemic. Our results show that the dissimilarity in consumption profiles between items increased, and the connectivity of the graphs decreased, highlighting the impact of the pandemic on hospital operations.

KEYWORDS. Graph Analytics. Healthcare Management. COVID-19.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

1. Introduction

According to the World Health Organization (WHO), the COVID-19 pandemic has infected more than 37 million and killed more than 700 thousand people. This significant number of cases and deaths caused disruptions in healthcare operations, overwhelming hospitals, straining medical resources, and challenging the health system capacity to provide adequate care to both COVID-19 and non-COVID-19 patients. To mitigate the impact of these disruptions and reduce deaths due to lack of resources in hospitals, time series analysis and forecasting have been employed to predict patient surges and optimize resource allocation [Borges and Nascimento, 2022; Francisquini et al., 2022; Junior et al., 2022].

During the COVID-19 pandemic, significant studies were published focusing on time series analysis to address various challenges posed by the crisis. For instance, Borges and Nascimento [2022] proposed a combined Prophet-LSTM method using exogenous variables, improving Intensive Care Unit (ICU) demand forecasting. Similarly, Junior et al. [2022] developed the *SafetyStock* system, a statistical modelling approach that combines historical data of personal protective equipment (PPE), current protocols for their uses, and epidemiological data related to COVID to build predictive models for PPE demand. Additionally, Francisquini et al. [2022] introduced a community-based time series anomaly detection method that leverages graph structures to successfully identify anomalies in the number of COVID-19 cases across different districts in São José dos Campos.

Despite significant advances in the time series literature related to COVID-19 data, we could not find any study "looking back" and analyzing how the COVID-19 pandemic altered the dynamic of hospital resource utilization. However, understanding this dynamic could improve the management of a future health crisis. Also, we could not find methods that contrast the interrelation of time series in different time frames. In this paper, we aim to analyze the data from resource utilization provided by the Municipal Hospital of São José dos Campos, contrasting differences in the relationship between resource utilization across different years contrasting non-pandemic and pandemic years.

In this paper, we conducted a topological analysis of graphs to compare the relationships between a set of time series across different years. First, we extracted meta-features to characterize each time series. Using these meta-features, we computed the Euclidean distances between pairs of time series to generate graphs for each year. Subsequently, we extracted graph features to analyze the topology of the graphs, allowing us to examine and compare the structural characteristics of the time series relationships over time. Using this methodology, we revealed significant deviations in the graph structure during the pandemic years, suggesting that the relationship between the resources became more dispersed and less connected during the pandemic, reinforcing the impact of the COVID-19 pandemic on hospital operations.

The remainder of the paper is organized as follows: Section 2 provides a brief literature review on time series characterization and graph characterization. Section 3 presents the methods applied in this work, detailing the meta-features extraction and the graph generation process. Section 4 displays the results obtained and provides a comprehensive data analysis. Finally, Section 5 outlines the main conclusions of this paper.

2. Literature review

This section provides a review of the related works in the field of time series analysis, focusing on multivariate time series, feature extraction and graph-based analysis.

2.1. Time series characterization / features

Time series characterization involves extracting meaningful features from data to summarize its underlying patterns, structures, and dynamics. This process is crucial for various analytical

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

tasks and has been applied across different domains, including time series classification [Fulcher and Jones, 2017; Lubba et al., 2019], clustering [Wang et al., 2006], and meta-learning [Talagala et al., 2018].

The Highly Comparative Time-Series Analysis (hctsa) toolkit [Fulcher and Jones, 2017] represents a comprehensive approach to time series analysis by providing a library of over 7,700 features for characterizing time series data. The hctsa library enables the extraction and comparison of thousands of time-series features, offering a rich set of descriptors that capture various statistical, information-theoretic, and model-based properties of time series. More recently, Lubba et al. [2019] refined this vast set of features into a group of 22 key features. The authors applied a feature selection pipeline to identify a small, canonical subset of features that each displays high performance across a range of tasks and exhibits complementary performance characteristics with each other, demonstrating comparable performance to hctsa on a set of 93 classification tasks.

Hyndman et al. [2019] developed a tool to extract features for characterizing time series data. Their work proposed diverse time series features, including distributional features, spectral features, autocorrelation measures, and more. These features have been successfully applied to various tasks, such as meta-learning for time series forecasting [Talagala et al., 2018] and time series clustering [Wang et al., 2006], highlighting their versatility and effectiveness in different analytical contexts.

Recently, Vallat and Walker [2021] applied entropy and fractal dimension-based features to classify sleep stages. Their study utilized spectral and fractal dimension features to capture the inherent complexity and variability of sleep data, enabling more accurate classification of different sleep stages. This approach underscores the power of meta-features in transforming raw time series data into meaningful descriptors that enhance the performance of classification algorithms. By leveraging these advanced feature extraction techniques, Vallat et al. demonstrated that incorporating meta-features can significantly improve the interpretability and accuracy of time series classification, paving the way for more sophisticated and effective analytical models in various domains.

2.2. Graph characterization

Graphs are well-known for their inherent capability to model relationships between vertices. Numerous researchers have applied this representation to capture and analyze diverse behaviors across various domains, including social networks, protein interactions, citation networks, webpages, and more [Barabási, 2013]. By leveraging the structural properties of graphs, these studies have uncovered complex patterns and dynamics within the data, providing valuable insights into the interconnected nature of these systems. Graphs offer a versatile and powerful tool for representing and understanding the multifaceted relationships within different data types.

Various methods have been employed to construct graphs and extract meaningful insights from time series data. Lacasa et al. [2008] proposed the visibility graph method, which converts time series data into a graph by representing each data point as a node and connecting nodes based on a visibility criterion, allowing the analysis of the properties of the time series. Anghinoni et al. [2019] developed a technique to extract six key features from time series, including noise, gradient, and extreme values (maximum and minimum). These features were then used to represent observations as graph nodes, connected based on their order of occurrence in time, facilitating the forecast of trends in stock market time series.

In another approach to time series analysis, Ferreira and Zhao [2016] focused on clustering time series using community detection in complex networks. Each time series sample was represented as a vertex in their study, with edge weights determined by various similarity measures. The authors compared similarity measures and community detection algorithms to cluster the time

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

series data effectively.

More recently, Mishra et al. [2022] proposed a novel graph-based time series data mining framework called Graft. This framework uses a piecewise aggregate approximation (PAA) algorithm for data reduction and symbolic representation. The symbolic data is then transformed into a graph-based representation that models co-occurrence patterns, providing a powerful tool for time series analysis focused on motif and anomaly detection.

3. Methods

In this study, we employed time series meta-features and graphs to investigate the interaction among multiple resource utilization time series over time. As far as we know, the literature is scarce on methods that contrast the interrelation of time series in different time frames. Our goal is to determine whether the relationship between time series modeling medical supplies differs between the periods before the COVID-19 pandemic and during the pandemic. Our analytical approach included time series data preprocessing, meta-feature extraction, feature selection, graph construction, and characterization.

3.1. Time series data preprocesing

The first step of our analysis is the series data preprocessing. Preprocessing transforms raw data into a format suitable for analysis and enhances the data quality by addressing various issues.

- **Data Integrity check**: In this step, we check the integrity of the data, ensuring that the data follows expected patterns or rules, checking for duplicates, and validating time order sequences to prevent data corruption.

- **Time series selection**: We deliberately excluded intermittent demand data from this study. This decision was based on the fact that such data, characterized by irregular and infrequent occurrences, can introduce significant challenges for analysis. Instead, we focused on resources with almost daily use, systematically excluding items with scarce demand patterns.

- **Noise elimination**: In this step, we performed a noise elimination based on the median and Median Absolute Deviation (MAD). For that, we remove the six MAD points from the median, minimizing the risk of erroneously discarding valid data points. After identifying and removing these outliers, we replaced them using linear interpolation to maintain the continuity of the dataset. The high threshold of six MADs was chosen to ensure we did not inadvertently remove significant data, especially those affected by COVID-19.

3.2. Time series feature extraction

One of the most important steps of time series data analysis is to extract patterns from data series to describe them. For this task, features can be useful since they can provide unique insights into the time series data, enhancing our understanding and facilitating further analysis. We employed the frameworks Antropy [Vallat, 2024], Catch22 [Lubba et al., 2019], TSfeatures [Hyndman et al., 2023] and TSfresh [Christ et al., 2018], widely employed in the literature to obtain feature values. The frameworks and the respective meta-features they extracted are presented next.

Antropy

Antropy calculates a set of entropy-based and fractal dimension features for time series. In the context of time series analysis, entropy and fractal dimensions measure the unpredictability or randomness of the data. Table 1 presents the meta-features extracted using Antropy.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

No	Meta-feature	No	Meta-feature
1	Approximate entropy	6	Katz fractal dimension
2	Detrended fluctuation	7	Petrosian fractal dimension
3	Higuchi fractal dimension	8	Spectral entropy
4	Hjorth Complexity	9	Singular value decomposition entropy
5	Hjorth Mobility		

Table 1: Features extracted by Antropy

Catch22

Catch-22 [Lubba et al., 2019] calculates a collection of 22 time series features initially designed for efficient time series classification. It represents a curated subset of features from the more extensive set known as Highly Comparative Time-Series Analysis (hctsa) toolkit [Fulcher and Jones, 2017] (hctsa) , which includes over 7,700 features. The selection of these 22 features was based on their ability to balance computational efficiency with discriminative power for time series data across a wide range of applications. Table 2 presents the meta-features extracted by Catch22.

No	Meta-feature	No	Meta-feature
1	5-bin histogram mode	14	Entropy of successive pairs in symbolized series
2	10-bin histogram mode	15	Histogram-based automutual information (lag 2, 5 bins)
3	Positive outlier normalized timing	16	Time reversibility
4	Negative outlier normalized timing	17	First minimum of the AMI function
5	First $\frac{1}{e}$ crossing of the ACF	18	Transition matrix column variance
6	First minimum of the ACF	19	Wang's periodicity metric
7	Power in lowest 20% frequencies	20	Goodness of exponential fit to embedding distance distribution
8	Centroid frequency	21	Rescaled range fluctuation analysis (low-scale scaling)
9	Error of 3-point rolling mean forecast	22	Detrended fluctuation analysis (low-scale scaling)
10	Change in autocorrelation timescale after incremental differencing		
11	Proportion of high incremental changes in the series		
12	Longest stretch of above-mean values		
13	Longest stretch of decreasing values		

Table 2: Features extracted by catch22 [Lubba et al., 2019]

TSfeatures

TSfeatures [Hyndman and Athanasopoulos, 2018] calculates a broad range of features from time series data, including trend, seasonality, autocorrelation, and other statistical properties, facilitating robust time series analysis and forecasting. Table 3 presents the 31 meta-features obtained using TSfeatures.

TSfresh

TSfresh (Time Series Feature Extraction on basis of Scalable Hypothesis tests) [Christ et al., 2018] is a Python library designed to automate extracting relevant features from time series data. Table 4 presents the 3 meta-features extracted using TSfresh.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

No	Meta-feature	No	Meta-feature
1	Alpha parameter from Holt decomposition	15	Mean of the absolute values of the remainder component of STL decomposition
2	Lagrange Multiplier test for autoregressive conditional heteroscedasticity	16	Variance of the absolute value of the remainder component of STL decomposition
3	Sum of squares of the first 12 autocorrelations of x_t^2	17	Sum of squares of the first 12 autocorrelations of z_t^2 (residuals of GARCH(1,1))
4	R^2 value of an autoregressive model applied to x_t^2	18	R^2 value of an AR model applied to z_t^2
5	β parameter from Holt decomposition	19	Hurst coefficient
6	Number of times a time series crosses the median line	20	Linearity of time series based on the coefficient of an orthogonal quadratic regression
7	Curvature of a time series calculated based on the coefficients of an orthogonal quadratic regression	21	Strength of peaks based on STL decomposition
8	First autocorrelation function for the differenced series	22	Seasonal period
9	Sum of squares of the first 10 autocorrelation function values of the differenced series	23	Strength of seasonal component in STL decomposition
10	First autocorrelation function for the twice-differenced series	24	Sum of the first period of the autocorrelation function of the time series
11	Sum of squares of the first 10 autocorrelation function values of the twice-differenced series	25	“Spikiness” of a time series
12	Shannon entropy of the spectral density of the data	26	Strength of trend in STL decomposition
13	First autocorrelation function of the remainder component of the STL decomposition	27	Strength of trough in STL decomposition
14	Sum of squares of the first 10 autocorrelation function values of the remainder component of the STL decomposition	28	Statistic for the Kwiatkowski et al. unit root test with linear trend and lag 1
		29	Z-tau version of Phillips & Perron unit root test with constant trend and lag 1
		30	First autocorrelation coefficient
		31	Sum of squares of the first 10 autocoefficients

Table 3: Meta features extracted by TSfeatures

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

No	Meta-feature
1	Expected value of the first digit when the first digit follows Benford's law
2	Kurtosis
3	Skewness

Table 4: Meta-features extracted by TSfresh

3.3. Feature selection

We performed a feature selection based on their correlation to reduce multicollinearity and potentially enhance interpretability. To achieve this, we first calculated the correlation coefficients between all extracted features. This step involved generating a correlation matrix that helped us identify highly correlated variables.

Subsequently, we constructed an undirected graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of vertices and $E = \{\langle v_i, v_j \rangle : v_i, v_j \in V$ and $\text{Abs}(\text{Corr}(v_i, v_j) > 0.7\}$ is the set of edges. Each vertex $v_i \in V$ represents a meta-feature, and each edge $\langle v_i, v_j \rangle \in E$ connects vertices v_i and v_j if the absolute value of the correlation between their corresponding features exceeds 0.7. Moreover, the cost associated with the edge $\langle v_i, v_j \rangle$ is defined as the absolute value of the correlation between the corresponding meta-features expressed as $c_{i,j} = \text{abs}(\text{corr}(v_i, v_j))$. Utilizing this graph, we then determined the optimal community structure by maximizing the weighted modularity score [Newman, 2006]. The weighted modularity of a given partition is given by the difference between the sum of weights inside communities and the expected sum of weights of links in these communities. The modularity maximization is a largely employed community detection strategy whose definition of the number of clusters of the partition with the highest modularity is not a required information.

Within each cluster, we selected the feature with the highest centrality value according the so-called to PageRank measures [Page et al., 1999]. PageRank is a centrality measure that helps identify the most influential feature within a cluster, ensuring that the selected features are less redundant and hold significant interpretative power. In this work, we calculated the PageRank for undirected weighted graphs using Igraph.

3.4. Graph generation and meta-feature extraction

We segmented the data based on the timestamp to analyze the time series, adopting a yearly granularity. This approach divided the data into six distinct datasets from 2016 through 2021. For each yearly segment, we extracted the selected time series meta-features. Subsequently, we constructed a graph $G = (V, E)$ for each year, where $V = R = \{r_1, r_2, \ldots, r_n\}$ represents the hospitalar items and $E = \{\langle r_i, r_j \rangle : r_i, r_j \in R$ and $||f(r_i) - f(r_j)||_2 \leq \tau\}$, i.e. items r_i and r_j are adjacent if the Euclidean distance between the features extracted from the items time series does not exceed the threshold τ. In addition, the cost associated with the edge $\langle r_i, r_j \rangle$ is precisely the Euclidean distance between the features extracted from its respective items $c_{i,j} = ||f(r_i) - f(r_j)||_2$. Here, $f(\cdot)$ is a function $f : \mathbb{R}^L \to \mathbb{R}^F$ that maps a time series with length L to a vector of meta-features of length F.

To compare the structural changes across different years, we extracted graph-based meta-features displayed in Table 5. By following this method, we can analyze and quantify the variations in the graph structure from one year to another, providing insights into different patterns of material utilization between the yearly datasets.

4. Data Analysis

In this section, we will examine the hospital data by applying the previously described methodology. Our first step involves describing the hospital data, specifically focusing on the distri-

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

Meta-feature	Description	Meta-feature	Description
min_V_cost	Min. vertex cost	median_edge_cost	Median of edge costs
max_V_cost	Max. vertex cost	avg_geodesic_dist	Average geodesic distance
avg_V_cost	Average vertex cost		
std_V_cost	Standard deviation of vertex costs	harmonic_geodesic_dist	Harmonic mean of geodesic distances
median_V_cost	Median of vertex costs	global_efficiency	Global efficiency
		network_vulnerability	Network vulnerability
nn_sum	Sum of the edge costs from each vertex to its nearest neighbor	clustering_coefficient	Clustering coefficient for weighted networks
		cluster_coef4trans	Clustering coefficient for transitivity
min_edge_cost	Lowest edge cost	maximum_degree	Maximum degree of vertices
avg_edge_cost	Average edge cost		
std_edge_cost	Standard deviation of edge costs	degree_dist_entropy	Entropy of the degree distribution

Table 5: Extracted graph meta features

bution and frequency of material utilization. This analysis is crucial to identifying and subsequently excluding items with intermittent demand from our dataset.

4.1. Hospital data

In this study, we utilized data provided by the Municipal Hospital of São José dos Campos, encompassing the consumption of various materials between 2016 and 2021 (2192 days). This dataset contains data from before and during the COVID-19 pandemic, containing 1,466 distinct materials ranging from medicines to personal protective equipment(PPE).

The histogram presented in Figure 1 shows the relationship between the item and the number of days that this item was used in the 2,192 days. The data reveal many items with intermittent demand, common in hospital settings where resource needs fluctuate. To focus on consistently used items, we implemented a frequency-based filter to exclude materials with sporadic consumption patterns. Specifically, we retained items that registered more than 2,000 usage instances throughout the period under review. This process of selective filtering honed our dataset to 187 materials.

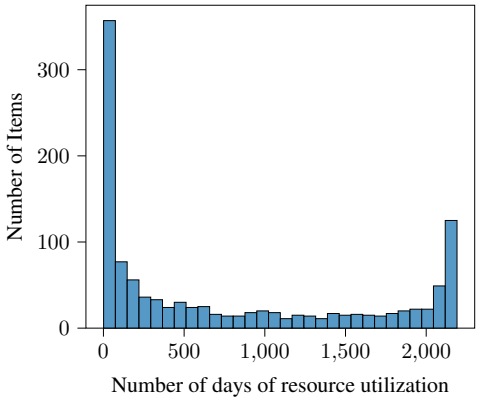


Figure 1: Histogram of Material Usage Frequency: relation between the number of days that materials were used vs. number of items corresponding to that number of days

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

4.2. Meta-feature selection

In the meta-feature selection step, we computed the time series meta-features across all the years from 2016 to 2021. This comprehensive approach allows us to capture overarching trends and patterns within the data. By analyzing the time series across multiple years, we ensure that our insights reflect the general behavior of the dataset. Building on the methodology outlined in Section 3.3, we generated a graph to visualize the relationships between features. This allowed us to eliminate highly correlated features, thus reducing redundancy. The PageRank centrality measure [Page et al., 1999] within our graph-based framework allowed us to pinpoint and select the most informative features. This selection strategy enhances the robustness and relevance of our analysis, removing the highly correlated features and keeping the "most informative" feature in each cluster.

This graph-based method reduced the number of features from 68 to 24. The selected features are: Feature 8 from Antropy; Features 1, 2, 3, 10, 11, 13, 16, 17, 18, 21 and 22 from catch22; Features 5, 7, 16, 17, 18, 20, 22, 23, 24, 28 and 29 from TSfeatures; and Feature 1 from TSfresh.

4.3. Graph generation and feature extraction

After the feature selection, we generated graphs where the hospital material corresponds to the nodes of the graph (as explained in Section 3.4). To determine an appropriate threshold τ, Figure 2 presents a histogram showing the distribution of the Euclidean distances between the features extracted from the time series of the items segmented into periods before and during the COVID-19 pandemic. The left panel, labeled "No COVID", encompasses the years 2016 to 2019, while the right panel, labeled "COVID", includes the years 2020 and 2021.

In the "No COVID" period, the distribution of Euclidean distances is skewed towards lower values, with most distances falling below 10. This suggests that the time series meta-features extracted from the items in these years are similar, leading to many close connections between items. The consistent pattern from 2016 to 2019 indicates stable relationships and patterns in the data before the pandemic. Conversely, the "COVID" period shows a slight shift in the distribution of Euclidean distances, particularly in 2021. While most distances still fall below 10, there is a noticeable increase in distances between 10 and 20. This shift implies that the relationships between features became more varied during the pandemic, likely reflecting changes in hospital operations and resource utilization in response to COVID-19.

To create the graphs for each year, we fixed the value of τ at 7.5, corresponding to the 75th percentile of the edge distances from 2016. This threshold ensures that we capture the majority of significant connections while excluding the more distant, less representative edges. We generated the graphs using this τ value, ensuring they reflect meaningful relationships between the features based on their Euclidean distances.

Subsequently, we extracted meta-features from the graphs generated for each year, which are presented in Table 6. This table reveals significant structural changes in the graph topology regarding the years 2020 and 2021, coinciding with the COVID-19 pandemic. Key observations include increases in vertex and edge costs, longer average geodesic distance, reduced global efficiency, higher network vulnerability, and decreased clustering coefficients. These changes in the graph structure indicate that the distance between vertices (items) increased, as captured by the Euclidean distance based on the meta-features extracted from their time series. This suggests that the relationships between items became more dispersed and less cohesive during the pandemic years.

The increase in the average vertex and edge costs suggests that the features representing the items showed more significant variability and deviation from each other. The longer average shortest paths and reduced global efficiency imply that the network became less interconnected and more fragmented. At the same time, the decreased clustering coefficients indicate a reduction in the local grouping of items, further emphasizing the loss of tight-knit clusters within the network.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

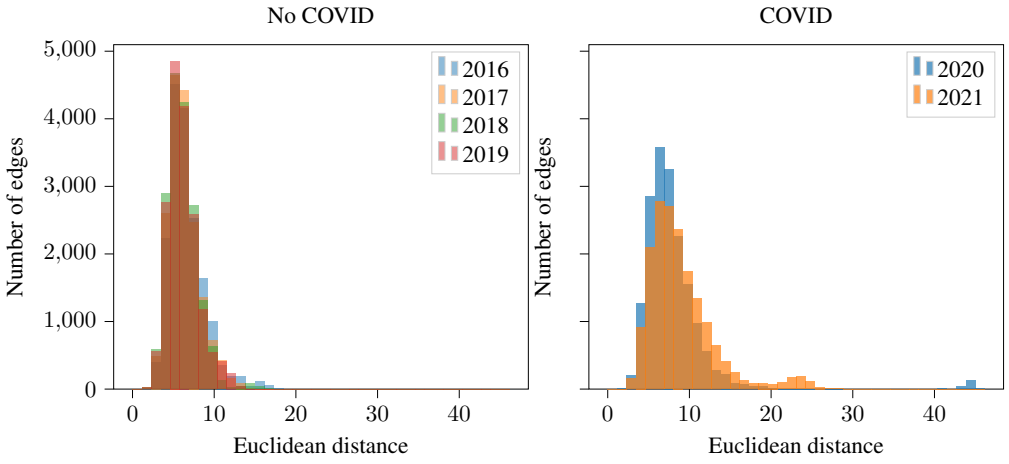


Figure 2: Histogram illustrating the distribution of Euclidean distances between features extracted from the time series of items, segmented into periods before and during the COVID-19 pandemic. The histograms use 40 bins with a bin width of 1.15. The left panel, labeled "No COVID", encompasses the years 2016 to 2019, while the right panel, labeled "COVID", includes the years 2020 and 2021.

Overall, these structural changes in the graph point to a significant impact of the COVID-19 pandemic on the underlying relationships and dynamics within the hospital's resource utilization network. The increased Euclidean distances between items meta-features suggest that resource usage patterns became less predictable, likely due to the unprecedented demands and operational challenges posed by the pandemic.

To better visualize the data, Figure 3 presents the relation between the two principal components according to the Principal Component Analysis (PCA). In this figure, we observe that the data points for the period unaffected by COVID-19 (2016 to 2019) cluster closely together. This proximity suggests that the underlying patterns and relationships in the data during these years were similar, indicating stable and consistent behavior in hospital resource utilization and operations before the pandemic. In contrast, the data points for 2020 and 2021 are distinctly separated from those of the pre-COVID years, with this separation primarily occurring along the PC1 axis. This indicates a significant shift in the underlying data patterns due to the impact of the COVID-19 pandemic, reflecting changes in hospital operations and resource utilization during this period.

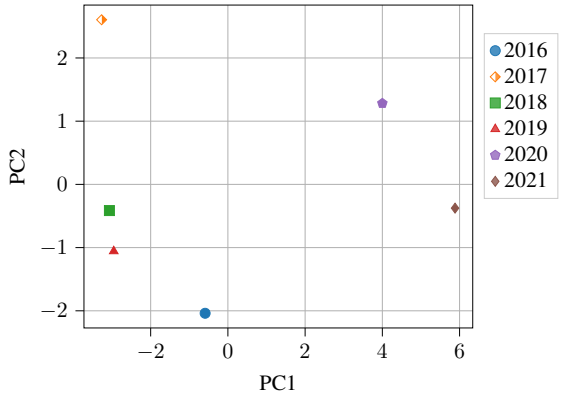


Figure 3: Projection of two principal components from time series meta-features, explaining 86.9% of variance (PC1: 73.9%, PC2: 13.0%). The plot highlights differences in resource utilization between pre-pandemic (2016-2019) and pandemic years (2020-2021) along the PC1 axis.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

	2016	2017	2018	2019	2020	2021
min_V_cost	0.00	4.55	0.00	0.00	0.00	0.00
max_V_cost	7.13	7.43	7.21	7.24	7.21	7.18
avg_V_cost	5.62	5.62	5.53	5.54	5.90	5.74
std_V_cost	0.85	0.60	0.73	0.73	0.92	1.38
median_V_cost	5.60	5.47	5.45	5.46	5.93	5.98
nn_sum	615.35	582.60	579.05	573.67	665.45	663.84
min_edge_cost	1.28	1.69	1.64	1.36	1.89	1.45
avg_edge_cost	5.53	5.49	5.46	5.45	5.82	5.84
std_edge_cost	1.08	1.11	1.14	1.12	1.09	1.08
median_edge_cost	5.58	5.54	5.50	5.48	5.95	6.00
avg_geodesic_dist	6.91	6.75	6.41	6.54	8.61	9.74
global_efficiency	0.84	0.89	0.90	0.89	0.75	0.64
harmonic_geodesic_dist	6.04	5.91	5.72	5.76	7.20	7.94
network_vulnerability	4.82e-03	1.07e-03	9.82e-04	1.92e-03	4.83e-03	3.56e-03
cluster_coef4trans	0.92	0.93	0.92	0.92	0.83	0.79
max_degreee	169	176	179	177	157	136
clustering_coefficient	0.65	0.67	0.66	0.66	0.63	0.58
degree_dist_entropy	5.45	5.25	5.50	5.45	6.24	6.18

Table 6: Meta-features extracted from graph (2016-2021)

5. Conclusions

In this study, we explored graph-based techniques for comparing the relationships between time series data over multiple years. By extracting meta-features from the time series, we could construct graphs representing the relationships between these series through Euclidean distances. We then analyzed the features extracted from these graphs to understand the structural changes over time, particularly in the context of the COVID-19 pandemic.

Our findings revealed significant deviations in the graph structure during the pandemic years, characterized by increased vertex and edge costs, higher average geodesic distance, reduced global efficiency, higher network vulnerability, and decreased clustering coefficient. This suggests that the relationship between resource usage became more dispersed and less cohesive during the pandemic compared with non-pandemic years, demonstrating the effectiveness of time series feature extraction integrated with graph representation for multivariate time series analysis.

Acknowledgment

We acknowledge the financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance Code 001, Grant: 88887.507037/2020-00), CNPq (grants 309385/2021-0, 403735/2021-1) and the health department of São José dos Campos for providing the COVID data. This research was conducted using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI), funded by FAPESP (grant 2013/07375-0).

References

Anghinoni, L., Zhao, L., Ji, D., and Pan, H. (2019). Time series trend detection and forecasting using complex network topology analysis. *Neural Networks*, 117:295–306.

Barabási, A.-L. (2013). Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375.

Borges, D. and Nascimento, M. C. V. (2022). Covid-19 icu demand forecasting: A two-stage prophet-lstm approach. *Applied Soft Computing*, 125:109181.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307:72–77.

Ferreira, L. N. and Zhao, L. (2016). Time series clustering via community detection in networks. *Information Sciences*, 326:227–242.

Francisquini, R., Lorena, A. C., and Nascimento, M. C. V. (2022). Community-based anomaly detection using spectral graph filtering. *Applied Soft Computing*, 118:108489.

Fulcher, B. D. and Jones, N. S. (2017). hctsa: A computational framework for automated time-series phenotyping using massive feature extraction. *Cell systems*, 5(5):527–531.

Hyndman, R., Kang, Y., Montero-Manso, P., O'Hara-Wild, M., Talagala, T., Wang, E., and Yang, Y. *tsfeatures: Time Series Feature Extraction*, 2023. https://pkg.robjhyndman.com/tsfeatures/, https://github.com/robjhyndman/tsfeatures.

Hyndman, R., Kang, Y., Montero-Manso, P., Talagala, T., Wang, E., Yang, Y., O'Hara-Wild, M., et al. (2019). tsfeatures: Time series feature extraction. *R package version*, 1(0).

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Junior, O. A. G., Nascimento, D. C., Russo, C. M., Henriques, M. J., Tomazella, C. P., Santos, M. O., Neves, D., Assad, D., Guerra, R., Bertazo, E. K., et al. (2022). Safety-stock: Predicting the demand for supplies in brazilian hospitals during the covid-19 pandemic. *Knowledge-Based Systems*, 247:108753.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., and Nuno, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105 (13):4972–4975.

Lubba, C. H., Sethi, S. S., Knaute, P., Schultz, S. R., Fulcher, B. D., and Jones, N. S. (2019). catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis. *Data Mining and Knowledge Discovery*, 33(6):1821–1852.

Mishra, K., Basu, S., and Maulik, U. (2022). Graft: A graph based time series data mining framework. *Engineering Applications of Artificial Intelligence*, 110:104695.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

Page, L., Brin, S., Motwani, R., Winograd, T., et al. (1999). The pagerank citation ranking: Bringing order to the web.

Talagala, T. S., Hyndman, R. J., Athanasopoulos, G., et al. (2018). Meta-learning how to forecast time series. *Monash Econometrics and Business Statistics Working Papers*, 6(18):16.

Vallat, R. (2024). Web page. `https://raphaelvallat.com/antropy/build/html/index.html`. Access: 2024-05-24.

Vallat, R. and Walker, M. P. (2021). An open-source, high-performance tool for automated sleep staging. *Elife*, 10:e70092.

Wang, X., Smith, K., and Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data mining and knowledge Discovery*, 13:335–364.

DOI: 10.59254/sbpo-2024-193760
https://proceedings.science/p/193760?lang=pt-br