

# Imputação Adaptativa em Espaço Latente para Séries Temporais de Vazão de Rede

Maria C. M. M. Ferreira<sup>1</sup>, Ismael S. F. de Castro<sup>1</sup>, Maria de L. Linhares<sup>1</sup>,  
Rafael L. Gomes<sup>1</sup>

<sup>1</sup>Universidade Estadual do Ceará (UECE)

{clara.mesquita, ismael.fonteles, malu.linhares}@aluno.uece.br

{rafa.lobes}@uece.br

**Abstract.** *Throughput monitoring is essential for Quality of Service (QoS), but measurement failures create gaps that compromise network performance forecasting. In this context, this work proposes an algorithm that reconstructs time series by combining Hankel matrices with latent space imputation. Experimental results using real-world data indicate that the proposal preserves network temporal dynamics better than traditional methods, reducing imputation errors and increasing the reliability of predictive models such as GRU and LSTM.*

**Resumo.** *O monitoramento de vazão é essencial para a Qualidade de Serviço (QoS), mas falhas de medição geram lacunas que comprometem a previsão de desempenho da rede. Dentro deste contexto, este trabalho propõe um algoritmo que reconstrói séries temporais combinando matrizes de Hankel e a imputação em espaço latente. Resultados experimentais, com dados reais, indicam que a proposta preserva a dinâmica temporal da rede superiormente aos métodos tradicionais, reduzindo erros de imputação e aumentando a confiabilidade de modelos preditivos como GRU e LSTM.*

## 1. Introdução

O monitoramento contínuo do desempenho de rede é essencial para garantir confiabilidade, eficiência e satisfação do usuário em sistemas distribuídos. Infraestruturas modernas de comunicação dependem de diferentes métricas de Qualidade de Serviço (QoS), como vazão, latência, jitter, perda de pacotes e disponibilidade, para avaliar e assegurar o desempenho dos serviços em aplicações e tecnologias diversas [Younas et al. 2023, Ghahramani et al. 2017, Brito et al. 2025]. Esses indicadores expressam o quanto a rede atende aos requisitos de usuários e aplicações, influenciando diretamente a efetividade de serviços sensíveis ao tempo, como computação em nuvem, transmissão de vídeo e sistemas de saúde conectados [Souza et al. 2024]. Nesse contexto, a análise dessas métricas ultrapassa a dimensão estritamente operacional e torna-se fundamental para o gerenciamento inteligente da rede, para o uso eficiente de recursos e para o suporte a aplicações orientadas por dados.

Em ambientes contemporâneos, o controle de QoS também está diretamente associado à capacidade de adaptação da infraestrutura a cargas dinâmicas e cenários heterogêneos. No domínio da computação em nuvem, por exemplo, a QoS é indispensável

para o cumprimento de acordos de nível de serviço (SLAs), permitindo alocação previsível de recursos e provisão confiável de serviços mesmo diante de flutuações de demanda [Zheng et al. 2013, Nobre et al. 2025, Lopes Gomes and Roberto Mauro Madeira 2012]. De forma semelhante, em sistemas veiculares e em arquiteturas baseadas em IoT, a gestão de QoS é crucial para manter baixa latência e conectividade estável sob restrições de mobilidade e largura de banda, afetando diretamente a segurança do usuário e a confiabilidade operacional [Alaya et al. 2021, Younas et al. 2023]. Em redes definidas por software (SDNs), por sua vez, a QoS pode ser empregada como referência para o controle de métricas como vazão, latência e largura de banda [Adedayo and Twala 2017, Gomes et al. 2016, Gomes et al. 2014]. Assim, sua observação contínua é particularmente relevante para os temas da WGRS, ao subsidiar mecanismos de gerenciamento autônomo, provisionamento de redes e planejamento de capacidade.

Entre as métricas de QoS, a vazão ocupa posição de destaque por representar o volume de dados efetivamente entregue pela rede em um determinado intervalo de tempo. Por refletir simultaneamente a utilização da largura de banda e a capacidade do sistema de sustentar a qualidade do serviço sob diferentes condições de tráfego, a vazão constitui uma medida central para avaliação de desempenho, detecção de congestionamentos, alocação de recursos e suporte à garantia de QoS. Não por acaso, diversos trabalhos concentram-se em estratégias para melhorar ou analisar essa métrica [Arquam and Kumari 2025, Ju and Zhang 2014, Liu et al. 2023, Liu et al. 2022, Ferreira et al. 2024]. Além disso, a predição de vazão vem se tornando especialmente importante em redes modernas, uma vez que permite antecipar gargalos, ajustar recursos proativamente e apoiar decisões de controle antes que a degradação de desempenho se materialize [Minovski et al. 2023].

Entretanto, a confiabilidade de qualquer análise ou mecanismo preditivo baseado em vazão depende criticamente da qualidade e da continuidade dos dados subjacentes. Séries temporais de vazão frequentemente apresentam valores ausentes ou corrompidos em decorrência de falhas de medição, interrupções de coleta ou inconsistências no processo de monitoramento. Tais problemas comprometem a interpretação do comportamento da rede, dificultam a estimação de largura de banda, prejudicam a identificação de tendências e podem induzir conclusões incorretas ou mal definidas [Khayati et al. 2020]. Apesar disso, muitos estudos orientados por dados em redes ainda tratam ausências de forma ad hoc, ou simplesmente as ignoram [Sharma et al. 2025], o que reduz a robustez das etapas posteriores de análise e predição.

Nesse sentido, a qualidade dos dados de monitoramento deve ser vista como um requisito central para a operação eficiente da rede. Para que seja possível analisar QoS e QoE, compreender o comportamento da infraestrutura, identificar anomalias e reduzir riscos de indisponibilidade, é necessário monitorar adequadamente as métricas de desempenho. Contudo, o monitoramento de rede envolve um conjunto complexo de estratégias voltadas à integridade, segurança e observabilidade da infraestrutura, constituindo uma tarefa desafiadora para administradores e operadores [Alkenani and Nassar 2022]. No contexto da WGRS, esse desafio se conecta diretamente à necessidade de construir soluções mais autônomas, confiáveis e orientadas a dados reais, capazes de sustentar decisões de gerenciamento mesmo diante de medições incompletas.

Para enfrentar esse problema, este trabalho propõe uma nova abordagem para imputação de valores ausentes em séries temporais de vazão, com o objetivo de melhorar a

qualidade dos dados disponíveis para análise e predição. Especificamente, propomos o algoritmo *Low-rank Adaptive Reconstruction via Hankel* (LARK), bem como uma variante simplificada denominada *Periodic Imputation via Latent Analysis* (PILA). A ideia central é preservar as estruturas temporais intrínsecas da série, como padrões periódicos e tendências, por meio de sua projeção em um espaço latente no qual seu comportamento fundamental possa ser representado de forma compacta. Diferentemente de técnicas convencionais de preenchimento, que frequentemente suavizam excessivamente os dados, a metodologia proposta utiliza análise espectral adaptativa e representação matricial de baixa ordem para manter a forma e a dinâmica da série temporal.

O processo de imputação é composto por quatro etapas principais. Primeiramente, realiza-se a detecção adaptativa de período por meio da Transformada Rápida de Fourier, com o objetivo de estimar a periodicidade dominante da série. Em seguida, a série temporal é incorporada em uma matriz, seja pela construção de uma matriz de trajetória de Hankel no LARK, seja por uma matriz alinhada ao período no PILA. Na terceira etapa, aplica-se a Decomposição em Valores Singulares (SVD) para extrair um espaço latente capaz de capturar a estrutura subjacente de baixa ordem da série. Por fim, utiliza-se interpolação ponderada por  $k$ -vizinhos mais próximos (KNN) nesse espaço latente para estimar os valores ausentes a partir de padrões temporais semelhantes. A combinação entre processamento de sinais e aprendizado de máquina produz, assim, uma abordagem de imputação ajustada à dinâmica inerente aos dados, buscando reconstruir segmentos faltantes sem sacrificar variabilidade nem características relevantes da série.

Os benefícios da metodologia proposta são evidenciados por meio de experimentos abrangentes usando dados reais da rede GEANT, nos quais a qualidade da imputação é avaliada segundo diferentes métricas.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve a solução proposta; a Seção 4 detalha os experimentos e discute os resultados obtidos; e, por fim, a Seção 5 apresenta as conclusões e direções para trabalhos futuros.

## 2. Trabalhos Relacionados

A forma mais simples de tratar dados faltantes consiste no uso de estratégias de imputação única, nas quais as lacunas são preenchidas com valores plausíveis, em vez de se descartar observações ou empregar abordagens mais complexas, como esquemas de imputação múltipla. Entre os métodos tradicionais dessa categoria, destacam-se substituição por média ou mediana, preenchimento progressivo ou regressivo, como no caso da interpolação linear ou por spline, além de técnicas de suavização baseadas em médias móveis. Tais métodos apresentam implementação simples, baixo custo computacional e, em muitos casos, desempenho satisfatório quando a taxa de ausência é reduzida ou quando as lacunas são curtas. Como exemplo, um estudo recente sobre séries temporais fisiológicas mostrou que métodos simples, como imputação pela média, KNN e interpolação spline, preservaram de maneira razoável a estrutura de autocorrelação dos dados em cenários com pequenas proporções de ausência [Niako et al. 2024]. Nesse mesmo estudo, essas técnicas simples produziram melhor acurácia preditiva em modelos ARIMA e LSTM aplicados posteriormente, quando apenas poucos pontos estavam ausentes [Niako et al. 2024]. Além disso, métodos baseados em suavização, como médias móveis exponencialmente ponderadas e

filtros de Kalman, também apresentaram desempenho robusto em diferentes cenários de ausência [Niako et al. 2024].

Para além dessas abordagens mais elementares, a literatura apresenta métodos mais sofisticados que procuram explorar a estrutura intrínseca dos dados temporais. De forma ampla, duas grandes classes de técnicas têm sido investigadas: métodos de completamento matricial e métodos baseados em correspondência de padrões, ou proximidade [Khayati et al. 2020]. A abordagem de completamento matricial formula o problema como a reconstrução de entradas ausentes em uma matriz parcialmente observada, assumindo que a matriz completa subjacente possui baixo posto. No contexto de múltiplas séries temporais relacionadas, ou mesmo de uma única série transformada em forma matricial, isso equivale a assumir que a dinâmica temporal pode ser representada em um subespaço de menor dimensionalidade.

Técnicas dessa classe geralmente constroem uma matriz a partir dos dados temporais e, em seguida, aplicam algoritmos como SVD ou minimização da norma nuclear para preencher os valores faltantes. Um exemplo particularmente relevante nesse contexto é o uso de matrizes de Hankel de baixo posto. Métodos baseados em Análise de Espectro Singular (*Singular Spectrum Analysis* – SSA) e suas variantes inserem-se nessa categoria. Nessas abordagens, uma única série temporal é embutida em uma matriz de Hankel de alta dimensionalidade, em que cada linha corresponde a uma janela temporal deslocada da série, de modo que tendências e periodicidades passam a se manifestar como estruturas aproximadamente de baixo posto. A imputação pode então ser realizada por meio da aproximação dessa matriz de Hankel por outra matriz de menor posto que melhor se ajusta às entradas observadas.

De forma mais ampla, métodos baseados em fatoração matricial, incluindo SVD, fatoração matricial não negativa e variantes de PCA robusta, têm apresentado desempenho superior ao de interpolações ingênuas, especialmente em cenários com lacunas em bloco, nos quais longas sequências contíguas de observações estão ausentes [Khayati et al. 2020].

Nesse contexto, Khattab et al. propuseram um algoritmo baseado em padrões denominado GMA (*Gap Matching Algorithm*), que exemplifica essa linha de pesquisa [Khattab et al. 2023]. O método inicialmente determina o período dominante da série por meio de análise de Fourier e, em seguida, para cada lacuna, identifica as sequências históricas mais semelhantes à esquerda e à direita do intervalo ausente, considerando janelas com tamanho igual a um período [Khattab et al. 2023]. A partir da reprodução desses padrões vizinhos, o método reconstrói o segmento faltante. Métodos como o GMA evidenciam a importância de preservar formas temporais locais, aspecto que por vezes é negligenciado por abordagens globais de fatoração matricial.

Apesar da amplitude da literatura sobre imputação em séries temporais, ainda há relativamente poucos trabalhos voltados especificamente para séries de vazão de rede ou métricas de QoS. Muitos estudos em monitoramento de redes assumem que os dados estão completos ou aplicam métodos convencionais de forma direta, sem investigar adaptações específicas ao domínio. Como exemplo, em um estudo recente sobre previsão de vazão em redes celulares 5G, os autores utilizaram apenas técnicas básicas de imputação de valores ausentes, como preenchimento pela mediana e codificação *one-hot* para lacunas

em atributos categóricos, tratando essa etapa apenas como pré-processamento do conjunto de dados [Sharma et al. 2025]. Embora essa estratégia tenha contribuído para melhorar a acurácia do modelo de aprendizado, a imputação em si não foi objeto central de estudo e permaneceu restrita ao uso de ferramentas convencionais.

De forma geral, dados de desempenho de rede apresentam desafios particulares. Essas séries podem exibir comportamento não estacionário, sazonalidades em múltiplas escalas, como ciclos diários e semanais de uso, além de mudanças abruptas decorrentes de eventos externos ou reconfigurações da infraestrutura. Tais características sugerem que abordagens de imputação especializadas, capazes de se adaptar a periodicidades variáveis e de preservar eventos abruptos ou transitórios, tendem a ser mais adequadas para fluxos de dados de rede. Este trabalho é motivado precisamente por essa lacuna na literatura.

### 3. Proposta

Esta seção descreve a metodologia adotada para a obtenção dos dados utilizados na modelagem da proposta e na configuração experimental, bem como apresenta o algoritmo proposto, empregado para realizar a imputação de dados no contexto do pipeline de predição proposto.

#### 3.1. Dados de vazão

Os conjuntos de dados analisados neste trabalho são oriundos da GÉANT Performance Measurement Platform (PMP), uma iniciativa pan-europeia voltada ao monitoramento do desempenho fim a fim em redes nacionais de ensino e pesquisa (*National Research and Education Networks* – NRENs). Cada conjunto de dados corresponde a medições de vazão coletadas por nós perfSONAR, as quais consistem em servidores especializados distribuídos em pontos estratégicos da rede e responsáveis pela execução contínua de testes de largura de banda entre diferentes domínios. O perfSONAR (*performance Service-Oriented Network Architecture*) constitui uma infraestrutura aberta e federada de monitoramento, mantida de forma colaborativa por GÉANT, ESnet, Internet2 e diversas NRENs ao redor do mundo. Sua natureza federada favorece a interoperabilidade e a consistência metodológica das medições, possibilitando análises interdomínio baseadas em formatos homogêneos de dados e em protocolos padronizados de medição, como o `iperf3` para vazão e o `owamp` para latência e perda de pacotes.

No ecossistema GÉANT, os nós PMP atuam como pontos de referência de alto desempenho posicionados no núcleo da rede. Os conjuntos de dados empregados nesta pesquisa contemplam medições provenientes de nós como `psmp-gn-bw-lis-pt.geant.org` (Lisboa, Portugal), `psmp-gn-bw-poz-pl.geant.org` (Poznan, Polônia) e `psmp-gn-bw-vie-at.geant.org` (Viena, Áustria). Esses nós operam sobre enlaces dedicados de backbone com capacidades entre 10 e 100 Gbps e seguem uma convenção padronizada de nomenclatura que codifica sua função (por exemplo, `bw`, associado a testes de largura de banda), sua localização geográfica e o domínio organizacional. Essa padronização facilita a identificação automatizada das características dos nós, favorecendo análises topológicas e geoespaciais dos caminhos de desempenho na rede.

De forma complementar, também foram considerados conjuntos de dados provenientes de redes nacionais e regionais de ensino e pesquisa (NRENs

e RENS) interconectadas ao backbone da GÉANT. Entre esses nós, destacam-se o nó turco `perfsonar-ankara.ulakbim.gov.tr`, operado pela TÜBİTAK-ULAKBİM (ULAKNET), o nó espanhol `perfsonar-sonda.rediris.es`, mantido pela RedIRIS, o nó luxemburguês `perfsonar.restena.lu`, gerenciado pela Fondation Restena, e o nó regional catalão `psmp-anella.csuc.cat`, operado pelo CSUC no contexto da “Anella Científica”. Cada um desses elementos funciona como ponto nacional ou regional de monitoramento, ampliando a visibilidade sobre o desempenho local da rede e sua interconexão com a infraestrutura continental.

Por fim, o conjunto de dados inclui um nó em nível de campus, `psmall.lut.ac.uk`, implantado na Loughborough University (Reino Unido), representando a borda da rede, onde usuários finais e instrumentos científicos efetivamente operam. Esse nó perfSONAR de pequeno porte exemplifica a camada inferior da arquitetura hierárquica da GÉANT, com foco em diagnósticos intra-campus e metropolitanos, em contraste com os testes de longa distância realizados no backbone. Em conjunto, esses dados fornecem uma visão multinível do ecossistema europeu de redes de pesquisa, abrangendo camadas institucionais, nacionais e transnacionais. Essa organização hierárquica dos nós possibilita a análise de padrões de vazão em diferentes contextos operacionais, contribuindo para estudos voltados à monitoração baseada em metrologia de redes, ao gerenciamento de desempenho e à compreensão da variabilidade inerente a infraestruturas distribuídas.

### 3.2. Pipeline de Predição

O pipeline tem início com os dados brutos coletados da rede GÉANT, os quais são inicialmente agregados em intervalos de 6 horas, de modo a compor o conjunto de dados principal utilizado nas análises subsequentes.

Após a agregação, é realizada uma etapa abrangente de Avaliação da Qualidade dos Dados. Essa etapa compreende três análises complementares: (i) análise de periodicidade, com o objetivo de identificar padrões temporais recorrentes; (ii) análise de dados faltantes, voltada à quantificação da extensão e da localização das lacunas nas séries, que variam entre 7% e 18% nos conjuntos analisados; e (iii) análise do maior intervalo contínuo sem falhas, utilizada para determinar o maior período confiável de observação disponível.

Com base nessa avaliação, executa-se a Estratégia de Imputação de Dados. Em um primeiro momento, realiza-se uma imputação geral sobre o conjunto de dados agregado. Em seguida, visando à construção de um subconjunto de alta qualidade para treinamento dos modelos, são extraídas amostras que apresentam apenas lacunas pouco significativas, definidas neste trabalho como, no máximo, 2 pontos ausentes. Posteriormente, um método simples de imputação é aplicado para preencher essas pequenas falhas e completar as amostras selecionadas.

Na etapa de Predição com Redes Neurais, um modelo de predição é desenvolvido e otimizado de forma sistemática. O processo se inicia com um estudo comparativo para identificação da arquitetura-base de melhor desempenho. Em seguida, essa arquitetura é refinada por meio de uma busca em grade com validação cruzada (*Cross-Validation Grid Search*), com o objetivo de ajustar seus hiperparâmetros.

Para avaliar a Robustez da proposta sob diferentes níveis de disponibilidade dos

dados, são geradas quatro cópias degradadas do conjunto de amostras limpas, induzindo-se artificialmente taxas de ausência de 5%, 10%, 15% e 20%. Esse intervalo foi definido de modo a abranger as taxas de dados faltantes observadas originalmente nos conjuntos analisados, situadas entre 7% e 18%. Em seguida, a imputação é realizada em cada uma dessas versões degradadas. Paralelamente, são conduzidos estudos para determinar os melhores parâmetros dos métodos de imputação e para estabelecer configurações de referência adequadas para os modelos de predição.

Por fim, a metodologia é concluída com uma etapa de Avaliação dos Modelos, organizada em duas frentes complementares. Na primeira, os métodos de imputação são avaliados de forma sistemática quanto à sua acurácia. Na segunda, realiza-se a avaliação final da predição, com o objetivo de mensurar o desempenho global e a capacidade preditiva do pipeline completo. Esse delineamento é particularmente relevante no contexto de monitoração e operação de redes, uma vez que permite investigar, de forma integrada, como a qualidade da reconstrução dos dados impacta tarefas subsequentes de análise de desempenho, planejamento de capacidade e suporte a mecanismos de gerenciamento mais autônomos.

### 3.3. Algoritmo Proposto

Métodos tradicionais de imputação frequentemente apresentam limitações diante das dinâmicas não lineares e dos padrões complexos encontrados em séries temporais de vazão de rede, falhando, em muitos casos, em preservar as estruturas subjacentes mais relevantes dos dados. Para tratar essa lacuna, este trabalho introduz o algoritmo LARK, uma abordagem proposta para imputação de valores ausentes em séries temporais de vazão. Adicionalmente, apresenta-se o método PILA, uma variante simplificada do algoritmo principal. Ambos os métodos exploram a estrutura intrinsecamente de baixo posto presente em séries temporais, diferindo, contudo, na forma de construção da matriz de representação utilizada na análise: o LARK emprega uma matriz de trajetória com estrutura de Hankel para capturar dinâmicas temporais mais complexas, enquanto o PILA utiliza uma matriz periódica voltada à modelagem da sazonalidade. Apesar dessa distinção, a principal característica de ambos reside na capacidade de preservar a estrutura fundamental dos dados em um espaço latente e utilizar essa representação para imputar os valores ausentes.

O processo de imputação, comum a ambos os métodos, está organizado em quatro etapas sequenciais. Inicialmente, realiza-se a estimação adaptativa do período dominante por meio da Transformada Rápida de Fourier (*Fast Fourier Transform* – FFT). Em seguida, a série é transformada em uma matriz de representação, utilizando-se, conforme o método, de uma estrutura de Hankel (LARK) ou um arranjo baseado em períodos (PILA). Na terceira etapa, aplica-se a Decomposição em Valores Singulares (*Singular Value Decomposition* – SVD) para extrair a estrutura latente dessa matriz. Por fim, os valores ausentes são imputados por meio de uma reconstrução baseada em vizinhos ponderados nesse espaço de baixa dimensionalidade. Cada uma dessas etapas foi concebida para preservar as dinâmicas subjacentes da série, ao mesmo tempo em que reduz os artefatos de suavização frequentemente introduzidos por métodos tradicionais de imputação.

A extração de um espaço latente representativo por meio de métodos de decomposição, como a SVD, não pode ser realizada diretamente sobre uma série temporal univariada unidimensional. Dessa forma, torna-se necessária uma etapa prévia de transformação

da série unidimensional  $y$  em uma matriz bidimensional de representação  $M$ , assumida como portadora de uma estrutura de baixo posto.

O primeiro passo crítico dessa transformação consiste na determinação da dimensão ótima da matriz, denotada por  $P$ . Esse parâmetro deve estar alinhado ao período dominante da série temporal, de modo a permitir que a matriz represente adequadamente os padrões cíclicos subjacentes aos dados. Para identificar esse período, emprega-se a Transformada Rápida de Fourier (FFT), uma abordagem consolidada e robusta para análise espectral. Dada a série pré-processada  $y_{norm}$ , de comprimento  $N$ , a Transformada Discreta de Fourier (DFT) produz um conjunto de coeficientes complexos  $Y_k$ :

$$Y_k = \sum_{n=0}^{N-1} y_{norm}(n) \cdot e^{-i2\pi kn/N}, \quad \text{para } k = 0, \dots, N-1$$

A partir desses coeficientes, calcula-se o espectro de potência do sinal,  $S(f)$ , definido como o quadrado do módulo dos coeficientes:

$$S(f) = |Y(f)|^2$$

Em seguida, as frequências  $f$  são convertidas em períodos segundo a relação  $T = 1/f$ . O algoritmo define como período ótimo  $P$  aquele associado à maior potência espectral, respeitando restrições previamente estabelecidas ( $P_{min}, P_{max}$ ):

$$P = \arg \max_{T \in [P_{min}, P_{max}]} (S(1/T))$$

O valor obtido para  $P$  constitui o parâmetro fundamental para a construção da matriz de representação.

Uma vez que a SVD não pode ser aplicada diretamente a uma série temporal unidimensional, a série de vazão  $y$  deve ser transformada em uma matriz bidimensional  $M$ , assumida como dotada de uma estrutura latente de baixo posto. A ideia central é tornar explícita essa estrutura por meio da construção de uma matriz representativa que preserve a dinâmica temporal da série. No caso do LARK, essa matriz assume a forma de uma matriz de Hankel. O princípio subjacente é que a dinâmica temporal implícita em uma série unidimensional passa a ser explicitada quando a série é embutida em um espaço de dimensão superior, permitindo à SVD separar de forma mais eficaz os componentes dinâmicos principais do ruído.

Dada a série de entrada  $y = (y_1, y_2, \dots, y_n)$  e considerando o tamanho de janela  $L = P$ , a matriz representativa  $H$  é construída pelo deslizamento dessa janela ao longo da série. A matriz resultante possui  $K = n - L + 1$  linhas e  $L$  colunas, sendo definida por:

$$H = \begin{pmatrix} y_1 & y_2 & \dots & y_L \\ y_2 & y_3 & \dots & y_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_K & y_{K+1} & \dots & y_n \end{pmatrix}$$

Cada linha  $H_i = (y_i, \dots, y_{i+L-1})$  representa, portanto, um vetor de estado composto por  $L$  observações consecutivas.

Como alternativa, o método PILA propõe a construção de uma matriz periódica  $M$ , organizada com base no período estimado  $P$ . Nesse caso, a série é reestruturada em uma matriz de dimensão  $T \times P$ , em que  $T = \lceil n/P \rceil$  representa o número de períodos completos ou parciais. Quando  $n$  não é múltiplo de  $P$ , a última linha é completada com valores NaN. Nessa representação, cada coluna  $M_{:,j}$  reúne os pontos de dados correspondentes à  $j$ -ésima fase do período.

Após a construção da matriz bidimensional de representação  $M$  – seja ela a matriz de trajetória  $H$  ou a matriz periódica  $M$  –, a etapa central consiste em sua decomposição para identificação do espaço latente. Para isso, emprega-se a Decomposição em Valores Singulares (SVD).

Entretanto, há um requisito computacional importante: a SVD não pode ser aplicada diretamente a matrizes com valores ausentes. Assim, antes da decomposição,  $M$  é imputada provisoriamente por meio do preenchimento de cada coluna com sua mediana, uma vez que a mediana apresenta maior robustez a valores extremos, frequentemente observados em dados de redes. Esse procedimento gera uma matriz completa, denotada por  $M_{filled}$ .

A SVD de  $M_{filled}$  é formalmente dada por:

$$M_{filled} = U\Sigma V^T$$

em que  $U$  e  $V$  são matrizes ortogonais contendo, respectivamente, os vetores singulares à esquerda e à direita, e  $\Sigma$  é uma matriz diagonal composta pelos valores singulares não negativos  $s_i$ , ordenados de forma decrescente.

A extração do espaço latente é realizada a partir da seleção de um posto ótimo  $r$ , capaz de capturar a estrutura principal dos dados e, simultaneamente, filtrar o ruído. Esse posto é determinado com base em um limiar de energia acumulada, denotado por  $E_{energy}$  (por exemplo, 0,9).

Neste trabalho, adota-se uma definição de energia distinta da formulação tradicional baseada na soma dos quadrados dos valores singulares. Com base em evidências empíricas, define-se a energia como a soma direta dos valores singulares. Assim, a energia acumulada para um posto  $k$  é calculada por:

$$\text{cum\_energy}(k) = \frac{\sum_{i=1}^k s_i}{\sum_{i=1}^{\min(T,P)} s_i}$$

O posto  $r$  é então definido como o menor valor de  $k$  que satisfaz o limiar de energia:

$$r = \arg \min_k \{ \text{cum\_energy}(k) \geq E_{energy} \}$$

A decisão de utilizar a soma direta  $s_i$ , em vez de  $s_i^2$ , constitui uma escolha deliberada de projeto. Observou-se que essa formulação evita a supervalorização dos primeiros componentes dominantes, a qual pode mascarar padrões mais sutis presentes nos dados. Como consequência, obtêm-se uma seleção de posto mais equilibrada e projeções mais estáveis no espaço latente.

Determinado o posto ótimo  $r$ , constrói-se a representação no espaço latente, denotada por  $Z$ . Essa projeção é formada a partir dos primeiros  $r$  vetores singulares à esquerda, ponderados pelos respectivos valores singulares, os quais expressam a relevância de cada componente:

$$Z = U_{:,1:r} \cdot \Sigma_{1:r,1:r}$$

Cada linha  $Z_i$  dessa nova matriz  $Z$  atua como um vetor-resumo de baixa dimensionalidade associado à linha correspondente  $M_i$  da matriz de representação, preservando seus padrões estruturais essenciais.

A imputação propriamente dita é então realizada por meio de uma abordagem de *K-Nearest Neighbors* (KNN) *no espaço latente*. Esse aspecto é central à proposta: ao calcular as distâncias em  $Z$ , em vez de operar diretamente sobre a matriz original  $M$ , a busca por vizinhos passa a considerar similaridade estrutural – isto é, linhas com padrões temporais ou sazonais semelhantes –, em vez de depender apenas de similaridades pontuais e potencialmente ruidosas.

O algoritmo percorre a matriz original  $M$  em busca de cada valor ausente  $M(i, j)$ . Para cada elemento faltante, recupera-se inicialmente o vetor-resumo  $Z_i$  correspondente à linha-alvo  $i$ . Em seguida, identifica-se o conjunto de linhas candidatas  $\mathcal{C}$  para as quais o valor na mesma coluna  $j$  é observado, ou seja,  $M(c, j) \neq \text{NaN}$ . A distância Euclidiana entre o vetor-alvo  $Z_i$  e cada vetor candidato  $Z_c$ , com  $c \in \mathcal{C}$ , é então calculada no espaço latente:

$$d(Z_i, Z_c) = \|Z_i - Z_c\|_2 + \epsilon$$

em que  $\epsilon$  é uma constante pequena, por exemplo,  $10^{-8}$ , empregada para evitar divisão por zero. Os  $k$  candidatos com menores distâncias são selecionados para compor o conjunto de vizinhos  $\mathcal{N}_k \subseteq \mathcal{C}$ .

Por fim, o valor ausente  $M(i, j)$  é imputado por meio de uma média ponderada pelo inverso da distância (*Inverse Distance Weighted* – IDW), de modo a atribuir maior influência aos vizinhos estruturalmente mais próximos no espaço latente:

$$M_{\text{imputed}}(i, j) = \frac{\sum_{c \in \mathcal{N}_k} w_c \cdot M(c, j)}{\sum_{c \in \mathcal{N}_k} w_c}$$

$$\text{em que } w_c = \frac{1}{d(Z_i, Z_c)}$$

Esse procedimento é repetido para todos os valores ausentes, resultando em uma matriz de representação completamente imputada, denotada por  $M_{\text{imputed}}$ .

#### 4. Experimentos

Esta seção apresenta uma avaliação abrangente do algoritmo proposto LARK em comparação com sua variante simplificada, PILA, e com diferentes técnicas consolidadas de imputação. A análise experimental foi estruturada em duas frentes complementares: (i) comparação direta da acurácia de imputação por meio de métricas clássicas de erro, como RMSE e MAE, sob diferentes níveis de esparsidade dos dados; e, (ii) avaliação da utilidade *downstream* dos dados imputados em tarefas de predição, utilizando modelos recorrentes GRU e LSTM.

Os resultados são analisados com medidas de erro: Root Mean Square Error (RMSE) e Mean Absolute Error (MAE). O uso de múltiplas métricas fornece uma visão mais abrangente do desempenho, como recomendado em estudos recentes [?].

Inicialmente, a qualidade de imputação foi analisada a partir de cenários com 5%, 10%, 15% e 20% de dados ausentes. Os resultados obtidos são apresentados na Tabela 1, permitindo observar não apenas o erro médio das técnicas, mas também sua estabilidade em diferentes níveis de degradação dos dados. Esse tipo de avaliação é particularmente relevante em contextos de monitoração de redes, nos quais falhas de medição e lacunas temporais podem comprometer atividades analíticas posteriores, como predição de desempenho, suporte a SLA e planejamento de capacidade.

Tabela 1. Resultados de Imputação

Técnica	5%		10%		15%		20%	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
LARK	6.73	1.08	<b>13.44</b>	<b>3.23</b>	<b>16.55</b>	<b>4.91</b>	<b>20.12</b>	<b>6.79</b>
	±4.62	±0.74	±8.29	±2.00	±8.47	±2.29	±10.20	±3.20
PILA	7.16	1.15	15.43	3.62	18.03	5.15	23.59	8.12
	±5.64	±0.84	±8.17	±2.04	±7.93	±2.18	±13.55	±4.76
HoltWinters	<b>5.47</b>	<b>0.93</b>	16.62	3.86	19.71	5.68	23.83	8.23
	±4.86	±0.84	±10.43	±2.47	±10.64	±2.98	±12.47	±4.32
ARIMA	6.26	1.02	17.84	4.26	22.36	6.55	25.74	9.15
	±4.52	±0.77	±11.96	±3.02	±11.54	±3.35	±12.63	±4.70
Kalman	8.17	1.28	163.00	30.17	173.79	38.17	169.94	38.87
	±4.77	±0.81	±159.55	±37.56	±164.95	±41.62	±136.57	±30.61
Linear	7.98	1.28	16.83	3.96	21.45	6.23	23.84	8.11
	±5.35	±0.88	±10.60	±2.56	±10.22	±2.78	±11.08	±3.69
Spline	10.51	1.64	21.74	5.14	31.91	8.77	36.87	12.15
	±9.08	±1.42	±13.95	±3.26	±17.17	±4.60	±20.42	±6.72

Em seguida, investigou-se a utilidade prática da imputação no contexto de predição de séries temporais de vazão. Para isso, modelos baseados em *Gated Recurrent Unit* (GRU) e *Long Short-Term Memory* (LSTM) foram treinados sobre os dados reconstruídos por cada técnica de imputação. Os resultados dessa etapa são apresentados na Tabela 2. Diferentemente de uma avaliação restrita à qualidade pontual da reconstrução, essa análise permite verificar se os valores imputados preservam de fato a estrutura temporal relevante para tarefas de inteligência de rede.

Os resultados experimentais apresentados evidenciam a superioridade do método proposto em termos de acurácia, onde o LARK obteve os menores valores de erro, destacando-se sobretudo à medida que o nível de esparsidade dos dados aumenta. Embora em 5% de dados ausentes o método Holt-Winters tenha apresentado o menor erro, o LARK manteve desempenho similar. Nos cenários mais desafiadores, entretanto, sua vantagem torna-se mais expressiva: em 10%, 15% e 20% de ausência, o método proposto alcançou os melhores resultados tanto em RMSE quanto em MAE.

Esse comportamento é particularmente relevante, pois demonstra que o LARK

Tabela 2. Resultados de Predição.

Técnica de Imputação	GRU		LSTM	
	RMSE	MAE	RMSE	MAE
LARK	<b>66.52</b> $\pm 23.82$	37.53 $\pm 16.59$	<b>66.85</b> $\pm 24.91$	37.68 $\pm 16.90$
PILA	66.66 $\pm 25.24$	<b>36.68</b> $\pm 17.72$	66.88 $\pm 25.72$	<b>37.15</b> $\pm 18.21$
Mean	66.78 $\pm 25.20$	36.94 $\pm 17.89$	67.88 $\pm 26.42$	38.42 $\pm 19.92$
ARIMA	69.67 $\pm 28.05$	41.20 $\pm 20.92$	71.83 $\pm 31.12$	42.95 $\pm 23.27$
HoltWinters	71.06 $\pm 28.78$	42.68 $\pm 20.44$	73.04 $\pm 32.06$	44.71 $\pm 23.16$
Linear	71.89 $\pm 27.99$	45.26 $\pm 22.65$	75.19 $\pm 31.97$	47.75 $\pm 25.03$
Spline	97.23 $\pm 36.33$	68.10 $\pm 32.42$	108.40 $\pm 43.05$	75.28 $\pm 35.78$
Kalman	105.78 $\pm 89.10$	68.55 $\pm 54.27$	111.65 $\pm 104.73$	72.31 $\pm 62.85$

não apenas apresenta boa acurácia em cenários favoráveis, mas também mantém desempenho robusto à medida que a degradação dos dados se intensifica. Em 20% de dados ausentes, por exemplo, o método superou a variante PILA com margem clara, além de apresentar menor variabilidade, o que indica maior estabilidade. Em contraste, métodos tradicionais como Kalman exibiram degradação severa a partir de 10% de ausência, alcançando erros excessivamente elevados e, portanto, pouco adequados para uso em pipelines analíticos confiáveis. De forma semelhante, abordagens como spline e interpolação linear mostraram perda progressiva de desempenho, sobretudo em cenários com maiores lacunas.

Os resultados de predição reforçam esse diagnóstico. Quando os dados imputados foram utilizados para treinar modelos recorrentes, o LARK produziu o menor RMSE tanto para GRU quanto para LSTM, evidenciando que sua reconstrução preserva de maneira mais efetiva a informação temporal necessária à tarefa preditiva. No caso do modelo GRU, o LARK atingiu RMSE de 66.52, superando métodos estatísticos amplamente utilizados, como ARIMA e Holt-Winters, bem como técnicas simples de preenchimento. Para LSTM, o comportamento foi semelhante, com o LARK novamente apresentando o melhor desempenho global em RMSE. Esse resultado é particularmente importante, pois mostra que a contribuição do LARK não se restringe à redução do erro de imputação em si. Em aplicações reais de gerenciamento e monitoração de redes, o valor de uma técnica de imputação está diretamente relacionado à sua capacidade de preservar padrões úteis para tarefas posteriores, como previsão de desempenho, identificação de degradações, análise de QoS e suporte à tomada de decisão operacional. Sob essa perspectiva, os resultados confirmam que o LARK reconstrói os dados de forma mais aderente à dinâmica real da série, possibilitando melhor desempenho dos modelos preditivos subsequentes.

## 5. Conclusão

Este trabalho apresentou uma nova abordagem para reconstrução de valores ausentes em séries temporais de vazão de rede, baseada na combinação entre análise espectral, incorporação matricial e interpolação em espaço latente. Ao explorar explicitamente a estrutura de baixo posto presente nos dados temporais, o método mostrou-se capaz de preservar padrões relevantes da série mesmo em cenários com níveis elevados de esparsidade.

Os resultados experimentais demonstraram, de forma consistente, que a proposta supera métodos tradicionais de imputação e sua variante simplificada em diferentes pers-

pectivas de avaliação. Em termos de acurácia direta, a proposta apresentou os menores erros na maior parte dos cenários analisados, com destaque para os casos de maior ausência de dados. Do ponto de vista prático, sua principal contribuição foi confirmada na etapa de predição *downstream*, em que os dados imputados pelo LARK permitiram aos modelos GRU e LSTM alcançar o melhor desempenho global. Esse achado reforça que a proposta não apenas preenche lacunas, mas preserva a dinâmica temporal necessária para tarefas posteriores de inteligência e gerenciamento de rede.

Como trabalho futuro, pretende-se estender a proposta para cenários multivariados, incorporando dependências entre diferentes métricas de desempenho, como vazão, latência, jitter e perda de pacotes. Essa extensão pode permitir uma modelagem mais abrangente do comportamento da rede, aproximando a solução de aplicações práticas em sistemas de gerenciamento inteligente.

### Agradecimentos

Os autores gostariam de agradecer ao CNPq (*N<sup>o</sup> 305946/2025-0 e N<sup>o</sup> 405940/2022-0*) e CAPES (*N<sup>o</sup> 88887.954253/2024-00 e N<sup>o</sup> 88887.972043/2024-00*) pelo apoio financeiro.

### Referências

- Adedayo, A. O. and Twala, B. (2017). Qos functionality in software defined network. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 693–699.
- Alaya, B., Khan, R., Moulahi, T., and el Khediri, S. (2021). Study on qos management for video streaming in vehicular ad hoc network (vanet). *Wireless Personal Communications*, 118:1–33.
- Alkenani, J. and Nassar, K. (2022). Network monitoring measurements for quality of service: A review. *Iraqi Journal for Electrical and Electronic Engineering*, 18:33–42.
- Arquam, M. and Kumari, S. (2025). *Maximization of Communication Network Throughput Using Dynamic Traffic Allocation Scheme*, page 116–127. Springer Nature Singapore.
- Brito, M. L. L., Ferreira, M. C. M., Portela, A. L. C., and Gomes, R. L. (2025). Ai-based estimation of bandwidth availability for data offloading in edge-cloud computing. *IEEE Networking Letters*, pages 1–1.
- Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*, pages 1–7.
- Ghahramani, M. H., Zhou, M., and Hon, C. T. (2017). Toward cloud computing qos architecture: analysis of cloud systems and cloud services. *IEEE/CAA Journal of Automatica Sinica*, 4(1):6–18.
- Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. (2014). A bandwidth-feasibility algorithm for reliable virtual network allocation. In *2014 IEEE 28th International Conference on Advanced Information Networking and Applications*, pages 504–511.

- Gomes, R. L., Bittencourt, L. F., Madeira, E. R., Cerqueira, E., and Gerla, M. (2016). Bandwidth-aware allocation of resilient virtual software defined networks. *Computer Networks*, 100:179–194.
- Ju, H. and Zhang, R. (2014). Throughput maximization in wireless powered communication networks. *IEEE Transactions on Wireless Communications*, 13(1):418–428.
- Khattab, A. A. R., Elshennawy, N. M., and Fahmy, M. (2023). Gma: Gap imputing algorithm for time series missing values. *Journal of Electrical Systems and Information Technology*, 10(1):41.
- Khayati, M., Lerner, A., Tymchenko, Z., and Cudré-Mauroux, P. (2020). Mind the gap: an experimental evaluation of imputation of missing values techniques in time series. *Proc. VLDB Endow.*, 13(5):768–782.
- Liu, X., Xu, B., Zheng, K., and Zheng, H. (2023). Throughput maximization of wireless-powered communication network with mobile access points. *IEEE Transactions on Wireless Communications*, 22(7):4401–4415.
- Liu, X., Yu, Y., Li, F., and Durrani, T. S. (2022). Throughput maximization for ris-uav relaying communications. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):19569–19574.
- Lopes Gomes, R. and Roberto Mauro Madeira, E. (2012). A traffic classification agent for virtual networks based on qos classes. *IEEE Latin America Transactions*, 10(3):1734–1741.
- Minovski, D., Ögren, N., Mitra, K., and Åhlund, C. (2023). Throughput prediction using machine learning in lte and 5g networks. *IEEE Transactions on Mobile Computing*, 22(3):1825–1840.
- Niako, N., Melgarejo, J. D., Maestre, G. E., and Vatcheva, K. P. (2024). Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with arima and lstm. *BMC Medical Research Methodology*, 24(1):320.
- Nobre, F. V. J., Silva, D. d. S., Ferreira, M. C. M. M., Brito, M. L. M. L., de Araújo, T. P., and Gomes, R. L. (2025). Time-weighted correlation approach to identify high delay links in internet service providers. *Journal of Internet Services and Applications*, 16(1):419–430.
- Sharma, A., Pandit, S., and Talluri, S. R. (2025). Throughput prediction of fifth-generation cellular system using hybrid feature selection and enhanced sequential decision tree machine learning algorithm. *Wireless Networks*, 31(3):3025–3042.
- Souza, M. S., Ribeiro, S. E. S. B., Lima, V. C., Cardoso, F. J., and Gomes, R. L. (2024). Combining regular expressions and machine learning for sql injection detection in urban computing. *Journal of Internet Services and Applications*, 15(1):103–111.
- Younas, M. I., Iqbal, M. J., Aziz, A., and Sodhro, A. H. (2023). Toward qos monitoring in iot edge devices driven healthcare—a systematic literature review. *Sensors*, 23(21).
- Zheng, Z., Wu, X., Zhang, Y., Lyu, M. R., and Wang, J. (2013). Qos ranking prediction for cloud services. *IEEE Transactions on Parallel and Distributed Systems*, 24(6):1213–1222.