

O Impacto da Pandemia do COVID-19 no Comportamento do Tráfego de Rede e no Processo de Predição

Rafael A. Menezes¹, Dyego H. L. Oliveira¹, Rafael L. Gomes¹

¹CCT – Universidade Estadual do Ceará (UECE)
Fortaleza – CE – Brasil

{menezes.almeida, dyego.leonel, rafael.l.gom}@larces.uece.br

Abstract. *The COVID-19 pandemic brought with it numerous changes, from the economy to social behavior. This new reality caused a profound impact on the network traffic of universities, companies, institutions, etc. Similarly, the traffic prediction process was also impacted, since the behavior of elastic demand for resources was indirectly affected. Within this context, this work presents an experimental analysis of the network traffic behavior, as well as the resource demand prediction of real datasets in the pre-pandemic period and during the pandemic. The results of the experiments carried out show the importance of stationary data identification in order to adjust the time series, minimizing prediction errors.*

Resumo. *A pandemia do COVID-19 trouxe consigo inúmeras mudanças, desde a economia até os comportamentos sociais. Esta realidade impactou diretamente no tráfego de rede das universidades, empresas, instituições, etc. Similarmente, o processo de predição de tráfego foi impactado também, visto que o comportamento da demanda elástica por recursos foi indiretamente afetado. Dentro deste contexto, este trabalho apresenta uma análise experimental sobre o comportamento do tráfego de rede, bem como a predição por demanda de recursos de conjuntos de dados reais no período Pré-Pandemia e durante a Pandemia. O resultado dos experimentos realizados sugerem que a identificação da estacionariedade dos dados é crucial para ajustar a série temporal analisada a fim de minimizarmos os erros de predição.*

1. Introdução

A Pandemia do COVID-19 provocou mudanças significativas em diversos cenários e em vários locais do mundo, demandando adaptações em diversas áreas, tal como a educação. A Universidade Estadual do Ceará (UECE), por exemplo, devido ao decreto estadual, paralisou suas atividades presenciais no mês de março de 2020. Desde então, o ensino se tornou remoto e o trabalho presencial dos servidores técnicos ocorre de forma escalonada por setores seguindo um percentual de ocupação orientado ao nível de alerta determinado. Semelhante abordagem ocorre nas demais universidades, empresas, instituições e órgãos governamentais por todo o Brasil e o Mundo [Grossi et al. 2020].

Estas alterações mudaram o comportamento social e de mobilidade existentes dentro da universidade, afetando a utilização e, conseqüentemente, a demanda de recursos físicos (água, luz, etc) e computacionais desta de forma direta. Por conta disso, o tráfego de dados da UECE sofreu uma grande transformação. A paralisação resultou em uma

queda expressiva de 82% na média do tráfego de dados, bem como alterou a demanda de uso ao longo do dia e dias da semana.

Concomitantemente, a predição de uso de tráfego de rede se faz importante, pois possibilita um fornecimento de internet personalizado (baseado no uso), garantindo uma boa experiência do usuário e uso adequado dos recursos. Contudo, técnicas de predição de tráfego comuns, não são eficientes no contexto de demanda elástica por recursos de rede ao longo do dia, enfrentados pelos provedores de serviço de internet (ISPs), pois a demanda não segue um comportamento padrão bem definido, ou seja, não segue uma série temporal bem definida. Recentemente, foi proposto por Oliveira et al. [Oliveira 2020] o modelo de Predição de Rede Adaptável (PRA), o qual decompõe os dados originais da série, testa a estacionariedade, para, por fim, corrigir os erros de ciclo. No entanto, ainda não se tem um entendimento do comportamento da demanda elástica no contexto da Pandemia do COVID-19.

Dentro deste contexto, este artigo apresenta uma análise do comportamento do tráfego de rede, bem como a predição por demanda de recursos de conjuntos de dados reais coletados na UECE, destacando os mecanismos necessários para aprimorar o processo de predição em cada cenário. Assim, este trabalho possui as seguintes contribuições: (I) Análise experimental comportamental dos conjuntos de dados de demanda por recursos de rede real Pré-Pandemia (Março a Agosto de 2019) e durante Pandemia (Abril a Setembro de 2020)¹; e, (II) Análise da capacidade de predição das técnicas existentes em relação a cada um dos conjuntos de dados.

2. Análise do Comportamento Pré-Pandemia e durante a Pandemia

O processo de predição é, de forma geral, dividido em três partes: coleta de dados necessários para predição, aplicação das técnicas de predição e a predição em si. Porém, adicionalmente, aplicaremos o modelo PRA neste processo, com o objetivo de garantir um ajuste nos dados a fim de adequar a análise ao contexto de demanda elástica.

Antes de tudo, é importante ressaltar alguns padrões encontrados em séries temporais: (1) Tendência (T), essa componente é vista quando a série temporal seguir alguma direção, crescente, decrescente e não exatamente linear; (2) Sazonalidade (S), padrão que se repete com uma certa frequência dentro de um período; e, (3) Ciclo (C), diferente da sazonalidade, este padrão se repete, mas não obedece uma periodicidade fixa.

Tendo conhecimento desses padrões, também é válido ressaltar que uma série temporal pode ser representada de forma aditiva, ou seja, sendo Z_t uma série temporal composta pela combinação de componentes de tendência, sazonalidade e ciclo, o modelo aditivo pode ser representado da forma $Z_t = T_t + S_t + E_t$, onde Z_t é o dado observado no período t , S_t é a componente sazonal do período t e E_t representa a parte não capturada do modelo no período t , chamada de erro (ou resíduo) [Pedro Guilherme Costa Ferreira 2017].

De volta ao processo de predição, o passo inicial foi a coleta de informações, que refere-se a um conjunto de dados que representam a utilização da largura de banda da UECE, coletados de hora em hora. De posse disto, constitui-se uma série temporal que deverá ser analisada e corrigida pelo modelo. Como o modelo sugere, a primeira fase

¹<https://github.com/rafaelalmeida2909/wpeif2021>

é a decomposição dos dados, que tem por objetivo extrair as componentes da série. As componentes são: sazonalidade (S_t), tendência (T_t) e Erro (E_t), estas são obtidas através das médias móveis.

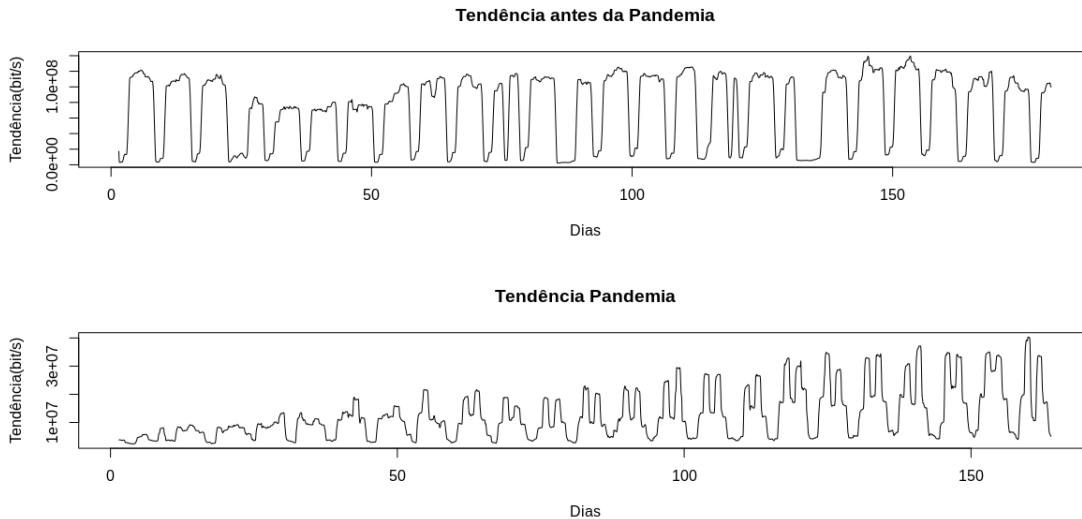


Figura 1. Tendência dos conjuntos de dados analisados

A Figura 1 ilustra a tendência em cada conjunto de dados. Na tendência da série referente ao período pré-pandêmico, percebemos que a média do tráfego obedece uma certa constância e possui quedas periódicas, já no caso da tendência do período pandêmico, notamos a componente crescente, mas também percebemos quedas periódicas. Essas quedas são os fins de semana, visto que o movimento na universidade nestes dias, em geral é bem menor, se comparado com os dias úteis. E a tendência crescente no período pandêmico, pode ser explicada pelo início em que os dados foram coletados, ou seja, no mês de abril (Ocorrência do *Lockdown*) até Setembro de 2020 (Faixa de alerta mais baixa do período de reabertura).

Portanto, damos início a segunda parte do modelo utilizado, faremos a remoção de erros de ciclo. Sabemos que devido a demanda elástica, a média de uso da largura de banda em dias regulares atingem valores bem diferentes dos dias considerados irregulares, como fins de semana e feriados. As técnicas de predição, em geral, não são capazes de identificar estas características, o que interfere de forma negativa nos resultados.

Dessa forma, os fins de semana serão removidos e adicionalmente um desenvolvido modelo de decomposição sazonal e de tendência, utilizando Loess [Dokumentov and Hyndman 2015], será aplicado às séries para remover valores que diferenciam drasticamente de todos os outros e estimar valores ocultos. A Figura 2 mostra a correção das séries de cada conjunto de dados, onde é possível perceber que os conjuntos de dados ficaram mais concisos, ou seja, grande parte dos ciclos irregulares foram removidos durante o processo do modelo PRA.

Após esta etapa dá-se início a terceira e última etapa do modelo, a análise de estacionariedade das séries. Os testes de estacionariedade de uma série temporal, tem por objetivo, observar de forma futura seu comportamento. Uma série não estacionária contém mudanças bruscas de inclinação em determinados períodos. Esta característica influencia de forma negativa na qualidade de predição das técnicas conhecidas. Consequentemente,

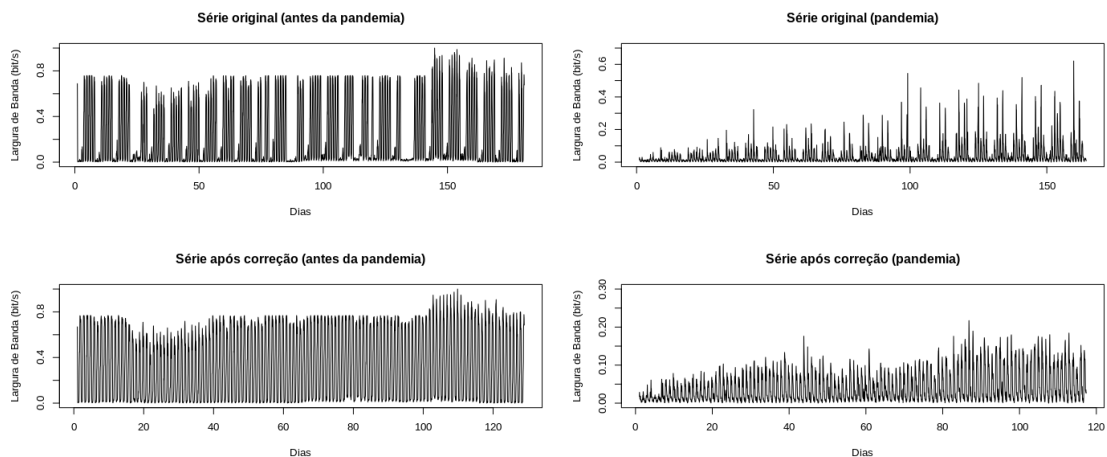


Figura 2. Correção das séries a partir da remoção de erros de ciclos

é importante verificar a estacionariedade e possivelmente tratá-la.

Um processo estocástico é estacionário se sua média e autocovariância (ou autocorrelação) são constantes ao longo do tempo. Isso significa que a série não contém tendências aparentes e que sua variação e o padrão dessa variação são constantes no tempo [Pedro Guilherme Costa Ferreira 2017]. Uma ferramenta que auxilia esta verificação é a função de autocorrelação (FAC). Segundo Enders [Enders 2015], um processo não estacionário apresenta um lento decaimento em sua FAC. Sabendo disto, o primeiro teste será a geração do gráfico das FACs das séries em questão, os quais são ilustrados na Figura 3.

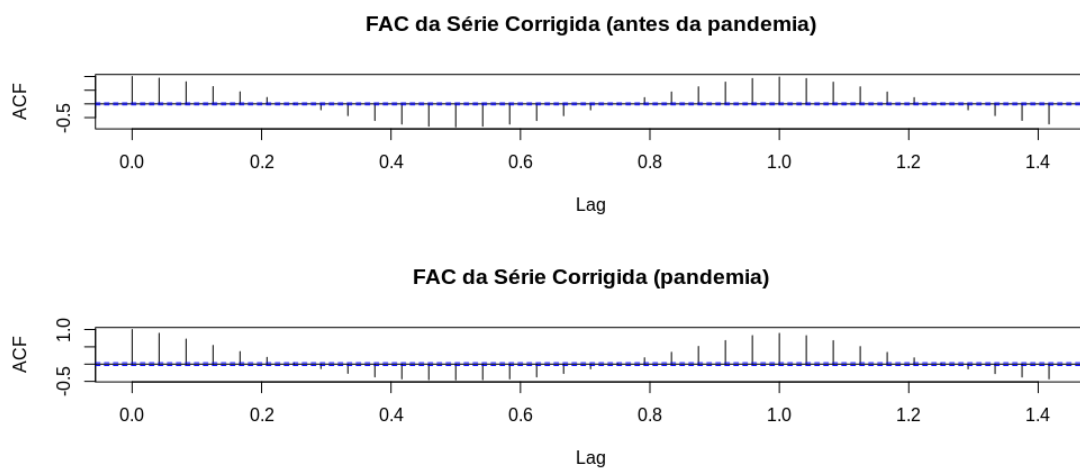


Figura 3. FACs das Séries

Tendo conhecimento do que foi dito, nota-se que os gráficos sugerem que as duas séries são estacionárias, visto que as duas apresentam quedas e subidas rápidas, características de funções senoidais. Entretanto, é importante continuar os testes para que a suspeita seja confirmada. O próximo teste será da constância da média e variância, ao longo das séries.

Vemos na Figura 4 que no caso da série temporal do período pré-pandêmico os dois testes indicaram estacionariedade, mas já no período pandêmico a não estacionariedade da série foi percebida. O teste final será o teste de Kwiatkowski-Phillips-Schmidt-

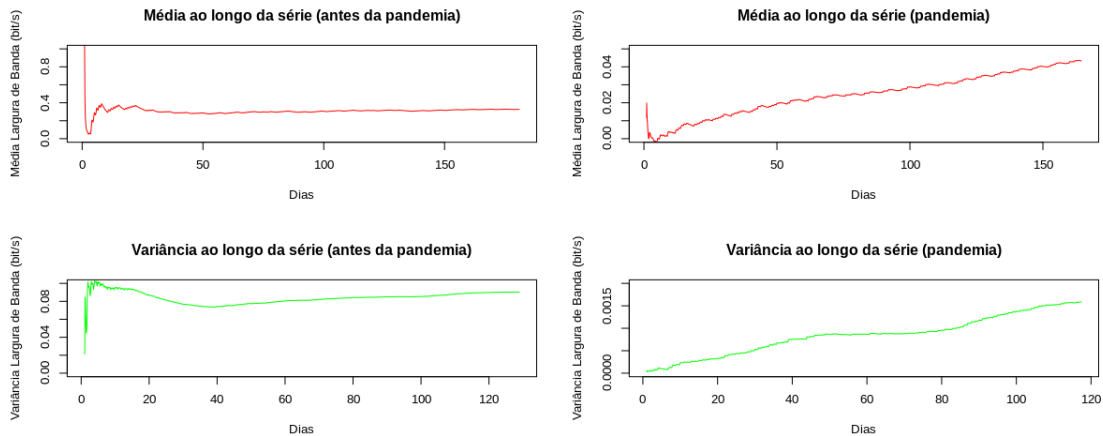


Figura 4. Médias e Variâncias das séries

Shin (KPSS) [i Silvestre et al. 2001] que verifica a estacionariedade dos dados. Os testes estatísticos KPSS confirmaram a estacionariedade da série pré-pandemia e a não estacionariedade da série pandêmica. A série pandêmica deve passar por transformações que garantam sua estacionariedade. Para resolvermos o problema da variância nos basearemos na teoria de Box&Jenkins, que diz que uma série temporal também deve ser estacionária na variância. Então, aplicaremos uma transformação logarítmica em toda a série. Já a média, pode ser resolvida através da diferenciação da série que é representada pela seguinte equação, onde t é um dado qualquer e Z é uma nova série criada a partir da série original Y (eliminando tendências indesejáveis): $Z_t = Y_t - Y_{t-1}$.

Após estas transformações, os testes foram refeitos e a série pandêmica demonstrou ser estacionária, assim como a pré-pandêmica. Dessa forma, finaliza-se o processo do modelo PRA e pode-se aplicar as técnicas de predição desejáveis. As técnicas de predição utilizadas foram clássicas, utilizamos o modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA), Rede Neural Auto-Regressiva e o modelo de suavização exponencial Holt Winters.

3. Resultados de Predição

Para avaliarmos a eficiência dos processos anteriormente aplicados, usaremos o Erro Quadrático Médio Raiz (RMSE) por um período T , seguindo a Equação 1, onde \hat{y}_t é o valor estimado e y_t é o valor real do uso da largura de banda no momento t . Os resultados são apresentados nas Tabelas 1 e 2.

$$\text{RMSE}(T) = \frac{1}{\sqrt{T}} \left(\sum_{t=1}^T (\hat{y}_t - y_t)^2 \right)^{\frac{1}{2}}. \quad (1)$$

A partir dos dados presentes na Tabela 1 é possível perceber que a aplicação do modelo PRA traz uma queda de 70% nos erros, em média, para dados estacionários. Similarmente, na Tabela 2 percebemos que as quedas nos erros foram semelhantes ao conjunto de dados pré-pandêmicos. Desta forma, nota-se que caso não fossem feitos os processos para tornar a série pandêmica estacionária, as quedas seriam bem inferiores em alguns modelos. De forma geral, tem-se que no modelo ARIMA a queda seria de

Tabela 1. RMSE do Conjunto de Dados Pré-Pandêmico

| Técnicas | ARIMA | NEURAL | HOLT |
|---------------|--------------|--------------|--------------|
| Média | 56774469 | 53486512 | 80656458 |
| Média com PRA | 14174426 | 14173623 | 23050529 |
| Resultado | Queda de 75% | Queda de 73% | Queda de 71% |

Tabela 2. RMSE do Conjunto de Dados Pandêmico

| Técnicas | ARIMA | NEURAL | HOLT |
|---------------|--------------|--------------|--------------|
| Média | 13135101 | 7251172 | 11507233 |
| Média com PRA | 3911780 | 4107046 | 4546957 |
| Resultado | Queda de 70% | Queda de 43% | Queda de 60% |

aproximadamente 47%, no modelo Neural a queda representaria 10% e no Holt Winters 19%.

4. Conclusão

A partir da análise experimental realizada neste trabalho, concluímos que apesar da eficácia do modelo PRA, é importante trabalharmos sempre com dados estacionários para mantermos uma média eficiente de predição. Portanto, faz-se necessário analisar a estacionariedade dos dados da série temporal e aplicar as transformações necessárias a cada caso, possibilitando a correção mais adequada possível para o processo de predição.

Como trabalhos futuros, iremos, a partir da análise experimental realizada, construir um modelo de predição que dinamicamente identifica séries estacionárias e não estacionárias, executando as ações de adequação necessárias para cada caso e possibilitando maximizar a capacidade de predição.

Referências

- Dokumentov, A. and Hyndman, R. J. (2015). STR: A Seasonal-Trend Decomposition Procedure Based on Regression. Monash Econometrics and Business Statistics Working Papers 13/15, Monash University, Department of Econometrics and Business Statistics.
- Enders, W. (2015). *Applied econometric time series*. Wiley, 4 edition.
- Grossi, M. G. R., Minoda, D. d. S. M., and FONSECA, R. G. P. (2020). Impacto da pandemia do covid-19 na educação: Reflexos na vida das famílias. *Teoria e Prática da Educação*, 23(3):150–170.
- i Silvestre, J. L. C., i Rosselló, A. S., and Ortuño, M. A. (2001). Unit root and stationarity tests' wedding. *Economics Letters*, 70(1):1–8.
- Oliveira, D. H. L. (2020). Modelo adaptativo para previsão de demanda por recursos de rede em provedores de internet modernos.
- Pedro Guilherme Costa Ferreira, Anna Carolina Barros, D. M. d. M. I. C. L. d. O. e. V. E. L. d. A. D. (2017). *Análise de séries temporais em R: curso introdutório*. Addison-Wesley, 1ª edition.