

Predição de Resiliência Econômica a partir de Métodos Supervisionados baseados em Grafos Dinâmicos

Marcus Araújo¹, Francisco A. Rodrigues¹, Elaine P. Sousa¹

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (ICMC)

marcus.santos.araujo@usp.br, {francisco, parros}@icmc.usp.br

Abstract. *The development of economic resilience has been a critical discussion for the last five years of the World Economic Forum. Moreover, since the Great Recession (2008), several articles focused on defining, measuring, and exploring factors that could differentiate economies that overcome rapidly severe interruptions from those that couldn't. This paper uses the United Nations' massive historical datasets to compare supervised methods under independent and dynamic graph-based representations, weighing the pros and cons of an approach that considers neighborhood information. The results report a 19% gain (F1-Score) in crisis and stability prediction.*

Resumo. *O desenvolvimento de resiliência econômica é uma das discussões priorizadas nas 5 últimas ocorrências do Fórum Econômico Mundial. Desde a Grande Recessão (2008), diversos trabalhos se dedicaram a definir, mensurar e recuperar fatores que expliquem a diferença entre economias que superaram desastres econômicos rapidamente, daquelas que demoram ou precisam de aporte externo. O presente trabalho utiliza bases históricas massivas da ONU para treinar diferentes modelos supervisionados sobre a economia de dezenas de países, apresentando as vantagens de utilizar uma modelagem que considere informações da vizinhança por meio de um grafo dinâmico. Resultados mostram um ganho de até 19% (F1-Score) na previsão de crise e estabilidade.¹*

1. Introdução

A resiliência é um "termo guarda chuva que expressa os fundamentos conceituais que permitem a sobrevivência de um determinado sistema (econômico) sob condições adversas"[Mai and Chan 2020]. Em particular, neste trabalho é considerada a variação comportamental da economia regional em relação aos choques econômicos.

Trabalhos ressaltam que ao se tratar de resiliência econômica, trata-se de um "sistema adaptativo altamente complexo, não linear e dependente dos estados anteriores"[Bristow and Healy 2018]. Sob essa perspectiva, existem bases teóricas em comum entre a Resiliência Econômica e a Resiliência em Sistemas Dinâmicos utilizadas na Engenharia e Ecologia [Punzo et al. 2020, Resbeut 2021]. E, mais recentemente, houve a expansão da definição formal de sistema dinâmico resiliente originalmente proposta por [Holling 1973] para conseguir lidar com múltiplos agentes [Gao et al. 2016]. Baseado nessa expansão, este trabalho explora a previsibilidade da resiliência econômica, diferenciando-se pelo uso de medidas de resiliência e bases de dados publicadas por órgãos econômicos mundiais (Banco Mundial e Organização das Nações Unidas).

¹Agradecimentos à CAPES e ao CNPq pelo apoio financeiro

O presente trabalho se organiza da seguinte maneira: a Seção 2 introduz conceitos relevantes para o aprendizado supervisionado de resiliência econômica por meio de grafos dinâmicos; a Seção 3 apresenta a proposta deste trabalho; a Seção 4 apresenta os resultados obtidos; e a Seção 5 apresenta as conclusões e trabalhos futuros.

2. Resiliência Econômica e Grafos Dinâmicos

O modelo de identificação e mensuração proposto por órgãos econômicos possuem uma relevância significativa [Hallegatte 2014], pois costumam considerar indicadores econômicos comuns a todos os países do globo e embasam diversos trabalhos recentes [Salignac et al. 2022, Yildirim et al. 2022]. A identificação da falha na sobrevivência do sistema econômico pode ser observada pela perda do potencial de produtividade (capital) da nação (medida usualmente pelo Produto Interno Bruto, PIB). A resiliência é constatada pelo retorno da produtividade aos níveis anteriores à crise (Figura 1).

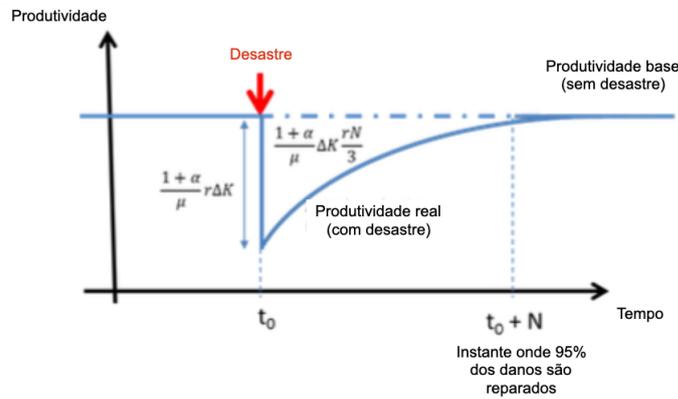


Figura 1. Recuperação de economia resiliente [Hallegatte 2014].

Fundamentalmente, o modelo do Banco Mundial [Hallegatte 2014] reduz a dinâmica de uma economia resiliente a dois estados: **Crise** e **Estabilidade**. Para mensurar a severidade do impacto, o modelo assume que a perda de produtividade seja reduzida a zero exponencialmente, e que 95% das perdas sejam reparadas em N anos, então o impacto entre o desastre e a recuperação ($\Delta Y(t)$, $t \in [t_0, t_0 + N]$, $\Delta Y \in R$) é definido pela eq. 1 e a perda total pela eq. 2. A mensuração da perda imediata $\Delta Y(t_0)$ de capital Y parte dos indicadores econômicos de crescimento ($1 + \alpha | \alpha \in R$), da produtividade marginal do capital ($r | r \in R$), da redução da produção ($\mu | \mu \in R$) e da redução do valor do capital ($\Delta K | \Delta K \in R$). O fator de reparação anual $3/N$ é interpretado como "tempo característico", e seu valor é derivado da taxa de recuperação final em tendência logarítmica (em 95%), onde $\exp(-3) \approx 0.05$, e, portanto, $\Delta Y(t_0 + N) = 0.05\Delta Y(t_0)$.

$$\Delta Y(t) = \frac{1 + \alpha}{\mu} r \Delta K e^{-\frac{3(t-t_0)}{N}} \tag{1}$$

$$\Delta Y = \int_{t_0}^{+\infty} \frac{1 + \alpha}{\mu} r \Delta K e^{-\frac{3(t-t_0)}{N}} dt \tag{2}$$

Essa abordagem, no entanto, se limita à mensuração de um único agente (país). Considerar múltiplos agentes e suas relações aumenta a complexidade e exige uma

formulação matemática mais sofisticada. Nesse sentido, avanços em áreas conceitual-mente correlacionadas apresentam trabalhos recentes que expandem a representação matemática primordial (em sistemas dinâmicos), a fim de considerar o caso com múltiplos agentes [Gao et al. 2016]. Para isso, utilizam um modelo matemático que representa agentes ($V = \{v_i\}$) e suas conexões ($E = \{e_{i,j}\}$), por meio de um Grafo $G = \{E, V\}$. Em cenários em que as variáveis são temporais, considera-se o Grafo Dinâmico (ou Grafo Temporal).

Os autores dividiram os fatores que influenciam o sistema entre: exclusivos de cada agente $F(v_i)$, e os que expressam influência sobre suas conexões $O(v_i, v_j)$. A função que representa a variação de estados de um agente ao longo do tempo é definida pela eq. 3, e indica resiliência quando seus valores retomam faixas consideradas estáveis para o domínio de aplicação.

$$f(v_i) = \frac{dv_i}{dt} = F(v_i) + \sum_{j=0}^{|V_i|} O(v_i, v_j) * e_{i,j} \quad (3)$$

Neste trabalho, o conjunto de agentes (V) representa os países, o conjunto de relacionamentos (E) representa os indicadores econômicos da relação entre dois países (e.g. Importação e Exportação), e $f(v_i)$ assume os valores referentes aos estados econômicos de cada país ($\Delta Y(t)$ associado a v_i), com faixas que representam "Crise" ou "Estabilidade" (eq. 1 e 2). Em cenários em que saber se um país retornará a estabilidade é mais relevante que saber a parametrização exata de $f(v_i)$, os métodos de aprendizado supervisionados podem ser particularmente úteis e, portanto, são explorados neste trabalho.

3. Previsão de Resiliência baseada em Grafos Dinâmicos

O presente trabalho se baseia nos avanços sobre sistemas dinâmicos [Gao et al. 2016] para medir a vantagem de considerar a vizinhança de um agente durante o aprendizado supervisionado sobre resiliência econômica [Hallegatte 2014]. Para isso, propõe que: 1) é possível utilizar métodos supervisionados para aproximar uma relação entre características dos países seus estados resultantes (Crise ou Estabilidade); e 2) é possível melhorar consideravelmente os resultados do aprendizado supervisionado sobre resiliência econômica ao considerar informações dos parceiros comerciais.

Para comparar as vantagens de considerar uma representação baseada em grafos dinâmicos, este trabalho se baseia no processo tradicional para aprendizado supervisionado utilizando vetores de características referentes aos países, e depois expande o processo para considerar sua vizinhança na rede financeira. A Figura 2 ilustra a metodologia adotada neste trabalho, que pode ser dividida em três fases: **Representação** (fase 1), **Seleção de Variáveis** (fase 2) e **Aprendizado Supervisionado** (fase 3).

O Conjunto de Entrada é composto por Indicadores Econômicos referentes aos países (e.g. PIB e taxa de desemprego) e seus relacionamentos (e.g. montante de importação e exportação). Nos parâmetros são definidos a medida de identificação e mensuração de resiliência econômica utilizada, e os métodos de vizinhança a serem considerados. A Representação é responsável por unir todos os indicadores econômicos referentes a um país em um vetor de características, calcular e separar as informações referentes ao estado econômico (Estabilidade ou Crise). A Seleção de Variáveis reduz o conjunto

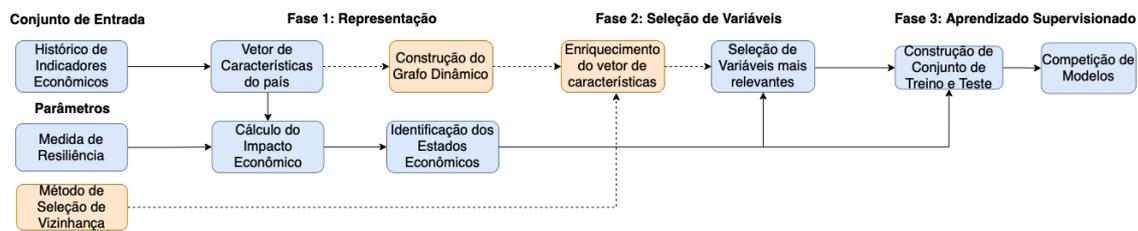


Figura 2. Metodologia para aprendizado supervisionado proposta sobre resiliência econômica, com principais contribuições destacadas em laranja.

de características para um subconjunto com alto poder preditivo e sem variáveis relacionadas ao cálculo dos estados econômicos. O Aprendizado Supervisionado subdivide o conjunto entre conjunto de treino e teste, considerando janelas temporais, ajusta os hiperparâmetros e executa os métodos supervisionados de interesse, retornando os respectivos resultados.

Para mensurar a diferença entre abordagens, são calculados: 1) os resultados obtidos utilizando a representação clássica (módulos destacados em azul na Figura 2) e o respectivo conjunto de variáveis $\{x_{observado}\}$; 2) resultados referentes ao acréscimo dos módulos que permitem o enriquecimento do conjunto com variáveis referentes aos vizinhos de cada país $\{x_{observado}^+\}$. Resumidamente, esses módulos adicionais (em laranja na Figura 2) interferem no Conjunto de Parâmetros, considerando Método de Seleção de Vizinhanças, e nas fases 1 e 2, com a Construção do Grafo Dinâmico a partir do histórico de relacionamento entre países e o Enriquecimento do Vetor de Características, respectivamente.

O Conjunto de Entrada utilizado nos experimentos deste trabalho são referentes ao histórico de indicadores econômicos da base para estudos econômicos *Penn World Table*², disponibilizado pela Universidade de Groningen em parceria com órgãos econômicos mundiais, e da base de dados que descreve o comércio de *commodities* mundial disponibilizado pela Organização das Nações Unidas (ONU)³. A primeira base de dados contém mais de 70 indicadores econômicos de 183 países, entre os anos de 1950 e 2019, somando 12.811 registros. A segunda, é composta por mais de 90 milhões de registros relativos aos montantes financeiros gastos com cada *commodity* e data de 1961 a 2019. Para o histórico de estados econômicos dos países, foram utilizados dados de PIB publicados pelo Banco Mundial⁴, com histórico entre os anos de 1995 e 2019, e acrescidos à base de históricos de indicadores econômicos da Universidade de Groningen. A Medida de Resiliência utilizada é a proposta pelo Banco Mundial [Hallegatte 2014], detalhada pelas eq. 1 e 2. Quanto aos Métodos de Seleção de Vizinhanças, são considerados: 1) os países fornecedores que representam mais de $p\%$ do fluxo de entrada $\mathcal{N}_{entrada}(v_i, p)$; 2) os países consumidores que representam mais de $p\%$ do fluxo de saída $\mathcal{N}_{saida}(v_i, p)$; e 3) os países cujos Produtos Internos Brutos têm uma relação de causalidade Granger com p-valor significativo $\mathcal{N}_{Granger}(v_i, p)$, onde p é um parâmetro testado no intervalo $[0; 1]$ para $\{\mathcal{N}_{entrada}, \mathcal{N}_{saida}\}$, e no intervalo $[0.01; 0.1]$ para $\mathcal{N}_{Granger}$.

²<https://www.rug.nl/ggdc/productivity/pwt/?lang=en>

³<https://comtrade.un.org>

⁴<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>

Na fase de Seleção de Variáveis, as mais relevantes são escolhidas a partir do seu Ganho de Informação Condicional [Wod 1985] entre a variável analisada e o estado econômico (Crise ou Estabilidade), eliminando valores fracos (abaixo de 0.3) ou suspeitos (acima de 0.5). Os métodos de aprendizado supervisionado utilizados foram os clássicos: AdaBoost, Isolation Forest, Random Forest, Naive Bayes, Gaussian Naive Bayes, Histogram Gradient Boosting [de Mello and Ponti 2018].

4. Resultados

O conjunto de dados analisado possui registros de 183 de países com séries anuais num período de 24 anos (1995 a 2019), 70 variáveis e 4.392 exemplos. Os melhores resultados (via *grid search*) foram referentes à janela temporal de 3 anos para treinamento, com uma distância de 1 ano entre as variáveis observadas e o estado econômico associado, e sem intersecção entre os períodos de treinamento ($[t, t + 3]$) e teste ($[t + 4, t + 7]$).

O conjunto de variáveis mais relevantes $\{x_{observado}^+\}$ é composto por 15 variáveis. Nele, estão contidos os nomes dos 3 principais parceiros comerciais do país, os valores dos índices de produtividade para o setor de serviços do país referentes aos últimos 3 anos, e os valores do mesmo índice para os seus parceiros comerciais ao longo dos últimos anos. Vale ressaltar que esse conjunto é o melhor para todos os tipos de seleção de vizinhança ($\mathcal{N}_{saida}, \mathcal{N}_{entrada}, \mathcal{N}_{Granger}$). O conjunto $\{x_{observado}\}$ possui apenas 4 variáveis relevantes (o nome do país e o indicador de serviços relacionado aos últimos 3 anos). Os respectivos resultados estão compilados na Tabela 1.

Tabela 1. Média e desvio padrão (F1-Score) das melhores configurações para cada método supervisionado.

Método Supervisionado	Abordagem	F1-Score
Random Forest (altura:30, estimadores:100)	$\mathcal{N}_{entrada}(v_i, 0.12)$	0.722(0.161)^a
Histogram Gradient Boosting	$\mathcal{N}_{entrada}(v_i, 0.12)$	0.717(0.165)^a
AdaBoostClassifier (estimadores:100)	$\mathcal{N}_{saida}(v_i, 0.12)$	0.605(0.169) ^b
Gaussian Naive Bayes	$\mathcal{N}_{entrada}(v_i, 0.12)$	0.569(0.169)
Naive Bayes	$\mathcal{N}_{saida}(v_i, 0.12)$	0.556(0.176)
Isolation Forest (estimadores:100)	$\mathcal{N}_{entrada}(v_i, 0.12)$	0.256(0.206) ^b

a) p-valor $\leq 5\%$; b) p-valor $\leq 10\%$

As métricas de seleção de vizinhança $\mathcal{N}_{entrada}$ e \mathcal{N}_{saida} apresentaram médias melhores para todos os métodos quando comparados à representação clássica. Em especial, para os métodos *Random Forest* e *Histogram Gradient Boosting* obtiveram ganho de 14% e 19%, respectivamente. Esses métodos também apresentaram evidências estatísticas de que seus resultados foram suficientemente diferentes da representação clássica (com p-valor inferior a 0.05). É relevante ressaltar que a informação com maior contribuição para a previsão dos estados é o nome do país parceiro, o que reforça a hipótese de que as conexões de um país podem revelar mais sobre sua resiliência econômica do que suas características domésticas. Também é interessante notar que, segundo essa abordagem, o nível mínimo de relacionamento para ser considerado um parceiro comercial determinante na previsão de resiliência de um país é somar 12% das importações ou exportações de *commodities*. Esse comportamento pode estar relacionado com o cálculo do PIB que considera a situação da balança comercial.

5. Conclusão

O presente trabalho introduziu uma abordagem para o processo de aprendizado supervisionado sobre modelos econômicos de resiliência, baseado em modelos de sistemas dinâmicos com múltiplos agentes e utilizando bases de dados econômicos públicas e disponibilizadas por órgãos econômicos mundiais. Além disso, os resultados apresentam evidências estatísticas que sugerem vantagem preditiva ao considerar informações das relações de importação e exportação de um país para antecipar crise ou estabilidade econômica.

Para trabalhos futuros, considera-se: A expansão de modelos supervisionados de classificação para regressão sobre intensidade e duração das crises econômicas; e A expansão para grafos com mais de um tipo de relacionamento (diferentes para cada *commodity*), com redes multiplex.

Referências

- Bristow, G. and Healy, A. (2018). *Economic crisis and the resilience of regions: A European study*. Edward Elgar Publishing.
- de Mello, R. F. and Ponti, M. A. (2018). Statistical learning theory. In *Machine Learning*, pages 75–128. Springer.
- Gao, J., Barzel, B., and Barabási, A.-L. (2016). Universal resilience patterns in complex networks. *Nature*, 530(7590):307–312.
- Hallegatte, S. (2014). Economic resilience: definition and measurement. *World Bank Policy Research Working Paper*, (6852).
- Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual review of ecology and systematics*, 4(1):1–23.
- Mai, X. and Chan, R. C. (2020). Detecting the intellectual pathway of resilience thinking in urban and regional studies: A critical reflection on resilience literature. *Growth and Change*, 51(3):876–889.
- Punzo, G., Tewari, A., Butans, E., Vasile, M., Purvis, A., Mayfield, M., and Varga, L. (2020). Engineering resilient complex systems: the necessary shift toward complexity science. *IEEE Systems Journal*, 14(3):3865–3874.
- Resbeut, M. (2021). *On Economic Resilience: A Theoretical Investigation of The Influence of Clusters*. PhD thesis, Université de Fribourg.
- Salignac, F., Hanoteau, J., and Ramia, I. (2022). Financial resilience: A way forward towards economic development in developing countries. *Social Indicators Research*, 160(1):1–33.
- Wod, I. (1985). Weight of evidence: A brief survey. *Bayesian statistics*, 2:249–270.
- Yildirim, E., Just, C., and Demir, I. (2022). Flood risk assessment and quantification at the community and property level in the state of iowa. *International Journal of Disaster Risk Reduction*, page 103106.