



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA
DOUTORADO ACADÊMICO EM FÍSICA

ANTONIO EDINALDO DE OLIVEIRA FERREIRA

EMERGÊNCIA E LEIS DE ESCALA EM SISTEMAS COMPLEXOS:
UMA ABORDAGEM ORIENTADA POR DADOS

FORTALEZA

2026

ANTONIO EDINALDO DE OLIVEIRA FERREIRA

EMERGÊNCIA E LEIS DE ESCALA EM SISTEMAS COMPLEXOS:
UMA ABORDAGEM ORIENTADA POR DADOS

Tese apresentada ao Programa de Pós-Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Física. Área de Concentração: Física da Matéria Condensada

Orientador: Prof. Dr. Cláudio Lucas Nunes de Oliveira

FORTALEZA

2026

ANTONIO EDINALDO DE OLIVEIRA FERREIRA

EMERGÊNCIA E LEIS DE ESCALA EM SISTEMAS COMPLEXOS:
UMA ABORDAGEM ORIENTADA POR DADOS

Tese apresentada ao Programa de Pós-Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Física. Área de Concentração: Física da Matéria Condensada

Aprovada em: 26/03/2026

BANCA EXAMINADORA

Prof. Dr. Cláudio Lucas Nunes de
Oliveira (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. César Ivan Nunes Sampaio Filho
Universidade Federal do Ceará (UFC)

Prof. Dr. Jeanlex Soares de Sousa
Universidade Federal do Ceará (UFC)

Prof. Dr. Jorge Luiz Bezerra de Araújo
Universidade de Fortaleza (Unifor)

Prof. Dr. Erneson Alves de Oliveira
Universidade de Fortaleza (Unifor)

A Deus, a minha família e aos meus amigos.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, cuja força e coragem me acompanharam durante toda a minha jornada acadêmica. Sou profundamente grato ao meu pai, Arnaldo, à minha mãe, Ivonete, e ao meu irmão, Ronaldo, pelo apoio inabalável que sempre me deram. Agradeço especialmente à minha namorada, Luanna, por estar continuamente ao meu lado, oferecendo apoio, motivação e acreditando no meu potencial em cada etapa dessa trajetória.

Expresso minha sincera gratidão ao professor Dr. Cláudio Lucas Nunes de Oliveira, cuja orientação tem sido decisiva ao longo de toda a minha formação acadêmica. Foram dez anos de trabalho conjunto, iniciados nos primeiros projetos de iniciação científica, passando pelo Trabalho de Conclusão de Curso na graduação, pela Dissertação de Mestrado e, agora, culminando na elaboração desta tese de Doutorado. Sua dedicação, paciência, apoio contínuo e confiança foram fundamentais para o desenvolvimento deste trabalho, que também resultou na produção de dois artigos científicos derivados desta pesquisa.

Registro minha profunda gratidão ao professor Dr. José Soares de Andrade Júnior e ao professor Dr. Humberto de Andrade Carmona, cuja colaboração foi essencial para o avanço deste trabalho. Sou especialmente grato a ambos pelo espaço concedido na Escola de Saúde Pública (ESP), que permitiu a continuidade da tese e favoreceu o planejamento de um novo artigo em conjunto. Agradeço também ao professor Dr. Jeanlex Soares de Sousa e ao professor Dr. Cesar Ivan Nunes Sampaio Filho pelo apoio e pelas contribuições importantes oferecidas ao longo deste processo.

Agradeço de maneira especial ao professor Dr. Jorge Luiz Bezerra de Araújo, cuja orientação e participação foram fundamentais para a consolidação desta pesquisa. Sua contribuição não apenas enriqueceu o desenvolvimento da tese, como também resultou na publicação de um artigo derivado de um dos resultados aqui apresentados, fortalecendo significativamente a relevância e o impacto deste trabalho.

Estendo meus agradecimentos aos professores do Departamento de Física — Renan Landim, Carlos William, Raimundo Nogueira, Rose Freire e Nildo Loiola — pelos ensinamentos e pelo apoio ao longo do curso de Doutorado.

Agradeço aos meus amigos, cuja presença, companheirismo e incentivo foram fundamentais ao longo de toda a minha trajetória acadêmica. Meu reconhecimento especial vai para Sergio Tanaka, Alehff Einstein, Gabriel de Sousa, Higo Barros, Rômulo Sampaio, Higor Monteiro, João Henrique, Alyson da Silva, Alyson Freitas, Márcio Antônio, Pedro

Bruno, Thiago Duarte, Damião Ferreira, Geovane Cavalcante, Pedro Dias, Manoel Vicente, Pedro Uchôa, Afonso Luiz e Lucas Saraiva (Professor), além de todo o grupo do laboratório de biofísica, que contribuíram com apoio constante e amizade sincera durante essa jornada.

Expresso também minha gratidão à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro contínuo durante o Doutorado, indispensável para a realização e o avanço desta pesquisa.

Por fim, agradeço à Universidade Federal do Ceará (UFC) pela oportunidade de desenvolver esta tese e por permitir a concretização do meu caminho acadêmico rumo ao título de Doutor em Física.

“Nunca confunda educação com inteligência,
você pode fazer um doutorado e ainda ser um
idiota.”

(Richard Feynman)

RESUMO

Neste trabalho, investigamos dois sistemas complexos de naturezas distintas sob uma mesma abordagem orientada por dados: (i) redes viscoelásticas imersas em fluido e (ii) o sistema urbano-econômico do Ceará, estruturado por fluxos multicamadas de comércio entre 184 municípios. Em ambos os casos, lidamos com muitos elementos interagentes, heterogeneidade estrutural e acoplamento multiescala, buscando inferir propriedades macroscópicas emergentes a partir de dados locais de alta dimensionalidade por meio de descritores coletivos. No estudo viscoelástico, consideramos um modelo de rede elástica com dissipação sublinear, caracterizada pelo parâmetro $\alpha \in [0, 1]$ na força de arrasto $f_a = \gamma v^\alpha$. Em redes homogêneas e rígidas, o expoente de arrasto α governa a transição entre diferentes regimes de relaxação, em uma faixa sublinear específica de α a resposta de força assume a forma de lei de potência, $F(t) \sim t^{-\beta}$, enquanto, fora dessa faixa, passam a prevalecer respostas de relaxação de tipo exponencial. Em redes heterogêneas gaussianas, a desordem amplia a faixa de α que gera relaxações em lei de potência, ao mesmo tempo em que mantém uma faixa robusta de valores de β , sensível sobretudo ao nível médio de rigidez, mas pouco à largura da distribuição. No sistema urbano-econômico, analisamos mais de 2,5 bilhões de transações formais (NF-e, 2016–2019) e mostramos que compras, vendas e comércio interno escalam superlinearmente com a população, com expoentes distintos. O indicador econômico RCA (*Revealed Comparative Advantage*), utilizado para caracterizar padrões de especialização produtiva regional, segue distribuições do tipo Zipf–Mandelbrot para as três camadas de fluxos, revelando hierarquias e limites de diversificação. O *Fitness* municipal, uma métrica econômica que mede o grau de diversificação e complexidade produtiva das cidades com base em seus produtos de exportação, também escala superlinearmente, indicando acúmulo desproporcional de capacidades estruturais, especialmente no lado da oferta e da recirculação interna. A integração desses resultados evidencia que leis de potência, emergência coletiva, hierarquias, transições críticas e organização multiescala são padrões recorrentes tanto em materiais viscoelásticos quanto em sistemas urbanos. Mostramos que a combinação de simulações numéricas, aprendizado de máquina, análise de redes, inferência bayesiana e técnicas multivariadas consolida a Ciência de Dados como um arcabouço unificado para revelar estruturas profundas em sistemas complexos aparentemente díspares.

Palavras-chave: sistemas complexos; ciência de dados; leis de potência; matéria mole; complexidade econômica.

ABSTRACT

In this work, we investigate two complex systems of distinct natures under a single data-driven approach: (i) viscoelastic networks immersed in a fluid and (ii) the urban–economic system of Ceará, structured by multilayer trade flows among 184 municipalities. In both cases, we deal with many interacting elements, structural heterogeneity, and multiscale coupling, seeking to infer emergent macroscopic properties from high-dimensional local data through collective descriptors. In the viscoelastic study, we consider a model of an elastic network with sublinear dissipation, characterized by the parameter $\alpha \in [0, 1]$ in the drag force $f_a = \gamma v^\alpha$. In homogeneous and rigid networks, the drag exponent α governs the transition between different relaxation regimes: within a specific sublinear range of α , the force response takes the form of a power law, $F(t) \sim t^{-\beta}$, whereas outside this range, exponential-type relaxation responses become predominant. In heterogeneous Gaussian networks, disorder broadens the range of α values that give rise to power-law relaxations, while preserving a robust interval of β values, which are mainly sensitive to the mean stiffness level and only weakly dependent on the width of its distribution. In the urban–economic system, we analyze more than 2.5 billion formal transactions (electronic invoices, NF-e, 2016–2019) and show that purchases, sales, and internal trade scale superlinearly with population, with distinct exponents. The economic indicator RCA (*Revealed Comparative Advantage*), used to characterize patterns of regional productive specialization, follows Zipf–Mandelbrot-type distributions for the three flow layers, revealing hierarchies and diversification limits. Municipal *Fitness*, an economic metric that measures the degree of diversification and productive complexity of cities based on their export products, also scales superlinearly, indicating a disproportionate accumulation of structural capabilities, especially on the supply side and in internal recirculation. The integration of these results shows that power laws, collective emergence, hierarchies, critical transitions, and multiscale organization are recurrent patterns in both viscoelastic materials and urban systems. We demonstrate that the combination of numerical simulations, machine learning, network analysis, Bayesian inference, and multivariate techniques consolidates Data Science as a unified framework for unveiling deep structures in seemingly disparate complex systems.

Keywords: complex systems; data science; power laws; soft matter; economic complexity.

LISTA DE FIGURAS

- Figura 1 – O diagrama de Venn mostra como ciência da computação, matemática/estatística e conhecimento da área se sobrepõem para formar a ciência de dados, destacando as regiões dedicadas a *softwares* tradicionais, análise de dados e IA/aprendizado de máquina. 33
- Figura 2 – Técnicas de pré-processamento de dados. 35
- Figura 3 – Similaridade entre a inteligência humana e artificial. 37
- Figura 4 – Representação esquemática do fluxo de transformação dos dados, ilustrando as etapas que vão da informação bruta à modelagem matemática e à construção de sistemas inteligentes. O diagrama evidencia o processo analítico que sustenta técnicas de aprendizado de máquina e aprendizado profundo, mostrando como padrões, relações e estruturas emergem progressivamente ao longo do tratamento dos dados. 38
- Figura 5 – A imagem apresenta um diagrama que organiza visualmente a relação entre diferentes áreas da Inteligência Artificial (IA), destacando onde se posicionam o *Machine Learning* (ML), o *Deep Learning* (DL) e a IA Generativa. As regiões aparecem dispostas em camadas concêntricas, evidenciando que o Deep Learning constitui um subconjunto do Machine Learning, enquanto ambos pertencem ao domínio mais amplo da IA. O diagrama também mostra a IA Generativa como uma área específica dentro do Deep Learning, indicando seu foco na produção de novos dados a partir de padrões previamente aprendidos. 39
- Figura 6 – Trata-se da sequência de elementos essenciais que descrevem como um programa de computador aprende a partir de dados para realizar uma tarefa específica, aprimorando seu desempenho à medida que desenvolve compreensão sobre o problema. 41
- Figura 7 – Cenários de aplicações do aprendizado de máquina. 42

- Figura 8 – Fluxograma que representa o processo de aprendizado de máquina, destacando etapas fundamentais como a coleta e o pré-processamento dos dados, o treinamento do modelo, a avaliação, o refinamento e a implementação. O processo é contínuo e inclui ciclos sucessivos de avaliação e melhoria. Também são apresentadas as categorias de métodos de aprendizado — supervisionado, semi-supervisionado, não supervisionado e por reforço — juntamente com os principais tipos de problemas que podem ser abordados, como regressão, classificação, clusterização, redução de dimensionalidade e extração de regras de associação. 45
- Figura 9 – Uma rede elástica cúbica de altura L encontra-se imersa em um fluido (não mostrado na figura). Cada partícula possui um diâmetro ρ e está interligada por molas de constante elástica k_{kl} com suas 12 vizinhas mais próximas. A rede é submetida à indentação por uma ponta esférica de diâmetro ρ_s de cima para baixo. 59
- Figura 10 – A curva representa a distribuição gaussiana das constantes elásticas k_{kl} em uma rede heterogênea. O pico na posição k_m corresponde à média da distribuição, indicando o valor mais provável para as constantes elásticas do sistema. A largura da curva, controlada pelo desvio padrão σ , expressa o grau de dispersão dos valores em torno da média. A região sombreada entre $[k_m - \sigma, k_m + \sigma]$ abrange cerca de 68% dos dados, correspondendo ao intervalo de um desvio padrão. Em seguida, os intervalos $[k_m - 2\sigma, k_m + 2\sigma]$ e $[k_m - 3\sigma, k_m + 3\sigma]$ englobam aproximadamente 95% e 99,7% dos valores, respectivamente, caracterizando os domínios de dois e três desvios padrão da média. 60

- Figura 11 – Composição setorial e distribuição regional da atividade comercial. (a) *Treemap* dos produtos comercializados agregados no nível NCM de dois dígitos, com cores indicando as seções correspondentes do HS. As áreas dos retângulos são proporcionais ao valor cumulativo de todas as transações entre empresas de 2016 a 2019, e as 15 principais categorias de produtos são identificadas com seus códigos NCM e respectivas participações em valor. (b) Breves descrições das categorias selecionadas de dois dígitos do sistema HS/NCM. A distribuição espacial dos fluxos totais entre empresas nos 184 municípios do Ceará é agregada em todos os produtos: (c) vendas totais para outros municípios, (d) transações totalmente intramunicipais e (e) compras totais provenientes de outros municípios. A intensidade da cor representa a magnitude relativa (vermelho = maior, azul = menor). 65
- Figura 12 – O painel (a) mostra as 15 categorias NCM de dois dígitos com maior valor acumulado ao longo de quatro anos, ordenadas em ordem decrescente. Os códigos HS/NCM e as descrições curtas das 15 principais categorias estão listados na Tabela 11(b). O painel (b) exhibe os fluxos georreferenciados entre os 184 municípios em 2016: os nós estão localizados nas sedes municipais; as arestas direcionadas representam ligações vendedor → comprador, mostrando apenas transações superiores a US\$10 milhões para maior clareza visual. . . 68
- Figura 13 – Estrutura matricial 184×1228 utilizada para o cálculo do RCA nas três camadas de comércio do Ceará. Cada célula representa o fluxo $S_{c,p}$ entre a cidade c e produto p . A faixa vertical em azul destaca o total estadual do produto ($S_p = \sum_c S_{c,p}$), enquanto a faixa horizontal em vermelho evidencia o total movimentado pela cidade ($S_c = \sum_p S_{c,p}$). A célula verde marca o valor específico $S_{c,p}$ cujo peso relativo é avaliado na métrica. O retângulo laranja superior indica o total agregado transacionado no estado ($S_t = \sum_{c,p} S_{c,p}$). O esquema ilustra como as diferentes somas marginais se articulam na matriz cidade-produto para compor a razão utilizada no cálculo do RCA. 71

Figura 14 – (a) Curvas numéricas de força para $\gamma = 80$ e $k = 800$ e vários valores de α , exibindo diferentes relaxações viscoelásticas. Os casos para $\alpha = 0.45$ e $\alpha = 0.85$ são mostrados nos painéis (b) e (c), respectivamente, onde as linhas tracejada e pontilhada representam ajustes com os modelos exponencial e de lei de potência das Eqs. 5.3 e 5.4, respectivamente.	78
Figura 15 – (a) Visualização gráfica do processo de clusterização utilizando o método <i>K-means</i> . Cada ponto representa um experimento computacional para uma dada combinação de k , γ e α , e as cores representam diferentes resultados de relaxação, a saber: lei de potência (PL) em azul, exponencial (EXP) em amarelo e comportamento transicional (TR) em vermelho. (b) Um gráfico de coordenadas paralelas mostra como diferentes valores dos parâmetros mesoscópicos da rede levam a distintos tipos de relaxação viscoelástica. Cada linha atravessa uma combinação de k , γ e α , juntamente com o comportamento reológico correspondente.	80
Figura 16 – A probabilidade de um dado valor de α levar a uma resposta do tipo lei de potência (barras azuis), exponencial (barras amarelas) ou transicional (barras vermelhas). As linhas sólidas mostram a probabilidade calculada pelo algoritmo KDE (<i>kernel density estimation</i>). O painel (a) apresenta resultados para toda a faixa de valores de k e γ , enquanto o painel (b) considera apenas redes com k ($k \geq 800$) e γ ($\gamma \geq 80$).	81
Figura 17 – Para aquelas redes classificadas como um material de lei de potência na Figura 30(b), mostramos a relação entre o expoente de relaxação β e o expoente de arrasto mesoscópico α	82
Figura 18 – Representações gráficas das curvas de força para $\gamma = 50$, $k_m = 500$, e $\alpha = 0.5$, utilizando 5 valores distintos de sementes aleatórias na rede viscoelástica. Observa-se variações mínimas no comportamento das curvas de relaxação, evidenciadas pela curva de força média <i>FM</i>	83

- Figura 19 – Apresentação gráfica da distribuição Gaussiana $P(k_{kl})$ para os parâmetros $\gamma = 50$, $k_m = 800$ e $\alpha = 0.5$, considerando diferentes valores do desvio padrão σ . A análise é realizada para 4 níveis distintos de variabilidade, revelando como a dispersão das constantes elásticas se altera na rede viscoelástica. Observa-se que, para $\sigma = 0$, a rede mantém uma estrutura completamente homogênea, enquanto os casos com $\sigma = [80, 160, 240]$ evidenciam um aumento progressivo da heterogeneidade, refletido no alargamento significativo da distribuição. 84
- Figura 20 – O diagrama de coordenadas paralelas ilustra como as variações nos parâmetros mesoscópicos da rede conduzem a diferentes padrões de relaxamento viscoelástico. Cada linha representa uma combinação específica de valores para k_m , γ , α , σ e para o comportamento reológico resultante, permitindo visualizar de forma integrada como cada parâmetro influencia a resposta global do sistema. 85
- Figura 21 – A probabilidade associada a um determinado valor de α resultar em uma resposta do tipo lei de potência (barras azuis), exponencial (barras amarelas) ou transicional (barras vermelhas) é apresentada no gráfico. As linhas contínuas correspondem à probabilidade estimada por meio do algoritmo KDE (*kernel density estimation*). No painel (a), exibimos os resultados considerando toda a faixa de variação dos parâmetros k_m , γ e σ . Já no painel (b), a análise é restrita às redes caracterizadas por valores mais elevados destes parâmetros, especificamente k_m ($k_m \geq 500$) e γ ($\gamma \geq 80$). 86
- Figura 22 – Nos casos em que a rede apresenta comportamento de lei de potência (sem qualquer filtragem, abrangendo todos os valores analisados), ilustramos a relação entre o expoente de relaxação β e o expoente de arrasto mesoscópico α na rede heterogênea com distribuição Gaussiana. 87
- Figura 23 – Nos casos em que a rede apresenta comportamento de lei de potência (sem filtragem, abrangendo todos os valores analisados), apresenta-se a correlação entre o expoente de relaxação β e o desvio padrão σ (expressando a fração de k_m) na distribuição Gaussiana. Observa-se que, mesmo para diferentes valores de σ , não há variação significativa nos valores de β 87

Figura 24 – Nos casos em que a rede manifesta um comportamento de lei de potência (sem filtragem, incluindo todos os valores analisados), apresentamos a visualização da relação entre o expoente de relaxação β e o parâmetro efetivo k_{ef} (ou k_m) na rede viscoelástica com distribuição Gaussiana. Observa-se que o expoente β varia de acordo com os diferentes valores de k_{ef} na rede.	88
Figura 25 – Esquema ilustrativo do uso do algoritmo <i>k-means</i> no contexto de aprendizado de máquina não supervisionado para <i>clusterizar</i> as curvas de força, destacando as etapas de pré-processamento, aplicação do algoritmo e interpretação dos resultados.	90
Figura 26 – Esquema da divisão do <i>dataframe</i> filtrado (<i>df filtro</i>) em 80% para o conjunto de treinamento e 20% para o conjunto de teste, considerando os três atributos de entrada (k, γ, α) e os três rótulos de saída (χ_E, χ_P, β). Essa configuração caracteriza explicitamente um modelo de aprendizado supervisionado. . . .	93
Figura 27 – Comparação dos modelos de predição <i>Random Forest Regressor</i> (RFR), Regressão Linear (RL) e <i>Decision Tree Regressor</i> (DTR) em relação aos valores de teste de χ_E . Observa-se que o modelo RFR apresenta melhor aderência aos dados de teste, aproximando-se mais dos valores observados. . .	95
Figura 28 – Escalonamento alométrico da atividade comercial municipal. (a–c) Relação entre a população municipal e o valor de todas as transações entre empresas em 2016: (a) vendas para outras cidades, (b) transações intramunicipais e (c) compras de outras cidades. Cada ponto representa um município. As linhas vermelhas representam ajustes bayesianos em escala log–log, com curvas pontilhadas e tracejadas indicando, respectivamente, os limites inferior e superior do intervalo de confiança de 95%. O expoente correspondente, β , captura o escalonamento alométrico anual para cada tipo de fluxo. (d) Estimativas anuais de β para vendas (triângulos verdes), transações intramunicipais (quadrados roxos) e compras (círculos laranja), com linhas tracejadas marcando as médias multianuais: $\langle \beta_S \rangle = 2.07 \pm 0.01$, $\langle \beta_I \rangle = 2.15 \pm 0.02$ e $\langle \beta_B \rangle = 1.54 \pm 0.01$. Os expoentes calculados a partir da agregação completa de quatro anos indicam que o comportamento alométrico das três camadas de comércio no Ceará é estável ao longo do tempo.	96

Figura 29 – Evolução temporal dos expoentes alométricos β (2016–2019) para cada um dos 15 principais grupos de produtos HS2 e para o agregado “Todos os produtos”. Os pontos representam os valores anuais de β estimados para cada produto e categoria de fluxo (compras, vendas e comércio intraurbano). As linhas tracejadas indicam a média ao longo dos anos, $\langle\beta\rangle$, correspondente a cada produto (ou agregado) e categoria. 97

Figura 30 – Expoentes de escalonamento alométrico por categoria de produto. Valores médios anuais do expoente alométrico $\langle\beta\rangle$ para os 15 grupos de produtos mais comercializados (códigos NCM de 2 dígitos). Para cada produto, β é estimado de forma independente para as vendas intermunicipais (β_S , triângulos verdes), transações intramunicipais (β_I , quadrados roxos) e compras intermunicipais (β_B , círculos laranja). As barras de erro representam o erro-padrão ao longo dos quatro anos (2016–2019). Linhas horizontais tracejadas indicam as médias gerais multi-produto para cada tipo de fluxo. Os resultados revelam diferenças sistemáticas de escalonamento entre os produtos: os fluxos intramunicipais exibem o comportamento superlinear mais forte, enquanto as compras apresentam o escalonamento mais fraco. 98

Figura 31 – Estrutura Zipf–Mandelbrot na RCA municipal. Gráficos rank–ordem da Vantagem Comparativa Revelada (RCA) para todos os municípios em 2016, apresentados separadamente para (a) vendas intermunicipais, (b) transações intramunicipais e (c) compras intermunicipais. A curva preta representa a distribuição entre todos os municípios e produtos. Cidades selecionadas são deslocadas horizontalmente para maior clareza visual — Fortaleza (maior), Crato (6ª maior), Caridade (90ª) e Guaramiranga (menor). Para $RCA \geq 1$ (linhas tracejadas), as distribuições seguem uma forma Zipf–Mandelbrot com expoentes $\alpha_S = 1.23$ (vendas), $\alpha_I = 1.24$ (intra) e $\alpha_B = 0.84$ (compras), indicando um decaimento mais acentuado e maior especialização em vendas e atividade interna. Em contraste, as compras apresentam uma cauda mais plana, consistente com um padrão mais diversificado de produtos demandados. 99

Figura 32 – Distribuições de rank-size dos valores de Vantagem Comparativa Revelada (RCA) para os municípios do Ceará nas três modalidades de fluxo: (a) vendas, (b) transações intramunicipais e (c) compras. Para cada categoria, são mostradas as curvas de 2016–2019, que exibem grande sobreposição e, portanto, robustez temporal. A linha laranja em cada painel representa o ajuste Zipf–Mandelbrot, $RCA(k) \sim (k + q)^{-\alpha}$, indicando um claro decaimento em lei de potência para $RCA > 1$. Os expoentes médios de Zipf ao longo do período são $\langle \alpha_S \rangle = 1.25$ (vendas), $\langle \alpha_I \rangle = 1.23$ (intramunicipal) e $\langle \alpha_B \rangle = 0.82$ (compras). 99

Figura 33 – Matrizes binárias de RCA por camada de comércio. As matrizes cidade–produto em 2016 mostram quais municípios exibem Vantagem Comparativa Revelada ($RCA \geq 1$) em (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Cada ponto verde marca um par cidade–produto no qual o município c se especializa no produto p . As linhas correspondem aos 184 municípios ordenados por Fitness crescente, e as colunas correspondem aos 1.228 produtos NCM de quatro dígitos ordenados por Complexidade crescente. Quatro municípios de referência são destacados: Fortaleza (FOR, mais populoso), Crato (CRA, 6º maior), Caridade (CAR, 90º) e Guaramiranga (GUA, o menor). Exemplos de produtos de menor a maior complexidade em cada camada incluem: em vendas, do NCM 7311 (contêineres para gases comprimidos ou liquefeitos) ao NCM 8904 (embarcações para reboque ou empurrar outras embarcações); em transações intraurbanas, do NCM 6904 (tijolos/ladrilhos cerâmicos) ao NCM 8478 (máquinas para preparação de tabaco); em compras, do NCM 2710 (óleos de petróleo) ao NCM 5108 (fios de pelos finos de animais). As matrizes revelam uma estrutura fortemente aninhada: municípios de alto Fitness participam tanto de produtos complexos quanto de produtos simples, enquanto municípios de baixo Fitness permanecem concentrados em um conjunto estreito de bens de baixa Complexidade. Notavelmente, apesar de ter a menor população, Guaramiranga posicionou-se consistentemente acima de muitos municípios em Fitness, refletindo uma diversificação impulsionada por sua economia orientada ao turismo. 101

- Figura 34 – Mesmo procedimento da Figura 33 para 2017. Extremos por camada: vendas — menor NCM 7311 (contêineres para gases comprimidos ou liquefeitos), maior NCM 8604 (veículos para inspeção e manutenção de ferrovias ou similares); mercado interno — menor NCM 8201 (pás, enxadas, picaretas, forcados, garfos, ovos e raspadores; machados, capões e ferramentas semelhantes com fio), maior NCM 2913 (derivados halogenados, sulfonados, nitrados ou nitrosados dos produtos da posição 29.12); compras — menor NCM 2710 (óleos de petróleo), maior NCM 2616 (minérios de metais preciosos e seus concentrados). 102
- Figura 35 – Mesmo procedimento da Figura 33 para 2018. Extremos por camada: vendas — menor NCM 0102 (animais vivos da espécie bovina), maior NCM 2610 (minérios de cromo e seus concentrados); mercado interno — menor NCM 0102 (animais vivos da espécie bovina), maior NCM 7105 (pó de diamante, pedras preciosas, semipreciosas ou sintéticas); compras — menor NCM 2202 (águas, incluindo águas minerais e águas gaseificadas), maior NCM 8524 (módulos de visualização de tela plana, mesmo que incorporem telas sensíveis ao toque). 103
- Figura 36 – Mesmo procedimento da Figura 33 para 2019. Extremos por camada: vendas — menor NCM 7311 (contêineres para gases comprimidos ou liquefeitos), maior NCM 6703 (cabelo disposto na mesma direção, fino, descolorido ou preparado de outra forma); mercado interno — menor NCM 0102 (animais vivos da espécie bovina), maior NCM 2307 (borras de vinho; tártaro em bruto); compras — menor NCM 2710 (óleos de petróleo), maior NCM 5005 (filamentos de resíduos de Seda, não embalados para venda a varejo). 104

Figura 37 – Trajetórias de 2016 a 2019 no plano Fitness–PIB. A coevolução do Fitness per capita e do Produto Interno Bruto per capita dos municípios é mostrada para as três camadas de comércio, com eixos em escala logarítmica: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Cada seta representa a trajetória de quatro anos de um município, de 2016 a 2019. Os cinco municípios mais populosos — Fortaleza (laranja), Caucaia (roxo), Juazeiro do Norte (verde), Maracanaú (vermelho) e Sobral (marrom) — são destacados para ilustrar os caminhos desenvolvimentistas heterogêneos observados no estado. As trajetórias revelam como a competitividade estrutural (Fitness) e os níveis de renda (PIB) evoluem conjuntamente em diferentes contextos transacionais, com alguns municípios exibindo ganhos simultâneos em ambas as dimensões, enquanto outros permanecem essencialmente estacionários. 105

Figura 38 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2016 nas três camadas de comércio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 2.03$, $\beta_I^F = 1.74$ e $\beta_B^F = 1.26$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.25$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%). (e) Razões anuais β^F/β^G para 2016–2019. O Fitness associado a vendas (verde) e circulação interna (roxo) cresce superlinearmente com a renda ($\beta_S^F/\beta^G \approx 1.62$ e $\beta_I^F/\beta^G \approx 1.39$). Em contraste, o Fitness associado a compras (laranja) cresce de forma aproximadamente proporcional ($\beta_B^F/\beta^G \approx 1$), indicando que capacidades estruturais se expandem de maneira desproporcional com a prosperidade apenas nos lados da oferta e da circulação interna. 106

Figura 39 – Comparação rank–rank entre o Fitness municipal e o PIB nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Cada ponto corresponde a um município, plotado por sua posição no ranking de PIB (eixo x) e no ranking de Fitness (eixo y). A linha vermelha e um ajuste de mínimos quadrados mostrado apenas para visualização. Os coeficientes de correlação de Spearman são 0.78 para vendas, 0.77 para transações internas e 0.72 para compras, indicando uma associação monotonicamente forte entre produto econômico e capacidade estrutural, com alinhamento ligeiramente mais fraco no lado da demanda. Essas correlações corroboram os resultados de escalonamento, que mostram que capacidades ligadas a oferta e circulação interna acompanham a renda mais de perto do que as capacidades de compra impulsionadas pela demanda.	107
Figura 40 – Diagrama de fluxo do algoritmo de Dinâmica Molecular.	133
Figura 41 – Condições periódicas de contorno (CPC) e convenção de imagem mínima.	135
Figura 42 – Ilustração das interações ligantes e não-ligantes entre as partículas.	136
Figura 43 – Gráfico do potencial de Lennard-Jones, destacando as influências de $\frac{1}{r^{12}}$ e $-\frac{1}{r^6}$ para valores de $\rho = 1$ e $\epsilon = 1$	141
Figura 44 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2017 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 1.61$, $\beta_I^F = 1.66$ e $\beta_B^F = 1.10$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.22$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%).	155

Figura 45 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2018 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 1.53$, $\beta_I^F = 1.66$ e $\beta_B^F = 1.30$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.22$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%). 156

Figura 46 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2019 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 1.50$, $\beta_I^F = 1.63$ e $\beta_B^F = 1.26$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.24$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%). 157

LISTA DE TABELAS

Tabela 1 – Conversão de unidades adimensionais para unidades reais no modelo Lennard-Jones para Argônio ($\frac{\epsilon}{k_B} = 119.8 K, \rho = 3.4 \times 10^{-10} m, M = 0.03994 kg/mol$).	142
Tabela 2 – Exemplos de Funções e Suas Transformadas de Laplace.	146

LISTA DE ABREVIATURAS E SIGLAS

APIs	<i>Application Programming Interfaces</i>
BRL	Real Brasileiro
CPC	Condições Periódicas de Contorno
DL	<i>Deep Learning</i>
DM	Dinâmica Molecular
DTR	<i>Decision Tree Regressor</i>
ECI	<i>Economic Complexity Index</i>
EDA	<i>Exploratory Data Analysis</i>
EXP	Exponencial
FCC	<i>Face-Centered Cubic</i>
FM	Força Média
GDPR	<i>General Data Protection Regulation</i>
HS	<i>Harmonized System</i>
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IGP	Índice Geral de Preços
IPECE	Instituto de Pesquisa e Estratégia Econômica do Ceará
KDE	<i>Kernel Density Estimation</i>
LGPD	Lei Geral de Proteção de Dados Pessoais
LLMs	<i>Large Language Models</i>
MCMC	Cadeias de Markov via Monte Carlo
ML	<i>Machine Learning</i>
NaN	<i>Not a Number</i>
NCM	Nomenclatura Comum do Mercosul
NF-e	Notas Fiscais Eletrônicas
PCA	<i>Principal Component Analysis</i>
PIB	Produto Interno Bruto
PL	Power Law
RCA	<i>Revealed Comparative Advantage</i>
RFR	<i>Random Forest Regressor</i>

RL	Regressão Linear
SQL	<i>Structured Query Language</i>
TR	Transição
USD	Dólar Americano

LISTA DE SÍMBOLOS

t	tempo
$\vec{r}(t)$	vetor posição
$r(t)$ ou r	módulo do vetor posição
$\vec{v}(t)$	vetor velocidade
$v(t)$ ou v	módulo do vetor velocidade
v_0	velocidade inicial
τ	tempo de relaxação do material viscoelástico
$\xi(t)$	força de flutuação
f_a	força de arrasto
γv^α	força de arrasto generalizada com o expoente α na velocidade
k_B	constante de Boltzmann
T	temperatura
γ	coeficiente da força de arrasto
C_a	coeficiente da força de arrasto com o expoente quadrático na velocidade
η	viscosidade do fluido
ρ	diâmetro da partícula esférica
ϕ	densidade do fluido
A_0	área de contato antes que o material sofra deformação
A	área após o material sofrer deformação
v_a	velocidade de referência
μ	expoente de Vogel
α	expoente da velocidade da força de arrasto / expoente da lei de zipf-Mandelbrot
L	número de camadas na direção z
N	número de partículas esféricas
k	constante de mola / <i>clusters</i> / rank
ℓ	distância de equilíbrio entre os centros das esferas

$\delta(t)$	indentação
δ_{max}	indentação máxima
γ^*	unidade reduzida do coeficiente de arrasto
ε	constante energética
ρ_s	diâmetro da ponta esférica
ϖ	expoente de rigidez do indentador
r_{ks}	distância entre o centro da partícula e o indentador
U	potencial de interação
k_{ef}	constante de mola efetiva
k_m	constante de mola média
σ	desvio padrão
λ	geometria do indentador
$F(t)$	curva de força numérica
$\mathcal{F}(t)$	curva de força analítica
ν	coeficiente de Poisson
$R_P(t)$	modelo de relaxação para materiais de lei de potência
$R_E(t)$	modelo de relaxação para materiais exponenciais
E_∞	módulo elástico em tempos suficientemente longos
τ_l	tempo de <i>loading</i>
ΔE	diferença entre o valor máximo, em $t = \tau_l$ e E_∞
β	expoente de decaimento da lei de potência / expoente de alometria
χ	erro quadrático médio
χ_P	erro quadrático médio da lei de potência
χ_E	erro quadrático médio da exponencial
N_p	número de pontos nas curvas de força
F	fitness
p	produto
c	cidade

$S_{c,p}^S$	valor das vendas externas para cada cidade c e produto p
$S_{c,p}^I$	valor das transações internas para cada cidade c e produto p
$S_{c,p}^B$	valor das compras externas para cada cidade c e produto p
S_c^S	valor das vendas externas sobre todos os produtos para cada cidade c
S_c^I	valor das transações internas sobre todos os produtos para cada cidade c
S_c^B	valor das compras externas sobre todos os produtos para cada cidade c
S_p	valor total estadual transacionado para o produto p para cada camada de comércio
S_t	valor total estadual transacionado para cada camada de comércio
q	parâmetro de deslocamento
Q	complexidade
β_S	expoente de alometria para a rede de vendas
β_I	expoente de alometria para a rede intramunicipal
β_B	expoente de alometria para a rede de compras
$\langle \beta \rangle$	valores médios anuais do expoente alométrico para cada camada de comércio
$\overline{\beta_B}$	valores médios dos 15 produtos mais relevantes anuais do expoente alométrico para cada camada de comércio
C_v	coeficiente de variação para cada camada de comércio
α_S	expoente da lei de Zipf-Mandelbrot para a rede de vendas
α_I	expoente da lei de Zipf-Mandelbrot para a rede intramunicipal
α_B	expoente da lei de Zipf-Mandelbrot para a rede de compras
$\langle \alpha \rangle$	valores médios anuais da lei de Zipf-Mandelbrot para cada camada de comércio
β^F	expoente de alometria do fitness para cada camada de comércio
β^G	expoente de alometria do PIB

SUMÁRIO

1	INTRODUÇÃO	29
2	CIÊNCIA DE DADOS	32
2.1	Análise de Dados e Abordagens Analíticas	34
2.2	Inteligência Artificial (IA)	37
2.3	<i>Machine Learning</i>	40
2.4	Principais Problemas Resolvidos e Métodos de <i>Machine Learning</i>	42
3	MACHINE LEARNING APLICADO A MATERIAIS VISCOELÁSTICOS	47
3.1	Modelo Mesoscópico para Materiais Viscoelásticos	47
3.2	Equação de Langevin	49
3.2.1	<i>Primeira Abordagem</i>	50
3.2.2	<i>Segunda Abordagem</i>	50
3.2.3	<i>Terceira Abordagem</i>	51
3.3	Força de arrasto	52
3.4	Abordagem Estrutural de Materiais Viscoelásticos em Multiescalas	57
3.5	Descrição do Modelo Estrutural de Materiais Viscoelásticos	58
3.5.1	<i>Redes Heterogêneas com Distribuição Gaussiana</i>	61
3.5.2	<i>Redes Homogêneas</i>	61
4	ANÁLISE DE DADOS APLICADO A CIÊNCIA DAS CIDADES	63
4.1	Notas Fiscais Eletrônicas, NF-e	65
4.2	Classificação do Produto	66
4.3	Construção de redes de comércio entre cidades-cidades e entre cidades- produtos	67
4.4	Alometria	70
4.5	<i>Revealed Comparative Advantage (RCA)</i>	70
4.5.1	<i>Distribuições de Zipf-Mandelbrot</i>	72
4.5.2	<i>Binarização RCA</i>	72
4.6	Fitness e Complexidade	73
4.7	Dados Demográficos	75
5	RESULTADOS E DISCUSSÕES	77
5.1	Redes Viscoelásticas Imersas em Fluido	77

5.1.1	<i>Redes Homogêneas</i>	77
5.1.2	<i>Redes Heterogêneas com Distribuição Gaussiana</i>	83
5.1.3	<i>Análise dos dados da estrutura homogênea por meio de leitura e interpretação dos resultados</i>	88
5.1.4	<i>Clusterização das curvas de força associadas aos testes EXP, TR e PL por meio do algoritmo de agrupamento k-means, utilizando técnicas de Machine Learning</i>	90
5.1.5	<i>Lidando com valores ausentes: técnica de imputação baseada no Random Forest Regressor (RFR)</i>	92
5.2	Ciência das Cidades	96
5.2.1	<i>Escalonamento alométrico dos fluxos comerciais municipais</i>	96
5.2.2	<i>Estrutura de Zipf-Mandelbrot das Capacidades Especializadas</i>	100
5.2.3	<i>Fitness e Complexidade</i>	101
5.2.4	<i>Limitações</i>	108
6	CONCLUSÃO	109
	REFERÊNCIAS	112
	APÊNDICES	128
	APÊNDICE A– MODELO NUMÉRICO	128
	APÊNDICE B– ALGUNS RESULTADOS OU PROPRIEDADES DAS TRANSFORMADAS DE LAPLACE	145
	APÊNDICE C– ALGORITMO DO FIT POR MEIO DO MÉTODO DE OTIMIZAÇÃO L-BFGS-B	150
	APÊNDICE D– MÉTODO DE INFERÊNCIA BAYESIANA PARA ESTIMAR EXPOENTES DE ESCALA	153
	APÊNDICE E– RESULTADOS DAS ALOMETRIAS DO FITNESS E PIB EM RELAÇÃO A POLUIÇÃO PARA OS ANOS (2017-2019)	155

1 INTRODUÇÃO

Sistemas complexos aparecem em contextos tão diversos quanto a matéria condensada, os organismos biológicos, os ecossistemas, as cidades e as redes de comércio (Mazzocchi, 2025; Modonesi *et al.*, 2025; LePoire *et al.*, 2025). O traço distintivo desses sistemas não é apenas o grande número de componentes, mas sobretudo a presença de interações distribuídas, não lineares e frequentemente heterogêneas, que produzem padrões coletivos que não são diretamente inferíveis a partir das propriedades individuais das partes (Mazzocchi, 2025; Mitchell, 2009; Newman, 2018; Barabási, 2016; Bentzur *et al.*, 2022; Gallegati *et al.*, 2025; Estrada, 2024). Comportamentos macroscópicos emergem de dinâmicas microscópicas complexas, dando origem a leis de potência, organização hierárquica, transições entre regimes e respostas multiescala que atravessam domínios físicos e sociais (Chen *et al.*, 2015; Milinkovic *et al.*, 2025).

Um dos desafios centrais da ciência contemporânea da complexidade consiste em compreender como tais regularidades emergem, por que são robustas frente a variações microscópicas e quais princípios estruturais sustentam sua organização multiescala. Estudos clássicos e recentes mostram que fenômenos como auto-organização, criticalidade e hierarquia aparecem em sistemas governados por regras microscópicas distintas, sugerindo a existência de mecanismos universais subjacentes (Bak, 1996; Goldenfeld; Kadanoff, 1999; Barabási, 2016; Mitchell, 2009; Sornette, 2006; Stanley *et al.*, 2000; Sieberer *et al.*, 2025; Adediran, 2024).

O avanço recente da Ciência de Dados transformou profundamente a investigação de sistemas complexos (Nawaz *et al.*, 2024; Chakrabarti *et al.*, 2023). A disponibilidade de microdados em larga escala, o avanço de simulações computacionais e o desenvolvimento de métodos estatísticos e de aprendizado de máquina consolidaram um arcabouço metodológico capaz de integrar modelagem teórica, simulação computacional e análise empírica (Bishop, 2006; Hastie *et al.*, 2009; Goodfellow *et al.*, 2016). Técnicas contemporâneas permitem identificar estruturas latentes, inferir parâmetros efetivos, detectar transições de regime e reconstruir dinâmicas coletivas a partir de dados de alta dimensão (Tang *et al.*, 2020; Fabiani, 2024).

Mais do que um conjunto de ferramentas, a Ciência de Dados consolidou-se como uma linguagem metodológica e analítica para descrever sistemas nos quais a complexidade estrutural e dinâmica impede abordagens puramente analíticas. Em particular, métodos de clusterização, redução de dimensionalidade, análise de redes e inferência probabilística têm se mostrado fundamentais para mapear regimes emergentes em sistemas físicos e socioeconômicos (Fabiani, 2024; Bentzur *et al.*, 2022; Dahia; Szabo, 2024).

É nesse horizonte que se insere o presente trabalho. Esta tese investiga dois sistemas complexos de naturezas distintas, um sistema viscoelástico mesoscópico e um sistema urbano-econômico baseado em fluxos de comércio municipal, sob um mesmo eixo metodológico: a inferência de organização macroscópica e de regimes emergentes em sistemas multiescala a partir de dados de alta dimensionalidade (Hramov *et al.*, 2025; Wang *et al.*, 2025b; Muñoz, 2018; Mahnoor *et al.*, 2024).

No primeiro domínio, analisamos um modelo viscoelástico composto por redes elásticas imersas em fluido, cuja dinâmica é descrita por uma equação de Langevin generalizada (Ferreira *et al.*, 2024). A interação entre elasticidade, arrasto e dissipação gera respostas coletivas não triviais, incluindo regimes sublineares de relaxação e comportamentos dependentes da heterogeneidade estrutural. A partir de simulações numéricas sistemáticas e varreduras no espaço de parâmetros, construímos um conjunto extenso de curvas de força e relaxação. Técnicas de clusterização não supervisionada (K-means) foram então empregadas para classificar automaticamente os regimes dinâmicos do material, demonstrando que propriedades macroscópicas podem ser identificadas por meio de padrões latentes extraídos das curvas temporais de força (Ferreira *et al.*, 2024). Essa abordagem evidencia como regimes reológicos emergem da acoplagem entre escalas micro e mesoscópicas (Ferreira *et al.*, 2024; De Araújo *et al.*, 2020; De Sousa *et al.*, 2020; De Sousa *et al.*, 2021; Zhu; Tree, 2023; Sun *et al.*, 2026).

No segundo domínio, investigamos a estrutura econômica do estado do Ceará a partir de microdados de notas fiscais eletrônicas emitidas entre 2016 e 2019. Reconstruímos redes multicamadas de comércio intermunicipal e aplicamos métricas de complexidade econômica, incluindo o método de Fitness (Tacchella *et al.*, 2012; Cristelli *et al.*, 2013; Caldarelli *et al.*, 2012; Operti *et al.*, 2018; Sahasranaman; Jensen, 2025; Servedio *et al.*, 2018), para inferir capacidades produtivas latentes e padrões de especialização (Mealy *et al.*, 2019; Wang *et al.*, 2025a; He; Li, 2025; Gallotti *et al.*, 2021; Bahrami *et al.*, 2023; Altmann, 2020; Zhang *et al.*, 2025b). Assim como no sistema físico, o objetivo central não é apenas descrever os fluxos observados, mas inferir a estrutura subjacente de capacidades e interdependências que organiza o sistema. A análise revela hierarquias produtivas, padrões de *nestedness* e organização multiescala consistentes com sistemas adaptativos complexos (Operti *et al.*, 2018; Straccamore *et al.*, 2023; Angelini *et al.*, 2024; Sahasranaman; Jensen, 2025; Stojkoski *et al.*, 2023; Aufiero *et al.*, 2024).

Apesar de pertencerem a domínios distintos, ambos os sistemas compartilham características estruturais fundamentais: heterogeneidade distribuída, ausência de escala dominante,

emergência de regimes coletivos e organização hierárquica. Em ambos os casos, o problema central consiste em inferir estrutura macroscópica a partir de dados locais, ruidosos, heterogêneos e de alta dimensionalidade. A tese demonstra que métodos de inferência baseados em dados permitem identificar regimes, reconstruir estruturas latentes e revelar regularidades universais que transcendem fronteiras disciplinares.

Assim, o fio condutor deste trabalho não é apenas a aplicação de técnicas de Ciência de Dados a problemas distintos, mas a demonstração de que a inferência de regimes emergentes constitui um princípio unificador na análise de sistemas complexos físicos e socioeconômicos. Ao aproximar materiais viscoelásticos e redes urbanas sob uma mesma lente metodológica, evidenciamos como organização multiescala, transições de regime e hierarquia estrutural podem ser descritas por um arcabouço computacional e estatístico comum.

A organização da tese reflete essa integração conceitual. O Capítulo 2 apresenta os fundamentos metodológicos de Ciência de Dados e de inferência em sistemas complexos. O Capítulo 3 desenvolve o modelo viscoelástico e apresenta os resultados da identificação de regimes dinâmicos, enquanto o Capítulo 4 reconstrói a rede econômica municipal e analisa a complexidade produtiva. O Capítulo 5 reúne, de forma integrada, os resultados e a discussão relativos aos dois domínios de investigação. Por fim, o Capítulo 6 apresenta as conclusões da tese, destacando os princípios unificadores revelados pela inferência orientada por dados em sistemas complexos multiescala.

2 CIÊNCIA DE DADOS

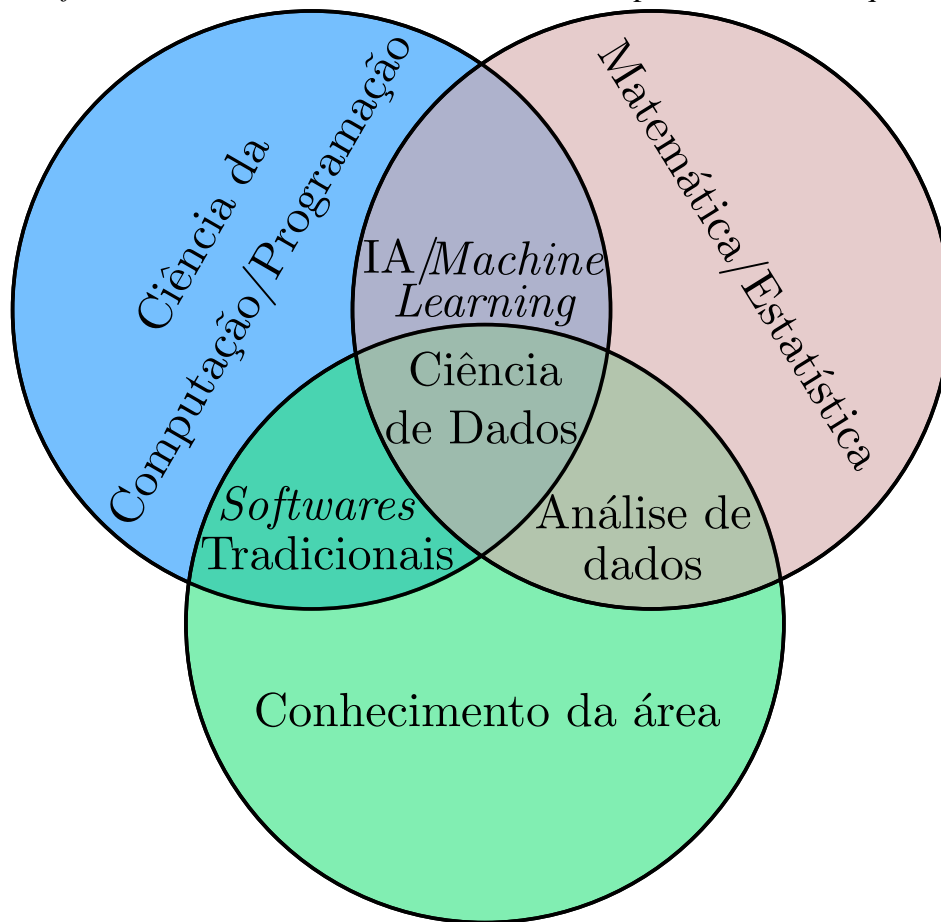
A ciência de dados é uma disciplina interdisciplinar voltada à extração de conhecimento e geração de valor a partir de grandes volumes de dados (*big data*) (Mayernik, 2023). Ela surge como resposta à crescente complexidade dos sistemas informacionais e à necessidade de transformar dados brutos em decisões fundamentadas, especialmente em contextos marcados por incerteza, variabilidade e alta dimensionalidade (Stekhoven; Bühlmann, 2012; Brady, 2019). A ciência de dados combina métodos da estatística, da ciência da computação e do conhecimento de domínio para construir modelos interpretáveis, gerar previsões e apoiar ações estratégicas (Cao, 2017; Provost; Fawcett, 2013a; Dhar, 2013).

Sua consolidação como campo autônomo está diretamente ligada à evolução da análise de dados e à emergência de um paradigma orientado a dados, no qual decisões são guiadas por evidências empíricas extraídas de grandes bases informacionais (Provost; Fawcett, 2013a; Berthold *et al.*, 2020). A ciência de dados não se limita à manipulação técnica de dados, mas envolve também a formulação de hipóteses, a modelagem de fenômenos e a construção de produtos de dados que promovem inteligência organizacional e inovação (Dhar, 2013; Naur, 1974).

A Figura 1 a seguir apresenta uma representação visual da interseção entre os três pilares que sustentam a ciência de dados: ciência da computação, matemática/estatística e conhecimento da área (Ullman, 2020; Conway, 2011; Kubben *et al.*, 2019). O diagrama de Venn destaca como a ciência de dados se posiciona no cruzamento dessas áreas, evidenciando sua natureza híbrida e integradora (Conway, 2011). A sobreposição entre estatística e computação dá origem ao aprendizado de máquina e à IA (Inteligência Artificial), enquanto a interseção entre estatística e conhecimento de domínio representa a análise de dados/pesquisa tradicional. Já a combinação entre computação e conhecimento da área está associada ao uso de *softwares* convencionais. A ciência de dados, portanto, emerge como o ponto de convergência entre essas três dimensões, incorporando técnicas computacionais, rigor estatístico e sensibilidade contextual.

A definição de ciência de dados enfatiza seu papel como uma disciplina orientada por dados, com foco na descoberta de conhecimento, na compreensão de fenômenos complexos e na construção de sistemas inteligentes (Cao, 2017; Brodie, 2024; Brady, 2019). Essa abordagem difere da estatística tradicional, que se concentra na inferência a partir de amostras, e da ciência da computação clássica, que prioriza a construção de algoritmos e estruturas de dados (Efron;

Figura 1 – O diagrama de Venn mostra como ciência da computação, matemática/estatística e conhecimento da área se sobrepõem para formar a ciência de dados, destacando as regiões dedicadas a *softwares* tradicionais, análise de dados e IA/aprendizado de máquina.



Fonte: Elaborado pelo autor.

Hastie, 2016). A ciência de dados, por sua vez, busca integrar essas competências para resolver problemas reais, muitas vezes mal definidos, utilizando dados como principal insumo (Cao, 2017).

Outro aspecto central é o ciclo de vida dos dados, que compreende etapas como aquisição, armazenamento, pré-processamento, análise, modelagem, interpretação e comunicação dos resultados (Berthold *et al.*, 2020; Provost; Fawcett, 2013a; Provost; Fawcett, 2013b). Cada uma dessas fases exige competências específicas e ferramentas adequadas, o que reforça o caráter multidisciplinar da área (Mayernik, 2023). A qualidade dos dados, sua relevância para o problema em questão e a clareza dos objetivos analíticos são fatores determinantes para o sucesso de qualquer projeto de ciência de dados (Berthold *et al.*, 2020; Provost; Fawcett, 2013a).

Por fim, destaca-se o conceito de produto de dados, entendido como qualquer entrega baseada em dados que agrega valor ao processo decisório (Loukides, 2011; Loukides, 2018). Esses produtos podem assumir diversas formas, desde *dashboards* interativos e relatórios

automatizados até modelos preditivos e sistemas de recomendação. O valor de um produto de dados está diretamente relacionado à sua capacidade de gerar conhecimento, promover ação e sustentar decisões com base em evidências (Dhar, 2013; Cao, 2017).

A ciência de dados segue crescendo com o desenvolvimento da inteligência artificial e modelos de linguagem como os LLMs (*Large Language Models*), que aumentam a capacidade de análise e entendimento dos dados (Gao *et al.*, 2025; Battineni *et al.*, 2026; Wang *et al.*, 2026). Além disso, a ética na ciência de dados torna-se cada vez mais crucial, com discussões sobre privacidade, transparência e o uso responsável dos dados (Anastasiadou, 2026; Kenig *et al.*, 2026; Leslie, 2019; Kelleher; Tierney, 2018; Fairfield; Shtein, 2014). A tendência observada na literatura indica que a ciência de dados tende a se tornar cada vez mais central para decisões estratégicas.

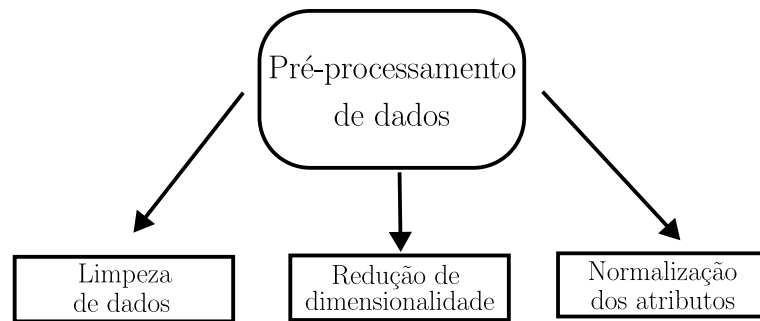
2.1 Análise de Dados e Abordagens Analíticas

A análise de dados constitui o núcleo operacional da ciência de dados, sendo responsável por transformar dados brutos em informações estruturadas, padrões relevantes e conhecimento útil para a tomada de decisão (Dash *et al.*, 2023; Brandt, 2014). Trata-se de um processo iterativo que envolve a exploração, transformação, modelagem e interpretação dos dados, com o objetivo de compreender fenômenos, apoiar decisões e resolver problemas complexos. Esse processo frequentemente serve de base para a construção de modelos mais sofisticados, como os utilizados em aprendizado de máquina.

O processo analítico é composto por diversas etapas fundamentais, que se articulam de forma sequencial e cíclica. Entre elas, destacam-se a coleta de dados, o pré-processamento e a análise exploratória, que são descritas a seguir.

A coleta de dados representa a etapa inicial e consiste na aquisição de informações relevantes a partir de fontes estruturadas, semiestruturadas ou não estruturadas. Essas fontes podem incluir bancos de dados relacionais, APIs (*Application Programming Interfaces*), sensores, arquivos públicos, *web scraping* e sistemas internos de informação (Cordeiro *et al.*, 2025; Brandt, 2014). A qualidade da coleta impacta diretamente a validade das análises, sendo comum o enfrentamento de desafios como viés de seleção, incompletude, heterogeneidade de formatos e questões éticas relacionadas à privacidade e à proteção de dados pessoais (Mezmir, 2020; O'Neil, 2016), especialmente em contextos regulados por legislações como a LGPD (Lei Geral de Proteção de Dados Pessoais) ou o GDPR (*General Data Protection Regulation*) (Voigt; von

Figura 2 – Técnicas de pré-processamento de dados.



Fonte: Elaborado pelo autor.

dem Bussche, 2024; Paula *et al.*, 2025). Ferramentas como Python e SQL (*Structured Query Language*) são amplamente utilizadas para automatizar esse processo, permitindo o acesso e a integração de múltiplas fontes de dados (Scikit-Learn Developers, 2017; Grus, 2015; Grippa; Kuzmichev, 2021).

Após a coleta, os dados brutos geralmente apresentam ruídos, valores ausentes, inconsistências e escalas incompatíveis. O pré-processamento é a etapa que prepara os dados para análise, garantindo que estejam limpos, coerentes e adequados ao objetivo do estudo. Essa fase pode envolver a imputação de valores faltantes, a remoção de duplicatas, a padronização de formatos e a transformação de variáveis. Técnicas como normalização e padronização são aplicadas para ajustar as escalas das variáveis e melhorar o desempenho de diversos modelos, embora o tratamento de *outliers* frequentemente exija procedimentos específicos (James *et al.*, 2013; Maharana *et al.*, 2022). Além disso, a redução de dimensionalidade, por meio de métodos como PCA (*Principal Component Analysis*), permite simplificar o conjunto de atributos, removendo ou combinando variáveis redundantes ou pouco informativas, o que pode melhorar a generalização e reduzir o custo computacional (James *et al.*, 2013; Maharana *et al.*, 2022; Wickham; Grolemond, 2017). A Figura 2 ilustra os principais eixos do pré-processamento de dados, destacando a importância da limpeza, da normalização e da simplificação das variáveis.

Com os dados preparados, inicia-se a análise exploratória, também conhecida como EDA (*Exploratory Data Analysis*). Essa etapa visa compreender a estrutura dos dados, identificar padrões, tendências, anomalias e relações entre variáveis (Tukey, 1977). A exploração começa com medidas descritivas como média, moda, variância, erro padrão, mediana, desvio padrão e quantis, que oferecem uma visão inicial da centralidade e da dispersão dos valores observados (Harbison; Simmons, 2024; Efron; Hastie, 2016). Em seguida, visualizações gráficas como histogramas, boxplots, *scatterplots* e mapas de calor (*heatmaps*) são utilizadas para revelar

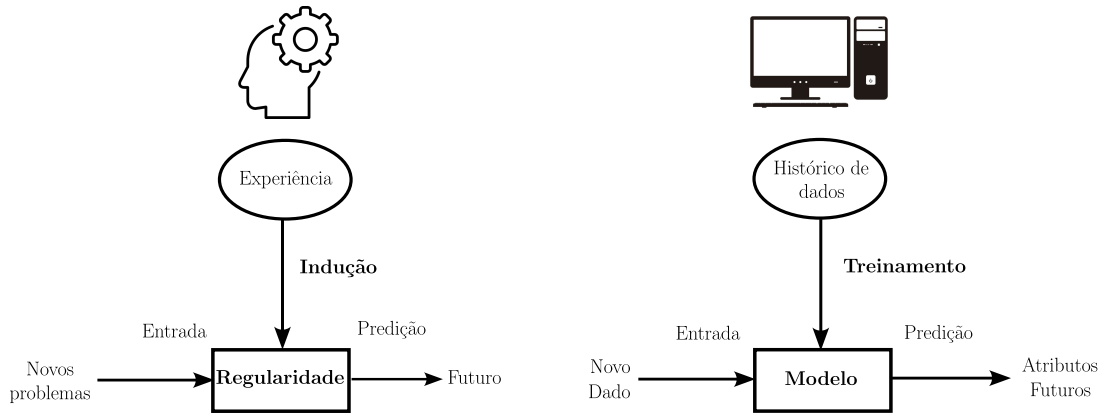
padrões mais profundos. Os histogramas ajudam a avaliar a forma da distribuição, enquanto os boxplots destacam a variabilidade e possíveis extremos. Os *scatterplots* permitem explorar relações bivariadas, e os mapas de calor evidenciam correlações em matrizes multidimensionais, facilitando a detecção de tendências espaciais ou temporais (Fisher; M., 2017; Unwin, 2015; James *et al.*, 2013).

A análise de correlações aprofunda essa compreensão ao quantificar a força e a direção das associações entre variáveis. Coeficientes como o de Pearson são utilizados para quantificar associações lineares entre variáveis, especialmente quando se pretende empregar procedimentos inferenciais baseados em pressupostos paramétricos, enquanto o de Spearman é mais apropriado quando o interesse recai sobre relações monotônicas, especialmente em contextos nos quais não se deseja assumir linearidade ou em que os dados podem ser tratados em termos de postos (James *et al.*, 2013; Harbison; Simmons, 2024; Chen *et al.*, 2016; Tu *et al.*, 2025). Essa distinção é importante para evitar inferências incorretas e garantir que os padrões identificados sejam robustos mesmo em contextos com violação de suposições estatísticas clássicas (Chen *et al.*, 2016; Efron; Hastie, 2016). Complementando esse processo, a detecção de *outliers* é fundamental para garantir que anomalias não distorçam as conclusões gerais da análise (Dash *et al.*, 2023). Métodos baseados em distância, como a distância de Mahalanobis, são úteis em espaços multivariados, enquanto procedimentos como o teste de Grubbs podem ser empregados na detecção de valores extremos em distribuições univariadas, desde que suas hipóteses de aplicação sejam razoavelmente atendidas (Kannan; Manoj, 2015).

A literatura especializada também classifica as abordagens analíticas em cinco categorias principais, que refletem diferentes níveis de complexidade e profundidade. A análise descritiva busca responder à pergunta “o que aconteceu?”, oferecendo uma visão panorâmica dos dados. A análise diagnóstica procura entender “por que algo aconteceu?”, investigando associações, dependências e possíveis fatores explicativos entre variáveis (Cao, 2017; Berthold *et al.*, 2020). A análise preditiva visa antecipar eventos futuros com base em dados históricos, utilizando modelos estatísticos e algoritmos computacionais (Cao, 2017; Dhar, 2013). Já a análise prescritiva propõe ações específicas a serem tomadas com base nas previsões realizadas, incluindo simulações e otimizações. Por fim, a análise avançada (*deep analytics*) busca compreender os mecanismos subjacentes aos fenômenos observados, integrando múltiplas fontes de dados e técnicas analíticas (Cao, 2017).

Essas abordagens não são mutuamente exclusivas, mas complementares. Em pro-

Figura 3 – Similaridade entre a inteligência humana e artificial.



Fonte: Elaborado pelo autor.

jetos reais, é comum que se inicie com uma análise descritiva, avance para a diagnóstica e, posteriormente, incorpore elementos preditivos e prescritivos. A escolha da abordagem depende do problema a ser resolvido, da disponibilidade de dados e dos objetivos analíticos. Além disso, a análise de dados deve ser conduzida com rigor metodológico, respeitando princípios como reprodutibilidade, transparência e ética (Anastasiadou, 2026; Leslie, 2019). A interpretação dos resultados deve considerar o contexto em que os dados foram gerados, evitando generalizações indevidas, extrapolações não justificadas ou conclusões precipitadas. A comunicação clara dos achados é parte essencial do processo, garantindo que os *insights* gerados sejam compreendidos e utilizados de forma eficaz (Berthold *et al.*, 2020).

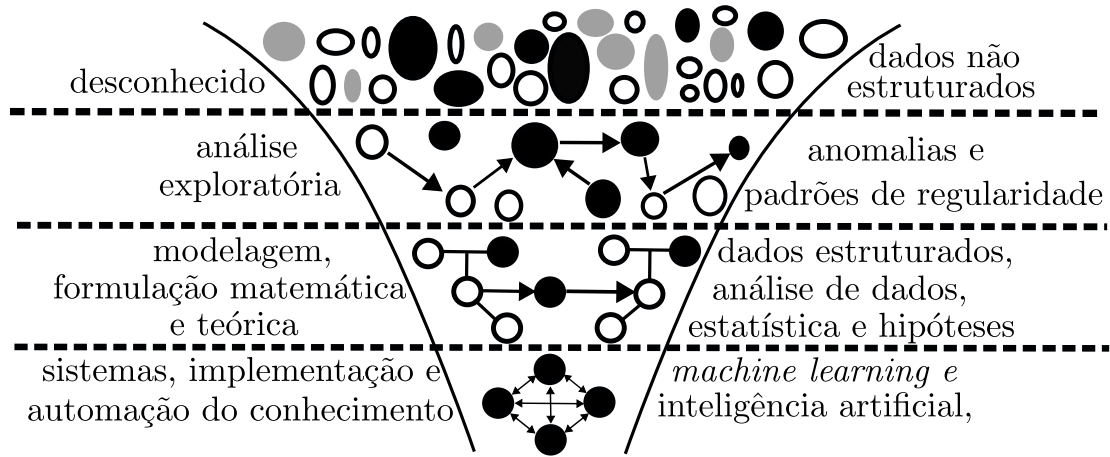
A análise de dados, portanto, é a base sobre a qual se constroem modelos mais avançados, como os utilizados em inteligência artificial e aprendizado de máquina, que serão explorados nas seções seguintes.

2.2 Inteligência Artificial (IA)

A inteligência artificial constitui um campo da ciência da computação composto por diversos subcampos inter-relacionados, cujo objetivo é desenvolver sistemas capazes de executar tarefas que normalmente exigiriam inteligência humana (Russell; Norvig, 2020). De forma geral, esses sistemas são projetados para analisar informações, identificar padrões em dados, resolver problemas e apoiar processos de tomada de decisão.

Na Figura 3, é possível perceber alguns pontos em comum entre a inteligência humana e a artificial. Ambas dependem da capacidade de processar informações, tomar decisões, resolver problemas e executar tarefas específicas. Apesar dessas semelhanças, os sistemas

Figura 4 – Representação esquemática do fluxo de transformação dos dados, ilustrando as etapas que vão da informação bruta à modelagem matemática e à construção de sistemas inteligentes. O diagrama evidencia o processo analítico que sustenta técnicas de aprendizado de máquina e aprendizado profundo, mostrando como padrões, relações e estruturas emergem progressivamente ao longo do tratamento dos dados.



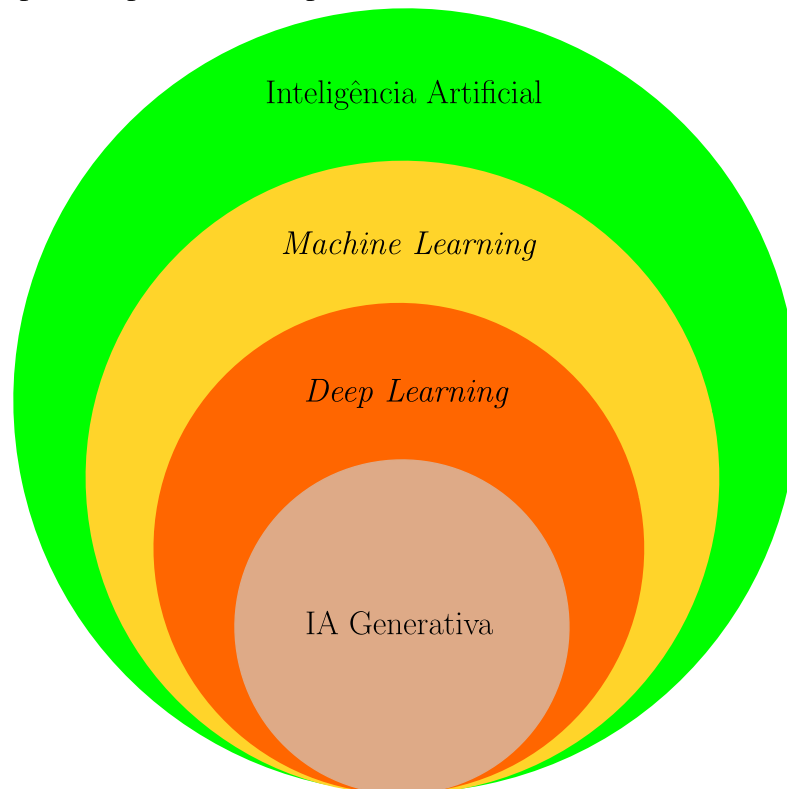
atuais de IA não apresentam compreensão genuína do mundo, nem atributos como consciência, emoções ou intuição, características associadas à inteligência humana. Além disso, a maior parte das aplicações contemporâneas de IA é altamente especializada, sendo projetada para executar tarefas específicas, especialmente no tratamento e análise de grandes conjuntos de dados a partir dos quais os modelos podem extrair padrões e realizar previsões.

Dentro desse contexto, destaca-se o aprendizado de máquina, ou ML (*Machine Learning*), área da inteligência artificial dedicada ao desenvolvimento de algoritmos capazes de aprender padrões a partir de dados e aprimorar seu desempenho à medida que são expostos a novas informações. Essa abordagem tem se mostrado essencial em diversas aplicações modernas, como sistemas de reconhecimento de voz, análise e interpretação de imagens e mecanismos de recomendação de conteúdos personalizados (Bishop, 2006; Goodfellow *et al.*, 2016).

Dentro do campo do aprendizado de máquina, destaca-se o aprendizado profundo (*Deep Learning*, DL), que utiliza redes neurais profundas compostas por múltiplas camadas hierárquicas de representação (Liu *et al.*, 2018). Esse tipo de abordagem possibilita que o sistema identifique padrões complexos e extraia características cada vez mais abstratas diretamente dos dados, utilizando arquiteturas inspiradas, em um nível abstrato, em princípios de processamento distribuído (Goodfellow *et al.*, 2016).

A Figura 4 ilustra, de forma esquemática, o percurso que transforma o desconhecido em conhecimento aplicado dentro do ciclo analítico da ciência de dados e da inteligência artificial. Na parte superior, aparecem elementos dispersos e heterogêneos, representando dados

Figura 5 – A imagem apresenta um diagrama que organiza visualmente a relação entre diferentes áreas da Inteligência Artificial (IA), destacando onde se posicionam o *Machine Learning* (ML), o *Deep Learning* (DL) e a IA Generativa. As regiões aparecem dispostas em camadas concêntricas, evidenciando que o Deep Learning constitui um subconjunto do Machine Learning, enquanto ambos pertencem ao domínio mais amplo da IA. O diagrama também mostra a IA Generativa como uma área específica dentro do Deep Learning, indicando seu foco na produção de novos dados a partir de padrões previamente aprendidos.



Fonte: Elaborado pelo autor.

não estruturados e informações ainda caóticas. À medida que esses elementos descem pelo funil conceitual, tornam-se gradualmente mais organizados: primeiro passam por uma etapa de análise exploratória, na qual padrões iniciais, correlações visuais e possíveis anomalias começam a emergir. Em seguida, a figura mostra a fase de modelagem e formulação matemática, simbolizada por estruturas mais articuladas e conectadas por relações funcionais. Na base do diagrama, encontram-se representações de redes e sistemas, expressando o estágio em que o conhecimento adquirido é formalizado, implementado e automatizado por métodos de *machine learning* e inteligência artificial. Assim, a imagem sintetiza o fluxo de transformação dos dados desde sua forma bruta até sua incorporação em sistemas capazes de aprender, inferir e generalizar.

A IA generativa, um ramo mais especializado dentro do *Deep Learning*, dedica-se à criação de novos conteúdos — como textos, imagens ou outros tipos de dados — a partir de informações previamente analisadas (Zhou; Lee, 2024). O diferencial dessa abordagem está na capacidade de produzir resultados sem comandos explícitos para cada tarefa, explorando padrões

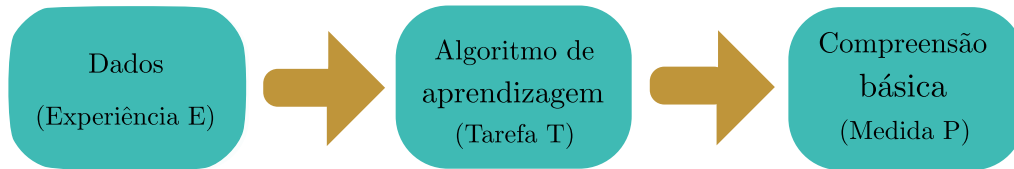
aprendidos em grandes volumes de dados por meio de técnicas de aprendizado supervisionado e não supervisionado. Essa competência em gerar novas informações amplia significativamente as possibilidades de inovação, abrangendo desde aplicações criativas até soluções práticas em variados setores industriais (Messer, 2024; Kusiak, 2024). Já na Figura 5, observa-se uma representação hierárquica que organiza as diversas camadas e dimensões que compõem o ecossistema da inteligência artificial.

2.3 *Machine Learning*

Esta seção apresenta os conceitos essenciais de *Machine Learning*, abordando sua classificação, o funcionamento geral dos processos que o compõem e os algoritmos mais utilizados. Na abordagem tradicional da ciência da computação, o desenvolvimento de soluções depende de linguagens de programação nas quais o programador especifica explicitamente as instruções que o computador deve seguir para executar uma tarefa específica. Nesse modelo clássico, grande parte do comportamento do programa é explicitamente especificada pelo programador. Em contraste, o aprendizado de máquina adota uma lógica diferente: em vez de codificar regras fixas, os algoritmos são projetados para permitir que o sistema aprenda diretamente a partir de dados e de certas suposições iniciais, tornando possível a execução de tarefas que não estão rigidamente detalhadas no código. De maneira análoga a processos de aprendizagem baseados em experiência, cujo aprendizado envolve tentativas, erros e um grande número de exemplos observados ao longo do treinamento, os modelos de ML também dependem de processos sucessivos de treinamento e ajuste para melhorar seu desempenho (Rodrigues *et al.*, 2021). Para que um modelo computacional adquira conhecimento, ele precisa receber informações estruturadas ou ser capaz de extrair padrões a partir dos dados de treinamento, refinando sua capacidade de tomada de decisão conforme se expõe a novos dados (Young *et al.*, 2018). Esse processo é o que caracteriza o aprendizado de máquina (Liu *et al.*, 2020). De acordo com a definição clássica proposta por Mitchell (Mitchell, 1997), considera-se que um programa aprendeu quando, após ser submetido a uma experiência (E), apresenta melhoria mensurável em uma classe de tarefas (T), segundo algum critério de desempenho (P); em síntese, o aprendizado ocorre quando o desempenho em T, avaliado por P, evolui como consequência direta da experiência adquirida.

A Figura 6 apresenta o fluxo geral do processo de *Machine Learning*, no qual os dados, correspondentes à Experiência (E), representam o conjunto de informações ou exemplos utilizados por um sistema computacional para aprender e aprimorar seu desempenho em uma ta-

Figura 6 – Trata-se da sequência de elementos essenciais que descrevem como um programa de computador aprende a partir de dados para realizar uma tarefa específica, aprimorando seu desempenho à medida que desenvolve compreensão sobre o problema.



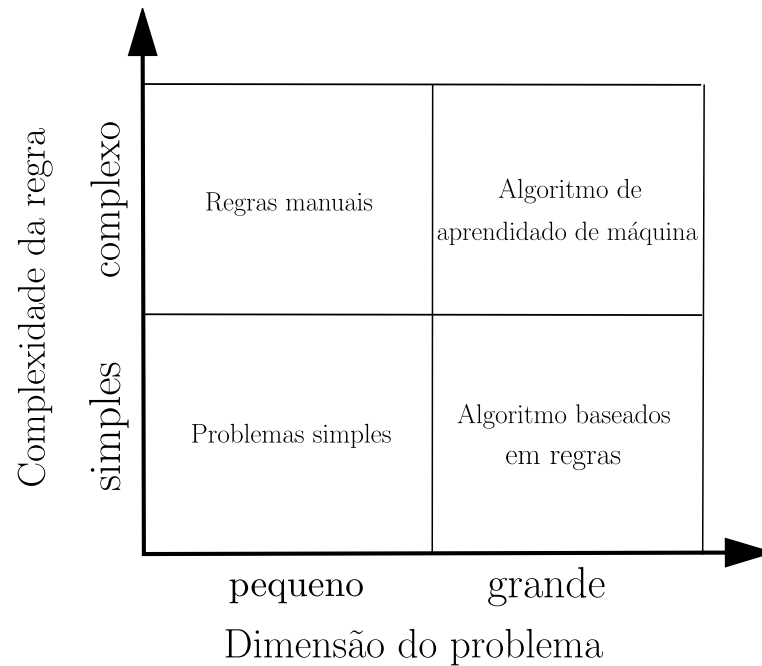
Fonte: Elaborado pelo autor.

refa específica. Os algoritmos de aprendizagem, associados à Tarefa (T), consistem em conjuntos de regras e procedimentos que orientam como o programa deve processar os dados e identificar padrões, permitindo a execução de funções como classificação, regressão ou agrupamento. Por fim, a Compreensão Básica, relacionada à Medida de Desempenho (P), corresponde aos critérios utilizados para avaliar o quanto o sistema está aprendendo e quão bem realiza a tarefa proposta, podendo envolver métricas como precisão, acurácia, erro médio ou outras medidas adequadas ao objetivo do modelo.

Na Figura 7, observa-se que a solução de um problema por meio de aprendizado de máquina pode variar em complexidade dependendo de como os dados são preparados e de quais resultados se deseja alcançar. Em muitos casos, o desafio envolve lidar com grandes volumes de informações sem uma distribuição claramente definida, o que torna mais difícil interpretar as análises e extrair conclusões assertivas. Nesse contexto, o aprendizado de máquina torna-se útil em diferentes cenários, como em situações em que as regras do problema são tão complexas que não podem ser descritas explicitamente, como ocorre no reconhecimento de voz; quando as regras evoluem ao longo do tempo, como na marcação de partes da fala, onde surgem constantemente novos termos e significados; ou ainda quando a distribuição dos dados se altera continuamente, exigindo a adaptação constante dos modelos, como na previsão de tendências de vendas de *commodities*.

Na prática do aprendizado de máquina, exemplos bem-sucedidos e malsucedidos desempenham um papel fundamental na adaptação e no refinamento do comportamento dos algoritmos, razão pela qual as etapas de coleta, seleção e preparação dos dados são consideradas cruciais. O avanço expressivo da capacidade computacional, aliado ao crescimento contínuo de grandes volumes de dados (*big data*), tornou possível implementar abordagens que antes eram inviáveis (Rodrigues *et al.*, 2021; Schleder; Fazzio, 2021). Esse cenário favorece o treinamento dos modelos, já que agora há disponibilidade tanto de dados suficientes quanto de infraestrutura

Figura 7 – Cenários de aplicações do aprendizado de máquina.



Fonte: Elaborado pelo autor.

computacional para sustentar processos de aprendizado mais complexos. Especialistas da área ressaltam que, em muitos casos, a preparação dos dados demanda mais tempo e esforço do que o próprio desenvolvimento dos algoritmos, dada sua importância para garantir resultados confiáveis e consistentes (Zhang; Lu, 2021).

2.4 Principais Problemas Resolvidos e Métodos de *Machine Learning*

O aprendizado de máquina baseia-se no princípio de que um algoritmo pode melhorar seu desempenho conforme é exposto a novos dados (Nilsson, 2014). De modo geral, os problemas tratados nessa área incluem tarefas como regressão, classificação e agrupamento (clusterização), além de técnicas analíticas amplamente utilizadas, como redução de dimensionalidade e extração de regras de associação (Erhan *et al.*, 2010; Géron, 2019). No caso da regressão, o objetivo é estimar uma variável resposta contínua a partir de um conjunto de variáveis de entrada, por meio de uma função do tipo $f : \mathbb{R}^n \rightarrow \mathbb{R}$, sendo amplamente aplicada, por exemplo, na previsão de preços de imóveis com base em características e tendências históricas (Nunes *et al.*, 2019). Já na classificação, busca-se identificar a qual de k categorias uma observação pertence, utilizando uma função que mapeia os dados de entrada para uma classe específica, representada por $f : \mathbb{R}^n \rightarrow \{1, 2, \dots, k\}$. Esse tipo de modelo tem grande relevância em áreas como filtragem

automática de mensagens, auxiliando na detecção de *spam* em serviços de e-mail (Hastie *et al.*, 2009), e também na saúde, apoiando sistemas de diagnóstico baseados em informações clínicas (Theerthagiri *et al.*, 2023).

Nos modelos de agrupamento (clusterização), o objetivo é organizar grandes conjuntos de dados não rotulados formando grupos que reúnem elementos com características semelhantes, ou seja, itens com maior similaridade entre si do que em relação a elementos de outros grupos. Esse tipo de abordagem é amplamente empregado em tarefas como categorização automática de documentos (Boley *et al.*, 1999) e na análise de interações em redes sociais (Pham *et al.*, 2011). Já a redução de dimensionalidade visa simplificar os dados ao combinar ou projetar variáveis em um espaço de menor dimensionalidade, preservando apenas as informações essenciais para a análise. Métodos como a Análise de Componentes Principais (PCA) permitem condensar o número de variáveis, facilitando a visualização, a análise e o processamento computacional, além de potencialmente melhorar o desempenho de alguns modelos de aprendizado (Hasan; Abdulazeez, 2021).

As regras de associação têm como propósito identificar padrões frequentes de coocorrência entre itens ou elementos presentes em grandes conjuntos de dados, utilizando algoritmos como *Apriori*, *Eclat* e *FP-Growth* (Srinadh, 2022). Esse tipo de técnica é especialmente relevante em análises de comportamento de compra, nas quais se avalia, por exemplo, a frequência com que clientes que compram o produto A também compram o produto B. Além de revelar padrões importantes de consumo, essas regras fornecem suporte para estratégias de marketing, definição de perfis de clientes e construção de sistemas de recomendação, contribuindo para uma experiência de compra mais personalizada e eficiente (Boley *et al.*, 1999).

Com relação à forma como os métodos de aprendizado são organizados, o aprendizado de máquina é frequentemente classificado em quatro categorias principais: aprendizado supervisionado, não supervisionado, semi-supervisionado e aprendizado por reforço (Géron, 2019; Conway; White, 2011; IGI Global; Information Resources Management Association, 2012). No caso do aprendizado supervisionado, o modelo é treinado a partir de dados rotulados, nos quais cada entrada possui uma saída associada. A partir desses exemplos, o algoritmo aprende a estabelecer a relação entre entradas e respostas esperadas, tornando-se capaz de realizar previsões sobre novos dados. Essa abordagem é amplamente utilizada em tarefas de classificação e regressão, já que a correspondência entre variáveis de entrada e saída facilita o processo de modelagem. No campo da segurança digital, técnicas supervisionadas são aplicadas

na identificação de mensagens fraudulentas, contribuindo para a detecção de mensagens de spam, aprimorando a filtragem e proteção de comunicações eletrônicas (Hastie *et al.*, 2009).

No aprendizado não supervisionado, não há disponibilidade de rótulos ou saídas previamente definidas, de modo que o algoritmo analisa os dados buscando estruturas internas, padrões recorrentes ou agrupamentos naturais sem rótulos explícitos associados às observações externa. Esse tipo de abordagem engloba métodos como técnicas de clusterização, procedimentos de redução de dimensionalidade e extração de regras de associação. Um exemplo relevante de aplicação é o uso de clusterização hierárquica na análise de dados de microarranjos de DNA, recurso que possibilita a identificação de padrões de expressão gênica, a separação de grupos funcionais de genes e a caracterização de respostas celulares diante de diferentes condições ambientais (Eisen *et al.*, 1998).

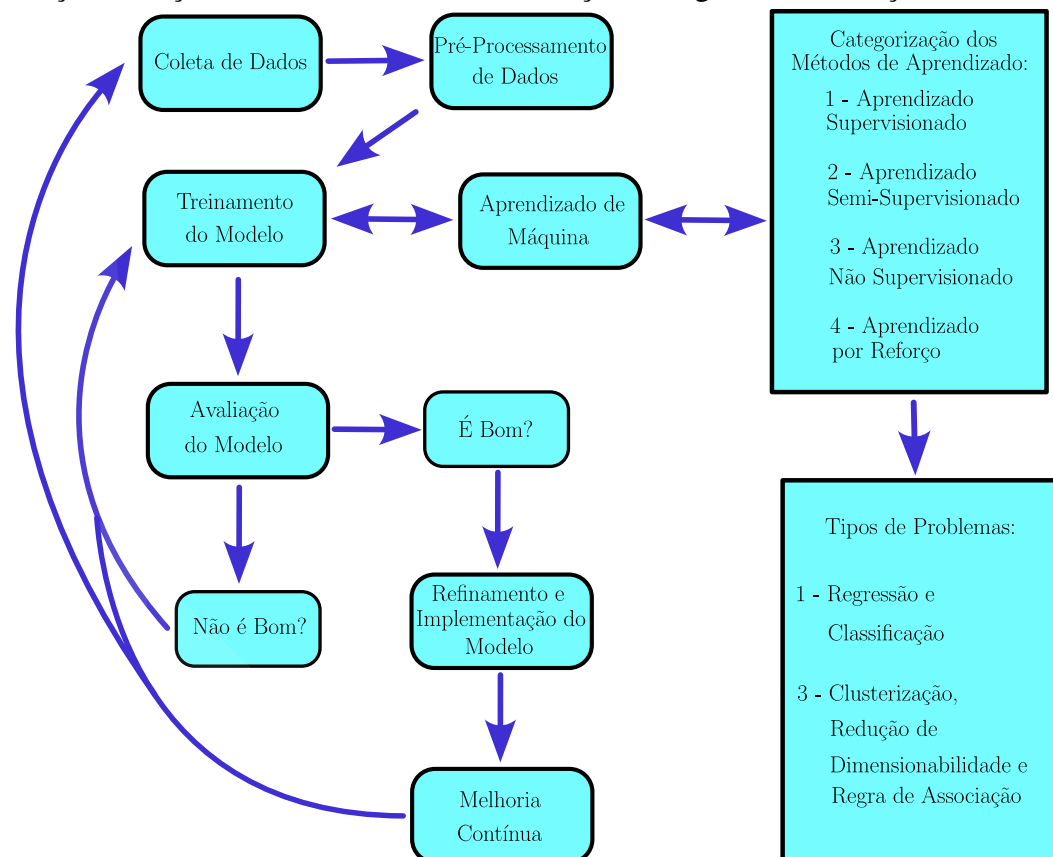
No aprendizado semi-supervisionado, trabalha-se com conjuntos de dados em que apenas uma pequena parte das amostras possui rótulos, enquanto a maior parcela permanece sem rotulagem. Esse tipo de abordagem é particularmente útil porque combina informações rotuladas e não rotuladas para aprimorar o desempenho do modelo, reduzindo a necessidade de grandes volumes de dados anotados, que muitas vezes são caros ou difíceis de obter. Entre as técnicas utilizadas, destaca-se a técnicas de propagação de rótulos em grafos, na qual estruturas baseadas em grafos permitem que as informações das amostras rotuladas sejam transmitidas para as não rotuladas, possibilitando alcançar resultados de classificação robustos mesmo quando poucos rótulos estão disponíveis (Zhu; Ghahramani, 2002).

Por fim, o aprendizado por reforço concentra-se em treinar modelos que tomam decisões de forma sequencial, aprendendo por meio de tentativa e erro quais ações conduzem às maiores recompensas ao longo do tempo. Esse tipo de abordagem se mostra particularmente eficaz em cenários dinâmicos e imprevisíveis, como a movimentação autônoma de robôs ou jogos que exigem estratégias complexas. Um exemplo emblemático dessa abordagem é o sistema AlphaGo, sistema que combinou técnicas avançadas de aprendizado por reforço, redes neurais profundas e algoritmos de busca em árvore que alcançou desempenho super-humano no jogo de tabuleiro Go, conhecido por sua elevada complexidade e pela enorme espaço de estados possíveis (Silver *et al.*, 2016).

A Figura 8 apresenta um fluxograma detalhado que representa todas as etapas envolvidas no desenvolvimento de modelos de *Machine Learning*, começando pela aquisição dos dados e avançando até o processo de melhoria contínua após a implantação. O ciclo se inicia na

fase de coleta, momento em que informações relevantes são reunidas a partir de diferentes fontes, sendo fundamental que esses dados possuam boa qualidade e estejam alinhados ao problema a ser solucionado.

Figura 8 – Fluxograma que representa o processo de aprendizado de máquina, destacando etapas fundamentais como a coleta e o pré-processamento dos dados, o treinamento do modelo, a avaliação, o refinamento e a implementação. O processo é contínuo e inclui ciclos sucessivos de avaliação e melhoria. Também são apresentadas as categorias de métodos de aprendizado — supervisionado, semi-supervisionado, não supervisionado e por reforço — juntamente com os principais tipos de problemas que podem ser abordados, como regressão, classificação, clusterização, redução de dimensionalidade e extração de regras de associação.



Fonte: Elaborado pelo autor a partir de dados de Ferreira (2022) e Géron (2019).

Após a coleta, realiza-se o pré-processamento dos dados para preparar os dados para análise, englobando atividades como limpeza, normalização e tratamento de *outliers* (Dash *et al.*, 2023), de modo a garantir que as informações estejam em condições adequadas para uma modelagem eficiente. Com o conjunto de dados devidamente estruturado, inicia-se o treinamento do modelo, fase em que algoritmos de aprendizado de máquina são aplicados para construir sistemas capazes de realizar previsões ou distinguir padrões relevantes.

Após o treinamento, o modelo passa pela etapa de avaliação, fase essencial para analisar seu desempenho e verificar se atende as métricas de desempenho previamente definidas.

Os resultados dessa análise determinam o caminho seguinte: caso o desempenho fique abaixo do aceitável, o fluxo retorna a etapas anteriores para novos ajustes, que podem envolver desde o pré-processamento até o próprio treinamento ou, se necessário, a ampliação do conjunto de dados. Quando o modelo apresenta resultados satisfatórios, ele segue para a fase de refinamento e posterior implementação, momento em que são aplicados ajustes finais para garantir seu funcionamento otimizado antes da utilização prática (Ferreira, 2022; Géron, 2019).

A etapa de melhoria contínua ressalta a necessidade de monitorar regularmente o desempenho do modelo, promovendo atualizações sempre que novos dados se tornem disponíveis ou quando ocorrerem alterações no ambiente de operação, assegurando assim que o sistema permaneça preciso e funcional ao longo do tempo (Ferreira, 2022; Géron, 2019). O fluxograma também organiza as principais categorias de métodos de aprendizado de máquina, abrangendo desde técnicas supervisionadas até abordagens não supervisionadas e baseadas em reforço, além de indicar diferentes classes de problemas que podem ser tratados, como clusterização, redução de dimensionalidade, extração de regras de associação e tarefas de classificação.

Neste trabalho, o foco concentra-se em técnicas supervisionadas e não supervisionadas, com ênfase em modelos de classificação e de clusterização aplicados à identificação de diferentes padrões de relaxação em materiais viscoelásticos. Esses métodos possibilitam, respectivamente, a criação de modelos apoiados em dados rotulados e a descoberta de estrutura e agrupamentos em conjuntos não rotulados. Além disso, são realizadas análises envolvendo grandes volumes de informações oriundas de notas fiscais eletrônicas, inseridas no contexto de *big data*.

3 MACHINE LEARNING APLICADO A MATERIAIS VISCOELÁSTICOS

3.1 Modelo Mesoscópico para Materiais Viscoelásticos

Comportamentos puramente elásticos e puramente viscosos representam casos limite das equações constitutivas de materiais (Lakes, 2017). Substâncias reais podem se deformar e escoar, mas um desses atributos geralmente domina o outro, dependendo das condições aplicadas. Essa dualidade sólido-líquido intriga pesquisadores desde pelo menos o século XIX. Naquela época, pioneiros como James Maxwell e Lord Kelvin propuseram modelos analíticos baseados em associações em série e paralelo de molas e amortecedores para explicar as características peculiares observadas em seda, fibras de vidro e fios de aço (Maxwell, 1867; Markovitz, 1977).

A resposta efetiva desses primeiros modelos invariavelmente apresenta decaimentos de relaxação exponenciais, independentemente da forma como molas e amortecedores são conectados. No entanto, essas abordagens simples são adequadas apenas para alguns materiais viscoelásticos atuais (De Sousa *et al.*, 2021). Nas últimas décadas, os materiais de matéria mole tornaram-se objeto de grande interesse, impulsionados tanto pelo desenvolvimento de materiais sintéticos quanto pelo avanço de técnicas experimentais que permitem analisar sistemas antes inacessíveis, como células vivas (De Sousa *et al.*, 2020; De Sousa *et al.*, 2021). Materiais de matéria mole são aqueles que podem ser facilmente deformados por forças relativamente fracas (como pequenas variações de temperatura, pressão ou campos elétricos) e que apresentam um grau de organização intermediário entre um sólido cristalino e um líquido comum (De Sousa *et al.*, 2020; De Sousa *et al.*, 2021).

O surgimento de materiais tão complexos impulsionou novos modelos teóricos e o aprimoramento de técnicas experimentais adequadas para explicar e controlar suas propriedades viscoelásticas (De Sousa *et al.*, 2020; De Sousa *et al.*, 2021; Carvalho *et al.*, 2021; Ramo *et al.*, 2018; Allen-Perkins *et al.*, 2017). Métodos de nanoindentação, como a Microscopia de Força Atômica, tornaram-se essenciais para caracterizar viscoelásticas em escalas micro e nanométricas, sondando materiais com indentadores de tamanho nanométrico (De Sousa *et al.*, 2017). A forma como a força de contato varia ao longo do tempo revela assinaturas reológicas do material. A caracterização de materiais viscoelásticos busca determinar a função de relaxação $R(t)$, que contém informações tanto qualitativas quanto quantitativas.

Funções de relaxação exponenciais e de lei de potência são os dois principais tipos de respostas investigadas experimentalmente. Géis de poliacrilamida (Song *et al.*, 2017; Calvet *et*

al., 2004) e soluções aquosas de surfactantes catiônicos (Rehage; Hoffmann, 1988), por exemplo, apresentam resposta do tipo exponencial, $R(t) \sim e^{-t/\tau}$, com um tempo de relaxação τ para que o material alcance uma nova configuração de equilíbrio. Por outro lado, células vivas (Efremov *et al.*, 2017; Wolff *et al.*, 2012), dispersões de microgéis (Ketz *et al.*, 1988), materiais vítreos moles (Sollich, 1998), géis de colágeno (Shayegan; Forde, 2013) e hidrogéis (Larson, 1999) apresentam um comportamento temporal do tipo lei de potência, $R(t) \sim t^{-\beta}$. Ambos os tipos de dinâmica de relaxação foram observados em coloides funcionalizados sobre substratos atrativos (Dias *et al.*, 2016; De Araújo *et al.*, 2017). Como observado em materiais elásticos (Moreira *et al.*, 2012; Oliveira *et al.*, 2014), parâmetros físicos macroscópicos estão intrinsecamente conectados às suas interações microscópicas, estruturas e escalas de tempo (Achar; Hanneken, 2009; Yucht *et al.*, 2013; Milkus; Zaccone, 2017).

Leis de potência e exponenciais surgem em muitos fenômenos físicos, tendo uma origem profunda em seus processos dinâmicos. Por exemplo, em sistemas descritos por entropias não extensivas, muitas variáveis físicas são descritas por distribuições de lei de potência em vez das tradicionais funções exponenciais presentes na entropia correspondente (Tsallis, 1988; Tsallis, 2009; Souza *et al.*, 2018). Distribuições canônicas exponenciais e de lei de potência emergem naturalmente dependendo da capacidade térmica do reservatório térmico ser constante ou divergir (Almeida, 2001). Além disso, leis de potência estão associadas a fenômenos emergentes, nos quais seus expoentes exibem comportamentos de escala próximos à criticalidade (Stanley, 1971; Seybold *et al.*, 2021). Sistemas com os mesmos expoentes críticos pertencem à mesma classe de universalidade, e um pequeno conjunto de classes de universalidade descreve quase todas as transições de fase materiais.

Um dos desafios da ciência dos materiais é conectar os mecanismos físicos em escalas microscópicas ao comportamento funcional macroscópico. Essa abordagem é especialmente relevante para a matéria mole, porque propriedades na escala molecular estão ligadas a flutuações conformacionais e composicionais nas escalas nanométrica e micrométrica e, além disso, abrangem muitas ordens de magnitude em comprimento (Praprotnik *et al.*, 2008; Qu *et al.*, 2011). A matéria mole possui estruturas ricas e diversas interações na mesoescala, onde a energia térmica por unidade de volume é negligenciável, em contraste com a alta densidade de energia armazenada nas ligações atômicas de estruturas cristalinas (Doi, 2013). Enquanto materiais de comportamento exponencial podem ser modelados por uma associação de molas e amortecedores, como o chamado modelo sólido linear padrão, materiais que exibem leis

de potência geralmente são descritos por reologia fracionária (West *et al.*, 2003; Jaishankar; McKinley, 2013) ou por modelos de reologia vítrea (Fabry *et al.*, 2001; Fabry *et al.*, 2003). No entanto, esses modelos não explicam a conexão entre as respostas macroscópicas e seus componentes elásticos e viscosos. Abordagens multiescala podem ajudar a revelar as relações constitutivas do material considerando suas interações e composições em escalas menores.

Aqui, desenvolvemos um modelo de materiais viscoelásticos composto por uma rede elástica de macromoléculas imersa em um fluido, a fim de estudar como interações mesoscópicas influenciam o comportamento reológico macroscópico em materiais moles (De Araújo *et al.*, 2020). Assumimos que forças de arrasto hidrodinâmico não lineares atuam entre a macromolécula e o fluido, onde a contribuição das interações elásticas e viscosas é controlada em nível mesoscópico. Ao alterar os parâmetros físicos das forças elásticas e de arrasto, obtemos materiais com relaxações exponenciais, de lei de potência ou um comportamento intermediário para respostas que não são claramente caracterizadas. Nossos resultados mostram que o comportamento exponencial é, de fato, o regime de deformação mais comum, sendo descrito pelo modelo sólido linear padrão (De Araújo *et al.*, 2020). Respostas de lei de potência são resultados excepcionais que surgem para um intervalo particular de forças de arrasto sublineares.

3.2 Equação de Langevin

Langevin desenvolveu um modelo para partículas coloidais em escalas mesoscópicas, onde a massa das partículas coloidais é significativamente maior que a massa atômica das partículas do fluido, ou seja, $m_{coloidal} \gg m_{fluido}$. Apesar dessa discrepância de massa, as partículas coloidais ainda seguem a distribuição de Maxwell. As frequentes colisões com as partículas do fluido geram uma força aleatória $\xi(t)$, sem direção preferencial (Tomé; de Oliveira, 2015; Kheifets, 2014; Langevin, 1908; Lemons; Gythiel, 1997).

A força resultante das colisões com as partículas do fluido apresenta variações rápidas, permitindo a suposição de que a média da força aleatória é zero. Além disso, assumimos que a correlação dessa força é modelada por uma função Delta de Dirac, indicando um ruído branco, ou seja, ausência de correlação entre impactos em diferentes momentos. Os efeitos da força aleatória são descritos pelos seus dois primeiros momentos,

$$\overline{\xi(t)} = 0, \quad (3.1)$$

$$\overline{\xi(t)\xi(t')} = \Gamma\delta(t-t'), \quad (3.2)$$

onde $\overline{\xi(t)}$ é a média sobre diferentes realizações da força aleatória $\xi(t)$.

3.2.1 Primeira Abordagem

Inicialmente, Langevin considerou somente a força aleatória $\xi(t)$ resultante das colisões das partículas com o banho térmico, conforme expresso pela segunda Lei de Newton

$$m \frac{dv}{dt} = \xi(t). \quad (3.3)$$

Calculando a média sobre todas as realizações, encontramos que $\overline{v(t)} = \overline{v_0}$, onde $\overline{v(t)}$ representa a média sobre diferentes realizações da velocidade.

Definindo $\langle v(t) \rangle$ como a média sobre todas as diferentes velocidades iniciais do sistema, com a suposição de que o sistema esteja inicialmente em equilíbrio, a velocidade média das partículas segue a distribuição de Maxwell-Boltzmann

$$\langle v(t) \rangle = \langle v_0 \rangle = 0. \quad (3.4)$$

Elevando ao quadrado e calculando a média sobre todas as realizações, obtemos

$$\overline{v^2(t)} = v_0^2 + \frac{\Gamma}{m^2} t. \quad (3.5)$$

Calculando a média sobre todas as condições iniciais, encontramos

$$\langle v^2(t) \rangle = \frac{k_B T}{m} + \frac{\Gamma}{m^2} t. \quad (3.6)$$

Esse resultado demonstra que a energia cinética aumenta com o tempo, indicando que a primeira abordagem é incorreta.

3.2.2 Segunda Abordagem

Nesta abordagem, Langevin aplicou a Lei de Newton somente à força de arrasto, expressa como $f_a = \gamma v$. A equação diferencial associada é formulada como

$$m \frac{dv}{dt} = -\gamma v \implies \int_{v_0}^v \frac{dv}{v} = - \int_0^t \frac{\gamma}{m} dt = \ln \left(\frac{v}{v_0} \right) = -\frac{\gamma}{m} t \implies v(t) = v_0 e^{-\frac{\gamma}{m} t}. \quad (3.7)$$

Calculando a média sobre todas as realizações em v_0 com a mesma condição inicial, encontramos

$$\overline{v(t)} = v_0 e^{-\frac{\gamma}{m} t}. \quad (3.8)$$

Calculando a média sobre todas as condições iniciais, obtemos

$$\langle v(t) \rangle = 0. \quad (3.9)$$

Elevando ao quadrado e calculando a média sobre todas as diferentes condições iniciais, temos

$$\langle v^2(t) \rangle = \langle v_0^2 e^{-\frac{2\gamma}{m}t} \rangle = \frac{k_B T}{m} e^{-\frac{2\gamma}{m}t}. \quad (3.10)$$

Essa expressão revela que a energia cinética diminui exponencialmente com o tempo, levando o sistema a não atingir o equilíbrio termodinâmico. Portanto, a segunda perspectiva também é incorreta.

3.2.3 Terceira Abordagem

Nesta abordagem, Langevin combinou a força de arrasto $f_a = \gamma v$ com a força de flutuação $\xi(t)$. A Lei de Newton é expressa como

$$m \frac{dv}{dt} = -\gamma v + \xi(t). \quad (3.11)$$

Resolvendo passo a passo, obtemos a solução de Langevin

$$\begin{aligned} \left(\frac{dv}{dt} + \frac{\gamma}{m} v = \frac{1}{m} \xi(t) \right) e^{\frac{\gamma}{m}t} &\implies \left(\frac{dv}{dt} + \frac{\gamma}{m} v \right) e^{\frac{\gamma}{m}t} = \frac{1}{m} \xi(t) e^{\frac{\gamma}{m}t} \\ &= \frac{d(v e^{\frac{\gamma}{m}t})}{dt} = \frac{1}{m} \xi(t) e^{\frac{\gamma}{m}t} \implies v(t) = v_0 e^{-\frac{\gamma}{m}t} + \frac{e^{-\frac{\gamma}{m}t}}{m} \int_0^t \xi(t') e^{\frac{\gamma}{m}t'} dt'. \end{aligned} \quad (3.12)$$

Calculando a média sobre todas as realizações em v_0 com a mesma condição inicial, encontramos

$$\overline{v(t)} = v_0 e^{-\frac{\gamma}{m}t}. \quad (3.13)$$

Calculando a média sobre todas as condições iniciais, obtemos

$$\langle v(t) \rangle = 0. \quad (3.14)$$

Calculando a média sobre todas as realizações para $\overline{v^2(t)}$, obtemos

$$\begin{aligned}
\overline{v^2(t)} &= v_0^2 e^{-\frac{2\gamma t}{m}} + \frac{e^{-\frac{2\gamma t}{m}}}{m^2} \int_0^t dt' \int_0^t dt'' \underbrace{\overline{\xi(t')\xi(t'')}}_{\Gamma \delta(t''-t')} e^{\frac{\gamma t'}{m}} e^{\frac{\gamma t''}{m}} + \frac{2v_0}{m} e^{-\frac{2\gamma t}{m}} \underbrace{\int_0^t \overline{\xi(t')} e^{\frac{\gamma t'}{m}} dt'}_0 \\
&= v_0^2 e^{-\frac{2\gamma t}{m}} + \frac{\Gamma e^{-\frac{2\gamma t}{m}}}{m^2} \int_0^t e^{\frac{2\gamma t'}{m}} dt' = v_0^2 e^{-\frac{2\gamma t}{m}} + \frac{\Gamma e^{-\frac{2\gamma t}{m}}}{m^2} \left[\frac{m}{2\gamma} \left(e^{\frac{2\gamma t}{m}} - 1 \right) \right] \\
&= v_0^2 e^{-\frac{2\gamma t}{m}} + \frac{\Gamma}{2\gamma m} \left(1 - e^{-\frac{2\gamma t}{m}} \right) = \left(v_0^2 - \frac{\Gamma}{2\gamma m} \right) e^{-\frac{2\gamma t}{m}} + \frac{\Gamma}{2\gamma m}.
\end{aligned} \tag{3.15}$$

Calculando a média sobre todas as condições iniciais sobre $\langle v^2(t) \rangle$, obtemos

$$\langle v^2(t) \rangle = \left(\langle v_0^2 \rangle - \frac{\Gamma}{2\gamma m} \right) e^{-\frac{2\gamma t}{m}} + \frac{\Gamma}{2\gamma m} = \left(\frac{k_B T}{m} - \frac{\Gamma}{2\gamma m} \right) e^{-\frac{2\gamma t}{m}} + \frac{\Gamma}{2\gamma m}, \tag{3.16}$$

conforme antecipado, ao longo de um intervalo de tempo suficientemente longo, o sistema tende a alcançar o equilíbrio termodinâmico. Assim, podemos usar o teorema da equipartição $\langle v^2(t) \rangle_{eq} = \frac{k_B T}{m}$, tal que

$$\langle v^2(t) \rangle_{eq} = \lim_{t \rightarrow \infty} \langle v^2(t) \rangle = \frac{\Gamma}{2\gamma m} = \frac{k_B T}{m} \implies \Gamma = 2\gamma k_B T. \tag{3.17}$$

A equação mencionada é significativa, pois estabelece uma relação importante entre as flutuações no sistema, representadas pelo fator Γ , e os processos de dissipação no sistema, representados pelo coeficiente de arrasto γ . Essa relação representa o teorema da flutuação-dissipação. Dessa forma, a terceira abordagem é a correta. Portanto, podemos escrever a equação de Langevin representada por

$$m \frac{d^2 \vec{r}}{dt^2} = -\gamma \vec{v} + \vec{\xi}(t). \tag{3.18}$$

3.3 Força de arrasto

Na compreensão dos materiais viscoelásticos, uma abordagem mesoscópica é indispensável, especialmente ao considerar as forças que encapsulam propriedades tanto elásticas quanto viscosas. Esta metodologia é crucial para representar adequadamente a complexidade inerente a tais materiais, que exibem características combinadas da elasticidade sólida e da viscosidade líquida. A integração de termos de potencial tanto conservativo quanto não-conservativo é essencial para a modelagem macroscópica do comportamento material, influenciado pelas interações microscópicas entre as partículas constituintes.

Neste contexto, o aspecto viscoso ressalta a variação na energia do sistema devido à presença de forças dissipativas, responsáveis pela conversão da energia mecânica em outras formas, como calor e deformação. A força de arrasto \vec{f}_a , atuante em partículas em movimento através de um fluido, é um exemplo clássico de força dissipativa. No regime viscoso, esta força segue a lei de Stokes, sendo proporcional à velocidade da partícula (Batchelor, 2000),

$$\vec{f}_a = -\gamma\vec{v}. \quad (3.19)$$

Aqui, γ é o coeficiente de arrasto, definido como $\gamma = 3\pi\rho\eta$ para partículas esféricas com diâmetro ρ , onde η é a viscosidade do fluido e \vec{v} é a velocidade da partícula no fluido. Esta formulação é comumente aplicada em estudos de movimento browniano (Coffey; Kalmykov, 2012; Maiocchi, 1990).

A força de arrasto \vec{f}_a para corpos rígidos em movimento rápido através de um fluido, ou quando experimentam um escoamento intenso do fluido, é expressa pela seguinte equação

$$\vec{f}_a = -\frac{1}{2}C_a\phi Av^2\hat{v}, \quad (3.20)$$

onde C_a representa o coeficiente de arrasto, dependente das características do objeto e de sua geometria, ϕ é a densidade do fluido, A é a área de referência frontal do objeto exposta ao fluxo do fluido e v é a velocidade do objeto em relação ao fluido. É compreendido que objetos flexíveis exibem comportamento distinto em comparação com objetos rígidos devido às deformações e interações envolvidas. Esse comportamento torna-se mais evidente em situações turbulentas, exigindo considerações adicionais (Chapman *et al.*, 2015).

Notavelmente, em situações envolvendo materiais viscoelásticos, cujo comportamento de deformabilidade é proeminente, a expressão 3.20 torna-se inviável (Vasco; Maciel, 2007). Em virtude disso, a adaptação da equação pode ser realizada, substituindo C_a por C_M , refletindo assim as características flexíveis dos materiais na hidrodinâmica/aerodinâmica (Verschoren *et al.*, 2016). Essa modificação visa abordar adequadamente o comportamento viscoelástico do material, introduzindo C_M como uma alternativa apropriada e adaptada à equação 3.20, tal que

$$\vec{f}_a = -\frac{1}{2}C_M\phi A\frac{A_0}{A_0}v^2\hat{v}, \quad (3.21)$$

na equação acima, $C_M = C_a\left(\frac{v}{v_a}\right)^{\alpha-2}$, em que A_0 representa a área de contato antes que o material sofra deformação devido ao fluxo, ou seja, quando está imerso no fluido com $v = 0m/s$. A é a área após o material sofrer deformação, v_a é a velocidade de referência utilizada para dimensionar o

coeficiente de arrasto C_M e α é o expoente da velocidade (Verschoren *et al.*, 2016). Verschoren considera essa velocidade de referência como $v_a = 1m/s$ (Verschoren *et al.*, 2016), enquanto Champtom conduz estudos com diferentes valores de v_a (Chapman *et al.*, 2015).

Incorporando dois elementos para representar a força de arrasto de materiais flexíveis, introduzimos um fator geométrico $\beta_a = \frac{A}{A_0}$ e um fator de coeficiente de arrasto $\alpha_a = (\frac{v}{v_a})^{\alpha-2}$. O fator de geometria β_a é empregado para levar em conta o impacto das variações na área, enquanto o fator de coeficiente de arrasto α_a é introduzido para examinar como a resistência de arrasto é influenciada pelas forças que atuam sobre o material flexível, podendo esta estar deformada ou reorganizada. Com isso, podemos reescrever a equação 3.21 da seguinte maneira

$$\vec{f} = -\frac{\vec{f}_a}{A_0} = \frac{1}{2}C_a\alpha_a\phi\beta_av^2\hat{v}, \quad (3.22)$$

podemos notar que quando os valores de $\alpha_a = 1$ e $\beta_a = 1$, a equação resultante é equivalente à equação 3.20, que é aplicável a corpos indeformáveis (Chapman *et al.*, 2015).

O expoente α fornece uma indicação útil de como a relação entre as mudanças no arrasto e a velocidade é influenciada pela reconfiguração. No entanto, não representa uma medida direta da força de arrasto real que atua sobre o material. Na prática, as previsões da força de arrasto são frequentemente realizadas utilizando a lei quadrática padrão (Luhar; Nepf, 2013). Entretanto, é importante observar que a área e o coeficiente de arrasto podem variar com a velocidade.

Vogel propôs que a força de arrasto, representada como $f_a \sim v^{2+\mu}$, fosse proporcional à velocidade ao quadrado, incluindo um expoente μ , conhecido como o expoente de Vogel (Vogel, 1989). Este expoente adicional é uma variável relevante em estudos aplicados à botânica. Ao compararmos com a equação 3.22, observamos que o expoente de Vogel é equivalente a $\mu = \alpha - 2$. Isso leva a força de arrasto a ser proporcional à velocidade com um expoente α , que é um parâmetro calibrado para considerar os efeitos da flexibilidade do material na resistência ao arrasto.

As origens físicas do regime sublinear ($\alpha < 1$), no entanto, têm naturezas completamente distintas e estão associadas à deformabilidade/adaptabilidade de corpos submetidos a forças de arrasto (Vogel, 1994). Em outras palavras, à medida que o valor de α diminui, a flexibilidade e/ou a natureza hidrodinâmica/aerodinâmica da amostra aumenta. Isso está relacionado ao regime de deformação, indicando que valores menores de α correspondem a materiais mais

flexíveis, enquanto valores maiores indicam um material mais rígido (regime elástico)(Luhar; Nepf, 2013; Vasco; Maciel, 2007; Ferreira, 2022; Ferreira *et al.*, 2024). Essa relação exerce um impacto direto na capacidade dos materiais de resistir às tensões externas e internas, por exemplo pela capacidade das células (materiais flexíveis) de resistirem a deformações de $10^2 KPa$, em contraste com a resistência de uma viga (material sólido) de $10^{11} GPa$ (Moeendarbary; Harris, 2014).

Muitos organismos vivos apresentam comportamentos de arrasto sublinear, onde a deformabilidade é uma estratégia de sobrevivência para proteger suas estruturas frágeis sob condições hidrodinâmicas. A caracterização dos expoentes de arrasto em sistemas deformáveis é uma prática comum em campos como botânica, aerodinâmica e hidrodinâmica (Vogel, 1989; Favier *et al.*, 2009; Jensen, 2003; Gosselin *et al.*, 2010; Chapman *et al.*, 2015). Expoentes de arrasto típicos para algas, por exemplo, podem ser tão baixos quanto 0.34 (Vogel, 1994). Por outro lado, plantas como tulipa, salgueiro, *ilex opalas* (*branch*) e *pinus taeda* (1 m de altura), que são mais rígidas, exibem expoentes próximos à unidade, com valores de $\alpha = 0.82$, $\alpha = 0.94$ (Vogel, 1989), $\alpha = 0.90$ e $\alpha = 0.87$ (Vogel, 1984), respectivamente.

Além disso, animais como pássaros adotam estratégias de voo que aumentam a velocidade e reduzem o consumo de energia durante os deslocamentos. Suas penas, por exemplo, se auto-adaptam durante o voo, contribuindo para um desempenho aerodinâmico otimizado (Favier *et al.*, 2009; Aldheeb *et al.*, 2016). Outro exemplo são as células, que possuem citoesqueleto e citoplasma apresentando interações elásticas e viscosas. Após sofrerem deformações, essas células se auto-adaptam ao meio. Consequentemente, essas estruturas complexas, que utilizam a deformabilidade para a autoadaptação ao meio, estão diretamente vinculadas ao expoente α . Este expoente α , por sua vez, relaciona-se diretamente com as deformações mesoscópicas dos constituintes viscoelásticos do material (Ferreira, 2022; Ferreira *et al.*, 2024).

Quando ocorre deformação no material, sua habilidade de manter o volume constante é preservada, embora sua geometria seja alterada. Segundo Gosselin, foram conduzidos experimentos para redução de arrasto em três placas flexíveis com geometrias distintas devido à reconfiguração. A análise revelou que, mesmo sob condições de fluxo idênticas, o expoente α apresentava valores distintos, em relação aos materiais como as placas retangulares, discos cortados em setores e discos rolando em cones apresentarem valores experimentais $\alpha = 0.60$, $\alpha = 0.70$ e $\alpha = 0.90$, respectivamente (Gosselin *et al.*, 2010).

As estruturas rígidas das asas em aeronaves, enquanto robustas, apresentam limi-

tações notáveis, incluindo a falta de adaptabilidade dinâmica às variadas condições de voo, restrições na otimização da eficiência aerodinâmica e dificuldades em ajustar-se a variações na carga. Diante desses obstáculos, um campo novo de pesquisa e inovação destaca as asas flexíveis como uma solução promissora. A natural flexibilidade dessas asas possibilita ajustes dinâmicos em pleno voo, aprimorando de forma contínua a aerodinâmica e a eficiência energética (Chin *et al.*, 2020). A melhoria no desempenho de elevação das asas flexíveis decorre da interação entre a deformação da membrana e a estrutura do fluido, resultando em maior deformação e vibração mais intensa (Guo *et al.*, 2021). Medições indicam que as asas flexíveis geram forças aerodinâmicas menores do que as asas rígidas em condições de estado estacionário (Zhao *et al.*, 2010). Além disso, essas asas flexíveis respondem eficazmente às variações de carga, reduzindo significativamente vibrações e turbulências, ao mesmo tempo que promovem uma notável adaptabilidade em diferentes fases do voo.

Quando objetos rígidos estão expostos a um fluxo com um número de Reynolds moderadamente alto, isso sugere cenários nos quais o arrasto seria proporcional ao quadrado da velocidade. Entretanto, para materiais rígidos perfeitamente hidrodinâmicos/aerodinâmicos, os valores do expoente α são inferiores a 2. Isso é exemplificado pela superfície plana (rígida) paralela ao fluxo, que possui um expoente $\alpha = 1.5$, um corpo aerodinâmico (fluxo laminar) com $\alpha = 1.5$ e um cilindro (eixo normal ao fluxo) com $\alpha = 1.61$ (Vogel, 1994). Outro exemplo notável é a aplicação da Lei de Stokes, onde geralmente se assume que a esfera é rígida (Batchelor, 2000). O experimento consiste em uma esfera pequena caindo lentamente em um fluxo viscoso a baixa velocidade e baixo número de Reynolds, sendo projetado para ocorrer em um regime de fluxo laminar ou de viscosidade. Nesse cenário, as camadas de fluido movem-se suavemente em caminhos paralelos, indicando que o fluido flui de forma contínua sobre a superfície da esfera à medida que esta se desloca pelo meio, resultando em efeitos hidrodinâmicos e um valor $\alpha = 1.0$. Assim, na ausência de flexibilidade (deformação), o expoente α pode resultar em valores inferiores a 2 para corpos rígidos, dependendo dos efeitos aerodinâmicos ou hidrodinâmicos do material.

No contexto de nosso estudo, a força de arrasto generalizada para materiais viscoelásticos é simplificada por

$$\vec{f}_a = -\gamma v^\alpha \hat{v}, \quad (3.23)$$

com γ representando o coeficiente de arrasto adaptado, incorporando todas as constantes relevantes. É importante notar que $\alpha = 0$ relaciona-se a forças de atrito constantes, $\alpha = 1$ a um regime

viscoso e $\alpha = 2$ a um regime turbulento. A unidade reduzida do coeficiente de arrasto A , γ^* , é definida como

$$\gamma^* = \frac{\rho^{\alpha+1}}{\epsilon t_0^\alpha} \gamma. \quad (3.24)$$

Esta formulação é fundamental para capturar as propriedades dinâmicas e a resposta de materiais viscoelásticos sob condições variadas, refletindo a interação complexa entre suas propriedades mecânicas inerentes e as forças externas atuantes.

3.4 Abordagem Estrutural de Materiais Viscoelásticos em Multiescalas

A complexidade de unir as escalas microscópicas e mesoscópicas com as macroscópicas na reologia dos materiais é um desafio teórico significativo. Essa complexidade surge da necessidade de entender como interações e características em escalas menores influenciam as propriedades observadas em escalas maiores. Esta abordagem, profundamente relevante em campos como polímeros, fluidos complexos e materiais compósitos, busca elucidar os mecanismos fundamentais que regem o comportamento reológico dos materiais e desenvolver modelos preditivos precisos (Praprotnik *et al.*, 2008; Qu *et al.*, 2011).

A modelagem multiescala é uma técnica inovadora que busca simular o comportamento dos materiais, integrando informações de escalas finas (microscópicas ou mesoscópicas) a uma escala macroscópica contínua. Esta metodologia inclui uma ampla gama de fenômenos, desde interações atômicas e moleculares até comportamentos de fases ou grãos em uma escala maior (Roters *et al.*, 2010; Fish *et al.*, 2021).

O desafio desses modelos está em mapear variáveis contínuas, como deformação ou temperatura, para fenômenos em escalas menores. Em sistemas heterogêneos, esta tarefa se torna mais complexa, principalmente ao abordar grandes deformações ou falhas materiais, que podem envolver processos históricos (Fish *et al.*, 2021). O desenvolvimento destes modelos se baseia na compreensão de que certas suposições ou leis constitutivas em uma escala podem não ser adequadas em outra, exigindo uma abordagem mais detalhada e baseada em princípios físicos. Métodos como o multigrid, baseados em homogeneização, têm se mostrado eficazes ao lidar com o comportamento oscilatório de sistemas heterogêneos (Fish; Belsky, 1995a; Fish; Belsky, 1995b; Miehe; Bayreuther, 2007).

Os métodos multiescala são classificados em duas categorias principais: *upscaling* e *resolved-scale*. Os métodos de *upscaling*, também conhecidos como métodos hierárquicos

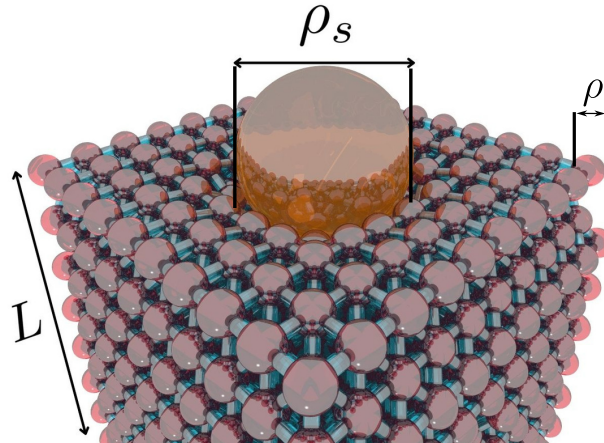
ou de homogeneização computacional, buscam unir escalas discretas e contínuas. Eles se subdividem em abordagens baseadas em matemática, física e dados. Cada abordagem fornece uma perspectiva única para integrar informações de uma escala a outra, seja por meio de equações derivadas, princípios físicos ou análise de dados obtidos em simulações de escalas finas (Fish *et al.*, 2021).

3.5 Descrição do Modelo Estrutural de Materiais Viscoelásticos

Nossa abordagem modela uma rede *Face-Centered-Cubic* (FCC) com dimensões $x \times y \times z$ dadas por $\rho L \sin(\pi/3) \times \rho L \sin(\pi/3) \times \rho L$, onde L é o número de camadas na direção z , composta por N partículas esféricas, cada uma com diâmetro ρ e massa m , conforme ilustrado na Figura 9. Cada partícula interage com seus doze vizinhos mais próximos por meio de um potencial elástico com uma constante de mola efetiva k . A rede de molas está imersa em um fluido viscoso, onde forças de arrasto atuam sobre as partículas em movimento. Nesta abordagem de descrição *coarse-grained* — isto é, uma modelagem em que cada partícula representa não átomos individuais, mas blocos maiores da estrutura, como segmentos de cadeias poliméricas ou agregados coloidais — as partículas podem representar macromoléculas presentes em cadeias poliméricas suspensas, agregações coloidais e outras estruturas responsáveis por suportar carga em materiais moles.

Realizamos experimentos computacionais de indentação para investigar as propriedades viscoelásticas dessa rede. Durante o processo, um indentador esférico rígido pressiona a rede a uma taxa constante, seguido por uma fase de reorganização da rede. Limitamos a profundidade máxima da indentação para evitar efeitos não lineares a menos de 10% da altura da rede (Garcia; Garcia, 2018), aplicando assim $\delta_{max} \simeq \rho$. O movimento de cada partícula na rede é descrito por uma equação de movimento, que inclui tanto a interação elástica com outras partículas e o indentador quanto a força de arrasto generalizada. Esta última é particularmente significativa, pois sua magnitude e direção são influenciadas pela velocidade da partícula e pelas propriedades do fluido circundante. A força de arrasto incorpora as contribuições viscosas e representa o modelo reológico do sólido linear padrão, assumindo um valor fixo de $\alpha = 1$ (De Araújo *et al.*, 2020). Além disso, abordamos redes heterogêneas com distribuição gaussiana das constantes elásticas, de forma que os valores de k se concentram em torno de um valor médio bem definido, permitindo analisar as propriedades mecânicas do sistema. Estas representações permitem modelar materiais com variações na rigidez em sua estrutura, oferecendo um panorama

Figura 9 – Uma rede elástica cúbica de altura L encontra-se imersa em um fluido (não mostrado na figura). Cada partícula possui um diâmetro ρ e está interligada por molas de constante elástica k_{kl} com suas 12 vizinhas mais próximas. A rede é submetida à indentação por uma ponta esférica de diâmetro ρ_s de cima para baixo.



Fonte: Elaborado pelo autor.

mais realista das propriedades viscoelásticas em comparação com redes homogêneas, onde uma única constante elástica é utilizada para todas as ligações.

A equação de movimento da partícula k -ésima, localizada na posição \vec{r}_k , é dada pela seguinte equação

$$m \frac{d^2 \vec{r}_k}{dt^2} = -\nabla U_k - \gamma v_k^\alpha \hat{v}_k, \quad (3.25)$$

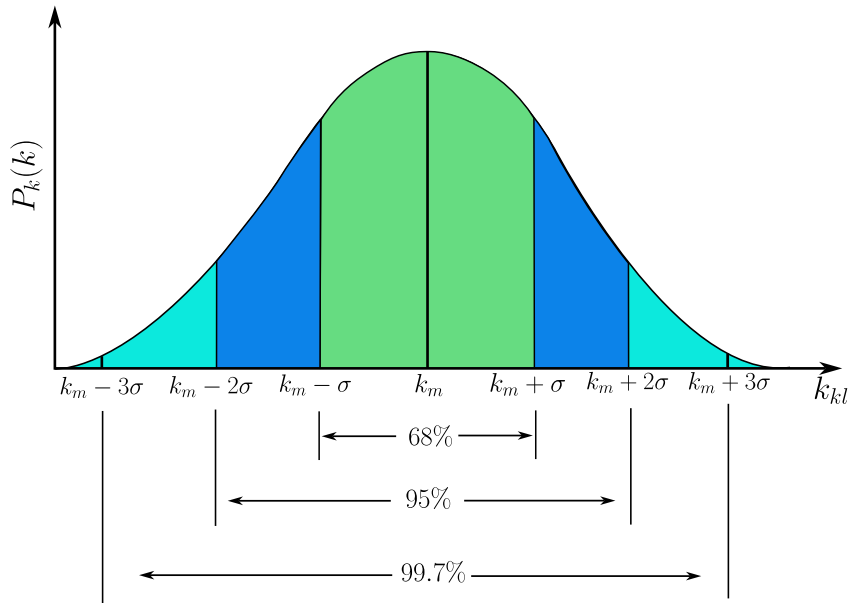
onde U_k representa o potencial de interação da partícula k com outras partículas e o indentador, sendo definido por

$$U_k = \sum_l \frac{k_{kl}(r_{kl} - \ell)^2}{2} + \varepsilon \left[\frac{\rho}{r_{ks} - (\rho_s - \rho)/2} \right]^\varpi. \quad (3.26)$$

O somatório na primeira parte abrange os vizinhos, onde $r_{kl} = |\vec{r}_k - \vec{r}_l|$ e ℓ representam, respectivamente, a distância entre os centros das partículas e a distância de equilíbrio das partículas k e l , com interações elásticas dadas por k_{kl} . O último termo da equação 3.26 descreve um potencial *hard core*, aplicado exclusivamente às partículas em contato com o indentador. Aqui, ε é um parâmetro de energia, e $r_{ks} = |\vec{r}_k - \vec{r}_s|$ indica a distância entre o centro da partícula e o indentador. O expoente ϖ deve ser suficientemente grande para preservar a rigidez do indentador. Mesmo quando um banho térmico é incorporado ao sistema, as forças de flutuações das partículas produzem resultados semelhantes na curva de força (De Araújo *et al.*, 2020). Portanto, a ausência do banho térmico na curva de força não interfere nos resultados propostos.

O último termo da equação (3.25) representa uma força de arrasto generalizada que age de maneira oposta à velocidade da partícula, $\vec{v}_k = v_k \hat{v}_k$, com uma magnitude determinada

Figura 10 – A curva representa a distribuição gaussiana das constantes elásticas k_{kl} em uma rede heterogênea. O pico na posição k_m corresponde à média da distribuição, indicando o valor mais provável para as constantes elásticas do sistema. A largura da curva, controlada pelo desvio padrão σ , expressa o grau de dispersão dos valores em torno da média. A região sombreada entre $[k_m - \sigma, k_m + \sigma]$ abrange cerca de 68% dos dados, correspondendo ao intervalo de um desvio padrão. Em seguida, os intervalos $[k_m - 2\sigma, k_m + 2\sigma]$ e $[k_m - 3\sigma, k_m + 3\sigma]$ englobam aproximadamente 95% e 99,7% dos valores, respectivamente, caracterizando os domínios de dois e três desvios padrão da média.



Fonte: Elaborado pelo autor.

por γv_k^α , em que γ e α estão associados à geometria da partícula e às propriedades do fluido em que as partículas estão imersas. Quando $\gamma = 0$, a dissipação desaparece, resultando em redes puramente elásticas, nas quais nosso modelo reproduz o conhecido comportamento Hertz para o contato mecânico (De Araújo *et al.*, 2020). Contudo, quando o atrito local se torna significativo ($\gamma > 0$), o modelo pode gerar comportamentos distintos de materiais viscoelásticos. Notavelmente, os valores típicos para α são 1 e 2, representando forças de arrasto que atuam em estruturas rígidas. O regime linear, conhecido como a lei de Stokes, surge para pequenos números de Reynolds, onde as forças viscosas predominam sobre as forças inerciais. Nesse caso, γ é proporcional à viscosidade do meio e ao diâmetro da partícula, exibindo um relaxamento numérico com um decaimento exponencial, representando o modelo reológico do sólido linear padrão (De Araújo *et al.*, 2020). Por outro lado, o arrasto quadrático é dominante para grandes números de Reynolds, onde γ é proporcional à densidade do meio e à área da seção transversal da partícula.

3.5.1 Redes Heterogêneas com Distribuição Gaussiana

Em nosso modelo, consideramos redes heterogêneas nas quais as constantes elásticas das ligações, k_{kl} , seguem uma distribuição gaussiana, conferindo ao sistema um caráter mais dinâmico e aderente ao comportamento real de materiais desordenados. Nessa perspectiva, cada ligação é tratada como uma variável aleatória, cujos valores flutuam em torno de uma média bem definida. A distribuição gaussiana, ou normal desempenha papel central na descrição dessas variações e é caracterizada por sua função de densidade de probabilidade $P(k)$, descrita por

$$P(k_{kl}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(k_{kl} - k_m)^2}{2\sigma^2}\right), \quad (3.27)$$

Nesta formulação, k_m corresponde à média da distribuição, representando o valor médio esperado para a constante elástica ao longo de toda a rede. O desvio padrão σ mede a variabilidade dos valores de k_m em torno dessa média, indicando o grau de dispersão das constantes elásticas. A largura da distribuição gaussiana é diretamente influenciada por σ , refletindo a extensão dos valores possíveis para k_{kl} . Dentro desse contexto, a constante elástica efetiva da rede, k_{ef} , coincide com a média da distribuição, de modo que $k_{ef} = k_m$.

Ao adotarmos o intervalo $[k_m - 3\sigma, k_m + 3\sigma]$, incluímos cerca de 99% dos valores associados à distribuição, o que garante a representação da maior parte das possíveis constantes elásticas presentes na rede, conforme ilustrado na Figura 10.

É importante destacar que o potencial de interação U_k associado a cada partícula k na rede continua sendo definido pela equação 3.26. Essa formulação leva em conta simultaneamente as interações entre as próprias partículas e a interação de cada partícula com o indentador, garantindo a coerência do modelo ao longo de todo o intervalo de constantes elásticas possíveis dentro da abordagem gaussiana. Essa estruturação detalhada permite compreender de forma mais aprofundada as propriedades viscoelásticas do material, considerando como a variação dessas constantes influencia o comportamento global do sistema.

3.5.2 Redes Homogêneas

Na rede homogênea, a uniformidade das ligações é a característica principal, com todas as ligações entre as partículas possuindo a mesma constante elástica k_{kl} , igual a um valor constante k . Isso garante que todas as interações na rede sejam caracterizadas pela mesma rigidez, simplificando significativamente a análise e modelagem do sistema. A uniformidade

das propriedades mecânicas por toda a rede contribui para uma representação mais regular e simplificada do material.

A equação de potencial de interação U_k para uma partícula k em uma rede homogênea é expressa por

$$U_k = \frac{k}{2} \sum_l (r_{kl} - \ell)^2 + \varepsilon \left[\frac{\rho}{r_{ks} - (\rho_s - \rho)/2} \right]^{\varpi}, \quad (3.28)$$

onde o primeiro termo reflete as interações elásticas entre as partículas, com r_{kl} representando a distância entre o centro das partículas k e l , e ℓ é a distância de equilíbrio entre elas. O segundo termo representa um potencial de repulsão forte, aplicado para partículas que entram em contato com o indentador, onde ρ é o raio da partícula, r_{ks} é a distância entre o centro da partícula k e o indentador, e ε e ϖ são parâmetros que determinam a força e a forma do potencial de repulsão.

4 ANÁLISE DE DADOS APLICADO A CIÊNCIA DAS CIDADES

Por que algumas cidades prosperam enquanto outras enfrentam dificuldades para se diversificar suas atividades econômicas? Compreender os motores do desenvolvimento regional e da transformação estrutural permanece como um desafio central na economia e nas políticas públicas. Nas últimas duas décadas, o conceito de complexidade econômica emergiu como um arcabouço robusto para explicar diferenças de prosperidade, enfatizando que o desenvolvimento depende não apenas de capital ou recursos, mas da diversidade e sofisticação das capacidades produtivas incorporadas em um território (Hidalgo; Hausmann, 2009; Hausmann *et al.*, 2014; Tacchella *et al.*, 2012; Pietronero *et al.*, 2017). Essas capacidades produtivas, muitas vezes não observáveis diretamente, são reveladas pela amplitude e complexidade dos bens que as localidades conseguem produzir e comercializar de forma competitiva (Cristelli *et al.*, 2013; Caldarelli *et al.*, 2012).

Embora métricas como o *Economic Complexity Index* (ECI) e o algoritmo de Fitness–Complexity tenham se tornado ferramentas centrais em diagnósticos de crescimento e formulação de políticas, sua aplicação tem permanecido quase totalmente restrita ao comércio internacional agregado no nível nacional (Hausmann *et al.*, 2014; Operti *et al.*, 2018; Pinheiro, 2025; Albeaik *et al.*, 2017; Zhang *et al.*, 2025a; Hausmann *et al.*, 2007; Zeković; Perić, 2024; Mealy *et al.*, 2019; Koch *et al.*, 2024). Esse foco no nível nacional pode mascarar fortes assimetrias regionais, onde grandes economias frequentemente abrigam cidades com capacidades, bases de conhecimento e estruturas industriais profundamente distintas (Operti *et al.*, 2018; Pinheiro, 2025). Além disso, fluxos de comércio interno — que representam a maior parte da atividade econômica na maioria dos países — permanecem amplamente invisíveis nessa literatura devido à escassez de dados de alta resolução (Sampaio Filho *et al.*, 2026). A lacuna resultante é uma compreensão limitada de como a complexidade e as dinâmicas de escala operam dentro dos sistemas urbanos nacionais.

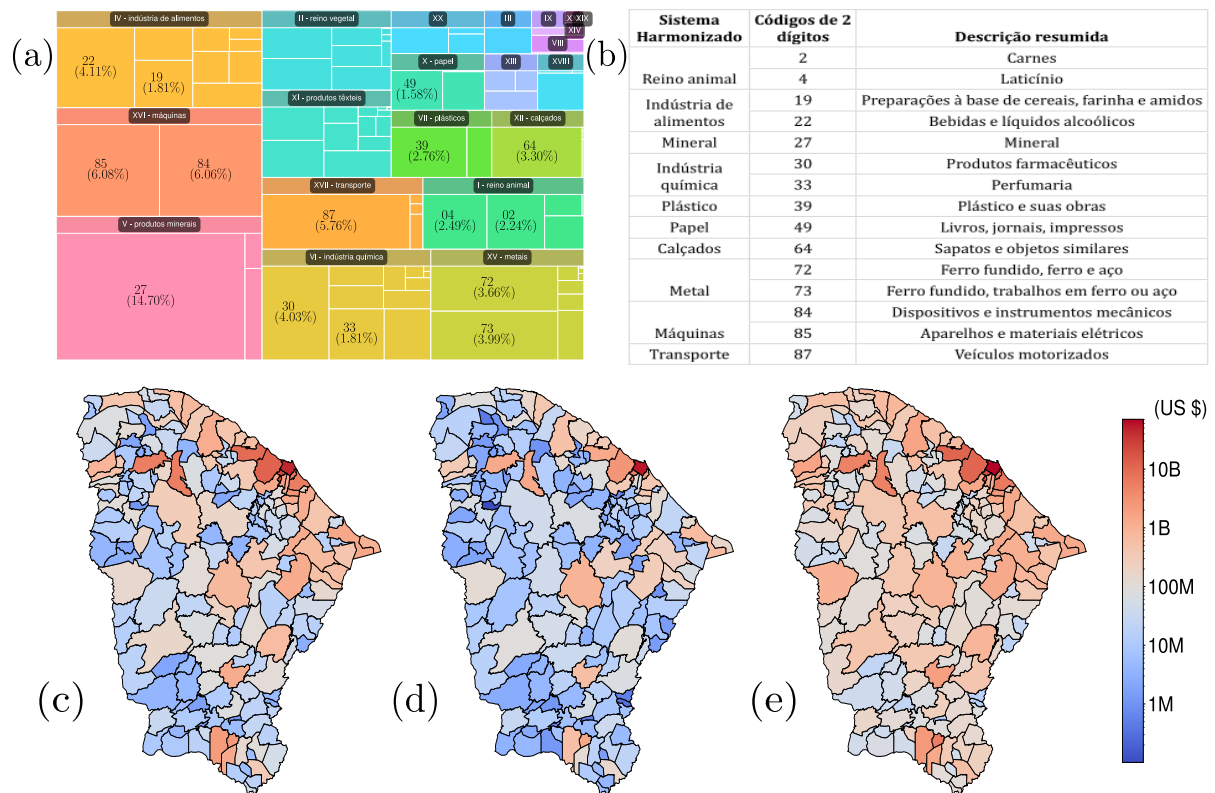
Aqui, enfrentamos essa lacuna ao analisar um conjunto abrangente de dados contendo aproximadamente 2.5 bilhões de notas fiscais eletrônicas (NF-e) trocadas entre empresas localizadas em 184 municípios do estado do Ceará, Brasil, no período de 2016 a 2019. Essas transações interfirmas registram origem, destino, produto (Nomenclatura Comum do Mercosul, NCM) e valor (Ministério da Fazenda, 2003; Portal Único Siscomex, 2024; Logcomex, 2024; Enotas, 2024; World Customs Organization, 2025; Receita Federal, 2021). Utilizando esse microconjunto de dados, reconstruímos redes comerciais direcionadas, ao nível de produto,

entre cidades. Desagregamos os fluxos em três camadas: (i) compras (demanda), (ii) vendas (produção) e (iii) comércio intramunicipal (autossuficiência). Essa estrutura nos permite mapear o comércio urbano em um nível de granularidade anteriormente inacessível (Sampaio Filho *et al.*, 2026).

Primeiro, quantificamos como esses fluxos escalam com a população das cidades. Todas as camadas exibem escalonamento superlinear, consistente com retornos crescentes à aglomeração, mas com expoentes distintos: as compras escalam em aproximadamente 1.5, enquanto as vendas e o comércio intramunicipal escalam de forma mais acentuada (em torno de 2.2 e 2.5), revelando uma divisão estrutural entre demanda e oferta — cidades maiores não apenas consomem mais, mas produzem e recirculam de maneira desproporcionalmente maior (Duranton; Puga, 2004; Rosen; Resnick, 1980; Jefferson, 1939). Em seguida, analisamos a especialização. A *Revealed Comparative Advantage* (RCA) cidade–produto (Mann *et al.*, 2023; Karbevskaja; Hidalgo, 2025; Hinloopen; van Marrewijk, 2001; Koch *et al.*, 2023) segue uma distribuição truncada de Zipf–Mandelbrot, indicando forte especialização na cauda superior e um limite de diversificação na cauda inferior, provavelmente moldado por fatores de escala, infraestrutura e restrições institucionais (Liu; Gao, 2019; Liu; Gao, 2022; Hinloopen; Van Marrewijk, 2011; Newman, 2005; Zipf, 1949; Gabaix, 1999; Bajracharya; Sultana, 2020; Soo, 2005). Aplicando o arcabouço de Fitness–Complexidade (Tacchella *et al.*, 2012; Cristelli *et al.*, 2013; Pietronero *et al.*, 2017; Pugliese *et al.*, 2016; Operti *et al.*, 2018; Morrison *et al.*, 2017), mostramos que o Fitness municipal aumenta de forma superlinear tanto com a população quanto com o Produto Interno Bruto (PIB) per capita, e que o Fitness e a renda exibem padrões de variação semelhantes ao longo do tempo.

Conceitualmente, estendemos a análise baseada em capacidades, tradicionalmente aplicada às exportações internacionais, para um ambiente intraestadual, trabalhando com microdados administrativos de municípios de um único estado. Empiricamente, revelamos uma divisão robusta entre escalonamento de demanda e oferta (compras/vendas/intramunicipal) e uma estrutura do tipo Zipf–Mandelbrot no RCA, a qual sinaliza limites de capacidade. Na prática, esses resultados fornecem um mapa detalhado de capacidades que abordagens em nível nacional não conseguem captar, oferecendo diagnósticos acionáveis para desenvolvimento regional e planejamento de resiliência.

Figura 11 – Composição setorial e distribuição regional da atividade comercial. (a) *Treemap* dos produtos comercializados agregados no nível NCM de dois dígitos, com cores indicando as seções correspondentes do HS. As áreas dos retângulos são proporcionais ao valor cumulativo de todas as transações entre empresas de 2016 a 2019, e as 15 principais categorias de produtos são identificadas com seus códigos NCM e respectivas participações em valor. (b) Breves descrições das categorias selecionadas de dois dígitos do sistema HS/NCM. A distribuição espacial dos fluxos totais entre empresas nos 184 municípios do Ceará é agregada em todos os produtos: (c) vendas totais para outros municípios, (d) transações totalmente intramunicipais e (e) compras totais provenientes de outros municípios. A intensidade da cor representa a magnitude relativa (vermelho = maior, azul = menor).



Fonte: Elaborado pelo autor.

4.1 Notas Fiscais Eletrônicas, NF-e

Este estudo utiliza microdados administrativos de alta resolução provenientes do sistema brasileiro de Nota Fiscal Eletrônica (NF-e), que registra todas as transações formais entre empresas em todo o país. Diferentemente das estatísticas agregadas de comércio nacional, o sistema NF-e fornece informações estruturadas em nível de transação para toda a economia, permitindo a reconstrução de redes econômicas detalhadas de produto–município e o estudo das dinâmicas comerciais subnacionais (Ministério da Fazenda, 2003).

Analisamos todas as transações envolvendo empresas localizadas no estado do Ceará (Nordeste do Brasil), utilizando acesso concedido pelo Instituto de Pesquisa e Estratégia Econô-

mica do Ceará (IPECE) (Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE), 2019). O conjunto de dados cobre o período de 2016 a 2019 e contém aproximadamente 2.5 bilhões de notas fiscais, correspondentes a cerca de US\$ 355 bilhões em valor transacionado. Identificadores sensíveis — como nomes de empresas, números de identificação fiscal e endereços exatos — foram anonimizados para preservar a confidencialidade, enquanto todas as variáveis essenciais para análise econômica (data da transação, valor, código do produto e município do vendedor/comprador) foram mantidas. Os produtos são classificados segundo a Nomenclatura Comum do Mercosul (NCM), totalmente harmonizada com o Sistema Harmonizado (HS) da Organização Mundial das Alfândegas. Isso garante compatibilidade com taxonomias internacionais de comércio e permite comparações com classificações setoriais padronizadas (Portal Único Siscomex, 2024; Logcomex, 2024; Enotas, 2024).

O Ceará é composto por 184 municípios e abrange aproximadamente 150 mil quilômetros quadrados. Segundo o Censo de 2022 (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022), o estado possui uma população de 8.8 milhões de habitantes, dos quais mais de 2.4 milhões residem na capital, Fortaleza — a quarta cidade mais populosa do Brasil. Fortaleza funciona como o centro econômico e logístico do Ceará, abrigando infraestrutura de grande porte e serviços avançados em educação superior, saúde e administração pública. É classificada como metrópole nacional na REGIC (Regiões de Influência das Cidades) do IBGE, exercendo influência que ultrapassa as fronteiras estaduais.

4.2 Classificação do Produto

A Nomenclatura Comum do Mercosul (NCM) é o sistema padronizado para classificar bens comercializados entre os países membros do Mercosul. Introduzida em 1995, ela harmoniza a identificação de produtos entre diferentes setores, facilita a compilação de estatísticas de comércio, agiliza procedimentos aduaneiros e dá suporte a políticas tributárias e regulatórias (Receita Federal, 2021). No Brasil, os códigos NCM são obrigatórios em notas fiscais eletrônicas, declarações aduaneiras e outros documentos oficiais de comércio. O sistema é totalmente alinhado ao Sistema Harmonizado (HS) da Organização Mundial das Alfândegas, garantindo comparabilidade internacional (World Customs Organization, 2025).

Cada código NCM é composto por oito dígitos organizados de forma hierárquica para transmitir níveis crescentes de especificidade. Por exemplo, o código 30031012 corresponde a amoxicilina ou seus sais, refinando progressivamente a classificação desde o capítulo de dois

dígitos para produtos farmacêuticos até um composto específico. Como algumas combinações de NCM não são utilizadas, o nível de dois dígitos (NCM2) abrange até 96 categorias amplas, enquanto o nível de quatro dígitos (NCM4) cobre 1228 posições. No nível mais agregado, o HS organiza os bens em 21 seções (Receita Federal, 2021; World Customs Organization, 2025).

Adotamos diferentes níveis de agregação dependendo do objetivo analítico. Para panoramas macro e análises de escalonamento, utilizamos códigos NCM de dois dígitos para capturar classes amplas de produtos. Para análises mais detalhadas, especialmente aquelas relacionadas à complexidade econômica e à especialização setorial, empregamos códigos NCM de quatro dígitos. A lista completa de códigos NCM e seus correspondentes códigos HS utilizados neste estudo é fornecida em (GitHub repository, 2025).

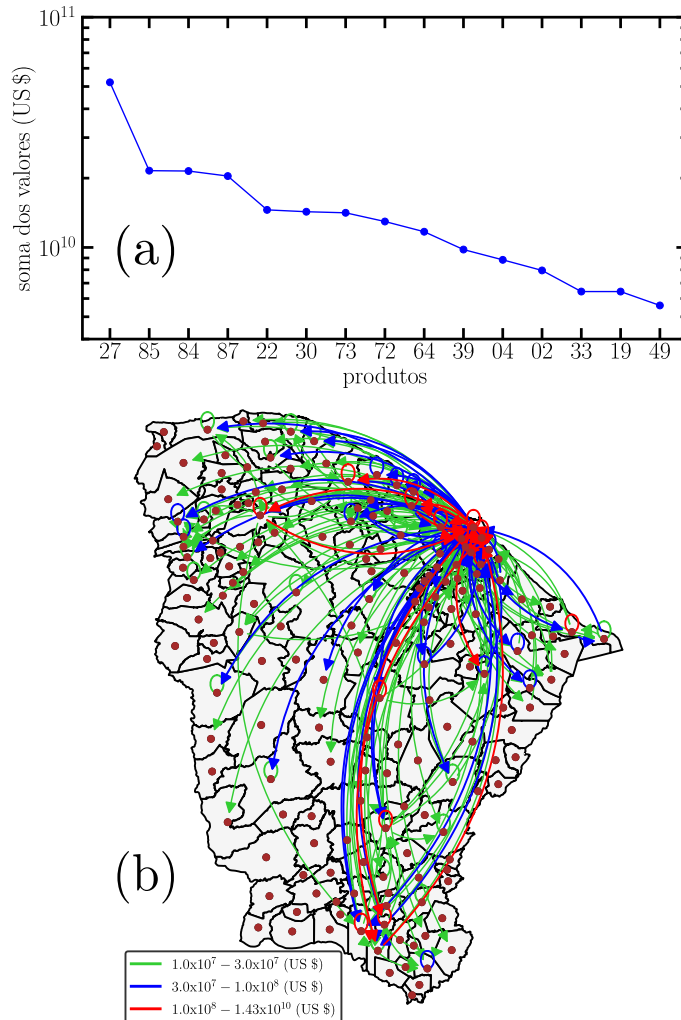
Todos os valores monetários foram deflacionados para preços de janeiro de 2025 utilizando o Índice Geral de Preços (IGP), publicado mensalmente pela Fundação Getulio Vargas (FGV) (Fundação Getulio Vargas (FGV), 2025). Os deflatores foram construídos a partir da série oficial do IGP e aplicados de forma uniforme a todas as transações de 2016 a 2019. Após a deflação, os valores foram convertidos de reais (BRL) para dólares americanos (USD) utilizando a taxa de câmbio comercial média de janeiro de 2025, conforme divulgada pelo Banco Central do Brasil (Banco Central do Brasil, 2025). Essa padronização garante plena comparabilidade ao longo da janela de agregação de quatro anos utilizada nos resultados principais.

A Figura 11(a) apresenta uma visão geral da distribuição dos produtos comercializados entre as categorias NCM de dois dígitos, destacando o peso relativo de cada setor no período de 2016–2019. A Figura 11(b) lista breves descrições das 15 principais categorias, ilustrando a diversidade de atividades econômicas capturadas pela classificação NCM/HS utilizada ao longo deste estudo. A Figura 12(a) mostra o valor transacionado cumulativo desses 15 principais produtos entre 2016 e 2019, ordenados do maior para o menor.

4.3 Construção de redes de comércio entre cidades-cidades e entre cidades-produtos

Agregamos todas as transações de NF-e no nível município-a-município para construir uma rede de comércio direcionada e ponderada, na qual os nós representam cidades e as arestas codificam o valor monetário dos bens trocados. Cada aresta é orientada do vendedor para o comprador, capturando a direção dos fluxos e revelando a geografia econômica interna do Ceará como um sistema interconectado de produção, circulação e consumo. Essa representação nos permite distinguir entre trocas intermunicipais, que refletem a estrutura conectiva dos mercados

Figura 12 – O painel (a) mostra as 15 categorias NCM de dois dígitos com maior valor acumulado ao longo de quatro anos, ordenadas em ordem decrescente. Os códigos HS/NCM e as descrições curtas das 15 principais categorias estão listados na Tabela 11(b). O painel (b) exhibe os fluxos georreferenciados entre os 184 municípios em 2016: os nós estão localizados nas sedes municipais; as arestas direcionadas representam ligações vendedor → comprador, mostrando apenas transações superiores a US\$10 milhões para maior clareza visual.



Fonte: Elaborado pelo autor.

regionais, e trocas intramunicipais, que capturam exclusivamente a atividade econômica local.

Para caracterizar os diferentes papéis que os municípios desempenham dentro desse sistema, separamos os fluxos em três camadas complementares. A camada de *vendas* compreende todos os fluxos de saída de empresas de um município para empresas de outras cidades, capturando o desempenho comercial externo e a capacidade produtiva local. A camada de *compras* reúne todos os fluxos de entrada provenientes de outros municípios, refletindo a demanda local e a dependência de fornecedores externos. A camada de *comércio interno* inclui transações em que tanto o vendedor quanto o comprador pertencem ao mesmo município, quantificando o tamanho e a circulação do mercado interno da cidade. Em conjunto, essas camadas fornecem

uma descrição abrangente da estrutura do comércio interno.

Os microdados de NF-e permitem a construção de matrizes município–produto detalhadas. Seja $x_{ij,p}$ o valor total do produto p transacionado do município i (vendedor) para o município j (comprador). Para fins de visualização, a Figura 12(b) ilustra a rede de comércio intermunicipal (agregada para todos os produtos) referente a 2016, exibindo apenas as arestas acima de US\$10000 para maior clareza. A partir desses fluxos, derivamos, para cada cidade c e produto p , as vendas externas $S_{c,p}^S$, as transações internas $S_{c,p}^I$ e as compras externas $S_{c,p}^B$, conforme segue

$$S_{c,p}^S = \sum_{\substack{j=1 \\ j \neq c}} x_{cj,p}, \quad S_{c,p}^I = x_{cc,p}, \quad S_{c,p}^B = \sum_{\substack{j=1 \\ j \neq c}} x_{jc,p}, \quad (4.1)$$

onde $c \in \{1, \dots, 184\}$ indexa os municípios do Ceará e as somas incluem todos os parceiros comerciais (dentro e fora do estado). Transações com empresas localizadas fora do Ceará são registradas quando o comprador ou vendedor pertence ao estado, permitindo identificar fluxos interestaduais associados às empresas cearenses. Cada matriz cidade–produto possui dimensões 184×96 no nível NCM2 ou 184×1228 no nível NCM4, dependendo da resolução desejada. Essas matrizes codificam o conjunto completo de produtos nos quais cada município é ativo como vendedor, comprador ou participante de comércio interno, e constituem a base para todos os cálculos de RCA, *Fitness* e *Complexity*.

A agregação ao longo dos produtos gera, para cada município, os totais correspondentes aos três tipos de fluxos,

$$S_c^S = \sum_p S_{c,p}^S, \quad S_c^I = \sum_p S_{c,p}^I, \quad S_c^B = \sum_p S_{c,p}^B, \quad (4.2)$$

representando o valor total que um município vende externamente, comercializa internamente e compra externamente. A Figura 11(c–e) mapeia a distribuição espacial desses fluxos agregados para o período de 2016-2019. A combinação de redes multicamadas de fluxos e matrizes cidade–produto fornece um arcabouço unificado para quantificar as posições estruturais dos municípios, suas capacidades de comercialização e seus padrões de especialização dentro da economia interna do Ceará.

4.4 Alometria

A alometria, tipicamente escrita como $A \sim L^\beta$, fornece uma formulação compacta para quantificar como uma atividade A escala com o tamanho do sistema L (West *et al.*, 1997; West; Brown, 2005). A alometria captura como uma quantidade varia de forma desproporcional com o tamanho do sistema, revelando se sistemas maiores são apenas versões ampliadas de sistemas menores ou se exibem comportamentos fundamentalmente distintos. Um expoente superlinear ($\beta > 1$) indica que cidades maiores geram desproporcionalmente mais atividade per capita, enquanto um escalonamento sublinear ($\beta < 1$) reflete retornos decrescentes à medida que o tamanho aumenta.

No contexto do comércio municipal, tratamos as cidades como sistemas adaptativos complexos cujo “metabolismo econômico” escala com a população, utilizada aqui como proxy para o tamanho urbano (Chen, 2014; Li *et al.*, 2015; Samaniego; Moses, 2008; Qiao *et al.*, 2024; D’Acci, 2025). Esse arcabouço nos permite identificar regimes de escalonamento superlinear, linear ou sublinear associados a diferentes tipos de atividade econômica. Para cada tipo de fluxo e para cada produto p , estimamos o expoente alométrico β utilizando regressão bayesiana D aplicada à relação

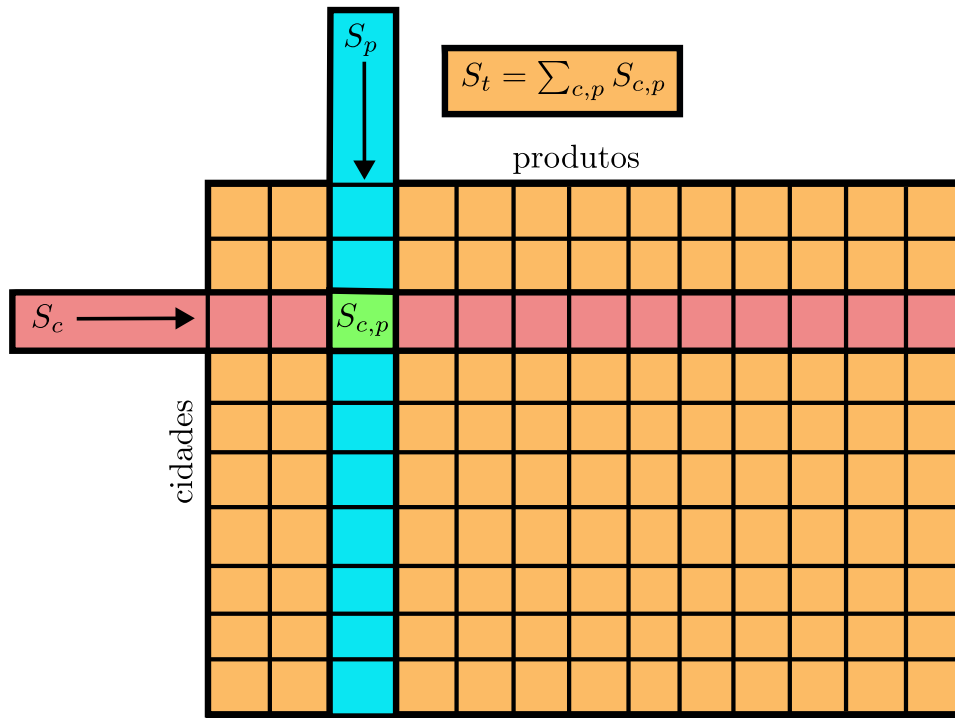
$$S \sim (\text{população})^\beta, \quad (4.3)$$

onde omitimos os índices matriciais S e o expoente β depende do produto sendo comercializado. Esse escalonamento é calculado separadamente para as três matrizes de fluxos — vendas intermunicipais, transações intramunicipais e compras intermunicipais — construídas com a Eq. (4.1). Ao analisar o agregado de todos os produtos, utilizamos os totais no nível municipal S_c derivados da Eq. (4.2).

4.5 Revealed Comparative Advantage (RCA)

A *Revealed Comparative Advantage* (RCA), introduzida por Balassa em 1965 (Mann *et al.*, 2023; Hinloopen; van Marrewijk, 2001; Liu; Gao, 2019; Liu; Gao, 2022), é amplamente utilizada para identificar setores nos quais uma economia apresenta vantagem relativa no comércio. Adaptamos a métrica para o nível municipal a fim de avaliar se a cidade c possui vantagem relativa na comercialização do produto p . Para cada camada de comércio — vendas

Figura 13 – Estrutura matricial 184×1228 utilizada para o cálculo do RCA nas três camadas de comércio do Ceará. Cada célula representa o fluxo $S_{c,p}$ entre a cidade c e produto p . A faixa vertical em azul destaca o total estadual do produto ($S_p = \sum_c S_{c,p}$), enquanto a faixa horizontal em vermelho evidencia o total movimentado pela cidade ($S_c = \sum_p S_{c,p}$). A célula verde marca o valor específico $S_{c,p}$ cujo peso relativo é avaliado na métrica. O retângulo laranja superior indica o total agregado transacionado no estado ($S_t = \sum_{c,p} S_{c,p}$). O esquema ilustra como as diferentes somas marginais se articulam na matriz cidade-produto para compor a razão utilizada no cálculo do RCA.



Fonte: Elaborado pelo autor.

intermunicipais, transações intramunicipais e compras intermunicipais — calculamos

$$RCA_{c,p} = \left(\frac{S_{c,p}}{S_c} \right) / \left(\frac{S_p}{S_t} \right), \quad (4.4)$$

onde $S_{c,p}$ e S_c foram definidos anteriormente, $S_p = \sum_c S_{c,p}$ é o total estadual para o produto p , e $S_t = \sum_{c,p} S_{c,p}$ é o valor total transacionado. A razão $S_{c,p}/S_c$ mede a importância do produto p no portfólio comercial da cidade c , enquanto S_p/S_t mede sua importância no portfólio agregado do estado. Um par cidade-produto com $RCA_{c,p} > 1$ indica especialização, significando que o produto está super-representado na cidade em relação à média estadual. As matrizes de RCA resultantes têm exatamente as mesmas dimensões das matrizes subjacentes $S_{c,p}$ e são analisadas tanto em forma contínua quanto binarizada para fins de análise de complexidade.

4.5.1 Distribuições de Zipf-Mandelbrot

Para caracterizar a distribuição das vantagens comparativas reveladas entre os municípios, examinamos a relação rank–size dos valores de RCA. Muitos processos econômicos e informacionais — como frequências linguísticas, tamanhos de firmas, produção de inovações e padrões de especialização de exportações — exibem comportamento *heavy-tailed* bem descrito por distribuições de Zipf ou Zipf-Mandelbrot (Mandelbrot, 1953; Mandelbrot, 1982). A forma Zipf–Mandelbrot generaliza a lei clássica de Zipf ao introduzir um parâmetro de deslocamento q , permitindo uma modelagem aprimorada da curvatura em baixos ranks

$$\text{RCA}(k) \sim (k + q)^{-\alpha}, \quad (4.5)$$

onde k é o rank e α determina a inclinação do decaimento. No contexto do comércio municipal, essa estrutura captura a coexistência de um pequeno número de pares cidade–produto altamente especializados e um grande número de pares próximos da média. O expoente α fornece uma medida resumida da desigualdade de especialização: valores elevados de α refletem uma hierarquia acentuada de capacidades, enquanto valores baixos de α refletem uma distribuição mais homogênea de vantagens.

Estimamos α e q para cada camada de comércio (vendas, compras e transações intramunicipais) utilizando mínimos quadrados não lineares aplicados ao regime $\text{RCA}_{c,p} \geq 1$, no qual o comportamento de lei de potência é empiricamente robusto.

4.5.2 Binarização RCA

Para analisar os padrões de especialização municipal e construir redes baseadas em capacidades, convertamos cada matriz contínua de RCA em uma matriz binária de incidência cidade–produto, atribuindo valor igual a um para $\text{RCA}_{c,p} \geq 1$ e zero caso contrário. Assim como na versão contínua, essas matrizes são construídas separadamente para vendas intermunicipais, transações intramunicipais e compras intermunicipais. A binarização da RCA é padrão na literatura de complexidade econômica (Hidalgo; Hausmann, 2009; Pietronero *et al.*, 2017; Tacchella *et al.*, 2012; Caldarelli *et al.*, 2012) e serve a diversos propósitos metodológicos.

Primeiro, o limiar $\text{RCA} \geq 1$ fornece um critério transparente e comparável para identificar especialização substantiva e reduz a sensibilidade a ruídos estatísticos próximos ao limiar, algo particularmente importante para municípios pequenos ou de baixo volume. A representação binária também é a entrada natural para modelos bipartidos cidade–produto

subjacentes ao arcabouço de Fitness–Complexity, permitindo detectar blocos de especialização, trajetórias de diversificação e adjacências de capacidades. Empiricamente, as distribuições rank–size que observamos seguem uma forma Zipf–Mandelbrot principalmente no regime $RCA \geq 1$, tornando esse limiar simultaneamente consistente em termos teóricos e orientado pelos dados.

Municípios extensos e diversificados tendem a apresentar muitos valores próximos ao limiar ($RCA \approx 1$) em uma ampla gama de produtos, refletindo portfólios equilibrados. Em contraste, municípios menores tipicamente exibem padrões polarizados, com valores elevados de RCA concentrados em poucos nichos e valores próximos de zero no restante, indicando especialização estreita e maior vulnerabilidade. O uso de $RCA \geq 1$ também mitiga “vantagens” espúrias que podem surgir devido a denominadores pequenos em cidades de baixa escala.

As matrizes binarizadas de RCA constituem a base para nossa análise subsequente de complexidade. Elas alimentam diretamente o cálculo do Fitness das cidades e da Complexidade dos produtos e sustentam a construção de redes de proximidade que mapeiam adjacências de capacidades. Essas ferramentas nos permitem caracterizar padrões de especialização, quantificar lacunas de diversificação e identificar gargalos estruturais que moldam as trajetórias evolutivas e a resiliência das economias municipais do Ceará.

4.6 Fitness e Complexidade

O modelo de Fitness–Complexity fornece uma abordagem não monetária, baseada em capacidades, para avaliar a competitividade estrutural, a diversificação e a sofisticação tecnológica das economias regionais. A modelagem da rede bipartida que liga municípios aos produtos nos quais exibem vantagem comparativa revelada (RCA) captura tanto a diversidade quanto a exclusividade dos portfólios locais de capacidades que sustentam o potencial de desenvolvimento (Tacchella *et al.*, 2012; Cristelli *et al.*, 2013; Caldarelli *et al.*, 2012).

Na formulação original (Tacchella *et al.*, 2012), o Fitness (F) de uma localidade resume a magnitude e a sofisticação de suas capacidades comercialização, isto é, sua habilidade de participar de atividades complexas. Em contraste, a Complexidade (Q) de um produto é definida de acordo com o nível de desenvolvimento das localidades que o produzem: um produto é considerado complexo quando somente localidades com alto Fitness conseguem produzi-lo, enquanto produtos produzidos também por localidades com baixo Fitness são classificados como menos complexos. Essa estrutura não linear e recursiva reflete a hierarquia empiricamente

observada das atividades econômicas: cidades de alto Fitness participam tanto de produtos sofisticados quanto de produtos básicos, enquanto cidades de baixo Fitness ficam restritas a bens mais simples (Cristelli *et al.*, 2013; Morrison *et al.*, 2017). Em comparação com medidas lineares, como o *Economic Complexity Index* (ECI), o algoritmo de Fitness–Complexity reproduz de forma mais fiel a estrutura triangular e aninhada de diversificação das matrizes comerciais empíricas e evita circularidades na inferência, isto é, situações em que países e produtos são definidos como complexos apenas por se referirem mutuamente (Pugliese *et al.*, 2016; Operti *et al.*, 2018).

Conceitualmente, Q aumenta quando um produto é vendido, comprado ou circulado apenas por cidades de alto fitness, e diminui à medida que se torna ubíquo entre localidades de menor capacidade. O arcabouço, portanto, vai além de simples contagens de variedades: ele captura a exclusividade, a sofisticação e o conteúdo informacional das capacidades locais, informando avaliações sobre capacidade de inovação, vulnerabilidade sistêmica e barreiras à atualização (Cristelli *et al.*, 2013; Caldarelli *et al.*, 2012).

Em nossa aplicação, computamos matrizes binarizadas de RCA separadamente para cada camada de comércio, isto é, vendas, compras e transações internas. A interpretação de Fitness permanece inalterada, mas a Complexidade torna-se específica por camada: na camada de vendas, (Q) reflete a sofisticação dos produtos que uma cidade é capaz de vender; na camada de compras, (Q) reflete a sofisticação dos insumos que uma cidade é capaz de adquirir; e na camada de comércio interno, (Q) reflete a sofisticação dos bens que podem ser produzidos e circulados localmente. O Fitness (F), portanto, resume as capacidades estruturais subjacentes, enquanto a Complexidade específica por camada captura a exclusividade e a sofisticação dos bens associados em cada dimensão comercial.

Operacionalmente, partimos de uma matriz de incidência binarizada RCA, onde $RCA = 1$ se a cidade c possui $RCA \geq 1$ no produto p , e 0 caso contrário. Construimos essa matriz para cada camada de comércio: vendas (proxy de produção), compras (proxy de demanda) e transações internas (circulação local). Com condições iniciais neutras, $\tilde{F}_c^{(0)} = \tilde{Q}_p^{(0)} = 1$ para todas as cidades e produtos, o Fitness e a Complexidade são computados por meio das iterações

acopladas (Tacchella *et al.*, 2012; Cristelli *et al.*, 2013),

$$\tilde{F}_c^{(n)} = \sum_p \text{RCA}_{cp} \cdot Q_p^{(n-1)}, \quad (4.6)$$

$$\tilde{Q}_p^{(n)} = \left[\sum_c \left(\text{RCA}_{cp} \cdot \frac{1}{F_c^{(n-1)}} \right) \right]^{-1}, \quad (4.7)$$

com normalização em cada etapa para evitar desvios de escala,

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c}, \quad (4.8)$$

$$Q_p^{(n)} = \frac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle_p}. \quad (4.9)$$

Aplicações subnacionais enfrentam desafios, incluindo matrizes esparsas, incidências voláteis e produtos de nicho. Trabalhos anteriores recomendam filtrar categorias raras, agregar códigos de produto ou importar complexidades de produtos em níveis mais amplos (nacional ou internacional) para estabilizar as estimativas (Operti *et al.*, 2018; Pugliese *et al.*, 2016). Apesar desses problemas, o Fitness tem se mostrado consistentemente associado à diversificação, à inovação e à resiliência: municípios de alto Fitness exibem estruturas econômicas mais robustas e adaptáveis, enquanto municípios de baixo Fitness permanecem vulneráveis devido a especializações estreitas (Cristelli *et al.*, 2013; Morrison *et al.*, 2017; Albeaik *et al.*, 2017). Em escala subnacional, o modelo de Fitness–Complexity frequentemente supera o ECI, produzindo rankings que refletem de forma mais plausível os padrões observados de sofisticação (Pietronero *et al.*, 2017; Albeaik *et al.*, 2017).

Aqui, computamos o Fitness e a Complexidade para todos os 184 municípios e para todas as categorias de produtos nas três camadas de comércio. Essa perspectiva de capacidades em múltiplas camadas mapeia o papel de cada cidade como produtora, consumidora e circuladora de bens. Ela fundamenta nossas análises subsequentes sobre trajetórias de diversificação, redes de proximidade e gargalos estruturais na economia regional do Ceará.

4.7 Dados Demográficos

Para contextualizar os padrões de comércio em função da escala urbana, mesclamos estimativas municipais de população do IBGE para 2016–2019 (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022). Esses dados fornecem a base demográfica necessária para testar relações alométricas (de escalonamento) entre fluxos econômicos e tamanho das cidades, no

contexto do marco teórico de escalonamento urbano (*urban scaling*) (Chen, 2014; Li *et al.*, 2015; Samaniego; Moses, 2008; Qiao *et al.*, 2024; D'Acci, 2025). Usando a população como proxy do tamanho urbano, avaliamos se os volumes de comércio em cada camada da rede — compras, vendas e transações intramunicipais (internas) — escalam de forma sublinear, linear ou superlinear com a população. Essas comparações revelam em que medida efeitos de aglomeração, concentração do comércio e acumulação de capacidades variam entre os municípios do Ceará.

5 RESULTADOS E DISCUSSÕES

5.1 Redes Viscoelásticas Imersas em Fluido

A solução numérica da Eq. (3.25) é realizada por meio de simulações de dinâmica molecular (Rapaport, 2004; De Araújo *et al.*, 2017)A, com condições de contorno periódicas aplicadas no plano horizontal, e a integração temporal é feita com o algoritmo velocity Verlet, com passo de tempo dt . A força é calculada como a soma de todas as colisões sobre o indentador em cada passo de tempo. Em amostras elásticas contínuas, o modelo de Hertz para contato mecânico estabelece que a força de contato segue uma lei de potência da indentação, $f \sim \delta^\lambda$, onde o expoente λ depende da geometria do indentador. Valores usuais são 1, 3/2 e 2 para indentadores planos, esféricos e cônicos, respectivamente (De Araújo *et al.*, 2020).

No entanto, em nosso modelo de rede, a razão entre ρ_s e ρ altera a forma como a rede percebe a deformação. Para indentadores do tamanho das partículas ($\rho_s \sim \rho$), o expoente de Hertz aproxima-se de 2, como ocorre com indentadores cônicos em amostras contínuas. Para indentadores muito maiores que as partículas ($\rho_s \gg \rho$), o expoente tende a 1, como em deformações causadas por um indentador plano. O expoente de Hertz iguala-se a 3/2 (como esperado para um indentador esférico) apenas para tamanhos intermediários de indentadores, especificamente para $\rho_s = 11\rho$, que é o valor utilizado neste trabalho. A relação entre λ e ρ_s para diferentes números de partículas é apresentada em (De Araújo *et al.*, 2020). Consulte (Elaborado pelo Autor, 2023) para os parâmetros utilizados.

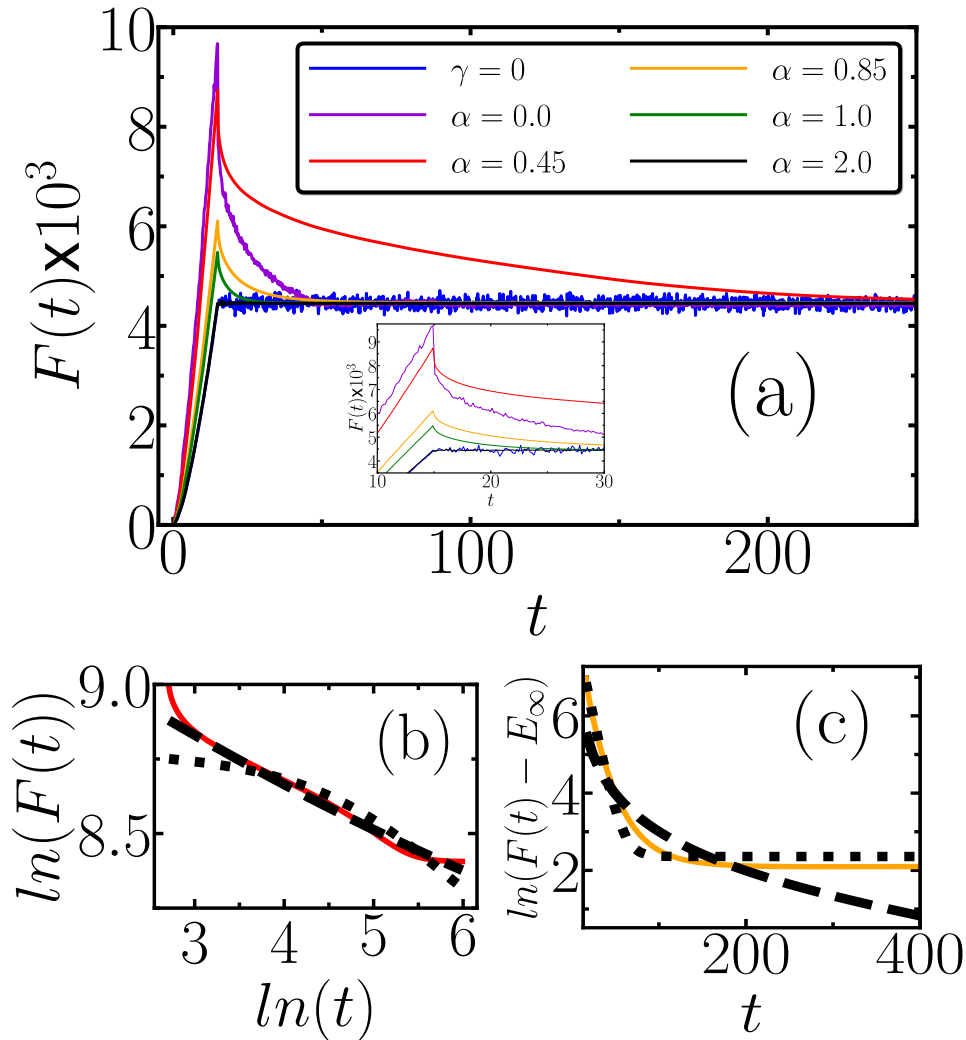
Cada conjunto de parâmetros mesoscópicos, k , γ e α , representa um material específico e define as respostas viscoelásticas macroscópicas.

5.1.1 Redes Homogêneas

Para investigar como as propriedades microscópicas levam ao comportamento reológico de toda a rede homogênea, realizamos simulações variando a rigidez de mola entre 100 e 1000, a constante de arrasto entre 10 e 100 e o expoente de arrasto entre 0.0 e 1.0, totalizando 2100 redes diferentes.

A Figura 14(a) exibe curvas típicas de força para $\gamma = 80$ e $k = 800$ para diferentes valores do expoente de arrasto. A força de relaxação $F(t)$ simula um ensaio de indentação, usado para caracterizar as respostas macroscópicas de materiais reais (De Sousa *et al.*, 2017; De Sousa *et al.*, 2021; De Sousa *et al.*, 2020). Durante a etapa de *loading*, o indentador deforma lentamente

Figura 14 – (a) Curvas numéricas de força para $\gamma = 80$ e $k = 800$ e vários valores de α , exibindo diferentes relaxações viscoelásticas. Os casos para $\alpha = 0.45$ e $\alpha = 0.85$ são mostrados nos painéis (b) e (c), respectivamente, onde as linhas tracejada e pontilhada representam ajustes com os modelos exponencial e de lei de potência das Eqs. 5.3 e 5.4, respectivamente.



Fonte: Elaborado pelo autor.

a rede até que a força de contato atinja seu valor máximo, que é inversamente proporcional a α . No estágio de *dwell*, quando o indentador permanece parado, $F(t)$ relaxa até que parte da energia mecânica seja dissipada por atrito (De Sousa *et al.*, 2017; De Sousa *et al.*, 2021; De Sousa *et al.*, 2020). A escala de tempo dessa dissipação depende de α de maneira complexa. Para $\alpha = 2$, a dissipação é rápida o suficiente para que o comportamento da resposta seja qualitativamente semelhante ao de uma rede perfeitamente elástica, ou seja, $\gamma = 0$. Para $\alpha < 1$, as forças de arrasto removem lentamente a energia da rede, resultando em um decaimento viscoelástico.

Em uma abordagem contínua, a força de contato dependente do tempo, referente à indentação de uma amostra viscoelástica por um indentador esférico, segue a integral de

convolução B(Ferreira, 2022; Ferreira *et al.*, 2024).

$$\mathcal{F}(t) = \int_0^t R(t-t') \frac{d\delta^{\frac{3}{2}}(t')}{dt'} dt', \quad (5.1)$$

onde $R(t)$ representa a função de relaxação e $\delta(t)$ é o histórico de profundidade de indentação. A força de contato na Eq. (5.1) é normalizada pela constante $4\sqrt{\frac{\rho_s}{2}} \delta_{max}^{\frac{3}{2}}/3(1-\nu^2)$, onde ν é o coeficiente de Poisson. Comportamentos típicos de relaxação para materiais viscoelásticos são descritos por

$$\begin{aligned} R_P(t) &= E_\infty + \Delta E t^{-\beta}, \\ R_E(t) &= E_\infty + \Delta E e^{-\frac{t}{\tau}}, \end{aligned} \quad (5.2)$$

onde $R_P(t)$ é o modelo de relaxação para materiais de lei de potência, com expoente β , e $R_E(t)$ representa materiais exponenciais com tempo de relaxação τ . Em ambos os modelos, E_∞ é o módulo elástico em tempos suficientemente longos, quando o material está completamente relaxado, e ΔE é a diferença entre o valor máximo, em $t = \tau_l$, e E_∞ . Assumimos que os materiais se tornam perfeitamente elásticos após certo tempo, embora esse platô elástico em longos tempos nem sempre seja observado em materiais reais. Essa suposição é necessária para evitar efeitos de alta pressão em nosso modelo numérico. Por exemplo, se a força externa for grande o suficiente para comprimir as pequenas partículas até que uma rede condensada seja formada, qualquer pressão adicional induziria a deformação dessas partículas. Entretanto, microdeformações são permitidas apenas devido ao movimento dos pseudoátomos, o qual é representado pelo regime de arrasto sublinear. Resolvendo a Eq. (5.1) com as Eqs. (5.2) na etapa de permanência (dwell) leva, respectivamente, a

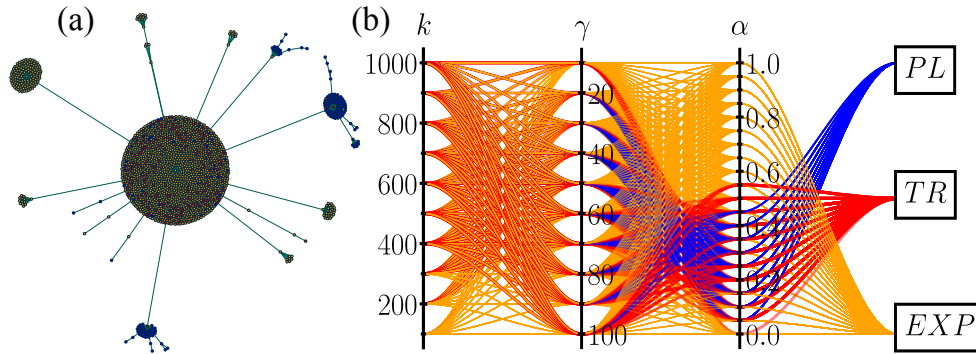
$$\mathcal{F}_P(t) = E_\infty + at^{-\beta}, \quad (5.3)$$

$$\mathcal{F}_E(t) = E_\infty + be^{-\frac{(t-\tau_l)}{\tau}} - ce^{-\frac{t}{\tau}}, \quad (5.4)$$

onde a constante $a \approx \frac{\Delta E}{\Gamma(1-\beta)}$ é obtida pela expansão da função beta incompleta(Ferreira, 2022; Ferreira *et al.*, 2024), $\Gamma(1-\beta)$ é a função gama, e $b = \frac{3\tau}{2\tau_l} \Delta E$ e $c = \frac{3\sqrt{\pi}}{4} (\frac{\tau}{\tau_l})^{\frac{3}{2}} \Delta E \operatorname{erfi}(\sqrt{\frac{\tau_l}{\tau}})$, onde $\operatorname{erfi}(\sqrt{\frac{\tau_l}{\tau}})$ é a função erro imaginária.

A análise da etapa de permanência (*dwell*) da curva de força permite mapear as propriedades mecânicas macroscópicas (E_∞ , ΔE , τ , β) com os parâmetros mesoscópicos (k , γ e α), como feito anteriormente para $\alpha = 1$ (De Araújo *et al.*, 2020). Aqui, focamos em

Figura 15 – (a) Visualização gráfica do processo de clusterização utilizando o método *K-means*. Cada ponto representa um experimento computacional para uma dada combinação de k , γ e α , e as cores representam diferentes resultados de relaxação, a saber: lei de potência (PL) em azul, exponencial (EXP) em amarelo e comportamento transicional (TR) em vermelho. (b) Um gráfico de coordenadas paralelas mostra como diferentes valores dos parâmetros mesoscópicos da rede levam a distintos tipos de relaxação viscoelástica. Cada linha atravessa uma combinação de k , γ e α , juntamente com o comportamento reológico correspondente.



Fonte: Elaborado pelo autor.

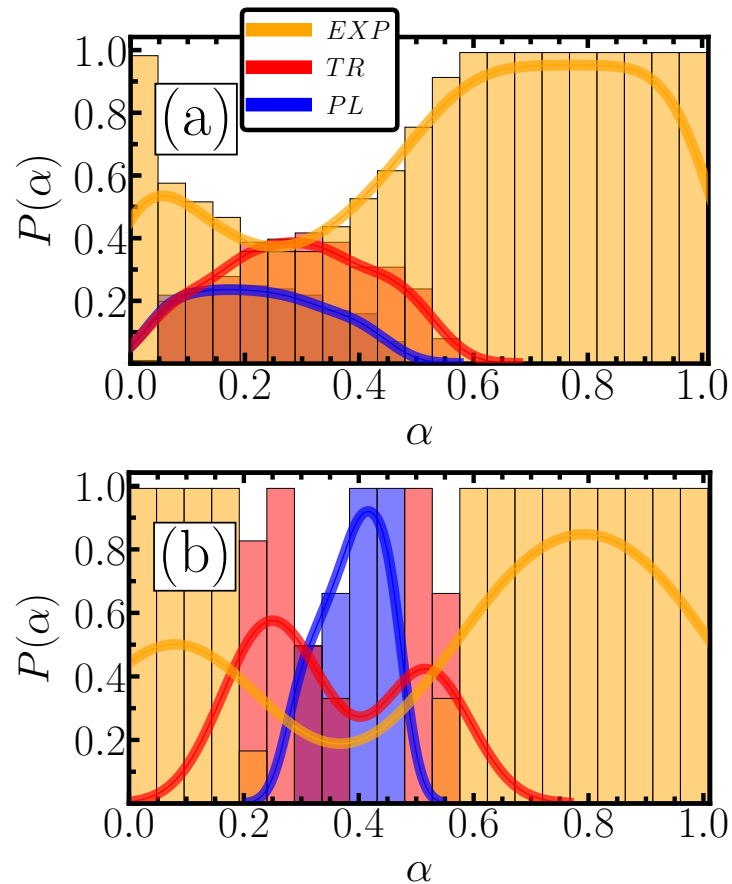
identificar o comportamento reológico qualitativo, em vez de descrever relações detalhadas entre os parâmetros. As Figuras 14(b) e 14(c) mostram as mesmas curvas de força em (a) para $\alpha = 0.45$ e $\alpha = 0.85$, respectivamente, onde as curvas são ajustadas tanto com $\mathcal{F}_P(t)$ quanto com $\mathcal{F}_E(t)$. Claramente, o caso $\alpha = 0.45$ é melhor ajustado com o modelo de lei de potência, enquanto $\alpha = 0.85$ é melhor ajustado por uma exponencial.

O erro quadrático médio χ determina o parâmetro de qualidade do ajuste entre a curva de força numérica obtida $F(t)$ e os modelos analíticos de força dados nas Eqs. 5.3 e 5.4 C,

$$\chi_M = \frac{1}{N_p} \sum_{i=1}^{N_p} [F(t_i) - \mathcal{F}_M(t_i)]^2, \quad (5.5)$$

onde N_p é o número de pontos nas curvas de força e M representa o tipo de modelo exponencial (E) ou lei de potência (P). Esse índice estatístico, calculado para cada combinação de k , γ , α , é utilizado para o método *K-means* como uma estratégia de agrupamento não supervisionado 5.1.4 (Géron, 2019; Sinaga; Yang, 2020) para classificar as curvas de força numéricas como apresentando um comportamento exponencial, comportamento de lei de potência ou até mesmo um regime transitório que não pode ser claramente classificado. Nesse processo de *Machine Learning*, a classificação do material leva em consideração não apenas os valores de χ_E e χ_P , mas também as tendências e distribuições no espaço de fase de χ_E e χ_P . A Figura 15(a) apresenta a visualização gráfica desse processo de clusterização. O comportamento exponencial é, de fato, o mais comum, como evidenciado pelo grande agrupamento na Figura 15(a), embora pequenos grupos de materiais do tipo lei de potência também sejam observados. O gráfico de coordenadas

Figura 16 – A probabilidade de um dado valor de α levar a uma resposta do tipo lei de potência (barras azuis), exponencial (barras amarelas) ou transicional (barras vermelhas). As linhas sólidas mostram a probabilidade calculada pelo algoritmo KDE (*kernel density estimation*). O painel (a) apresenta resultados para toda a faixa de valores de k e γ , enquanto o painel (b) considera apenas redes com k ($k \geq 800$) e γ ($\gamma \geq 80$).

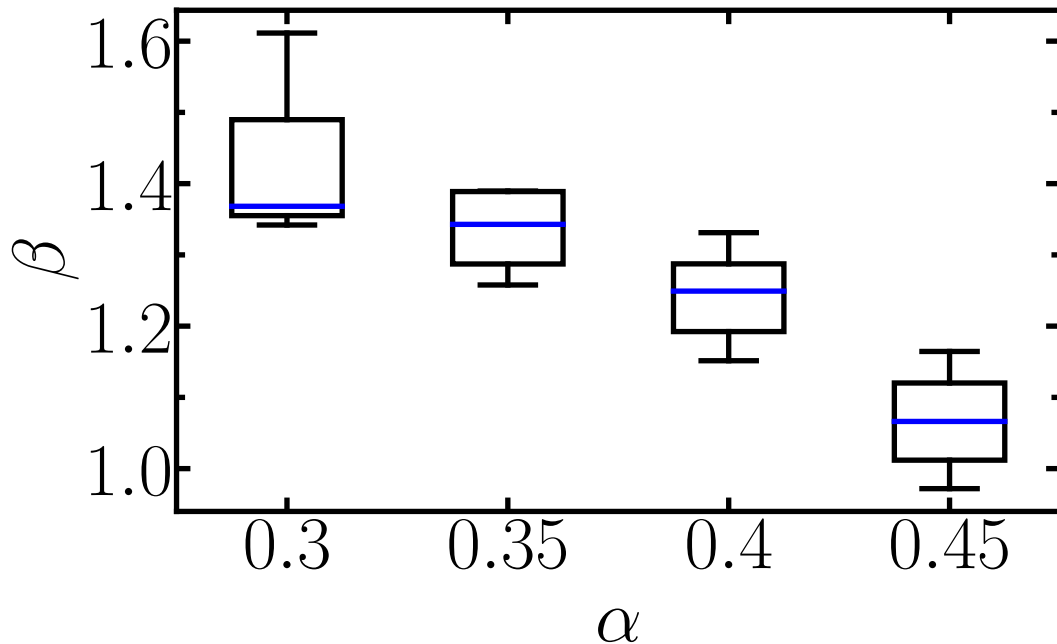


Fonte: Elaborado pelo autor.

paralelas na Figura 15(b) resume os resultados reológicos. Cada linha passa por toda combinação de k , γ e α , terminando em uma das três caixas que representam o comportamento de resposta da rede.

Para entender por que pequenos agrupamentos de materiais de lei de potência surgem na 15(a), devemos investigar o impacto de considerar regimes de arrasto sublinear. A Figura 16 mostra a probabilidade normalizada $P(\alpha)$ de encontrar comportamento PL, EXP ou TR para um dado valor de α . No painel (a), $P(\alpha)$ é calculada para todos os valores de k e γ considerados aqui. As três distribuições se sobrepõem fortemente, tornando a classificação um exercício desafiador. No painel (b), por outro lado, removemos valores pequenos de k e γ . Para redes que não são muito macias ($k \geq 800$), e não muito elásticas ($\gamma \geq 80$), essas distribuições se separam em diferentes regiões de α , e o expoente de arrasto torna-se o controlador central para caracterizar o comportamento macroscópico. Comportamentos exponenciais são encontrados

Figura 17 – Para aquelas redes classificadas como um material de lei de potência na Figura 30(b), mostramos a relação entre o expoente de relaxação β e o expoente de arrasto mesoscópico α .

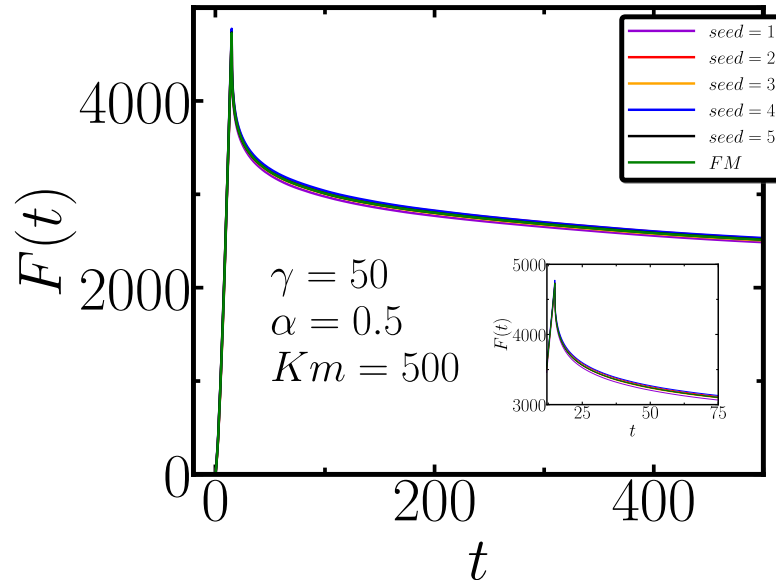


Fonte: Elaborado pelo autor.

principalmente para $\alpha \lesssim 0.2$ e $\alpha \gtrsim 0.55$, enquanto a relaxação segue uma lei de potência para α entre 0.3 e 0.45. Para valores de α em que há interseção, a sobreposição das probabilidades de PL e EXP deixa dúvidas na classificação do material como exponencial ou lei de potência. O efeito das forças de arrasto de Stokes (para $\alpha = 1$) entre macromoléculas e o solvente no qual estão imersas desempenha um papel nas propriedades viscoelásticas de materiais coloidais e poliméricos (Khair; Brady, 2005; Xi; Bai, 2016; Chen *et al.*, 2022). No entanto, até onde sabemos, este trabalho fornece uma possível explicação da mecânica subjacente que controla a resposta viscoelástica em matéria mole em termos de arrastos sublineares. Trabalhos recentes têm utilizado forças de arrasto modificadas em interações de mesoescala de matéria mole, mas novamente com intenções distintas (Zhao *et al.*, 2023; Zhang *et al.*, 2023).

Para os casos em que a rede apresenta um comportamento de lei de potência, mostramos na Figura 17 a relação entre o expoente de relaxação macroscópica β e o expoente de arrasto α . Para cada valor de α , há dispersão nos valores de β , cujo valor médio situa-se entre 1.05 e 1.35. Materiais reais que exibem relaxação do tipo lei de potência geralmente apresentam expoentes menores e exibem desordem estrutural e metastabilidade (Sollich *et al.*, 1997; Fabry *et al.*, 2001; Fabry *et al.*, 2003; Jaishankar; McKinley, 2013). Investigações computacionais de redes bidimensionais desordenadas obtiveram expoentes de relaxamento entre 0.5 e 0.75, dependendo da disposição da rede (Milkus; Zaccane, 2017). Nossas simulações apresentam

Figura 18 – Representações gráficas das curvas de força para $\gamma = 50$, $k_m = 500$, e $\alpha = 0.5$, utilizando 5 valores distintos de sementes aleatórias na rede viscoelástica. Observa-se variações mínimas no comportamento das curvas de relaxação, evidenciadas pela curva de força média FM .



Fonte: Elaborado pelo autor.

expoentes de relaxamento macroscópicos acima de 1.0, principalmente porque nosso modelo sólido viscoelástico é estruturalmente ordenado e estável, restringindo as respostas viscoelásticas a regimes de relaxamento mais rápidos.

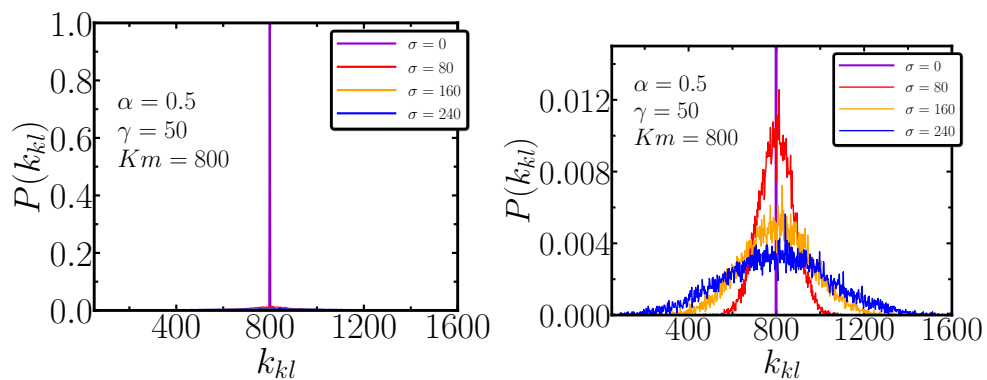
5.1.2 Redes Heterogêneas com Distribuição Gaussiana

Para investigar como as propriedades microscópicas influenciam o comportamento reológico da rede heterogênea no caso Gaussiano, realizamos um conjunto abrangente de simulações. Variamos a constante elástica média $k_m = [50, 100, 500, 800]$ (4 valores), e consideramos 3 níveis de heterogeneidade definidos por $\sigma = [10\%, 20\%, 30\%]$ de k_m . Além disso, exploramos 6 valores para a constante de arrasto (entre 50 e 100) e 15 valores para o expoente de arrasto (entre 0.1 e 0.8). Para cada combinação de parâmetros, utilizamos 5 *seeds* aleatórias distintas, totalizando 5400 simulações. As *seeds* correspondem aos valores iniciais que determinam a sequência de números pseudoaleatórios usada na simulação. Ao variar essas *seeds*, garantimos que, para uma mesma combinação de parâmetros, diferentes realizações estatísticas sejam geradas, permitindo avaliar a robustez dos resultados em relação às flutuações aleatórias do sistema.

A Figura 18 apresenta curvas representativas de força para $\gamma = 50$, $k_m = 500$, e $\alpha = 0.5$ obtidas a partir de diferentes valores de sementes (*seeds*) na rede viscoelástica. Observa-

se que a escolha de distintas sementes aleatórias exerce influência mínima sobre as curvas de relaxação. A forte similaridade entre elas é evidenciada pela sobreposição quase perfeita em torno da curva média FM . Esse comportamento demonstra que a rede é essencialmente insensível às variações das *seeds*, produzindo respostas de força praticamente idênticas.

Figura 19 – Apresentação gráfica da distribuição Gaussiana $P(k_{kl})$ para os parâmetros $\gamma = 50$, $k_m = 800$ e $\alpha = 0.5$, considerando diferentes valores do desvio padrão σ . A análise é realizada para 4 níveis distintos de variabilidade, revelando como a dispersão das constantes elásticas se altera na rede viscoelástica. Observa-se que, para $\sigma = 0$, a rede mantém uma estrutura completamente homogênea, enquanto os casos com $\sigma = [80, 160, 240]$ evidenciam um aumento progressivo da heterogeneidade, refletido no alargamento significativo da distribuição.



Fonte: Elaborado pelo autor.

A Figura 19 apresenta uma análise visual das distribuições dos valores de σ em relação a k_m na rede viscoelástica, considerando os parâmetros $\gamma = 50$, $k_m = 500$, e $\alpha = 0.5$. No primeiro painel, destacamos o caso homogêneo, correspondente a $\sigma = 0$, no qual a distribuição aparece praticamente uniforme, dificultando a visualização das demais curvas devido à forte sobreposição. No segundo painel, exibimos as distribuições associadas a $\sigma = [80, 160, 240]$, e aplicamos um zoom para melhorar a visualização, evidenciando de forma mais clara a heterogeneidade crescente da rede.

De forma análoga ao procedimento adotado para a estrutura homogênea, realizamos o ajuste das curvas de força numérica utilizando as equações 5.3 e 5.4. Nesse processo, repetimos a minimização das curvas de força tomando o erro médio quadrático χ como métrica de avaliação. A partir disso, determinamos os parâmetros de erro definidos na expressão 5.5 e conduzimos procedimentos semelhantes, utilizando técnicas de aprendizado de máquina, como o *K-means*.

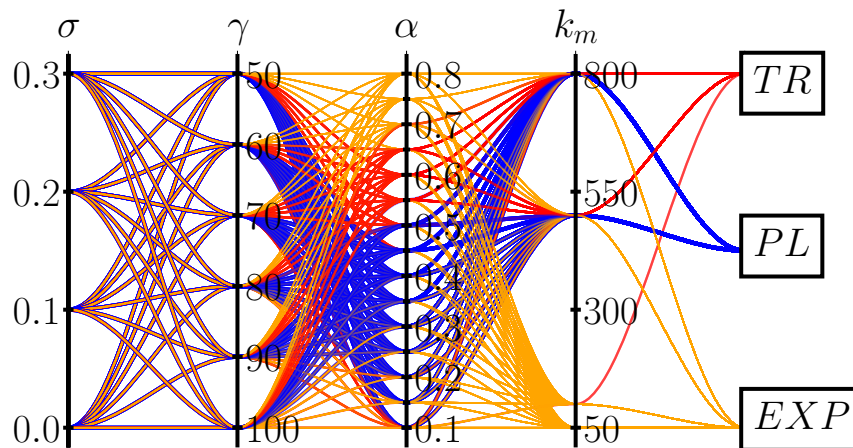
Os resultados obtidos são sintetizados no diagrama de coordenadas paralelas apresentado na Figura 20, que organiza visualmente as propriedades reológicas da rede. Cada trajetória nesse diagrama representa uma combinação específica de k_m , γ , α e σ , convergindo para uma das três regiões que caracterizam os diferentes regimes de resposta mecânica do sistema. Observa-se

que, para valores de k_m na faixa entre $[50, 100]$, não se identifica o comportamento tipicamente associado a uma lei de potência. Ressalta-se ainda que o parâmetro k_m , também conhecido como k_{ef} , desempenha um papel importante na classificação desses materiais.

A Figura 21 apresenta a probabilidade normalizada $P(\alpha)$ associada à ocorrência dos comportamentos PL, EXP ou TR em função de um determinado valor de α . No painel (a), $P(\alpha)$ é calculada considerando todas as combinações de k_m , σ e γ avaliadas, resultando em distribuições amplamente sobrepostas, o que dificulta a distinção clara entre os diferentes regimes reológicos. No painel (b), ao remover os valores mais baixos de k_m na faixa $[50, 100]$ e ao impor $\gamma \geq 80$ mantendo todos os valores de σ , observa-se uma separação mais nítida das distribuições ao longo do eixo de α . Nesse cenário, o expoente de arrasto emerge como o principal parâmetro responsável por caracterizar o comportamento macroscópico. Verifica-se que respostas exponenciais predominam para $\alpha \lesssim 0.45$ e $\alpha \gtrsim 0.7$, enquanto a relaxação segue uma lei de potência para valores de α entre aproximadamente 0.2 e 0.55. Nas regiões onde esses intervalos se sobrepõem, a coincidência entre as distribuições de PL e EXP gera incerteza na classificação da resposta do material, que pode se manifestar tanto como exponencial quanto como lei de potência.

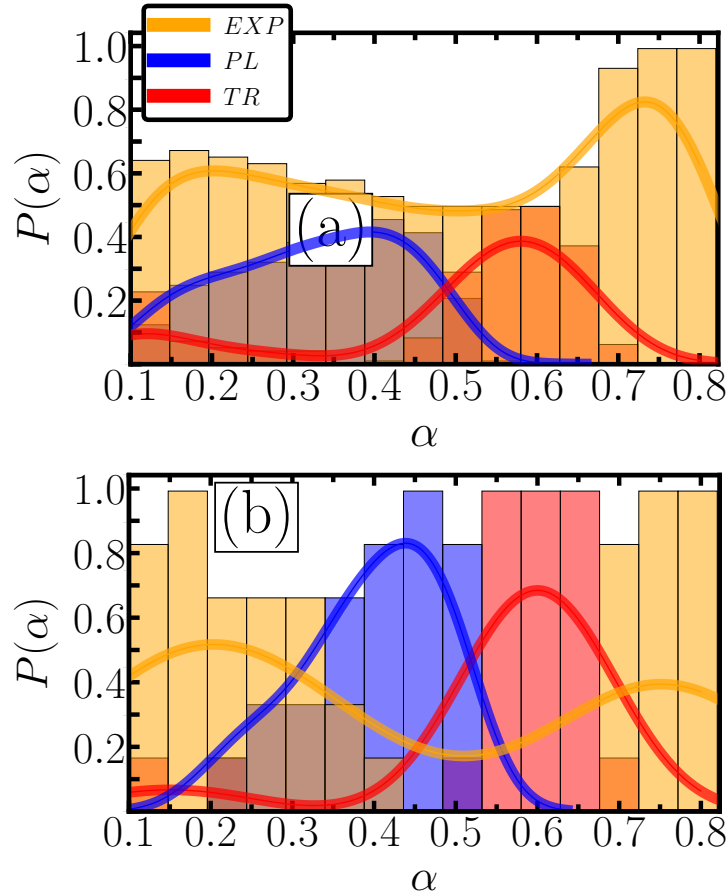
Nos casos em que a rede exibe um comportamento de lei de potência (sem qualquer filtro, englobando todas as combinações de parâmetros), a Figura 22 mostra a relação entre o expoente de relaxação macroscópica β e o expoente de arrasto. Observa-se que, para cada valor de α , a dispersão de β é reduzida, concentrando-se em um intervalo médio entre aproximadamente

Figura 20 – O diagrama de coordenadas paralelas ilustra como as variações nos parâmetros mesoscópicos da rede conduzem a diferentes padrões de relaxamento viscoelástico. Cada linha representa uma combinação específica de valores para k_m , γ , α , σ e para o comportamento reológico resultante, permitindo visualizar de forma integrada como cada parâmetro influencia a resposta global do sistema.



Fonte: Elaborado pelo autor.

Figura 21 – A probabilidade associada a um determinado valor de α resultar em uma resposta do tipo lei de potência (barras azuis), exponencial (barras amarelas) ou transicional (barras vermelhas) é apresentada no gráfico. As linhas contínuas correspondem à probabilidade estimada por meio do algoritmo KDE (*kernel density estimation*). No painel (a), exibimos os resultados considerando toda a faixa de variação dos parâmetros k_m , γ e σ . Já no painel (b), a análise é restrita às redes caracterizadas por valores mais elevados destes parâmetros, especificamente k_m ($k_m \geq 500$) e γ ($\gamma \geq 80$).



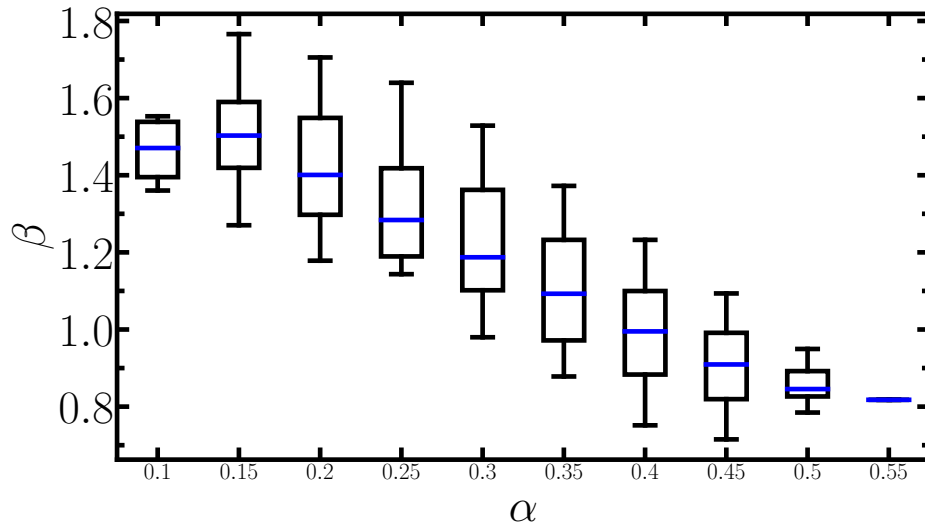
Fonte: Elaborado pelo autor.

0.81 e 1.5.

A análise da dependência de β em relação ao desvio padrão σ , apresentada na Figura 23, revela que σ , evidencia que a heterogeneidade gerada pela distribuição Gaussiana das constantes elásticas k_{kl} não altera o valor de β . Mesmo aumentando a dispersão das constantes elásticas por meio de σ , o expoente de relaxação permanece essencialmente inalterado. Por outro lado, a Figura 24 indica que β apresenta variação quando o parâmetro efetivo $k_{ef} = k_m$ é modificado, mostrando que a rigidez média da rede exerce influência direta na dinâmica de relaxação.

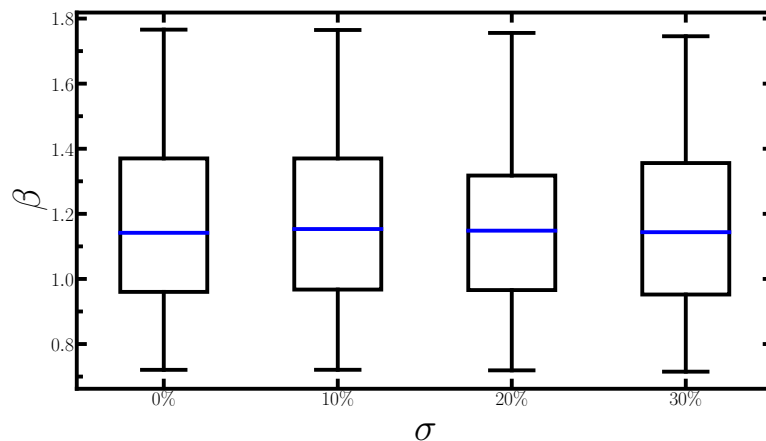
Esses resultados sugerem que, para distribuições Gaussianas, a heterogeneidade, representada por σ não desempenha papel significativo na determinação de β . Contudo, esse

Figura 22 – Nos casos em que a rede apresenta comportamento de lei de potência (sem qualquer filtragem, abrangendo todos os valores analisados), ilustramos a relação entre o expoente de relaxação β e o expoente de arrasto mesoscópico α na rede heterogênea com distribuição Gaussiana.



Fonte: Elaborado pelo autor.

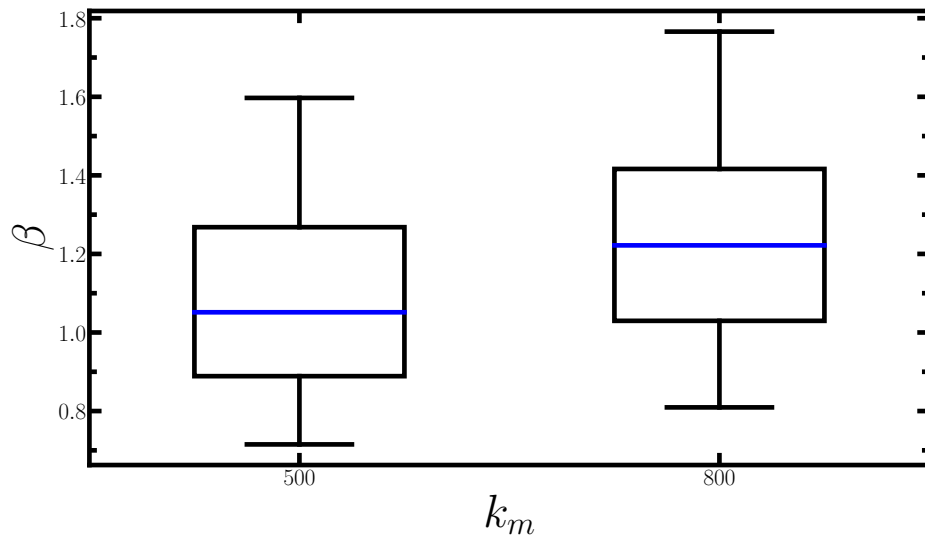
Figura 23 – Nos casos em que a rede apresenta comportamento de lei de potência (sem filtragem, abrangendo todos os valores analisados), apresenta-se a correlação entre o expoente de relaxação β e o desvio padrão σ (expressando a fração de k_m) na distribuição Gaussiana. Observa-se que, mesmo para diferentes valores de σ , não há variação significativa nos valores de β .



Fonte: Elaborado pelo autor.

cenário pode ser diferente para outros tipos de distribuições, como bimodal ou estruturas altamente heterogêneas, exemplificadas pela distribuição em lei de potência, que estabelece ordens de grandeza distintas para as constantes elásticas no material viscoelástico.

Figura 24 – Nos casos em que a rede manifesta um comportamento de lei de potência (sem filtragem, incluindo todos os valores analisados), apresentamos a visualização da relação entre o expoente de relaxação β e o parâmetro efetivo k_{ef} (ou k_m) na rede viscoelástica com distribuição Gaussiana. Observa-se que o expoente β varia de acordo com os diferentes valores de k_{ef} na rede.



Fonte: Elaborado pelo autor.

5.1.3 Análise dos dados da estrutura homogênea por meio de leitura e interpretação dos resultados

Após utilizarmos o método de otimização *L-BFGS-B* C para ajustar os modelos teóricos EXP e PL às 2100 curvas de força numéricas referentes ao caso da estrutura homogênea, obtivemos um conjunto de atributos que compõem a base de dados analisada. Esse conjunto inclui: (i) os parâmetros de entrada (α, γ, k); (ii) as constantes ajustadas dos modelos teóricos; e (iii) parâmetros estatísticos associados às funções erro ou funções de perda (χ_E, χ_P).

Para realizar o pré-processamento e a análise exploratória, organizamos esses resultados em uma estrutura tabular (um *dataframe*) e utilizamos ferramentas de análise de dados em *Python*, em particular a biblioteca *pandas*, amplamente empregada para manipulação e exploração de dados tabulares. Essa biblioteca oferece estruturas de alto desempenho e métodos robustos para limpeza, transformação e análise dos dados, possibilitando extrair *insights* relevantes e, em etapas posteriores, construir modelos de previsão para novos valores de entrada (Scikit-Learn Developers, 2017).

Inicialmente, inspecionamos a base de dados para verificar tipos de variáveis, número de entradas por coluna e presença de valores ausentes ou inconsistentes. Nessa análise preliminar, identificamos 179 valores ausentes (*NaN*) no parâmetro χ_E e 37 *outliers*, detectados por meio de

filtros aplicados ao *dataframe*. Esses *outliers* correspondem a pontos que se afastam de forma significativa do padrão geral da distribuição, não seguindo o comportamento esperado dos dados.

Diante disso, atualizamos o conjunto de dados original para uma versão filtrada, na qual esses 216 registros foram removidos antes da aplicação dos modelos de *Machine Learning*. Essa etapa de limpeza resultou na exclusão de aproximadamente 10,2% do total de 2100 amostras, contribuindo para reduzir o impacto de observações ruidosas ou pouco confiáveis sobre o treinamento e a avaliação dos modelos.

Em seguida, examinamos a dimensionalidade do conjunto de dados, isto é, o número de colunas e o conteúdo informativo de cada uma delas. A identificação de variáveis pouco relevantes ou redundantes é fundamental, pois a redução de dimensionalidade auxilia na mitigação de problemas de *overfitting* em modelos de *Machine Learning*, especialmente quando o número de características é elevado em comparação ao número de amostras. Essa simplificação também melhora o desempenho computacional e facilita a interpretação dos resultados.

Com base nessa análise, removemos colunas consideradas menos informativas para os objetivos deste trabalho. Em particular, descartamos parte dos parâmetros das constantes ajustadas dos modelos teóricos, mantendo apenas o expoente β da lei de potência, que é de maior interesse físico. De forma análoga, eliminamos alguns parâmetros estatísticos redundantes, preservando apenas χ_E e χ_P para a análise dos erros de ajuste, além de manter inalterados os parâmetros de entrada (α, γ, k) . Após essas etapas, o *dataframe* filtrado passou a conter apenas os atributos considerados mais relevantes do ponto de vista físico e preditivo.

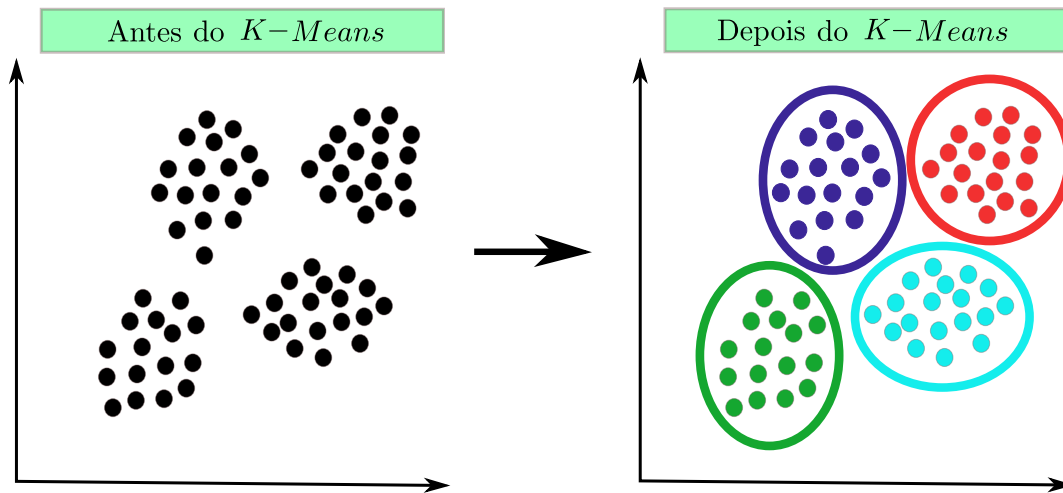
Por fim, aplicamos uma etapa de normalização aos atributos de saída χ_E e χ_P . A normalização foi realizada por meio da transformação linear

$$X_i^{\text{norm}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \quad (5.6)$$

em que X_i representa cada valor observado na coluna X , e X_{\min} e X_{\max} são, respectivamente, os valores mínimo e máximo dessa coluna. Dessa forma, ambos os parâmetros de erro foram ajustados para uma mesma faixa numérica, entre 0 e 1.

A normalização é particularmente importante porque muitos algoritmos de aprendizado de máquina, como regressão linear, *k-means* e redes neurais, são sensíveis à escala dos atributos. Quando uma variável apresenta valores numericamente muito maiores que as demais, ela tende a dominar o processo de ajuste, introduzindo um viés indesejado no modelo. Além disso, a normalização contribui para melhorar a convergência dos algoritmos de otimização utilizados no treinamento, tornando o processo mais estável e eficiente. Um benefício adicional

Figura 25 – Esquema ilustrativo do uso do algoritmo *k-means* no contexto de aprendizado de máquina não supervisionado para *clusterizar* as curvas de força, destacando as etapas de pré-processamento, aplicação do algoritmo e interpretação dos resultados.



Fonte: Elaborado pelo autor.

é a maior clareza na interpretação dos resultados, já que atributos na mesma escala permitem comparar, de maneira mais direta, a importância relativa de cada variável no comportamento final do modelo.

5.1.4 Clusterização das curvas de força associadas aos testes EXP, TR e PL por meio do algoritmo de agrupamento *k-means*, utilizando técnicas de Machine Learning

O agrupamento *k-means* é uma técnica de aprendizado não supervisionado cujo objetivo consiste em particionar um conjunto de dados contendo n observações em k clusters, de modo que cada observação seja atribuída ao grupo cujo centróide apresenta a menor distância em relação a ela, atuando como um protótipo representativo do cluster. A semelhança entre objetos é usualmente avaliada por meio da distância euclidiana entre seus vetores de características, de forma que valores menores dessa distância indicam maior similaridade entre os elementos. O algoritmo busca minimizar a soma das distâncias euclidianas ao quadrado entre cada objeto e o centróide do cluster ao qual foi atribuído, assegurando que os elementos agrupados em um mesmo cluster apresentem alta similaridade interna (Grus, 2015; Géron, 2019).

Como as amostras não possuem rótulos previamente definidos, empregamos o algoritmo de aprendizado não supervisionado para modelar diretamente o conjunto de dados de entrada e realizar a *clusterização*, conforme ilustrado na Figura 25. Nesse contexto, é necessário estabelecer, a priori, o número de grupos finais k nos quais os dados serão particionados. Uma

vez escolhido o valor de k , o algoritmo *k-means* procede à divisão das n observações em k *clusters* por meio das seguintes etapas:

- definição inicial de k centróides, que atuam como protótipos dos *clusters*;
- atribuição de cada objeto de dados ao *cluster* cujo centróide é mais próximo, em termos de distância euclidiana;
- atualização dos centróides, calculados como a média dos objetos atualmente atribuídos a cada *cluster*;
- repetição dos passos de atribuição e atualização até que não haja mais mudanças na composição dos *clusters* ou até que seja atingido um número máximo de iterações;
- obtenção de uma partição final em k grupos, na qual os objetos dentro de um mesmo *cluster* apresentam alta similaridade interna, enquanto a semelhança entre objetos de *clusters* distintos é relativamente baixa.

Para aplicar o *k-means* às curvas de força, utilizamos a implementação disponibilizada na biblioteca *scikit-learn*, amplamente empregada em problemas de *Machine Learning*. Nessa implementação, configuramos o número de *clusters* como $k = 3$, refletindo os três tipos de materiais que desejamos classificar: EXP (exponencial), TR (transição) e PL (lei de potência). A inclusão da categoria TR é justificada pelo fato de algumas curvas apresentarem comportamentos intermediários, que não podem ser associados de forma inequívoca apenas aos padrões EXP ou PL. Assim, a definição de três *clusters* permite representar de maneira mais fiel os diferentes regimes observados nos dados.

Além do número de *clusters*, outros hiperparâmetros relevantes foram ajustados na configuração do algoritmo, como o método de inicialização dos centróides, o número de inicializações independentes e o estado aleatório (*random state*) associado ao gerador de números pseudoaleatórios. Este último garante a reprodutibilidade dos resultados, permitindo que diferentes execuções do algoritmo produzam a mesma partição dos dados.

A *clusterização* foi realizada utilizando como atributos os parâmetros de erro χ_E e χ_P , previamente normalizados na etapa de pré-processamento. Esses dois parâmetros constituem um espaço bidimensional no qual cada curva de força é representada como um ponto, e o algoritmo *k-means* particiona esse espaço em três regiões associadas aos diferentes *clusters*. Ao final do processo, cada curva é atribuída a um dos três grupos, de acordo com o centróide ao qual está mais próxima.

Os centróides dos *clusters* obtidos podem ser interpretados como valores médios (em

termos de χ_E e χ_P) característicos de cada regime (EXP, TR e PL). Além disso, as etiquetas de *cluster* associadas a cada observação foram incorporadas ao *dataframe* filtrado como uma nova coluna categórica, denotada por CLUSTER. Dessa forma, o conjunto de dados final utilizado nas análises passou a conter apenas as colunas $[k, \gamma, \alpha, \chi_E, \chi_P, \beta, \text{CLUSTER}]$, nas quais os parâmetros de entrada (k, γ, α), o expoente β da lei de potência, os erros de ajuste (χ_E, χ_P) e a classificação atribuída pelo algoritmo de *k-means* são reunidos em uma estrutura única, adequada para as etapas posteriores de interpretação física e avaliação dos modelos.

5.1.5 Lidando com valores ausentes: técnica de imputação baseada no *Random Forest Regressor (RFR)*

Na etapa de leitura e interpretação dos dados da estrutura homogênea, realizamos inicialmente uma limpeza correspondente a aproximadamente 10,2% do total de 2100 curvas de força. Essa remoção incluiu registros com valores ausentes ou inconsistentes, resultando em um *dataframe* filtrado, denotado por `dfiltro`, contendo apenas as observações consideradas confiáveis para a análise.

Entretanto, a exclusão de dados pode reduzir a representatividade da amostra e afetar o poder estatístico das análises subsequentes, especialmente em contextos de *Machine Learning*, nos quais o tamanho do conjunto de dados influencia diretamente o desempenho dos modelos. Nesse cenário, torna-se desejável recuperar, tanto quanto possível, a informação associada às amostras descartadas, estimando os valores ausentes de forma sistemática por meio de técnicas de imputação de dados (Scikit-Learn Developers, 2017).

Nosso objetivo consiste em estimar os valores faltantes dos rótulos de saída (χ_E, χ_P, β) a partir dos parâmetros de entrada (k, γ, α). Para isso, adotamos uma abordagem baseada em modelos de aprendizado de máquina supervisionado, treinados sobre `dfiltro`, no qual todos os atributos de entrada e de saída já se encontram completamente definidos. De maneira geral, o procedimento segue os seguintes passos:

1. separar, em `dfiltro`, os atributos de entrada (k, γ, α) e os rótulos de saída (χ_E, χ_P, β) em estruturas distintas;
2. dividir o conjunto de dados com rótulos completos em subconjuntos de treinamento e de teste;
3. treinar diferentes algoritmos de regressão supervisionada nos dados de treinamento;
4. avaliar o desempenho preditivo de cada modelo nos dados de teste, por meio de métricas

Figura 26 – Esquema da divisão do *dataframe* filtrado (*dfiltro*) em 80% para o conjunto de treinamento e 20% para o conjunto de teste, considerando os três atributos de entrada (k, γ, α) e os três rótulos de saída (χ_E, χ_P, β). Essa configuração caracteriza explicitamente um modelo de aprendizado supervisionado.

Conjuntos		Atributos			Rótulos		
		k	γ	α	χ_E	χ_P	β
treinamento	1	k_1	γ_1	α_1	χ_{E1}	χ_{P1}	β_1
	2	k_2	γ_2	α_2	χ_{E2}	χ_{P2}	β_2
	3	k_3	γ_3	α_3	χ_{E3}	χ_{P3}	β_3
	4	k_4	γ_4	α_4	χ_{E4}	χ_{P4}	β_4
testes	5	k_5	γ_5	α_5	χ_{E5}	χ_{P5}	β_5

Fonte: Elaborado pelo autor.

adequadas;

5. selecionar o modelo com melhor desempenho e utilizá-lo para imputar os valores ausentes nas amostras originalmente removidas.

Uma vez que *dfiltro* já está livre de valores ausentes, é possível utilizá-lo diretamente em algoritmos de *Machine Learning* supervisionado. Nessa etapa, construímos dois subconjuntos: um de treinamento, utilizado para ajustar os modelos, e outro de teste, empregado para avaliar a capacidade de generalização desses modelos a novos dados. Essa separação é uma etapa fundamental em aprendizado de máquina, pois permite verificar se o modelo não está apenas “decorando” os dados de treinamento (*overfitting*), mas de fato aprendendo padrões que se estendem a observações não vistas.

No presente trabalho, adotamos uma divisão em que 80% das amostras de *dfiltro* foram designadas ao conjunto de treinamento e 20% ao conjunto de teste, conforme ilustrado na Figura 26. A partir disso, definimos:

- um conjunto de atributos de entrada, contendo apenas as colunas (k, γ, α);
- um conjunto de rótulos de saída, contendo apenas as colunas (χ_E, χ_P, β).

Essa estrutura caracteriza um problema de regressão multivariada, em que o objetivo é prever simultaneamente três saídas contínuas a partir de três entradas contínuas.

Como nosso modelo se baseia em aprendizado supervisionado, avaliamos três algoritmos de regressão amplamente utilizados na literatura:

- Regressão Linear (RL): modelo paramétrico que busca estabelecer uma relação linear entre as variáveis de entrada e as variáveis de saída, representada por uma combinação linear dos atributos;
- *Decision Tree Regressor* (DTR): modelo baseado em árvore de decisão, no qual o espaço de atributos é particionado recursivamente de acordo com regras de divisão que maximizam a redução do erro em cada nó. Cada folha da árvore corresponde a uma previsão para os dados que caem naquela região do espaço;
- *Random Forest Regressor* (RFR): modelo de aprendizado de conjunto (*ensemble*), composto por várias árvores de decisão de regressão. Cada árvore é treinada a partir de um subconjunto aleatório dos dados de treinamento e de um subconjunto aleatório dos atributos, o que aumenta a diversidade entre as árvores. A previsão final é obtida pela média das previsões individuais das árvores, o que tende a reduzir o *overfitting* e melhorar o desempenho preditivo.

Os três modelos foram ajustados utilizando o conjunto de treinamento e avaliados no conjunto de teste por meio do coeficiente de determinação R^2 , métrica padrão para problemas de regressão. A comparação dos resultados mostrou que:

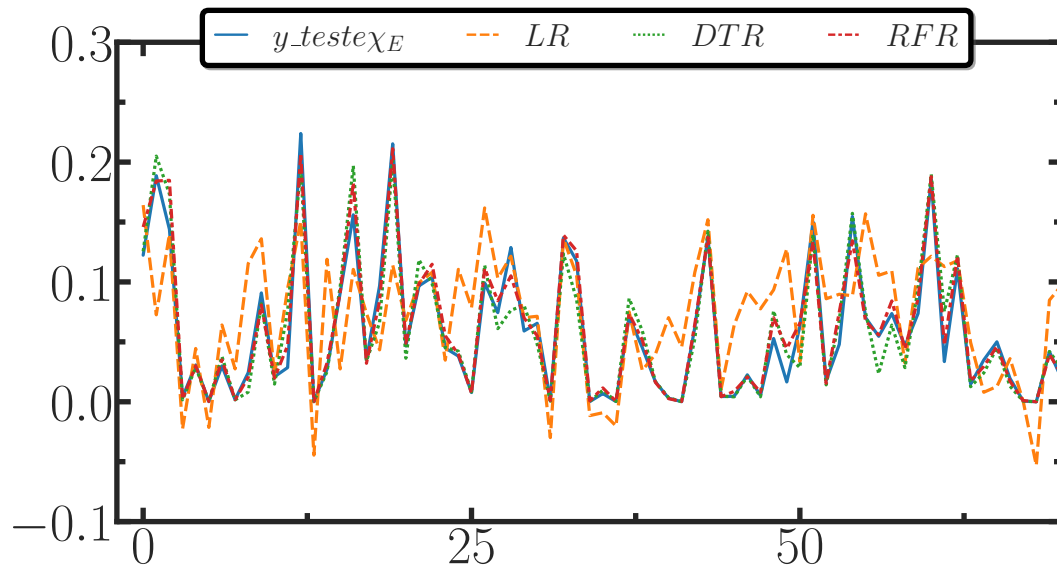
- o modelo RFR apresentou o melhor desempenho preditivo, com $R^2 \approx 0.93$, isto é, cerca de 93% da variância dos dados de teste explicada pelo modelo;
- o modelo DTR obteve desempenho intermediário, com $R^2 \approx 0.84$;
- o modelo baseado em Regressão Linear apresentou desempenho significativamente inferior, com $R^2 \approx 0.22$.

Os resultados, ilustrados na Figura 27, indicam que o *Random Forest Regressor* é o mais adequado entre os modelos considerados para realizar previsões dos rótulos (χ_E, χ_P, β) a partir dos atributos (k, γ, α) . Diante desse desempenho superior, adotamos o RFR como modelo de referência para a imputação dos valores ausentes.

Em seguida, utilizamos o modelo RFR treinado sobre `dfiltro` para estimar os rótulos (χ_E, χ_P, β) das amostras originalmente removidas na etapa de limpeza, isto é, das curvas de força que compunham os 10,2% dos dados faltantes. Para isso:

1. organizamos os parâmetros de entrada (k, γ, α) dessas amostras em um novo *dataframe*, denominado `dfaltantes`;
2. aplicamos o modelo RFR ajustado para prever os valores de (χ_E, χ_P, β) em cada linha de `dfaltantes`;

Figura 27 – Comparação dos modelos de predição *Random Forest Regressor* (RFR), *Regressão Linear* (RL) e *Decision Tree Regressor* (DTR) em relação aos valores de teste de χ_E . Observa-se que o modelo RFR apresenta melhor aderência aos dados de teste, aproximando-se mais dos valores observados.



Fonte: Elaborado pelo autor.

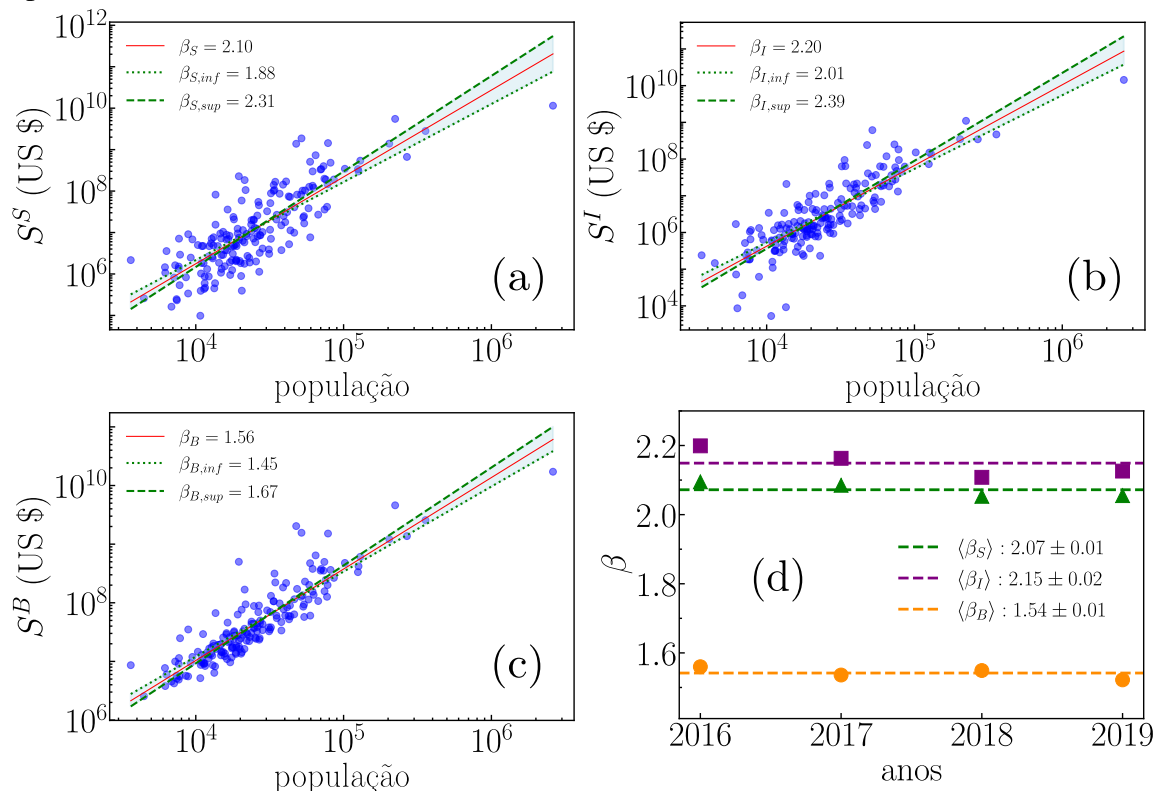
3. incorporamos as previsões de (χ_E, χ_P, β) como novas colunas em *dfaltantes*.

Dessa forma, *dfaltantes* passou a conter seis colunas, correspondentes a $[k, \gamma, \alpha, \chi_E, \chi_P, \beta]$, isto é, os mesmos atributos presentes em *dfiltro*, porém com rótulos imputados por meio do modelo RFR.

Como etapa adicional, associamos a cada uma dessas novas amostras um rótulo de *cluster* (CLUSTER), de forma consistente com a classificação previamente obtida para *dfiltro*. Para isso, utilizamos o modelo de *k-means* ajustado anteriormente, aplicando-o aos pares (χ_E, χ_P) de *dfaltantes*. Assim, cada curva de força imputada foi atribuída ao *cluster* mais próximo no espaço de erros, completando a informação categórica sobre o tipo de comportamento (EXP, PL ou TR).

Por fim, concatenamos os *dataframes* *dfiltro* (com dados originais completos) e *dfaltantes* (com dados imputados), obtendo um novo *dataframe* consolidado, denotado por *df*, contendo novamente as 2100 amostras iniciais. Esse *dataframe* final reúne, em uma única estrutura dada por $[k, \gamma, \alpha, \chi_E, \chi_P, \beta, \text{CLUSTER}]$, combinando a imputação supervisionada via *Random Forest Regressor* para os valores ausentes e a clusterização não supervisionada via *k-means* para a classificação dos três tipos de materiais (EXP, PL e TR). Essa reconstrução completa da base de dados permite a realização de análises posteriores com maior robustez estatística e melhor aproveitamento das informações disponíveis.

Figura 28 – Escalonamento alométrico da atividade comercial municipal. (a–c) Relação entre a população municipal e o valor de todas as transações entre empresas em 2016: (a) vendas para outras cidades, (b) transações intramunicipais e (c) compras de outras cidades. Cada ponto representa um município. As linhas vermelhas representam ajustes bayesianos em escala log–log, com curvas pontilhadas e tracejadas indicando, respectivamente, os limites inferior e superior do intervalo de confiança de 95%. O expoente correspondente, β , captura o escalonamento alométrico anual para cada tipo de fluxo. (d) Estimativas anuais de β para vendas (triângulos verdes), transações intramunicipais (quadrados roxos) e compras (círculos laranja), com linhas tracejadas marcando as médias multianuais: $\langle\beta_S\rangle = 2.07 \pm 0.01$, $\langle\beta_I\rangle = 2.15 \pm 0.02$ e $\langle\beta_B\rangle = 1.54 \pm 0.01$. Os expoentes calculados a partir da agregação completa de quatro anos indicam que o comportamento alométrico das três camadas de comércio no Ceará é estável ao longo do tempo.



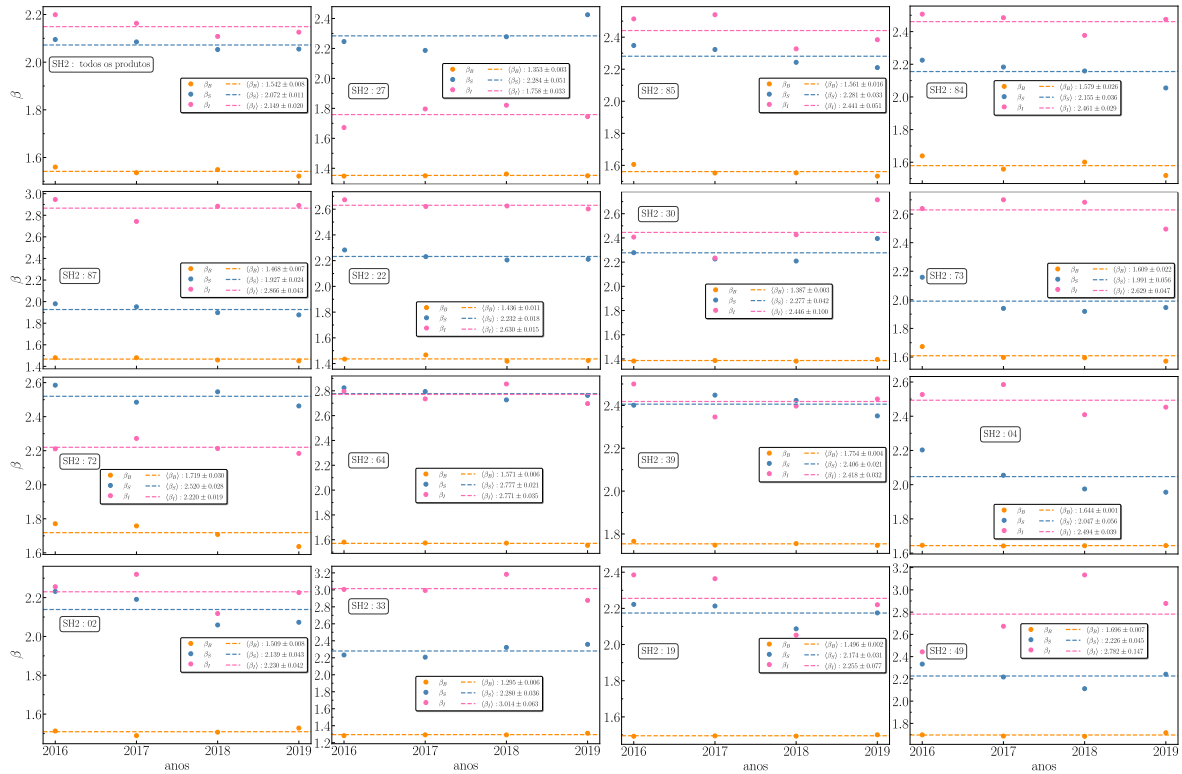
Fonte: Elaborado pelo autor.

5.2 Ciência das Cidades

5.2.1 Escalonamento alométrico dos fluxos comerciais municipais

Os expoentes alométricos para os valores transacionados foram inicialmente estimados para o conjunto agregado de produtos. A Figura 28 apresenta os resultados de 2016: os painéis (a–c) exibem as regressões bayesianas em escala log–log entre a população municipal e os valores totais de comércio para vendas intermunicipais, transações intramunicipais e compras intermunicipais. As compras crescem de forma superlinear com $\beta = 1.56$ (intervalo credível de 95% ([1.45, 1.67])), enquanto as vendas e o comércio interno exibem crescimento ainda mais

Figura 29 – Evolução temporal dos expoentes alométricos β (2016–2019) para cada um dos 15 principais grupos de produtos HS2 e para o agregado “Todos os produtos”. Os pontos representam os valores anuais de β estimados para cada produto e categoria de fluxo (compras, vendas e comércio intraurbano). As linhas tracejadas indicam a média ao longo dos anos, $\langle\beta\rangle$, correspondente a cada produto (ou agregado) e categoria.

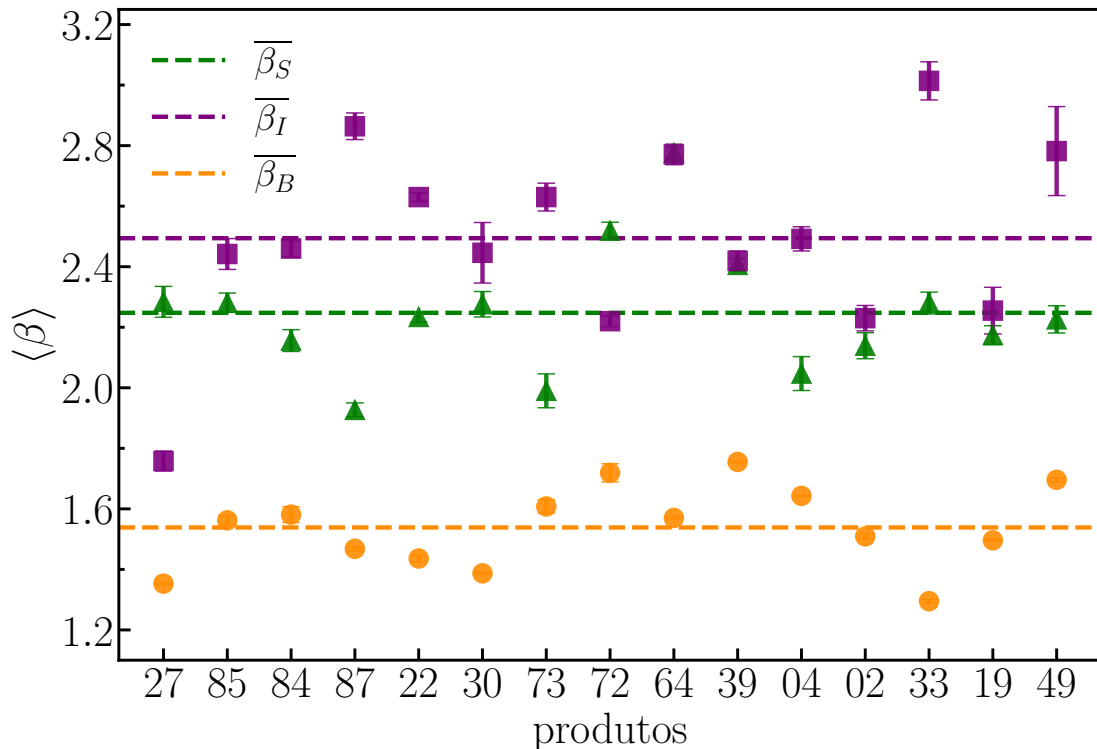


acentuado, com $\beta_S = 2.09$ ([1.88, 2.31]) e $\beta_I = 2.20$ ([2.00, 2.39]). O painel (d) apresenta as estimativas ano a ano de 2016 a 2019. Resultados para as 15 principais categorias de produtos são fornecidos na Figura 29. Como as estimativas anuais exibem variabilidade moderada — especialmente para produtos de baixo volume — adotamos médias multianuais como indicadores mais estáveis do comportamento de escalonamento.

Dois padrões principais emergem. Primeiro, a intensidade de escalonamento difere fortemente entre os tipos de fluxo: as compras aumentam moderadamente com a população, enquanto as vendas e o comércio interno crescem de forma muito mais rápida, com expoentes bem acima de dois. Segundo, essa divergência reflete uma assimetria estrutural entre demanda e oferta: o consumo cresce mais rapidamente do que a população, mas a produção e a circulação interna se amplificam desproporcionalmente em cidades maiores. Esses padrões são consistentes com forças de aglomeração, especialização e redes densas entre empresas, que se fortalecem de maneira desproporcional em municípios mais prósperos e populosos.

A heterogeneidade específica por fluxo também emerge no nível de produto. A

Figura 30 – Expoentes de escalonamento alométrico por categoria de produto. Valores médios anuais do expoente alométrico $\langle \beta \rangle$ para os 15 grupos de produtos mais comercializados (códigos NCM de 2 dígitos). Para cada produto, β é estimado de forma independente para as vendas intermunicipais (β_S , triângulos verdes), transações intramunicipais (β_I , quadrados roxos) e compras intermunicipais (β_B , círculos laranja). As barras de erro representam o erro-padrão ao longo dos quatro anos (2016–2019). Linhas horizontais tracejadas indicam as médias gerais multi-produto para cada tipo de fluxo. Os resultados revelam diferenças sistemáticas de escalonamento entre os produtos: os fluxos intramunicipais exibem o comportamento superlinear mais forte, enquanto as compras apresentam o escalonamento mais fraco.

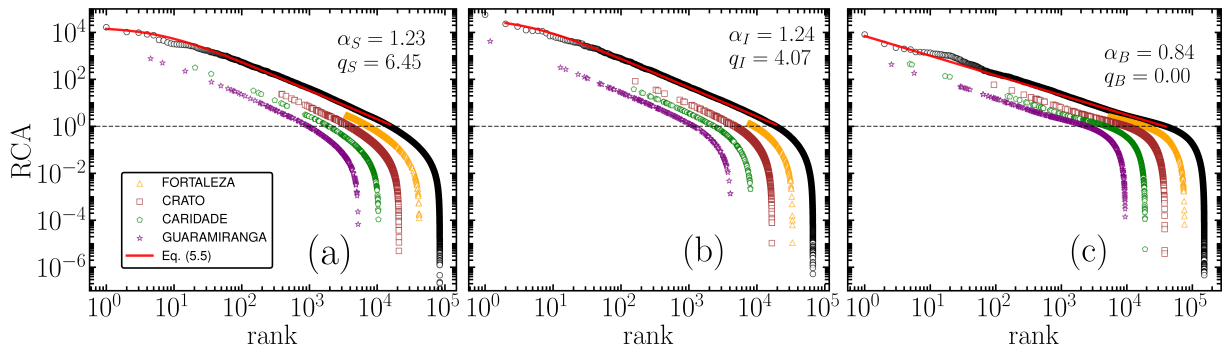


Fonte: Elaborado pelo autor.

Figura 30 apresenta os expoentes médios de escalonamento $\langle \beta \rangle$ para as 15 principais categorias NCM, com barras de erro refletindo a variabilidade entre anos. As médias multianuais são $\overline{\beta_B} = 1.54 \pm 0.03$ para compras, $\overline{\beta_S} = 2.24 \pm 0.05$ para vendas e $\overline{\beta_I} = 2.47 \pm 0.08$ para transações intramunicipais. Os coeficientes de variação são mais baixos para compras ($C_v^B = 0.08$), indicando que a atividade de aquisição escala de forma relativamente uniforme entre os produtos. Em contraste, vendas e fluxos internos mostram dispersão mais ampla ($C_v^S = 0.09$; $C_v^I = 0.13$), refletindo maior sensibilidade às capacidades de comercializações locais, à estrutura das cadeias de suprimento e ao contexto institucional.

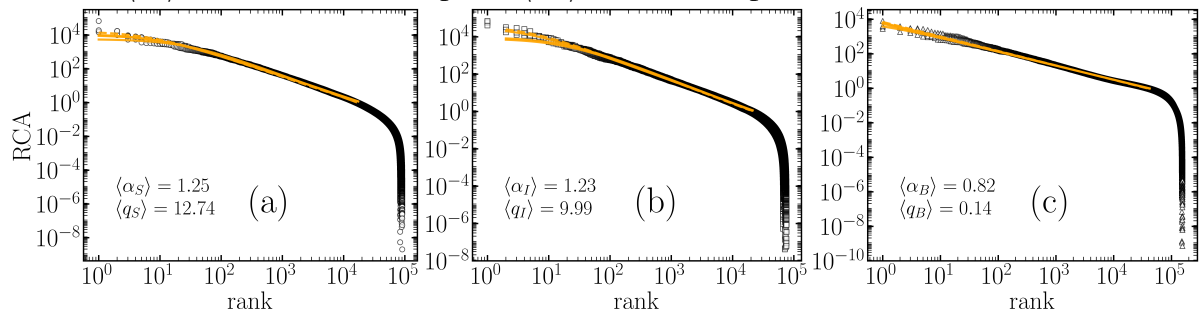
Tomados em conjunto, esses resultados revelam um desequilíbrio sistemático entre demanda e oferta na escala urbana. Cidades maiores não são apenas mercados maiores; elas são desproporcionalmente mais fortes como produtoras e como agentes de comércio interno. Essa assimetria ajuda a explicar desequilíbrios comerciais persistentes entre municípios. Ela destaca

Figura 31 – Estrutura Zipf–Mandelbrot na RCA municipal. Gráficos rank–ordem da Vantagem Comparativa Revelada (RCA) para todos os municípios em 2016, apresentados separadamente para (a) vendas intermunicipais, (b) transações intramunicipais e (c) compras intermunicipais. A curva preta representa a distribuição entre todos os municípios e produtos. Cidades selecionadas são deslocadas horizontalmente para maior clareza visual — Fortaleza (maior), Crato (6ª maior), Caridade (90ª) e Guaramiranga (menor). Para $RCA \geq 1$ (linhas tracejadas), as distribuições seguem uma forma Zipf–Mandelbrot com expoentes $\alpha_S = 1.23$ (vendas), $\alpha_I = 1.24$ (intra) e $\alpha_B = 0.84$ (compras), indicando um decaimento mais acentuado e maior especialização em vendas e atividade interna. Em contraste, as compras apresentam uma cauda mais plana, consistente com um padrão mais diversificado de produtos demandados.



Fonte: Elaborado pelo autor.

Figura 32 – Distribuições de rank-size dos valores de Vantagem Comparativa Revelada (RCA) para os municípios do Ceará nas três modalidades de fluxo: (a) vendas, (b) transações intramunicipais e (c) compras. Para cada categoria, são mostradas as curvas de 2016–2019, que exibem grande sobreposição e, portanto, robustez temporal. A linha laranja em cada painel representa o ajuste Zipf–Mandelbrot, $RCA(k) \sim (k + q)^{-\alpha}$, indicando um claro decaimento em lei de potência para $RCA > 1$. Os expoentes médios de Zipf ao longo do período são $\langle \alpha_S \rangle = 1.25$ (vendas), $\langle \alpha_I \rangle = 1.23$ (intramunicipal) e $\langle \alpha_B \rangle = 0.82$ (compras).



Fonte: Elaborado pelo autor.

possíveis alavancas de política: expandir portfólios locais de capacidades e reduzir resistências de coordenação, por meio do desenvolvimento de fornecedores, do aprimoramento logístico e da capacitação da força de trabalho, tende a gerar os ganhos mais significativos em vendas e atividade do mercado interno, onde os retornos de escalonamento são mais acentuados (Duranton; Puga, 2004).

5.2.2 *Estrutura de Zipf-Mandelbrot das Capacidades Especializadas*

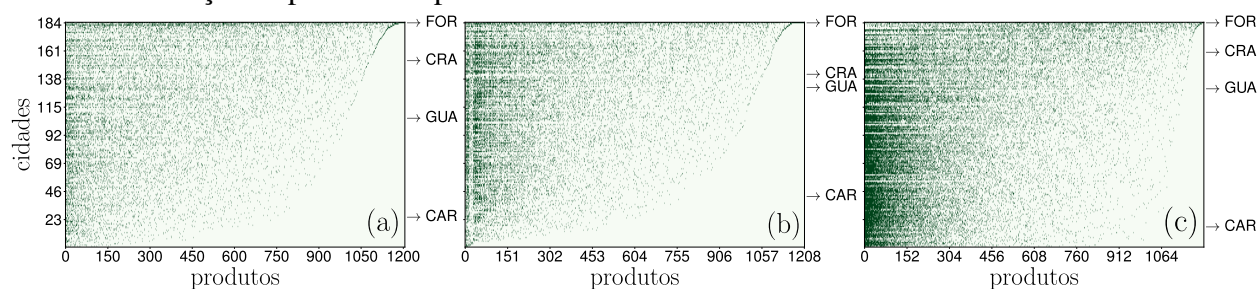
Classificamos todos os pares cidade–produto por valores de RCA em ordem decrescente. Para $RCA > 1$, as distribuições seguem a lei de Zipf–Mandelbrot (Liu; Gao, 2019; Mandelbrot, 1953; Mandelbrot, 1982), conforme descrito na Eq. 4.5. O padrão resultante, um Zipf truncado composto por uma lei de potência seguida por um corte abrupto, captura a tensão fundamental entre diversificação e especialização. O trecho em lei de potência reflete a aquisição cumulativa e autorreforçada de capacidades em um conjunto restrito de atividades, enquanto o corte sinaliza limites estruturais à diversificação impostos por escala, qualificação e condições institucionais. Do ponto de vista da complexidade, esses padrões zipfianos revelam desigualdades persistentes na especialização setorial e o espaço limitado de diversificação disponível para municípios pequenos e médios. Entretanto, as três camadas de comércio exibem regimes de escalonamento distintos.

A Figura 31 apresenta as curvas correspondentes de rank-size para 2016. As compras exibem o menor expoente ($\alpha_B = 0.84$), produzindo uma curva mais plana e uma distribuição mais homogênea de vantagem comparativa entre os municípios. As vendas e as transações intraurbanas apresentam expoentes mais altos ($\alpha_S = 1.23$; $\alpha_I = 1.24$), indicando uma queda mais acentuada e maior concentração de RCA em um subconjunto reduzido de pares cidade–produto. Essa hierarquia é compatível com a presença de polos centralizados de produção, indústrias localmente especializadas e redes logísticas assimétricas que reforçam a concentração do lado da oferta em comparação com a demanda.

As curvas de rank-size de RCA também revelam um gradiente explícito de especialização diversificação entre os municípios. Cidades grandes e diversificadas, como Fortaleza (a maior) e Crato (a sexta maior), exibem curvas curtas e fortemente concentradas próximas de $RCA \approx 1$, refletindo portfólios amplos de produtos que espelham a estrutura do estado. Em contraste, municípios pequenos e altamente especializados, como Caridade (90°) e Guaramiranga (o menor), exibem caudas longas com $RCA \gg 1$ em alguns produtos de nicho e $RCA \lesssim 1$ nos demais, consistentes com conjuntos estreitos de capacidades. Essa heterogeneidade transversal se agrega na estrutura estatal de Zipf–Mandelbrot para $RCA \geq 1$: a diversificação comprime a vantagem comparativa, enquanto a especialização produz valores extremos.

Especializações fortes dominam a cauda direita. Apenas 18–28% dos pares cidade–produto satisfazem $RCA \geq 1$, o que significa que a maior parte dos pares está abaixo da média estadual. Assim, a especialização genuína concentra-se em uma minoria de capacida-

Figura 33 – Matrizes binárias de RCA por camada de comércio. As matrizes cidade–produto em 2016 mostram quais municípios exibem Vantagem Comparativa Revelada ($RCA \geq 1$) em (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Cada ponto verde marca um par cidade–produto no qual o município c se especializa no produto p . As linhas correspondem aos 184 municípios ordenados por Fitness crescente, e as colunas correspondem aos 1.228 produtos NCM de quatro dígitos ordenados por Complexidade crescente. Quatro municípios de referência são destacados: Fortaleza (FOR, mais populoso), Crato (CRA, 6º maior), Caridade (CAR, 90º) e Guaramiranga (GUA, o menor). Exemplos de produtos de menor a maior complexidade em cada camada incluem: em vendas, do NCM 7311 (contêineres para gases comprimidos ou liquefeitos) ao NCM 8904 (embarcações para reboque ou empurrar outras embarcações); em transações intraurbanas, do NCM 6904 (tijolos/ladrilhos cerâmicos) ao NCM 8478 (máquinas para preparação de tabaco); em compras, do NCM 2710 (óleos de petróleo) ao NCM 5108 (fios de pelos finos de animais). As matrizes revelam uma estrutura fortemente aninhada: municípios de alto Fitness participam tanto de produtos complexos quanto de produtos simples, enquanto municípios de baixo Fitness permanecem concentrados em um conjunto estreito de bens de baixa Complexidade. Notavelmente, apesar de ter a menor população, Guaramiranga posicionou-se consistentemente acima de muitos municípios em Fitness, refletindo uma diversificação impulsionada por sua economia orientada ao turismo.



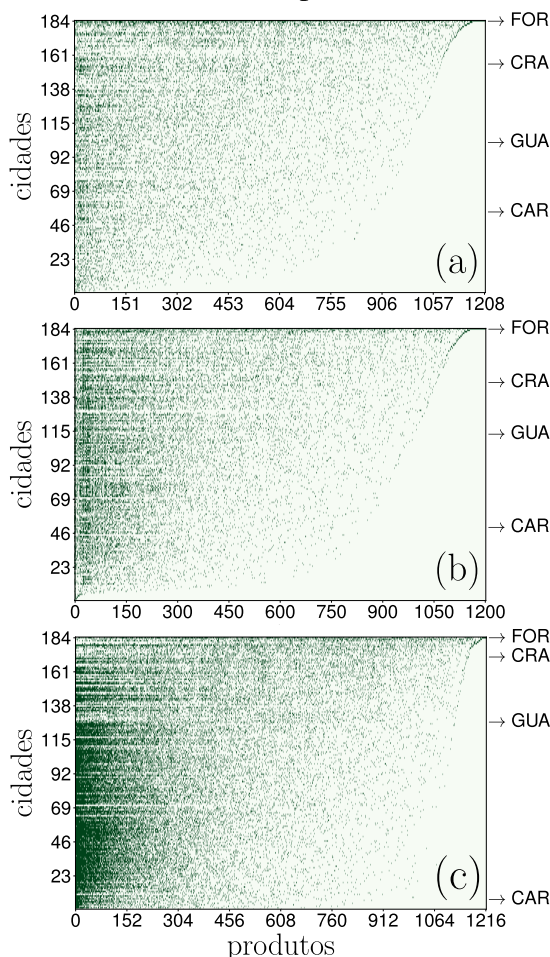
Fonte: Elaborado pelo autor.

des. Esses padrões são notavelmente estáveis ao longo do tempo, com os dados praticamente colapsando uns sobre os outros para 2016–2019, como mostrado na Figura 32. Os expoentes médios de Zipf–Mandelbrot são $\langle \alpha_S \rangle = 1.25$, $\langle \alpha_I \rangle = 1.23$ e $\langle \alpha_B \rangle = 0.82$. Essa persistência temporal sugere que a especialização relativa é ancorada por fundamentos de evolução lenta, como capacidades acumuladas, estrutura das cadeias de suprimento e resistências institucionais, que se transformam em horizontes plurianuais mesmo quando os volumes totais de comércio flutuam.

5.2.3 *Fitness e Complexidade*

A Figura 33 revela uma estrutura hierárquica pronunciada de cidade–produto em todas as camadas de comércio. Municípios de alto Fitness ocupam regiões amplas da matriz, incluindo muitos produtos de alta complexidade, enquanto municípios de baixo Fitness exibem incidências esparsas concentradas em bens ubíquos e de baixa complexidade. Essa hierarquia aninhada aparece em vendas intermunicipais, transações intramunicipais (internas) e compras

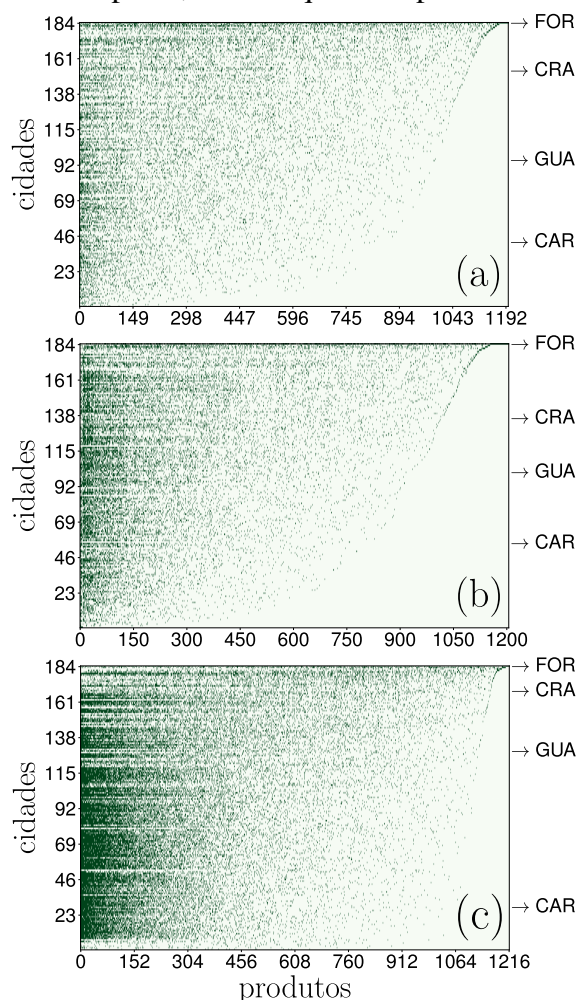
Figura 34 – Mesmo procedimento da Figura 33 para 2017. Extremos por camada: vendas — menor NCM 7311 (contêineres para gases comprimidos ou liquefeitos), maior NCM 8604 (veículos para inspeção e manutenção de ferrovias ou similares); mercado interno — menor NCM 8201 (pás, enxadas, picaretas, forcados, garfos, ovos e raspadores; machados, capões e ferramentas semelhantes com fio), maior NCM 2913 (derivados halogenados, sulfonados, nitrados ou nitrosados dos produtos da posição 29.12); compras — menor NCM 2710 (óleos de petróleo), maior NCM 2616 (minérios de metais preciosos e seus concentrados).



Fonte: Elaborado pelo autor.

intermunicipais, embora com diferenças visíveis em densidade. As camadas do lado da oferta (vendas e circulação interna) formam faixas mais espessas e contínuas entre cidades diversificadas, indicando que esses municípios não apenas produzem uma maior variedade de bens, como também movimentam internamente essa produção em larga escala, o que sugere capacidades de comercialização mais desenvolvidas e um circuito econômico interno mais ativo. Em contraste, a camada de compras é mais fina e distribuída de maneira mais difusa, refletindo uma demanda que escala com o tamanho urbano, mas se diversifica menos. Os casos destacados na Figura 33 (Fortaleza-FOR, Crato-CRA, Caridade-CAR, Guaramiranga-GUA) ilustram esse gradiente: Fortaleza, a maior cidade do estado, abrange tanto bens complexos quanto bens simples em todas as camadas, enquanto municípios menores se concentram em nichos mais

Figura 35 – Mesmo procedimento da Figura 33 para 2018. Extremos por camada: vendas — menor NCM 0102 (animais vivos da espécie bovina), maior NCM 2610 (minérios de cromo e seus concentrados); mercado interno — menor NCM 0102 (animais vivos da espécie bovina), maior NCM 7105 (pó de diamante, pedras preciosas, semipreciosas ou sintéticas); compras — menor NCM 2202 (águas, incluindo águas minerais e águas gaseificadas), maior NCM 8524 (módulos de visualização de tela plana, mesmo que incorporem telas sensíveis ao toque).

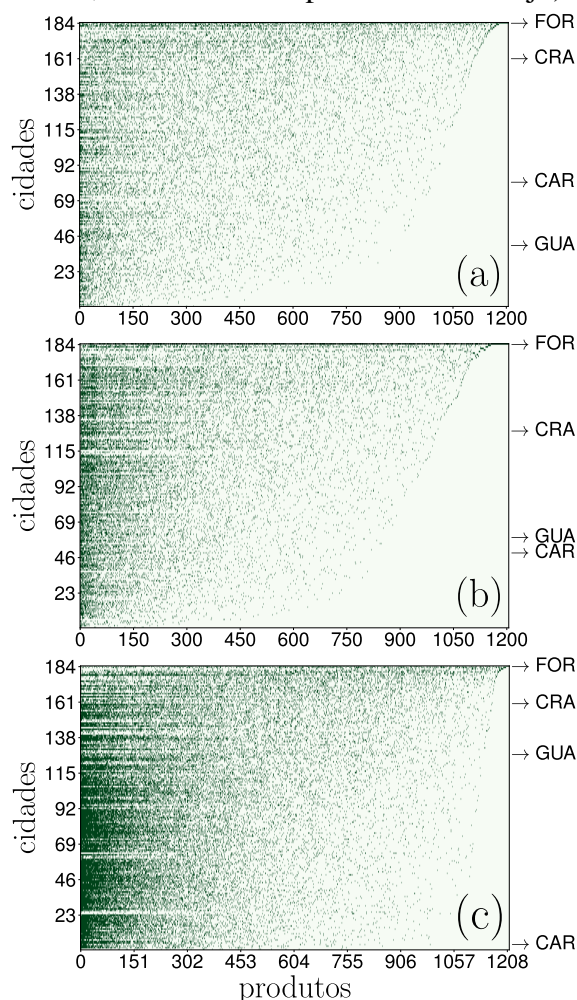


Fonte: Elaborado pelo autor.

estreitos, com portfólios comercializados menos diversificados. Essa configuração reflete uma maior especialização setorial, o que, embora possa ser uma estratégia compatível com a escala dessas cidades, também as torna mais sensíveis a choques que atinjam diretamente os setores em que são especializadas. As análises das matrizes binárias de RCA por camada de comércio referentes aos demais anos encontram-se disponíveis nas Figuras 34 (2017), 35 (2018) e 36 (2019).

Municípios do Ceará apresentam trajetórias heterogêneas no plano Fitness-per-capita versus PIB-per-capita entre 2016 e 2019 (Figura 37), indicando que competitividade estrutural e renda coevoluem, mas de forma desigual entre as diferentes camadas de comércio. Grandes cidades, especialmente Fortaleza, deslocam-se ao longo de fronteiras de alto Fitness e alta renda,

Figura 36 – Mesmo procedimento da Figura 33 para 2019. Extremos por camada: vendas — menor NCM 7311 (contêineres para gases comprimidos ou liquefeitos), maior NCM 6703 (cabelo disposto na mesma direção, fino, descolorido ou preparado de outra forma); mercado interno — menor NCM 0102 (animais vivos da espécie bovina), maior NCM 2307 (borras de vinho; tártaro em bruto); compras — menor NCM 2710 (óleos de petróleo), maior NCM 5005 (filamentos de resíduos de Seda, não embalados para venda a varejo).

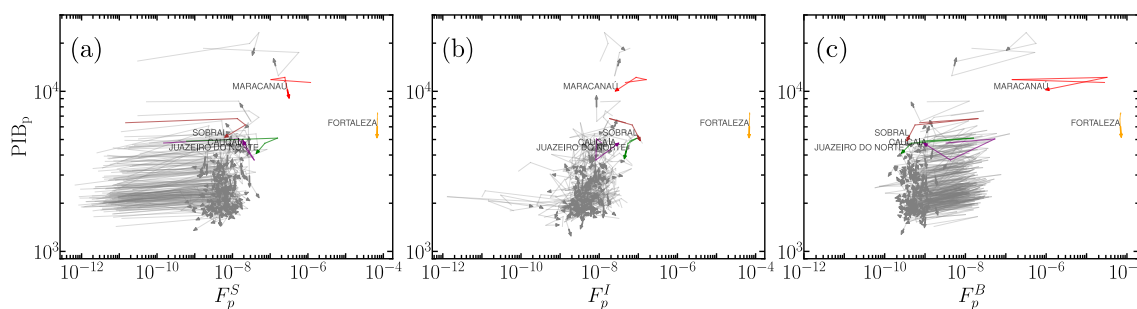


Fonte: Elaborado pelo autor.

enquanto municípios de médio e pequeno porte exibem trajetórias mais curtas e irregulares. A relação é mais forte nas camadas de vendas e intra-cidade, onde a acumulação de capacidades acompanha de perto o crescimento da renda, e mais fraca na camada de compras, no qual aumentos de demanda não são acompanhados por avanços comparáveis em sofisticação estrutural. Esses padrões apontam para persistentes lacunas de capacidades e sugerem que a atualização é dependente da trajetória, avançando mais rapidamente onde redes produtivas e circulação interna já são densas.

As propriedades de escalonamento do Fitness fornecem uma quantificação mais

Figura 37 – Trajetórias de 2016 a 2019 no plano Fitness–PIB. A coevolução do Fitness per capita e do Produto Interno Bruto per capita dos municípios é mostrada para as três camadas de comércio, com eixos em escala logarítmica: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Cada seta representa a trajetória de quatro anos de um município, de 2016 a 2019. Os cinco municípios mais populosos — Fortaleza (laranja), Caucaia (roxo), Juazeiro do Norte (verde), Maracanaú (vermelho) e Sobral (marrom) — são destacados para ilustrar os caminhos desenvolvimentistas heterogêneos observados no estado. As trajetórias revelam como a competitividade estrutural (Fitness) e os níveis de renda (PIB) evoluem conjuntamente em diferentes contextos transacionais, com alguns municípios exibindo ganhos simultâneos em ambas as dimensões, enquanto outros permanecem essencialmente estacionários.



Fonte: Elaborado pelo autor.

formal dessas relações. Expressando a alometria como

$$F \sim (\text{população})^{\beta^F}, \quad (5.7)$$

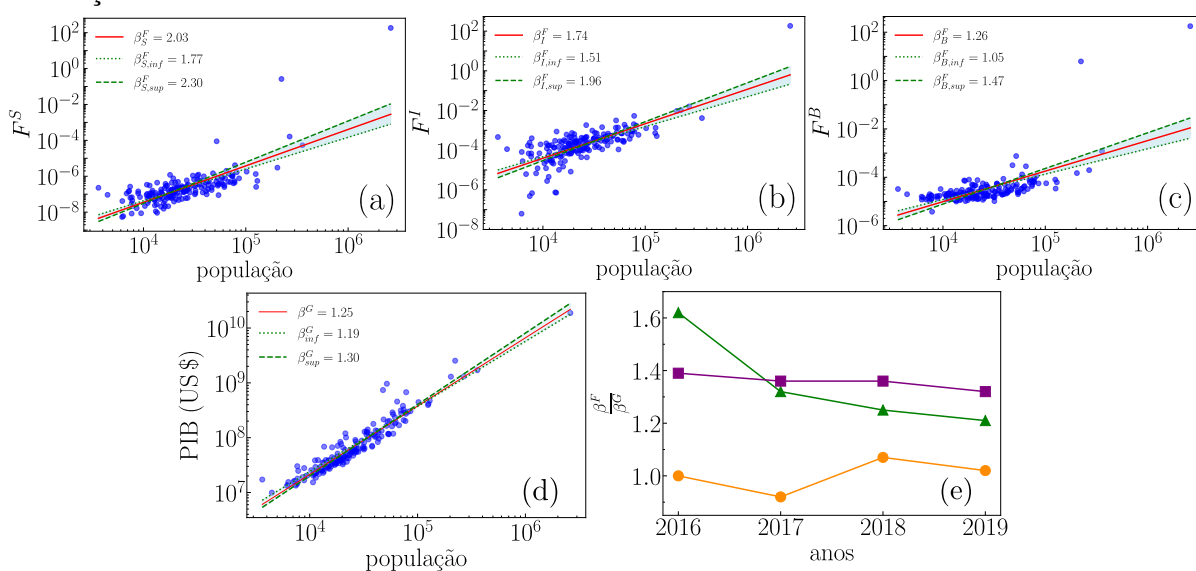
observamos um comportamento claramente superlinear em todas as camadas de comércio, mas com inclinações marcadamente diferentes (Figuras 46(a–c)). Para 2016, os expoentes são $\beta_S^F = 2.03$ para vendas intermunicipais, $\beta_I^F = 1.74$ para transações intra-cidade, e um valor substancialmente mais fraco de $\beta_B^F = 1.26$ para compras intermunicipais. Resultados para 2017–2019, apresentados no apêndice E, exibem a mesma estrutura qualitativa: o Fitness cresce muito mais rapidamente com a população nas camadas de oferta e recirculação, enquanto permanece próximo da proporcionalidade na camada de compras.

Como o Produto Interno Bruto também escala com a população com expoente β^G (Figura 46(d)), a relação entre Fitness e renda pode ser reescrita como

$$F \sim (\text{PIB})^{\beta^F/\beta^G}. \quad (5.8)$$

Nesse contexto, porém, é mais informativo interpretar o Fitness como uma medida estrutural que antecede a renda: isto é, variações nas capacidades de comercialização (capturadas pelo Fitness) estão associadas a variações sistemáticas no PIB. A razão β^F/β^G é fortemente superlinear para vendas, moderadamente superlinear para o comércio interno, e aproximadamente linear para compras. Assim, dobrar a renda municipal mais do que dobra o Fitness associado às atividades

Figura 38 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2016 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 2.03$, $\beta_I^F = 1.74$ e $\beta_B^F = 1.26$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.25$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%). (e) Razoes anuais β^F/β^G para 2016–2019. O Fitness associado a vendas (verde) e circulação interna (roxo) cresce superlinearmente com a renda ($\beta_S^F/\beta^G \approx 1.62$ e $\beta_I^F/\beta^G \approx 1.39$). Em contraste, o Fitness associado a compras (laranja) cresce de forma aproximadamente proporcional ($\beta_B^F/\beta^G \approx 1$), indicando que capacidades estruturais se expandem de maneira desproporcional com a prosperidade apenas nos lados da oferta e da circulação interna.

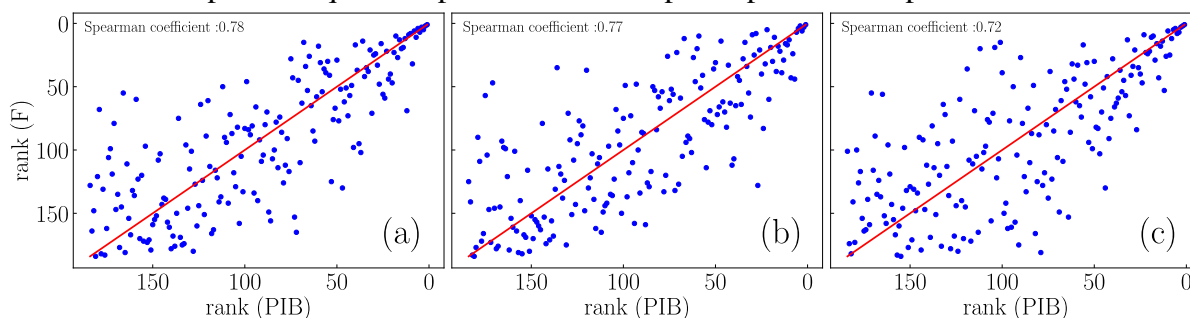


Fonte: Elaborado pelo autor.

de venda e circulação interna, enquanto gera aumentos aproximadamente proporcionais no Fitness relacionado às compras. Essa assimetria entre camadas indica que a acumulação de capacidades é impulsionada principalmente pela produção e pela recirculação interna, ao passo que as capacidades do lado da demanda escalam de forma mais uniforme com o tamanho do mercado.

A associação entre os rankings de Fitness e de PIB oferece uma perspectiva adicional sobre essa relação. Em todas as redes, há correlação positiva entre Fitness e PIB, como mostrado na Figura 39, com coeficientes de Spearman são 0.78 para vendas, 0.77 para transações internas e 0.72 para compras, indicando uma forte associação monotônica entre as ordenações das cidades. Quando consideramos o ajuste linear entre Fitness e PIB, os coeficientes de determinação são $R^2=0.61$ para vendas, $R^2=0.59$ para transações internas e $R^2=0.52$ para compras, o que revela uma relação estatisticamente robusta, porém com dispersão considerável em torno da tendência

Figura 39 – Comparação rank–rank entre o Fitness municipal e o PIB nas três camadas de comércio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Cada ponto corresponde a um município, plotado por sua posição no ranking de PIB (eixo x) e no ranking de Fitness (eixo y). A linha vermelha e um ajuste de mínimos quadrados mostrado apenas para visualização. Os coeficientes de correlação de Spearman são 0.78 para vendas, 0.77 para transações internas e 0.72 para compras, indicando uma associação monotonicamente forte entre produto econômico e capacidade estrutural, com alinhamento ligeiramente mais fraco no lado da demanda. Essas correlações corroboram os resultados de escalonamento, que mostram que capacidades ligadas a oferta e circulação interna acompanham a renda mais de perto do que as capacidades de compra impulsionadas pela demanda.



Fonte: Elaborado pelo autor.

média. A correlação mais fraca na camada de compras reflete seu caráter mais difuso e orientado pela demanda, consistente com a escala quase linear do Fitness associado às atividades de compra em relação à renda. Por outro lado, o alinhamento mais substancial nas redes de vendas e de comércio interno reflete seu comportamento superlinear e reforça a interpretação de que as capacidades estruturais estão mais estreitamente vinculadas às atividades do lado da oferta e da recirculação interna. Em conjunto, esses padrões baseados em rankings corroboram os resultados alométricos e de complexidade: cidades mais ricas ocupam consistentemente posições de maior Fitness de comercialização e circulação interna, enquanto o componente de compras isoladamente é um indicador mais fraco de sofisticação estrutural.

De modo geral, as evidências mostram que população e renda não se traduzem de forma uniforme em ganhos de capacidades nas diferentes camadas de comércio. A comercialização e a circulação interna constituem os principais canais pelos quais novas capacidades emergem e se expandem, beneficiando-se de forma desproporcional de forças de aglomeração, como bases de fornecedores mais profundas, transbordamentos de conhecimento e redes interempresas densas. Em contraste, a atividade de compras escala de maneira mais uniforme com o tamanho do mercado e desempenha um papel mais limitado na formação de capacidades. Esses resultados destacam a natureza assimétrica da acumulação de capacidades em economias urbanas e reforçam a importância de considerar múltiplas dimensões do comércio ao avaliar a competitividade regional.

5.2.4 Limitações

Os dados de NF-e capturam principalmente transações formais de bens entre empresas; atividades informais e muitos serviços estão sub-representados. Empresas com múltiplos estabelecimentos podem registrar operações entre municípios a partir de unidades centralizadas de faturamento, o que pode distorcer alguns fluxos. Além disso, categorias HS4 podem agrupar itens heterogêneos. Esses fatores provavelmente geram estimativas conservadoras de concentração; ainda assim, os padrões de escalonamento e as distribuições do tipo Zipf–Mandelbrot são pronunciados e robustos ao longo dos anos e entre diferentes níveis de agregação de produtos.

6 CONCLUSÃO

Nesta tese, investigamos dois sistemas complexos de naturezas distintas, um físico mesoscópico baseado em redes viscoelásticas e um urbano-econômico estruturado a partir de fluxos de multicamadas de comércio entre municípios, e mostramos que ambos exibem regularidades de escala e leis de potência como assinaturas estruturais de sua organização coletiva. Ao longo do trabalho, demonstramos que uma abordagem orientada por dados, combinada à modelagem de sistemas complexos, funciona como a ponte metodológica capaz de revelar, comparar e interpretar fenômenos emergentes nesses contextos tão diferentes. Embora esses sistemas pertençam a domínios distintos, concluímos que ambos exibem características universais da complexidade: múltiplos agentes interagindo localmente, relações não lineares, heterogeneidade estrutural, dinâmica multiescala e o surgimento de regularidades coletivas que não podem ser reduzidas à soma das partes.

No sistema viscoelástico, analisamos inicialmente o caso homogêneo, no qual todas as ligações possuem a mesma constante elástica. Concluímos que a emergência de relaxações do tipo *power-law* emerge, no modelo, em associação direta com a dissipação não linear quando $0.30 \lesssim \alpha \lesssim 0.45$, desde que $k \geq 800$ e $\gamma \geq 80$. Dentro desse intervalo crítico, verificamos que a relaxação coletiva segue $F(t) \sim t^{-\beta}$, com β diminuindo de aproximadamente 1.38 para 1.07 à medida que α aumenta de 0.30 para 0.45, evidenciando uma relação sistemática entre a dissipação mesoscópica e o regime de relaxação macroscópico. Esse resultado confirma que, mesmo em uma rede estruturalmente ordenada, a combinação entre interação elástica e dissipação sublinear gera propriedades macroscópicas emergentes que não podem ser previstas a partir dos componentes individuais.

Ao introduzirmos heterogeneidade Gaussiana nas interações elásticas, com $k_{kl} \sim \mathcal{N}(k_m, \sigma)$, constatamos que a complexidade se intensifica. Para $k_m \geq 500$ e $\gamma \geq 80$, independentemente de σ , observamos que a separação entre regimes EXP e PL torna-se mais nítida por meio do parâmetro α . Verificamos que respostas exponenciais predominam para $\alpha \lesssim 0.45$ e $\alpha \gtrsim 0.70$, enquanto regimes *power-law* emergem principalmente no intervalo $0.20 \lesssim \alpha \lesssim 0.55$. Nas regiões em que esses intervalos se sobrepõem, identificamos coexistência estatística entre respostas EXP e PL, produzindo uma região de transição classificatória compatível com a presença de fronteiras entre regimes dinâmicos distintos. Para todos os casos classificados como PL, observamos que β se concentra aproximadamente no intervalo $0.81 \lesssim \beta \lesssim 1.50$. Concluímos, assim, que o expoente de arrasto α funciona como um verdadeiro parâmetro crítico, regulando transições

emergentes e organizando a dinâmica coletiva do material.

Em todas essas análises, empregamos técnicas de Ciência de Dados como clusterização *k-means*, RFR para imputação, diagramas de coordenadas paralelas e a exploração de um espaço paramétrico com mais de 5400 simulações para a rede heterogênea Gaussiana e 2100 simulações para a rede homogênea para mapear regimes, identificar fronteiras entre comportamentos distintos, caracterizar propriedades multiescala, reduzir incertezas classificatórias e interpretar padrões emergentes. Os resultados obtidos reforçam que a interação entre modelagem física e Ciência de Dados é essencial para revelar padrões profundos em sistemas complexos materiais.

No sistema urbano-econômico, analisamos fluxos organizados em três camadas, compras, vendas e circulação interna, entre 184 municípios. Verificamos que os fluxos seguem relações alométricas do tipo $S \sim \text{population}^\beta$, com $\beta \approx 1.54$ para compras, $\beta \approx 2.07$ para vendas e $\beta \approx 2.15$ para circulação interna. Esses resultados nos permitiram concluir que cidades maiores não são apenas versões ampliadas de cidades menores, mas sistemas qualitativamente distintos, exibindo retornos crescentes à escala e forte concentração funcional.

Também constatamos que as distribuições de Vantagem Comparativa Revelada seguem a lei de Zipf–Mandelbrot, descrita por $\text{RCA}(\text{rank}) \sim (\text{rank} + q)^{-\alpha}$, revelando hierarquia e especialização estrutural estáveis ao longo do tempo. Ademais, observamos que o Fitness municipal escala superlinearmente com a população, seguindo $F \sim \text{population}^{\beta^F}$, com expoentes distintos entre as camadas, o que reforça que capacidades produtivas se acumulam de forma desproporcional em centros urbanos maiores. Esse conjunto de evidências demonstra que a economia urbana opera como um sistema adaptativo complexo: agentes heterogêneos interagem localmente, gerando padrões globais emergentes; retroalimentações reforçam especializações produtivas; capacidades evoluem de forma *path-dependent*; e a estrutura da rede se reorganiza no tempo em resposta a mudanças demográficas, logísticas e econômicas. Tais resultados confirmam que cidades constituem sistemas complexos onde heterogeneidade, hierarquia, *nestedness*, dependência de trajetória e emergência de padrões coletivos desempenham papéis fundamentais.

Novamente, foi por meio de técnicas de Ciência de Dados aplicadas a grandes volumes de dados (*big data*), regressões bayesianas, análise de redes, extração de expoentes, ordenamento Zipf–Mandelbrot e medidas iterativas como Fitness-Complexity que nos permitiu revelar as estruturas profundas do sistema urbano, identificar regularidades macroscópicas e mapear como capacidades produtivas e fluxos comerciais se organizam.

Ao integrar os dois estudos, observamos que o elo conceitual desta tese não reside apenas na presença de leis de potência em ambos os sistemas, mas no problema científico comum que eles representam: a inferência de propriedades coletivas emergentes em sistemas compostos por muitos elementos interagentes a partir de dados locais de alta dimensionalidade. Concluimos que leis de potência aparecem de forma universal, tanto na dinâmica de materiais viscoelásticos quanto na organização econômico-urbana. Em ambos os domínios, observamos fenômenos genuinamente multiescala, nos quais processos microscópicos, mesoscópicos e macroscópicos se acoplam para produzir regimes coletivos emergentes. Nesses contextos, observamos comportamentos sem uma escala característica dominante, compatíveis com leis de potência e outras regularidades do tipo *scale-free*. Múltiplas escalas interagem simultaneamente e nenhuma prevalece, resultando em comportamentos de cauda pesada e padrões robustos de auto-organização.

Identificamos, ainda, a presença de fronteiras críticas entre regimes, mecanismos de cooperação entre múltiplos elementos, reorganização coletiva e forte dependência de parâmetros mesoscópicos, como o expoente de arrasto α e o expoente de relaxação β no sistema viscoelástico, bem como o expoente alométrico β e o expoente de Zipf–Mandelbrot α no sistema urbano. Em síntese, demonstramos que a complexidade não é um atributo isolado, mas um fenômeno estrutural que atravessa sistemas físicos e socioeconômicos, manifestando-se por meio de leis de potência, heterogeneidade, emergência e organização multiescala.

Por fim, concluimos que a Ciência de Dados, ao lidar com grandes volumes de dados, espaços paramétricos amplos, algoritmos iterativos, métodos de classificação e análise de redes torna-se indispensável para revelar a organização profunda desses sistemas. Assim, esta tese demonstra que fenômenos tão distintos podem ser compreendidos a partir de princípios unificadores da teoria de sistemas complexos e que seu estudo exige, necessariamente, uma abordagem orientada por dados.

REFERÊNCIAS

- ACHAR, B. N. N.; HANNEKEN, J. W. Microscopic formulation of fractional calculus theory of viscoelasticity based on lattice dynamics. **Physica Scripta**, v. 2009, n. T136, p. 014011, 2009.
- ADEDIRAN, E. Complexity theory; unraveling the fabric of intricacy. **SSRN**, v. 1, p. 1–17, 2024.
- ALBEAIK, S.; KALTENBERG, M.; ALSALEH, M.; HIDALGO, C. A. Improving the economic complexity index. **arXiv preprint arXiv:1707.05826**, 2017.
- ALDHEEB, M. A.; ASRAR, W.; SULAEMAN, E.; OMAR, A. A. A review on aerodynamics of nonflapping bird wings. **Journal of Aerospace Technology and Management**, v. 8, p. 7–17, 2016.
- ALLEN, M. P. Introduction to molecular dynamics simulation. In: **Computational Soft Matter: From Synthetic Polymers to Proteins**. [S. l.]: Computational Soft Matter: From Synthetic Polymers to Proteins, 2004. v. 23, n. 1, p. 1–28.
- ALLEN, M. P.; TILDESLEY, D. J. **Computer Simulation of Liquids**. [S. l.]: Oxford University Press, 2017.
- ALLEN-PERKINS, A.; DE ASSIS, T. A.; PASTOR, J. M.; ANDRADE, R. F. S. Relaxation time of the global order parameter on multiplex networks: The role of interlayer coupling in kuramoto oscillators. **Physical Review E**, v. 96, n. 4, p. 042312, 2017.
- ALMEIDA, M. P. Generalized entropies from first principles. **Physica A: Statistical Mechanics and its Applications**, v. 300, n. 3-4, p. 424–432, 2001.
- ALTMANN, E. G. Spatial interactions in urban scaling laws. **PLoS ONE**, v. 15, n. 12, p. e0243390, 2020.
- ANASTASIADOU, S. D. Exploring AI ethics through factorial correspondence analysis. In: **Modern International Developments in Data Analysis and Multivariate Statistical Analysis**. [S. l.]: IGI Global Scientific Publishing, 2026. p. 495–528.
- ANGELINI, O.; GABRIELLI, A.; TACCHELLA, A.; ZACCARIA, A.; PIETRONERO, L.; DI MATTEO, T. Forecasting the countries' gross domestic product growth: The case of technological fitness. **Chaos, Solitons & Fractals**, v. 184, p. 115006, 2024.
- AUFIERO, S.; DE MARZO, G.; SBARDELLA, A.; ZACCARIA, A. Mapping job fitness and skill coherence into wages: an economic complexity analysis. **Scientific Reports**, v. 14, n. 1, p. 11752, 2024.
- BAHRAMI, F.; SHAHMORADI, B.; NOORI, J.; TURKINA, E.; BAHRAMI, H. Economic complexity and the dynamics of regional competitiveness: a systematic review. **Competitiveness Review: An International Business Journal**, v. 33, n. 4, p. 711–744, 2023.
- BAJRACHARYA, P.; SULTANA, S. Rank-size distribution of cities and municipalities in bangladesh. **Sustainability**, v. 12, n. 11, p. 4643, 2020.
- BAK, P. **How Nature Works: The Science of Self-Organized Criticality**. [S. l.]: Springer, 1996.

- BANCO CENTRAL DO BRASIL. **Correção de valor por índices de preços**. 2025. <https://www3.bcb.gov.br/CALCIDADAOPublico/exibirFormCorrecaoValores.do?method=exibirFormCorrecaoValores>. Acesso em: 27-06-2025.
- BARABÁSI, A.-L. **Network Science**. [S. l.]: Cambridge University Press, 2016.
- BATCHELOR, G. K. **An Introduction to Fluid Dynamics**. [S. l.]: Cambridge University Press, 2000.
- BATTINENI, G.; CHINTALAPUDI, N.; DHULIPALLA, V. R.; AMENTA, F. From diagnostics to education: Multi-domain evaluation of llm chatbots in neurology. **Journal of Taibah University Medical Sciences**, v. 21, n. 1, p. 15–24, 2026.
- BENTZUR, A.; ALON, S.; S-OPHIR, G. Behavioral neuroscience in the era of genomics: tools and lessons for analyzing high-dimensional datasets. **International Journal of Molecular Sciences**, v. 23, n. 7, p. 3811, 2022.
- BERTHOLD, M. R.; BORGELT, C.; HÖPPNER, F.; KLAWONN, F.; SILIPO, R. **Guide to Intelligent Data Science**. [S. l.]: Springer, 2020.
- BILLOTTA, K. G. S. Trabalho de Conclusão de Curso de Graduação, **Auto-organização de partículas coloidais do tipo patchy**. 2018.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. [S. l.]: Springer, 2006.
- BOLEY, D.; GINI, M.; GROSS, R.; HAN, E.-H. S.; HASTINGS, K.; KARYPIS, G.; KUMAR, V.; MOBASHER, B.; MOORE, J. Partitioning-based clustering for web document categorization. **Decision Support Systems**, v. 27, n. 3, p. 329–341, 1999.
- BRADY, H. E. The challenge of big data and data science. **Annual Review of Political Science**, v. 22, n. 1, p. 297–323, 2019.
- BRANDT, S. **Data Analysis: Statistical and Computational Methods for Scientists and Engineers**. [S. l.]: Springer, 2014.
- BRODIE, M. L. A framework for understanding data science. **arXiv preprint arXiv:2403.00776**, 2024.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. **SIAM Journal on Scientific Computing**, v. 16, n. 5, p. 1190–1208, 1995.
- CALDARELLI, G.; CRISTELLI, M.; GABRIELLI, A.; PIETRONERO, L.; SCALA, A.; TACCHHELLA, A. A network analysis of countries' export flows: firm grounds for the building blocks of the economy. **PLoS ONE**, v. 7, n. 10, p. e47278, 2012.
- CALVET, D.; WONG, J. Y.; GIASSON, S. Rheological monitoring of polyacrylamide gelation: Importance of cross-link density and temperature. **Macromolecules**, v. 37, n. 20, p. 7762–7771, 2004.
- CAMARGO, R. F. **Cálculo Fracionário e Aplicações**. Tese (Doutorado) – Universidade Estadual de Campinas, 2009.

- CAO, L. Data science: A comprehensive overview. **ACM Computing Surveys**, v. 50, n. 3, p. 1–42, 2017.
- CARVALHO, G. M.; ARAÚJO, N. A. M.; PATRÍCIO, P. Wound opening in a thin incompressible viscoelastic tissue. **Physical Review E**, v. 104, n. 1, p. 015001, 2021.
- CHAKRABARTI, A. S.; BAKAR, K. S.; CHAKRABORTI, A. **Data Science for Complex Systems**. [S. l.]: Cambridge University Press, 2023.
- CHAPMAN, J. A.; WILSON, B. N.; GULLIVER, J. S. Drag force parameters of rigid and flexible vegetal elements. **Water Resources Research**, v. 51, n. 5, p. 3292–3302, 2015.
- CHEN, D.-G.; CHEN, J.; LU, X.; YI, G. Y.; YU, H. **Advanced Statistical Methods in Data Science**. [S. l.]: Springer, 2016.
- CHEN, L.; TONER, J.; LEE, C. F. Critical phenomenon of the order–disorder transition in incompressible active fluids. **New Journal of Physics**, v. 17, p. 042002, 2015.
- CHEN, S.; PIRHADI, E.; YONG, X. Viscoelastic necking dynamics between attractive microgels. **Journal of Colloid and Interface Science**, v. 618, p. 283–289, 2022.
- CHEN, Y. An allometric scaling relation based on logistic growth of cities. **Chaos, Solitons & Fractals**, v. 65, p. 65–77, 2014.
- CHIN, A. W.; TRUONG, S. S.; SPIVEY, N. D. X-56a structural dynamics ground testing overview and lessons learned. In: American Institute of Aeronautics and Astronautics. **AIAA Scitech 2020 Forum**. [S. l.], 2020.
- COFFEY, W. T.; KALMYKOV, Y. P. **The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering**. [S. l.]: World Scientific, 2012.
- CONWAY, D. Data science in the u.s. intelligence community. **IQT Quarterly**, v. 2, n. 4, p. 24–27, 2011.
- CONWAY, D.; WHITE, J. M. **Machine Learning for Email: Spam Filtering and Priority Inbox**. [S. l.]: O’Reilly Media, 2011.
- CORDEIRO, D.; LOPEZOSA, C.; GUALLAR, J. A methodological framework for ai-driven textual data analysis in digital media. **Future Internet**, v. 17, n. 2, p. 59, 2025.
- COSTA, E. S. Trabalho de Conclusão de Curso de Graduação, **Aplicação de dinâmica molecular em sistemas com interação de potenciais competitivos**. 2019.
- CRISTELLI, M.; GABRIELLI, A.; TACHELLA, A.; CALDARELLI, G.; PIETRONERO, L. Measuring the intangibles: a metrics for the economic complexity of countries and products. **PLoS ONE**, v. 8, n. 8, p. e70726, 2013.
- D’ACCI, L. S. The allometry of housing prices in urban scaling laws and an equalised dwellers’ utility across settlement sizes. **Environment and Planning B: Urban Analytics and City Science**, v. 0, n. 0, p. 23998083251342626, 2025.
- DAHIA, S. S.; SZABO, C. Detecting emergent behavior in complex systems: A machine learning approach. In: **Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation**. [S. l.: s. n.], 2024. p. 81–87.

- DASH, C. S. K.; BEHERA, A. K.; DEHURI, S.; GHOSH, A. An outliers detection and elimination framework in classification task of data mining. **Decision Analytics Journal**, v. 6, p. 100164, 2023.
- DE ARAÚJO, J. L. B. **Auto-organização de partículas patchy carregadas**. Tese (Doutorado) – Universidade Federal do Ceará, 2018.
- DE ARAÚJO, J. L. B.; DE SOUSA, J. S.; FERREIRA, W. P.; OLIVEIRA, C. L. N. Viscoelastic multiscaling in immersed networks. **Physical Review Research**, v. 2, n. 3, p. 033222, 2020.
- DE ARAÚJO, J. L. B.; MUNARIN, F. F.; FARIAS, G. A.; PEETERS, F. M.; FERREIRA, W. P. Structure and reentrant percolation in an inverse patchy colloidal system. **Physical Review E**, v. 95, n. 6, p. 062606, 2017.
- DE ARAÚJO, N. A. M.; DIAS, C. S.; DA GAMA, M. M. T. Nonequilibrium self-organization of colloidal particles on substrates: adsorption, relaxation, and annealing. **Journal of Physics: Condensed Matter**, v. 29, n. 1, p. 014001, 2017.
- DE SOUSA, F. B.; BABU, P. K. V.; RADMACHER, M.; OLIVEIRA, C. L. N.; DE SOUSA, J. S. Multiple power-law viscoelastic relaxation in time and frequency domains with atomic force microscopy. **Journal of Physics D: Applied Physics**, v. 54, n. 3, p. 335401, 2021.
- DE SOUSA, J. S.; FREIRE, R. S.; SOUSA, F. D.; RADMACHER, M.; SILVA, A. F. B.; RAMOS, M. V.; MONTEIRO-MOREIRA, A. C. O.; MESQUITA, F. P.; MORAES, M. E. A.; MONTENEGRO, R. C.; OLIVEIRA, C. L. N. Double power-law viscoelastic relaxation of living cells encodes motility trends. **Scientific Reports**, v. 10, n. 1, p. 4749, 2020.
- DE SOUSA, J. S.; SANTOS, J. A. C.; BARROS, E. B.; ALENCAR, L. M. R.; CRUZ, W. T.; RAMOS, M. V.; MENDES FILHO, J. Analytical model of atomic-force-microscopy force curves in viscoelastic materials exhibiting power law relaxation. **Journal of Applied Physics**, v. 121, n. 3, p. 034901, 2017.
- DELHOMMELLE, J.; MILLIE, P. Inadequacy of the lorentz-berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation. **Molecular Physics**, v. 99, n. 8, p. 619–625, 2001.
- DHAR, V. Data science and prediction. **Communications of the ACM**, v. 56, n. 12, p. 64–73, 2013.
- DIAS, C. S.; BRAGA, C.; ARAÚJO, N. A. M.; GAMA, M. M. T. Relaxation dynamics of functionalized colloids on attractive substrates. **Soft Matter**, v. 12, n. 5, p. 1550–1557, 2016.
- DING, Y.; LUSHI, E.; LI, Q. **Investigation of Quasi-Newton Methods for Unconstrained Optimization**. [S. l.], 2004.
- DOI, M. **Soft Matter Physics**. [S. l.]: Oxford University Press, 2013.
- DURANTON, G.; PUGA, D. Micro-foundations of urban agglomeration economies. In: **Handbook of Regional and Urban Economics**. [S. l.]: Elsevier, 2004. v. 4, p. 2063–2117.
- EFREMOV, Y. M.; WANG, W.-H.; HARDY, S. D.; GEAHLEN, R. L.; RAMAN, A. Measuring nanoscale viscoelastic parameters of cells directly from afm force-displacement curves. **Scientific Reports**, v. 7, n. 1, p. 1541, 2017.

EFRON, B.; HASTIE, T. **Computer Age Statistical Inference: Algorithms, Evidence and Data Science**. [S. l.]: Cambridge University Press, 2016.

EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**, v. 95, n. 25, p. 14863–14868, 1998.

ELABORADO PELO AUTOR. **Parâmetros de simulação do modelo**. 2023. As unidades adimensionais de comprimento, massa e energia são dadas, respectivamente, por $\rho = 1$, $m = 1$, $\varepsilon = 1$, e o *time-step* $dt = 0.001t_0$, onde a unidade de tempo tem a seguinte relação $t_0 = \sqrt{\frac{m\sigma^2}{\varepsilon}}$. Além disso, outros parâmetros de comprimento são especificados como $\rho_s = 10\rho$, $\ell = 1\rho$, e $L = 15\rho$, enquanto alguns expoentes são atribuídos como $\nu = 0.5$ e $\varpi = 400$. O número de partículas na rede é $N = 2748$.

ENOTAS. **O que é NCM e como usar em suas notas fiscais**. 2024. <https://enotas.com.br/blog/ncm-nomenclatura-comum-do-mercosul/>. Acesso em 21-04-2024.

ERHAN, D.; BENGIO, Y.; COURVILLE, A.; MANZAGOL, P.-A.; VINCENT, P.; BENGIO, S. Why does unsupervised pre-training help deep learning? **Journal of Machine Learning Research**, v. 11, p. 625–660, 2010.

ESTRADA, E. What is a complex system, after all? **Foundations of Science**, v. 29, p. 1143–1170, 2024.

FABIANI, G. **Machine learning-based Modelling and Numerical Analysis of the Emergent Dynamics of Complex Systems**. Tese (Doutorado) – University of Naples Federico II, 2024.

FABRY, B.; MAKSYM, G. N.; BUTLER, J. P.; GLOGAUER, M.; NAVAJAS, D.; FREDBERG, J. J. Scaling the microrheology of living cells. **Physical Review Letters**, v. 87, n. 14, p. 148102, 2001.

FABRY, B.; MAKSYM, G. N.; BUTLER, J. P.; GLOGAUER, M.; NAVAJAS, D.; TABACK, N. A.; MILLET, E. J.; FREDBERG, J. J. Time scale and other invariants of integrative mechanical behavior in living cells. **Physical Review E**, v. 68, n. 4, p. 041914, 2003.

FAIRFIELD, J.; SHTEIN, H. Big data, big problems: Emerging issues in the ethics of data science and journalism. **Journal of Mass Media Ethics**, v. 29, n. 1, p. 38–51, 2014.

FAVIER, J.; DAUPTAIN, A.; BASSO, D.; BOTTARO, A. Passive separation control using a self-adaptive hairy coating. **Journal of Fluid Mechanics**, v. 627, p. 451–483, 2009.

FERREIRA, A. E. O. **Regime de Arrasto Sublinear em Escalas Mesoscópicas de Materiais Viscoelásticos**. Dissertação (Mestrado) – Universidade Federal do Ceará, 2022.

FERREIRA, A. E. O.; DE ARAÚJO, J. L. B.; FERREIRA, W. P.; DE SOUSA, J. S.; OLIVEIRA, C. L. N. Sublinear drag regime at mesoscopic scales in viscoelastic materials. **PLoS ONE**, v. 19, n. 3, p. e0299296, 2024.

FISH, J.; BELSKY, V. Multigrid method for periodic heterogeneous media. part 1: Convergence studies for one-dimensional case. **Computer Methods in Applied Mechanics and Engineering**, v. 126, n. 1-2, p. 1–16, 1995.

FISH, J.; BELSKY, V. Multigrid method for periodic heterogeneous media. part 2: Multiscale modeling and quality control in multidimensional case. **Computer Methods in Applied Mechanics and Engineering**, v. 126, n. 1-2, p. 17–38, 1995.

FISH, J.; WAGNER, G. J.; KETEN, S. Mesoscopic and multiscale modelling in materials. **Nature Materials**, v. 20, n. 6, p. 774–786, 2021.

FISHER, D.; MEYER, M. **Making Data Visual: A Practical Guide to Using Visualization for Insight**. [S. l.]: O’Reilly Media, 2017.

FRENKEL, D.; SMIT, B. **Understanding Molecular Simulation: From Algorithms to Applications**. [S. l.]: Academic Press, 2002.

FUNDAÇÃO GETULIO VARGAS (FGV). **General Price Index (IGP)**. 2025. <https://portalibre.fgv.br>. Acesso em: 23-06-2025.

GABAIX, X. Zipf’s law for cities: An explanation. **Quarterly Journal of Economics**, Oxford University Press, v. 114, n. 3, p. 739–767, 1999.

GALLEGATI, G.; GALLEGATI, M.; LANDINI, S. **Complexity in Economics**. [S. l.]: Cambridge University Press, 2025.

GALLOTTI, R.; SACCO, P.; DE DOMENICO, M. Complex urban systems: Challenges and integrated solutions for the sustainability and resilience of cities. **Complexity**, v. 2021, n. 1, p. 1782354, 2021.

GAO, M.; HU, X.; YIN, X.; RUAN, J.; PU, X.; WAN, X. Llm-based nlg evaluation: Current status and challenges. **Computational Linguistics**, v. 51, n. 2, p. 661–687, 2025.

GARCIA, P. D.; GARCIA, R. Determination of the elastic moduli of a single cell cultured on a rigid support by force microscopy. **Biophysical Journal**, v. 114, n. 12, p. 2923–2932, 2018.

GELMAN, A. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). **Bayesian Analysis**, v. 1, n. 3, p. 515–534, 2006.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. [S. l.]: O’Reilly Media, 2019.

GITHUB REPOSITORY. **Scaling and Complexity in Urban Trade**. 2025. <https://github.com/edn007/Scaling-and-Complexity-in-Urban-Trade>. Acesso em: 07-06-2025.

GOLDENFELD, N.; KADANOFF, L. P. Simple lessons from complexity. **Science**, v. 284, n. 5411, p. 87–89, 1999.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S. l.]: MIT Press, 2016.

GOSSELIN, F.; EL LANGRE, E.; DE LANGRE, E.; ALMEIDA, B. A. M. Drag reduction of flexible plates by reconfiguration. **Journal of Fluid Mechanics**, v. 650, p. 319–341, 2010.

GRIFFIN, C. **Numerical Optimization: Penn State Math 555 Lecture Notes**. 2012.

GRIPPA, V. M.; KUZMICHEV, S. **Learning MySQL**. [S. l.]: O’Reilly Media, 2021.

GRUS, J. **Data Science from Scratch**. [S. l.]: O’Reilly Media, 2015.

GUO, Q.; HE, X.; WANG, Z.; WANG, J. Effects of wing flexibility on aerodynamic performance of an aircraft model. **Chinese Journal of Aeronautics**, v. 34, n. 9, p. 133–142, 2021.

HARBISON, L.; SIMMONS, K. Fundamentals of descriptive statistics. **Journal of Dental Hygiene**, v. 98, n. 5, p. 51–54, 2024.

HASAN, B. M. S.; ABDULAZEEZ, A. M. A review of principal component analysis algorithm for dimensionality reduction. **Journal of Soft Computing and Data Mining**, v. 2, n. 1, p. 20–30, 2021.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: Data mining, inference, and prediction**. [S. l.]: Springer, 2009.

HAUSMANN, R.; HIDALGO, C. A.; BUSTOS, S.; COSCIA, M.; SIMOES, A.; YILDIRIM, M. A. **The Atlas of Economic Complexity: Mapping Paths to Prosperity**. [S. l.]: MIT Press, 2014.

HAUSMANN, R.; HWANG, J.; RODRIK, D. What you export matters. **Journal of Economic Growth**, v. 12, n. 1, p. 1–25, 2007.

HE, J.; LI, H. A review of trade geography research: Theoretical foundations, spatial patterns, and frontier dynamics. **Geographical Research Bulletin**, v. 4, p. 96–119, 2025.

HIDALGO, C. A.; HAUSMANN, R. The building blocks of economic complexity. **Proceedings of the National Academy of Sciences**, v. 106, n. 26, p. 10570–10575, 2009.

HINLOOPEN, J.; VAN MARREWIJK, C. On the empirical distribution of the balassa index. **Weltwirtschaftliches Archiv**, v. 137, n. 1, p. 1–35, 2001.

HINLOOPEN, J.; VAN MARREWIJK, C. Power laws and comparative advantage. **Applied Economics**, v. 44, n. 12, p. 1483–1507, 2011.

HRAMOV, A. E.; GHOSH, D.; PISARCHIK, A. N.; PAVLOV, A.; STOYANOV, D.; ZAIKIN, A.; KURKIN, S.; SANJUÁN, M. Artificial intelligence and complex networks meet natural sciences. **European Physical Journal Special Topics**, v. 234, n. 15, p. 3547–3562, 2025.

IGI GLOBAL; INFORMATION RESOURCES MANAGEMENT ASSOCIATION (Ed.). **Machine Learning: Concepts, Methodologies, Tools and Applications**. [S. l.]: IGI Global, 2012.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Cidades e Estados: Ceará**. 2022. <https://www.ibge.gov.br/cidades-e-estados/ce.html>. Acesso em: 21-04-2024.

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ (IPECE). **Consulta aos mapas e legislação**. 2019. <https://www.ipece.ce.gov.br/consulta-aos-mapas-e-legislacao/>. Acesso em: 21-04-2024.

JAISHANKAR, A.; MCKINLEY, G. H. Power-law rheology in the bulk and at the interface: Quasi-properties and fractional constitutive equations. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 469, n. 2149, p. 20120284, 2013.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning**. [S. l.]: Springer, 2013.

- JEFFERSON, M. The law of the primate city. **Geographical Review**, v. 29, n. 2, p. 226–232, 1939.
- JENSEN, K. S. Drag and reconfiguration of freshwater macrophytes. **Freshwater Biology**, v. 48, n. 2, p. 271–283, 2003.
- KANNAN, K. S.; MANOJ, K. Outlier detection in multivariate data. **Applied Mathematical Sciences**, v. 9, n. 47, p. 2317–2324, 2015.
- KARBEVSKA, L.; HIDALGO, C. A. Mapping global value chains at the product level. **EPJ Data Science**, v. 14, n. 1, p. 21, 2025.
- KELLEHER, J. D.; TIERNEY, B. **Data Science**. [S. l.]: MIT Press, 2018.
- KENIG, N.; VIVES, A. M.; ECHEVERRIA, J. M. Ethics for artificial intelligence in medicine. In: **Business Ethics in the Healthcare Industry**. [S. l.]: Springer Nature Switzerland, 2026. p. 661–675.
- KETZ, R. J.; PRUD'HOMME, R. K.; GRAESSLEY, W. W. Rheology of concentrated microgel solutions. **Rheologica Acta**, v. 27, n. 5, p. 531–539, 1988.
- KHAIR, A. S.; BRADY, J. F. “microviscoelasticity” of colloidal dispersions. **Journal of Rheology**, v. 49, n. 6, p. 1449–1481, 2005.
- KHEIFETS, A. S. **High-sensitivity tracking of optically trapped particles in gases and liquids: observation of Brownian motion in velocity space**. Dissertação (Mestrado) – The University of Texas at Austin, 2014.
- KOCH, P.; STOJKOSKI, V.; HIDALGO, C. A. The role of immigrants, emigrants, and locals in the historical formation of european knowledge agglomerations. **Regional Studies**, v. 58, n. 9, p. 1659–1673, 2023.
- KOCH, P.; STOJKOSKI, V.; HIDALGO, C. A. Augmenting the availability of historical GDP per capita estimates through machine learning. **Proceedings of the National Academy of Sciences**, v. 121, n. 39, p. e2402060121, 2024.
- KREMER, K.; ROBBINS, M. O.; GRETT, G. S. Phase diagram of yukawa systems: Model for charge-stabilized colloids. **Physical Review Letters**, v. 57, n. 21, p. 2694, 1986.
- KUBBEN, P.; DUMONTIER, M.; DEKKER, A. **Fundamentals of Clinical Data Science**. [S. l.]: Springer, 2019.
- KUSIAK, A. Generative artificial intelligence in smart manufacturing. **Journal of Intelligent Manufacturing**, v. 36, n. 1, p. 1–3, 2024.
- LAKES, R. S. **Viscoelastic Solids**. [S. l.]: CRC Press, 2017.
- LANGEVIN, P. Sur la théorie du mouvement brownien. **Comptes Rendus de l'Académie des Sciences**, v. 146, n. 530-533, p. 530, 1908.
- LARSON, R. G. **The Structure and Rheology of Complex Fluids**. [S. l.]: Oxford University Press, 1999.

- LEMONS, D. S.; GYTHIEL, A. Paul langevin's 1908 paper "on the theory of brownian motion" [«sur la théorie du mouvement brownien,» c. r. acad. sci. (paris) 146, 530–533 (1908)]. **American Journal of Physics**, v. 65, n. 11, p. 1079–1081, 1997.
- LEPOIRE, D. J.; GRININ, L.; KOROTAYEV, A. Complexity in big history. an introductory exploration. **Journal of Big History**, v. 8, n. 3, p. 98–139, 2025.
- LESLIE, D. Understanding artificial intelligence ethics and safety. **arXiv preprint arXiv:1906.05684**, 2019.
- LI, X.; WANG, X.; ZHANG, J.; WU, L. Allometric scaling, size distribution and pattern formation of natural cities. **Palgrave Communications**, v. 1, n. 1, p. 15017, 2015.
- LIU, B.; GAO, J. Understanding the non-gaussian distribution of revealed comparative advantage index and its alternatives. **International Economics**, v. 158, p. 1–11, 2019.
- LIU, B.; GAO, J. Normality in the distribution of revealed comparative advantage index for international trade and economic complexity. **Applied Sciences**, v. 12, n. 3, p. 1125, 2022.
- LIU, D. C.; NOCEDAL, J. On the limited memory bfgs method for large scale optimization. **Mathematical Programming**, v. 45, n. 1, p. 503–528, 1989.
- LIU, J.; PAN, Y.; LI, M.; CHEN, Z.; TANG, L.; LU, C.; WANG, J. Applications of deep learning to MRI images: a survey. **Big Data Mining and Analytics**, v. 1, n. 1, p. 1–18, 2018.
- LIU, L.; OUYANG, W.; WANG, X.; FIEGUTH, P.; CHEN, J.; LIU, X.; PIETIKÄINEN, M. Deep learning for generic object detection: a survey. **International Journal of Computer Vision**, v. 128, n. 2, p. 261–318, 2020.
- LOGCOMEX. **Consulta de NCM Logcomex**. 2024. <https://ncm.logcomex.com/>. Acesso em 21-04-2024.
- LONG, A. W.; FERGUSON, A. L. Rational design of patchy colloids via landscape engineering. **Molecular Systems Design & Engineering**, v. 3, n. 1, p. 49–65, 2018.
- LOUKIDES, M. **The Evolution of Data Products**. [S. l.]: O'Reilly Media, 2011.
- LOUKIDES, M. **What is Data Science?** [S. l.]: MIT Press, 2018.
- LUHAR, M.; NEPF, H. M. From the blade scale to the reach scale: A characterization of aquatic vegetative drag. **Advances in Water Resources**, v. 51, p. 305–316, 2013.
- MAHARANA, K.; MONDAL, S.; NEMADE, B. A review: Data pre-processing and data augmentation techniques. **Global Transitions Proceedings**, v. 3, n. 1, p. 91–99, 2022.
- MAHNOOR; SHAFI, I.; CHAUDHRY, M.; MONTERO, E. C.; ALVARADO, E. S.; DIEZ, I. T.; SAMAD, M. D. A.; ASHRAF, I. A review of approaches for rapid data clustering: Challenges, opportunities, and future directions. **IEEE Access**, v. 12, p. 138086–138120, 2024.
- MAIOCCHI, R. The case of brownian motion. **The British Journal for the History of Science**, v. 23, n. 3, p. 257–283, 1990.
- MANDELBROT, B. B. An informational theory of the statistical structure of language. **Communication Theory**, v. 84, n. 21, p. 486–502, 1953.

- MANDELBROT, B. B. **The Fractal Geometry of Nature**. [S. l.]: W. H. Freeman, 1982.
- MANN, S.; BECIU, S.; ARGHIROIU, G. A. Colouring the balassa index: A hermeneutic approach towards romanian meat imports. **Ciência Rural**, v. 53, p. e20210811, 2023.
- MARKOVITZ, H. Boltzmann and the beginning of linear viscoelasticity. **Transactions of the Society of Rheology**, v. 21, n. 2, p. 381–398, 1977.
- MAXWELL, J. C. On the dynamical theory of gases. **Philosophical Transactions of the Royal Society of London**, v. 157, p. 49–88, 1867.
- MAYERNIK, M. S. Data science as an interdisciplinary: historical parallels from information science. **Data Science Journal**, v. 22, n. 1, p. 1–18, 2023.
- MAZZOCCHI, F. An investigation into the notion of complex systems. **Foundations of Science**, p. 1–20, 2025.
- MEALY, P.; FARMER, J. D.; TEYTELBOYM, A. Interpreting economic complexity. **Science Advances**, v. 5, n. 1, p. eaau1705, 2019.
- MESSER, U. Co-creating art with generative artificial intelligence: Implications for artworks and artists. **Computers in Human Behavior: Artificial Humans**, v. 2, n. 1, p. 100056, 2024.
- MEZMIR, E. A. Qualitative data analysis: an overview of data reduction, data display and interpretation. **Research on Humanities and Social Sciences**, v. 10, n. 21, p. 15–27, 2020.
- MIEHE, C.; BAYREUTHER, C. G. On multiscale fe analyses of heterogeneous structures: From homogenization to multigrid solvers. **International Journal for Numerical Methods in Engineering**, v. 71, n. 10, p. 1135–1180, 2007.
- MILINKOVIC, B.; BARNETT, L.; CARTER, O.; SETH, A. K.; ANDRILLON, T. Capturing the emergent dynamical structure in biophysical neural models. **PLOS Computational Biology**, v. 21, n. 5, p. e1012572, 2025.
- MILKUS, R.; ZACCONE, A. Atomic-scale origin of dynamic viscoelastic response and creep in disordered solids. **Physical Review E**, v. 95, n. 2, p. 023001, 2017.
- MINISTÉRIO DA FAZENDA. **Portal da Nota Fiscal Eletrônica**. 2003. <https://www.nfe.fazenda.gov.br/portal/principal.aspx>. Acesso em 21-03-2025.
- MITCHELL, M. **Complexity: A Guided Tour**. [S. l.]: Oxford University Press, 2009.
- MITCHELL, T. M. **Machine Learning**. [S. l.]: McGraw-Hill, 1997.
- MODONESI, C.; SERAFINI, S.; GIULIANI, A. Urban systems between the environment, human health and society: An overview. **Systems**, v. 13, n. 6, p. 487, 2025.
- MOEENDARBARY, E.; HARRIS, A. R. Cell mechanics: principles, practices, and prospects. **Wiley Interdisciplinary Reviews: Systems Biology and Medicine**, v. 6, n. 5, p. 371–388, 2014.
- MOREIRA, A. A.; OLIVEIRA, C. L. N.; HANSEN, A.; ARAÚJO, N. A. M.; HERRMANN, H. J.; ANDRADE, J. S. Fracturing highly disordered materials. **Physical Review Letters**, v. 109, n. 25, p. 255701, 2012.

- MORRISON, G.; BULDYREV, S. V.; IMBRUNO, M.; ARRIETA, O. A. D.; RUNGI, A.; RICCABONI, M.; PAMMOLLI, F. On economic complexity and the fitness of nations. **Scientific Reports**, v. 7, n. 1, p. 15332, 2017.
- MUÑOZ, M. A. Colloquium: Criticality and dynamical scaling in living systems. **Reviews of Modern Physics**, v. 90, n. 3, p. 031001, 2018.
- NAUR, P. **Concise Survey of Computer Methods**. [S. l.]: Petrocelli/Charter, 1974.
- NAWAZ, S.; SALEEM, M.; KUSMARTSEV, F. V.; ANJUM, D. H. Major role of multiscale entropy evolution in complex systems and data science. **Entropy**, v. 26, n. 4, p. 330, 2024.
- NEWMAN, M. **Networks**. [S. l.]: Oxford University Press, 2018.
- NEWMAN, M. E. J. Power laws, pareto distributions and Zipf's law. **Contemporary Physics**, v. 46, n. 5, p. 323–351, 2005.
- NILSSON, N. J. **Principles of Artificial Intelligence**. [S. l.]: Morgan Kaufmann Publishers Inc., 2014.
- NUNES, D. B.; NETO, J. P. B.; FREITAS, S. M. Modelo de regressão linear múltipla para avaliação do valor de mercado de apartamentos residenciais em Fortaleza, CE. **Ambiente Construído**, v. 19, n. 1, p. 89–104, 2019.
- OHTA, H.; HAMAGUCHI, S. Molecular dynamics evaluation of self-diffusion in yukawa systems. **Physics of Plasmas**, v. 7, n. 11, p. 4506–4514, 2000.
- OLIVEIRA, C. L. N.; BATES, J. H. T.; SUKI, B. A network model of correlated growth of tissue stiffening in pulmonary fibrosis. **New Journal of Physics**, v. 16, n. 6, p. 065022, 2014.
- OLIVEIRA, P. H. G. **Uma revisão bibliográfica sobre algoritmos de minimização sem restrições de grande porte**. 2009. https://abepro.org.br/biblioteca/enegep2009_TN_STO_107_711_14061.pdf. Acesso em: 21-06-2023.
- O'NEIL, C. **Weapons of Math Destruction**. [S. l.]: Crown Publishing, 2016.
- OPERTI, F. G.; PUGLIESE, E.; ANDRADE, J. S.; PIETRONERO, L.; GABRIELLI, A. Dynamics in the fitness-income plane: Brazilian states vs world countries. **PLoS ONE**, v. 13, n. 6, p. e0197616, 2018.
- PAULA, G. S.; KUNIHIRO, R. K.; DOS SANTOS, L. M. A lei geral de proteção de dados (LGPD) e seus efeitos na internet: implicações legais e desafios para a privacidade digital. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 11, n. 6, p. 1864–1872, 2025.
- PEÑA, J. J.; MENÉNDEZ, A.; GARCÍA-RAVELO, J.; MORALES, J. Mie-type potential from a class of multiparameter exponential-type potential: Bound state solutions in d dimensions. **Journal of Physics: Conference Series**, v. 633, n. 1, p. 012025, 2015.
- PHAM, M. C.; CAO, Y.; KLAMMA, R.; JARKE, M. A clustering approach for collaborative filtering recommendation using social network analysis. **Journal of Universal Computer Science**, v. 17, n. 4, p. 583–604, 2011.

PIETRONERO, L.; CRISTELLI, M.; GABRIELLI, A.; MAZZILLI, D.; PUGLIESE, E.; TACHELLA, A.; ZACCARIA, A. Economic complexity: “buttarla in caciarà” vs a constructive approach. **arXiv preprint arXiv:1709.05272**, 2017.

PINHEIRO, C. Relatedness and economic complexity as tools for industrial policy: Insights and limitations. **Structural Change and Economic Dynamics**, v. 72, p. 1–10, 2025.

PORTAL ÚNICO SISCOMEX. **Exportar NCM**. 2024. <https://portalunico.siscomex.gov.br/classif/#/nomenclatura/tabela?perfil=publico>. Acesso em 21-04-2024.

POWELL, M. J. D. How bad are the bfgs and dfg methods when the objective function is quadratic?. **Mathematical Programming**, v. 34, n. 1, p. 34–47, 1986.

PRAPROTNIK, M.; SITE, L. D.; KREMER, K. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. **Annual Review of Physical Chemistry**, v. 59, n. 1, p. 545–571, 2008.

PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51–59, 2013.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. [S. l.]: O’Reilly Media, 2013.

PUGLIESE, E.; ZACCARIA, A.; PIETRONERO, L. On the convergence of the fitness-complexity algorithm. **European Physical Journal Special Topics**, v. 225, n. 10, p. 1893–1911, 2016.

QIAO, W.; YIN, S.; HUANG, X. Assessing the imbalances in growth between urban land and urban population and the influencing factors: an allometric growth perspective. **Land**, v. 13, n. 10, p. 1657, 2024.

QU, Z.; GARFINKEL, A.; WEISS, J. N.; NIVALA, M. Multi-scale modeling in biology: How to bridge the gaps between scales? **Progress in Biophysics and Molecular Biology**, v. 107, n. 1, p. 21–31, 2011.

RAHMAN, A. Correlations in the motion of atoms in liquid argon. **Physical Review**, v. 136, n. 2A, p. A405, 1964.

RAMO, N. L.; PUTTLITZ, C. M.; TROYER, K. L. The development and validation of a numerical integration method for non-linear viscoelastic modeling. **PLoS ONE**, v. 13, n. 1, p. e0190137, 2018.

RAPAPORT, D. C. **The Art of Molecular Dynamics Simulation**. [S. l.]: Cambridge University Press, 2004.

RECEITA FEDERAL. **Nomenclatura Comum do Mercosul**. 2021. <https://www.gov.br/receitafederal/pt-br/assuntos/aduana-e-comercio-exterior/classificacao-fiscal-de-mercadorias/download-ncm-nomenclatura-comum-do-mercosul>. Acesso em 21-03-2025.

REHAGE, H.; HOFFMANN, H. Rheological properties of viscoelastic surfactant systems. **Journal of Physical Chemistry**, v. 92, n. 16, p. 4712–4719, 1988.

RODRIGUES, J. F.; FLOREA, L.; DE OLIVEIRA, M. C. F.; DIAMOND, D.; OLIVEIRA, O. N. Big data and machine learning for materials science. **Discover Materials**, v. 1, n. 1, p. 12, 2021.

ROSEN, K. T.; RESNICK, M. The size distribution of cities: An examination of the pareto law and primacy. **Journal of Urban Economics**, v. 8, n. 2, p. 165–186, 1980.

ROTTERS, F.; EISENLOHR, P.; HANTCHERLI, L.; TJAHOJANTO, D. D.; BIELER, T. R.; RAABE, D. Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. **Acta Materialia**, v. 58, n. 4, p. 1152–1211, 2010.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach** (Pearson Series in Artificial Intelligence). [S. l.]: Pearson, 2020.

SAHASRANAMAN, A.; JENSEN, H. J. Amalgamating urban scaling and economic complexity for new metrics of urban fitness. **Social Indicators Research**, v. 176, n. 3, p. 931–945, 2025.

SALVATIER, J.; WIECKI, T. V.; FONNESBECK, C. Probabilistic programming in python using PyMC3. **PeerJ Computer Science**, v. 2, p. e55, 2016.

SAMANIEGO, H.; MOSES, M. E. Cities as organisms: Allometric scaling of urban road networks. **Journal of Transport and Land Use**, v. 1, n. 1, p. 21–39, 2008.

SAMPAIO FILHO, C. I. N.; PIRES, R. S.; CARMONA, H. A.; ANDRADE JR., J. S. Mapping inter-city trade networks to maximum entropy models using electronic invoice data. **Chaos, Solitons & Fractals**, v. 202, p. 117593, 2026.

SCHLEDER, G. R.; FAZZIO, A. Machine learning na física, química, e ciência de materiais: Descoberta e design de materiais. **Revista Brasileira de Ensino de Física**, v. 43, p. e20200407, 2021.

SCHNABEL, T.; VRABEC, J.; HASSE, H. Unlike lennard–jones parameters for vapor–liquid equilibria. **Journal of Molecular Liquids**, v. 135, n. 1-3, p. 170–178, 2007.

SCIKIT-LEARN DEVELOPERS. **scikit-learn User Guide, Release 0.18.2**. [S. l.].

SERVEDIO, V. D. P.; BUTTÀ, P.; MAZZILLI, D.; TACHELLA, A.; PIETRONERO, L. A new and stable estimation method of country economic fitness and product complexity. **Entropy**, v. 20, n. 10, p. 783, 2018.

SEYBOLD, H. J.; LIMA, I. A.; ARAÚJO, A. D. The critical behavior of the clogging process in a porous medium. **Physics of Fluids**, v. 33, n. 9, p. 093609, 2021.

SHAYEGAN, M.; FORDE, N. R. Microrheological characterization of collagen systems: From molecular solutions to fibrillar gels. **PLoS ONE**, v. 8, n. 8, p. e70590, 2013.

SIEBERER, L. M.; BUCHHOLD, M.; MARINO, J.; DIEHL, S. Universality in driven open quantum matter. **Reviews of Modern Physics**, v. 97, p. 025004, 2025.

SILVER, D.; HUANG, A.; MADDISON, C. J.; GUEZ, A.; SIFRE, L.; DRIESSCHE, G. van den; SCHRITTWIESER, J.; ANTONOGLU, I.; PANNEERSHELVA, V.; LANCTOT, M.; DIELEMAN, S.; GREWE, D.; NHAM, J.; KALCHBRENNER, N.; SUTSKEVER, I.; LILLICRAP, T.; LEACH, M.; KAVUKCUOGLU, K.; GRAEPEL, T.; HASSABIS, D. Mastering the game of go with deep neural networks and tree search. **Nature**, v. 529, n. 7587, p. 484–489, 2016.

SINAGA, K. P.; YANG, M. Unsupervised k-means clustering algorithm. **IEEE Access**, v. 8, p. 80716–80727, 2020.

SODRÉ, U. **Transformadas de Laplace – Notas de Aulas**. 2003.

SOLLICH, P. Rheological constitutive equation for a model of soft glassy materials. **Physical Review E**, v. 58, n. 1, p. 738, 1998.

SOLLICH, P.; LEQUEUX, F.; HÉBRAUD, P.; CATES, M. E. Rheology of soft glassy materials. **Physical Review Letters**, v. 78, n. 10, p. 2020, 1997.

SONG, R.; JIANG, G.; WANG, K. Gelation mechanism and rheological properties of polyacrylamide crosslinking with polyethyleneimine and its plugging performance in air-foam displacement. **Journal of Applied Polymer Science**, v. 135, n. 5, p. 45778, 2017.

SOO, K. T. Zipf's law for cities: A cross-country investigation. **Regional Science and Urban Economics**, v. 35, n. 3, p. 239–263, 2005.

SORNETTE, D. **Critical Phenomena in Natural Sciences**. [S. l.]: Springer, 2006.

SOUZA, A. M. C.; ANDRADE, R. F. S.; NOBRE, F. D.; CURADO, E. M. F. Thermodynamic framework for compact q -gaussian distributions. **Physica A: Statistical Mechanics and its Applications**, v. 491, p. 153–166, 2018.

SRINADH, V. Evaluation of apriori, fp-growth and eclat association rule mining algorithms. **International Journal of Health Sciences**, v. 6, n. 2, p. 7475–7485, 2022.

STANLEY, H. E. **Introduction to Phase Transitions and Critical Phenomena**. [S. l.]: Oxford University Press, 1971.

STANLEY, H. E.; AMARAL, L. A. N.; GOPIKRISHNAN, P.; IVANOV, P. C.; KEITT, T. H.; PLEROU, V. Scale invariance and universality: organizing principles in complex systems. **Physica A: Statistical Mechanics and its Applications**, v. 281, n. 1, p. 60–68, 2000.

STEKHOFEN, D. J.; BÜHLMANN, P. Missforest—non-parametric missing value imputation for mixed-type data. **Bioinformatics**, v. 28, n. 1, p. 112–118, 2012.

STOJKOSKI, V.; KOCH, P.; HIDALGO, C. A. Multidimensional economic complexity and inclusive green growth. **Communications Earth & Environment**, v. 4, n. 1, p. 130, 2023.

STRACCAMORE, M.; BRUNO, M.; MONECHI, B.; LORETO, V. Urban economic fitness and complexity from patent data. **Scientific Reports**, v. 13, n. 1, p. 3655, 2023.

SUN, H.; FAN, X.; ZHAN, M.; GUO, J.; ZHANG, J. Exploring the densification mechanism of heterogeneous porous pbx composites by multiscale dem modeling: A perspective of mesoscopic deformation and microscopic multiphase synergistic evolution. **Journal of Materials Science & Technology**, v. 265, p. 64–75, 2026.

TACCHELLA, A.; CRISTELLI, M.; CALDARELLI, G.; GABRIELLI, A.; PIETRONERO, L. A new metrics for countries' fitness and products' complexity. **Scientific Reports**, v. 2, n. 1, p. 723, 2012.

TANG, Y.; KURTHS, J.; LIN, W.; OTT, E.; KOCAREV, L. Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, v. 30, n. 6, p. 063151, 2020.

- TEODORO, G. S. **Cálculo Fracionário e as Funções de Mittag-Leffler**. Dissertação (Mestrado) – Universidade Estadual de Campinas, 2014.
- THEERTHAGIRI, P.; RUBY, A. U.; VIDYA, J. Diagnosis and classification of the diabetes using machine learning algorithms. **SN Computer Science**, v. 4, n. 1, p. 72, 2023.
- TOMÉ, T.; DE OLIVEIRA, M. J. **Stochastic Dynamics and Irreversibility**. [S. l.]: Springer, 2015.
- TSALLIS, C. Possible generalization of boltzmann-gibbs statistics. **Journal of Statistical Physics**, v. 52, n. 1, p. 479–487, 1988.
- TSALLIS, C. Nonadditive entropy and nonextensive statistical mechanics – an overview after 20 years. **Brazilian Journal of Physics**, v. 39, p. 337–356, 2009.
- TU, S.; LI, C.; SHEPHERD, B. E. Between- and within-cluster spearman rank correlations. **Statistics in Medicine**, v. 44, n. 3-4, p. e10326, 2025.
- TUKEY, J. W. **Exploratory Data Analysis**. [S. l.]: Addison-Wesley, 1977.
- ULLMAN, J. D. The battle for data science. **IEEE Data Engineering Bulletin**, v. 43, n. 2, p. 8–14, 2020.
- UNWIN, A. **Graphical Data Analysis with R**. [S. l.]: Chapman and Hall/CRC, 2015.
- VASCO, J. R. G.; MACIEL, G. F. Interação onda-vegetação submersa: Parte ii: Modelo físico e numérico. **Revista Brasileira de Recursos Hídricos**, v. 12, n. 1, p. 163–175, 2007.
- VERLET, L. Computer “experiments” on classical fluids. i. thermodynamical properties of lennard-jones molecules. **Physical Review**, v. 159, n. 1, p. 98, 1967.
- VERSCHOREN, V.; MEIRE, D.; SCHOELYNCK, J.; BUIS, K.; DE BAL, K.; TROCH, P.; MEIRE, P.; TEMMERMAN, S. Resistance and reconfiguration of natural flexible submerged vegetation in hydrodynamic river modelling. **Environmental Fluid Mechanics**, v. 16, n. 1, p. 245–265, 2016.
- VOGEL, S. Drag and flexibility in sessile organisms. **American Zoologist**, v. 24, n. 1, p. 37–44, 1984.
- VOGEL, S. Drag and reconfiguration of broad leaves in high winds. **Journal of Experimental Botany**, v. 40, n. 8, p. 941–948, 1989.
- VOGEL, S. **Life in Moving Fluids: The Physical Biology of Flow**. [S. l.]: Princeton University Press, 1994.
- VOIGT, P.; BUSSCHE, A. von dem. **The EU General Data Protection Regulation (GDPR)**. [S. l.]: Springer, 2024.
- WANG, C.; WEI, S.; QI, J. Matchnav: Llm-based enhanced description and instruction matching in vision-and-language navigation. **Information Fusion**, v. 125, p. 103444, 2026.
- WANG, J.; LI, Y.; WEI, H.; WU, Q. The impact and spatiotemporal heterogeneity of differentiated industrial land supply regarding industrial total factor productivity. **Land**, v. 14, n. 12, p. 2435, 2025.

WANG, X.; LU, D.; WU, Z.; XU, W.; HOU, H.; HU, Y.; MORENO, Y. Predicting the critical behavior of complex dynamic systems via learning the governing mechanisms. **Chaos, Solitons & Fractals**, v. 198, p. 116515, 2025.

WEST, B. J.; BOLOGNA, M.; GRIGOLINI, P. **Physics of Fractal Operators**. [S. l.]: Springer, 2003.

WEST, G. B.; BROWN, J. H. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. **Journal of Experimental Biology**, v. 208, n. 9, p. 1575–1592, 2005.

WEST, G. B.; BROWN, J. H.; ENQUIST, B. J. A general model for the origin of allometric scaling laws in biology. **Science**, v. 276, n. 5309, p. 122–126, 1997.

WICKHAM, H.; GROLEMUND, G. **R for Data Science**. [S. l.]: O'Reilly Media, 2017.

WOLFF, L.; FERNÁNDEZ, P.; KROY, K. Resolving the stiffening-softening paradox in cell mechanics. **PLoS ONE**, v. 7, n. 7, p. e40063, 2012.

WORLD CUSTOMS ORGANIZATION. **Harmonized System**. 2025. <https://www.wcotradetools.org/en/harmonized-system>. Acesso em 15-04-2025.

XI, L.; BAI, X. Marginal turbulent state of viscoelastic fluids: A polymer drag reduction perspective. **Physical Review E**, v. 93, n. 4, p. 043118, 2016.

YOUNG, T.; HAZARIKA, D.; PORIA, S.; CAMBRIA, E. Recent trends in deep learning based natural language processing. **IEEE Computational Intelligence Magazine**, v. 13, n. 3, p. 55–75, 2018.

YUCHT, M. G.; SHEINMAN, M.; BROEDERSZ, C. P. Dynamical behavior of disordered spring networks. **Soft Matter**, v. 9, n. 29, p. 7000–7006, 2013.

ZEKOVIĆ, S.; PERIĆ, A. Regional industrial policy in the western balkans: Neither specialization nor spatialization?. **Eastern European Economics**, v. 62, n. 4, p. 479–504, 2024.

ZHANG, C.; LU, Y. Study on artificial intelligence: The state of the art and future prospects. **Journal of Industrial Information Integration**, v. 23, p. 100224, 2021.

ZHANG, L.; CHEN, W.; ZHANG, Q.; YOU, K.; DIAO, G. Improving economic complexity index: Insights from value added. **Economic Analysis and Policy**, v. 85, p. 1391–1408, 2025.

ZHANG, Y.; LIN, Y.; ZHENG, G.; LIU, Y.; SUKIENNIK, N.; XU, F.; XU, Y.; LU, F.; WANG, Q.; LAI, Y.; TIAN, L.; LI, N.; FANG, D.; WANG, F.; ZHOU, T.; LI, Y.; ZHENG, Y.; WU, Z.; GUO, H. Metacity: Data-driven sustainable development of complex cities. **The Innovation**, v. 6, n. 2, p. 100775, 2025.

ZHANG, Z.; BERTIN, V.; ESSINK, M.; ZHANG, H.; FARES, N.; SHEN, Z.; BICKEL, T.; SALEZ, T.; MAALI, A. Unsteady drag force on an immersed sphere oscillating near a wall. **Journal of Fluid Mechanics**, v. 977, p. A21, 2023.

ZHAO, J.; LIU, G.; GUO, X.; LU, H. A modified stokes law-based solid-solid drag model for the gas-fluidization of binary mixtures with different sizes and densities. **Powder Technology**, v. 425, p. 118533, 2023.

ZHAO, L.; HUANG, Q.; DENG, X.; SANE, S. P. Aerodynamic effects of flexibility in flapping wings. **Journal of the Royal Society Interface**, v. 7, n. 44, p. 485–497, 2010.

ZHOU, E.; LEE, D. Generative artificial intelligence, human creativity, and art. **PNAS Nexus**, v. 3, n. 3, p. pgae052, 2024.

ZHU, C.; BYRD, R. H.; LU, P.; NOCEDAL, J. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. **ACM Transactions on Mathematical Software (TOMS)**, v. 23, n. 4, p. 550–560, 1997.

ZHU, Q.; TREE, D. R. Simulations of morphology control of self-assembled amphiphilic surfactants. **Journal of Polymer Science**, v. 61, n. 12, p. 1214–1240, 2023.

ZHU, X.; GHAHRAMANI, Z. **Learning from Labeled and Unlabeled Data with Label Propagation**. [S. l.], 2002.

ZIPF, G. K. **Human Behavior and the Principle of Least Effort**. [S. l.]: Addison-Wesley Press, 1949.

APÊNDICE A – MODELO NUMÉRICO

O anexo aborda o uso da dinâmica molecular (DM) como uma técnica de simulação computacional para estudar o movimento de partículas em escalas atômicas ou mesoscópicas. A DM é aplicada aqui em uma abordagem multi-escala, visando relacionar propriedades observadas em escalas mesoscópicas com aquelas em escalas macroscópicas. Este método é crucial para entender como um conjunto de partículas em pequena escala traduz-se em propriedades de materiais viscoelásticos em escalas maiores. Esta compreensão é fundamental para controlar e compreender o comportamento dos materiais em nível mesoscópico, integrando a dinâmica e as interações em diferentes escalas.

A modelagem computacional multi-escala é uma ferramenta valiosa para pesquisa e desenvolvimento de materiais, fornecendo insights detalhados sobre o comportamento dos materiais em diferentes níveis de escala. Esta abordagem é essencial para o design de materiais com propriedades específicas, adaptando suas características para atender às necessidades específicas. A aplicabilidade dessa técnica estende-se a vários campos da ciência, incluindo biofísica, ciência dos materiais, engenharia, química, sistemas complexos e matéria mole.

A.1 Equações de Movimento das Partículas

As posições das partículas são determinadas por cálculos numéricos baseados na segunda lei de Newton. Esta abordagem permite replicar resultados significativos da mecânica analítica considerando as interações entre os corpos. A força resultante que atua em cada partícula é dada por:

$$\vec{F}_k(t) = m\vec{a}_k(t) = m \frac{d^2 \vec{r}_k(t)}{dt^2}, \quad (\text{A.1})$$

onde o vetor de posição $\vec{r}_k(t)$ para cada partícula k é expresso em coordenadas tridimensionais, integrando a dinâmica e as interações das partículas no espaço. A trajetória de um grupo de N partículas é representada em um espaço de $3N$ dimensões, com cada ponto ao longo dessa trajetória definido por um vetor de $3N$ elementos. As posições iniciais e velocidades das partículas ($\vec{r}_k(0), \vec{v}_k(0)$) são essenciais para iniciar a simulação e prever suas trajetórias futuras.

A discretização das trajetórias futuras é feita em intervalos regulares Δt , permitindo o acompanhamento e cálculo das mudanças no sistema ao longo do tempo. O algoritmo de Verlet é

um método computacional reconhecido para calcular a posição futura das partículas com base nas equações cinemáticas do movimento. Este algoritmo é particularmente útil para a simulação de DM, pois proporciona uma maneira eficiente de integrar as equações de movimento ao longo do tempo (Verlet, 1967). Esta abordagem multi-escala na DM fornece uma compreensão profunda do comportamento dos materiais em diferentes escalas, essencial para avanços na ciência dos materiais e em outras áreas relacionadas.

A.1.1 Discretização do Algoritmo de Verlet

O algoritmo de Verlet é uma técnica de integração numérica amplamente selecionada devido à sua capacidade de manter uma ótima conservação de energia e estabilidade da simulação ao longo do tempo, mesmo em cenários nos quais outros métodos tradicionais podem apresentar oscilações numéricas. A essência do método reside na atualização iterativa das posições das partículas com base em aproximações sucessivas da aceleração. Isso é realizado em passos discretos, o que proporciona uma abordagem robusta para a predição precisa das trajetórias das partículas ao longo do tempo. O algoritmo se concentra na relação entre duas expansões em série de Taylor do vetor posição para um tempo anterior $\vec{r}(t - \Delta t)$ e um tempo posterior $\vec{r}(t + \Delta t)$,

$$\vec{r}_k(t + \Delta t) \simeq \vec{r}_k(t) + \frac{d\vec{r}_k(t)}{dt} \Delta t + \frac{d^2\vec{r}_k(t)}{dt^2} \frac{\Delta t^2}{2} + \frac{d^3\vec{r}_k(t)}{dt^3} \frac{\Delta t^3}{6} + O(t^4), \quad (\text{A.2})$$

$$\vec{r}_k(t - \Delta t) \simeq \vec{r}_k(t) - \frac{d\vec{r}_k(t)}{dt} \Delta t + \frac{d^2\vec{r}_k(t)}{dt^2} \frac{\Delta t^2}{2} - \frac{d^3\vec{r}_k(t)}{dt^3} \frac{\Delta t^3}{6} + O(t^4), \quad (\text{A.3})$$

então ao somarmos as duas expressões mencionadas, isolarmos $\vec{r}_k(t + \Delta t)$ e substituímos $\frac{d^2\vec{r}_k(t)}{dt^2} = \vec{a}_k(t)$, obtemos

$$\vec{r}_k(t + \Delta t) = 2\vec{r}_k(t) + \vec{r}_k(t - \Delta t) + \vec{a}_k(t)\Delta t^2 + O(t^4), \quad (\text{A.4})$$

com isso, ao empregar diferenças simetricamente centradas para aproximar $\vec{a}_k(t)$, conseguimos alcançar uma maior precisão no método. Para calcular as velocidades, realizamos a subtração das expressões A.2 com a A.3 tal que

$$\frac{d\vec{r}_k}{dt} = \vec{v}_k(t) \simeq \frac{\vec{r}_k(t + \Delta t) - \vec{r}_k(t - \Delta t)}{2\Delta t} + O(t^2). \quad (\text{A.5})$$

Portanto, podemos descrever o algoritmo de Verlet como

(Algoritmo de Verlet)

Dadas as posições $\vec{r}_k(t - \Delta t)$ e $\vec{r}_k(t)$,

1. Determinar $\vec{a}_k(t)$ em função de $\{\vec{r}_i(t)\}$,
2. $\vec{r}_k(t + \Delta t) \leftarrow 2\vec{r}_k(t) + \vec{r}_k(t - \Delta t) + \vec{a}_k(t)\Delta t^2$,
3. $\vec{v}_k(t) \leftarrow [\vec{r}_k(t + \Delta t) - \vec{r}_k(t - \Delta t)]/(2\Delta t)$.

Entretanto, o algoritmo de Verlet apresenta algumas limitações notáveis. Uma delas é a considerável ordem de erro da velocidade, que é da ordem de $O(t^2)$. Além disso, as equações encontram problemas quando $t = 0$, especialmente devido à expressão $(\vec{r}_k(0 - \Delta t))$ (Allen; Tildesley, 2017). A determinação da velocidade é impactada pela presença do termo $2\Delta t$ no denominador, especialmente quando Δt é escolhido como um valor menor. Uma complicação adicional surge da necessidade de examinar a posição e a velocidade das partículas em instantes distintos, ou seja, nos intervalos de tempo $t - \Delta t$ e t . Isso implica que a velocidade no tempo t não pode ser calculada até que a coordenada no tempo $t + \Delta t$ seja estabelecida, introduzindo complexidades na sincronização das grandezas físicas.

A.1.2 Discretização de Velocity-Verlet

A discretização de Velocity-Verlet aborda eficazmente a questão do reescalonamento das velocidades presente no método de Verlet, ao mesmo tempo que assegura a conservação de energia. Ao reescrever a série de Taylor até a segunda ordem para a expressão A.2, temos

$$\vec{r}_k(t + \Delta t) = \vec{r}_k(t) + \vec{v}_k(t)\Delta t + \vec{a}_k(t)\frac{\Delta t^2}{2}, \quad (\text{A.6})$$

note que o vetor posição depende do vetor velocidade. Portanto, iremos realizar um novo desenvolvimento em série de Taylor para determinar as velocidades. Nesse processo, substituímos $\vec{r}_k(t + \Delta t) \rightarrow \vec{v}_k(t + \Delta t)$, resultando em

$$\vec{v}_k(t + \Delta t) = \vec{v}_k(t) + \vec{a}_k(t)\Delta t + \frac{d\vec{a}_k(t)}{dt}\frac{\Delta t^2}{2}, \quad (\text{A.7})$$

podemos encontrar o termo $\frac{d\vec{a}_k(t)}{dt}$ pela aproximação em série de Taylor de $\vec{a}_k(t + \Delta t)$ até a primeira ordem, resultando em

$$\vec{a}_k(t + \Delta t) = \vec{a}_k(t) + \frac{d\vec{a}_k(t)}{dt}\Delta t, \quad (\text{A.8})$$

e assim o valor de $\frac{d\vec{a}_k(t)}{dt}$ é dado

$$\frac{d\vec{a}_k(t)}{dt} = \frac{\vec{a}_k(t + \Delta t) - \vec{a}_k(t)}{\Delta t}, \quad (\text{A.9})$$

e reescrevendo a expressão A.7 com o valor de $\frac{d\vec{a}_k(t)}{dt}$, obtemos

$$\begin{aligned} \vec{v}_k(t + \Delta t) &= \vec{v}_k(t) + \vec{a}_k(t)\Delta t + \left[\frac{\vec{a}_k(t + \Delta t) - \vec{a}_k(t)}{\Delta t} \right] \frac{\Delta t^2}{2} \\ &= \vec{v}_k(t) + \frac{[\vec{a}_k(t + \Delta t) + \vec{a}_k(t)]}{2} \Delta t. \end{aligned} \quad (\text{A.10})$$

Assim, as equações A.6 e A.10 descrevem o método de discretização de Velocity-Verlet que permitem prever as futuras posições e velocidades das partículas. Essa previsão ocorre ao avançarmos no tempo, indo de $(\vec{r}_k(t), \vec{v}_k(t))$ para $(\vec{r}_k(t + \Delta t), \vec{v}_k(t + \Delta t))$ (Allen; Tildesley, 2017; Allen, 2004).

A.1.3 Algoritmo de Velocity-Verlet

O algoritmo de Velocity-Verlet, popularmente conhecido como o algoritmo leap-frog, é uma extensão do algoritmo de Verlet e é fundamentado na expansão em série de Taylor para a resolução de equações de movimento, notadamente na equação A.6. Ao examinar os dois últimos termos dessa equação e destacar o termo Δt , podemos reestruturar o procedimento da seguinte forma

$$\vec{r}_k(t + \Delta t) = \vec{r}_k(t) + \underbrace{\left[\vec{v}_k(t) + \vec{a}_k(t) \frac{\Delta t}{2} \right]}_{\vec{v}_k(t + \frac{\Delta t}{2})} \Delta t, \quad (\text{A.11})$$

com isso, o termo em destaque representa a expansão da série de Taylor na velocidade em $(t + \frac{\Delta t}{2})$, então

$$\vec{v}_k\left(t + \frac{\Delta t}{2}\right) = \left[\vec{v}_k(t) + \vec{a}_k(t) \frac{\Delta t}{2} \right], \quad (\text{A.12})$$

reescrevendo a equação A.11, podemos expressar a equação da posição da seguinte forma

$$\vec{r}_k(t + \Delta t) = \vec{r}_k(t) + \vec{v}_k\left(t + \frac{\Delta t}{2}\right) \Delta t, \quad (\text{A.13})$$

reorganizando a equação A.10, podemos expressar o vetor velocidade como

$$\vec{v}_k(t + \Delta t) = \underbrace{\left[\vec{v}_k(t) + \vec{a}_k(t) \frac{\Delta t}{2} \right]}_{\vec{v}_k(t + \frac{\Delta t}{2})} + \frac{\vec{a}_k(t + \Delta t)}{2} \Delta t. \quad (\text{A.14})$$

reescrevendo a expressão A.14, temos

$$\vec{v}_k(t + \Delta t) = \vec{v}_k \left(t + \frac{\Delta t}{2} \right) + \frac{\vec{a}_k(t + \Delta t)}{2} \Delta t. \quad (\text{A.15})$$

Portanto, o algoritmo de Velocity-Verlet é descrito como

(Algoritmo de Velocity-Verlet)

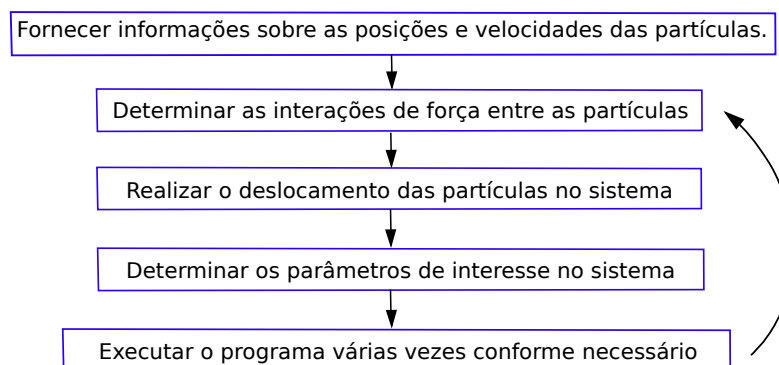
Dada a posição $\vec{r}_k(t)$ e a velocidade $\vec{v}_k(t)$,

1. Determinar $\vec{a}_k(t)$ em função de $\{\vec{r}_k(t)\}$,
2. $\vec{v}_k(t + \frac{\Delta t}{2}) \leftarrow \vec{v}_k(t) + \vec{a}_k(t) \frac{\Delta t}{2}$,
3. $\vec{r}_k(t + \Delta t) \leftarrow \vec{r}_k(t) + \vec{v}_k(t + \frac{\Delta t}{2}) \Delta t$,
4. Determinar $\vec{a}_k(t + \Delta t)$ em função de $\{\vec{r}_k(t + \Delta t)\}$,
5. $\vec{v}_k(t + \Delta t) \leftarrow \vec{v}_k(t + \frac{\Delta t}{2}) + \frac{\vec{a}_k(t + \Delta t)}{2} \Delta t$.

Assim, esse método não apresenta problemas em $t = 0$, pois as posições e velocidades das partículas são consideradas simultaneamente. Além disso, as equações de velocidade não são afetadas por divisões envolvendo valores menores. Uma desvantagem do método é a necessidade de realizar dois cálculos numéricos das acelerações em um único ciclo de integração das equações diferenciais. Apesar disso, o algoritmo de Velocity-Verlet destaca-se pela sua eficiência na descrição dos movimentos das partículas.

A simulação de Dinâmica Molecular (DM) pode ser dividida em três etapas distintas, sendo essas fases associadas cronologicamente. Inicialmente, ocorre a inicialização do sistema. Em seguida, realiza-se o cálculo das forças de interação entre as partículas no sistema, e, por último, efetua-se o cálculo dos parâmetros de interesse. O algoritmo é construído conforme o fluxograma ilustrado na Figura 40. No contexto deste trabalho, os resultados obtidos visam explicar a relaxação de materiais viscoelásticos, que alteram sua configuração ao longo do tempo por meio de um sistema físico em desequilíbrio.

Figura 40 – Diagrama de fluxo do algoritmo de Dinâmica Molecular.



Fonte: Elaborado pelo autor.

A.1.4 Condições Periódicas de Contorno (CPC)

Nas simulações de dinâmica molecular, um desafio comum é lidar com os efeitos de superfície, especialmente em amostras de tamanho limitado. Em pequenas amostras, uma fração significativa de partículas localiza-se na superfície, o que pode levar a resultados distorcidos devido ao efeito de borda. Por exemplo, em um cubo de 10^3 , quase metade das moléculas está na superfície. Em um cubo maior de 100^3 , aproximadamente 6% das substâncias estão nas faces externas. Essas partículas superficiais experimentam forças diferentes das partículas internas, o que pode afetar a precisão das simulações (Allen, 2004).

Para superar esses desafios, as condições periódicas de contorno (CPC) são frequentemente utilizadas. Na CPC, a caixa cúbica de simulação é replicada em todas as direções para formar uma rede infinita. Isso significa que quando uma molécula se move para fora da caixa original, ela reaparece na face oposta, criando a ilusão de um sistema infinito. Essa técnica elimina efetivamente as superfícies que poderiam influenciar o comportamento das moléculas, permitindo que a simulação reflita mais precisamente um sistema macroscópico.

Para implementar a CPC, as posições das partículas são ajustadas conforme elas atravessam as fronteiras da caixa. Se uma partícula cruza uma das fronteiras da caixa de lado L , sua posição é corrigida subtraindo ou adicionando L , dependendo da direção em que atravessou a fronteira. Essa abordagem garante que as partículas permaneçam efetivamente dentro dos limites da caixa de simulação, enquanto simulam um ambiente sem limites físicos. Uma representação bidimensional desse sistema periódico está ilustrada na Figura 41.

Essas correções são aplicadas a cada componente (x , y , z) da posição da partícula,

como expresso pelas seguintes equações

$$\left\{ \begin{array}{l} \text{se } x_k \geq \frac{L}{2} \longrightarrow x_k = x_k - L, \text{ se } x_k \leq -\frac{L}{2} \longrightarrow x_k = x_k + L, \end{array} \right. \quad (\text{A.16})$$

$$\left\{ \begin{array}{l} \text{se } y_k \geq \frac{L}{2} \longrightarrow y_k = y_k - L, \text{ se } y_k \leq -\frac{L}{2} \longrightarrow y_k = y_k + L, \end{array} \right. \quad (\text{A.17})$$

$$\left\{ \begin{array}{l} \text{se } z_k \geq \frac{L}{2} \longrightarrow z_k = z_k - L, \text{ se } z_k \leq -\frac{L}{2} \longrightarrow z_k = z_k + L. \end{array} \right. \quad (\text{A.18})$$

Essas correções garantem a continuidade do movimento das partículas e são fundamentais para a realização de simulações precisas em dinâmica molecular, especialmente em sistemas de tamanho limitado.

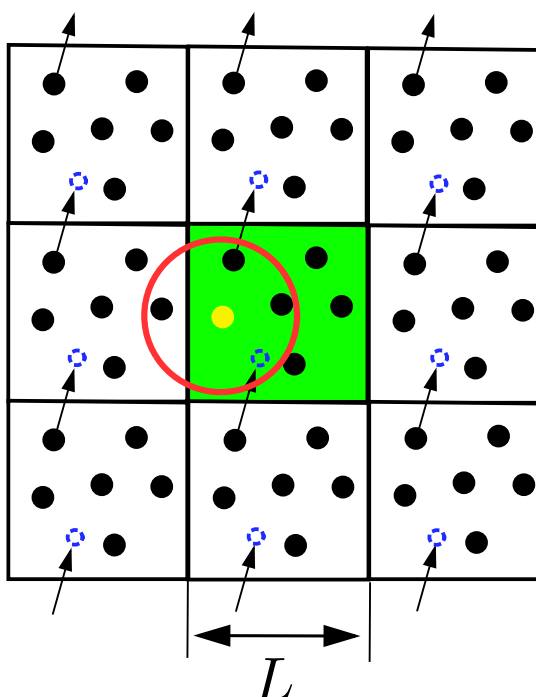
A.1.5 Convenção de Mínima Imagem e Potencial de Truncamento

Ao lidar com simulações de dinâmica molecular em sistemas sujeitos a condições periódicas de contorno (CPC), um aspecto crítico é o cálculo eficiente da energia potencial e das forças conservativas atuando sobre as moléculas. Devido à natureza infinita da CPC, teoricamente seria necessário considerar interações entre uma molécula e um número ilimitado de suas imagens em caixas replicadas. No entanto, isso é inviável na prática devido às limitações computacionais.

Para contornar esse problema, o método de convenção de imagem mínima é comumente empregado. Essa abordagem simplifica o cálculo ao considerar as interações de uma molécula apenas com outras moléculas localizadas dentro da mesma caixa de simulação ou em suas imagens mais próximas. Assim, cada molécula interage apenas com outras moléculas cujos centros estão dentro de uma distância que não excede metade do comprimento da caixa de simulação em qualquer direção.

O cálculo da energia potencial e das forças em sistemas com muitas partículas ainda pode ser computacionalmente exigente, mesmo com a convenção de imagem mínima. Para sistemas grandes, como aqueles com 1000 partículas ou mais, o número de interações a serem calculadas pode ser enorme. Para mitigar esse desafio, uma técnica comum é aplicar um "raio de corte" ou "distância de corte" nas interações. Isso significa que o potencial entre duas moléculas é considerado zero quando a distância entre elas excede um valor pré-determinado, denominado raio de corte r_c . Consequentemente, apenas as interações entre moléculas separadas por distâncias menores que r_c são levadas em conta nos cálculos.

Figura 41 – Condições periódicas de contorno (CPC) e convenção de imagem mínima.



Fonte: Elaborado pelo autor.

Essa abordagem de corte reduz significativamente o número de cálculos necessários, tornando as simulações mais viáveis do ponto de vista computacional. É especialmente eficaz em sistemas onde o potencial de interação diminui rapidamente com a distância, como é o caso em muitas interações moleculares.

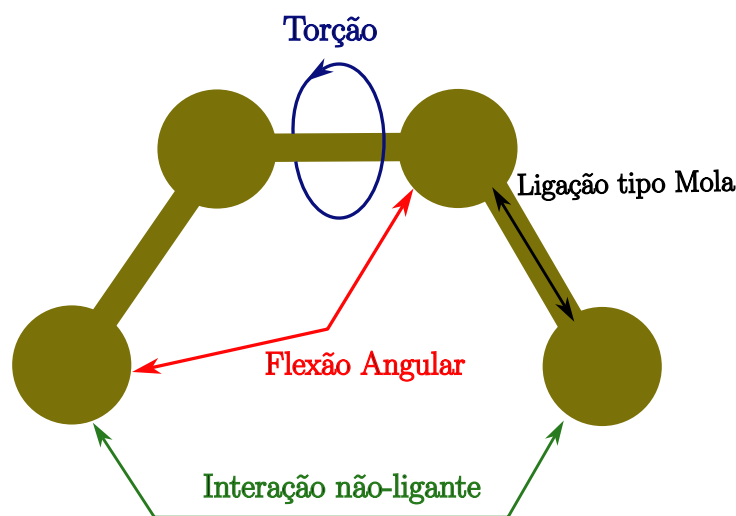
A Figura 41 ilustra a aplicação do raio de corte em uma molécula específica (destacada em amarelo), onde apenas as moléculas dentro do raio de corte são consideradas para interações. Esse método é essencial para garantir a eficiência computacional em simulações de dinâmica molecular, permitindo o estudo de sistemas complexos em escalas de tempo e tamanhos relevantes para aplicações práticas.

A.2 Potenciais de Interações

A caracterização de materiais em escalas mesoscópicas é profundamente influenciada pelas interações entre suas partículas constituintes. Estas interações são fundamentais para definir as propriedades físicas dos materiais, incluindo elasticidade, viscosidade e condutividade elétrica. O estudo detalhado destas interações em escalas mesoscópicas revela insights essenciais sobre o funcionamento das forças conservativas e sua influência no comportamento coletivo do material.

A escolha do potencial de interação apropriado é crucial para representar corre-

Figura 42 – Ilustração das interações ligantes e não-ligantes entre as partículas.



Fonte: Elaborado pelo autor.

tamente as forças atuantes e os resultados teóricos ou experimentais desejados. As forças conservativas \vec{F}_k em um sistema de partículas podem ser derivadas da energia potencial total $U(\vec{r}^N)$ do sistema, conforme a equação

$$\vec{F}_k = -\nabla U(\vec{r}^N) = -\frac{\partial U(\vec{r}^N)}{\partial r_k^N} = -\left(\frac{\partial U(\vec{r}^N)}{\partial x_k}, \frac{\partial U(\vec{r}^N)}{\partial y_k}, \frac{\partial U(\vec{r}^N)}{\partial z_k}\right). \quad (\text{A.19})$$

As interações em sistemas moleculares podem ser divididas em intramoleculares (ligações) e intermoleculares (não-ligantes). As ligações intramoleculares, que conectam os átomos dentro de uma molécula, são frequentemente modeladas usando potenciais que simulam molas para descrever as ligações e ângulos diedros para as rotações. Por outro lado, as interações entre partículas não diretamente ligadas são frequentemente descritas pelo potencial de Lennard-Jones, que representa tanto forças atrativas quanto repulsivas. Para partículas carregadas, como íons, a lei de Coulomb é usada para descrever as forças eletrostáticas.

Uma análise aprofundada das interações potenciais é crucial para entender a dinâmica molecular de um material. Vários potenciais são empregados, dependendo das características específicas das partículas e de suas interações. A Figura 42 pode ilustrar uma representação visual destas interações, enfatizando a complexidade e a diversidade das forças que governam o comportamento dos sistemas moleculares. Assim, a compreensão destas interações é fundamental para projetar materiais com propriedades específicas e para avançar no campo da ciência dos materiais.

$$U = \underbrace{\sum_{\text{ligações}} \frac{k_{kl}(r_{kl} - \ell)^2}{2} + \sum_{\text{ângulos}} \frac{k_{klm}(\theta_{klm} - \theta_0)^2}{2} + \sum_{\text{torções}} \sum_q \frac{k_{klmn}(1 + \cos(q\varphi_{klmn} - \Theta_q))}{2}}_{\text{interações de ligação}} \quad (\text{A.20})$$

$$+ \underbrace{\sum_k \sum_{l>k} \left(\frac{q_k q_l}{4\pi\epsilon_0} \frac{e^{-r_{kl}\xi}}{r_{kl}} + \frac{A_{kl}}{r_{kl}^m} - \frac{B_{kl}}{r_{kl}^n} \right)}_{\text{interações não-ligantes}} \quad (\text{A.21})$$

Na equação A.20, a primeira parte representa a combinação de todas as ligações entre partículas, onde as interações elásticas são modeladas como molas. Aqui, $r_{kl} = |\vec{r}_k - \vec{r}_l|$ indica a distância entre substâncias k e l , ℓ é a distância de equilíbrio entre os centros das mesmas, e k_{ij} é a constante elástica associada. O segundo termo da equação diz respeito à soma de todos os ângulos de ligação, em que θ_{klm} é o vetor que descreve as coordenadas de três partículas e θ_0 é o ângulo de ligação inicial. A terceira componente abrange a soma de todos os ângulos de torção, envolvendo um vetor φ_{klmn} com quatro coordenadas de partículas. Esse potencial incorpora uma expansão periódica de ordem q em torno da rotação entre as amostras l e m , e Θ_q representa o ângulo de fase de torção.

Na equação A.21, a primeira parte engloba a soma de todos os potenciais de Yukawa associados às interações entre partículas carregadas, sendo uma extensão do potencial de Coulomb. Nesse contexto, q_k e q_l representam as cargas elétricas das partículas, r_{kl} é a distância entre elas, ϵ_0 é a permissividade do vácuo, e ξ é o comprimento de blindagem, indicando a rapidez com que o potencial de Yukawa diminui com a distância. Em essência, o potencial de Yukawa modifica o potencial de Coulomb para considerar o efeito de blindagem elétrica em meios condutores, refletido pelo termo $e^{-r_{kl}\xi}$. Esse termo atenua a força das interações à medida que a distância entre as partículas aumenta, representando como a presença de outras partículas reduz a influência do campo elétrico de uma partícula específica.

O potencial de Yukawa apresenta aplicabilidade em diversos contextos, servindo como um modelo simplificado para partículas carregadas em plasmas empoeirados, permitindo a construção de modelos mais realistas para representar esses plasmas em diferentes condições (Ohta; Hamaguchi, 2000). Além disso, é empregado em suspensões coloidais com cargas diluídas, como esferas de látex em água, micelas e microemulsões (Kremer *et al.*, 1986). Este potencial, predominantemente repulsivo, sugere que as partículas tendem a se afastar umas das

outras. Em sistemas de auto-organização, onde se explora a formação de estruturas organizadas, a repulsão entre as partículas desempenha um papel crucial ao estabelecer espaçamentos e prevenir aglomerações excessivas (De Araújo, 2018).

A segunda componente da equação A.21 refere-se à contribuição das forças repulsivas, representadas por $\frac{A_{kl}}{r_{kl}^m}$, entre as partículas k e l , onde r_{kl} é a distância entre as partículas e m é um expoente relacionado ao alcance do potencial. Por outro lado, a última componente da equação representa a contribuição das forças atrativas, indicadas por $-\frac{B_{kl}}{r_{kl}^n}$, entre as partículas, sendo n é outro expoente relacionado ao alcance do potencial. As constantes A_{kl} e B_{kl} são ajustadas empiricamente para representar as propriedades específicas das interações entre as partículas consideradas. Essas constantes podem ter dimensões de energia vezes volume elevado a um certo expoente, resultando em uma unidade de energia. O potencial de Mie $U_M(r_{kl})$ pode ser representado somando essas duas contribuições, uma para interações atrativas de longas distâncias e outra para interações repulsivas em curtas distâncias (Peña *et al.*, 2015). Essa soma é dada por

$$U_M(r_{kl}) = \frac{A_{kl}}{r_{kl}^m} - \frac{B_{kl}}{r_{kl}^n}, \quad (\text{A.22})$$

de maneira mais abrangente, é possível representar o potencial de Mie de forma generalizada ao substituir os parâmetro $A_{kl} = \left(\frac{m}{m-n}\right)\left(\frac{m}{n}\right)^{\left(\frac{n}{m-n}\right)}\nu\rho^m$ e $B_{kl} = \left(\frac{m}{m-n}\right)\left(\frac{m}{n}\right)^{\left(\frac{n}{m-n}\right)}\nu\rho^n$. Aqui, o ν representa uma constante de energia e ρ é o diâmetro das partículas. Esses parâmetros do potencial podem ser ajustados de acordo com as características específicas das partículas em análise, tal que

$$U_M(r_{kl}) = \left(\frac{m}{m-n}\right)\left(\frac{m}{n}\right)^{\left(\frac{n}{m-n}\right)}\nu\left[\left(\frac{\rho}{r_{kl}}\right)^m - \left(\frac{\rho}{r_{kl}}\right)^n\right]. \quad (\text{A.23})$$

Para obter o potencial de Lennard-Jones a partir do potencial de Mie, é possível estabelecer relações específicas entre os parâmetros. O potencial de Lennard-Jones pode ser considerado como um caso especial do potencial de Mie, onde os expoentes são definidos como $m = 12$ e $n = 6$ (Peña *et al.*, 2015). Ao substituímos esses valores no potencial de Mie, obtemos a expressão característica do potencial de Lennard-Jones

$$U_{LJ}(r_{kl}) = 4\nu\left[\left(\frac{\rho}{r_{kl}}\right)^{12} - \left(\frac{\rho}{r_{kl}}\right)^6\right], \quad (\text{A.24})$$

e, considerando que $4v$ é uma constante energética, podemos reescrever isso como $\varepsilon = 4v$, onde ε é outra constante de energia que absorve o fator 4, tal que

$$U_{LJ}(r_{kl}) = \varepsilon \left[\left(\frac{\rho}{r_{kl}} \right)^{12} - \left(\frac{\rho}{r_{kl}} \right)^6 \right]. \quad (\text{A.25})$$

O potencial de Lennard-Jones tem uma ampla gama de aplicações em diversos campos da ciência. Em sistemas coloidais, especialmente aqueles constituídos por partículas coloidais carregadas, conhecidas como *patchy*, este potencial é empregado para modelar as interações entre as partículas, fornecendo uma base essencial para a compreensão da auto-organização e do comportamento agregado desses sistemas complexos (De Araújo, 2018; Billotta, 2018). Além disso, em estudos moleculares, o potencial de Lennard-Jones é uma ferramenta valiosa para descrever as interações entre átomos ou moléculas. No contexto do argônio líquido, esse potencial desempenha um papel crucial na análise das propriedades termodinâmicas e estruturais desse líquido nobre (Rahman, 1964; Verlet, 1967). Em sistemas nos quais interações de potenciais competitivos estão presentes, como nos casos do potencial de Lennard-Jones e do potencial de Yukawa (Costa, 2019).

Ao considerar o limite $r_{kl} \rightarrow 0$ na expressão A.25, obtemos

$$\lim_{r_{kl} \rightarrow 0} U_{LJ}(r_{kl}) = \varepsilon \left[\left(\frac{\rho}{r_{kl}} \right)^{12} - \left(\frac{\rho}{r_{kl}} \right)^6 \right] \simeq \varepsilon \left(\frac{\rho}{r_{kl}} \right)^{12}, \quad (\text{A.26})$$

esse resultado indica que quando r_{kl} se aproxima de zero, o termo dominante é $\left(\frac{\rho}{r_{kl}} \right)^{12}$, caracterizando uma interação de repulsão de curto alcance, também conhecida como *soft-core*. Isso reflete o princípio de exclusão de Pauli, que estabelece que partículas não podem ocupar a mesma posição, resultando em uma repulsão significativa em distâncias muito curtas. Já para o caso em que o limite $r_{kl} \rightarrow \infty$ em A.25, obtemos

$$\lim_{r_{kl} \rightarrow \infty} U_{LJ}(r_{kl}) = \varepsilon \left[\left(\frac{\rho}{r_{kl}} \right)^{12} - \left(\frac{\rho}{r_{kl}} \right)^6 \right] \simeq -\varepsilon \left(\frac{\rho}{r_{kl}} \right)^6, \quad (\text{A.27})$$

nesse cenário, o termo dominante é $-\left(\frac{\rho}{r_{kl}} \right)^6$, representando uma força atrativa de longo alcance. Esse comportamento é característico da interação de van der Waals, onde a atração entre as partículas ocorre em distâncias maiores.

A força \vec{F}_{klLJ} associada ao potencial de Lennard-Jones é dada

$$\begin{aligned}\vec{F}_{klLJ} &= F_x \hat{i} + F_y \hat{j} + F_z \hat{k} = -\frac{dU_{LJ}(r_{kl})}{dr_{kl}} = \frac{12\varepsilon}{\rho} \left[\left(\frac{\rho}{r_{kl}} \right)^{13} - \frac{1}{2} \left(\frac{\rho}{r_{kl}} \right)^7 \right] \frac{x}{r_{kl}} \hat{i} \\ &+ \frac{12\varepsilon}{\rho} \left[\left(\frac{\rho}{r_{kl}} \right)^{13} - \frac{1}{2} \left(\frac{\rho}{r_{kl}} \right)^7 \right] \frac{y}{r_{kl}} \hat{j} + \frac{12\varepsilon}{\rho} \left[\left(\frac{\rho}{r_{kl}} \right)^{13} - \frac{1}{2} \left(\frac{\rho}{r_{kl}} \right)^7 \right] \frac{z}{r_{kl}} \hat{k} \\ &= \frac{12\varepsilon}{\rho^2} \left[\left(\frac{\rho}{r_{kl}} \right)^{14} - \frac{1}{2} \left(\frac{\rho}{r_{kl}} \right)^8 \right] \vec{r}_{kl}.\end{aligned}\quad (\text{A.28})$$

Uma análise adicional que podemos conduzir é a determinação do ponto de mínimo no potencial de Lennard-Jones, alcançado ao igualar a primeira derivada a zero. Isso nos leva à seguinte equação

$$\frac{dU_{LJ}(r_{kl})}{dr_{kl}} = \frac{12\varepsilon}{\rho} \left[\left(\frac{\rho}{r_{kl}} \right)^{13} - \frac{1}{2} \left(\frac{\rho}{r_{kl}} \right)^7 \right] = 0, \quad (\text{A.29})$$

temos que a distância r_{min} é dada por

$$\frac{r^{13}}{r^7} = 2 \frac{\rho^{13}}{\rho^7} \longrightarrow r_{min} = 2^{\frac{1}{6}} \rho \simeq 1.122\rho, \quad (\text{A.30})$$

assumindo que ε e ρ são iguais a 1, então o potencial mínimo $U_{LJ_{min}}(2^{\frac{1}{6}})$ é dado por

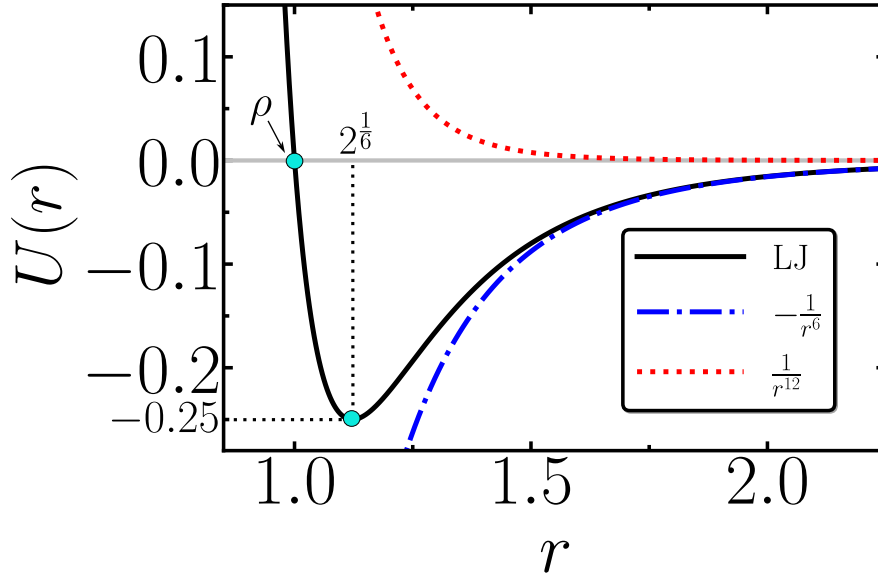
$$U_{LJ_{min}} = \left(\frac{1}{2^{\frac{1}{6}}} \right)^{12} - \left(\frac{1}{2^{\frac{1}{6}}} \right)^6 = -\frac{1}{4}, \quad (\text{A.31})$$

na Figura 43, é possível visualizar o potencial de Lennard-Jones e as forças atrativas e repulsivas de maneira isolada, assumindo que ε e ρ são iguais a 1.

Na equação A.25, os parâmetros estão definidos para partículas idênticas. Contudo, ao lidar com interações de Lennard-Jones entre partículas distintas, é essencial determinar os parâmetros do potencial correspondentes. Estes parâmetros incluem tipicamente a profundidade do potencial, representada por ε , e o parâmetro de tamanho ou diâmetro da partícula, representado por ρ . A regra de Berthelot é frequentemente empregada para calcular a profundidade média do potencial, denotada como ε_{ab} , quando tratamos de interações entre dois tipos diferentes de partículas, referidas como a e b . Esta regra utiliza a média geométrica dos parâmetros ε_a e ε_b , como expresso na seguinte equação

$$\varepsilon_{ab} = \sqrt{\varepsilon_a \varepsilon_b}, \quad (\text{A.32})$$

Figura 43 – Gráfico do potencial de Lennard-Jones, destacando as influências de $\frac{1}{r^{12}}$ e $-\frac{1}{r^6}$ para valores de $\rho = 1$ e $\varepsilon = 1$.



Fonte: Elaborado pelo autor.

por outro lado, a regra de Lorentz é aplicada para determinar parâmetro de tamanho médio ρ_{ab} . Essa regra utiliza a média aritmética dos parâmetros ρ_a e ρ_b , como mostrado abaixo

$$\rho_{ab} = \frac{\rho_a + \rho_b}{2}, \quad (\text{A.33})$$

com base nas regras de mistura para diferentes tipos de partículas, a expressão ajustada para o potencial de Lennard-Jones é fornecida por

$$U_{LJ}(r_{kl}) = \varepsilon_{ab} \left[\left(\frac{\rho_a}{r_{kl} - \Delta_{ab}} \right)^{12} - \left(\frac{\rho_a}{r_{kl} - \Delta_{ab}} \right)^6 \right], \quad (\text{A.34})$$

onde $\Delta_{ab} = ((\rho_a + \rho_b)/2 - \rho_a)$ representa um ajuste na distância entre as partículas, levando em conta os diâmetros médios das mesmas. Essa correção é essencial para considerar a influência das diferentes geometrias e tamanhos das partículas na interação, deslocando o potencial de ação entre as superfícies das substâncias (Long; Ferguson, 2018; Delhommelle; Millie, 2001; Schnabel *et al.*, 2007). É importante notar que, se as partículas forem idênticas (com $a = b$), então $\Delta_{aa} = 0$, $\rho_a = \rho$, e $\varepsilon_{aa} = \varepsilon$, retornando assim à expressão original A.25.

A.2.1 Unidades reduzidas (ou adimensionais)

As unidades reduzidas, ou adimensionais, em simulações computacionais são adotadas para simplificar e universalizar a representação de grandezas físicas em escalas interatômicas

ou mesoscópicas, onde os valores das unidades físicas são extremamente pequenos. A escolha dessas unidades reduzidas visa simplificar as equações, eliminando constantes físicas e proporcionando uma descrição mais simplificada dos fenômenos físicos.

Os resultados de simulações realizadas em unidades reduzidas podem sempre ser convertidos de volta para unidades reais. Por exemplo, consideremos uma simulação em um modelo Lennard-Jones em que a temperatura adimensional é representada por $T^* = 1$ e a pressão adimensional por $P^* = 1$. Se desejamos comparar esses resultados com dados experimentais para o argônio, onde $\frac{\epsilon}{k_B} = 119.8K$ (ϵ é a profundidade do poço do potencial de Lennard-Jones e k_B é a constante de Boltzmann), $\rho = 3.4 \times 10^{-10}m$, e $M = 0.03994kg/mol$ (Frenkel; Smit, 2002), podemos usar a tabela de tradução de unidades na Tabela 1 para converter os parâmetros da simulação de volta para unidades reais do Sistema Internacional (SI).

Tabela 1 – Conversão de unidades adimensionais para unidades reais no modelo Lennard-Jones para Argônio ($\frac{\epsilon}{k_B} = 119.8K$, $\rho = 3.4 \times 10^{-10}m$, $M = 0.03994kg/mol$).

Quantidade	Unidade Reduzida		Unidade Real
Temperatura	$T^* = 1$	\leftrightarrow	$T = 119.8K$
Densidade	$d^* = 1$	\leftrightarrow	$d = 1680kg/m^3$
Tempo	$\Delta t^* = 0005$	\leftrightarrow	$\Delta t = 1.09 \times 10^{-14}s$
Pressão	$P^* = 1$	\leftrightarrow	$P = 41.9MPa$

Fonte: (Frenkel; Smit, 2002), com adaptações.

Essa comparação ressalta a importância das unidades reduzidas, especialmente quando os dados experimentais apresentam valores na ordem de 10^{-14} ou ainda menores. Utilizar essas unidades facilita a detecção de erros computacionais, pois operações que envolvem quantidades muito pequenas correm o risco de resultar em erros de overflow ou underflow ao utilizar a multiplicação de ponto flutuante padrão.

Ao utilizar essas unidades reduzidas, as equações de movimento e os potenciais de interação tornam-se mais simplificados, pois muitas constantes e parâmetros específicos do sistema são absorvidos nessas unidades. Isso facilita a comparação entre diferentes sistemas e escalas, permitindo uma compreensão mais geral dos fenômenos em estudo.

Neste estudo, focamos nos potenciais de ligação descritos pela equação A.20 e no termo das forças repulsivas de curto alcance do potencial de Lennard-Jones modificado, conforme apresentado na equação A.34, para diferentes partículas, mas introduzindo o expoente ϖ para controlar a dureza. Considere um sistema físico composto por N substâncias. A energia total do sistema, representada por $E = K + U + U_{ext}$, onde $K = \frac{mv^2}{2}$ é a energia cinética das partículas, U

está relacionado às forças conservativas e U_{ext} é o termo de energia associado às forças externas, pode ser expressa como

$$E = \sum_k \frac{mv_k^2}{2} + \sum_k \sum_{k>l} \frac{k_{kl}(r_{kl} - \ell)^2}{2} + \varepsilon \sum_k \sum_{k>l} \left(\frac{\rho_a}{r_{kl} - \frac{(\rho_b - \rho_a)}{2}} \right)^{\overline{\omega}} + U_{ext}, \quad (\text{A.35})$$

ao reduzir a unidade de energia E_k através da constante de energia ε associada a cada partícula k , podemos expressar cada termo da equação acima de maneira adimensional, temos

$$\frac{E_k}{\varepsilon} = \frac{mv_k^2}{2\varepsilon} + \sum_l \frac{k_{kl}(r_{kl} - \ell)^2}{2\varepsilon} + \sum_l \left(\frac{\rho_a}{r_{kl} - \frac{(\rho_b - \rho_a)}{2}} \right)^{\overline{\omega}} + \frac{U_{ext}}{\varepsilon}, \quad (\text{A.36})$$

dessa forma, as unidades adimensionais surgirão quando definirmos,

- Energia adimensional $E^* = \frac{E_k}{\varepsilon}$;
- Constante da mola adimensional $k^* = \frac{k_{kl}\rho^2}{\varepsilon} \rightarrow k_{kl} = \frac{\varepsilon k^*}{\rho^2}$;
- Tempo adimensional $t^* = \frac{t}{t_0}$;
- Posição adimensional para partículas idênticas $r^* = \frac{r}{\rho} \rightarrow r = \rho r^*$;
- Posição adimensional para partículas distintas $r^* = \frac{r_{kl} - (\rho_b - \rho_a)/2}{\rho_a}$;
- Velocidade adimensional $v^* = \frac{v_k t_0}{\rho} \rightarrow v_k = \frac{\rho v^*}{t_0}$;
- Potencial de ligação adimensional $U_{ligante}^* = \frac{U_{ligante}}{\varepsilon}$;
- Potencial da força conservativa adimensional $U^* = \frac{U_k}{\varepsilon} \rightarrow U_k = \varepsilon U^*$;
- Termo de energia externa adimensional $U_{ext}^* = \frac{U_{ext}}{\varepsilon}$.

É importante notar que a constante de energia ε possui dimensões de massa multiplicada pelo quadrado do comprimento dividido pelo quadrado do tempo $[ML^2T^{-2}]$, e, portanto, deve incluir um termo de massa m na energia cinética adimensional, tal que

$$E^* = \frac{m\rho^2}{t_0^2\varepsilon} \frac{m^*v^{*2}}{2} + \sum_l \frac{k^*(r_{kl} - \ell)^2}{2\rho^2} + \sum_l \left(\frac{1}{r^*} \right)^{\overline{\omega}} + U_{ext}^* = \frac{m\rho^2}{t_0^2\varepsilon} K^* + U^* + U_{ext}^*. \quad (\text{A.37})$$

Note que os parâmetros multiplicando a energia cinética adimensional K^* resultam em 1. Portanto, podemos determinar o parâmetro de tempo característico t_0 da simulação, tal que

$$\frac{m\rho^2}{t_0^2\varepsilon} = 1 \longrightarrow t_0 = \sqrt{\frac{m\rho^2}{\varepsilon}}. \quad (\text{A.38})$$

Com isso, podemos encontrar as equações de movimento de Newton com as unidades reduzidas, utilizando o resultado da expressão A.38 para $\frac{m\rho^2}{t_0^2\varepsilon} = 1$, obtendo

$$m \frac{d^2 r_k}{dt^2} = m \frac{dv_k}{dt} = \frac{m\rho}{t_0^2} m^* \frac{dv_k^*}{dt^*} = \frac{m\rho}{t_0^2} m^* \frac{d^2 r_k^*}{dt^{*2}} = -\frac{\varepsilon}{\rho} \frac{\partial U^*}{\partial r_k^*} = \frac{\varepsilon}{\rho} F_k^* = F_k, \quad (\text{A.39})$$

tal que

$$F^* = \frac{\rho F_k}{\varepsilon} = \frac{\rho}{\varepsilon} \left(\frac{m\rho}{t_0^2} m^* \frac{d^2 r_k^*}{dt^{*2}} \right) = \left(\frac{m\rho^2}{t_0^2 \varepsilon} \right) m^* \frac{d^2 r_k^*}{dt^{*2}} = m^* \frac{d^2 r_k^*}{dt^{*2}}, \quad (\text{A.40})$$

utilizando as seguintes relações: $\frac{d}{dt} = \frac{dt^*}{dt} \frac{d}{dt^*} = \frac{1}{t_0} \frac{d}{dt^*}$; $v_k = \frac{\rho v^*}{t_0}$; $\frac{\partial}{\partial r_k} = \frac{\partial r_k^*}{\partial r_k} \frac{\partial}{\partial r_k^*} = \frac{1}{\rho} \frac{\partial}{\partial r_k^*}$; $U_k = \varepsilon U^*$.

APÊNDICE B – ALGUNS RESULTADOS OU PROPRIEDADES DAS TRANSFORMADAS DE LAPLACE

B.1 Transformada de Laplace

Seja uma função $f(t)$ integrável definida para todo número real $t > 0$, então a transformada de Laplace de $f(t)$, que denotamos por $\mathcal{L}[f(t)]$ ou $\tilde{F}(s)$, é definida pela integral imprópria (Teodoro, 2014)

$$\mathcal{L}[f(t)] = \int_0^{\infty} f(t)e^{-st} dt = \tilde{F}(s), \quad (\text{B.1})$$

com s sendo um parâmetro complexo, onde trabalhamos apenas com a parte real $Re(s) > 0$, com isso, podemos escolher $s \in \mathbb{R}$.

A transformada inversa de Laplace de uma função $\tilde{F}(s)$, geralmente denotada como $f(t)$, é o processo matemático de recuperar a função original no domínio do tempo, dada

$$f(t) = \mathcal{L}^{-1}[\tilde{F}(s)], \quad (\text{B.2})$$

o processo de cálculo da transformada inversa de Laplace envolve a aplicação de técnicas matemáticas, como a integral complexa de inversão (também conhecida como integral de Bromwich) ou tabelas de transformadas de Laplace inversas, dependendo da complexidade da transformada de Laplace e da função que está sendo analisada. A integral de Bromwich é uma ferramenta poderosa para recuperar a função original no domínio do tempo a partir de sua transformada de Laplace, dada pela seguinte fórmula

$$f(t) = \mathcal{L}^{-1}[\tilde{F}(s)] = \begin{cases} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{st} \tilde{F}(s) ds, & \text{com } t > 0, c \in \mathbb{R} \\ 0, & \text{com } t < 0, \end{cases} \quad (\text{B.3})$$

a escolha de c é crítica e deve ser tal que todas as singularidades do integrando estejam à sua esquerda, ou seja, $Re(s) > c$, isso garante que a integral convirja corretamente. Se o integrando possui pontos de ramificação (pontos singulares nos quais a função se comporta de maneira singular, como raízes múltiplas), podemos precisar usar o chamado “contorno de Bromwich modificado”. Nesse caso, o caminho de integração é adaptado para evitar os pontos de ramificação. Essa modificação do caminho de integração é necessária para calcular a inversa da Transformada de Laplace de funções com pontos de ramificação. O método da integral de Bromwich é especialmente útil quando a inversão direta da Transformada de Laplace usando tabelas de transformadas de Laplace inversas não é possível ou conveniente.

Tabela 2 – Exemplos de Funções e Suas Transformadas de Laplace.

Funções, $f(t)$, $g(t)$	Propriedades das Transformadas de Laplace, $\tilde{F}(s)$, $\tilde{G}(s)$
$f(t) = 1$	$\mathcal{L}[1] = \int_0^{\infty} 1 \cdot e^{-st} dt = \lim_{t \rightarrow \infty} \left(\frac{1 - e^{-st}}{s} \right) = \frac{1}{s}$, com $s > 0$
$f(t) = e^{-at}$	$\mathcal{L}[e^{-at}] = \int_0^{\infty} e^{-(a+s)t} dt = \lim_{t \rightarrow \infty} \left(\frac{1 - e^{-(s+a)t}}{s+a} \right) = \frac{1}{s+a}$, com $s+a > 0$
$f(t) = t^n$	$\mathcal{L}[t^n] = \frac{\Gamma(n+1)}{s^{n+1}} \Rightarrow \mathcal{L}^{-1} \left[\frac{\Gamma(n+1)}{s^{n+1}} \right] = t^n$, com $s > 0$, $n > -1$
$f(t) = \delta(t-a)$	$\mathcal{L}[\delta(t-a)] = e^{-as}$, com $a > 0$
$f(t) + g(t)$	$\mathcal{L}[f(t) + g(t)] = \mathcal{L}[f(t)] + \mathcal{L}[g(t)] \Rightarrow$ $\mathcal{L}^{-1}[\tilde{F}(s) + \tilde{G}(s)] = \mathcal{L}^{-1}[\tilde{F}(s)] + \mathcal{L}^{-1}[\tilde{G}(s)]$
$kf(t)$	$\mathcal{L}[kf(t)] = k\mathcal{L}[f(t)] \Rightarrow$ $\mathcal{L}^{-1}[k\tilde{F}(s)] = k\mathcal{L}^{-1}[\tilde{F}(s)]$, onde k é um escalar
$\frac{f(t)}{t}$	$\mathcal{L}^{-1} \left[\int_s^{\infty} \tilde{F}(x) dx \right] = \frac{f(t)}{t}$, onde $s > 0$

Fonte: Elaborado pelo autor.

B.2 Convolução e Transformadas de Laplace

A multiplicação das transformadas de Laplace não equivale à transformada de Laplace do produto de funções. No entanto, introduzimos a seguir o conceito de convolução, que é uma operação conveniente, na qual essa propriedade se mantém verdadeira. Em outras palavras, a transformada do produto de convolução é igual ao produto das transformadas individuais. Veremos adiante que essa propriedade é de extrema importância no cálculo das transformadas de Laplace de integrais fracionárias (Camargo, 2009; Sodr , 2003). Definindo a convol o de $f(t)$ e $g(t)$, denotada como $(f * g)(t)$ dada pela seguinte f rmula

$$(f * g)(t) = \int_0^t f(t')g(t-t')dt' = \int_0^t f(t-t')g(t')dt', \quad (\text{B.4})$$

em seguida, calculando a transformada de Laplace da integral de convolução, temos

$$\mathcal{L}[(f * g)(t)] = \int_0^\infty e^{-st} dt \int_0^t f(t')g(t-t')dt', \quad (\text{B.5})$$

realizando uma mudança de variável, $t - t' = \tau'$, podemos reescrever como

$$\mathcal{L}[(f * g)(t)] = \int_{-t'}^\infty d\tau' \int_0^t e^{-s(\tau'+t')} f(t')g(\tau')d\tau', \quad (\text{B.6})$$

denotando $g(\tau') = 0$ para $\tau' < 0$, e o limite superior $t \rightarrow \infty$, então

$$\mathcal{L}[(f * g)(t)] = \int_0^\infty e^{-s\tau'} g(\tau')d\tau' \int_0^\infty e^{-st'} f(t')dt' = \tilde{F}(s)\tilde{G}(s), \quad (\text{B.7})$$

a transformada de Laplace do produto de convolução é igual ao produto das transformadas.

Em particular, se $g(t) = u(t) = 1$, então $\tilde{G}(s) = \frac{1}{s}$, com $s > 0$, temos

$$\mathcal{L}[(f * u)(t)] = \frac{\tilde{F}(s)}{s}, \quad (\text{B.8})$$

como $u(t') = 1$ para $t' > 0$, logo

$$\begin{aligned} \frac{\tilde{F}(s)}{s} &= \mathcal{L}[(f * u)(t)] = \mathcal{L} \left[\int_0^t f(t-t')u(t')dt' \right] = \\ &= \mathcal{L} \left[\int_0^t f(t-t')dt' \right] = \mathcal{L} \left[\int_0^t f(t')dt' \right], \end{aligned} \quad (\text{B.9})$$

tomando as transformadas $f = f(t)$, $g(t) = h(t) = u(t) = 1$, respectivamente, são definidas por $F(s)$ e $G(s) = H(s) = \frac{1}{s}$, com $s > 0$, logo

$$\frac{\tilde{F}(s)}{s^2} = \mathcal{L} \left[\int_0^t \left(\int_0^{t'} f(v)dv \right) dt' \right]. \quad (\text{B.10})$$

B.3 Transformada de Laplace da derivada de uma função

Sejam $f(t)$ uma função contínua nos reais e $f'(t)$ uma função contínua nos reais por partes para $t > 0$ e não negativas, logo

$$\mathcal{L}[f'(t)] = s\mathcal{L}[f(t)] - f(0) = s\tilde{F}(s) - f(0), \quad (\text{B.11})$$

podemos demonstrar por meio da integração por partes dada

$$\begin{aligned}
\mathcal{L}[f'(t)] &= \int_0^{\infty} e^{-st} f'(t) dt \\
&= e^{-st} f(t) \Big|_0^{\infty} - \int_0^{\infty} (-se^{-st}) f(t) dt \\
&= -f(0) + s \int_0^{\infty} e^{-st} f(t) dt \\
&= -f(0) + s\mathcal{L}[f(t)] \\
&= s\tilde{F}(s) - f(0),
\end{aligned} \tag{B.12}$$

supondo as mesmas condições para $f''(t)$ seja contínua por partes nos reais e não negativa para $t > 0$, temos

$$\begin{aligned}
\mathcal{L}[f''(t)] &= s\mathcal{L}[f'(t)] - f'(0) \\
&= s(s\mathcal{L}[f(t)] - f(0)) - f'(0) \\
&= s^2\mathcal{L}[f(t)] - sf(0) - f'(0) \\
&= s^2\tilde{F}(s) - sf(0) - f'(0),
\end{aligned} \tag{B.13}$$

realizando de forma análoga, caso $f(t), f'(t), f''(t), \dots, f^{n-1}(t)$ sejam contínuas e $f^n(t)$ seja contínua por partes, temos

$$\begin{aligned}
\mathcal{L}[f^n(t)] &= \mathcal{L}\left[\frac{d^n f(t)}{dt^n}\right] = s^n\tilde{F}(s) - s^{n-1}f(0) - \\
&\quad s^{n-2}f'(0) - \dots - s^1 f^{n-2}(0) - s^0 f^{n-1}(0),
\end{aligned} \tag{B.14}$$

onde $n > 0$ é um número inteiro. Com isso, podemos reescrever a expressão B.14 substituindo os termos $f(t) = \varepsilon(t)$ e $\tilde{F}(s) = \tilde{\varepsilon}(s)$, além da simplificação da fórmula, obtemos

$$\mathcal{L}[\varepsilon^n(t)] = \mathcal{L}\left[\frac{d^n \varepsilon(t)}{dt^n}\right] = s^n\tilde{\varepsilon}(s) - \sum_{k=0}^{n-1} s^{n-k-1} \varepsilon^k(0). \tag{B.15}$$

B.4 Derivadas de Transformadas de Laplace

Derivando ambos os membros da igualdade da expressão B.1 em relação a s , temos

$$\frac{d\tilde{F}(s)}{ds} = \int_0^{\infty} (-t)f(t)e^{-st} dt, \tag{B.16}$$

pode ser reescrito como

$$\frac{d\tilde{F}(s)}{ds} = \mathcal{L}[(-t)f(t)], \tag{B.17}$$

realizando de forma análoga as derivadas sucessivas de $\tilde{F}(s)$, obtemos

$$(-1)^n \frac{d^n \tilde{F}(s)}{ds^n} = \mathcal{L}[(t^n)f(t)]. \quad (\text{B.18})$$

APÊNDICE C – ALGORITMO DO FIT POR MEIO DO MÉTODO DE OTIMIZAÇÃO *L-BFGS-B*

Problemas de otimização envolvem a maximização ou minimização de funções em um domínio de dados específico. Diversos métodos são usados para lidar com esses problemas. Um deles é o *Newton's Method*, que atualiza iterativamente as variáveis da seguinte maneira

$$x_{1+k} = x_k - (\nabla^2 f_k)^{-1} \nabla f_k, \quad k = 0, 1, 2, \dots \quad (\text{C.1})$$

onde x_k é o vetor de variáveis na iteração k , ∇f_k é um vetor gradiente na iteração k e $\nabla^2 f_k$ é matriz Hessiana na iteração k . Ela descreve as segundas derivadas parciais da função objetivo em relação às variáveis. A matriz Hessiana é importante porque representa a curvatura da função e é usada para determinar a direção de descida.

Este método é aplicado à resolução de problemas de minimização não linear sem restrições, em que tentamos encontrar x que minimize $f(x)$, onde $x \in \mathbb{R}^n$. Contudo, esse método apresenta desvantagens, uma vez que requer o cálculo analítico da segunda derivada e a inversão do Hessiano em cada iteração e requer mais informações sobre a função do que apenas o gradiente, tornando-o menos geral em comparação com os *Quasi-Newton Methods*.

Os *Quasi-Newton Methods*, como o *BFGS*, não requerem o cálculo direto da matriz Hessiana B_k ou a matriz inversa do Hessiano H_k , onde $[B_k]^{-1} = [\nabla^2 f_k]^{-1} = H_k$, temos

$$B_k = \nabla^2 f_k = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_k \partial x_1} & \frac{\partial^2 f}{\partial x_k \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_k^2} \end{bmatrix}. \quad (\text{C.2})$$

Em vez disso, eles aproximam as matrizes usando informações sobre as iterações anteriores, juntamente com os gradientes $g_k = \nabla f_k$, tornando-os mais eficientes. Dado x_0 e aproximação hessiana inicial B_0 (aproximação hessiana inversa H_0), $k \rightarrow 0$. Para $k = 0, 1, 2, \dots$, temos

$$x_{k+1} - x_k \rightarrow s_k \quad \text{e} \quad (\text{C.3})$$

$$g_{k+1} - g_k \rightarrow y_k, \quad (\text{C.4})$$

onde s_k e y_k são vetores de atualização. Dessa forma, os valores de B_{k+1} e H_{k+1} são atualizados de acordo com os *Quasi-Newton Methods* (Griffin, 2012; Ding *et al.*, 2004).

O *BFGS* (Broyden-Fletcher-Goldfarb-Shanno) é um dos *Quasi-Newton method* mais populares e amplamente utilizados em otimização numérica, reconhecido por sua eficiência e pela capacidade de autocorreção das atualizações das inversas dos Hessianos H_{k+1} . Ele é considerado rápido e eficaz em comparação com outros *Quasi-Newton Methods*, como o *DFP* (Davidon–Fletcher–Powell) (Powell, 1986).

C.1 *L-BFGS-B*

L-BFGS-B é uma variante do *Quasi-Newton method*, projetada para resolver problemas de otimização não linear sujeitos a limites nas variáveis, economizando memória e recursos computacionais. O método é conhecido por sua boa capacidade de convergência, especialmente em problemas de otimização não linear e é particularmente útil quando a matriz Hessiana é difícil de obter ou para problemas complexos. Este método pode ser aplicado a problemas com limites nas variáveis para que possamos minimizar a função não linear $\min[f(x)]$, representados por i e s , indicando os limites inferior e superior das variáveis dentro do intervalo $[i \leq x \leq s]$ (Byrd *et al.*, 1995; Zhu *et al.*, 1997). Vale mencionar que nem todas as variáveis precisam ter limites definidos, tornando o algoritmo adequado para problemas irrestritos. Uma característica distintiva do *L-BFGS-B* é que ele não exige o conhecimento prévio da matriz Hessiana da função, sendo necessário apenas fornecer o gradiente. Isso faz com que o algoritmo seja apropriado para a resolução de problemas de otimização complexos.

O algoritmo *L-BFGS-B* representa uma extensão do método conhecido como *BFGS*, que se concentra em problemas de otimização sem restrições (Liu; Nocedal, 1989). Uma das melhorias mais notáveis do *L-BFGS-B* é sua capacidade de lidar com problemas de otimização que têm limites ou restrições nas variáveis de decisão, garantindo que as soluções encontradas estejam dentro desses limites. Este método de otimização, o *L-BFGS-B*, tem como objetivo a minimização da função $\min_H(\|H - H_k\|)$, enquanto atende às condições impostas pelas equações $H y_k = s_k$ e $H = H^T$, para garantir a simetria e a positividade definida das matrizes, é importante assegurar que a condição de curvatura seja atendida (Ding *et al.*, 2004; Oliveira, 2009), o que implica na seguinte relação

$$s_k^T y_k > 0. \quad (\text{C.5})$$

Ao descrever o método *L-BFGS-B*, é relevante destacar suas origens no método *BFGS*. Os pontos a cada iteração que serão denotadas por x_k , e define-se $s_k = x_{k+1} - x_k$ e $y_k = g_{k+1} - g_k$. Cada iteração do método *BFGS* segue o formato

$$x_{k+1} = x_k - \alpha_k H_k \nabla f_k, \quad k = 0, 1, 2, \dots, \quad (\text{C.6})$$

onde α_k denota o comprimento do passo de descida, e H_k é atualizado a cada iteração usando a fórmula

$$H_{k+1} = V_k^T H_k V_k + \rho_k s_k s_k^T, \quad (\text{C.7})$$

$$\rho_k = \frac{1}{y_k^T s_k}, \quad V_k = I - \rho_k y_k s_k^T, \quad V_k^T = I - \rho_k s_k y_k^T, \quad (\text{C.8})$$

onde ρ_k é um escalar, I é a matriz identidade, a matriz H_{k+1} é obtida atualizando H_k com base no par (s_k, y_k) .

No contexto da análise de curvas de força resultantes das simulações, o *L-BFGS-B* é utilizado para ajustar modelos analíticos aos dados da simulação. Isso é feito minimizando uma função de perda χ que mede a discrepância entre os dados da simulação (Ys) e os dados do modelo analítico ajustado (Ya)

$$\chi = \frac{1}{M} \sum_{i=1}^M (Ys_i - Ya_i)^2, \quad (\text{C.9})$$

onde M representa o número de dados, o valor de χ indica quão bem o modelo analítico se ajusta aos dados da simulação, sendo que valores menores indicam um ajuste mais próximo entre os pontos da simulação e a curva de ajuste. Portanto, aplicamos o método de otimização *L-BFGS-B* na região *dwell* nas curvas de força numéricas, para determinar o ajuste ideal por meio da função de perda χ .

APÊNDICE D – MÉTODO DE INFERÊNCIA BAYESIANA PARA ESTIMAR EXPOENTES DE ESCALA

Para quantificar como os indicadores econômicos urbanos — compras, vendas e comércio interno (intra-cidade) — escalam com o tamanho populacional, adotamos uma formulação bayesiana da relação padrão de escalonamento urbano. Para o município c ,

$$S_c = A \cdot (\text{população}_c)^\beta, \quad (\text{D.1})$$

em que S_c é o indicador econômico da cidade c , população_c é sua população, A é uma constante e β é o expoente alométrico de interesse. Tomando logaritmos, obtemos uma especificação linear,

$$\log S_c = \log A + \beta \cdot \log (\text{população}_c) + \varepsilon_c, \quad (\text{D.2})$$

onde ε_c captura desvios em relação à lei de escalonamento. Essa formulação permite estimativas robustas de β e é bem adequada a modelos bayesianos.

No paradigma bayesiano, os parâmetros são variáveis aleatórias com distribuições a priori que codificam crenças iniciais (Salvatier *et al.*, 2016; Gelman, 2006). Utilizamos priors fracamente informativas para que os dados conduzam principalmente a inferência, evitando ao mesmo tempo estimativas extremas: o intercepto $\rho = \log A$ e o expoente β seguem

$$\rho \sim \mathcal{N}(0, 10), \quad (\text{D.3})$$

$$\beta \sim \mathcal{N}(0, 10). \quad (\text{D.4})$$

A variância residual é imposta como positiva por meio de um prior half-normal,

$$\sigma \sim \text{HalfNormal}(1), \quad (\text{D.5})$$

o que fornece flexibilidade ao modelo ao mesmo tempo em que garante a coerência dos parâmetros.

A média condicional na escala logarítmica é

$$\mu_c = \rho + \beta \cdot \log (\text{população}_c), \quad (\text{D.6})$$

e a verossimilhança é

$$\log S_c \sim \mathcal{N}(\mu_c, \sigma). \quad (\text{D.7})$$

Essa formulação conecta os valores observados aos parâmetros latentes por meio da verossimilhança, permitindo a inferência a posteriori para ρ , β e σ .

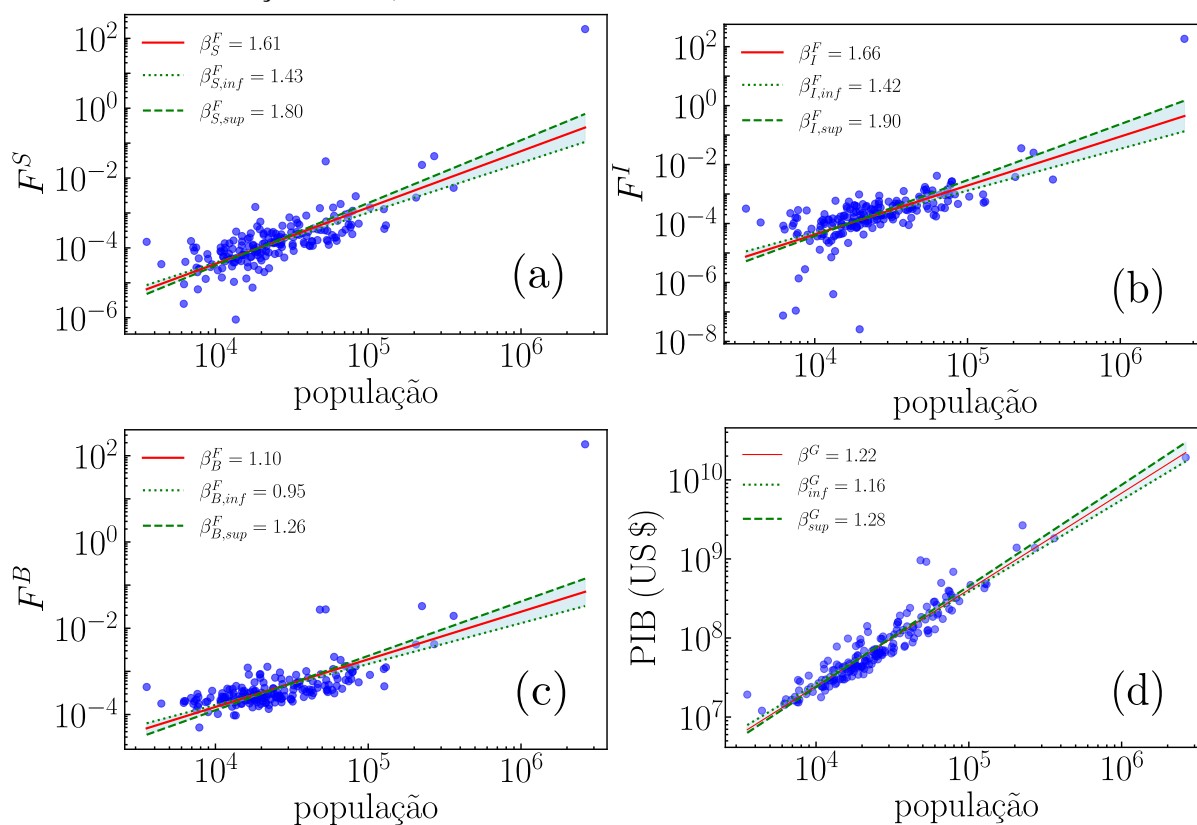
A amostragem a posteriori é realizada por meio de Cadeias de Markov via Monte Carlo (MCMC), utilizando quatro cadeias independentes, cada uma executada por 6.000 iterações; as primeiras 3.000 iterações são descartadas como burn-in, resultando em 3.000 amostras a posteriori. Para promover estabilidade numérica e convergência, definimos a taxa alvo de aceitação em 0.95, fixamos a semente aleatória (42) e paralelizamos a execução em quatro núcleos.

As amostras a posteriori resultantes caracterizam a incerteza em todos os parâmetros e são utilizadas para calcular intervalos credíveis de 95%, visualizar distribuições a posteriori e avaliar diagnósticos de convergência.

Essa abordagem bayesiana fornece estimativas robustas e interpretáveis dos expoentes de escalonamento urbano, permitindo uma análise fundamentada de como a atividade econômica se concentra com a população nos municípios do Ceará. O mesmo arcabouço se estende naturalmente para modelos hierárquicos capazes de acomodar heterogeneidades regionais ou dinâmicas temporais.

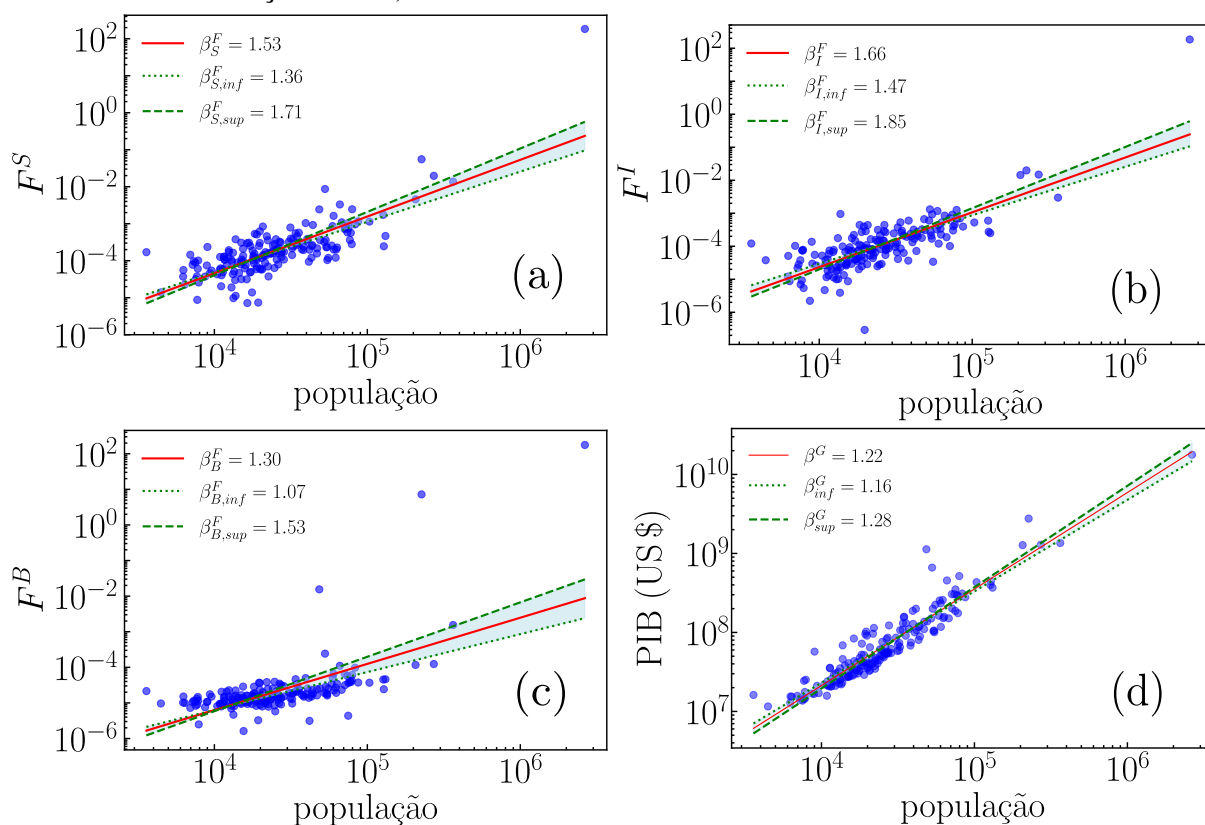
**APÊNDICE E – RESULTADOS DAS ALOMETRIAS DO FITNESS E PIB EM
RELAÇÃO A POLULAÇÃO PARA OS ANOS (2017-2019)**

Figura 44 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2017 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 1.61$, $\beta_I^F = 1.66$ e $\beta_B^F = 1.10$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.22$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%).



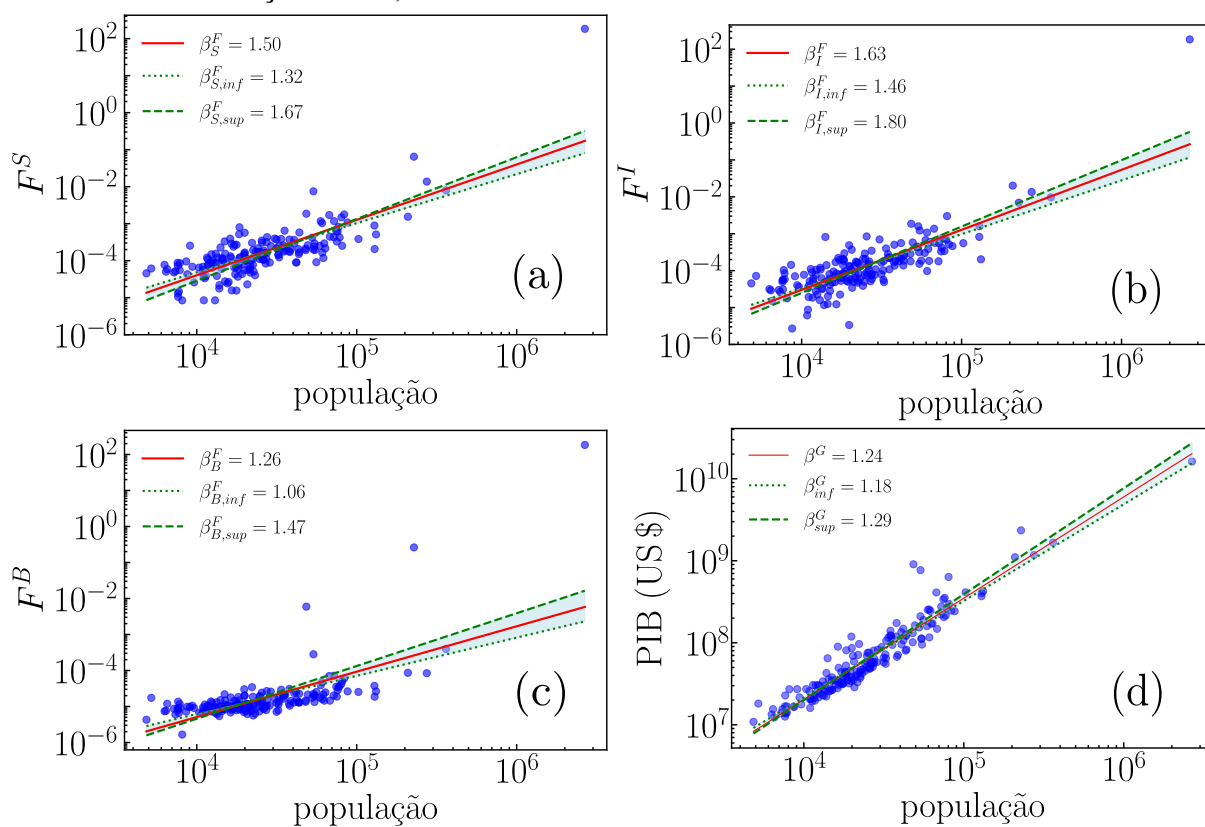
Fonte: Elaborado pelo autor.

Figura 45 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2018 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 1.53$, $\beta_I^F = 1.66$ e $\beta_B^F = 1.30$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.22$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%).



Fonte: Elaborado pelo autor.

Figura 46 – Escalonamento do Fitness com população e renda. (a–c) Fitness municipal versus a população em 2019 nas três camadas de comercio: (a) vendas intermunicipais, (b) transações intramunicipais (internas) e (c) compras intermunicipais. Os expoentes de escalonamento β^F são estimados por regressão bayesiana em escala log–log (linhas vermelhas), com intervalos de confiança de 95% indicados pelas linhas pontilhadas (inferior) e tracejadas (superior). O Fitness escala de forma superlinear com a população em todas as categorias, com $\beta_S^F = 1.50$, $\beta_I^F = 1.63$ e $\beta_B^F = 1.26$. (d) O PIB em função da população para o mesmo período fornece uma relação de escalonamento urbano com expoente $\beta^G = 1.24$ (as linhas pontilhada e tracejada indicam o intervalo de confiança de 95%).



Fonte: Elaborado pelo autor.