



**Universidade  
Estadual do Paraná**  
Campus Apucarana

---

JOÃO PEDRO DE SOUZA OLIVO TARDIVO

**ESTUDO EXPERIMENTAL DO MÉTODO NETWORK  
SCALE-UP EM DIFERENTES TOPOLOGIAS DE GRAFOS  
ALEATÓRIOS**

---

APUCARANA-PR

2025



JOÃO PEDRO DE SOUZA OLIVO TARDIVO

**ESTUDO EXPERIMENTAL DO MÉTODO NETWORK  
SCALE-UP EM DIFERENTES TOPOLOGIAS DE GRAFOS  
ALEATÓRIOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual do Paraná para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Nicollas Mocelin Sdroievski

**APUCARANA-PR**

**2025**

Ficha catalográfica elaborada pelo Sistema de Bibliotecas da UNESPAR e  
Núcleo de Tecnologia de Informação da UNESPAR, com Créditos para o ICMC/USP  
e dados fornecidos pelo(a) autor(a).

de Souza Olivo Tardivo, João Pedro  
ESTUDO EXPERIMENTAL DO MÉTODO NETWORK SCALE-UP EM  
DIFERENTES TOPOLOGIAS DE GRAFOS ALEATÓRIOS / João  
Pedro de Souza Olivo Tardivo. -- Apucarana-PR, 2025.  
59 f.

Orientador: Nicollas Mocelin Sdroievski.  
Trabalho de Conclusão de Curso, Ciência da  
Computação - Universidade Estadual do Paraná, 2025.

1. Grafos Aleatórios. 2. Network Scale-Up. 3.  
Simulação Computacional. 4. Análise de Erro. 5.  
Estimadores de Rede.. I - Mocelin Sdroievski,  
Nicollas (orient). II - Título.

JOÃO PEDRO DE SOUZA OLIVO TARDIVO

**ESTUDO EXPERIMENTAL DO MÉTODO NETWORK  
SCALE-UP EM DIFERENTES TOPOLOGIAS DE GRAFOS  
ALEATÓRIOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual do Paraná para obtenção do título de Bacharel em Ciência da Computação.

**BANCA EXAMINADORA**

---

Prof. Dr. Nicollas Mocelin Sdroievski  
Universidade Estadual do Paraná  
Orientador

---

Prof. Dr. Lisandro Rogério Modesto  
Universidade Estadual do Paraná

---

Prof. Kleber Marcio de Souza  
Universidade Estadual do Paraná

Apucarana-PR, 01 de dezembro de 2025



*Dedico este trabalho aos meus pais e ao meu irmão,  
que sempre me deram suporte e estiveram ao meu lado.*





## AGRADECIMENTOS

Agradeço meu orientador, Prof. Dr. Nicollas Mocelin Sdroievski, por todo o suporte oferecido ao longo desta caminhada acadêmica. Sua paciência, conduta exemplar e os apontamentos técnicos sempre pertinentes foram fundamentais para o direcionamento e refinamento desta pesquisa.

Ao coordenador do Trabalho de Conclusão de Curso, Prof. Dr. José Luis Seixas Junior, expresso minha gratidão pela disponibilização e manutenção do modelo estrutural deste trabalho, bem como pelas valiosas observações quanto à organização do texto.

Por fim, estendo meu reconhecimento aos pesquisadores Sergio Díaz-Aranda *et al.*, autores do estudo “*Error Bounds for the Network Scale-Up Method*”. A fundamentação teórica e os resultados apresentados em seu trabalho serviram como alicerce indispensável para a reprodução, simulação e análise crítica desenvolvidas neste TCC.



*“Todos os modelos estão errados, mas alguns são úteis.”*  
*(George E. P. Box)*



TARDIVO, JPSO. **Estudo Experimental do Método Network Scale-Up em Diferentes Topologias de Grafos Aleatórios**. 57 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual do Paraná, Apucarana–PR, 2025.

## RESUMO

A estimação do tamanho de populações de difícil acesso é um desafio central em diversas áreas, para o qual o Método Network Scale-Up (NSUM) se apresenta como uma ferramenta proeminente. A acurácia deste método, contudo, depende da escolha entre seus estimadores e da topologia da rede social subjacente. Este trabalho conduz um estudo experimental comparativo para avaliar a robustez e a precisão dos estimadores Média das Razões (MoR) e Razão das Somas (RoS) do NSUM sob diferentes topologias. Por meio de simulação computacional, o desempenho dos estimadores foi avaliado em três modelos de grafos aleatórios: Erdős-Rényi, Barabási-Albert e Watts-Strogatz. Os resultados demonstram que em redes com distribuição de grau homogênea como Erdős-Rényi e Watts-Strogatz, os estimadores MoR e RoS apresentam desempenho virtualmente idêntico. Contudo, em redes do tipo *scale-free* como Barabási-Albert, caracterizadas pela heterogeneidade de grau, o estimador RoS revela uma superioridade, apresentando menor risco de erro significativo e maior confiabilidade. Conclui-se que a heterogeneidade da distribuição de grau é o fator estrutural determinante para a escolha do estimador, com o RoS sendo a opção mais robusta em cenários que emulam redes sociais realistas com a presença de *hubs*.

**Palavras-chave:** Network Scale-Up. Grafos Aleatórios. Estimadores de Rede. Análise de Erro. Simulação Computacional.



TARDIVO, JPSO. **Experimental Study of the Network Scale-Up Method in Different Random Graph Topologies**. 57 p. Final Project (Bachelor of Science in Computer Science) – State University of Paraná, Apucarana-PR, 2025.

## ABSTRACT

Estimating the size of hard-to-reach populations is a key challenge across several fields, for which the Network Scale-Up Method (NSUM) stands out as a prominent tool. The accuracy of this method, however, depends on the choice of its estimators and the underlying social network topology. This work conducts a comparative experimental study to assess the robustness and precision of NSUM's Mean of Ratios (MoR) and Ratio of Sums (RoS) estimators under different topologies. Through computational simulation, the performance of the estimators was evaluated on three random graph models: Erdős-Rényi, Barabási-Albert, and Watts-Strogatz. The results demonstrate that in networks with a homogeneous degree distribution such as Erdős-Rényi and Watts-Strogatz, the MoR and RoS estimators exhibit virtually identical performance. However, in scale-free networks such as Barabási-Albert, characterized by degree heterogeneity, the RoS estimator reveals a superior performance, showing a lower risk of significant error and greater reliability. It is concluded that the heterogeneity of the degree distribution is the decisive structural factor for estimator selection, with RoS being the more robust option in scenarios that emulate realistic social networks with the presence of hubs.

**Keywords:** Network Scale-Up. Random Graphs. Network Estimators. Error Analysis. Computational Simulation.





## LISTA DE ILUSTRAÇÕES

Figura 1 – Erro médio dos estimadores MoR e RoS em função do tamanho da amostra para redes Erdős-Rényi. . . . .	43
Figura 2 – Distribuição do erro dos estimadores MoR e RoS para redes Erdős-Rényi, mostrando mediana, quartis e <i>outliers</i> . . . . .	44
Figura 3 – Probabilidade de o erro do estimador exceder 5% em função do tamanho da amostra para redes Erdős-Rényi. . . . .	45
Figura 4 – Análise de viés dos estimadores MoR e RoS para redes Barabási-Albert. A figura é organizada por tamanho de rede (colunas) e tipo de estimador (linhas). . . . .	46
Figura 5 – Erro médio dos estimadores MoR e RoS em função do tamanho da amostra para redes Barabási-Albert, facetado pelo tipo de estimador (linhas) e tamanho da rede (colunas). . . . .	47
Figura 6 – Probabilidade de o erro do estimador exceder 5% em função do tamanho da amostra para redes Barabási-Albert. . . . .	48
Figura 7 – Análise de viés dos estimadores MoR (linhas contínuas) e RoS (linhas tracejadas) para redes Watts-Strogatz. . . . .	49
Figura 8 – Erro médio dos estimadores MoR e RoS em função do tamanho da amostra para redes Watts-Strogatz. . . . .	50



## LISTA DE TABELAS

Tabela 1	–	Descrição formal das etapas de simulação e análise. . . . .	33
Tabela 2	–	Parâmetros para os experimentos com redes de Erdős-Rényi. . . . .	38
Tabela 3	–	Parâmetros para os experimentos com redes de Barabási-Albert. . . . .	40
Tabela 4	–	Parâmetros para os experimentos com redes de Watts-Strogatz. . . . .	41



## LISTA DE ABREVIATURAS E SIGLAS

NSUM	Network Scale-Up Method
ARD	Aggregated Relational Data
MoR	Mean of Ratios
RoS	Ratio of Sums
CSV	Comma-Separated Values



# SUMÁRIO

1	INTRODUÇÃO . . . . .	23
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	25
2.1	O Método Network Scale-Up (NSUM) . . . . .	25
2.2	Erdős–Rényi . . . . .	28
2.3	Barabási–Albert . . . . .	30
2.4	Watts–Strogatz . . . . .	31
3	MÉTODO DE PESQUISA . . . . .	33
3.1	Métricas de Avaliação . . . . .	34
4	EXPERIMENTOS . . . . .	37
4.1	Redes de Erdős–Rényi . . . . .	38
4.2	Redes de Barabási–Albert . . . . .	39
4.3	Redes de Watts–Strogatz . . . . .	40
5	RESULTADOS . . . . .	43
5.1	Rede Erdős–Rényi . . . . .	43
5.2	Rede Barabási–Albert . . . . .	45
5.3	Rede Watts–Strogatz . . . . .	48
6	CONCLUSÃO . . . . .	53
	REFERÊNCIAS . . . . .	55





# 1 INTRODUÇÃO

A estimação do tamanho de populações de difícil acesso, como grupos afetados por desastres naturais, portadores de doenças estigmatizadas ou membros de redes clandestinas, representa um desafio fundamental para a epidemiologia, as ciências sociais e a formulação de políticas públicas [1]. Métodos de enumeração direta são frequentemente inviáveis devido à natureza dispersa ou sigilosa desses grupos. Nesse contexto, a análise de redes sociais emerge como uma abordagem poderosa, permitindo inferir características de uma população inteira a partir das conexões locais de um subconjunto de seus membros [2].

Uma das técnicas mais proeminentes para essa finalidade é o Método Network Scale-Up (NSUM), uma abordagem de estimação indireta que utiliza Dados Relacionais Agregados, do inglês *Aggregated Relational Data* (ARD) [3]. O método baseia-se em inquirir uma amostra de indivíduos sobre o tamanho de suas redes pessoais e quantos membros da população-alvo eles conhecem. A partir desses dados, dois estimadores são predominantemente utilizados para calcular a prevalência da população oculta, a Média das Razões, do inglês *Mean of Ratios* (MoR), que pondera cada respondente igualmente, e a Razão das Somas, do inglês *Ratio of Sums* (RoS), que implicitamente dá mais peso aos indivíduos mais conectados [4].

A precisão do NSUM, contudo, não é absoluta e sua performance é intrinsecamente dependente da topologia subjacente da rede social [5]. A escolha entre os estimadores MoR e RoS não é trivial, e trabalhos teóricos sugerem que a presença de heterogeneidade na distribuição de conectividade, como a existência de *hubs*, vértices altamente conectados, pode introduzir vieses e afetar a robustez de cada estimador de maneira distinta [6, 4]. Enquanto limitantes de erro analíticos fornecem uma base teórica, a validação empírica de seu comportamento em diferentes estruturas de rede é crucial para compreender sua aplicabilidade prática.

O objetivo central deste trabalho é, portanto, conduzir um estudo experimental e comparativo para avaliar a acurácia e a robustez dos estimadores MoR e RoS sob diferentes topologias de grafos aleatórios. Para isso, foi desenvolvida uma metodologia de pesquisa baseada em simulação computacional que espelha o protocolo experimental de Díaz-Aranda et al. [4], garantindo a comparabilidade direta dos resultados, o desempenho dos estimadores será sistematicamente medido em três modelos de rede canônicos, cada um representando uma propriedade estrutural distinta: o modelo de Erdős-Rényi, como linha de base de uma rede homogênea e aleatória; o modelo de Barabási-Albert, para investigar o impacto da heterogeneidade de grau em redes *scale-free*; e o modelo de Watts-Strogatz, para analisar o efeito da alta clusterização local característica de redes de "mundo

pequeno".

Este trabalho está estruturado da seguinte forma. O Capítulo 2 apresenta a fundamentação teórica, detalhando o método NSUM, seus estimadores e os três modelos de grafos aleatórios utilizados como bancada de testes. O Capítulo 3 descreve a metodologia de pesquisa, incluindo a organização da simulação e as métricas de avaliação de desempenho. No Capítulo 4, são apresentados os parâmetros específicos e a configuração dos experimentos para cada topologia de rede. Os resultados quantitativos obtidos nas simulações são detalhados e analisados no Capítulo 5. Por fim, o Capítulo 6 sintetiza as conclusões do estudo, discute as implicações dos resultados e aponta direções para trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

As redes sociais podem ser modeladas formalmente como grafos [7, 8], onde os indivíduos são representados por vértices e as relações entre eles por arestas. Essa abstração permite o uso de ferramentas matemáticas e computacionais para analisar padrões de conexão, difusão de informação, e estrutura relacional em diferentes contextos [9]. No caso de redes sociais humanas, as arestas geralmente representam interações como amizade, comunicação, colaboração ou influência, facilitando estudos sobre dinâmica social e comportamental [2].

O presente trabalho foca na análise de um método de estimação populacional que opera sobre essas redes, tomando como referência central o arcabouço teórico e experimental estabelecido recentemente por Díaz-Aranda et al. [4]. Para validar os limitantes analíticos propostos pelos autores e avaliar a robustez do método, é fundamental testá-lo em ambientes controlados que simulam diferentes estruturas de rede observadas no mundo real. Esta fundamentação teórica, portanto, está organizada da seguinte forma: primeiro, apresenta-se o método central de estudo, o NSUM e seus estimadores MoR e RoS, conforme as definições do trabalho base; em seguida, justifica-se a necessidade de modelos de grafos aleatórios para a análise de erro; por fim, detalham-se os três modelos de rede sintética que servirão como bancadas de teste para a avaliação empírica.

### 2.1 O Método Network Scale-Up (NSUM)

O NSUM, emerge como uma abordagem indireta para estimar o tamanho de populações de difícil acesso [1, 10] em cenários onde a enumeração direta através de listas administrativas ou inquéritos convencionais é inviável, seja devido à raridade do grupo, estigmatização associada, ou ausência de uma amostragem completa, o NSUM oferece uma alternativa metodológica. A técnica explora a estrutura inerente das redes sociais dos indivíduos para inferir características populacionais agregadas [11].

O princípio fundamental do NSUM reside na utilização de ARD [3]. Estes dados são coletados por meio de perguntas diretas aos respondentes de um inquérito amostral, tipicamente na forma "Quantos X você conhece?", onde X representa membros da população de interesse. Essencialmente, o método assume que a rede pessoal de um respondente reflete, em certa medida, a composição da população geral.

A formulação básica [12] do estimador NSUM requer duas informações principais de cada respondente  $i$ , o número de pessoas que ele conhece pertencentes ao grupo de interesse  $k$ , denotado por  $y_{ik}$ , e o tamanho total de sua rede pessoal, ou seu grau, denotado por  $d_i$ . O método então calcula [13] a proporção de membros do grupo  $k$  na população

geral  $p_k = N_k/N$ , onde  $N_k$  é o tamanho desconhecido do grupo e  $N$  o tamanho total da população estimando essa proporção  $\hat{p}_k$  através da razão entre o número total de indivíduos do grupo  $k$  conhecidos por todos os respondentes na amostra e o tamanho total das redes de todos os respondentes.

A estimativa do tamanho da população  $N_k$  é então obtida multiplicando-se essa proporção estimada pelo tamanho total conhecido da população  $N$ , resultando em  $\hat{N}_k = \hat{p}_k \times N$ . Este estimador básico, contudo, fundamenta-se em pressupostos significativos, como a mistura aleatória na rede [5], com ausência de homofilia ou barreiras sociais sistemáticas, a transmissão perfeita de informação, onde o respondente sabe que o contato pertence ao grupo  $k$ , a ausência de erros de recordação por parte do respondente, e a acurácia na estimação do grau  $d_i$ .

O trabalho [4] propõe um modelo geral para a construção de redes aleatórias, que serve como um ambiente controlado para o estudo de erros. Nesse modelo, a rede  $G$  sobre uma população  $V$  é gerada através de um processo estocástico independente para cada vértice  $v \in V$ , em duas etapas, na primeira, o vértice  $v$  sorteia um grau  $R_v$  de uma distribuição de graus predefinida  $P_{deg}$ , em segundo lugar, dado  $R_v$ , o vértice seleciona  $R_v$  vizinhos de forma uniforme e aleatória do conjunto de todos os outros vértices possíveis  $V \setminus \{v\}$ .

Neste *framework*, as variáveis de interesse para o NSUM, como a proporção de vizinhos ocultos de cada indivíduo, tornam-se dependentes entre si. Essa dependência impede a aplicação direta de limitantes de concentração padrão, que tipicamente exigem independência das variáveis aleatórias [14]. O desafio analítico, portanto, reside em derivar limitantes de erro válidos mesmo na presença dessa correlação, o que os autores abordam demonstrando correlação negativa e utilizando independência condicional.

Na prática, desvios desses pressupostos podem introduzir vieses consideráveis. A heterogeneidade na propensão dos indivíduos a conhecer membros do grupo-alvo gera efeitos de barreira e superdispersão. Erros de transmissão ocorrem quando o respondente desconhece a afiliação do contato ao grupo, comum em populações estigmatizadas. Erros de recordação sistemáticos, onde o respondente subestima ou superestima o número de conhecidos em grupos grandes ou pequenos, respectivamente, também afetam a precisão [15].

A estimação precisa do grau  $d_i$  de cada respondente é outro desafio central. Diversas técnicas foram propostas, incluindo métodos de enumeração direta, inviáveis em larga escala, métodos de soma por categorias relacionais, e o próprio método *scale-up* aplicado a populações de tamanho conhecido, como pessoas com nomes comuns, para calibrar a estimação do grau [5]. A definição operacional de "conhecer" alguém é crucial e pode impactar significativamente os resultados.

Modelos estatísticos mais sofisticados foram desenvolvidos para mitigar o impacto dessas fontes de erro [16, 6, 17]. Abordagens hierárquicas e Bayesianas buscam modelar explicitamente a superdispersão, corrigir erros de recordação através de curvas de calibração, ajustar para efeitos de barreira considerando a estrutura social, e incorporar incerteza na estimação do grau e nos erros de transmissão. Esses refinamentos tornam o NSUM uma ferramenta mais robusta, embora demandem dados adicionais ou pressupostos mais complexos.

O presente trabalho analisa duas abordagens, a Média das Razões (MoR) e a Razão das Somas (RoS), para calcular a proporção estimada da população oculta,  $\hat{p}_k$  [4]. Embora ambos utilizem os mesmos dados de entrada, o grau do respondente e o número de contatos no grupo-alvo, eles diferem fundamentalmente na maneira como agregam essa informação, o que acarreta implicações diretas na robustez e no viés do resultado final.

O estimador MoR, representa uma abordagem no nível do indivíduo. Para cada respondente  $v$  na amostra  $S$ , calcula-se primeiro a proporção da população oculta dentro de sua rede pessoal. Esta proporção individual,  $Y_v$ , é simplesmente a razão entre o número de seus vizinhos na população oculta  $C_v$  e seu grau total  $R_v$ . A estimativa final é então a média aritmética simples dessas proporções individuais sobre toda a amostra. Formalmente:

$$\hat{p}_{\text{MoR}} = \frac{1}{|S|} \sum_{v \in S} Y_v = \frac{1}{|S|} \sum_{v \in S} \frac{C_v}{R_v} \quad (2.1)$$

A intuição por trás do MoR é que cada respondente fornece uma estimativa independente da prevalência da população oculta, e a melhor estimativa geral é a média dessas "opiniões" [6]. Este método, no entanto, é sensível a respondentes com grau muito baixo,  $R_v$  pequeno, pois suas proporções individuais podem ser extremamente voláteis, e se torna indefinido para indivíduos com grau zero  $R_v = 0$ .

Em contrapartida, o estimador RoS, adota uma abordagem no nível da amostra. Em vez de calcular proporções individuais, o RoS primeiro agrega toda a informação disponível. Ele soma o número de contatos na população oculta reportado por todos os respondentes e divide este total pela soma dos graus de todos os respondentes. Matematicamente, o estimador é definido como:

$$\hat{p}_{\text{RoS}} = \frac{\sum_{v \in S} C_v}{\sum_{v \in S} R_v} \quad (2.2)$$

A lógica do RoS é tratar o conjunto de todas as conexões reportadas pela amostra como um "super grafo" e calcular a prevalência do grupo oculto nesse agregado. Uma consequência direta é que os respondentes com maior grau, os *hubs*, têm uma influência maior no resultado final, pois seus contatos, tanto  $C_v$  quanto  $R_v$ , contribuem com mais

peso para os somatórios [13]. Este comportamento é frequentemente vantajoso em redes heterogêneas, como as Scale-Free, onde a informação dos *hubs* pode ser mais representativa da estrutura geral da rede.

A distinção entre esses dois estimadores é, portanto, fundamental. O MoR pondera cada respondente de forma igual, enquanto o RoS pondera cada aresta, ou conexão social, de forma igual [6]. A escolha entre eles não é trivial e, como demonstrado por análises teóricas e simulações, sua performance relativa depende intrinsecamente da topologia da rede social subjacente [6].

Nessa sentida, a análise de métodos de estimação, como o NSUM, sobre diferentes topologias de rede exige um passo além da descrição estática. Como esses métodos operam sobre amostras da população, a incerteza é um elemento inerente. Para quantificar o desempenho e os erros de um estimador, é necessário tratar a própria estrutura da rede não como fixa, mas como o resultado de um processo aleatório [18]. Essa necessidade motiva a transição de um grafo particular para o conceito de modelos de grafos aleatórios, que definem uma distribuição de probabilidade sobre um conjunto de grafos possíveis [19].

Desta forma, a modelagem de grafos aleatórios é uma ferramenta fundamental para a compreensão de propriedades estruturais de redes complexas [18, 20]. Esses modelos fornecem abstrações matemáticas que permitem a geração sintética de grafos com características estatísticas controladas, sendo amplamente utilizados em experimentos computacionais, análise de algoritmos e estudos de fenômenos emergentes em redes [21]. Entre os modelos mais relevantes destacam-se o modelo de Erdős–Rényi e o modelo de Barabási–Albert, com avanços recentes em suas aplicações para estimação em NSUM [4].

Ademais, é crucial notar que os teoremas que fundamentam os limitantes de erro para esses estimadores frequentemente necessitam que os grafos subjacentes não possuam vértices isolados, ou seja, com grau zero, visto que  $R_v = 0$  tornaria a razão do estimador MoR indefinida. Essa premissa apresenta um desafio específico para certos modelos probabilísticos, como o modelo de Erdős–Rényi, onde a probabilidade de gerar vértices isolados, embora possa ser pequena, não é nula.

## 2.2 Erdős–Rényi

O modelo de Erdős–Rényi [22, 23] se apresenta em duas variantes principais. A primeira,  $G(n, M)$ , considera o conjunto de todos os grafos não direcionados simples com  $n$  vértices e exatamente  $M$  arestas, escolhendo uniformemente um grafo dessa coleção. A segunda, mais comum na literatura contemporânea, é o modelo  $G(n, p)$ , no qual cada uma das  $\binom{n}{2}$  possíveis arestas entre pares de vértices é incluída de forma independente com probabilidade  $p$ . Esta independência torna o modelo particularmente conveniente para análises probabilísticas, como o estudo da conectividade ou da emergência de componentes

gigantes.

Uma vez estabelecido o modelo probabilístico para a rede, a tarefa seguinte é avaliar a confiabilidade de um estimador. O objetivo central é garantir que a probabilidade de o erro da estimativa ultrapassar um limiar aceitável seja pequena. Para derivar tais garantias, uma ferramenta analítica de grande poder é o Limitante da União. Para um conjunto de "eventos ruins"  $A_1, A_2, \dots, A_k$  cada  $A_i$  sendo o evento onde a estimativa em uma amostra  $i$  é imprecisa, o limitante afirma que a probabilidade de que ao menos um deles ocorra é, no máximo, a soma de suas probabilidades individuais:

$$P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(A_i) \quad (2.3)$$

A grande vantagem deste limitante é sua generalidade, pois não requer independência entre os eventos. Ele nos permite calcular um limite superior conservador para a probabilidade de erro, fornecendo um controle rigoroso sobre a performance do método em estudo [24].

Apesar de distintos na definição, os modelos  $G(n, M)$  e  $G(n, p)$  compartilham comportamento assintótico semelhante sob certas condições. Especificamente, quando o número esperado de arestas em  $G(n, p)$ , dado por  $\binom{n}{2}p$ , é próximo de  $M$ , diversas propriedades monótonas, como conectividade ou existência de ciclos hamiltonianos, ocorrem com probabilidade similar à medida que  $n \rightarrow \infty$ . Tal equivalência é formalizada por resultados como os de Luczak [18], que provam convergência de propriedades entre os dois modelos sob restrições assintóticas apropriadas.

Entretanto, uma limitação crítica do modelo Erdős–Rényi reside na sua distribuição de grau. Por se tratar de um processo com inclusão de arestas independente e homogênea, a distribuição de grau dos vértices segue uma distribuição binomial, o que resulta em pouca variabilidade entre os graus dos vértices. Esse comportamento é inconsistente com o que se observa em muitas redes reais, caracterizadas por alta heterogeneidade de grau e presença de vértices altamente conectados [25]. Essa inadequação motiva a adoção de modelos mais realistas, como o modelo de configuração e os baseados em leis de potência, para representar com maior fidelidade as propriedades topológicas de redes naturais e sociais [26].

Adicionalmente, é pertinente notar que a obtenção de limitantes de erro analíticos para o estimador RoS neste modelo envolve desafios teóricos substancialmente maiores em comparação ao MoR. Enquanto o MoR opera com a média de razões individuais, o RoS constitui uma razão de somas de variáveis aleatórias, introduzindo dependências estocásticas complexas entre o numerador e o denominador. Consequentemente, a demonstração de limitantes para o RoS não permite a aplicação direta de desigualdades de concentração padrão, exigindo o emprego de técnicas analíticas mais sofisticadas para tratar a

correlação inerente à estrutura do estimador.

## 2.3 Barabási–Albert

O modelo de Barabási–Albert, introduzido em 1999 por Albert-László Barabási e Réka Albert [27], representa um marco na modelagem de redes complexas ao capturar de forma eficaz a natureza heterogênea da conectividade observada em diversas redes reais. Diferente de modelos clássicos como Erdős–Rényi, que assumem uma probabilidade uniforme e independente de ligação entre pares de vértices, o modelo de Barabási–Albert incorpora um mecanismo dinâmico de ligação preferencial, pelo qual vértices com maior grau têm maior probabilidade de receber novas conexões ao longo da evolução da rede.

A construção da rede inicia-se com um pequeno conjunto conectado de vértices [28]. Em cada passo temporal, um novo vértice é adicionado e conectado a  $m$  vértices já existentes, sendo a probabilidade de ligação com um vértice  $v$  proporcional ao seu grau atual  $k_v$ . Ou seja, a chance de um novo nó se conectar a um nó existente é dada por:

$$P(v) = \frac{k_v}{\sum_{u \in V} k_u} \quad (2.4)$$

Esse processo iterativo gera redes cujas distribuições de grau apresentam o que se chama de lei de potência [29]. Nessa distribuição, a fração de vértices com grau  $k$ , denotada por  $P(k)$ , decai de forma polinomial, isto é,  $P(k)$  é proporcional a  $k^{-c}$ , onde  $c$  é um expoente positivo. Tal comportamento implica que, embora a maioria dos nós possua um grau relativamente baixo, há uma cauda pesada na distribuição que permite a presença de alguns vértices com grau muito elevado. Esse padrão contrasta fortemente com distribuições exponenciais, em que a probabilidade diminui de maneira muito mais rápida conforme  $k$  aumenta.

Estudos empíricos e teóricos indicam que, para muitas redes reais, o expoente costuma estar em torno de 3, ou seja, tipicamente temos:

$$P(k) \sim k^{-3} \quad (2.5)$$

Esse resultado destaca o fenômeno “os ricos ficam mais ricos”, caracterizando redes onde poucos nós acumulam a maior parte das conexões, enquanto a maioria dos nós apresenta graus modestos. Essa propriedade é observada em sistemas como a internet, redes de colaboração científica e redes de interação social, com aplicações modernas em sustentabilidade e inferência de redes [30].

Desta forma, este modelo não segue estritamente o processo de geração em duas etapas do modelo teórico considerado por Díaz-Aranda et al. [4], pois seu mecanismo de



ligação preferencial é dinâmico e não se baseia em uma seleção uniforme. No entanto, a distribuição de grau resultante pela lei de potência pode ser utilizada como a distribuição  $P_{\text{deg}}$  em simulações que buscam emular redes *Scale-Free* sob o arcabouço teórico geral. Ele serve para testar o comportamento dos estimadores em redes heterogêneas, onde a hipótese central é que o estimador RoS, ao ponderar pelos graus, corrigirá o viés introduzido pelos *hubs*.

O modelo de Barabási–Albert oferece uma explicação teórica para a ubiquidade das leis de potência [31] em sistemas complexos e destaca o papel da dinâmica de crescimento em conjunto com a preferência por popularidade como mecanismos formadores de estruturas topológicas emergentes. Ainda que simplificado, ele permanece um dos modelos mais influentes e frequentemente utilizados em estudos de redes, com extensões recentes para modelagem generativa e análise de escalabilidade [32].

## 2.4 Watts–Strogatz

Proposto por Duncan Watts e Steven Strogatz em 1998 [33], o modelo de Watts Strogatz foi desenvolvido para descrever uma classe de redes que se situa entre as redes perfeitamente regulares e as completamente aleatórias. O modelo captura com sucesso uma propriedade ubíqua em muitas redes sociais e biológicas do mundo real, o fenômeno de "mundo pequeno".

Essas redes são caracterizadas simultaneamente por um alto coeficiente de agrupamento ou *clustering* [34], indicando que os amigos de um indivíduo também tendem a ser amigos entre si, e por um baixo comprimento médio do caminho, significando que quaisquer dois indivíduos na rede podem ser alcançados através de uma pequena sequência de intermediários.

O processo de geração de uma rede Watts–Strogatz parte de uma estrutura determinística e introduz aleatoriedade de forma controlada [35]. O algoritmo é definido por três parâmetros, o número de vértices  $n$ , o grau médio inicial  $k$ , e a probabilidade de religação  $p$ . O processo ocorre em duas etapas.

Primeiramente ocorre a construção de uma rede regular, que inicia-se com uma rede em anel, onde os  $n$  vértices são dispostos em um círculo e cada vértice é conectado aos seus  $k/2$  vizinhos mais próximos em cada direção [9]. Essa estrutura inicial é altamente ordenada, com um coeficiente de agrupamento elevado e um comprimento de caminho longo.

Em seguida, ocorre a etapa da religação aleatória de arestas, onde cada aresta da rede é percorrida sequencialmente. Com uma probabilidade  $p$ , uma das extremidades da aresta é desconectada de seu vizinho original e reconectada a um outro vértice da rede, escolhido uniformemente ao acaso, com a restrição de evitar laços, ou seja, arestas de um

vértice para si mesmo, e arestas múltiplas.

O parâmetro  $p$  atua como um botão de sintonia que interpola entre a ordem e a aleatoriedade. Quando  $p = 0$ , nenhuma aresta é religada e a rede permanece uma estrutura regular. Quando  $p = 1$ , todas as arestas são religadas, e a rede resultante se aproxima de um grafo aleatório de Erdős–Rényi [36].

O *insight* fundamental de Watts e Strogatz foi que, para valores intermediários de  $p$ , mesmo muito pequenos, a introdução de algumas poucas "arestas de atalho" de longa distância é suficiente para reduzir drasticamente o comprimento médio do caminho, enquanto a estrutura local, e consequentemente o alto coeficiente de agrupamento, permanece largamente preservada.

No contexto deste trabalho, a inclusão do modelo de Watts–Strogatz é de particular importância e representa um desafio ao arcabouço teórico de Díaz-Aranda et al. [4]. Os teoremas de limitantes de erro apresentados pelos autores fundamentam-se estritamente em um processo gerador onde, dado o grau de um vértice, seus vizinhos são selecionados uniformemente ao acaso dentre toda a população. O modelo de Watts–Strogatz, por sua vez, não satisfaz essa premissa de geração estocástica.

A estrutura de redes "mundo pequeno" emerge de uma topologia inicial regular, onde as conexões possuem fortes correlações locais que são apenas parcialmente desfeitas pelo processo de religação. Como resultado, a hipótese de independência condicional na escolha dos vizinhos é violada. Portanto, os teoremas analíticos que garantem a precisão dos estimadores em grafos aleatórios genéricos não são diretamente aplicáveis a esta classe de redes, uma vez que as premissas matemáticas sobre as quais foram demonstrados não se sustentam nesta topologia específica.

Essa limitação teórica torna a avaliação empírica indispensável. A presença de alta coesão local pode introduzir vieses sistemáticos que não são capturados pelos modelos Erdős–Rényi ou Barabási–Albert. Assim, os experimentos com redes Watts–Strogatz permitirão analisar a robustez dos estimadores MoR e RoS em cenários onde as garantias teóricas formais estão ausentes, aproximando a análise de estruturas sociais reais caracterizadas por transitividade elevada.

### 3 MÉTODO DE PESQUISA

O objetivo central deste trabalho é conduzir uma avaliação empírica rigorosa e multifacetada da acurácia e da robustez dos estimadores MoR e RoS, pertencentes ao NSUM. Para atingir este objetivo, foi desenvolvida uma metodologia de pesquisa fundamentada em simulação computacional.

A escolha por simulação permite a criação de ambientes de teste controlados, nos quais as propriedades estruturais da rede e os parâmetros do processo de amostragem podem ser sistematicamente variados, possibilitando uma análise aprofundada do comportamento dos estimadores. Este capítulo detalha o arcabouço metodológico que sustenta a pesquisa, descrevendo a estratégia de avaliação e as métricas utilizadas para quantificar o desempenho.

A análise em redes sintéticas é estruturada em torno de três modelos generativos, cada um selecionado para testar uma hipótese estrutural distinta. A investigação inicia-se com o modelo de Erdős-Rényi, que estabelece uma linha de base com sua estrutura puramente aleatória e homogênea.

Em seguida, o modelo de Barabási-Albert é utilizado para introduzir a heterogeneidade na distribuição de graus e a presença de hubs, características centrais de redes scale-free. Por fim, o modelo de Watts-Strogatz permite a análise do efeito de alta clusterização local, uma propriedade definidora de redes de "mundo pequeno".

Este fluxo de trabalho completo, que abrange desde a configuração da rede até a análise dos resultados, é executado de forma padronizada sobre cada uma das topologias investigadas. A Tabela 1 descreve formalmente cada etapa deste processo, detalhando sua função e o principal resultado gerado em cada passo.

Tabela 1 – Descrição formal das etapas de simulação e análise.

Passo	Descrição da Etapa	Resultado Principal
1	Configuração da Rede	Grafo ( $G$ ) com topologia definida
2	Definição da Pop. Oculta	Conjunto de vértices ( $H$ ) e $\rho$ real
3	Amostragem de Vértices	Subconjunto de vértices ( $S$ )
4	Coleta de Dados Relacionais	Pares ( $R_v, C_v$ ) para cada $v \in S$
5	Cálculo dos Estimadores	Estimativas $\hat{\rho}_{\text{MoR}}$ e $\hat{\rho}_{\text{RoS}}$
6	Avaliação de Desempenho	Valores das métricas de erro
7	Agregação Estatística	Média, desvio padrão e outras estatísticas

A fase inicial de configuração do cenário estabelece a verdade fundamental contra a qual as estimativas serão comparadas. Esta fase compreende duas etapas críticas. A primeira é a geração ou carregamento da estrutura da rede, que serve como o universo social

da simulação. Esta etapa envolve a geração de um grafo a partir de um modelo generativo específico, como Erdős-Rényi, utilizando um conjunto de parâmetros pré-definidos.

A segunda etapa é a definição da população oculta  $H$  dentro desta rede. Um subconjunto de vértices é selecionado para compor esta população, cuja prevalência real,  $\rho = |H|/|V|$ , é calculada e armazenada como o valor de referência.

Subsequentemente, a fase de execução da simulação emula o processo de pesquisa de campo do NSUM. O processo inicia-se com a amostragem de um subconjunto de vértices  $S$  da população total  $V$ , realizado através de uma seleção aleatória uniforme sem reposição. Esta etapa simula a limitação prática de não poder inquirir toda a população.

Para cada vértice  $v$  na amostra  $S$ , são coletadas as informações relacionais, o seu número total de vizinhos de entrada,  $R_v$ , e o número de vizinhos de entrada que pertencem à população oculta,  $C_v$ . Com base neste conjunto de dados agregados da amostra, são então calculadas as estimativas da prevalência da população oculta,  $\hat{\rho}_{\text{MoR}}$  e  $\hat{\rho}_{\text{RoS}}$ , utilizando as formulações matemáticas de cada estimador.

Por fim, ocorre a fase de avaliação. As estimativas  $\hat{\rho}_{\text{MoR}}$  e  $\hat{\rho}_{\text{RoS}}$  obtidas na fase anterior são comparadas com a prevalência real  $\rho$ , que foi estabelecida na fase de configuração. O resultado desta etapa é um conjunto de valores que quantificam a acurácia de cada estimador para uma única execução do processo de amostragem. Para garantir a significância estatística, o processo é repetido múltiplas vezes para cada configuração de parâmetros, permitindo a agregação dos resultados e a análise do desempenho médio, da variabilidade e de outras propriedades estatísticas dos estimadores.

### 3.1 Métricas de Avaliação

A seleção de um conjunto apropriado de métricas é de suma importância, pois são estas que permitem traduzir os resultados brutos da simulação em conclusões significativas sobre a acurácia, a confiabilidade e a robustez de cada método. Uma única métrica é, em geral, insuficiente para capturar a complexidade do comportamento de um estimador.

Logo, foi adotada uma abordagem de avaliação empregando quatro métricas distintas, cada uma projetada para iluminar uma faceta particular do desempenho, a magnitude do erro, o risco de falha significativa, a presença de viés sistemático e a estabilidade geral da estimação.

A métrica primária utilizada para quantificar a acurácia de uma única estimativa é o erro relativo, aqui denotado por  $E_M$ . Dada uma prevalência real da população oculta,  $\rho$ , e uma prevalência estimada,  $\hat{\rho}$ , a simples diferença aritmética entre elas não é uma medida ideal, pois seu significado é dependente da escala de  $\rho$ . Para superar esta limitação, a métrica de erro é definida como a razão entre o maior e o menor valor entre a estimativa

e a realidade, conforme a equação:

$$E_M = \max\left(\frac{\hat{\rho}}{\rho}, \frac{\rho}{\hat{\rho}}\right) \quad (3.1)$$

Esta formulação apresenta propriedades desejáveis para a análise. Primeiramente, a métrica é adimensional e normalizada, onde um valor de  $E_M = 1.0$  representa uma estimativa perfeita, e valores superiores a 1.0 quantificam a magnitude do erro. Em segundo lugar, ela trata superestimções e subestimções de forma simétrica; por exemplo, uma estimativa que seja o dobro do valor real ( $\hat{\rho} = 2\rho$ ) resulta no mesmo erro de uma estimativa que seja a metade do valor real ( $\hat{\rho} = 0.5\rho$ ), sendo  $E_M = 2.0$  em ambos os casos.

Nos experimentos, o erro médio, calculado como a média aritmética dos valores de  $E_M$  sobre múltiplas execuções, será utilizado como o indicador principal do desempenho esperado de um estimador.

Embora o erro médio seja um indicador útil da tendência central do desempenho, ele pode mascarar a frequência de erros de grande magnitude. Um estimador pode ter um erro médio baixo, mas ainda assim produzir, com uma frequência não desprezível, estimativas severamente imprecisas.

Para capturar esta dimensão de risco e confiabilidade, foi definida a métrica de probabilidade de erro elevado. Esta métrica quantifica a probabilidade de que a métrica de erro  $E_M$  exceda um limiar de tolerância pré-definido,  $1 + \epsilon$ . Formalmente, a métrica é:

$$P[E_M > 1 + \epsilon] \quad (3.2)$$

Neste trabalho, será utilizado um valor de  $\epsilon = 0.05$ , que corresponde a um limiar de erro de 5%. A métrica, portanto, responde a uma questão de grande relevância prática: "Qual é a probabilidade de que uma única aplicação do estimador resulte em um erro superior a 5%?". Um valor baixo para esta métrica indica um estimador mais confiável e previsível, cujos resultados raramente se desviam de forma substancial do valor real, sendo um critério de avaliação de robustez mais rigoroso que o erro médio isoladamente.

Ademais, o viés, ou *bias*, de um estimador quantifica a sua tendência sistemática de produzir estimativas que são, em média, diferentes do valor verdadeiro do parâmetro. Enquanto a métrica  $E_M$  foca na magnitude do erro, o viés foca na sua direção. Ele é formalmente definido como o valor esperado da diferença entre a estimativa e o valor real:

$$\text{Bias} = E[\hat{\rho} - \rho] \quad (3.3)$$

Em um contexto de simulação, o viés é aproximado pela média aritmética desta diferença sobre um grande número de execuções. Um viés positivo indica uma tendência

sistemática à superestimação, enquanto um viés negativo indica uma tendência à subestimação. Um estimador ideal é não-enviesado, apresentando um viés próximo de zero.

A análise do viés é particularmente importante neste trabalho, pois certas topologias de rede, como as *scale-free* com seus hubs de alto grau, podem introduzir distorções sistemáticas que afetam os estimadores de forma distinta. Esta métrica, portanto, atua como uma ferramenta de diagnóstico para identificar falhas estruturais na lógica de um estimador quando confrontado com estruturas de rede específicas.

Para obter uma compreensão completa e granular do comportamento do erro, para além de medidas de resumo como a média ou a probabilidade, é fundamental analisar a sua distribuição estatística completa. A metodologia empregada para esta análise é a visualização da distribuição da métrica de erro  $E_M$  através de diagramas de caixa, ou *boxplots*.

O *boxplot* revela simultaneamente múltiplas propriedades da distribuição. A linha central representa a mediana, uma medida de tendência central mais robusta a valores extremos do que a média. A altura da caixa, que representa o intervalo interquartil, quantifica a dispersão dos 50% centrais dos dados, servindo como um indicador da consistência do estimador, caixas mais curtas implicam em maior consistência.

Além disso, o diagrama identifica explicitamente os valores atípicos ou *outliers*, que são os pontos de dados que se situam muito além do corpo principal da distribuição. A frequência e a magnitude destes *outliers* são um indicador direto da instabilidade de um estimador.

Um método que gera um número elevado de *outliers* com erros de grande magnitude é considerado menos robusto, mesmo que sua mediana ou média de erro seja baixa. Esta análise visual, portanto, complementa as métricas quantitativas, fornecendo um diagnóstico qualitativo profundo sobre a confiabilidade geral de cada método.

## 4 EXPERIMENTOS

Todos os experimentos foram implementados utilizando a linguagem de programação Python, versão 3.12.4. A execução foi suportada por um ecossistema de bibliotecas de código aberto amplamente estabelecidas. A biblioteca NetworkX foi a ferramenta central para a criação, manipulação e armazenamento das estruturas de grafo. Para as operações numéricas, geração de números aleatórios e processos de amostragem, utilizou-se a biblioteca NumPy.

A organização, agregação e exportação dos dados brutos gerados pelas simulações foram gerenciadas pela biblioteca Pandas, que permitiu a estruturação dos resultados em formato tabular CSV para análise subsequente. A visualização dos resultados foi realizada com as bibliotecas Matplotlib e Seaborn.

Para garantir a total reprodutibilidade dos resultados, uma semente pseudoaleatória fixa foi definida no início de cada execução do script de simulação. Isso assegura que a sequência de números aleatórios utilizada na geração dos grafos, na seleção da população oculta e na amostragem dos vértices seja a mesma em diferentes execuções, permitindo que os resultados possam ser replicados de forma idêntica.

Adicionalmente, foi implementado um sistema de *cache* para os grafos sintéticos. Uma vez que a geração de redes de grande escala é um processo computacionalmente intensivo, cada grafo gerado foi salvo em disco. Em execuções subsequentes, o sistema verifica a existência do grafo em *cache* e o carrega diretamente, otimizando o tempo de execução sem comprometer a integridade do experimento.

O procedimento de execução para todos os experimentos com redes sintéticas seguiu um fluxo padronizado para garantir a consistência. Para cada combinação de parâmetros específica de um experimento, foram geradas 50 instâncias de grafos independentes. Dentro de cada uma dessas instâncias, uma única população oculta foi selecionada aleatoriamente. Subsequentemente, o processo de amostragem foi executado, e para cada tamanho de amostra  $|S|$  definido, foram extraídas 20 amostras distintas de vértices.

Esta estrutura aninhada, com 50 repetições no nível do grafo e 20 no nível da amostragem, totalizando 1.000 execuções para cada ponto de dados, foi projetada para garantir que os resultados agregados sejam estatisticamente robustos e que o desempenho médio não seja influenciado por artefatos de uma única configuração aleatória. Em cada execução, as estimativas  $\rho_{\text{MoR}}$  e  $\rho_{\text{RoS}}$  foram calculadas e armazenadas para a análise posterior.

## 4.1 Redes de Erdős-Rényi

O objetivo desta análise é estabelecer o desempenho dos estimadores em uma topologia puramente estocástica, que serve como uma referência fundamental. Neste modelo, cada possível aresta direcionada entre dois vértices é formada com uma probabilidade  $p$  independente, resultando em uma rede com estrutura homogênea e desprovida de padrões complexos.

A configuração paramétrica para este experimento foi definida para permitir a avaliação de múltiplos fatores, como a escalabilidade em relação ao tamanho da rede e o impacto do esforço de amostragem. Os parâmetros específicos utilizados estão detalhados na Tabela 2. A escolha de dois tamanhos de rede  $n$ , 100.000 e 1.000.000 de vértices, visa observar se o desempenho dos estimadores se mantém consistente em diferentes ordens de magnitude.

O grau médio esperado foi fixado em 30, um valor que gera redes esparsas, porém conectadas, análogo ao observado em muitas redes sociais de grande escala. A prevalência da população oculta  $\rho$  foi definida em 5%, um valor representativo de muitos cenários de populações minoritárias ou de difícil acesso.

A variação do tamanho da amostra  $|S|$ , de 100 a 10.000, foi definida para reproduzir o delineamento experimental adotado pelos autores no artigo de referência [4]. Esta escolha permite mapear a curva de convergência dos estimadores à medida que mais informação é coletada da rede, possibilitando uma comparação direta com os resultados originais. Estes valores foram selecionados para cobrir um espectro que vai de uma amostragem muito esparsa até uma amostragem mais substancial, refletindo um intervalo de esforços práticos em pesquisas de campo.

Tabela 2 – Parâmetros para os experimentos com redes de Erdős-Rényi.

Parâmetro	Valor(es)
Número de Vértices ( $n$ )	100.000, 1.000.000
Grau Médio Esperado	30
Prevalência da Pop. Oculta ( $\rho$ )	0.05 (5%)
Tamanhos de Amostra ( $ S $ )	100, 500, 1.000, 5.000, 10.000
Número de Instâncias de Grafo	50
Amostras por Instância	20

A execução deste experimento seguiu o procedimento padrão detalhado no início deste capítulo. A aplicação desta metodologia para cada combinação de parâmetros especificada na Tabela 2 resultou na geração de um conjunto de dados abrangente. Estes dados formam a base para a análise de desempenho dos estimadores em redes homogêneas, cujos resultados quantitativos são apresentados no capítulo subsequente de resultados.



## 4.2 Redes de Barabási-Albert

O objetivo deste experimento é avaliar a robustez dos estimadores MoR e RoS em redes *scale-free*, cuja estrutura é caracterizada por uma distribuição de grau que segue uma lei de potência. Este tipo de rede, gerado pelo mecanismo de conexão preferencial, resulta na emergência de um pequeno número de vértices com um grau muito elevado, os chamados *hubs*, enquanto a maioria dos vértices possui poucas conexões.

Esta topologia é amplamente reconhecida como um modelo mais fiel para muitas redes sociais, tecnológicas e biológicas do mundo real, tornando este experimento um passo crucial para avaliar a aplicabilidade prática dos estimadores.

Uma consideração metodológica importante neste experimento diz respeito à natureza do grafo gerado. O modelo canônico de Barabási-Albert, conforme implementado na biblioteca NetworkX, produz um grafo não direcionado. Contudo, a formulação dos estimadores, conforme detalhado no capítulo de Metodologia, baseia-se no conceito de vizinhança de entrada *in-neighbors*, uma propriedade inerente a grafos direcionados. Para compatibilizar o modelo topológico com os requisitos dos estimadores, uma modelagem de grafo direcionado simétrico foi adotada.

É importante destacar que, embora o trabalho utilizado como referência apresente resultados para redes do tipo *scale-free*, os autores não especificam o algoritmo gerador utilizado. Neste trabalho, adotou-se explicitamente o modelo de Barabási-Albert, amplamente reconhecido na literatura para a geração de redes com distribuição de grau em lei de potência. Ao definir inequivocamente o modelo gerador, este experimento contribui para a transparência e facilita a reprodutibilidade dos resultados para esta classe de topologia.

A partir do grafo não direcionado gerado, foi criado um grafo direcionado onde cada aresta não direcionada  $\{u, v\}$  foi convertida em um par de arestas direcionadas,  $(u, v)$  e  $(v, u)$ . Sob esta modelagem, o grau de entrada de qualquer vértice  $v$  no grafo direcionado é precisamente igual ao seu grau total no grafo não direcionado original, representando de forma fiel a noção de conexões em uma rede social intrinsecamente recíproca.

Os parâmetros para esta série de experimentos, detalhados na Tabela 3, foram selecionados para investigar a interação entre a topologia *scale-free* e outras variáveis. A variação do número de vértices  $n$  mantém o objetivo de analisar a escalabilidade. O parâmetro  $m$ , fixado em 30, define o número de arestas que cada novo vértice forma durante o processo de crescimento da rede.

Diferentemente do experimento anterior, a prevalência da população oculta  $\rho$  foi variada em três níveis distintos. Esta variação é fundamental para investigar se a presença de hubs impacta de forma diferente a estimação de populações raras versus populações mais comuns, uma questão central para a aplicabilidade dos estimadores.

Os tamanhos de amostra  $|S|$  e a prevalência  $\rho$  seguiram a mesma metodologia de variação aplicada ao experimento anterior e ao trabalho base. Esta manutenção dos parâmetros de amostragem é fundamental para isolar o efeito da mudança de topologia no desempenho dos estimadores, em busca de manter a consistência comparativa entre os cenários.

Tabela 3 – Parâmetros para os experimentos com redes de Barabási-Albert.

Parâmetro	Valor(es)
Número de Vértices ( $n$ )	25.000, 250.000
Parâmetro de Conexão ( $m$ )	30
Prevalência da Pop. Oculta ( $\rho$ )	0.10, 0.20, 0.30
Tamanhos de Amostra ( $ S $ )	100, 500, 1.000, 5.000, 10.000
Número de Instâncias de Grafo	50
Amostras por Instância	20

O fluxo de execução para este experimento também seguiu o procedimento padrão descrito no início deste capítulo, utilizando os parâmetros definidos na Tabela 3. A avaliação do impacto desta estrutura heterogênea sobre o desempenho dos estimadores será detalhada no capítulo de resultados.

### 4.3 Redes de Watts–Strogatz

Este conjunto de experimentos investiga o desempenho dos estimadores NSUM em topologias com a propriedade de "mundo pequeno". As redes de Watts-Strogatz são caracterizadas por possuírem, simultaneamente, um caminho médio curto entre quaisquer dois vértices e um alto coeficiente de clusterização.

Esta última propriedade, em particular, modela a tendência observada em muitas redes sociais onde os amigos de uma pessoa também tendem a ser amigos entre si, formando comunidades locais coesas. O objetivo deste experimento é, portanto, avaliar se a alta redundância de conexões em vizinhanças locais impacta a eficiência dos estimadores em coletar informações sobre a população oculta.

Assim como no modelo de Barabási-Albert, a implementação canônica do modelo de Watts-Strogatz gera um grafo não direcionado. Para manter a consistência metodológica com os experimentos anteriores e atender aos requisitos de formulação dos estimadores, foi adotada a mesma abordagem de modelagem.

Cada grafo não direcionado gerado foi convertido em um grafo direcionado simétrico, onde a vizinhança de entrada de um vértice corresponde ao conjunto completo de seus vizinhos na estrutura original.

Os parâmetros para este experimento, descritos na Tabela 4, foram definidos para explorar as características únicas do modelo. O parâmetro  $k$  determina o grau médio

inicial de cada vértice, fixado em 60 para manter uma densidade de rede comparável à dos experimentos anteriores. O parâmetro que determina a probabilidade de religação  $p$ , foi variado em dois níveis.

Um valor baixo  $p = 0.01$  gera redes com uma estrutura altamente ordenada e clusterizada, próxima de um anel regular. Um valor mais alto  $p = 0.1$  introduz maior aleatoriedade na rede, diminuindo a clusterização local, mas preservando a propriedade de "mundo pequeno".

Esta variação permite investigar se o desempenho do estimador é sensível ao nível de ordem versus aleatoriedade na estrutura da rede. Os demais parâmetros, como o tamanho da rede  $n$ , a prevalência da população oculta  $\rho$  e os tamanhos de amostra  $|S|$ , foram mantidos consistentes com os experimentos anteriores para facilitar a análise comparativa.

Tabela 4 – Parâmetros para os experimentos com redes de Watts-Strogatz.

Parâmetro	Valor(es)
Número de Vértices ( $n$ )	25.000, 250.000
Grau Médio ( $k$ )	60
Prob. de Religação ( $p$ )	0.01, 0.1
Prevalência da Pop. Oculta ( $\rho$ )	0.10, 0.20, 0.30
Tamanhos de Amostra ( $ S $ )	100, 500, 1.000, 5.000, 10.000
Número de Instâncias de Grafo	50
Amostras por Instância	20

A execução deste experimento seguiu o procedimento padrão detalhado no início deste capítulo, utilizando os parâmetros definidos na Tabela 4. A análise quantitativa dos dados gerados será apresentada no capítulo de resultados.



## 5 RESULTADOS

### 5.1 Rede Erdős-Rényi

A principal conclusão extraída desta seção é que, no contexto de uma rede aleatória, os estimadores MoR e RoS apresentam desempenho virtualmente idêntico em todas as métricas avaliadas. A avaliação de desempenho foi realizada por meio do erro médio  $E_M$ , onde um valor de 1.0 indica uma estimativa perfeita. A Figura 1 exibe o erro médio para os estimadores.

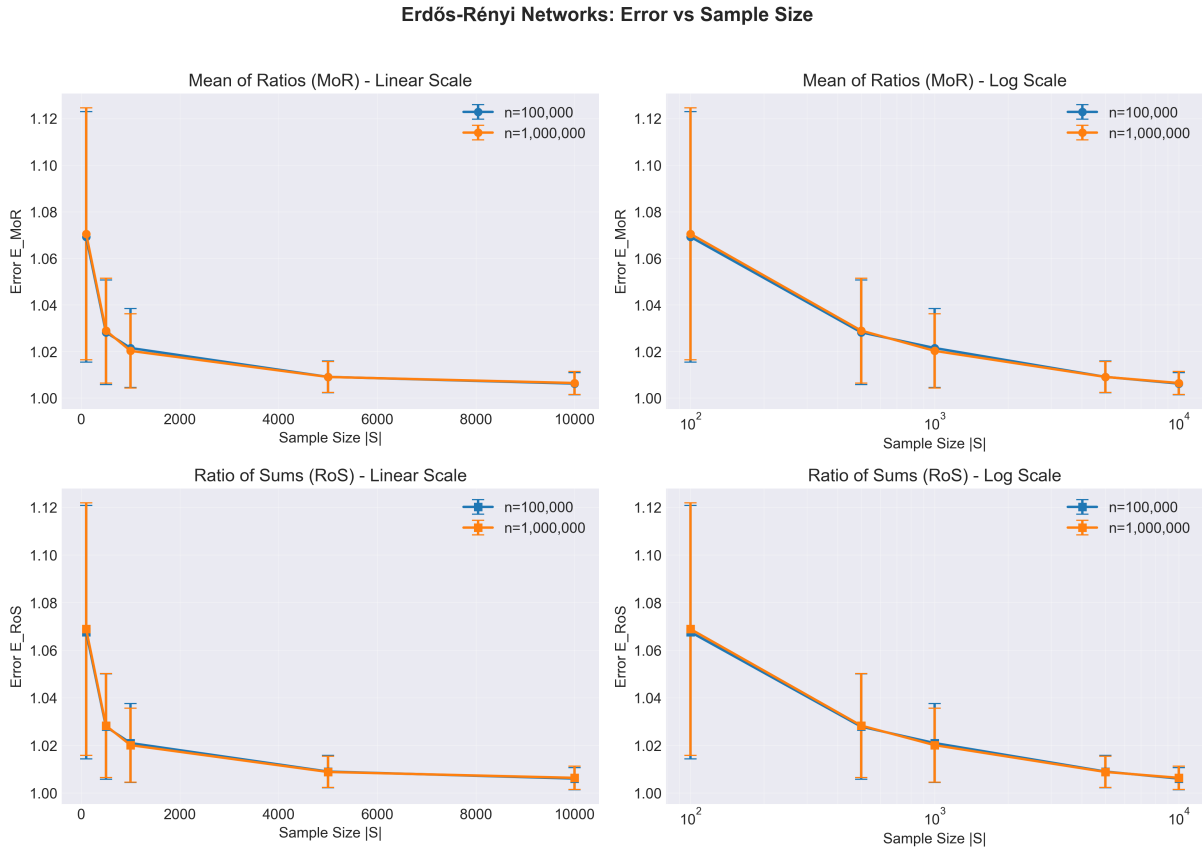


Figura 1 – Erro médio dos estimadores MoR e RoS em função do tamanho da amostra para redes Erdős-Rényi.

Consistentemente, o erro médio decresce à medida que o tamanho da amostra aumenta, um comportamento esperado que demonstra a convergência dos estimadores. Notavelmente, a performance dos estimadores MoR e RoS é quase idêntica em todas as condições testadas. O relatório sumário indica, por exemplo, que para  $n=1.000.000$  e  $|S|=1.000$ , o erro médio do MoR foi de 1.020257, enquanto o do RoS foi de 1.020096, uma diferença estatisticamente insignificante.

Para uma análise mais aprofundada da estabilidade, a Figura 2 apresenta a distribuição completa do erro através de *boxplots*.

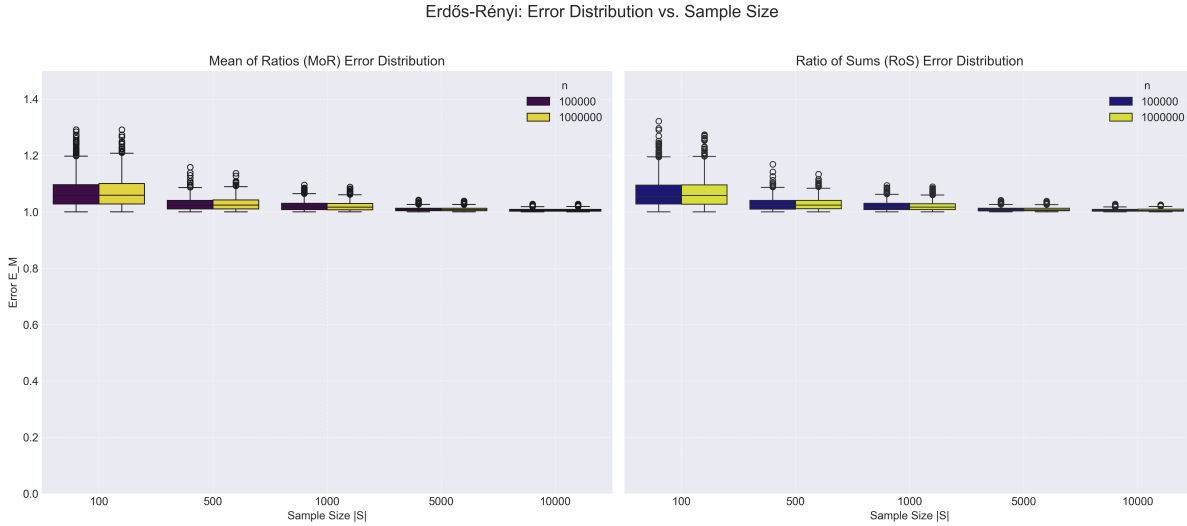


Figura 2 – Distribuição do erro dos estimadores MoR e RoS para redes Erdős-Rényi, mostrando mediana, quartis e *outliers*.

As distribuições de erro para MoR e RoS são visualmente indistinguíveis. Ambos apresentam mediana próxima de 1.0 e uma redução na dispersão e no número de *outliers* com o aumento do tamanho da amostra. A análise conjunta do erro médio e de sua distribuição demonstra que não há vantagem prática na escolha de um estimador sobre o outro em termos de acurácia ou consistência no contexto de redes Erdős-Rényi.

A análise de risco quantifica a probabilidade de se obter uma estimativa com um erro superior a um limiar de 5%, representada por  $P[E > 1.05]$ . A Figura 3 ilustra este perfil de risco.

O risco de uma estimativa significativamente incorreta é elevado para amostras pequenas, acima de 50% para  $|S|=100$ , mas decresce drasticamente, aproximando-se de zero para amostras com 5.000 ou mais nós. Assim como nas métricas anteriores, os perfis de risco para MoR e RoS são idênticos. A escolha do estimador não altera a confiabilidade do resultado, apenas o aumento do tamanho da amostra é uma estratégia eficaz para mitigar o risco.

Desta forma, demonstra-se que, para uma rede de topologia aleatória, ambos os estimadores são não viesados, apresentam perfis de erro e risco idênticos, e sua performance é ditada primariamente pelo tamanho da amostra.

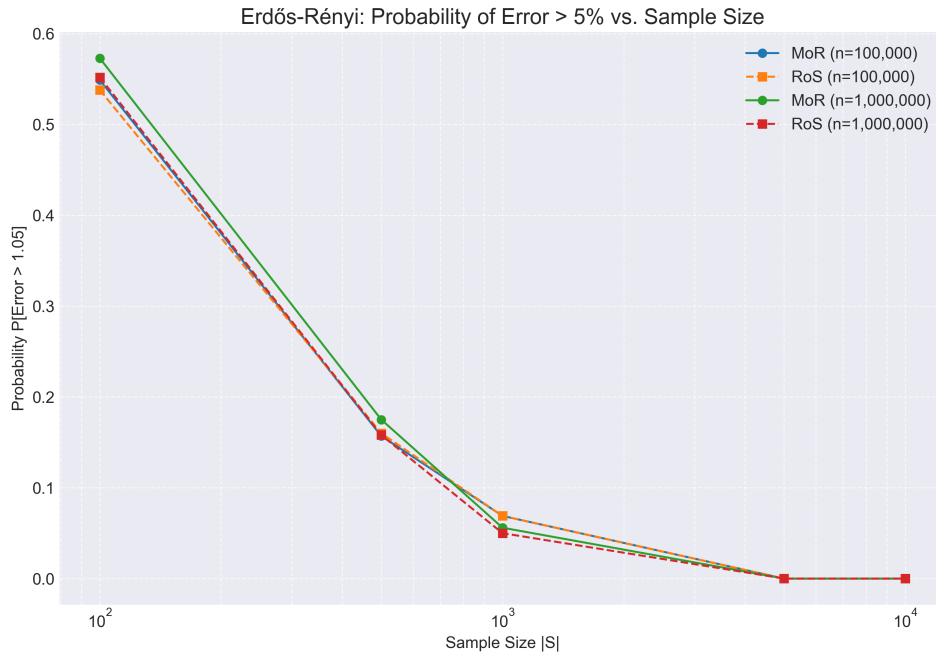


Figura 3 – Probabilidade de o erro do estimador exceder 5% em função do tamanho da amostra para redes Erdős-Rényi.

## 5.2 Rede Barabási-Albert

A hipótese central para esta seção é que tal heterogeneidade estrutural irá expor diferenças de desempenho entre os estimadores MoR e RoS. Os resultados a seguir indicam que, embora algumas métricas não revelem uma divergência drástica, a análise de risco demonstra uma vantagem quantificável e consistente do estimador RoS.

A presença de *hubs* é teoricamente uma fonte de viés para o estimador MoR, que trata a informação de cada nó amostrado com peso igual, podendo sub-representar a contribuição dos hubs. A Figura 4 apresenta a análise de viés para os estimadores no modelo Barabási-Albert.

Contrariamente à forte expectativa teórica de um viés negativo para o MoR, os resultados da simulação não demonstram um viés sistemático e significativo para nenhum dos estimadores. Conforme a Figura 4 e o relatório sumário, os valores de viés médio para ambos os estimadores são da ordem de  $10^{-4}$  e flutuam entre valores positivos e negativos dependendo dos parâmetros de prevalência ( $\rho$ ) e tamanho da rede ( $n$ ).

Para  $n = 25.000$  e  $\rho = 0.10$ , o MoR apresenta um pequeno viés positivo, enquanto para  $\rho = 0.20$ , o viés é negativo. Este comportamento sugere que, para os parâmetros de geração de rede utilizados nesta simulação, o efeito estrutural que induz o viés pode não ser suficientemente pronunciado para se manifestar de forma clara e consistente.

A avaliação do erro médio, apresentada na Figura 5, revela uma pequena, porém

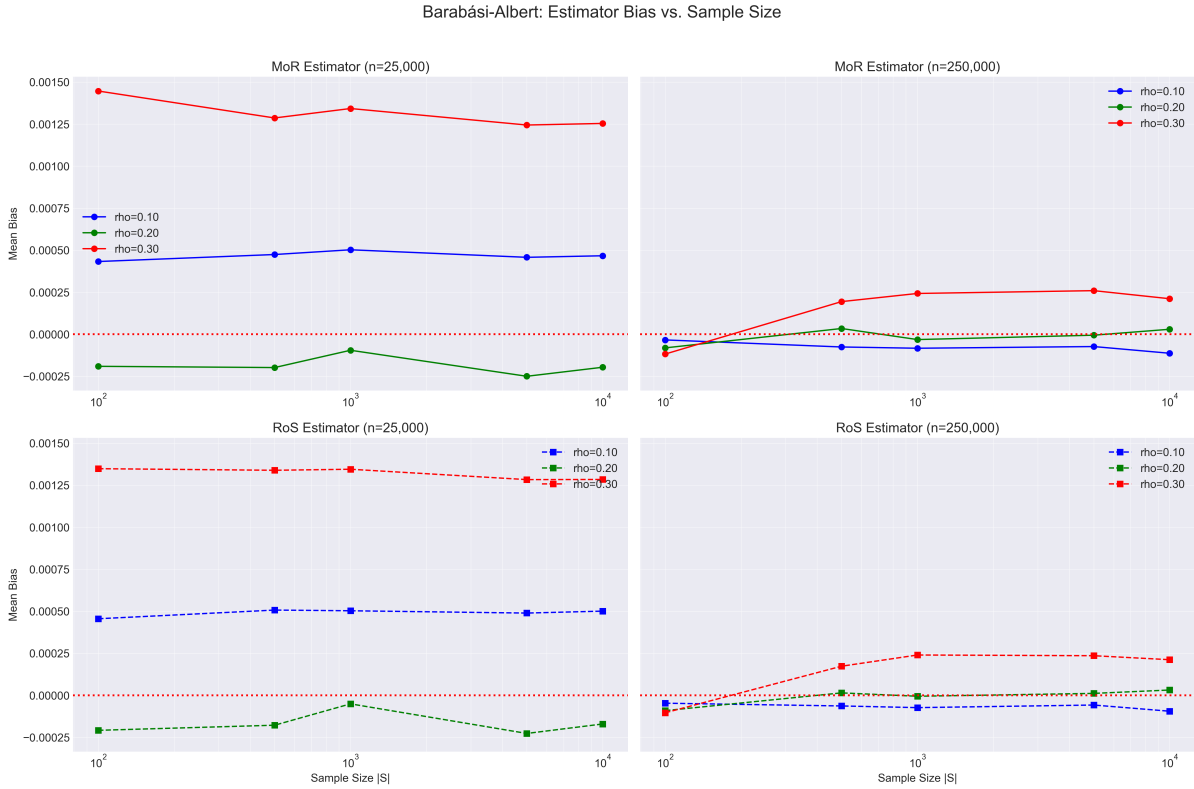


Figura 4 – Análise de viés dos estimadores MoR e RoS para redes Barabási-Albert. A figura é organizada por tamanho de rede (colunas) e tipo de estimador (linhas).

consistente, vantagem do estimador RoS.

Em todos os cenários, as curvas do estimador RoS se posicionam ligeiramente abaixo das curvas do estimador MoR. Para  $n = 250.000$ ,  $\rho = 0.10$  e  $|S| = 500$ , o erro médio do MoR é de 1.016849, enquanto o do RoS é de 1.014876. Embora a diferença seja modesta, sua consistência em múltiplas condições indica uma superioridade marginal do RoS em termos de acurácia média.

A métrica que revela a diferença mais significativa entre os estimadores é a probabilidade de erro superior a 5%, conforme ilustrado na Figura 6.

Nesta análise, o estimador RoS demonstra uma vantagem clara e consistente sobre o MoR. Para qualquer condição de  $n$  e  $\rho$ , a curva de risco do RoS está posicionada abaixo da curva do MoR, especialmente para tamanhos de amostra menores. Esta diferença quantifica o RoS como uma escolha mais confiável, oferecendo uma menor probabilidade de produzir uma estimativa com erro significativo.

Diferentemente dos resultados para a rede Erdős-Rényi, a topologia *scale-free* do modelo Barabási-Albert foi suficiente para revelar uma diferença de desempenho entre os estimadores. A análise de risco demonstrou a superioridade do estimador RoS. A sua capacidade de ponderar a informação pela conectividade dos nós, mesmo que o efeito



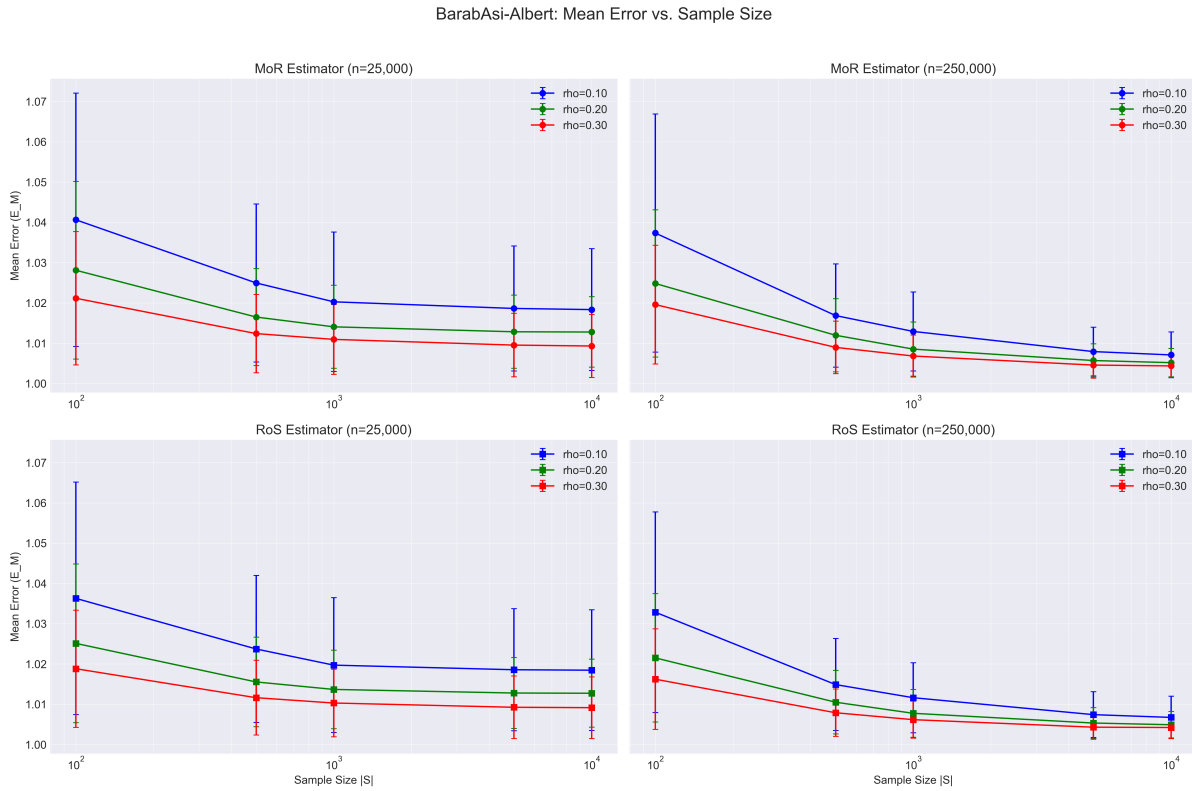


Figura 5 – Erro médio dos estimadores MoR e RoS em função do tamanho da amostra para redes Barabási-Albert, facetado pelo tipo de estimador (linhas) e tamanho da rede (colunas).

sobre o erro médio seja modesto, resulta em um estimador mais confiável e com menor probabilidade de falha.

Estes resultados empíricos, que apontam para uma maior confiabilidade do estimador RoS em redes *scale-free*, corroboram a análise teórica desenvolvida no artigo de referência [4]. O trabalho original estabelece limitantes de erro analíticos que são específicos para diferentes topologias de rede, reconhecendo que a estrutura da rede subjacente impacta diretamente o desempenho dos estimadores.

Especificamente para redes *scale-free*, o artigo deriva um limitante superior de erro aprimorado para o estimador RoS, que é mais apertado do que os limitantes mais genéricos aplicáveis a redes aleatórias em geral. A existência de um teorema dedicado a este caso sinaliza que a estrutura de conectividade heterogênea, com a presença de *hubs*, permite uma análise teórica mais refinada para o RoS. Esta vantagem teórica encontra sua contraparte empírica em nossos resultados de simulação.

Embora as diferenças no erro médio e no viés não tenham se mostrado proeminentes, a métrica mais diretamente relacionada a um limitante superior de erro é a probabilidade de o erro exceder um determinado limiar. A superioridade clara do RoS na análise de risco, documentada na Figura 6, é, portanto, a manifestação prática mais

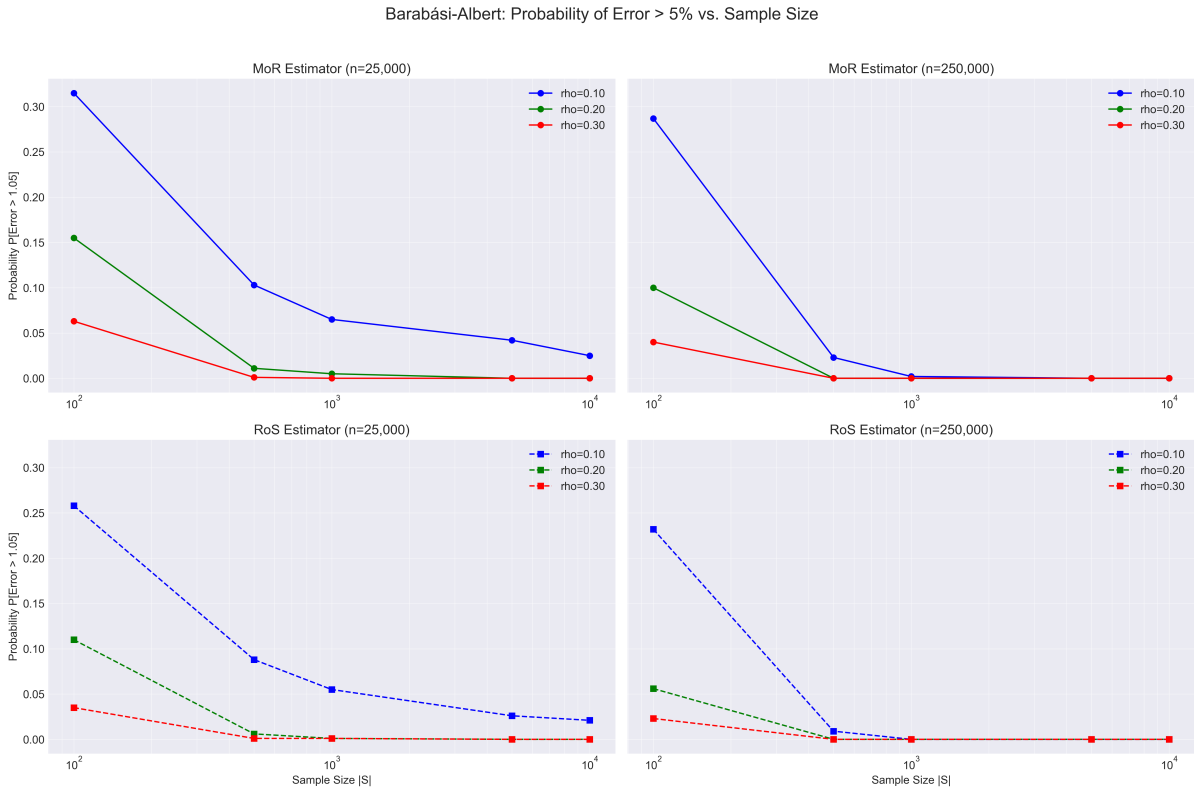


Figura 6 – Probabilidade de o erro do estimador exceder 5% em função do tamanho da amostra para redes Barabási-Albert.

direta da vantagem teórica formalizada pelo artigo de referência. A convergência entre o limitante teórico mais restrito e a menor probabilidade de erro observada na prática confere um suporte robusto à conclusão de que o estimador RoS é, de fato, mais adequado para a análise de redes com distribuição de grau do tipo *scale-free*.

### 5.3 Rede Watts-Strogatz

Este modelo é particularmente interessante por interpolar entre uma rede regular tipo anel com alto coeficiente de agrupamento e uma rede aleatória, através do parâmetro de probabilidade de religação  $p$ . Foram testados cenários com baixo valor de  $p = 0.01$ , que preserva a estrutura local e o alto agrupamento, e outro com um valor maior de  $p = 0.1$ , que torna a rede mais próxima de uma topologia aleatória.

Avaliar a performance dos estimadores neste modelo apresenta alguns desafios, pois a análise teórica do artigo base [4] pode não se aplicar diretamente a grafos gerados por este processo. A teoria de erro proposta naquele trabalho parte da definição de *random networks*, na qual os vizinhos de cada nó são selecionados de forma uniforme e aleatória do restante da rede.

O modelo Watts-Strogatz, contudo, viola esta premissa, tendo em vista que sua

geração parte de uma estrutura de rede altamente regular, um anel, e aplica um processo de religação local. Este método preserva um alto coeficiente de agrupamento, uma característica de “mundo pequeno” que não emerge de um processo de seleção de vizinhos puramente aleatório e uniforme.

Portanto, esta seção investiga empiricamente o comportamento dos estimadores MoR e RoS em um cenário que se desvia das suposições teóricas. A principal característica estrutural das redes Watts-Strogatz que se mantém é a sua distribuição de grau regular e homogênea, similar à de um grafo aleatório e distinta da heterogeneidade do modelo Barabási-Albert.

A hipótese para esta seção, então, é que, mesmo com o processo de seleção de vizinhos não-uniforme, a forma da distribuição de graus continua sendo o fator dominante na determinação do viés dos estimadores. Espera-se que, devido à ausência de *hubs*, o desempenho dos estimadores MoR e RoS seja similar, assim como observado no modelo Erdős-Rényi.

A Figura 7 apresenta a análise de viés para os estimadores no modelo Watts-Strogatz, organizada pelos parâmetros de tamanho da rede  $n$  e probabilidade de religação  $p$ .

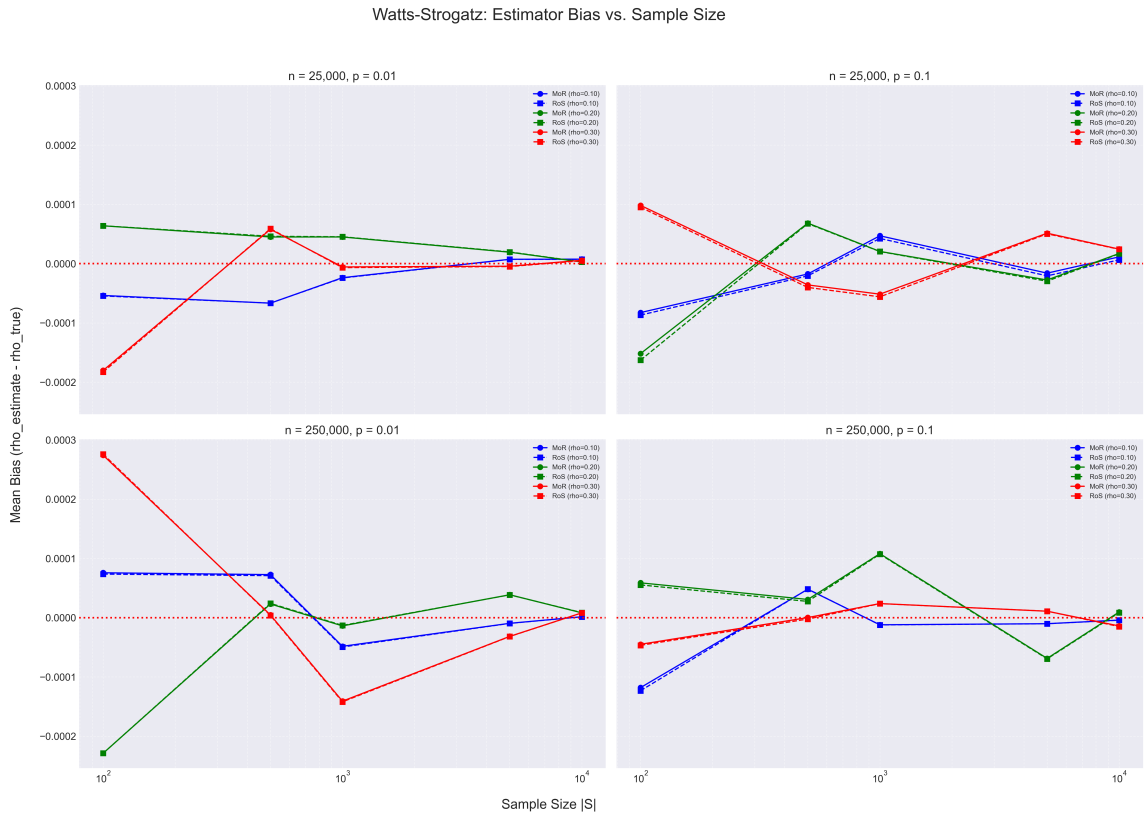


Figura 7 – Análise de viés dos estimadores MoR (linhas contínuas) e RoS (linhas tracejadas) para redes Watts-Strogatz.

Em conformidade com o comportamento observado no modelo Erdős-Rényi, ambos os estimadores se mostram não viesados no modelo Watts-Strogatz. Em todos os quatro *subplots*, que cobrem as variações de  $n$  e  $p$ , as linhas de viés para MoR e RoS flutuam em torno da linha de referência  $y = 0$ . Este resultado, embora não constitua uma validação direta da teoria do artigo de base devido às diferenças no processo de geração do grafo, é o desfecho esperado.

A ausência de viés sistemático reforça a hipótese de que a principal causa do viés do estimador MoR é a heterogeneidade da distribuição de graus, ou seja, a presença de *hubs*, e não outras propriedades estruturais como o alto coeficiente de agrupamento. Uma vez que o modelo Watts-Strogatz possui uma distribuição de grau homogênea, era de se esperar que os estimadores tivessem um desempenho similar e não viesado, o que os dados confirmam.

As demais métricas de avaliação, erro médio, distribuição do erro e probabilidade de erro significativo, apresentam um comportamento uníssono e conclusivo. A Figura 8 é representativa de todos os resultados.

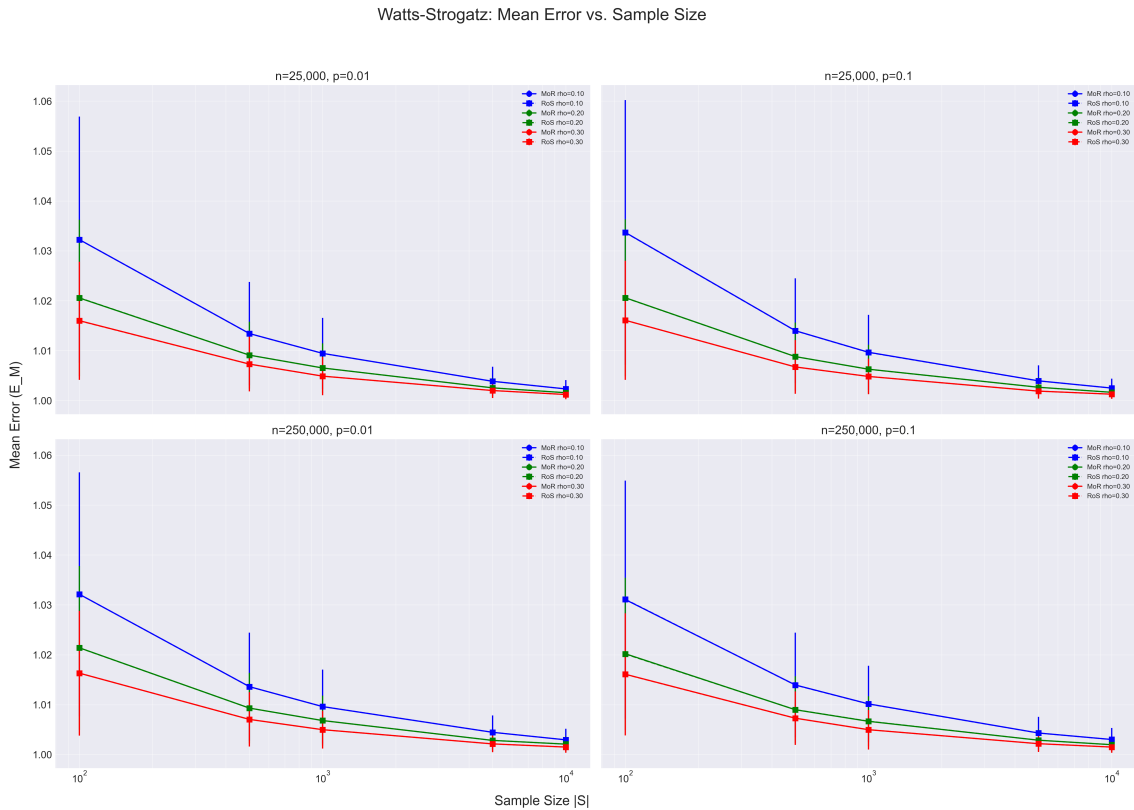


Figura 8 – Erro médio dos estimadores MoR e RoS em função do tamanho da amostra para redes Watts-Strogatz.

As linhas e distribuições para os estimadores MoR e RoS são praticamente indistinguíveis. Os valores de erro, desvio padrão e probabilidade são idênticos até a quarta

ou quinta casa decimal na maioria dos casos. Por exemplo, para  $n = 250.000$ ,  $p = 0.01$ ,  $\rho = 0.10$  e  $|S| = 500$ , o erro médio do MoR é de 1.013611, e o do RoS é de 1.013609. A probabilidade de erro significativo para a mesma condição é de 0.009000 para ambos.

Assim, conclui-se que a performance dos estimadores MoR e RoS é idêntica em redes com distribuição de grau regular, independentemente de a rede possuir uma estrutura local com alto coeficiente de agrupamento  $p = 0.01$  ou ser mais aleatória  $p = 0.1$ . Isso demonstra que a propriedade estrutural que diferencia o desempenho dos estimadores não é o agrupamento local representado pela característica de "mundo pequeno", mas sim a heterogeneidade da distribuição de conectividade.



## 6 CONCLUSÃO

Este trabalho propôs-se a investigar uma questão central para a aplicação do NSUM, a influência da topologia da rede no desempenho e na confiabilidade de seus dois principais estimadores, o MoR e o RoS sob diferentes topologias de grafos aleatórios por meio de simulações computacionais medindo o erro estimativo em cada cenário.

Para responder a esta questão, foi conduzida uma análise empírica multifacetada em três modelos de rede distintos. Os resultados obtidos nos modelos de Erdős-Rényi e Watts-Strogatz estabeleceram uma linha de base fundamental. Em ambas as topologias, caracterizadas por uma distribuição de grau homogênea, o desempenho dos estimadores MoR e RoS mostrou-se virtualmente idêntico em todas as métricas de avaliação, erro médio, viés, distribuição do erro e risco de falha.

Esta equivalência em redes aleatórias e de mundo pequeno demonstrou que propriedades como a *clusterização* local não são o fator determinante que diferencia os estimadores. A divergência de desempenho, portanto, deveria emergir em uma topologia com características estruturais distintas.

A análise no modelo de Barabási-Albert, que gera redes do tipo *scale-free*, revelou esta diferenciação. Embora a diferença no erro médio tenha sido modesta e um viés sistemático forte não tenha sido consistentemente observado nos parâmetros testados, a análise de risco validou o resultado teórico do estimador RoS. Neste cenário específico, o RoS apresentou uma probabilidade consistentemente menor de produzir estimativas com erro significativo, especialmente com amostras de menor tamanho.

Os resultados, em conjunto, permitem levantar um hipótese ao problema inicial. A superioridade de um estimador sobre o outro não é absoluta, mas sim dependente da estrutura da rede. A propriedade que se mostrou crucial foi a heterogeneidade da distribuição de grau. A presença de *hubs*, característica central das redes *scale-free*, é o fator que torna o estimador RoS tecnicamente superior em alguns cenários, não necessariamente por uma acurácia média drasticamente maior, mas por sua maior confiabilidade e menor risco.

Como limitação, nota-se que o viés teoricamente esperado para o MoR no modelo Barabási-Albert não se manifestou de forma proeminente, o que pode ser atribuído aos parâmetros de geração de rede específicos utilizados. Trabalhos futuros poderiam explorar uma gama mais ampla destes parâmetros para investigar as condições em que este viés se torna mais acentuado.

Adicionalmente, futuras investigações poderiam estender esta análise para outros modelos de rede, como os baseados em blocos estocásticos, e incluir a avaliação de estra-

tégias de amostragem mais complexas, comparando seus efeitos sobre o desempenho dos estimadores.



## REFERÊNCIAS

- [1] MCCORMICK, T. H.; ZHENG, T. Latent demographic profile estimation in hard-to-reach groups. *Ann. Appl. Stat.*, Institute of Mathematical Statistics, v. 6, n. 4, p. 1795–1813, dez. 2012.
- [2] JACKSON, M. O. *Social and Economic Networks*. [S.l.]: Princeton University Press, 2010. ISBN 9780691148205.
- [3] BREZA, E. et al. Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, v. 110, n. 8, p. 2454–84, August 2020. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.20170861>.
- [4] DÍAZ-ARANDA, S. et al. Error bounds for the network scale-up method. In: *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*. New York, NY, USA: Association for Computing Machinery, 2025. (KDD '25), p. 498–509. ISBN 9798400714542. Disponível em: <https://doi.org/10.1145/3711896.3736940>.
- [5] LAGA, I.; BAO, L.; NIU, X. Thirty years of the network scale-up method. *Journal of the American Statistical Association*, ASA Website, v. 116, n. 535, p. 1548–1559, 2021. PMID: 37994314. Disponível em: <https://doi.org/10.1080/01621459.2021.1935267>.
- [6] KUNKE, J. P. et al. *Comparing the Robustness of Simple Network Scale-Up Method (NSUM) Estimators*. 2024. Disponível em: <https://arxiv.org/abs/2303.07490>.
- [7] FIENBERG, S. E. A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics*, ASA Website, v. 21, n. 4, p. 825–839, 2012. Disponível em: <https://doi.org/10.1080/10618600.2012.738106>.
- [8] NEWMAN, M. *Networks*. [S.l.]: Oxford University Press, 2018. ISBN 9780198805090.
- [9] BARABÁSI, A.; PÓSFAL, M. *Network Science*. Cambridge University Press, 2016. ISBN 9781107076266. Disponível em: <https://networksciencebook.com/>.
- [10] MALTIEL, R. et al. Estimating population size using the network scale up method. *Ann. Appl. Stat.*, Institute of Mathematical Statistics, v. 9, n. 3, p. 1247–1277, set. 2015.
- [11] MCCORMICK, T. H.; SALGANIK, M. J.; ZHENG, T. How many people do you know?: Efficiently estimating personal network size. *J. Am. Stat. Assoc.*, Informa UK Limited, v. 105, n. 489, p. 59–70, mar. 2010.
- [12] MCCORMICK, T. H. *THE NETWORK SCALE-UP METHOD*. *The Oxford Handbook of Social Networks*. [S.l.]: Oxford University Press, 2020. 153 p.

- [13] DÍAZ-ARANDA, S. et al. Performance analysis of nsum estimators in social-network topologies. *The American Statistician*, 2025. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/00031305.2024.2421361>.
- [14] SEYMOUR, R.; LI, S. L.; . . . . A partially pooled network scale-up method model: Detailed estimation of child sexual exploitation material trafficking prevalence in philippine municipalities. *Journal of the Royal Statistical Society C (advance article)*, 2025. Disponível em: <https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlaf033/8151526>.
- [15] –, authors not specified in snippet. *Estimating Foreign Born Scientists and Engineers (FBSE)*. [S.l.], 2023. Disponível em: [https://www.americasdatahub.org/wp-content/uploads/2025/04/Estimating\\_Foreign\\_Born\\_Scientists\\_and\\_Engineers\\_November\\_2023-final-report.pdf](https://www.americasdatahub.org/wp-content/uploads/2025/04/Estimating_Foreign_Born_Scientists_and_Engineers_November_2023-final-report.pdf).
- [16] HABECKER, P.; DOMBROWSKI, K.; KHAN, B. Improving the network scale-up estimator: Incorporating means of sums, recursive back estimation, and sampling weights. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 12, p. e0143406, 2015.
- [17] FEEHAN, D. M. et al. Quantity versus quality: A survey experiment to improve the network scale-up method. *American journal of epidemiology*, Oxford University Press, v. 183, n. 8, p. 747–757, 2016.
- [18] JANSON, S.; LUCZAK, T.; RUCINSKI, A. *Random graphs*. [S.l.]: John Wiley & Sons, 2011.
- [19] LOVÁSZ, L. *Large Networks and Graph Limits*. Providence, RI: American Mathematical Society, 2012. v. 60. (Colloquium Publications, v. 60). ISBN 978-0-8218-9085-1. Disponível em: <https://bookstore.ams.org/coll-60/>.
- [20] SNIJDERS, T. A. B. et al. New specifications for exponential random graph models. *Statistica Neerlandica (preprint)*, 2006. Disponível em: <https://www.stats.ox.ac.uk/~snijders/SnijdersPattisonRobinsHandcock2006.pdf>.
- [21] BABKIN, S. et al. Large-scale estimation of random graph models with local dependence. *Computational Statistics & Data Analysis*, v. 152, p. 107029, 2020. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7282802/>.
- [22] ERDÖS, P.; RÉNYI, A. On random graphs i. *Publ. math. debrecen*, v. 6, n. 290-297, p. 18, 1959.
- [23] BRITO, G.; DUMITRIU, I.; HARRIS, K. D. Spectral gap in random bipartite biregular graphs and applications. *arXiv preprint*, arXiv:1804.07808, 2018. Disponível em: <https://arxiv.org/abs/1804.07808>.
- [24] MITZENMACHER, M.; UPFAL, E. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge, UK: Cambridge University Press, 2005. Disponível em: <https://www.cambridge.org/core/books/probability-and-computing/3A5B47DB315FC64B9256C5C8131C5EFA>.

- [25] SHAHRAEINI, M. A comprehensive approach to synthetic distribution grid generation: Erdős-rényi to barabási-albert. *AUT Journal of Electrical Engineering*, v. 57, n. 1, p. 85–100, 2025. Disponível em: <[https://eej.aut.ac.ir/article\\_5469.html](https://eej.aut.ac.ir/article_5469.html)>.
- [26] SAMSEL, M. J.; FRONCZAK, A.; FRONCZAK, P. Universality classes in the time evolution of epidemic outbreaks on complex networks. *Phys. Rev. E*, v. 112, n. 2, p. 024308, 2025. Disponível em: <<https://doi.org/10.1103/PhysRevE.112.024308>>.
- [27] BARABÁSI, A. L.; ALBERT, R. Emergence of scaling in random networks. *science*, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999.
- [28] ZADOROZHNYI, V. N.; YUDIN, E. B. Structural properties of the scale-free Barabasi-Albert graph. *Automation and Remote Control*, v. 73, n. 4, p. 702–716, abr. 2012.
- [29] CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. J. Power-law distributions in empirical data. *SIAM review*, SIAM, v. 51, n. 4, p. 661–703, 2009.
- [30] ABRAMOV, A. et al. Barabási-albert-based network growth model to sustainable urban planning. *Sustainability*, v. 17, n. 3, p. 1095, 2025.
- [31] EASLEY, D.; KLEINBERG, J. *Networks, crowds, and markets: Reasoning about a highly connected world*. [S.l.]: Cambridge university press Cambridge, 2010. v. 1.
- [32] WILLIAMS, R.; NALISNICK, E. T.; HOLBROOK, A. Scalable generative modeling of weighted graphs. *arXiv preprint*, arXiv:2507.23111, 2025.
- [33] WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *Nature*, v. 393, n. 6684, p. 440–442, 1998. Disponível em: <<https://www.nature.com/articles/30918>>.
- [34] HUMPHRIES, M. D.; GURNEY, K. Network ‘small-world-ness’: A quantitative method for determining canonical network equivalence. *PLOS ONE*, v. 3, n. 4, p. e2051, 2008. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002051>>.
- [35] KLEMM, K.; EGUÍLUZ, V. M. Highly clustered scale-free networks. *Physical Review E*, v. 65, n. 3, p. 036123, 2002. Disponível em: <<https://doi.org/10.1103/PhysRevE.65.036123>>.
- [36] BOLLOBÁS, B. *Random Graphs (Second Edition)*. Cambridge, UK: Cambridge University Press, 2001. ISBN 0-521-80920-7. Disponível em: <<https://www.cambridge.org/core/books/random-graphs/E21023008001CFA182CE666F5028489F>>.