

ANTONIO NEWTON LICCIARDI JUNIOR

UM MODELO DE REDE COMPLEXA
COM APLICAÇÃO EM REDES SOCIAIS
COM CLASSES ECONÔMICAS

São Paulo
2024

ANTONIO NEWTON LICCIARDI JUNIOR

**UM MODELO DE REDE COMPLEXA
COM APLICAÇÃO EM REDES SOCIAIS
COM CLASSES ECONÔMICAS**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do Título de Doutor em Ciências.

São Paulo
2024

ANTONIO NEWTON LICCIARDI JUNIOR

UM MODELO DE REDE COMPLEXA
COM APLICAÇÃO EM REDES SOCIAIS
COM CLASSES ECONÔMICAS

Versão corrigida

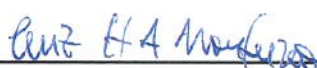
Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do Título de Doutor em Ciências.

Área de Concentração:
Engenharia de Sistemas

Orientador:
Dr. Luiz Henrique Alves Monteiro

São Paulo
2024

Tese de autoria de Antonio Newton Licciardi Junior, sob o título 'UM MODELO DE REDE COMPLEXA COM APLICAÇÃO EM REDES SOCIAIS COM CLASSES ECONÔMICAS', apresentada 'Escola Politécnica da Universidade de São Paulo, para obtenção do título de Doutor em Ciências pelo Programa de Pós-Graduação em Engenharia Elétrica, na área de concentração Engenharia de Sistemas, aprovada em 01 de fevereiro de 2024, pela comissão julgadora constituída pelos doutores:



Prof. Dr. Luiz Henrique Alves Monteiro
Escola Politécnica da USP
Presidente



Prof. Dr. José Roberto Castilho Piqueira
Escola Politécnica da USP



Prof. Dr. Iberê Luiz Caldas
Instituto de Física da USP



Prof. Dr. Rogério de Oliveira
Universidade Presbiteriana Mackenzie



Prof. Dr. Pedro Henrique Triguís Schimit
Universidade Uninove

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo-na-publicação

Licciardi Junior, A. N.

Um modelo de rede complexa com aplicação em redes sociais com classes econômicas / A. N. Licciardi Junior -- São Paulo, 2024.

78 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Telecomunicações e Controle.

1.Engenharia de Sistemas de Computação 2.Simulação (Estatística)
3.Geometria e Modelagem Computacional 4.Algoritmos Geométricos
I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II.t.

Aos meus pais Antonio e Tereza (em memória), que me ensinaram os primeiros passos, os valores que carrego e a importância da perseverança. Por todas as lições, visíveis e invisíveis, e pelo amor incondicional que me permitiu almejar os necessários voos da vida.

À minha amada esposa, Tania, minha companheira de todas as horas, que com paciência e carinho, segurou minha mão nos momentos de dúvida e celebrou comigo cada conquista. Seu apoio foi o alicerce que me permitiu persistir nos momentos mais desafiadores.

E ao nosso precioso filho, Antonio, luz que ilumina meus dias e me motiva a ser uma versão melhor de mim mesmo a cada dia. Por ser a lembrança constante do que realmente importa e do futuro que desejamos construir.

A vocês, e aos demais membros desta família, minha eterna gratidão. Esta conquista é nossa.

AGRADECIMENTOS

Primeiramente, minha profunda gratidão ao Professor Doutor Luiz Henrique Monteiro. Sua paciência sem precedentes, empatia ímpar e compreensão foram cruciais em minha jornada. Suas ideias sempre tão iluminadas e brilhantes, aliadas ao amor incondicional pelo trabalho, não apenas me inspiraram, mas também pavimentaram o caminho para a conclusão desta Tese. Aprendi assuntos fascinantes sob sua tutela e sou e serei eternamente grato pela sua confiança, amizade e apoio que tornaram possível a realização deste trabalho.

À Escola Politécnica, renomada por sua excelência acadêmica e dedicação à pesquisa de ponta, expresso minha profunda gratidão. A paciência, orientação e auxílio incontestes dos professores e funcionários foram fundamentais em minha jornada. Cada sala de aula, cada corredor, ressoa com o comprometimento e a paixão pela educação, e é um privilégio fazer parte desta comunidade. A todos que, direta ou indiretamente, contribuíram para a minha formação nesta instituição, minha sincera gratidão.

Um agradecimento especial e carinhoso aos Professores Sidnei Martini e André Hirakawa (em memória). Suas orientações cuidadosas e atentas foram essenciais para aperfeiçoar meus estudos, até a fase de orientação do Professor Luiz.

Ao Professor Doutor Paulo Garcia, estendo meus sinceros agradecimentos por sua inestimável amizade de décadas, bom-senso e constante incentivo. Seu espírito positivo foi uma luz que guiou esta empreitada acadêmica.

Aos meus colegas, amigos e companheiros de jornada, tanto no ambiente de trabalho quanto na vida, meu mais sincero agradecimento. Vocês desempenham um papel fundamental em minha evolução como ser humano, e valorizo cada momento compartilhado.

Por fim, mas certamente não menos importante, minha família. Este percurso, longo e repleto de desafios, questões pessoais e situações familiares primordiais, só foi superado graças ao suporte incondicional e ao amor que vocês me ofereceram. Em cada página deste trabalho, há um pouco da força e do carinho que me proporcionaram.

*“Deus não joga dados.”
(Albert Einstein)*

RESUMO

Redes que representam as interações sociais presenciais numa comunidade têm natureza complexa. Nesta tese, propõe-se um modelo de grafo complexo para expressar os contatos cotidianos que ocorrem numa comunidade. Nesses grafos, as conexões entre indivíduos geograficamente mais próximos são privilegiadas. No algoritmo desenvolvido para gerar tais grafos, os parâmetros são o número de nós, o grau médio do grafo, o raio em que as conexões podem ocorrer e o expoente que caracteriza a distribuição de graus, assumida obedecer a uma lei de potência. A variação do valor desse expoente faz a rede transitar de uma rede tipo mundo-pequeno para uma rede livre-de-escala. Curiosamente, a distribuição de graus desses grafos gerados computacionalmente é melhor descrita por uma lei de potência para valores do expoente tipicamente encontrados em redes do mundo real. Então, o algoritmo proposto é adaptado para levar em conta que sociedades são estratificadas em classes econômicas. A partir de dados sociodemográficos e de padrões de contato diário referentes a França, Peru e Zimbábue, realizam-se simulações numéricas que permitem quantificar diferentes aspectos das conexões interpessoais nesses países. No algoritmo modificado, as conexões são mediadas tanto pelo fator econômico quanto pela distância geográfica. Assim, privilegiam-se contatos entre indivíduos de mesma classe socioeconômica e geograficamente mais próximos. Diversas métricas topológicas são calculadas numericamente, a fim de caracterizar os grafos relacionados a esses três países. Os resultados obtidos podem ajudar a compreender o impacto da estratificação social na troca de mercadorias, na difusão de informações e na propagação de infecções.

Palavras-chaves: classe socioeconômica, coeficiente de clusterização, efeito mundo-pequeno, grafo livre-de-escala, interação social, medidas de centralidade, redes complexas.

ABSTRACT

Networks that represent face-to-face social interactions in a community are complex in nature. In this thesis, a model of complex graph is proposed to express the daily contacts that occur in a community. In these graphs, the connections between geographically closer individuals are privileged. In the algorithm developed to generate such graphs, the parameters are the number of nodes, the average degree of the graph, the radius in which the connections can occur and the exponent that characterizes the degree distribution, assumed to obey a power law. The variation in the value of this exponent causes the network to shift from a small-world type network to a scale-free network. Interestingly, the degree distribution of these computationally generated graphs is best described by a power law for exponent values typically found in real-world networks. Then, the proposed algorithm is adapted to take into account that societies are stratified into economic classes. By using sociodemographic data and daily contact patterns for France, Peru and Zimbabwe, numerical simulations are performed to quantify different aspects of the interpersonal connections in these countries. In the modified algorithm, connections are mediated by both the economic factor and geographic distance. Thus, contacts between individuals of the same socioeconomic class and who are geographically closer are privileged. Several topological metrics are numerically calculated in order to characterize the graphs related to these three countries. The obtained results can help to understand the impact of social stratification on the exchange of goods, on the dissemination of information and on the spread of infections.

Keywords: centrality measures, clustering coefficient, complex networks, scale-free graph, small-world effect, social interaction, socioeconomic class.

LISTA DE FIGURAS

2.1	Vizinhança de Von Neumann num reticulado bidimensional para $r = 1$ e $r = 2$. A célula considerada está na cor preta. Já as células vizinhas estão na cor cinza.	14
2.2	Vizinhança de Moore num reticulado bidimensional para $r = 1$ e $r = 2$. A célula considerada está na cor preta. Já as células vizinhas estão na cor cinza.	14
2.3	Acima: mapa ilustrativo do problema das sete pontes de Königsberg. Esse é o desenho original da publicação de Euler. Abaixo: um possível grafo, que corresponde à uma redução do problema. Nesse grafo, o conjunto $\{A,B,C,D\}$ representa as diversas porções de terra (A e D são as ilhas e B e C são as margens do rio) conectadas por sete pontes, que são arestas do grafo pertencentes ao conjunto $\{a,b,c,d,e,f,g\}$	15
2.4	Exemplo de um grafo não orientado com 6 nós e 5 arestas. O grafo é desconectado (ou não percolado). O subconjunto dos nós $\{1, 3, 4, 6\}$ forma um subgrafo conectado, assim como o subconjunto dos nós $\{2, 5\}$. O grafo desse exemplo é utilizado para ilustrar diversos cálculos de métricas definidas ao longo desta seção.	16
2.5	Exemplos de grafos regulares com seis nós. No grafo da esquerda, cada nó possui grau 1; no grafo central, cada nó possui grau 2 (há duas arestas conectadas a cada nó); no grafo da direita, cada nó possui grau 3.	17
2.6	Exemplo de rede aleatória, segundo a proposta de Erdős e Rényi, em que 10 nós são conectados por 9 arestas.	18
2.7	Exemplo de construção de rede mundo-pequeno, segundo a proposta de Watts e Strogatz. Da esquerda para direita, há três grafos: o primeiro, à esquerda, é um grafo regular, já que a probabilidade de reconexão é nula. No segundo grafo, ao centro, há uma probabilidade de reconexão não nula, tal que $0 < p_L < 1$. No terceiro grafo, à direita, tem-se $p_L = 1$, o que corresponde a um grafo de Erdős e Rényi. Figura adaptada de Maier (2019).	19
3.1	Valores possíveis de inicialização da rede complexa. Escala vertical em log. A curva tracejada representa o valor de k_{max} ; a curva em linha sólida, o valor de $\langle k \rangle$; e a linha pontilhada, o valor $\langle k \rangle_{alvo} = 10$ (essa linha foi plotada apenas por referência).	33

4.1	Distribuição de graus $p(k)$ em um gráfico log-log (base 10) para $N = 4900$ indivíduos, raio de Moore $r = 10$, expoente $\gamma = 3$ e $\langle k \rangle_{alvo} = \{5, 10, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100\}$. As escalas horizontais não são iguais nos 14 gráficos, a fim de facilitar a visualização de distribuição dos pontos nas diversas décadas em que ocorrem.	36
4.2	Centralidades médias de grau C_g , de proximidade C_p e de intermediação C_i , comprimento médio do caminho mais curto $\langle \ell \rangle$ e coeficiente de clusterização médio $\langle c \rangle$ em função de $\langle k \rangle$	38
4.3	Matriz de correlação das métricas $C_g, C_p, C_i, C_a, \langle l \rangle, \langle c \rangle, \langle k \rangle, \langle k_2 \rangle, \langle k_3 \rangle$ e Δk , considerando os grafo gerados a partir da escolha $\langle k \rangle_{alvo}$. As células cujos valores de correlação são mais próximos de 1 possuem um tom azul mais intenso. As células cujos valores de correlação são mais próximos de -1 possuem um tom vermelho mais intenso.	40
4.4	Matriz de correlação das métricas C_g, C_p, C_i, C_a e $\langle c \rangle$ em função de $\langle k \rangle$ para os 100 nós mais representativos. As células cujos valores de correlação são mais próximos de 1 possuem um tom azul mais intenso. As células cujos valores de correlação são mais próximos de -1 possuem um tom vermelho mais intenso.	41
4.5	Distribuição de graus $p(k)$ em função de k em gráficos log-log. Nesses gráficos, $N = 4900, r = 10, \langle k \rangle_{alvo} = 15$ e $\gamma = \{0,3; 0,5; 0,7; 1; 2; 2,5; 3; 5; 6; 7; 8; 9; 10\}$. As escalas horizontais e as verticais não coincidem, a fim de facilitar a visualização das curvas plotadas.	42
4.6	Variação de Δk em função de γ . Lembre que Δk é intervalo de graus em que o grafo obedece à lei de potência.	44
4.7	Matriz de correlação das métricas $C_g, C_p, C_i, C_a, \langle \ell \rangle, \langle c \rangle, \langle k \rangle, \langle k_2 \rangle, \langle k_3 \rangle, \Delta k$ e γ , considerando os grafos obtidos variando-se o expoente γ . O código de cores é o mesmo usado nas matrizes anteriores similares.	45
4.8	Matriz de correlação das métricas C_g, C_p, C_i, C_a e $\langle c \rangle$, em função de γ para os 100 nós mais representativos. Como nas outras matrizes de correlação, azul mais intenso denota correlação mais próxima de 1; vermelho mais intenso denota correlação mais próxima de -1	46
4.9	Distribuição de graus $p(k)$ em um gráfico log-log para $N = 4900$ e $\langle k \rangle_{alvo} = 20$. Para $\gamma = 2$, geram-se grafos com $r = \{5, 10, 20, 30, 40, 60\}$; para $\gamma = 3$, geram-se grafos com $r = \{5, 20, 60\}$. As escalas horizontais não são iguais nesses nove gráficos, a fim de facilitar a visualização das distribuições mostradas.	47

4.10	Matriz de correlação das métricas $C_g, C_p, C_i, C_a, \langle \ell \rangle, \langle c \rangle, \langle k \rangle, \langle k_2 \rangle, \langle k_3 \rangle, r$ e γ , considerando os grafos obtidos variando-se o raio r e o expoente γ . Como antes, azul mais intenso indica correlação mais próxima de 1, vermelho mais intenso indica correlação mais próxima de -1	49
4.11	Matriz de correlação das métricas C_g, C_p, C_i, C_a e $\langle c \rangle$ em função de r e γ . O código de cores é o mesmo até então usado.	50
4.12	Sumário de tendências das métricas $C_g, C_p, C_i, C_a, \langle \ell \rangle, \langle c \rangle, \langle k_2 \rangle, \langle k_3 \rangle$ e Δk em função de $\langle k \rangle_{alvo}, \gamma$ e r , considerando todos os nós do grafo.	50
4.13	Sumário de tendências das métricas C_g, C_p, C_i, C_a e $\langle c \rangle$ em função de $\langle k \rangle_{alvo}, \gamma$ e r , considerando os 100 nós mais relevantes.	51
6.1	Figura ilustrativa das ligações entre os indivíduos em cada país. Os reticulados gerados pelo algoritmo 5.1 são convertidos em gráficos tridimensionais, com os nós distribuídos em três planos horizontais para facilitar a visualização das três classes e suas conexões. Os eixos x e y representam a posição única de cada indivíduo no reticulado, cujos números foram deixados da forma proposital em fonte menor. Os nós da cor vermelha representam os indivíduos da classe alta (A) e estão alocados no plano $z = 3$. Os nós da cor verde representam os indivíduos da classe média (B) e estão alocados no plano $z = 2$. Já os nós da cor azul representam os indivíduos da classe baixa (C) e estão alocados no plano $z = 1$. As arestas entre os indivíduos foram desenhadas na cor cinza.	60
6.2	Distribuição de graus dos grafos de cada país numa escala-log-log (base 10). Os pontos em preto são os dados obtidos diretamente dos grafos. A função ajustada $A_0 k^{-\gamma_0}$ corresponde à linha vermelha e ela é um bom ajuste para $k_{min} \leq k \leq 2\langle k \rangle_{alvo}$. A função ajustada $A_1 k^{-\gamma_1} 10^{\delta k^{10}}$ corresponde à linha azul para $k_{min} \leq k \leq k_{max}$	65

LISTA DE TABELAS

2.1	Divisão dos indivíduos por classe social (em porcentagem).	11
2.2	Centralidades de intermediação do subgrafo maior da figura 2.4.	24
2.3	Centralidades de proximidade do subgrafo maior da figura 2.4.	25
2.4	Centralidades de autovetor do grafo da figura 2.4.	26
2.5	Centralidades de grau do grafo da figura 2.4.	26
2.6	Algumas propriedades topológicas dos modelos de redes mais famosos. Adaptado de Schimit (2010) e Monteiro (2014).	27
4.1	Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i , de autovetor C_a , comprimento médio do caminho mais curto $\langle \ell \rangle$, coeficiente de clusterização médio $\langle c \rangle$ e grau médio $\langle k \rangle$ em função de $\langle k \rangle_{alvo}$	35
4.2	Momentos de segunda $\langle k_2 \rangle$ e de terceira ordem $\langle k_3 \rangle$, desvio padrão de uma distribuição de Poisson $\langle k \rangle^{1/2}$, $\sigma^{1/2}$ e Δk dos grafos gerados, em função de $\langle k \rangle_{alvo}$	39
4.3	Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a e coeficiente de clusterização médio $\langle c \rangle$ em função de $\langle k \rangle$ para os 100 nós mais relevantes.	41
4.4	Centralidades de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a , comprimento médio do caminho mais curto $\langle \ell \rangle$, coeficiente de clusterização médio $\langle c \rangle$ e grau médio $\langle k \rangle$ em função do expoente γ	43
4.5	Momentos de segunda e terceira ordens, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, desvio padrão de uma distribuição de Poisson $\langle k \rangle^{1/2}$, variância $\sigma^{1/2}$ e Δk dos grafos gerados.	44
4.6	Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a e coeficiente de clusterização médio $\langle c \rangle$ em função de γ para os 100 nós mais relevantes.	46
4.7	Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a , comprimento médio do caminho mais curto $\langle \ell \rangle$, coeficiente de clusterização médio $\langle c \rangle$ e grau médio $\langle k \rangle$ em função de γ e r	47
4.8	Momentos de segunda e terceira ordens, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, desvio padrão de uma distribuição de Poisson $\langle k \rangle^{1/2}$, variância $\sigma^{1/2}$ e Δk dos grafos gerados.	48
4.9	Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a e coeficiente de clusterização médio $\langle c \rangle$ em função de r e γ para os 100 nós mais relevantes.	49
6.1	Média diária de contatos por indivíduo para os países estudados.	57

6.2	Valores dos parâmetros utilizados na geração dos grafos de contatos sociais para França, Zimbábue e Peru.	59
6.3	Valores médios de $\langle k \rangle$, $\langle k_A \rangle$, $\langle k_B \rangle$, $\langle k_C \rangle$, $\langle \ell \rangle$, $\langle c \rangle$, C_g , C_p , C_i e C_a e os respectivos desvios padrões.	61
6.4	Porcentagem das classes sociais considerando os 100 maiores valores de $\langle k \rangle$, $\langle c \rangle$, C_p , C_i e C_a para os três países.	63
6.5	Valores da tabela 6.4 divididos pA , pB e pC . Chama-se essa divisão de razão normalizada.	63
6.6	Valores de A_0 , γ_0 e do erro quadrático médio (EQM) da regressão linear de $p(k)$ para os três países simulados.	66
6.7	Valores de k_{min} , $\langle k \rangle$, k_{max} , k_{cut} e Δk para os três países simulados.	66
6.8	Valores de A_1 , γ_1 , δ e do EQM da regressão não linear para os três países.	66

SUMÁRIO

1	Introdução	1
2	Referencial teórico	6
2.1	Introdução histórica sobre a prática científica: deducionismo, reducionismo, complexidade	6
2.2	Sistemas complexos	9
2.3	Classes socioeconômicas e medidas de desigualdade	10
2.3.1	IDH	11
2.3.2	Coefficiente de Gini	12
2.4	Vizinhanças de Von Neumann e de Moore	13
2.5	Grafos	15
2.5.1	Redes regulares	17
2.5.2	Redes completamente aleatórias	17
2.5.3	Redes mundo-pequeno	19
2.5.4	Redes livres-de-escala	21
2.5.5	Métricas topológicas	22
3	Um modelo de rede de contatos sociais	29
3.1	O algoritmo	30
4	Simulação do algoritmo de contatos sociais	34
4.1	Variação do grau desejado $\langle k \rangle_{alvo}$	35
4.1.1	Considerando os 100 nós mais relevantes	40
4.2	Variação do expoente γ	41
4.2.1	Considerando os 100 nós mais relevantes	45
4.3	Variação do raio de vizinhança r	45
4.3.1	Considerando os 100 nós mais relevantes	49
4.4	Resumo dos resultados	50
4.5	Características do modelo proposto para gerar redes complexas	51
5	Redes sociais com classes socioeconômicas	53
5.1	Incluindo classes sociais	54
5.2	Modelo de rede social livre-de-escala	55

6	Grafos de contatos sociais para três países	57
6.1	Estudos de campo sobre contatos sociais	57
6.2	Considerações sobre a matriz de pesos	58
6.3	Resultados para os três países	59
6.4	Distribuição de graus dos três países	64
6.5	Discussão dos resultados	66
7	Conclusão	68
	Referências	70

1 INTRODUÇÃO

As redes sociais humanas são estruturas topológicas intrincadas que representam os contatos sociais presenciais e virtuais entre indivíduos (Barabasi, 2016; Mitzenmacher, 2004; Vytla et al., 2021). Por estrutura intrincada entende-se como algo que é complexo no sentido de ser rico em detalhes (Heylighen, 1988). Usualmente, tais estruturas não são bem compreendidas isolando suas partes, como em experimentos tipicamente realizados em laboratório por reducionismo (Cencini et al., 2021; Góes, Guérios e Góes, 2019; Morin, 2014). Estudar redes sociais sob a ótica da teoria da complexidade pode aumentar nossa compreensão sobre a emergência de padrões em grafos, sobre a dinâmica de propagação de infecções e sobre o fluxo de informações em sociedades (Boucher, 2015; Landherr, Friedl e Heidemann, 2010; Monteiro, 2014; Wellman, Quan-Haase e Harper, 2020).

A ciência da complexidade abrange a biologia, a computação, a engenharia, a física, a matemática, a sociologia. Ela se ocupa, dentre outros assuntos, do estudo dos modelos de redes e de suas propriedades topológicas (Góes, Guérios e Góes, 2019; Prigogine e Stengers, 1996; Vasconcellos, Rodrigues e Luzzi, 2015). Em geral, estruturas complexas como as redes sociais são representadas por meio da grafos, cujos nós (ou vértices) são os indivíduos e as arestas (ou conexões) simbolizam os contatos que ocorrem entre tais indivíduos (Albert e Barabási, 2002; Boccaletti et al., 2006; Monteiro, 2014; Newman, 2003; Strogatz, 2001).

No presente trabalho, propõe-se um modelo de redes de contatos sociais presenciais, que são construídas por meio de algoritmos computacionais. O modelo proposto apresenta características de redes livres-de-escala e de redes mundo-pequeno (Barabási e Albert, 1999; Newman, 2003; Watts e Strogatz, 1998).

Redes livres-de-escala são estruturas em que a distribuição de conexões entre os indivíduos segue uma lei de potência, caracterizada pela equação $p(k) = Ak^{-\gamma}$, em que $p(k)$ representa a porcentagem de indivíduos com k conexões, e A e γ são constantes positivas (Barabasi, 2016; Monteiro, 2014). Esse padrão implica que, embora a maioria dos indivíduos tenha poucas conexões, uma pequena fração deles, conhecidos como *hubs*, possui um número muito elevado de conexões, excedendo significativamente a média. Esses *hubs* são cruciais na rede, atuando como pontos centrais na disseminação de informações ou de doenças contagiosas (Clauset, Shalizi e Newman, 2009; Newman, 2003). Dois exemplos clássicos de redes livre-de-escala que podem ser citados são os grafos que representam

e-mails trocados (Ebel, Mielsch e Bornholdt, 2002), úteis em estudos sobre difusão de ideias, e os grafos que expressam os contatos sexuais (Liljeros et al., 2001), que são relevantes no estudo de infecções sexualmente transmissíveis. Exemplos mais atuais incluem sistemas de reputação baseados em opinião em aplicações de *e-commerce* (Urena et al., 2019) e a identificação de influenciadores em plataformas digitais (Wen e Deng, 2020).

Redes mundo-pequeno são aquelas que apresentam grande conectividade local (Watts e Strogatz, 1998; Monteiro, 2014), de modo que muitos vizinhos de um nó estão conectados entre si. Essa propriedade se reflete num coeficiente de clusterização alto. Um exemplo clássico é a rede de atores de Hollywood, na qual dois atores estão conectados se atuaram juntos num mesmo filme (Watts e Strogatz, 1998). Nessa rede, o coeficiente de clusterização é em torno de 0,8; isto é, em média, 80% dos vizinhos de um nó estão diretamente conectados (em outras palavras, 80% dos atores que trabalharam com um certo ator A em diversos filmes também trabalharam juntos em outros filmes, sem a participação de A).

Tanto redes livres-de-escala como redes mundo-pequeno apresentam o efeito mundo pequeno, segundo o qual a distância entre dois nós quaisquer é muito menor do que o número de nós. Isso decorre do fato de que, na construção dessas redes, há um certo nível de aleatoriedade na criação das conexões entre os nós (Barabási e Albert, 1999; Watts e Strogatz, 1998). Essa aleatoriedade cria atalhos no grafo, encurtando distâncias. Esse efeito ocorre em redes sociais reais e tem sido bastante estudado (Barabasi, 2016; Chen, Zhang e Huang, 2007; Mullick e Sen, 2020; Reggiani e Vinciguerra, 2007; Travers e Milgram, 1969). O nome desse efeito deriva da expressão “como este mundo é pequeno” normalmente dita quando dois desconhecidos descobrem, numa conversa, que têm um amigo em comum.

Grafos de redes sociais também têm sido construídos a partir da integração de dados acadêmicos em grande escala de múltiplas fontes (Zhou et al., 2018), do fluxo turístico internacional (Chung et al., 2020), dos casos de infectados por COVID-19 (Kwan et al., 2021). Há também modelos de redes sociais com múltiplas camadas (Fan et al., 2022) e com conexões dinâmicas (Hoffman et al., 2020).

Para estudos teóricos sobre dinâmica social, é relevante dispor de um modelo matemático capaz de gerar grafos que reproduzam o padrão de conexões encontrado numa rede social real, que se reflete, por exemplo, no valor do grau médio $\langle k \rangle$ e na distribuição de graus $p(k)$. Nesta tese, apresenta-se um algoritmo capaz de gerar grafos com o grau médio desejado e com a distribuição de grau desejada. Os modelos clássicos da literatura, que serviram de inspiração, são também aqui apresentados (Barabasi, 2016; Barabási e Oltvai, 2004; Chen, Zhang e Huang, 2007; Fortunato, 2010; Newman, 2003; Watts, 2004;

Wen e Deng, 2020) e as novidades do modelo proposto são destacadas.

Apesar de haver diversos modelos de redes sociais na literatura, esses modelos não consideram as classes socioeconômicas de seus indivíduos (Boucher, 2015; Centola, 2010; Landherr, Friedl e Heidemann, 2010; Newman, 2010; Wang e Chen, 2003). Além disso, não é usual ajustar tais modelos (Barabasi, 2016; Boccaletti et al., 2006; Clauset, Shalizi e Newman, 2009; Mitzenmacher, 2004; Newman, 2003; Wang e Chen, 2003) para que exibam um determinado $p(k)$ ou que tenham um dado valor do expoente γ , no caso em que $p(k) = Ak^{-\gamma}$, que podem ser característicos de um determinado país.

Estudar redes sociais complexas com enfoque em classes sociais pode ser fundamental para se compreender a dinâmica e as estruturas subjacentes que moldam as interações humanas em diversas sociedades (Voutsas et al., 2021). Uma classe social pode ser definida como um grupo de indivíduos que possuem posição socioeconômica semelhante (Grusky, 2019; Holst, Fessler e Niehoff, 2021; Lenski, 2013; Levine, 2006). O estudo de redes com classes sociais permite evidenciar e compreender aspectos atrelados às desigualdades econômicas. Além disso, pode-se analisar como uma classe influencia outra, afetando a propagação de informações e de ideias na rede (Holst, Fessler e Niehoff, 2021; Grusky, 2019; Levine, 2006).

Em geral, as pessoas interagem principalmente com familiares, amigos e vizinhos, que geralmente pertencem à mesma classe social. Essa propriedade da rede é chamada de homofilia (Chodrow, 2017; McPherson, Smith-Lovin e Cook, 2001; Xu et al., 2019), pela qual nós com características similares (no caso, condição financeira semelhante) tendem a estar mais conectados do que nós com características diferentes. O impacto na qualidade de vida das classes socioeconômicas em diferentes países tem sido avaliado por meio do Índice de Desenvolvimento Humano (IDH) da ONU e do coeficiente de Gini, e ambos têm implicações nas relações sociais. Essas duas medidas são levadas em conta nesta tese.

Evidentemente, encontros presenciais entre indivíduos pertencentes a diferentes classes também ocorrem, por exemplo, no local de trabalho, num parque, num metrô, num shopping center. Interações entre classes sociais distintas ou não também podem ocorrer em ambientes virtuais (Boucher, 2015; Noble e Tynes, 2016; Xu et al., 2019); entretanto, elas não são aqui consideradas.

Há poucos estudos de campo que determinam como os contatos entre classes sociais ocorrem. Tais estudos são usados, por exemplo, para a criação de dados artificiais para mais de 150 países, a fim de viabilizar investigações apoiadas em inteligência artificial, para diversos fins (Prem et al., 2021). É certo que a relação entre interações sociais e classes sociais carece de melhor entendimento. Nesse contexto, cabe citar um artigo de revisão (Vytla et al., 2021) que analisa estudos publicados com base em dados da pandemia de

COVID-19. Nesse artigo, enfatiza-se a imprecisão e a pouca maturidade de algoritmos de *machine-learning* para explorar aspectos complexos das redes sociais e da dinâmica de propagação de doenças, oferecendo resultados de pouca valia. Uma fonte de problemas pode ser os modelos consagrados de redes complexas mencionados acima, empregados em estudos de redes sociais, mas que nem sempre representam bem a realidade estudada.

É nesse cenário de pesquisa que se encaixa o modelo de contatos sociais proposto. No modelo, incorpora-se a existência de diferentes classes socioeconômicas. Além disso, ele exhibe o efeito mundo pequeno (pois, numa rede com N nós, o menor caminho médio $\langle \ell \rangle$ entre dois nós é tal que $\langle \ell \rangle \ll N$) e pode apresentar a distribuição de graus $p(k)$ e o grau médio $\langle k \rangle$ desejados. Aqui, considera-se que a rede é livre-de-escala; entretanto, essa não é uma limitação do modelo proposto. Dados reais sobre características sociais e econômicas obtidos em estudos de campo são levados em conta, considerando um país africano (Zimbábue) (Melegaro et al., 2017), um país europeu (França) (Béraud et al., 2015) e um país sul-americano (Peru) (Grijalva et al., 2015). Com esses dados reais e utilizando o modelo proposto, grafos representativos dos três países escolhidos são construídos computacionalmente. Então, calculam-se e comparam-se diversas métricas topológicas, como centralidade de grau, centralidade de proximidade, centralidade de intermediação, centralidade de autovetor, menor caminho médio, coeficiente de clusterização médio. A partir desses resultados, discute-se como a topologia da rede influencia a propagação de informações em cada país. Em resumo, esta tese tem dois objetivos:

- objetivo secundário: desenvolver um algoritmo capaz de gerar um grafo com $\langle k \rangle$ e $p(k)$ desejados;
- objetivo principal: adaptar e aplicar esse algoritmo para criar grafos adequados para representar redes sociais com várias classes econômicas.

O algoritmo aqui proposto tem como ponto de partida aquele elaborado por Monteiro, Paiva e Piqueira (2006) e por Schimit e Monteiro (2009), em estudos de depressão alastrante e de epidemiologia, respectivamente. Nessas referências, entretanto, $\langle k \rangle$ e $p(k)$ não são definidos de antemão e não se consideram que os nós que compõem a rede podem ser de tipos (por exemplo, de classes sociais) diferentes.

Este trabalho encontra-se dividido da seguinte forma. No capítulo 2, são apresentados conceitos básicos de grafos, de IDH, de coeficiente de Gini e de correlação de dados. No capítulo 3, descreve-se o modelo de rede complexa proposto e apresenta-se o algoritmo empregado em sua implementação computacional. No capítulo 4, analisam-se três séries de simulações realizadas com o intuito de compreender como as características topológicas dos grafos gerados dependem dos parâmetros do modelo. No capítulo 5, o algoritmo proposto no capítulo 3 é adaptado a fim de se levar em conta a existência de

diferentes classes socioeconômicas. Nessa adaptação, usa-se uma matriz \mathbf{f} que afeta a probabilidade de contato entre indivíduos de diferentes classes socioeconômicas. No capítulo 6, são mostrados os resultados relativos aos três países citados. No capítulo 7, trata-se das conclusões e das considerações finais desta tese.

Dois artigos foram elaborados a partir do desenvolvimento desta tese:

- “*A network model of social contacts with small-world and scale-free features, tunable connectivity, and geographic restrictions*” (Licciardi e Monteiro, 2024);
- “*A complex network model for a society with socioeconomic classes*” (Licciardi e Monteiro, 2022).

2 REFERENCIAL TEÓRICO

“Nos complexos entrelaçamentos dos grafos sociais, cada conexão é um potencial caminho para uma nova descoberta.”

-- Anônimo

Neste capítulo, reúnem-se os principais conceitos teóricos utilizados no estudo de redes complexas. Na seção 2.1, faz-se uma ligeira introdução histórica a respeito da prática científica, chegando até o desenvolvimento da ciência da complexidade. Exemplos de sistemas complexos, em particular, de redes sociais, são abordados na seção 2.2. Na seção 2.3, trata-se de classes socioeconômicas, Índice de Desenvolvimento Humano (IDH) da ONU e coeficiente de Gini. Na seção 2.4, descrevem-se maneiras de definir vizinhança em uma rede bidimensional. Como indivíduos e suas relações sociais podem ser representados por grafos, apresentam-se, na seção 2.5, as métricas habitualmente empregadas para caracterizar grafos e que são normalmente calculadas em estudos de redes complexas. Nessa seção, apresenta-se também uma medida de correlação, que é usada principalmente no capítulo 4 desta tese.

2.1 Introdução histórica sobre a prática científica: deducionismo, reducionismo, complexidade

O entendimento sobre a prática das ciências pode ser resumido de maneira simplista pelo seguinte breve retrospecto histórico. Um personagem importante no início da história das ciências é Aristóteles (384-322 a.C.), filósofo clássico que foi discípulo de Platão e tutor de Alexandre, o Grande (Grant, 2006). As incursões dedutivas de Aristóteles abrangeram diversas áreas do conhecimento. A ciência aristotélica era de cunho qualitativo e dedutivo, apoiada em “verdades evidentes”. Ela influenciou os ensinamentos em escolas, principalmente até a Idade Média. Sua abordagem alcançou destaque, por exemplo, na Astronomia e na Física, ao buscar respostas sobre o funcionamento do universo.

Após a Idade Média, aumentaram os registros que contrariavam a forma aristotélica de pensar (Grant, 2006). É relevante mencionar as contribuições de Galileu Galilei (1564-1642), cujos apontamentos, experimentos e interpretações da realidade que observava criaram um novo paradigma sobre o estudo e a prática das ciências, conhecido como rea-

lismo científico (Naess, 2015). Famoso crítico das traduções de textos científicos antigos, Koyré (1943) afirma que Galileu teve sua iniciação no deducionismo e que seus experimentos científicos seriam inicialmente preparados apenas para endossar suas induções. Koyré, porém, destaca a nova forma de fazer ciência que Galileu havia propiciado. Talvez, por isso, Galileu seja mencionado em textos históricos como o “pai da Ciência Moderna”. A partir de Galileu, a Física passou a ser compreendida como uma ciência baseada em experimentações mensuráveis, passíveis de serem traduzidas em leis matemáticas.

Contribuições essenciais foram também realizadas por Isaac Newton (1642-1727), conhecido por ter realizado trabalhos fundamentais na Física e na Matemática, como seus estudos para a compreensão da dinâmica celeste. Seus textos consolidaram a prática do realismo científico (Patomäki, 2021), que pode ser assim resumido: a partir de observações de um fenômeno, formulam-se questionamentos; em seguida, elaboram-se experimentos que levam a respostas para as questões formuladas. Na obra de Newton, bem como nas obras de outros cientistas da época, passou a ser notório o princípio de análise científica, que parte do recorte de um certo problema identificado no ambiente, o que pode ser chamado de objeto de investigação. Sobre esse objeto, aplicam-se métodos a fim de desvendar o comportamento e o funcionamento de suas partes componentes. Isso permite a análise de suas partes, visando a descrição do funcionamento do objeto como um todo (Fenzl e Machado, 2009). Essa maneira de investigar fenômenos em ciências é conhecida como reducionismo.

No final do século XIX, a fim de celebrar os 60 anos do rei Oscar II (1829-1907), da Suécia e Dinamarca, foi oferecido um prêmio para a resolução de diversos problemas, sendo que um deles tratava da estabilidade do sistema solar (Barrow-Green, 1994). O matemático Henri Poincaré (1854-1912) se propôs a enfrentar esse desafio; porém, ao perceber a dificuldade de realizar o tratamento direto desse problema, reduziu-o ao estudo de apenas três corpos interagindo pela gravidade. Com esse trabalho, Poincaré recebeu o prêmio em 1889, no qual estabelece a impossibilidade de se precisar o comportamento exato desse sistema, pois sua evolução temporal pode ser drasticamente afetada pelas condições iniciais de posição e velocidade dos três corpos. Poincaré foi o primeiro matemático a evidenciar a possibilidade de comportamento caótico em um sistema determinista. Cabe lembrar que determinismo é uma teoria filosófica em que os acontecimentos são explicados por relações de causalidade (Earman, 2004).

Para que um sistema de tempo contínuo possa exibir comportamento caótico, no sentido estabelecido por Poincaré, é necessário que (Cencini et al., 2021; Monteiro, 2019): (1) o sistema deve não ser linear, o que implica que sua resposta não é proporcional à entrada; (2) ele é sensível a perturbações mínimas em seu estado; ou seja, uma alteração

mínima no presente pode causar, com o passar do tempo, mudanças no futuro que são imprevisíveis na prática; (3) ele deve ser ao menos de terceira ordem; ou seja, descrito por três variáveis de estado.

O que se conhece hoje como Teoria do Caos é o resultado de diversos esforços que ocorreram no final do século XIX e ao longo do século XX. Um trabalho relevante do meteorologista Edward Lorenz (1917-2008) foi publicado em 1963, no qual se analisam sistemas dinâmicos deterministas e não lineares ligados à circulação atmosférica (Lorenz, 1963). Por meio de simulações computacionais, Lorenz observou comportamentos complexos, chamados posteriormente de caóticos (Monteiro, 2019; Williams, 1997). Em sistemas meteorológicos, pequenas mudanças nas condições iniciais que podem resultar em alterações significativas no comportamento do sistema ficaram popularmente conhecidas como “efeito borboleta”. Caos é considerado um comportamento complexo.

A Ciência da Complexidade surgiu como um campo transdisciplinar de estudo, tentando compreender a natureza dos sistemas complexos. Não há definição única para sistemas complexos. Uma definição possível é: um sistema complexo é aquele composto por muitas partes interconectadas e interdependentes, cuja estrutura e dinâmica são influenciadas pelas interações entre essas partes de modo não trivial (Góes, Guérios e Góes, 2019). A origem e o desenvolvimento da Ciência da Complexidade são motivados por uma confluência de ideias e de técnicas provenientes de diversas disciplinas, incluindo biologia, ciências sociais, computação, engenharia, física, matemática. Vários autores podem ser citados por suas contribuições. Dentre eles, destacam-se Edgar Morin, Ilya Prigogine e Isabelle Stengers (Góes, Guérios e Góes, 2019; Prigogine e Stengers, 1996; Vasconcellos, Rodrigues e Luzzi, 2015). Em particular e adicionalmente, é interessante mencionar o entendimento de complexidade de Heylighen (1988):

Voltemos à palavra latina original “complexus”, que significa “entrelaçado”, “torcido”, “junto”. Isso pode ser interpretado da seguinte forma: para ter um complexo, precisa-se de: 1) duas ou mais partes ou elementos diferentes; 2) essas partes devem de alguma forma ser conectadas entre si, de forma que seja difícil separá-las. Aqui, encontramos a dualidade básica entre as partes que são ao mesmo tempo distintas e conectadas. É claro que um complexo não pode ser analisado ou separado em um conjunto de elementos independentes sem destruí-lo. Portanto, devemos concluir que o método reducionista não é adequado para ser usado na compreensão de entidades complexas. Isso explica a conotação de difícil, que um sistema complexo recebeu em períodos posteriores. Na verdade, o método racional geral de lidar com problemas consiste em analisar o domínio do problema. Se esse domínio for complexo, então, por definição, ele resistirá à análise” (tradução livre).

Note que esse autor parte da origem da palavra para buscar uma definição para sistema complexo. Na seção a seguir, são listados alguns exemplos típicos de sistemas complexos.

2.2 Sistemas complexos

Quando se busca por exemplos de sistemas complexos, de acordo com os conceitos abordados na seção anterior, alguns comumente mencionados são: uma rede de transporte integrada em uma cidade (ônibus, metrô, trem), uma rede de telefonia de diversas operadoras, uma rede neural biológica (Monteiro, 2014). Nesses exemplos, as partes constituintes talvez possam ser descritas por relações simples; porém, o conjunto possui comportamentos próprios e intrincados, que emergem das relações entre seus componentes, de modo que a análise das partes separadas (reducionismo) não é capaz de levar à compreensão da evolução do conjunto (Williams, 1997).

Existem diversos sistemas complexos envolvendo humanos, nos quais as interações entre os indivíduos e as características emergentes desses sistemas são de difícil previsão com base apenas nas propriedades dos indivíduos isolados. Alguns exemplos são:

- **Redes sociais:** as redes sociais são sistemas complexos compostos por indivíduos interconectados por meio de relações sociais, decorrentes, por exemplo, de amizade ou parentesco. Essas redes podem ser estudadas usando conceitos da teoria de grafos e podem exibir propriedades emergentes (Albert e Barabási, 2002; Watts, 2004), como a formação de comunidades e a polarização de opiniões.
- **Ecosistemas:** os ecossistemas são sistemas complexos que envolvem interações entre organismos vivos, como plantas, animais, microrganismos e seu ambiente físico. O estudo da ecologia de comunidades e das redes tróficas é um exemplo de como a estrutura e as interações nesses sistemas podem levar a comportamentos imprevisíveis (May, 2019; Pimm, 1982). Por exemplo, é difícil prever as consequências de se introduzir uma nova espécie num ecossistema ou as consequências de uma espécie ser extinta.
- **Mercados financeiros:** os mercados financeiros são sistemas complexos que envolvem interações entre investidores, instituições financeiras e ativos financeiros. Esses sistemas podem exibir comportamentos coletivos, como flutuações de preços e crises financeiras, que não podem ser previstas com base no comportamento individual dos participantes (Farmer e Foley, 2009; Mandelbrot, 1963).

Neste trabalho, propõe-se e analisa-se um modelo de rede social complexa que considera classes socioeconômicas e relações de vizinhança entre seus indivíduos (Licciardi e Monteiro, 2022). Na próxima seção, apresenta-se o conceito de classe socioeconômica utilizado e, na seção 2.4, explicita-se o conceito de vizinhança adotado no modelo.

2.3 Classes socioeconômicas e medidas de desigualdade

O conceito de classes socioeconômicas mudou ao longo do tempo, passando por diferentes abordagens e interpretações. Inicialmente, o estudo das classes sociais tinha uma perspectiva mais estritamente econômica, baseada na posição ocupacional e na renda das pessoas. No entanto, ao longo das décadas, o conceito foi ampliado e passou a incorporar outras dimensões, como educação, status social, estilo de vida, poder político (Lynch e Kaplan, 2000).

Uma das principais contribuições para o desenvolvimento do conceito de classes sociais foi feita por Karl Marx (1818-1883) e Friedrich Engels (1820-1895) em suas obras clássicas. Eles enfatizaram a importância das relações de produção e da propriedade dos meios de produção no surgimento das classes sociais (Marx e Engels, 2010). A partir desse ponto de partida, outros teóricos sociais expandiram e refinaram esse conceito.

Pierre Bourdieu (1930-2002) desenvolveu o conceito de “capital social” que engloba diferentes formas de recursos sociais, econômicos e culturais que afetam a posição e as oportunidades das pessoas na sociedade. Esses recursos podem incluir o nível de educação, o acesso à rede de contatos, o conhecimento cultural, os padrões de consumo. Bourdieu destacou a importância do capital social no estabelecimento de desigualdades sociais (Bourdieu, 1986).

Outra abordagem importante foi proposta por Max Weber (1864-1920), que enfatizou a influência do status social e do poder na estratificação social. Weber considerava que a classe social não se baseava apenas na posição econômica. Em sua obra, ele identificou diferentes tipos de estratificação, como a classe econômica, o status social e o poder político, que interagem e influenciam a posição e o prestígio das pessoas na sociedade (Lynch e Kaplan, 2000; Weber, 2003).

Na sociologia moderna (Wright, 2006), classe social se refere à divisão socioeconômica, para o mundo capitalista. Entre os grupos existe uma hierarquia de importância de funções e de divisão de trabalhos, o que é conhecido como o fenômeno da estratificação social. Dentro dessa perspectiva, de acordo com o Programa das Nações Unidas para o Desenvolvimento (UNDP, 2022), a Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2019) e a PewResearch (2021), há um senso comum em categorizar a classe social de um indivíduo de acordo com a média líquida de proventos que os ocupantes do seu lar recebem mensalmente. Assim, os indivíduos são divididos em cinco faixas: pobres, baixa renda, média renda, média renda alta, alta renda. Para efeito das simulações do algoritmo computacional e das análises realizadas no capítulo 6, considera-se como

Classe C os indivíduos pobres ou de baixa renda, como Classe B os indivíduos de renda média ou média-alta e como Classe A os de alta renda. Com base nessa classificação, constrói-se a tabela 2.1, que apresenta a porcentagem dos indivíduos acordo com a classe social, para os países França, Zimbábue e Peru.

Tabela 2.1. Divisão dos indivíduos por classe social (em porcentagem).

País	Classe A	Classe B	Classe C
França	26	60	14
Zimbábue	0,5	8	91,5
Peru	4	42	54

Fonte: UNDP (2022), OCDE (2019) e PewResearch (2021).

Outro conceito sociológico relevante atrelado aos relacionamento entre indivíduos em classes sociais é o de homofilia. Tal conceito influencia o estabelecimento de diversos tipos de relações nas redes sociais, como trabalho, amizade, casamento e transferência de informações (McPherson, Smith-Lovin e Cook, 2001). A consequência é que similaridades entre indivíduos, como suas características socioeconômicas, favorecem a criação de laços sociais. A homofilia tem sido investigada analisando, por exemplo, dados de telefones celulares em Cingapura (Xu et al., 2019), a segregação étnico-racial em Detroit (Chodrow, 2017), padrões de amizade em escolas americanas de ensino médio (Boucher, 2015), a dimensão sociocultural nas áreas urbanas da Holanda (Gent, Das e Musterd, 2019). Nessas análises, porém, as medidas estatísticas básicas comumente utilizadas para caracterizar a estrutura topológica de um rede social não foram computadas. Para o modelo proposto nesta tese, essas medidas (definidas na seção 2.5) são calculadas e analisadas no capítulo 6 para os três países considerados.

Para realizar uma comparação de classes socioeconômicas de diferentes países, é interessante a compreensão de dois índices: o IDH, Índice de Desenvolvimento Humano, e o coeficiente de Gini.

2.3.1 IDH

O IDH foi desenvolvido como uma alternativa aos métodos convencionais de medir a riqueza de um país apenas pelo seu Produto Interno Bruto (PIB). Introduzido em 1990 pelos economistas Mahbub ul Haq e Amartya Sen, o IDH é uma iniciativa do Programa das Nações Unidas para o Desenvolvimento (PNUD) (Haq, 1995).

No IDH, há três dimensões:

1. Saúde: a expectativa de vida ao nascer, por exemplo, em países com sistemas de saúde robustos, como Japão e Suíça, ultrapassa os 80 anos, enquanto que em nações

afetadas por conflitos ou epidemias, como algumas na África Subsaariana, pode ser inferior a 60 anos (Alsalem et al., 2020).

2. Educação:

- Expectativa de anos de escolaridade: em países nórdicos como a Finlândia, as crianças costumam passar 20 anos na escola. Em contraste, em países com acesso limitado à educação de boa qualidade, esse número pode ser drasticamente menor (UNESCO, 2020).
 - Média de anos de estudo para a população adulta: países com histórico de educação universalizada, como a Alemanha, têm uma média elevada, enquanto que países em desenvolvimento ou com grandes desafios educacionais apresentam médias mais baixas (Will e Homuth, 2020).
3. Padrão de vida: o PIB per capita ajustado pela paridade do poder de compra (PPP) pode variar amplamente, desde valores elevados de nações como os EUA ou Noruega até valores muito mais baixos de países em desenvolvimento (DIMFR, 2021).

Após calcular os índices individuais para cada uma dessas três dimensões, a média harmônica é usada para determinar o IDH geral. Essa abordagem evita que o progresso excepcional em uma dimensão compense a falta de progresso nas outras duas (Klugman, Rodríguez e Choi, 2011; Larsen e Marx, 2000).

O IDH é uma medida que permite visualizar e comparar o desenvolvimento humano, mas pode ser usado como ferramenta de política social, incentivando países a priorizar o bem-estar humano. Por exemplo, na década de 1990, Botsuana, apesar de ter um PIB crescente devido à sua rica indústria de diamantes, passou a ter uma expectativa de vida em declínio por causa da epidemia de HIV/AIDS. Isso se refletiu no seu IDH, evidenciando a necessidade de intervenções específicas (Leckie, 2000). Cabe mencionar que existem críticas ao IDH. Um dos argumentos principais é que ele é um índice simplista, incapaz de aferir as desigualdades internas dos países. A fim de contornar essa limitação, o PNUD introduziu o IDH ajustado à desigualdade (Harttgen e Klasen, 2012).

O IDH fornece um quadro útil para avaliar e comparar o bem-estar humano além das métricas econômicas tradicionais. No presente trabalho, o IDH é utilizado para estabelecer um fator relativo de relacionamento de classes socioeconômicas entre os diferentes países analisados. Isso tornou possível as simulações apresentadas no capítulo 6.

2.3.2 Coeficiente de Gini

O coeficiente de Gini é medida estatística usada para determinar a desigualdade da distribuição de renda ou da riqueza dentro de uma população. Introduzido pelo estatístico

Corrado Gini (1884-1965) em 1912 (Gini, 1912), tornou-se uma das ferramentas mais usadas para analisar a desigualdade econômica.

Esse coeficiente é expresso como um número entre 0 e 1 (ou entre 0 e 100%). O caso em que vale zero representa perfeita igualdade (todos têm a mesma renda), enquanto que o caso em que vale um indica perfeita desigualdade (uma pessoa apenas possui toda a renda/riqueza e os demais têm nada) (Gujarati e Porter, 2017).

Para se calcular o coeficiente de Gini, utiliza-se frequentemente a curva de Lorenz. Essa curva descreve a porcentagem acumulada da renda total da população em função da porcentagem acumulada da população, ordenada do mais pobre ao mais rico. A curva de Lorenz é comparada à linha reta de igualdade perfeita (com inclinação de 45°), e o coeficiente de Gini é calculado como a área entre a curva de Lorenz e essa linha de igualdade, dividida pela área total sob a linha de igualdade (Yitzhaki, 1994).

O coeficiente de Gini é usado por diversas instituições, como o Banco Mundial e a OCDE, para monitorar a evolução da desigualdade de renda ao longo do tempo e comparar os níveis de desigualdade entre diferentes países (WorldBank, 2017). Apesar de fornecer uma visão geral da desigualdade, esse coeficiente não esclarece como essa desigualdade ocorre. Por exemplo, dois países com o mesmo coeficiente de Gini podem ter desigualdades concentradas em diferentes partes de suas distribuições de renda (Atkinson, 2016). Nesta tese, tanto o coeficiente de Gini como o IDH são usados para se estabelecer um fator relativo de comparação de classes socioeconômicas entre os três países considerados. Esse fator propicia a modelagem de redes sociais com classes socioeconômicas apresentada no capítulo 6.

O algoritmo de rede complexa proposto também leva em consideração o distanciamento geográfico entre os indivíduos. Essa questão é melhor explicada na próxima seção.

2.4 Vizinhanças de Von Neumann e de Moore

No modelo proposto de rede social, as relações de vizinhança entre os indivíduos refletem os contatos sociais presenciais que esses indivíduos realizam. A superfície em que os membros dessa comunidade vivem corresponde a um reticulado bidimensional (Wolfram, 1984). Cada célula desse reticulado equivale a um indivíduo da rede. Nesse reticulado, podem-se definir diversos padrões de conectividade entre as células. Dois padrões de conectividade regular comumente adotados são a vizinhança de Von Neumann e a vizinhança de Moore, muito utilizados em estudos com autômatos celulares (Wolfram, 2018).

Em um reticulado bidimensional, a vizinhança de Von Neumann é aquela estabelecida com as células contidas dentro de um losango, cujo centro é a célula considerada. A figura 2.1 ilustra as vizinhanças de Von Neumann de raios $r = 1$ e $r = 2$. A vizinhança de Moore de raio r de uma célula é formada pelas células que pertencem à matriz quadrada de lado $2r + 1$ centrada nessa célula. Como exemplo, as vizinhanças de Moore de raios $r = 1$ e $r = 2$ estão representadas na figura 2.2.

Figura 2.1. Vizinhança de Von Neumann num reticulado bidimensional para $r = 1$ e $r = 2$. A célula considerada está na cor preta. Já as células vizinhas estão na cor cinza.

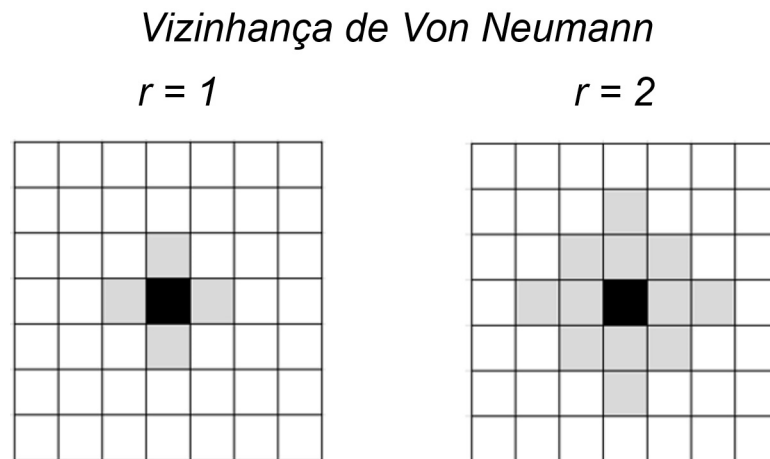
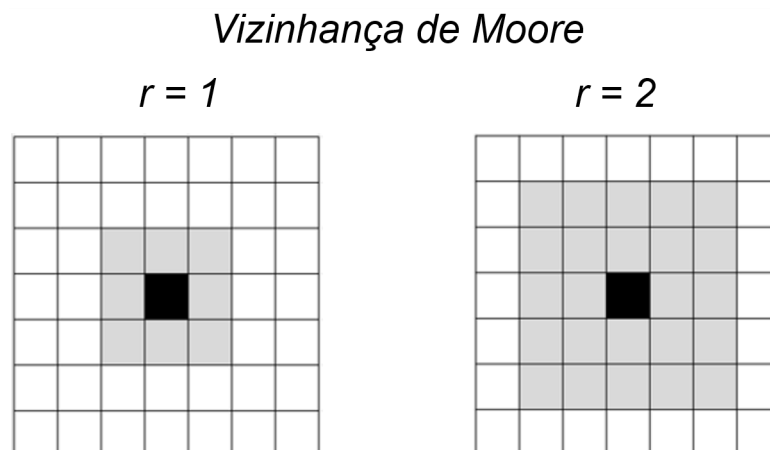


Figura 2.2. Vizinhança de Moore num reticulado bidimensional para $r = 1$ e $r = 2$. A célula considerada está na cor preta. Já as células vizinhas estão na cor cinza.



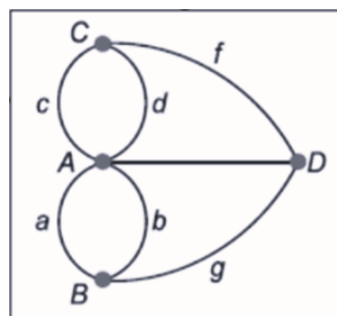
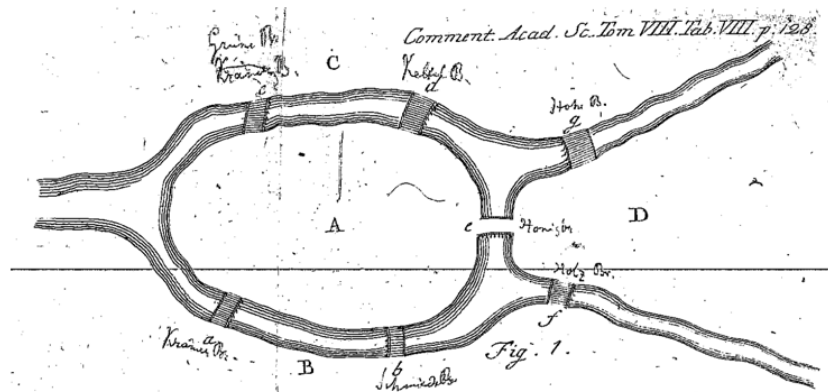
No modelo aqui proposto, a vizinhança de um nó é formada de acordo com uma regra probabilista, dentro de uma vizinhança de Moore de raio r .

Como já mencionado, grafos têm sido usados para representar redes sociais. Grafos são o tema da próxima seção.

2.5 Grafos

O nascimento da Teoria de Grafos ocorreu em 1736, quando Leonhard Euler (1707-1783) analisou o Problema das Sete Pontes de Königsberg, ilustrado na figura 2.3. Nessa análise, Euler (1736) simplificou a representação geográfica, considerando as ilhas e as margens dos rios como nós interligados por arestas, que correspondem às pontes entre essas porções de terra (Monteiro, 2014). Duzentos anos mais tarde, a ideia de grafos foi usada por Linus Pauling (1936) para descrever o movimento de partículas em redes físicas. Grafos têm sido utilizados em estudos de sistemas complexos (Heylighen, 1988; Monteiro, 2014). A seguir, introduzem-se alguns conceitos básicos.

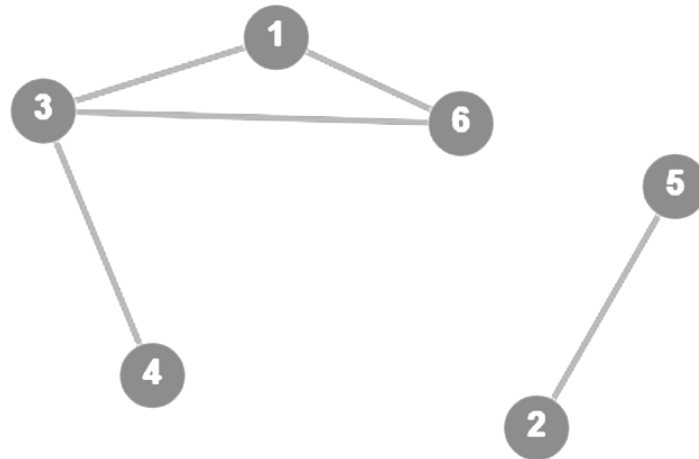
Figura 2.3. Acima: mapa ilustrativo do problema das sete pontes de Königsberg. Esse é o desenho original da publicação de Euler. Abaixo: um possível grafo, que corresponde à uma redução do problema. Nesse grafo, o conjunto $\{A,B,C,D\}$ representa as diversas porções de terra (A e D são as ilhas e B e C são as margens do rio) conectadas por sete pontes, que são arestas do grafo pertencentes ao conjunto $\{a,b,c,d,e,f,g\}$.



Fonte: (Euler, 1741; Vera, 2017).

Um grafo $G = (V, A)$ é formado pelo par de conjuntos V e A , sendo V o conjunto de N vértices ou nós e A o conjunto de M arestas ou conexões entre os vértices. Um possível exemplo de grafo é ilustrado na figura 2.4. Considere as seguintes definições (Boccaletti et al., 2006; Monteiro, 2014; Newman, 2003):

Figura 2.4. Exemplo de um grafo não orientado com 6 nós e 5 arestas. O grafo é desconectado (ou não percolado). O subconjunto dos nós $\{1, 3, 4, 6\}$ forma um subgrafo conectado, assim como o subconjunto dos nós $\{2, 5\}$. O grafo desse exemplo é utilizado para ilustrar diversos cálculos de métricas definidas ao longo desta seção.



- Vizinhança de um nó. Vizinhos são os nós que se encontram conectados por uma aresta. Para o grafo da figura 2.4, são vizinhos os pares não orientados de nós do conjunto $\{(1,6), (1,3), (2,5), (3,4), (3,6)\}$.
- Caminho entre dois nós. O caminho entre dois nós é uma sequência de nós com a restrição de que existe aresta entre um nó e o seguinte. No grafo da figura 2.5, $\{(1,3), (3,4)\}$ representa um caminho entre os nós 1 e 4.
- Distância entre dois nós. A distância entre dois nós é calculada como o menor número de arestas que ligam esses nós por algum caminho. Na figura 2.4, a distância entre os nós, 1 e 4, é 2, embora exista um outro caminho (maior) entre eles (que passa pelo nó 6).
- Percolação. O grafo mostrado na figura 2.4 é desconexo ou não percolado, uma vez que não existem caminhos, por meio de suas arestas, que interconectem quaisquer dois nós. Por exemplo, nessa figura, não há caminho entre os nós 1 e 2.
- Grau de um nó. O grau de um nó corresponde ao número de arestas que têm esse nó em uma de suas extremidades. Para o grafo exemplificado, pode-se verificar que o grau do nó 3 é três, e do nó 5 é um.

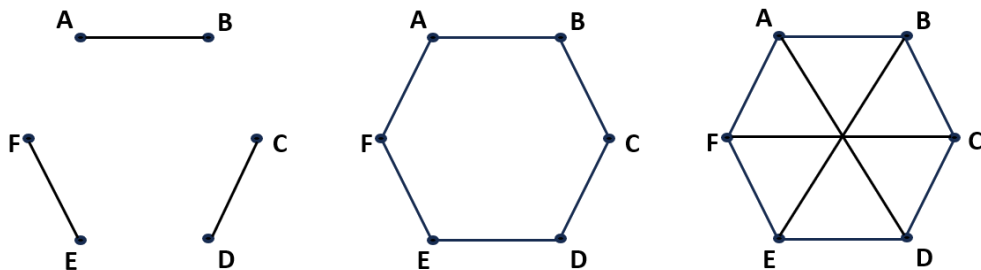
Neste trabalho, consideram-se somente grafos não direcionados, ou seja, aqueles em que o par (i, j) e o par (j, i) correspondem à mesma arestas (por exemplo, na figura acima, os pares $(3, 4)$ e $(4, 3)$ representam a mesma aresta) e não ponderados (ou seja, todas as arestas têm peso unitário).

Diversos trabalhos têm sido realizados a fim de compreender a estrutura, a dinâmica e a evolução de redes reais (Barabasi, 2016; Boccaletti et al., 2006; Monteiro, 2014; Newman, 2003; Voutsas et al., 2021; Watts e Strogatz, 1998) como a internet, redes neurais, redes de distribuição de energia elétrica. Uma maneira de analisar a estrutura de tais redes consiste em calcular e comparar certas medidas topológicas, que refletem propriedades locais ou globais. Redes também surgem em fenômenos sociais, como redes de amigos, de contatos sexuais e de colaboração em trabalhos (como filmes ou artigos científicos). A seguir, descrevem-se alguns tipos de grafos.

2.5.1 Redes regulares

Modelos de redes artificiais têm sido criados com o intuito de permitir a modelagem de fenômenos reais. Um modelo comum e simples consiste em uma rede regular, que pode ser entendida como aquela em que cada nó do grafo apresenta o mesmo grau (o mesmo número de conexões) e o mesmo padrão de conectividade (Monteiro, 2014; Voutsas et al., 2021). A figura 2.5 mostra exemplos de grafos regulares.

Figura 2.5. Exemplos de grafos regulares com seis nós. No grafo da esquerda, cada nó possui grau 1; no grafo central, cada nó possui grau 2 (há duas arestas conectadas a cada nó); no grafo da direita, cada nó possui grau 3.



Embora sejam bastante empregadas na modelagem de fenômenos, como processos de difusão, dinâmica de populações e formação de padrões, as redes com topologias regulares são inadequadas para representar a conectividade presente na maioria dos sistemas complexos reais, caracterizados, dentre outros aspectos, por um grande número de nós com diversos graus. Por isso, outros modelos de rede surgiram. Três modelos interessantes são: a rede aleatória de Erdős e Rényi (1959), a rede mundo-pequeno de Watts e Strogatz (1998) e a rede livre-de-escala de Barabási e Albert (1999).

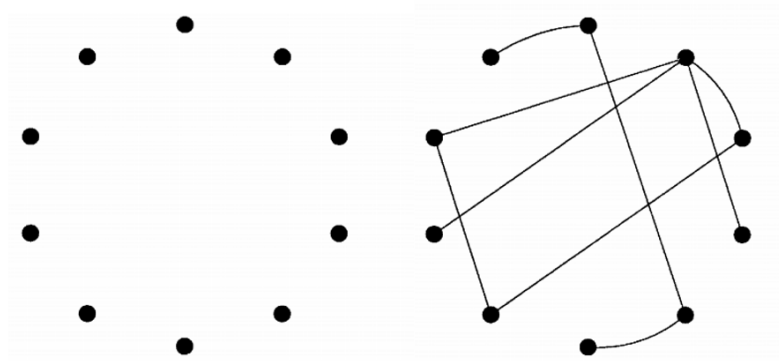
2.5.2 Redes completamente aleatórias

Paul Erdős e Alfréd Rényi propuseram modelos de grafos totalmente aleatórios (Erdős e Rényi, 1959; Monteiro, 2014). Em um dos modelos, N nós são conectados por

M arestas escolhidas aleatoriamente dentre as $N(N-1)/2$ arestas possíveis. Em um outro modelo, cada par de nós é conectado com uma determinada probabilidade predefinida p . Dessa forma, para o primeiro modelo, são fornecidos o número de nós e o número de arestas. Já no segundo, são fornecidos o número de nós e a probabilidade de conexão. A figura 2.6 mostra um grafo com $N = 10$ nós em que são realizadas $M = 9$ conexões aleatórias. Ou, de maneira equivalente, a probabilidade de conexão nesse grafo vale:

$$p = \frac{M}{\frac{N(N-1)}{2}} = \frac{1}{5} \quad (2.1)$$

Figura 2.6. Exemplo de rede aleatória, segundo a proposta de Erdős e Rényi, em que 10 nós são conectados por 9 arestas.



O modelo de redes completamente aleatórias proposto por Erdős e Rényi possui algumas deficiências que limitam sua capacidade de representar redes do mundo real. Algumas limitações desse modelo são:

- Homogeneidade estrutural: no modelo de Erdős-Rényi, todas as arestas têm a mesma probabilidade de existir, resultando em uma rede com uma distribuição de grau de Poisson (Erdős e Rényi, 1959; Monteiro, 2014). No entanto, em muitas redes reais, como redes sociais, a distribuição de graus é diferente, de modo que alguns poucos nós têm muitas conexões, enquanto a maioria tem poucas. Essa heterogeneidade estrutural afeta a dinâmica e as propriedades emergentes das redes (Barabási e Albert, 1999; Newman, 2003).
- Baixa clusterização: o modelo de Erdős-Rényi não favorece a criação de agrupamentos ou comunidades nas redes. Em redes reais, é comum observar agrupamentos (*clusters*), nos quais dois nós têm uma probabilidade maior de se conectarem entre si se ambos são vizinhos de um terceiro nó. Essa estrutura de agrupamentos tem um papel importante em vários fenômenos, como a propagação de informações e a dinâmica de contágio (Newman e Girvan, 2004; Watts e Strogatz, 1998).

Redes mundo-pequeno, tratadas a seguir, têm alta clusterização.

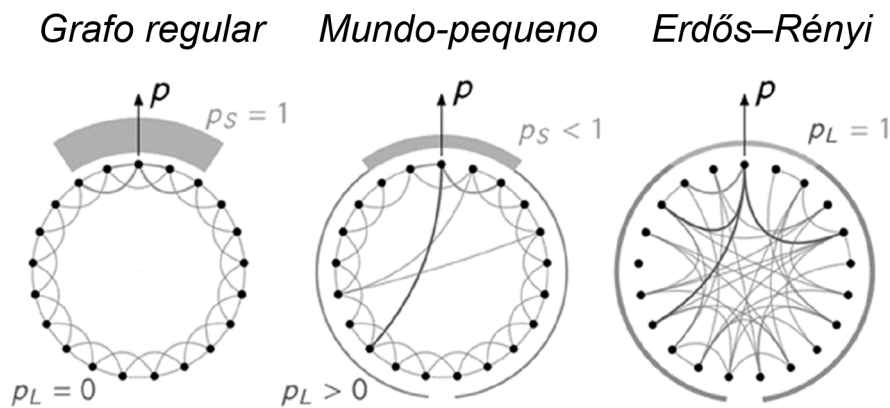
2.5.3 Redes mundo-pequeno

Para a construção de redes mundo-pequeno (*small-world*), parte-se de uma rede regular com formato de anel, em que cada nó está conectado a $2m$ vizinhos, com m à direita e m à esquerda (Watts e Strogatz, 1998). A figura 2.7 à esquerda ilustra o caso em que $m = 2$. Então, considera-se uma probabilidade p_L de um nó ter uma ligação refeita com nós diferentes dos vizinhos iniciais.

Uma rede mundo-pequeno é criada da seguinte maneira:

- percorre-se a rede no sentido horário a partir de um dado nó;
- cada ligação passa a ter probabilidade p_L de ser desfeita e reconectada a outro nó da rede (de modo que $p_S = 1 - p_L$ é a probabilidade de manter a ligação);
- não se permite autoconexão;
- não se pode ligar com outro nó ao qual o atual já se encontra conectado (evitando arestas múltiplas entre dois nós).

Figura 2.7. Exemplo de construção de rede mundo-pequeno, segundo a proposta de Watts e Strogatz. Da esquerda para direita, há três grafos: o primeiro, à esquerda, é um grafo regular, já que a probabilidade de reconexão é nula. No segundo grafo, ao centro, há uma probabilidade de reconexão não nula, tal que $0 < p_L < 1$. No terceiro grafo, à direita, tem-se $p_L = 1$, o que corresponde a um grafo de Erdős e Rényi. Figura adaptada de Maier (2019).



Esse procedimento permite criar conexões de “longa distância”, fazendo com que a distância média entre os nós caia de maneira drástica, mesmo para pequenos valores de p_L (da ordem de 1%). Esse é o conhecido fenômeno mundo-pequeno (*small-world*) (Barrat et al., 2004; Maier, 2019; Monteiro, 2014; Watts e Strogatz, 1998). Conforme a probabilidade p_L aumenta, a rede passa a ser mais semelhante a uma rede aleatória. Para $p_L = 1$, gera-se uma rede completamente aleatória.

Existem outras propostas de modelo de rede mundo-pequeno, como a de Newman e Watts (1999), em que novas arestas são adicionadas de modo aleatório, sem a remoção de quaisquer conexões existentes da estrutura regular original. Há também outra variação do modelo de Watts-Strogatz, que pode ser obtida adicionando-se à rede regular alguns nós centrais que conectados aos nós originais da rede, em uma configuração do tipo estrela. Se apenas um elemento central é adicionado, a rede assim criada é denominada de *smallest world* (Newman e Watts, 1999).

Modelos de redes mundo-pequeno têm fenômenos reais como inspiração (Barabási e Oltvai, 2004; Monteiro, 2014). Como exemplo, pode-se citar o experimento que objetivava quantificar a estrutura da rede de contatos sociais formada pela sociedade americana, relatado por Travers e Milgram, em 1969. Nesse experimento, certos moradores de Omaha (cidade do estado de Nebraska) e de Boston (cidade do estado de Massachusetts) deveriam fazer chegar uma carta a um morador de Boston. A carta poderia ser entregue diretamente a esse morador ou enviada a um possível conhecido dele. As cartas que chegaram, passaram, em média, por cinco pessoas intermediárias (Monteiro, 2014; Travers e Milgram, 1969), sugerindo que o menor caminho médio era em torno de seis arestas.

O modelo de rede mundo-pequeno proposto por Watts e Strogatz também tem sido objeto de críticas, pois apresenta limitações como:

- Homogeneidade estrutural local: a construção da rede começa com uma rede regular, na qual cada nó está conectado a seus vizinhos mais próximos. Esse ponto de partida, por si só, é bastante artificial. Em seguida, ocorrem uma religação probabilística das arestas para criar atalhos e reduzir o menor caminho médio. No entanto, esse processo não leva em consideração uma possível heterogeneidade estrutural local, ou seja, a tendência de alguns nós atraírem naturalmente mais conexões do que outros. Em redes reais, a heterogeneidade estrutural local é comum e afeta a dinâmica e as propriedades da rede (Barrat et al., 2004; Newman, 2010).
- Dependência da regra de religação: a regra probabilística de religação de arestas foi escolhida de maneira arbitrária, sem uma justificativa baseada em evidências de redes reais. Outras regras poderiam ter sido escolhidas, o que afetaria as propriedades do grafo final obtido (Holme e Kim, 2002; Pastor-Satorras e Vespignani, 2004).
- Número de nós fixo: como no modelo de Erdős e Rényi, no modelo de Watts e Strogatz o número de nós da rede se mantém fixo; entretanto, muitas redes reais são abertas, ou seja, ganham e perdem nós e arestas com o passar do tempo.
- Distribuição de graus: a rede de Watts-Strogatz tem uma distribuição de graus do tipo Poisson (Barrat et al., 2004; Monteiro, 2014); entretanto, diversas redes do mundo real têm distribuição de graus livre de escala.

Essa última observação leva ao próximo assunto.

2.5.4 Redes livres-de-escala

Esse modelo foi proposto para explicar a estrutura de algumas redes reais, como a *world wide web* (WWW) (Barabási e Albert, 1999; Barabási e Oltvai, 2004; Monteiro, 2014). Agora, o número de nós não permanece fixo, como nos modelos das duas seções anteriores. No modelo de Barabási e Albert, o número de nós cresce continuamente e a probabilidade de um novo nó se conectar aos já existentes é proporcional aos graus dos nós existentes. Portanto, nós com maior grau têm maiores probabilidades de estabelecer uma conexão com o nó recém-chegado. Essa propriedade também é conhecida como lei do “rico fica mais rico” (*richer get richer*) ou lei da ligação preferencial (*preferential attachment*), previamente introduzida por outros autores (Price, 1976; Simon, 1955; Yule, 1925). Nessa rede, o grau médio não é o grau mais encontrado. De fato, há muitos nós com grau pequeno, um número intermediário de nós com grau médio e um número pequeno de nós com grau grande. Nos modelos de redes aleatórias e de redes do tipo *small-world*, existem muitos nós com grau intermediário e um pequeno número de nós com grau muito alto ou muito baixo (Barabási e Albert, 1999; Barabasi, 2016; Monteiro, 2014). Para o modelo de Barabási e Albert, constatou-se que a distribuição de graus $p(k)$ obedece a uma lei de potências dada por:

$$p(k) = Ak^{-\gamma} \quad (2.2)$$

sendo A e γ constantes positivas. Tais redes são ditas livres-de-escala (*scale free*), pois efetuando-se a substituição de k por ck , com c constante, não se altera a maneira pela qual essa distribuição depende da variável k , já que $p(ck) = A'k^{-\gamma}$, com $A' = Ac^{-\gamma}$.

Conforme estabelecido na literatura sobre redes do mundo real, o expoente γ costuma variar entre 1,4 e 6, ficando tipicamente entre 2 e 3 (Albert, Jeong e Barabási, 1999; Goh et al., 2002; Dorogovtsev, Goltsev e Mendes, 2008; Landherr, Friedl e Heidemann, 2010; Monteiro, 2014).

O modelo de grafo livre-de-escala proposto por Barabási-Albert também tem sido objeto de questionamentos. Os principais são:

- Desconsideração de outros mecanismos de crescimento: no modelo, assume-se que o crescimento da rede é exclusivamente baseado no mecanismo de ligação preferencial, no qual novos nós se conectam preferencialmente aos nós existentes com alta conectividade. No entanto, em algumas redes reais, existem outros mecanismos de crescimento em ação, como a formação de comunidades, interações locais e restrições geográficas. Esses mecanismos alternativos podem afetar a formação e a evolução

das redes (Bianconi e Barabási, 2001; Dorogovtsev e Mendes, 2002). Aliás, a própria lei de ligação preferencial é, na prática, irrealista, pois ela exige que o novo nó conheça os graus de todos os nós presentes na rede (a fim de calcular a probabilidade com que pode se ligar a cada nó).

- Baixa clusterização: embora o modelo seja útil para entender redes complexas com distribuição de graus livre-de-escala, ele não captura certas características importantes presentes em redes reais. Por exemplo, nesse modelo, a clusterização é muito mais baixa do que a observada em redes reais (Caldarelli, 2007; Newman, 2003).

Por ora, foram apresentados os principais modelos de redes complexas e suas limitações. A seguir, definem-se as medidas mais utilizadas para a caracterização de grafos não direcionados.

2.5.5 Métricas topológicas

A distribuição de graus $p(k)$ expressa como a fração de indivíduos com grau k varia com k . O grau médio $\langle k \rangle$ da rede é dado por:

$$\langle k \rangle = \frac{\sum_{i=1}^N k_i}{N} = \sum_{k=k_{min}}^{k_{max}} kp(k) \quad (2.3)$$

em que k_{min} e k_{max} correspondem ao menor grau e ao maior grau encontrados na rede, respectivamente (Albert e Barabási, 2002; Boccaletti et al., 2006; Monteiro, 2014; Newman, 2003). Como exemplo, para o grafo da figura 2.4, os graus dos nós 1, 2, 3, 4, 5 e 6 são, respectivamente 2, 1, 3, 1, 1 e 2. Nesse caso:

$$\langle k \rangle = \frac{2 + 1 + 3 + 1 + 1 + 2}{6} = \frac{5}{3} \quad (2.4)$$

A distribuição de graus para o grafo da figura 2.4 é dada por: $p(1) = 3/6$, $p(2) = 2/6$ e $p(3) = 1/6$, de modo que $p(1) + p(2) + p(3) = 1$. O grau médio também pode ser obtido de $\langle k \rangle = 2M/N$. Para o exemplo, $\langle k \rangle = 10/6 = 5/3$, como já calculado.

A análise de $p(k)$ traz informações relevantes sobre a estrutura da rede (Barabasi, 2016; Newman, 2003). Uma distribuição de graus que segue uma distribuição de Poisson ou tipo-Poisson é típica de redes totalmente aleatórias ou mundo-pequeno (Barrat et al., 2004; Erdős e Rényi, 1959; Monteiro, 2019). Em tais redes, o grau mais provável (o de maior $p(k)$) é o grau médio. Já em redes livres-de-escala, para as quais vale a equação (2.2), o grau mais prevalente é um grau pequeno, menor que o grau médio, e poucos nós apresentam um número elevado de conexões (Barabasi, 2016; Monteiro, 2019).

Outra medida relevante é o comprimento médio do caminho mais curto, dada por:

$$\langle \ell \rangle = \sum_{i=1}^N \sum_{j=1}^N \frac{\ell_{ij}}{N(N-1)} \quad (2.5)$$

em que ℓ_{ij} é a distância entre os nós i e j (com $\ell_{ii} = 0$) (Albert e Barabási, 2002; Boccaletti et al., 2006; Monteiro, 2014; Newman, 2003). Considerando a figura 2.4, pode-se calcular $\langle \ell \rangle$ para o subgrafo $G' = \{1, 3, 4, 6\}$ como:

$$\langle l \rangle = \frac{1 + 1 + 2 + 1 + 1 + 2 + 1 + 1 + 1 + 1 + 2 + 2}{12} = \frac{4}{3} \quad (2.6)$$

Para o grafo da figura 2.4 como um todo, tem-se que $\langle l \rangle \rightarrow \infty$, uma vez que não há arestas entre o nó 2 ou o nó 5 e os demais nós.

De modo geral, em redes complexas, $\langle \ell \rangle$ é um parâmetro que permite avaliar:

- Eficiência da rede. O comprimento médio do caminho mais curto pode indicar quão eficientemente um dado ou outra entidade é transferida através da rede. Redes com $\langle \ell \rangle$ “pequenos” são frequentemente mais eficientes na transmissão de informações (Latora e Marchiori, 2001). Esse aspecto é explorado no capítulo 6.
- Fenômeno de mundo-pequeno. O conceito do mundo-pequeno refere-se à ideia de que, mesmo em redes grandes, qualquer nó pode ser alcançado passando por apenas alguns poucos nós intermediários (Watts e Strogatz, 1998).
- Coesão da rede. Quanto menor $\langle \ell \rangle$, mais coesa é a rede, no sentido de que os nós estão mais próximos uns dos outros, facilitando a comunicação ou a interação entre eles (Barabasi, 2016).

O coeficiente de clusterização (ou agregação) c_i do nó i é dado pela fração entre o número b_i de arestas entre vizinhos de i e o número máximo de arestas que podem existir entre os vizinhos de i . Se o nó i possui grau k_i , então o máximo número de arestas entre seus vizinhos é $k_i(k_i - 1)/2$. Portanto:

$$c_i = \frac{2b_i}{k_i(k_i - 1)} \quad (2.7)$$

se $k_i > 1$ (Albert e Barabási, 2002; Boccaletti et al., 2006; Monteiro, 2014; Newman, 2003; Strogatz, 2001). Se $k_i = 0$ ou 1, então $c_i = 0$. O coeficiente de clusterização médio $\langle c \rangle$ é dado por:

$$\langle c \rangle = \sum_{i=1}^N \frac{c_i}{N} \quad (2.8)$$

Considerando o subgrafo $G' = \{1, 3, 4, 6\}$ na figura 2.4, tem-se $c_1 = 1$, $c_3 = 1/3$, $c_4 = 0$ e $c_6 = 1$, de modo que $\langle c \rangle = 7/12$.

O coeficiente de clusterização médio $\langle c \rangle$ é uma medida que representa o nível com que os nós de uma rede tendem a criar agrupamentos. Quanto maior $\langle c \rangle$, mais os nós estão localmente conectados, ou seja, mais ligações existem entre os vizinhos de um dado nó (Newman, 2003; Watts e Strogatz, 1998).

Há diversas medidas de centralidade. Essas medidas são definidas de forma a permitir uma classificação que forneça a importância de um nó em relação a uma certa característica na rede (Albert e Barabási, 2002).

A centralidade de intermediação (*betweenness centrality*) afere a influência que um nó tem sobre o fluxo de informações em um grafo. Geralmente, é usada para localizar nós que servem como ponte de uma parte de um grafo para outra parte (Jeong et al., 2000). A centralidade de intermediação $C_i(i)$ do nó i é definida como:

$$C_i(i) = \frac{2}{(N-1)(N-2)} \sum_{j=1}^{N-1} \sum_{l=j+1}^N \frac{g_{jl}(i)}{g_{jl}} \quad (2.9)$$

em que g_{jl} é o número de caminhos mais curtos entre os nós j e l e $g_{jl}(i)$ é o número de caminhos mais curtos entre os nós j e l que passam por i (Boccaletti et al., 2006; Monteiro, 2014; Wen e Deng, 2020). A centralidade de intermediação média é dada por:

$$C_i = \sum_{i=1}^N \frac{C_i(i)}{N} \quad (2.10)$$

Para o subgrafo $G' = \{1, 3, 4, 6\}$ mostrado na figura 2.4, obtém-se a tabela 2.2 com as centralidades de intermediação de cada nó. Note que apenas o nó 3 tem centralidade não nula, pois esse nó está no caminho mais curto entre os nós 1 e 4 e entre os nós 4 e 6.

Tabela 2.2. Centralidades de intermediação do subgrafo maior da figura 2.4.

Nó i	$C_i(i)$
1	0
3	2/3
4	0
6	0

A centralidade de proximidade (*closeness centrality*) $C_p(i)$ do nó i é definida como (Boccaletti et al., 2006; Monteiro, 2014; Wen e Deng, 2020):

$$C_p(i) = \frac{N-1}{\sum_{j+1}^N \ell_{ij}} \quad (2.11)$$

A centralidade de proximidade de um nó é tanto maior quanto mais próximo esse nó estiver dos demais, de modo a minimizar o somatório que aparece no denominador da definição dessa medida. O termo $N - 1$ que aparece no numerador é um fator de normalização. Obviamente, a centralidade de proximidade média é dada por:

$$C_p = \sum_{i=1}^N \frac{C_p(i)}{N} \quad (2.12)$$

A tabela 2.3 mostra os valores de $C_p(i)$ para o maior subgrafo da figura 2.4.

Tabela 2.3. Centralidades de proximidade do subgrafo maior da figura 2.4.

Nó i	$C_p(i)$
1	3/4
3	1
4	3/5
6	3/4

A centralidade de proximidade oferece uma percepção sobre o quão rapidamente um nó pode interagir com outros nós da rede. Um valor mais alto de centralidade de proximidade significa que o nó pode disseminar informações de forma mais eficiente. No maior subgrafo da figura 2.4, o nó 3 possui centralidade de proximidade máxima. Por outro lado, um valor baixo de centralidade de proximidade implica que o nó está mais distante dos outros, o que pode dificultar sua capacidade de se comunicar (Newman, 2003; Newman, 2010). Observe que o nó 4 é o menos central por essa métrica.

A centralidade de autovetor (*eigenvector centrality*) $C_a(i)$ do nó i é calculada por:

$$C_a(i) = \frac{1}{\lambda} \sum_{j=1}^N x_{ij} C_a(j) \quad (2.13)$$

sendo λ o maior autovalor da matriz de adjacências X formada pelos elementos x_{ij} , tal que $x_{ij} = 1$ se os nós i e j estão conectados e $x_{ij} = 0$ se não estão (Monteiro, 2014; Newman, 2003; Wen e Deng, 2020). O valor médio dessa medida é dado por:

$$C_a = \sum_{i=1}^N \frac{C_a(i)}{N} \quad (2.14)$$

Para o grafo da figura 2.4, estão mostrados na tabela 2.4 os valores de centralidade de autovetor dos seis nós. Para esse grafo, $C_a \simeq 0,3231$. A centralidade de autovetor de um nó é maior quanto maiores forem as centralidades dos nós aos quais ele está conectado.

Segundo essa métrica, um nó é mais central se ele está conectado a nós que são muito conectados (Landherr, Friedl e Heidemann, 2010; Newman, 2003; Newman, 2010). No grafo da figura 2.4, o nó 3 é o que tem maior centralidade de autovetor. Os nós 2 e 5 têm centralidade nula, pois não se conectam com outros nós.

Tabela 2.4. Centralidades de autovetor do grafo da figura 2.4.

Nó i	$C_a(i)$
1	0,5227
2	0
3	0,6116
4	0,2818
5	0
6	0,5227

A centralidade de grau $C_g(i)$ do nó i é definida como:

$$C_g(i) = \frac{k_i}{N - 1} \quad (2.15)$$

sendo k_i o grau do nó i e $N - 1$ o fator de normalização, correspondente ao maior grau que um nó pode apresentar (Boccaletti et al., 2006; Monteiro, 2014; Wen e Deng, 2020). Portanto, tanto k_i quanto $C_g(i)$ aumentam conforme o nó i ganha arestas. O valor médio dessa centralidade é dado por:

$$C_g = \sum_{i=1}^N \frac{C_g(i)}{N} \quad (2.16)$$

de modo que $C_g = \langle k \rangle / (N - 1)$. Para o grafo da figura 2.4, obtêm-se os valores de $C_g(i)$ apresentados na tabela 2.5.

Tabela 2.5. Centralidades de grau do grafo da figura 2.4.

Nó i	$C_g(i)$
1	2/5
2	1/5
3	3/5
4	1/5
5	1/5
6	2/5

Um nó com valor alto de centralidade de grau é um nó com muitas conexões diretas e, por-

tanto, tal nó é potencialmente mais influente, especialmente em termos da disseminação imediata de informações ou doenças (Landherr, Friedl e Heidemann, 2010; Newman, 2003; Newman, 2010). O nó 3 da figura 2.4 é o que apresenta maior centralidade de grau.

O n -ésimo momento da distribuição de graus de uma rede é definido como (Barabasi, 2016; Newman, 2003):

$$\langle k_n \rangle = \frac{\sum_{i=1}^N (k_i)^n}{N} \quad (2.17)$$

Em geral, são de particular interesse o segundo e o terceiro momentos de uma distribuição. O segundo momento aparece no cálculo da variância σ , dada por:

$$\sigma = \langle k_2 \rangle - \langle k \rangle^2 \quad (2.18)$$

que indica o quão dispersos estão os graus na rede. O segundo momento é relevante em análises de robustez e de vulnerabilidade. Numa rede com alto segundo momento, existem nós com graus muito altos, tornando a rede mais suscetível a ataques direcionados a esses nós altamente conectados (Barabasi, 2016; Newman, 2003). O terceiro momento, por outro lado, oferece *insights* sobre a assimetria da distribuição de graus. Em uma rede livre-de-escala, esse momento tende a divergir para o infinito, conforme a rede cresce (Barabasi, 2016; Newman, 2003).

A tabela 2.6 reúne informações sobre a distribuição de graus $p(k)$, o comprimento médio do caminho mais curto $\langle \ell \rangle$ e o coeficiente de clusterização médio $\langle c \rangle$ para os três modelos de rede apresentados nesta seção.

Tabela 2.6. Algumas propriedades topológicas dos modelos de redes mais famosos. Adaptado de Schimit (2010) e Monteiro (2014).

	Redes mundo-pequeno	Redes totalmente aleatórias	Redes livres-de-escala
$p(k)$	$p(k) \approx \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$	$p(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$	$p(k) = Ak^{-\gamma}$
$\langle \ell \rangle$	$\langle \ell \rangle \approx \frac{\log(N)}{\log(\langle k \rangle)}$	$\langle \ell \rangle \approx \frac{\log(N)}{\log(\langle k \rangle)}$	$\langle \ell \rangle \approx \frac{\log(N)}{\log(\langle k \rangle)}$
$\langle c \rangle$	$\langle c \rangle \gg \frac{\langle k \rangle}{N}$	$\langle c \rangle = \frac{\langle k \rangle}{N}$	$\langle c \rangle \approx \frac{\langle k \rangle}{N}$

Já havia sido mencionado que, na rede de Erdős e Rényi, a distribuição $p(k)$ é a de Poisson; nas redes de Watts e Strogatz, $p(k)$ se assemelha a de Poisson; nas redes de Barabási e Albert, $p(k)$ segue uma lei de potência. Devido à presença de arestas aleatórias nos três modelos citados na tabela acima, pode-se estimar $\langle \ell \rangle$ pela expressão mostrada.

O coeficiente de clusterização médio $\langle c \rangle$ para redes aleatórias ou livres-de-escala é muito menor do que para redes mundo-pequeno.

Nas simulações com o algoritmo que cria redes complexas, é interessante examinar a correlação entre as diversas propriedades topológicas das redes, quando se varia o valor de um parâmetro do modelo, como o raio que define a vizinhança ou o expoente da lei de potência de sua distribuição de graus. A correlação refere-se à relação estatística entre duas variáveis contínuas, X e Y . Ela pode ser quantificada pelo coeficiente de correlação r_{xy} , que varia de -1 a $+1$, indicando correlação negativa perfeita e correlação positiva perfeita, respectivamente (Mukaka, 2012). Esse coeficiente r_{xy} , para as variáveis x e y de interesse, pode ser calculado pelo método de Pearson quando os dados respeitam certa distribuição estatística. Assim:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.19)$$

de modo que r_{xy} é o coeficiente de correlação de Pearson, x_i e y_i são os valores individuais das variáveis X e Y , \bar{x} e \bar{y} são as médias dos valores de X e Y respectivamente, e n representa o número de pares de dados. Assim, o somatório é calculado para cada par de dados, de $i = 1$ até n .

Valores de $|r_{xy}| \geq 0,8$ implicam que as variáveis x e y são, em geral, consideradas fortemente correlacionadas (positiva ou negativamente). Já valores de $|r_{xy}| \leq 0,2$ implicam que as variáveis x e y são fracamente ou pouco correlacionadas. É válido comentar que uma correlação alta ($|r_{xy}| \approx 1$) não implica uma relação causal entre as variáveis.

Neste capítulo, foram apresentados conceitos que possibilitam o entendimento do algoritmo proposto para criar redes complexas e das análises realizadas. Esses são os temas dos próximos capítulos.

3 UM MODELO DE REDE DE CONTATOS SOCIAIS

Redes de interações sociais entre indivíduos, grupos e organizações costumam apresentar simultaneamente propriedades de redes mundo-pequeno e de redes livres-de-escala. Essas propriedades têm sido analisadas e discutidas na literatura (Barabasi, 2016; Landherr, Friedl e Heidemann, 2010; Newman, 2003; Newman e Watts, 1999). Das redes mundo-pequeno, as redes sociais reais apresentam um alto coeficiente de clusterização médio; das redes livres-de-escala, elas apresentam uma distribuição de graus que segue uma lei de potência. Observe que, portanto, os modelos de Watts-Strogatz e Barabási-Albert não são, isoladamente, adequados. O primeiro leva a uma distribuição de graus do tipo Poisson; o segundo leva a um grafo com baixo coeficiente de clusterização médio.

O modelo de rede complexa introduzido neste capítulo almeja ter $\langle c \rangle$ grande e $p(k) \sim k^{-\gamma}$. Aqui, $\langle c \rangle$ grande significa $\langle c \rangle \gg \langle c \rangle_{random}$, sendo $\langle c \rangle_{random} = \langle k \rangle / N$. Assim, $\langle c \rangle$ grande significa clusterização muito maior do que aquela encontrada numa rede aleatória com os mesmos valores de $\langle k \rangle$ e N .

O modelo aqui proposto se inspira naquele originalmente usado por Monteiro, Paiva e Piqueira (2006) num estudo sobre propagação de disparos neuronais e posteriormente empregado por Schimit e Monteiro (2009) numa investigação sobre propagação de doenças contagiosas. Assim, assume-se que cada indivíduo interage com outros indivíduos dentro de uma região geográfica limitada, que pode até ser pequena em relação à rede que representa a sociedade como um todo. Essa hipótese intuitiva se baseia na observação de que os indivíduos geralmente se relacionam mais com parentes que moram perto e com vizinhos; assim, as ligações são preferencialmente locais. Entretanto, também há contatos sociais de “longa” distância (dentro daquela região limitada) com outras pessoas em lugares diversos, como escolas, escritórios, parques, shoppings, etc. (Grusky, 2019; Levine, 2006; Noble e Tynes, 2016).

Vale lembrar que existem modelos de redes, que se formam segundo a regra da ligação preferencial (Albert e Barabási, 2002; Barabasi, 2016; Barabási e Albert, 1999; Dorogovtsev e Mendes, 2002), que levam à criação de grafos livre-de-escala. No entanto, tal modelo de rede complexa não justifica como um novo indivíduo, que se conectará à rede existente, pode conhecer toda a topologia da rede pré-existente a fim de determinar quais são os nós preferenciais, para então sortear a qual se conectará.

No modelo aqui proposto, impõe-se que os nós podem estabelecer um número especificado de conexões, restritas à área da sua vizinhança. Essas conexões simbolizam os contatos presenciais que ocorrem de modo usual. Em tais contatos, podem-se trocar informações, mercadorias e até infecções. Esse aspecto do modelo está em concordância com trabalhos que tratam de relações sociais reais (Grusky, 2019; Landherr, Friedl e Heidemann, 2010; Pastor-Satorras e Vespignani, 2004; Caldarelli, 2007). O número de conexões de cada indivíduo é atribuído de maneira probabilista aos indivíduos, de maneira que a distribuição de graus da rede seja livre-de-escala.

Na próxima seção, detalha-se como a rede é gerada. No capítulo 4, apresentam-se os resultados de simulações que permitem melhor compreender o impacto dos parâmetros do modelo sobre a estrutura do grafo construído. Por enquanto, assume-se que todos os indivíduos pertencem à mesma classe (socioeconômica).

3.1 O algoritmo

Considere um reticulado bidimensional, como aquele ilustrado na figura 2.2, no qual cada indivíduo corresponde a uma célula desse reticulado. Por simplicidade, assumamos que a borda superior está conectada à borda inferior e que a borda esquerda está conectada à borda direita (Licciardi e Monteiro, 2022; Monteiro, 2014; Schimit, 2010). Dessa maneira, todos os indivíduos são idênticos em relação a questões de vizinhança, independentemente da posição física que ocupam no reticulado. Em outras palavras, a posição geográfica de cada indivíduo é, por si só, irrelevante.

Por simplicidade, suponhamos que os vizinhos de cada indivíduo se localizam no interior de uma região com vizinhança de Moore de raio r , como mostrado na seção 2.4. Agora, é necessário estabelecer o critério de formação de conexões entre a célula central mostrada na 2.2 e as demais células contidas em sua vizinhança. É razoável impor que a probabilidade de conexão diminua com a distância entre os indivíduos (Schimit, 2010; Schimit e Monteiro, 2009). Assim, assume-se que a probabilidade q_i de conexão é dada por:

$$q_i = \frac{2(r + 1 - i)}{r(r + 1)} \quad (3.1)$$

em que $i = 1, 2, \dots, r$ representa o raio de Moore em que a célula a ser conectada está, em relação à célula central. Note que q_i diminui com i . Por exemplo, para $r = 4$, então $q_1 = 4/10$, $q_2 = 3/10$, $q_3 = 2/10$ e $q_4 = 1/10$. Nessa expressão para q_i , o denominador apenas assegura a normalização. De fato:

$$\sum_{i=1}^r q_i = 1 \quad (3.2)$$

O valor de r é o mesmo para todos os nós que compõem o grafo a ser construído seguindo esse modelo. Note que o número máximo de vizinhos que um nó pode ter é:

$$(2r + 1)^2 - 1 = 4r^2 + 4r \quad (3.3)$$

O grau k de cada indivíduo é assim determinado. Primeiro, toma-se um número aleatório x entre 0 e 1, segundo uma distribuição uniforme. Então, o valor de k é obtido a partir de $p(k) = x$, com $p(k) = Ak^{-\gamma}$. Para o algoritmo, tem-se o interesse que $\langle k \rangle$ do reticulado apresente um grau predeterminado.

A condição de normalização de $p(k)$, dada por:

$$\sum_{k=k_{min}}^{k_{max}} p(k) = 1 \quad (3.4)$$

impõe que a constante A seja igual a:

$$A = \frac{1}{\sum_{k=k_{min}}^{k_{max}} k^{-\gamma}} \quad (3.5)$$

em que k_{min} e k_{max} são o grau mínimo e o grau máximo da rede.

Pela definição da distribuição de grau, tem-se que:

$$p(k) = Ak^{-\gamma} = \frac{n_k}{N} \quad (3.6)$$

sendo n_k o número de nós que têm grau k numa rede com N nós. Combinando as duas expressões acima, chega-se a:

$$n_k = \frac{Nk^{-\gamma}}{\sum_{k=k_{min}}^{k_{max}} k^{-\gamma}} \quad (3.7)$$

Na implementação do algoritmo, n_k é tomado com a parte inteira do número real calculado a partir da equação (3.7). Evidentemente:

$$\sum_{k=k_{min}}^{k_{max}} n_k = N \quad (3.8)$$

Para o modelo proposto, deseja-se que $p(k)$ seja livre-de-escala e que a rede apresente um grau médio especificado, chamado de $\langle k \rangle_{alvo}$. Combinando as equações (2.3), (3.6) e (3.7), obtém-se:

$$\langle k \rangle = \frac{\sum_{k=k_{min}}^{k_{max}} k^{-\gamma+1}}{\sum_{k=k_{min}}^{k_{max}} k^{-\gamma}} \quad (3.9)$$

Para o grafo a ser criado, deseja-se que:

$$n_{k_{max}} \geq 1 \quad (3.10)$$

isto é, há pelo menos um nó com $k = k_{max}$. Consequentemente:

$$k_{max} \leq \left(\frac{N}{\sum_{k=k_{min}}^{k_{max}} k^{-\gamma}} \right)^{1/\gamma} \quad (3.11)$$

As equações e (3.9) e (3.11) são usadas para construir um grafo com $\langle k \rangle \simeq \langle k \rangle_{alvo}$. Observe que o raio r deve ser tal que $(2r + 1)^2 - 1 \geq k_{max}$, isto é, deve existir uma quantidade de indivíduos na vizinhança de um nó que permita o aparecimento de pelo menos um nó conectado a k_{max} vizinhos.

A título de exemplo, considere $\gamma = 2,5$ e um reticulado bidimensional com $N = 10000$ células. Assuma que se deseja uma rede com $\langle k \rangle_{alvo} = 10$. Nesse caso, as equações (3.9) e (3.11) se tornam:

$$\langle k \rangle = \frac{\sum_{k=k_{min}}^{k_{max}} k^{-1,5}}{\sum_{k=k_{min}}^{k_{max}} k^{-2,5}} \quad (3.12)$$

e:

$$k_{max} \leq \left(\frac{10000}{\sum_{k=k_{min}}^{k_{max}} k^{-2,5}} \right)^{1/2,5} \quad (3.13)$$

Computacionalmente, realizam-se os seguintes passos, a fim de determinar os valores de k_{min} e k_{max} que podem gerar uma rede livre-de-escala com $\langle k \rangle = \langle k \rangle_{alvo}$:

- tomam-se sucessivos números inteiros para k_{min} , tal que $1 \leq k_{min} \leq \langle k \rangle_{alvo}$;
- para cada k_{min} , calcula-se k_{max} obedecendo à restrição imposta pela equação (3.13), considerando o sinal de igualdade;
- para cada par de valores de k_{min} e k_{max} , determina-se $\langle k \rangle$ usando a equação (3.12).

O gráfico mostrado na figura 3.1 ilustra o resultado desses cálculos. Observe nessa figura que para $k_{min} = 3$, tem-se $\langle k \rangle \simeq \langle k \rangle_{alvo}$. Portanto, toma-se $k_{min} = 3$ e, consequentemente, $k_{max} = 161$.

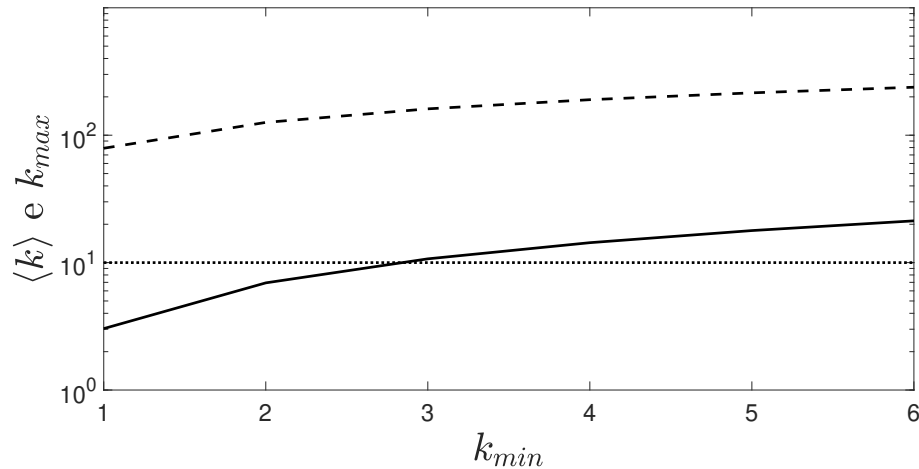
O algoritmo 3.1 corresponde ao pseudocódigo do modelo implementado computacionalmente para gerar redes complexas. Esse algoritmo sequencialmente cria $M \simeq N \langle K \rangle_{alvo} / 2$ conexões, de forma a obter $\langle k \rangle = \langle k \rangle_{alvo}$, com $p(k) = Ak^{-\gamma}$. Vale lembrar que $\langle k \rangle = 2M/N$ (Albert e Barabási, 2002; Newman, 2003; Boccaletti et al., 2006; Monteiro, 2014), sendo M o número de arestas, conforme tratado na seção 2.5.

Como observação final, lembre que a curva $p(k) = Ak^{-\gamma}$, num gráfico log-log, torna-se uma reta, pois:

$$\log p(k) = \log A - \gamma \log k \quad (3.14)$$

em que $-\gamma$ é o coeficiente angular dessa reta. Gráficos log-log de $p(k)$ são apresentados no próximo capítulo, para diferentes valores de r , γ e $\langle k \rangle_{alvo}$.

Figura 3.1. Valores possíveis de inicialização da rede complexa. Escala vertical em log. A curva tracejada representa o valor de k_{max} ; a curva em linha sólida, o valor de $\langle k \rangle$; e a linha pontilhada, o valor $\langle k \rangle_{alvo} = 10$ (essa linha foi plotada apenas por referência).



Algoritmo 3.1 Implementação da rede complexa de contatos sociais

- 1: Escolha valores para N , r , γ e $\langle k \rangle_{alvo}$
 - 2: A partir de $\langle k \rangle_{alvo}$, determine k_{min} e k_{max}
 - 3: Calcule n_k para $k_{min} \leq k \leq k_{max}$
 - 4: Atribua para cada nó um valor de k , respeitando as quantidades n_k de nós com grau k calculadas no passo anterior
 - 5: Crie (de modo probabilista) conexões entre os nós, de acordo com a probabilidade q_i com $1 \leq i \leq r$ (autoconexões ou conexões múltiplas entre nós não são permitidas)
 - 6: Caso existam subgrafos isolados, eles são ligados aleatoriamente ao grafo principal
 - 7: Após a criação de um grafo conexo, compute $\langle k \rangle$. Em função das restrições de vizinhança, acrescidas da restrição de n_k , o grau médio pode estar abaixo de $\langle k \rangle_{alvo}$. Se isso acontecer, retorne ao passo 5 e construa um novo grafo
 - 8: Caso, após 10 tentativas, $\langle k \rangle$ continue abaixo de $\langle k \rangle_{alvo}$, repita a rotina a partir do passo 2, adicionando 1 ao k_{min} até então usado
 - 9: O processo iterativo é encerrado somente quando $\langle k \rangle$ for igual ou ligeiramente maior do que $\langle k \rangle_{alvo}$
-

Após a rede ser construída, são calculadas as métricas apresentadas na subseção 2.5.5. Os resultados de simulações com esse modelo são mostrados no próximo capítulo.

4 SIMULAÇÃO DO ALGORITMO DE CONTATOS SOCIAIS

Neste capítulo, várias séries de simulações são conduzidas usando o algoritmo de redes de contatos sociais com conectividade ajustável proposto do capítulo anterior. O objetivo é compreender o impacto dos parâmetros do modelo sobre as métricas topológicas das redes complexas criadas.

Nas simulações, utiliza-se uma matriz bidimensional quadrada com 70 linhas e 70 colunas, para gerar um reticulado com $N = 4900$ indivíduos. Esse número de indivíduos oferece um tempo de computação do algoritmo razoável, tipicamente 8 a 9 horas para cada grafo, incluindo a etapa de geração do grafo e o cálculo das métricas topológicas, em *software* interpretado como Matlab, em um computador i7 de 11^a geração e 16 GB de memória RAM. Este número de indivíduos não apresenta variabilidade estatística relevante nas métricas calculadas, se repete-se a montagem do grafo com os mesmos valores de parâmetros, com o número de algoritmos significativos apresentados nas tabelas a seguir. As simulações realizadas contemplam:

- Variação do grau médio $\langle k \rangle_{alvo}$

Para $\gamma = 3$ e $r = 10$ (de modo que há, no máximo, 440 vizinhos, de acordo com a equação (3.3)), constroem-se grafos com $\langle k \rangle_{alvo} = \{5, 10, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100\}$. Quatorze gráficos log-log para $p(k)$ são apresentados na figura 4.1. As análises desses grafos são relatadas na seção 4.1.

- Variação do expoente γ

Para $\langle k \rangle_{alvo} = 15$ e $r = 10$, constroem-se grafos com $\gamma = \{0,3; 0,5; 0,7; 1; 2; 3; 5; 6; 7; 8; 9; 10\}$. Treze gráficos log-log para $p(k)$ são mostrados na figura 4.5. As análises são reportadas na seção 4.2.

- Variação do raio r da vizinhança de Moore

Para $\langle k \rangle_{alvo} = 20$, constroem-se grafos com $\gamma = 2$ e $r = \{5, 10, 20, 30, 40, 60\}$ e com $\gamma = 3$ e $r = \{5, 20, 60\}$. Nove gráficos log-log para $p(k)$ são exibidos na figura 4.9. As análises desses resultados se encontram na seção 4.3.

4.1 Variação do grau desejado $\langle k \rangle_{alvo}$

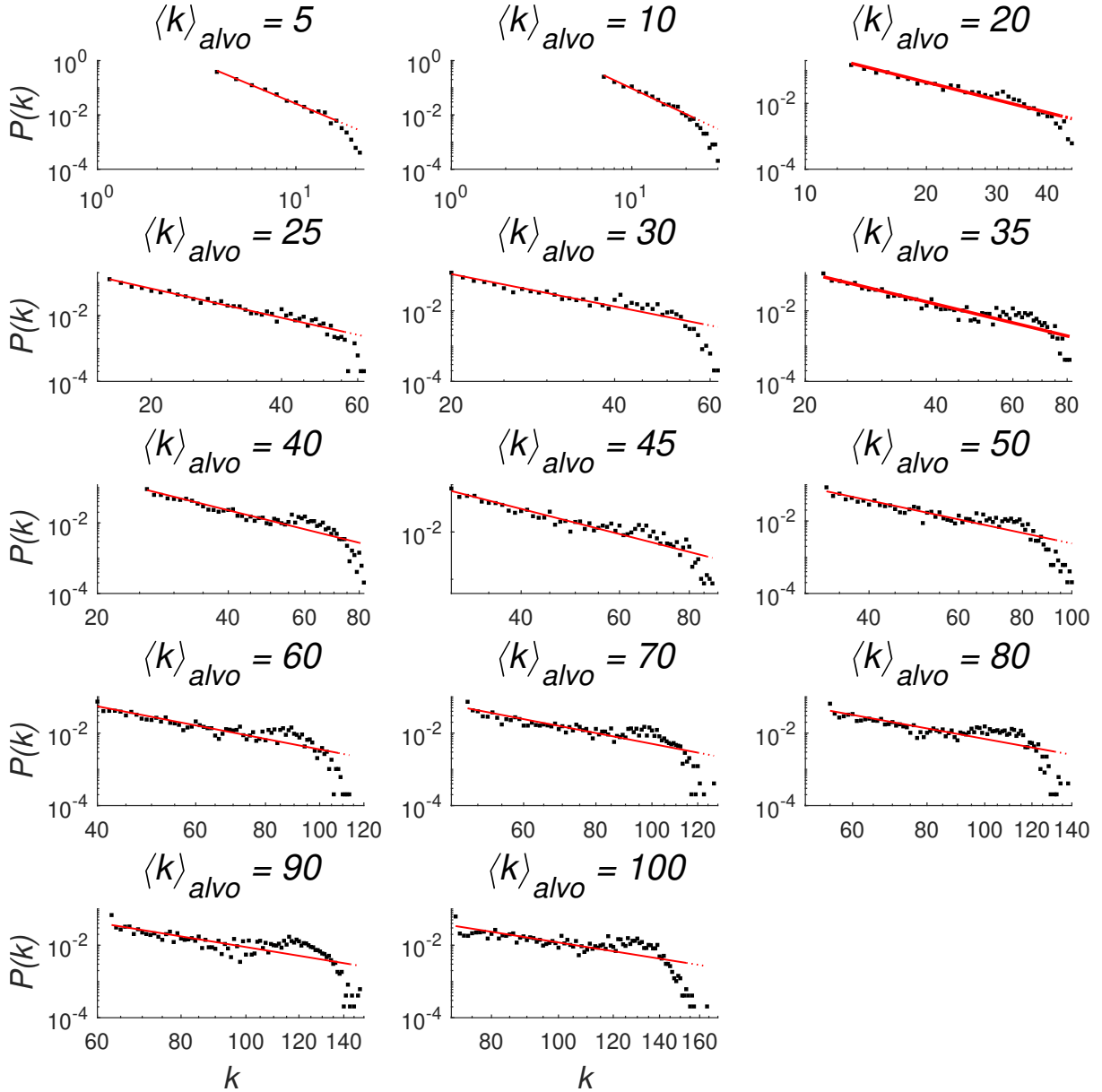
A figura 4.1 contém quatorze gráficos que apresentam a distribuição de graus $p(k)$ em função de k , em gráficos log-log (base 10) para $\langle k \rangle_{alvo} = \{5, 10, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100\}$, com $r = 10$ e $\gamma = 3$. Os gráficos também contêm a regressão linear com coeficiente angular -3 , que caracterizaria o aspecto livre-de-escala pretendido com o algoritmo proposto. Pode-se observar nesses 14 gráficos um comportamento de cauda à direita, relacionado aos maiores graus da rede complexa (Barabasi, 2016; Barabási e Albert, 1999; Clauset, Shalizi e Newman, 2009; Mitzenmacher, 2004). Esse comportamento é encontrado em redes livre-de-escala reais e é analisado na seção 4.2, nas simulações em que o expoente γ varia. O cálculo das centralidades e de outras métricas topológicas são apresentados na tabela 4.1.

Observa-se na tabela 4.1 que o valor de $\langle k \rangle$ é ligeiramente superior ao $\langle k \rangle_{alvo}$ desejado, para todas as redes construídas, com exceção de $\langle k \rangle_{alvo} = 5$. Esse valor, todavia, é muito baixo, quando se consideram redes de contatos sociais reais (Béraud et al., 2015; Grijalva et al., 2015; Melegaro et al., 2017). Para esse valor de $\langle k \rangle_{alvo}$, já na primeira execução do algoritmo 3.1, é produzida uma rede com $\langle k \rangle \approx 6$, um erro de 20% acima do desejado. Nos demais casos, porém, o algoritmo gera grafos com o grau médio bastante próximo do pretendido.

Tabela 4.1. Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i , de autovetor C_a , comprimento médio do caminho mais curto $\langle \ell \rangle$, coeficiente de clusterização médio $\langle c \rangle$ e grau médio $\langle k \rangle$ em função de $\langle k \rangle_{alvo}$.

$\langle k \rangle_{alvo}$	C_g	C_p	C_i	C_a	$\langle \ell \rangle$	$\langle c \rangle$	$\langle k \rangle$
5	0,00123	0,15	0,00115	0,0002	6,6	0,03	6,0
10	0,00213	0,19	0,00086	0,0002	5,2	0,04	10,4
15	0,00308	0,21	0,00073	0,0002	4,5	0,05	15,1
20	0,00411	0,23	0,00065	0,0002	4,1	0,06	20,1
25	0,00513	0,25	0,00060	0,0002	3,9	0,07	25,1
30	0,00615	0,26	0,00057	0,0002	3,8	0,08	30,1
35	0,00716	0,27	0,00054	0,0002	3,6	0,10	35,0
40	0,00819	0,27	0,00053	0,0002	3,5	0,10	40,1
45	0,00922	0,28	0,00051	0,0002	3,4	0,11	45,1
50	0,01021	0,29	0,00050	0,0002	3,4	0,12	50,0
60	0,01226	0,29	0,00048	0,0002	3,3	0,14	60,0
70	0,01441	0,30	0,00047	0,0002	3,2	0,16	70,5
80	0,01635	0,30	0,00046	0,0002	3,2	0,18	80,0
90	0,01839	0,31	0,00045	0,0002	3,1	0,20	90,0
100	0,02047	0,31	0,00044	0,0002	3,1	0,22	100,3

Figura 4.1. Distribuição de graus $p(k)$ em um gráfico log-log (base 10) para $N = 4900$ indivíduos, raio de Moore $r = 10$, expoente $\gamma = 3$ e $\langle k \rangle_{alvo} = \{5, 10, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100\}$. As escalas horizontais não são iguais nos 14 gráficos, a fim de facilitar a visualização de distribuição dos pontos nas diversas décadas em que ocorrem.



Quanto ao coeficiente de clusterização médio, sabe-se (como mostrado na tabela 2.6) que $\langle c \rangle = \langle k \rangle / N$ para redes completamente aleatórias, $\langle c \rangle \approx \langle k \rangle / N$ para redes livre-de-escala, e $\langle c \rangle \gg \langle k \rangle / N$ para redes mundo-pequeno. Para as redes obtidas com o algoritmo proposto, note, por exemplo, que para $\langle k \rangle_{alvo} = 10$, tem-se que $\langle k \rangle_{alvo} / N = 0,002$ que é muito menor do que $\langle c \rangle = 0,04$ encontrado no grafo; para $\langle k \rangle_{alvo} = 45$, tem-se que $\langle k \rangle_{alvo} / N = 0,009$ que é muito menor do que $\langle c \rangle = 0,12$ encontrado no grafo.

Portanto, o algoritmo proposto leva a grafos com coeficiente de clusterização “grande”, como em redes mundo-pequeno.

De acordo com a literatura, uma estimativa para o comprimento médio do caminho mais curto é $\langle \ell \rangle = \log N / \log \langle k \rangle$. Por exemplo, para $\langle k \rangle = 50$, tem-se $\langle \ell \rangle = \log N / \log \langle k \rangle = 2,17$, que é um número próximo, mas menor do que o mostrado na tabela 4.1 (na tabela, $\langle \ell \rangle = 3,4$). Nosso algoritmo restringe a região onde as conexões podem ocorrer, o que contribui para aumentar $\langle \ell \rangle$, se comparado a uma rede totalmente aleatória.

Conforme mostrado na tabela 4.1, aumentar $\langle k \rangle$ causa diminuição em $\langle \ell \rangle$ e aumento em $\langle c \rangle$. Esse resultado é esperado, pois aumentar o grau médio dos nós implica aumentar o número de arestas, o que deve reduzir a distância entre os nós e aumentar a clusterização local. Como curiosidade, esse comportamento também é observado em redes mundo-pequeno ultrapequeno. Tais redes são caracterizadas pelo surgimento de alguns nós com alta clusterização (Barabasi, 2016; Mullick e Sen, 2020; Zheng et al., 2014).

Na tabela 4.1, verifica-se que a centralidade de grau média C_g cresce linearmente com $\langle k \rangle$, como esperado pela definição. A centralidade de proximidade média C_p aumenta com $\langle k \rangle$, o que é consistente com a diminuição de $\langle \ell \rangle$. Em relação à centralidade de intermediação média C_i , nota-se que ela diminui com o aumento de $\langle k \rangle$, refletindo que a quantidade de nós intermediários entre quaisquer dois nós diminui devido ao aumento em C_p . A centralidade de autovalor média $C_a = 0,0020 \approx 1/4900$ fica inalterada. A variação das demais métricas em função de $\langle k \rangle$ são mostradas na figura 4.2.

Os momentos de segunda e terceira ordem, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, respectivamente, e o valor de Δk de cada grafo simulado são apresentados na tabela 4.2. Os momentos são calculados de acordo com a equação (2.17) e Δk é o intervalo em que k obedece à equação (2.2). Esse intervalo é dado por:

$$\Delta k = k_{cut} - k_{min} \quad (4.1)$$

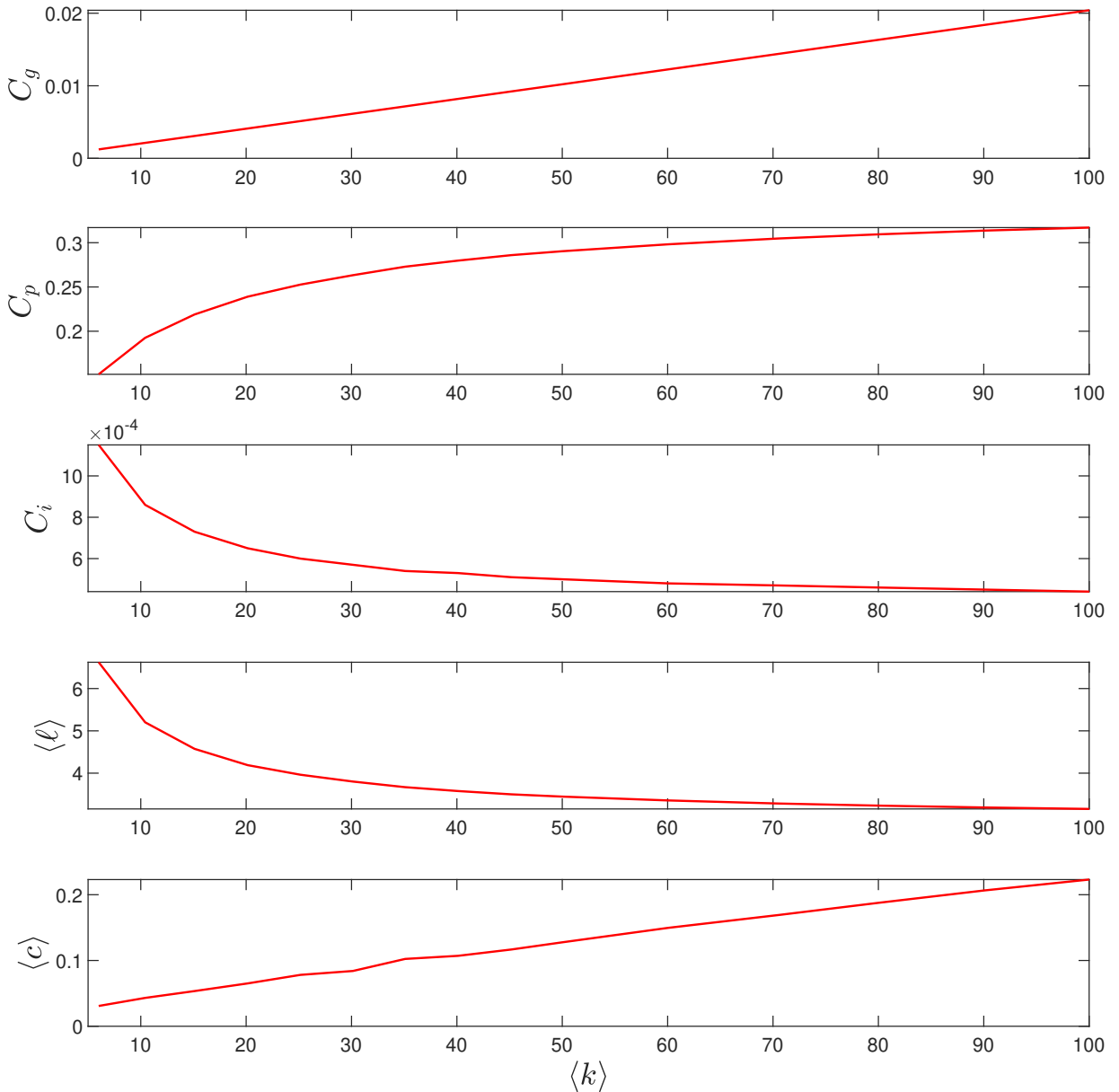
em que k_{min} é o menor grau que ocorre na rede e k_{cut} é definido como o grau de corte, determinado a partir do método dos mínimos quadrados. Assim, k_{cut} é o valor de k acima do qual $\log p(k)$ se desvia da declividade $-\gamma$ desejada, face o fenômeno da cauda (Barabasi, 2016; Barabási e Albert, 1999; Clauset, Shalizi e Newman, 2009; Mitzenmacher, 2004). De fato, diversas distribuições de grau de redes reais são habitualmente melhores descritas por:

$$p(k) = Ak^{-\gamma} 10^{-k/k_{cut}} \quad (4.2)$$

que corresponde a lei de potência com um *cutoff* exponencial, ou:

$$\log p(k) = \log A - \gamma \log k - k/k_{cut} \quad (4.3)$$

Figura 4.2. Centralidades médias de grau C_g , de proximidade C_p e de intermediação C_i , comprimento médio do caminho mais curto $\langle \ell \rangle$ e coeficiente de clusterização médio $\langle c \rangle$ em função de $\langle k \rangle$.



como pode ser encontrado na literatura (Barabasi, 2016; Clauset, Shalizi e Newman, 2009; Mitzenmacher, 2004).

Redes livres-de-escala tendem a ter o segundo e o terceiro momentos divergindo se o número de nós cresce indefinidamente (Barabasi, 2016; Barabási e Albert, 1999; Clauset, Shalizi e Newman, 2009; Mitzenmacher, 2004). Para redes reais finitas, nas quais o fenômeno da cauda é observado, esses momentos tendem a ser grandes se comparados a $\langle k \rangle$, como pode ser observado na tabela 4.2. O momento de segunda ordem pode ser usado para o cálculo da variância, de acordo com a equação (2.18). Em redes em que $p(k)$

segue a distribuição de Poisson, o desvio padrão da rede é $\sigma^{1/2} = \langle k \rangle^{1/2}$ (Barabasi, 2016).

Tabela 4.2. Momentos de segunda $\langle k_2 \rangle$ e de terceira ordem $\langle k_3 \rangle$, desvio padrão de uma distribuição de Poisson $\langle k \rangle^{1/2}$, $\sigma^{1/2}$ e Δk dos grafos gerados, em função de $\langle k \rangle_{alvo}$.

$\langle k \rangle_{alvo}$	$\langle k_2 \rangle$	$\langle k_3 \rangle$	$\langle k \rangle^{1/2}$	$\sigma^{1/2}$	Δk
5	63	797	2,2	6,2	12
10	170	2983	3,1	8,4	16
20	506	13946	4,4	10,3	29
25	922	35666	5,0	17,2	40
30	1133	44646	5,4	15,3	38
35	1424	66979	5,9	14,1	60
40	1910	98269	6,3	17,6	54
45	2368	133057	6,7	18,5	56
50	3371	225251	7,0	29,5	61
60	4780	373800	7,7	34,4	68
70	6905	628725	8,3	44,8	72
80	8467	840877	8,9	45,5	76
90	10079	1084947	9,4	44,5	80
100	13567	1642120	10	59,7	83

Por exemplo, para $\langle k \rangle_{alvo} = 45$, tem-se que $\langle k_2 \rangle^{1/2} \simeq 49 > \langle k \rangle^{1/2} \simeq 6,7$. Portanto, a variância de $p(k)$ dos grafos gerados é bastante diferente daquela de redes completamente aleatórias, e alta se comparada ao $\langle k \rangle$ da rede, como encontrado em redes livres-de-escala.

Em relação ao intervalo Δk , nota-se que quanto maior o $\langle k \rangle_{alvo}$, maior é o Δk devido ao efeito da cauda ser observado em graus cada vez mais altos. Para o grafo com $\langle k \rangle = 100$, um Δk de quase duas décadas é encontrado, o que é característico de redes reais livres-de-escala, apesar de o número de indivíduos na redes geradas ser relativamente pequeno ($N = 4900$), quando comparado a redes reais descritas na literatura (Albert e Barabási, 2002; Barabasi, 2016; Barabási e Albert, 1999; Clauset, Shalizi e Newman, 2009; Dorogovtsev e Mendes, 2002; Mitzenmacher, 2004; Newman, 2003).

Considerando os dados apresentados nas tabelas 4.1 e 4.2, é possível calcular as correlações a partir da equação (2.19). Na figura 4.3, apresenta-se a matriz de correlação de C_g , C_p , C_i , C_a , $\langle \ell \rangle$, $\langle c \rangle$, $\langle k \rangle$, $\langle k_2 \rangle$, $\langle k_3 \rangle$ e Δk . As células da matriz, cujos valores podem variar no intervalo $[-1, 1]$, são pintadas com cores. Quanto mais próximo de 1 for a correlação de duas variáveis, a célula apresenta um azul mais intenso. Quanto mais próximo de -1 for a correlação de duas variáveis, a célula apresenta um vermelho mais intenso.

Observa-se que $\langle k \rangle$ está fortemente correlacionado positivamente com C_g , como esperado, e também com C_p e $\langle c \rangle$. Esses resultados confirmam a ideia de que quanto maior o número médio de conexões por nó, mais próximos e clusterizados os nós estão. Ainda, $\langle k \rangle$ está fortemente correlacionado positivamente com Δk ; ou seja, quanto maior o

Figura 4.3. Matriz de correlação das métricas C_g , C_p , C_i , C_a , $\langle l \rangle$, $\langle c \rangle$, $\langle k \rangle$, $\langle k_2 \rangle$, $\langle k_3 \rangle$ e Δk , considerando os grafo gerados a partir da escolha $\langle k \rangle_{alvo}$. As células cujos valores de correlação são mais próximos de 1 possuem um tom azul mais intenso. As células cujos valores de correlação são mais próximos de -1 possuem um tom vermelho mais intenso.

	C_g	C_p	C_i	C_a	$\langle l \rangle$	$\langle c \rangle$	$\langle k_2 \rangle$	$\langle k_3 \rangle$	Δk	$\langle k \rangle$
C_g	1									
C_p	0,87	1								
C_i	-0,76	-0,98	1							
C_a	-0,36	-0,31	0,23	1						
$\langle l \rangle$	-0,76	-0,98	1	0,23	1					
$\langle c \rangle$	1	0,89	-0,78	-0,36	-0,78	1				
$\langle k_2 \rangle$	0,96	0,72	-0,59	-0,33	-0,59	0,95	1			
$\langle k_3 \rangle$	0,92	0,64	-0,51	-0,32	-0,51	0,9	0,99	1		
Δk	0,95	0,95	-0,87	-0,42	-0,87	0,96	0,84	0,77	1	
$\langle k \rangle$	1	0,87	-0,76	-0,36	-0,76	1	0,96	0,92	0,95	1

número médio de conexões dos nós, maior o intervalo em que $p(k)$ respeita a lei de potência dada pela equação (2.2). Ainda, pode-se dizer que $\langle k \rangle$ está fortemente e negativamente correlacionado com C_i e com $\langle l \rangle$; uma vez que, quanto maior o número médio de arestas por nó, menos nós intermediários são necessários para estabelecer conexões entre dois nós e menor a distância entre os mesmos.

4.1.1 Considerando os 100 nós mais relevantes

Um aspecto interessante de analisar é o comportamento das centralidades e do coeficiente de clusterização considerando apenas os 100 indivíduos mais relevantes, ou seja, os 100 indivíduos que apresentam os maiores valores dessas métricas.

Os valores das médias das centralidades e do coeficiente de clusterização para esses 100 indivíduos é mostrado na tabela 4.3. A partir dessa tabela, podem-se calcular as correlações. Na figura 4.4, apresenta-se a matriz de correlação para os 100 indivíduos com maiores valores de C_g , C_p , C_i , C_a e $\langle c \rangle$. O código de cores é o mesmo usado na figura 4.3.

Como nas análises em que se levam em conta todos os nós, nota-se que $\langle k \rangle$ está fortemente correlacionado positivamente com a média das 100 maiores centralidades de grau C_g e está também fortemente correlacionado positivamente com a média das 100 maiores centralidades de proximidade. Ainda, $\langle k \rangle$ está fortemente correlacionado negativamente com a média dos 100 maiores valores das centralidades de intermediação e de autovetor. Ainda, $\langle c \rangle$ aumenta, quando comparado com o grafo completo, mas a correlação com $\langle k \rangle$ diminui, se comparado com o valor apresentado na figura 4.3.

Das análises realizadas ao longo desta seção, constata-se que o algoritmo foi capaz de criar um grafo com o grau médio desejado (exceto para $\langle k \rangle_{alvo} = 5$), com a distribuição

Tabela 4.3. Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a e coeficiente de clusterização médio $\langle c \rangle$ em função de $\langle k \rangle$ para os 100 nós mais relevantes.

$\langle k \rangle_{alvo}$	C_g	C_p	C_i	C_a	$\langle c \rangle$	$\langle k \rangle$
5	0,0034	0,17	0,0107	0,0009	0,26	6,0
10	0,0050	0,21	0,0059	0,0007	0,17	10,4
15	0,0065	0,24	0,0044	0,0006	0,15	15,1
20	0,0086	0,26	0,0034	0,0006	0,15	20,1
25	0,0110	0,27	0,0033	0,0008	0,16	25,2
30	0,0114	0,28	0,0023	0,0005	0,15	30,1
35	0,0152	0,29	0,0026	0,0006	0,19	35,1
40	0,0153	0,30	0,0020	0,0005	0,18	40,1
45	0,0169	0,30	0,0019	0,0005	0,19	45,2
50	0,0182	0,31	0,0018	0,0005	0,20	50,0
60	0,0211	0,31	0,0016	0,0004	0,21	60,1
70	0,0233	0,32	0,0014	0,0004	0,23	70,6
80	0,0255	0,32	0,0013	0,0004	0,25	80,1
90	0,0280	0,32	0,0013	0,0004	0,27	90,1
100	0,0299	0,33	0,0012	0,0004	0,28	100,3

Figura 4.4. Matriz de correlação das métricas C_g , C_p , C_i , C_a e $\langle c \rangle$ em função de $\langle k \rangle$ para os 100 nós mais representativos. As células cujos valores de correlação são mais próximos de 1 possuem um tom azul mais intenso. As células cujos valores de correlação são mais próximos de -1 possuem um tom vermelho mais intenso.

	C_g	C_p	C_i	C_a	$\langle c \rangle$	$\langle k \rangle$
C_g	1					
C_p	0,9	1				
C_i	-0,77	-0,96	1			
C_a	-0,85	-0,87	0,86	1		
$\langle c \rangle$	0,68	0,3	-0,06	-0,35	1	
$\langle k \rangle$	0,99	0,85	-0,72	-0,84	0,73	1

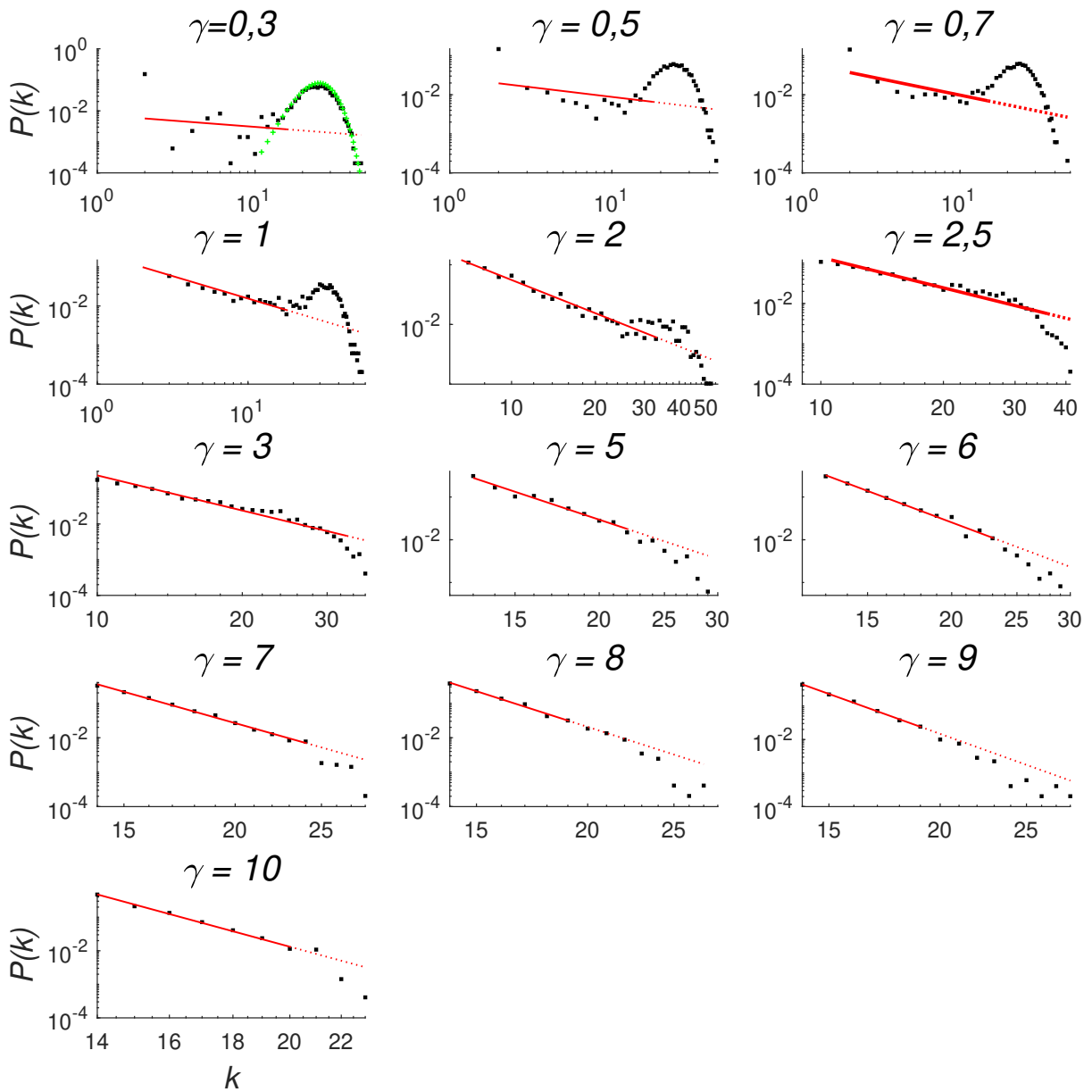
de graus exibindo comportamento livre-de-escala (ao menos até $k = k_{cut}$), com $\langle \ell \rangle$ “pequeno” (isto é, $\ell \ll N$) e com coeficiente de clusterização “grande” (isto é, $\langle c \rangle \gg \langle k \rangle / N$), aproximando o grafo gerado computacionalmente de redes sociais encontradas no mundo real. Assim, o algoritmo proposto gera grafos que combinam características de redes mundo-pequeno e redes livres-de-escala.

4.2 Variação do expoente γ

A figura 4.5 mostra a distribuição de graus $p(k)$ em função de k em gráficos log-log para $\gamma = \{0,3; 0,5; 0,7; 1; 2; 2,5; 3; 5; 6; 7; 8; 9; 10\}$, com $r = 10$ e $\langle k \rangle_{alvo} = 15$. Os gráficos

também contêm a regressão linear que melhor se ajusta ao valor do expoente simulado. Para o gráfico com $\gamma = 0,3$, nota-se que o comportamento da cauda se assemelha a uma distribuição de Poisson ajustada com $\langle k \rangle = 26$, exibida com cruzeiros ('+') em verde. Cálculos de centralidades e outras métricas topológicas são mostrados na tabela 4.4.

Figura 4.5. Distribuição de graus $p(k)$ em função de k em gráficos log-log. Nesses gráficos, $N = 4900$, $r = 10$, $\langle k \rangle_{alvo} = 15$ e $\gamma = \{0,3; 0,5; 0,7; 1; 2; 2,5; 3; 5; 6; 7; 8; 9; 10\}$. As escalas horizontais e as verticais não coincidem, a fim de facilitar a visualização das curvas plotadas.



Como mencionado na subseção 2.5.4, para redes reais, encontra-se na literatura $1,4 < \gamma < 6$ (Barabasi, 2016; Dorogovtsev, Goltsev e Mendes, 2008; Landherr, Friedl e Heidemann, 2010; Pastor-Satorras e Vespignani, 2004; Caldarelli, 2007). Barabasi (2016)

Tabela 4.4. Centralidades de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a , comprimento médio do caminho mais curto $\langle \ell \rangle$, coeficiente de clusterização médio $\langle c \rangle$ e grau médio $\langle k \rangle$ em função do expoente γ .

γ	C_g	C_p	C_i	C_a	$\langle \ell \rangle$	$\langle c \rangle$	$\langle k \rangle$
0,3	0,00439	0,25	0,00063	0,0002	4,1	0,07	21,5
0,5	0,00411	0,24	0,00064	0,0002	4,1	0,07	20,2
0,7	0,00386	0,24	0,00065	0,0002	4,2	0,07	18,9
1,0	0,00404	0,24	0,00065	0,0002	4,2	0,11	19,8
2,0	0,00308	0,23	0,00070	0,0002	4,4	0,09	15,1
2,5	0,00322	0,23	0,00070	0,0002	4,5	0,06	15,8
3,0	0,00308	0,22	0,00073	0,0002	4,6	0,05	15,1
5,0	0,00315	0,22	0,00074	0,0002	4,6	0,05	15,4
6,0	0,00312	0,22	0,00074	0,0002	4,6	0,05	15,3
7,0	0,00320	0,22	0,00074	0,0002	4,6	0,05	15,7
8,0	0,00313	0,21	0,00075	0,0002	4,7	0,04	15,3
9,0	0,00306	0,21	0,00075	0,0002	4,7	0,04	15,0
10,0	0,00308	0,21	0,00076	0,0002	4,7	0,04	15,1

e outros autores anteriores a ele (Monteiro, 2014) demonstraram que uma rede que cresce segundo a regra da ligação preferencial exibem uma distribuição de graus livre de escala em que $\gamma = 3$.

Nas simulações realizadas, escolheram-se alguns valores de γ , tal que $\gamma \leq 1$, levando a uma distribuição de graus que se distancia de uma livre de escala, como pode ser visto na figura acima. Nesses caso, não faz muito sentido realizar uma regressão linear, uma vez que a dispersão é visivelmente alta. Por outro lado, pode-se notar que, conforme γ aumenta, a parte de $p(k)$ que lembra uma curva de Poisson vai se deslocando para a direita (para graus maiores) e acaba se tornando o que é chamado de cauda. Para $\gamma = 0,3$ esse comportamento da cauda faz com que $\langle k \rangle$ assumam um valor de 21,5, muito superior ao esperado $\langle k \rangle_{alvo} = 15$. Entretanto, conforme γ se aproxima de 1, $\langle k \rangle$ diminui e passa-se a ter $\langle k \rangle \simeq \langle k \rangle_{alvo}$ para $\gamma \geq 2$.

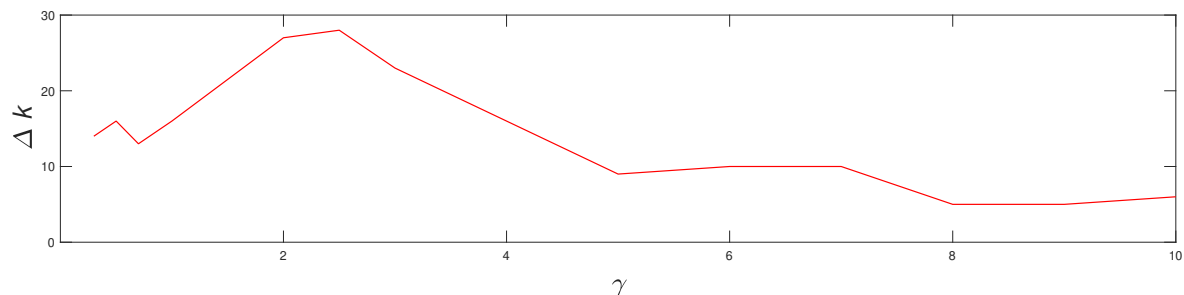
A análise da tabela 4.4 sugere que o aumento de γ causa aumentos em C_i e $\langle \ell \rangle$ e diminuição em $\langle c \rangle$. Ainda, os gráficos da figura 4.5 exibem comportamento livre-de-escala para $\gamma \geq 2$ e têm alta clusterização, como redes mundo-pequeno. Por exemplo, para $\gamma = 3$, $\langle c \rangle \simeq 0,05 \gg \langle k \rangle / N \simeq 0,003$.

Os momentos de segunda e terceira ordem, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, e o Δk de cada gráfico simulado são apresentados na tabela 4.5. Assim como nos casos analisados na seção anterior, observa-se que o momento de segunda ordem $\langle k_2 \rangle$ é grande quando comparado ao de redes totalmente aleatória. Por exemplo, para $\gamma = 2,5$, tem-se $\langle k_2 \rangle^{(1/2)} = 19 > \langle k \rangle^{(1/2)} \simeq 3,8$.

Tabela 4.5. Momentos de segunda e terceira ordens, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, desvio padrão de uma distribuição de Poisson $\langle k \rangle^{1/2}$, variância $\sigma^{1/2}$ e Δk dos grafos gerados.

γ	$\langle k \rangle$	$\langle k_2 \rangle$	$\langle k_3 \rangle$	$\langle k \rangle^{1/2}$	$\sigma^{1/2}$	Δk
0,3	15	423	16185	3,8	20,2	14
0,5	15	320	10847	3,8	17,5	16
0,7	15	430	16275	3,8	20,4	13
1,0	15	673	29318	3,8	25,7	16
2,0	15	531	19413	3,8	22,7	27
2,5	15	361	9016	3,8	18,6	28
3,0	15	287	6012	3,8	16,5	23
5,0	15	269	4799	3,8	15,9	9
6,0	15	280	5061	3,8	16,3	10
7,0	15	269	4739	3,8	16,0	10
8,0	15	255	4334	3,8	15,5	5
9,0	15	262	4502	3,8	15,7	5
10,0	15	236	3727	3,8	14,9	6

Figura 4.6. Variação de Δk em função de γ . Lembre que Δk é intervalo de graus em que o grafo obedece à lei de potência.



Com os valores de γ e de Δk apresentados na tabela 4.5, traça-se o gráfico mostrado na figura 4.6. Essa figura revela, de forma surpreendente, que valores mais altos de Δk ocorrem entre $2 \leq \gamma \leq 3$, com máximo em $\gamma = 2,5$. Essa é a faixa de valores de γ mais prevalente em redes reais, de acordo com a literatura (Barabasi, 2016; Dorogovtsev, Goltsev e Mendes, 2008; Landherr, Friedl e Heidemann, 2010; Monteiro, 2014; Pastor-Satorras e Vespignani, 2004; Travers e Milgram, 1969).

Considerando os dados apresentados nas tabelas 4.4 e 4.5, podem-se calcular as correlações por meio da equação (2.19). A figura 4.7 apresenta a matriz de correlação de C_g , C_p , C_i , C_a , $\langle \ell \rangle$, $\langle c \rangle$, $\langle k \rangle$, $\langle k_2 \rangle$, $\langle k_3 \rangle$, Δk e γ . Como anteriormente, quanto mais próximo de 1 for a correlação de duas variáveis, mais próxima a célula fica de um azul mais intenso; quanto mais próximo de -1 , mais próxima a célula fica de um vermelho mais intenso.

Figura 4.7. Matriz de correlação das métricas C_g , C_p , C_i , C_a , $\langle \ell \rangle$, $\langle c \rangle$, $\langle k \rangle$, $\langle k_2 \rangle$, $\langle k_3 \rangle$, Δk e γ , considerando os grafos obtidos variando-se o expoente γ . O código de cores é o mesmo usado nas matrizes anteriores similares.

	C_g	C_p	C_i	C_a	$\langle \ell \rangle$	$\langle c \rangle$	$\langle k_2 \rangle$	$\langle k_3 \rangle$	Δk	$\langle k \rangle$	γ
C_g	1										
C_p	0,94	1									
C_i	-0,94	-1	1								
C_a	-0,17	-0,12	0,11	1							
$\langle \ell \rangle$	-0,94	-1	1	0,11	1						
$\langle c \rangle$	0,64	0,77	-0,77	-0,37	-0,77	1					
$\langle k_2 \rangle$	0,53	0,68	-0,68	-0,41	-0,68	0,96	1				
$\langle k_3 \rangle$	0,65	0,77	-0,77	-0,39	-0,77	0,98	0,99	1			
Δk	0,09	0,37	-0,39	-0,1	-0,39	0,49	0,49	0,44	1		
$\langle k \rangle$	1	0,94	-0,94	-0,17	-0,94	0,64	0,53	0,65	0,09	1	
γ	-0,72	-0,89	0,9	0,05	0,9	-0,76	-0,69	-0,73	-0,68	-0,72	1

Observe que γ está fortemente correlacionado positivamente com C_i e com $\langle \ell \rangle$. Como se observa na figura 4.7, conforme se aumenta γ , o número de nós com grau grande (os *hubs*) diminui. Dessa forma, a quantidade de nós intermediários cresce, refletindo um aumento de C_i , bem como de $\langle \ell \rangle$, já que os indivíduos passam a ficar mais distantes uns dos outros. O expoente γ também está fortemente negativamente correlacionado a C_p , o que endossa a análise realizada.

4.2.1 Considerando os 100 nós mais relevantes

A partir das simulações realizadas em que se varia o expoente γ , é interessante analisar o comportamento das centralidades e do coeficiente de clusterização considerando apenas os 100 indivíduos mais representativos, ou seja, os 100 indivíduos que apresentam os maiores valores para essas métricas.

Os valores das médias das centralidades e do coeficiente de clusterização para esses indivíduos são mostrado na tabela 4.6. Na figura 4.8, apresenta-se a matriz de correlação correspondente, adotando o código de cores já explicado.

Percebe-se que γ está fortemente correlacionado negativamente com a médias das 100 maiores centralidades de grau e de proximidade. Esse resultado reforça a ideia de que aumentar γ diminui a probabilidade de existir nós que atuem como grandes *hubs* (Barabasi, 2016; Dorogovtsev, Goltsev e Mendes, 2008; Landherr, Friedl e Heidemann, 2010).

4.3 Variação do raio de vizinhança r

A figura 4.9 contém os gráficos log-log da distribuição de graus $p(k)$ em função de k , para diferentes combinações de γ e r . Em todos, $\langle k \rangle_{alvo} = 20$. Há seis gráficos em que

Tabela 4.6. Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a e coeficiente de clusterização médio $\langle c \rangle$ em função de γ para os 100 nós mais relevantes.

γ	C_g	C_p	C_i	C_a	$\langle c \rangle$
0,3	0,0078	0,26	0,0020	0,0006	0,78
0,5	0,0076	0,26	0,0021	0,0005	0,89
0,7	0,0074	0,26	0,0024	0,0005	0,82
1,0	0,0096	0,26	0,0029	0,0009	1,00
2,0	0,0097	0,26	0,0056	0,0015	0,28
2,5	0,0072	0,25	0,0042	0,0008	0,17
3,0	0,0065	0,24	0,0044	0,0006	0,15
5,0	0,0054	0,23	0,0033	0,0004	0,13
6,0	0,0052	0,23	0,0034	0,0004	0,12
7,0	0,0050	0,23	0,0032	0,0004	0,12
8,0	0,0046	0,23	0,0030	0,0004	0,12
9,0	0,0044	0,23	0,0031	0,0003	0,12
10,0	0,0042	0,23	0,0030	0,0003	0,11

Figura 4.8. Matriz de correlação das métricas C_g , C_p , C_i , C_a e $\langle c \rangle$, em função de γ para os 100 nós mais representativos. Como nas outras matrizes de correlação, azul mais intenso denota correlação mais próxima de 1; vermelho mais intenso denota correlação mais próxima de -1 .

	γ	C_g	C_p	C_i	C_a	$\langle c \rangle$
γ	1					
C_g	-0,88	1				
C_p	-0,94	0,94	1			
C_i	0,06	0,2	-0,11	1		
C_a	-0,56	0,84	0,64	0,65	1	
$\langle c \rangle$	-0,77	0,7	0,86	-0,54	0,24	1

$\gamma = 2$ e $r = \{5, 10, 20, 30, 40, 60\}$ e três gráficos em que $\gamma = 3$ e $r = \{5, 20, 60\}$. Esses gráficos também mostram a regressão linear que melhor se ajusta ao expoente γ escolhido. Valores das centralidades e das outras métricas topológicas são apresentados na tabela 4.7.

Na tabela 4.7, pode-se perceber que o grau médio das rede complexas geradas está muito próximo ao grau médio desejado, isto é, $\langle k \rangle \simeq \langle k \rangle_{alvo} = 20$. Ainda, o algoritmo foi capaz de criar grafos com o expoente γ escolhido, como pode ser notado na figura 4.9.

Aumentar o raio r diminui a clusterização, pois as ligações podem ocorrer entre indivíduos cada vez mais distantes. Entretanto, mesmo para $r = 60$ (que é um valor grande, pois o reticulado é 70×70), o valor de $\langle c \rangle$ dos grafos é maior do que para uma

Figura 4.9. Distribuição de graus $p(k)$ em um gráfico log-log para $N = 4900$ e $\langle k \rangle_{alvo} = 20$. Para $\gamma = 2$, geram-se grafos com $r = \{5, 10, 20, 30, 40, 60\}$; para $\gamma = 3$, geram-se grafos com $r = \{5, 20, 60\}$. As escalas horizontais não são iguais nesses nove gráficos, a fim de facilitar a visualização das distribuições mostradas.

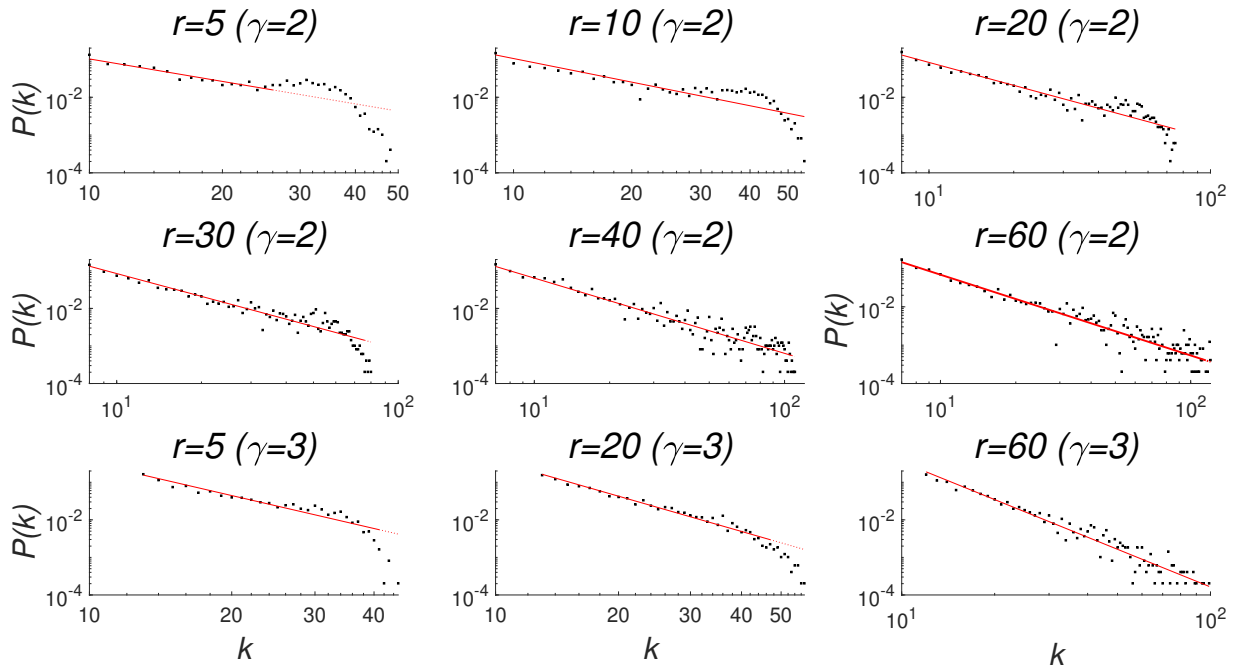


Tabela 4.7. Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a , comprimento médio do caminho mais curto $\langle \ell \rangle$, coeficiente de clusterização médio $\langle c \rangle$ e grau médio $\langle k \rangle$ em função de γ e r .

γ	r	C_g	C_p	C_i	C_a	$\langle \ell \rangle$	$\langle c \rangle$	$\langle k \rangle$
2	5	0,00409	0,17	0,00103	0,0002	6,0	0,19	20,0
2	10	0,00417	0,24	0,00064	0,0002	4,2	0,08	20,4
2	20	0,00414	0,30	0,00048	0,0002	3,4	0,04	20,3
2	30	0,00420	0,32	0,00044	0,0002	3,2	0,03	20,6
2	40	0,00410	0,32	0,00044	0,0002	3,1	0,03	20,1
2	60	0,00410	0,33	0,00042	0,0002	3,1	0,02	20,1
3	5	0,00409	0,16	0,00103	0,0002	6,1	0,17	20,0
3	20	0,00413	0,29	0,00050	0,0002	3,4	0,03	20,3
3	60	0,00409	0,32	0,00044	0,0002	3,2	0,01	20,0

rede totalmente aleatória. Por exemplo, para $\gamma = 3$ e $r = 60$, obteve-se $\langle c \rangle \simeq 0,01 > \langle k \rangle / N \simeq 0,004$. Ainda, aumentar γ de 2 para 3 também diminui $\langle c \rangle$, como já havia sido comentado na seção anterior.

No que diz respeito a $\langle \ell \rangle$, a expressão $\langle \ell \rangle \approx \log(N) / \log(\langle k \rangle)$ fornece uma estimativa razoável. Por exemplo, $\langle \ell \rangle \approx \log(N) / \log(\langle k \rangle) \simeq 2,84$ para $N = 4900$ e $\langle k \rangle_{alvo} = 20$. Os valores de ℓ mostrados na tabela 4.7 são maiores, mas dessa mesma ordem de grandeza.

Ainda, na tabela 4.7, verifica-se que a centralidade de grau tem pouco de sensibi-

lidade com a variação de r , pois está atrelada a $\langle k \rangle_{alvo}$. A centralidade de proximidade aumenta à medida que r aumenta, coerente a diminuição de $\langle \ell \rangle$. De fato, aumentar a área onde as conexões podem ser realizadas encurta os caminhos entre os nós da rede. Já a centralidade de intermediação diminui com o aumento de r , pois o número de intermediários entre qualquer par de nós diminui.

A figura 4.9 mostra que o comportamento tipo-Poisson da cauda diminui conforme r aumenta. Esse resultado se explica pelo fato de que cada indivíduo sofre menos restrições geográficas com o aumento de r , reduzindo o efeito de mundo-pequeno.

Os momentos de segunda e terceira ordem, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, respectivamente, o desvio padrão de Poisson $\langle k \rangle^{1/2}$, a raiz quadrada da variância σ e o intervalo Δk , no qual a rede exhibe comportamento livre-de-escala, são apresentados na tabela 4.8.

Tabela 4.8. Momentos de segunda e terceira ordens, $\langle k_2 \rangle$ e $\langle k_3 \rangle$, desvio padrão de uma distribuição de Poisson $\langle k \rangle^{1/2}$, variância $\sigma^{1/2}$ e Δk dos grafos gerados.

γ	r	$\langle k_2 \rangle$	$\langle k_3 \rangle$	$\langle k \rangle^{1/2}$	$\sigma^{1/2}$	Δk
2	5	498	15865	4,4	21,9	16
2	10	547	17901	4,4	23,0	46
2	20	626	26388	4,4	24,6	67
2	30	880	46314	4,4	29,3	68
2	40	1096	82111	4,4	32,8	100
2	60	1276	108706	4,4	35,4	109
3	5	522	14442	4,4	22,4	28
3	20	726	26356	4,4	26,6	34
3	60	515	18070	4,4	22,3	89

Como nos dois casos previamente analisados, em que variaram $\langle k \rangle_{alvo}$ e γ , aqui também se observa que $\sigma^{1/2}$ é grande comparado com o de redes puramente aleatórias. O aumento de r leva, em geral, a aumentos de $\langle k_2 \rangle$, $\langle k_3 \rangle$ e Δk . Para $r = 60$, Δk possui um tamanho de maior ou próximo a duas décadas.

A figura 4.10 apresenta a matriz de correlação das métricas C_g , C_p , C_i , C_a , $\langle \ell \rangle$, $\langle c \rangle$, r , $\langle k_2 \rangle$, $\langle k_3 \rangle$, Δk , r e γ . É possível observar que a variação de γ entre 2 e 3 está fracamente correlacionada com as demais métricas apresentadas. Já o raio r está positivamente fortemente correlacionado com o intervalo Δk e com C_p . Quanto maior r , maior a área da vizinhança de um certo nó, onde residem os candidatos à conexão; por consequência, espera-se que C_p aumente. O aparecimento de *hubs* com grau elevado aumenta o efeito livre-de-escala e o Δk pode ocupar duas décadas, como encontrado em redes reais da literatura (Barabasi, 2016; Clauset, Shalizi e Newman, 2009; Dorogovtsev, Goltsev e Mendes, 2008; Landherr, Friedl e Heidemann, 2010; Pastor-Satorras e Vespignani, 2004; Caldarelli, 2007). Por outro lado, r está negativa e fortemente correlacionado com C_i , com

$\langle \ell \rangle$ e com $\langle c \rangle$. Com a diminuição da distância geográfica entre os nós vizinhos, aumenta o número de intermediários, o comprimento do menor caminho entre os nós e a clusterização local.

Figura 4.10. Matriz de correlação das métricas $C_g, C_p, C_i, C_a, \langle \ell \rangle, \langle c \rangle, \langle k \rangle, \langle k_2 \rangle, \langle k_3 \rangle, r$ e γ , considerando os grafos obtidos variando-se o raio r e o expoente γ . Como antes, azul mais intenso indica correlação mais próxima de 1, vermelho mais intenso indica correlação mais próxima de -1 .

	C_g	C_p	C_i	C_a	$\langle \ell \rangle$	$\langle c \rangle$	$\langle k \rangle$	$\langle k_2 \rangle$	$\langle k_3 \rangle$	Δk	r	γ
C_g	1											
C_p	0,29	1										
C_i	-0,37	-0,99	1									
C_a	0,05	0,24	-0,18	1								
$\langle \ell \rangle$	-0,37	-0,99	1	-0,18	1							
$\langle c \rangle$	-0,33	-0,98	0,99	-0,16	0,99	1						
$\langle k \rangle$	0,01	0,64	-0,58	0,05	-0,58	-0,55	1					
$\langle k_2 \rangle$	-0,11	0,59	-0,53	0,05	-0,53	-0,49	0,98	1				
$\langle k_3 \rangle$	-0,06	0,84	-0,79	0,29	-0,79	-0,78	0,73	0,78	1			
Δk	-0,21	0,79	-0,74	0,22	-0,74	-0,75	0,61	0,65	0,9	1		
r	-0,33	-0,16	0,16	-0,22	0,16	0,04	-0,41	-0,44	-0,26	0,02	1	
γ												1

4.3.1 Considerando os 100 nós mais relevantes

Como realizado nas duas seções anteriores, agora analisa-se o comportamento das centralidades e do coeficiente de clusterização considerando apenas os 100 indivíduos mais representativos, ou seja, aqueles que têm os 100 maiores valores para essas métricas topológicas. Na tabela 4.9, são exibidos os resultados. A figura 4.11 contém a respectiva matriz de correlação.

Tabela 4.9. Centralidades médias de grau C_g , de proximidade C_p , de intermediação C_i e de autovetor C_a e coeficiente de clusterização médio $\langle c \rangle$ em função de r e γ para os 100 nós mais relevantes.

r	γ	C_g	C_p	C_i	C_a	$\langle c \rangle$
5	2	0,0085	0,17	0,0060	0,0014	0,36
10	2	0,0101	0,27	0,0038	0,0010	0,21
20	2	0,0138	0,36	0,0039	0,0010	0,15
30	2	0,0145	0,38	0,0041	0,0009	0,11
40	2	0,0198	0,39	0,0058	0,0018	0,12
60	2	0,0221	0,40	0,0064	0,0017	0,11
5	3	0,0079	0,17	0,0058	0,0014	0,30
20	3	0,0098	0,33	0,0033	0,0006	0,09
60	3	0,0136	0,36	0,0041	0,0011	0,06

Nota-se que a variação do expoente γ , com $\gamma = 2$ ou 3 , está pouco correlacionada

com as médias dos 100 maiores valores das quatro centralidades e do coeficiente de clusterização agregação. Já o raio r está positiva e fortemente correlacionado com C_g e C_p e negativa e fortemente correlacionado com $\langle c \rangle$, o que está de acordo com a análise realizada anteriormente.

Figura 4.11. Matriz de correlação das métricas C_g , C_p , C_i , C_a e $\langle c \rangle$ em função de r e γ . O código de cores é o mesmo até então usado.

	r	γ	C_g	C_p	C_i	C_a	$\langle c \rangle$
r	1						
γ	0,02	1					
C_g	0,81	-0,44	1				
C_p	0,8	-0,22	0,82	1			
C_i	0,12	-0,25	0,34	-0,23	1		
C_a	0,34	-0,34	0,57	0,03	0,91	1	
$\langle c \rangle$	-0,76	-0,14	-0,59	-0,92	0,47	0,23	1

4.4 Resumo dos resultados

A figura 4.12 sumariza as relações entre os parâmetros da rede e as métricas calculadas nas seções 4.1, 4.2 e 4.3, considerando todos os nós do grafo. Essa figura apresenta as relações obtidas por meio de um diagrama com flechas. A flecha \uparrow indica aumento da métrica com o aumento do parâmetro (correlação positiva). A flecha \downarrow indica diminuição da métrica com o aumento do parâmetro (correlação negativa). O símbolo - denota ausência de variação significativa (ausência de correlação significativa).

Figura 4.12. Sumário de tendências das métricas C_g , C_p , C_i , C_a , $\langle \ell \rangle$, $\langle c \rangle$, $\langle k_2 \rangle$, $\langle k_3 \rangle$ e Δk em função de $\langle k \rangle_{alvo}$, γ e r , considerando todos os nós do grafo.

		C_g	C_p	C_i	C_a	$\langle \ell \rangle$	$\langle c \rangle$	$\langle k_2 \rangle$	$\langle k_3 \rangle$	Δk
$\langle k \rangle_{alvo}$	\uparrow	\uparrow	\uparrow	\downarrow	-	\downarrow	\uparrow	\uparrow	\uparrow	\uparrow
γ	\uparrow	\downarrow	\downarrow	\uparrow	-	\uparrow	\downarrow	\downarrow	\downarrow	\downarrow
r	\uparrow	-	\uparrow	\downarrow	-	\downarrow	\downarrow	\uparrow	\uparrow	\uparrow

A figura 4.13 sumariza as relações entre as métricas e os parâmetros da rede considerando os 100 nós mais relevantes. A simbologia é a mesma usada na figura 4.12.

Adicionalmente, usa-se a cor vermelha para as métricas que apresentaram um comportamento diferente daquele mostrado na figura 4.12.

Figura 4.13. Sumário de tendências das métricas C_g , C_p , C_i , C_a e $\langle c \rangle$ em função de $\langle k \rangle_{alvo}$, γ e r , considerando os 100 nós mais relevantes.

		C_g	C_p	C_i	C_a	$\langle c \rangle$
$\langle k \rangle_{alvo}$	↑	↑	↑	↓	↓	↑
γ	↑	↓	↓	-	↓	↓
r	↑	↑	↑	-	-	↓

4.5 Características do modelo proposto para gerar redes complexas

Dentre os diversos aspectos apresentados e discutidos neste capítulo, cabe fazer os seguintes destaques:

1. Ajuste de $\langle k \rangle_{alvo}$ e γ

O algoritmo proposto é capaz de, em geral, gerar grafos com os valores escolhidos para $\langle k \rangle_{alvo}$ e γ . Não se encontrou na literatura um algoritmo com essa característica.

2. Simplicidade

O modelo necessita de poucos parâmetros (relembrando, os parâmetros são N , $\langle k \rangle_{alvo}$, γ e r). A ideia de se estabelecer uma área onde as conexões podem ser feitas é intuitiva. Entretanto, essa ideia já havia sido usada em outros modelos.

3. Transição de redes mundo-pequeno para redes livre-de-escala

As simulações permitem olhar para o comportamento de cauda que aparece em $p(k)$ segundo uma perspectiva diferente. No modelo aqui proposto, conforme γ passa de valores menores do que um para valores maiores do que um, $p(k)$ passa de uma distribuição do tipo Poisson para uma do tipo livre-de-escala, em que a cauda pode ser vista como um resquício da distribuição de Poisson. Isso decorre das restrições de conectividade geográfica dos indivíduos. Tal fenômeno não parece ter sido relatado na literatura, pois os trabalhos geralmente dão foco nas redes que são ou livre-de-escala ou mundo-pequeno (Barabasi, 2016; Mullick e Sen, 2020; Reggiani

e Vinciguerra, 2007; Zheng et al., 2014). Estudos sobre a transição entre esses dois modelos de redes complexas não foram encontrados na literatura.

Parte desses resultados foram relatados num artigo publicado em periódico internacional (Licciardi e Monteiro, 2024).

No próximo capítulo, o modelo até então usado é adaptado para levar em consideração as classes socioeconômicas dos indivíduos. Então, no capítulo 6, são apresentados os resultados das simulações para três países: França, Peru e Zimbábue.

5 REDES SOCIAIS COM CLASSES SOCIOECONÔMICAS

Neste capítulo, é proposto um modelo de redes sociais complexas que considera relacionamentos de indivíduos pertencentes a diferentes classes socioeconômicas. O modelo aqui apresentado é uma variação do algoritmo 3.1 mostrado no capítulo 3. As principais características desse modelo são:

1. as conexões são principalmente locais (com os indivíduos geograficamente mais próximos), mas são permitidas ligações com nós mais distantes, dentro do raio de Moore r , como explicado no capítulo 2 (Licciardi e Monteiro, 2022; Monteiro, Paiva e Piqueira, 2006; Schimit e Monteiro, 2009);
2. a rede é construída para que a distribuição de graus $p(k)$ seja livre-de-escala, sendo gerada a partir de uma variação do algoritmo descrito no capítulo 3. Ainda, o grafo gerado deve ter o grau médio desejado. Esse valor de $\langle k \rangle_{alvo}$ pode ser obtido na literatura e caracteriza o país estudado (Béraud et al., 2015; Grijalva et al., 2015; Melegaro et al., 2017);
3. as conexões entre os nós representam contatos presenciais entre familiares, amigos, vizinhos. Essas conexões correspondem, em geral, a contatos dentro de uma mesma classe social, o que está de acordo com a propriedade de homofilia das redes sociais (Boucher, 2015; Chodrow, 2017; McPherson, Smith-Lovin e Cook, 2001; Xu et al., 2019). Entretanto, são também permitidas conexões entre classes sociais diferentes, como as que podem ocorrer, por exemplo, durante um passeio em locais públicos e nos locais de trabalho (Béraud et al., 2015; Grusky, 2019; Noble e Tynes, 2016).

As características apresentadas nos itens 1 e 2 acima foram usadas no algoritmo 3.1 e são comentadas brevemente no início da próxima seção. Para que seja possível incorporar o que está descrito no item 3, ajustes são propostos e efetuados. O algoritmo final é apresentado na seção 5.2. Esse algoritmo pode ser utilizado para a construção de redes complexas baseadas nas características socioeconômicas de regiões ou países. Nesta tese, são considerados três países: Zimbábue, Peru e França. Assim, para o bem da diversidade, lida-se com um país africano, um sul-americano e um europeu.

5.1 Incluindo classes sociais

A fim de representar indivíduos vivendo numa região geográfica, usa-se um reticulado bidimensional com condição de contorno periódica, como feito no capítulo 3. Assim, considere um reticulado quadrado de lado \sqrt{N} , de modo que a rede possui N indivíduos. As bordas inferior e superior desse reticulado estão conectadas, bem com as bordas esquerda e direita. Assim, todos os indivíduos são geograficamente equivalentes.

Na inicialização, a distribuição dos indivíduos de diversas classes sociais pelo reticulado é aleatória, respeitando apenas a porcentagem deles na população. Essa porcentagem é definida pelos parâmetros pA , pB e pC , apresentados na tabela 2.1. Esses parâmetros representam as porcentagens da população nas classes A, B e C, respectivamente, nos países mencionados. Como dito no capítulo 3, cada célula do reticulado corresponde a um indivíduo, que agora pertence a uma classe socioeconômica. A probabilidade de conexão dos indivíduos, dentro de uma vizinhança de Moore de raio r , continua sendo dada pela equação (3.1).

No modelo apresentado neste capítulo, a chance de dois indivíduos se conectarem depende não apenas da distância, mas depende também da classe social a que eles pertencem. Assume-se que eles possuem maior chance de conexão se pertencerem à mesma classe social, face ao conceito de homofilia (Boucher, 2015; Chodrow, 2017; Licciardi e Monteiro, 2022; McPherson, Smith-Lovin e Cook, 2001; Xu et al., 2019). Para que o modelo de rede complexa possa ponderar de modo diferente a conexão entre diferentes classes, estabelecem-se pesos de conexão entre as classes. Esses pesos são representados pela matriz de \mathbf{f} , dada por:

$$\mathbf{f} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \quad (5.1)$$

em que f_{mj} representa o peso de um indivíduo da classe m estabelecer conexão com um indivíduo da classe j . Assume-se que a classe A é representada pelo número 3; a classe B, pelo número 2; e a classe C, pelo número 1. Ainda, assume-se que \mathbf{f} é uma matriz simétrica. Por exemplo, $f_{12} = f_{21}$ corresponde ao peso da conexão entre dois indivíduos, um da classe B e outro da classe C. A ideia é de que quanto maior a ponderação, maior a probabilidade p_{mj} de indivíduos das classes m e j se conectarem.

Considere que um indivíduo da classe $m = \{1, 2, 3\}$ tenha n_j vizinhos da classe $j = \{1, 2, 3\}$. Por exemplo, n_3 é o número de vizinhos que pertencem à classe A. Nesta tese, define-se a probabilidade p_{mj} desse indivíduo da classe m se conectar a um indivíduo

da classe j como:

$$p_{mj} = \frac{f_{mj}n_j}{\sum_{z=1}^3 f_{mz}n_z} \quad (5.2)$$

Evidentemente:

$$\sum_{j=1}^3 p_{mj} = 1 \quad (5.3)$$

A probabilidade P_{mji} de um indivíduo da classe m se conectar a um indivíduo da classe j (com $m, j = 1, 2, 3$) que se encontra a uma distância com raio de Moore i (com $i = 1, 2, \dots, r$) é dada por:

$$P_{mji} = p_{mj}q_i = \frac{f_{mj}n_j}{\sum_{z=1}^3 f_{mz}n_z} \left(\frac{2(r+1-i)}{r(r+1)} \right) \quad (5.4)$$

levando-se em conta que q_i é dada pela equação (3.1) e p_{mj} é dada pela equação (5.2). Evidentemente:

$$\sum_{i=1}^r \sum_{j=1}^3 P_{mji} = 1 \quad (5.5)$$

Note que a equação (5.5) pode ser facilmente adaptada para incluir um número maior ou menor de classes socioeconômicas.

Na próxima seção, apresenta-se a versão final do algoritmo, que considera uma distribuição de graus que obedece à lei de potência $p(k) = Ak^{-\gamma}$, com o expoente γ sendo um parâmetro ajustável. Tipicamente, em redes sociais, tem-se $2 \leq \gamma \leq 3$ (Barabasi, 2016; Barabási e Albert, 1999; Barabási, Albert e Jeong, 2000; Clauset, Shalizi e Newman, 2009; Licciardi e Monteiro, 2022; Mitzenmacher, 2004; Voutsas et al., 2021; Wang e Chen, 2003).

5.2 Modelo de rede social livre-de-escala

Nas simulações apresentadas no capítulo 6, toma-se $\gamma = 2,5$. De acordo com a figura 4.6 mostradas na seção 4.2, esse é valor de γ que maximiza Δk ; ou seja, que maximiza o intervalo em k para o qual o grafo obedece à lei de potência. Ainda, deve-se realizar um cálculo iterativo de k_{min} e k_{max} para que o grafo tenha o grau médio $\langle k \rangle_{alvo}$ previamente escolhido, como explicado no capítulo 3.

A seguir, apresenta-se o algoritmo computacional usado para gerar grafos com conectividade ajustável que representam redes de contatos sociais presenciais, em que esses contatos são influenciados pela distância geográfica entre os indivíduos e pela classe socioeconômica às quais esses indivíduos pertencem. Agora, além dos parâmetros N , r , γ e $\langle k \rangle_{alvo}$, devem-se fornecer valores para os parâmetros pA , pB e pC (referentes às porcentagens da cada classe na população) e para os elementos da matriz \mathbf{f} (que pondera as conexões entre as diferentes classes sociais).

O algoritmo mostrado abaixo é uma variação do algoritmo 3.1.

Algoritmo 5.1 Implementação da rede complexa de contatos sociais

- 1: Escolha valores para pA , pB , pC e para os elementos que constituem a matriz \mathbf{f}
 - 2: Escolha valores para N , r , γ e $\langle k \rangle_{alvo}$
 - 3: A partir de $\langle k \rangle_{alvo}$, determine k_{min} e k_{max}
 - 4: Calcule n_k para $k_{min} \leq k \leq k_{max}$
 - 5: Atribua para cada nó um valor de k , respeitando as quantidades n_k de nós com grau k calculadas no passo anterior
 - 6: Crie (de modo probabilista) conexões entre os nós, de acordo com a probabilidade P_{mji} com $1 \leq i \leq r$ e $m, j = 1, 2, 3$ (autoconexões ou conexões múltiplas entre nós não são permitidas)
 - 7: Caso existam subgrafos isolados, eles são ligados aleatoriamente ao grafo principal
 - 8: Após a criação de um grafo conexo, compute $\langle k \rangle$. Em função das restrições de vizinhança, acrescidas da restrição de n_k , o grau médio pode estar abaixo de $\langle k \rangle_{alvo}$. Se isso acontecer, retorne ao passo 6 e construa um novo grafo
 - 9: Caso, após 10 tentativas, $\langle k \rangle$ continue abaixo de $\langle k \rangle_{alvo}$, repita a rotina a partir do passo 3, adicionando 1 ao k_{min} até então usado
 - 10: O processo iterativo é encerrado somente quando $\langle k \rangle$ for igual ou ligeiramente maior que $\langle k \rangle_{alvo}$
-

No próximo capítulo, o algoritmo 5.1 é utilizado para a geração de grafos que representam contatos sociais presenciais em três países de três continentes diferentes.

6 GRAFOS DE CONTATOS SOCIAIS PARA TRÊS PAÍSES

Neste capítulo, relatam-se as simulações realizadas empregando o algoritmo de redes complexas descrito no capítulo anterior. Os estudos que tratam de levantamentos de campo sobre contatos sociais nos países mencionados são brevemente descritos na seção 6.1. Na seção 6.2, são feitas considerações sobre as matrizes de pesos \mathbf{f} utilizadas nas simulações. Na seção 6.3, listam-se os valores dos parâmetros usados na geração dos grafos, que são caracterizados calculando as métricas já apresentadas. Na seção 6.4, analisa-se a distribuição de graus dos grafos gerados. Na seção 6.5, discutem-se os resultados.

6.1 Estudos de campo sobre contatos sociais

Na literatura, encontram-se trabalhos de pesquisa de campo que se baseiam em anotações em diários a respeito dos contatos sociais realizados. Assim, os indivíduos, que aceitam participar da pesquisa, passam a efetuar anotações sobre suas conexões sociais em um diário, que ao final da pesquisa é analisado. Por meio desses (raros) estudos, é possível obter o número médio de conexões por indivíduo. Nesta tese, são utilizados os dados obtidos nas seguintes pesquisas de campo:

- artigo de Béraud et al. (2015), sobre a França;
- artigo de Melegaro et al. (2017), sobre o Zimbábue;
- artigo de Grijalva et al. (2015), sobre o Peru.

A tabela 6.1 apresenta a média de contatos por indivíduo, o que, no algoritmo proposto, corresponde ao valor de $\langle k \rangle_{alvo}$.

Tabela 6.1. Média diária de contatos por indivíduo para os países estudados.

País	Número médio de conexões
França	16,5
Zimbábue	10,9
Peru	16,1

Fonte: (Béraud et al., 2015; Grijalva et al., 2015; Melegaro et al., 2017).

6.2 Considerações sobre a matriz de pesos

Para que a construção dos grafos de redes sociais com classes socioeconômicas, é necessário especificar os pesos da matriz \mathbf{f} do modelo. Por simplicidade, considera-se que essa matriz pode ser escrita em função de um único coeficiente f , com $f > 1$. Assim, assume-se que:

$$\mathbf{f} = \begin{bmatrix} f & f/2 & 1 \\ f/2 & f & f/2 \\ 1 & f/2 & f \end{bmatrix} \quad (6.1)$$

Observe que os maiores pesos valem f e se referem a conexões entre indivíduos da mesma classe. Essa escolha está de acordo com os resultados de estudos sobre homofilia 2.3 (Boucher, 2015; Chodrow, 2017; Currarini, Jackson e Pin, 2009; Gent, Das e Musterd, 2019; McPherson, Smith-Lovin e Cook, 2001; Xu et al., 2019). Os menores pesos valem 1 e correspondem a conexões entre as classes A e C, socioeconomicamente mais distantes entre si. Conexões entre classes A e B e conexões entre classes B e C são ponderadas pelo peso $f/2$.

Ainda, assume-se que f diminui de acordo com o IDH (o Índice de Desenvolvimento Humano da ONU apresentado na subseção 2.3.1) e aumenta com o coeficiente de Gini (apresentado na subseção 2.3.2). No relatório do Programa das Nações Unidas para o Desenvolvimento (UNDP, 2022), verifica-se que:

$$\mathbf{IDH}_{França} > \mathbf{IDH}_{Peru} > \mathbf{IDH}_{Zimbábue} \quad (6.2)$$

e que:

$$\mathbf{Gini}_{França} < \mathbf{Gini}_{Peru} < \mathbf{Gini}_{Zimbábue} \quad (6.3)$$

Portanto:

$$f_{França} < f_{Peru} < f_{Zimbábue} \quad (6.4)$$

Escolhe-se $f = 4$ para a França, $f = 10$ para o Peru e $f = 40$ para o Zimbábue. Assim, a matriz de pesos possui valores mais próximos para a França e mais distantes para o Zimbábue, o que é consistente com o fato de a França ser um país menos desigual, do ponto de vista socioeconômico, do que o Zimbábue. O Peru ficaria no meio termo.

Explicitamente, tem-se que:

$$\mathbf{f}_{França} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 4 \end{bmatrix} \quad (6.5)$$

$$\mathbf{f}_{Peru} = \begin{bmatrix} 10 & 5 & 1 \\ 5 & 10 & 5 \\ 1 & 5 & 10 \end{bmatrix} \quad (6.6)$$

$$\mathbf{f}_{Zimbábue} = \begin{bmatrix} 40 & 20 & 1 \\ 20 & 40 & 20 \\ 1 & 20 & 40 \end{bmatrix} \quad (6.7)$$

Na próxima seção, são apresentados os parâmetros utilizados na construção das redes complexas dos três países considerados, utilizando o algoritmo 5.1. São também apresentados os resultados das análises dessas redes.

6.3 Resultados para os três países

A tabela 6.2 reúne os valores dos parâmetros usados para gerar as redes complexas para os três países, usando o algoritmo 5.1. Para os três países, toma-se $N = 10000$, $\gamma = 2,5$ e $r = 10$. Tomar $r = 10$ implica haver $((2 \times 10)^2 - 1) = 440$ indivíduos que podem estabelecer conexões com cada indivíduo do reticulado.

Tabela 6.2. Valores dos parâmetros utilizados na geração dos grafos de contatos sociais para França, Zimbábue e Peru.

país	N	r	γ	$\langle k \rangle_{alvo}$	pA	pB	pC	\mathbf{f}
França	10000	10	2,5	17	26	60	14	$\begin{bmatrix} 4 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 4 \end{bmatrix}$
Zimbábue	10000	10	2,5	11	0,5	8	91,5	$\begin{bmatrix} 40 & 20 & 1 \\ 20 & 40 & 20 \\ 1 & 20 & 40 \end{bmatrix}$
Peru	10000	10	2,5	16	4	42	54	$\begin{bmatrix} 10 & 5 & 1 \\ 5 & 10 & 5 \\ 1 & 5 & 10 \end{bmatrix}$

Figura 6.1. Figura ilustrativa das ligações entre os indivíduos em cada país. Os reticulados gerados pelo algoritmo 5.1 são convertidos em gráficos tridimensionais, com os nós distribuídos em três planos horizontais para facilitar a visualização das três classes e suas conexões. Os eixos x e y representam a posição única de cada indivíduo no reticulado, cujos números foram deixados da forma proposital em fonte menor. Os nós da cor vermelha representam os indivíduos da classe alta (A) e estão alocados no plano $z = 3$. Os nós da cor verde representam os indivíduos da classe média (B) e estão alocados no plano $z = 2$. Já os nós da cor azul representam os indivíduos da classe baixa (C) e estão alocados no plano $z = 1$. As arestas entre os indivíduos foram desenhadas na cor cinza.

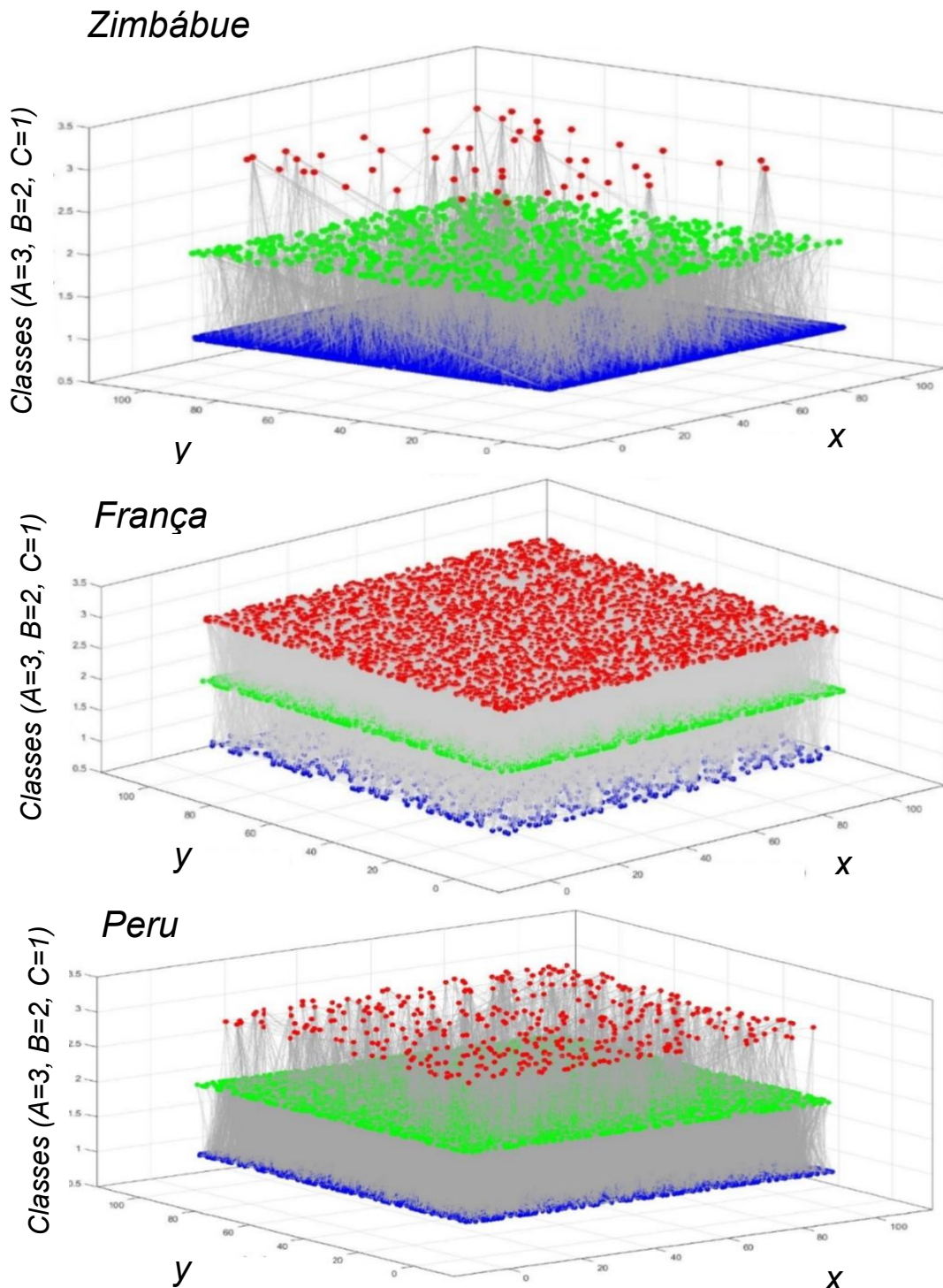


Tabela 6.3. Valores médios de $\langle k \rangle$, $\langle k_A \rangle$, $\langle k_B \rangle$, $\langle k_C \rangle$, $\langle \ell \rangle$, $\langle c \rangle$, C_g , C_p , C_i e C_a e os respectivos desvios padrões.

	Zimbábue	França	Peru
$\langle l \rangle$	$6,22 \pm 0,03$	$5,40 \pm 0,01$	$5,48 \pm 0,01$
$\langle c \rangle$	$0,044 \pm 0,001$	$0,0493 \pm 0,0002$	$0,045 \pm 0,003$
$\langle k \rangle$	$11,4 \pm 0,3$	$17,00 \pm 0,03$	$16,10 \pm 0,06$
$\langle k_A \rangle$	$5,9 \pm 0,3$	$16,50 \pm 0,09$	$11,6 \pm 1,2$
$\langle k_B \rangle$	$8,9 \pm 0,2$	$17,7 \pm 0,1$	$16,07 \pm 0,04$
$\langle k_C \rangle$	$11,6 \pm 0,1$	$15,3 \pm 0,1$	$16,4 \pm 0,2$
C_g	$0,00110 \pm 0,00001$	$0,001700 \pm 0,000003$	$0,001600 \pm 0,000006$
C_p	$0,1610 \pm 0,0009$	$0,1852 \pm 0,0004$	$0,1827 \pm 0,0004$
C_i	$0,000522 \pm 0,000003$	$0,000440 \pm 0,000001$	$0,000448 \pm 0,000001$
C_a	$0,0001$	$0,0001$	$0,0001$

O valor $N = 10000$ (ou seja, um reticulado 100×100) foi escolhido para reduzir a variabilidade estatística das simulações (que ocorrem de maneira mais evidente se são utilizados reticulados com 1000 indivíduos ou menos). A figura 6.1 mostra uma versão tridimensional desse reticulado para cada país. Nesses gráficos tridimensionais, destacam-se as classes sociais e suas ligações. Os pontos em vermelho representam os indivíduos da classe A; os pontos em verde, os da classe B; e pontos em azul, os da classe C. As relações entre os nós são mostradas como linhas em cinza. O gráfico para o Zimbábue sugere que as conexões entre os nós da classe A são realizadas por meio de outras classes e não entre si. Tal característica pode ser consequência da pouca representatividade estatística dessa classe no Zimbábue (apenas 0,5% da população).

Na tabela 6.3, apresentam-se os valores de $\langle k \rangle$, $\langle k_A \rangle$, $\langle k_B \rangle$, $\langle k_C \rangle$, $\langle \ell \rangle$, $\langle c \rangle$, C_g , C_p , C_i e C_a e os respectivos desvios padrões para os três países simulados. As métricas $\langle k_A \rangle$, $\langle k_B \rangle$, $\langle k_C \rangle$ denotam, respectivamente, os graus médios das classes A, B e C. Note que os desvios são pequenos, da ordem de 1% (por isso, a fim de facilitar a leitura dos resultados, eles não foram apresentados no capítulo 4).

Vários comentários podem ser feitos analisando essa tabela. O primeiro resultado a se destacar é que os valores de grau médio $\langle k \rangle$ das redes geradas está bem próximo dos valores $\langle k \rangle_{alvo}$ apresentados na tabela 6.1, que correspondem aos valores encontrados na literatura (Béraud et al., 2015; Grijalva et al., 2015; Melegaro et al., 2017). Esse resultado endossa a viabilidade do algoritmo 5.1 para a geração de redes sociais complexas. Consistentemente, $C_g \simeq \langle k \rangle / (N - 1)$ para os três países.

Para o Zimbábue, o grau médio da classe C é maior do que da classe B que, por sua vez, é maior do que da classe A. Para o Peru, o grau médio das classes B e C é maior do que da classe A. Para a França, o grau médio das três classes é similar.

Em relação a $\langle \ell \rangle$, pode-se inferir que $\langle \ell \rangle_{França} \approx \langle \ell \rangle_{Peru} < \langle \ell \rangle_{Zimbábue}$. Assim, na média, os indivíduos estão mais distantes no Zimbábue do que na França e no Peru. Uma vez que C_p aumenta diminuindo $\langle \ell \rangle$, a tabela 6.3 apresenta de forma consistente que $C_{pFrança} \approx C_{pPeru} > C_{pZimbábue}$. Essas relações sugerem que as informações são transmitidas mais rapidamente na França e no Peru do que no Zimbábue.

Sobre o coeficiente de clusterização, nota-se que $\langle c \rangle_{França} > \langle c \rangle_{Peru} \approx \langle c \rangle_{Zimbábue}$ e, para a centralidade de intermediação, $C_{iFrança} \approx C_{iPeru} < C_{iZimbábue}$. Assim, o valor de $\langle c \rangle$ não distingue o Peru do Zimbábue e o valor de C_i não distingue a França do Peru. Essas desigualdades sugerem que os vizinhos de um indivíduo estão mais conectados na França e há menos indivíduos controlando o fluxo de informações em Zimbábue. Em relação aos valores da centralidade de autovalor, os valores de C_a são idênticos para os três países.

Comparando $\langle \ell \rangle$ encontrado nos grafos com os valores obtidos a partir das fórmulas aproximadas mostradas na tabela 2.6, observa-se que:

$$\begin{aligned} \text{Zimbábue: } \langle \ell \rangle &\simeq 6,22 > \frac{\log 10000}{\log 11,4} \simeq 3,78 \\ \text{França: } \langle \ell \rangle &\simeq 5,40 > \frac{\log 10000}{\log 17} \simeq 3,25 \\ \text{Peru: } \langle \ell \rangle &\simeq 5,48 > \frac{\log 10000}{\log 16} \simeq 3,31 \end{aligned}$$

Portanto, os valores de $\langle \ell \rangle$ do grafo são apenas cerca de 50% maiores do que aqueles que a rede deveria apresentar caso fosse totalmente aleatória. Em relação a $\langle c \rangle$, tem-se:

$$\begin{aligned} \text{Zimbábue: } \langle c \rangle &\simeq 0,044 \gg \frac{11,4}{10000} \simeq 0,00114 \\ \text{França: } \langle c \rangle &\simeq 0,0493 \gg \frac{17}{10000} \simeq 0,0017 \\ \text{Peru: } \langle c \rangle &\simeq 0,045 \gg \frac{16,1}{10000} \simeq 0,00161 \end{aligned}$$

de modo que a clusterização média da rede é cerca de 30 vezes maior do que aquela de uma rede puramente aleatória. Os valores de $\langle \ell \rangle$ e de $\langle c \rangle$ do grafo são afetados pela escolha do raio de Moore r .

As três redes geradas numericamente exibem percolação, de modo que não há nós ou subgrafos isolados. Em relação a esse assunto, é válido mencionar o artigo de Cohen et al. (2000). Esse artigo se propõe estudar a resiliência da internet a ataques que poderiam deixar fora do ar um grande número de nós da rede. Para redes livres-de-escala com

$\gamma < 3$, é improvável não ocorrer percolação. A probabilidade de percolação p_c pode ser aproximada analiticamente por:

$$p_c = 1 - \frac{1}{k_0 - 1} \quad (6.8)$$

em que $k_0 \approx N^{2/3}$. Usando essa fórmula para redes com $N = 10000$, então $p_c = 99,8\%$.

Na tabela 6.4, são apresentadas as composições socioeconômicas dos indivíduos que apresentam os 100 maiores valores de $\langle k \rangle$, $\langle c \rangle$, C_p , C_i e C_a . Os desvios foram calculados e variam de 0 a 10%. Eles foram omitidos para facilitar a legibilidade das médias apresentadas.

Tabela 6.4. Porcentagem das classes sociais considerando os 100 maiores valores de $\langle k \rangle$, $\langle c \rangle$, C_p , C_i e C_a para os três países.

	Zimbábue			França			Peru		
	%C	%B	%A	%C	%B	%A	%C	%B	%A
100 maiores $\langle k \rangle$	100	0	0	0	99,3	0,7	71,3	28,7	0
100 maiores $\langle c \rangle$	81,3	15,7	3	23,7	50	26,3	45	26	29
100 maiores C_p	99,3	0,7	0	4	79,3	16,7	66	33,7	0,3
100 maiores C_i	99,3	0,7	0	2,7	86,3	11	72	28	0
100 maiores C_a	99,7	0,3	0	1,7	78,3	20	69,7	30	0,3

Os resultados apresentados nessa tabela sugerem que o fluxo de informações é controlado principalmente pela classe média na França e pela classe baixa no Peru e no Zimbábue; no entanto, a classe média tem maior influência no Peru do que no Zimbábue. Apesar da predominância da classe média da França e da classe baixa no Peru e no Zimbábue, as proporções apresentadas na tabela 6.4 para cada métrica topológica são diferentes dos dados sociodemográficos mostrados na tabela 2.1.

Tabela 6.5. Valores da tabela 6.4 divididos pA , pB e pC . Chama-se essa divisão de razão normalizada.

	Zimbábue			França			Peru		
	%C	%B	%A	%C	%B	%A	%C	%B	%A
100 maiores $\langle k \rangle$	1,1	0	0	0	1,7	0	1,3	0,7	0
100 maiores $\langle c \rangle$	0,9	2	6	1,7	0,8	1	0,8	0,6	7,3
100 maiores $\langle C_p \rangle$	1,1	0,1	0	0,3	1,3	0,6	1,2	0,8	0,1
100 maiores $\langle C_i \rangle$	1,1	0,1	0	0,2	1,4	0,4	1,3	0,7	0
100 maiores $\langle C_a \rangle$	1,1	0	0	0,1	1,3	0,8	1,3	0,7	0,1

Na tabela 6.5, apresentam-se os valores da tabela 6.4 divididos pelas respectivas porcentagens pA , pB e pC que compõem os países como um todo. Esses números são

aqui chamados de razão normalizada. O objetivo dessa análise é fazer uma comparação mais acurada das métricas topológicas. Por exemplo, para o Peru, os 100 maiores valores de $\langle k \rangle$ são tais que 71,3% estão na classe C, que corresponde a 54% da população total. Calculando a razão normalizada $71,3/54$, obtém-se 1,3 que é o número que aparece na primeira linha e na sétima coluna da tabela 6.5. Esse resultado por ser interpretado assim: entre os indivíduos com os 100 maiores valores de $\langle k \rangle$, há cerca de 30% a mais de indivíduos na classe C do que na distribuição de classes desse país.

Em relação à tabela 6.5, pode-se dizer que:

- (a) Sobre o grau médio: para a França, só há representantes da classe B. Para o Peru, a proporção de indivíduos da classe C aumenta e a da classe B diminui, em relação às porcentagens que caracterizam o país como um todo.
- (b) Sobre a clusterização: a razão normalizada indica que, no Zimbábue, a proporção de indivíduos com alto $\langle c \rangle$ na classe A é 6 vezes a proporção encontrada no país como um todo, na classe B é 2 vezes maior e na classe C é cerca de 10% menor. Na França, há um aumento na proporção da classe C e diminuição na classe B; a proporção da classe A não se altera, em relação àquela encontrada no país. No Peru, há um aumento expressivo na proporção da classe A e diminuição nas proporções das classes B e C.
- (c) Com relação às centralidades: para as três métricas, aumentam a proporção da classe C no Zimbábue e no Peru e a proporção da classe B na França; as demais classes ou diminuem ou não têm representantes. Esses resultados reforçam a ideia de que há indivíduos da classe B na França e da classe C no Zimbábue e no Peru agem como *hubs* de informação.

Na próxima seção 6.4, trata-se da distribuição de graus $p(k)$.

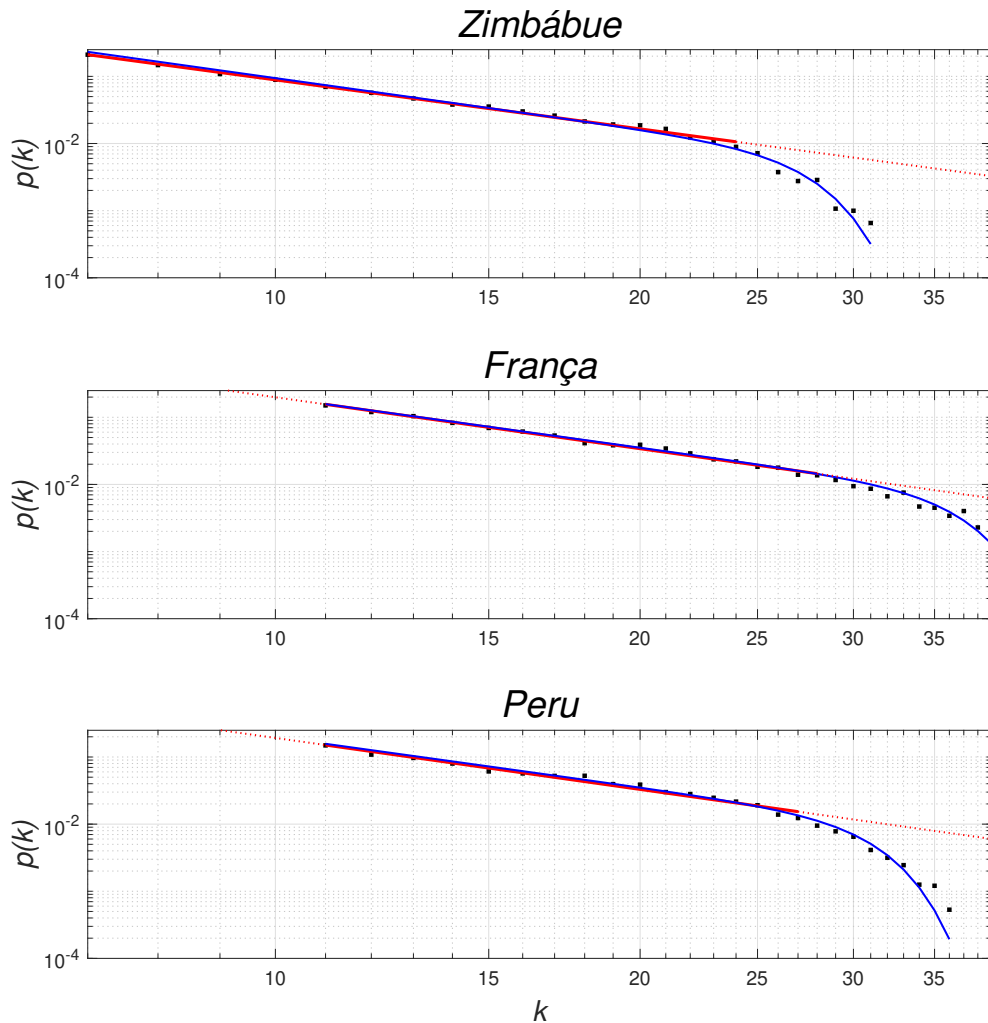
6.4 Distribuição de graus dos três países

Nesta seção, as curvas de $p(k)$ para Zimbábue, França e Peru são ajustadas por meio de regressões lineares e não lineares (Licciardi e Monteiro, 2022). Assim como no capítulo 4, fazem-se gráficos log-log, pois uma rede em que $p(k) = Ak^{-\gamma}$ para todo k leva, num gráfico log-log, a uma reta com coeficiente angular $-\gamma$, como mencionado no final do capítulo 3. A figura 6.2 mostra os gráficos de $p(k)$ para cada país. Os pontos em preto são os valores de $p(k)$ obtidos a partir dos grafos. A linha vermelha corresponde à função ajustada $A_0k^{-\gamma_0}$. A curva em azul corresponde à função ajustada $A_1k^{-\gamma_1}10^{\delta k^{10}}$.

A tabela 6.6 contém os valores dos parâmetros A_0 e γ_0 e o erro quadrático médio (EQM) correspondente à função linear ajustada no gráfico log-log de cada país. Observa-

se que $\gamma_0 \simeq \gamma = 2,5$, que foi o valor previamente escolhido para esse expoente. Além disso, observa-se que EQM é “pequeno”. A tabela 6.7 apresenta os valores de k_{min} , $\langle k \rangle$, k_{max} , k_{cut} e Δk . Nota-se que $\langle k \rangle \simeq \langle k \rangle_{alvo}$ para os três países.

Figura 6.2. Distribuição de graus dos grafos de cada país numa escala-log-log (base 10). Os pontos em preto são os dados obtidos diretamente dos grafos. A função ajustada $A_0 k^{-\gamma_0}$ corresponde à linha vermelha e ela é um bom ajuste para $k_{min} \leq k \leq 2\langle k \rangle_{alvo}$. A função ajustada $A_1 k^{-\gamma_1} 10^{\delta k^{10}}$ corresponde à linha azul para $k_{min} \leq k \leq k_{max}$.



O ajuste linear no gráfico log-log pode ser considerado uma primeira aproximação para $p(k)$. Observe que para $k > 2\langle k \rangle_{alvo}$, essa função distancia-se cada vez mais dos dados obtidos na simulação. Esse desvio é, de fato, esperado e encontrado em diversos outros trabalhos (Barabasi, 2016; Barabási e Albert, 1999; Clauset, Shalizi e Newman, 2009; Dorogovtsev e Mendes, 2002; Newman, 2003; Newman e Girvan, 2004).

A fim de se obter um ajuste mais adequado para $p(k)$, principalmente para $k > 2\langle k \rangle_{alvo}$, toma-se a função $A_1 k^{-\gamma_1} 10^{\delta k^{10}}$, que é não linear no gráfico log-log. Os valores dos parâmetros A_1 , γ_1 e δ para os três países são apresentados na tabela 6.8.

Tabela 6.6. Valores de A_0 , γ_0 e do erro quadrático médio (EQM) da regressão linear de $p(k)$ para os três países simulados.

	Zimbábue	França	Peru
A_0	23,4	69,2	67,6
γ_0	2,43	2,54	2,55
EQM	0,0092	0,0063	0,0094

Tabela 6.7. Valores de k_{min} , $\langle k \rangle$, k_{max} , k_{cut} e Δk para os três países simulados.

	Zimbábue	França	Peru
k_{min}	7	11	11
$\langle k \rangle$	11,4	17,1	16
k_{max}	31	39	36
k_{cut}	24	29	26
Δk	17	18	15

Essa nova curva ajustada apresenta um comportamento livre-de-escala para valores baixos ou intermediários de k , como no ajuste linear. Note que o EQM é similar ao do ajuste linear e que $\gamma_1 \simeq \gamma = 2,5$, como desejado. Esse tipo de ajuste não linear (ou similar) já foi usado em outros trabalhos (Ebel, Mielsch e Bornholdt, 2002; Newman, 2003).

Tabela 6.8. Valores de A_1 , γ_1 , δ e do EQM da regressão não linear para os três países.

	Zimbábue	França	Peru
A_1	23,4	69,2	67,6
γ_1	2,50	2,56	2,50
δ	$-1,0 \times 10^{-15}$	$-8,7 \times 10^{-17}$	$-4,4 \times 10^{-16}$
EQM	0,0076	0,0035	0,0051

6.5 Discussão dos resultados

Uma informação pode afetar a percepção da realidade e os processos de tomadas de decisão (Abib, 2010; Gascuel, 2021; Gorla, Somers e Wong, 2010). Na verdade, uma informação pode se tornar uma questão de vida ou morte (Jones, 2014). Nesse contexto, pode-se citar, por exemplo, o esforço realizado por autoridades para evitar a disseminação de notícias falsas sobre a pandemia de COVID-19. Por exemplo, na França, foi lançado um *site* para fornecer informação confiável sobre o uso de drogas durante a pandemia (Larrouquere et al., 2020). No Peru, criar e espalhar notícias falsas sobre o COVID-19 podia ser punido com sentença de prisão (Alvarez-Risco et al., 2020). No Zimbábue, usu-

se um serviço de redes sociais para difundir informações confiáveis sobre esse vírus (Bowles, Larreguy e Liu, 2020). Os resultados obtidos por meio de simulações computacionais, como os apresentados neste capítulo, podem ajudar a compreender como a comunicação interpessoal é influenciada pela estratificação social e que medidas podem ser tomadas para favorecer as ações de divulgação de notícias e fatos relevantes.

As simulações realizadas, cujos resultados estão condensados em parte na tabela 6.3, sugerem que no Zimbábue as informações transmitidas em contatos presenciais levam mais tempo para circular e sofrem maior influência de nós intermediadores. Já no Peru e na França, as informações circulam mais rapidamente.

A tabela 6.4 mostra que a maioria dos nós com maiores centralidades e grau médio pertencem à classe C no Peru e no Zimbábue e à classe B na França. Note que a maioria desses nós *top* 100 pertence à classe mais prevalente em cada país. Portanto, campanhas de educação e de saúde pública (muito frequentes, por exemplo, durante a pandemia de COVID-19) dirigidas especificamente para essa classe mais prevalente podem espalhar mais rapidamente uma mensagem de interesse público.

A tabela 6.5 revela que, no que diz respeito à clusterização, a participação de indivíduos da classe A no Zimbábue e no Peru e da classe C na França, são mais relevantes no *top* 100 do que na população como um todo. Esse fato deve ser também considerado na elaboração de campanhas que visam divulgar informações. Maior clusterização implica maior compartilhamento local de informações.

A distribuição de graus para os três países foi bem ajustada para $k_{min} < k < k_{max}$ pela função $A_1 k^{-\gamma_1} 10^{\delta k^{10}}$. Diversas tentativas foram feitas até se encontrar essa função que levou a um bom ajuste. Para $k > 2\langle k \rangle_{alvo}$, a distribuição do graus nitidamente se desvia daquela prevista pela lei de potência pura, indicando que nós com grau grande (os *hubs* de informação) são mais raros no grafo do que essa lei prevê.

Esses comentários feitos para a disseminação de informação são também válidos, ou até mesmo mais válidos, para a propagação de infecções, pois informações são também transmitidas por contatos virtuais, mas infecções certamente não. Lembre que o modelo proposto é apenas para contatos presenciais, pois privilegiam-se conexões entre indivíduos mais próximos.

7 CONCLUSÃO

Nesta tese, foram propostos algoritmos computacionais para gerar grafos que podem representar redes de contatos sociais presenciais, pelos quais se compartilham informações, se trocam mercadorias, se transmitem infecções.

Os grafos gerados pelo algoritmo 3.1 foram analisados no capítulo 4. As principais características desse algoritmo são:

- o grau médio $\langle k \rangle$ e o expoente γ da distribuição de graus $p(k) \sim k^{-\gamma}$ são parâmetros desse algoritmo, de modo que os grafos gerados apresentam (em quase todos os casos mostrados no capítulo 4) os valores de $\langle k \rangle$ e o expoente γ desejados;
- devido a restrições geográficas aplicadas na criação das arestas (que só são criadas na região delimitada pelo raio r), o grafo tende a ter alta clusterização (ou seja, $\langle c \rangle \gg \langle k \rangle / N$).

Portanto, o algoritmo 3.1 pode ser usado para gerar grafos que apresentam características de redes livre-de-escala ($p(k) \sim k^{-\gamma}$) e de redes mundo-pequeno ($\langle c \rangle$ “grande”). Observe, entretanto, que o algoritmo desenvolvido não é específico para redes em que $p(k)$ segue uma lei de potência. De fato, ele é adaptável para criar grafos com qualquer distribuição de graus expressa por uma função ou obtida por meio de levantamentos estatísticos.

Um resultado inesperado foi relatado na seção 4.2. Nessa seção, mostrou-se que, conforme γ passa de valores menores do que um para valores maiores do que um, a distribuição de graus dos grafos gerados passa de uma que se assemelha à distribuição de Poisson para uma distribuição livre-de-escala, em que a famosa cauda dessa distribuição, encontrada em diversos trabalhos, é aqui interpretada como um resíduo da distribuição do tipo-Poisson que surge para $\gamma < 1$.

Outro resultado inesperado, também relatado na seção 4.2, refere-se à relação entre γ e Δk (o intervalo em k para o qual a expressão $p(k) = Ak^{-\gamma}$ ajusta bem a distribuição de graus nos grafos). Curiosamente, Δk tem maiores valores no intervalo de γ que tipicamente caracteriza redes reais, que é $2 \leq \gamma \leq 3$. Esta relação sugere a seguinte conjectura: redes sociais apresentam $2 \leq \gamma \leq 3$ porque a conectividade nessas redes é influenciada pela localização espacial de seus nós. Geralmente, as coordenadas espaciais dos nós são negligenciadas em modelos de rede. Por exemplo, considere a rede de atores e atrizes mencionada no capítulo 1. Em geral, a maioria dos artistas envolvidos

em um filme é do mesmo país, mas esse fato é ignorado no processo de construção do grafo. Outra rede social que evidentemente sofre influência da localização espacial é a rede de contatos sexuais (Liljeros et al., 2001). Nesses exemplos, é óbvio que a questão geográfica pode afetar a formação das conexões. Portanto, a localização espacial dos nós constituintes deveria ser levada em conta em modelos teóricos de redes sociais baseadas em contatos presenciais.

No capítulo 5, o algoritmo 3.1 foi adaptado para levar em conta que os indivíduos que compõem uma rede social pertencem a classes socioeconômicas diferentes e que isso afeta a probabilidade de eles se conectarem (como comprovam estudos sobre homofilia). Essa foi motivação para se elaborar o algoritmo 5.1. As características adicionais desse algoritmo são:

- as porcentagens pA , pB e pC de indivíduos nas classes A, B e C, respectivamente, são parâmetros do algoritmo;
- a criação das arestas é afetada pela ponderação expressa pela matriz \mathbf{f} que especifica os pesos com que as classes se afetam. Esses pesos refletem o IDH e/ou o coeficiente de Gini do país estudado, de modo que quanto mais próximos são os elementos da matriz \mathbf{f} , menos socialmente desigual é o país considerado.

Por simplicidade, escreveu-se a matriz \mathbf{f} em termos de um único parâmetro f . Certamente, outras escolhas para os elementos dessa matriz poderiam ter sido feitas.

Com os valores de parâmetros adotados nesta tese, mostrou-se que na França e no Peru a informação tende a circular mais rapidamente via encontros presenciais do que no Zimbábue. Mostrou-se também que a classe C influencia mais a difusão de informação no Peru e no Zimbábue; na França, esse papel é desempenhado pela classe B.

Vários estudos podem ser conduzidos a partir desta tese. Por exemplo, o modelo de grafo proposto pode ser usado em investigações sobre propagação de doenças ou de boatos; temas comuns em pesquisas de sistemas complexos. Ainda, cada nó pode ter uma dinâmica própria, de modo que o grafo proposto pode ser empregado em estudos sobre sincronismo e a formação de padrões espaço-temporais.

Nesta tese, faltou avaliar o impacto que a matriz \mathbf{f} tem sobre o grafo; ou seja, como os elementos dessa matriz afetam a estrutura topológica da rede. A dedução analítica das propriedades topológicas da rede também foi deixado para trabalho futuro. Ainda, o modelo aqui proposto pode ser modificado para incluir contatos sociais virtuais, o que alteraria o caráter de conexões localizadas exclusivamente dentro de uma área geográfica em torno de cada indivíduo da rede.

REFERÊNCIAS

- ABIB, G. A qualidade da informação para a tomada de decisão sob a perspectiva do sensemaking: uma ampliação do campo. **Ciência da Informação**, v. 39, p. 73–82, 2010.
- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, v. 74, p. 47–97, 2002.
- ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Diameter of the world-wide web. **Nature**, v. 401, p. 130–131, 1999.
- ALSALEM, K. et al. Predicting life expectancy at birth. In: IEEE. **2020 2nd International Conference on Computer and Information Sciences (ICCIS)**. Rajasthan, 2020. p. 1–6.
- ALVAREZ-RISCO, A. et al. The Peru approach against the COVID-19 infodemic: insights and strategies. **American Journal of Tropical Medicine and Hygiene**, v. 103, p. 583, 2020.
- ATKINSON, A. B. **Desigualdade: O que Pode Ser Feito?** São Paulo, SP: Leya, 2016.
- BARABASI, A.-L. The scale-free property. In: **Network Science**. Cambridge: Cambridge University Press, 2016. p. 1–15.
- BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, p. 509–512, 1999.
- BARABÁSI, A.-L.; ALBERT, R.; JEONG, H. Scale-free characteristics of random networks: the topology of the world-wide web. **Physica A**, v. 281, p. 69–77, 2000.
- BARABÁSI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell’s functional organization. **Nature Reviews Genetics**, v. 5, p. 101–113, 2004.
- BARRAT, A. et al. The architecture of complex weighted networks. **Proceedings of the National Academy of Sciences of USA**, v. 101, p. 3747–3752, 2004.
- BARROW-GREEN, J. Oscar II’s Prize competition and the error in Poincaré’s memoir on the three body problem. **Archive for History of Exact Sciences**, v. 48, p. 107–131, 1994.
- BÉRAUD, G. et al. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. **PLoS ONE**, v. 10, p. e0133203, 2015.

- BIANCONI, G.; BARABÁSI, A.-L. Competition and multiscaling in evolving networks. **Europhysics Letters**, v. 54, p. 436, 2001.
- BOCCALETTI, S. et al. Complex networks: structure and dynamics. **Physics Reports**, v. 424, p. 175–308, 2006.
- BOUCHER, V. Structural homophily. **International Economic Review**, v. 56, p. 235–264, 2015.
- BOURDIEU, P. The forms of capital. In: RICHARDSON, J. G. (Ed.). **Handbook of Theory and Research for the Sociology of Education**. Westport, C.N.: Greenwood Press, 1986. p. 241–258.
- BOWLES, J.; LARREGUY, H.; LIU, S. Countering misinformation via Whatsapp: preliminary evidence from the COVID-19 pandemic in Zimbabwe. **PLoS One**, v. 15, n. e0240005, 2020.
- CALDARELLI, G. **Scale-free Networks: Complex Webs in Nature and Technology**. Oxford: Oxford University Press, 2007.
- CENCINI, M. et al. **A Random Walk in Physics: Beyond Black Holes and Time-Travels**. New York: Springer, 2021.
- CENTOLA, D. The spread of behavior in an online social network experiment. **Science**, v. 329, p. 1194–1197, 2010.
- CHEN, Y. W.; ZHANG, L. F.; HUANG, J. P. The Watts-Strogatz network model developed by including degree distribution: theory and computer simulation. **Journal of Physics A**, v. 40, p. 8237, 2007.
- CHODROW, P. S. Structure and information in spatial segregation. **Proceedings of the National Academy of Sciences of USA**, v. 114, p. 11591–11596, 2017.
- CHUNG, M. G. et al. International tourism dynamics in a globalized world: a social network analysis approach. **Journal of Travel Research**, v. 59, p. 387–403, 2020.
- CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. J. Power-law distributions in empirical data. **SIAM Review**, v. 51, p. 661–703, 2009.
- COHEN, R. et al. Resilience of the internet to random breakdowns. **Physical Review Letters**, v. 85, p. 4626–4628, 2000.
- CURRARINI, S.; JACKSON, M. O.; PIN, P. An economic model of friendship: homophily, minorities, and segregation. **Econometrica**, v. 77, p. 1003–1045, 2009.
- DIMFR. **World economic outlook, April 2021**. International Monetary Fund, 2021. Disponível em: <https://www.elibrary.imf.org/display/book/9781513575025/9781513575025.xml>.

DOROGOVTSEV, S. N.; GOLTSEV, A. V.; MENDES, J. F. Critical phenomena in complex networks. **Reviews of Modern Physics**, v. 80, p. 1275, 2008.

DOROGOVTSEV, S. N.; MENDES, J. F. Evolution of networks. **Advances in Physics**, v. 51, p. 1079–1187, 2002.

EARMAN, J. **Determinism**. Winter 2016. Metaphysics Research Lab, Stanford University, 2004. Disponível em: <https://plato.stanford.edu/archives/win2016/entries/determinism-causal/>.

EBEL, H.; MIELSCH, L.-I.; BORNHOLDT, S. Scale-free topology of e-mail networks. **Physical Review E**, v. 66, p. 035103, 2002.

ERDŐS, P.; RÉNYI, A. On random graphs I. **Publicaniones Mathematicae**, v. 6, p. 290–297, 1959.

EULER, L. Solutio problematis ad geometriam situs pertinentis. **Commentarii Academiae Scientiarum Petropolitanae**, p. 128–140, 1741.

FAN, J. F. et al. Epidemics on multilayer simplicial complexes. **Proceedings of the Royal Society A**, v. 478, p. 20220059, 2022.

FARMER, J. D.; FOLEY, D. The economy needs agent-based modelling. **Nature**, v. 460, p. 685–686, 2009.

FENZL, N.; MACHADO, J. A. **A Sustentabilidade de Sistemas Complexos**. Belém: UFPA, 2009.

FORTUNATO, S. Community detection in graphs. **Physics Reports**, v. 486, p. 75–174, 2010.

GASCUEL, H.-M. **Diffusion de l'Information et Prise de Décision Collective**. Tese (Doutorado) — Université Paul Sabatier-Toulouse III, 2021.

GENT, W. V.; DAS, M.; MUSTERD, S. Sociocultural, economic and ethnic homogeneity in residential mobility and spatial sorting among couples. **Environment and Planning A**, v. 51, p. 891–912, 2019.

GINI, C. **Variabilità e Mutabilità**. Bologna: Tipografia di Paolo Cuppini, 1912.

GÓES, H. C.; GUÉRIOS, E. C.; GÓES, A. R. T. Pensamento complexo de Edgar Morin e matemática - caminhos futuros. **Revista Espacios**, v. 40, n. 31, p. 16–30, 2019.

GOH, K.-I. et al. Classification of scale-free networks. **Proceedings of the National Academy of Sciences of USA**, v. 99, p. 12583–12588, 2002.

GORLA, N.; SOMERS, T. M.; WONG, B. Organizational impact of system quality, information quality, and service quality. **Journal of Strategic Information Systems**,

v. 19, p. 207–228, 2010.

GRANT, E. **Science and Religion, 400 BC to AD 1550: From Aristotle to Copernicus**. Baltimore, MD: Johns Hopkins University Press, 2006.

GRIJALVA, C. G. et al. A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. **PLoS ONE**, v. 10, p. e0118457, 2015.

GRUSKY, D. **Social Stratification, Class, Race, and Gender in Sociological Perspective**. New York: Routledge, 2019.

GUJARATI, D. N.; PORTER, D. C. **Basic Econometrics**. New York: McGraw-Hill Education, 2017.

HAQ, M. U. **Reflections on Human Development**. Oxford: Oxford University Press, 1995.

HARTTGEN, K.; KLASSEN, S. A household-based human development index. **World Development**, v. 40, p. 878–899, 2012.

HEYLIGHEN, F. Building a science of complexity. In: FATMI, H. A. (Ed.). **1988 Annual Conference of the Cybernetic Society**. London: Cybernetics Society, King's College, London, 1988. p. 1–22.

HOFFMAN, M. et al. A model for the dynamics of face-to-face interactions in social groups. **Network Science**, v. 8, p. S4–S25, 2020.

HOLME, P.; KIM, B. J. Growing scale-free networks with tunable clustering. **Physical Review E**, v. 65, p. 026107, 2002.

HOLST, H.; FESSLER, A.; NIEHOFF, S. COVID-19, social class and work experience in Germany: inequalities in work-related health and economic risks. **European Societies**, v. 23, p. S495–S512, 2021.

JEONG, H. et al. The large-scale organization of metabolic networks. **Nature**, v. 407, p. 651–654, 2000.

JONES, M. A matter of life and death. **MIS Quarterly**, v. 38, p. 895–925, 2014.

KLUGMAN, J.; RODRÍGUEZ, F.; CHOI, H.-J. The HDI 2010: new controversies, old critiques. **Journal of Economic Inequality**, v. 9, p. 249–288, 2011.

KOYRÉ, A. Galileo and the scientific revolution of the seventeenth century. **The Philosophical Review**, v. 52, p. 333–348, 1943.

KWAN, T. H. et al. Mining relationships between transmission clusters from contact tracing data: an application for investigating COVID-19 outbreak. **Journal of the**

- American Medical Informatics Association**, v. 28, p. 2385–2392, 2021.
- LANDHERR, A.; FRIEDL, B.; HEIDEMANN, J. A critical review of centrality measures in social networks. **Business & Information Systems Engineering**, v. 2, p. 371–385, 2010.
- LARROUQUERE, L. et al. Genesis of an emergency public drug information website by the French society of pharmacology and therapeutics during the COVID-19 pandemic. **Fundamental & Clinical Pharmacology**, v. 34, p. 389–396, 2020.
- LARSEN, R. J.; MARX, M. L. **Introduction to Mathematical Statistics and its Applications**. Englewood Cliffs, N.J.: Prentice Hall, 2000.
- LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. **Physical Review Letters**, v. 87, p. 198701, 2001.
- LECKIE, S. **UNDP Human Development Report 2000, Housing Rights**. 2000. Disponível em: <https://hdr.undp.org/system/files/documents/leckiepdf.pdf>.
- LENSKI, G. E. **Power and Privilege: A Theory of Social Stratification**. Chapel Hill: University of North Carolina Press, 2013.
- LEVINE, R. F. **Social Class and Stratification: Classic Statements and Theoretical Debates**. Oxford: Rowman & Littlefield, 2006.
- LICCIARDI, A. N.; MONTEIRO, L. H. A. A complex network model for a society with socioeconomic classes. **Mathematical Biosciences and Engineering**, v. 19, p. 6731–6742, 2022.
- LICCIARDI, A. N.; MONTEIRO, L. H. A. A network model of social contacts with small-world and scale-free features, tunable connectivity, and geographic restrictions. **Mathematical Biosciences and Engineering**, v. 21, p. 4801–4813, 2024.
- LILJEROS, F. et al. The web of human sexual contacts. **Nature**, v. 411, p. 907–908, 2001.
- LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the Atmospheric Sciences**, v. 20, p. 130–141, 1963.
- LYNCH, J.; KAPLAN, G. Socioeconomic position. In: BERKMAN; KAWACHI (Ed.). **Social Epidemiology**. Oxford: Oxford University Press, 2000. p. 13–35.
- MAIER, B. F. Generalization of the small-world effect on a model approaching the Erdős–Rényi random graph. **Scientific Reports**, v. 9, p. 9268, 2019.
- MANDELBROT, B. The variation of certain speculative prices. **The Journal of Business**, v. 36, p. 394–419, 1963.

MARX, K. H.; ENGELS, F. Manifesto do partido comunista. **Germinal: Marxismo e Educação em Debate**, Universidade Estadual de Londrina, Londrina, v. 2, p. 215–240, 2010.

MAY, R. M. **Stability and Complexity in Model Ecosystems**. Princeton: Princeton University Press, 2019.

MCPHERSON, M.; SMITH-LOVIN, L.; COOK, J. M. Birds of a feather: homophily in social networks. **Annual Review of Sociology**, v. 27, p. 415–444, 2001.

MELEGARO, A. et al. Social contact structures and time use patterns in the Manicaland province of Zimbabwe. **PLoS ONE**, v. 12, p. e0170459, 2017.

MITZENMACHER, M. A brief history of generative models for power law and lognormal distributions. **Internet Mathematics**, v. 1, p. 226–251, 2004.

MONTEIRO, L. H. A. **Sistemas Dinâmicos Complexos**. São Paulo: Editora Livraria da Física, 2014.

MONTEIRO, L. H. A. **Sistemas Dinâmicos**. São Paulo: Editora Livraria da Física, 2019.

MONTEIRO, L. H. A.; PAIVA, D.; PIQUEIRA, J. R. C. Spreading depression in mainly locally connected cellular automaton. **Journal of Biological Systems**, v. 14, p. 617–629, 2006.

MORIN, E. **Complex Thinking for a Complex World – About Reductionism, Disjunction and Systemism**. 2014. Disponível em: <https://api.semanticscholar.org/CorpusID:142543732>.

MUKAKA, M. M. A guide to appropriate use of correlation coefficient in medical research. **Malawi Medical Journal**, v. 24, p. 69–71, 2012.

MULLICK, P.; SEN, P. Active-absorbing phase transition and small-world behaviour in ising model on finite addition type networks in two dimensions. **Journal of Complex Networks**, v. 8, p. cnz046, 2020.

NAESS, A. **Galileu Galilei: Um Revolucionário e seu Tempo**. São Paulo: Editora Schwarcz-Companhia das Letras, 2015.

NEWMAN, M. E. J. The structure and function of complex networks. **SIAM Review**, v. 45, p. 167–256, 2003.

NEWMAN, M. E. J. **Networks: An Introduction**. Oxford: Oxford University Press, 2010.

NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical Review E**, v. 69, p. 026113, 2004.

NEWMAN, M. E. J.; WATTS, D. J. Scaling and percolation in the small-world network model. **Physical Review E**, v. 60, p. 7332, 1999.

NOBLE, S. U.; TYNES, B. M. **The Intersectional Internet: Race, Sex, Class, and Culture Online**. New York: Peter Lang, 2016.

OCDE. **Under Pressure: The Squeezed Middle Class**. 2019. Disponível em: https://www.oecd-ilibrary.org/social-issues-migration-health/under-pressure-the-squeezed-middle-class_689afed1-en.

PASTOR-SATORRAS, R.; VESPIGNANI, A. **Evolution and Structure of the Internet: A Statistical Physics Approach**. Cambridge: Cambridge University Press, 2004.

PATOMÄKI, H. Scientific realism. In: MORIN, J. F.; OLSSON, C.; ATIKCAN, E. O. (Ed.). **Research Methods in the Social Sciences: An AZ of Key Concepts**. Oxford: Oxford University Press, 2021. p. 251–253.

PEWRESEARCH. **Worldwide Population by Income**. 2021. Disponível em: <https://www.pewresearch.org/global/interactives/global-population-by-income/>.

PIMM, S. L. Food webs. In: **Food Webs. Population and Community Biology**. Dordrecht: Springer, 1982. p. 1–11.

PREM, K. et al. Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era. **PLoS Computational Biology**, v. 17, p. e1009098, 2021.

PRICE, D. S. A general theory of bibliometric and other cumulative advantage processes. **Journal of the American Society for Information Science**, v. 27, p. 292–306, 1976.

PRIGOGINE, I.; STENGERS, I. **La Fin des Certitudes: Temps, Chaos et les Lois de la Nature**. Paris: Odile Jacob, 1996.

REGGIANI, A.; VINCIGUERRA, S. Network connectivity models: An overview and empirical applications. In: _____. **Network Science, Nonlinear Science and Infrastructure Systems**. Boston, MA: Springer US, 2007. p. 147–165. ISBN 978-0-387-71134-8. Disponível em: https://doi.org/10.1007/0-387-71134-1_7.

SCHIMIT, P. H. T. **Modelagem e Controle de Propagação de Epidemias Usando Autômatos Celulares e Teoria de Jogos**. Tese (Doutorado) — Escola Politécnica, Universidade de São Paulo, 2010.

SCHIMIT, P. H. T.; MONTEIRO, L. H. A. On the basic reproduction number and the topological properties of the contact network: an epidemiological study in mainly locally connected cellular automata. **Ecological Modelling**, v. 220, p. 1034–1042, 2009.

SIMON, H. A. On a class of skew distribution functions. **Biometrika**, v. 42, p. 425–440, 1955.

STROGATZ, S. H. Exploring complex networks. **Nature**, v. 410, p. 268–276, 2001.

TRAVERS, J.; MILGRAM, S. An experimental study of the small world problem. **Sociometry**, v. 32, p. 425–443, 1969.

UNDP. **Human Development Report 2021-22**. New York: United Nations, 2022. Disponível em: <http://report.hdr.undp.org>.

UNESCO. **Global Education Monitoring Report 2020. Inclusion and Education: All Means All**. Paris: United Nations Educational, Scientific and Cultural Organization, 2020.

URENA, R. et al. A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. **Information Sciences**, v. 478, p. 461–475, 2019.

VASCONCELLOS, R.; RODRIGUES, C. G.; LUZZI, R. Complexidade, auto-organização e informação em sistemas dinâmicos. **Revista Brasileira de Ensino de Física**, v. 37, p. 2314, 2015.

VERA, A. M. **Propriedades de Redes Complexas de Telecomunicações**. Tese (Doutorado) — Escola de Engenharia de São Carlos, Universidade de São Paulo, 2017.

VOUTSA, V. et al. A stylised view on structural and functional connectivity in dynamical processes in networks. **arXiv:2106.03523**, 2021.

VYTILA, V. et al. Mathematical models for predicting COVID-19 pandemic: a review. **Journal of Physics**, v. 1797, p. 012009, 2021.

WANG, X. F.; CHEN, G. Complex networks: small-world, scale-free and beyond. **IEEE Circuits and Systems Magazine**, v. 3, n. 1, p. 6–20, 2003.

WATTS, D. J. **Six Degrees: The Science of a Connected Age**. New York: W. W. Norton & Company, 2004.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **Nature**, v. 393, p. 440–442, 1998.

WEBER, M. **Economy and Society: An Outline of Interpretive Sociology**. New York: Bedminster Press, 2003.

WELLMAN, B.; QUAN-HAASE, A.; HARPER, M.-G. The networked question in the digital era: how do networked, bounded, and limited individuals connect at different stages in the life course? **Network Science**, v. 8, p. 291–312, 2020.

WEN, T.; DENG, Y. Identification of influencers in complex networks by local information dimensionality. **Information Sciences**, v. 512, p. 549–562, 2020.

WILL, G.; HOMUTH, C. Education of refugee adolescents at the end of secondary school: the role of educational policies, individual and familial resources. **Soziale Welt**, v. 71, p. 160–200, 2020.

WILLIAMS, G. **Chaos Theory Tamed**. London: CRC Press, 1997.

WOLFRAM, S. Cellular automata as models of complexity. **Nature**, v. 311, p. 419–424, 1984.

WOLFRAM, S. **Cellular Automata and Complexity: Collected Papers**. Champaign, IL: Westview Press, 2018.

WORLDBANK. **Atlas of Sustainable Development Goals 2017: From World Development Indicators**. Washington, DC: The World Bank, 2017.

WRIGHT, E. O. **Sociology 929 Alternative Foundations of Class Analysis**. 2006. Notas de Aula em [Sociology 929]. Disponibilizado por [University of Wisconsin Madison]. Disponível em: <https://www.sccc.wisc.edu/soc/faculty/pages/wright/soc929-2006.pdf>.

XU, Y. et al. Quantifying segregation in an integrated urban physical-social space. **Journal of the Royal Society Interface**, v. 16, p. 20190536, 2019.

YITZHAKI, S. Economic inequality and Lorenz curves. **Journal of Economic Perspectives**, v. 28, p. 45–72, 1994.

YULE, G. U. A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis. **Philosophical Transactions of the Royal Society of London, Series B**, v. 213, p. 21–87, 1925.

ZHENG, B. et al. A simple model clarifies the complicated relationships of complex networks. **Scientific Reports**, v. 4, p. 6197, 2014.

ZHOU, X. K. et al. Academic influence aware and multidimensional network analysis for research collaboration navigation based on scholarly big data. **IEEE Transactions on Emerging Topics in Computing**, v. 9, p. 246–257, 2018.