



Judson de Oliveira Moura

**Controlando o grau médio na construção de
redes complexas**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Física pelo Programa de Pós-graduação em Física, do Departamento de Física da PUC-Rio.

Orientador : Profa. Celia Beatriz Anteneodo de Porto

Co-orientador: Dr. Allan Rodrigues Vieira

Rio de Janeiro,
novembro de 2021

Judson de Oliveira Moura

Controlando o grau médio na construção de redes complexas

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Física da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Profa. Celia Beatriz Anteneodo de Porto

Orientador

Departamento de Física – PUC-Rio

Dr. Allan Rodrigues Vieira

Co-orientador

Departamento de Física – PUC-Rio

Prof. Silvio da Costa Ferreira Junior

UFV

Prof. Marlon Ferreira Ramos

UERJ

Dra. Vivian de Araujo Dornelas Nunes

UNESP

Rio de Janeiro, 16 de novembro de 2021

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Judson de Oliveira Moura

Graduado em Física pelo Instituto Federal do Rio de Janeiro - IFRJ (Rio de Janeiro, Brasil) em 2018.

Ficha Catalográfica

Moura, Judson O.

Controlando o grau médio na construção de redes complexas / Judson de Oliveira Moura; orientador: Celia Beatriz Anteneodo de Porto; co-orientador: Allan Rodrigues Vieira. – Rio de Janeiro: PUC-Rio, Departamento de Física, 2021.

v., 79 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Física.

Inclui bibliografia

1. Física – Teses. 2. Redes complexas;. 3. redes livres de escala;. 4. conectividade média;. 5. coeficiente de aglomeração;. 6. correlação de graus..
I. Anteneodo, Celia. II. Vieira, Allan. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Física. IV. Título.

CDD: 530

Dedicado a minha mãe, Neide Rodrigues de Oliveira.

Agradecimentos

Agradeço a Deus por ter guiado meu caminho até este momento. Aos meus pais, Antônio Luiz e Neide, por mesmo estando longe ter dado total apoio e motivação para a conclusão deste mestrado.

Em especial, gostaria de agradecer minha noiva Érica por estar comigo me apoiando nesses últimos três anos, ela foi parte fundamental para conclusão dessa pesquisa.

Também gostaria de agradecer minha orientadora, prof. Celia Anteneodo, por toda a paciência nesses tempos difíceis de pandemia. Obrigado por juntamente com meu co-orientador, Dr. Allan Vieira, propor esse tema novo pra mim até então, em que aprendi e evolui bastante durante a pesquisa.

Agradeço também ao meu co-orientador, Dr. Allan Vieira, por toda disponibilidade e paciência pra tirar minhas dúvidas durante esse tempo que trabalhamos juntos.

Agradeço ao meu primo Altair Moura por ter me ajudado com linguagem de programação no início e durante esses anos de pesquisa.

Agradeço a PUC-Rio juntamente com todos os professores que tive o prazer de encontrar durante essa caminhada. Agradeço também por ter me proporcionado fazer esse mestrado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Moura, Judson O.; Anteneodo, Celia; Vieira, Allan. **Controlando o grau médio na construção de redes complexas**. Rio de Janeiro, 2021. 79p. Dissertação de Mestrado – Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro.

A construção de redes complexas é de grande importância para o estudo de modelos de agentes e sistemas dinâmicos, a exemplo dos modelos de opinião, de epidemias, sistemas de osciladores ou mapas acoplados, etc., que usam grafos como substrato das interações entre os elementos do sistema. Essas dinâmicas dependem fortemente das características topológicas da rede de interações, portanto, é fundamental construir redes com propriedades estruturais bem definidas. Uma das propriedades de grande importância é o grau médio, primeiro momento da distribuição de graus. Nos casos em que a distribuição de graus decai como uma lei de potência, o seu expoente é outra grandeza relevante, relacionada à possibilidade de ter vértices muito conectados. Além disso, procura-se evitar as correlações. Dentro deste quadro, estudamos os efeitos que certas características da distribuição de graus têm nas propriedades da rede, construída mediante o modelo de configuração. Para cada valor do expoente da lei de potência, fixamos os graus mínimo, máximo, e médio, comparando o efeito destes parâmetros nas redes resultantes, através do coeficiente de agrupamento e da correlação de graus entre sítios vizinhos.

Palavras-chave

Redes complexas; redes livres de escala; conectividade média; coeficiente de aglomeração; correlação de graus.

Abstract

Moura, Judson O.; Anteneodo, Celia (Advisor); Vieira, Allan (Co-Advisor). **Controlling the average degree in building complex networks**. Rio de Janeiro, 2021. 79p. Dissertação de mestrado – Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro.

The construction of complex networks is of great importance for the study of agent-based models and dynamical systems, such as opinion models, epidemics, oscillator systems or coupled maps, etc., that use graphs as a substrate to represent the interaction paths. The dynamics can strongly depend on the topological characteristics of the interaction network, therefore, it is essential to build networks with well-defined structural properties. One of the properties of great importance is the average degree, the first moment of degree distribution. In cases where the degree distribution decays like a power law, its exponent is another relevant quantity, related to the possibility of hubs. Within this framework, we study the effects that certain characteristics of the degree distribution have on the properties of the network, built using the configuration model. For each value of the power-law exponent, we fix the minimum, maximum, and average degrees, comparing the effect of these parameters on the resulting networks, through the clustering coefficient and the degree-degree correlation between neighboring sites.

Keywords

Complex networks; free scale networks; clustering coefficient; mean connectivity; degree-degree correlation.

Sumário

1	Introdução	17
2	Fundamentos teóricos de redes complexas	19
2.1	Conceitos básicos	19
2.2	Propriedades das redes	22
3	Modelos de redes: revisão e estudo comparativo	31
3.1	Modelo de configuração (MC)	31
3.2	Redes de Erdős-Rényi (ER)	34
3.3	Gerando redes com distribuição de Poisson utilizando o modelo de configuração	45
3.4	Redes livres de escala	51
3.5	Gerando redes livres de escala com o modelo de configuração	52
4	Controlando o grau médio	62
4.1	Métodos para controlar o grau médio	62
5	Conclusões e comentários finais	74
	Referências bibliográficas	76

Lista de figuras

Figura 2.1	(a) Rede com $N = 6$ vértices e $M = 7$ arestas. (b) Representação de arestas múltiplas e auto-arestas.	20
Figura 2.2	Representação por matriz de adjacência e lista de adjacência. Ambos os casos representam a rede da figura 2.1(a).	22
Figura 2.3	Representação das componentes de uma rede contendo seis vértices. Nesse caso existem três componentes.	25
Figura 2.4	Rede desconexa de tamanho $N = 200$ com 10 componentes.	26
Figura 2.5	Ilustração de como calcular $k_{nn,i}$ para o vértice 1. Nesse caso o grau do vértice 1 é $k_1 = 4$. Os graus dos vértices vizinhos a 1 são $k_2 = 3, k_3 = 3, k_4 = 2, k_5 = 1$. Logo $k_{nn,i} = (3 + 3 + 2 + 1)/4 = 2.25$.	27
Figura 2.6	(a) Mostra todas as possíveis ligações da vizinhança do vértice $i = 1$, nesse caso $c_1 = 0$, pois não há conexões entre seus vizinhos. Em (b), dentre todas as conexões possíveis, os vértices $i = 3$ e $i = 4$ estão unidos por uma aresta, nesse caso existe uma ligação dentre as três possíveis, então $c_1 = 1/3$.	29
Figura 3.1	Meias arestas para a sequência de graus $(4, 1, 2, 2, 1)$.	31
	2.6(a)	31
	2.6(b)	31
Figura 3.2	Primeira aresta gerada a partir do modelo de configuração.	32
Figura 3.3	Em (a), os vértices três e quatro foram sorteados e uma conexão foi colocada entre eles. Em (b), é mostrada a rede finalizada, após completar todo o processo do modelo de configuração.	33
Figura 3.4	Rede de Erdős-Rényi. $N = 10$ e $p = 0.4$	36
Figura 3.5	Histograma do número de arestas gerado pelo modelo $G(N, p)$, com $N = 10^3$ e $\langle k \rangle = 6$. O histograma em azul representa a distribuição do número de arestas M gerado por 10^3 redes. A curva em vermelho é a distribuição gaussiana (Eq. 3-5) com média $\mu = 2996.82$ e desvio padrão $\sigma = 51,76$, obtidos da sequência de graus.	37
Figura 3.6	Rede criada a partir do modelo $G(N, M)$ e sua matriz de adjacência associada. O tamanho da rede equivale a $N = 10$ e o número de arestas $M = 7$.	38

- Figura 3.7 Distribuições de graus para redes de ER, geradas pelos modelos $G(N, p)$ e $G(N, M)$, com tamanho $N = 10^4$ e número total de arestas $M = 3 \times 10^4$, portanto $\langle k \rangle = 6$. Os pontos são a média sobre 10^2 redes e a curva em vermelho é a distribuição de Poisson da equação 3-8. 39
- Figura 3.8 Distribuição de aglomeração para o modelo $G(N, p)$. Nas duas figuras superiores (a) e (b), $N = 10^3$ e $\langle k \rangle = 6$. Para as figuras (c) e (d), $N = 10^4$ e $\langle k \rangle = 6$. As duas figuras à direita (b) e (d), são ampliações das figuras (a) e (c). 10^2 redes foram geradas para as distribuições acima. 41
- Figura 3.9 Distribuição de aglomeração para o modelo $G(N, M)$. Nas duas figuras superiores (a) e (b), $N = 10^3$ e $M = 3 \times 10^3$. Para as figuras (c) e (d), $N = 10^4$ e $M = 3 \times 10^4$. As duas figuras à direita, (b) e (d), são ampliações das figuras (a) e (c). 10^2 redes foram geradas para as distribuições acima. 42
- Figura 3.10 Cada figura mostra a distribuição de frequências para o C global de cada rede para os valores de $\langle k \rangle = 4, 6, 8, 12$. Para cada figura foram geradas 10^2 redes. Figura baseada no modelo $G(N, M)$ 43
- Figura 3.11 As figuras acima mostram o $k_{nn}(k)$ e o $c(k)$ médio para a rede de ER utilizando tanto o modelo $G(N, p)$, quanto o modelo $G(N, M)$. As figuras a e b representam o modelo $G(N, p)$, enquanto c e d, o modelo $G(N, M)$. Para os dois modelos foram geradas rede com valores médio $\langle k \rangle = 4, 6, 8, 12$ e tamanho $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. Para cada modelo foi feito uma média sobre 10^2 redes. 44
- Figura 3.12 Cada figura mostra a distribuição para o coeficiente de correlação de Pearson r de cada rede. Com valores de $\langle k \rangle = 4, 6, 8, 12$. Para cada figura foram geradas 10^2 redes. Figura baseada no modelo $G(N, M)$ 45
- Figura 3.13 Distribuição de graus de uma rede de ER utilizando o modelo de configuração. $N = 10^4$, $\langle k \rangle = 6$. A figura exhibe a média sobre 10^2 redes. A curva vermelha é a mesma distribuição de Poisson da figura 3.7. 46

- Figura 3.14 Distribuição de aglomeração para rede de ER utilizando modelo de configuração. Nas duas figuras superiores (a) e (b), $N = 10^3$ e $\langle k \rangle = 6$. Para as figuras (c) e (d), $N = 10^4$ e $\langle k \rangle = 6$. As duas figuras a direita (b) e (d), são ampliações das figuras (a) e (c). 10^2 redes foram geradas para as distribuições acima. 47
- Figura 3.15 Distribuição de frequências para o C global do modelo de configuração para os valores de $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. Para cada figura foram geradas 10^2 redes. 48
- Figura 3.16 As figuras acima mostram o $k_{nn}(k)$ e o $c(k)$ para a rede de ER utilizando o MC. As rede foram geradas com valores de $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. Média sobre 10^2 redes. 49
- Figura 3.17 Cada figura mostra a distribuição para o coeficiente de correlação de Pearson r para valores de $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. Para cada figura 10^2 redes foram geradas. 50
- Figura 3.18 Distribuição de graus pelo sorteio aleatório de k . Os parâmetros usados foram $N = 10^4$, $k_{min} = 4$, $k_{max} = 2000$, $\gamma \sim 3,45$ e $\langle k \rangle = 6$. A figura mostra a média sobre 100 realizações. A curva em vermelho é a distribuição $P(k) \sim k^{-\gamma}$. 53
- Figura 3.19 Histogramas do grau médio para simulação de 10^2 redes. Em (a), $N = 2 \times 10^3$ e em (b) $N = 10^4$. Para ambas as figuras foram utilizados $k_{min} = 3$, $k_{max} = 2 \times 10^3$ e $\gamma = 3,45$, desta forma o grau médio flutua em torno de 6. 53
- Figura 3.20 Histogramas do grau médio para simulação de 10^2 redes. Em (a), $N = 2 \times 10^3$ e em (b) $N = 10^4$. Para ambas as figuras foram utilizados $k_{min} = 4$, $k_{max} = 2 \times 10^3$ e $\gamma = 2,73$, desta forma o grau médio flutua em torno de 6. 54
- Figura 3.21 Histogramas do coeficiente de aglomeração global para simulação de 10^2 redes. Em (a), $N = 2 \times 10^3$ e em (b) $N = 10^4$. Para ambas as figuras foram utilizados $k_{min} = 4$, $k_{max} = 2 \times 10^3$ e $\gamma = 3,45$, desta forma o grau médio flutua em torno de 6. 55

Figura 3.22 Coeficiente de aglomeração local para redes em lei de potência construídas utilizando o modelo de configuração. O símbolos vermelhos são os c_i , coeficiente de aglomeração local para cada sítio, em 100 realizações (redes). Os símbolos cinza resultam de agrupar o coeficiente c_i por grau (a linha horizontal é a média, C dada pela equação 2-17). As linhas inclinadas, correspondem a $2t/(k(k-1))$, onde t é o número de triângulos. O número de triângulos por grau k está relacionado com o termo e_i na equação 2-15. As duas figuras superiores (a) e (b) são representações em diferentes escalas, para redes com $N = 2 \times 10^3$, $\gamma \sim 3,45$, $k_{min} = 4$ e $k_{max} = 2 \times 10^3$. Para as figuras (c) e (d), $N = 10^4$, $\gamma \sim 3,45$ e $k_{min} = 4$, $k_{max} = 2000$. 56

Figura 3.23 Grau médio dos vizinhos mais próximos dos vértices de grau k , $k_{nn}(k)$, (a) e coeficiente de aglomeração médio $c(k)$ (b). As curvas correspondem a diferentes valores de γ . As redes possuem tamanho $N = 10^4$. As figuras representam a média sobre 10^2 redes. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. 57

Figura 3.24 Coeficiente de correlação de Pearson r para redes em lei de potência. As figuras apresentam diferentes valores de γ . Em todos os casos $N = 10^4$ e foram geradas 10^2 redes. 58

Figura 3.25 Grau médio dos vizinhos mais próximos dos vértices com grau k , $k_{nn}(k)$ (a) e coeficiente de aglomeração médio $c(k)$ (b). As curvas correspondem a diferentes valores de γ . As redes possuem tamanho $N = 10^4$ com $k_{min} = 2$, $k_c = \sqrt{10^4}$. As figuras representam a média sobre 10^2 redes. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. 59

Figura 3.26 Coeficiente de correlação de Pearson r pra rede em lei de potência. As figuras apresentam diferentes valores de γ e $N = 10^4$. Nesse caso, foi utilizado o corte estrutural no grau máximo da rede, isto é, $k_c = \sqrt{10^4}$. Histogramas gerados por 10^2 redes. 60

Figura 3.27 Grau médio global $\langle k \rangle$ em função de γ . Na curva em azul foi utilizado $k_{min} = 2$ e $k_{max} = N$. Para a curva em laranja $k_{min} = 2$ e $k_c = \sqrt{N}$. Para ambas as curvas $N = 10^4$ 61

- Figura 4.1 Distribuição do grau médio, fixando $k_{min} = 2$, $k_{max} = \sqrt{N}$ e mudando o expoente γ . Os símbolos preenchidos representam $N = 10^4$ enquanto os vazios $N = 10^5$. A figura representa a média sobre 10^3 realizações. 63
- Figura 4.2 Limitação para fixar $\langle k \rangle$ no plano k_{min} - k_{max} , considerando faixas para $\langle k \rangle$. As linhas pretas correspondem aos valores $\langle k \rangle = 3, 4, \dots, 10$, para diferentes valores de γ . A faixa roxa foi usada somente como pano de fundo para figura. 64
- Figura 4.3 Distribuição de graus com $\gamma = 3.50$ e $\langle k \rangle = 4$, $k_{min} = 2$, $k_{max} = k_c = \sqrt{N}$, $N = 10^4$ e $a = 0.0604$. A figura foi gerada pela média sobre 10 redes. 65
- Figura 4.4 A figura (a) exhibe $k_{nn}(k)$ enquanto a figura (b) $c(k)$. Resultados da média de 100 redes geradas usando o MC, a partir do método 1, com $\langle k \rangle = 4$, $k_{min} = 2$, $k_{max} = 100$ e $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. 66
- Figura 4.5 Histograma de $\langle k \rangle$ usando o método 1. Os valores de referência são $\langle k \rangle = 4$, $k_{min} = 2$, $k_{max} = 100$ e $N = 10^4$. 67
- Figura 4.6 A figura (a) exhibe valores para $k_{nn}(k)$ enquanto a figura (b) valores para $c(k)$ (média sobre 100 redes geradas usando o MC, a partir do método 1). As redes foram geradas com valores de $\langle k \rangle = 4, 6, 12$, $\gamma = 2.5, 3.5$, $k_{max} = 100$ e $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. 68
- Figura 4.7 Curva em azul: Distribuição de graus com $\gamma = 3.50$ e $\langle k \rangle = 12$, $k_{min} = 4$, $k_{max} = k_c = \sqrt{N}$, $t = 8.16$ e $N = 10^4$. Curva em vermelho: Distribuição de graus com $\gamma = 3.50$ gerada através do sorteio aleatório de k sem controlar o grau médio. Os parâmetros utilizados foram $k_{min} = 4$, $k_{max} = k_c = \sqrt{N}$ e $N = 10^4$. Ambas as curva apresentam a média sobre 100 redes. 69
- Figura 4.8 A figura (a) mostra os resultados para $k_{nn}(k)$ enquanto (b) os resultados para $c(k)$ (média sobre 100 redes). A redes foram geradas para $k_{min} = 2, \dots, 7$ e $\gamma = 3.5, 2.5$ e $\langle k \rangle = 12$. Em azul a rede tem $\gamma = 2.5$ enquanto em vermelho $\gamma = 3.5$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. 70

Figura 4.9 Histograma $p(k) = \lfloor NP(k) \rfloor / N$. Os parâmetros usados foram $\gamma \sim 3,45$ e $\langle k \rangle = 6$. Nesse caso, originariamente $N = 10^4$, porém com o processo de truncamento o total de vértices se reduziu para $\sum_k \lfloor NP(k) \rfloor = 9961$. A curva vermelha representa a distribuição teórica.

71

Figura 4.10 Distribuição de graus após a rede ser completada. Os parâmetros usados foram $\gamma \sim 3.45$, $k_{min} = 4$, $k_{max} = 2 \times 10^3$, $N = 10^4$ e $\langle k \rangle = 6$. A curva em vermelho é a distribuição $P(k) \sim k^{-\gamma}$ e os pontos em azul representam a média sobre 10^3 redes.

73

Lista de tabelas

Tabela 2.1	Representação de uma rede com 5 vértices através de uma lista de adjacência. Como observado, o vértice 1 tem ligações com os vértices 3 e 4. O vértice 2 não possui ligações. Já os vértices 3, 4 e 5 possuem ligações.	21
2.1(a)		21
2.1(b)		21
Tabela 3.1	Esses valores correspondem a uma sequência de graus prescrita por uma distribuição de Poisson com média $\langle 6 \rangle$. Os valores foram obtidos das equações 3-1 e 3-2, e entre parênteses os valores medidos em redes geradas pelo MC.	34
Tabela 4.1	Valores de a (método 1) para valores de γ .	66

Lista de Abreviaturas

BA – Barabasi-Albert

ER – Erdős-Rényi

MC – modelo de configuração

1

Introdução

A teoria de redes complexas tem como substrato a teoria de grafos, que surgiu por volta de 1735 quando Euler propôs a solução do problema das sete pontes de Königsberg [1]. Um grafo é uma representação matemática de como elementos de um conjunto estão conectados entre si. Cada elemento do conjunto é chamado de vértice (ou nó) e as ligações entre eles, arestas. As redes complexas são um tipo de grafo com propriedades estruturais não triviais, que veremos posteriormente, diferentemente de grafos completamente aleatórios ou completamente regulares [2, 3].

No mundo moderno, as redes complexas constituem grande parte do cotidiano das pessoas. Por exemplo, notamos que dependemos fortemente de energia elétrica para problemas simples de nosso cotidiano, e para usufruir dessa energia elétrica, estamos conectados em uma rede de estações de energia que constitui um complexo sistema de distribuição. Esse sistema pode ser representado mediante uma rede. Outro exemplo são as redes sociais. As pessoas podem ser consideradas nós de uma rede, enquanto as amizades entre elas são as ligações que conectam esses nós. Investigar relações de amizades entre pessoas pode trazer conhecimento do quão rápido opiniões se espalham numa determinada sociedade. Diversos outros sistemas do nosso dia a dia podem ser estudados através de redes complexas, como uma rede de gasodutos de gás natural, a rede de internet mundial, rede de vasos sanguíneos, rede de colaboração entre cientistas, dentre muitas outras [4, 5, 6, 7]. Portanto, as redes complexas representam uma maneira de representar um dado sistema onde não apenas os elementos constituintes mas também suas relações serão relevantes para a descrição do todo. Está é a ideia básica do que é um Sistema Complexo [8], em que estados coletivos emergem da interação de suas unidades constituintes.

Um dos modelos mais simples e famoso de redes foi proposto por volta de 1960 por Erdős e Rényi [9, 10] onde as ligações entre um par de elementos são aleatórias. Esse modelo até hoje é regularmente usado como comparação e referência em diversos estudos de sistemas complexos. Porém novos modelos foram propostos na tentativa de entender e reproduzir redes reais. Esses modelos possuem características que não aparecem no modelo de Erdős e

Rényi, ou seja, podem apresentar topologias bastante distintas das redes aleatórias.

Um dos trabalhos nesse sentido foi proposto por Watts e Strogatz [2], contendo o conceito de rede de mundo pequeno, onde a topologia não é totalmente regular (em que todos os nós possuem o mesmo grau¹) nem completamente aleatória, mas sim algo intermediário entre esses dois extremos.

Contemporâneo ao trabalho de Watts e Strogatz, foi apresentado por Barabási e Albert [11] um modelo de rede onde as ligações entre seus vértices apresentam uma distribuição de graus que decai como uma lei de potência. Isso ocorre pois novos elementos adicionados à rede tendem a se conectar preferencialmente a elementos que possuem um alto número de ligações.

Ainda refletindo sobre as causas por trás da popularidade das redes complexas, podemos citar sua flexibilidade e capacidade de generalização ao representar os mais variados sistemas, tanto naturais quanto artificiais. No entanto, mesmo com o grande avanço obtido pela teoria, ainda existem limitações, e são constantemente realizados esforços para definir métricas que permitam analisar redes reais, bem como critérios para caracterização e classificação de tais redes [12], e também para a construção de redes sintéticas [13].

Dentro deste cenário, nosso trabalho tem como objetivo comparar diferentes procedimentos para gerar redes e também tentar aprimorar esses métodos com o objetivo de controlar com precisão, particularmente a conectividade média. Por ser um dos parâmetros de controle nas transições de fase que ocorrem em redes [2, 14, 15], a conectividade média pode variar muito dependendo dos demais parâmetros que são fixados.

Este trabalho está estruturado da seguinte forma. No capítulo 2, apresentamos conceitos básicos de teoria de grafos, além dos fundamentos teóricos necessários para nossa análise. No capítulo 3, foi feito um estudo de alguns modelos de redes, como as redes de Erdős-Rényi e as livres de escala através do modelo de configurações. No capítulo 4, apresentamos um novo método para controlar com precisão o grau médio (primeiro momento da distribuição de graus) de redes livre de escala. Também mostramos resultados da aplicação desse método, para valores diferentes do expoente da lei de potência, fixando os graus mínimo, máximo, e médio, a fim de observar os efeitos desses parâmetros no coeficiente de agrupamento e na correlação de graus entre sítios vizinhos. No capítulo 5, encerramos com algumas conclusões e comentários finais sobre os resultados.

¹O grau da rede será definido na seção 2.2.1

2

Fundamentos teóricos de redes complexas

Neste capítulo apresentaremos conceitos básicos e propriedades de redes complexas que serão utilizados posteriormente. Na seção 2.1 serão apresentados conceitos básicos de redes, como vértices e arestas, além de mostrar duas maneiras de armazenar esses grafos. Na seção 2.2 descreveremos algumas medidas estatísticas e métricas necessárias para nossa análise, como a distribuição de graus, momentos da distribuição de graus e coeficiente de aglomeração.

2.1

Conceitos básicos

As redes complexas são um tipo de grafo com propriedades estruturais não triviais, diferentemente de grafos completamente aleatórios ou completamente regulares [2, 3]. Como dito na introdução, um grafo é uma estrutura composta por uma coleção de pontos (nós ou vértices) ligados por arcos (arestas). Assim, um grafo $\mathcal{G} = G(N, M)$ é representado por um conjunto de N pontos $\mathcal{N} = \{n_1, n_2, n_3, \dots, n_N\}$ ligados por um número M de arestas $\mathcal{M} = \{e_1, e_2, e_3, \dots, e_M\}$. A existência de conexão entre os vértices i e j em \mathcal{N} é representada por um par $e_k = (i, j)$, conseqüentemente, a descrição completa de uma rede pode ser especificada dando o número N de vértices e uma lista de todas as arestas [16]. Um exemplo é dado na figura 2.1(a), onde $N = 6$ e as arestas são $(1, 5)$, $(1, 2)$, $(2, 3)$, $(2, 4)$, $(3, 5)$, $(3, 4)$ e $(3, 6)$. Essa forma de representação é muitas vezes chamada de *lista de arestas* e em alguns casos é usada em implementações computacionais, no entanto não é prática para o tratamento matemático [17], veremos em seguida outras representações. Os grafos também podem ser direcionados e ser dado peso às arestas, mas estas possibilidades não serão consideradas aqui. Além disso, neste trabalho serão considerados apenas grafos simples, ou seja, grafos que não possuem arestas múltiplas ou auto-arestas, como mostrado na figura 2.1(b).

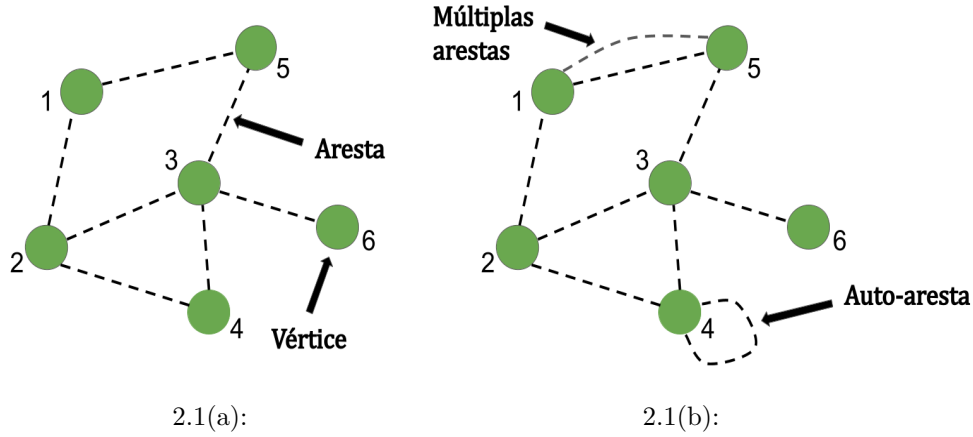


Figura 2.1: (a) Rede com $N = 6$ vértices e $M = 7$ arestas. (b) Representação de arestas múltiplas e auto-arestas.

Na literatura, existem duas principais maneiras de armazenar um grafo: através de uma matriz de adjacência ou de uma lista de adjacência [18, 19, 20]. Essas estruturas são melhores representações de uma rede do que a lista de arestas, como veremos em seguida, e são mais adequadas para nossos propósitos ao longo desse trabalho.

Uma matriz de adjacência A é uma matriz quadrada $N \times N$ e cada elemento $A_{ij} = A_{ji}$ representa a conexão entre o par de vértices ij . Nesse caso, se $A_{ij} = 1$, existe uma aresta entre os vértices i e j , caso contrário $A_{ij} = 0$. Logo, trata-se de uma matriz simétrica. Consequentemente, a matriz de adjacência A pode ser definida como:

$$A_{ij} = \begin{cases} 1 & \text{se existir ligação entre os vértices } i \text{ e } j, \\ 0 & \text{caso contrário,} \end{cases} \quad (2-1)$$

e $A_{ii} = 0$ pois não são permitidas auto-conexões.

Ou seja, trataremos apenas a existência ($A_{ij} = 1$) ou não ($A_{ij} = 0$) de uma dada conexão, sem dar peso nem orientação à mesma. No entanto poderíamos indicar diferentes valores para as conexões A_{ij} (peso) e em uma rede direcionada $A_{ij} \neq A_{ji}$ em geral, ou seja, se o vértice i está conectado com o vértice j isso não implica no oposto. Portanto, nesse caso, a matriz de adjacência A não seria simétrica, mas não consideraremos estes casos.

A segunda forma de representar uma rede é a lista de adjacência, onde apenas representa-se quais vértices estão conectados entre si. A lista de adjacência possui N linhas (uma para cada vértice) e cada linha contém todos os rótulos i dos vértices que estão conectados. Um exemplo para uma pequena rede pode ser visto na tabela 2.1.

Cada uma dessas duas representações tem suas vantagens e desvantagens.

Vértices	Arestas
1	3, 4
2	
3	4, 1
4	5, 1, 3
5	4

Tabela 2.1: Representação de uma rede com 5 vértices através de uma lista de adjacência. Como observado, o vértice 1 tem ligações com os vértices 3 e 4. O vértice 2 não possui ligações. Já os vértices 3, 4 e 5 possuem ligações.

Em uma matriz de adjacência, pode-se adicionar ou remover arestas facilmente, ou, se queremos verificar se existe uma aresta entre o par de vértices i e j , é preciso somente inspecionar o elemento A_{ij} da matriz. Contudo, ela ocupa um espaço da ordem $\mathcal{O}(N^2)$ [21], mesmo para redes esparsas com poucas arestas, pois para redes grandes a representação por matriz de adjacência será composta principalmente por zeros. Outra desvantagem surge quando se quer encontrar os vértices adjacentes, pois os vizinhos de um vértice i são os elementos diferentes de zero na i -ésima linha da matriz, logo é necessário percorrer todos os elementos dessa linha da matriz.

Em uma lista de adjacência (ver o exemplo da tabela 2.1), cada linha i terá uma sub-lista com vizinhos do sítio i . Normalmente, existe uma composição de N sub-listas de adjacência, uma para cada vértice. Para descobrir se há uma aresta entre o par de vértices i e j , é preciso visitar a sub-lista de adjacência de i , e então procurar o elemento j nessa sub-lista. Nesse caso, o tempo de busca é reduzido substancialmente em comparação ao da matriz de adjacência, pois não é preciso percorrer todos os elementos da rede para encontrar uma conexão e sim somente os elementos com os quais o vértice i está conectado [18, 22]. No exemplo de lista de adjacência dado na tabela 2.1, cada aresta aparece duas vezes. A presença de uma aresta ente 1 e 4 indica que o vértice 4 está listado como adjacente ao vértice 1 e o vértice 1 também está listado como adjacente ao vértice 4. Consequentemente, para reproduzir M arestas, é necessário armazenar $2M$ inteiros, o que é mais eficiente que armazenar os N^2 elementos da matriz de adjacência, pois $2M < N^2$. A figura 2.2 mostra o mapeamento da rede 2.1 em matriz de adjacência e lista de adjacência.

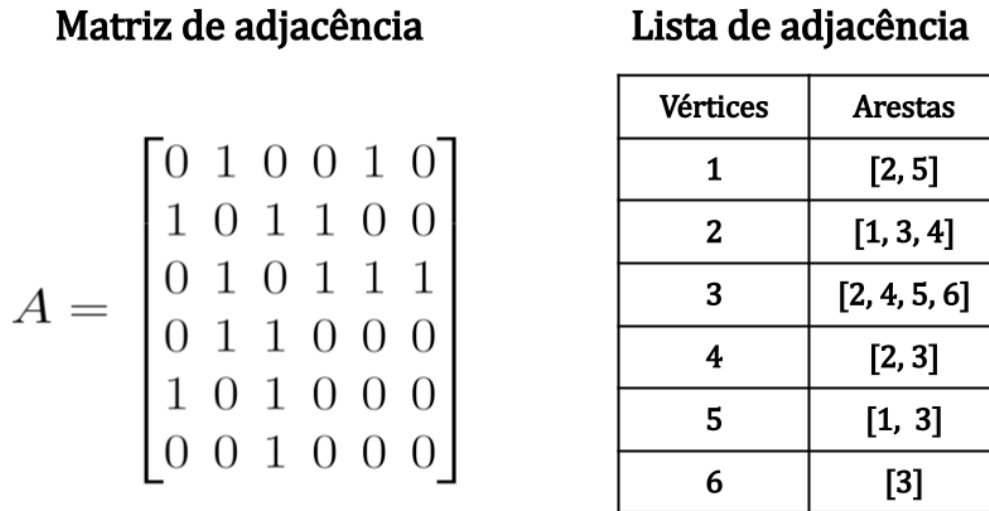


Figura 2.2: Representação por matriz de adjacência e lista de adjacência. Ambos os casos representam a rede da figura 2.1(a).

2.2

Propriedades das redes

2.2.1

Grau

Uma medida simples para a descrição de uma rede é fornecida pelo *grau* de um vértice, que é a quantidade de nós ligados a ele. Tomando como exemplo a rede mostrada na figura 2.1(a), o vértice 1 possui grau 2, o vértice 2 possui grau 3, e assim por diante. Deste modo, para cada vértice i da rede estará associado um grau k_i . O grau k_i pode ser escrito em termos da matriz de adjacência como

$$k_i = \sum_{j=1}^N A_{ij}. \quad (2-2)$$

Sendo que cada aresta está conectada a dois vértices, ao somar os graus k_i sobre todos os vértices i da rede, o resultado do somatório será $2M$, isto é,

$$\sum_{i=1}^N k_i = 2M, \quad (2-3)$$

pois cada ligação será contada duas vezes.

O grau médio ou *conectividade média*, sobre todos os vértices da rede, é

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (2-4)$$

2.2.1.1

Distribuição de graus

Uma das propriedades mais fundamentais em teoria de redes complexas é a distribuição de graus. Em redes não direcionadas, a distribuição de graus é definida como a probabilidade de um vértice, escolhido ao acaso, ter um determinado grau k e pode ser expressa como $P(k)$. Uma vez que a rede tenha N vértices, vamos definir N_k como o número de vértices com grau k . O valor de N_k satisfaz a condição $\sum_k N_k = N$, o que permite definir, no limite termodinâmico,

$$P(k) = \lim_{N \rightarrow \infty} \frac{N_k}{N}. \quad (2-5)$$

Como a equação 2-5 define uma distribuição de probabilidade, esta função deve ser normalizada ($\sum_k P(k) = 1$).

Outra maneira de representar a distribuição de graus, que contém a mesma informação, é a *sequência de graus*. A sequência de graus de uma rede, nada mais é que um vetor $(k_1, k_2, k_3, \dots, k_N)$ de graus onde o índice de cada elemento representa o vértice. Para melhor visualização desse conceito, imagine que a rede representada na figura 2.1(a) possui graus fixos, dessa forma sua sequência de graus é representada da seguinte maneira $(2, 3, 4, 2, 2, 1)$, onde $k_1 = 2$ indica que o vértice com rótulo 1 tem grau dois, $k_2 = 3$ indica que o vértice com rótulo 2 tem grau três, e assim por diante.

2.2.1.2

Momentos da distribuição de graus

Os momentos das distribuições são de grande interesse em redes, pois a partir deles é possível calcular diferentes métricas.

O primeiro momento da distribuição de graus é o grau médio $\langle k \rangle$, apresentado na equação 2-4. Usando a distribuição $P(k)$, é dado por

$$\langle k \rangle = \sum_k k P(k). \quad (2-6)$$

O segundo momento da distribuição é

$$\langle k^2 \rangle = \sum_k k^2 P(k). \quad (2-7)$$

Em geral, o m -ésimo momento é dado por

$$\langle k^m \rangle = \sum_k k^m P(k), \quad (2-8)$$

onde tipicamente m é um inteiro positivo, entretanto em alguns cálculos em redes valores negativos e racionais também são relevantes [15]. Entretanto, o objetivo central deste trabalho será controlar o primeiro momento.

2.2.2

Caminhos e distâncias

Dois conceitos úteis em teoria de redes são caminhos e a distâncias. Um *caminho* é definido como qualquer sequência de vértices onde cada vértice seguinte na sequência está conectado por uma aresta. Pode-se dizer que um caminho é uma rota que passa de vértice em vértice através da rede, no entanto esse caminho só pode ser feito onde há ligações. Geralmente, existem diversos caminhos de um vértice i para um vértice j em uma rede. O comprimento de um dado caminho é o número de arestas percorridas através deste caminho. Já a distância entre dois vértices é o comprimento mínimo.

Neste trabalho não temos o objetivo de medir caminhos ou distâncias entre vértices, no entanto, a partir desses conceitos é possível definir outra grandeza que está associada à rede estar completamente conectada ou não. Essa grandeza é chamada *componente* e pode ser medida percorrendo um caminho que contém todos os vértices daquela componente da rede usando um algoritmo de busca em profundidade, como descrito na próxima seção.

2.2.2.1

Componentes em uma rede

Tecnicamente, uma componente em um grafo não direcionado é um conjunto onde todo pare de vértices do conjunto está conectado por um caminho e desconectados do resto da rede [23]. Um vértice totalmente desconexo na rede também representa uma componente. De maneira geral, uma rede pode ter uma ou mais componentes. Na figura 2.3, a rede apresentada exhibe três componentes, isto é, três subgrafos sem conexão entre eles. Por exemplo, não existe uma aresta ligando os nós 3 e 5 tornando assim a rede desconectada. Se todos os vértices da rede tivessem conexões uns com os outros, então a rede seria totalmente conectada, ou seja, estaria formada por apenas uma componente.

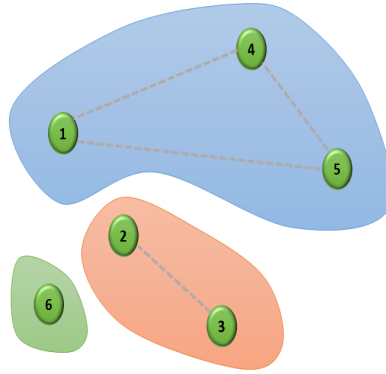


Figura 2.3: Representação das componentes de uma rede contendo seis vértices. Nesse caso existem três componentes.

A maneira mais difundida para calcular componentes de uma rede é usando o algoritmo de *busca em profundidade* ou do inglês Depth-First Search (DFS). Esse algoritmo é reconhecido como uma poderosa técnica para resolver diversos problemas em redes, como encontrar caminhos entre dois vértices, encontrar componentes fortemente conectadas, entre outros [24, 25].

O algoritmo DFS consiste em percorrer a rede por suas arestas, de vértice em vértice, até que não haja mais vértice a serem visitados. Esse caminho pode ser inicializado a princípio em qualquer vértice da rede, e utilizamos marcadores à medida que um vértice é visitado.

Para calcular o número de componentes da rede todos os vértices serão inicializados como não visitados, em seguida a busca em profundidade é feita a partir de um vértice qualquer. A ideia é usar uma variável para contar as componentes, isto é, depois de o algoritmo ter sido executado uma vez será contado 1 nessa variável. Após percorrer todos os vértices, é verificado se ainda existe algum vértice não visitado, se sim, novamente o algoritmo DFS é executado em um desses vértices e após percorrer todos os vértices, o valor 1 é aumentado na variável de contagem. Esse processo é feito até que todos os vértices da rede estejam visitados.

A rede da figura 2.4 apresenta a ilustração de uma rede desconexa com uma componente maior, ou seja, maior parte da rede, juntamente com algumas componentes menores não conectadas à componente maior. Para calcular as componentes foi usado o algoritmo DFS que contou 10 componentes.

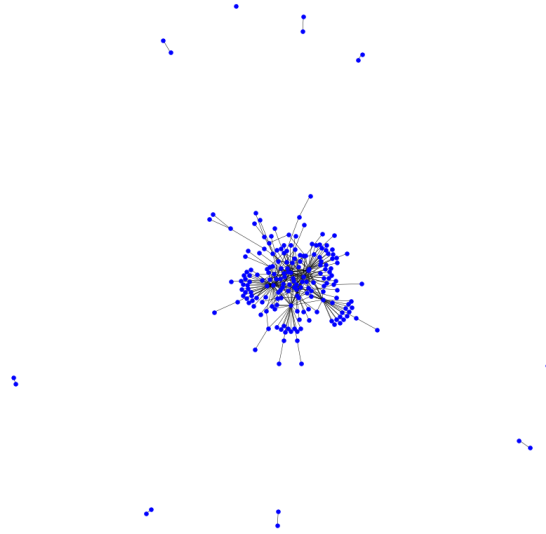


Figura 2.4: Rede desconexa de tamanho $N = 200$ com 10 componentes.

2.2.3

Correlação de graus e assortatividade

Em uma rede de colaboração de atores de cinema, os atores representam vértices e as relações adquiridas em grupos que formam os elencos dos filmes são consideradas as arestas. Nesse caso, atores melhor relacionados, isto é, com grau alto, tendem a se conectar a atores com graus semelhantes, e o mesmo acontece para atores que possuem poucas relações [26]. Nessa situação, a correlação é chamada *assortativa*, como observado em geral em redes sociais. Em contrapartida, o padrão de ligações presentes em muitas redes tecnológicas e biológicas, é que vértices com alto grau se conectam a vértices com baixo grau mais facilmente. Esse tipo de correlação é conhecida como *dissassortativa* [27].

Na literatura, diversos trabalhos propõem maneiras de calcular a correlação entre graus [26, 28, 29, 30]. Um caminho é através da probabilidade condicional $P(k'|k)$ de que um vértice tenha grau k' dado que está conectado a outro com grau k . Uma abordagem mais direta, que está relacionada a essa probabilidade $P(k'|k)$, é calcular o grau médio dos vizinhos de um vértice i , ou seja,

$$k_{nn,i} = \frac{1}{k_i} \sum_j k_j A_{ij}. \quad (2-9)$$

Por meio da figura 2.5, é possível visualizar de forma prática como calcular $k_{nn,i}$.

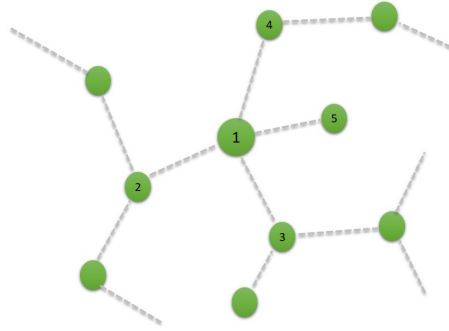


Figura 2.5: Ilustração de como calcular $k_{nn,i}$ para o vértice 1. Nesse caso o grau do vértice 1 é $k_1 = 4$. Os graus dos vértices vizinhos a 1 são $k_2 = 3, k_3 = 3, k_4 = 2, k_5 = 1$. Logo $k_{nn,i} = (3 + 3 + 2 + 1)/4 = 2.25$.

Por outro lado é, conveniente tomar a média sobre todos os vértices com o mesmo grau k , logo

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i|k_i=k} k_{nn,i}, \quad (2-10)$$

onde N_k representa o número de vértices com grau k e a soma é realizada sobre todos os vértices i com graus iguais, $k_i = k$. Considerando a probabilidade condicional, $k_{nn}(k)$ pode ser escrito da seguinte maneira,

$$k_{nn}(k) = \sum_{k'} k' P(k'|k). \quad (2-11)$$

A função $k_{nn}(k)$ define o nível de assortividade da rede. Se essa função for crescente, a rede é assortativa, se for decrescente, a rede é dissassortativa. Caso seja constante, a rede não possui correlação entre graus.

Para redes não correlacionadas, a probabilidade condicional não depende de k [13] e pode ser escrita da seguinte forma,

$$P(k'|k)^{nc} = \frac{k' P(k')}{\langle k \rangle}. \quad (2-12)$$

Considerando as equações 2-11, e 2-12, definimos também a quantidade

$$k_{nn}^{nc}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}. \quad (2-13)$$

Outra forma de medir correlação de graus é através do *coeficiente de correlação de Pearson* r . Essa quantidade também computa a tendência de vértices vizinhos terem graus semelhantes ou diferentes e pode ser definida como [30],

$$r = \frac{\langle k \rangle \sum_k k^2 k_{nn}(k) P(k) - \langle k^2 \rangle^2}{\langle k \rangle \langle k^3 \rangle - \langle k^2 \rangle^2}. \quad (2-14)$$

O coeficiente r varia no intervalo $[-1, 1]$. Se $r > 0$, a rede será considerada *assortativa*, enquanto, se $r < 0$, será *disassortativa*. Substituindo $k_{nn}^{nc}(k) = \langle k^2 \rangle / \langle k \rangle$ na equação 2-14 obtêm-se $r = 0$, isto é, a rede é não correlacionada.

2.2.4

Coeficiente de aglomeração

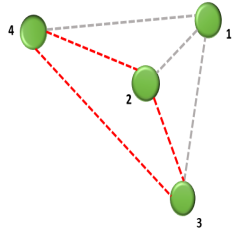
O coeficiente de aglomeração, ou de agrupamento, considera a vizinhança local de uma rede, porém buscando a existência de triângulos entre três vértices, quer dizer, três vértices ligados entre si. Essa propriedade é conhecida como *transitividade* e é definida como a probabilidade de que dois vizinhos de um vértice da rede também estejam conectados entre si [17].

Em um contexto de redes sociais, por exemplo, se um vértice i está conectado por uma aresta ao j , e j está conectado ao vértice k , então existe uma alta probabilidade de i estar conectado a k . Isso pode significar que o amigo do amigo também será amigo.

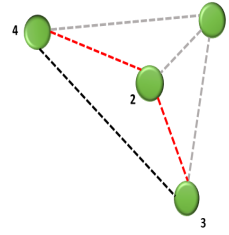
É possível medir o coeficiente de aglomeração para um vértice i , como a proporção de conexões entre os vértices na vizinhança de i dividido pelo maior número de arestas possível entre seus vértices vizinhos. Sendo assim, se e_i é o número de arestas entre os vizinhos de i , e $k_i(k_i - 1)/2$ é o número total de conexões possíveis, então o coeficiente de aglomeração local c_i será

$$c_i = \frac{2e_i}{k_i(k_i - 1)}. \quad (2-15)$$

Note que definido dessa forma, não é possível calcular c_i para $k_i = 1$ ou $k_i = 0$, então a equação 2-15 só é válida para $k_i > 1$. A figura 2.6 mostra um exemplo de como calcular o coeficiente de aglomeração. Nesse caso, o vértice 1 está conectado aos vértices 2, 3 e 4 (arestas em cinza), as arestas em vermelho na figura 2.6(a) são as possíveis conexões que podem ocorrer entre os vizinhos de 1, calculado via $k_i(k_i - 1)/2$. Na figura 2.6(b) temos uma conexão (aresta em preto) dentre as três possíveis, consequentemente o coeficiente de aglomeração para o vértice 1 seria $c_1 = 1/3$.



2.6(a):



2.6(b):

Figura 2.6: (a) Mostra todas as possíveis ligações da vizinhança do vértice $i = 1$, nesse caso $c_1 = 0$, pois não há conexões entre seus vizinhos. Em (b), dentre todas as conexões possíveis, os vértices $i = 3$ e $i = 4$ estão unidos por uma aresta, nesse caso existe uma ligação dentre as três possíveis, então $c_1 = 1/3$.

Tendo em vista que qualquer rede é totalmente descrita por sua matriz de adjacência, é possível definir o coeficiente de aglomeração para um grafo simples não direcionado em termos de A como

$$c_i = \frac{1}{k_i(k_i - 1)} \sum_{j,k} A_{ij} A_{jk} A_{ki}, \quad (2-16)$$

onde $k_i = \sum_j A_{ij}$. Caso, k_i seja zero ou um, então $c_i = 0$.

Uma vez que o coeficiente de aglomeração local é definido, pode-se agora calcular o coeficiente de aglomeração da rede como um todo, isto é, o coeficiente de aglomeração global C da rede. Assim C é definido como a média aritmética sobre todos os vértices da rede [2],

$$C = \frac{1}{N} \sum_i c_i. \quad (2-17)$$

É importante considerar que uma rede aleatória difere muito de uma rede real por conta de seu coeficiente de aglomeração. No geral, redes reais tendem a ter coeficiente de aglomeração muito alto, enquanto em uma rede aleatória tende a zero [31].

Para uma rede aleatória não correlacionada [32]

$$C^{mc} = \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{N \langle k \rangle^3}. \quad (2-18)$$

Ao calcular o coeficiente de aglomeração obtêm-se uma noção importante

sobre as correlações da rede. De acordo com [14], existe uma relação direta entre o coeficiente de correlação de Pearson e o coeficiente de aglomeração prescrito na equação 2-17.

3

Modelos de redes: revisão e estudo comparativo

Neste capítulo faremos uma revisão de alguns modelos de redes a começar na seção 3.1 pelo modelo de configuração, uma poderosa ferramenta para criar redes. Na seção 3.2 será abordada a rede de Erdős-Rényi e duas construções possíveis para esse modelo, chamadas de $G(N, p)$ e $G(N, M)$. Na seção 3.3, o objetivo será criar redes de Erdős-Rényi utilizando o modelo de configuração para então comparar com os modelos anteriores. Na seção 3.4, apresentaremos as redes livres de escalas, que também serão construídas utilizando o modelo de configuração. Para todas as redes será feita uma análise comparativa da estatística de suas métricas, considerando principalmente o grau médio entre vizinhos $k_{nn}(k)$ (Eq. 2-10), o coeficiente de aglomeração local c_i (Eq. 2-15), o coeficiente de aglomeração global C (Eq. 2-17) e o coeficiente de correlação de Pearson r (Eq. 2-14).

3.1

Modelo de configuração (MC)

O modelo de configuração consiste em produzir redes a partir de uma dada sequência de graus, ligando os vértices aleatoriamente. Essa forma de criar redes aleatórias é diferente do modelo de Erdős-Rényi, descrito na seção

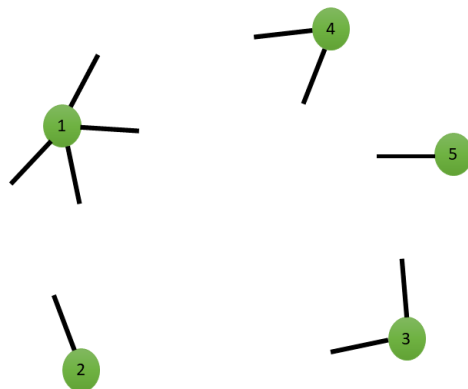


Figura 3.1: Meias arestas para a sequência de graus (4, 1, 2, 2, 1).

3.2, e se aplica também para qualquer distribuição $P(k)$ [33].

No modelo de configuração, a cada vértice de grau k é associado um número k de *meias arestas* como mostra a figura 3.1. Partindo deste ponto, o segundo passo é sortear duas meias arestas aleatoriamente e conectá-las.

Para exemplificar o processo, a figura 3.1 apresenta um total de 10 meias arestas pré-definidas pela sequência de graus $(4, 1, 2, 2, 1)$. Para sortear duas delas será necessário definir uma lista de meias arestas, isto é, se na primeira entrada da sequência de graus temos $k_1 = 4$, as quatro primeiras entradas da listas de meias arestas serão representadas pelo índice do vértice (1 nesse caso) quatro vezes. Para a sequência dada $(1, 1, 1, 1, 2, 3, 3, 4, 4, 5)$. Na figura 3.2, duas meias arestas foram escolhidas de forma aleatória. Após o sorteio de duas entradas da lista de meias arestas, uma ligação foi feita entre essas duas meias arestas.

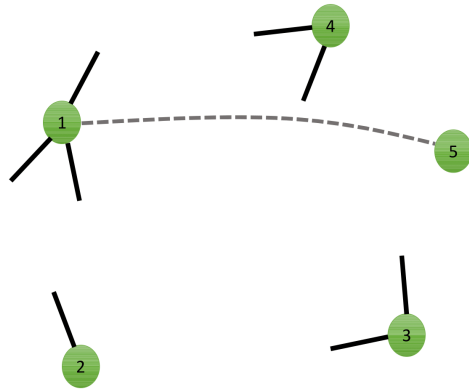
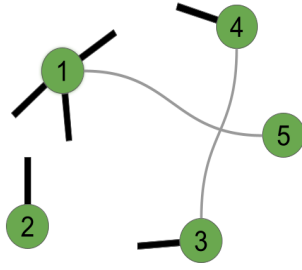
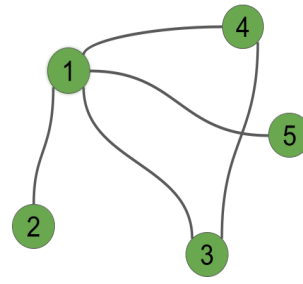


Figura 3.2: Primeira aresta gerada a partir do modelo de configuração.

Considerando que nesse processo há um total de $\sum_i k_i = 2M$ meias arestas, ao realizar o primeiro sorteio sobram $2M - 2$, ou seja, no início tínhamos no total 10, após o primeiro sorteio sobraram 8. Como a cada sorteio, duas meias arestas são retiradas, o número total delas necessariamente precisa ser par, para que a rede esteja completamente conectada ao final do processo. As etapas citadas acima, ocorrem até que todas as meias arestas estejam totalmente conectadas, como na figura 3.3(b).



3.3(a):



3.3(b):

Figura 3.3: Em (a), os vértices três e quatro foram sorteados e uma conexão foi colocada entre eles. Em (b), é mostrada a rede finalizada, após completar todo o processo do modelo de configuração.

Assim sendo, para uma mesma sequência de graus é possível ter diversas configurações para as conexões das meias arestas. Logo, o modelo de configuração é definido como o conjunto em que todas as combinações entre pares de meias arestas apresentam igual probabilidade, isto é, dada qualquer sequência de graus, é possível conectar as meias arestas de par em par de forma uniformemente aleatória [34].

Definir o modelo desta forma implica em permitir que auto-arestas e múltiplas arestas ocorram, pois impedi-las privilegiaria conexões entre vértices com determinados graus. Consequentemente, isso permite solucionar muitas propriedades relacionadas a esse modelo [31].

Segundo Newman [17], se um algoritmo para gerar a rede impedisse a criação de auto-arestas ou múltiplas arestas, a distribuição de correspondências (entre pares) não seria uniforme.

Permitir auto-conexões e conexões múltiplas, nesse caso, não afeta a solução do problema, pois à medida que a rede se torna grande, o número de auto-arestas e arestas múltiplas não depende do tamanho da rede. Para evidenciar esse fato, Newman ainda propõe [17] que o número esperado de arestas múltiplas e de auto-arestas pode ser calculado pelas equações 3-1 e 3-2, respectivamente, isto é,

$$\frac{1}{2M^2} \sum_{ij} k_i k_j (k_i - 1)(k_j - 1) = \frac{1}{2} \left[\frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} \right]^2, \quad (3-1)$$

e

$$\sum_i \frac{k_i(k_i - 1)}{4M} = \frac{\langle k^2 \rangle - \langle k \rangle^2}{2\langle k \rangle}. \quad (3-2)$$

Deste modo, para N suficientemente grande, o valor esperado de auto-arestas e arestas múltiplas permanece constante conforme a rede cresce em número de vértices, desde que $\langle k^2 \rangle$ seja constante e finito.

A tabela 3.1 mostra o valor esperado de arestas múltiplas e auto arestas de acordo com as equações 3-1 e 3-2, respectivamente, conforme N aumenta. As redes foram criadas utilizando o modelo de configuração. A sequência de graus foi estabelecida usando uma distribuição de Poisson com $\langle k \rangle = 6$. Como visto em 3.1, os valores referentes a auto e múltiplas arestas permanecem bem próximos de valores constantes conforme N cresce.

N	Múltiplas arestas	Auto arestas
100	17,8 (17,5)	2,98 (2,94)
500	18,7 (18,4)	3,06 (2,44)
1000	18,6 (17,1)	3,05 (2,98)
5000	17,7 (17,8)	2,97 (3,16)
10000	18,3 (17,4)	3,02 (3,10)

Tabela 3.1: Esses valores correspondem a uma sequência de graus prescrita por uma distribuição de Poisson com média $\langle 6 \rangle$. Os valores foram obtidos das equações 3-1 e 3-2, e entre parênteses os valores medidos em redes geradas pelo MC.

Em teoria de grafos, auto conexões e múltiplas conexões são comuns [35], no entanto dificilmente elas aparecem em redes reais. Consequentemente, para os propósitos deste trabalho, auto conexões e múltiplas conexões serão evitadas, isto é, a cada duas meias arestas sorteadas é verificado se já existe uma conexão entre os vértices correspondentes para evitar arestas múltiplas, se sim, outras duas meias arestas são sorteadas novamente. Para evitar auto arestas, os sorteios das meias arestas são necessariamente de vértices diferentes.

3.2

Redes de Erdős-Rényi (ER)

Um importante avanço na pesquisa sobre redes aleatórias ocorreu por volta da década de 1950, quando dois matemáticos húngaros, Paul Erdős e Alfred Rényi publicaram o artigo “On the evolution of random graphs”. Nesse artigo, Erdős e Rényi analisaram os grafos de forma exclusivamente estocástica, diferente do que até então era feito na matemática discreta e na sociologia. A partir dessa premissa, apresentaram um modelo de rede baseado em conexões aleatórias uniformemente distribuídas [9, 36].

Um grafo aleatório, como proposto por ER, é usualmente gerado utilizando um modelo conhecido como $G(N, p)$, onde é fixado um número N de vértices e uma probabilidade p de ocorrer ligações entre esses vértices. Nessa rede, não fixamos o número de arestas, assim é possível que a rede não tenha arestas, ou que haja arestas entre pares distintos de vértices.

Se tratando de propriedades estatísticas das redes, não seria interessante gerar apenas uma única rede, mas sim um *ensemble*. Sendo assim, uma definição técnica para um grafo aleatório não é feita em termos de uma única rede, mas em termos de um *ensemble*, em outras palavras, uma distribuição de probabilidade sobre todas as redes possíveis [31].

Mais especificamente, $G(N, p)$ é o conjunto de redes com N vértices e probabilidade p de dois desses vértices estarem conectados. Tais grafos são chamados simples, pois não possuem várias arestas ligando um par de vértices, ou auto arestas, sendo nula a probabilidade para ambos os casos. Dessa forma, segundo o modelo $G(N, p)$, todos os grafos com N nós e M arestas (não fixadas a priori) têm probabilidade de ocorrer igual a

$$P(G) = p^M (1 - p)^{\binom{M}{2} - M}. \quad (3-3)$$

Uma maneira de construir a rede de ER de N vértices é criando uma matriz de adjacência de $N \times N$ como segue. Os elementos A_{ij} com $i > j$ da matriz adotam o valor 1 com probabilidade p e 0 com probabilidade $1 - p$. Depois simetrizamos o restante da matriz, fazendo $A_{ji} = A_{ij}$. Com essa matriz, estabelecemos as conexões (ou não) entre cada par ij . Isto evita auto-conexões (pois A_{ii}) e conexões múltiplas. A matriz de adjacência 3-4 foi obtida utilizando este algoritmo.

Isto é equivalente a considerar N vértices inicialmente desconectados, percorrer cada par de vértices (i, j) com $i > j$ e colocar uma ligação entre esses vértices com probabilidade independente p [32].

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (3-4)$$

Uma representação da rede gerada pela matriz 3-4 pode ser visualizada na figura 3.4.

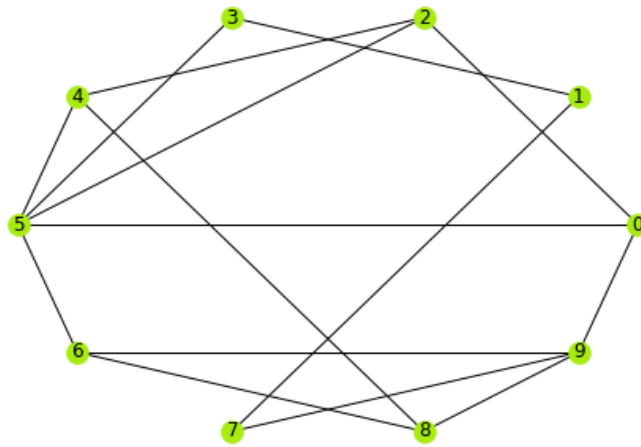


Figura 3.4: Rede de Erdős-Rényi. $N = 10$ e $p = 0.4$

Para o modelo $G(N, p)$, o número de arestas não é definido. Consequentemente, ao produzir um determinado número de redes, os valores M do número de arestas estarão dispersos em torno de uma média. A figura 3.5 demonstra esse fato a partir de um conjunto gerado por mil redes com $\langle k \rangle = 6$ e $N = 10^3$. Como observado, as flutuações do número M de arestas obedecem uma distribuição gaussiana

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}. \quad (3-5)$$

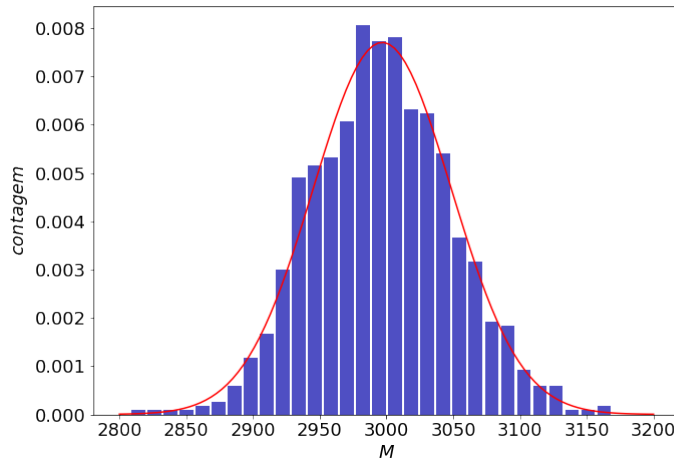


Figura 3.5: Histograma do número de arestas gerado pelo modelo $G(N, p)$, com $N = 10^3$ e $\langle k \rangle = 6$. O histograma em azul representa a distribuição do número de arestas M gerado por 10^3 redes. A curva em vermelho é a distribuição gaussiana (Eq. 3-5) com média $\mu = 2996.82$ e desvio padrão $\sigma = 51,76$, obtidos da sequência de graus.

Por outro lado, também é possível construir uma rede de ER não em termos da probabilidade p de haver uma aresta, mas sim considerando um número fixo de arestas M . Esse modelo é conhecido como $G(N, M)$. Assim sendo, a rede é inicializada com um número N de vértices desconexos, nesse momento dois vértices i e j são escolhidos aleatoriamente e uma aresta é colocada entre eles. Nesse caso, uma aresta só é colocada entre pares ij distintos e em pares ainda não conectados. Deste modo auto-arestas e arestas múltiplas são evitadas. Esse procedimento ocorre até que as M arestas estejam na rede. Obviamente a conectividade média será definida, de acordo com as equações 2-3 e 2-4, como

$$\langle k \rangle = \frac{2M}{N}. \quad (3-6)$$

A figura 3.6 mostra uma rede gerada pelo algoritmo citado acima.

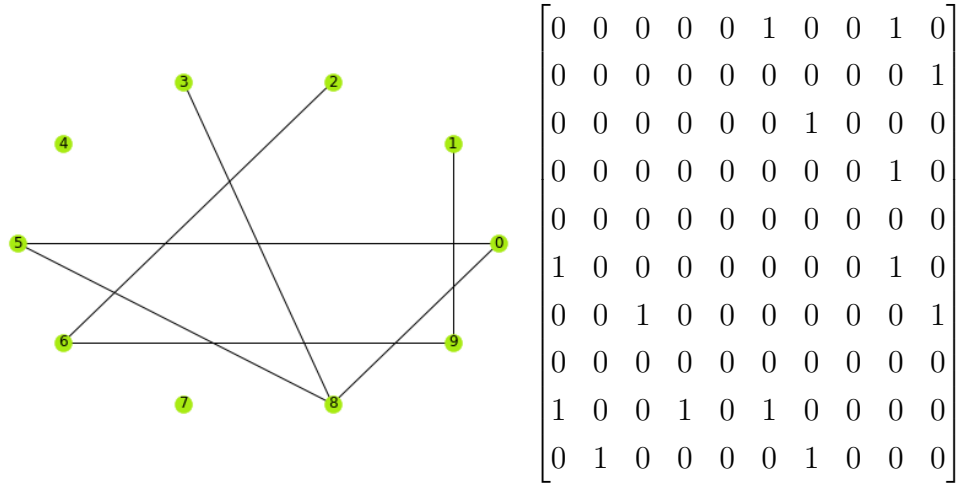


Figura 3.6: Rede criada a partir do modelo $G(N, M)$ e sua matriz de adjacência associada. O tamanho da rede equivale a $N = 10$ e o número de arestas $M = 7$.

3.2.1

Distribuição de graus

Para o modelo $G(N, p)$, cada nó na rede é conectado com probabilidade p a qualquer um dos $N - 1$ nós restantes. Assim, o parâmetro p está contido num intervalo $[0, 1]$.

Considerando o tamanho da rede quando N é suficientemente grande, o número médio de arestas se mantém fixo [17] e equivalente a

$$\langle k \rangle = p(N - 1), \quad (3-7)$$

onde p é probabilidade de conexão entre dois vértices e N o número de vértices da rede.

Outra característica importante é que para o limite em que N é suficientemente alto e p razoavelmente pequeno, a distribuição de graus do modelo de ER é bem aproximada pela distribuição de Poisson, isto é,

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}. \quad (3-8)$$

Os pontos na figura 3.7 são produzidos por simulações que criam tanto o modelos $G(N, p)$ e $G(N, M)$. A curva em vermelho é a distribuição de Poisson, eq. 3-8. É possível observar que o desvio é bem pequeno para redes da ordem de $N = 10^4$.

A figura 3.7 mostra a distribuição de graus para o modelo $G(N, p)$, com $N = 10^4$ e $\langle k \rangle = 6$. Para o modelo $G(N, M)$ foi calculada a distribuição de graus levando em conta $N = 10^4$, $M = 3 \times 10^4$ e $\langle k \rangle = 6$.

De acordo com as figuras 3.7, ambos os modelos $G(N, p)$ e $G(N, M)$, evidenciam uma distribuição de Poisson para os graus da rede 3-8.

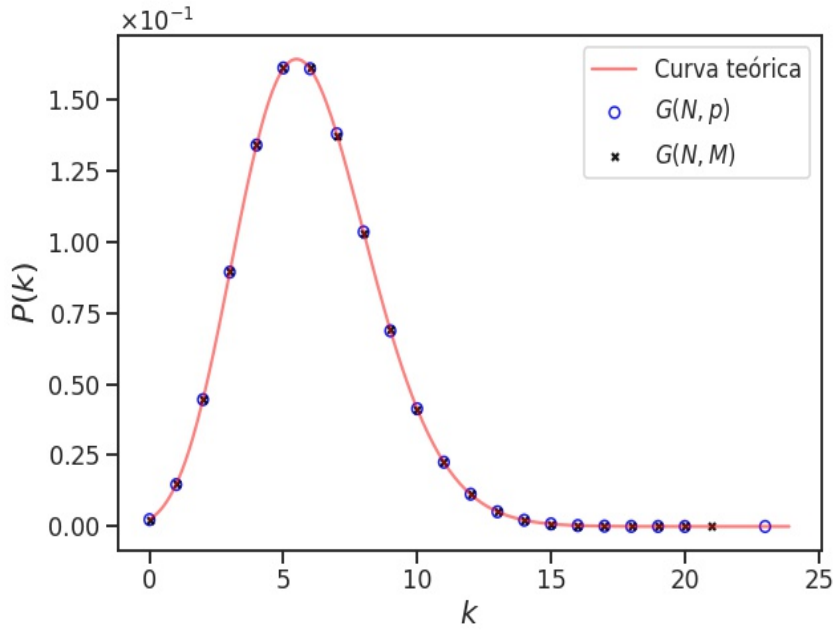


Figura 3.7: Distribuições de graus para redes de ER, geradas pelos modelos $G(N, p)$ e $G(N, M)$, com tamanho $N = 10^4$ e número total de arestas $M = 3 \times 10^4$, portanto $\langle k \rangle = 6$. Os pontos são a média sobre 10^2 redes e a curva em vermelho é a distribuição de Poisson da equação 3-8.

3.2.2

Coefficiente de aglomeração

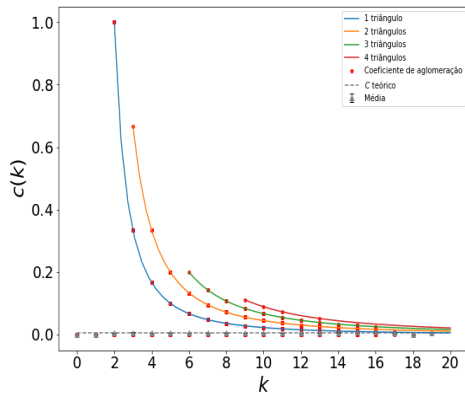
No modelo de ER a probabilidade de que dois vértices quaisquer da rede estejam conectados é exatamente a mesma, isto é, pela equação 3-7 $p = \langle k \rangle / (N - 1)$. Disto posto, a probabilidade de dois vértices estarem conectados é a mesma caso tenham um vizinho mútuo ou não. Consequentemente, $C = p$, e o coeficiente de aglomeração global [37] pode ser definido como,

$$C = \frac{\langle k \rangle}{N - 1}. \quad (3-9)$$

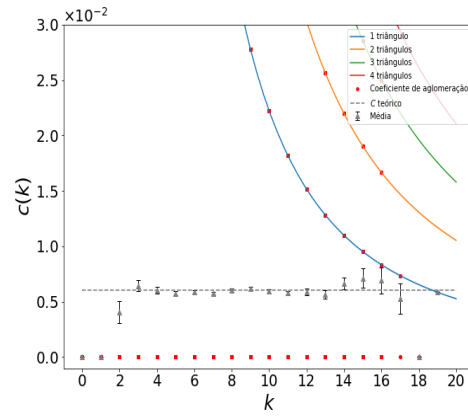
Uma medida local desse coeficiente pode ser obtido para cada vértice a partir da equação 2-15. Nas figuras em 3.8, é possível observar a distribuição do coeficiente de aglomeração local agrupados pelos valores de k da rede de ER utilizando o modelo $G(N, p)$. Na figura 3.8(a), a rede tem tamanho $N = 10^3$ e grau médio $\langle k \rangle = 6$. A figura 3.8(b) corresponde a uma ampliação da figura 3.8(a), onde é possível observar o valor médio de $c(k)$ para todos os nós da rede com determinado grau k . Nas figuras 3.8 (c) e (d), tem-se uma rede com $N = 10^4$ e grau médio $\langle k \rangle = 6$ onde também é representado a curva para $c(k)$ médio.

A princípio pode-se notar que à medida que o grau da rede aumenta, o coeficiente de aglomeração local tende a zero como prevê a equação 2-15. Também é possível observar que para a primeira rede com $N = 10^3$, há maior probabilidade de conexão entre vértices vizinhos, de acordo valor médio observado na figura. Consequentemente, dentro das cem redes geradas com $N = 10^3$, vértices com grau 10 podem apresentar até 4 triângulos sendo formado entre seus vértices vizinhos, enquanto as redes com $N = 10^4$, além de menor probabilidade de conexão entre vértices vizinhos, a formação de triângulos chega a no máximo 3 para graus maiores que 10, por exemplo.

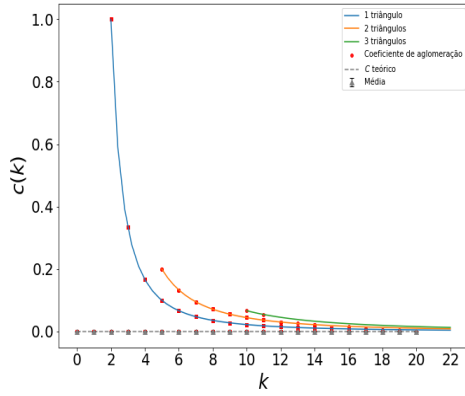
Para a rede de ER com $N = 10^4$, o coeficiente de aglomeração global teórico pode ser calculado via equação 3-9, ou seja, $C \approx 6 \times 10^{-4}$, o valor obtido pela simulação representado na 3.8 foi de $C \approx 5,8 \times 10^{-4}$.



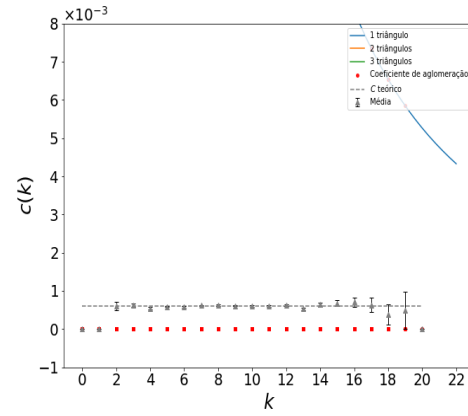
3.8(a):



3.8(b):



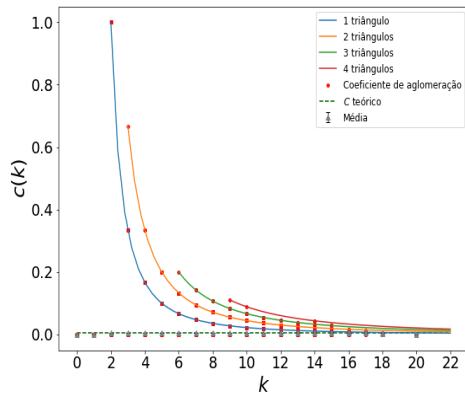
3.8(c):



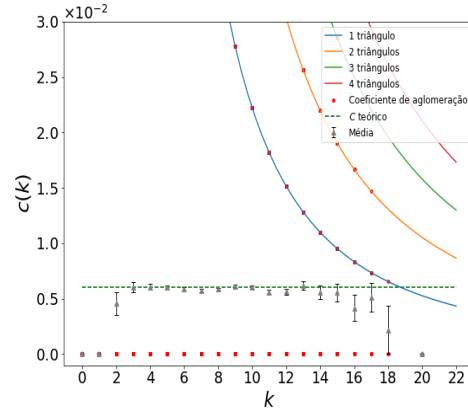
3.8(d):

Figura 3.8: Distribuição de aglomeração para o modelo $G(N, p)$. Nas duas figuras superiores (a) e (b), $N = 10^3$ e $\langle k \rangle = 6$. Para as figuras (c) e (d), $N = 10^4$ e $\langle k \rangle = 6$. As duas figuras à direita (b) e (d), são ampliações das figuras (a) e (c). 10^2 redes foram geradas para as distribuições acima.

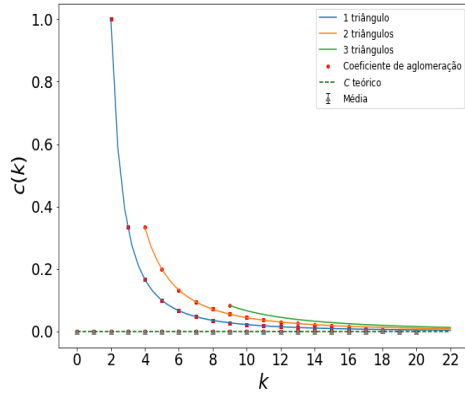
Com o objetivo de confrontar os modelos $G(N, p)$ e $G(N, M)$, assim como na seção 3.2.1, foi criada uma figura análoga a 3.8.



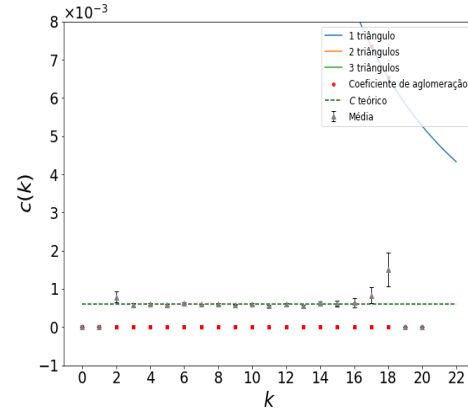
3.9(a):



3.9(b):



3.9(c):



3.9(d):

Figura 3.9: Distribuição de aglomeração para o modelo $G(N, M)$. Nas duas figuras superiores (a) e (b), $N = 10^3$ e $M = 3 \times 10^3$. Para as figuras (c) e (d), $N = 10^4$ e $M = 3 \times 10^4$. As duas figuras à direita, (b) e (d), são ampliações das figuras (a) e (c). 10^2 redes foram geradas para as distribuições acima.

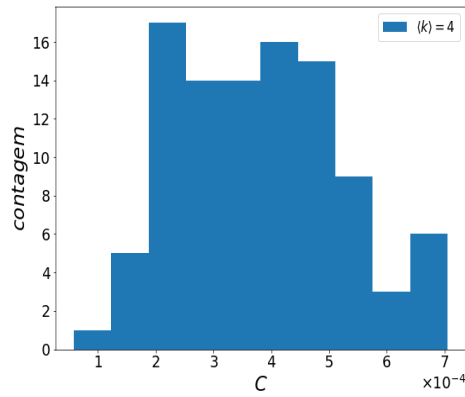
De fato mudanças não são perceptíveis ao analisar as figuras em 3.9, até mesmo ao calcular o coeficiente de aglomeração global da rede (modelo $G(N, M)$), encontramos um valor de $C \approx 5.9 \times 10^{-4}$, bem próximo do modelo $G(N, p)$.

3.2.3

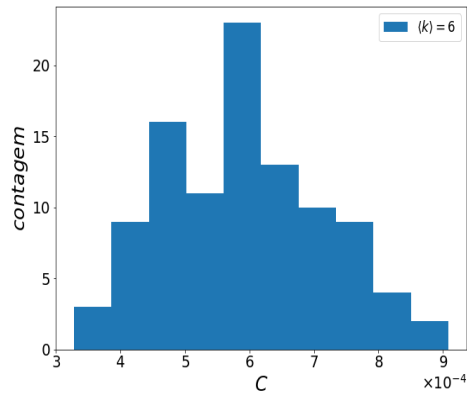
Distribuições de outras métricas dos modelos $G(N, p)$ e $G(N, M)$

Para uma visão mais geral do modelo de ER, as figuras 3.10 mostram a distribuição dos valores de C global calculado via equação 2-17. É possível verificar que para os graus médios $\langle k \rangle = 4, 6, 8, 12$, as redes apresentam

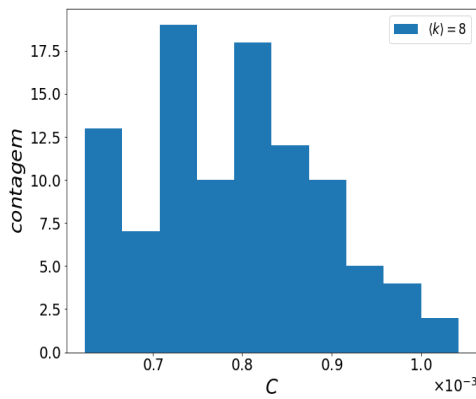
flutuações de C em torno do valor teórico prescrito anteriormente na equação 3-9.



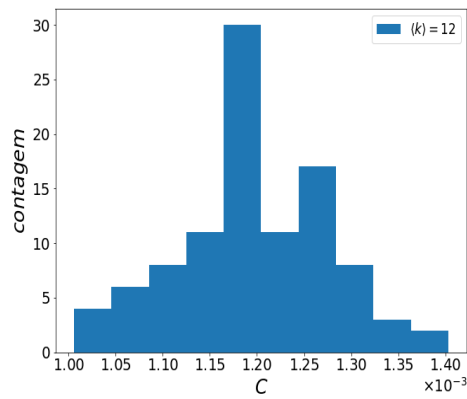
3.10(a):



3.10(b):



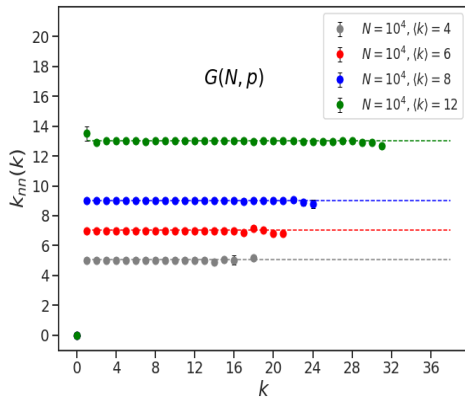
3.10(c):



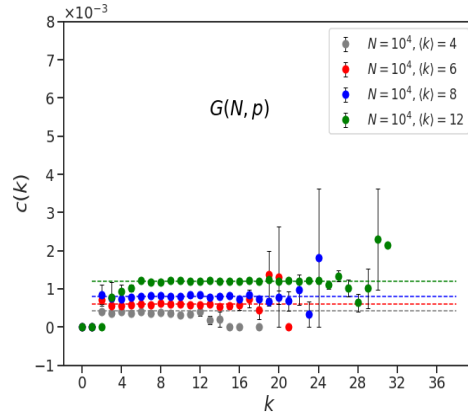
3.10(d):

Figura 3.10: Cada figura mostra a distribuição de frequências para o C global de cada rede para os valores de $\langle k \rangle = 4, 6, 8, 12$. Para cada figura foram geradas 10^2 redes. Figura baseada no modelo $G(N, M)$

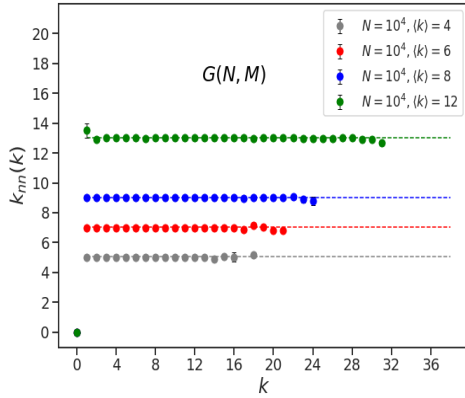
Para medir as correlações é possível calcular o $k_{nn}(k)$, $c(k)$, e o coeficiente de correlação de Pearson r para as rede de ER. A figura 3.11 trata os dois modelos até agora apresentados, $G(N, p)$ e $G(N, M)$. É possível ver que para todas as redes com diferentes valores de $\langle k \rangle$, tanto o $k_{nn}(k)$ quanto o $c(k)$ não dependem de k , isto é, não possuem correlação de graus.



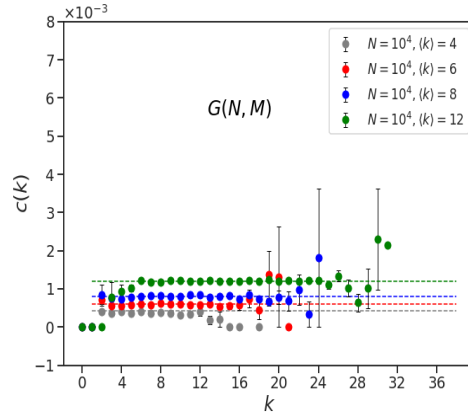
3.11(a):



3.11(b):



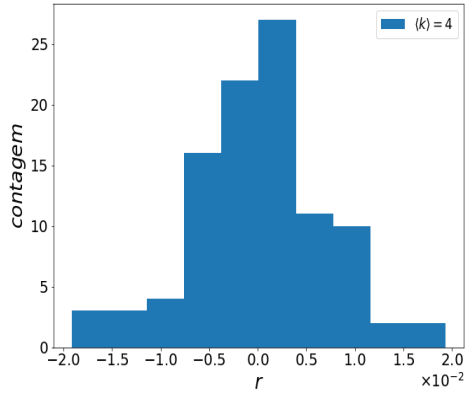
3.11(c):



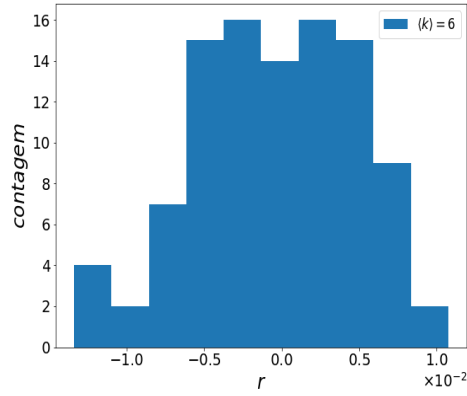
3.11(d):

Figura 3.11: As figuras acima mostram o $k_{nn}(k)$ e o $c(k)$ médio para a rede de ER utilizando tanto o modelo $G(N, p)$, quanto o modelo $G(N, M)$. As figuras a e b representam o modelo $G(N, p)$, enquanto c e d, o modelo $G(N, M)$. Para os dois modelos foram geradas rede com valores médio $\langle k \rangle = 4, 6, 8, 12$ e tamanho $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. Para cada modelo foi feito uma média sobre 10^2 redes.

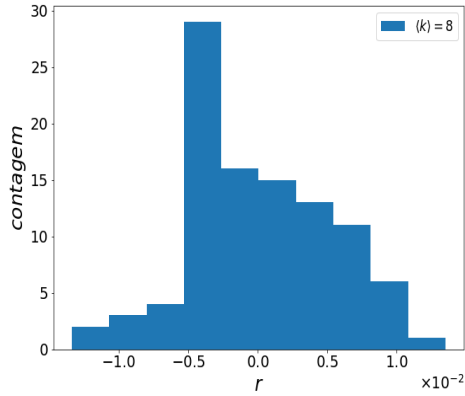
Uma visão global sobre as correlações pode ser observada através do coeficiente de correlação de Pearson r , definido na equação 2-14, como é mostrado na figura 3.12. Nesse caso, é possível observar que para as 10^2 redes construídas, a distribuição de r gira em torno de zero, mostrando que as redes são descorrelacionadas.



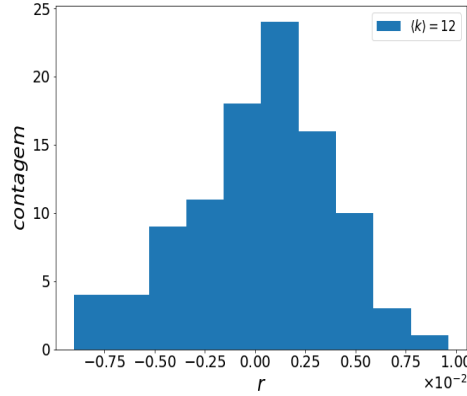
3.12(a):



3.12(b):



3.12(c):



3.12(d):

Figura 3.12: Cada figura mostra a distribuição para o coeficiente de correlação de Pearson r de cada rede. Com valores de $\langle k \rangle = 4, 6, 8, 12$. Para cada figura foram geradas 10^2 redes. Figura baseada no modelo $G(N, M)$

Além dos modelos citados acima, podemos construir redes de ER a partir do modelo de configuração apresentado na seção 3.1. Essa forma de abordar as redes de ER será apresentada na próxima seção 3.3.

3.3

Gerando redes com distribuição de Poisson utilizando o modelo de configuração

Outra forma de produzir redes de ER é utilizando o modelo de configuração. Como abordado na seção 3.1, o modelo de configuração consiste em construir redes com sequência de graus prescrita anteriormente. Para o caso da rede de ER, essa sequência terá tamanho N e seus graus serão distribuídos de acordo a distribuição de Poisson $P(k)$ onde $k \in [k_{min}, k_{max}]$ Eq. 3-8. Estabe-

lecida a sequência de graus, é criada uma lista de adjacência onde as ligações em cada sítio da rede são colocados de forma uniformemente aleatória. Auto conexões e conexões múltiplas são evitadas. Durante o processo de criação da rede, caso sobre uma meia aresta no final, descartamos a rede e criamos outra, uma vez que para os casos considerados poucos casos ocorrem.

3.3.1

Distribuição de graus

Uma vez que a rede é criada é possível medir a distribuição de graus calculada através da equação 2-5. Na figura 3.13 foi estabelecido $k_{min} = 0$ e $k_{max} = 35$. Além disso, o número de vértices foi definido como $N = 10^4$ e o grau médio como $\langle k \rangle = 6$.

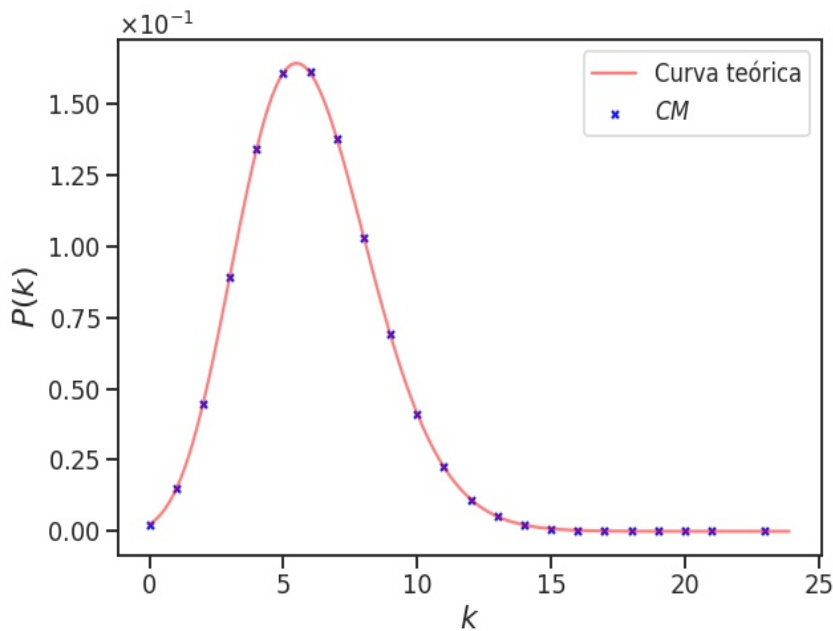


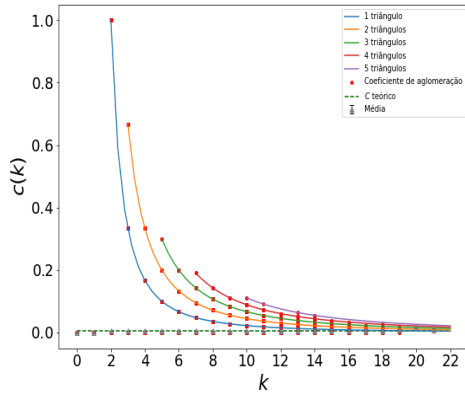
Figura 3.13: Distribuição de graus de uma rede de ER utilizando o modelo de configuração. $N = 10^4$, $\langle k \rangle = 6$. A figura exhibe a média sobre 10^2 redes. A curva vermelha é a mesma distribuição de Poisson da figura 3.7.

3.3.2

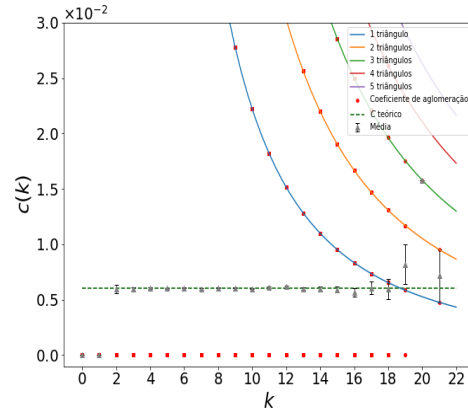
Coeficiente de aglomeração

A título de comparação, as figuras 3.14 mostram a distribuição do coeficiente de aglomeração local calculado via equação 2-15 além de sua média. Na figura 3.14(a) possui tamanho $N = 10^3$ e grau médio $\langle k \rangle = 6$. Já na figura 3.14(c), $N = 10^4$ com grau médio também $\langle k \rangle = 6$. As figura (b) e (d) são ampliações de (a) e (c). Ao utilizar o modelo de configuração, vemos maior

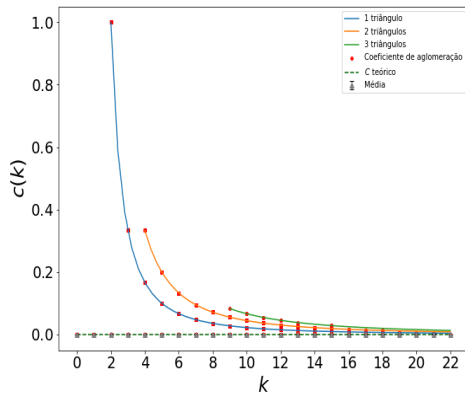
formação de triângulos quando $N = 10^3$, chegando a ter formação de cinco triângulos para vértices que possuem grau 10, por exemplo.



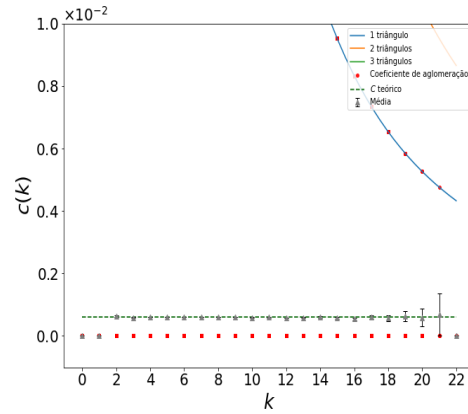
3.14(a):



3.14(b):



3.14(c):



3.14(d):

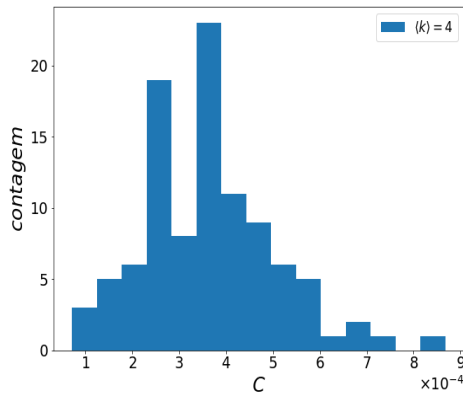
Figura 3.14: Distribuição de aglomeração para rede de ER utilizando modelo de configuração. Nas duas figuras superiores (a) e (b), $N = 10^3$ e $\langle k \rangle = 6$. Para as figuras (c) e (d), $N = 10^4$ e $\langle k \rangle = 6$. As duas figuras a direita (b) e (d), são ampliações das figuras (a) e (c). 10^2 redes foram geradas para as distribuições acima.

Como previsto $c(k)$ diminui à medida que k cresce, porém é visível que em média a aglomeração não depende de k , indicando que as redes são descorrelacionadas [38]. Além disso, é mostrado o valor teórico C calculado através da equação 3-9.

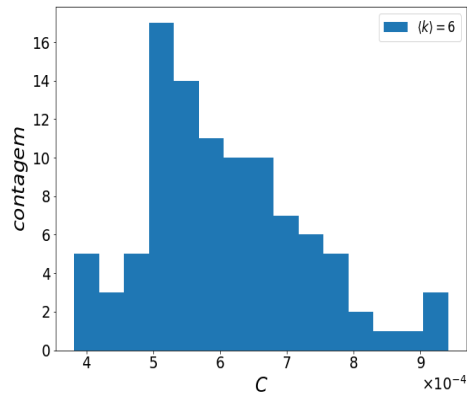
3.3.3

Distribuições de outras métricas da rede ER

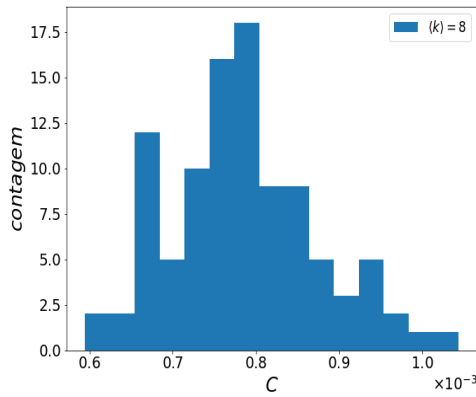
Nas figuras 3.15 temos a distribuição do coeficiente de aglomeração global para todas as cem redes geradas. Para cada medida $\langle k \rangle$ é possível calcular o C global através da equação 3-9 e encontrar valores teóricos de $C_4 = 4 \times 10^{-4}$, $C_6 = 6 \times 10^{-4}$, $C_8 = 8 \times 10^{-4}$, $C_{12} = 12 \times 10^{-4}$. Os índices nos valores de C indicam o grau médio, isto é, $\langle k \rangle = 4, 6, 8, 12$. Ao observar a figura 3.15 pode-se notar que os histogramas flutuam em torno desses valores teóricos



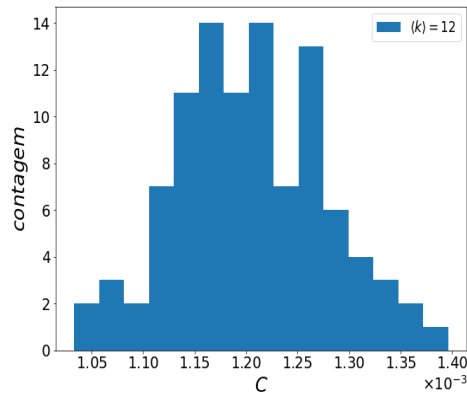
3.15(a):



3.15(b):



3.15(c):

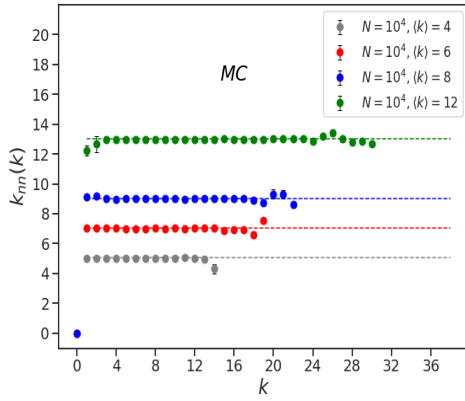


3.15(d):

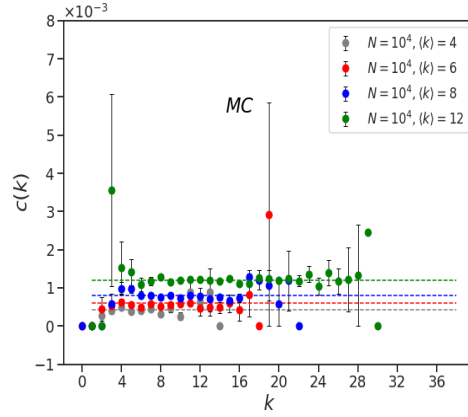
Figura 3.15: Distribuição de frequências para o C global do modelo de configuração para os valores de $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. Para cada figura foram geradas 10^2 redes.

Como anteriormente, utilizamos três métricas para indicar correlações nas redes criadas. O $k_{nn}(k)$, o coeficiente de correlação de Pearson r e o coeficiente de aglomeração. Na figura 3.16(a) é possível observar o gráfico

de $k_{nn}(k)$ e o coeficiente de aglomeração médio $c(k)$ para $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. Como para os outros modelos construídos, não foi possível identificar relação de k_{nn} e c com k , isto é, as redes também são descorrelacionadas.



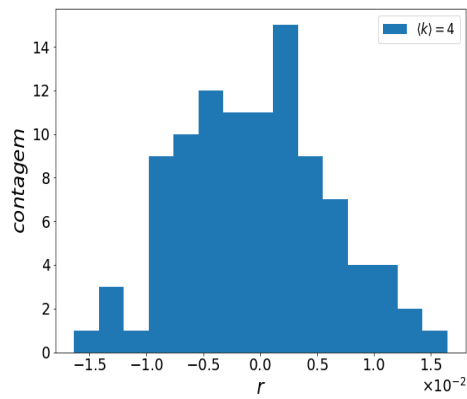
3.16(a):



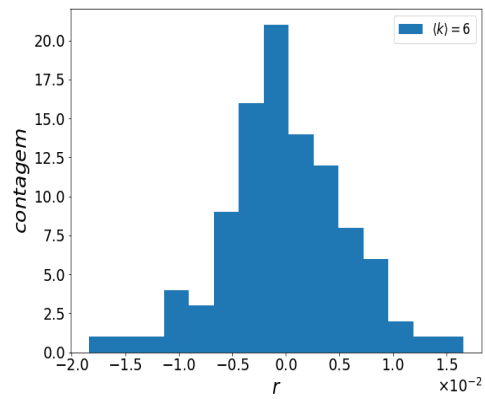
3.16(b):

Figura 3.16: As figuras acima mostram o $k_{nn}(k)$ e o $c(k)$ para a rede de ER utilizando o MC. As rede foram geradas com valores de $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18. Média sobre 10^2 redes.

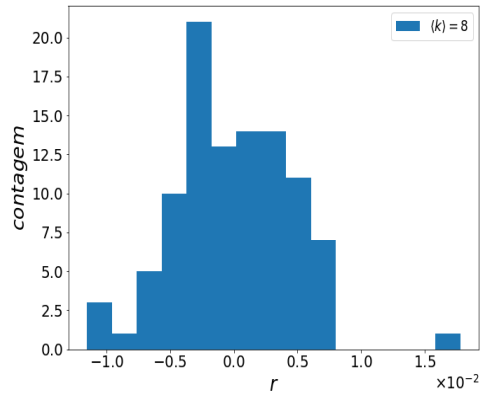
A figura 3.17 apresenta o coeficiente de correlação de Pearson r para as redes citadas acima. Todas as correlações flutuam em um intervalo muito pequeno em torno de $r = 0$, confirmando o que já constatamos na figura 3.16 que a correlação é muito fraca.



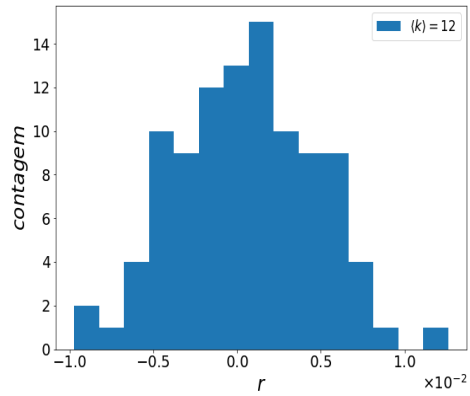
3.17(a):



3.17(b):



3.17(c):



3.17(d):

Figura 3.17: Cada figura mostra a distribuição para o coeficiente de correlação de Pearson r para valores de $\langle k \rangle = 4, 6, 8, 12$ e $N = 10^4$. Para cada figura 10^2 redes foram geradas.

3.4

Redes livres de escala

Outro aspecto importante na teoria de redes complexas emerge do fato que diversos sistemas reais possuem distribuição de graus descrita por uma *lei de potência*, isto é,

$$P(k) \sim k^{-\gamma}, \quad k \in [k_{\min}, k_{\max}], \quad (3-10)$$

onde k_{\min} é o grau mínimo da rede e k_{\max} o grau máximo.

Na tentativa de explicar a topologia da internet (*world wide web* - *www*), Waxman [39], em 1988, propôs um modelo teórico em que um número N de vértices era disposto aleatoriamente em um plano bidimensional e conectados com uma probabilidade que decai com a distância euclidiana entre eles. No entanto esse modelo acaba gerando uma distribuição das conexões similares às redes de Erdős-Rényi. Já em 1999, Barabási e Albert [3], além de propor um modelo que era descrito por uma distribuição em lei de potência, observaram que outros sistemas seguem essa topologia, como a rede formada pelos padrões de citações em publicações científicas ou a rede gerada pela colaboração de atores no cinema.

O modelo de Barabási e Albert (BA) difere das redes aleatórias de Erdős-Rényi (ER) e das redes de pequeno mundo de Watts e Strogatz (WS), pois trata-se de um modelo de crescimento de redes. Nos modelos de crescimento de redes, cada novo vértice adicionado conecta-se (ou não) com os vértices previamente adicionados seguindo alguma lei de formação. No modelo de BA, cada novo vértice adicionado à rede se conecta com um vértice pré-existente, com probabilidade proporcional à sua conectividade.

Após o modelo de BA ser apresentado [3], diversos outros modelos surgiram na tentativa de explicar essas redes que decaem como uma lei de potência [40, 41, 42].

Para gerar redes em lei de potência, utilizaremos o MC. Consequentemente se faz necessário construir uma sequência de graus onde a distribuição estará de acordo com a equação 3-10. Nesse tipo de rede com distribuição de graus descrita por lei de potência, os momentos dependem fortemente de γ . Na ampliação do caso contínuo da distribuição $P(k)$, o m -ésimo momento pode ser calculado como

$$\langle k^m \rangle \simeq \frac{\int_{k_{\min}}^{k_{\max}} k^m / k^\gamma dk}{\int_{k_{\min}}^{k_{\max}} 1/k^\gamma dk}. \quad (3-11)$$

Em particular, se $\gamma \leq 3$, o segundo momento diverge, $\langle k^2 \rangle \rightarrow \infty$, no limite do tamanho da rede infinito. Para esse regime, as redes em lei potência são especificamente chamadas de *livres de escala*. Para o caso em que $\gamma \leq 2$, o primeiro momento também diverge, isto é, $\langle k \rangle \rightarrow \infty$ [11].

Se queremos controlar por exemplo o grau médio (primeiro momento da distribuição de graus), podemos fixar alguns parâmetros da distribuição com valores pre especificados e encontrar algum outro parâmetro definido como variável, resolvendo a equação 3-11, eventualmente por métodos numéricos. Por exemplo, se γ é a variável a ser descoberta, fixando os parâmetros $\langle k \rangle = 6$, $k_{min} = 4$ e $k_{max} = 2000$, encontramos um valor de $\gamma \sim 3,45$. Essa seria uma forma de fixar o primeiro momento da rede.

3.5

Gerando redes livres de escala com o modelo de configuração

Uma das propriedades fundamentais de redes complexas é sua sequência de graus. Para utilizarmos o modelo de configuração o primeiro passo é criar uma sequência de graus a partir de uma probabilidade $P(k) \sim k^{-\gamma}$.

Isto posto, sorteamos aleatoriamente um valor k_i no intervalo $[k_{min}, k_{max}]$ e cada k_i sorteado é colocado na rede com probabilidade $P(k_i)$ calculada via equação 3-10. Esse processo é feito obedecendo duas condições: a primeira é que a $\sum k_i$ necessariamente precisa ser par, pois para formar uma conexão pelo modelo de configuração, duas meias arestas são selecionadas por vez e a segunda condição é que o número de vezes que k_i entra na sequência é igual ao número de vértices da rede N .

Para obter uma visão sobre as propriedades da rede, se faz necessário conectar os graus da sequência utilizando o modelo de configuração descrito na seção 3.1.

3.5.1

Distribuição de graus

Através da sequência de graus é possível visualizar a distribuição $P(k)$ de graus da rede. A figura 3.18 mostra a distribuição de graus para uma rede com $P(k) \sim k^{-\gamma}$.

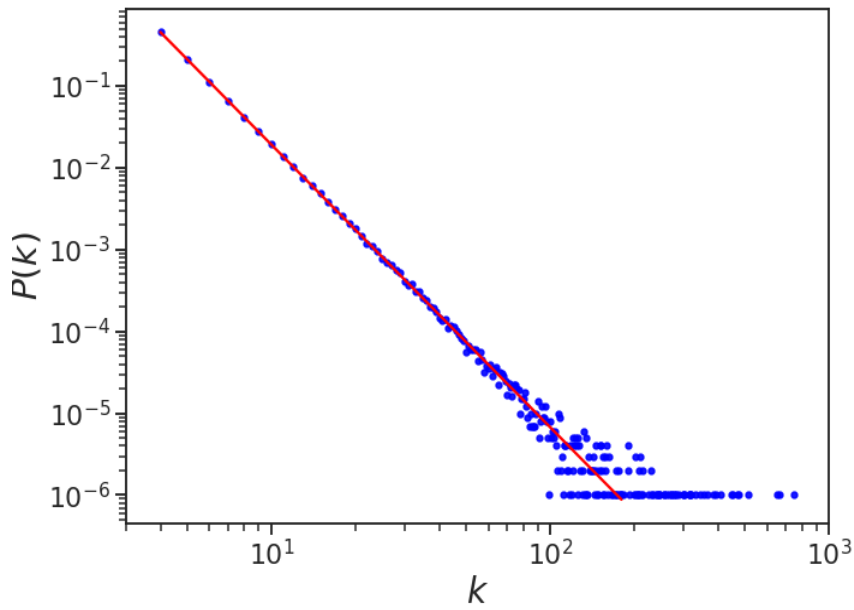
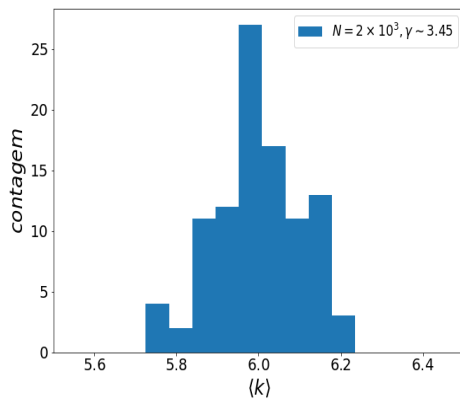
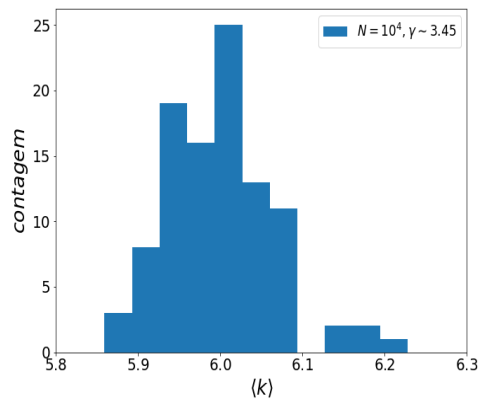


Figura 3.18: Distribuição de graus pelo sorteio aleatório de k . Os parâmetros usados foram $N = 10^4$, $k_{min} = 4$, $k_{max} = 2000$, $\gamma \sim 3,45$ e $\langle k \rangle = 6$. A figura mostra a média sobre 100 realizações. A curva em vermelho é a distribuição $P(k) \sim k^{-\gamma}$.

As Figs. 3.19 e 3.20 mostram as flutuações do grau médio, quando os parâmetros foram escolhidos de modo a ter $\langle k \rangle = 6$.

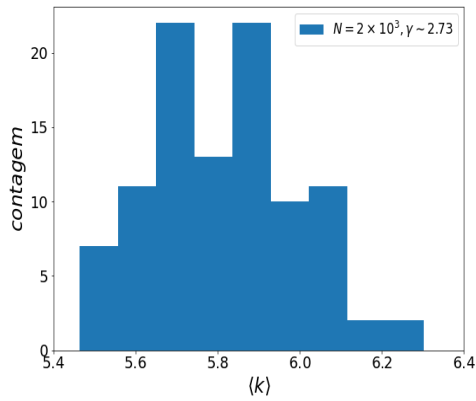


3.19(a):

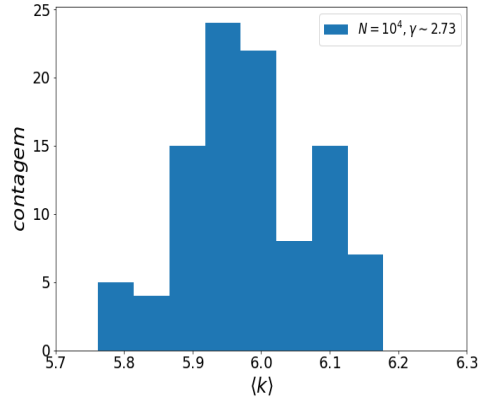


3.19(b):

Figura 3.19: Histogramas do grau médio para simulação de 10^2 redes. Em (a), $N = 2 \times 10^3$ e em (b) $N = 10^4$. Para ambas as figuras foram utilizados $k_{min} = 3$, $k_{max} = 2 \times 10^3$ e $\gamma = 3,45$, desta forma o grau médio flutua em torno de 6.



3.20(a):



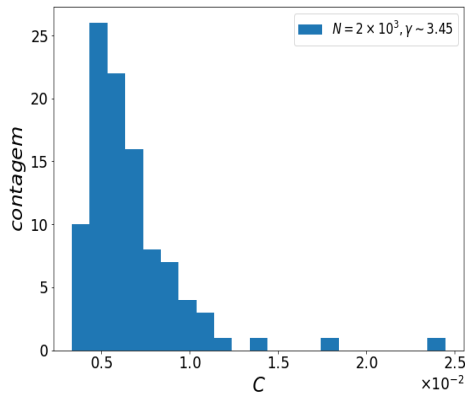
3.20(b):

Figura 3.20: Histogramas do grau médio para simulação de 10^2 redes. Em (a), $N = 2 \times 10^3$ e em (b) $N = 10^4$. Para ambas as figuras foram utilizados $k_{min} = 4$, $k_{max} = 2 \times 10^3$ e $\gamma = 2,73$, desta forma o grau médio flutua em torno de 6.

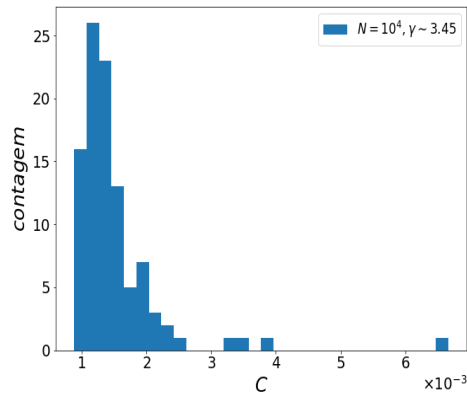
3.5.2

Flutuações no coeficiente de aglomeração

Para as realizações da figura 3.18, medimos o grau médio, obtendo um desvio relativo de 10% em torno do valor desejado $\langle k \rangle = 6$. Para cada realização construímos a rede correspondente usando o MC. Para o conjunto das redes geradas, a figura 3.21 mostra a distribuição do coeficiente de aglomeração global C . Como é possível observar, os valores de C para 3.21(a) variam em um intervalo de $[0.4 \times 10^{-2}, 2.5 \times 10^{-2}]$, enquanto 3.21(b), C flutua entre $[1 \times 10^{-3}, 7 \times 10^{-3}]$.



3.21(a):

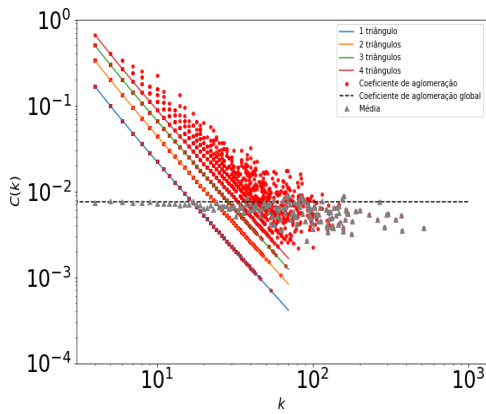


3.21(b):

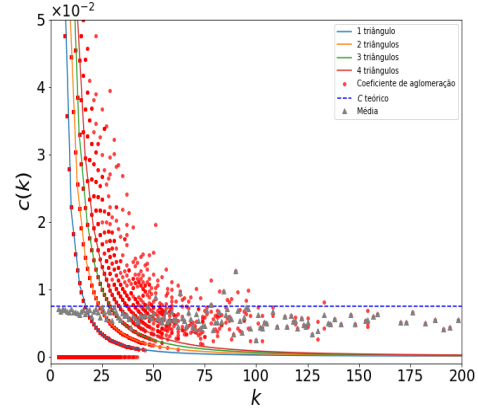
Figura 3.21: Histogramas do coeficiente de aglomeração global para simulação de 10^2 redes. Em (a), $N = 2 \times 10^3$ e em (b) $N = 10^4$. Para ambas as figuras foram utilizados $k_{min} = 4$, $k_{max} = 2 \times 10^3$ e $\gamma = 3,45$, desta forma o grau médio flutua em torno de 6.

A figura 3.22 representa o coeficiente de aglomeração local para cada grau k das redes em lei de potência construídas utilizando o modelo de configuração. O símbolos vermelhos são os c_i , coeficiente de aglomeração local para cada sítio. Os símbolos cinza resultam de agrupar o coeficiente c_i por grau, $c(k)$, enquanto a linha horizontal é a média C , dada pela equação 2-17. As linhas inclinadas, correspondem a $2t/(k(k-1))$, onde t é número de triângulos. O número de triângulos por grau k está relacionado com o termo e_i na equação 2-15.

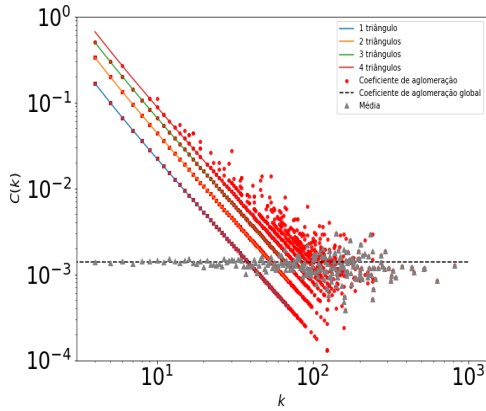
Em cinza temos a média de $c(k)$. O pontos em vermelho representam a sobreposição de 10^2 redes. As figuras 3.22(a) e 3.22(b) são representações das mesmas redes com parâmetros $N = 2 \times 10^3$, $\gamma \sim 3,45$, $k_{min} = 4$ e $k_{max} = 2 \times 10^3$. Nas figuras 3.22(c) e 3.22(d), $N = 10^4$, $\gamma \sim 3,45$, $k_{min} = 4$ e $k_{max} = 2000$.



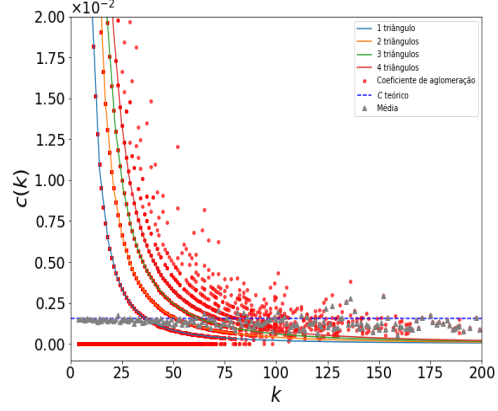
3.22(a):



3.22(b):



3.22(c):



3.22(d):

Figura 3.22: Coeficiente de aglomeração local para redes em lei de potência construídas utilizando o modelo de configuração. Os símbolos vermelhos são os c_i , coeficiente de aglomeração local para cada sítio, em 100 realizações (redes). Os símbolos cinza resultam de agrupar o coeficiente c_i por grau (a linha horizontal é a média, C dada pela equação 2-17). As linhas inclinadas, correspondem a $2t/(k(k-1))$, onde t é o número de triângulos. O número de triângulos por grau k está relacionado com o termo e_i na equação 2-15. As duas figuras superiores (a) e (b) são representações em diferentes escalas, para redes com $N = 2 \times 10^3$, $\gamma \sim 3,45$, $k_{min} = 4$ e $k_{max} = 2 \times 10^3$. Para as figuras (c) e (d), $N = 10^4$, $\gamma \sim 3,45$ e $k_{min} = 4$, $k_{max} = 2000$.

Nas figura 3.22(a) é possível verificar um leve decaimento no valor médio de $c(k)$ à medida que k aumenta. Esse decaimento indica que existem correlações nas redes aqui construídas [43]. Isso ocorre pois o grau máximo k_{max} da rede tem a mesma ordem de grandeza de N , introduzindo assim correlações. Com o intuito de evidenciar as correlações nas redes, iremos discutir na seção

3.5.3 as correlações para diferentes valores de γ .

3.5.3

Distribuições de outras métricas

Primeiramente, vemos na figura 3.23, que, para k fixo, $k_{nn}(k)$ aumenta quando o expoente γ diminui, devido à maior probabilidade de nós muito conectados.

Além disso, fica visível que para os casos em que $\gamma \geq 3$ a curva é praticamente uma linha horizontal, indicando que $k_{nn}(k)$ depende fracamente de k . Estes resultados reproduzem o comportamento observado na referência [13], obtidos para $N = 10^5$. Entretanto quando a rede é livre de escala ($\gamma < 3$), com nodos muito conectados, o grau médio dos vizinhos diminui quando k cresce.

Em comparação com os resultados para redes descorrelacionadas (linhas tracejadas), observamos que no caso $\gamma = 3.5$, o resultado é bem próximo ao valor teórico, entretanto para os outros valores ocorre uma separação que aumenta quando γ diminui.

Veremos depois que estes efeitos podem ser reduzidos impondo restrições no tamanho da rede.

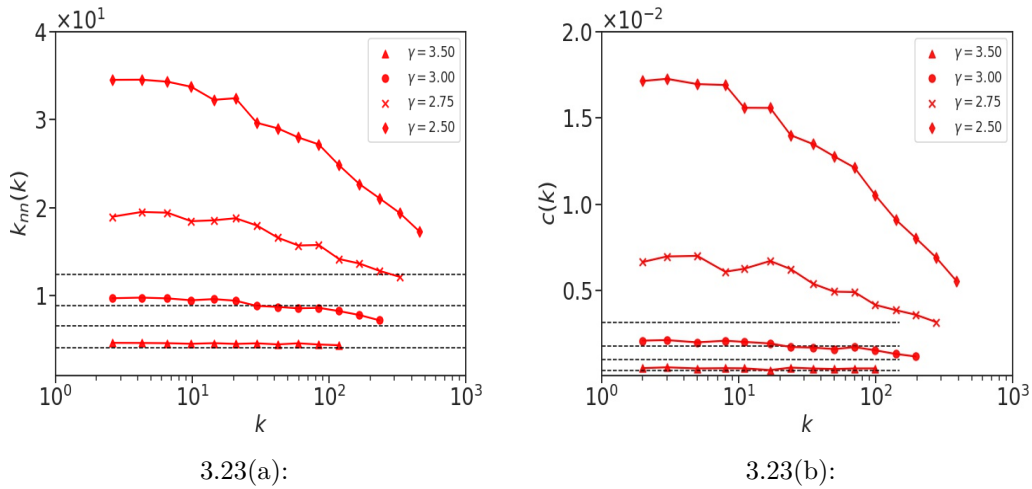
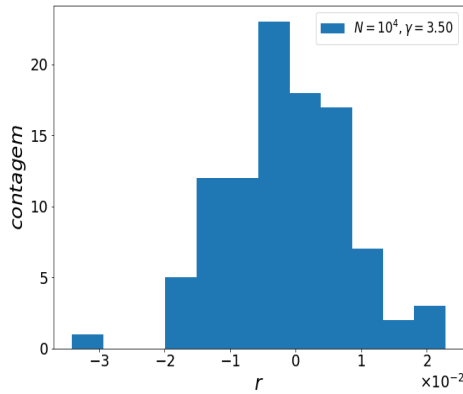


Figura 3.23: Grau médio dos vizinhos mais próximos dos vértices de grau k , $k_{nn}(k)$, (a) e coeficiente de aglomeração médio $c(k)$ (b). As curvas correspondem a diferentes valores de γ . As redes possuem tamanho $N = 10^4$. As figuras representam a média sobre 10^2 redes. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18.

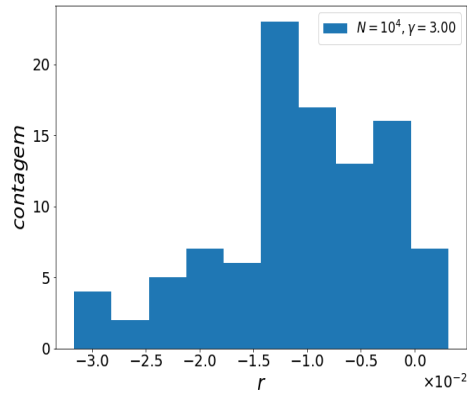
Para construir redes livres de escala não correlacionadas utilizando o modelo de configuração, é possível definir um corte estrutural para k_{max} ao criar redes

com distribuição $P(k) \sim k^{-\gamma}$. Nesse caso, para redes com tamanho N o grau máximo da rede é definido como $k_{max} = k_c(N) \sim N^{1/2}$, dessa forma o modelo de configuração cria redes não correlacionadas mesmo para $\gamma \leq 3$ [13, 44].

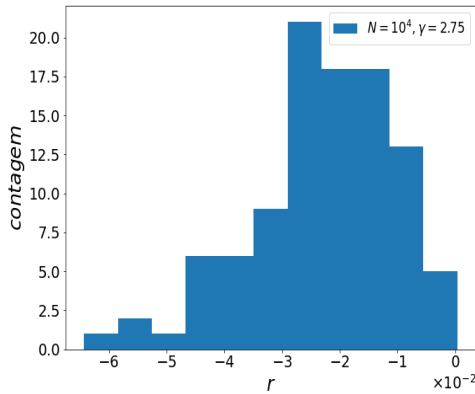
Outra forma de observar as correlações mostradas na figura 3.23 é através do coeficiente de correlação de Pearson 2-14.



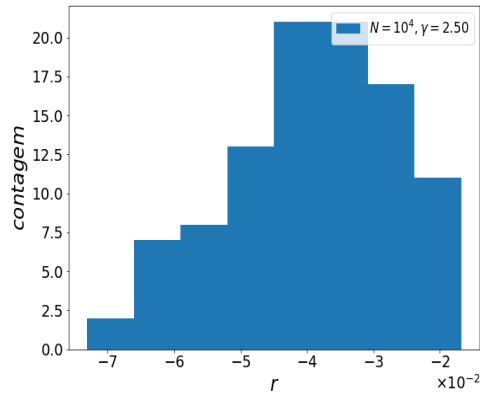
3.24(a):



3.24(b):



3.24(c):



3.24(d):

Figura 3.24: Coeficiente de correlação de Pearson r para redes em lei de potência. As figuras apresentam diferentes valores de γ . Em todos os casos $N = 10^4$ e foram geradas 10^2 redes.

Como discutido com relação à figura 3.23, à medida que γ diminui, as correlações são mais dissassortativas. Logo, o coeficiente de correlação de Pearson tende a flutuar com valores negativos. Na figura 3.23, a curva em que $\gamma = 3.0$ não é um bom indicativo de correlações, no entanto fica visível na figura 3.24 (b) que o coeficiente r flutua em um intervalo negativo, apontando correlação dissassortativa.

A figura 3.25 mostra que mesmo para as redes livres de escala $2 < \gamma \leq 3$, quando impomos um corte estrutural em $k_{max} = k_c$, como proposto anteriormente [13, 44] não são introduzidas correlações nas redes.

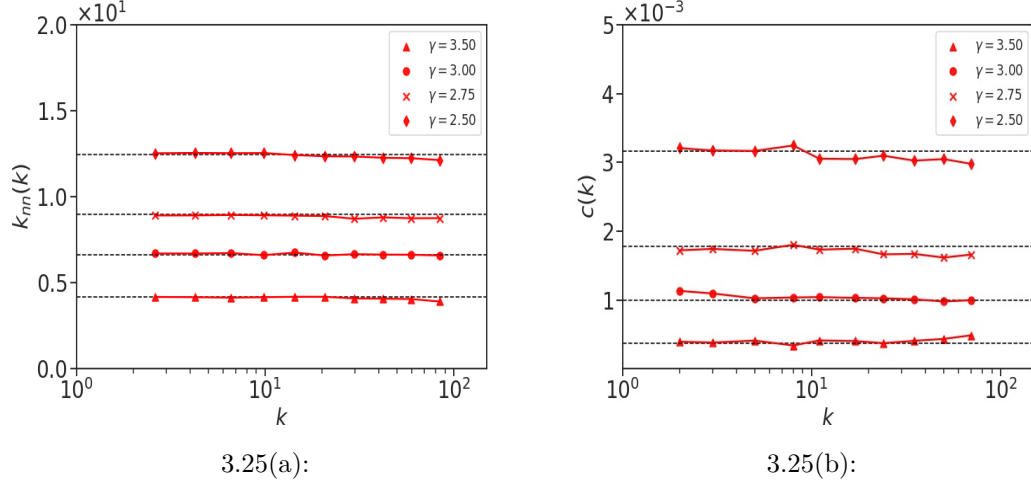
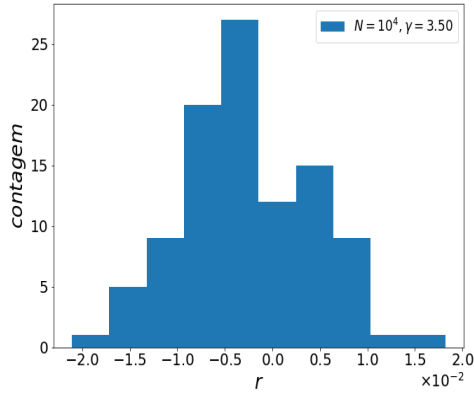


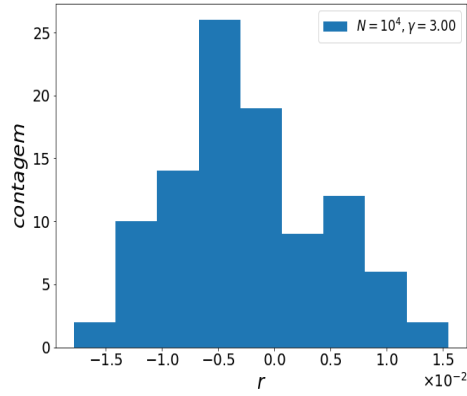
Figura 3.25: Grau médio dos vizinhos mais próximos dos vértices com grau k , $k_{nn}(k)$ (a) e coeficiente de aglomeração médio $c(k)$ (b). As curvas correspondem a diferentes valores de γ . As redes possuem tamanho $N = 10^4$ com $k_{min} = 2$, $k_c = \sqrt{10^4}$. As figuras representam a média sobre 10^2 redes. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18.

A figura 3.25 apresenta curvas em que $k_{nn}(k)$ e $c(k)$ médio, não dependem de k . Este resultado foi observado na referência [13], para $N = 10^5$.

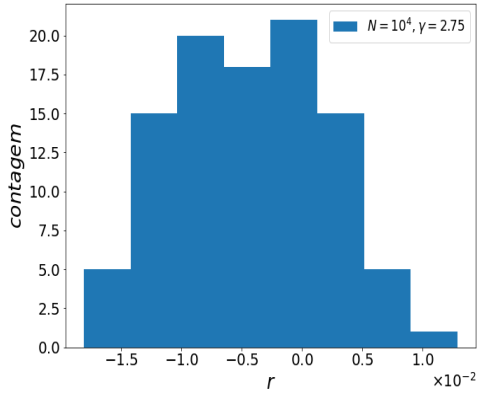
Uma figura semelhante a 3.24 pode ser feita para as redes utilizando o corte estrutural $k_c = \sqrt{N}$, assim como em 3.26.



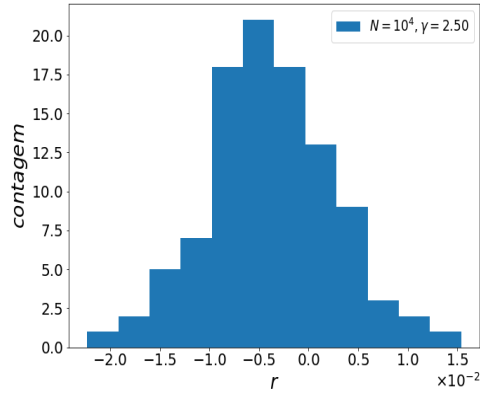
3.26(a):



3.26(b):



3.26(c):



3.26(d):

Figura 3.26: Coeficiente de correlação de Pearson r pra rede em lei de potência. As figuras apresentam diferentes valores de γ e $N = 10^4$. Nesse caso, foi utilizado o corte estrutural no grau máximo da rede, isto é, $k_c = \sqrt{10^4}$. Histogramas gerados por 10^2 redes.

Nesse caso, as figuras em 3.26 indicam valores de negativos de r bem maiores que na figura 3.24. Como é possível observar, o coeficiente r se restringe a quantidades num intervalo mais próximo de $r = 0$.

A figura 3.27 mostra como o grau máximo afeta o primeiro momento $\langle k \rangle$ da rede, sendo a discrepância 10% no pior caso ($\gamma = 2, 5$).

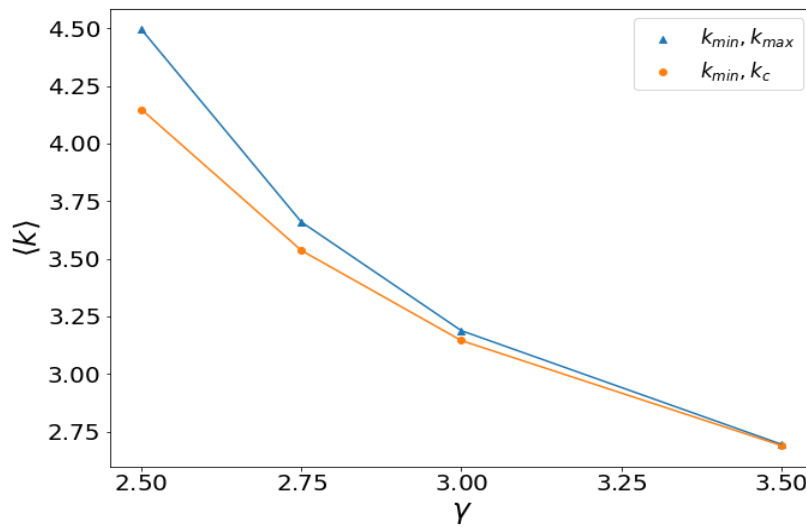


Figura 3.27: Grau médio global $\langle k \rangle$ em função de γ . Na curva em azul foi utilizado $k_{min} = 2$ e $k_{max} = N$. Para a curva em laranja $k_{min} = 2$ e $k_c = \sqrt{N}$. Para ambas as curvas $N = 10^4$

É possível verificar em 3.27 que correlações na rede afetam o grau médio $\langle k \rangle$ e à medida que $\gamma > 3$, é indiferente utilizar o modelo de configuração tradicional ou o modelo de configuração com o corte estrutural k_c .

4

Controlando o grau médio

Sabendo que a evolução de uma dinâmica em redes pode depender do seu grau médio, e que este ainda pode flutuar para uma dada realização da rede, propomos um método que permite fixar $\langle k \rangle$.

Na seção 4.1, mostramos as limitações na escolha dos parâmetros de uma rede quando se quer selecionar o grau médio. Nas subseções 4.1.1 e 4.1.2 propomos duas maneiras diferentes de controlar o grau médio para uma rede com $P(k)$ em lei de potência, alterando essa distribuição para os valores menores de k e preservando a cauda. Definimos um procedimento para eliminar as flutuações no grau médio, na subseção 4.1.3. Também realizamos um estudo estatístico de diversas métricas calculadas sobre redes geradas através do modelo de configuração, em base às distribuições de graus propostas.

4.1

Métodos para controlar o grau médio

Para ajustar o grau médio da rede pode-se alterar algum dos seguintes parâmetros: o menor grau k_{min} , o maior grau k_{max} e o expoente γ da distribuição. Entretanto, como mostrado na seção 3.5.1, o valor resultante do grau médio pode flutuar bastante (ver figuras 3.19 e 3.20). Ainda, se fixamos k_{min} , k_{max} e mudamos γ , surge uma grande variação nos valores de $\langle k \rangle$, como mostrado na figura 4.1.

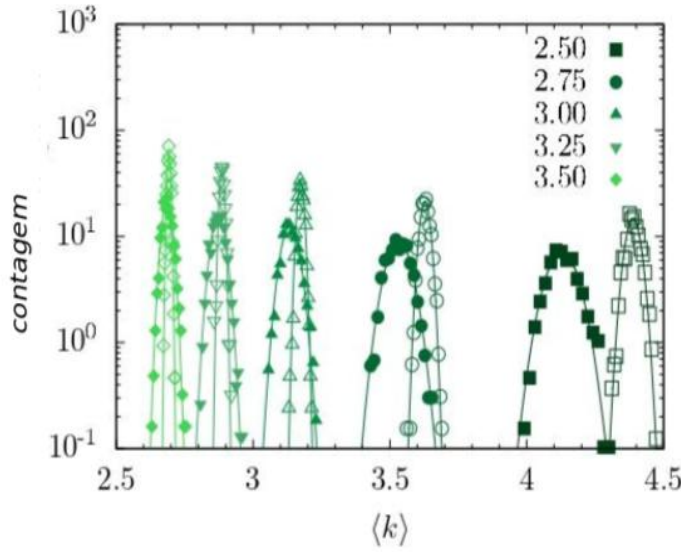


Figura 4.1: Distribuição do grau médio, fixando $k_{min} = 2$, $k_{max} = \sqrt{N}$ e mudando o expoente γ . Os símbolos preenchidos representam $N = 10^4$ enquanto os vazios $N = 10^5$. A figura representa a média sobre 10^3 realizações.

Para não introduzir correlações na rede, o modelo de configuração descorrelacionado [13] prevê que o grau máximo da rede deve ser menor que um limiar $k_c \sim N^{1/2}$. Como o grau máximo da rede é estabelecido por k_c , só restam dois valores a serem alterados, isto é, k_{min} e γ .

Fixando γ , ao avaliar $\langle k \rangle$ usando a equação 2-6 e uma distribuição $P(k) \sim k^{-\gamma}$, nos deparamos com uma grande limitação para os valores que $\langle k \rangle$ assume. Por exemplo, nas figuras em 4.2 (supondo k_{min} e k_{max} reais), cada faixa representa um intervalo de valores para $\langle k \rangle$ iniciando em 3. A primeira faixa em rosa representa o intervalo que começa em $\langle k \rangle = 3$ e termina em $\langle k \rangle = 4$. As curvas em preto (delimitando as faixas) são os valores inteiros para $\langle k \rangle$, isto é, a primeira curva corresponde a $\langle k \rangle = 3$, a segunda $\langle k \rangle = 4$, seguindo até a última com $\langle k \rangle = 10$. Além disso, cada gráfico representa um valor para γ , isto é, a figura 4.2(a) $\gamma = 2.50$, 4.2(b) $\gamma = 2.75$, 4.2(c) $\gamma = 3.00$ e 4.2(d) $\gamma = 3.50$.

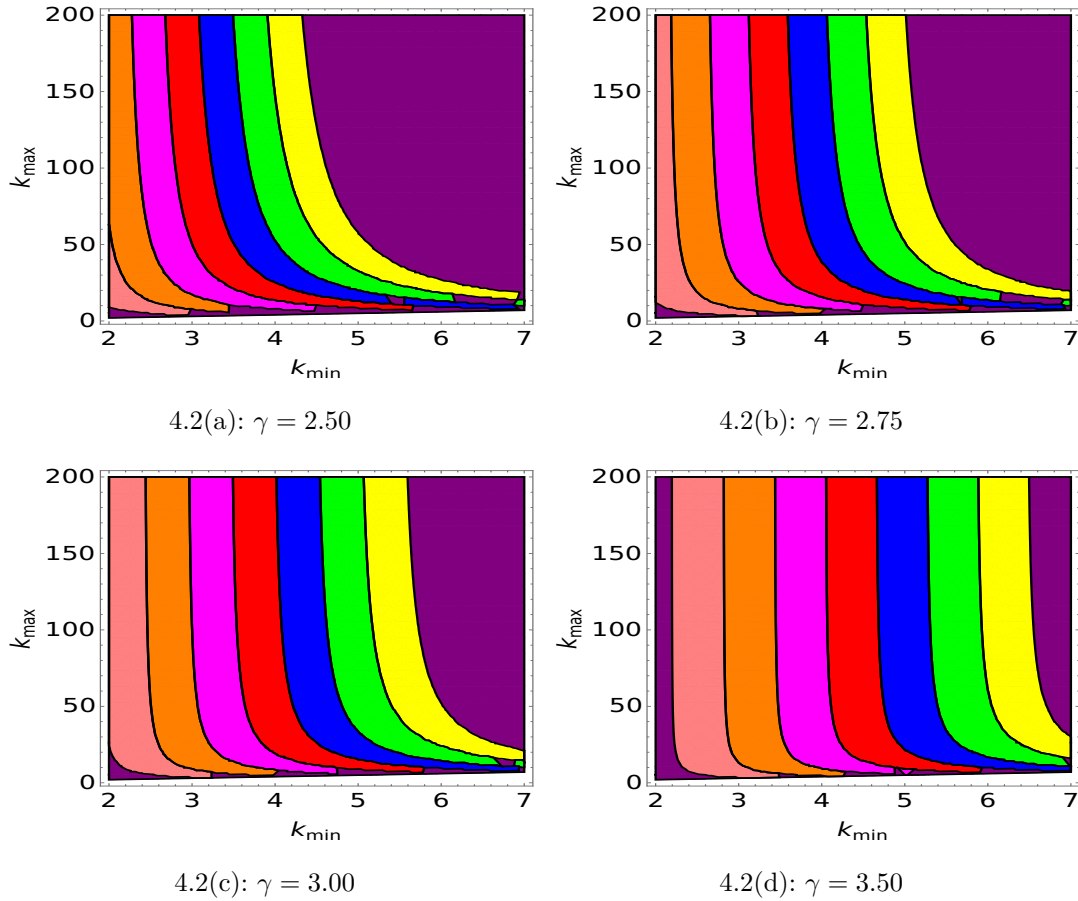


Figura 4.2: Limitação para fixar $\langle k \rangle$ no plano k_{min} - k_{max} , considerando faixas para $\langle k \rangle$. As linhas pretas correspondem aos valores $\langle k \rangle = 3, 4, \dots, 10$, para diferentes valores de γ . A faixa roxa foi usada somente como pano de fundo para figura.

Considerando a figura 4.2(c), para criar uma rede com $\langle k \rangle = 4$ e $\gamma = 3.00$ os valores extremos de k seriam $k_{min} = 3$ e $k_{max} \simeq 20$. Para $\gamma = 3.50$, na figura 4.2(d), seria impossível utilizar $k_{min} = 2$ para obter $\langle k \rangle = 4$.

Portanto, apresentaremos dois métodos com o intuito de contornar as limitações exibidas acima, que chamaremos de *método 1* e *método 2*.

4.1.1

Método 1

O procedimento consiste em controlar o grau médio da rede alterando a quantidade de sítios com grau mínimo k_{min} . Nesse caso, foi necessário introduzir um novo parâmetro $a \in (0, 1]$ na distribuição, da seguinte forma

$$P(k) = \begin{cases} \frac{Aa}{k_{min}^\gamma}, & k = k_{min}, \\ \frac{A}{k^\gamma}, & k_{min} < k \leq k_{max}, \end{cases} \quad (4-1)$$

onde A é nova a constante de normalização. A figura 4.3 mostra a distribuição $P(k)$ inserindo o parâmetro a .

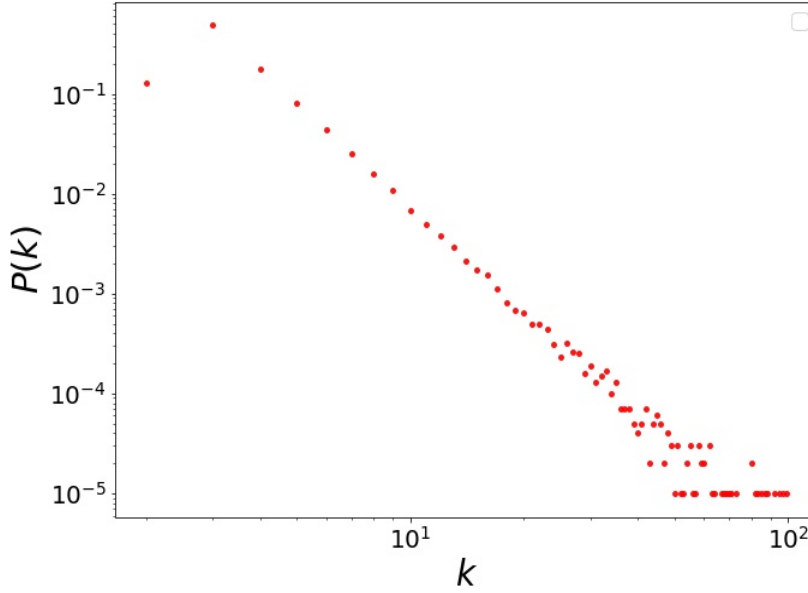


Figura 4.3: Distribuição de graus com $\gamma = 3.50$ e $\langle k \rangle = 4$, $k_{min} = 2$, $k_{max} = k_c = \sqrt{N}$, $N = 10^4$ e $a = 0.0604$. A figura foi gerada pela média sobre 10 redes.

Da maneira que $P(k)$ é definido na equação 4-1, podemos rescrever o grau médio da rede como,

$$\langle k \rangle = \frac{\frac{a}{k_{min}^{\gamma-1}} + \sum_{k=(k_{min}+1)}^{k_{max}} \frac{1}{k^{\gamma-1}}}{\frac{a}{k_{min}^{\gamma}} + \sum_{k=(k_{min}+1)}^{k_{max}} \frac{1}{k^{\gamma}}}. \quad (4-2)$$

Uma vez fixado os valores de k_{min} , k_{max} , $\langle k \rangle$ e γ , é possível isolar o valor a na equação 4-2, tal que,

$$a = \frac{k_{min}^{\gamma}}{\langle k \rangle - k_{min}} \sum_{k=k_{min}+1}^{k_{max}} \frac{k - \langle k \rangle}{k^{\gamma}}. \quad (4-3)$$

Escolhendo $\langle k \rangle = 4$, $k_{min} = 2$ e $k_{max} = k_c = N^{1/2} = \sqrt{10^4} = 100$, obtém-se a tabela 4.1.

γ	a
2.55	1.000
2.75	0.597
3.00	0.308
3.5	0.0604

Tabela 4.1: Valores de a (método 1) para valores de γ .

A figura 4.4 apresenta $k_{nn}(k)$ e $c(k)$ para o **método 1**. É possível observar que tanto os valores de $k_{nn}(k)$ quanto os de $c(k)$ não dependem de k . Além disso, os valores obtidos estão de acordo com o previsto para o caso decorrelacionado (linhas tracejadas). Na figura 4.4(a) e 4.4(b) foram usados valores de $k_{min} = 2$, $N = 10^4$ para os valores de γ apresentados na tabela 4.1.

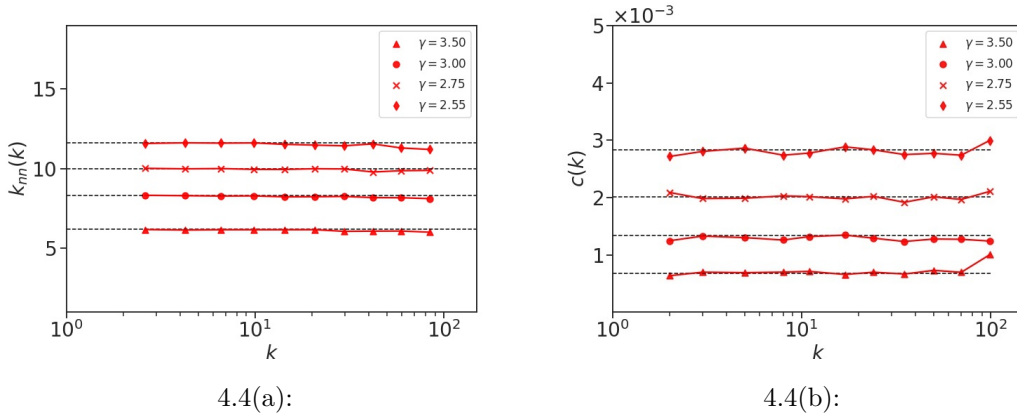
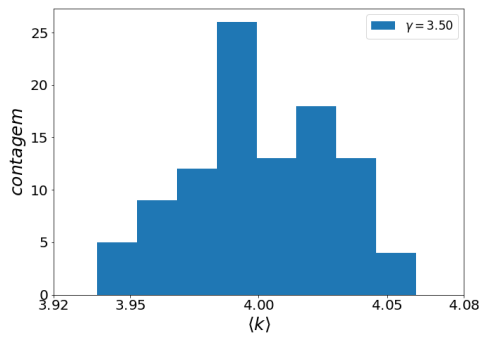
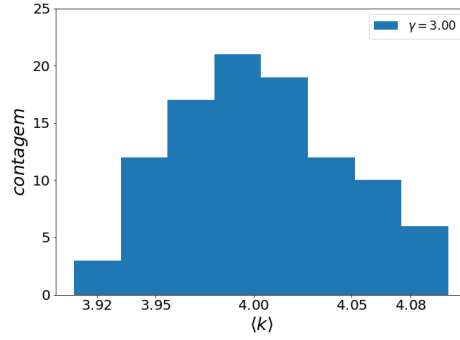


Figura 4.4: A figura (a) exibe $k_{nn}(k)$ enquanto a figura (b) $c(k)$. Resultados da média de 100 redes geradas usando o MC, a partir do método 1, com $\langle k \rangle = 4$, $k_{min} = 2$, $k_{max} = 100$ e $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18.

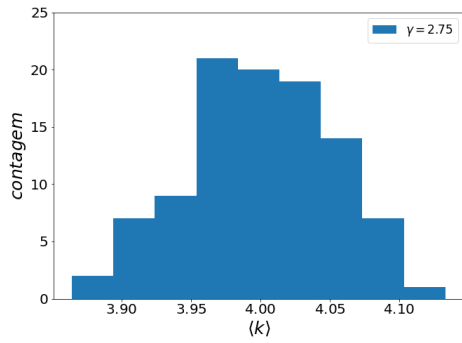
A figura 4.5 evidencia o fato de que, para a sequência de graus gerada com o sorteio através de $P(k)$, o grau médio resultante sofre flutuações em torno do valor desejado $\langle k \rangle = 4$, de até 2.5%, diminuindo com o aumento do expoente γ .



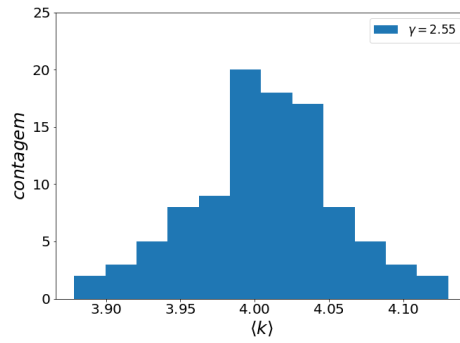
4.5(a):



4.5(b):



4.5(c):



4.5(d):

Figura 4.5: Histograma de $\langle k \rangle$ usando o método 1. Os valores de referência são $\langle k \rangle = 4$, $k_{min} = 2$, $k_{max} = 100$ e $N = 10^4$.

A figura 4.6 mostra $k_{nn}(k)$ e o coeficiente de aglomeração (média sobre 100 amostras), para $\langle k \rangle = 4, 6, 12$ e $\gamma = 2.5, 3.5$. Também neste caso a concordância com os cálculos sem correlações é boa.

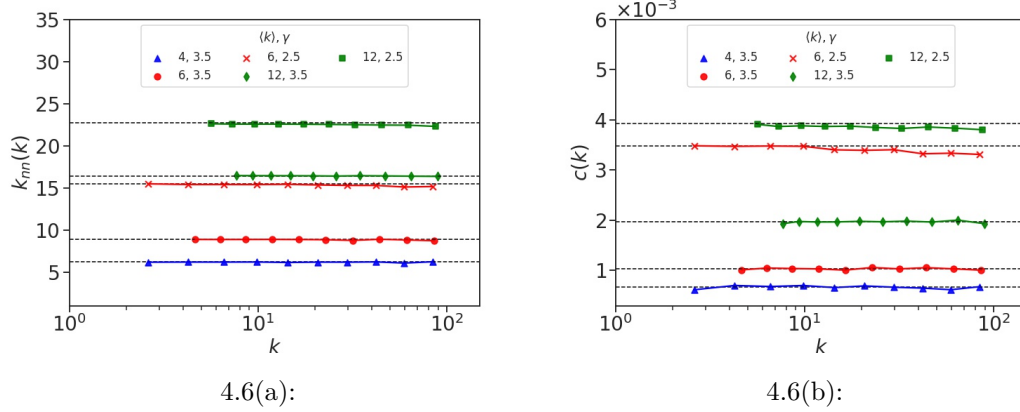


Figura 4.6: A figura (a) exibe valores para $k_{nn}(k)$ enquanto a figura (b) valores para $c(k)$ (média sobre 100 redes geradas usando o MC, a partir do método 1). As redes foram geradas com valores de $\langle k \rangle = 4, 6, 12$, $\gamma = 2.5, 3.5$, $k_{max} = 100$ e $N = 10^4$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18.

4.1.2 Método 2

Inserimos um platô inicial na distribuição. Para tal, a nova distribuição $P(k)$ foi definida mediante um parâmetro $t \geq 0$ da seguinte forma

$$P(k) = \begin{cases} \frac{A}{(k_{min}+t)^\gamma}, & k_{min} \leq k \leq \tilde{k} \\ \frac{A}{k^\gamma}, & \tilde{k} < k \leq k_{max}, \end{cases} \quad (4-4)$$

onde A é a nova constante de normalização e $\tilde{k} = k_{min} + \lfloor t \rfloor$, onde C fornece o comprimento do platô. A figura 4.7 apresenta a distribuição $P(k)$ (equação 4-4) para o método 2.

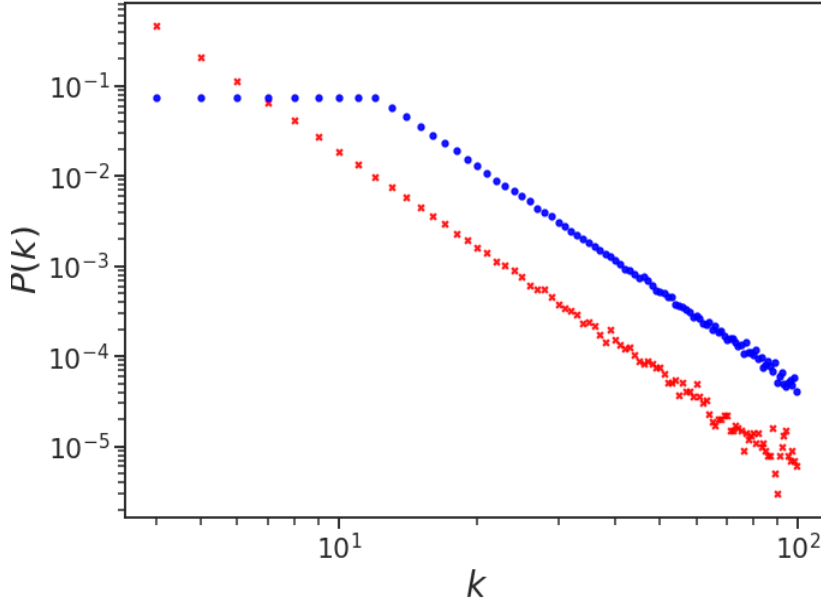


Figura 4.7: Curva em azul: Distribuição de graus com $\gamma = 3.50$ e $\langle k \rangle = 12$, $k_{min} = 4$, $k_{max} = k_c = \sqrt{N}$, $t = 8.16$ e $N = 10^4$. Curva em vermelho: Distribuição de graus com $\gamma = 3.50$ gerada através do sorteio aleatório de k sem controlar o grau médio. Os parâmetros utilizados foram $k_{min} = 4$, $k_{max} = k_c = \sqrt{N}$ e $N = 10^4$. Ambas as curvas apresentam a média sobre 100 redes.

Construindo $P(k)$ desta maneira, o grau médio pode ser escrito da seguinte forma,

$$\langle k \rangle = \frac{p(t) \sum_{i=k_{min}}^{\tilde{k}} k + \sum_{k+1}^{k_{max}} k^{1-\gamma}}{p(t) + \sum_{k+1}^{k_{max}} k^{-\gamma}}, \quad (4-5)$$

onde $p(t) = (k_{min} + t)^{-\gamma}$ é a altura do platô relativa a $(\tilde{k} + 1)^{-\gamma}$.

É possível observar na figura 4.7 um platô para os 9 primeiros pontos do gráfico, onde foram utilizados os parâmetros $\gamma = 3.50$, $\langle k \rangle = 12$, $k_{min} = 4$, $t = 8.16$ e $N = 10^4$.

A figura 4.8 apresenta resultados de $k_{nn}(k)$ e $c(k)$ para valores de $k_{min} = 2, 3, \dots, 7$ e $\gamma = 3.5, 2.5$ para $\langle k \rangle = 12$. As linhas em azul equivalem a $\gamma = 2.5$ enquanto as linhas em vermelho a $\gamma = 3.5$.

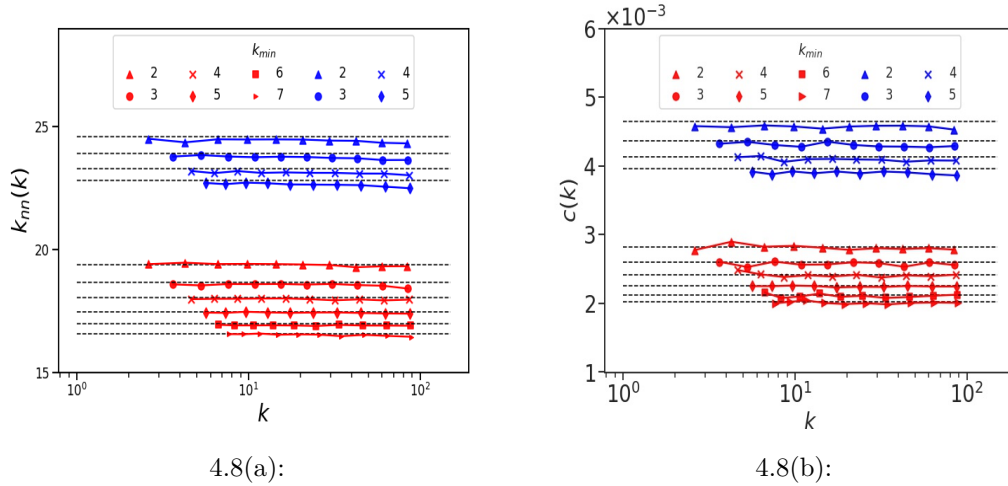


Figura 4.8: A figura (a) mostra os resultados para $k_{nn}(k)$ enquanto (b) os resultados para $c(k)$ (média sobre 100 redes). As redes foram geradas para $k_{min} = 2, \dots, 7$ e $\gamma = 3.5, 2.5$ e $\langle k \rangle = 12$. Em azul a rede tem $\gamma = 2.5$ enquanto em vermelho $\gamma = 3.5$. As linhas tracejadas correspondem às respectivas previsões teóricas em ausência de correlações, dadas pelas equações 2-13 e 2-18.

As figuras em 4.8 mostram que esse modelo não introduz correlações nas redes variando k_{min} e mantendo $k_{max} = k_c = \sqrt{N}$. Tanto $k_{nn}(k)$ quanto $c(k)$ não dependem de k .

4.1.3 Restringindo as flutuações do grau médio

Os dois métodos propostos acima cumprem o papel de controlar o grau médio da rede. No entanto, como exemplificado na figura 4.5, o $\langle k \rangle$ pode ainda flutuar entorno do valor pré-fixado. Isso ocorre pois o processo em que a sequência de graus é construída é aleatório. Nessa subseção, será proposta uma forma de fixar o grau médio de modo de eliminar as flutuações nessa métrica.

Para gerar uma sequência de graus utilizamos a distribuição $P(k)$. Sendo assim, o valor esperado para o número de vértices da rede com grau k é $N_k = NP(k)$. No entanto, esse número não é inteiro, em geral. Por outro lado, para cada dada realização da sequência N_k é inteiro. Sendo assim, quando somamos $\sum_k N_k/N$ temos o valor médio da conectividade que flutua em torno do valor desejado.

Para obter valores inteiros para N_k será proposto realizar primeiramente o truncamento dessa quantidade que será representado por $\lfloor NP(k) \rfloor$. Desta maneira, pode-se produzir uma lista com todos o graus da rede, onde em vez

de usar $P(k)$ para o sorteio, agora deterministicamente $\lfloor NP(k) \rfloor$ é o número de vezes que o grau k vai aparecer na sequência, em uma primeira etapa. Este procedimento produz um número menor de vértices e de arestas do que produziria sem truncar. Isto será corrigido numa segunda etapa.

A figura 4.9 mostra a “distribuição” intermediária de graus $p(k) \sim k^{-\gamma}$ após esse processo. Para essa rede foi utilizado inicialmente $N = 10^4$, $\gamma \sim 3,45$ e $\langle k \rangle = 6$. Com a redução da rede, a quantidade de vértices se reduziu a 9961. A medida que k aumenta fica mais visível que $p(k)$ retorna quantidades abaixo da linha teórica.

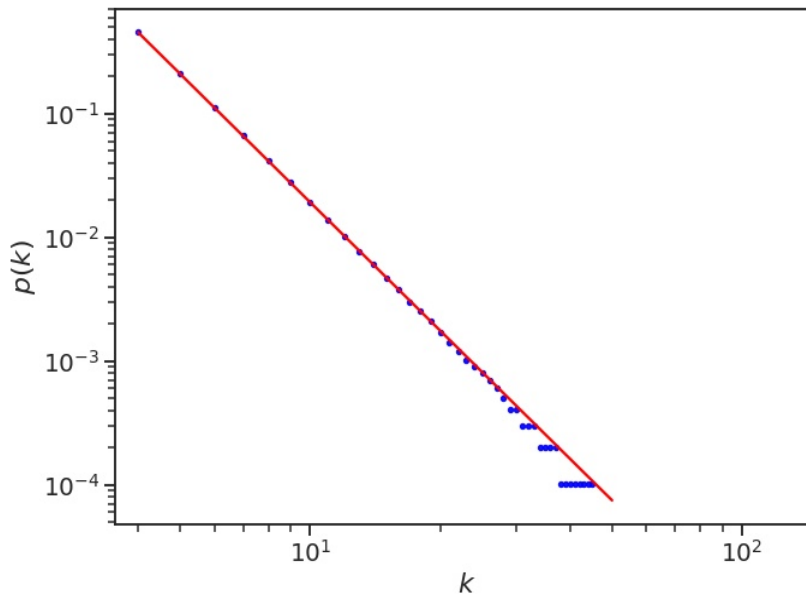


Figura 4.9: Histograma $p(k) = \lfloor NP(k) \rfloor / N$. Os parâmetros usados foram $\gamma \sim 3,45$ e $\langle k \rangle = 6$. Nesse caso, originariamente $N = 10^4$, porém com o processo de truncamento o total de vértices se reduziu para $\sum_k \lfloor NP(k) \rfloor = 9961$. A curva vermelha representa a distribuição teórica.

Com o objetivo de manter o tamanho N da rede e o grau médio $\langle k \rangle$ fixo, serão preenchidos os valores faltantes da sequência. Nesse caso, a medida $N - \sum_k \lfloor NP(k) \rfloor$ informa o número total de nós que faltam para completar a rede.

Dada a distribuição $P(k)$ onde $\sum_{k_{min}}^{k_{max}} P(k) = 1$, multiplicando o número de vértices N pela probabilidade de um grau k ocorrer na rede, os valores truncados serão, consequentemente, menores que os valores não truncados, sendo $0 \leq \tilde{p}(k) \equiv NP(k) - \lfloor NP(k) \rfloor \leq 1$. Este resto pode ser usado como probabilidade para sortear os próximos elementos k para completar a sequência da seguinte maneira:

1. Se $\sum_k [NP(k)] < N$ e $\sum_k k < \langle k \rangle N$, é realizado um sorteio aleatório de k , com $k \in [k_{min}, k_{max}]$ que entrará na sequência de graus obedecendo a probabilidade truncada definida anteriormente, isto é, $\tilde{p}(k)$.
2. Se falta apenas um vértice para chegar a N , a soma $\sum_k k$ ainda não chegou a $\langle k \rangle N$, o número de graus k faltantes está entre $[k_{min}, k_{max}]$ e ainda não foi preenchido da sequência, então k é adicionado.
3. Se completamos os vértices da rede até o valor N , porém o total de arestas ultrapassaram $\langle k \rangle N$, é retirado um vértice com grau mais comum da sequência de graus.
4. Se ultrapassamos o máximo de arestas possíveis e ainda faltam vértices a serem preenchidos é sorteado um k_i e um k_j , onde necessariamente $k_j > k_i$ e k_i já esteja preenchido na rede. Após o sorteio é feita uma troca de k_j por k_i para que reduza o número de ligações já preenchidos sem alterar o valor de N .
5. Caso faltem ligações, porém já completamos N , novamente sorteamos $[k_i, k_j]$, onde necessariamente $k_i > k_j$. Nesse caso sorteamos até que k_i já esteja preenchido na rede e então fazemos a troca de k_i por k_j .

A figura 4.10 apresenta a distribuição de graus da rede preenchida a partir da rede gerada na figura 4.9 seguindo o processo citado acima. A figura mostra o resultado de uma média sobre 10^3 redes onde o grau médio é exatamente igual a 6.

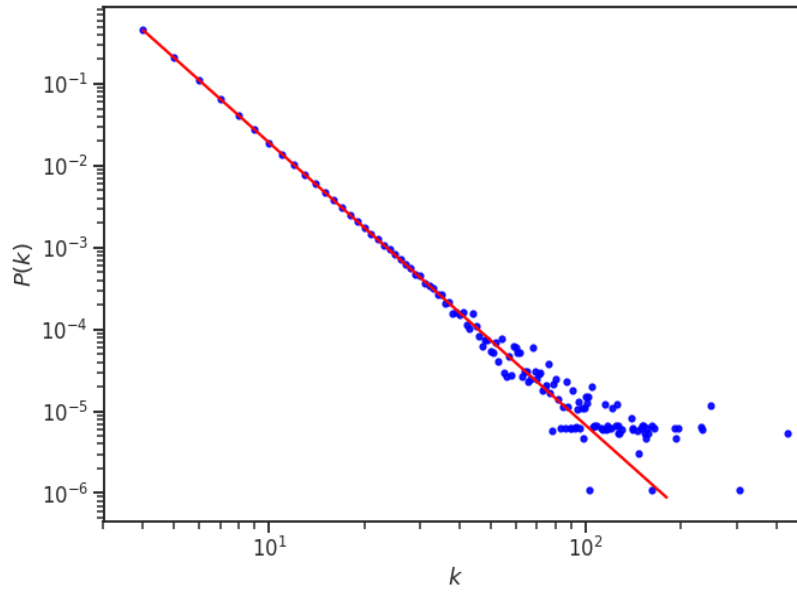


Figura 4.10: Distribuição de graus após a rede ser completada. Os parâmetros usados foram $\gamma \sim 3.45$, $k_{min} = 4$, $k_{max} = 2 \times 10^3$, $N = 10^4$ e $\langle k \rangle = 6$. A curva em vermelho é a distribuição $P(k) \sim k^{-\gamma}$ e os pontos em azul representam a média sobre 10^3 redes.

A rede construída na figura 4.10 tem $k_{min} = 4$, $k_{max} = 2 \times 10^3$ e $N = 10^4$ para que $\langle k \rangle = 6$.

5

Conclusões e comentários finais

Diversos sistemas reais da natureza podem ser representados por redes complexas. Assim sendo, a teoria de redes, propondo modelos que geram redes e criando métricas para caracterizá-las, tem se tornado uma das áreas mais promissoras, devido a sua alta capacidade de integração com diversas áreas da ciência.

No capítulo 2 revisamos conceitos básicos de redes. No capítulo 3 revisamos os modelos de Erdős-Rényi e livres de escala, gerando redes artificiais mediante o modelo de configuração, e explorando características estruturais e estatísticas para cada uma das redes criadas, principalmente a dispersão de valores do coeficiente de aglomeração, do grau médio entre vizinhos $k_{nn}(k)$, do coeficiente de Pearson e do grau médio.

No capítulo 4, introduzimos a ideia de controlar o grau médio $\langle k \rangle$, pois a evolução dinâmica em redes pode depender dessa métrica. Isto é importante especialmente para redes com distribuição de graus em lei de potência, e para tamanhos de rede não muito grandes. Logo, propomos dois tipos de distribuições de graus para controlar o grau médio, que chamamos de *método 1* e *método 2*.

No primeiro método introduzimos um parâmetro a na distribuição de graus que permite diminuir a probabilidade do grau mínimo k_{min} da rede sem alterar as características do restante da distribuição. Desta forma conseguimos controlar o grau médio da rede e encontrar o valor a através da equação 4-3. No método 2, é atribuída a mesma probabilidade a todos os nodos com baixa conectividade, formando um platô na distribuição de graus. Em ambos os casos a cauda em lei de potência é preservada.

Para os dois modelos, consideramos vários casos com diferentes valores do grau médio e do expoente γ da distribuição de graus. No método 2 também consideramos várias extensões do platô. Analisamos principalmente o grau médio entre vizinhos k_{nn} e o coeficiente de aglomeração médio $c(k)$.

Chegamos à conclusão de que os procedimentos propostos permitem controlar o grau médio sem introduzir novas correlações nas redes geradas pelo modelo de configuração.

Apesar dos dois métodos controlarem o grau médio com baixo desvio, em

um ensemble de redes o grau médio flutua em torno do valor pré-estabelecido. Para eliminar este desvio introduzimos no final do capítulo 4, um método, para fixar essa métrica com precisão.

Referências bibliográficas

- [1] BIGGS, N.; LLOYD, E. K. ; WILSON, R. J. **Graph Theory, 1736-1936**. Oxford University Press, 1986.
- [2] WATTS, D. J.; STROGATZ, S. H. **nature**. Collective dynamics of 'small-world' networks, journal, v.393, n.6684, p. 440–442, 1998.
- [3] BARABÁSI, A.-L.; ALBERT, R. **science**. Emergence of scaling in random networks, journal, v.286, n.5439, p. 509–512, 1999.
- [4] CARVALHO, R.; BUZNA, L.; BONO, F.; GUTIÉRREZ, E.; JUST, W. ; ARROWSMITH, D. **Physical review E**. Robustness of trans-European gas networks, journal, v.80, n.1, p. 016106, 2009.
- [5] BERNERS-LEE, T.; CAILLIAU, R.; GROFF, J.-F. ; POLLERMANN, B. **Internet Research**. World-Wide Web: the information universe, journal, 1992.
- [6] CONWAY, E. M.; COLLEN, D. ; CARMELIET, P. **Cardiovascular research**. Molecular mechanisms of blood vessel growth, journal, v.49, n.3, p. 507–521, 2001.
- [7] NEWMAN, M. E. **Proceedings of the national academy of sciences**. The structure of scientific collaboration networks, journal, v.98, n.2, p. 404–409, 2001.
- [8] THURNER, S.; HANEL, R. ; KLIMEK, P. **Introduction to the theory of complex systems**. Oxford University Press, 2018.
- [9] ERDŐS, P.; RÉNYI, A. **Publ. Math. Inst. Hung. Acad. Sci.** On the evolution of random graphs, journal, v.5, n.1, p. 17–60, 1960.
- [10] ERDŐS, P.; RÉNYI, A. **Acta Mathematica Academiae Scientiarum Hungarica**. On the strength of connectedness of a random graph, journal, v.12, n.1-2, p. 261–267, 1964.
- [11] BARABÁSI, A.-L. **science**. Scale-free networks: a decade and beyond, journal, v.325, n.5939, p. 412–413, 2009.

- [12] COSTA, L. D. F.; RODRIGUES, F. A.; TRAVIESO, G. ; VILLAS BOAS, P. R. **Advances in physics**. Characterization of complex networks: A survey of measurements, journal, v.56, n.1, p. 167–242, 2007.
- [13] CATANZARO, M.; BOGUNÁ, M. ; PASTOR-SATORRAS, R. **Physical review e**. Generation of uncorrelated random scale-free networks, journal, v.71, n.2, p. 027103, 2005.
- [14] RAMOS, M.; ANTENEODO, C. **Journal of Statistical Mechanics: Theory and Experiment**. Random degree–degree correlated networks, journal, v.2013, n.02, p. P02024, 2013.
- [15] VIEIRA, A. R.; PERALTA, A. F.; TORAL, R.; MIGUEL, M. S. ; ANTENEODO, C. **Phys. Rev. E**. Pair approximation for the noisy threshold q -voter model, journal, v.101, p. 052131, May 2020.
- [16] WEST, D. B.; OTHERS. **Introduction to graph theory**, volume 2. Prentice hall Upper Saddle River, 2001.
- [17] NEWMAN, M. **Networks: An Introduction**. OUP Oxford, 2010.
- [18] SINGH, H.; SHARMA, R. **International Journal of Computers & Technology**. Role of adjacency matrix & adjacency list in graph theory, journal, v.3, n.1, p. 179–183, 2012.
- [19] WEISSTEIN, E. W. <https://mathworld.wolfram.com/>. Adjacency matrix, journal, 2007.
- [20] RAHMAN, M. S.; OTHERS. **Basic graph theory**. Springer, 2017.
- [21] Adjacency list matchings: an ideal genotype for cycle covers. t. In: PROCEEDINGS OF THE 9TH ANNUAL CONFERENCE ON GENETIC AND EVOLUTIONARY COMPUTATION, p. 1203–1210, 2007.
- [22] NEWMAN, M. E. **The new palgrave encyclopedia of economics**. The mathematics of networks, journal, v.2, n.2008, p. 1–12, 2008.
- [23] SEDGEWICK, R.; WAYNE, K. **Algorithms**. Addison-wesley professional, 2011.
- [24] EVEN, S. **Graph algorithms**. Cambridge University Press, 2011.
- [25] NEWMAN, M. E. J. **Networks: an introduction**. Oxford; New York: Oxford University Press, 2010.

- [26] NEWMAN, M. E. **Physical review letters**. Assortative mixing in networks, journal, v.89, n.20, p. 208701, 2002.
- [27] NEWMAN, M. E. **Physical review E**. Mixing patterns in networks, journal, v.67, n.2, p. 026126, 2003.
- [28] XULVI-BRUNET, R.; SOKOLOV, I. M. **Physical Review E**. Reshuffling scale-free networks: From random to assortative, journal, v.70, n.6, p. 066102, 2004.
- [29] MENCHE, J.; VALLERIANI, A. ; LIPOWSKY, R. **Physical review E**. Asymptotic properties of degree-correlated scale-free networks, journal, v.81, n.4, p. 046103, 2010.
- [30] SERRANO, M. A.; BOGUNÁ, M.; PASTOR-SATORRAS, R. ; VESPIGNANI, A. **Large scale structure and dynamics of complex networks: From information technology to finance and natural sciences**. Correlations in complex networks, journal, p. 35–66, 2007.
- [31] NEWMAN, M. **Networks**. Oxford university press, 2018.
- [32] Random graph generation for scheduling simulations. t. 2010.
- [33] A critical point for random graphs with a given degree sequence. t. In: **THE STRUCTURE AND DYNAMICS OF NETWORKS**, p. 240–258. Princeton University Press, 2011.
- [34] NADAKUDITI, R. R.; NEWMAN, M. E. **Physical Review E**. Spectra of random graphs with arbitrary expected degrees, journal, v.87, n.1, p. 012803, 2013.
- [35] BOLLOBÁS, B. **Modern graph theory**, volume 184. Springer Science & Business Media, 2013.
- [36] BOLLOBÁS, B.; BÉLA, B. **Random graphs**. Number 73. Cambridge university press, 2001.
- [37] NEWMAN, M. E.; OTHERS. **Handbook of graphs and networks**. Random graphs as models of networks, journal, v.1, p. 35–68, 2003.
- [38] CALDARELLI, G. **Large scale structure and dynamics of complex networks: from information technology to finance and natural science**, volume 2. World Scientific, 2007.

- [39] WAXMAN, B. M. **IEEE journal on selected areas in communications**. Routing of multipoint connections, journal, v.6, n.9, p. 1617–1622, 1988.
- [40] MITZENMACHER, M. **Internet mathematics**. A brief history of generative models for power law and lognormal distributions, journal, v.1, n.2, p. 226–251, 2004.
- [41] BIANCONI, G.; BARABÁSI, A.-L. **EPL (Europhysics Letters)**. Competition and multiscaling in evolving networks, journal, v.54, n.4, p. 436, 2001.
- [42] NEWMAN, M. E. **SIAM review**. The structure and function of complex networks, journal, v.45, n.2, p. 167–256, 2003.
- [43] DOROGOVTSSEV, S. **Physical Review E**. Clustering of correlated networks, journal, v.69, n.2, p. 027104, 2004.
- [44] BOGUNÁ, M.; PASTOR-SATORRAS, R. ; VESPIGNANI, A. **The European Physical Journal B**. Cut-offs and finite size effects in scale-free networks, journal, v.38, n.2, p. 205–209, 2004.