

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
UFRGS

Tese de Doutorado

Soluções analíticas para o modelo de Barabási-Albert de crescimento de redes.

Ricardo Melo Ferreira

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal do Rio Grande do Sul, sob orientação da Profa. Rita Maria Cunha de Almeida e coorientação do Prof. Leonardo Gregory Brunnet.

Tese de doutorado apresentada em Porto Alegre, quatro de Outubro de 2016

Resumo

Redes Complexas podem ser utilizadas para estudar uma variedade de assuntos, desde problemas tecnológicos a de saúde pública. Um dos modelos de redes complexas mais famoso é o modelo de Barabási-Albert, proposto inicialmente para reproduzir propriedades da rede mundial de computadores. Este modelo constrói redes através da sucessiva adição de nós, seguindo o princípio da adesão preferencial. Este princípio afirma que a probabilidade de um nó receber um vizinho é proporcional ao número de vizinhos que esse nó já possui. O principal resultado deste modelo é uma distribuição de grau que segue uma lei de potência. Uma vez que diversas redes reais apresentam distribuições de grau que se aproximam de uma lei de potência o princípio de adesão preferencial do modelo de Barabási-Albert passou a ser considerado como um dos principais mecanismos por trás da formação das redes reais. A comparação dessas redes com o modelo de Barabási-Albert é controversa, uma vez que suas distribuições de grau não são perfeitamente aproximadas por uma lei de potência. Entretanto, redes geradas através do próprio modelo apresentam desvios devido a efeitos de tamanho finito. Nesse contexto, soluções analíticas capazes de descrever o modelo de Barabási-Albert para redes pequenas são bem vindas, pois permitiriam compreender a extensão dos efeitos de tamanho finito e, portanto, poderiam ser comparadas com redes reais pequenas. Neste trabalho apresentamos um método que permite obter uma descrição analítica do modelo de Barabási-Albert, válida também para redes pequenas, e apresentamos essas soluções.

Abstract

Complex Networks can be used to study a variety of subjects ranging from tecnologic to public health problems. Barabási-Albert model for complex networks was initially proposed to describe the world wide web. This model builds networks successively adding nodes following a preferential attachment mechanism, where the probability of a node to receive a new neighbour is proportional to the number of neighbours the node already has. Its main result is a degree distribution following a power-law. Since many real networks have degree distributions similar to power-laws, the Barabási-Albert model was considered a main mechanism underlying real netowrks formation. The comparison between these networks and the Barabási-Albert model is controversial, since their degree distributions are not perfect power-laws. However, networks grown following the model algorithm show deviations from a power-law due to finite-size effects. In this context, analytical solutions describing Barabási-Albert model for small networks are welcome, as a tool to verify the extent of the finite-size effects in small real networks, and therefore could be compared to small real networks. In this work we present a method to obtain an analytical description for Barabási-Albert networks, valid for small systems, and present these solutions.

Sumário

1	Introdução	5
2	Redes Complexas	9
2.1	Medidas	11
3	Modelo de Barabási-Albert	18
3.1	Modelos de adesão mais gerais	21
3.2	Soluções Analíticas	23
4	Equações mestras e obtenção das soluções analíticas	31
4.1	Equações Mestras	31
4.2	Equação diferencial	34
5	Resultados	39
6	Conclusão	47
6.1	Perspectivas	48
	Referências Bibliográficas	50

Capítulo 1

Introdução

Redes complexas possuem um amplo espectro de aplicações. No campo de tecnologia podem ser utilizadas na procura de redes robustas a falhas para comunicações ou transmissão de energia [1], no campo da saúde pública como base para simulações de propagação de epidemias [2–4], no estudo da malha aérea [5], em ferramentas de diagnóstico [6], na avaliação do sistema legislativo [7] e na evolução das redes de associação de proteínas [8], para citar alguns exemplos. Tanta variedade de assuntos é possível pois uma rede é um modelo simplificado de um sistema complexo, composta de dois elementos. Os nós, que representam os elementos das redes, e as ligações entre os nós, que representam alguma interação entre os elementos. Essa visão reducionista dos sistemas complexos permite que informações estatísticas relevantes sejam extraídas da estrutura de interações subjacente ao sistema. No caso de propagação de epidemias, por exemplo, cada nó pode representar um indivíduo, e as ligações representam contatos pessoais desse indivíduo. Cada indivíduo tem uma dinâmica particular, um sistema imunológico e exposição à epidemia diferente dos demais. As interações pessoais, dependendo da sua natureza, podem estar presentes em instantes diferentes, como os contatos familiares e os do local de trabalho. Entretanto, ao estudar a propagação de epidemias em uma rede podemos ignorar as idiosincrasias de cada nó e interação, e procurar comportamentos estatísticos globais.

O mais conhecido modelo para crescimento de redes em geral é o de Barabási-Albert. Neste modelo considera-se apenas a aquisição de novos nós que associam-se preferencialmente a genes

com maior número de conexões, isto é, maior grau [9, 10]. O resultado são redes que assintoticamente apresentam uma distribuição de número de ligações, ou grau, que obedece uma lei de potência com expoente $\gamma = -3$ e por isso chamadas de redes livre de escala. A distribuição na forma de uma lei de potência é uma característica fundamental do modelo de Barabási-Albert, se compararmos, por exemplo, com o modelo de rede aleatória proposto por Erdős e Rényi [11] no qual a distribuição de grau segue uma distribuição de poisson [12]. Isto significa que em uma rede gerada pelo modelo de Barabási-Albert são encontrados nós com grau alto em maior número do que o esperado caso as ligações fossem distribuídas aleatoriamente.

Diversas redes reais apresentam uma distribuição de grau que diminui mais lentamente que uma exponencial, ou seja, de cauda longa, e portanto foram comparadas com o modelo de Barabási-Albert [10, 13]. O modelo pode então ser um arquétipo de redes que possuam esse tipo de distribuição de grau, não necessariamente como a dinâmica por trás das redes reais, mas sim indicando um princípio subjacente a geração de tais distribuições. Entretanto a comparação de redes reais com o modelo de Barabási-Albert é criticada, principalmente levando em consideração desvios que as redes apresentam de uma lei de potência [14–16]. Porém, mesmo simulações do algoritmo de crescimento do modelo apresentam desvios de uma lei de potência devido a efeitos de tamanho finito [17], e, uma vez que redes reais geralmente apresentam um número pequeno de nós [13, 18], esses efeitos devem ser levados em consideração.

Outras medidas podem descrever com mais detalhes a topologia da rede, como as correlações entre os graus que estão ligados, ou a formação de triângulos que acontece quando dois vizinhos de um nó são ligados também entre si. A clusterização, por exemplo, também descreve a vizinhança de um nó e pode ser importante para a dinâmica de crescimento da rede [8], para estudo de resistência à falhas [1] e propagação de epidemias [2, 19]. No modelo de Barabási-Albert a clusterização costuma ser desprezada, pois tende a zero quando o tamanho da rede vai para infinito, mas, como mencionado anteriormente, redes reais estão longe desse limite, e portanto a clusterização desse modelo pode ser relevante nesse caso.

Nesse contexto resultados analíticos para o modelo de Barabási-Albert tornam-se importantes. Estes resultados permitiriam obter as medidas topológicas esperadas para uma rede pequena, e

ainda decidir se desvios da lei de potência são devidos aos efeitos de tamanho ou se o mecanismo de crescimento da rede é incompatível com o modelo. A maior parte dos resultados encontrados na literatura estudam a distribuição de grau no limite de tempo infinito [9, 20–23], com um trabalho que encontra uma correção para redes de tamanho finito [24]. Além disso a clusterização é encontrada de forma indireta [25] e para redes muito grandes [26]. Permanece então a lacuna para uma descrição da distribuição de ligações e de triângulos em função dos graus para tempo finito.

Este trabalho foi motivado pelas questões acima que surgiram durante um trabalho anterior no qual propusemos o modelo de Duplicação-Aquisição que reproduz as redes de associação de proteínas [8]. Nesse modelo novos nós surgem na rede através de dois mecanismos. O primeiro é similar ao do modelo de Barabási-Albert, representando surgimento *de novo*, e o segundo é inspirado no modelo de Duplicação-Divergência [27, 28], representando a duplicação de genes. Ao estudarmos redes de associação de proteínas a primeira questão foi a adequação do modelo de Barabási-Albert para descrevê-las, uma vez que o mecanismo mais biologicamente plausível para o crescimento dessas redes é a duplicação gênica. A segunda questão surgiu quando levamos em consideração a topologia da rede para escolher os nós a serem duplicados, onde os escolhidos são os menos conectados e clusterizados. Além disso as redes de associação de proteínas são pequenas, as maiores tendo da ordem de 10000 nós. Percebemos então a necessidade de resultados analíticos que contemplassem também outras medidas da estrutura da rede e fossem válidos para redes pequenas, a fim de compararmos com as redes reais.

Este texto está estruturado da seguinte forma. Vamos primeiro introduzir algumas medidas de redes complexas que utilizaremos para caracterização de redes. A seguir apresentaremos o modelo de Barabási-Albert e alguns modelos de dinâmicas adesão de nós muito similares que podem ser considerados variações do primeiro. Neste mesmo capítulo mostramos alguns dos principais resultados analíticos já conhecidos, para comparar com os resultados que iremos obter. No quarto capítulo escreveremos as equações mestras para o número de nós com determinado grau, para o número de ligações entre dois nós em função de seus graus, e para o número de triângulos com vértices de determinados graus, e apresentaremos um método de solução analítica

das equações diferenciais associadas que permite obter essas quantidades no limite de tempo finito. As soluções então obtidas serão comparadas no capítulo seguinte com resultados de simulação e com a iteração numérica das equações master. Finalmente, nas conclusões, vamos discutir a relevância e as aplicações desses resultados, e a comparação deles com os já existentes na literatura.

Capítulo 2

Redes Complexas

Neste capítulo vamos apresentar o conceito de redes complexas e dar uma breve introdução da sua história. A seguir apresentaremos algumas medidas que utilizaremos para caracterizar o modelo de Barabási-Albert. Não pretendemos fazer uma revisão extensa das medidas topológicas existentes para redes complexas, pois sairíamos rapidamente do escopo desse texto. Entretanto, para o leitor interessado indicamos algumas referências úteis para prosseguir no estudo.

Um grafo é definido formalmente como um par ordenado (V, E) consistindo de um conjunto V de vértices e um conjunto de arestas E , disjunto de V , juntamente com uma função ψ que associa cada aresta de E com um par não ordenado de vértices de V . [29]. Grafos são coloquialmente chamados de redes, com seus vértices e arestas denominados de nós e ligações, respectivamente. Redes podem ser construídas de diversas formas: as ligações podem ter pesos, representando a intensidade da interação entre os vértices, ou podem ser direcionadas do nó i para o j , representando o sentido da interação, como numa rede de citações bibliográficas, onde i cita j . Os nós também podem ser de mais de um tipo, como numa rede de colaboração entre atores, onde um tipo de nós representariam os autores e outro, os atos de interação, como filmes.

O precursor da teoria de grafos foi Leonhard Euler, que em 1735 resolveu o problema das pontes de Königsberg [3]. Esta cidade prussiana, localizada onde hoje está localizada Kaliningrado, possuía no rio Pregel duas ilhas conectadas às margens e entre si por sete pontes (veja esquema na figura 2.1). O problema que intrigava os habitantes de Königsberg se tratava de encontrar um

caminho que permitisse atravessar uma única vez cada uma das pontes. Euler encontrou a solução simplificando o problema. Uma vez que o caminho entre uma ponte e outra não é relevante, cada massa de terra pode ser substituída por um ponto, com ligações entre eles representando as pontes. As margens e ilhas do rio Pregel foram, então, representadas por uma rede, que pode ser observada na figura 2.2. Euler apontou que, excetuando-se o ponto de partida e o de chegada,

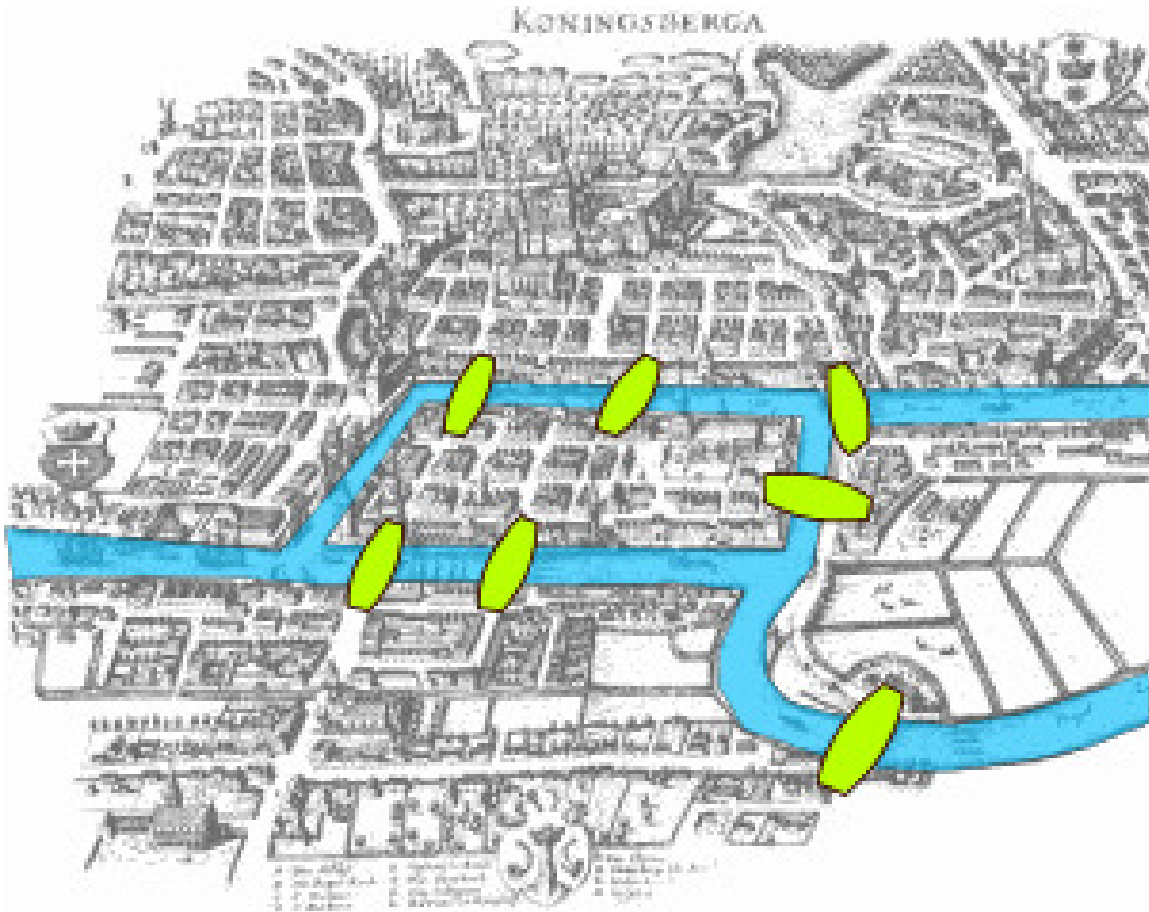


Figura 2.1: Mapa de Königsberg com as pontes em destaque. [30]

em todos os outros seria necessário entrar e sair e, portanto, estes têm que ter um número par de conexões. Logo era impossível fazer um caminho passando-se uma única vez por cada ponte, já que todos os pontos nessa rede teriam um número ímpar de ligações.

Apesar de existirem propostas de medidas da complexidade de uma rede [31], redes complexas são geralmente tratadas como a intersecção entre teoria de grafos e mecânica estatística, apresentando propriedades que não são encontradas em grafos, como estruturas de comunidades,

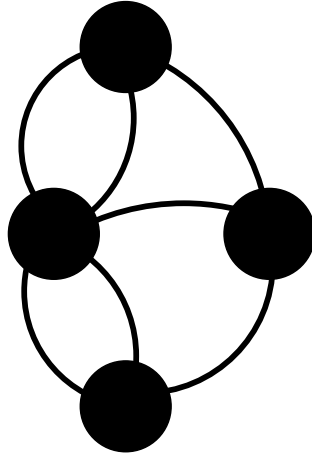


Figura 2.2: Rede simbolizando Königsberg onde as massas de terra são representadas por nós e pontes por ligações.

distribuições de grau, *hubs* ou *clusters* ou até mesmo dinâmicas de adição ou remoção de nós e ligações [32]. Redes complexas podem descrever uma série de fenômenos no mundo real como propagação de epidemias, colaboração de atores em filmes, linguística, redes de distribuição elétrica, internet e redes de associação proteica [13]. Utilizar uma única abordagem para problemas tão variados somente é possível porque os elementos e as interações são representados por nós e ligações sem distinção de sua natureza. Ao longo deste trabalho, nos referiremos às redes complexas simplesmente como redes, e consideraremos redes com ligações sem peso, não direcionadas e com um único tipo de nó.

2.1 Medidas

Apresentaremos algumas medidas utilizadas no estudo de redes que serão úteis posteriormente [33]. A definição de outras medidas, bem como generalizações destas para casos mais específicos é simples e pode ser encontrada em Costa *et al.* [32]. Uma rede com n nós pode ser completamente definida por uma matriz quadrada de elementos A_{ij} , chamada matriz de associação, representativa das ligações entre cada par de nós da rede: $A_{i,j} = 1$ indica a existência de uma ligação entre os nós i e j ; $A_{i,j} = 0$ a sua inexistência. Como as ligações não são direcionadas a matriz de associação é simétrica.

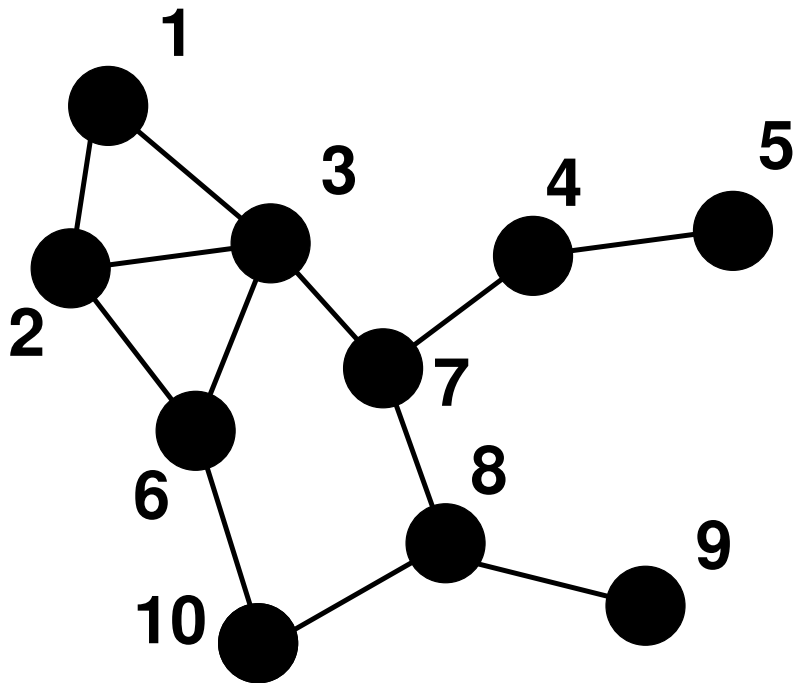


Figura 2.3: Rede esquemática apresentando nós e ligações. A numeração dos nós pode ser aleatória e representa apenas um índice do nó.

Vamos considerar, por exemplo a rede da figura 2.3. Sua matriz de associação pode ser dada por

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}. \quad (2.1)$$

Nas sessões a seguir iremos utilizar essa rede exemplo para apresentar algumas medidas topológicas das redes complexas.

2.1.1 Grau

A primeira característica de um nó i é o seu grau k_i , que representa o seu número de ligações, ou de vizinhos, e é dado por

$$k_i = \sum_{j=1}^n A_{i,j} . \quad (2.2)$$

Se tomarmos o nó 6 da nossa rede exemplo podemos ver, na figura 2.4 que o número de vizinhos desse nó é 3, portanto $k_6 = 3$.

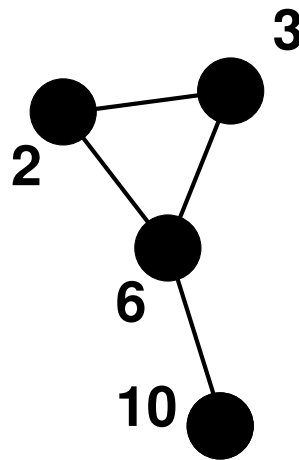


Figura 2.4: Detalhe do nó 6 da rede esquemática

2.1.2 Distribuição de Graus

As ligações entre os nós formam a estrutura da rede, mas para uma rede com um grande número de nós não podemos analisar o grau de cada nó independentemente. Definimos, então, $N(k, t)$ como o número de nós com grau k no tempo t

$$N(k, t) = \sum_{i=1}^n \delta_{k_i, k} . \quad (2.3)$$

A distribuição de grau então é definida pelo número de nós com grau k dividido pelo número total de nós $\mathcal{N}(t)$

$$P(k, t) = \frac{N(k, t)}{\mathcal{N}(t)}. \quad (2.4)$$

A distribuição de grau pode também ser interpretada como a probabilidade de um nó escolhido aleatoriamente ter grau k . Esta fração pode também ser utilizada para determinar a estrutura da rede. Por exemplo, uma rede aleatória possui uma distribuição de grau dada pela distribuição de Poisson [3], e redes com uma distribuição de grau que seguem uma lei de potência, na forma $P(k, t) \propto k^{-\gamma}$ ($\gamma > 0$), são as chamadas redes livres de escala (*scale-free*).

2.1.3 Ligações entre graus

Outra forma de caracterizar estatisticamente a rede é procurar correlações entre os graus. Assim como $N(k, t)$ contém o número de nós de grau k , podemos armazenar a informação sobre as ligações entre nós de grau k e k' na função $L(k, k', t)$. Uma restrição que impomos sobre esta função é

$$\sum_{k, k'} L(k, k', t) = \frac{1}{2} \sum_{i, j=1}^n A_{ij} = \mathcal{L}(t), \quad (2.5)$$

onde $\mathcal{L}(t)$ é o número total de ligações no tempo t . Para satisfazer essa equação, $L(k, k', t)$ é definida como o número de ligações entre nós de graus k e k' se $k = k'$ e metade do número de ligações entre nós de graus k e k' se $k \neq k'$. Como as ligações não são direcionadas, essa função é simétrica frente à troca entre k e k' . Segue da definição que

$$\sum_{k'} L(k, k', t) = \frac{kN(k, t)}{2}. \quad (2.6)$$

2.1.4 Ciclos

Um ciclo é um caminho fechado, percorrido através das ligações entre os nós. Seu comprimento é medido pelo número (mínimo) de ligações necessárias para se voltar a um nó de partida sem repetir passagens sobre outro nó qualquer. Em rede sociais, por exemplo, é frequente a presença de ciclos de tamanho três, ou triângulos, e é fácil perceber que isto ocorre. Se escolhermos dois amigos de determinado indivíduo, há uma grande probabilidade de eles serem também amigos entre si.

Assim como definimos $N(k, t)$ e $L(k, k', t)$ vamos definir $\Delta(k, k', k'', t)$ como o número de triângulos com vértices de graus k, k' e k'' no tempo t . De forma análoga à $L(k, k', t)$, $\Delta(k, k', k'', t)$ será simétrico e

$$\sum_{k, k', k''} \Delta(k, k', k'', t) = \mathcal{D}(t), \quad (2.7)$$

onde $\mathcal{D}(t)$ é o número total de triângulos presentes na rede no tempo t . Portanto, para evitar que cada triângulo seja contabilizado mais de uma vez, $\Delta(k, k', k'', t)$ é definido como: o número de triângulos entre nós de graus k, k' e k'' , se $k = k' = k''$; um sexto dos triângulos entre nós de graus k, k' e k'' , se $k \neq k' \neq k''$; e um terço dos triângulos entre nós de graus k, k' e k'' , se $k \neq k' = k''$. Na rede esquemática da figura 2.3 temos dois triângulos, um entre os nós 1, 2 e 3, e outro entre os nós 2, 3, e 6. Se observarmos o grau desses nós vemos que

$$\begin{aligned} \Delta(2, 3, 4, t) &= \frac{1}{6} \\ \Delta(3, 4, 4, t) &= \frac{1}{3}. \end{aligned} \quad (2.8)$$

Queremos que a condição 2.7 seja satisfeita. São possíveis seis permutações nos graus do primeiro triângulo, portanto esse triângulo seria somado seis vezes, então somamos $1/6$ em cada um dos termos.

Uma forma de medir a presença destes ciclos de tamanho três é a clusterização. O número total de ciclos de grau três (N_3) que poderiam passar por um nó i é a combinação de seus vizinhos tomados dois a dois, $N_3 = k_i(k_i - 1)/2$. Já o número de ciclos de tamanho três que passam por

este nó é o número de ligações entre os seus vizinhos, ou seja, o número de triângulos $\Delta_i(t)$ aos quais o nó i pertence. A clusterização $C_i(t)$ do nó é então dada por

$$C_i(t) = \frac{2\Delta_i(t)}{k_i(k_i - 1)}. \quad (2.9)$$

Analisando novamente o nó 6 na nossa rede exemplo, podemos ver na figura 2.5 que esse nó pertence somente a um triângulo, mas que o número total de triângulos possíveis, indicados pelas linhas tracejadas, é três. Portanto a clusterização do nó 6 é

$$C_3(t) = \frac{1}{3}. \quad (2.10)$$

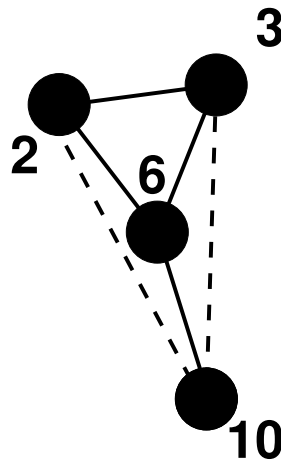


Figura 2.5: Detalhe do nó 6 da rede esquemática evidenciando o número de triângulos possíveis com as linhas tracejadas.

Mas assim como o grau, a clusterização do nó é uma medida local, e portanto não é útil para caracterizar grandes redes. Podemos definir a clusterização média da rede

$$\langle C(t) \rangle = \frac{1}{\mathcal{N}(t)} \sum_{i=1}^n C_i(t), \quad (2.11)$$

mas novamente perdemos informação sobre a estrutura da rede. Utilizaremos então a clusterização

média por grau, dada por

$$\begin{aligned}
\langle C(t) \rangle_k &= \frac{1}{N(k, t)} \sum_{i=1}^n \delta_{k, k_i} C_i(t) \\
&= \frac{1}{N(k, t)} \sum_{i=1}^n \delta_{k, k_i} \frac{2\Delta_i(t)}{k(k-1)}.
\end{aligned} \tag{2.12}$$

Uma vez que $\sum_{i=1}^n \delta_{k, k_i} \frac{2\Delta_i(t)}{k(k-1)} = \sum_{k', k''} \Delta(k, k', k'', t)$, podemos reobter

$$\begin{aligned}
\langle C(t) \rangle_k &= \frac{1}{N(k, t)} \sum_{i=1}^n \delta_{k, k_i} \frac{2\Delta_i(t)}{k(k-1)} \\
&= \frac{2 \sum_{k', k''} \Delta(k, k', k'', t)}{N(k, t) k(k-1)}.
\end{aligned} \tag{2.13}$$

As medidas definidas acima permitem compreender a estrutura da rede, uma vez que apresentam o número de nós com determinado grau, como os nós de graus diferentes se ligam entre si, e sobre a vizinhança em função do grau. Vamos, então, obter essas medidas de forma analítica para caracterizarmos o modelo de Barabási-Albert para redes de tamanho finito.

Capítulo 3

Modelo de Barabási-Albert

Aqui apresentaremos em mais detalhe o modelo de Barabási-Albert e alguns modelos que introduzem variações. Apesar de serem modelos distintos, seguem o mesmo princípio de adesão preferencial, e alguns desses modelos são utilizados para a obtenção de soluções analíticas, que serão vistas a seguir.

Os modelos de adesão constroem redes adicionando novos nós com um certo número de ligações que pode ou não ser fixo. Estas ligações distribuem-se na rede existente de várias formas: aleatoriamente; como uma função do grau; ou de acordo com outra função de atração dos nós. Estes modelos buscam reproduzir diversas redes, como internet ou citações, e costumam dar origem a distribuições de grau na forma de leis de potência.

O modelo de adesão mais conhecido, e um dos mais simples, é o de Barabási-Albert [9,10]. Foi uma primeira proposta para explicar diversas redes, como redes de associação proteica, internet, colaboração de atores, entre outras. Neste modelo cada novo nó é adicionado com um número m de ligações, e cada vizinho j é escolhido com probabilidade proporcional ao seu grau k_j . Assim o nó com maior grau tem maior probabilidade de receber uma nova ligação. Esse mecanismo é chamado de adesão preferencial, e é de fácil justificativa na WWW, por exemplo, onde uma página mais conhecida tem uma probabilidade maior de receber uma referência do que uma pouco conhecida. Este modelo produz, para um alto número de nós, uma rede livre de escala com uma distribuição de grau que se aproxima de $P(k) = 2m^2/k^3$ com expoente $\gamma = 3$.

Nas figuras abaixo apresentaremos as medidas para o modelo de Barabási-Albert. Foram realizadas médias sobre 10000 simulações, com $m = 3$ e com a condição inicial de $m + 2$ nós completamente ligados entre si. Na figura 3.1 temos o número de nós $N(k, t)$ para o modelo de Barabási-Albert para três valores de t . Podemos observar que $N(k, t)$ decai como uma lei de potência com k , conforme o estabelecido. Além disso o número de nós cresce com o tempo, tanto o número total quanto o valor para cada grau individualmente. Podemos ainda observar que para valores maiores de k o comportamento de $N(k, t)$ desvia da reta associada com uma lei de potência.

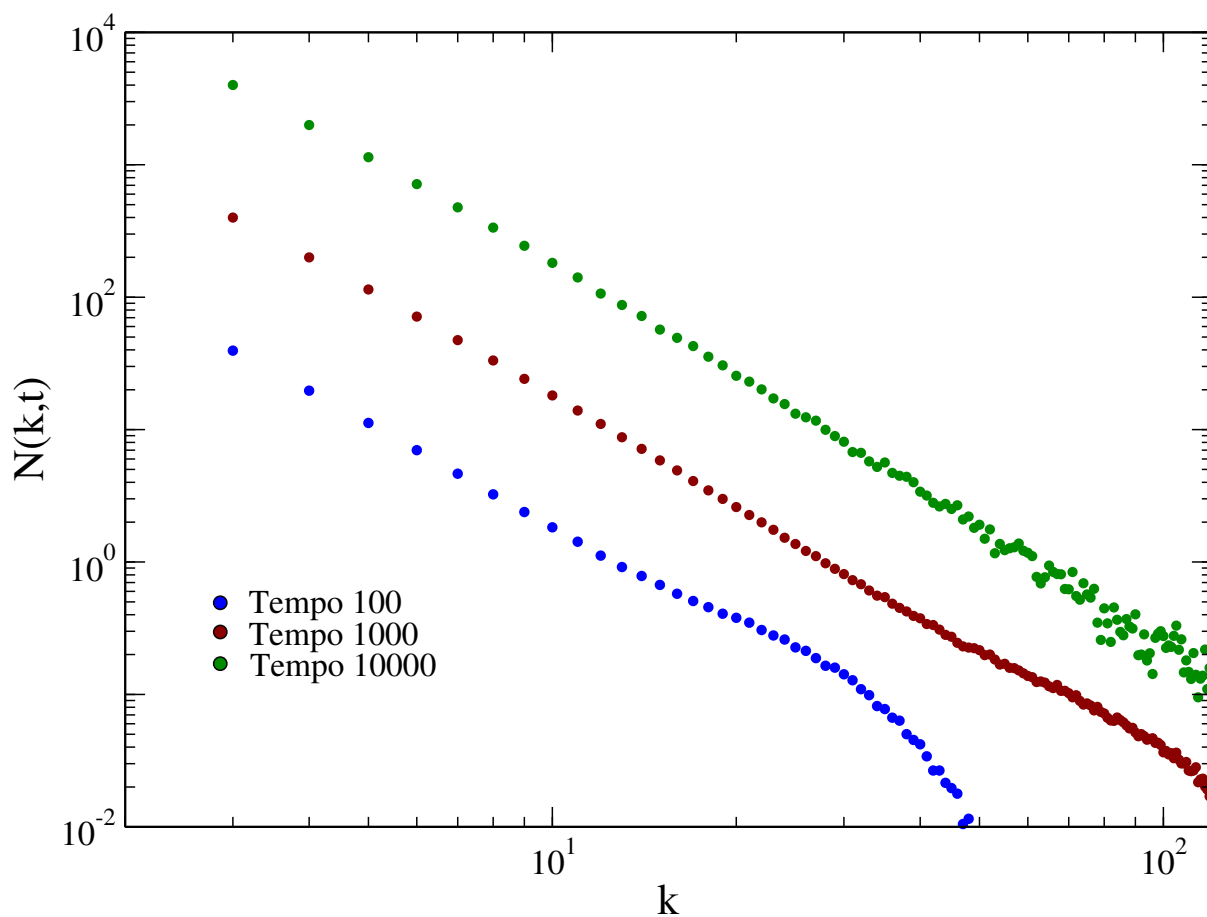


Figura 3.1: Número de nós com grau k para o modelo de Barabási-Albert obtidos através da média sobre 10000 simulações para três valores de t . Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$.

Como $L(k, k', t)$ depende de mais uma variável a visualização mais natural seria na forma de uma figura tridimensional. Na figura 3.2 podemos observar $L(k, k', 10000)$.

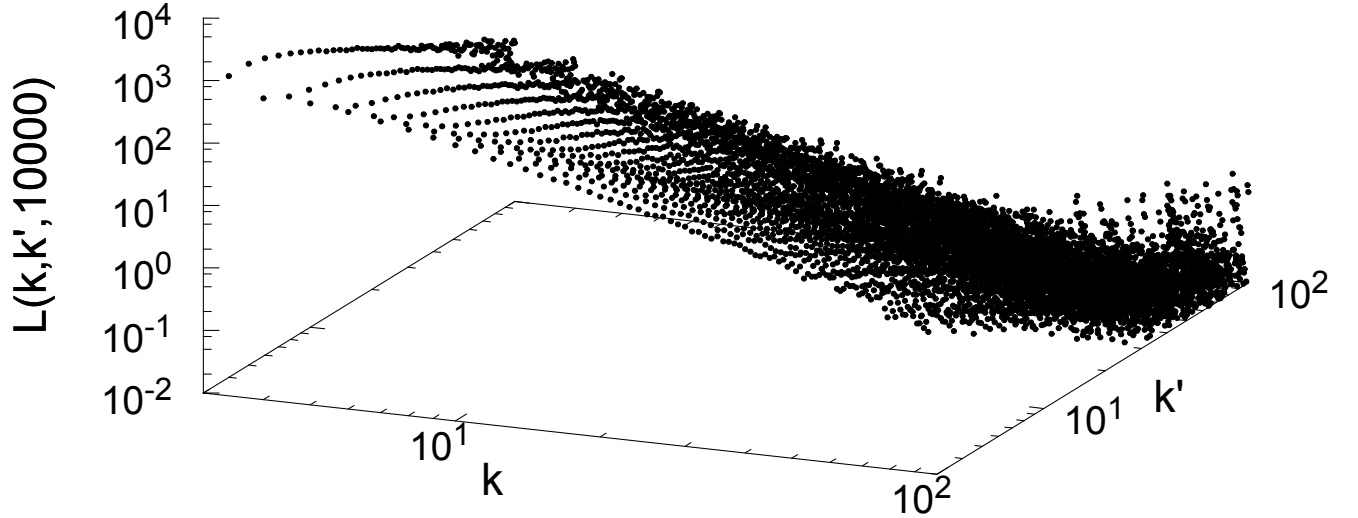


Figura 3.2: Número de ligações entre nós com grau k e k' para o modelo de Barabási-Albert obtidos através da média sobre 10000 simulações para $t = 10000$.

Para facilitar a comparação de $L(k, k', t)$ para tempos diferentes optamos por selecionar dois valores de k' para visualizarmos na forma de uma curva, ao invés de uma superfície. Na figura 3.3 temos $L(k, k', t)$ para três valores de t . Para cada tempo selecionamos os valores de $k' = 10$ e $k' = 50$, portanto teremos duas curvas para cada valor de t .

Para visualizar $\Delta(k, k', k'', t)$ optamos por construir uma função que calcula a densidade de triângulos sobre um octante da superfície esférica localizada no espaço gerado por k , k' e k'' . Definimos, então r_k dado por

$$r_k = \sqrt{k^2 + k'^2 + k''^2}, \quad (3.1)$$

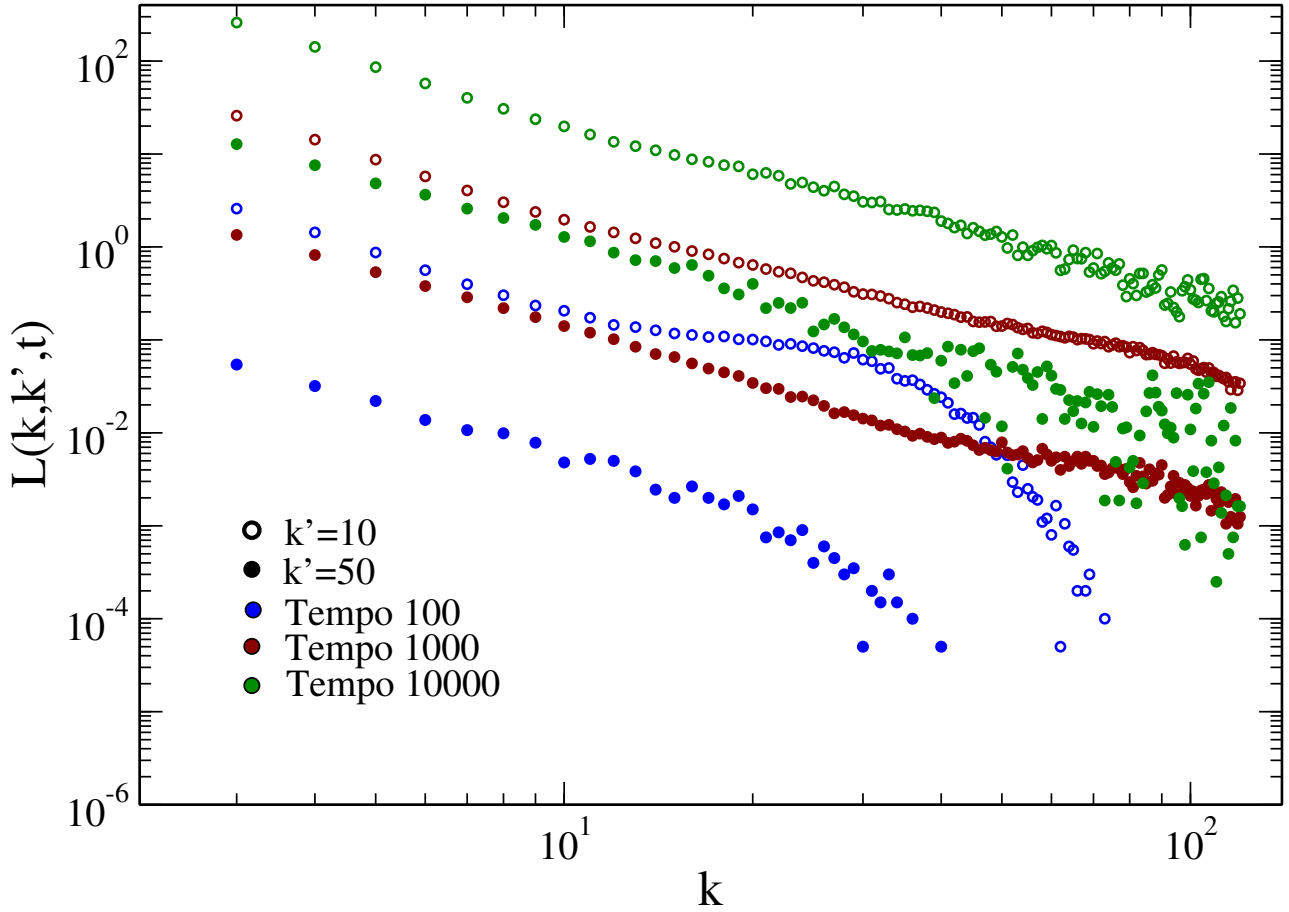


Figura 3.3: Número de ligações entre nós com grau k e k' para o modelo de Barabási-Albert obtidos através da média sobre 10000 simulações para três valores de t , com valores fixos de k' . Os pontos vazados representam $k' = 10$ e os sólidos $k' = 50$. Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$.

e então

$$\rho(r, t) = \delta_{r, r_k} \Delta(k, k', k'', t) \frac{2}{\pi r^2}, \quad (3.2)$$

onde δ_{r, r_k} é a delta de Kronecker, que vale um quando r for igual à r_k e zero se forem diferentes.

A figura 3.4 apresenta o valor de $\rho(r)$ para três tempos distintos.

3.1 Modelos de adesão mais gerais

O modelo de Barabási-Albert na verdade é similar ao proposto por Price em 1965. Price tentava reproduzir as redes de citações [34] e descreveu o princípio da vantagem cumulativa, onde um

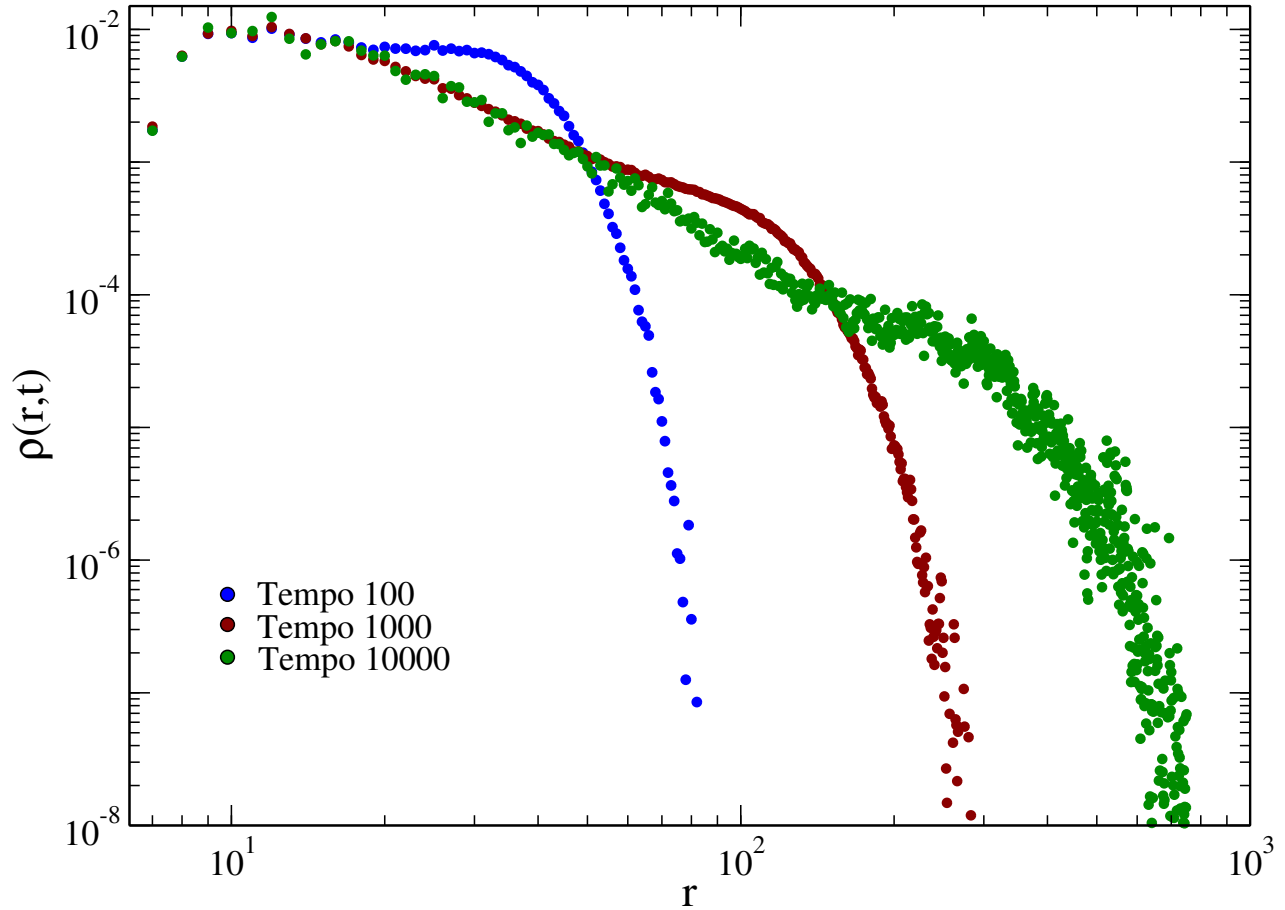


Figura 3.4: Densidade superficial de triângulos para o modelo de Barabási-Albert em função do raio r da esfera no espaço gerado por k, k' e k'' para três valores de t . Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$.

artigo com mais citações tem maior chance de obter uma nova citação. Como é uma rede de citações, as ligações são direcionadas do nó mais novo ao mais antigo, e o modelo também da origem a uma rede livre de escala. Entretanto, apesar da similaridade entre os modelos, o princípio da adesão preferencial só ganhou popularidade após a sua publicação por Barabási, quando recebeu várias generalizações.

Foram propostas variações na regra de escolha do nó que recebe as ligações. Dorogovtsev *et. al.* [21] propôs uma probabilidade de adesão proporcional ao grau mais uma constante, já no modelo de Krapivsky *et. al.* [35,36], essa probabilidade é dada por uma potência do grau. Além de alguma função do grau, também é possível considerar a idade do nó, um *fitness* [37], ou então adicionar alguma outra quantidade, como no modelo de Papadopoulos *et. al.* [38] que

considera a similaridade. Neste modelo cada nó é distribuído em um espaço circular em uma posição definida por um ângulo aleatório θ e pelo raio s dado pelo tempo de adição do nó. Cada novo nó se conecta com m nós com o menor produto $\Delta s \Delta \theta$, onde Δs é a distância no tempo, medida pelo número de nós adicionados entre os dois nós, e $\Delta \theta$ é a distância angular. Assim cada nó se conecta aos nós com θ similar.

Variações no número de ligações m de um novo nó também foram consideradas. Huberman *et. al.* [39] e Goh *et. al.* [40] discutem o efeito de uma flutuação em m , enquanto Dorogovtsev *et. al.* [41] considera uma rede onde $m = n^\alpha$, ou seja, o crescimento é acelerado. Além disso, diversos modelos foram propostos permitindo adição ou deleção de ligações existentes [12, 21, 36, 42–46].

Além dessas variações, também foram propostos modelos tentando obter uma clusterização maior que a apresentada pelo modelo de Barabási-Albert, procurando refletir as redes de associação proteína-proteína. Por exemplo o modelo proposto por Holme e Kim [47] em que, uma vez que um primeiro nó tenha sido selecionado para receber uma ligação do novo nó, as ligações restantes devem ser distribuídas pelos vizinhos daquele nó, quando possível, com o objetivo de aumentar o número de triângulos.

3.2 Soluções Analíticas

Diversas redes reais podem ser aproximadas por uma lei de potência e, uma vez que o modelo de Barabási-Albert reproduz essa característica através de uma regra simples, diversas tentativas foram feitas para caracterizar analiticamente esse modelo [10, 13]. Neste capítulo mostraremos alguns resultados analíticos encontrados na literatura para algumas grandezas que caracterizam o modelo de Barabási-Albert no limite assintótico. Para obter essas quantidades é considerado o mecanismo de crescimento dessa rede: a cada passo de tempo, um nó i é adicionado no tempo t_i . Além disso, em um determinado tempo t a rede terá um número de nós igual a $t + N_0$ sendo N_0 o número de nós iniciais, então para tempos grandes o tamanho da rede pode ser aproximado por t .

É possível obter uma solução para a distribuição de grau do modelo de Barabási-Albert utili-

zando uma aproximação de campo médio [9]. Vamos definir Π_i como a probabilidade do nó i ser escolhido para receber uma ligação. No modelo de adesão preferencial essa probabilidade é dada por

$$\Pi_i = \frac{k_i}{\sum_{j=1}^n k_j}, \quad (3.3)$$

que pode ser escrita como função do grau, e não do nó:

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^n k_j}. \quad (3.4)$$

O aumento de k_i é proporcional à probabilidade desse nó receber uma nova ligação $\Pi(k_i)$, então, considerando k e t contínuos, podemos escrever

$$\frac{\partial k_i}{\partial t} = A\Pi(k_i) = \frac{Ak_i}{\sum_{j=1}^{m_0+t-1} k_j}. \quad (3.5)$$

O somatório em (3.5) é efetuado até $m_0 + t - 1$ pois, como a cada Δt é adicionado um nó, dada uma rede com m_0 nós iniciais, este será o número de nós no tempo $t - 1$. Além disso, a cada $\Delta t = 1$ são adicionadas m ligações, e o grau de m nós aumenta em um. Portanto temos $A = m$ e $\sum_{j=1}^{m_0+t-1} k_j \approx 2m(t - 1)$, onde desprezamos as conexões iniciais, caso existam. A equação (3.5) fica, então

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \quad (3.6)$$

Cada nó i surge no tempo t_i com m ligações, então, com essa condição inicial, a solução é

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{0.5}. \quad (3.7)$$

Usando a equação 3.7 podemos definir a probabilidade $P(k_i(t) < k)$ de um nó ter grau menor

que k como

$$P(k_i(t) < k) = P\left(t_i > \frac{m^2 t}{k^2}\right). \quad (3.8)$$

Como a cada passo Δt adicionamos um nó, a probabilidade do nó i ter sido adicionado no tempo t_i é

$$P(t_i) = \frac{1}{m_0 + t_i}, \quad (3.9)$$

então podemos reescrever

$$P\left(t_i > \frac{m^2 t}{k^2}\right) = 1 - P\left(t_i \leq \frac{m^2 t}{k^2}\right) = 1 - \frac{m^2 t}{k^2(m_0 + t)}. \quad (3.10)$$

A distribuição de grau, então, pode ser obtida:

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{k^3(m_0 + t)}, \quad (3.11)$$

que reproduz corretamente o expoente $\gamma = 3$.

Dorogovtsev *et. al.* [21] propõem uma equação mestra para um modelo semelhante, com a diferença que as ligações são direcionadas. É definido q_i como o número de ligações apontando para o nó i . A probabilidade de um nó i receber uma ligação é definida como

$$A_i = A + q_i. \quad (3.12)$$

A_i é chamada atração do nó i , e A é a probabilidade de um novo nó receber uma ligação. A cada passo de tempo surgem m ligações, mas o modelo não deixa explícito a origem dessas ligações. Podem ser do novo nó, de nós anteriores, ou mesmo externas à rede. A atratividade total A_Σ da rede é

$$A_\Sigma = (m + A)t = (1 + a)t/m, \quad (3.13)$$

então a probabilidade de um nó i receber uma ligação é dada por A_i/A_Σ . A solução também permite que mais de uma ligação sejam adicionadas ao mesmo nó. Então, a probabilidade do nó i receber l ligações é dada por

$$\Pi_s^{ml} = \binom{m}{l} \left(\frac{A_i}{A_\Sigma} \right)^l \left(1 - \frac{A_i}{A_\Sigma} \right)^{m-l}. \quad (3.14)$$

A equação mestra obtida por Dorogovtsev *et. al.* para a distribuição de grau de um nó i é:

$$P(q, i, t+1) = \sum_{l=0}^m \Pi_s^{ml} = \sum_{l=0}^m \binom{m}{l} \left(\frac{q-l-am}{(1+a)m} \right)^l \left(1 - \frac{q-l-am}{(1+a)m} \right)^{m-l} P(q-l, i, t). \quad (3.15)$$

A distribuição de grau para toda a rede $P(q, t)$ é dada por $P(q, t) = \sum_{i=1}^t P(q, i, t)$. Assumindo que $P(q) = P(q, t \rightarrow \infty)$, é possível obter

$$P(q) = (1+a) \frac{\Gamma[(m+1)a+1]}{\Gamma[ma]} \frac{\Gamma[q+ma]}{\Gamma[q+2+(m+1)a]}, \quad (3.16)$$

onde a função $\Gamma[a]$ é a extensão da função fatorial definida como

$$\Gamma[a] = \int_0^\infty x^{a-1} e^{-x} dx. \quad (3.17)$$

Caso $a = 1$, ou seja, $A = q + m$, temos um caso equivalente ao modelo de Barabási-Albert no qual os nós surgem com m ligações, e a distribuição torna-se

$$P(q) = \frac{2m(m+1)}{(q+m)(q+m+1)(q+m+2)}, \quad (3.18)$$

resultado que é coerente com o resultado conhecido da lei de potência com $\gamma = 3$. Além disso, com a variação do parâmetro a os autores encontram distribuições de grau que seguem leis de potência com expoentes dados por $\gamma = 2 + a$.

Krapivsky *et. al.* [23] obtêm resultados similares utilizando a equação de taxa de reação de uma reação química, com base na concentração de seus reagentes. Uma aproximação baseada nessa equação foi utilizada para obter a distribuição de grau para um modelo de adesão onde os

novos nós são adicionados com uma ligação e a probabilidade de um nó de grau k receber uma ligação é dada por

$$A_k = k^\alpha . \quad (3.19)$$

Para $\alpha = 1$ os autores encontram,

$$P(k) = \frac{4}{k(k+1)(k+2)} , \quad (3.20)$$

além de demonstrar que o número de ligações por grau cresce linearmente. Além disso, explorando o parâmetro α mostra-se que, para $0 < \alpha < 1$, a distribuição segue uma *stretched exponential* e, para $\alpha > 1$, um único nó se liga a praticamente todos os outros nós da rede. Usando o mesmo método da equação de taxa de reação [35] os autores obtiveram a fração de nós de grau k ligados com nós de grau k' , dada por

$$l_{k,k'} = \frac{4(k' - 1)}{k(k+1)(k+k')(k+k'+1)(k+k'+2)} + \frac{12(k' - 1)}{k(k+k'-1)(k+k')(k+k'+1)(k+k'+2)} . \quad (3.21)$$

Essa equação não é simétrica pois as ligações são direcionadas.

Todos os resultados apresentados acima foram obtidos no limite $t \rightarrow \infty$, o único trabalho que conhecemos que obtêm resultado da distribuição de grau no tempo finito foi publicado por Fotouhi e Rabat [24]. Nesse trabalho os autores escrevem a equação mestra para $N(k, t)$, e, tomando o limite de tempo contínuo obtêm

$$\frac{\partial N(k, t)}{\partial t} = \frac{\alpha m}{\sum_k k N(k, t)} ((k-1)N(k-1, t) - kN(k, t)) + \alpha \delta_{k,m} , \quad (3.22)$$

onde α é o número de nós adicionados a cada passo de tempo. Os autores então utilizam a transformada Z , com a função geradora $\Phi(z, t) = \sum_k z^{-k} N(k, t)$, para obter a seguinte equação

diferencial

$$\frac{\partial \Phi(z, t)}{\partial t} - \frac{\alpha m(z-1)}{N_0 k_0 + 2\alpha m t} \frac{\partial \Phi(z, t)}{\partial z} = \alpha z^{-b}. \quad (3.23)$$

Esta equação é resolvida através do método da equação característica e, após efetuar a transformação inversa os autores obtêm

$$N(k, t) = \sum_{l=1}^k N(l, 0)(1-c)^l c^{k-l} \binom{k-1}{l-1} + \frac{N_0 k_0 + 2\alpha m t}{m} \frac{m(m+1)}{k(k+1)(k+2)} u(k-m) - \frac{N_0 k_0}{m} \sum_{l=m}^k \frac{m(m+1)}{l(l+1)(l+2)} (1-c)^l c^{k-l} \binom{k-1}{l-1}, \quad (3.24)$$

onde $u(k-m)$ é a função de Heavside e c é definido como

$$c = 1 - \sqrt{\frac{N_0 k_0}{N_0 k_0 + 2\alpha m t}}. \quad (3.25)$$

Os autores demonstram que o resultado concorda com a literatura no limite $t \rightarrow \infty$, mas a expressão é válida no intervalo $1 \leq k \leq N_0$.

A clusterização média por grau foi obtida por Fronczak *et. al.* [26] para um modelo de Barabási-Albert considerando a evolução da clusterização C_i de um nó específico i . Esta clusterização depende apenas do grau k_i e da vizinhança do nó. A clusterização C_i sofre modificação através de dois mecanismos: quando recebe uma ligação, já que a clusterização é dependente do grau (eq. 2.9), e quando um novo nó se ligar a ele e a x vizinhos, formando x triângulos. A evolução de C_i é dada por

$$\frac{dC_i}{dt} = \sum_{x=0}^{m-1} \tilde{p}_{ix} \Delta C_{ix}, \quad (3.26)$$

onde ΔC_{ix} é a variação da clusterização quando o novo nó se liga ao nó i e a x vizinhos, dada por

$$\Delta C_{ix} = \frac{2(\Delta_i + x)}{k_i(k_i + 1)} - \frac{2\Delta_i}{k_i(k_i - 1)} = -\frac{2C_i}{k + 1} + \frac{2x}{k_i(k_i + 1)}, \quad (3.27)$$

e \tilde{p}_{ix} é a probabilidade dessas ligações ocorrerem, dada por

$$\tilde{p}_{ix} = \frac{k_i}{2t} \binom{m-1}{x} \left(\frac{k_i \ln t}{4t} \right)^x \left(1 - \frac{k_i \ln t}{4t} \right)^{m-1-x}. \quad (3.28)$$

Lembrando que Δ_i é o número de triângulos que têm o nó i como vértice, e sua relação com a clusterização é dada por $C_i = 2\Delta_i/k_i(k_i - 1)$. Resolvendo a equação diferencial (3.26) sujeita à condição que no tempo t_i do surgimento do nó i sua clusterização é dada por

$$C_i(t_i) = \frac{m^2}{8(m-1)} \frac{(\ln t_i)^2}{t_i}, \quad (3.29)$$

os autores obtiveram

$$C_i(t) = \frac{(m-1)}{8(\sqrt{t} + \sqrt{t_i}/m)^2} \left[(\ln t)^2 \frac{4m}{(m-1)^2} (\ln t_i)^2 \right]. \quad (3.30)$$

A clusterização da rede, definida como $C = \frac{\sum_i c_i}{t}$, sendo t o tamanho da rede, é dada por

$$C_i(t) = \frac{m^2(m-1)^2}{4(m-1)} \left[(\ln t)^2 \frac{4m}{(m-1)^2} (\ln t_i)^2 \right]. \quad (3.31)$$

Bianconi e Capocci obtiveram $\langle N_h(t) \rangle$ o número médio de ciclos de qualquer tamanho h para o modelo de Barabási-Albert. Em particular no caso de ciclos de tamanho três, ou triângulos, esse resultado permite obter a clusterização média total da rede, mas não a clusterização média por grau. Os autores obtiveram esse resultado considerando as probabilidades de dois nós estarem conectados por um caminho auto-evitante, e de receberem, ambos, uma nova ligação. Desta forma eles obtém a seguinte equação diferencial

$$\frac{\partial \langle N_h(t) \rangle}{\partial t} = \frac{m}{2t} (h-1) \langle N_{h-1}(t) \rangle. \quad (3.32)$$

Considerando apenas as probabilidades de ligação é possível escrever

$$\langle N_3(t) \rangle = \frac{1}{3!} \left[\frac{m}{2} \log(t) \right]^3. \quad (3.33)$$

Os autores resolvem, então, a equação diferencial iterativamente e encontram

$$\langle N_h(t) \rangle = \left[\frac{m}{2} \log(t) \right]^h \left[1 + O\left(\frac{2}{m \log(t)}\right) \right] \approx \left[\frac{m}{2} \log(t) \right]^h . \quad (3.34)$$

Com exceção da solução obtida por Foutouhi e Rabat [24] e do número total de ciclos todas as soluções analíticas apresentadas são válidas no limite $t \rightarrow \infty$, inclusive a clusterização, onde uma aproximação foi utilizada para obter a clusterização inicial do nó i . Esta aproximação é válida quando se está interessado em modelar redes grandes, onde o modelo de Barabási-Albert se encontra na vizinhança do estado assintótico. Apenas para ilustrar esse ponto, Barabási *et al.* [9] afirmam que a distribuição de grau independe do tempo baseando esta afirmação no estado final de simulações com 10^5 , 1.5×10^5 e 2×10^5 nós. Estes tamanhos de redes são muito superiores a várias redes reais [13, 18]. Outra consideração de tamanho finito é apresentada por Dorogovtsev e Mendes [22], que tratam de obter o grau máximo k_{cut} para o qual a aproximação de lei de potência vale. O valor obtido é

$$k_{cut} = t^{\frac{1}{\gamma-1}} , \quad (3.35)$$

que para o modelo de Barabási-Albert é \sqrt{t} . Podemos ver na figura 3.1 que, de fato, para valores de k superiores a esse limite ocorre um desvio de uma lei de potência devido, segundo este resultado analítico, aos efeitos de tamanho finito.

Os resultados analíticos para o modelo de Barabási-Albert apresentam, em sua maioria, a distribuição de grau no limite $t \rightarrow \infty$. Como procuramos a relevância dos efeitos de tamanho finito é importante obter soluções analíticas que contenham a evolução temporal, para que possamos comparar estes resultados com redes de tamanho pequeno. Vamos então procurar descrever o número de nós em função do grau, relacionado com a distribuição de grau, já conhecida, as ligações e os triângulos em função dos graus dos nós, ambos desconhecidos na literatura. Com esses resultados é possível descrever adequadamente o modelo de Barabási-Albert para redes pequenas e também investigar a extensão dos efeitos de tamanho finito.

Capítulo 4

Equações mestras e obtenção das soluções analíticas

Neste capítulo apresentaremos um método que permite obter as equações de evolução das quantidades apresentadas no capítulo 2 para o modelo de Barabási-Albert, através da construção das equações mestras que reproduzem o mecanismo de crescimento do modelo. Consideraremos, por simplicidade, que o estado inicial da rede consiste em $N_0 = m + 2$ nós todos ligados entre si, portanto com $k_0 = m + 1$ ligações. O número de ligações dessa rede é dado por $N_0 k_0 / 2$. A generalização dos resultados para diferentes estados iniciais é simples de ser obtida.

4.1 Equações Mestras

Como uma rede pode ser descrita pelo seu conjunto de nós e ligações, é natural supor que a regra de crescimento do modelo contenha toda a informação necessária para escrever as equações de evolução da rede. Vamos construir as equações de evolução para o modelo de Barabási-Albert a partir da probabilidade de adesão proporcional ao grau, portanto $\Pi(k) = kN(k, t) / \sum k'N(k', t)$. O denominador nesta expressão contabiliza o dobro do número total de ligações, pois cada ligação é contada duas vezes. Como, no modelo de Barabási-Albert, a cada passo de tempo m ligações

são adicionadas, o número total de ligações é dado por $mt + N_0k_0/2$. Então

$$\Pi(k) = \frac{kN(k, t)}{2mt + N_0k_0}. \quad (4.1)$$

Para obter o número de ligações com o tempo $N(t)$, basta considerar que a cada passo de tempo um novo nó é adicionado. Portanto, o número de nós em um certo tempo t é o número de nós adicionados ao longo da evolução mais o número de nós iniciais e, então, temos

$$N(t) = t + N_0. \quad (4.2)$$

A equação mestra que descreve a evolução temporal de $N(k, t)$, ou seja, o valor esperado para o número de nós com grau k no tempo t , é dada por

$$N(k, t + 1) = N(k, t) + \delta_{k,m} + m\Pi(k - 1) - m\Pi(k). \quad (4.3)$$

Cada termo dessa equação representa uma parte da dinâmica do modelo responsável pela alteração do valor de $N(k, t)$. O segundo termo no lado direito da equação descreve o surgimento de um nó com grau m a cada passo de tempo, uma vez que se $k = m$ valor de $N(m, t)$ aumentará em um devido a esse termo. Quando um nó de grau k receber uma ligação passará a ser de grau $k + 1$ e $N(k, t)$ diminuirá. A probabilidade de isso acontecer é dada pelo produto entre a probabilidade $\Pi(k)$ deste nó receber uma ligação, o número m de ligações adicionadas e o número $N(k, t)$ de nós com este grau. Esta diminuição é representada pelo último termo na equação. O terceiro termo do lado direito descreve o equivalente aumento de $N(k, t)$ quando um nó de grau $k - 1$ recebe uma ligação. Substituindo a probabilidade $\Pi(k)$ dada pela equação 4.1 temos

$$N(k, t + 1) = N(k, t) + \delta_{k,m} + m \frac{(k - 1)N(k - 1, t)}{(2mt + N_0k_0)} - m \frac{kN(k, t)}{(2mt + N_0k_0)}. \quad (4.4)$$

Podemos interpretar essa equação de forma que o termo $\delta_{k,m}$ representaria a fonte de nós com grau m e os dois últimos termos descrevem um fluxo de nós de grau $k - 1$ para grau k e de grau k para $k + 1$. É interessante notar que na equação 4.3 poderíamos substituir uma probabilidade

$\Pi(k)$ diferente e teríamos a equação mestra para outro modelo de adesão.

De forma análoga à utilizada para obter a equação (4.4) podemos obter a equação para o número de ligações $L(k, k', t)$ entre nós de graus k e k' no tempo t

$$\begin{aligned}
L(k, k', t + 1) &= L(k, k', t) + \delta_{k,m} \frac{m}{2} \Pi(k' - 1) N(k' - 1, t) + \delta_{k',m} \frac{m}{2} \Pi(k - 1) N(k - 1, t) \\
&\quad + m \Pi(k - 1) L(k - 1, k', t) + m \Pi(k' - 1) L(k, k' - 1, t) \\
&\quad - m [\Pi(k) + \Pi(k')] L(k, k', t) .
\end{aligned} \tag{4.5}$$

Quando o novo nó se liga a um nó de grau $k' - 1$ este passa a ser de grau k e $L(m, k', t)$ recebe meia ligação para evitar que cada ligação seja contabilizada duas vezes. A probabilidade dessa ligação acontecer é o produto entre o número m de ligações do nó novo, a probabilidade $\Pi(k' - 1)$ de um nó de grau $(k' - 1)$ receber essa ligação e o número $N(k' - 1, t)$ de nós com esse grau. Essa dinâmica é representada pelo segundo termo do lado direito da equação, e o terceiro representa o caso simétrico. Os outros termos da equação descrevem nós de grau $k - 1$ ($k' - 1$) recebendo uma ligação e passando ao grau k (k') ou quando nós de grau k (k') passam a ser de grau $k + 1$ ($k' + 1$). Novamente podemos substituir a probabilidade $\Pi(k)$ dada pela equação 4.1 e obter a equação

$$\begin{aligned}
L(k, k', t + 1) &= L(k, k', t) + \delta_{k,m} m \frac{(k' - 1) N(k' - 1, t)}{2(2mt + N_0 k_0)} + \delta_{k',m} m \frac{(k - 1) N(k - 1, t)}{2(2mt + N_0 k_0)} \\
&\quad + \frac{m(k - 1)}{(2mt + N_0 k_0)} L(k - 1, k', t) + \frac{m(k' - 1)}{(2mt + N_0 k_0)} L(k, k' - 1, t) \\
&\quad - \frac{m(k + k')}{(2mt + N_0 k_0)} L(k, k', t) .
\end{aligned} \tag{4.6}$$

Analisando as duas equações anteriores vemos que em ambas existem dois tipos de termos: os relativos à criação de ligações ou nós quando $k = m$; e os relativos ao fluxo de ligações ou nós do grau $k - 1$ para k quando um nó de grau $k - 1$ recebe uma das novas ligações. Além disso, por simetria na equação para o número de ligações, para cada termo em k existe um equivalente para k' . Seguindo o mesmo método utilizado para as equações anteriores podemos construir a equação para número de triângulos $\Delta(k, k', k'', t)$ com vértices em nós de graus k, k' e k'' no tempo t . Para

que surja um triângulo é necessário que o novo nó se ligue com dois nós de graus $k' - 1$ e $k'' - 1$ com uma ligação entre si. Então o termo de criação de triângulos é

$$\delta_{k,m} m \Pi(k' - 1)(m - 1) \Pi(k'' - 1) m(m - 1) L(k' - 1, k'' - 1, t), \quad (4.7)$$

com dois termos equivalentes quando $k' = m$ ou $k'' = m$. Os termos $m \Pi(k' - 1)$ e $(m - 1) \Pi(k'' - 1)$ referem-se à probabilidade de nós de graus k e k' serem sorteados, respectivamente, enquanto o termo $m(m - 1)$ é o número máximo de triângulos que podem surgir a cada passo de tempo. O termo de fluxo é semelhante aos das equações para $N(k, t)$ e $L(k, k', t)$, ou seja, quando um nó de grau $k - 1$ recebe um vizinho, passa a ser de grau k , e $\Delta(k, k', k'', t)$ recebe o número de triângulos relativos àquele nó. Então a equação para $\Delta(k, k', k'', t + 1)$, já utilizando a probabilidade dada pela equação 4.1, fica

$$\begin{aligned} \Delta(k, k', k'', t + 1) = & \Delta(k, k', k'', t) + \delta_{k,m} \frac{m(k' - 1)(m - 1)(k'' - 1)}{6(2mt + N_0 k_0)^2} m(m - 1) L(k' - 1, k'' - 1, t) \\ & + \delta_{k',m} \frac{m(k - 1)(m - 1)(k'' - 1)}{6(2mt + N_0 k_0)^2} m(m - 1) L(k - 1, k'' - 1, t) \\ & + \delta_{k'',m} \frac{m(k - 1)(m - 1)(k' - 1)}{6(2mt + N_0 k_0)^2} m(m - 1) L(k - 1, k' - 1, t) \\ & + \frac{m(k - 1)}{(2mt + N_0 k_0)} \Delta(k - 1, k', k'', t) + \frac{m(k' - 1)}{(2mt + N_0 k_0)} \Delta(k, k' - 1, k'', t) \\ & + \frac{m(k'' - 1)}{(2mt + N_0 k_0)} \Delta(k, k', k'' - 1, t) \\ & - \frac{m(k + k' + k'')}{(2mt + N_0 k_0)} \Delta(k, k', k'', t). \end{aligned} \quad (4.8)$$

4.2 Equação diferencial

Vamos escrever equações diferenciais associadas às equações mestras para o modelo de Barabási-Albert considerando que, para $t \gg 1$, podemos tomar o tempo como uma variável contínua

$$\frac{\partial N(k, t)}{\partial t} = \delta_{k,m} + m \frac{(k - 1)N(k - 1, t)}{(2mt + N_0 k_0)} - m \frac{kN(k, t)}{(2mt + N_0 k_0)}. \quad (4.9)$$

A equação (4.9) pode ser interpretada como um conjunto de equações diferenciais acopladas

da seguinte forma

$$\frac{\partial N(k, t)}{\partial t} + m \frac{kN(k, t)}{(2mt + N_0k_0)} = \delta_{k,m} + m \frac{(k-1)N(k-1, t)}{(2mt + N_0k_0)}, \quad (4.10)$$

já que a equação para o número de nós com grau k depende do grau $k-1$. O lado direito da equação representa a parte homogênea, que será idêntica qualquer que seja o grau k , somente a não-homogeneidade irá variar para cada uma das equações. Se utilizarmos, portanto, o método de funções de Green para resolver essas equações a função de Green será sempre a mesma, facilitando o processo.

Dada a rede inicial, ou seja, N_0 nós com grau $m+1$, e pelas características da evolução do modelo, sabemos que $N(m-1, t) = 0$ para qualquer tempo. Então podemos escrever a equação diferencial para o grau m

$$\frac{\partial N(m, t)}{\partial t} + m \frac{mN(m, t)}{(2mt + N_0k_0)} = 1, \quad (4.11)$$

sujeita à condição inicial $N(m, 0) = 0$. Esta é uma equação diferencial não homogênea em t , que pode ser resolvida usando funções de Green (para mais detalhes da obtenção da solução, ver apêndice A), resultando em

$$N(m, t) = \frac{N_0k_0 + 2mt}{m(m+2)} - \frac{(N_0k_0)^{1+m/2}}{m(m+2)(N_0k_0 + 2mt)^{m/2}}. \quad (4.12)$$

Quando $t \gg N_0k_0$ o segundo termo se torna desprezível, e temos como solução

$$N(m, t) = \frac{N_0k_0 + 2mt}{m(m+2)}. \quad (4.13)$$

Utilizando a solução para $N(m, t)$, 4.13, podemos construir a equação para $k = m+1$

$$\frac{\partial N(m+1, t)}{\partial t} + m \frac{(m+1)N(m+1, t)}{(2mt + N_0k_0)} = m \frac{mN(m, t)}{(2mt + N_0k_0)}. \quad (4.14)$$

Novamente temos uma equação diferencial não homogênea, que resolvemos da mesma forma da

anterior. Aplicando esse método é possível obter $N(k, t)$ para qualquer grau, e então obter a regra geral

$$N(k, t) = \frac{(N_0 k_0 + 2mt)(m + 1)}{k(k + 1)(k + 2)}. \quad (4.15)$$

Essa equação é coerente com os resultados obtidos por Dorogovtsev *et. al.* [21] e Krapivsky *et. al.* [23]. Para comparar as soluções escrevemos

$$P(k, t) = \frac{N(k, t)}{N(t)} = \frac{N(k, t)}{t} = \frac{(N_0 k_0/t + 2m)(m + 1)}{k(k + 1)(k + 2)}, \quad (4.16)$$

e no limite $t \rightarrow \infty$ temos

$$P(k) = \frac{2m(m + 1)}{k(k + 1)(k + 2)}, \quad (4.17)$$

a mesma solução obtida pelos autores, que consideraram o limite $t \rightarrow \infty$. Para obtermos essa solução apenas assumimos assumimos que $t \gg N_0 k_0$ e desprezamos os termos que caem rapidamente com t , portanto este resultado é válido longe do limite $t \rightarrow \infty$.

Resolvendo as equações para $L(k, k', t)$ da mesma forma, utilizando $L(m, m, t)$ e $N(m, t)$ para obtermos $L(m + 1, m, t)$ e $L(m, m + 1, t)$, chegamos à expressões indicando que $L(k, k', t)$ segue a seguinte recorrência

$$L(k, k', t) = \begin{cases} \frac{(k-1)L(k-1, k', t) + (k'-1)L(k, k'-1, t)}{k+k'+2} & k, k' \neq m \\ \frac{(k-1)L(k-1, k', t) + (k-1)N(k-1, t)/2}{k+m+2} & k \neq m, k' = m \\ 0 & k, k' = m \end{cases} \quad (4.18)$$

Essa recorrência reflete indiretamente o mecanismo de crescimento da rede, uma vez que $L(k, k', t)$ depende de $L(k - 1, k', t)$ e $L(k, k' - 1, t)$ que dependem cada um de $L(k - 2, k', t)$ e $L(k, k' - 2, t)$ e ambos de $L(k - 1, k' - 1, t)$. A figura 4.1 indica a dependência do termo $L(k, k', t)$ de alguns dos termos de menor grau. As flechas apontam dos termos de menor grau para os termos de maior grau que eles compõem, onde as vermelhas indicam todos os necessários para compor

$L(k-1, k', t)$, e as azuis os que compõem $L(k, k' - 1, t)$. Observe que o número setas saindo de cada termo indica o número de vezes que esse termo apareceria caso escrevêssemos todas as componentes de recorrência. Essa forma de construção remete ao triângulo de Pascal e, de fato, se abirmos a recorrência, os coeficientes de cada um dos termos que compõem $L(k, k', t)$ será dado pelo binômio de Newton. A solução para $L(k, k', t)$ é dada por

$$L(k, k', t) = \left[\sum_{i=1}^{k'-m} \binom{k-m+i-1}{i-1} \frac{(m+1)(2mt+N_0k_0)}{2(k'+1-i)(k'+2-i)} \frac{(k-1)!}{(m-1)!} \frac{(k'-1)!}{(k'-i)!} \frac{(k'+m-i+2)!}{(k+k'+2)!} + \sum_{i=1}^{k-m} \binom{k'-m+i-1}{i-1} \frac{(m+1)(2mt+N_0k_0)}{2(k+1-i)(k+2-i)} \frac{(k'-1)!}{(m-1)!} \frac{(k-1)!}{(k-i)!} \frac{(k+m-i+2)!}{(k+k'+2)!} \right]. \quad (4.19)$$

Esta solução é dependente do tempo, crescendo linearmente com este, da mesma forma que $N(k, t)$. Isso ocorre devido ao termo com $N(k, t)$ na equação master que introduz a dependência temporal.

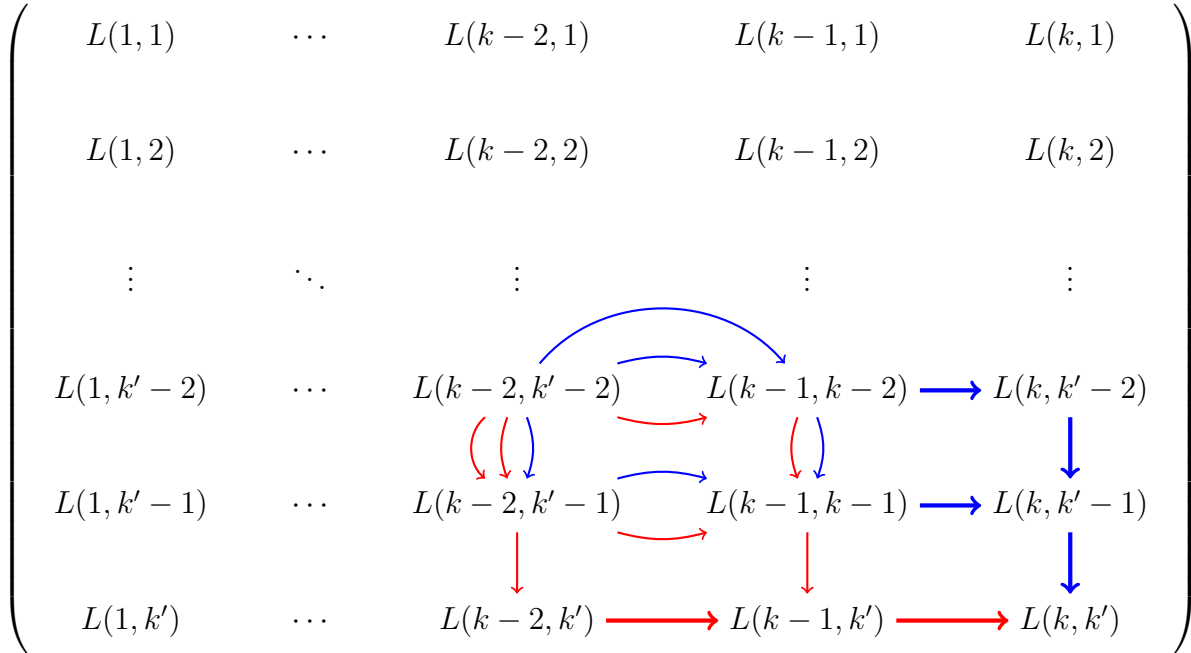


Figura 4.1: Figura esquemática que representa a recorrência, apresentando a dependência de $L(k, k', t)$ dos termos de menor grau. As setas apontam dos termos de grau menor para o maior, onde as vermelhas indicam todos os necessários para compor $L(k-1, k', t)$, e as azuis os que compõem $L(k, k' - 1, t)$.

Através do mesmo método, resolvendo as equações diferenciais para $\Delta(k, k', k'', t)$, encontra-

mos que $\Delta(k, k', k'', t)$ satisfaz

$$\Delta(k, k', k'', t) = \begin{cases} \frac{(k-1)\Delta(k-1, k', k'', t) + (k'-1)\Delta(k, k'-1, k'', t) + (k''-1)\Delta(k, k', k''-1, t)}{k+k'+k''} & k, k', k'' \neq m \\ \frac{((k-1)\Delta(k-1, k', k'', t) + (k'-1)\Delta(k, k'-1, k'', t) + m(m-1)^2(k-1)(k'-1)L(k-1, k'-1, t)/6}{k+k'+m} & k, k' \neq m, k'' = m \\ \frac{((k-1)\Delta(k-1, k', k'', t))}{k+m+m} & k \neq m, k', k'' = m \\ 0 & k, k', k'' = m \end{cases} \quad (4.20)$$

Assim como os coeficientes dos termos que compõem $L(k, k', t)$ seguem o estabelecido pelo triângulo de Pascal, neste caso os coeficientes de $\Delta(k, k', k'', t)$ seguem o tetraedro de Pascal. Resolvendo essa recorrência obtemos

$$\begin{aligned} \Delta(k, k', k'', t) = & \left[\sum_{i=0}^{k'-m} \sum_{j=0}^{k''-m} \frac{(k-m+i+j)!}{(k-m-2)!i!j!} \frac{(k-1)!}{(m-1)!} \frac{1}{6} m(m-1)^2 \frac{(k'-1)!}{(k'-i-2)!} \frac{(k''-1)!}{(k''-j-2)!} \frac{(m+k'-i+k''-j)!}{(k+k'+k'')!} \frac{L(k'-i, k''-j, t)}{(2mt+N_0k_0)} + \right. \\ & \sum_{i=0}^{k-m} \sum_{j=0}^{k''-m} \frac{(k'-m+i+j)!}{(k'-m-2)!i!j!} \frac{(k'-1)!}{(m-1)!} \frac{1}{6} m(m-1)^2 \frac{(k-1)!}{(k-i-2)!} \frac{(k''-1)!}{(k''-j-2)!} \frac{(m+k-i+k''-j)!}{(k+k'+k'')!} \frac{L(k-i, k''-j, t)}{(2mt+N_0k_0)} + \\ & \left. \sum_{i=0}^{k-m} \sum_{j=0}^{k'-m} \frac{(k''-m+i+j)!}{(k''-m-2)!i!j!} \frac{(k''-1)!}{(m-1)!} \frac{1}{6} m(m-1)^2 \frac{(k-1)!}{(k-i-2)!} \frac{(k'-1)!}{(k'-j-2)!} \frac{(m+k-i+k'-j)!}{(k+k'+k'')!} \frac{L(k-i, k'-j, t)}{(2mt+N_0k_0)} \right] \quad (4.21) \end{aligned}$$

Observe que esta solução é independente do tempo, pois cada termo $L(k, k', t)$ que deveria introduzir a dependência temporal é dividida pelo termo $2mt + N_0k_0$. Este resultado é contraintuitivo, já que é a solução para um modelo de redes em crescimento. Esta discussão será retomada no próximo capítulo quando iremos comparar as soluções analíticas com os resultados de simulação e de iteração numérica das equações mestras.

Todas as soluções são obtidas como dependentes do tempo, ao contrário das soluções obtidas por outros autores, válidas apenas no limite $t \rightarrow \infty$. Isso significa que essas equações são válidas para tamanho finito, com a única restrição de $t \gg N_0k_0$. Entretanto, isso não é preocupante, uma vez que a maioria das simulações são conduzidas com N_0 e k_0 muito pequenos, da ordem da unidade ou dezena. Os resultados de Krapivsky *et. al.* [23], por exemplo, são obtidos com $N_0 = k_0 = 1$.

Capítulo 5

Resultados

No capítulo anterior obtivemos as equações mestras para o modelo de Barabási-Albert, resolvemos as equações diferenciais associadas, e obtivemos as equações de evolução das quantias $N(k, t)$, $L(k, k', t)$, e $\Delta(k, k', k'', t)$. Para avaliar os resultados obtidos iremos comparar a iteração da equação mestra e a solução analítica com a média de simulações do modelo de Barabási-Albert. Foram efetuadas médias sobre 10^4 simulações. As iterações para as equações mestras foram efetuadas em um notebook com processador Intel i5, com 16 gigabytes de memória RAM, levando cerca de 6 horas para serem efetuadas. As simulações das redes foram efetuadas em um computador com processador Intel i7 e 8 gigabytes de memória do cluster Ada do Instituto de Física da UFRGS. As dez mil amostras levam cerca de uma semana. Todas os programas foram escritos em Fortran 90. A primeira até aproximadamente o grau $k_{cut} = \sqrt{t} = \sqrt{10000} = 100$, obtido por Dorogovtsev e Mendes [22] como o limite para o qual a rede obedece uma lei de potência. A interpretação de k_{cut} é que apenas um nó em média teria grau $k > k_{cut}$. A segunda figura apresenta resultados para graus maiores para observarmos o desvio do comportamento de lei de potência devido ao efeito de tamanho finito.

Na figura 5.1 temos o resultado para $N(k, t)$ obtido através de simulação, iteração, e da solução analítica para três tempos distintos. Podemos observar que para k pequeno as três curvas coincidem. Para valores de $k > \sqrt{t}$ começa a ocorrer um desvio entre as simulações, iterações e soluções. Isto se deve a efeitos de tamanho finito e era esperado devido ao resultado

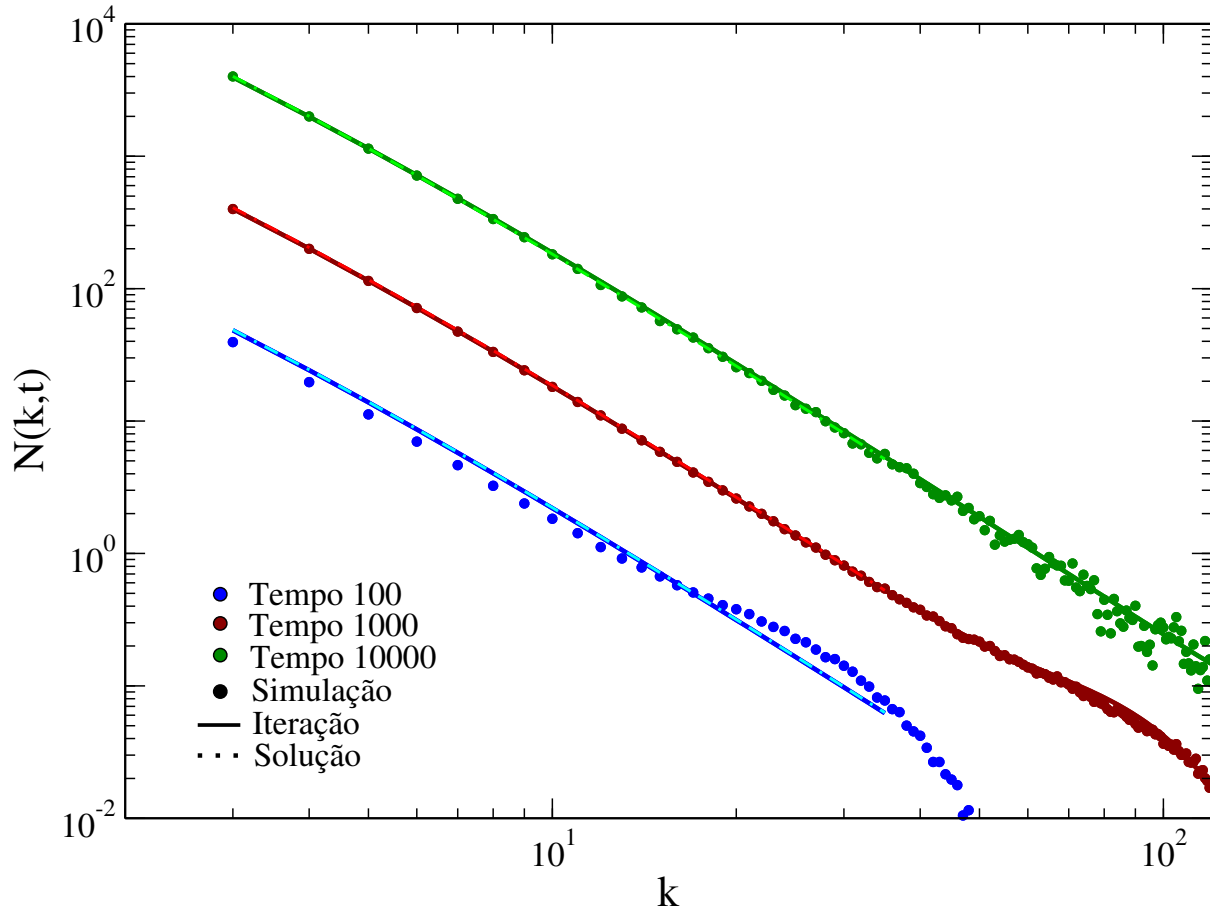


Figura 5.1: Número de nós com grau k obtidos através de simulação, iteração, e da solução analítica para $t = 100$, $t = 1000$ e $t = 10000$. Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$. As linhas contínuas apresentam os resultados de iteração da equação mestra e as tracejadas o resultado obtido analiticamente, ambas seguindo o código de cores estabelecido para os dados de simulação

de Dorogovtsev e Mendes [22]. Para visualizarmos melhor o desvio a figura 5.2 apresenta as mesmas curvas para valores maiores de k .

Para analisarmos a dependência do desvio com \sqrt{t} na figura 5.3 reescalamos o grau por $1/\sqrt{t}$ e $N(k, t)$ por \sqrt{t} . Podemos observar nessa figura que a queda dos valores de iteração e de simulação coincidem, e além disso que os desvios começam ocorrer após $\frac{k}{\sqrt{t}} = 1$, o que concorda com a escala esperada.

Para visualizar $L(k, k', t)$ novamente escolheremos dois valores de k' assim como fizemos no capítulo 3. Na figura 5.4 temos os resultados de simulação, iteração e solução analítica para $L(k, k', t)$ para três tempos distintos e para valores de $k' = 10$ e $k' = 50$. Novamente podemos observar que para k e k' pequenos os resultados coincidem. As curvas relativas a $t = 100$ e $k' = 50$

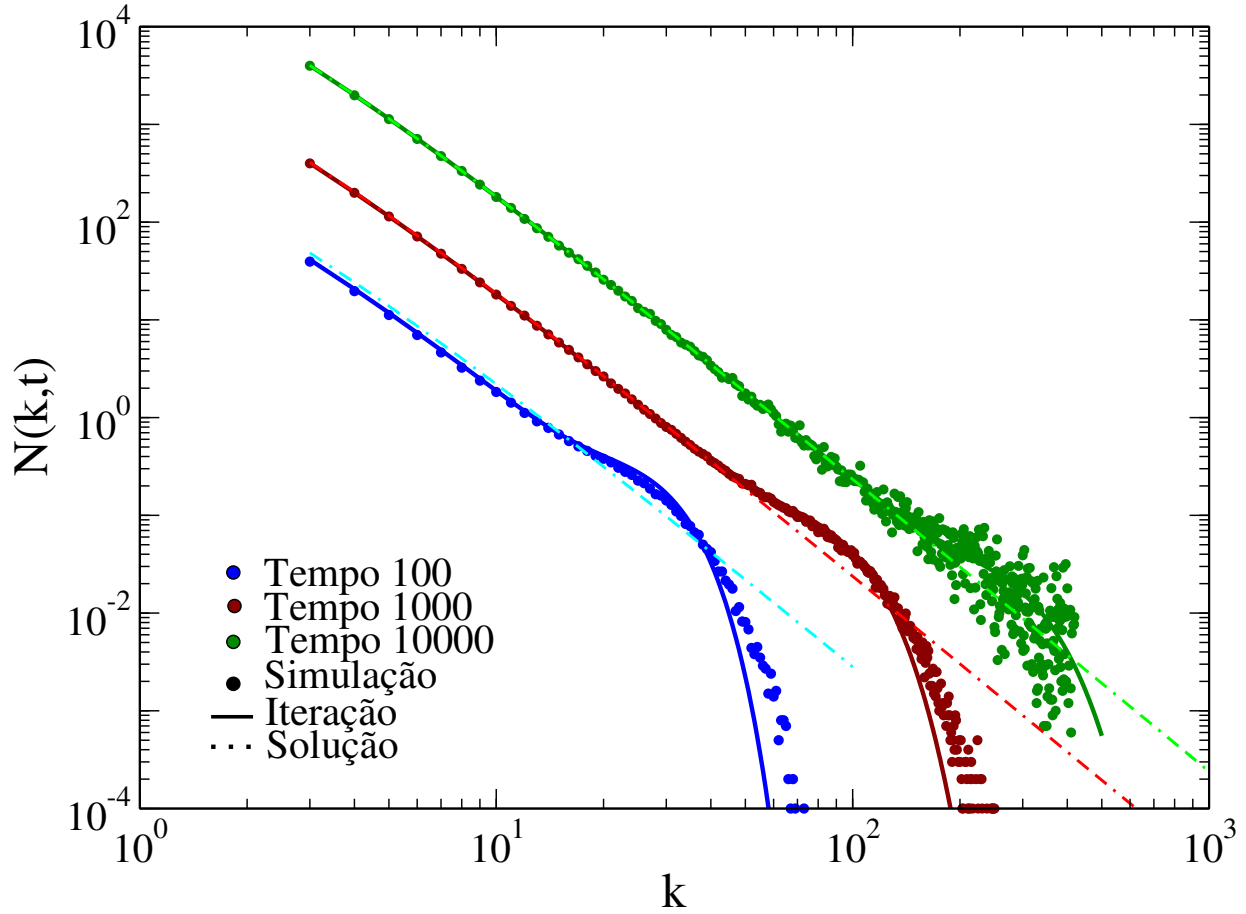


Figura 5.2: Número de nós com grau k obtidos através de simulação, iteração, e da solução analítica para $t = 100$, $t = 1000$ e $t = 10000$ e $k < 300$. Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$. As linhas contínuas apresentam os resultados de iteração da equação mestra e as tracejadas o resultado obtido analiticamente, ambas seguindo o código de cores estabelecido para os dados de simulação

apresentam uma grande divergência, entretanto $k' = 50 > \sqrt{100}$, portanto está fora do limite já estabelecido. Para valores de $k > \sqrt{t}$ começa a ocorrer um desvio também para os outros valores de k' e t .

Na figura 5.5 reescalamos k e k' por $1/\sqrt{t}$ e $L(k, k', t)$ por t . Agora escolhemos valores de $\frac{k'}{\sqrt{t}} = 0.5$ e $\frac{k'}{\sqrt{t}} = 1$. Podemos observar mais uma vez que as quedas ocorrem juntas, e que o desvio começa após $\frac{k}{\sqrt{t}} = 1$, o que indica que a escala é a mesma.

A solução obtida para $\Delta(k, k', k'', t)$ no capítulo 4 é independente do tempo, o que parece ser um resultado inconsistente para um modelo de crescimento de redes. Para verificar a validade

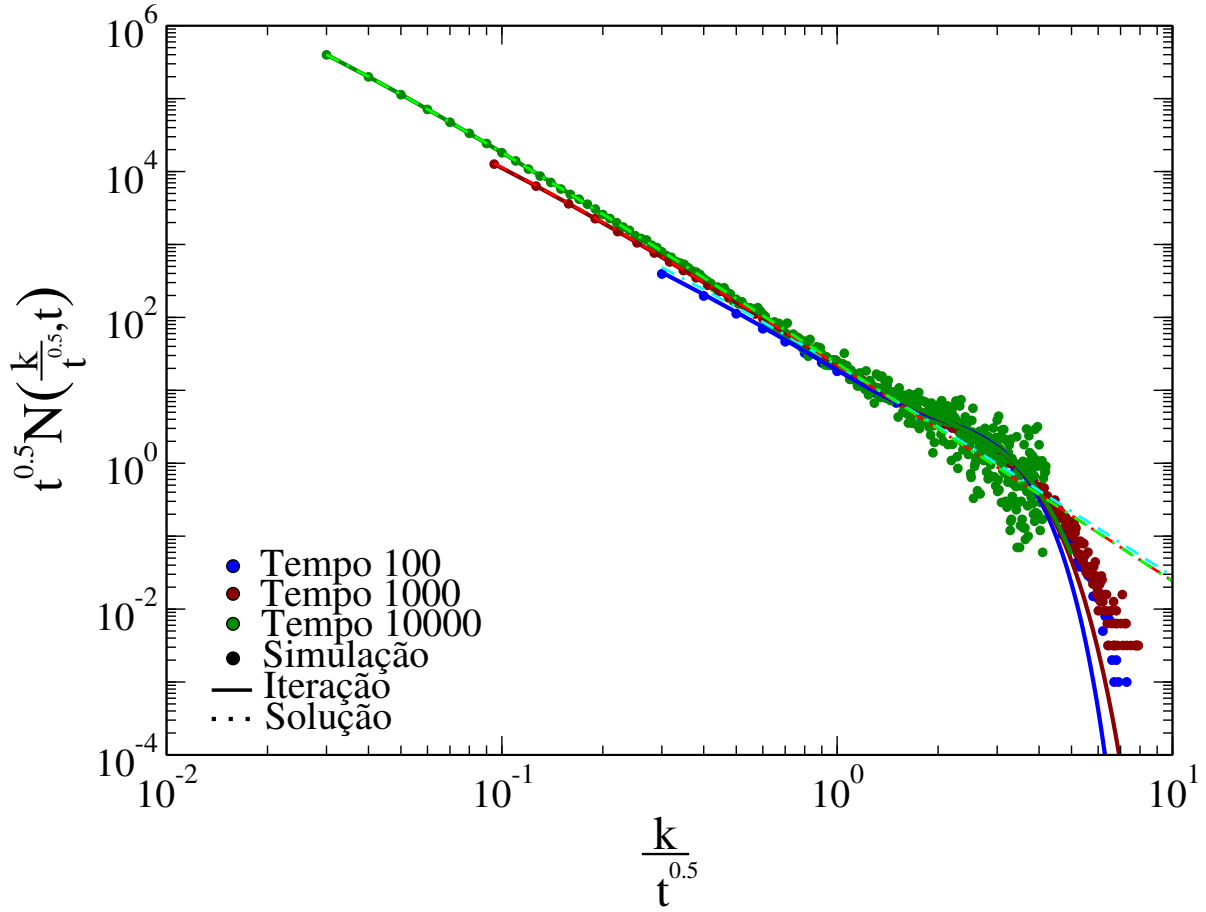


Figura 5.3: Reescala de $N(k, t)$ para simulação, iteração, e da solução analítica. Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$. As linhas contínuas apresentam os resultados de iteração da equação mestra e as tracejadas o resultado obtido analiticamente, ambas seguindo o código de cores estabelecido para os dados de simulação

dessa solução vamos utilizar a densidade superficial de triângulos definida no capítulo 3 como

$$\rho(r, t) = \delta_{r, r_k} \Delta(k, k', k'', t) \frac{2}{\pi r^2}, \quad (5.1)$$

onde

$$r_k = \sqrt{k^2 + k'^2 + k''^2}. \quad (5.2)$$

A figura 5.6 apresenta a comparação entre a simulação e a iteração para três tempos distintos e a solução independente do tempo. Como a solução é independente do tempo ela está representada

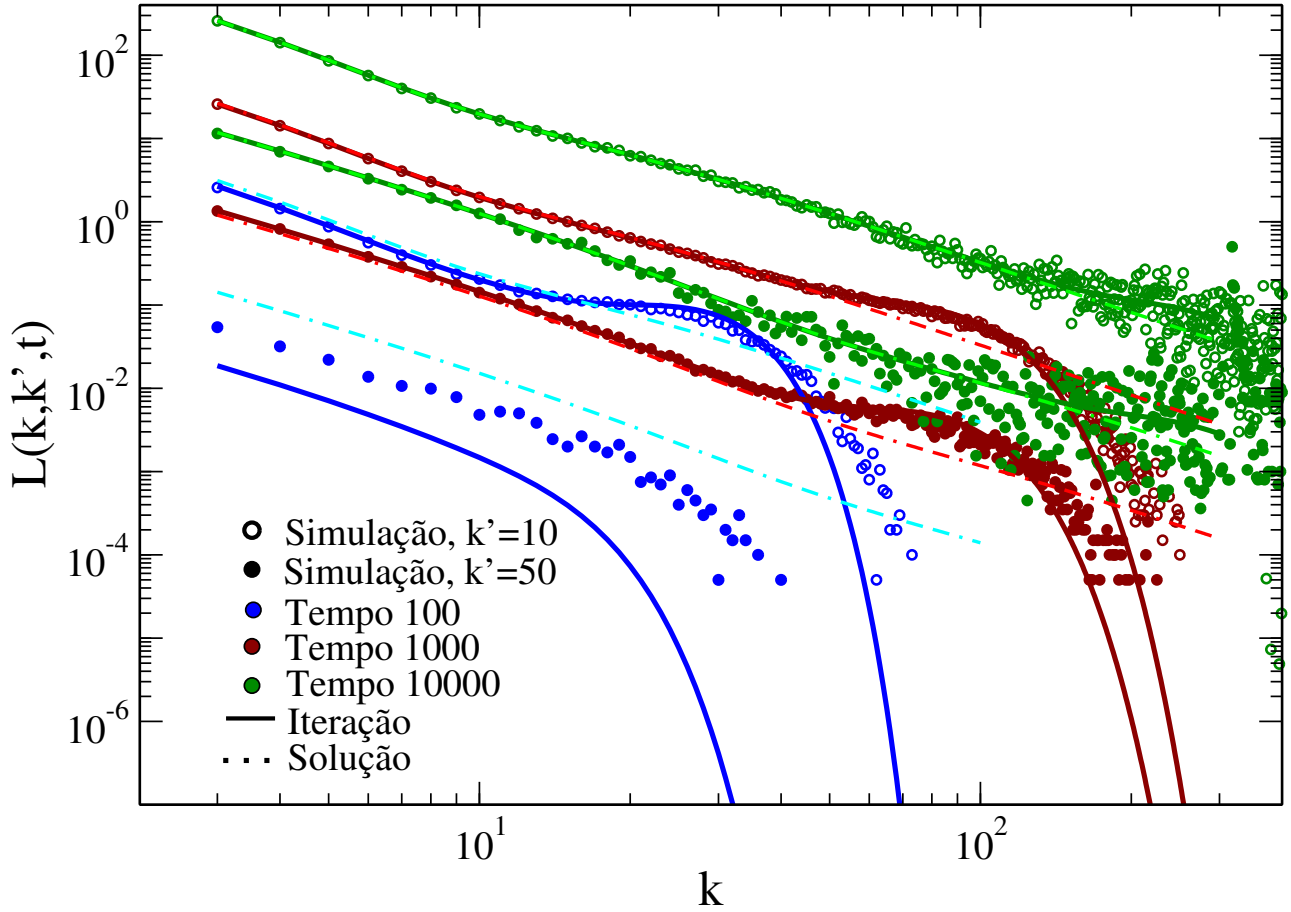


Figura 5.4: Número de ligações entre nós com grau k e k' obtidos através de simulação, iteração, e da solução analítica para $t = 100$, $t = 1000$ e $t = 10000$. Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$, os pontos vazados equivalem a $k' = 10$ e os sólidos a $k' = 50$. As linhas contínuas apresentam os resultados de iteração da equação mestra e as tracejadas o resultado obtido analiticamente, ambas seguindo o código de cores estabelecido para os dados de simulação

por uma única linha tracejada laranja. Podemos observar que, apesar da solução ser independente do tempo ela concorda com os resultados para r pequeno.

Para verificar o intervalo de validade da solução analítica a figura 5.7 apresenta os dados de simulação e iteração com o raio reescalado por $1/\sqrt{t}$, e, como $\Delta(k, k', k'', t)$ escala com $t^{3/2}$, $\rho(r)$ foi reescalado pelo mesmo valor. Foi também adicionada três curvas da solução analítica que, apesar de ser independente do tempo, foi reescalada por tempos equivalentes aos da simulação e iteração. Podemos novamente observar que a queda dos valores de simulação são coerentes com a escala temporal definida. Além disso a solução analítica descreve os resultados até r/\sqrt{t} um pouco maior do que um.

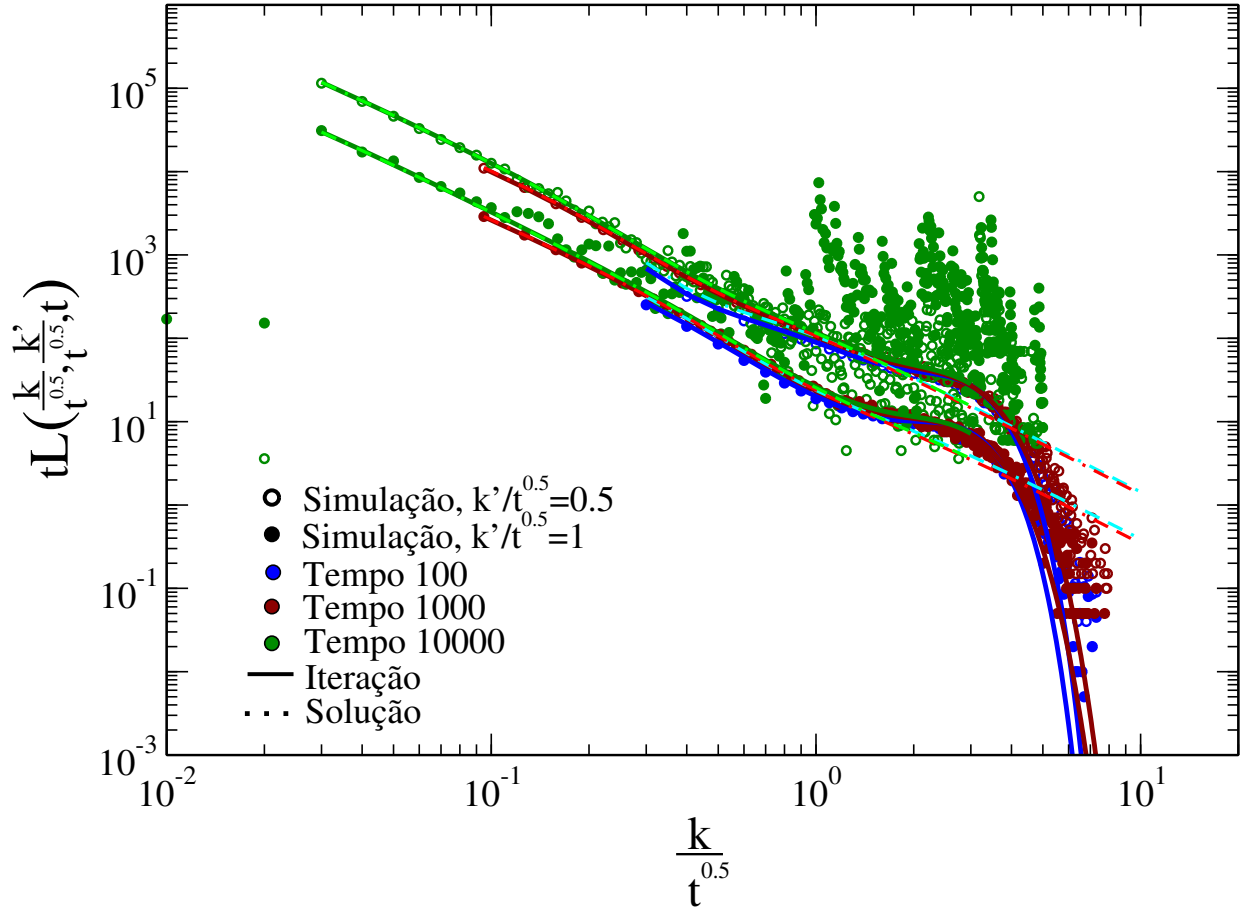


Figura 5.5: Reescala de $L(k, k', t)$ para simulação, iteração, e da solução analítica para $t = 100$, $t = 1000$ e $t = 10000$. Os pontos azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$, os pontos vazados equivalem a $k'/\sqrt{t} = 0.5$ e os sólidos a $k'/\sqrt{t} = 1$. As linhas contínuas apresentam os resultados de iteração da equação mestra e as tracejadas o resultado obtido analiticamente, ambas seguindo o código de cores estabelecido para os dados de simulação

Vimos que através do método aplicado conseguimos obter expressões analíticas para as principais quantias necessárias para caracterizar as redes obtidas com o modelo de Barabási-Albert, $N(k, t)$, $L(k, k', t)$ e $\Delta(k, k', k'', t)$. Esses resultados foram obtidos através de simulação, iteração, e analiticamente. Os resultados analíticos obtidos são muito próximos dos obtidos através de simulação, e, por apresentarem explicitamente a dependência no tempo, podem ser utilizados para descrever os resultados fora do estado assintótico. Além disso as iterações são precisas o suficiente para obtermos esses resultados inclusive para modelos sem solução analítica.

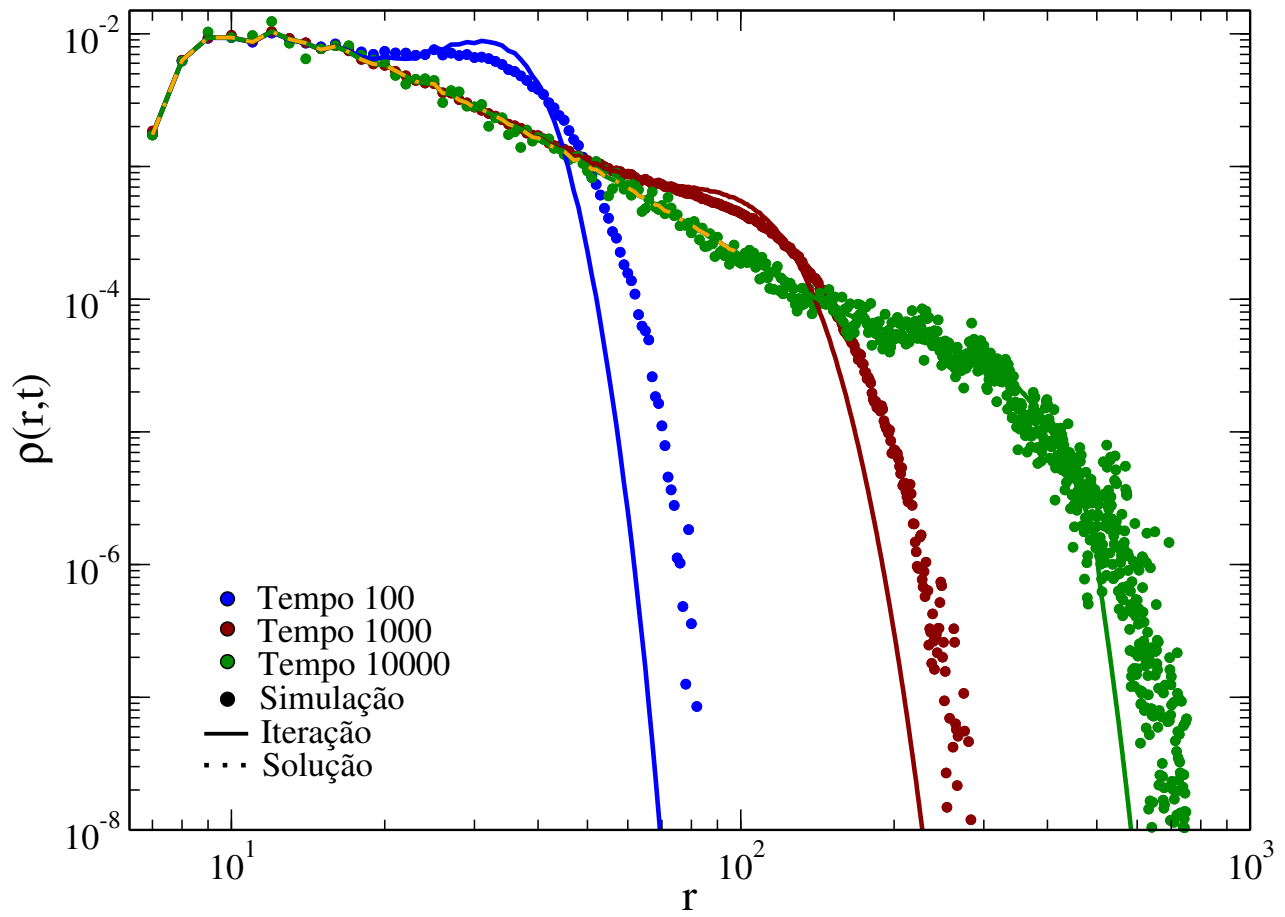


Figura 5.6: Densidade superficial de triângulos em função do raio de simulação, iteração, e da solução analítica. Os pontos e linhas contínuas azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$, para simulação e iteração. A linha tracejada em laranja apresenta o resultado independente do tempo obtido analiticamente.

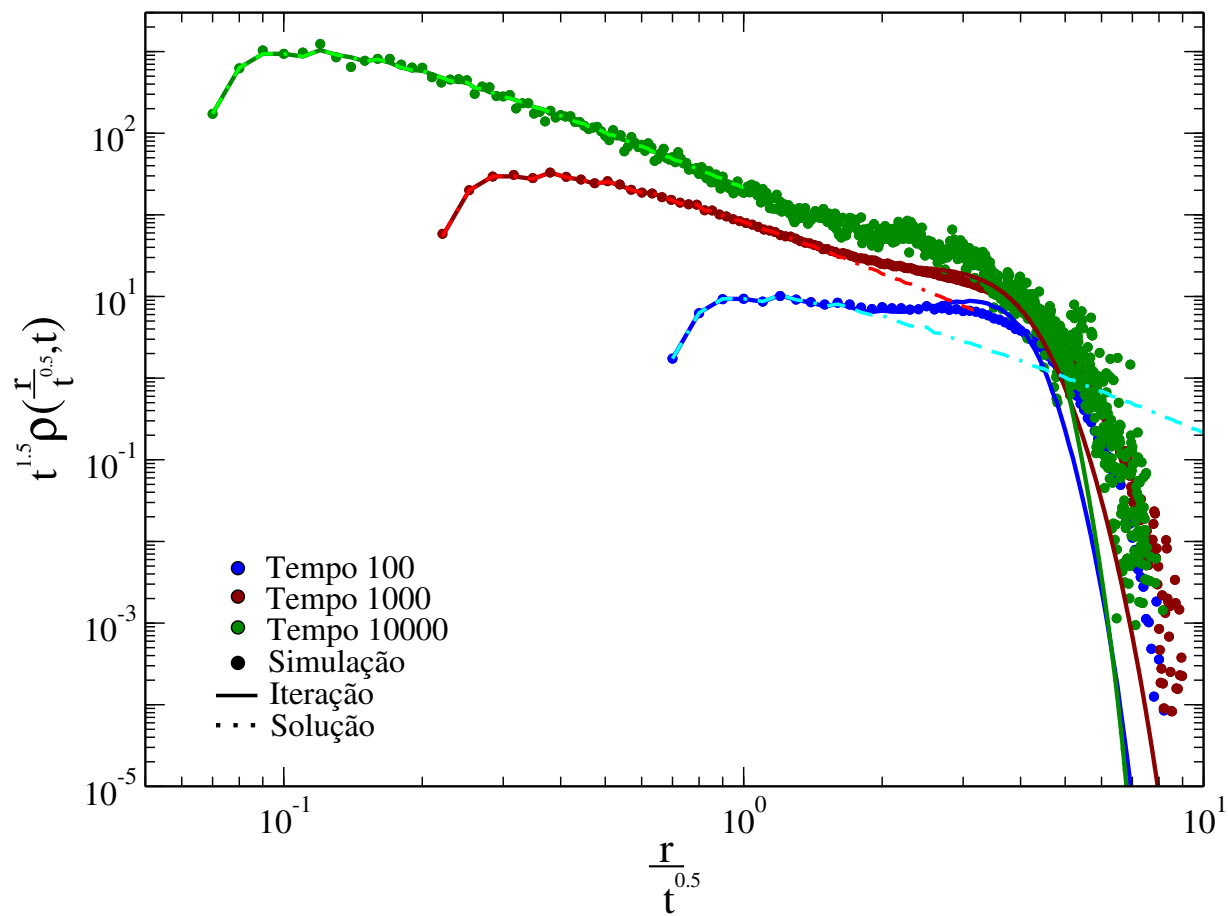


Figura 5.7: Densidade superficial de triângulos em função de r/\sqrt{t} simulação, iteração, e da solução analítica. Os pontos e linhas contínuas azuis representam $t = 100$, os vermelhos $t = 1000$ e os verdes $t = 10000$, para simulação e iteração. As linhas tracejadas em representam a solução analítica independente do tempo reescalada pela mesma escada da simulação e iteração.

Capítulo 6

Conclusão

Neste trabalho apresentamos as equações mestras para o número de ligações $L(k, k', t)$, dada pela Eq. 4.6 e para o número de triângulos $\Delta(k, k', k'', t)$, Eq. 4.8, além da equação para o número de nós $N(k, t)$, Eq. 4.4, que já era conhecida. Essas equações podem ser iteradas numericamente para se conhecer o valor das distribuições. Uma vantagem desse método é que a iteração é numericamente menos custosa que a simulação de um grande número de redes para se obter as médias de $N(k, t)$, $L(k, k', t)$ e $\Delta(k, k', k'', t)$. A precisão dessas equações pode ser observada nas figuras apresentadas no capítulo 5, onde se mostra que até mesmo os efeitos de tamanho finito são previstos com fidelidade.

A partir das equações mestras obtivemos as equações diferenciais associadas e desenvolvemos um método de solução iterativa, que pode ser utilizado em outros conjuntos de equações acopladas, desde que seja possível ordená-las de forma adequada. Neste caso em particular o operador diferencial é independente do grau, e portanto a função de Green é sempre da mesma forma, facilitando as soluções. Essa particularidade pode se reproduzir em outros modelos de crescimento de redes, e portanto é possível que esse método seja aplicado também a esses modelos.

As soluções de $N(k, t)$, $L(k, k', t)$ e $\Delta(k, k', k'', t)$ obtidas através desse método são válidas para $t \gg N_0 k_0$, ou seja, tempo suficiente para as condições iniciais perderem relevância frente à dinâmica. Todas as soluções obtidas são válidas até $k = \sqrt{t}$, o limite apresentado por Dorogovtsev e Mendes [22] a partir do qual os efeitos de tamanho finito prevalecem. Portanto esses resultados

podem ser utilizados para discutir se uma rede real pequena pode ou não ser representada pelo modelo de Barabási-Albert, uma vez que, até o limite $k = \sqrt{t}$, as redes devem ser descritas pelas soluções analíticas.

A solução de $\Delta(k, k', k'', t)$, apesar de ser independente do tempo, descreve com bastante precisão as simulações dentro do limite de tamanho finito. Isso acontece pois, se construirmos a imagem mental de $\Delta(k, k', k'', t)$ como um objeto que cresce no espaço tridimensional gerado por k , k' e k'' , a dinâmica de crescimento ocorre na fronteira desse objeto, já que os nós com maior probabilidade de receberem ligações estão na fronteira. Por isso para grau pequeno a dinâmica converge rapidamente para o estado estacionário.

A solução obtida para $\Delta(k, k', k'', t)$ não permite que seja calculada a clusterização média por grau para o modelo de Barabási-Albert, apesar desta grandeza estar diretamente relacionada com a clusterização, como vimos na equação 2.13. Isso ocorre porque é necessário realizar um somatório para todos os valores possíveis de k' e k'' , e portanto são somados valores acima de \sqrt{t} , com isto o cálculo da clusterização a partir do resultado obtido só seria possível no limite $t \rightarrow \infty$. Entretanto, $\Delta(k, k', k'', t)$ oferece uma visão da estrutura de vizinhança ainda mais detalhada, já que estabelece uma relação para o grau dos três vértices do triângulo. A solução para $\Delta(k, k', k'', t)$ e $L(k, k', t)$ é inédita na literatura, enquanto que a solução para $N(k, t)$ concorda com os resultados já existentes na literatura, dentro dos respectivos limites de validade.

Este trabalho obteve os seguintes resultados: apresenta equações mestras e soluções analíticas para o modelo de Barabási-Albert no limite de tempo finito; oferece uma ferramenta para o estudo de redes reais pequenas; apresenta uma medida analítica da vizinhança dos nós de determinado grau, complementar à clusterização; e introduz um método que pode ser utilizado para descrever analiticamente outros modelos de crescimento de redes.

6.1 Perspectivas

Ao longo da realização deste trabalho percebemos oportunidades de perspectivas futuras, que apresentaremos a seguir:

- a mais imediata é já uma das motivações do trabalho, e consiste em comparar os resultados obtidos com redes reais que tenham a comparação com o modelo de Barabási-Albert contestada;
- podemos utilizar o método através do qual obtivemos as soluções analíticas para outros modelo de crescimento de redes, em particular para generalizações do modelo de Barabási-Albert, nos quais a adaptação do método é mais simples;
- podemos ainda utilizar as soluções obtidas para retornar ao estudo de redes de interação de proteínas, que motivou este trabalho;
- podemos ainda procurar obter resultados analíticos para um modelo de crescimento genérico, de forma que possamos investigar os mecanismos de crescimento de redes reais, escolhendo *a posteriori* o mecanismo que reproduz a rede observada.

Referências Bibliográficas

- [1] Ash, J. and Newth, D., *Physica A: Statistical Mechanics and its Applications* **380** (2007) 673.
- [2] Eguíluz, V. and Klemm, K., *Physical Review Letters* **89** (2002) 108701.
- [3] Newman, M. E. J., *SIAM Review* **45** (2003) 167.
- [4] Pastor-Satorras, R. and Vespignani, A., *Physical Review E* **65** (2002) 035108.
- [5] Guimerà, R. and Amaral, L. A. N., *The European Physical Journal B - Condensed Matter* **38** (2004) 381.
- [6] Rybarczyk-Filho, J. L. et al., *Nucleic acids research* (2010) 1.
- [7] Cho, W. K. T. and Fowler, J. H., *Journal of Politics* **72** (2010) 124.
- [8] Ferreira, R. M. et al., *PLoS ONE* **8** (2013) e56579.
- [9] Barabási, a. L., Albert, R., and Jeong, H., *Physica A: Statistical Mechanics and its Applications* **272** (1999) 173.
- [10] Barabási, A.-L. and Oltvai, Z. N., *Nature Reviews. Genetics* **5** (2004) 101.
- [11] Erdős, P. and Rényi, A., *Publ. Math. Inst. Hungar. Acad. Sci* **5** (1960) 17.
- [12] Albert, R. and Barabasi, A.-L., *Physical Review Letters* **85** (2000) 5234.
- [13] Albert, R. and Barabási, A. L., *Reviews of modern physics* **74** (2002) 47.

- [14] Przulj, N., Corneil, D. G., and Jurisica, I., *Bioinformatics* **20** (2004) 3508.
- [15] Khanin, R. and Wit, E., *Journal of Computational Biology* **13** (2006) 810.
- [16] Fox Keller, E., *BioEssays* **27** (2005) 1060.
- [17] Dorogovtsev, S., Goltsev, A., and Mendes, J., *Physical Review E* **65** (2002) 066122.
- [18] Dunne, J. a., Williams, R. J., and Martinez, N. D., *Proceedings of the National Academy of Sciences of the United States of America* **99** (2002) 12917.
- [19] Newman, M., *Physical Review E* **68** (2003) 026121.
- [20] Barabási, A.-L. and Albert, R., *Science* **286** (1999) 509.
- [21] Dorogovtsev, S. N., Mendes, J. F., and Samukhin, a. N., *Physical review letters* **85** (2000) 4633.
- [22] Dorogovtsev, S. and Mendes, J., *Advances in physics* **51** (2002) 1079.
- [23] Krapivsky, P. L., Redner, S., and Leyvraz, F., *Physical review letters* **85** (2000) 4629.
- [24] Fotouhi, B. and Rabbat, M. G., *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **88** (2013).
- [25] Bianconi, G. and Capocci, A., *Physical Review Letters* **90** (2003) 078701.
- [26] Fronczak, A., Fronczak, P., and Hołyst, J., *Physical Review E* **68** (2003) 046126.
- [27] Vázquez, A., Flammini, A., Maritan, A., and Vespignani, A., *Complexus* (2002) 38.
- [28] Vázquez, A., *Physical Review E* **67** (2003) 1.
- [29] Bondy, J.-A. and Murty, U. S. R., *Graph theory*, Graduate texts in mathematics, Springer, New York, London, 2007.
- [30] Barabási, A.-L. and Frangos, J., *Linked: the new science of networks science of networks*, Basic Books, 2014.

- [31] Kim, J. and Wilhelm, T., *Physica A: Statistical Mechanics and its Applications* **387** (2008) 2637.
- [32] Costa, L. F., Rodrigues, F. A., Travieso, G., and Boas, P. R. V., *Advances in Physics* **56** (2007) 167.
- [33] Colizza, V., Flammini, A., Maritan, A., and Vespignani, A., *Physica A: Statistical Mechanics and its Applications* **352** (2005) 1.
- [34] De Solla Price, D. J., *Science* **149** (1965) 510.
- [35] Krapivsky, P. L. and Redner, S., *Physical review. E, Statistical, nonlinear, and soft matter physics* **63** (2001) 066123.
- [36] Krapivsky, P. L., Rodgers, G. J., and Redner, S., *Physical Review Letters* **86** (2001) 5401.
- [37] Bianconi, G. and Barabási, A.-L., *Europhysics Letters* **54** (2001) 436.
- [38] Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M., and Krioukov, D., *Nature* **489** (2012) 537.
- [39] Adamic, L. A. and Huberman, B. A., *Nature* **401** (1999) 131.
- [40] Goh, K., Kahng, B., and Kim, D., *Physical review letters* **88** (2002) 108701.
- [41] Dorogovtsev, S. and Mendes, J., *Physical Review E* **63** (2001) 4.
- [42] Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E., and Strogatz, S. H., *Physical review. E, Statistical, nonlinear, and soft matter physics* **64** (2001) 041902.
- [43] Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N., *Physical Review E* **64** (2001) 066110.
- [44] Rodgers, G. and Darby-Dowman, K., *The European Physical Journal B* **23** (2001) 267.
- [45] Tadić, B., *Physica A: Statistical Mechanics and its Applications* **293** (2001) 273.

- [46] Tadić, B., *Physica A: Statistical Mechanics and its Applications* **314** (2002) 278.
- [47] Holme, P. and Kim, B. J., *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **65** (2002) 2.

Apêndice A

Solução da equação master para $N(m, t)$

No capítulo 4 apresentamos a equação diferencial para $N(k, t)$ para o modelo de Barabási-Albert, obtida através da equação master

$$\frac{\partial N(k, t)}{\partial t} + m \frac{kN(k, t)}{(2mt + N_0k_0)} = \delta_{k,m} + m \frac{(k-1)N(k-1, t)}{(2mt + N_0k_0)}. \quad (\text{A.1})$$

Esta equação tem a forma de uma espécie de recorrência, uma vez que a equação diferencial para $N(k, t)$ depende de $N(k-1, t)$. Como, por construção, na nossa rede o menor grau é m , podemos resolver a equação para $N(m, t)$, utilizando-a então para resolver a equação para $N(m+1, t)$, e assim sucessivamente. A equação diferencial para $N(m, t)$ então fica

$$\frac{\partial N(m, t)}{\partial t} + m \frac{mN(m, t)}{(2mt + N_0k_0)} = 1, \quad (\text{A.2})$$

sujeita a condição inicial $N(m, 0) = 0$, uma vez que $N(m-1, t) = 0$, já que inicialmente todos os nós tem grau maior que m , e a dinâmica do modelo não admite a diminuição de grau.

A solução da equação homogênea associada à A.2 é

$$N_h(m, t) = A(N_0k_0 + 2mt)^{-m/2} \quad (\text{A.3})$$

A condição inicial homogênea leva à solução trivial. Para resolver a equação diferencial não

homogênea utilizamos funções de Green. Como a solução da equação homogênea é dada por $A(N_0k_0 + 2mt)^{-k_0/2}$, temos que a função de green deve satisfazer

$$g(m, t, t') = \begin{cases} 0, & t < t' \\ g(m, t, t') = A(N_0k_0 + 2mt)^{-m/2}, & t > t' \end{cases}. \quad (\text{A.4})$$

Em $t = t'$ temos

$$A(N_0k_0 + 2mt')^{-m/2} = 1, \quad (\text{A.5})$$

e portanto a função de green associada fica

$$G(m, t, t') = \left(\frac{N_0k_0 + 2mt'}{N_0k_0 + 2mt} \right)^{m/2}. \quad (\text{A.6})$$

Resolvendo

$$N_2(m, t) = \int_0^t dt' 1 \left(\frac{N_0k_0 + 2mt'}{N_0k_0 + 2mt} \right)^{m/2}, \quad (\text{A.7})$$

temos a solução para $N(m, t)$

$$N(k_0, t) = \frac{N_0k_0 + 2mt}{m(m+2)} - \frac{(N_0k_0)^{1+m/2}}{m(m+2)(N_0k_0 + 2mt)^{m/2}} \quad (\text{A.8})$$

aproximando para $t \gg N_0k_0$, temos

$$N(m, t) = \frac{N_0k_0 + 2mt}{m(m+2)}. \quad (\text{A.9})$$