

UNIVERSIDADE FEDERAL DO PARANÁ

DAVID REKSIDLER JÚNIOR

CLIQUE MÁXIMA EM GRAFOS LEI DE POTÊNCIA

CURITIBA PR

2020

DAVID REKSIDLER JÚNIOR

CLIQUE MÁXIMA EM GRAFOS LEI DE POTÊNCIA

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Murilo Vicente Gonçalves da Silva.

CURITIBA PR

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

R381p Reksidler Júnior, David
Clique máxima em grafos lei de potência [recurso eletrônico] / David Reksidler Júnior. –
Curitiba, 2020.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-
Graduação em Informática, 2020.

Orientador: Murilo Vicente Gonçalves da Silva.

1. Teoria dos grafos. 2. Otimização combinatória. 3. Distribuição (Teoria da probabilidade). I.
Universidade Federal do Paraná. II. Silva, Murilo Vicente Gonçalves da. III. Título.

CDD: 519.24

Bibliotecária: Vanusa Maciel CRB- 9/1928

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **DAVID REKSIDLER JUNIOR** intitulada: **Clique máxima em grafos lei de potência**, sob orientação do Prof. Dr. MURILO VICENTE GONÇALVES DA SILVA, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 31 de Agosto de 2020.

Assinatura Eletrônica

01/09/2020 15:31:30.0

MURILO VICENTE GONÇALVES DA SILVA

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

01/09/2020 09:55:02.0

RENATO JOSÉ DA SILVA CARMO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

01/09/2020 10:53:23.0

LEANDRO MIRANDA ZATESKO

Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ)

Se o acaso é inevitável e a sorte é fundamental, então o diferencial pode não ser o talento (ainda que ele seja necessário), mas a persistência. (Mlodinow, 2009)

AGRADECIMENTOS

Durante a realização deste trabalho várias pessoas me ajudaram, sem as quais o mesmo não teria sido possível. A todas elas, as quais seria exaustivo aqui enunciar, a minha profunda gratidão. A algumas delas, pelo apoio especial, gostaria de agradecer especialmente.

Agradeço à minha esposa, Renata, pelo companheirismo, parceria, apoio e incentivo durante todo o mestrado.

Agradeço à toda a minha família, em especial meus pais, David e Patrícia, e meus irmãos, Kelvyn e Gabriel, por sempre acreditarem no meu potencial e incentivarem o meu desenvolvimento.

Agradeço ao meu orientador, Murilo Vicente Gonçalves da Silva, pelos ensinamentos, conselhos e pela paciência depositada em mim durante todo o período de orientação.

Agradeço aos professores, Renato Carmo, André Luís Vignatti, André Luiz Pires Guedes e Leandro Miranda Zatesko pelas importantes contribuições e críticas que possibilitaram o desenvolvimento do meu trabalho.

RESUMO

Com o avanço na capacidade de processamento e armazenamento de grandes quantidades de dados, foi-se observado que muitas redes de grande porte advindas de situações práticas, desde a World Wide Web até redes sociais e biológicas, apresentam uma distribuição lei de potência nos graus dos vértices. Segundo essa distribuição, o número de vértices com um certo grau k é proporcional a $k^{-\beta}$, onde β é o expoente que caracteriza a intensidade da lei de potência. Tais redes, modeladas matematicamente por grafos de lei de potência, tiveram um grande aumento no interesse de estudo dentro da área de computação teórica. Muitos pesquisadores acreditam que pode ser mais fácil resolver alguns problemas de otimização combinatória em grafos de lei de potência do que em outros tipos de grafos em geral. Grande parte dos estudos se utilizam de heurísticas gulosas em grafos de lei de potência para obter algoritmos eficientes que computem uma solução próxima da ótima para os problemas de otimização. Grafos de lei de potência podem ser gerados e analisados por meio de modelos de grafos aleatórios. Embora o desempenho dos algoritmos para esses problemas esteja sendo bastante explorado experimentalmente na literatura, há pouco material teórico que analise esse comportamento. A presente dissertação se insere neste contexto teórico, analisando o problema da clique máxima a fim de evidenciar a eficiência de algoritmos gulosos em grafos de lei de potência para tal problema computacional. Fazendo uso de um modelo de grafo aleatório de ligação preferencial, provamos que a probabilidade de aparecimento uma clique em um grafo lei de potência construído pelo modelo decai exponencialmente como uma função que depende do tamanho da clique. Esta é a principal ligação entre o nosso resultado e a eficiência dos algoritmos gulosos em grafos de lei de potência, dado que de maneira geral, algoritmos gulosos seguem uma certa ordenação de vértices em sua solução, geralmente começando dos vértices de grau alto.

Palavras-chave: Teoria dos grafos. Otimização combinatória. Teoria da probabilidade.

ABSTRACT

With the advance in the capacity to process and store large amounts of data, it was observed that many large networks arising from practical situations, from the World Wide Web to social and biological networks, present a power-law distribution in the degrees of vertices. According to this distribution, the number of vertices with a certain degree k is proportional to $k^{-\beta}$, wherein β is the exponent that characterizes the intensity of the power-law. Such networks, mathematically modeled by power-law graphs, had a great increase in the interest of study within the theoretical computing area. Many researchers believe that it may be easier to solve some combinatorial optimization problems in power-law graphs than in other types of graphs in general. Most studies use greedy heuristics in power-law graphs to obtain efficient algorithms that compute a solution close to the optimum for optimization problems. Power-law graphs can be generated and analyzed using random graph models. Although the performance of the algorithms for these problems is being experimentally explored in the literature, there is little theoretical material to analyze this behavior. The present dissertation is inserted in this theoretical context, analyzing the problem of maximum clique in order to show the efficiency of greedy algorithms in power-law graphs for such computational problem. Using a preferential attachment random graph model, we prove that the probability of a clique in a power-law graph built by the model decays exponentially as a function that depends on the clique size. This is the main link between our result and the efficiency of greedy algorithms in power-law graphs, given that in general, greedy algorithms follow a certain order of vertices in their solution, usually starting from the high degree vertices.

Keywords: Graph theory. Combinatorial optimization. Probability theory.

LISTA DE FIGURAS

2.1	Distribuição normal	16
2.2	Distribuição Lei de Potência	16
3.1	Construção do grafo no modelo $ACL(\alpha, \beta)$	21
3.2	Construção do grafo para $m > 1$ no modelo $PA_t(m, \delta)$	26

SUMÁRIO

1	INTRODUÇÃO	9
2	FUNDAMENTAÇÃO TEÓRICA.	11
2.1	GRAFOS	11
2.2	TEORIA DE PROBABILIDADE.	12
2.3	NOTAÇÃO ASSINTÓTICA E FUNÇÕES ESPECIAIS.	15
2.4	GRAFOS DE LEI DE POTÊNCIA.	16
2.5	PROBLEMAS DE OTIMIZAÇÃO E ALGORITMOS GULOSOS	17
2.5.1	CLIQUE MÁXIMA.	17
2.6	CONCLUSÃO	18
3	MODELOS DE GRAFOS ALEATÓRIOS	19
3.1	O MODELO $G_{n,p}$.	19
3.2	AJUSTANDO A DISTRIBUIÇÃO DE GRAUS EM GRAFOS ALEATÓRIOS	20
3.2.1	MODELO DE CONFIGURAÇÃO	20
3.2.2	MODELO DE GRAFO ALEATÓRIO GENERALIZADO	21
3.3	MODELOS DE LIGAÇÃO PREFERENCIAL.	22
3.4	CONCLUSÃO	26
4	PROBLEMAS DE OTIMIZAÇÃO EM GRAFOS DE LEI DE POTÊNCIA	28
4.1	REVISÃO DA LITERATURA	28
4.1.1	ESTUDOS EXPERIMENTAIS.	28
4.1.2	ESTUDOS TEÓRICOS.	28
4.2	RESULTADOS ORIGINAIS	31
4.3	CONCLUSÃO	41
5	CONCLUSÃO	42
	REFERÊNCIAS	43
	APÊNDICE A – GRAFOS	45
A.1	GRAFOS LEI DE POTÊNCIA	45
	APÊNDICE B – TEORIA DE PROBABILIDADE	47
B.1	INTERSECÇÃO DE EVENTOS	47

1 INTRODUÇÃO

Grafos que apresentam distribuição lei de potência nos graus dos vértices, comumente chamados de *grafos de lei de potência*, são grafos em que a quantidade de vértices com grau k é proporcional a $k^{-\beta}$, onde β é uma constante, chamada de expoente da lei de potência. Esta constante caracteriza a intensidade da curva da distribuição de graus.

A partir de diversos estudos experimentais em cima de grandes bases de dados de grafos, observou-se que muitos grafos de diferentes domínios apresentam a distribuição lei de potência nos graus de seus vértices. Este fenômeno foi observado por diversos autores de diferentes áreas de estudo (Faloutsos et al., 1999; Guelzim et al., 2002; Eubank et al., 2004).

Apresentar uma definição precisa, e ao mesmo tempo útil, do que é um grafo lei de potência é uma tarefa que exige algum cuidado. Em particular, se quisermos incluir na classe dos grafos de lei de potência apenas os grafos que possuam a sequência de graus cujos valores sigam precisamente o requerido por uma regra de lei de potência, possivelmente pouquíssimos grafos advindos de algumas aplicações práticas passariam no teste. Portanto, neste trabalho não vamos apresentar uma definição determinística para grafos de lei de potência, e sim uma definição baseada em modelos de grafos aleatórios.

Teóricos da computação e matemáticos estudam grafos aleatórios desde a década de 60, em particular com os trabalhos de Erdős e Rényi (1960). Nestes modelos permite-se aleatoriedade na geração das arestas, possibilitando assim, uma análise de comportamento sobre grafos fazendo uso de modelos aleatórios para representá-los. Entretanto estes modelos, embora relevantes e muito estudados, podem não capturar com exatidão grafos advindos de certas aplicações práticas, pelo fato de induzirem uma distribuição normal nos graus dos vértices. Portanto, uma das tarefas relevantes no estudo de grafos aleatórios é ajustar os modelos para que os graus esperados dos vértices dos grafos em questão possam ter outras distribuições diferentes da distribuição normal.

Mesmo antes do recente interesse em grafos de lei de potência, muitos autores já haviam proposto modelos de grafos aleatórios com distribuições ajustáveis de graus, como por exemplo em Bender e Canfield (1978) e Molloy e Reed (1995). Mais recentemente, Aiello et al. (2001) propuseram um modelo capaz de gerar especificamente grafos de lei de potência onde é possível ajustar características do grafo gerado através da escolha de certos parâmetros α e β que guiam o tamanho do grafo a intensidade da lei de potência.

Muitos problemas computacionais podem ser resolvidos se utilizando de grafos, e dentre eles, estão os problemas de otimização. Tais problemas de otimização são amplamente estudados devido à sua importância em aplicações e situações práticas. Alguns destes problemas pertencem à classe de complexidade NP-difícil e o grau de dificuldade para computar uma solução ótima é alto, supondo que $P \neq NP$.

Recentemente, pesquisadores observaram que alguns destes problemas podem ser mais fáceis de resolver em grafos de lei de potência do que em outros grafos em geral (Park e Lee, 2001; Eubank et al., 2004; da Silva et al., 2013; Rogiski, 2016). O foco do presente trabalho, de maneira mais ampla, é analisar o motivo desta eficiência de um tipo particular de estratégia, especificamente a utilização de algoritmos gulosos em tais grafos. Algoritmos gulosos, que geralmente são simples e intuitivos, mas nem sempre retornam a resposta ótima, apresentam resultados muito bons se aplicados em grafos de lei de potência. Estudos experimentais mostram que para alguns problemas NP-Difíceis, os algoritmos gulosos computam soluções ótimas ou próximas da ótima se aplicados em grafos de lei de potência. Entretanto, há pouco material teórico explicando o motivo desse comportamento.

Portanto, este trabalho se insere neste ramo teórico, podendo especificar dois principais objetivos:

- Analisar o uso de probabilidade e técnicas aleatorizadas, que são amplamente utilizadas na construção de modelos de grafos aleatórios, apresentando ao leitor alguns exemplos de modelos.
- Propor uma análise sobre um problema computacional específico, no caso da presente dissertação o problema da clique máxima.

Utilizando um modelo de grafo aleatório de ligação preferencial, provamos que a probabilidade de existir uma clique em um grafo lei de potência construído pelo modelo, decai exponencialmente como uma função que depende do tamanho da clique. Como geralmente heurísticas gulosas seguem uma determinada ordenação de vértices em sua solução, por exemplo, começando dos vértices de grau alto, esta prova é a principal ligação entre o nosso resultado e o bom desempenho observado de algoritmos gulosos em grafos de lei de potência.

O presente trabalho está dividido em cinco capítulos. Primeiramente nos Capítulos 2 e 3 é apresentado a fundamentação teórica necessária para desenvolver os resultados desta dissertação. Os capítulos apresentam definições e teoremas relacionados a área de teoria da probabilidade, grafos, lei de potência, algoritmos gulosos e modelos de grafos aleatórios. Decidimos separar o Capítulo de modelos de grafos aleatórios do resto do referencial teórico dada a importância do tópico para o estudo. No Capítulo 4, após apresentarmos os resultados obtidos em trabalhos relacionados, apresentamos os resultados originais que obtivemos. Por fim, o Capítulo 5 apresenta a conclusão desta dissertação.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados conceitos relacionados à área de teoria da probabilidade, grafos, lei de potência, algoritmos gulosos e o problema da clique máxima. Estas definições são utilizadas no desenvolvimento e análises da presente dissertação.

2.1 GRAFOS

Nesta parte do referencial teórico, são apresentadas algumas definições básicas envolvendo grafos. As definições apresentadas são baseadas no livro de West (2001).

Seja S um conjunto e k um inteiro, denota-se por $\binom{S}{k}$ o conjunto de subconjuntos de S de tamanho k .

Definição 2.1. (Grafos (não direcionados)). Um *grafo (não direcionado)* G é um par de conjuntos $(V(G), E(G))$, onde $V(G)$ é um conjunto finito em que cada elemento é chamado de *vértice* de G , e $E(G) \subseteq \binom{V(G)}{2}$ em que cada elemento é chamado de *aresta* de G .

Definição 2.2. (Relações de Vizinhaça entre vértices). Se $u, v \in V(G)$ e $\{u, v\} \in E(G)$, dizemos que existe uma aresta *entre* u e v . Uma aresta $a \in E(G)$ é *incidente* a um vértice u se $u \in a$. Se existe uma aresta entre u e v , dizemos que u é *adjacente* a v . A *vizinhaça* de um vértice u , denotada por $N(u)$, é o conjunto de todos os vértices que são adjacentes a u . Se $a = \{u, v\} \in E(G)$, dizemos que u e v são as *pontas* da aresta a .

Definição 2.3. (Grau). O *grau* de um vértice u , denotado por $d(u)$, representa a quantidade de arestas incidentes ao mesmo, ou seja, $d(u) = |N(u)|$.

Definição 2.4. (Grau máximo e mínimo). O *grau máximo* de um grafo G é definido como:

$$\Delta(G) = \max\{d(u) : u \in V(G)\},$$

o *grau mínimo* é:

$$\delta(G) = \min\{d(u) : u \in V(G)\}$$

Definição 2.5. (Grafos Direcionados). Um *grafo direcionado* G é um par de conjuntos $(V(G), E(G))$, onde $V(G)$ é um conjunto finito de vértices, e $E(G) \subseteq \binom{V(G)}{2}$ em que cada elemento é um par ordenado de vértices chamado de aresta. O par ordenado determina a *direção* da aresta, logo, $(u, v) \in E(G)$ significa que a direção da aresta é saindo de u e chegando em v . O primeiro vértice do par ordenado é chamado de *cauda* da aresta enquanto o segundo é chamado de *cabeça* da aresta.

Para os grafos direcionados, a noção de grau é diferenciada em: *grau de entrada* e *grau de saída*. Respectivamente, a quantidade de arestas que chegam e a quantidade de arestas que partem do vértice.

Definição 2.6. (Grau de entrada e de saída). O *grau de entrada* de um vértice u , denotado por $d^-(u)$, representa a quantidade de arestas com cabeça u . O *grau de saída* de um vértice u , denotado por $d^+(u)$, representa a quantidade de arestas com cauda u .

Definição 2.7. (Subgrafo). Um grafo G' é um *subgrafo* de G se

$$\begin{aligned} V(G') &\subseteq V(G), \\ E(G') &\subseteq E(G). \end{aligned}$$

Definição 2.8. (Subgrafo induzido por um conjunto de vértices). Dado um conjunto $S \subseteq V(G)$, denota-se $G[S]$ o *subgrafo induzido* por S em G , em que $G[S] = (S, E(G) \cap \binom{S}{2})$.

Definição 2.9. (Multigrafo). Um *multigrafo* é um grafo que pode conter mais de uma aresta entre um par de vértices. Assim, dois vértices podem ser conectados por mais de uma aresta. Dizemos que duas arestas são *paralelas* se contem o mesmo par de vértices. Se duas pontas de uma aresta coincidem no mesmo vértice, a aresta é um *loop*.

Definição 2.10. (Emparelhamento). Dado um grafo G , um emparelhamento M em G é um subconjunto $M \subseteq E(G)$ de modo que as arestas de M não compartilhem um vértice comum.

Definição 2.11. (Cobertura por Vértices). Uma *cobertura por vértices* em um grafo G é um conjunto de vértices $C \subseteq V(G)$ tal que toda aresta $e \in E(G)$ tem pelo menos uma ponta em C .

Definição 2.12. (Conjunto Independente). Dado um grafo G , um subconjunto $V' \subseteq V(G)$ é um *conjunto independente* de G se para todo par de vértices $u, v \in V'$, temos que $\{u, v\} \notin E(G)$.

Definição 2.13. (Clique). Dado um grafo G , um subconjunto $V' \subseteq V(G)$ é uma *clique* de G se para todo par de vértices $u, v \in V'$, temos que $\{u, v\} \in E(G)$. Se V' é uma clique de cardinalidade máxima em G , o tamanho de V' é denotado por $\omega(G)$.

Definição 2.14. (Grafo completo). Um grafo G é dito *completo* se $V(G)$ é uma clique. Um grafo completo com n vértices é denotado por K_n .

2.2 TEORIA DE PROBABILIDADE

Nesta seção são definidos alguns axiomas e teoremas de probabilidade utilizados nesta dissertação. As definições são baseadas na obra de Mitzenmacher e Upfal (2017).

Definição 2.15. (Espaço de Probabilidade Discreto). Um *espaço de probabilidade* possui três componentes:

- Um espaço amostral Ω , que é o conjunto de todos os resultados possíveis do processo aleatório modelado pelo espaço de probabilidade.
- Um conjunto de conjuntos \mathcal{F} representando os eventos, onde cada conjunto em \mathcal{F} é um subconjunto do espaço amostral Ω .
- Uma função de probabilidade que relaciona os elementos de \mathcal{F} ao conjunto dos números reais, $Pr : \mathcal{F} \rightarrow \mathbb{R}$.
 - Para qualquer evento $E \in \mathcal{F}$, $0 \leq Pr(E) \leq 1$.
 - $Pr(\Omega) = 1$.
 - Para qualquer sequência de eventos mutuamente disjuntos:

$$Pr\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n Pr(E_i). \quad (2.1)$$

Eventos são conjuntos e, por isso, são utilizadas notações da teoria dos conjuntos para expressar relações entre os eventos. Escreve-se $E_1 \cap E_2$ quando se quer representar a ocorrência dos dois eventos. Já $E_1 \cup E_2$ representa a ocorrência de pelo menos um dos dois eventos, ou seja, pode ocorrer somente um deles ou os dois. A subtração de conjuntos, $E_1 - E_2$, é usada para representar a ocorrência do evento E_1 e a não-ocorrência do evento E_2 . A negação de um evento, \overline{E} , é igual a $\Omega - E$.

Exemplo 2.1. Em um lançamento de um dado de 6 lados, o espaço amostral é composto por todos os possíveis resultados do lançamento, neste caso os 6 valores do dado, ou seja, $\Omega = \{\square, \square, \square, \square, \square, \square\}$. Seja E_1 o evento em que o valor do lançamento seja maior que 3, ou seja, $E_1 = \{\square, \square, \square\}$. Seja também E_2 o evento que o valor seja par, ou seja, $E_2 = \{\square, \square, \square\}$. Note que neste exemplo $E_1 \cap E_2$ é o evento em que o valor do lançamento é maior que 3 e par, ou seja, $E_1 \cap E_2 = \{\square, \square\}$.

A partir da Equação 2.1, é possível generalizar a equação para eventos não necessariamente disjuntos. A probabilidade da união dos eventos é no máximo a soma das probabilidades e a generalização fica como segue:

Lema 2.1. (Limitante da União). Dados os eventos E_1, E_2, \dots, E_n temos que

$$Pr\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n Pr(E_i). \quad (2.2)$$

Em alguns casos que não se tenha certeza sobre a disjunção dos eventos, pode-se descobrir o valor exato da união das probabilidades com o seguinte lema.

Lema 2.2. (Princípio da inclusão-exclusão). Dados os eventos E_1, E_2, \dots, E_n temos que

$$\begin{aligned} Pr\left(\bigcup_{i=1}^n E_i\right) &= \sum_{i=1}^n Pr(E_i) \\ &\quad - \sum_{i<j} Pr(E_i \cap E_j) \\ &\quad + \sum_{i<j<k} Pr(E_i \cap E_j \cap E_k) \\ &\quad - \dots \\ &\quad + (-1)^{l+1} \sum_{i_1<i_2<\dots<i_l} Pr(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_l}). \end{aligned} \quad (2.3)$$

Observe que acima se faz necessário sempre retirar a probabilidade que fica na intersecção dos eventos, pois quando se soma eventos que se intersectam, soma-se a mesma probabilidade mais de uma vez.

Definição 2.16. (Independência de Eventos). Dados os eventos E_1, E_2, \dots, E_n , eles são *independentes* se e somente se

$$Pr\left(\bigcap_{i=1}^n E_i\right) = \prod_{i=1}^n Pr(E_i). \quad (2.4)$$

Pode-se pensar na ocorrência de um evento afetar a ocorrência de outro. Neste caso, pode-se calcular a probabilidade de um evento acontecer, dado que outro evento já aconteceu da seguinte forma:

Definição 2.17. (Probabilidade Condicional). A *probabilidade condicional* do evento E_1 dado que E_2 ocorreu é dada por

$$Pr(E_1|E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)}, \quad (2.5)$$

dado que $Pr(E_2) > 0$. A probabilidade será positiva, se de alguma forma o evento E_2 influenciar a ocorrência do evento E_1 . De certa forma, a equação é intuitiva, pois se quer saber a probabilidade da ocorrência de $E_1 \cap E_2$ dentro do conjunto do evento E_2 . Como E_2 define um espaço amostral restrito, é preciso normalizar as probabilidades dividindo por $Pr(E_2)$, validando o espaço amostral de modo que a soma das probabilidades dos eventos seja 1.

Utilizando o conceito de independência apresentado na Equação 2.4, caso os eventos E_1 e E_2 sejam independentes, $Pr(E_1|E_2) = Pr(E_1)$, pois a independência garante que a ocorrência do primeiro evento não afete a ocorrência do segundo.

No estudo de eventos aleatórios, também é interessante associar valores à ocorrência dos eventos além de ter somente a informação da ocorrência. Para isso utilizam-se variáveis aleatórias.

Definição 2.18. (Variável Aleatória). Uma *variável aleatória* (*v.a.*) X sobre um espaço amostral Ω é uma função $X : \Omega \rightarrow \mathbb{R}$.

Seja $x \in \mathbb{R}$, o evento $X = x$ representa o conjunto $\{e \in \Omega : X(e) = x\}$, ou seja, a união de todos os eventos em que a variável apresenta valor x . Logo, $Pr(X = x) = \sum_{e \in \Omega: X(e)=x} Pr(e)$.

Exemplo 2.2. Seja X a variável aleatória que represente a soma do valor do lançamento de dois dados. Note que aqui o espaço amostral é composto pelas 36 possíveis jogadas dos dois dados. Já a variável pode assumir 11 possíveis valores das somas dos resultados dos lançamentos, $X \in \{2, 3, \dots, 12\}$. Logo, a probabilidade de a soma ser igual a 3, ou seja, $\{(\square, \square), (\square, \square)\}$, é igual a $Pr(X = 3) = 2/36 = 1/18$.

O conceito de independência visto acima também é válido quando se fala de variáveis aleatórias.

Definição 2.19. (Independência de Variáveis Aleatórias). Dada as (*v.a.*) X_1, X_2, \dots, X_n , elas são *independentes* se e somente se

$$Pr\left(\bigcap_{i=1}^n X_i = x_i\right) = \prod_{i=1}^n Pr(X_i = x_i). \quad (2.6)$$

Um conceito muito importante se falando de variáveis aleatórias é a denominada *esperança*. A esperança pode ser entendida como a média dos valores assumidos pela variável aleatória, onde cada valor é ponderado pela probabilidade da variável assumir este valor.

Definição 2.20. (Esperança). A *esperança* de uma variável aleatória X é dada por,

$$E[X] = \sum_i i \cdot Pr(X = i), \quad (2.7)$$

onde a soma é sobre todos os valores assumidos por X .

Exemplo 2.3. Considere o lançamento de dois dados, como visto no Exemplo 2.2, com a esperança, consegue-se calcular o valor esperado dessa soma. Neste caso,

$$E[X] = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \dots + 12 \cdot \frac{1}{36} = 7. \quad (2.8)$$

2.3 NOTAÇÃO ASSINTÓTICA E FUNÇÕES ESPECIAIS

Definição 2.21. (Notação $O(1)$). Uma função $f : \mathbb{N} \rightarrow \mathbb{R}$ é assintoticamente limitada superiormente por uma constante se existem $c_f > 0$ e $n_f \in \mathbb{N}$ tais que $f(n)$ em valor absoluto nunca ultrapassa c_f de n_f em diante, isto é $|f(n)| \leq c_f$ para todo $n \geq n_f$. O conjunto das funções assintoticamente limitadas superiormente por uma constante é denotado por $O(1)$.

Definição 2.22. (Notação $O(f(n))$). Uma função $g : \mathbb{N} \rightarrow \mathbb{R}$ é assintoticamente limitada superiormente por $f : \mathbb{N} \rightarrow \mathbb{R}$ se existem $c > 0$ e $n_c \in \mathbb{N}$ tais que $|g(n)| \leq c|f(n)|$ para todo $n \geq n_c$. O conjunto das funções assintoticamente limitadas superiormente por $f(n)$ é denotado por $O(f(n))$.

Durante o texto iremos escrever $f = O(g(n))$, porém fica subentendido que tal função f pertence ao conjunto de funções $O(g(n))$.

Definição 2.23. (Notação $o(1)$). Uma função $f : \mathbb{N} \rightarrow \mathbb{R}$ é assintoticamente nula se para todo $c > 0$ existe $n_c \in \mathbb{N}$ tal que $f(n)$ em valor absoluto nunca ultrapassa c_f de n_f em diante, isto é $|f(n)| \leq c$ para todo $n \geq n_c$. O conjunto das funções assintoticamente nulas é denotado por $o(1)$.

Definição 2.24. (Valor aproximado). Uma função $g(n)$ é aproximadamente igual a $f(n)$, ou seja, $g(n) \approx f(n)$ se $g(n) = (1 + o(1))f(n)$.

Definição 2.25. (Função Zeta de Riemann (Titchmarsh et al., 1986)) Seja $t \in \mathbb{C}$, a função $\zeta(t)$ é definida como:

$$\zeta(t) = \sum_{n=1}^{\infty} \frac{1}{n^t}$$

onde a função converge quando a parte real de t é maior que 1.

Definição 2.26. (Função Gama) A função gama $\Gamma(n)$ é definida como uma extensão da função fatorial para o conjunto dos números complexos e reais. Para qualquer $n \in \mathbb{C}$, com parte real positiva, a função gama é definida como uma integral,

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt$$

Os dois seguintes teoremas envolvendo a função gama serão relevantes nesta dissertação.

Teorema 2.1. (Hofstad, 2010) Se n for um número inteiro, a função gama se relaciona com a função fatorial da seguinte forma:

$$\Gamma(n) = (n-1)!$$

Teorema 2.2. (Fórmula de Duplicação de Legendre (Arfken e Weber, 1999)) Seja $z \in \mathbb{C}$, com parte real positiva, então,

$$\Gamma(z)\Gamma\left(z + \frac{1}{2}\right) = 2^{1-2z}\sqrt{\pi}\Gamma(2z)$$

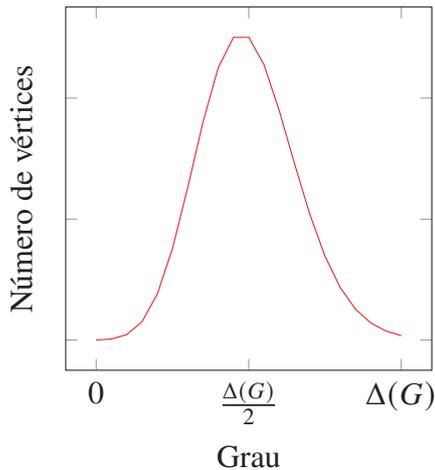


Figura 2.1: Distribuição normal

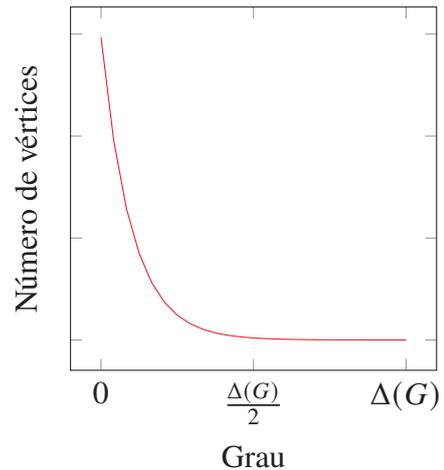


Figura 2.2: Distribuição Lei de Potência

2.4 GRAFOS DE LEI DE POTÊNCIA

Barabási e Albert (1999) enunciaram a hipótese de que a distribuição lei de potência presente em grafos advém do fato de que muitas redes na natureza tendem a começar pequenas e crescem com o tempo e, se o processo seguir uma regra denominada *ligação preferencial*, a lei de potência inevitavelmente aparecerá nestes grafos. Tal regra define que novos vértices que vão sendo adicionados ao grafo se conectam preferencialmente a vértices que já possuem bastante conexões.

Definição 2.27. (Lei de Potência). A lei de potência é uma relação entre dois valores, onde uma mudança em um dos valores resulta em uma mudança proporcional ao outro valor. Tal relação entre dois valores x e y pode ser escrita da seguinte forma

$$y = a \cdot x^k$$

onde a (constante de proporcionalidade) e k (expoente) são constantes.

Em grafos, a distribuição dos graus segue uma lei de potência quando a proporção de vértices com determinado grau k é

$$P_{deg}(k) \approx k^{-\beta}.$$

As Figuras 2.1 e 2.2 demonstram a diferença entre a distribuição normal e a distribuição lei de potência sobre grafos.

Enquanto na distribuição normal os vértices em sua maioria se concentram em uma média, numa distribuição lei de potência, que é uma contínua função decrescente, existe uma diferença grande entre a quantidade de vértices com grau muito alto se comparado à quantidade de vértices com grau muito baixo.

Como foi dito, inicialmente a teoria de grafos lei potência se desenvolveu utilizando ferramentas heurísticas não muito apropriadas para uma análise do ponto de vista da comunidade de computação teórica. Em consequência disso, pesquisadores propuseram modelos matemáticos formais que tornam essas análises possíveis. Existem modelos probabilísticos e determinísticos propostos na literatura. Alguns exemplos de modelos determinísticos podem ser encontrados em Chauhan et al. (2016) e Brach et al. (2016). Os modelos abordados na presente dissertação são

probabilísticos e podem ser classificados em três tipos: Os modelos de configuração (Bollobás, 1980; Aiello et al., 2001), modelos de grafo aleatório generalizado (Chung e Lu, 2002; Britton et al., 2006; Vignatti e da Silva, 2016) e os modelos de ligação preferencial (Barabási e Albert, 1999; Bollobás et al., 2001; Hofstad, 2010; Mitzenmacher e Upfal, 2017). Tais modelos serão definidos detalhadamente no Capítulo 3.

É importante destacar que embora haja uma definição formal para classificar um grafo como pertencente à classe de grafos de lei de potência, os modelos estudados não necessariamente geram tais grafos. Entretanto, como a maioria trabalha com conceitos de probabilidade, tais modelos possuem uma alta probabilidade de gerarem grafos com distribuição lei de potência.

Na seção seguinte, será apresentado o problema de otimização estudado no presente trabalho, para posteriormente serem apresentados os modelos analisados.

2.5 PROBLEMAS DE OTIMIZAÇÃO E ALGORITMOS GULOSOS

Na literatura, existem diversos estudos experimentais que evidenciam a eficiência de algoritmos gulosos aplicados a problemas de otimização combinatória em grafos de lei de potência (Park e Lee, 2001; Eubank et al., 2004; da Silva et al., 2013; Rogiski, 2016).

De uma maneira breve, *algoritmos gulosos* são aqueles que tomam decisões “míopes” com o objetivo de encontrar uma solução ótima global. Eles avaliam a melhor opção a cada passo, escolhendo a opção que pareça mais adequada. Tais algoritmos desconsideram implicações futuras de suas tomadas de decisões e também o problema maior como um todo, pois uma vez tomada a decisão, ela nunca é reconsiderada. Geralmente são implementados de maneira simples e intuitiva, porém muitos são utilizados para resolver problemas complexos.

Dentro do escopo desta dissertação, foi analisado o problema da clique máxima em grafos de lei de potência. O problema é descrito na subseção seguinte. O problema é NP-Difícil, ou seja, é difícil encontrar um algoritmo que compute a solução ótima em tempo polinomial de execução. Este problema é um problema clássico entre os problemas de otimização combinatória e é muito utilizado em diversas áreas de aplicação.

2.5.1 CLIQUE MÁXIMA

Abaixo definimos o problema computacional da clique máxima.

Definição 2.28. (CLIQUE MÁXIMA) O *problema da clique máxima*, denotado CLIQUE MÁXIMA, refere-se a: Dado como entrada um grafo G , obter como saída uma clique C de G tal que $|C| = \omega(G)$.

Dois algoritmos gulosos são bem conhecidos para esse problema. O primeiro algoritmo, conhecido como "Worst out", segue como: Dado um grafo G , escolha um vértice v com o menor grau e remova de G . Repita a mesma estratégia no grafo restante iterativamente até que ele se torne uma clique ou o número de vértices se torne menor que 1.

Algoritmo 1 "Worst Out"

Entrada: Grafo G .

Saída: Clique de G .

- 1: $n = |V(G)|$
 - 2: Ordene os vértices em $V(G)$ em ordem não-decrescente de grau, de tal modo que $d(v_i) \leq d(v_{i+1})$, para todo $i \in \{1, \dots, n\}$.
 - 3: **while** G não for uma clique e $|V(G)| > 1$ **do**
 - 4: Escolha um vértice v de grau mínimo em G .
 - 5: Remova v de G .
 - 6: **end while**
 - 7: **return** G
-

O outro algoritmo, conhecido como "Best in", possui uma abordagem simetricamente inversa ao algoritmo anterior. Dado um grafo G , ordene os vértices em ordem decrescente de grau. Iterativamente, para cada vértice, começando com o mais alto grau, remova os vértices não adjacentes ao selecionado até que uma clique seja encontrada. Retorne a maior clique do grafo encontrada.

Algoritmo 2 "Best In"

Entrada: Grafo G .

Saída: Clique de G .

- 1: $n = |V(G)|$
 - 2: Ordene os vértices em $V(G)$ em ordem decrescente de grau, de tal modo que $d(v_i) \geq d(v_{i+1})$, para todo $i \in \{1, \dots, n\}$.
 - 3: **while** G não for uma clique e $|V(G)| > 1$ **do**
 - 4: Escolha um vértice v de grau máximo em G .
 - 5: Remova os vértices $V(G) - N(v)$ de G .
 - 6: **end while**
 - 7: **return** G
-

2.6 CONCLUSÃO

Neste capítulo foram apresentados conceitos acerca de teoria da probabilidade, grafos, lei de potência, algoritmos gulosos e o problema da clique máxima. Estas definições são importantes para o entendimento da presente dissertação, pois servem como base para as conclusões alcançadas.

3 MODELOS DE GRAFOS ALEATÓRIOS

Formalmente, não faz sentido dizer que um dado objeto matemático, como um grafo, é aleatório. Quando dizemos que estamos utilizando grafos aleatórios, o que queremos dizer é que estamos utilizando grafos amostrados de um conjunto de grafos de acordo com uma distribuição de probabilidade. Neste capítulo são apresentados alguns modelos de grafos aleatórios normalmente utilizados no estudo de grafos de lei de potência. A ideia central neste contexto é definir os modelos aleatórios de maneira que os grafos amostrados possuam com alta probabilidade a distribuição de graus seguindo a lei de potência.

3.1 O MODELO $G_{n,p}$

O estudo de grafos aleatórios foi iniciado por Paul Erdős e Alfréd Rényi, numa sequência de publicações na década de 60 (Erdős e Rényi, 1960, 1961). Os autores introduziram um modelo, que é comumente referenciado na literatura como modelo Erdős-Rényi. O modelo é utilizado para gerar grafos aleatórios e possui duas variações. De maneira simplificada, o modelo $G_{n,p}$ define que a probabilidade p de existir uma aresta entre dois vértices de um grafo G com n vértices é escolhida de maneira uniformemente aleatória para todas as $\binom{n}{2}$ possíveis arestas.

Definição 3.1. (Modelo $G_{n,p}$). Seja \mathcal{G} o conjunto de todos os grafos possíveis sobre o conjunto de vértices $[n]$. O modelo $G_{n,p}$ gera um grafo $G \in \mathcal{G}$ com a seguinte probabilidade

$$Pr(G) = p^{|E(G)|} (1-p)^{\binom{n}{2}-|E(G)|}.$$

A outra variação, denotada por $G_{n,m}$, escolhe um grafo aleatoriamente de forma uniforme de uma coleção de grafos com um número m de arestas, ou seja, entre $\binom{n}{m}$ possibilidades. Em outras palavras, o modelo amostra um grafo $G \in \mathcal{G}$ com probabilidade $Pr(G) = 1/\binom{n}{m}$. Essa segunda variação do modelo é equivalente a se pensar da seguinte forma, dado um grafo G com nenhuma aresta, o modelo adiciona arestas de forma aleatória e uniforme até que $|E(G)| = m$.

Segundo Aiello et al. (2001), muitas aplicações práticas têm sido efetivamente modeladas por modelos de grafos aleatórios. Levando em conta a ausência de grandes quantidades de dados e recursos, durante boa parte do tempo em que desenvolveu-se pesquisa em teoria de grafos aleatórios, não era muito fácil fazer previsões a respeito de propriedades de grafos advindos de aplicações e se tais grafos poderiam ser modelados adequadamente utilizando os modelos de Erdős-Rényi. No entanto, com a aquisição de dados cada vez mais fácil e rápida, é aberta a possibilidade de estudar propriedades desses grandes grafos.

Outros modelos começaram a ser propostos na literatura, com o objetivo de capturar propriedades de diversos tipos de grafos por meio da utilização de grafos aleatórios. De maneira geral, são definidos processos capazes de produzir grafos com uma sequência de graus pré-definida com alta probabilidade.

Os modelos estudados no presente trabalho podem ser classificados em três grandes tipos: modelo de configuração, modelo de grafo aleatório generalizado (GRG) e o modelo de ligação preferencial. Alguns dos modelos propostos permitem que a distribuição dos graus seja configurada através de parâmetros, já outros, geram apenas uma certa distribuição característica pela maneira que foram implementados.

3.2 AJUSTANDO A DISTRIBUIÇÃO DE GRAUS EM GRAFOS ALEATÓRIOS

Nas próximas seções iremos apresentar modelos de grafos aleatórios que permitem que a distribuição dos graus dos vértices do grafo construído seja ajustada. Existem diversas maneiras de se construir grafos aleatórios ajustando os graus dos vértices, porém, no escopo da presente dissertação, escolhemos duas classes de modelos que são bastante utilizadas na literatura. Estas são a dos modelos de configuração e a dos modelos de grafo aleatório generalizado.

3.2.1 MODELO DE CONFIGURAÇÃO

O *modelo de configuração* foi introduzido em Bollobás (1980) com o objetivo de estudar grafos aleatórios com uma sequência determinada de grau. O modelo permite especificar qual é a distribuição dos graus dos vértices. Neste modelo os graus dos vértices é um parâmetro fixo utilizado na construção do grafo. Dada a sequência de graus d_1, d_2, \dots, d_n , representando o grau que os vértices terão no grafo construído no processo aleatório. Para construir o grafo, se um vértice i deve ter grau d_i então o vértice i recebe d_i “meias arestas” e de maneira uniformemente aleatória, o modelo escolhe duas metades e as liga formando uma aresta no grafo, iterando até o grafo não possuir mais as meias arestas.

No final da construção o grafo terá a distribuição dos graus que foi configurada. Uma característica do modelo, é que o número de meias arestas deve ser par, para assim ser possível ligar todas as meias arestas. Outra característica específica deste modelo é que o grafo construído pode conter *loops* e multiarestas, por exemplo, caso na escolha aleatória das meias arestas a se conectar o modelo escolha duas metades que partem do mesmo vértice gerando um *loop*. Um modelo de configuração pode construir um grafo lei de potência se a sequência de graus passada como parâmetro também seguir uma lei de potência.

Iremos apresentar o modelo de configuração proposto por Aiello et al. (2001), pois é um modelo bastante utilizado na literatura (Shen et al., 2012; Gast e Hauptmann, 2014; Gast et al., 2015; Eubank et al., 2004). Em Aiello et al. (2001), os autores propuseram um modelo de grafo aleatório, conhecido como modelo ACL, que estende o modelo de configuração, sendo capaz de construir grafos de lei de potência.

Definição 3.2. ((α, β) -Grafos de Lei de Potência). Um grafo (α, β) -GLP é um multigrafo $G_{\alpha, \beta}$ não direcionado, possivelmente contendo *loops* e com grau máximo $\Delta(G_{\alpha, \beta}) = \lfloor e^{\alpha/\beta} \rfloor$, onde para cada $i = 1, 2, \dots, \Delta$ e $G_{\alpha, \beta}$ possua y_i vértices com grau i , onde

$$y_i = \begin{cases} \lfloor \frac{e^\alpha}{i^\beta} \rfloor & , \text{ se } i > 1 \text{ ou } \sum_{i=1}^{\Delta} \lfloor \frac{e^\alpha}{i^\beta} \rfloor \text{ é par} \\ \lfloor e^\alpha \rfloor + 1 & , \text{ caso contrário.} \end{cases} \quad (3.1)$$

Na Equação 3.1 os valores i e y_i satisfazem $\log y_i = \alpha - \beta \log i$. Além disto, α é o logaritmo do tamanho do grafo e β é um parâmetro que permite guiar a intensidade da distribuição lei de potência presente nos graus dos vértices. O número n de vértices e o número m de arestas do grafo $G_{\alpha, \beta}$ podem ser calculados da seguinte forma:

$$n = \sum_{i=1}^{\lfloor e^{\alpha/\beta} \rfloor} \lfloor \frac{e^\alpha}{i^\beta} \rfloor \approx \begin{cases} \zeta(\beta) e^\alpha & , \text{ se } \beta > 1 \\ \alpha e^\alpha & , \text{ se } \beta = 1 \\ \frac{e^\alpha}{1-\beta} & , \text{ se } 0 < \beta < 1 \end{cases}$$

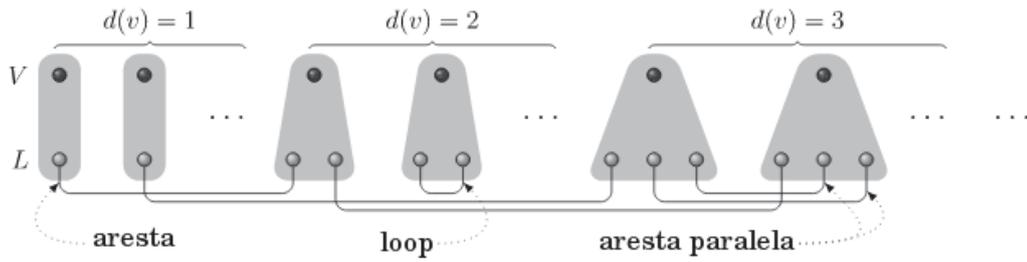


Figura 3.1: Construção do grafo no modelo $ACL(\alpha, \beta)$
 Fonte: (Gast e Hauptmann, 2014)

$$m = \frac{1}{2} \sum_{i=1}^{\lfloor e^{\alpha/\beta} \rfloor} i \lfloor \frac{e^\alpha}{i^\beta} \rfloor \approx \begin{cases} \frac{1}{2} \zeta(\beta - 1) e^\alpha & , \text{ se } \beta > 2 \\ \frac{1}{4} \alpha e^\alpha & , \text{ se } \beta = 2 \\ \frac{1}{2} \frac{e^{\frac{2\alpha}{\beta}}}{2-\beta} & , \text{ se } 0 < \beta < 2 \end{cases}$$

relembrando que $\zeta(t) = \sum_{n=1}^{\infty} \frac{1}{n^t}$ é a função Zeta de Riemann.

Definição 3.3. (Modelo $ACL(\alpha, \beta)$). No modelo $ACL(\alpha, \beta)$ a construção de (α, β) -GLP é a distribuição $P(\alpha, \beta)$ no conjunto de todos os (α, β) -GLP, que é obtida da seguinte forma:

1. Gere um conjunto L com $d(v)$ cópias distintas de cada vértice v .
2. Gere um emparelhamento aleatório nos elementos de L .
3. Para cada par de vértices u e v , o número de arestas ligando u e v em G é igual ao número de arestas em L que juntam cópias de u a cópias de v (veja a Figura 3.1 para um exemplo).

3.2.2 MODELO DE GRAFO ALEATÓRIO GENERALIZADO

O *modelo de grafo aleatório generalizado* (GRG) proposto por Britton et al. (2006) é uma adaptação do modelo $G_{n,p}$, apresentado na Definição 3.1. O modelo GRG tem como objetivo estender tal modelo, permitindo uma distribuição dos graus dos vértices ajustável. Uma vantagem do modelo é que as arestas são geradas de maneira independentes entre si, facilitando algumas análises. No modelo, cada vértice $i \in V(G)$ é relacionado a um peso w_i do vetor de pesos w , de maneira que a probabilidade das arestas serem incidentes a i sejam guiadas por esse peso.

Definição 3.4. (Modelo GRG). Dado um vetor de pesos $w = (w_1, w_2, \dots, w_n)$, o grafo aleatório com n vértices é construído tal que a probabilidade de um vértice i se conectar a um vértice j é igual a

$$Pr(ij \in E) = \frac{w_i w_j}{l_n + w_i w_j}, \forall i \neq j$$

onde $l_n = \sum_{i \in V} w_i$, é a soma dos pesos de todos os vértices do grafo.

Este modelo permite especificar a distribuição dos graus dos vértices, para isso, basta configurar os pesos para a distribuição desejada. Se os pesos seguirem uma distribuição lei de potência, os graus dos vértices tenderão a apresentar uma distribuição lei de potência.

Derivado do modelo GRG, em Vignatti e da Silva (2016) os autores propuseram uma adaptação para construção de grafos de lei de potência aleatórios. Como a configuração dos graus nesse tipo de modelo se dá pela definição dos pesos, os autores escolhem os pesos fazendo uso de uma adaptação do modelo proposto por Aiello et al. (2001), descrito na seção anterior. Essa estratégia foi adotada, pois os pesquisadores observaram a melhor eficácia em se realizar análises probabilísticas em um modelo que os eventos fossem independentes como no GRG. De maneira simplificada, um grafo é construído a partir de uma adaptação do modelo ACL, e os graus dos vértices são transformados na sequência de pesos que é usada como parâmetro para construção do modelo GRG. Como foi descrito anteriormente, é possível gerar uma distribuição lei de potência com o modelo ACL utilizando os parâmetros corretos, logo, o grafo construído pelo modelo GRG também constrói um grafo lei de potência.

Definição 3.5. (Modelo GRG-ACL). Considere a utilização do Modelo GRG, o vetor de pesos w é construído usando princípios semelhantes a Aiello et al. (2001). Seja $\alpha = \ln\left(\frac{|V|}{\zeta(\beta)}\right)$, $\beta > 2$ e $\Delta = \lfloor e^\alpha \rfloor$. Para cada $j = 1, 2, \dots, \Delta$, então

$$y_j = \begin{cases} \lfloor \frac{e^\alpha}{j^\beta} \rfloor & , \text{ se } n > 1 \\ \lfloor e^\alpha \rfloor & , \text{ caso contrário} \end{cases}$$

Agora w é construído a partir de y_j para cada $j = 1, 2, \dots, \Delta$. Desta maneira, existem y_j vértices com peso j (note que no modelo ACL existem y_j vértices com grau j).

Em Chung e Lu (2002) os autores apresentam uma outra implementação do modelo GRG.

Definição 3.6. (Modelo Chung-Lu). Dado um vetor de pesos $w = (w_1, w_2, \dots, w_n)$, o multigrafo aleatório com n vértices é construído tal que a probabilidade de um vértice i se conectar a um vértice j é igual a

$$Pr(ij \in E) = \frac{w_i w_j}{l_n}, \text{ onde } l_n = \sum_{i \in V} w_i.$$

Percebe-se que não é necessário que i e j sejam distintos, dessa forma, há possibilidade de *loops*.

3.3 MODELOS DE LIGAÇÃO PREFERENCIAL

Alguns grafos advindos de situações práticas, além de possuírem a distribuição dos graus dos vértices seguindo uma lei de potência, também apresentam duas propriedades quanto à evolução dos grafos. Uma dessas propriedades é que os grafos crescem continuamente pela adição de novos vértices, proporcionando a característica de crescimento da rede. A outra propriedade é que esses novos vértices que vão sendo adicionados, se conectam preferencialmente a vértices que já possuam bastante conexões (Barabási e Albert, 1999).

A principal diferença entre os modelos apresentados anteriormente e os *modelos de ligação preferencial* é que a dinâmica de criação de um grafo em um modelo de ligação preferencial tenta simular uma dinâmica real de crescimento de um grafo, onde os vértices são adicionados sequencialmente proporcionando o crescimento, e para cada vértice adicionado são conectadas uma ou mais arestas sobre o vértice.

Os novos vértices adicionados ao grafo recebem uma aresta para vértices já inseridos com probabilidade proporcional ao grau destes, ou seja, vértices com maior grau possuem maior

probabilidade de serem adjacentes ao novo vértice. Outros modelos como o de Erdős-Renyi também podem ser pensados como modelos que evoluem com o tempo, onde arestas são adicionadas a medida que o tempo evolui, porém, nestes modelos as probabilidades das arestas serem ou não adicionadas são iguais até o fim da construção do grafo.

Segundo Barabási e Albert (1999), modelos que definam o crescimento e a ligação preferencial para construção do grafo são capazes de reproduzir grafos observados em situações práticas, pois depois de muitos passos a proporção de vértices com um determinado grau deve obedecer a uma lei de potência. Os autores sugeriram o seguinte modelo:

Definição 3.7. (Modelo de Ligação Preferencial Barabási e Albert). Começando com um número pequeno (m_0) de vértices, a cada passo é adicionado um novo vértice com $m (\leq m_0)$ arestas que ligam o novo vértice a m vértices diferentes já presentes no grafo. Para incorporar a ligação preferencial, é assumido que a probabilidade de um novo vértice estar conectado a um vértice i depende do grau k_i daquele vértice, de modo que $\prod(k_i) = k_i / \sum_j k_j$. Após t etapas, o modelo leva a uma rede aleatória com $t + m_0$ vértices e $m.t$ arestas.

Tendo em vista a falta de formalização desse modelo, que foi baseado puramente em argumentos heurísticos, e não apresenta uma base matemática adequada, em Bollobás et al. (2001) os autores propõem um modelo de ligação preferencial apresentando a formalização necessária.

Definição 3.8. (Modelo de Ligação Preferencial de Bollobás et al.). Seja v_1, v_2, \dots, v_n o conjunto de vértices do grafo G . Por indução, é definido o seguinte processo de construção do grafo aleatório: $(G_m^t)_{t>0}$, onde m é o número de arestas adicionada a cada passo t . Observe que G_1^t é um grafo direcionado em $\{v_i : 1 \leq i \leq t\}$. O processo começa com G_1^0 , que é o grafo sem nenhum vértice ou com G_1^1 onde grafo possui um vértice e um loop. Dado G_1^{t-1} , o grafo G_1^t é construído adicionando um novo vértice v_t que recebe uma aresta para outro vértice v_i já presente no grafo, sendo i escolhido aleatoriamente seguindo a seguinte equação,

$$Pr(i = s) = \begin{cases} d_{G_1^{t-1}}(v_s) / (2t - 1) & 1 \leq s \leq t - 1 \\ 1 / (2t - 1) & s = t \end{cases}$$

onde $d_{G_1^{t-1}}(v_s)$ é o grau do vértice v_s no grafo G_1^{t-1} .

Este modelo segue o paradigma da ligação preferencial, pois o novo vértice v_t recebe uma aresta que incide em um determinado vértice v_i com probabilidade proporcional ao grau de v_i no instante anterior a t . Vale ressaltar que, para $m > 1$, são adicionadas m arestas por instante t , uma de cada vez, já sendo contabilizadas nos cálculos para as próximas arestas.

No livro de Mitzenmacher e Upfal (2017), os autores apresentam uma versão própria do modelo de ligação preferencial. O modelo proposto foi baseado em um exemplo no qual a maioria dos autores se baseia para explicar o paradigma da ligação preferencial, que é um modelo para a World Wide Web, onde os vértices são páginas e as arestas direcionadas são links, que levam de uma página a outra.

O modelo possui uma configuração inicial de 2 vértices e a cada passo que o grafo vai sendo construído, um novo vértice é adicionado ao grafo junto com uma aresta que incide em outro vértice já existente. O modelo é definido somente para uma aresta sendo adicionada a cada instante para simplificar as análises. Para implementar o paradigma de ligação preferencial, a nova aresta é incidente preferencialmente a um vértice com maior grau.

Definição 3.9. (Modelo de Ligação Preferencial Mitzenmacher e Upfal). Dado γ , um novo vértice é adicionado ao grafo com uma aresta que tem probabilidade γ de incidir a um vértice escolhido de maneira aleatória e uniforme e probabilidade $1 - \gamma$ de copiar a ponta de outra aresta já existente.

Se a probabilidade for $1 - \gamma$, então a nova aresta incide em um vértice existente de maneira proporcional ao grau de entrada do vértice escolhido. Os seguintes Lemas 3.1 e 3.2 e o Teorema 3.1 foram apresentados em Mitzenmacher e Upfal (2017).

Lema 3.1. Seja $X_j(t)$ o número de vértices com grau de entrada j quando existem t vértices no grafo, então a probabilidade de X_j crescer do tempo t para o próximo tempo $t + 1$ é,

$$Pr(X_j(t+1) > X_j(t)) = \frac{\gamma X_{j-1}(t)}{t} + \frac{(1-\gamma)(j-1)X_{j-1}(t)}{t}, \forall j \geq 1.$$

Demonstração. O primeiro termo se refere a probabilidade da nova aresta incidir em um vértice aleatoriamente com grau de entrada $j - 1$, elevando o grau de tal vértice para j e aumentando a quantidade de elementos em X_j . O modelo possui probabilidade γ de se ligar a um vértice aleatoriamente multiplicado pela quantidade $X_{j-1}(t)$ de vértices com grau $j - 1$, dividido pelo tamanho do espaço amostral que neste caso é total de vértices t .

O segundo termo se refere a probabilidade $1 - \gamma$ do modelo decidir copiar uma aresta do grafo multiplicado pela quantidade $(j - 1)X_{j-1}(t)$ de arestas de vértices com grau $j - 1$, tudo isso dividido pelo total de arestas que também é igual a t . \square

Lema 3.2. A probabilidade de X_j decrescer do tempo t para o próximo tempo $t + 1$ é,

$$Pr(X_j(t+1) < X_j(t)) = \frac{\gamma X_j(t)}{t} + \frac{(1-\gamma)jX_j(t)}{t}, \forall j \geq 1$$

Demonstração. O primeiro termo se refere a probabilidade da nova aresta incidir em um vértice aleatoriamente com grau de entrada j , elevando o grau de tal vértice para $j + 1$ e diminuindo a quantidade de elementos em X_j . De maneira equivalente a prova anterior, esse termo é a probabilidade γ do modelo decidir se ligar a um vértice aleatoriamente multiplicado pela quantidade $X_j(t)$ de vértices com grau j , dividido pelo total de vértices que é t .

O segundo termo se refere a probabilidade $1 - \gamma$ do modelo decidir copiar uma aresta do grafo multiplicado pela quantidade $jX_j(t)$ de arestas de vértices com grau j , dividido pelo total de arestas t . \square

Teorema 3.1. Seja $\Delta(X_j(t)) = X_j(t+1) - X_j(t)$, representando a variação do número de vértices com grau de entrada j do tempo t para $t + 1$, então, o valor esperado desse aumento, dado os valores que provocam o crescimento e decréscimo é

$$E[\Delta(X_j(t)) | X_{j-1}(t), X_j(t)] = \frac{\gamma X_{j-1}(t)}{t} + \frac{(1-\gamma)(j-1)X_{j-1}(t)}{t} - \frac{\gamma X_j(t)}{t} - \frac{(1-\gamma)jX_j(t)}{t}$$

Demonstração. Com base nos Lemas 3.1 e 3.2,

$$\begin{aligned} E[\Delta(X_j(t)) | X_{j-1}(t), X_j(t)] &= Pr(X_j(t+1) > X_j(t)) - Pr(X_j(t+1) < X_j(t)) \\ &= \frac{\gamma X_{j-1}(t)}{t} + \frac{(1-\gamma)(j-1)X_{j-1}(t)}{t} - \frac{\gamma X_j(t)}{t} - \frac{(1-\gamma)jX_j(t)}{t} \end{aligned}$$

\square

Observe que como o modelo adiciona a cada passo somente um vértice e uma nova aresta partindo desse vértice, o grafo construído é uma árvore. Nesta dissertação estamos interessados em um modelo mais completo. Uma alternativa seria aumentar o número de vértices adicionados em cada passo e/ou aumentar o número de arestas que são relacionadas com os vértices. O modelo proposto por Hofstad (2010) supre tais necessidades.

Definição 3.10. (Modelo de Ligação Preferencial Hofstad ($PA_t(m, \delta)$)) O modelo produz uma sequência de grafos denotada por $\{PA_t(m, \delta)\}_{t=1}^{\infty}$, onde para cada t , o modelo produz um grafo de t vértices e mt arestas. O parâmetro δ permite ajustar a intensidade da distribuição lei de potência nos graus dos vértices do grafo construído pelo processo aleatório. Note que um grafo gerado pelo modelo $PA_1(1, \delta)$ consiste em um grafo amostrado com um único vértice com um loop. O conjunto de vértices de um grafo amostrado de $PA_1(1, \delta)$ é igual a $\{v_1^{(1)}, v_2^{(1)}, \dots, v_t^{(1)}\}$. Consecutivamente em $PA_t(1, \delta)$, a regra de crescimento para se obter $PA_{t+1}(1, \delta)$, sendo $D_i(t)$ o grau de um vértice i em $PA_t(1, \delta)$, é descrita como segue:

1. Adicione um vértice v_{t+1} contendo uma aresta em $PA_{t+1}(1, \delta)$.
2. A aresta é conectada ao próprio vértice v_{t+1} com probabilidade $(1+\delta)/(t(2+\delta)+(1+\delta))$, ou é conectada a outro vértice $v_i \in PA_t(1, \delta)$ com probabilidade $(D_i(t) + \delta)/(t(2+\delta) + (1+\delta))$, onde $\delta > -1$ é um parâmetro do modelo.

Considere dois vértices u e v , utilizaremos a seguinte notação $u \rightarrow v$ para o evento do aparecimento de uma aresta entre os vértices u e v . Por exemplo, a probabilidade descrita no passo 2 da Definição 3.10 pode ser apresentada como:

$$Pr(v_{t+1}^{(1)} \rightarrow v_i^{(1)} | PA_t(1, \delta)) = \begin{cases} \frac{1+\delta}{t(2+\delta)+1+\delta}, & \text{se } i = t + 1 \\ \frac{D_i(t)+\delta}{t(2+\delta)+1+\delta}, & \text{se } i \in [t] \end{cases}$$

Para valores de $m > 1$, o modelo é definido usando como base o modelo com $m = 1$, da seguinte forma:

1. Comece com $PA_{mt}(1, \delta/m)$, sendo os vértices do grafo denotados por $\{v_1^{(1)}, v_2^{(1)}, \dots, v_{mt}^{(1)}\}$.
2. Identifique $\{v_1^{(1)}, v_2^{(1)}, \dots, v_m^{(1)}\}$ em $PA_{mt}(1, \delta/m)$ para ser $v_1^{(m)}$ em $PA_t(m, \delta)$.
3. Identifique $\{v_{m+1}^{(1)}, v_{m+2}^{(1)}, \dots, v_{2m}^{(1)}\}$ em $PA_{mt}(1, \delta/m)$ para ser $v_2^{(m)}$ em $PA_t(m, \delta)$.
4. E assim consecutivamente até $\{v_{(j-1)m+1}^{(1)}, v_{(j-1)m+2}^{(1)}, \dots, v_{jm}^{(1)}\}$ em $PA_{mt}(1, \delta/m)$ ser $v_j^{(m)}$ em $PA_t(m, \delta)$ (veja a Figura 3.2 para um exemplo).

O procedimento acima é chamado de colapso dos vértices, onde uma aresta em $PA_{mt}(1, \delta/m)$ está ligada ao vértice $v_k^{(1)}$ com probabilidade proporcional ao seu peso, onde o peso é igual ao seu grau mais δ/m . Portanto, uma aresta em $PA_t(m, \delta)$ é ligada a um vértice $v_j^{(m)}$ com probabilidade proporcional à soma dos pesos dos vértices $\{v_{(j-1)m+1}^{(1)}, v_{(j-1)m+2}^{(1)}, \dots, v_{jm}^{(1)}\}$ de $PA_{mt}(1, \delta/m)$. Como a soma dos graus dos vértices $\{v_{(j-1)m+1}^{(1)}, v_{(j-1)m+2}^{(1)}, \dots, v_{jm}^{(1)}\}$ é igual ao grau do vértice $v_j^{(m)}$, esta probabilidade é proporcional ao grau de $v_j^{(m)}$ mais δ . É importante observar que, durante a construção do grafo, para valores de $m \geq 2$, os graus são atualizados após cada aresta ser adicionada ao grafo.

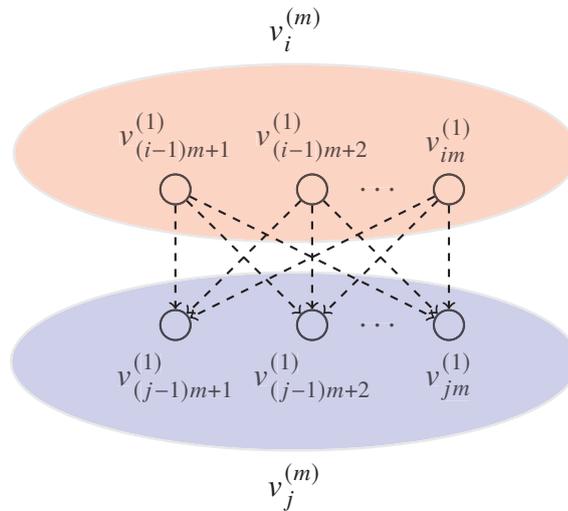


Figura 3.2: Construção do grafo para $m > 1$ no modelo $PA_t(m, \delta)$

Em Hofstad (2010), o autor também analisa o grau esperado de um vértice fixo no grafo. O modelo é diferente do modelo proposto por Mitzenmacher e Upfal (2017), mas ainda sim é possível perceber semelhanças nas análises dos modelos. Uma diferença é que o Modelo de Ligação Preferencial Mitzenmacher e Upfal permite apenas a adição de uma aresta em cada tempo da construção do grafo aleatório.

Teorema 3.2. Seja $m = 1$ e $\delta > -1$, o valor esperado do grau do i -ésimo vértice no tempo t é igual a

$$E[D_i(t) + \delta] = (1 + \delta) \frac{\Gamma(t + 1)\Gamma(i - 1/(2 + \delta))}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(i)}$$

Para o presente trabalho, optamos por trabalhar com um caso particular do modelo $PA_t(m, \delta)$. Mais precisamente iremos considerar sempre $\delta = 0$ conforme exposto abaixo.

Definição 3.11. (Modelo de Ligação Preferencial ($PA_t(m)$)) O modelo segue as definições do modelo de Hofstad, porém a probabilidade de um novo vértice $v_{t+1}^{(1)}$ adicionado no grafo no tempo $t + 1$ se conectar a outro vértice $v_i^{(1)}$ é igual a

$$Pr(v_{t+1}^{(1)} \rightarrow v_i^{(1)} | PA_t(1)) = \begin{cases} \frac{1}{2t+1}, & \text{se } i = t + 1 \\ \frac{D_i(t)}{2t+1}, & \text{se } i \in [t] \end{cases}$$

Corolário 3.1. Segue do Teorema 3.2, para o caso de $\delta = 0$, o valor esperado do grau do i -ésimo vértice no tempo t é igual a

$$E[D_i(t)] = \frac{\Gamma(t + 1)\Gamma(i - 1/2)}{\Gamma(t + 1/2)\Gamma(i)}$$

3.4 CONCLUSÃO

Neste capítulo foram apresentados os modelos de grafos aleatórios analisados na presente dissertação. Decidimos separar este conceito dos demais com o objetivo de dar melhor ênfase ao

assunto. Os modelos de grafos aleatórios, além de serem importantes para o entendimento desta dissertação, também guiam nossas análises buscando evidências acerca do bom desempenho de algoritmos gulosos em grafos de lei de potência.

4 PROBLEMAS DE OTIMIZAÇÃO EM GRAFOS DE LEI DE POTÊNCIA

Neste capítulo serão apresentados resultados envolvendo problemas de otimização em grafos de lei de potência. Na Seção 4.1 revisaremos resultados da literatura que lidam com os problemas da coloração de vértices, clique máxima, cobertura mínima por vértices, conjunto dominante e conjunto independente máximo. Na Seção 4.2 apresentamos os resultados originais desta dissertação, que envolvem o problema da clique máxima.

4.1 REVISÃO DA LITERATURA

Nesta seção é apresentada a revisão da literatura referente a trabalhos investigando o comportamento de algoritmos para problemas de otimização em grafos de lei de potência. Na Seção 4.1.1 apresentamos trabalhos experimentais e na Seção 4.1.2 apresentamos trabalhos teóricos.

4.1.1 ESTUDOS EXPERIMENTAIS

Nesta seção apresentamos trabalhos experimentais mostrando que heurísticas gulosas atingem soluções muito próximas da ótima em grafos de lei de potência, para alguns problemas de otimização como o da coloração de vértices, clique máxima e o da cobertura mínima por vértices. Isto explica em parte porque problemas computacionalmente difíceis em algumas circunstâncias admitem soluções eficientes na prática.

Eubank et al. (2004) realizaram um estudo sobre grafos advindos de interações sociais. O grafo foi gerado utilizando a abordagem de geração de grafos aleatórios proposta por Aiello et al. (2001). Os autores utilizaram de heurísticas gulosas para obter algoritmos eficientes para o problema do conjunto dominante no grafo. Em Park e Lee (2001), os autores propuseram uma nova abordagem para prevenção de ataques (DDoS), e mostraram que há uma relação entre a performance da abordagem e a distribuição lei de potência na topologia do grafo estudado.

Em da Silva et al. (2013), os autores realizaram uma análise do problema da cobertura mínima por vértices, utilizando-se de 25 grafos de lei de potência provenientes de diversas áreas. Os grafos apresentam uma grande diferença no número de vértices e arestas de um para outro, permitindo uma análise profunda da eficiência da abordagem gulosa. Para realizar a análise, além de utilizar o algoritmo guloso, também foi utilizado um algoritmo de 2-aproximação, pois durante as execuções não foi possível os cálculos para todos os grafos, em particular os com maiores números de vértices e arestas. A partir disto, foi possível comparar os resultados da heurística gulosa com os obtidos pelo algoritmos de aproximação mostrando melhores resultados pelo guloso.

Os problema da coloração de vértices e da clique máxima foram analisados detalhadamente em Rogiski (2016), onde a autora investiga os limitantes do problema da clique máxima para se obter limitantes inferiores para o problema da coloração.

4.1.2 ESTUDOS TEÓRICOS

Na literatura existem também alguns estudos teóricos investigando por que problemas de otimização podem ser mais fáceis de resolver em grafos de lei de potência do que em grafos em geral. Estes estudos buscam dar sustentação matemática aos resultados experimentais ao bom desempenho de algoritmos para dois problemas de otimização em grafos de lei de potência, mais

precisamente, os problemas envolvidos são o problema de cobertura por vértices e o problema do conjunto dominante. A grande maioria dos autores se utilizam da característica de que grafos de lei de potência têm poucos vértices com grau muito alto para aplicar heurísticas gulosas nos mesmos. A presente dissertação se insere neste contexto, investigando o problema da clique máxima em grafos de lei de potência.

No primeiro trabalho desta linha, Janson et al. (2010) estudam o tamanho da maior clique presente em grafos lei de potência, provando que este tamanho difere dependendo do valor do parâmetro α utilizado pelo modelo que construiu o grafo. A partir disso, os autores demonstram o seguinte teorema:

Teorema 4.1. (Janson et al., 2010) Seja $\omega(G(n, \alpha))$ o tamanho da maior clique de um grafo construído por $G(n, \alpha)$, então, para qualquer $\alpha > 0$ existe um algoritmo que, com alta probabilidade, encontra em $G(n, \alpha)$ uma clique de tamanho $(1 - o(1))\omega(G(n, \alpha))$ em tempo polinomial.

Ainda os autores discutem em como o tamanho da maior clique é em grafos gerados por modelos de ligação preferencial, onde segundo eles, este tamanho é limitado superiormente por uma constante absoluta ou cresce muito lentamente com n , fazendo com que o tamanho da maior clique seja pequeno. Até onde entendemos não parece ser direta a transposição do resultado dos autores para o modelo em que estamos trabalhando nesta dissertação.

Em Shen et al. (2012) os autores mostram que o problema do conjunto dominante e do conjunto independente máximo são mais fáceis de serem computados em grafos de lei de potência se utilizando de heurísticas gulosas. Para demonstrar isso, os autores utilizam o modelo $ACL(\alpha, \beta)$ apresentado na Definição 3.3. Mais precisamente, os autores demonstram os seguintes teoremas:

Teorema 4.2. (Shen et al., 2012) Seja um grafo amostrado no modelo $ACL(\alpha, \beta)$ com parâmetro $\beta > 2$. Então, o algoritmo para o problema do conjunto dominante, que seleciona os vértices de maneira gulosa em ordem decrescente de graus alcança um fator de aproximação de

$$\left(\zeta(\beta) - \frac{1}{2}\right) / \left(\zeta(\beta) - \sum_{i=1}^{t_0} \frac{1}{i^\beta}\right)$$

onde t_0 é uma constante que satisfaz,

$$\sum_{i=t_0}^{\Delta} i \lfloor \frac{e^\alpha}{i^\beta} \rfloor \geq \sum_{i=1}^{t_0} \lfloor \frac{e^\alpha}{i^\beta} \rfloor.$$

Teorema 4.3. (Shen et al., 2012) Seja um grafo amostrado no modelo $ACL(\alpha, \beta)$ com parâmetro $\beta > 1$. Para o problema do conjunto independente máximo, o algoritmo apresentado pelos autores tem fator de aproximação de

$$\frac{1}{2\zeta(\beta)}.$$

Em Gast e Hauptmann (2014), os autores desenvolveram um algoritmo de aproximação para o problema da cobertura mínima por vértices em grafos de lei de potência, com uma taxa de aproximação esperada menor que 2. Observe que, para o caso geral, conjectura-se não ser possível atingir tal fator de aproximação. Para isso, os autores utilizaram grafos gerados a partir do modelo proposto por Aiello et al. (2001) que gera grafos (α, β) -GLP.

A estratégia dos autores é estimar o número esperado de vértices que são adjacentes aos vértices de grau 1 e 2. Os autores utilizam um algoritmo de programação linear com arredondamento que retorna os conjuntos $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ de vértices de forma que \mathcal{X}_1 obrigatoriamente deve estar em uma cobertura ótima, \mathcal{X}_3 não deve estar nessa cobertura ótima e \mathcal{X}_2 é o conjunto de vértices que deve se decidir estar ou não na solução ótima. Os autores provam o seguinte resultado

Teorema 4.4. (Gast e Hauptmann, 2014) Seja G um grafo amostrado no modelo $ACL(\alpha, \beta)$ e seja $V^* = \bigcup_{v:d(v) \in \{1,2\}} (\{v\} \cup N(v))$, o conjunto de vértices de grau 1, 2 e seus vizinhos:

$$E[\mathcal{X}_2(V^*)] \geq \left(\frac{e^\alpha}{2^\beta}\right) \left(\frac{\zeta(\beta) - 1 - \frac{1}{2^\beta}}{\zeta(\beta - 1)}\right)$$

Com isso, os autores alcançam o seguinte fator de aproximação utilizando um algoritmo guloso:

Teorema 4.5. (Gast e Hauptmann, 2014) Seja um grafo amostrado no modelo $ACL(\alpha, \beta)$ com parâmetro $\beta > 2$, o algoritmo apresentado pelos autores para o problema da cobertura mínima por vértices tem fator de aproximação de

$$2 - \frac{\zeta(\beta) - 1 - \frac{1}{2^\beta}}{2^\beta \zeta(\beta - 1) \zeta(\beta)}$$

Em Gast et al. (2015), os autores provam os primeiros limites inferiores para o problema do conjunto dominante mínimo em grafos (α, β) -GLP. Embora acredite-se que não exista algoritmo de aproximação com fator melhor que $(1 - o(1)) \ln(n)$ para o problema do conjunto dominante mínimo (Feige, 1998). Ainda os autores apresentam uma nova aproximação dos limites superior e inferior para o problema, se assumirmos que o grafo é amostrado com uma distribuição lei de potência do modelo proposto por Aiello et al. (2001).

Teorema 4.6. (Gast et al., 2015) Seja $d = \min\{d' | \text{vol}([d', \Delta]) > e^\alpha\}$ e seja $\text{vol}([a, b]) = \sum_{j=a}^b j \lfloor \frac{e^\alpha}{j^\beta} \rfloor$. Seja um grafo amostrado no modelo $ACL(\alpha, \beta)$ com parâmetro $2 < \beta \leq 2,729$, o algoritmo apresentado pelos autores para o problema do conjunto dominante mínimo em (α, β) -GLP tem fator de aproximação de

$$\frac{\zeta(\beta) - 1}{\zeta(\beta) - \sum_{j=1}^{d-1} \frac{1}{j^\beta}}$$

Teorema 4.7. (Gast et al., 2015) Seja um grafo amostrado no modelo $ACL(\alpha, \beta)$ com parâmetro $\beta > 2,729$, o algoritmo para o problema do conjunto dominante mínimo tem fator de aproximação de

$$\frac{\zeta(\beta) - \frac{\zeta(\beta-1)}{2}}{1 - \frac{\zeta(\beta-1)}{2}}$$

Em Vignatti e da Silva (2016) os autores também estudam a aproximabilidade para o problema da cobertura mínima por vértices em grafos de lei de potência. Os autores utilizam um algoritmo de 2-aproximação junto com uma etapa de pré-processamento. Para esta análise, os autores utilizam o modelo descrito na Definição 3.5, porém mostram que os resultados seguem

válidos para o modelo proposto por Chung e Lu (2002). Com isso obtiveram um fator de aproximação esperado melhor que o obtido por Gast e Hauptmann (2014), além de utilizarem um algoritmo mais simples. Para isso, os autores calculam o tamanho do conjunto de vértices adjacentes aos vértices de grau 1, da seguinte maneira:

Teorema 4.8. (Vignatti e da Silva, 2016) Seja V_1 o conjunto de vértices de grau 1 e seja Li a função polilogarítmica definida pela série $Li_s(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^s}$, então,

$$E[|N(V_1)|] \gtrsim |V| \left(1 - \frac{Li_{\beta}(1/e)}{\zeta(\beta)} - \frac{Li_{\beta-1}(1/e)}{\zeta(\beta)} \right) \left(1 - \frac{Li_{\beta}(e^{-\rho(\beta)})}{\zeta(\beta)} \right)$$

O algoritmo utilizado pelos autores se aproveita do fato de que a quantidade de vizinhos dos vértices de grau 1 é alta, e aplica um algoritmo de 2-aproximação nos vértices do grafo exceto os de grau 1 e seus vizinhos. Note que a inserção de vizinhos de vértices de grau 1 na solução, não altera a solução ótima. Quando aplicado em grafos de lei de potência gerados pelos modelos aleatórios especificados o fator de aproximação do algoritmo é $\left(1 + \frac{Li(e^{-\rho(\beta)})}{\zeta(\beta)} \right)$.

Os estudos apresentados serviram como base para o presente trabalho, onde seguimos as metodologias dos trabalhos teóricos para realizar as análises dos problemas de otimização e dos modelos de grafos aleatórios.

4.2 RESULTADOS ORIGINAIS

Nesta seção são apresentados os resultados obtidos nesta dissertação. O objetivo destes resultados é prover sustentação matemática para o bom desempenho do algoritmo guloso apresentado na Seção 2.5.1 quando aplicado ao problema da clique máxima em grafos de lei de potência. O modelo de grafos aleatórios que utilizamos é o modelo $PA_t(m)$, descrito na Definição 3.11.

Ressaltamos que o resultado obtido pode ser aplicado tanto no contexto em que temos um grafo lei de potência criado dinamicamente quanto em um contexto tradicional, em que o grafo de entrada é fixo. Isso é possível, pois o modelo utilizado não apenas provê um grafo lei de potência obtido aleatoriamente, mas também captura a criação de um grafo lei de potência supondo-se que o mesmo seja construído usando o paradigma de ligação preferencial.

Em linhas gerais o nosso resultado é o seguinte: dado ℓ e k , a probabilidade de uma clique de tamanho ℓ conter os vértices $\{v_k, v_{k+1}, \dots, v_{k+\ell-1}\}$ do grafo decai exponencialmente como uma função que depende de ℓ e k . Esta é a principal ligação entre o nosso resultado e a qualidade de um algoritmo guloso em grafos de lei de potência, uma vez que algoritmos gulosos procuram utilizar primeiramente vértices de grau alto em sua solução. Note que vértices de grau alto significa k pequeno e portanto probabilidade alta de estar contido numa clique de tamanho ℓ .

De maneira um pouco mais precisa, dado ℓ , para um grafo G amostrado do modelo $PA_t(m)$, considere os vértices $v_k, v_{k+1}, \dots, v_{k+\ell-1}$ para qualquer $k < t$, lembrando que $d(v_k) > d(v_{k+1}) > \dots > d(v_{k+\ell-1})$ e também considere o evento $E_{k,\ell}$, em que $G[\{v_k, \dots, v_{k+\ell-1}\}]$ é um K_{ℓ} . O que vamos mostrar é que :

$$Pr(E_{k,\ell}) = O\left(\frac{\ell}{k^{\binom{\ell}{2}}}\right)$$

Para analisarmos o problema da clique máxima no modelo $PA_t(m)$, precisamos primeiramente calcular a probabilidade de aparecer uma clique no modelo. Portanto, primeiramente

precisamos analisar como são construídas as arestas. Segundo a Definição 3.11, para $m = 1$, dois vértices v_i e v_j possuem uma aresta entre si com probabilidade:

$$Pr(v_i^{(1)} \rightarrow v_j^{(1)}) = \frac{D_j(i-1)}{2(i-1)+1}, \forall i > j \geq 1 \quad (4.1)$$

Porém como já foi discutido anteriormente, um grafo amostrado no modelo $PA_t(m)$ com valor de $m = 1$, como acontece também no modelo de Mitzenmacher da Seção 3.3, acaba sendo sempre uma árvore. Neste caso, não apresentaria nenhuma clique de tamanho maior do que 2, o que torna o modelo pouco interessante em nosso cenário.

Para resolver isto, é preciso expandir essa definição para qualquer valor de $m > 1$. Como para tais valores, o modelo gera um grafo com $m = 1$ e depois os vértices são agrupados m a m , basta somarmos as probabilidades das aparições das arestas de cada um dos vértices de um grupo se conectarem aos vértices de outro para sabermos a probabilidade de haver uma aresta entre os dois vértices resultantes de cada grupo. Generalizando a equação (4.1) para valores de $m > 1$:

Lema 4.1. Dado um grafo G do modelo $PA_t(m)$ e $v_i, v_j \in V(G)$ e $m > 1$, então

$$Pr(v_i^{(m)} \rightarrow v_j^{(m)}) \leq \sum_{l=(i-1)m+1}^{im} \sum_{k=(j-1)m+1}^{jm} \frac{D_k(l-1)}{2(l-1)+1} \quad (4.2)$$

Demonstração. A probabilidade de dois vértices se conectarem para valores de $m > 1$ é a combinação de todas as possíveis arestas entre os vértices em $m = 1$ que durante a construção do grafo são agrupados para se tornarem os vértices em $m > 1$. Logo,

$$\begin{aligned} Pr(v_i^{(m)} \rightarrow v_j^{(m)}) &\leq Pr(v_{(i-1)m+1}^{(1)} \rightarrow v_{(j-1)m+1}^{(1)}) + Pr(v_{(i-1)m+1}^{(1)} \rightarrow v_{(j-1)m+2}^{(1)}) + \dots \\ &+ Pr(v_{(i-1)m+1}^{(1)} \rightarrow v_{jm}^{(1)}) \\ &+ Pr(v_{(i-1)m+2}^{(1)} \rightarrow v_{(j-1)m+1}^{(1)}) + Pr(v_{(i-1)m+2}^{(1)} \rightarrow v_{(j-1)m+2}^{(1)}) + \dots \\ &+ Pr(v_{(i-1)m+1}^{(1)} \rightarrow v_{jm}^{(1)}) \\ &\dots \\ &+ Pr(v_{im}^{(1)} \rightarrow v_{(j-1)m+1}^{(1)}) + Pr(v_{im}^{(1)} \rightarrow v_{(j-1)m+2}^{(1)}) + \dots + Pr(v_{im}^{(1)} \rightarrow v_{jm}^{(1)}) \\ &= \sum_{l=(i-1)m+1}^{im} \sum_{q=(j-1)m+1}^{jm} Pr(v_l^{(1)} \rightarrow v_q^{(1)}) \end{aligned}$$

de acordo com a equação 4.1:

$$Pr(v_i^{(m)} \rightarrow v_j^{(m)}) \leq \sum_{l=(i-1)m+1}^{im} \sum_{q=(j-1)m+1}^{jm} \frac{D_q(l-1)}{2(l-1)+1}$$

□

A variável aleatória mais importante em nossa análise é a que nos dá o grau esperado de cada um dos vértices v_i do grafo. O Teorema 3.2 nos provê tal valor esperado.

A principal dificuldade de trabalhar com as probabilidades de criação de arestas obtidas no Lema 4.1 é o fato de que o valor esperado $D_q(l-1)$ é definido em termos da função Gama.

Um dos nossos primeiros objetivos vai ser lidar com esta função ao calcularmos o valor esperado de cada vértice do grafo. Faremos isso no Lema 4.3. Entretanto precisamos antes provar o seguinte lema.

Lema 4.2. Sejam $i, t \in \mathbb{N}$ tal que $i > 1$ e $t > i + 1$, então:

$$\frac{(i \cdot (i+1) \cdot \dots \cdot (t-1))2^{t-i+1}i}{((2i-1) \cdot (2i+1) \cdot \dots \cdot (2t-1))} < 1$$

Demonstração. Provaremos por indução em $t - i$.

Base: $t - i = 2$

$$\begin{aligned} & \frac{i(i+1)2^3i}{(2i-1)(2i+1)(2i+3)} \\ &= \frac{8(i^2+i)i}{(4i^2-1)(2i+3)} \\ &= \frac{8i^3+8i^2}{8i^3+12i^2-2i-3} \end{aligned}$$

como $i > 2$,

$$\frac{8i^3+8i^2}{8i^3+12i^2-2i-3} < 1$$

Hipótese: Suponha que, para $t - i = k$, o seguinte seja verdade:

$$\frac{(i \cdot (i+1) \cdot \dots \cdot (t-1))2^{t-i+1}i}{((2i-1) \cdot (2i+1) \cdot \dots \cdot (2t-1))} < 1$$

ou, de maneira equivalente,

$$\frac{(i \cdot (i+1) \cdot \dots \cdot (i+k-1))2^{k+1}i}{((2i-1) \cdot (2i+1) \cdot \dots \cdot (2i+(2k+1)))} < 1 \quad (4.3)$$

Passo indutivo: Nosso objetivo agora é provar que a fórmula 4.3 também é válida para $k+1$.

Note que $2(i+k) < 2(i+k+1)$ e portanto,

$$\begin{aligned} & \frac{(i \cdot (i+1) \cdot \dots \cdot (i+k-1))2^{k+1}i}{((2i-1) \cdot (2i+1) \cdot \dots \cdot (2i+(2k+1)))} \cdot \frac{2(i+k)}{2(i+k+1)} < 1 \\ & \frac{(i \cdot (i+1) \cdot \dots \cdot (i+k-1))2^{k+1}i}{((2i-1) \cdot (2i+1) \cdot \dots \cdot (2i+(2k+1)))} \cdot \frac{(i+k)2}{2i+(2k+2)} < 1 \\ & \frac{(i \cdot (i+1) \cdot \dots \cdot (i+k-1) \cdot (i+k))2^{k+2}i}{((2i-1) \cdot (2i+1) \cdot \dots \cdot (2i+(2k+1)) \cdot (2i+(2k+2)))} < 1 \end{aligned}$$

Ou seja, a fórmula 4.3 também vale para $t - i = k + 1$. □

A seguir, provamos um lema que nos provê um limitante superior para o grau esperado de um dado vértice em função de i e de t , ou seja do momento que o vértice foi inserido e do momento t do processo do processo aleatório de geração do grafo.

Lema 4.3. Sejam $i, t \in \mathbb{N}$ tal que $i > 1$ e $t > i + 1$, então:

$$E[D_i(t)] \leq \frac{t}{i}$$

Demonstração. Vamos provar que a seguinte desigualdade é verdadeira:

$$E[D_i(t)] \cdot \frac{i}{t} \leq 1$$

Note que provar a desigualdade acima é equivalente a provar o enunciado do teorema. Agora, usando a Definição 3.1, temos

$$E[D_i(t)] \cdot \frac{i}{t} = \frac{\Gamma(t+1)\Gamma\left(i-\frac{1}{2}\right)}{\Gamma\left(t+\frac{1}{2}\right)\Gamma(i)} \cdot \frac{i}{t} = \frac{\Gamma(t+1)\Gamma\left(i-1+\frac{1}{2}\right)}{\Gamma\left(t+\frac{1}{2}\right)\Gamma(i)} \cdot \frac{i}{t}$$

Usando o Teorema 2.2, podemos expandir os termos da função gama da seguinte maneira,

$$\begin{aligned} E[D_i(t)] \cdot \frac{i}{t} &= \frac{\Gamma(t+1) \frac{2^{1-2(i-1)}\sqrt{\pi}\Gamma(2(i-1))}{\Gamma(i-1)}}{\Gamma\left(t+\frac{1}{2}\right)\Gamma(i)} \cdot \frac{i}{t} = \frac{\Gamma(t+1) \frac{2^{1-2(i-1)}\sqrt{\pi}\Gamma(2(i-1))}{\Gamma(i-1)}}{\frac{2^{1-2t}\sqrt{\pi}\Gamma(2t)\Gamma(i)}{\Gamma(t)}} \cdot \frac{i}{t} \\ &= \frac{\Gamma(t+1)2^{1-2(i-1)}\sqrt{\pi}\Gamma(2(i-1))}{\Gamma(i-1)} \cdot \frac{\Gamma(t)}{2^{1-2t}\sqrt{\pi}\Gamma(2t)\Gamma(i)} \cdot \frac{i}{t} \\ &= \frac{\Gamma(t+1)2^{1-2(i-1)}\sqrt{\pi}\Gamma(2(i-1))\Gamma(t)}{2^{1-2t}\sqrt{\pi}\Gamma(2t)\Gamma(i)\Gamma(i-1)} \cdot \frac{i}{t} \\ &= \frac{\Gamma(t+1) \frac{2}{2^{2i-2}}\Gamma(2(i-1))\Gamma(t)}{\frac{2}{2^{2i}}\Gamma(2t)\Gamma(i)\Gamma(i-1)} \cdot \frac{i}{t} = \frac{\Gamma(t+1)2^{2t}\Gamma(2i-2)\Gamma(t)}{2^{2i-2}\Gamma(2t)\Gamma(i)\Gamma(i-1)} \cdot \frac{i}{t} \end{aligned}$$

Usando o Teorema 2.1, temos

$$\begin{aligned} E[D_i(t)] \cdot \frac{i}{t} &= \frac{t!2^{2t}(2i-3)!(t-1)!}{2^{2i-2}(2t-1)!(i-1)!(i-2)!} \cdot \frac{i}{t} \\ &= \frac{(t-1)!2^{2t}(2i-3)!(t-1)!i}{2^{2i-2}(2t-1)!(i-1)!(i-2)!} \end{aligned} \quad (4.4)$$

Sabendo que $2i-3$ é ímpar, podemos reescrever $(2i-3)!$ da seguinte forma,

$$\begin{aligned} (2i-3)! &= 1 \cdot 2 \cdot 3 \cdot \dots \cdot (2i-4) \cdot (2i-3) \\ &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i-3)(2 \cdot 4 \cdot 6 \cdot \dots \cdot 2i-4) \\ &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i-3)(2 \cdot 1 \cdot 2 \cdot 2 \cdot 2 \cdot 3 \cdot \dots \cdot 2 \cdot (i-2)) \\ &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i-3)2^{(2i-4)/2}(i-2)! \\ &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i-3)2^{i-2}(i-2)! \end{aligned} \quad (4.5)$$

De maneira análoga, como $2t - 1$ também é ímpar, temos

$$\begin{aligned}
 (2t - 1)! &= 1 \cdot 2 \cdot 3 \cdot \dots \cdot (2t - 2) \cdot (2t - 1) \\
 &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)(2 \cdot 4 \cdot 6 \cdot \dots \cdot 2t - 2) \\
 &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)(2 \cdot 1 \cdot 2 \cdot 2 \cdot 2 \cdot 3 \cdot \dots \cdot 2 \cdot (t - 1)) \\
 &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)2^{(2t-2)/2}(t - 1)! \\
 &= (1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)2^{t-1}(t - 1)!
 \end{aligned} \tag{4.6}$$

Retomando a Fórmula 4.4, sabemos que

$$E[D_i(t)] \cdot \frac{i}{t} = \frac{(t - 1)!2^{2t}(2i - 3)!(t - 1)!i}{2^{2i-2}(2t - 1)!(i - 1)!(i - 2)!}$$

Agora substituindo as Fórmulas 4.5 e 4.6 na fórmula acima, temos:

$$\begin{aligned}
 E[D_i(t)] \cdot \frac{i}{t} &= \frac{(t - 1)!2^{2t}(2i - 3)!(t - 1)!i}{2^{2i-2}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)2^{t-1}(t - 1)!(i - 1)!(i - 2)!} \\
 &= \frac{(t - 1)!2^{2t}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i - 3)2^{i-2}(i - 2)!i}{2^{2i-2}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)2^{t-1}(i - 1)!(i - 2)!} \\
 &= \frac{(t - 1)!2^{2t}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i - 3)2^{i-2}i}{2^{2i-2}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)2^{t-1}(i - 1)!} \\
 &= \frac{(t - 1)!2^{2t}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i - 3)i}{2^i(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)2^{t-1}(i - 1)!} \\
 &= \frac{(t - 1)!2^{t+1}(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2i - 3)i}{2^i(1 \cdot 3 \cdot 5 \cdot \dots \cdot 2t - 1)(i - 1)!} \\
 &= \frac{(1 \cdot 2 \cdot 3 \cdot \dots \cdot (t - 1))(1 \cdot 3 \cdot 5 \cdot \dots \cdot (2i - 3))2^{t+1}i}{(1 \cdot 2 \cdot 3 \cdot \dots \cdot i - 1)(1 \cdot 3 \cdot 5 \cdot \dots \cdot (2t - 1))2^i}
 \end{aligned}$$

como $t > i$ podemos simplificar da seguinte maneira,

$$\begin{aligned}
 E[D_i(t)] \cdot \frac{i}{t} &= \frac{(i \cdot (i + 1) \cdot \dots \cdot (t - 1))(1 \cdot 3 \cdot 5 \cdot \dots \cdot (2i - 3))2^{t+1}i}{(1 \cdot 3 \cdot 5 \cdot \dots \cdot (2t - 1))2^i} \\
 &= \frac{(i \cdot (i + 1) \cdot \dots \cdot (t - 1))2^{t+1}i}{((2i - 1) \cdot (2i + 1) \cdot \dots \cdot (2t - 1))2^i}
 \end{aligned}$$

Usando o Lema 4.2,

$$E[D_i(t)] \cdot \frac{i}{t} = \frac{(i \cdot (i + 1) \cdot \dots \cdot (t - 1))2^{t+1}i}{((2i - 1) \cdot (2i + 1) \cdot \dots \cdot (2t - 1))2^i} < 1$$

□

Sabendo a probabilidade de aparecer uma aresta entre dois vértices para qualquer valor de m , podemos calcular a probabilidade de aparecer uma clique num grafo gerado pelo modelo $PA_t(m)$, multiplicando as probabilidades dos eventos de aparecer as arestas. Como estamos trabalhando no modelo $PA_t(m)$, sabemos que a maior clique que pode aparecer no grafo tem tamanho $\ell = m + 1$. Nosso objetivo é calcular a probabilidade de tal clique aparecer.

Entretanto, para facilitar o entendimento de como isto será feito, primeiramente vamos calcular a probabilidade do aparecimento de um K_3 por $G[\{v_i, v_{i+1}, v_{i+2}\}]$, para $m = 2$.

Teorema 4.9. Considere um grafo G amostrado do modelo $PA_t(2)$ e seja $v_i, v_{i+1}, v_{i+2} \in V(G)$, então a probabilidade de $G[\{v_i, v_{i+1}, v_{i+2}\}]$ ser um K_3 é

$$Pr(E_{i,3}) = O\left(\frac{1}{i^3}\right)$$

Demonstração.

A probabilidade de as 3 arestas da clique aparecerem é no máximo

$$Pr(E_{i,3}) \leq Pr(v_{i+1}^{(2)} \rightarrow v_i^{(2)}) \cdot Pr(v_{i+2}^{(2)} \rightarrow v_{i+1}^{(2)}) \cdot Pr(v_{i+2}^{(2)} \rightarrow v_i^{(2)})$$

A partir do Lema 4.1, temos que

$$\begin{aligned} Pr(E_{i,3}) &\leq \left(\sum_{l=((i+1)-1)2+1}^{2(i+1)} \sum_{q=(i-1)2+1}^{2i} \frac{D_q(l-1)}{2(l-1)+1} \right) \\ &\quad \cdot \left(\sum_{l=((i+2)-1)2+1}^{2(i+2)} \sum_{q=(i+1)-1)2+1}^{2(i+1)} \frac{D_q(l-1)}{2(l-1)+1} \right) \\ &\quad \cdot \left(\sum_{l=((i+2)-1)2+1}^{2(i+2)} \sum_{q=(i-1)2+1}^{2i} \frac{D_q(l-1)}{2(l-1)+1} \right) \end{aligned}$$

Simplificando os índices da fórmula acima, obtemos,

$$\begin{aligned} Pr(E_{i,3}) &\leq \left(\sum_{l=2i+1}^{2i+2} \sum_{q=2i-1}^{2i} \frac{D_q(l-1)}{2l-1} \right) \\ &\quad \cdot \left(\sum_{l=2i+3}^{2i+4} \sum_{q=2i+1}^{2i+2} \frac{D_q(l-1)}{2l-1} \right) \\ &\quad \cdot \left(\sum_{l=2i+3}^{2i+4} \sum_{q=2i-1}^{2i} \frac{D_q(l-1)}{2l-1} \right) \end{aligned} \tag{4.7}$$

O que vamos fazer agora é substituir na fórmula 4.7 o limitante superior para o valor esperado do grau de um dado vértice, obtido no Lema 4.3,

$$\begin{aligned} Pr(E_{i,3}) &\leq \left(\sum_{l=2i+1}^{2i+2} \sum_{q=2i-1}^{2i} \frac{(l-1)/q}{2l-1} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \sum_{q=2i+1}^{2i+2} \frac{(l-1)/q}{2l-1} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \sum_{q=2i-1}^{2i} \frac{(l-1)/q}{2l-1} \right) \\ &= \left(\sum_{l=2i+1}^{2i+2} \sum_{q=2i-1}^{2i} \frac{l-1}{q(2l-1)} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \sum_{q=2i+1}^{2i+2} \frac{l-1}{q(2l-1)} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \sum_{q=2i-1}^{2i} \frac{l-1}{q(2l-1)} \right) \\ &= \left(\sum_{l=2i+1}^{2i+2} \frac{l-1}{2l-1} \sum_{q=2i-1}^{2i} \frac{1}{q} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \frac{l-1}{2l-1} \sum_{q=2i+1}^{2i+2} \frac{1}{q} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \frac{l-1}{2l-1} \sum_{q=2i-1}^{2i} \frac{1}{q} \right) \end{aligned}$$

$$= \left(\sum_{l=2i+1}^{2i+2} \frac{l-1}{2l-1} \frac{4i-1}{2i(2i-1)} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \frac{l-1}{2l-1} \frac{4i+3}{2(i+1)(2i+1)} \right) \cdot \left(\sum_{l=2i+3}^{2i+4} \frac{l-1}{2l-1} \frac{4i-1}{2i(2i-1)} \right)$$

Desenvolvendo os somatórios da fórmula acima, obtemos

$$\begin{aligned} Pr(E_{i,3}) \leq & \left(\frac{i(16i+12)+1}{i(16i+16)+3} \frac{4i-1}{2i(2i-1)} \right) \cdot \left(\frac{i(16i+44)+29}{i(16i+48)+35} \frac{4i+3}{2(i+1)(2i+1)} \right) \\ & \cdot \left(\frac{i(16i+44)+29}{i(16i+48)+35} \frac{4i-1}{2i(2i-1)} \right) \end{aligned} \quad (4.8)$$

Note que o terceiro fator que aparece no lado direito da fórmula 4.8 é maior que os outros dois. Portanto,

$$\begin{aligned} Pr(E_{i,3}) & \leq \left(\frac{i(16i+44)+29}{i(16i+48)+35} \frac{4i-1}{2i(2i-1)} \right)^3 \\ & = \left(\mathcal{O}\left(\frac{1}{i}\right) \right)^3 \\ & = \mathcal{O}\left(\frac{1}{i^3}\right) \end{aligned}$$

□

Agora generalizando a probabilidade para uma clique de qualquer tamanho. Observe que se $m < \ell - 1$ não existe uma clique de tamanho ℓ em um grafo construído pelo modelo $PA_t(m)$. Neste trabalho iremos considerar o menor valor de m que seja adequado, ou seja, $m = \ell - 1$.

Teorema 4.10. Seja $m = \ell - 1$, considere um grafo G amostrado do modelo $PA_t(m)$ e seja $v_i, v_{i+1}, \dots, v_{i+\ell-1} \in V(G)$, então a probabilidade de $G[\{v_i, v_{i+1}, \dots, v_{i+\ell-1}\}]$ ser um K_ℓ é

$$Pr(E_{i,\ell}) = \mathcal{O}\left(\frac{\ell}{i^{\binom{\ell}{2}}}\right)$$

Demonstração.

A probabilidade de que todas as arestas da clique aparecerem é no máximo

$$\begin{aligned} Pr(E_{i,\ell}) & \leq Pr(v_{i+1}^{(m)} \rightarrow v_i^{(m)}) \\ & \cdot Pr(v_{i+2}^{(m)} \rightarrow v_{i+1}^{(m)}) \cdot Pr(v_{i+2}^{(m)} \rightarrow v_i^{(m)}) \\ & \cdot \dots \\ & \cdot Pr(v_{i+\ell-1}^{(m)} \rightarrow v_{i+\ell-2}^{(m)}) \cdot Pr(v_{i+\ell-1}^{(m)} \rightarrow v_{i+\ell-3}^{(m)}) \cdot \dots \cdot Pr(v_{i+\ell-1}^{(m)} \rightarrow v_i^{(m)}) \end{aligned}$$

A partir do Lema 4.1, temos que

$$\begin{aligned} Pr(E_{i,\ell}) & \leq \left(\sum_{l=((i+1)-1)m+1}^{m(i+1)} \sum_{q=(i-1)m+1}^{mi} \frac{D_q(l-1)}{2(l-1)+1} \right) \\ & \cdot \left(\sum_{l=((i+2)-1)m+1}^{m(i+2)} \sum_{q=((i+1)-1)m+1}^{m(i+1)} \frac{D_q(l-1)}{2(l-1)+1} \right) \end{aligned}$$

$$\begin{aligned}
& \cdot \left(\sum_{l=((i+2)-1)m+1}^{m(i+2)} \sum_{q=(i-1)m+1}^{mi} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \dots \\
& \cdot \left(\sum_{l=((i+\ell-1)-1)m+1}^{m(i+\ell-1)} \sum_{q=((i+\ell-2)-1)m+1}^{m(i+\ell-2)} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \left(\sum_{l=((i+\ell-1)-1)m+1}^{m(i+\ell-1)} \sum_{q=((i+\ell-3)-1)m+1}^{m(i+\ell-3)} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \dots \cdot \left(\sum_{l=((i+\ell-1)-1)m+1}^{m(i+\ell-1)} \sum_{q=(i-1)m+1}^{mi} \frac{D_q(l-1)}{2(l-1)+1} \right)
\end{aligned}$$

Como estamos assumindo sempre o menor valor possível de m para o aparecimento de uma clique de tamanho ℓ , neste caso $\ell = m + 1$, podemos deixar a fórmula acima somente em função de m onde está ℓ , portanto

$$\begin{aligned}
Pr(E_{i,\ell}) & \leq \left(\sum_{l=((i+1)-1)m+1}^{m(i+1)} \sum_{q=(i-1)m+1}^{mi} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \left(\sum_{l=((i+2)-1)m+1}^{m(i+2)} \sum_{q=((i+1)-1)m+1}^{m(i+1)} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \left(\sum_{l=((i+2)-1)m+1}^{m(i+2)} \sum_{q=(i-1)m+1}^{mi} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \dots \\
& \cdot \left(\sum_{l=((i+m+1-1)-1)m+1}^{m(i+m+1-1)} \sum_{q=((i+m+1-2)-1)m+1}^{m(i+m+1-2)} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \left(\sum_{l=((i+m+1-1)-1)m+1}^{m(i+m+1-1)} \sum_{q=((i+m+1-3)-1)m+1}^{m(i+m+1-3)} \frac{D_q(l-1)}{2(l-1)+1} \right) \\
& \cdot \dots \cdot \left(\sum_{l=((i+m+1-1)-1)m+1}^{m(i+m+1-1)} \sum_{q=(i-1)m+1}^{mi} \frac{D_q(l-1)}{2(l-1)+1} \right)
\end{aligned}$$

Agora vamos substituir na fórmula acima o limitante superior para o valor esperado do grau de um dado vértice, obtido no Lema 4.3,

$$\begin{aligned}
Pr(E_{i,\ell}) & \leq \left(\sum_{l=mi+1}^{mi+m} \sum_{q=mi-m+1}^{mi} \frac{(l-1)/q}{2l-1} \right) \\
& \cdot \left(\sum_{l=mi+m+1}^{mi+2m} \sum_{q=mi+1}^{mi+m} \frac{(l-1)/q}{2l-1} \right)
\end{aligned}$$

$$\begin{aligned}
& \cdot \left(\sum_{l=mi+m+1}^{mi+2m} \sum_{q=mi-m+1}^{mi} \frac{(l-1)/q}{2l-1} \right) \\
& \dots \\
& \cdot \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \sum_{q=mi+m^2-2m+1}^{mi+m^2-m} \frac{(l-1)/q}{2l-1} \right) \\
& \cdot \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \sum_{q=mi+m^2-3m+1}^{mi+m^2-2m} \frac{(l-1)/q}{2l-1} \right) \\
& \dots \cdot \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \sum_{q=mi-m+1}^{mi} \frac{(l-1)/q}{2l-1} \right)
\end{aligned}$$

Desenvolvendo os somatórios da fórmula acima, obtemos

$$\begin{aligned}
Pr(E_{i,\ell}) & \leq \left(\sum_{l=mi+1}^{mi+m} \frac{l-1}{2l-1} \sum_{q=mi-m+1}^{mi} \frac{1}{q} \right) \\
& \cdot \left(\sum_{l=mi+m+1}^{mi+2m} \frac{l-1}{2l-1} \sum_{q=mi+1}^{mi+m} \frac{1}{q} \right) \\
& \cdot \left(\sum_{l=mi+m+1}^{mi+2m} \frac{l-1}{2l-1} \sum_{q=mi-m+1}^{mi} \frac{1}{q} \right) \\
& \dots \\
& \cdot \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \frac{l-1}{2l-1} \sum_{q=mi+m^2-2m+1}^{mi+m^2-m} \frac{1}{q} \right) \\
& \cdot \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \frac{l-1}{2l-1} \sum_{q=mi+m^2-3m+1}^{mi+m^2-2m} \frac{1}{q} \right) \\
& \dots \cdot \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \frac{l-1}{2l-1} \sum_{q=mi-m+1}^{mi} \frac{1}{q} \right)
\end{aligned}$$

Sabendo que o último fator da fórmula acima é maior que os outros. Portanto,

$$Pr(E_{i,\ell}) \leq \left(\sum_{l=mi+m^2-m+1}^{mi+m^2} \frac{l-1}{2l-1} \sum_{q=mi-m+1}^{mi} \frac{1}{q} \right)^{\binom{\ell}{2}} \quad (4.9)$$

Podemos reescrever $\sum_{q=mi-m+1}^{mi} \frac{1}{q}$ da seguinte forma:

$$\begin{aligned} \sum_{q=mi-m+1}^{mi} \frac{1}{q} &= \frac{1}{mi-m+1} + \frac{1}{mi-m+2} + \dots + \frac{1}{mi} \\ &\leq \frac{1}{mi-m+1} + \frac{1}{mi-m+1} + \dots + \frac{1}{mi-m+1} \\ &= (m+2) \frac{1}{mi-m+1} = \frac{m+2}{mi-m+1} \end{aligned}$$

Lembrando que neste modelo, temos que m é uma constante, portanto:

$$\sum_{q=mi-m+1}^{mi} \frac{1}{q} = O\left(\frac{1}{i}\right) \quad (4.10)$$

De maneira análoga, também podemos reescrever $\sum_{l=mi+m^2-m+1}^{mi+m^2} \frac{l-1}{2l-1}$ da seguinte forma:

$$\begin{aligned} \sum_{l=mi+m^2-m+1}^{mi+m^2} \frac{l-1}{2l-1} &= \frac{(mi+m^2-m+1)-1}{2(mi+m^2-m+1)-1} + \frac{(mi+m^2-m+2)-1}{2(mi+m^2-m+2)-1} + \dots + \frac{(mi+m^2)-1}{2(mi+m^2)-1} \\ &= \frac{mi+m^2-m}{2mi+2m^2-2m+1} + \frac{mi+m^2-m+1}{2mi+2m^2-2m+3} + \dots + \frac{mi+m^2-1}{2mi+2m^2-1} \\ &\leq \frac{mi+m^2-m}{2mi+2m^2-2m+1} + \frac{mi+m^2-m}{2mi+2m^2-2m+1} + \dots + \frac{mi+m^2-m}{2mi+2m^2-2m+1} \\ &= (m+2) \frac{mi+m^2-m}{2mi+2m^2-2m+1} \\ &= \frac{m^2i+m^3-m^2+2mi+2m^2-2m}{2mi+2m^2-2m+1} \\ &= \frac{m^2i+m^3+2mi+m^2-2m}{2mi+2m^2-2m+1} \\ &= O(m) = O(\ell) \end{aligned} \quad (4.11)$$

Agora substituindo as fórmulas 4.10 e 4.11 na fórmula 4.9, temos:

$$Pr(E_{i,\ell}) \leq \left(O(\ell) O\left(\frac{1}{i}\right) \right)^{\binom{\ell}{2}}$$

E portanto,

$$Pr(E_{i,\ell}) = O\left(\frac{\ell}{i^{\binom{\ell}{2}}}\right)$$

□

4.3 CONCLUSÃO

Neste capítulo, foram apresentados os estudos experimentais e teóricos que serviram como base para a presente dissertação. Ainda neste capítulo provamos que a probabilidade de aparecer uma clique em um grafo gerado pelo modelo adaptado do modelo $PA_t(m)$ decai exponencialmente em função do tamanho da clique e dos vértices escolhidos, ou seja, quanto maior a clique, menor a probabilidade dela existir no grafo.

5 CONCLUSÃO

Recentemente, diversos estudos demonstraram que alguns problemas computacionais são resolvidos de maneira mais eficiente em grafos de lei de potência do que em outros grafos em geral. Ainda que já exista bastante pesquisa experimental acerca disso, há pouco material teórico que analise este fato. Na presente dissertação analisamos o problema da clique máxima com o objetivo de encontrar evidências para o bom desempenho que heurísticas gulosas apresentam em grafos de lei de potência. Para isso, se fez necessário o estudo de modelos de grafos aleatórios que são úteis neste contexto para analisar o processo de construção de um grafo lei de potência.

Obtivemos uma relação de probabilidade de aparecimento de cliques em grafos de lei de potência construídos pelo modelo de grafo aleatório de ligação preferencial $PA_t(m)$, que foi adaptado do modelo proposto por Hofstad (2010). Tais modelos de grafo aleatório auxiliam o estudo a respeito de grafos de lei de potência.

Durante o desenvolvimento do resultado, optou-se por simplificar os parâmetros δ e m do modelo. Neste caso, em um estudo futuro seria interessante analisar as relações de tais parâmetros com os resultados obtidos. Além disso, é aberta a possibilidade de se analisar a utilização de outros tipos de modelos de grafos aleatórios para as análises realizadas.

REFERÊNCIAS

- Aiello, W., Chung, F. e Lu, L. (2001). A random graph model for power law graphs. *Experimental Mathematics*, 10(1):53–66.
- Arfken, G. B. e Weber, H. J. (1999). *Mathematical methods for physicists*. American Association of Physics Teachers.
- Barabási, A.-L. e Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Barabási, A.-L. e Bonabeau, E. (2003). Scale-free networks. *Scientific american*, 288(5):60–69.
- Bender, E. A. e Canfield, E. R. (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307.
- Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316.
- Bollobás, B., Riordan, O., Spencer, J. e Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290.
- Brach, P., Cygan, M., Łącki, J. e Sankowski, P. (2016). Algorithmic complexity of power law networks. Em *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, páginas 1306–1325. SIAM.
- Britton, T., Deijfen, M. e Martin-Löf, A. (2006). Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, 124(6):1377–1397.
- Chauhan, A., Friedrich, T. e Rothenberger, R. (2016). Greed is good for deterministic scale-free networks. *arXiv preprint arXiv:1610.04217*.
- Chung, F. e Lu, L. (2002). Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145.
- da Silva, M. O., Gimenez-Lugo, G. A. e da Silva, M. V. G. (2013). Vertex cover in complex networks. *International Journal of Modern Physics C*, 24(11):1350078.
- Erdős, P. e Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60.
- Erdős, P. e Rényi, A. (1961). On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, 12(1-2):261–267.
- Eubank, S., Kumar, V., Marathe, M. V., Srinivasan, A. e Wang, N. (2004). Structural and algorithmic aspects of massive social networks. *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, páginas 718–727.
- Faloutsos, M., Faloutsos, P. e Faloutsos, C. (1999). On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*, 29(4):251–262.

- Feige, U. (1998). A threshold of $\ln n$ for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652.
- Gast, M. e Hauptmann, M. (2014). Approximability of the vertex cover problem in power-law graphs. *Theoretical Computer Science*, 516:60–70.
- Gast, M., Hauptmann, M. e Karpinski, M. (2015). Inapproximability of dominating set on power law graphs. *Theoretical Computer Science*, 562:436–452.
- Guelzim, N., Bottani, S., Bourguine, P. e Képès, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nature genetics*, 31(1):60.
- Hofstad, R. V. D. (2010). *Random Graphs and Complex Networks*, páginas 543–555. Cambridge University Press.
- Janson, S., Łuczak, T. e Norros, I. (2010). Large cliques in a power-law random graph. *Journal of Applied Probability*, 47(4):1124–1135.
- Mitzenmacher, M. e Upfal, E. (2017). *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press.
- Mlodinow, L. (2009). *O andar do bêbado: Como o acaso determina nossas vidas*. Zahar.
- Molloy, M. e Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180.
- Park, K. e Lee, H. (2001). On the effectiveness of route-based packet filtering for distributed dos attack prevention in power-law internets. *ACM SIGCOMM computer communication review*, 31(4):15–26.
- Rogiski, R. (2016). Desempenho do algoritmo guloso na coloração de vértices em grafos de sistemas complexos. B.S. thesis, Universidade Tecnológica Federal do Paraná.
- Salganik, M., Dodds, P. e Watts, D. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(1):854–856.
- Shen, Y., Nguyen, D. T., Xuan, Y. e Thai, M. T. (2012). New techniques for approximating optimal substructure problems in power-law graphs. *Theoretical Computer Science*, 447:107–119.
- Titchmarsh, E. C., Titchmarsh, E. C. T., Heath-Brown, D. et al. (1986). *The theory of the Riemann zeta-function*. Oxford University Press.
- Vignatti, A. L. e da Silva, M. V. G. (2016). Minimum vertex cover in generalized random graphs with power law degree distribution. *Theoretical Computer Science*, 647(1):101–111.
- West, D. B. (2001). *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River.

APÊNDICE A – GRAFOS

A.1 GRAFOS LEI DE POTÊNCIA

A teoria dos grafos é um ramo da matemática, cujo objeto de estudo pode ser visto de maneira ampla como um conjunto de entidades quaisquer e as relações que estas entidades possuem entre si. Dados provenientes de diferentes áreas podem ser mapeados em grafos, de maneira que podemos assim, utilizar-se de propriedades conhecidas de grafos e realizar uma análise independente do contexto original (West, 2001).

Com o avanço da tecnologia, houve um aumento na capacidade de processamento e armazenamento de dados. A partir da observação desses dados, pesquisadores evidenciaram características comuns, presentes em grafos advindos de diversas situações práticas.

Abaixo estão alguns exemplos desses grafos segundo (Barabási e Bonabeau, 2003):

Tabela A.1: Exemplos de grafos advindos de situações práticas

Grafos	Vértices	# Arestas
Hollywood	Atores	Atuações no mesmo filme
Internet	Roteadores	Conexões físicas
Rede reguladora de proteína	Proteínas	Interações entre proteínas
Colaborações de pesquisa	Pesquisadores	Co-autoria de artigos
World Wide Web	Páginas Web	URLs
Metabolismo celular	Moléculas envolvidas na queima de alimentos para energia	Participação na mesma reação bioquímica

Barabási e Bonabeau observaram que uma importante característica presente nesses grafos é a distribuição do grau dos vértices do grafos seguirem uma distribuição lei de potência (Barabási e Bonabeau, 2003). Ou seja, existem poucos vértices com grau alto e muitos vértices com grau baixo. Em consequência disso, os autores sinalizam a aparição de “hubs” no grafo, que são os poucos vértices com grau muito alto. Tais “hubs”, tornam o grafo extremamente vulnerável a ataques coordenados, onde a remoção de um vértice “hub” pode acarretar numa grande perda ou ganho, dependendo da situação, para o sistema modelado pelo grafo.

Exemplo A.1. Um exemplo de relação que apresenta distribuição lei de potência é a popularidade, onde somente algumas pessoas conseguem alcançar algum reconhecimento e a maioria é pouco conhecida. A tendência com o tempo, é que as pessoas mais populares se tornem ainda mais populares, pois quanto mais conhecida a pessoa é, maior a probabilidade de se ouvir o nome dela em uma conversa.

Segundo (Mitzenmacher e Upfal, 2017), são exemplos que seguem a lei de potência aqueles em que a distribuição não está tão concentrada na média, como tamanho de cidades, popularidade e a riqueza entre as famílias.

Os estágios iniciais de qualquer ascensão que obedece a lei de potência é relativamente frágil. Pois, no início, todas as instâncias possuem probabilidades bem parecidas de realizarem

conexões umas com as outras, e a medida que a popularidade de uma instância vai aumentando a lei de potência aumentará ainda mais.

Exemplo A.2. É possível observar essa fragilidade inicial do processo num experimento realizado por (Salganik et al., 2006). No experimento foi criado um site de download de músicas onde visitantes podiam visualizar o número de vezes que cada música foi baixada. Sem as quantidades iniciais de download, no começo os visitantes agiam aleatoriamente na escolha de suas músicas, mas a medida de que as contagens foram se atualizando, as músicas que estavam com um número maior de downloads ficavam com um número maior ainda, pois eram mais baixadas por outros visitantes. Foi realizado esse experimento em 8 cópias paralelas do site, onde cada cópia evoluiu de forma diferente conforme as ações dos primeiros usuários. De fato, descobriu-se que a popularidade das diferentes músicas variou consideravelmente nas diferentes cópias paralelas, embora as melhores músicas nunca tenham acabado na parte inferior e as piores músicas nunca tenham acabado no topo. Uma nova versão do site também foi aplicada, onde não havia informação sobre o número de downloads para os usuários. Foi possível observar que houve uma variação significativamente menor na relação de músicas diferentes.

APÊNDICE B – TEORIA DE PROBABILIDADE

B.1 INTERSECÇÃO DE EVENTOS

Uma boa forma de pensar na intersecção de eventos é através do seguinte texto extraído do livro do (Mlodinow, 2009):

“Antes de seguirmos em frente, vale a pena atentarmos para um detalhe importante: a condição que diz que “se dois eventos possíveis, A e B, forem independentes”. Suponha que uma companhia aérea tenha 1 lugar restante num voo, e ainda restem 2 passageiros por chegar. Suponha que, a partir da experiência, a companhia saiba que existe uma chance de $2/3$ de que um passageiro que reservou um voo se apresente para viajar. Utilizando a regra da multiplicação, a funcionária da companhia poderá concluir que existe uma chance de $2/3 \times 2/3$, ou cerca de 44%, de que ela tenha que lidar com um cliente insatisfeito. A chance de que nenhum cliente apareça e de que o avião tenha que voar com um lugar vago, por outro lado, é de $1/3 \times 1/3$, ou apenas 11%, aproximadamente. Mas isso só ocorre se presumirmos que os passageiros são independentes. Se, por exemplo, eles estiverem viajando juntos, a análise acima estará errada. A chance de que ambos queiram viajar é de $2/3$, igual à de um passageiro independente. É importante lembrar que só calculamos a probabilidade combinada a partir das probabilidades simples por meio da multiplicação quando os eventos não têm nenhuma relação entre si”.