

Aparecimento de cliques em grafos aleatórios lei de potência no modelo de ligação preferencial

David Reksidler Jr.¹, Murilo V. G. da Silva¹

¹DINF – Universidade Federal do Paraná (UFPR)
Curitiba, Brazil

juniorrek@ufpr.br, murilo@inf.ufpr.br

Abstract. *In this paper we present preliminary results on the occurrence of cliques of a given size in power law graphs sampled from a distribution of random graphs based on the preferential attachment paradigm. We derive explicitly a formula for the probability of a clique of a prescribed size to appear in a graph sampled from the referred distribution. These results are part of a more general project aiming to investigate the performance of algorithms for the maximum clique problem in power law graphs.*

Resumo. *Neste artigo, apresentamos estudos preliminares sobre o surgimento de cliques de diferentes tamanhos em grafos aleatórios lei de potência amostrados segundo o modelo de ligação preferencial. Derivamos explicitamente a fórmula para a probabilidade de uma clique de tamanho prescrito aparecer na amostragem. Estes resultados estão inseridos em um projeto em andamento mais amplo que visa investigar o desempenho de algoritmos para o problema de clique máxima em grafos lei de potência.*

Keywords: Cliques; Power Law Graphs; Random Graphs

Palavras-chave: Cliques; Grafos lei de potência; Grafos Aleatórios

1. Introdução

Com o avanço na capacidade de processamento e armazenamento de grandes quantidades de dados, foi observado em diversos estudos [Faloutsos et al. 1999, Guelzim et al. 2002, Eubank et al. 2004] que muitas redes complexas de grande porte, advindas de diversas situações práticas, desde a *World Wide Web* até redes sociais e biológicas, apresentam uma distribuição lei de potência nos graus de seus vértices. Tais redes, chamadas neste artigo de *grafos lei de potência*, de maneira simplificada, apresentam poucos vértices com grau alto e muitos vértices com grau baixo. Formalmente, em um grafo lei de potência, o número de vértices com um determinado grau k é $k^{-\beta}$, onde $\beta > 0$ é o expoente que caracteriza a lei de potência e k é uma constante de proporcionalidade.

Do ponto de vista algorítmico, a relevância deste fato é a evidência crescente de que alguns problemas de otimização parecem admitir soluções mais eficientes em grafos lei de potência do que em outros grafos em geral. Estudos empíricos

[Park and Lee 2001, Silva et al. 2013, Rogiski 2016, Pinto and da Silva 2017] mostram que algoritmos gulosos atingem soluções muito próximas da ótima em tais grafos, em particular, melhores que as garantias de aproximação que se aplicam a grafos no caso geral para problemas como o da coloração de vértices, clique máxima e o da cobertura mínima por vértices. Isso explica em parte porque problemas computacionalmente difíceis em algumas circunstâncias admitem soluções eficientes na prática. Recentemente alguns estudos analíticos [Shen et al. 2012, Gast and Hauptmann 2014, Gast et al. 2015, Vignatti and da Silva 2016] têm procurado esclarecer este fato e dar sustentação matemática a tal fenômeno.

O presente trabalho se insere neste contexto, investigando o problema da clique máxima em grafos lei de potência, dado a falta de análise teórica presente na literatura acerca de tal problema. O primeiro passo para entender o comportamento de algoritmos para clique máxima é compreender o fenômeno do aparecimento de cliques em modelos de grafos aleatórios que seguem a lei de potência. Neste estudo preliminar investigamos o aparecimento de cliques de diversos tamanhos em um modelo conhecido como *modelo de ligação preferencial*. O objetivo disso é estabelecer as bases para o prosseguimento deste projeto que visa o esclarecimento de quão rápido decai a quantidades de cliques nestes grafos à medida que o tamanho requerido para tal clique cresce e como isso impacta no desempenho de algoritmos para o problema da clique máxima.

2. Modelos de Grafos Aleatórios

O estudo de modelos de grafos aleatórios iniciou-se na década de 60 com [Erdős and Rényi 1960]. Os pesquisadores buscavam analisar redes complexas, propondo modelos matemáticos capazes de prever algumas características de tais redes muito antes que dados massivos de redes de grande escala estivessem disponíveis, como é o caso hoje. Tais modelos podem não capturar com exatidão fenômenos de muitos grafos advindos de aplicações práticas, pelo fato de induzirem uma distribuição normal (ao invés de, por exemplo, lei de potência) nos graus dos vértices. Ainda assim, a simplicidade dos modelos de Erdos e Rényi resultou na ascensão de um grande interesse de estudo de grafos aleatórios estabelecendo as bases da pesquisa na área. [Aiello et al. 2001, Bollobás et al. 2001, Mitzenmacher and Upfal 2017, Hofstad 2010].

Atualmente, os modelos propostos no estudo de redes de grande escala e que acomodam o fenômeno da lei de potência, são comumente divididos em três grupos: *grafos aleatório generalizado (GRG)*, *modelo de configuração* e *modelo de ligação preferencial*.

O *modelo de ligação preferencial*, que é o modelo utilizado neste trabalho, é inspirado em uma das hipóteses para o aparecimento do fenômeno lei de potência. A hipótese é que redes na natureza tendem a começar pequenas e crescem com o tempo seguindo um processo denominado *ligação preferencial* [Barabási and Albert 1999]. A regra define que novos vértices que vão sendo adicionados ao grafo, se conectam preferencialmente a vértices que já possuem bastantes conexões. De maneira mais precisa, neste trabalho é utilizado o modelo proposto por [Hofstad 2010] para estudar o fenômeno do aparecimento de cliques de diversos tamanhos em grafos lei de potência.

3. O Modelo de Ligação Preferencial

Em modelos de ligação preferencial os vértices são adicionados sequencialmente com uma ou mais arestas, proporcionando o crescimento do grafo. Essas arestas são conectadas a vértices pré-existentes no grafo com probabilidade proporcional ao grau desses vértices, favorecendo ainda mais o crescimento dos graus dos vértices que já tem grau elevado, levando o fenômeno da lei de potência emergir.

Definição 3.1 (Modelo de Ligação Preferencial [Hofstad 2010]) O modelo produz uma sequência de grafos não direcionados denotada por $(PA_t^{(m,\delta)})_{t>1}$, onde para cada tempo t é gerado um grafo com t vértices, onde cada vértice possui m arestas resultando em mt arestas no total. O parâmetro δ do modelo controla a proeminência da lei de potência nos grafos gerados.

Inicialmente, o modelo é definido em termos de $m = 1$, onde os grafos gerados consistem em uma coleção de árvores. O conjunto de vértices do modelo $PA_t^{(1,\delta)}$ é denotado como $\{v_1^{(1)}, v_2^{(1)}, \dots, v_t^{(1)}\}$. Seja $D_i(t)$ o grau do vértice $v_i^{(1)}$ no tempo t , onde um laço aumenta o grau em 2. Por definição, $PA_1^{(1,\delta)}$ consiste em um grafo de um único vértice com um laço. A regra para se obter o grafo $PA_{t+1}^{(1,\delta)}$ a partir de $PA_t^{(1,\delta)}$ é descrita abaixo usando a seguinte notação: $v_i^{(m)} \rightarrow v_j^{(m)}$ denota o surgimento de uma aresta do vértice $v_i^{(m)}$ para o vértice $v_j^{(m)}$. Segue a regra define o modelo aleatório:

1. Adicione um vértice v_{t+1} contendo uma aresta em $PA_{t+1}^{(1,\delta)}$.
2. A aresta é conectada a outro vértice com probabilidade:

$$Pr(v_{t+1}^{(1)} \rightarrow v_i^{(1)}) = \begin{cases} \frac{1+\delta}{(2+\delta)t+1+\delta}, & \text{se } i = t + 1 \\ \frac{D_i(t)+\delta}{(2+\delta)t+1+\delta}, & \text{se } 0 < i < t + 1 \end{cases}, \forall \delta \geq -1$$

Para valores de $m > 1$, o modelo é definido em termos do modelo para $m = 1$, da seguinte forma: Fixa-se $\delta \geq -m$; Amostra-se um grafo do modelo $PA_{mt}^{(1,\delta/m)}$ e denota seus vértices como $\{v_1^{(1)}, v_2^{(1)}, \dots, v_{mt}^{(1)}\}$; Colapsa-se os m primeiros vértices de $PA_{mt}^{(1,\delta/m)}$ em um único vértice que será o primeiro vértice de $PA_t^{(m,\delta)}$, ou seja, $v_1^{(m)} = \{v_1^{(1)}, v_2^{(1)}, \dots, v_m^{(1)}\}$. Ao fazer isso, todas as arestas que são incidentes em qualquer um dos m primeiros vértices de $PA_{mt}^{(1,\delta/m)}$ serão incidentes no novo vértice $v_1^{(m)}$.

A seguir colapsa-se os próximos m vértices para se tornarem $v_2^{(m)} = \{v_{m+1}^{(1)}, v_{m+2}^{(1)}, \dots, v_{2m}^{(1)}\}$, e assim consecutivamente até que $v_t^{(m)} = \{v_{(t-1)m+1}^{(1)}, v_{(t-1)m+2}^{(1)}, \dots, v_{mt}^{(1)}\}$. Ou seja, para $m > 1$ o modelo gera um grafo em termos de $m = 1$ e logo depois converte-o para o valor de m desejado. A Figura 1 ilustra o processo de geração do grafo para um valor de $m = 2$.

Para entender $PA_t^{(m,\delta)}$ em termos de $PA_{mt}^{(1,\delta/m)}$ é necessário observar que uma aresta em $PA_{mt}^{(1,\delta/m)}$ é conectada a um determinado vértice $v_k^{(1)}$ com probabilidade proporcional ao peso do vértice, onde o peso é igual ao grau do vértice somado com δ/m . Como os vértices em $PA_{mt}^{(1,\delta/m)}$ são colapsados para formarem os vértices de $PA_t^{(m,\delta)}$,

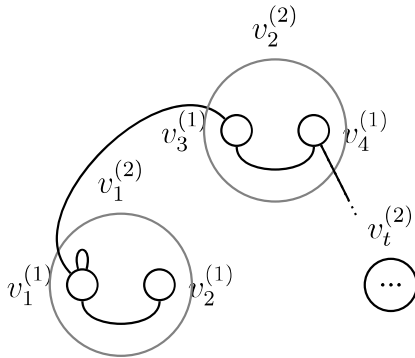


Figura 1. Exemplo com $m = 2$. As arestas presentes na figura foram criadas no passo inicial do processo, em que $m = 1$.

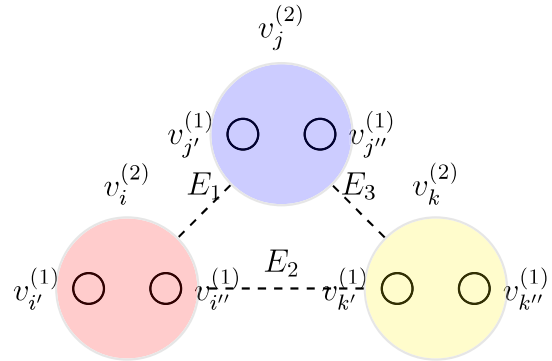


Figura 2. Eventos para o aparecimento de um K_3 . Os três vértices finais da clique são os coloridos com as cores vermelha, azul e amarela.

então uma aresta em $PA_t^{(m,\delta)}$ é conectada a um determinado vértice $v_k^{(m)}$ com probabilidade proporcional à soma dos pesos dos m vértices em $PA_{mt}^{(1,\delta/m)}$. Dado que a soma dos graus dos m vértices em $PA_{mt}^{(1,\delta/m)}$ é igual ao grau do vértice $v_k^{(m)}$ em $PA_t^{(m,\delta)}$, então a probabilidade da aresta conectar no vértice $v_k^{(m)}$ é igual ao seu grau somado com δ . É importante ressaltar que durante a geração do grafo, para valores de $m > 1$, os graus são atualizados após cada aresta ser adicionada ao grafo.

4. Aparecimento de cliques em grafos aleatórios lei de potência

Agora é apresentado o cálculo da probabilidade do aparecimento de cliques em grafos amostrados do modelo apresentado na Seção 3. Para facilitar o entendimento, inicialmente realizamos o cálculo para cliques de tamanho 3 e, em seguida generalizamos para cliques arbitrárias.

Para que haja ciclos no grafo (e consequentemente cliques de tamanho maior que 2), precisa-se de $m > 1$ e $t > 2$. Seja G um grafo amostrado do modelo $PA_3^{(2,\delta)}$. Como o modelo gera grafos inicialmente em termos de $m = 1$, primeiramente tem-se um grafo com mt vértices onde cada vértice liga-se com apenas um outro vértice. Logo, para que apareça uma clique, dados três vértices $v_i^{(m)}, v_j^{(m)}, v_k^{(m)}$ distintos, é necessário que aconteçam os seguintes eventos: $E_1 =$ pelo menos um par de vértices conectado entre $\{v_{j'}^{(1)}, v_{j''}^{(1)}\}$ e $\{v_{i'}^{(1)}, v_{i''}^{(1)}\}$; $E_2 =$ pelo menos um par de vértices conectado entre $\{v_{k'}^{(1)}, v_{k''}^{(1)}\}$ e $\{v_{i'}^{(1)}, v_{i''}^{(1)}\}$; $E_3 =$ pelo menos um par de vértices conectado entre $\{v_{k'}^{(1)}, v_{k''}^{(1)}\}$ e $\{v_{j'}^{(1)}, v_{j''}^{(1)}\}$. A Figura 2 ilustra tais eventos.

Como mencionado anteriormente, a probabilidade de um vértice $v_i^{(2)}$ ter uma aresta com o vértice $v_j^{(2)}$ é proporcional à soma dos pesos desses vértices se conectarem para $m = 1$. Em outras palavras, $Pr(E_1) = Pr(v_j^{(2)} \rightarrow v_i^{(2)})$ é igual a

$Pr(v_{j'}^{(1)} \rightarrow v_{i'}^{(1)}) + Pr(v_{j''}^{(1)} \rightarrow v_{i''}^{(1)}) + Pr(v_{j''}^{(1)} \rightarrow v_{i'}^{(1)}) + Pr(v_{j'}^{(1)} \rightarrow v_{i''}^{(1)})$. O Lema a seguir generaliza a probabilidade de existir uma aresta entre um par qualquer de vértices para $m > 1$.

Lema 4.1 A probabilidade de $v_i^{(m)}$ ser adjacente a $v_j^{(m)}$ é dada por

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

A prova do Lema 4.1 vem da formalização do argumento esboçado nos parágrafos anteriores. A seguir o resultado principal deste trabalho:

Teorema 4.1 Seja $K_x \subseteq G$ o evento de uma clique de tamanho x aparecer em um grafo no modelo $(PA_t^{(m,\delta)})_{t>1}$. Temos que

$$Pr(K_x \subseteq G) = \sum_{\substack{S_i \in [t] \\ |S|=t-x}} \prod_{\substack{0 < j,k \\ j,k \in [t] \setminus S_i}} Pr(v_j^{(m)} \rightarrow v_k^{(m)})$$

Prova: Começaremos calculando a probabilidade para uma clique de tamanho 3 dados 3 vértices distintos $v_i^{(m)}, v_j^{(m)}, v_k^{(m)}$. Como mencionado anteriormente, para a clique aparecer é necessário que aconteçam os seguintes eventos $Pr(E_1) \cdot Pr(E_2) \cdot Pr(E_3)$. Utilizando o Lema 4.1, temos que a probabilidade dos 3 vértices formarem uma clique é igual a $Pr(v_j^{(m)} \rightarrow v_i^{(m)}) \cdot Pr(v_k^{(m)} \rightarrow v_i^{(m)}) \cdot Pr(v_k^{(m)} \rightarrow v_j^{(m)})$. Porém essa é a probabilidade para uma clique considerando somente os 3 vértices fixos. Vamos agora considerar todos os possíveis conjuntos de 3 vértices do grafo.

A equação generalizada para cliques de tamanho 3 em qualquer grafo e para qualquer conjunto de vértices é igual a $Pr(K_3 \subseteq G) = \sum_{\substack{S_i \in [t] \\ |S|=t-3}} \prod_{\substack{0 < j,k \\ j,k \in [t] \setminus S_i}} Pr(v_j^{(m)} \rightarrow v_k^{(m)})$,

onde o índice do produtório percorre os vértices de um determinado subconjunto de vértices, onde esse subconjunto é dado pelo somatório, que itera sobre todas as possíveis combinações de vértices do grafo. Isso nos fornece a probabilidade do aparecimento de cliques de tamanho 3. Para generalizar esta fórmula para cliques de qualquer tamanho, o tamanho do conjunto S é igual a $|S| = t - x$, onde x é o tamanho da clique desejada. \square

5. Conclusões e Trabalhos Futuros

Ainda que preliminares, os resultados obtidos contribuem para analisar o surgimento de cliques de diversos tamanhos em grafos aleatórios lei de potência. Pode-se observar que é necessário aumentar muito o número de arestas de um dado grafo para possivelmente surgir uma clique, dado que a probabilidade vai decrescendo a medida que o tamanho da clique aumenta. Na opinião dos autores, isso deve ajudar no desempenho de algoritmos de otimização em tais grafos lei de potência. Para confirmar ou refutar esta hipótese, o passo seguinte neste projeto será obter resultados analíticos a respeito do tamanho da maior clique do grafo e realizar experimentos com o problema da clique máxima.

Referências

- Aiello, W., Chung, F., and Lu, L. (2001). A random graph model for power law graphs. *Experimental Mathematics*, 10(1):53–66.
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290.
- Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60.
- Eubank, S., Kumar, V., Marathe, M. V., Srinivasan, A., and Wang, N. (2004). Structural and algorithmic aspects of massive social networks. *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 718–727.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*, 29(4):251–262.
- Gast, M. and Hauptmann, M. (2014). Approximability of the vertex cover problem in power-law graphs. *Theoretical Computer Science*, 516:60–70.
- Gast, M., Hauptmann, M., and Karpinski, M. (2015). Inapproximability of dominating set on power law graphs. *Theoretical Computer Science*, 562:436–452.
- Guelzim, N., Bottani, S., Bourguin, P., and 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*, pages 543–555. Cambridge University Press.
- Mitzenmacher, M. and Upfal, E. (2017). *Probability and Computing Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press.
- Park, K. and Lee, H. (2001). On the effectiveness of route-based packet filtering for distributed dos attack prevention in power-law internets. 31(4):15–26.
- Pinto, G. P. and da Silva, L. A. (2017). Algoritmos de clique máximo em redes complexas. Technical report, Universidade Tecnológica Federal do Paraná.
- Rogiski, R. (2016). Desempenho do algoritmo guloso na coloração de vértices em grafos de sistemas complexos. B.S. thesis, Universidade Tecnológica Federal do Paraná.
- Shen, Y., Nguyen, D. T., Xuan, Y., and Thai, M. T. (2012). New techniques for approximating optimal substructure problems in power-law graphs. *Theoretical Computer Science*, 447:107–119.
- Silva, M. O. D., Gimenez-Lugo, G. A., and Silva, M. V. D. (2013). Vertex cover in complex networks. *International Journal of Modern Physics C*, 24(11):1350078.
- Vignatti, A. L. and 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.