

Análise empírica do espectro de grafo dos principais modelos teóricos de redes complexas

Empirical analysis of the graph spectrum of the main theoretical models of complex networks

RESUMO

Omar Condori Lopez
omarlopez@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Andre Yoshiaki Kashiwabara
kashiwabara@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Redes Complexas estão presentes em quase tudo ao nosso redor, desde interações químicas entre proteínas até relações sociais entre indivíduos, portanto estudar estas redes tem se tornado cada vez mais importante. Estudos mostram que é possível obter informações importantes sobre a rede, assim como entender seu comportamento e estrutura, através do espectro do grafo que a representa, este espectro pode ser obtido extraindo os autovalores da matriz que representa o grafo. No entanto, redes complexas do mundo real são muito grandes e obter seu espectro tem um custo computacional muito alto. Portanto existe a necessidade de otimizar os métodos de obtenção do seu espectro, o objetivo desta pesquisa é analisar a forma do espectro dos principais modelos teóricos de redes complexas. Em termos de validação os espectros foram gerados usando diferentes parâmetros para cada modelo, com o fim de comparar sua influência na forma do espectro. Foi obtida a forma do espectro de cada modelo teórico e sua respectiva estimativa de densidade por kernel, foi constatado que independentemente do tamanho da rede a forma do seu espectro é mantida, deste modo é possível diferenciar as redes a partir da forma do seu espectro.

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

PALAVRAS-CHAVE: Distribuição. Probabilidade. Autovalor.



ABSTRACT

Complex networks are present in almost everything around us, from chemical interactions between proteins to social relationships between individuals, so studying these networks has become increasingly important. Studies show that it is possible to obtain important information about the network, as well as understand its behavior and structure, through the spectrum of the graph that represents it, this spectrum can be obtained by extracting the eigenvalues from the matrix that represents the graph. However, complex real-world networks are very large and obtaining their spectrum has a very high computational cost. Therefore, there is a need to optimize the methods of obtaining its spectrum, the objective of this research is to analyze the shape of the spectrum of the main theoretical models of complex networks. In terms of validation, the spectrum will be generated by varying the parameters of each model, comparing the influence of each parameter on the shape of the spectrum. The spectrum shape of each theoretical model was obtained and its respective density estimate per kernel, it was found that regardless of the size of the network, the shape of its spectrum is maintained, so it is possible to differentiate the networks from the shape of their spectrum.

KEYWORDS: Distribution. Probability. Eigenvalue.

INTRODUÇÃO

Redes complexas são essencialmente grafos, os quais podem representar instâncias encontradas em diferentes domínios. Como por exemplo as redes sociais nas quais pessoas e suas relações de amizade (VILLAZON-TERRAZAS; APARICIO; ALVAREZ, 2015), ou proteínas e suas interações químicas (COSTA; RODRIGUES; CRISTINO, 2008), são nós e arestas respectivamente, deste modo existe um interesse crescente no tema devido a sua interdisciplinaridade que podem envolver física, biologia, sociologia, ciência da computação entre outros (JAVED, 2018).

Por meio dos grafos é possível analisar a rede e extrair suas características. Grafos podem ser classificados a partir de métricas topológicas como tamanho, densidade, grau, centralidade e assortatividade. Contudo pesquisadores acreditam ainda que se desconhece a estrutura das redes complexas profundamente (VAN, 2010). Por outro lado, estudos mostram que a saída para analisar a rede como um todo está no espectro do grafo, o qual é obtido através do cálculo dos autovalores da matriz que representa o grafo. O espectro do grafo possui informações importantes sobre a estrutura e o comportamento da rede, as quais podem ser utilizadas para prever o comportamento da rede (SARKAR; JALAN, 2018).

No entanto, redes complexas do mundo real possuem um número de nós e arestas significativamente grande. Por exemplo, a quantidade de páginas web na internet e suas interligações com outras páginas. O custo computacional para obter os autovalores e gerar o espectro é alto, chegando a ter uma complexidade $O(n^3)$ no pior dos casos (COHEN-STEINER, 2018).

Neste contexto, existe a necessidade de otimizar o processo de obtenção do espectro de redes grandes, portanto pesquisadores estudam métodos de aproximação do espectro do grafo com o objetivo de reduzir sua complexidade (COHEN-STEINER, 2018).

O objetivo desta pesquisa é analisar a influência dos parâmetros de cada modelo teórico de redes complexas, em seu respectivo espectro e demonstrar de maneira empírica, que é possível identificar cada modelo teórico pela forma do seu espectro.

MATERIAL E MÉTODOS

Para realizar o estudo empírico dos espectros do grafo foram escolhidos os seguintes modelos teóricos: Erdős-Rényi, Barabási-Albert e Watts-Strogatz. A linguagem Python foi escolhida para implementação dos modelos, devido a sua curva de aprendizado e sua vasta coleção de bibliotecas. A biblioteca *Networkx* foi utilizada para gerar os espectros dos grafos de cada modelo teórico.

O espectro de cada modelo foi gerado a partir da matriz Laplaciana normalizada, escolhida por fornecer informações estruturais importantes sobre a rede (DAS; PANIGRAHI, 2017), da qual foram obtidos seus autovalores, com os quais foi possível representar sua distribuição por meio de um histograma. Para melhor visualização foi utilizada também a estimativa de densidade Kernel (KDE).

Para cada conjunto de parâmetros foram realizadas com repetições a fim de aumentar a precisão dos resultados obtidos.

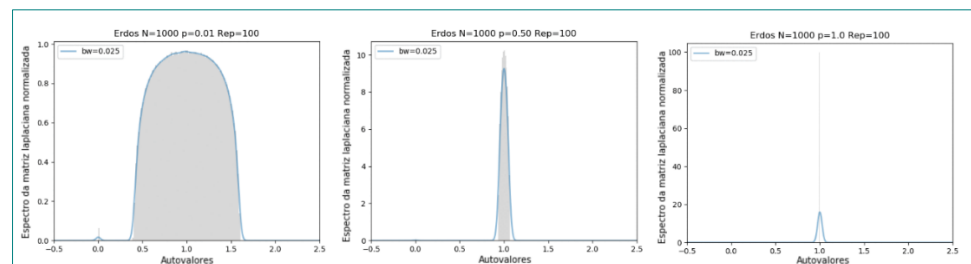
RESULTADOS E DISCUSSÃO

Inicialmente foram gerados os espectros da matriz Laplaciana normalizada de cada um dos modelos anteriormente citados, nos quais foram analisados a influência de cada parâmetro em relação aos outros. Utilizou-se a largura de banda igual a 0,025 para obter a estimativa de densidade por Kernel, este parâmetro foi escolhido por se adequar melhor aos espectros gerados.

A Figura 1 representa os espectros dos grafos gerados a partir do modelo de Erdős–Rényi. Este modelo possui dois parâmetros: N que indica o número de nós do grafo e p que indica a probabilidade de cada novo nó estabelecer uma ligação com outro. Deste modo à medida que p aumenta o grafo se torna regular, é possível observar que isso é refletido em seu espectro o que faz com que seus autovalores se aproximem de um.

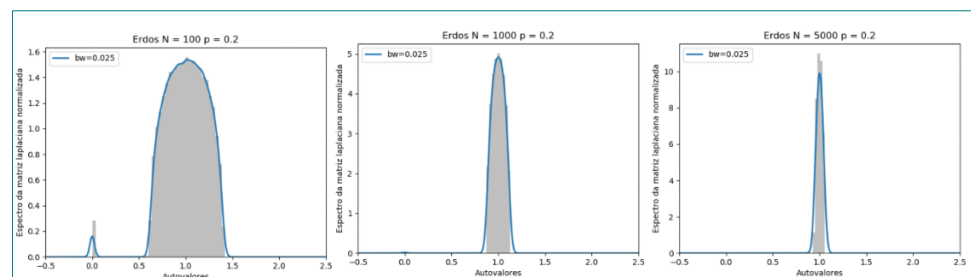
Os espectros da Figura 2 foram gerados a partir do mesmo modelo, a probabilidade de ligação fixada em 20% ($p=0,2$) e o número de nós alterado. A forma do espectro obtida neste teste se manteve, porém em escalas diferentes.

Figura 1 – Espectros do modelo Erdős–Rényi variando p.



Fonte: Autoria Própria (2020).

Figura 2 – Espectros do modelo Erdős–Rényi variando N.

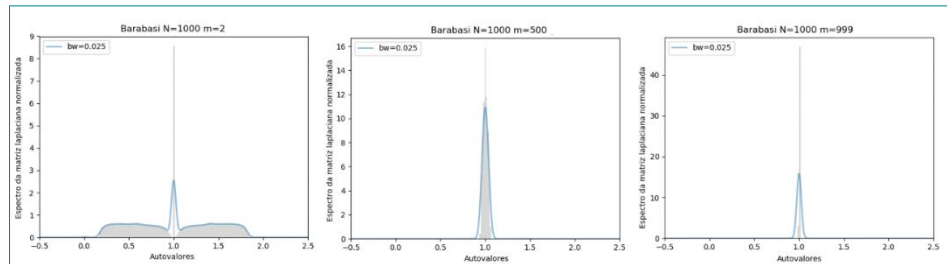


Fonte: Autoria Própria (2020).

As figuras a seguir representam os espectros dos grafos gerados a partir do modelo de Barabási–Albert, a Figura 3 representa os espectros gerados de um grafo com mil nós ($N=1000$), e a variação do parâmetro m (número de ligações que cada nó faz usando a ligação preferencial). Pode-se observar que a distribuição de autovalores é inversamente proporcional a m, o qual gera autovalores próximos de um, portanto o espectro tem uma forma mais fina em relação aos anteriores.

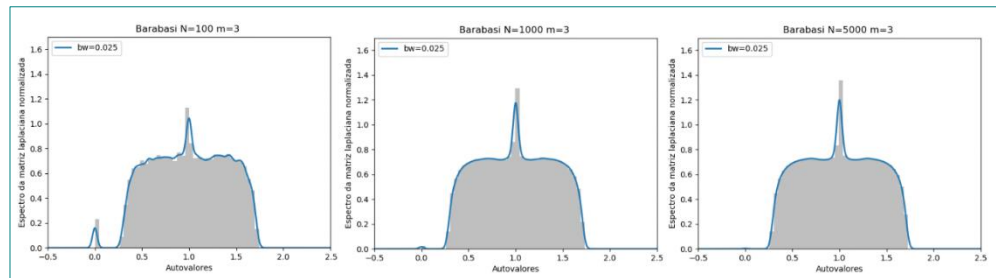
Por outro lado, quando o número de nós varia e o parâmetro m é fixo (neste caso $m=3$) representado na figura 4, é evidente que independentemente da variação do número de nós a forma do espectro é semelhante.

Figura 3 – Espectros do modelo Barabási–Albert variando m .



Fonte: Autoria Própria (2020).

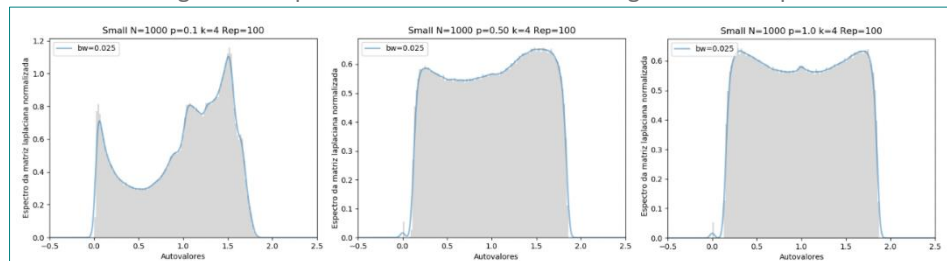
Figura 4 – Espectros do modelo Barabási–Albert variando N .



Fonte: Autoria Própria (2020).

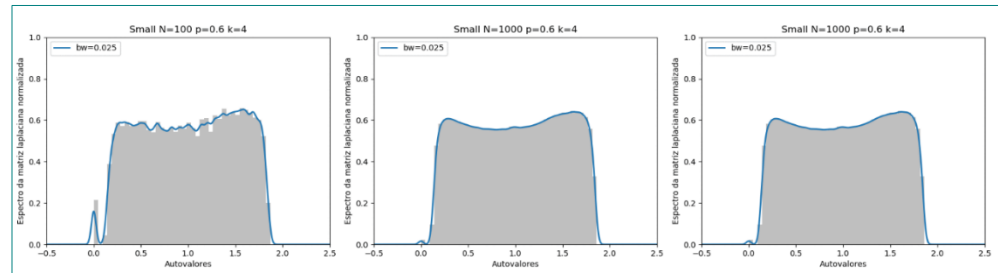
Na sequência estão os espectros obtidos do modelo de Watts–Strogatz, a Figura 5 mostra o espectro do grafo com 1000 nós, e a variação da probabilidade (p) de um par de nós sofrer uma ligação nova, é possível notar que conforme a probabilidade aumenta a forma do espectro, acaba tomando uma forma mais regular. Quando a probabilidade é fixada ($p = 0.5$) como mostrado na Figura 6, assim como no modelo anterior, a variação de N não influencia significativamente sua forma espectral.

Figura 5 – Espectros do modelo Watts–Strogatz variando p .



Fonte: Autoria Própria (2020).

Figura 6 – Espectros do modelo Watts–Strogatz variando N.



Fonte: Aatoria Própria (2020).

A obtenção dos autovalores para cada espectro teve um tempo médio de 5 minutos para cada realização em um notebook padrão (Intel i5 com 8Gb RAM). Utilizar o mesmo método se torna inviável computacionalmente para redes reais com milhares de nós, portanto métodos de aproximação são necessários para viabilizar o estudo destes espectros.

Cohen-Steiner apresentou um método de aproximação usando passeios aleatórios em um grafo a fim de obter seus momentos espectrais (COHEN-STEINER, 2018), os quais podem ser usados para chegar no espectro aproximado do grafo. Este método foi implementado e adaptado para grafos com pesos, obtendo os momentos espectrais com complexidade próxima de $O(1)$ para o modelo de Barabási–Albert (DIAS; KASHIWABARA, 2019), deste modo há evidências de que é possível otimizar o processo de aproximação dos espectros.

CONCLUSÃO

Apresentou-se, nesta pesquisa, um estudo empírico dos principais modelos teóricos de redes complexas e o processo de geração de seus espectros. Uma análise gráfica da influência dos parâmetros na forma do espectro, foi possível visualizar que cada modelo teórico possui uma forma espectral característica, diferenciando-se dos outros modelos independentemente do número de nós, também foram discutidos os problemas em relação a complexidade computacional. Deste modo, seria interessante em trabalhos futuros explorar estas características, e implementar algoritmos de aproximação de espectros de redes complexas.

AGRADECIMENTOS

À Fundação Araucária que através do PIBIC fomenta a pesquisa no Paraná, à UTFPR que me proporcionou a oportunidade de realizar a iniciação científica e finalmente ao meu orientador pela ajuda, ensinamento e preciosas sugestões.

REFERÊNCIAS

COHEN-STEINER, D. **Approximating the spectrum of a graph**. Proceedings of the 24th Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. p. 1263-1271, 2018

COSTA, L. da F.; RODRIGUES, F. A.; CRISTINO, A.S. **Complex networks: the key to systems biology**. Genet. Mol. Biol., São Paulo, v. 31, n. 3, p. 591-601, 2008.

DAS, A.; PANIGRAHI, P. **Normalized Laplacian spectrum of different type of coronas of two regular graphs**. Kragujevac journal of mathematics, v. 41, p. 57-69, 2017.

DIAS, G.; KASHIWABARA, A.Y. **A Probabilistic Algorithm to Estimate the Spectral Moments of Large Undirected Weighted Graphs**. 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), p783-787, 2019.

JAVED, M.A. **Community detection in networks: A multidisciplinary review**. Journal of Network and Computer Applications, v. 108, p. 87-111, 2018.

SARKAR, C.; JALAN, S. **Spectral properties of complex networks**. Chaos: An Interdisciplinary Journal of Nonlinear Science, v. 28, n. 10, p. 102101, 2018.

VAN, M. P. **Graph spectra for complex networks**. Cambridge University Press, p. 10, 2010.

VILLAZON-TERRAZAS, J.; APARICIO, S.; ALVAREZ, G. **Study on Twitter as a complex network**. The Third International Conference on Digital Enterprise and Information Systems (DEIS2015), p. 19, 2015.