

12º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2021

ANÁLISE DA REDE DE INTERAÇÃO DE PROTEÍNAS DA BACTÉRIA *KLEBSIELLA PNEUMONIAE* USANDO MEDIDAS DE CENTRALIDADE

GABRIEL FERNANDES JUSTINO¹, RODRIGO HENRIQUE RAMOS²

¹ Técnico Informática para Internet Integrado ao Ensino Médio, voluntário PIVICT, IFSP, Campus São Carlos, gabriel.justino@aluno.ifsp.edu.br.

² Doutorando em Ciência da Computação, Professor do Instituto Federal de São Paulo, Campus São Carlos, ramos@ifsp.edu.br.

Área de conhecimento (Tabela CNPq): 1.03.03.00-6 Metodologia e Técnicas da Computação

RESUMO: Com os avanços nas tecnologias de Sequenciamento de Nova Geração houve um aumento significativo na disponibilização de dados ômicos. As Rede de Interação de Proteínas modelam as interações celulares e podem ser utilizadas para analisar características genômicas dos organismos modelados. A base de dados STRING disponibiliza a rede de vários organismos, dentre eles o da bactéria *Klebsiella Pneumoniae*. Essa bactéria é resistente à maioria dos antibióticos e causa sérios problemas nos infectados. Partindo da hipótese de que a rede de interação de proteínas dessa bactéria tem características particulares, foi realizado uma análise usando as medidas de centralidade Grau, *Closeness* e *Betweenness*. A distribuição de Grau e de *Betweenness* teve o comportamento esperado para redes livres de escala. O *Closeness* apresentou valores próximos de zero em cerca de 40% dos vértices e um comportamento linear nos demais. Os valores do *Closeness* indicam a existência de vários vértices periféricos que se conectam apenas com vizinhos próximos. Essa descoberta levanta a possibilidade de outras análises que consideram a distância entre vértices, assim como a comparação com as redes de outras bactérias.

PALAVRAS-CHAVE: dados ômicos; rede de interação de proteínas; redes complexas; medidas de centralidade; *klebsiella pneumoniae*.

KLEBSIELLA PNEUMONIAE BACTERIA PROTEIN INTERACTION NETWORK ANALYSIS USING CENTRALITY METHODS

ABSTRACT: With the advances in Next Generation Sequencing technologies, there has been a significant increase in the availability of omics data. Protein Interaction Networks model the cell's interactions and can be used to analyze genomic characteristics of modeled organisms. The STRING database provides networks of various organisms, including the bacteria *Klebsiella Pneumoniae*. This bacteria is resistant to most antibiotics and causes severe problems for those infected. Based on the hypothesis that the protein interaction network of this bacteria has particular characteristics, an analysis was carried out using Degree, *Closeness*, and *Betweenness* centrality measures. The distribution of degree and betweenness had the expected behavior for scale-free networks. The *Closeness* presented zero values in about 40% of the vertices and a linear behavior in the others. The *Closeness* values indicate the existence of several peripheral vertices that connect only to the nearest ones. This finding raises the possibility of other analyses that consider the distance between vertices and a comparison with networks of other bacteria.

KEYWORDS: omic data; protein network; complex networks; centrality measures; *klebsiella pneumoniae*.

INTRODUÇÃO

Com os avanços nas tecnologias de Sequenciamento de Nova Geração, houve um grande aumento de dados biológicos relacionados a genômica, transcriptômica, proteômica, metabolômica e farmacogenômica. Esses dados são chamados na literatura de “dados ômicos” e permitem análises computacionais que contribuem em estudos de doenças complexas e funções biológicas (MISRA, 2019). A relação existente entre genes e proteínas pode ser representada por meio de Redes de Interação Proteína-Proteína. A Medicina de Redes aplica, entre outras coisas, métodos de Redes Complexas para realizar análises nessas redes a fim de descobrir padrões e características que contribuem para maior entendimento do problema estudado (BARABÁSI; GULBAHCE; LOSCALZO, 2011). Existem diversas bases de dados ômicos, das que tratam de rede de interação de proteína vale destacar o Reactome, STRING e HINT, que são criadas e mantidas com curadoria de especialistas. Essas redes são de livre acesso e atualizadas frequentemente.

A base de dados STRING disponibiliza online¹ redes de 14.094 organismos, dentre eles a bactéria *Klebsiella Pneumoniae*. Essa bactéria é caracterizada como “superbactéria”, já que é resistente à maior parte dos antibióticos e quando infecta o ser humano é capaz de gerar infecções graves, como pneumonia ou meningite (PEREIRA, 2016).

Considerando o exposto e motivado pela hipótese de que bactérias resistentes possuem redes de interação particulares, o objetivo desta pesquisa é realizar uma análise topológica da rede de interação de proteínas da bactéria *Klebsiella Pneumoniae* usando medidas centralidade.

MATERIAL E MÉTODOS

Após um estudo de revisão bibliográfica sobre bactérias resistentes que estão disponíveis na base STRING, foi escolhida a *Klebsiella Pneumoniae* devido ao seu alto índice de resistência à antibióticos e impacto à saúde humana.

A rede foi estudada usando a linguagem Python e a biblioteca NetworkX no ambiente Anaconda. A biblioteca Pandas foi usada para armazenar os valores das centralidades calculadas pelo NetworkX. Por fim, a biblioteca Matplotlib foi utilizada para gerar os gráficos das medidas e fornecer uma análise visual das distribuições.

Medidas de centralidade caracterizam numericamente cada vértice da rede e são utilizadas para analisar a rede de diferentes perspectivas. As medidas utilizadas neste trabalho foram: Grau, *Closeness* e *Betweenness*. O Grau indica quantos vizinhos um nó possui; O *Betweenness* mede gargalos, pontuando os nós que frequentemente aparecem nos caminhos mais curtos entre todos os pares de nós. O *Closeness* mede o comprimento médio dos menores caminho de cada nó para todos os outros nós da rede.

RESULTADOS E DISCUSSÃO

A rede de interação de proteínas da bactéria *Klebsiella Pneumoniae* possui 3.546 vértices e 12.529 arestas. Apesar de ter grau médio de 7, a rede tem hubs com mais de 100 vizinhos. Esse comportamento é esperado em redes livres de escala (BARABÁSI; GULBAHCE; LOSCALZO, 2011). A Figura 1 apresenta uma representação visual da rede usando o algoritmo de *layout Kamada Kawai*. A Figura 2 apresenta a distribuição ordenada das três medidas utilizadas para analisar a rede. No eixo X de cada sub figura estão os 3.546 vértices e no eixo Y o valor de cada medida. A primeira distribuição é a de grau. É possível observar um comportamento livre de escala, ou seja, o grau médio não é representativo já que muitos vértices possuem grau consideravelmente diferente da média 7. A segunda distribuição é de *Betweenness*. A grande maioria dos vértices possuem valor zero ou próximo de zero nessa média. Esse comportamento é esperado, pois os poucos hubs concentram o fluxo dos menores

¹ <https://string-db.org/>

caminhos. A terceira e última distribuição se refere ao *Closeness*. Da posição 1500 até 3500 existe um crescimento praticamente linear, porém os últimos vértices apresentam um crescimento mais acentuado. Os vértices de 1 a 1500 têm *closeness* praticamente nulos. Esses vértices estão na periferia da rede. A Figura 1 apresenta um indício visual desse comportamento ao se analisar os vértices mais externos.

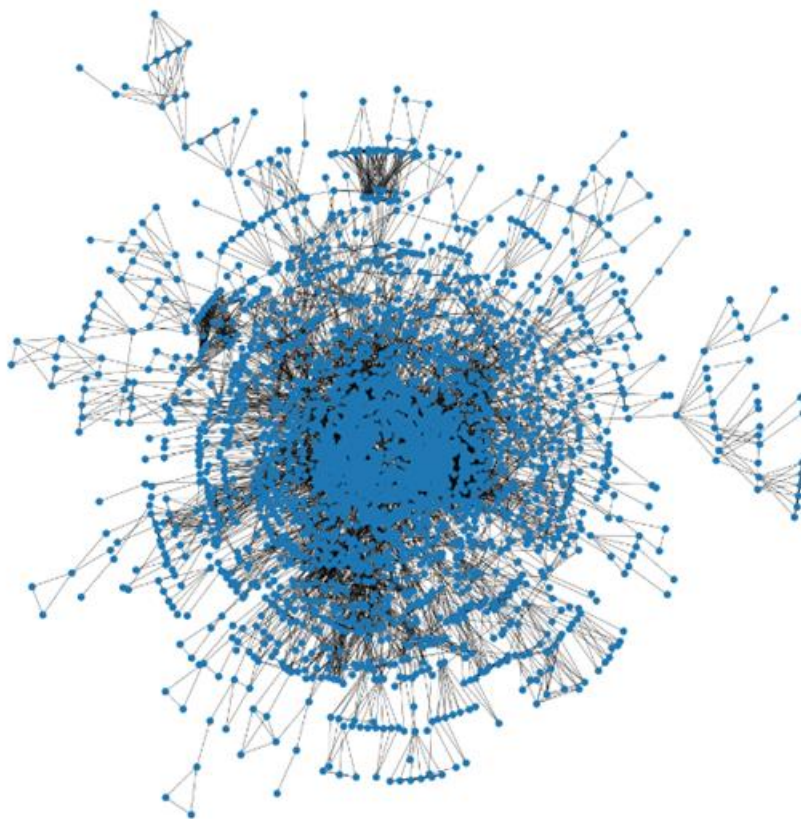


FIGURA 1. Representação gráfica da rede de interação de proteínas da bactéria *Klebsiella Pneumoniae*.

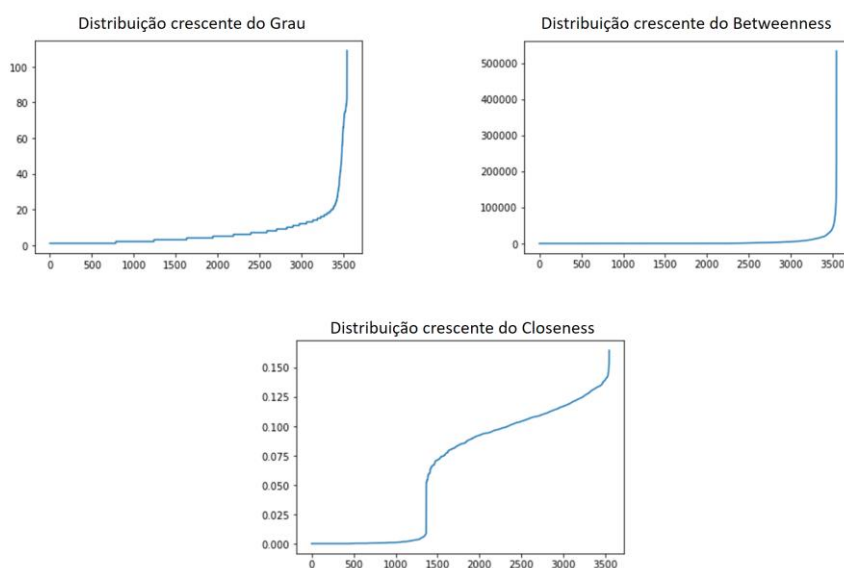


FIGURA 2. Gráficos com medidas de centralidade Grau, *Betweenness* e *Closeness* referentes a rede de interação de proteínas da bactéria *Klebsiella Pneumoniae*.

CONCLUSÕES

A distribuição de Grau e de *Betweenness* teve o comportamento esperado para redes livres de escala. O *Closeness* apresentou valores próximos de zero em cerca de 40% dos vértices e um comportamento linear nos demais. Os valores do *Closeness* indicam a existência de vários vértices periféricos que se conectam apenas com vizinhos próximos. Essa descoberta levanta a possibilidade de outras análises que consideram a distância entre vértices, assim como a comparação com as redes de outras bactérias.

REFERÊNCIAS

BARABÁSI, Albert-László; GULBAHCE, Natali; LOSCALZO, Joseph. Network medicine: a network-based approach to human disease. **Nature reviews genetics**, v. 12, n. 1, p. 56-68, 2011.

ESPINDOLA, Foued Salmen et al. Recursos de bioinformática aplicados às ciências ômicas como genômica, transcriptômica, proteômica, interatômica e metabolômica, 2010. Disponível em: <<http://www.seer.ufu.br/index.php/biosciencejournal/article/view/7146/4861>>

MISRA, Biswapriya B. et al. Integrated omics: tools, advances and future approaches. **Journal of molecular endocrinology**, v. 62, n. 1, p. R21-R45, 2019.

PEREIRA, P. S. **Caracterização molecular de *Klebsiella pneumoniae* produtoras de carbapenemases do tipo KPC-2, NDM-1 e OXA-370 isoladas no Brasil**. 2016. 213f. Tese (Doutorado em Biologia Celular e Molecular) - Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Rio de Janeiro, RJ, 2016.