

Relatório Prático

Disciplina de Redes Complexas - PESC - COPPE - UFRJ

Vinícius W. Salazar, Prof. Daniel R. Figueiredo

Outubro de 2019

Introdução

Nesse relatório vamos explorar a análise de redes de alguns conjuntos de dados do livro “Network Science”, do professor Albert-László Barabási. Os dados podem ser encontrados **nesta página**. Para cada conjunto, vamos caracterizar métricas como grau, distância e tamanho de componentes conexas. Para isso, foram usadas as bibliotecas NetworkX v2.4¹, Pandas v0.25.4² e Matplotlib v3.1.1³.

Conjunto 1 - proteínas

Uma rede representando interações proteína-proteína em leveduras. Cada vértice representa uma proteína, que estão conectadas se interagem fisicamente dentro da célula. Dados originais:

- Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., ... & Vidal, M. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898), 104-110.

Nosso arquivo de input é no formato .tsv com duas colunas, onde cada coluna representa um vértice e cada linha representa uma aresta.

Temos um grafo $G_1 = (V, E)$ com $V = 2017$ e $E = 2929$. Vamos analisar algumas métricas:

Graus de G_1

- **Máximo** ($n = 5$): [91, 82, 81, 52, 46]
- **Mínimo** ($n = 5$): [1, 1, 1, 1, 1]
- **Média** ≈ 2.904
- **Mediana** = 2.0
- **Desvio padrão** ≈ 4.884

A distribuição de cauda pesada do grau de G_1 pode ser observada na **Figure 1**.

Densidade de $G_1 \approx 1.44 \times 10^{-3}$

Componentes conexas

G_1 possui **185 componentes conexas**, sendo que a maior componente conexa (CC_{G_1}) tem **1646 vértices**, enquanto a segunda e terceira menor têm 6 e 5, com a **mediana = 2** para o número de vértices em cada componente conexa., demonstrando outro **padrão de cauda pesada**.

Como G_1 não é conectado, vamos medir o diâmetro e calcular o baricentro de CC_{G_1} :

- $diam(CC_{G_1}) = 14$
- $baricentro(CC_{G_1}) = V_{1356}$

Adamic-Adar

Podemos computar o índice Adamic-Adar $AA_{(i,j)}$ para um par de vértices (i,j) em G_1 , por exemplo:

$$AA_{(6,249)} \approx 0.71$$

Conjunto 2 - metabólitos

Uma rede representando reações metabólicas em bactérias *E. coli*. Cada vértice é um metabólito, e cada link direcionado $A \rightarrow B$ representa uma reação onde A é um reagente e B é um produto. Dados originais:

- Schellenberger, J., Park, J. O., Conrad, T. M., & Palsson, B. Å. (2010). BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC bioinformatics, 11(1), 213.

Nosso arquivo de input é um .tsv com duas colunas, onde cada coluna representa um vértice e cada linha representa uma aresta **direcionada**.

Temos um grafo $G_2 = (V, E)$ com $V = 1039$ e $E = 5801$.

Graus de G_2

- **Máximo** ($n = 5$): [906, 522, 337, 289, 279]
- **Mínimo** ($n = 5$): [1, 1, 2, 2, 2]
- **Média** ≈ 11.167
- **Mediana** = 6.0
- **Desvio padrão** ≈ 37.621

A distribuição de cauda pesada dos **graus de entrada** de G_2 pode ser observada na **Figure 2**.

Densidade de $G_2 \approx 5.38 \times 10^{-3}$

Componentes conexas

G_2 não é fortemente conexo, porém é fracamente conexo. Possui **147 componentes fortemente conexas**, cuja maior é representada por:

- $CC_{G_2} = (893, 5435)$
- $diam(CC_{G_2}) = 8$
- $\$baricentro(CC_{G_2}) = V_{592}$

Caminho mais curto

Podemos computar um caminho mais curto $S_{(i,j)}$ dentro de CC_{G_2} como por exemplo $S_{(991,749)} = [991, 735, 589, 1003, 750, 749]$. Note que o caminho inverso $S_{(j,i)}$ pode ser diferente ou nem existir. Nesse caso, é mais curto: $S_{(749,991)} = [749, 589, 992, 991]$.

Referências

1. Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, “Exploring network structure, dynamics, and function using NetworkX”, in Proceedings of the 7th Python in Science Conference (SciPy2008), Gael Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008
2. Wes McKinney. Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56 (2010)
3. John D. Hunter. Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95 (2007), DOI: 10.1109/MCSE.2007.55