

Lista 3 - Similaridade de Redes Complexas

Redes Complexas para Ciência da Computação

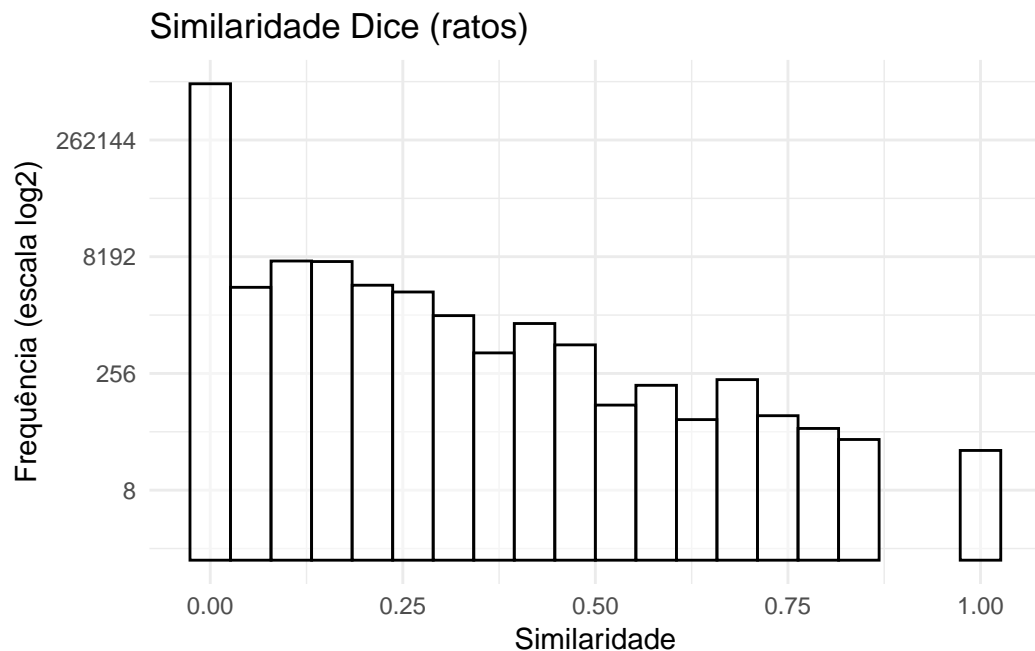
Luben Miguel, Rodrigo Lassance

Para os exercícios da lista, foram escolhidas as redes [mammalia-voles-bhp-trapping](#) e [ia-fb-messages](#). No caso da primeira, como a rede apresenta pesos, todo valor diferente de 0 foi igualado a 1.

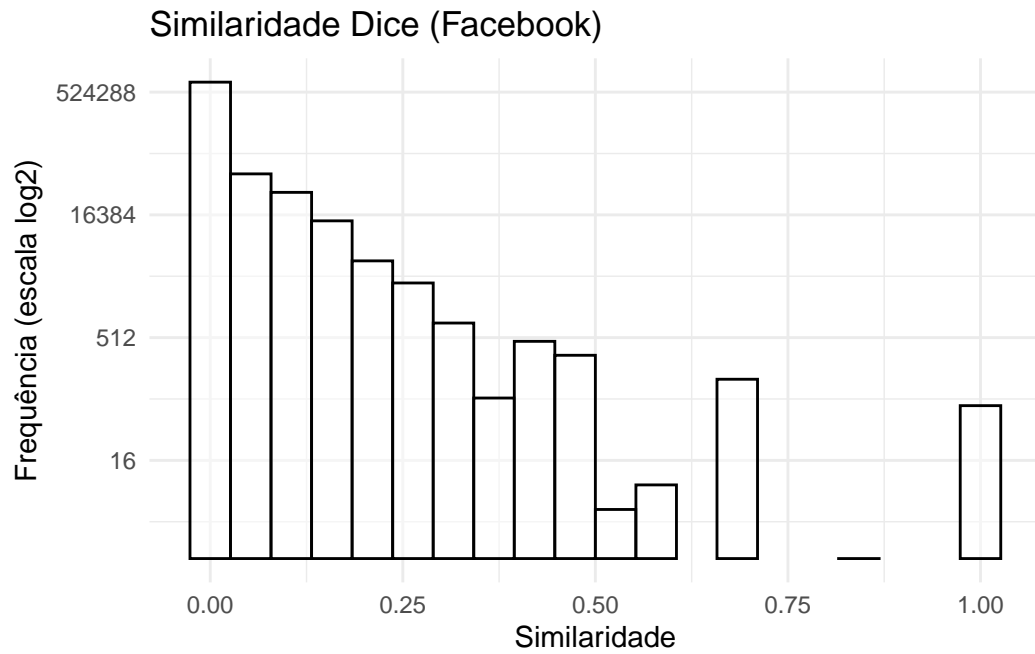
Exercício 1

Considerando a medida de similaridade obtida a partir da distância euclidiana (similaridade Dice) - cuja fórmula é $S_{ij} = 2 \cdot \frac{n_{ij}}{(k_i + k_j)}$ para os vértices i e j - obtemos a seguir o histograma das similaridades considerando todos os pares de vértices para as duas redes de interesse.

- Rede de ratos



- Rede de mensagens do Facebook



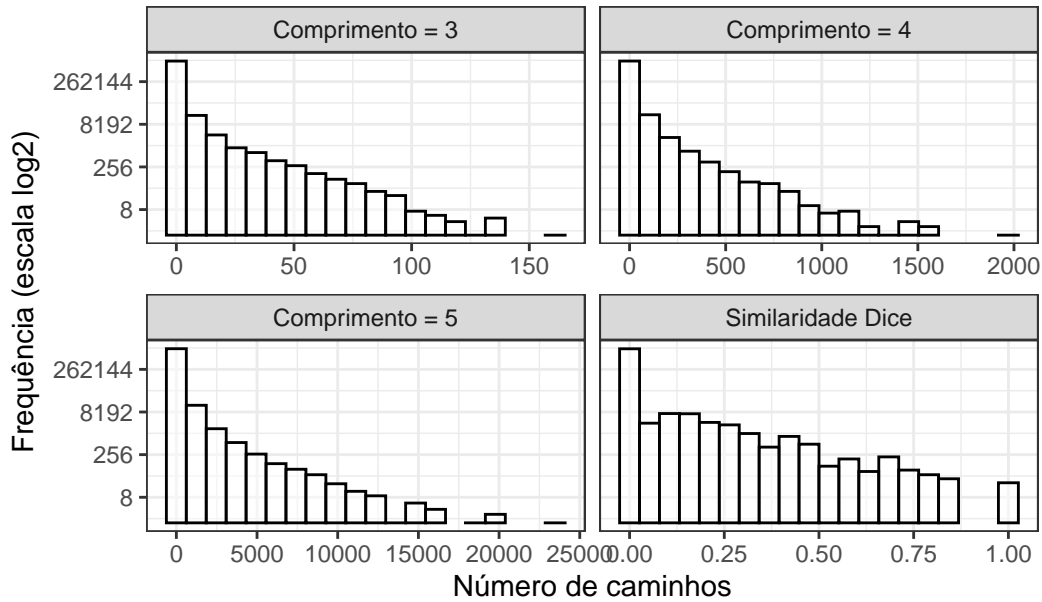
Podemos observar que a rede de ratos apresenta maiores valores de similaridade, porém também possui uma concentração maior de valores próximos de 0.

Exercício 2

A seguir, obtemos o número de caminhos de comprimento 3, 4 e 5 para cada par de vértices (i, j) para cada rede de interesse.

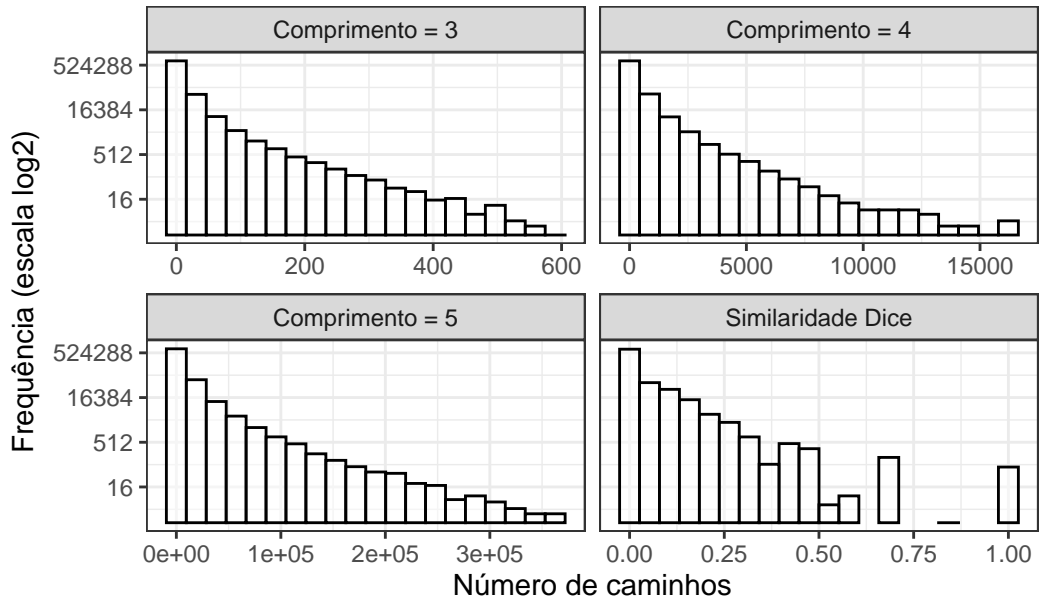
- Rede de ratos

Distribuição do número de caminhos para a rede de ratos



- Rede de mensagens no Facebook

Distribuição do número de caminhos para a rede de mensa



Vemos que, para ambas as redes, a distribuição do número de caminhos decresce suavemente para todos os comprimentos tomados. Além disso, elas parecem apresentar um comportamento similar ao da distribuição da similaridade, tendo apenas certas diferenças na cauda. Isso pode

estar relacionado ao fato de que dois vértices muito similares tendem a possuir um número maior de vizinhos compartilhados e, por consequência, têm um número maior de possíveis caminhos a serem tomados nos diversos comprimentos. Por outro lado, para vértices não muito similares, o número de caminhos não será tão grande quanto para vértices mais similares, por terem vizinhos mais distantes e/ou poucos vizinhos compartilhados. Assim, é esperado que ocorra um certo espelhamento nas distribuições do número de caminhos para diferentes comprimentos e da similaridade entre vértices. Porém, nota-se como principal diferença entre as distribuições a taxa de decrescimento, sendo ela mais rápida entre as curvas do número de caminhos do que na de similaridade.

Exercício 3

Para calcular uma medida de similaridade regular, consideraremos a medida $\sigma = \sum_{m=1}^{\infty} (\alpha A)^m$. Para tal, tomaremos $\alpha < \frac{1}{\lambda_1}$ sendo λ_1 o maior autovalor de A . Os maiores autovalores para as redes de ratos e do Facebook são:

Tabela 1: Limiares para ambas redes

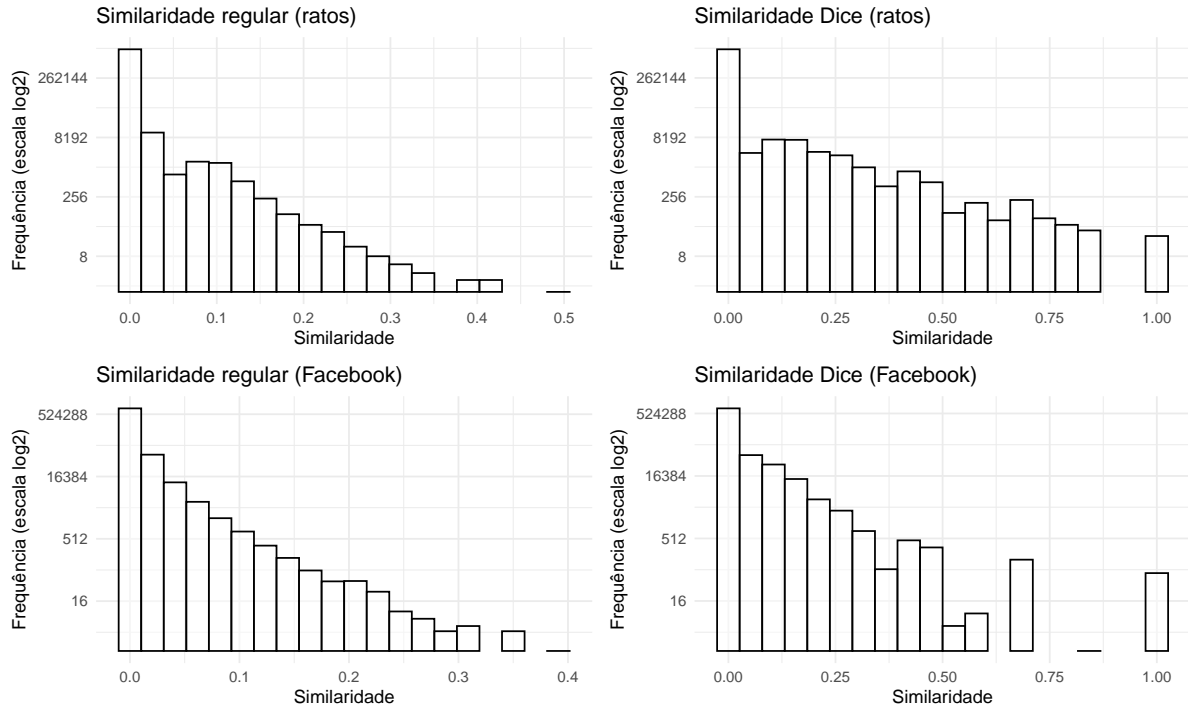
Rede	Limiares
Ratos	0.0825
Facebook	0.0377

Podemos tomar assim $\alpha_{ratos} = 0.07$ e $\alpha_{fb} = 0.035$ para cada rede e computar a matriz σ iterativamente. Tomamos como critério de convergência $\max |\sigma^{(n)} - \sigma^{(n-1)}| \leq \varepsilon$, fixando $\varepsilon = 10^{-6}$.

Notamos pelo gráfico, para a rede de ratos, que a distribuição da similaridade regular decresce mais rapidamente do que a regularidade estrutural. Ou seja, mesmo nos casos em que ambos os vértices de interesse compartilham vizinhos entre si, os vizinhos de um dos vértices tendem a ser menos similares ao outro. O mesmo argumento pode ser aplicado para a rede de mensagens do Facebook, visto que novamente a similaridade regular tem maior decrescimento na distribuição do que a similaridade estrutural.

Exercício 4

Em um contexto em que se pode representar a informação contida em um objeto através de um vetor de *embeddings*, torna-se possível comparar diferentes objetos por meio das diferenças entre as suas representações. Desse modo, trazendo essa noção de *embeddings* para o contexto de redes complexas, o método **node2vec** permite representar um vértice a partir dos vizinhos que ele possui. Cada vértice seria representado por um vetor binário, sendo que se o vértice i é



vizinho do vértice j , então o i -ésimo vetor receberia o valor 1 na j -ésima entrada. Uma vez que se tenha uma representação para cada vértice, o processo de comparação pode ser resumido a uma avaliação da similaridade entre esses dois vetores. Para estes casos, a similaridade cosseno é uma das métricas mais comumente utilizadas.

Exercício 5

Consideraremos neste caso os pares de vértices (400, 434) e (675, 615) da rede de ratos. Para obter o primeiro tempo de passagem médio de i para j , consideraremos que a probabilidade de ir de um vértice para o outro é $1/k_i$, sendo i o vértice de saída e k_i seu grau. Simularemos $B = 1000$ caminhos do vértice i ao j e armazenaremos o tamanho de cada caminho realizado. Ao final, calcularemos o caminho/tempo médio de chegar ao vértice j partindo do vértice i . Depois, repetiremos o processo de j para i . A tabela abaixo resume os resultados:

Tabela 2: MFPT(i, j) e MFPT(j, i) para dois pares de vértices distintos

Vértice.i	Vértice.j	MFPT.i.j.	MFPT.j.i.
400	434	391.235	2937.692
675	615	1088.248	1830.095

Vemos que $MFPT(i, j) \neq MFPT(j, i)$ neste exemplo da rede de ratos. De fato, essa desigualdade é esperada. Em um exemplo em que i e j compartilham apenas um vizinho, porém o número de vizinhos de i é menor que o de j , é esperado que, caso não se passe para o vizinho compartilhado em um primeiro passo, o processo demorará muito mais para voltar a j do que voltar para i . Neste caso, visualiza-se isso pelo número médio de passos de j para i em comparação ao de i para j .