

FGV EMap  
João Pedro Jerônimo

Ciência de Redes  
Revisão para A1

Rio de Janeiro  
2025

# Conteúdo

1	Grafos .....	3
2	Medidas de Centralidade .....	6
3	Redes Aleatórias .....	13
3.1	Ideia Inicial .....	14
3.2	Evolução das Redes Aleatórias .....	15
3.3	Distribuição de tamanhos de Cluster .....	16
3.4	Mundos pequenos .....	16
3.5	Coeficiente de Clustering .....	16
3.6	Conclusão .....	17
4	Evoluções de Redes .....	18
4.1	Anexação Uniforme .....	19
4.2	Anexação Preferencial .....	20

**Grafos**

De antemão valhe ressaltar que essa matéria, por mais que seja chamada de **Ciência de Redes**, o termo **rede** se refere a um grafo, não ao tipo específico de grafo que se é visto em **Fluxo em Redes** quando estudamos matemática discreta. Então que já fique esclarecido de antemão que, ao citarmos redes, estamos nos referindo a um grafo no geral, desde que o contrário seja explicitado

Essa sessão será apenas algumas definições que não foram passadas no curso de Matemática Discreta, então conceitos que forem citados sobre grafos e não houver definição nesse resumo, a mesma estará no recap de Matemática Discreta. Aqui segue algumas notações sobre grafos para que não fique confuso:

- $G(V, E) :=$  Grafo com conjunto de vértices  $V$  e de arestas  $E$  (edges)
- $N(v) :=$  Vizinhança do vértice  $v$  (Neighbourhood)
- $\delta(v) :=$  Grau do vértice  $v$
- $K_n :=$  Grafo completo com  $n$  vértices
- $K_{m,n} :=$  Grafo completo bipartido com  $m$  vértices no primeiro conjunto e  $n$  vértices no segundo
- $X(G) :=$  Número cromático de  $G$
- $X'(G) :=$  Número cromático por arestas de  $G$

**Definição 1.1** (Grau Médio): Dado um grafo não-dirigido  $G(V, E)$ , o grau médio de  $G$  é:

$$\delta_{\text{med}}(G) := \frac{1}{|V|} \sum_{v_i \in V} \delta(v_i) \quad (1)$$

Se  $G$  é dirigido, podemos definir os graus médios de entrada e saída

$$\delta_{\text{med}}^{\text{in}}(G) := \frac{1}{|V|} \sum_{v_i \in V} \delta^{\text{in}}(v_i) \quad \text{Entrada} \quad (2)$$

$$\delta_{\text{med}}^{\text{out}}(G) := \frac{1}{|V|} \sum_{v_i \in V} \delta^{\text{out}}(v_i) \quad \text{Saída} \quad (3)$$

**Definição 1.2** (Distribuição do Grau): A distribuição do grau de um Grafo  $G(V, E)$  é a distribuição da variável aleatória  $X$ , sendo  $X$  o grau do vértice que eu escolho ao acaso

**Para os teoremas a seguir e daqui em diante, consideremos a matriz de incidência de forma que  $A_{ij} = 1$  se a aresta  $j$  se conecta no vértice  $i$  e, 0 do contrário (-1 se  $G$  for dirigido).**

**Teorema 1.1:** Dado um grafo  $G(V, E)$  e sua matriz de incidência  $A$ , temos que:

$$\text{n}^\circ \text{ de ciclos} = |E| - \text{posto}(A) \quad (4)$$

*Demonstração:*

$$\begin{aligned} \text{posto}(A) + \dim(N(A)) &= |E| \\ \Leftrightarrow |E| - \text{posto}(A) &= \dim(N(A)) \end{aligned} \quad (5)$$

Porém, a dimensão do núcleo de  $A$  é a quantidade de ciclos no grafo, então eu tenho que:

$$\text{n}^\circ \text{ de ciclos} = |E| - \text{posto}(A) \quad (6)$$

□

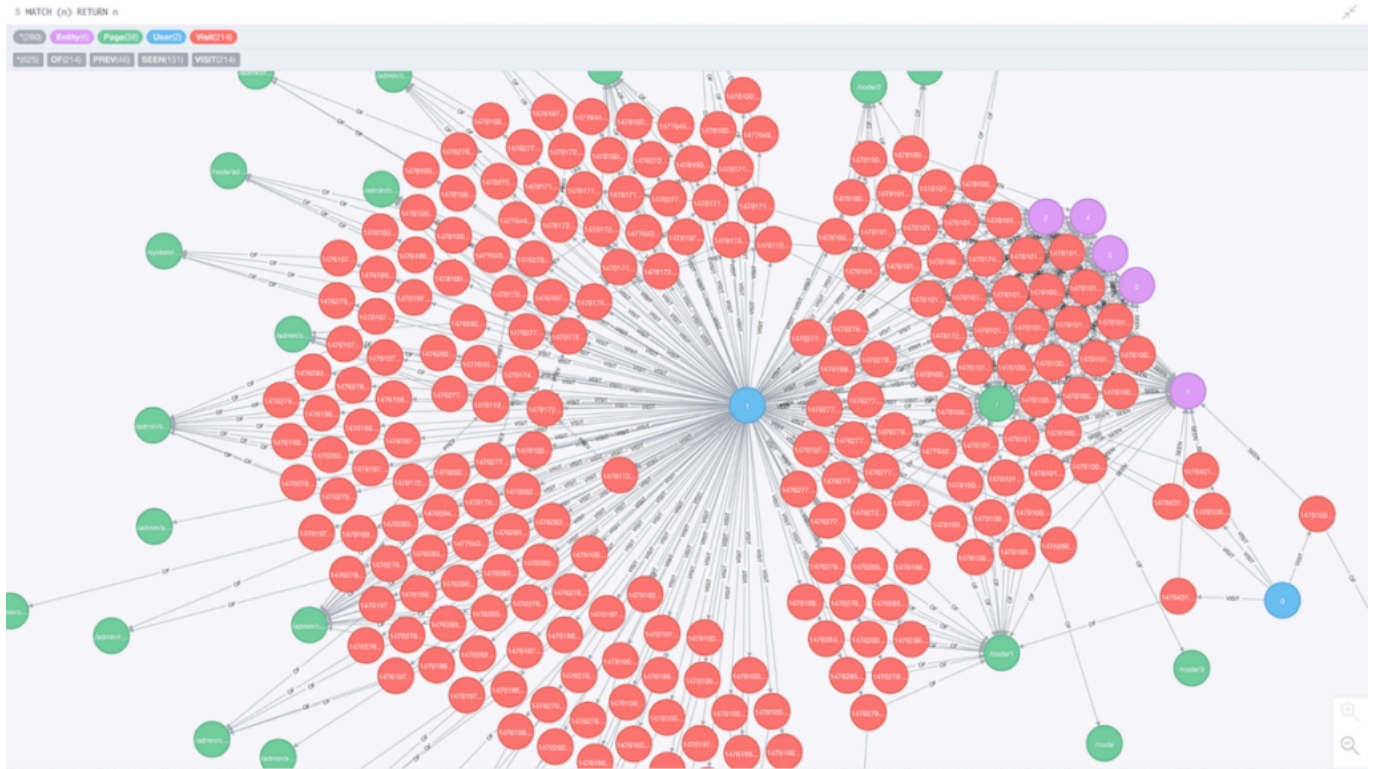
**Definição 1.3** (Coeficiente de Clustering): Dado um grafo  $G(V, E)$ , o coeficiente de clustering de um nó  $v \in V$  é:

$$C(v) := \frac{2E_v}{\delta(v)(\delta(v) - 1)} \quad (7)$$

onde  $E_v$  é a quantidade de arestas ligadas aos nós vizinhos

# **Medidas de Centralidade**

Quando estamos vendo aplicações reais de grafos, é muito comum querermos ver o “quão importante” um nó é no contexto que estamos analisando. Por exemplo, se nosso grafo representa as conexões entre servidores que um pacote pode percorrer, faz muito sentido querermos ver qual o servidor que quase todos os pacotes percorrem



Imagine que esse é o grafo que estávamos falando (Não importa o que ele representa de verdade, só finge que é o caso que falamos), então o nó azul tem uma importância MUITO grande, mas como podemos medir isso? Nem sempre o grafo vai tá arrumadinho assim pra gente. Daí que surgem as medidas de Centralidade.

**Definição 2.1** (Farness/'Lonjura'): Dado um grafo  $G(V, E)$ , a farness de um vértice  $v_i$  é dada por

$$L(v_i) := \sum_{v_i \neq v_j \in V} d(v_i, v_j) \quad (8)$$

onde  $d(v_i, v_j)$  é o tamanho do menor caminho entre  $v_i$  e  $v_j$

Essa medida mede o quão longe o nó está dos outros, de forma que, quanto maior essa medida é, menos importante o meu nó é (Depende do contexto analisado)

**Definição 2.2** (Closeness/Proximidade): Dado um grafo  $G(V, E)$ , a proximidade/closeness do vértice  $v_i \in V$  é dada por:

$$C(v_i) := \frac{|V|}{L(v_i)} \quad (9)$$

Por convenção, se  $v_i$  e  $v_j$  estão em componentes conexas separadas em  $G$ , então  $d(v_i, v_j) = \infty$ , o que torna a definição de antes inútil, então podemos redefinir como:

$$C(v_i) := \frac{1}{|V|} \sum_{v_i \neq v_j \in V} \frac{1}{d(v_i, v_j)} \quad (10)$$

**Definição 2.3** (Betweenness/Intermediação): Dado um grafo  $G(V, E)$  e  $P(v_i, v_j)$  o conjunto de todos os menores caminhos possíveis entre  $v_i$  e  $v_j$ , então a intermediação de  $v_i$  é:

$$B(v_i) := \sum_{v_s, v_t \in V} \frac{|c \in P(v_s, v_t); v_i \in c|}{|P(v_s, v_t)|} \quad (11)$$

Saindo um pouco dessas definições, vamos tentar pensar em alguma medida mais básica e intuitiva. Uma medida bem padrão que podemos pensar logo de cara é simplesmente o grau do vértice, já que, quanto mais vértices ele se ligar, mais importante ele é! Em muitas literaturas sobre redes o grau do vértice é chamado de **Centralidade de Grau**.

Um outro pensamento que pode surgir a partir desse é: “Poxa, meu vértice tem um grau alto, então ele é importante, mas eu quero valorizar aqueles vértices que se conectam com ele, afinal, se ele é importante, os vértices que estão diretamente ligados nele também são, não é?”, e esse pensamento não está errado! É dessa ideia que surge a centralidade por autovetor. Funciona assim: Vamos inicialmente assumir que todos os nossos vértices  $v_i$  tem importância  $x_i^{(0)} = 1$ , o que não me é muito útil agora, porém, vamos tentar fazer uma nova estimativa baseada nos vizinhos, que tal a nova centralidade do vértice  $v_i$  ser a soma da centralidade dos vizinhos? Isso faz com que a importância do  $v_i$  se baseie no quão importante são seus vizinhos! Eu posso expressar isso com uma fórmula:

$$x_i^{(1)} = \sum_j A_{ij} x_j^{(0)} \quad (12)$$

Onde  $A$  é minha matriz de adjacência. Se meu nó  $v_i$  não é vizinho de  $v_j$ , então  $A_{ij} = 0$  o que faz com que minha centralidade  $x_j^{(0)}$  não seja somada. Posso reformular isso de forma matricial:

$$x^{(1)} = Ax^{(0)} \quad (13)$$

onde  $x^{(k)}$  é o vetor com entradas  $x_i^{(k)}$ . Se fizermos esse processo várias vezes, depois de  $k$  passos, vamos ter algo do tipo:

$$x^{(k)} = A^k x^{(0)} \quad (14)$$

Tomemos a liberdade, então, de escrever  $x^{(0)}$  como uma combinação linear dos autovetores  $w_j$  de  $A$  de forma que

$$x^{(0)} = \sum_{j=1}^n c_j w_j \quad (15)$$



Para alguma escolha apropriada de  $c_j$ . Então temos:

$$x^{(k)} = A^k \sum_{j=1}^n c_j w_j = \sum_{j=1}^n c_j \lambda_j^k w_j = \lambda_1^k \sum_{j=1}^n c_j \left( \frac{\lambda_j}{\lambda_1} \right)^k w_j \quad (16)$$

De forma que  $\lambda_j$  são os autovalores de  $A$  e  $\lambda_1$  pode ser, sem perda de generalização, o maior de todos em módulo. Como  $\lambda_i/\lambda_1 < 1 \forall \lambda_i$  com  $i \neq j$ , então:

$$\lim_{k \rightarrow \infty} \sum_{j=1}^n c_j \lambda_j^k w_j = c_1 \lambda_1 w_1 \quad (17)$$

Ou seja, o vetor de centralidades que limita as centralidades que eu fiz antes é proporcional ao autovetor associado ao maior autovalor de  $A$ , que é equivalente a dizer que o vetor de centralidades  $x$  satisfaz:

$$Ax = \lambda_1 x \quad (18)$$

**Definição 2.4** (Centralidade Autovalor): Seja  $r$  um vetor com as centralidades dos vértices  $v_i$  de uma rede  $G$  de forma que  $r_i$  = centralidade de  $v_i$ , então:

$$Ar = \lambda_1 r \quad (19)$$

onde  $\lambda_1$  é o maior autovalor de  $A$

Agora temos outro problema. Quando temos um grafo dirigido, essa medida de centralidade autovalor já não funciona, já que se um nó não tem nenhuma aresta apontando para ele (Apenas saem arestas dele), ele não terá sequer uma centralidade, e isso afeta não só esse vértice como os vértices que ele aponta, que não terão nenhuma “pontuação” adicionada por serem apontados por esse vértice, e isso não pode ocorrer, já que não faz muito sentido na maioria das aplicações práticas. O que podemos fazer para contornar isso? Então entra a solução a seguir:

$$x_i = \alpha \sum_j A_{ij} x_j + \beta \quad (20)$$

Onde  $\alpha$  e  $\beta$  são constantes positivas. O primeiro termo é a centralidade autovetor que vimos antes, porém o termo  $\beta$  garante que os nós que comentei anteriormente (Sem grau de entrada) possuam uma pontuação e possam contribuir para a pontuação dos nós que eles apontam. Essa medida é interessante por conta do termo  $\alpha$  que balanceia o termo constante e a medida de centralidade autovetor. Podemos expressar isso de forma matricial:

$$x = \alpha Ax + \beta \mathbf{1} \quad (21)$$

Onde  $\mathbf{1} = (1, \dots, 1)$ . Se rearranjarmos para  $x$ , obtemos:

$$x = \beta(I - \alpha A)^{-1} \mathbf{1} \quad (22)$$

Normalmente colocamos  $\beta = 1$  pois não estamos interessados em saber o valor exato das centralidades, mas saber quais vértices são ou não mais ou menos centrais.

$$x = -\alpha \left( A - \frac{1}{\alpha} I \right)^{-1} \quad (23)$$

Perceba que eu quero que  $A - \frac{1}{\alpha} I$  seja invertível, e isso acontece quando  $\frac{1}{\alpha} \neq \lambda_j$  onde  $\lambda_j$  são os autovalores de  $A$ . Ou seja, o meu  $\alpha$  não é completamente arbitrário, eu vou ter que analisar

o contexto da minha aplicação. Porém, muito comumente, se é utilizado  $\alpha = \frac{1}{\lambda_1}$  com  $\lambda_1$  sendo o maior autovalor

**Definição 2.5** (Centralidade de Katz): Dado uma rede  $G(V, E)$  e duas constantes  $\alpha, \beta > 0$ , o vetor de centralidades de katz de todos os nós em  $V$  é:

$$K(V) = \beta(I - \alpha A)^{-1} \mathbf{1} \quad (24)$$

Onde  $A$  é a matriz de adjacência de  $G$ . ( $K(V) \in \mathbb{R}^{|V|}$ )

Um outro tipo de medida surge quando queremos responder a questão: “Se eu estou navegando entre meus nós, ao longo prazo, qual é o nó que eu mais vou percorrer/parar nele?”. Um exemplo são páginas na internet que referenciam entre si, daí surge o nome da medida: **PageRank**. O que fazemos essencialmente é transformar a rede em uma cadeia de markov. Por exemplo:

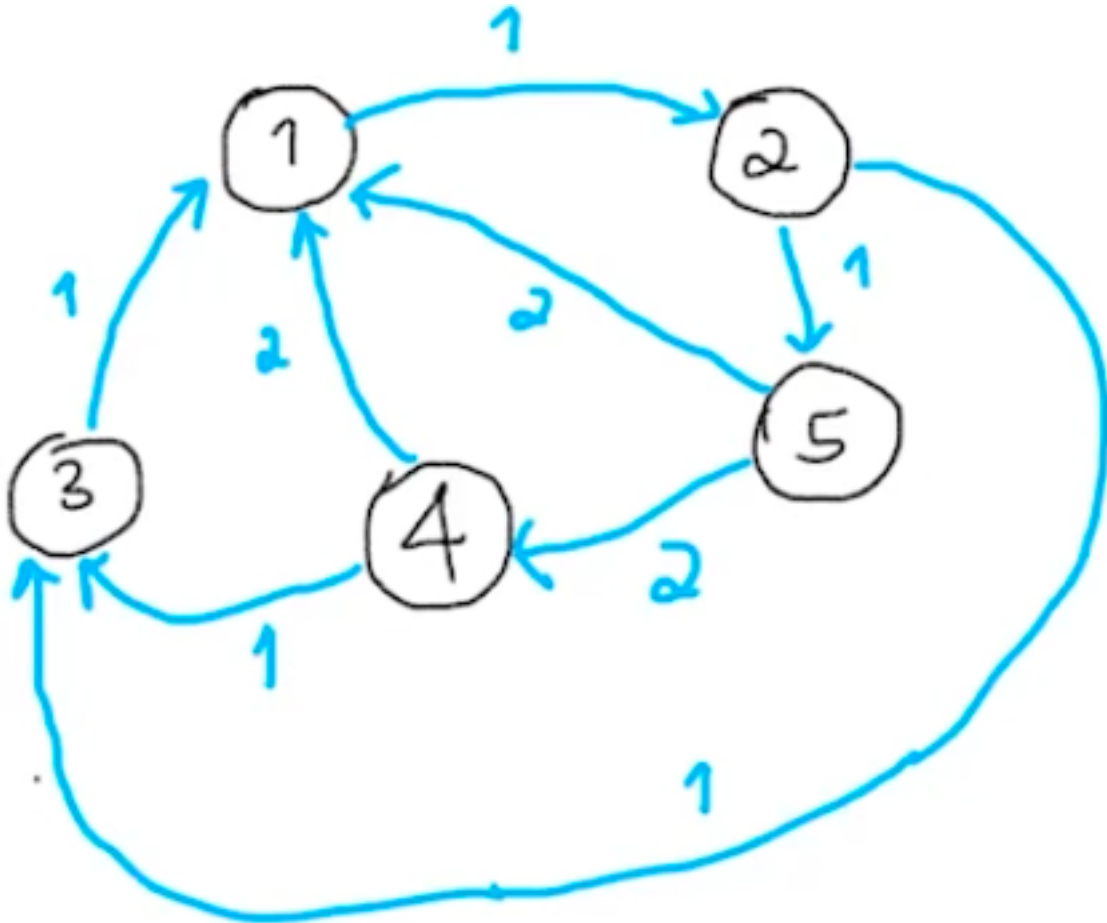


Figura 1: Grafo de Exemplo 1

Vamos supor que estamos no nó 4 e queremos escolher aleatoriamente entre os nós 3 e 1 para irmos, como podemos ver na distribuição dos pesos (Nesse exemplo, isso indica que a página 4 tem 2 links referenciando a página 1 e apenas 1 link referenciando a página 3), então teríamos:

$$\begin{aligned} \mathbb{P}(4 \rightarrow 3) &= \frac{1}{3} \\ \mathbb{P}(4 \rightarrow 1) &= \frac{2}{3} \end{aligned} \quad (25)$$

E fazemos isso definindo uma matriz estocástica  $H$  de tal forma que:



Definimos um  $\alpha \in (0, 1)$ , onde podemos interpretar  $\alpha$  como a chance do meu navegador permanecer no meu nó. Definimos então nossa nova matriz de chances da seguinte forma:

$$\mathbb{G} = \alpha H + (1 - \alpha)C \quad (29)$$

De forma que  $C$  é uma matriz  $n \times n$  com todas as entradas iguais a  $1/n$  para representar um dirigido onde todos os nós apontam para todos os outros nós (Representando a ideia de que eu posso ir para o nó que eu quiser). Porém, há uma propriedade que, se eu tenho uma combinação convexa entre duas matrizes estocásticas/markovianas, então o resultado é uma matriz markoviana. Ou seja, eu ainda posso aplicar a mesma ideia de antes do vetor  $p_0$  inicial e aplicar o limite, assim, eu vou obter meu vetor de centralidades  $r$ , de tal forma que

$$\lim_{t \rightarrow \infty} \mathbb{G}^t p_0 = r \quad (30)$$

**Definição 2.6** (PageRank): Sejam a matriz  $\mathbb{G}$  como definida anteriormente e o vetor inicial  $p_i$  sendo a  $i$ -ésima coluna de  $\mathbb{G}$ , então o vetor de centralidades PageRank  $r$  onde a  $k$ -ésima entrada é a centralidade de  $v_k$ , então:

$$r = \lim_{t \rightarrow \infty} \mathbb{G}^t p_0 \quad (31)$$

# **Redes Aleatórias**

### 3.1 Ideia Inicial

Também chamadas de **Redes Erdős-Renyi** ou **Redes de Poisson**, são tipos de redes que vão se montando aleatoriamente. Por exemplo, imagine que você está em uma festa e o anfitrião está fornecendo um vinho da melhor qualidade, mas ele não avisou ninguém. Um convidado curioso, por acidente, provou desse vinho e **adorou**, então ele vai contar para as pessoas da festa. A pergunta é, para quem ele vai falar? Ele vai falar para todos? Vai sobrar vinho para você?

Em cima disso conseguimos montar as redes aleatórias, onde cada par de nós (Aresta) é formado de acordo com uma **probabilidade**

**Definição 3.1.1** (Rede Aleatória): Uma rede aleatória é um grafo  $G(V, E)$  de  $|V| = N$  nós onde cada par de nós é conectado por uma probabilidade  $p$

Como cada aresta tem uma probabilidade  $p$  de aparecer, podemos interpretar ela como ela aparecer ou não sendo uma variável indicadora, de forma que o número total de arestas segue uma distribuição binomial. Ou seja, a probabilidade a quantidade de arestas ser  $L = l$  é:

$$\mathbb{P}(L = l) = \binom{\binom{N}{2}}{l} p^l (1-p)^{\frac{N(N-1)}{2} - l} \quad (32)$$

Podemos aplicar a mesma ideia para o grau de um vértice também:

$$\mathbb{P}(\delta(v) = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad (33)$$

Já que meu vértice pode se ligar a  $N - 1$  vértices com probabilidade  $p$ , então isso vira a soma das variáveis indicadoras que são 1 quando o meu vértice se liga com outro vértice, de forma que eu tenho a soma de  $N - 1$  variáveis de bernoulli

Com isso, nós temos:

$$\delta_{\text{med}}(G) = (N-1)p \quad (34)$$

E podemos obter também a variância dos graus

$$\mathbb{V}(\delta(v)) = (N-1)p(1-p) \quad (35)$$

Então, apenas para resumir, temos que:

$$\begin{aligned} \text{Número de arestas } L &\sim \text{Bin}\left(\binom{N}{2}, p\right) \\ \text{Grau do vértice } \delta(v) &\sim \text{Bin}(N-1, p) \end{aligned} \quad (36)$$

Porém, em redes reais, elas são **esparsas**, ou seja, eu tenho **muitos** nós e graus pequenos. E lembra qual é a distribuição que é a binomial com  $n$  muito grande? Exato, a **Poisson**! Essas redes aleatórias também são chamadas de **redes de poisson**

$$\mathbb{P}(\delta(v) = k) = e^{-\delta_{\text{med}}(G)} \frac{\delta_{\text{med}}(G)^k}{k!} \quad (37)$$

Ou seja, para  $N$  muito grande e  $k$  pequeno com relação a  $N$ , podemos estimar de forma que:

$$\text{Grau do vértice } \delta(v) \sim \text{Poisson}(\delta_{\text{med}}(G)) \quad (38)$$

### 3.2 Evolução das Redes Aleatórias

Conforme iniciamos um grafo com um grau médio 0 e vamos aumentando ele aos poucos, nós percebemos que a partir de um ponto chave, os nós começam a se agrupar em algo que chamamos de **componente gigante**, que seria a maior componente conexa da rede.

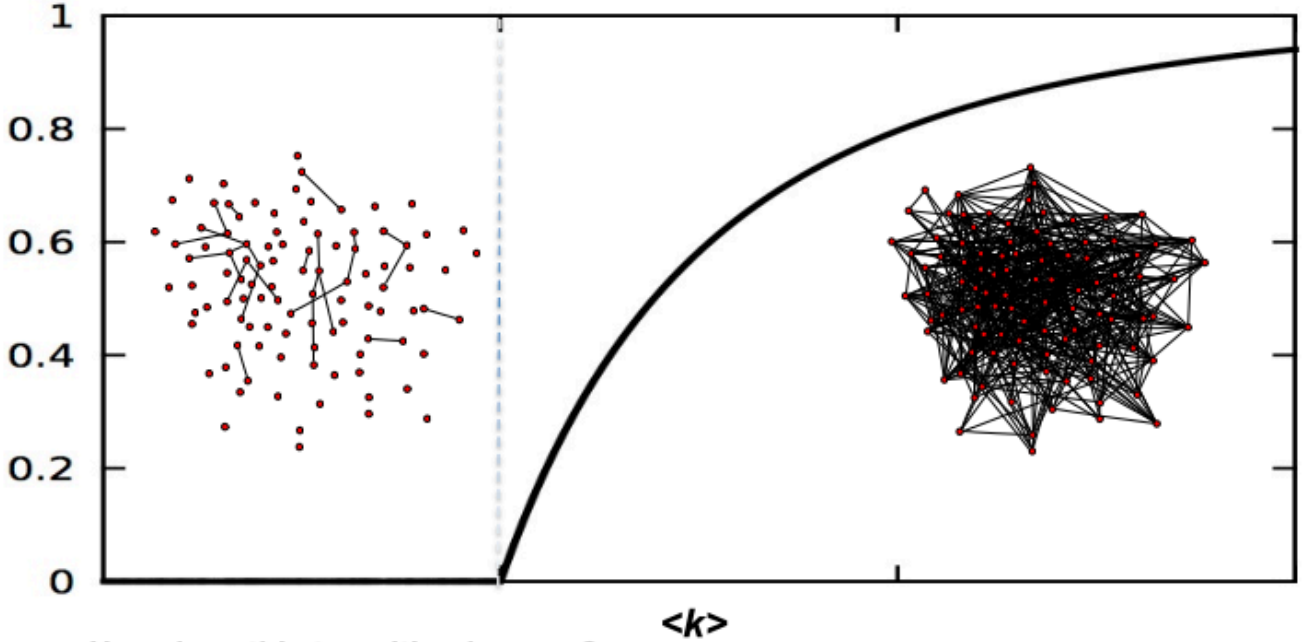


Figura 3: Gráfico que mostra a fração de nós dentro de uma grande componente conexa em função do grau médio

Quanto  $\delta_{\text{med}}(G) < 1$ , então a quantidade de nós na componente gigante é desprezível em relação à quantidade de nós na rede, porém, a partir de  $\delta_{\text{med}}(G) = 1$ , isso indica que temos, pelo menos,  $\frac{n}{2}$  componentes conexas, o que já começa a fazer uma diferença no gráfico.

Dado uma rede  $G(V, E)$ , vamos definir a fração de nós que **não está** na componente gigante como:

$$u = 1 - \frac{N_G}{|V|} \quad (39)$$

De forma que  $N_G$  é a quantidade de nós dentro dessa componente gigante, vamos definir essa componente como  $\Psi \subseteq V$ . Se um nó  $v_i \in \Psi$ , então ele deve estar interligado com outro nó  $v_j$ , que também deve satisfazer  $v_j \in \Psi$ . Por isso, se  $v_i \notin \Psi$ , então isso pode ocorrer por duas razões:

- $\{v_i, v_j\} \notin E$ . A probabilidade de isso acontecer é  $1 - p$
- $\{v_i, v_j\} \in E$ , porém  $v_j \notin \Psi$ . A probabilidade de isso acontecer é  $pu$

Então temos:

$$\mathbb{P}(v_i \notin \Psi) = 1 - p + pu \quad (40)$$

Então a probabilidade de que  $v_i$  não esteja linkado a  $\Psi$  por qualquer nó é de  $(1 - p + pu)^{|V| - 1}$ , já que temos outros  $|V| - 1$  nós que poderiam fazer com que  $v_i$  se interligasse a componente gigante.

Sabemos que  $u$  é a fração de nós que não está em  $\Psi$ , para qualquer  $p$  e  $|V|$ , a solução da equação

$$u = (1 - p + pu)^{|V| - 1} \quad (41)$$

nos dá o tamanho da componente gigante por meio de  $N_G = |V|(1 - u)$ . Usando  $p = \frac{\delta_{\text{med}}(G)}{|V|-1}$  e tirando log de ambos os lados, para  $\delta_{\text{med}}(G) \ll |V|$  (Grau médio **muito** menor que  $|V|$ ), obtemos:

$$\ln(u) \approx (|V| - 1) \ln \left[ 1 - \frac{\delta_{\text{med}}(G)}{|V| - 1} (1 - u) \right]$$

Tiramos exponencial e obtemos: (42)

$$u \approx \exp \left\{ -\frac{\delta_{\text{med}}(G)}{1 - u} \right\}$$

Se denotarmos  $S = \frac{N_G}{|V|}$ , obtemos que:

$$S = 1 - e^{-\delta_{\text{med}}(G) \cdot S} \quad (43)$$

### 3.3 Distribuição de tamanhos de Cluster

Queremos também ter uma noção da probabilidade de um nó  $v_i$  qualquer estar em um cluster (Grupo de nós na rede) de tamanho  $s$ . No livro do Newman, ele nos mostra que essa probabilidade é:

$$\mathbb{P}(v_i \in \Psi_{|\Psi|=s}) = e^{-\delta_{\text{med}}(G) \cdot s} \frac{(\delta_{\text{med}}(G) \cdot s)^{s-1}}{s!} \quad (44)$$

### 3.4 Mundos pequenos

Mundos pequenos (Small worlds) são grafos em que, independente da quantidade de vértices, a distância entre dois nós aleatórios costuma ser muito pequeno. Um exemplo é um modelo que cada nó representa todas as pessoas do mundo e as arestas indicam se elas já interagiram e se conhecem ou não (Impressionantemente), tanto que existe a teoria dos 6 graus de distância entre as pessoas

*Vídeo sobre o assunto (Clique aqui)*

E se quisermos ter uma noção de o quão **não-relacionadas** duas pessoas são em uma rede social? Podemos calcular sua distância, obviamente, mas alguns algoritmos ficam computacionalmente inviáveis. Podemos então estimar uma distância média entre dois nós selecionados aleatoriamente no grafo

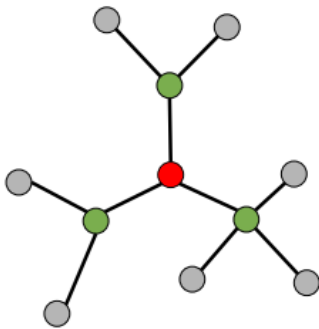


Figura 4: Grafo Árvore

Redes aleatórias costumam ter uma topologia de árvore com praticamente um número constante de graus. Perceba então que eu posso escrever a quantidade de nós  $|V|$  como:

$$\begin{aligned} |V| &= 1 + \delta_{\text{med}}(G) + \dots + \delta_{\text{med}}(G)^{d_{\text{max}}} \\ &= \frac{\delta_{\text{med}}(G)^{d_{\text{max}}} - 1}{\delta_{\text{med}}(G) - 1} \end{aligned} \quad (45)$$

Então vamos ter que:

$$d_{\text{max}} \approx \frac{\log |V|}{\log(\delta_{\text{med}}(G))} \quad (46)$$

### 3.5 Coeficiente de Clustering

Indica o quão agrupado um nó está dentro de uma rede



**Definição 3.5.1** (Coeficiente de Clustering): Dado uma rede  $G(V, E)$ , o coeficiente de clustering de um nó  $v_i \in V$  é definido como:

$$\text{Cluster}(v_i) := \frac{2 \cdot \mathbb{L}(v_i)}{\delta(v_i)(\delta(v_i) - 1)} \quad (47)$$

Onde  $\mathbb{L}(v_i)$  é quantas arestas **entre si** os **vizinhos** de  $v_i$  possuem e  $\frac{\delta(v_i)(\delta(v_i)-1)}{2}$  é a quantidade **máxima** de arestas que poderiam estar interligando os vizinhos de  $v_i$  (Quantidade de arestas em um grafo completo  $K_{\delta(v_i)}$ )

Em redes aleatórias, como as arestas são independentes e tem a mesma probabilidade  $p$  de aparecer, temos:

$$\mathbb{L}(v_i) \approx p \frac{\delta(v_i)(\delta(v_i) - 1)}{2} \Rightarrow \text{Cluster}(v_i) = p = \frac{\delta_{\text{med}}(G)}{|V|} \quad (48)$$

Só que sabemos que, em redes aleatórias, para que esse número seja alto, a probabilidade em si das arestas tem que ser alto, porém, se  $p$  é alto, então a rede aleatória em si será um grande aglomerado, seria um único cluster enorme. Essa característica é um forte indicativo, por exemplo, de que redes como as **redes sociais não são** redes aleatórias

### 3.6 Conclusão

Como conclusão, temos que redes aleatórias **não representam bem as redes da vida real**. Não existem redes na natureza que são corretamente descritas como **redes aleatórias**. Então por que estudar elas? Na verdade, veremos posteriormente que, mesmo elas sendo erradas e irrelevantes, elas são **muito úteis**

# **Evoluções de Redes**

Vimos as redes aleatórias onde os graus dos nós tinham distribuição de Poisson. Mas e se eu quisesse fazer uma rede com distribuição diferente? Muitos pacotes de grafos e redes utilizam de **configuration models**, que são funções que recebem a quantidade de nós da rede e um **vetor** que representa a **função de distribuição** dos graus dos nós

Voltando ao assunto sobre **evoluções**, eu estou interessado em pensar um jeito intuitivo/natural de como as redes vão evoluir com o passar do tempo.

Então vamos imaginar o seguinte cenário. Eu tenho uma rede inicial  $G_0(V_0, E_0)$  e a cada unidade de tempo  $t$  eu vou ter uma nova rede  $G_t(V_t, E_t)$ , de forma que a cada unidade de tempo, eu vou adicionar um novo nó em  $V_{t-1}$  e novas arestas em  $E_{t-1}$ . Qual é a distribuição do grau médio desses nós? O que podemos inferir dessa rede?

## 4.1 Anexação Uniforme

Vamos imaginar uma **anexação uniforme**. Nesse caso, cada nó inserido sempre terá um grau de  $m$ . Ou seja, a **probabilidade** de um link do meu novo nó inserido se interligar ao vértice  $v_i$  é igual a  $m/i$  (A chance de ele se ligar com uma das arestas é  $1/i$ , logo, como eu posso me ligar com  $m$  arestas diferentes, todas independentes entre si, a probabilidade total vai ser  $m/i$ ), logo:

$$\delta(v_j, t = i) := \text{Grau de } v_j \text{ no momento } i \quad (49)$$

Com isso, podemos interpretar esse grau como uma **variável aleatória**. Temos que o grau de  $v_i$  no momento inicial  $i$  é fixa como  $m$ . Então a quantidade de arestas no momento  $i + 1$  pode ser escrita como:

$$\delta(v_i, i + 1) = m + \mathbb{I}_{i+1}(1) + \mathbb{I}_{i+1}(2) + \dots + \mathbb{I}_{i+1}(m) \quad (50)$$

Onde  $\mathbb{I}_j(k)$  é a variável indicadora que diz se, no momento  $j$ , a aresta  $k$  do **novo nó que está sendo adicionado na rede** foi adicionado ou não no nosso nó. Podemos reescrever como a soma de uma única variável aleatória de distribuição binomial também. Você pode ter reparado que eu utilizei  $i$  tanto no  $v_i$  quanto no  $i$ . Vou utilizar isso pois eu estou supondo que, na nossa análise, estamos saindo do último nó adicionado (Uma aproximação razoável do modelo real, obviamente que nem todos os nós vão ser adicionados com essa anexação, já que antes de eu iniciar essa abordagem, já vai ter uma rede “preexistente”)

Porém, queremos ter uma **noção** de como isso vai ser ao longo prazo, podemos então tirar a esperança disso.

$$\mathbb{E}[\delta(v_i, i + 1)] = m + \frac{m}{i} \quad (51)$$

Porém, isso é apenas para um único passo, queremos generalizar para vários passos. Vamos supor então que estamos saindo do  $i$ -ésimo nó adicionado e estamos no momento  $t$ :

$$\delta(v_i, t) = m + \sum_{k=i+1}^t \sum_{j=1}^m \mathbb{I}_k(j) \quad (52)$$

$$\begin{aligned} \mathbb{E}[\delta(v_i, t)] &= m + \sum_{k=i+1}^t \sum_{j=1}^m \mathbb{E}[\mathbb{I}_k(j)] \\ &= m + \sum_{k=i+1}^t \sum_{j=1}^m \frac{j}{k-1} \end{aligned} \quad (53)$$

$$\begin{aligned}
\mathbb{E}[\delta(v_i, t)] &= m \cdot \sum_{k=i+1}^t \frac{1}{k-1} \\
\mathbb{E}[\delta(v_i, t)] &\approx m + m \ln\left(\frac{t-1}{i-1}\right) \\
\frac{\mathbb{E}[\delta(v_i, t)]}{m} - 1 &\approx \ln\left(\frac{t-1}{i-1}\right) \\
\exp\left(\frac{\mathbb{E}[\delta(v_i, t)] - m}{m}\right) &\approx \frac{t-1}{i-1} \\
\exp\left(-\frac{\mathbb{E}[\delta(v_i, t)] - m}{m}\right) &\approx \frac{i-1}{t-1} \\
\exp\left(-\frac{\mathbb{E}[\delta(v_i, t)] - m}{m}\right) &\approx \frac{i}{t}
\end{aligned} \tag{54}$$

No intervalo  $[0, t]$ , temos a seguinte estruturação:



Figura 5: Intervalo de  $i/t$

Então, o que encontramos foi a fração de nós que tem grau maior que  $v_i$ . Então temos que:

$$\mathbb{P}(\delta_t(v_i) \leq k) = 1 - e^{-\frac{k-m}{m}} \tag{55}$$

Logo, temos uma distribuição **Exponencial**

## 4.2 Anexação Preferencial

Na anexação **uniforme**, cada novo nó podia se ligar com um dos nós anteriores com mesma probabilidade. Nessa abordagem, os nós de **maior grau** terão uma **maior preferência** para serem escolhidos (Não é uma obrigatoriedade). Podemos expressar, então da seguinte forma. Antes, vamos fazer duas definições rápidas:

**Definição 4.2.1:** Dada uma rede  $G(V, E)$ , o conjunto  $E_t$  é definido como o conjunto de arestas da rede no momento  $t$

**Definição 4.2.2:** Dado uma rede  $G(V, E)$ , a função  $\delta_t : V \rightarrow \mathbb{N}$  é a função que retorna o grau de um vértice em um momento  $t$  do tempo

Voltando, queremos então, antes de tudo, saber qual que é a probabilidade do nó que vai ser adicionado se ligar com um vértice  $v_i$ , então:

$$\mathbb{P}(\{v_i, v_{t+1}\} \in E_{t+1}) = \frac{\delta_t(v_i)}{\sum_{j=1}^t \delta_t(v_j)} \tag{56}$$

Queremos achar uma distribuição para os graus dos nós. Vamos tentar achar, então, uma taxa de crescimento do grau dos nós:

$$\frac{\delta_{t+1}(v_i) - \delta_t(v_i)}{\Delta t} = m \cdot \frac{\delta_t(v_i)}{\sum_{j=1}^t \delta_t(v_j)} \approx \frac{d(\delta_t(v_i))}{dt} \quad (57)$$

Porém, sabemos que  $\sum_{j=1}^t \delta_t(v_j) = 2 |E_t|$ , então vamos obter:

$$\frac{d(\delta_t(v_i))}{dt} = m \cdot \frac{\delta_t(v_i)}{2 |E_t|} = m \cdot \frac{\delta_t(v_i)}{2tm} = \frac{\delta_t(v_i)}{2t} \quad (58)$$

Logo, obtemos uma EDO para resolver. Vamos chamar  $\delta_t(v_i)$  de  $k$  apenas para facilitar a visualização:

$$\begin{aligned} \frac{dk}{dt} &= \frac{k}{2t} \quad (\delta_t(v_i) = m) \\ \frac{dk}{k} &= \frac{dt}{2t} \\ \int \frac{dk}{k} &= \int \frac{dt}{2t} \\ \ln k &= \frac{1}{2} \ln t + C \\ \delta_t(v_i) &= t^{1/2} \cdot D \end{aligned} \quad (59)$$

Resolvendo para o caso  $\delta_i(i) = m$ , temos:

$$\begin{aligned} m &= i^{1/2} D \Rightarrow D = m \cdot i^{-1/2} \\ \Rightarrow \delta_t(v_i) &= m \left( \frac{t}{i} \right)^{1/2} \end{aligned} \quad (60)$$

Queremos então calcular  $\mathbb{P}(\delta_t(v_i) \leq k)$ . Na média, todos os nós **posteriores** ao nó  $v_i$  tem grau menor do que  $k$ , então precisamos apenas inverter aquela equação de antes, assim, vamos obter:

$$\begin{aligned} \delta_t(v_i)^2 &= m^2 \frac{t}{i} \\ \Leftrightarrow i &= \frac{m^2 \cdot t}{\delta_t(v_i)^2} \end{aligned} \quad (61)$$

Assim, conseguimos obter a **fração de nós com grau maior que  $v_i$** , que são justamente os nós anteriores a ele ( $i/t$ ) que é  $m^2 \delta_t(v_1)^{-2}$ . Temos então que:

$$\mathbb{P}(\delta_t(v_i) \leq k) = 1 - m^2 k^{-2} \quad (62)$$

Temos também que a densidade vai ser:

$$f_K(k) = 2m^2 k^{-3} \quad (63)$$

Percebemos, então, que a variável aleatória  $K$ , que representa o grau de um nó na rede, tem a distribuição **Pareto(2,  $m$ )**