

FGV EMap
João Pedro Jerônimo

Ciência de Redes

Revisão para A2

Rio de Janeiro
2025

Conteúdo

- 1 Correlação de Graus 3
 - 1.1 Descobrimdo os Critérios 4
 - 1.2 Average Next Neighbour Degree 6
 - 1.3 Cutoffs 8
- 2 Percolação e Robustez 10
 - 2.1 Percolação 11
 - 2.2 Robustez 12
 - 2.3 Critério de Molloy-Reed 14
 - 2.4 Limite Crítico 14
 - 2.5 Tolerância a Ataques 17
 - 2.6 Melhorando a Robustez 19

Correlação de Graus

Nesse capítulo, temos o objetivo de entender como nós se relacionam dentro de uma rede. Imagine uma rede que os nós são pessoas e as arestas são direcionadas e apontam para nós que aquela pessoa conhece. Quando paramos para reparar, nós com muitos graus (Que muitas pessoas conhecem, logo, celebridades), tendem a apontar para outros nós com muitos graus de entrada (Celebridades conhecem muitas celebridades). Porém, muitas vezes esses critérios podem nem estar explícitos na rede em si, e queremos estudar todos os casos possíveis nesse capítulo

1.1 Descobrimos os Critérios

Dado uma rede $G(V, E)$, vamos supor que conhecemos a distribuição dos graus dessa rede ($\mathbb{P}(\delta(v_i) = k) = p(k)$ para simplificação de notação), queremos ver se o grau do nó tem alguma influência em como eles se ligam. Então temos que ver a probabilidade de que um nó de grau k' se ligue com um de grau k , ou seja:

$$\mathbb{P}(\{v_i, v_j\} \in E \mid \delta(v_i) = k, \delta(v_j) = k') = p(\{k, k'\} \in E) \quad (1)$$

Vamos escrever dessa forma para simplificar notação. Vamos ter que:

$$p(\{k, k'\} \in E) = \frac{kk'}{2|E|} \quad (2)$$

Onde $k'/2L$ é a probabilidade do meu nó com grau k se ligar com um nó de grau k' e eu multiplico por k pois meu nó possui k arestas ligadas a ele. Essa probabilidade indica que nós com graus maiores tem mais probabilidade de se ligar com outros nós de grau grande

Definição 1.1.1 (Matriz de Correlação por Grau): A matriz de correlação por grau é a matriz que a entrada e_{ij} é a probabilidade de encontrar nós de graus i e j nas pontas de uma aresta selecionada aleatoriamente

Essa matriz nos traz informações diretas da relação **linear** entre os nós. Podemos dividir e classificar as redes de 3 formas de acordo com a relação entre seus graus

- **Rede Assortativa:** Nós com graus parecidos intelrigam entre si
- **Redes Neutras:** Os nós não apresentam um padrão para intelrigarem entre si
- **Redes Desassortativas:** Nós com graus maiores se ligam com graus menores e vice-versa

Podemos então avaliar a matriz de covariância ou a correlação entre os dados. Porém a relação entre esses pontos pode não ser linear, o que pode dar falsas impressões sobre a correlação

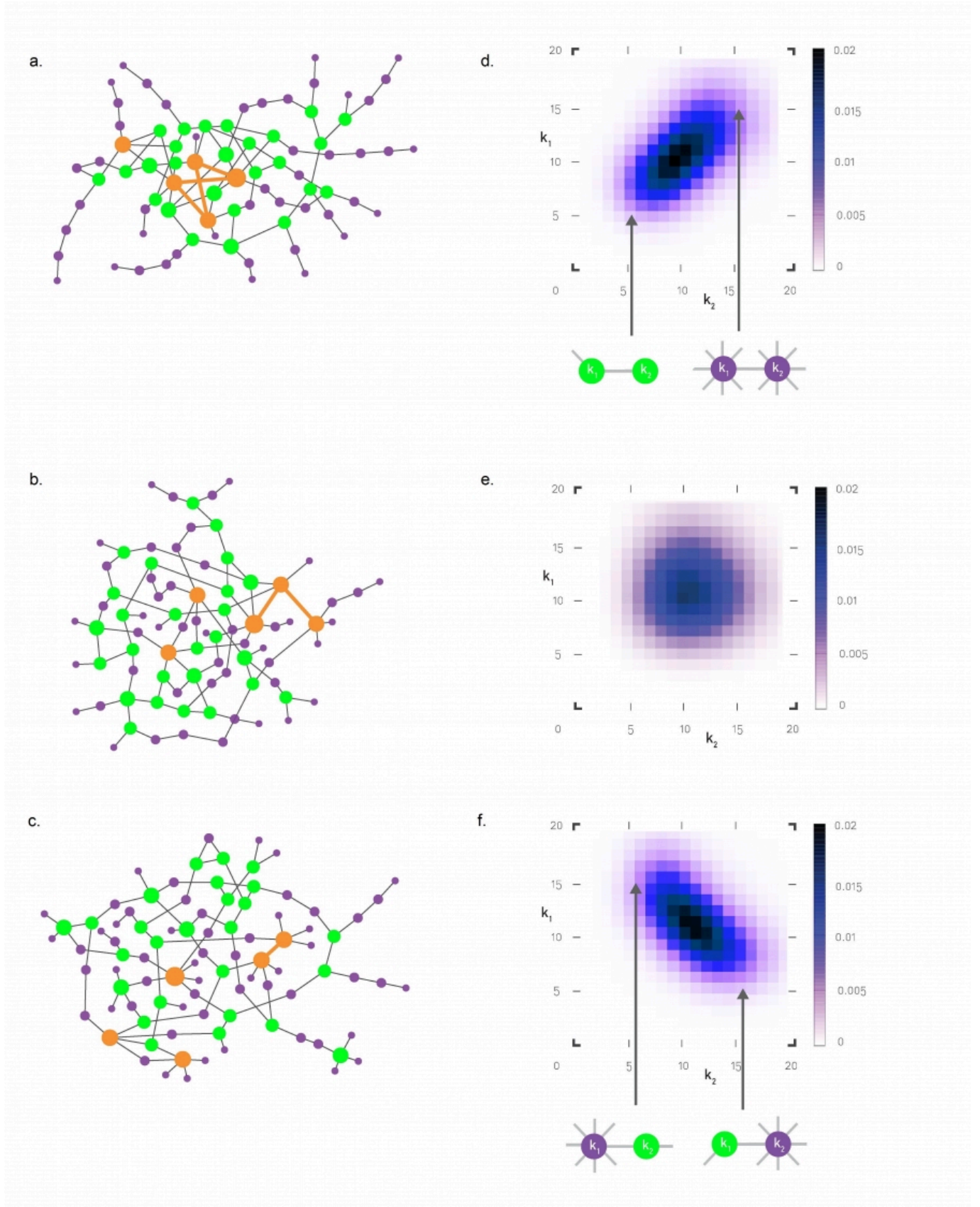


Figura 1: Matriz de covariância de graus com 3 redes. A primeira é assortativa, a segunda é neutra e a última é desassortativa

A matriz de correlação acaba por representar uma distribuição conjunta em i e j de forma que:

$$\sum_{i,j} e_{ij} = 1 \quad (3)$$

então definindo um pouco melhor essa probabilidade, defina K como a bariável aleatória que representa o grau de um nó selecionado aleatoriamente. Seja também $\{v_i, v_j\} \in E$ uma aresta escolhida aleatoriamente dentre todas as do conjunto E . Para simplificar notações, também definimos $A = \delta(v_i)$ e $B = \delta(v_j)$. Então vamos ter que:

$$e_{ij} = \mathbb{P}(A = i, B = j) = \mathbb{P}(A = i|B = j) \cdot \mathbb{P}(B = j) = \mathbb{P}(B = j|A = i) \cdot \mathbb{P}(A = i) \quad (4)$$

vamos então calcular cada um desses termos (Até onde for possível). Primeiro vamos calcular $\mathbb{P}(A = i)$

$$\mathbb{P}(A = i) = \frac{iN_i}{2|E|} \times \frac{1}{N} = \frac{ip(i)}{\mathbb{E}[K]} \quad (5)$$

onde N_i é a quantidade de nós com grau i e eu multiplico por i pois eu posso escolher qualquer uma das i arestas presentes no nó que eu gostaria que fosse escolhido (O de grau i). E divido isso pela quantidade total de nós. Eu posso ainda, dividir o termo de cima e o de baixo por N , obtendo então a expressão na direita

Se os eventos A e B **não são independentes**, seria necessário saber a distribuição entre eles dois para calcular $\mathbb{P}(A = i|B = j)$. Porém, se A e B são independentes, então $\mathbb{P}(A = i|B = j) = \mathbb{P}(A = i) \cdot \mathbb{P}(B = j)$, então, no caso de A e B independentes, obtemos:

$$\mathbb{P}(A = i|B = j) = \mathbb{P}(A = i) \cdot \mathbb{P}(B = j) = \frac{ip(i)jp(j)}{(\mathbb{E}[K])^2} \quad (6)$$

1.2 Average Next Neighbour Degree

É interessante fazer o gráfico de uma matriz de covariância, porém, por ela perder essas informações antes mencionadas, as vezes ela pode ser ineficiente ou ser de difícil leitura se houver muitos vértices, então criamos a seguinte definição:

Definição 1.2.1 (Average Next Neighbour Degree): Seja $G(V, E)$ uma rede, A a matriz de adjacência da mesma e $v_i \in V$, então:

$$k_{nn}(v_i) := \frac{1}{\delta(v_i)} \sum_{j=1}^{|V|} A_{ij} \delta(v_j) \quad (7)$$

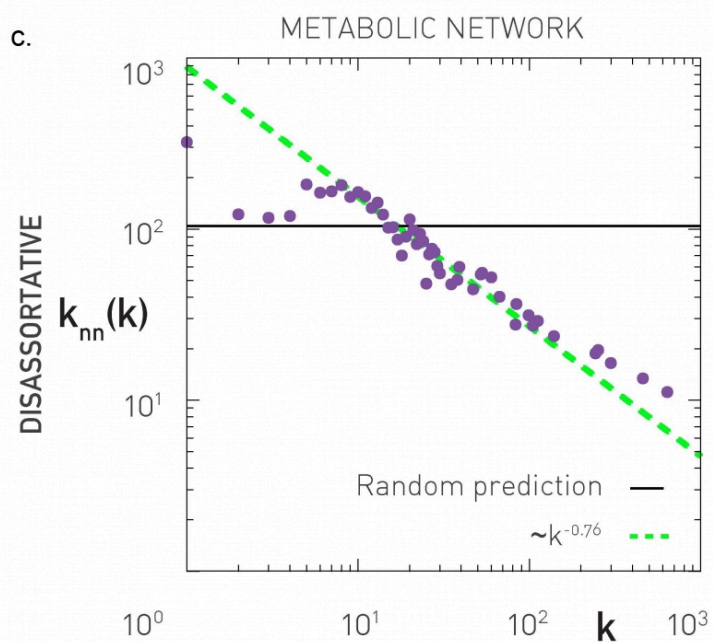
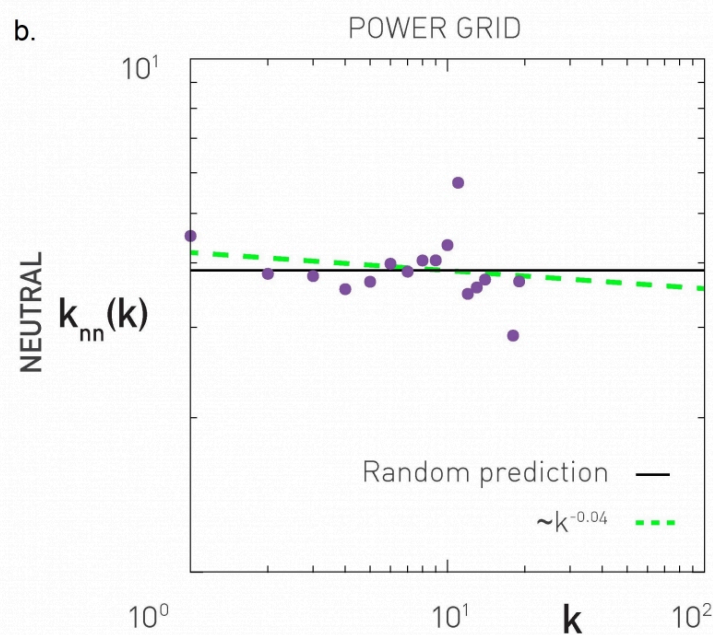
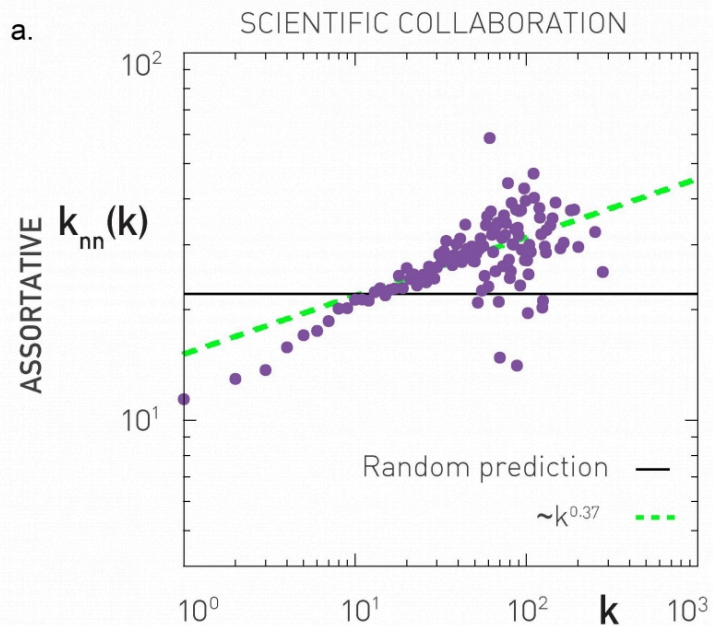
Essa definição é o k_{nn} para um vértice específico, mas e se quisermos o valor médio dessa medida para nós de grau k ?

$$k_{nn}(k) = \mathbb{E}[B|A = k] = \sum_{k'} k' \mathbb{P}(B = k'|A = k) \quad (8)$$

Podemos ver o que acontece com essa medida quando A e B são independentes, logo, analisar o caso das **redes neutras**

$$k_{nn} = \sum_{k'} k' \mathbb{P}(B = k') = \sum_{k'} k' \frac{k'p(k')}{\mathbb{E}[K]} = \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} \quad (9)$$

Perceba que, quando consideramos o caso das redes neutras, **a média dos vizinhos de um nó não depende em quantos nós ele tem ou de informações sobre como o nó se relaciona com os outros, mas única e exclusivamente de características globais da rede**. A mesma lógica vale para as redes assortativas e desassortativas:



- **Assortativas:** A medida tende a crescer conforme aumentamos os valores de k já que quanto maior o grau, maior a probabilidade de os graus se interligarem
- **Desassortativas:** Nós de grau menor tendem a interligar com nós de grau maior, o que acaba por fazer com que as probabilidades sejam altas em valores de k muito pequenos, isso faz com que conforme aumentamos o k , o valor da medida diminua

O livro aponta que, de acordo com o gráfico mostrado anteriormente, que podemos aproximar a medida do k_{nn} como:

$$k_{nn}(k) = \beta k^\mu \quad (10)$$

de forma que:

- $\mu > 0 \Rightarrow$ Rede assortativa
- $\mu \approx 0 \Rightarrow$ Rede neutra
- $\mu < 0 \Rightarrow$ Rede desassortativa

1.3 Cutoffs

A gente até agora ta assumindo que as redes analisadas são simples, ou seja, entre dois nós, só pode existir **no máximo** uma aresta. Porém, em algumas redes, quando calculamos o número esperado de arestas entre nós

$$\mathbb{E}_{kk'} = e_{kk'} \mathbb{E}[K]N \quad (11)$$

esse valor é **maior que 1**. Então como fazemos para achar o valor de k que faz com que esse valor seja > 1 ?

Primeiro, vamos definir

$$r_{kk'} = \frac{E_{kk'}}{m_{kk'}} \quad (12)$$

onde $E_{kk'}$ é o número de arestas entre todos os nós de grau k e de k' e $m_{kk'} = \min\{kN_k, k'N_{k'}, N_kN_{k'}\}$ é a maior quantidade **possível** de restas que ligam nós de grau k e k' . Vou explicar o porquê dessa fórmula para $m_{kk'}$

- Cada nó com grau k só pode se ligar a **no máximo** k nós de grau k' , então:

$$E_{kk'} \leq kN_k \quad (13)$$

- O mesmo argumento vale para os nós de grau k'

$$E_{kk'} \leq k'N_{k'} \quad (14)$$

- Eu também não consigo ligar mais que o número máximo disponível de pares entre os dois grupos de graus

$$E_{kk'} \leq N_kN_{k'} \quad (15)$$

Ou seja, temos que $r_{kk'} \leq 1 \forall k, k'$. Só que como mencionei anteriormente, por conta do valor esperado “não-esperado” de mais que uma aresta entre dois nós, pode ocorrer que $r_{kk'} > 1$, então queremos encontrar k_s tal que:

$$r_{k_s k_s} = 1 \quad (16)$$

Mas antes de achar k_s , vamos reescrever $r_{kk'}$ de outra forma e fazer uma análise. Perceba que, quando $k > Np(k')$ ou $k' > Np(k)$, os efeitos das arestas múltiplas aparecem, transformando a expressão em:

$$r_{kk'} = \frac{\mathbb{E}_{kk'}}{m_{kk'}} = \frac{\mathbb{E}[K]e_{kk'}}{Np(k)p(k')} \quad (17)$$

Para redes livre-de-escala, as condições são satisfeitas para $k, k' > (a |V|)^{\frac{1}{\gamma+1}}$ onde a depende de $p(k)$. Só que esse valor é abaixo do cutoff natural, de forma que temos CERTEZA de que, se k e k' são maiores que isso, não podemos garantir que o valor esperado dos links será menor que 1. Olhando o caso específico das redes neutras, sabemos que:

$$e_{kk'} = \frac{kk'p(k)p(k')}{\mathbb{E}[K]^2} \quad (18)$$

Logo $r_{kk'}$ vira:

$$r_{kk'} = \frac{kk'}{\mathbb{E}[K]N} \quad (19)$$

Ou seja, chegamos que $k_s \propto (\mathbb{E}[K]N)^{\frac{1}{2}}$

Tendo isso em mente, conseguimos dividir as **redes livre-de-escala** em dois regimes ao compararmos k_{\max} e k_s (Lembrando: $k_{\max} \propto N^{\frac{1}{\gamma-1}}$)

- **Sem cutoff estrutural:** Para redes aleatórias e livre-de-escala com $\gamma \geq 3$, o expoente de $k_{\max} \leq \frac{1}{2} \Rightarrow k_{\max} \leq k_s$, logo os valores esperados dos links entre nós será **sempre** menor ou igual a 1
- **Dissassortividade estrutural:** Para redes livre-de-escala com $\gamma < 3$, temos $-\frac{1}{1-\gamma} > \frac{1}{2}$, logo k_s pode ser menor que k_{\max} , logo, há nós entre k_s e k_{\max} que podem ter $E_{kk'} > 1$

Percolação e Robustez

No contexto de redes, um problema muito comum são nós pararem de funcionar e atrapalharem o fluxo da informação implícita na rede. Por exemplo, pode acontecer de a rede de internet ter falhas aleatórias dentro dela, então alguns roteadores falham. Nosso objetivo aqui é entender quando que essas falhas (Sejam propositalis ou não) afetam o fluxo da rede. A **Percolação** nos diz como e quando uma rede colapsa, enquanto a **Robustez** vai medir e indicar como prevenir esse colapso

2.1 Percolação

Em nosso contexto, vamos definir da seguinte forma. Queremos olhar a **capacidade da rede de se manter conexa após perder uma fração de seus nós**. Para começar, não vamos partir para a teoria em si, vamos analisar **um caso específico** primeiro para pegar a noção

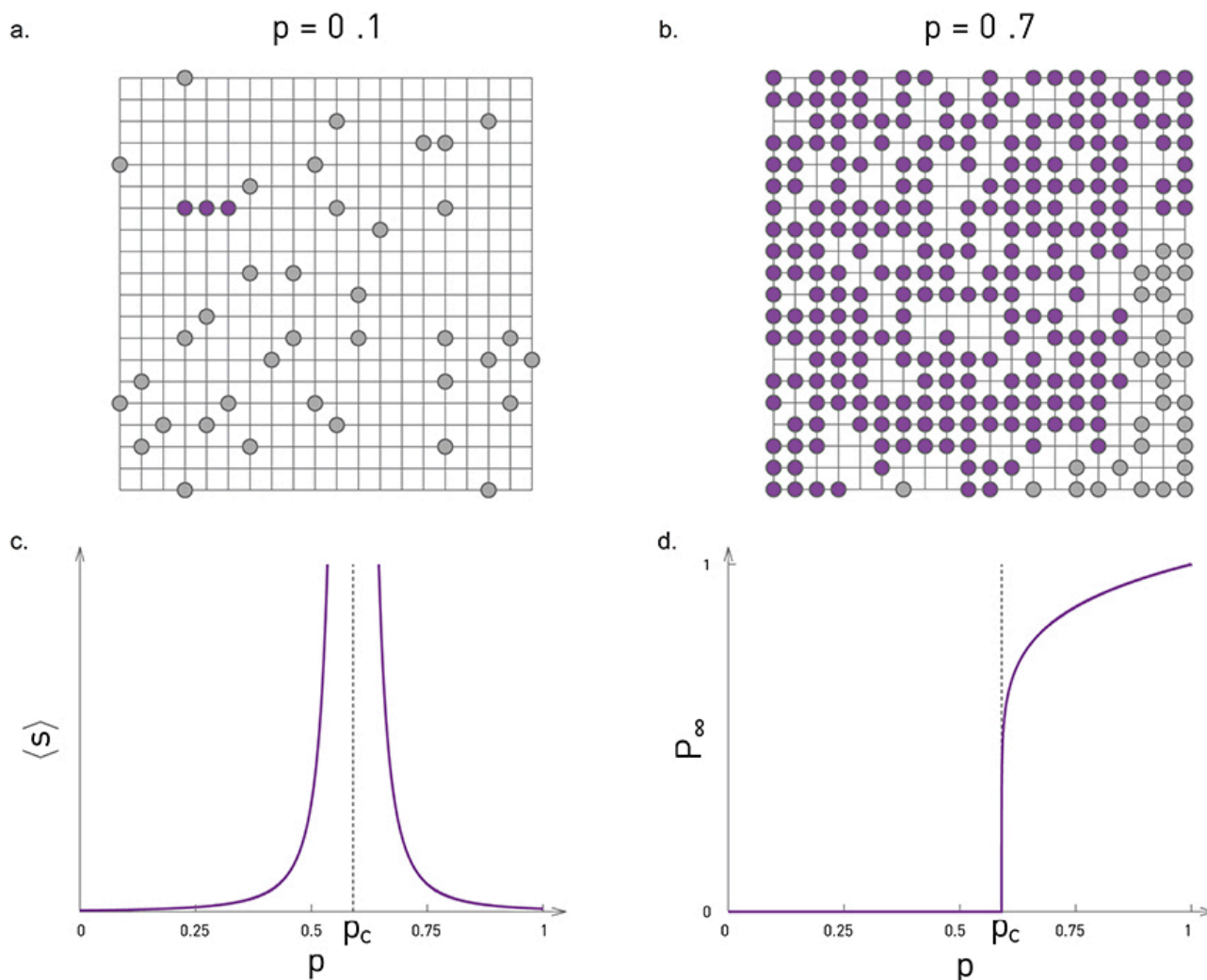


Figura 3: Rede quadrada onde cada cruzamento representa um nó (Pode ou não existir)

Vamos imaginar um grid onde cada intersecção de linhas é um nó que pode ou não existir com probabilidade p e há uma aresta entre dois nós se eles forem vizinhos. Podemos fazer então duas perguntas:

- Qual o tamanho esperado do maior cluster?
- Qual o tamanho médio dos clusters?

Olhando o gráfico presente na Figura 3, o tamanho médio dos clusters não muda gradativamente de acordo com o valor de p , mas ele se explode conforme se aproxima de um valor crítico p_c . Isso ocorre porque, conforme $p \rightarrow p_c$, os pequenos clusters se aglutinam e formam uma componente maior muito grande

Então, de acordo com o observado podemos fazer algumas definições e observações:

- Tamanho médio de clusters:

$$\mathbb{E}[S] \propto |p - p_c|^{-\gamma_p} \quad (20)$$

- Parâmetro de ordem (Probabilidade de um nó selecionado aleatoriamente pertencer ao maior cluster):

$$p_\infty \propto (p - p_c)^{\beta_p} \quad (21)$$

- Correlação de tamanho (Distância média entre dois nós do mesmo cluster):

$$\xi \propto |p - p_c|^{-v} \quad (22)$$

Perceba que, em p_c , o maior cluster tem tamanho infinito, assim ele cobre todo o quadriculado. γ_p, β_p e v são chamados de expoentes críticos e a teoria da percolação diz que eles são universais (Não dependem da natureza do grid, pode ser triangular, hexagonal, enfim)

Agora que entendemos um pouco melhor o comportamento do nosso caso específico, vamos analisar ele como uma rede em si. Imagine que vamos remover uma fração f dos nós da rede antes mencionada. Conforme aumentamos a fração f , em algum momento, a componente gigante vai se desfazer e, para algum f_c , vale que $\forall f > f_c \Rightarrow p_\infty = 0$, logo, não há mais uma componente gigante

Para redes aleatórias, **sob falhas aleatórias**, compartilham os mesmos expoentes críticos que uma rede de percolação dimensional-infinita:

$$\gamma_p = 1 \quad \beta_p = 1 \quad v = \frac{1}{2} \quad (23)$$

Já em uma rede livre-de-escala, os expoentes são:

$$\beta_p = \begin{cases} \frac{1}{3-\gamma} & \text{se } 2 < \gamma < 3 \\ \frac{1}{\gamma-3} & \text{se } 3 < \gamma < 4 \\ 1 & \text{se } 4 < \gamma \end{cases} \quad \gamma_p = \begin{cases} 1 & \text{se } \gamma > 3 \\ -1 & \text{se } 2 < \gamma < 3 \end{cases} \quad (24)$$

Perceba que no regime $2 < \gamma < 3$ **sempre há uma componente gigante**. Temos também a relação da **quantidade de componentes de tamanho s** (n_s)

$$n_s \propto s^{-\tau} e^{-s/s^*} \quad (25)$$

$$s^* \propto |p - p_c|^{-\sigma} \quad (26)$$

$$\tau = \begin{cases} \frac{5}{2} & \text{se } \gamma > 4 \\ \frac{2\gamma-3}{\gamma-2} & \text{se } 2 < \gamma < 4 \end{cases} \quad (27)$$

$$\sigma = \begin{cases} \frac{3-\gamma}{\gamma-2} & \text{se } 2 < \gamma < 3 \\ \frac{\gamma-3}{\gamma-2} & \text{se } 3 < \gamma < 4 \\ \frac{1}{2} & \text{se } \gamma > 4 \end{cases} \quad (28)$$

2.2 Robustez

Vimos um exemplo específico com redes quadriculadas, mas e se não seguirmos esse padrão? O que fazemos? Na verdade a ideia é bem parecida! Isso nos vai revelar um comportamento muito interessante sobre as redes livre-de-escala.

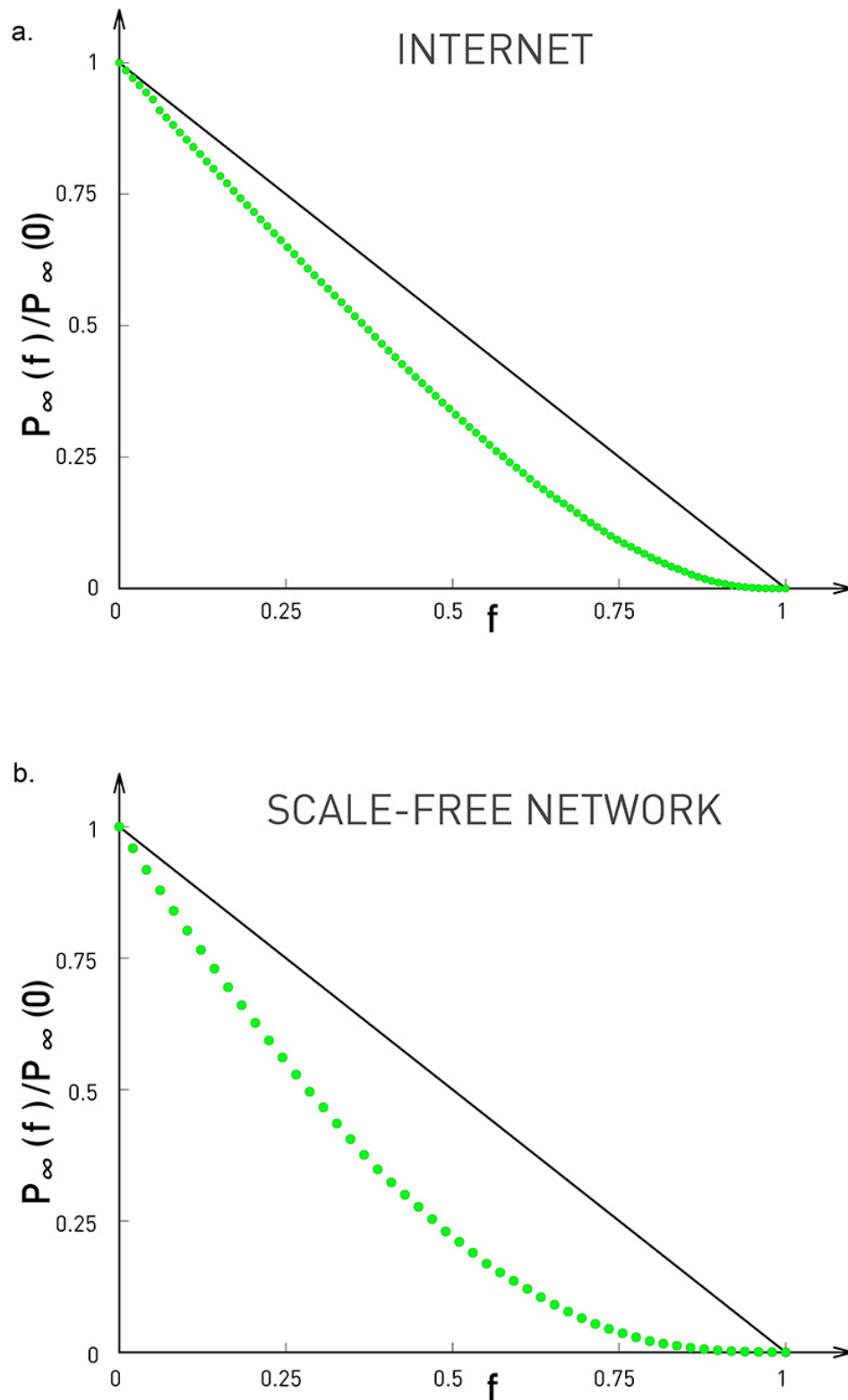


Figura 4: Comparação do tamanho relativo de uma rede livre-de-escala qualquer e da rede de internet conforme removemos uma fração **aleatória** f de seus nós

Os gráficos acima indicam uma resistência muito forte das redes livre-de-escala contra falhas aleatórias dentro da mesma. Será que podemos encontrar um meio matemático de entender o porquê que isso acontece?

2.3 Critério de Molloy-Reed

Teorema 2.3.1 (Critério de Molloy-Reed): Dado que K é a variável aleatória que representa o grau de um nó selecionado aleatoriamente dentro de uma rede $G(V, E)$. Para que uma componente gigante exista dentro dessa rede, ela deve satisfazer:

$$\frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} \geq 2 \quad (29)$$

Demonstração: Para que minha rede tenha uma componente gigante, um nó **da componente** deve ter grau médio maior ou igual a 2, já que caso contrário, quer dizer que muitos nós possuem apenas uma ou menos conexões. Defina $\mathbb{P}(\delta(v_i) = k_i | v_i \leftrightarrow v_j)$ como a probabilidade de que v_i tem grau k dado que ele se liga com j e **j está na componente gigante**. Por questões de simplificação de notação, chamemos a probabilidade antes definida como $\mathbb{P}(k_i | i \leftrightarrow j)$. Temos que:

$$\mathbb{E}[K = k_i | i \leftrightarrow j] = \sum_{k_i} k_i \mathbb{P}(k_i | i \leftrightarrow j) \geq 2 \quad (30)$$

Vamos calcular alguns termos. Sabemos que:

$$\mathbb{P}(k_i | i \leftrightarrow j) = \frac{\mathbb{P}(i \leftrightarrow j | k_i) \cdot \mathbb{P}(k_i)}{\mathbb{P}(i \leftrightarrow j)} \quad (31)$$

E também temos que:

$$\mathbb{P}(i \leftrightarrow j) = \frac{|E|}{\binom{|V|}{2}} = \frac{\mathbb{E}[K]}{|V| - 1} \quad (32)$$

Além de que:

$$\mathbb{P}(i \leftrightarrow j | k_i) = \frac{k_i}{|V| - 1} \quad (33)$$

Então, substituindo, vamos ter:

$$\mathbb{E}[K = k_i | i \leftrightarrow j] = \sum_{k_i} k_i \frac{k_i \mathbb{P}(k_i)}{\mathbb{E}[K]} = \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} \geq 2 \quad (34)$$

□

Olhando o caso específico de **redes aleatórias**, nós vamos obter que:

$$\frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} \geq 2 \Leftrightarrow \mathbb{E}[K] \frac{1 + \mathbb{E}[K]}{\mathbb{E}[K]} \geq 2 \Leftrightarrow \mathbb{E}[K] \geq 1 \quad (35)$$

O que coincide com os resultados vistos no primeiro resumo

2.4 Limite Crítico

Vamos agora utilizar do critério visto anteriormente para entender o porquê de as redes livre-de-escala serem robustas a falhas aleatórias. Primeiro de tudo, ao remover uma fração dos nós de uma rede **aleatoriamente**, existem duas consequências:

- Altera o grau de alguns nós $[k' \leq k]$

- Muda a distribuição dos graus $[p_k \rightarrow p'_{k'}]$

Vamos primeiro descobrir a nova distribuição dos graus após a remoção da fração f . Vamos fixar que estamos analisando um nó $v \in V$ que, antes da remoção, tem grau k , ou seja, tem k vizinhos. Para saber quantos vizinhos vão sobrar após remover a fração, definimos uma variável indicadora para cada um dos vizinhos do nó v

$$\mathbb{I}_j = \begin{cases} 1 & \text{se o vizinho NÃO foi removido com probabilidade } 1 - f \\ 0 & \text{se o vizinho foi removido com probabilidade } f \end{cases} \quad (36)$$

então a quantidade de vizinhos de v APÓS A REMOÇÃO dos f nós é:

$$\sum_{j=1}^k \mathbb{I}_j \quad (37)$$

e como isso é uma soma de bernoullis independentes, essa soma nos dá uma distribuição Binomial($k, 1 - f$), então temos que:

$$\mathbb{P}(K' = k' | K = k) = \binom{k}{k'} f^{k-k'} (1-f)^{k'} \quad (38)$$

Considere que K é a variável aleatória do grau de um nó selecionado aleatoriamente na rede ANTERIOR à remoção da fração f e K' é selecionando um nó na rede POSTERIOR à remoção da fração. Então para achar a nova distribuição dos graus após as remoções, apenas fazemos:

$$\begin{aligned} \mathbb{P}(K' = k') &= \sum_i^{\infty} \mathbb{P}(K' = k' | K = i) \mathbb{P}(K = i) \\ &= \sum_i^{\infty} \mathbb{P}(K = i) \binom{i}{k'} f^{i-k'} (1-f)^{k'} \end{aligned} \quad (39)$$

Agora vamos assumir que sabemos $\mathbb{E}[K]$ e $\mathbb{E}[K^2]$ (distribuição original), e queremos calcular $\mathbb{E}[K']$ e $\mathbb{E}[K'^2]$, então fazemos:

$$\mathbb{E}[K'] = \mathbb{E}[\underbrace{\mathbb{E}[K' | K]}_{\text{Bin}(k, 1-f)}] = \mathbb{E}[(1-f)K] = (1-f)\mathbb{E}[K] \quad (40)$$

$$\begin{aligned} \mathbb{E}[K'^2] &= \mathbb{V}[K'] + (\mathbb{E}[K'])^2 \\ &= \mathbb{V}[K'] + (1-f)^2 (\mathbb{E}[K])^2 \\ \mathbb{V}[K'] &= \mathbb{E}[\mathbb{V}[K' | K]] + \mathbb{V}[\mathbb{E}[K' | K]] \\ &= (1-f)f\mathbb{E}[K] + (1-f)^2\mathbb{V}[K] \end{aligned} \quad (41)$$

$$\begin{aligned} \Rightarrow \mathbb{E}[K'^2] &= (1-f)f\mathbb{E}[K] + (1-f)^2(\mathbb{V}[K] + \mathbb{E}[K]^2) \\ &= (1-f)f\mathbb{E}[K] + (1-f)^2\mathbb{E}[K^2] \end{aligned} \quad (42)$$

Agora que sabemos os momentos da distribuição após a remoção dos f , podemos aplicar o critério de Molloy-Reed na rede posterior à remoção

$$\begin{aligned}
\frac{\mathbb{E}[K'^2]}{\mathbb{E}[K']} = 2 &\Leftrightarrow \frac{(1-f)f\mathbb{E}[K] + (1-f)^2\mathbb{E}[K^2]}{(1-f)\mathbb{E}[K]} = 2 \\
f + \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} - f \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} &= 2 \\
f \left(1 - \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} \right) &= 2 - \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} \\
f &= 1 - \frac{1}{\frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} - 1}
\end{aligned} \tag{43}$$

Note que a fração limite depende única e exclusivamente das informações da distribuição. Se olharmos o caso **específico** das redes aleatórias:

$$f_c = 1 - \frac{1}{\mathbb{E}[K]} \tag{44}$$

Ou seja, quanto mais densa a rede, maior a fração crítica de nós necessários para a remoção. Agora, para redes livre-de-escala, vamos fazer um passo-a-passo diferente. Vamos primeiro calcular o m -ésimo momento do grau de uma rede livre-de-escala:

$$\begin{aligned}
\mathbb{E}[K^m] &= (\gamma - 1) k_{\min}^{\gamma-1} \int_{k_{\min}}^{k_{\max}} k^{m-\gamma} dk \\
&= \frac{\gamma - 1}{m - \gamma + 1} k_{\min}^{\gamma-1} [k^{m-\gamma+1}]_{k_{\min}}^{k_{\max}} \\
&= \frac{\gamma - 1}{m - \gamma + 1} k_{\min}^{\gamma-1} [k_{\max}^{m-\gamma+1} - k_{\min}^{m-\gamma+1}]
\end{aligned} \tag{45}$$

Agora calculamos o limite crítico f_c

$$\kappa = \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} = \frac{(2 - \gamma) k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{(3 - \gamma) k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}} \tag{46}$$

então, obtemos:

$$\kappa = \left| \frac{2-\gamma}{3-\gamma} \right| \begin{cases} k_{\min} & \text{se } \gamma > 3 \\ k_{\max}^{3-\gamma} k_{\min}^{\gamma-2} & \text{se } 2 < \gamma < 3 \\ k_{\max} & \text{se } 1 < \gamma < 2 \end{cases} \tag{47}$$

daí, utilizando todas as contas que vimos agora relebrando o fato, visto no último resumo, que:

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}} \tag{48}$$

vamos obter que:

$$\begin{aligned}
f_c &= 1 - \frac{1}{\kappa - 1} \\
&= 1 - \frac{C}{N^{\frac{3-\gamma}{\gamma-1}}}
\end{aligned} \tag{49}$$

2.5 Tolerância a Ataques

Até agora, vimos apenas falhas aleatórias, na rede de internet, por exemplo, se alguns roteadores falharem aleatoriamente, a rede tem estrutura suficiente para se manter, porém, e se um ataque planejado for feito e derrubar todos os pontos com maiores conexões? Esse é o contexto que vamos analisar, em vez de retirarmos nós aleatoriamente, vamos retirar sempre os nós com **maior grau**

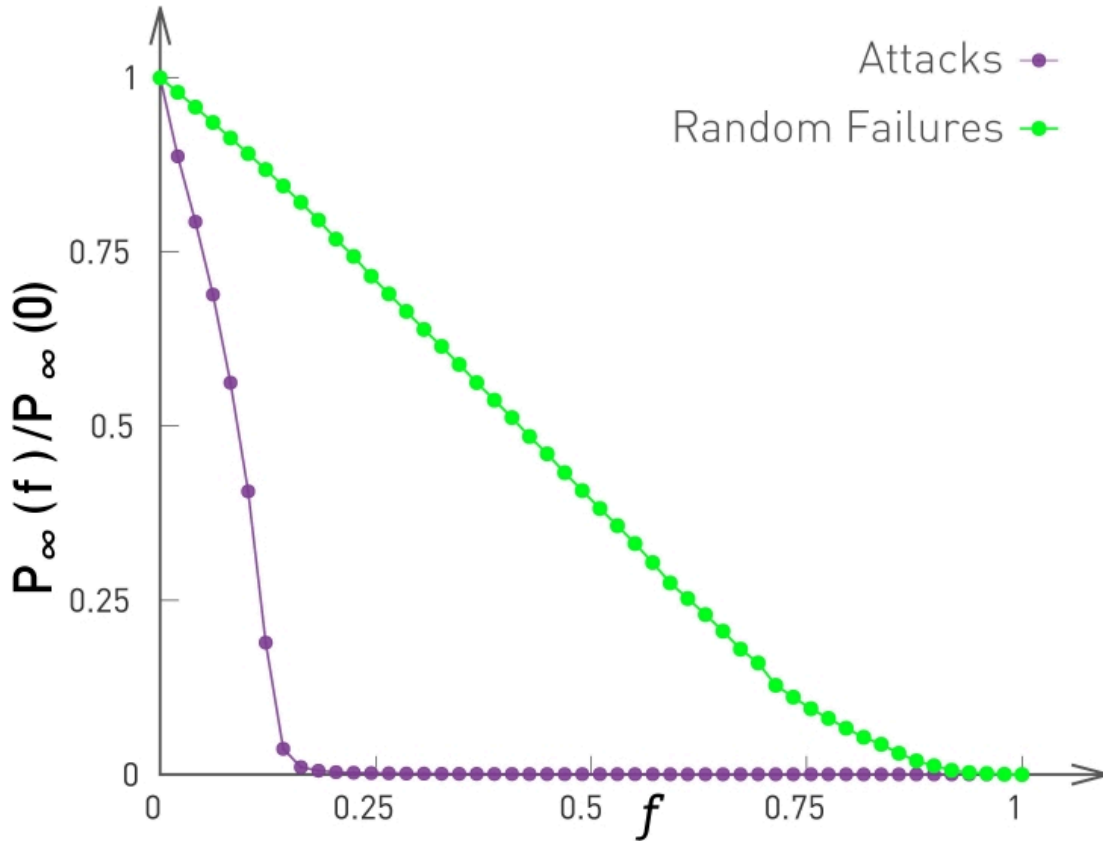


Figura 5: Tamanho relativo do maior cluster conforme retiramos os nós seguindo o regime de falhas aleatórias e de ataques em uma rede **livre-de-escala**

Perceba que, para redes livre-de-escala, elas são **muito** mais suscetíveis a ataques direcionados. O que faz bastante sentido, já que a presença de hubs é muito marcante nesse tipo de rede.

Agora a gente gostaria de entender como que a remoção dos hubs afeta a rede e o limite para que ela colapse. Para entender isso, devemos saber que remover um hub afeta a rede de duas formas:

- Muda o grau máximo de k_{\max} para k'_{\max}
- A distribuição de graus muda de p_k para p'_k

Lembrando que, para o problema das redes livre-de-escala, temos:

$$\begin{aligned}
 p_k &= ck^{-\gamma} \\
 k &\in \{k_{\min}, \dots, k_{\max}\} \\
 c &\approx \frac{\gamma - 1}{k_{\min}^{-\gamma+1} - k_{\max}^{-\gamma+1}}
 \end{aligned} \tag{50}$$

Após removermos a fração f de nós, temos que:

$$\begin{aligned}
 f &= \int_{k'_{\max}}^{k_{\max}} p_k dk \\
 &= (\gamma - 1) k_{\min}^{\gamma-1} \int_{k'_{\max}}^{k_{\max}} k^{-\gamma} dk \\
 &= (\gamma - 1) k_{\min}^{\gamma-1} \left[\frac{k^{1-\gamma}}{1-\gamma} \right]_{k'_{\max}}^{k_{\max}} \\
 &= k_{\min}^{\gamma-1} (k_{\max}^{1-\gamma} - k_{\max}'^{1-\gamma})
 \end{aligned} \tag{51}$$

Normalmente em redes livre-de-escala, o cutoff natural é MUITO grande ($k_{\max} \gg k'_{\max}$), ou seja, o termo $k_{\max}^{1-\gamma}$ é desprezível, logo, a expressão se torna:

$$f \approx \left(\frac{k'_{\max}}{k_{\min}} \right)^{1-\gamma} \tag{52}$$

Então chegamos na relação:

$$k'_{\max} \approx k_{\min} f^{\frac{1}{1-\gamma}} \tag{53}$$

Agora vamos olhar para a segunda consequência, que é a alteração da distribuição dos graus da rede. Na ausência da correlação entre os nós, vamos assumir que os links dos hubs se ligam aleatoriamente entre si. Vamos agora tentar encontrar a fração de **links** removidos da rede:

$$\begin{aligned}
 \tilde{f} &= \frac{\int_{k'_{\max}}^{k_{\max}} k p_k dk}{\mathbb{E}[K]} \\
 &= \frac{c}{\mathbb{E}[K]} \int_{k'_{\max}}^{k_{\max}} k^{-\gamma+1} dk \\
 &= \frac{1}{\mathbb{E}[K]} \cdot \frac{1-\gamma}{2-\gamma} \cdot \frac{k_{\max}'^{-\gamma+2} - k_{\max}^{-\gamma+2}}{k_{\min}^{-\gamma+1} - k_{\max}^{-\gamma+2}}
 \end{aligned} \tag{54}$$

Como sabemos que o cutoff natural costuma ser muito maior, então podemos ignorar k_{\max} e, usando o fato que:

$$\mathbb{E}[K] \approx \frac{\gamma-1}{\gamma-2} k_{\min} \tag{55}$$

vamos obter que:

$$\tilde{f} \approx \left(\frac{k'_{\max}}{k_{\min}} \right)^{-\gamma+2} \tag{56}$$

e assim, juntando com a equação (53), vamos obter que:

$$\tilde{f} \approx f^{\frac{2-\gamma}{1-\gamma}} \tag{57}$$

O que essa fração nos diz? Conforme $\gamma \rightarrow 2$, temos que $\tilde{f} \rightarrow 1$, ou seja, em redes que $\gamma \approx 2$, remover uma pequena parte dos hubs já remove quase todas as arestas (O que é compatível com a teoria vista no último resumo).

Agora vamos encontrar a nova distribuição dos graus! Usando o mesmo raciocínio visto para as falhas aleatórias, vamos assumir que K é a variável aleatória de um nó selecionado aleatoriamente antes de remover a fração f e K' após remover a fração de vértices. Sabemos que, dado que eu selecionei um vértice que tem $K = k$, ele possui exatamente k vizinhos, e depois da remoção dos vértices, cada um dos vértices vizinhos ao que escolhi **pode** ou **não** sobreviver e não ser removidos (com probabilidade $1 - \tilde{f}$), logo, temos uma soma de k variáveis de bernoulli que representam quantos vizinhos meu nó tem após o ataque na rede:

$$\begin{aligned}\mathbb{P}(K' = k' | K = k) &= \binom{k}{k'} \tilde{f}^{k-k'} (1 - \tilde{f})^{k'} \\ \Rightarrow \mathbb{P}(K' = k') &= \sum_{k'=k_{\min}}^{k'_{\max}} \mathbb{P}(K = k) \binom{k}{k'} \tilde{f}^{k-k'} (1 - \tilde{f})^{k'}\end{aligned}\quad (58)$$

Agora que eu tenho essas informações, posso tentar achar o critério de Molloy-Reed da rede:

$$\kappa = \frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} = \frac{2 - \gamma}{3 - \gamma} k_{\min} \left(\frac{f^{\frac{3-\gamma}{1-\gamma}-1}}{f^{\frac{2-\gamma}{1-\gamma}-1}} \right) \quad (59)$$

Então conseguimos chegar no limiar crítico da fração de nós

$$f_c^{\frac{2-\gamma}{1-\gamma}} = 2 + \frac{2 - \gamma}{3 - \gamma} k_{\min} \left(f_c^{\frac{3-\gamma}{1-\gamma}-1} \right) \quad (60)$$

Perceba que, se $\gamma \rightarrow \infty$, então $f_c \rightarrow 1 - \frac{1}{k_{\min}-1}$

2.6 Melhorando a Robustez

Certo, temos uma rede, é possível melhorar a sua tolerância, tanto a ataques, quanto a falhas aleatórias? Um primeiro pensamento que opdemos ter é conectar todos os nós periféricos em um hub, além de conectar eles entre si. Porém, na vida real, isso pode não ser aplicável, tendo em vista que, se cada aresta tem um custo para ser mantida, o custo de manutenção da rede pode exceder o viável

Da para maximizar a robustez para ataques e falhas aleatórias sem alterar o custo? Queremos aumentar o limite f_c , então temos que aumentar $\mathbb{E}[K^2]$ sem alterar o custo médio $\mathbb{E}[K]$. Isso vai ocorrer em uma distribuição **bimodal** onde todo nó tem grau k_{\min} ou k_{\max} , seguindo a seguinte distribuição:

$$p(k) = (1 - r)\delta(k - k_{\min}) + r\delta(k - k_{\max}) \quad (61)$$

onde r é a fração de nós com grau k_{\max} . Então vamos querer maximizar:

$$f_c^{\text{tot}} = f_c^{\text{rand}} + f_c^{\text{targ}} \quad (62)$$

onde f_c^{rand} é o limite crítico de falhas aleatórias e f_c^{targ} o limite crítico dos ataques direcionados. Dado a distribuição bimodal citada anteriormente:

$$\begin{aligned}\mathbb{E}[K] &= (1 - r)k_{\min} + rk_{\max} \\ \mathbb{E}[K^2] &= (1 - r)k_{\min}^2 + rk_{\max}^2\end{aligned}\quad (63)$$

substituindo isso em f_c^{rand}

$$f_c^{\text{rand}} = 1 - \frac{1}{\frac{\mathbb{E}[K^2]}{\mathbb{E}[K]} - 1} = \frac{\mathbb{E}[K]^2 - 2rk_{\max}\mathbb{E}[K] - 2(1-r)\mathbb{E}[K] + rk_{\max}^2}{\mathbb{E}[K]^2 - 2rk_{\max}\mathbb{E}[K] - (1-r)\mathbb{E}[K] + rk_{\max}^2} \quad (64)$$

Agora, para achar f_c^{targ} , vamos fazer uma análise mais cuidadosa

$$\begin{aligned} f_c^{\text{targ}} > r &\Rightarrow \text{Todos os hubs foram removidos} \\ \Rightarrow f_c^{\text{targ}} &= r + \frac{1-r}{\mathbb{E}[K] - rk_{\max}} \left(\mathbb{E}[K] \frac{\mathbb{E}[K] - rk_{\max} - 2(1-r)}{\mathbb{E}[K] - rk_{\max} - (1-r)} - rk_{\max} \right) \end{aligned} \quad (65)$$

$$\begin{aligned} f_c^{\text{targ}} < r &\Rightarrow \text{Sobrou alguns hubs} \\ \Rightarrow f_c^{\text{targ}} &= \frac{\mathbb{E}[K]^2 - 2r\mathbb{E}[K]k_{\max} + rk_{\max}^2 - 2(1-r)\mathbb{E}[K]}{k_{\max}(k_{\max} - 1)(1-r)} \end{aligned} \quad (66)$$

E nós estamos procurando o valor de k que maximiza f_c^{tot} . Usando as equações encontradas para f_c^{targ} e f_c^{rand} , descobrimos que podemos aproximar k_{\max} por:

$$\begin{aligned} k_{\max} &\approx Ar^{-\frac{2}{3}} \\ A &= \left[\frac{2(\mathbb{E}[K])^2(\mathbb{E}[K] - 1)^2}{2\mathbb{E}[K] - 1} \right]^{\frac{1}{3}} \end{aligned} \quad (67)$$

então obtemos que, para r pequeno:

$$f_c^{\text{tot}} = 2 - \frac{1}{\mathbb{E}[K] - 1} - \frac{3\mathbb{E}[K]}{A^2} r^{\frac{1}{3}} + O\left(r^{\frac{2}{3}}\right) \quad (68)$$

Para uma rede com N nós, o máximo de f_c^{tot} ocorre com $r = \frac{1}{N}$

$$\Rightarrow k_{\max} = AN^{\frac{2}{3}} \quad (69)$$

ou seja, em redes em que apenas 1 nó possui grau k_{\max} enquanto o resto possui k_{\min}