Introdução às Redes de Interação – MO804 (MC908)
Redes dinâmicas: principais modelos

Prof. Dr. Ruben Interian

Instituto de Computação, UNICAMP

Resumo

1. Objetivo
2. Lei de potência
3. Modelo BA

Resumo

1. Objetivo
2. Lei de potência
3. Modelo BA

Objetivo

- **Lei de potência**: origem, propriedades e importância.
- **Modelo BA**, ou de conexão preferencial: algoritmo e características.

Resumo

1. Objetivo
2. Lei de potência
3. Modelo BA

Lei de potência

- Consideremos alguma **rede real**, por exemplo, a rede de hiperlinks na **Web**, ou a rede usuários e seguidores de uma **rede social**.

Lei de potência

- Consideremos alguma **rede real**, por exemplo, a rede de hiperlinks na **Web**, ou a rede usuários e seguidores de uma **rede social**.
- O **grau** (neste caso, de entrada) de um vértice v_i é a **popularidade** de v_i.

Lei de potência

- Consideremos alguma **rede real**, por exemplo, a rede de hiperlinks na **Web**, ou a rede usuários e seguidores de uma **rede social**.
- O **grau** (neste caso, de entrada) de um vértice v_i é a **popularidade** de v_i.
- Suponha que temos uma **foto** (*snapshot*) da rede em um determinado instante.

Lei de potência

- Consideremos alguma **rede real**, por exemplo, a rede de hiperlinks na **Web**, ou a rede usuários e seguidores de uma **rede social**.
- O **grau** (neste caso, de entrada) de um vértice v_i é a **popularidade** de v_i.
- Suponha que temos uma **foto** (*snapshot*) da rede em um determinado instante.
- O que podemos esperar da quantidade de vértices (usuários, ou páginas) que possuem grau k, para cada k? Ou seja queremos saber como a **popularidade é distribuída** entre os vértices.

Lei de potência

Primeira ideia: supor que a distribuição dos graus é aproximadamente normal.

- **Lembrando**: A distribuição normal **ocorre naturalmente** em diversas áreas de ciências naturais e ciências exatas.

Lei de potência

Primeira ideia: supor que a distribuição dos graus é aproximadamente normal.

- **Lembrando**: A distribuição normal **ocorre naturalmente** em diversas áreas de ciências naturais e ciências exatas.
- Uma quantidade que é resultado da soma de muitas variáveis independentes menores, é **aproximadamente normal** ($\Rightarrow$ Teorema Central do Limite).

Lei de potência

Primeira ideia: supor que a distribuição dos graus é aproximadamente normal.

- **Lembrando**: A distribuição normal **ocorre naturalmente** em diversas áreas de ciências naturais e ciências exatas.
- Uma quantidade que é resultado da soma de muitas variáveis independentes menores, é **aproximadamente normal** ($\Rightarrow$ Teorema Central do Limite).
- **Consequência**: se diferentes usuários/páginas decidem com quem ter vínculos de forma independente e aleatória, então o grau/popularidade de cada página é a soma de variáveis aleatórias independentes, e **deveria ter** distribuição normal.

Lei de potência

- Fato sobre distribuições normais: a probabilidade de observar um valor que fica a x **desvios padrões** da média **decresce exponencialmente** quando x aumenta.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Ou, no caso normal padrão:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Lei de potência

- Fato sobre distribuições normais: a probabilidade de observar um valor que fica a x **desvios padrões** da média **decresce exponencialmente** quando x aumenta.

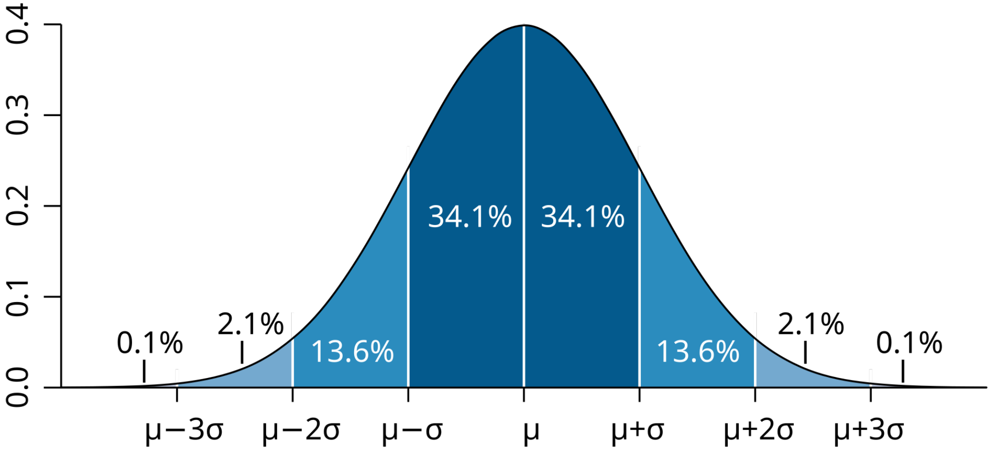
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Ou, no caso normal padrão:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

- Se acreditamos neste modelo, então o número de usuários/páginas com grau k deveria **decrescer exponencialmente** a medida de k aumenta.

Lei de potência



Lei de potência

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Distribuição das observações no caso normal:

- 68.2% estão no intervalo $+/-$ um desvio padrão da média;
- 95.4% estão no intervalo $+/-$ dois desvios padrões da média; e
- 99.7% estão no intervalo $+/-$ três desvios padrões da média.

Lei de potência

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Distribuição das observações no caso normal:

- 68.2% estão no intervalo $+/-$ um desvio padrão da média;
- 95.4% estão no intervalo $+/-$ dois desvios padrões da média; e
- 99.7% estão no intervalo $+/-$ três desvios padrões da média.

A expressão e^{-x} para $x = 100$ já é $4 \cdot 10^{-44}$. E dentro da fórmula da PDF da distribuição normal temos $e^{-x^2/2}$, muito menor.

Lei de potência

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Distribuição das observações no caso normal:

- 68.2% estão no intervalo $+/-$ um desvio padrão da média;
- 95.4% estão no intervalo $+/-$ dois desvios padrões da média; e
- 99.7% estão no intervalo $+/-$ três desvios padrões da média.

A expressão e^{-x} para $x = 100$ já é $4 \cdot 10^{-44}$. E dentro da fórmula da PDF da distribuição normal temos $e^{-x^2/2}$, muito menor.

Não devemos ter observações distando muitos desvios padrões da média, ou para isso precisaríamos de um dataset gigante!

Lei de potência

Mas, o que observamos **empiricamente, na realidade?**

Lei de potência

Mas, o que observamos **empiricamente, na realidade?** – **Exatamente o inverso! Há observações a dezenas e centenas de desvios padrões da média!**

- A fração de **páginas na WEB** com grau de entrada k é proporcional, aproximadamente, a $\frac{1}{k^2}$. Se $k = 100$, este valor é $0.0001 = \frac{1}{10000}$.

Lei de potência

Mas, o que observamos **empiricamente, na realidade?** – **Exatamente o inverso! Há observações a dezenas e centenas de desvios padrões da média!**

- A fração de **páginas na WEB** com grau de entrada k é proporcional, aproximadamente, a $\frac{1}{k^2}$. Se $k = 100$, este valor é $0.0001 = \frac{1}{10000}$.
- A parcela de **celulares** que recebem k chamadas por dia é proporcional, aproximadamente, a $\frac{1}{k^2}$.

Lei de potência

Mas, o que observamos **empiricamente, na realidade?** – **Exatamente o inverso! Há observações a dezenas e centenas de desvios padrões da média!**

- A fração de **páginas na WEB** com grau de entrada k é proporcional, aproximadamente, a $\frac{1}{k^2}$. Se $k = 100$, este valor é $0.0001 = \frac{1}{10000}$.
- A parcela de **celulares** que recebem k chamadas por dia é proporcional, aproximadamente, a $\frac{1}{k^2}$.
- A parcela de **livros comprados** por k pessoas é proporcional, aproximadamente, a $\frac{1}{k^3}$.

Lei de potência

Mas, o que observamos **empiricamente, na realidade?** – **Exatamente o inverso! Há observações a dezenas e centenas de desvios padrões da média!**

- A fração de **páginas na WEB** com grau de entrada k é proporcional, aproximadamente, a $\frac{1}{k^2}$. Se $k = 100$, este valor é $0.0001 = \frac{1}{10000}$.
- A parcela de **celulares** que recebem k chamadas por dia é proporcional, aproximadamente, a $\frac{1}{k^2}$.
- A parcela de **livros comprados** por k pessoas é proporcional, aproximadamente, a $\frac{1}{k^3}$.
- A parcela de **artigos científicos** que recebem k citações é proporcional, aproximadamente, a $\frac{1}{k^3}$.

Lei de potência

Mas, o que observamos **empiricamente, na realidade?** – **Exatamente o inverso! Há observações a dezenas e centenas de desvios padrões da média!**

- A fração de **páginas na WEB** com grau de entrada k é proporcional, aproximadamente, a $\frac{1}{k^2}$. Se $k = 100$, este valor é $0.0001 = \frac{1}{10000}$.
- A parcela de **celulares** que recebem k chamadas por dia é proporcional, aproximadamente, a $\frac{1}{k^2}$.
- A parcela de **livros comprados** por k pessoas é proporcional, aproximadamente, a $\frac{1}{k^3}$.
- A parcela de **artigos científicos** que recebem k citações é proporcional, aproximadamente, a $\frac{1}{k^3}$.

Parece que a distribuição tem uma característica conhecida como **cauda longa**: há muito mais dados ao longo da cauda, quando comparamos a uma Gaussiana.

Lei de potência

Resumo: redes reais não parecem ter distribuições dos graus dos vértices parecidas à normal. A normal funciona apenas em redes aleatórias!

As **distribuições dos graus em redes reais** são mais parecidas a $\frac{1}{k^\alpha}$.

Lei de potência

Resumo: redes reais não parecem ter distribuições dos graus dos vértices parecidas à normal. A normal funciona apenas em redes aleatórias!

As **distribuições dos graus em redes reais** são mais parecidas a $\frac{1}{k^\alpha}$.

Uma relação funcional do tipo $y = b\frac{1}{x^\alpha} = bx^{-\alpha}$ é chamada **lei de potência**, onde:

- x é a variável independente,
- y é a variável dependente (a frequência),
- b é a constante de proporcionalidade, e
- α é outra constante (o expoente).

Lei de potência

$$y = b\frac{1}{x^{\alpha}} = bx^{-\alpha}$$

Primeira propriedade importante da lei de potência:

$$\log y = \log(bx^{-\alpha}) = \log b + \log x^{-\alpha} = -\alpha \log x + \log b.$$

Lei de potência

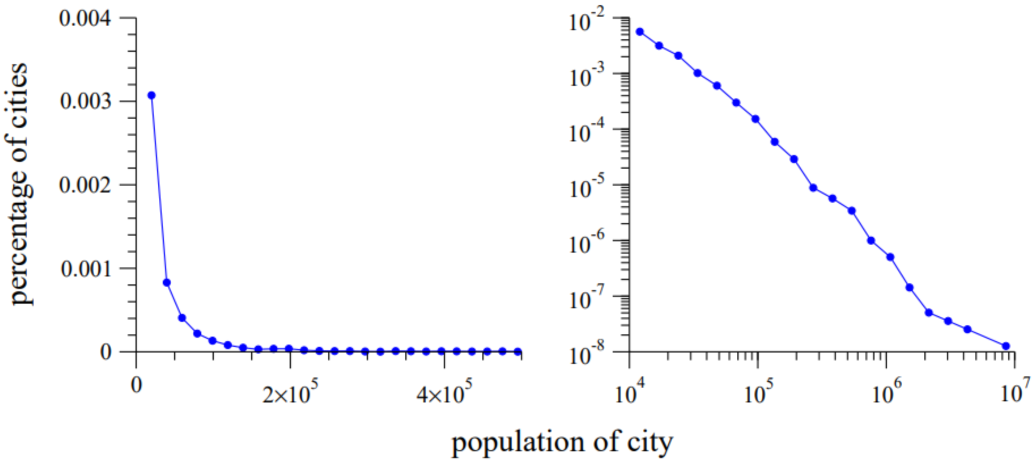
$$y = b\frac{1}{x^{\alpha}} = bx^{-\alpha}$$

Primeira propriedade importante da lei de potência:

$$\log y = \log(bx^{-\alpha}) = \log b + \log x^{-\alpha} = -\alpha \log x + \log b.$$

Essa propriedade fornece um teste relativamente simples (mas não muito rigoroso), sobre se os seus dados seguem uma **lei de potência**: se o gráfico dos valores de $\log x$ e $\log y$ é, aproximadamente, uma **reta**, temos um indicador de que a distribuição é uma **lei de potência**.

Lei de potência



Lei de potência

$$y = b\frac{1}{x^{\alpha}} = bx^{-\alpha}$$

Segunda propriedade importante da lei de potência:

$$y(cx) = b(cx)^{-\alpha} = c^{-\alpha} \cdot bx^{-\alpha} = c^{-\alpha}y(x) \; \propto \; y(x)$$

Lei de potência

$$y = b\frac{1}{x^{\alpha}} = bx^{-\alpha}$$

Segunda propriedade importante da lei de potência:

$$y(cx) = b(cx)^{-\alpha} = c^{-\alpha} \cdot bx^{-\alpha} = c^{-\alpha}y(x) \;\propto\; y(x)$$

Esta propriedade é chamada de **invariância de escala**, e significa que ao multiplicar o argumento x por um fator constante c, obtemos apenas a própria função em um outro nível de escala (**autossimilaridade**).

Lei de potência

A **lei de potência** é observada em diversos domínios, e pode ser considerada uma lei geral da natureza. Em particular, se uma quantidade x reflete algum tipo de **popularidade** (mas não apenas nesses casos!), geralmente surge a **lei de potência**.

- Pessoas "famosas" e influenciadores – número de seguidores;
- Livros – quantidade de vendas;
- Sites na Internet – links para o site;
- Artigos científicos – número de citações;
- Cidades – número de habitantes;
- Músicas em um serviço de *streaming* – número de downloads;
- Palavras em textos – frequências de uso;
- Rendimentos dos contribuintes – valor de ingressos anuais;
- Tamanho das erupções solares, força dos terremotos . . .

Lei de potência

Curiosidades: Suponha que temos a lei de potência $x^{-\alpha}$.

- O valor da média está bem definido (é finito) somente se $\alpha > 2$. Existe variância finita somente se $\alpha > 3$.

Lei de potência

Curiosidades: Suponha que temos a lei de potência $x^{-\alpha}$.

- O valor da média está bem definido (é finito) somente se $\alpha > 2$.
  Existe variância finita somente se $\alpha > 3$.
- **Interpretação**: se $\alpha \leq 3$, então ao pegar amostras finitas (por exemplo, graus de vértices em redes cada vez maiores), teremos nós com graus cada vez maiores, e a variância irá aumentar.

Lei de potência

Curiosidades: Suponha que temos a lei de potência $x^{-\alpha}$.

- O valor da média está bem definido (é finito) somente se $\alpha > 2$. Existe variância finita somente se $\alpha > 3$.
- **Interpretação**: se $\alpha \leq 3$, então ao pegar amostras finitas (por exemplo, graus de vértices em redes cada vez maiores), teremos nós com graus cada vez maiores, e a variância irá aumentar.
- Na maioria dos **casos reais**, o expoente fica exatamente no intervalo $(2, 3)$, o que significa que a média é bem definida, mas a variância não.

Lei de potência

Curiosidades: Suponha que temos a lei de potência $x^{-\alpha}$.

- O valor da média está bem definido (é finito) somente se $\alpha > 2$.
  Existe variância finita somente se $\alpha > 3$.
- **Interpretação**: se $\alpha \leq 3$, então ao pegar amostras finitas (por exemplo, graus de vértices em redes cada vez maiores), teremos nós com graus cada vez maiores, e a variância irá aumentar.
- Na maioria dos **casos reais**, o expoente fica exatamente no intervalo $(2, 3)$, o que significa que a média é bem definida, mas a variância não.
- Este fato implica na ocorrência de observações raras e difíceis de prever (chamadas "cisnes negros") que estão além do escopo das nossas expectativas "normais". Exemplos: **popularidade inesperada** de um usuário, uma noticia ou um vídeo.

Lei de potência

Um caso específico da lei de potência leva ao famoso **Princípio de Pareto**, ou **regra 80/20**:

Lei de potência

Um caso específico da lei de potência leva ao famoso **Princípio de Pareto**, ou **regra 80/20**:

- "80% dos efeitos vêm de 20% das causas": ao corrigir 20% dos bugs (mais relatados), eliminamos 80% dos erros presentes no sistema [Microsoft].

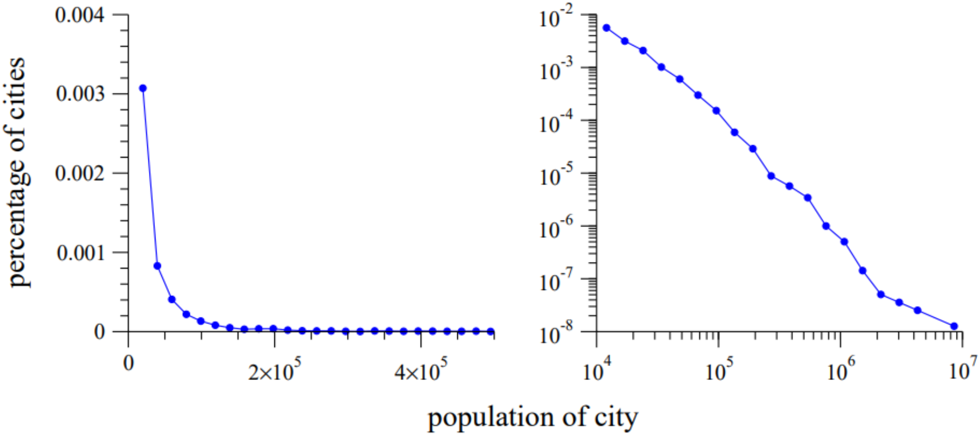
Lei de potência

Um caso específico da lei de potência leva ao famoso **Princípio de Pareto**, ou **regra 80/20**:

- "80% dos efeitos vêm de 20% das causas": ao corrigir 20% dos bugs (mais relatados), eliminamos 80% dos erros presentes no sistema [Microsoft].
- O 20% das pessoas mais ricas recebem 80% da renda.

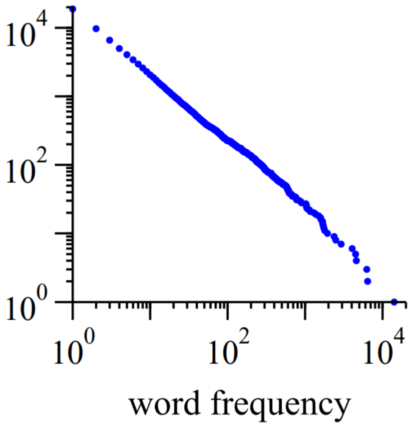
Lei de potência: Exemplos

Distribuições reais que seguem a Lei de Potencia: População das cidades



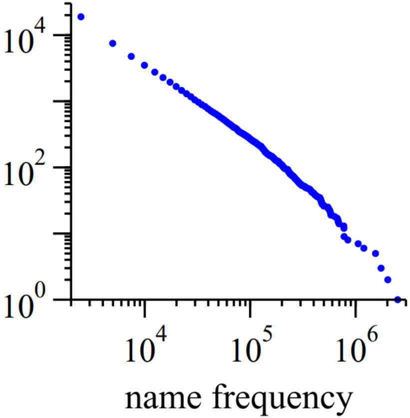
Lei de potência: Exemplos

Distribuições reais que seguem a Lei de Potencia: Palavras no texto, frequências



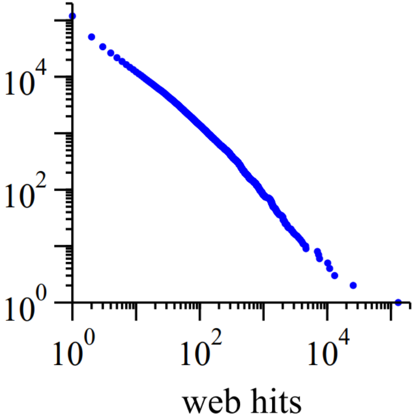
Lei de potência: Exemplos

Distribuições reais que seguem a Lei de Potencia: Sobrenomes, frequências



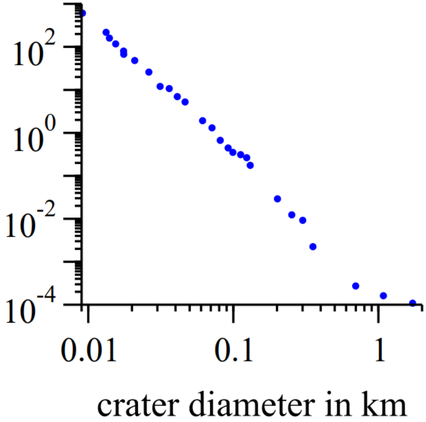
Lei de potência: Exemplos

Distribuições reais que seguem a Lei de Potencia: Acessos a sites



Lei de potência: Exemplos

Distribuições reais que seguem a Lei de Potencia: Diâmetros das crateras lunares



Resumo

1 Objetivo

2 Lei de potência

3 Modelo BA

Redes dinâmicas: modelo BA

Lembrando: O modelo de **conexão preferencial**, ou Barabási-Albert (BA), é um modelo **estocástico e populacional**.

Redes dinâmicas: modelo BA

Lembrando: O modelo de **conexão preferencial**, ou Barabási-Albert (BA), é um modelo **estocástico e populacional**.

- As **distribuições dos graus** dos vértices em muitas redes reais obedecem a **Lei de Potencia**.

Redes dinâmicas: modelo BA

Lembrando: O modelo de **conexão preferencial**, ou Barabási-Albert (BA), é um modelo **estocástico e populacional**.

- As **distribuições dos graus** dos vértices em muitas redes reais obedecem a **Lei de Potencia**.
- **Ideia da conexão preferencial**: quanto **maior** o grau ("popularidade") do nó, **maior** é a chance de ele receber novas arestas (mais rápido o grau aumentará).

Redes dinâmicas: modelo BA

Lembrando: O modelo de **conexão preferencial**, ou Barabási-Albert (BA), é um modelo **estocástico e populacional**.

- As **distribuições dos graus** dos vértices em muitas redes reais obedecem a **Lei de Potencia**.
- **Ideia da conexão preferencial**: quanto **maior** o grau ("popularidade") do nó, **maior** é a chance de ele receber novas arestas (mais rápido o grau aumentará).
- **Na prática**, isso significa que um vértice **novo** w na rede cria uma aresta com um vértice existente v_i com probabilidade proporcional ao grau $d(v_i) = k_i$.

Redes dinâmicas: modelo BA

- Se P_i é a probabilidade do evento "o vértice novo w escolheu v_i", então:

$$P_i = \frac{k_i}{\sum_j k_j}.$$

Redes dinâmicas: modelo BA

- Se P_i é a probabilidade do evento "o vértice novo w escolheu v_i", então:

$$P_i = \frac{k_i}{\sum_j k_j}.$$

- A expressão $\sum_j k_j$ é igual a $2|E|$ em grafos não direcionados (ou $|E|$ em grafos direcionados, onde consideramos os graus de entrada).

Redes dinâmicas: modelo BA

- Se P_i é a probabilidade do evento "o vértice novo w escolheu v_i", então:

$$P_i = \frac{k_i}{\sum_j k_j}.$$

- A expressão $\sum_j k_j$ é igual a $2|E|$ em grafos não direcionados (ou $|E|$ em grafos direcionados, onde consideramos os graus de entrada).
- Cada vértice adiciona uma aresta a algum vértice existente com probabilidade P_i.

Redes dinâmicas: modelo BA

- Se P_i é a probabilidade do evento "o vértice novo w escolheu v_i", então:

$$P_i = \frac{k_i}{\sum_j k_j}.$$

- A expressão $\sum_j k_j$ é igual a $2|E|$ em grafos não direcionados (ou $|E|$ em grafos direcionados, onde consideramos os graus de entrada).
- Cada vértice adiciona uma aresta a algum vértice existente com probabilidade P_i.
- Podemos adicionar não um, mas m vínculos para cada vértice. A essência do modelo e dos resultados **não muda** neste caso.

Redes dinâmicas: modelo BA

- Se P_i é a probabilidade do evento "o vértice novo w escolheu v_i", então:

$$P_i = \frac{k_i}{\sum_j k_j}.$$

- A expressão $\sum_j k_j$ é igual a $2|E|$ em grafos não direcionados (ou $|E|$ em grafos direcionados, onde consideramos os graus de entrada).
- Cada vértice adiciona uma aresta a algum vértice existente com probabilidade P_i.
- Podemos adicionar não um, mas m vínculos para cada vértice. A essência do modelo e dos resultados **não muda** neste caso.
- Existem variantes do modelo BA para grafos direcionados e não direcionados. Inicialmente veremos o modelo mais simples para o caso **não direcionado**.

Redes dinâmicas: modelo BA (simplificado)

BA (N)

1: $E \leftarrow \{(1,2)\}$
2: **para cada** $i = 3, \ldots, N$ **faça**
3: Escolher um vértice j, onde $1 \leq j \leq i-1$, com probabilidade P_j.
4: $E \leftarrow E \cup \{(i,j)\}$
devolva E

Redes dinâmicas: modelo BA (simplificado)

BA (N)

$E \leftarrow \{(1,2)\}$
para cada $i = 3, \ldots, N$ **faça**
    Escolher um vértice j, onde $1 \leq j \leq i-1$, com probabilidade P_j.
    $E \leftarrow E \cup \{(i,j)\}$
devolva E

- **Sutileza**: Iniciamos no vértice 3. Motivo: o 1^{ro} vértice não recebe arestas. Mas ao colocar o 2^{do} vértice, precisamos colocar a aresta para o primeiro!

Redes dinâmicas: modelo BA (simplificado)

BA (N)

1: $E \leftarrow \{(1,2)\}$
2: **para cada** $i = 3, \ldots, N$ **faça**
3: Escolher um vértice j, onde $1 \leq j \leq i-1$, com probabilidade P_j.
4: $E \leftarrow E \cup \{(i,j)\}$
devolva E

- **Sutileza**: Iniciamos no vértice 3. Motivo: o 1^{ro} vértice não recebe arestas. Mas ao colocar o 2^{do} vértice, precisamos colocar a aresta para o primeiro!
- Este modelo é **equivalente** a sempre começar com um caminho de tamanho 3!

Redes dinâmicas: modelo BA (simplificado)

BA (N)

1: $E \leftarrow \{(1,2)\}$
2: **para cada** $i = 3, \ldots, N$ **faça**
3: Escolher um vértice j, onde $1 \leq j \leq i-1$, com probabilidade P_j.
4: $E \leftarrow E \cup \{(i,j)\}$
devolva E

- **Sutileza**: Iniciamos no vértice 3. Motivo: o 1^{ro} vértice não recebe arestas. Mas ao colocar o 2^{do} vértice, precisamos colocar a aresta para o primeiro!
- Este modelo é **equivalente** a sempre começar com um caminho de tamanho 3!
- **Alternativa**: começar com outro grafo, e ir adicionando vértices e arestas.

Redes dinâmicas: modelo BA

Características: o modelo gera uma rede **livre de escala** cuja distribuição de graus segue uma lei de potência $bx^{-\alpha}$. Vamos **provar** este fato.

Redes dinâmicas: modelo BA

Características: o modelo gera uma rede **livre de escala** cuja distribuição de graus segue uma lei de potência $bx^{-\alpha}$. Vamos **provar** este fato.

- O comportamento dos graus dos vértices vai ser definido pela função real $k_i(t)$ que indica o valor esperado do grau no momento de tempo t (iteração i, quando temos $N(t)$ vértices).

Redes dinâmicas: modelo BA

Características: o modelo gera uma rede **livre de escala** cuja distribuição de graus segue uma lei de potência $bx^{-\alpha}$. Vamos **provar** este fato.

- O comportamento dos graus dos vértices vai ser definido pela função real $k_i(t)$ que indica o valor esperado do grau no momento de tempo t (iteração i, quando temos $N(t)$ vértices). A velocidade de crescimento (a derivada) dessa função é:

$$\frac{dk_i}{dt} = mP_i = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

- No nosso caso, $m = 1$, mas podemos estudar o modelo com qualquer número de vínculos adicionados por cada vértice.

Redes dinâmicas: modelo BA

- O número de arestas após t iterações é mt, e a soma de graus é $2mt$:

$$\frac{dk_i}{dt} = m\frac{k_i}{2mt} = \frac{k_i}{2t}.$$

Redes dinâmicas: modelo BA

- O número de arestas após t iterações é mt, e a soma de graus é $2mt$:

$$\frac{dk_i}{dt} = m\frac{k_i}{2mt} = \frac{k_i}{2t}.$$

- Reescrevemos a equação e integramos:

$$\frac{dk_i}{k_i} = \frac{dt}{2t}$$

$$\ln k_i = \frac{\ln t}{2} + C$$

Redes dinâmicas: modelo BA

$$k_i = e^{\frac{1}{2}\ln t + C} = C' t^{\frac{1}{2}}$$

Redes dinâmicas: modelo BA

$$k_i = e^{\frac{1}{2}\ln t + C} = C' t^{\frac{1}{2}}$$

- Para calcular o valor de C', usamos o fato que no momento ou iteração t_i, quando o i-ésimo vértice aparece, $k_i(t_i) = C' t_i^{\frac{1}{2}} = m$:

Redes dinâmicas: modelo BA

$$k_i = e^{\frac{1}{2}\ln t + C} = C' t^{\frac{1}{2}}$$

- Para calcular o valor de C', usamos o fato que no momento ou iteração t_i, quando o i-ésimo vértice aparece, $k_i(t_i) = C' t_i^{\frac{1}{2}} = m$:

$$C' = m / t_i^{\frac{1}{2}}$$

$$k_i = \left(\frac{t}{t_i}\right)^{\frac{1}{2}} m.$$

- Essa fórmula mostra como evolui o grau do vértice no tempo.

Redes dinâmicas: modelo BA

- Agora, precisamos analisar a **distribuição** dos valores de k_i. Primeiramente, vamos calcular a função de distribuição acumulada $F(k_i)$:

$$F(k_i) = P(k_i(t) \leq k)$$

Redes dinâmicas: modelo BA

- Agora, precisamos analisar a **distribuição** dos valores de k_i. Primeiramente, vamos calcular a função de distribuição acumulada $F(k_i)$:

$$F(k_i) = P(k_i(t) \leq k)$$

$$\left(\frac{t}{t_i}\right)^{\frac{1}{2}} m \leq k$$

$$\frac{t}{t_i} \leq \left(\frac{k}{m}\right)^2$$

$$t_i \geq t\left(\frac{m}{k}\right)^2$$

Redes dinâmicas: modelo BA

$$P\left(t_i \geq t\left(\frac{m}{k}\right)^2\right) = 1 - P\left(t_i < t\left(\frac{m}{k}\right)^2\right)$$

Redes dinâmicas: modelo BA

$$P\left(t_i \geq t\left(\frac{m}{k}\right)^2\right) = 1 - P\left(t_i < t\left(\frac{m}{k}\right)^2\right)$$

- Todos os nós são adicionados em intervalos de tempo iguais: $P(t_i) = \frac{1}{t}$.

$$1 - \frac{1}{t}t\left(\frac{m}{k}\right)^2 = 1 - \left(\frac{m}{k}\right)^2 = F(k_i)$$

Redes dinâmicas: modelo BA

$$P\left(t_i \geq t\left(\frac{m}{k}\right)^2\right) = 1 - P\left(t_i < t\left(\frac{m}{k}\right)^2\right)$$

- Todos os nós são adicionados em intervalos de tempo iguais: $P(t_i) = \frac{1}{t}$.

$$1 - \frac{1}{t}t\left(\frac{m}{k}\right)^2 = 1 - \left(\frac{m}{k}\right)^2 = F(k_i)$$

- Falta apenas encontrar a derivada da distribuição acumulada para calcular a PDF:

$$F'(k_i) = -m^2(k^{-2})' = 2m^2k^{-3} = \frac{2m^2}{k^3}.$$

Redes dinâmicas: modelo BA

Resumo:

- O expoente α da lei de potência desta variante do modelo é igual a 3 – temos lei de potência.

Redes dinâmicas: modelo BA

Resumo:

- O expoente α da lei de potência desta variante do modelo é igual a 3 – temos lei de potência.
- Porém, é um modelo limitado, pois gera redes com apenas um valor em específico de α.

Redes dinâmicas: modelo BA

Resumo:

- O expoente α da lei de potência desta variante do modelo é igual a 3 – temos lei de potência.
- Porém, é um modelo limitado, pois gera redes com apenas um valor em específico de α.

Como **generalizar** o modelo para que gere distribuições de graus com valores de α diferentes?

Redes dinâmicas: modelo BA (generalizado)

Generalização do modelo Barabási-Albert:

BA (N, p)

$E \leftarrow \{(1,2)\}$
para cada $i = 3, \ldots, N$ **faça**
    se Random$(0,1) \leq p$ **então**
        Escolher um vértice aleatório j, onde $1 \leq j \leq i-1$.
    senão
        Escolher um vértice j, onde $1 \leq j \leq i-1$, com probabilidade P_j.
    $E \leftarrow E \cup \{(i,j)\}$
devolva E

Redes dinâmicas: modelo BA (generalizado)

Generalização do modelo Barabási-Albert:

BA (N, p)

$E \leftarrow \{(1,2)\}$
para cada $i = 3, \ldots, N$ **faça**
    se Random$(0,1) \leq p$ **então**
        Escolher um vértice aleatório j, onde $1 \leq j \leq i-1$.
    senão
        Escolher um vértice j, onde $1 \leq j \leq i-1$, com probabilidade P_j.
    $E \leftarrow E \cup \{(i,j)\}$
devolva E

Este modelo foi originalmente definido e estudado para **grafos direcionados**!

Redes dinâmicas: modelo BA

Características: o modelo **generalizado** gera distribuições de graus que, *com alta probabilidade*, seguem a lei de potência com valores de α de aproximadamente $1 + \frac{1}{1-p}$.

Redes dinâmicas: modelo BA

Características: o modelo **generalizado** gera distribuições de graus que, *com alta probabilidade*, seguem a lei de potência com valores de α de aproximadamente $1 + \frac{1}{1-p}$.

Consequências:

- Se $p \approx 0$ (quase não há vínculos aleatórios), então $\alpha \approx 2$ para grafos **direcionados**. Para os não direcionados, já vimos que $\alpha = 3$ se $p = 0$.

Redes dinâmicas: modelo BA

Características: o modelo **generalizado** gera distribuições de graus que, *com alta probabilidade*, seguem a lei de potência com valores de α de aproximadamente $1 + \frac{1}{1-p}$.

Consequências:

- Se $p \approx 0$ (quase não há vínculos aleatórios), então $\alpha \approx 2$ para grafos **direcionados**. Para os não direcionados, já vimos que $\alpha = 3$ se $p = 0$.
- Se $p \approx 1$ (quase todos os vínculos são aleatórios), então α é grande (tende a ∞ quando $p \to 1$), e não há conexão preferencial. Este é o caso das redes aleatórias que já estudamos.

Redes dinâmicas: modelo BA

Características: o modelo **generalizado** gera distribuições de graus que, *com alta probabilidade*, seguem a lei de potência com valores de α de aproximadamente $1 + \frac{1}{1-p}$.

Consequências:

- Se $p \approx 0$ (quase não há vínculos aleatórios), então $\alpha \approx 2$ para grafos **direcionados**. Para os não direcionados, já vimos que $\alpha = 3$ se $p = 0$.
- Se $p \approx 1$ (quase todos os vínculos são aleatórios), então α é grande (tende a ∞ quando $p \to 1$), e não há conexão preferencial. Este é o caso das redes aleatórias que já estudamos.

De novo, podemos adicionar uma quantidade diferente de vínculos ($\neq 1$) em cada iteração. A essência do modelo e dos resultados **não muda** neste caso.

Redes dinâmicas: modelo BA

Exemplos de situações nas quais **surge a conexão preferencial** em redes:

- Pesquisadores, artigos **mais conhecidos**, recebem mais crédito (citações) do que um pesquisador/artigo comparativamente desconhecido, mesmo se o seu trabalho for similar. O crédito irá usualmente para pesquisadores que já são famosos.

Redes dinâmicas: modelo BA

Exemplos de situações nas quais **surge a conexão preferencial** em redes:

- Pesquisadores, artigos **mais conhecidos**, recebem mais crédito (citações) do que um pesquisador/artigo comparativamente desconhecido, mesmo se o seu trabalho for similar. O crédito irá usualmente para pesquisadores que já são famosos.

  Caso do César Lattes: Embora Lattes tenha sido o principal pesquisador e primeiro autor do histórico artigo da Nature descrevendo o *meson pi*, Cecil Powell foi o único agraciado com o Prêmio Nobel de Física em 1950.

Redes dinâmicas: modelo BA

Exemplos de situações nas quais **surge a conexão preferencial** em redes:

- Rede de **interação de proteínas**: os mecanismos evolutivos, como a duplicação genética e o surgimento de novas funções genéticas, levam a redes de interação de proteínas de cauda longa, embora a distribuição específica hoje seja contestada.

Redes dinâmicas: modelo BA

Exemplos de situações nas quais **surge a conexão preferencial** em redes:

- Rede de **interação de proteínas**: os mecanismos evolutivos, como a duplicação genética e o surgimento de novas funções genéticas, levam a redes de interação de proteínas de cauda longa, embora a distribuição específica hoje seja contestada.

  Essas redes podem parecer estáticas (herdamos genes e proteínas dos nossos pais), mas elas **não são**: durante milhões de anos, o número de genes cresceu de uns poucos para mais de 20.000 genes presentes em uma célula humana hoje.

Redes dinâmicas: modelo BA

Exemplos de situações nas quais **surge a conexão preferencial** em redes:

- Rede de **interação de proteínas**: os mecanismos evolutivos, como a duplicação genética e o surgimento de novas funções genéticas, levam a redes de interação de proteínas de cauda longa, embora a distribuição específica hoje seja contestada.

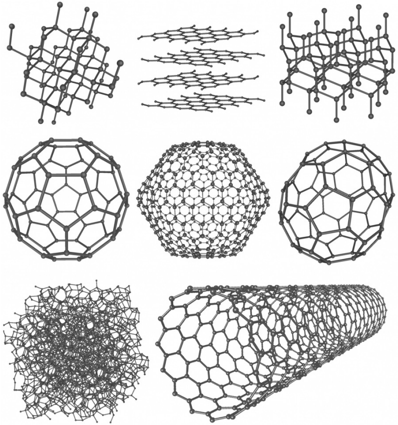
  Essas redes podem parecer estáticas (herdamos genes e proteínas dos nossos pais), mas elas **não são**: durante milhões de anos, o número de genes cresceu de uns poucos para mais de 20.000 genes presentes em uma célula humana hoje.

- **Redes sociais**: usuários escolhem de quem serão seguidores.

Redes dinâmicas: modelo BA

Limitações do modelo BA:

nem todas as redes são livres de escala.



Redes dinâmicas: modelo BA

Limitações do modelo BA:
coeficientes de clusterização.

$$C_i \xrightarrow[n \to \infty]{} 0$$

Redes dinâmicas: modelo BA

Em resumo: O modelo Barabási-Albert gera redes **livres de escala** (*scale-free*), nas quais nós cujo grau é **várias vezes maior** do que a média (**hubs**) são muito mais frequentes.

Redes dinâmicas: modelo BA

Em resumo: O modelo Barabási-Albert gera redes **livres de escala** (*scale-free*), nas quais nós cujo grau é **várias vezes maior** do que a média (**hubs**) são muito mais frequentes.

1. Os **graus** dos vértices são dissimilares – a sua distribuição segue a lei de potência.

Redes dinâmicas: modelo BA

Em resumo: O modelo Barabási-Albert gera redes **livres de escala** (*scale-free*), nas quais nós cujo grau é **várias vezes maior** do que a média (**hubs**) são muito mais frequentes.

1. Os **graus** dos vértices são dissimilares – a sua distribuição segue a lei de potência.
2. Os **coeficientes de clusterização** são baixos (a medida que n cresce, $C_i \to 0$).

Redes dinâmicas: modelo BA

Em resumo: O modelo Barabási-Albert gera redes **livres de escala** (*scale-free*), nas quais nós cujo grau é **várias vezes maior** do que a média (**hubs**) são muito mais frequentes.

1. Os **graus** dos vértices são dissimilares – a sua distribuição segue a lei de potência.
2. Os **coeficientes de clusterização** são baixos (a medida que n cresce, $C_i \to 0$).
3. Os **caminhos mais curtos** são **pequenos** (na média crescem como $\log n$).

Redes dinâmicas: modelo BA

Valores das **principais métricas** em redes reais (dados de **diferentes fontes**).

Dataset	$diam$	$\overline{dist(v,w)}$	α	$C = \overline{C_i}$
Facebook	15	4.27	2.4	0.221
YouTube	21	5.10	2.0	0.136
Emails	13	3.39	2.1	0.497

Redes dinâmicas: modelo BA

Barabási e Albert fizeram a diferença com os seus trabalhos: eles explicaram um **grande conjunto de fenômenos reais**, tanto naturais como comportamentais, e revelaram o **mecanismo** de conexão preferencial por trás deste comportamento.

Material bibliográfico

M. Newman: "The structure and function of complex networks".

M. Newman: "Power laws, Pareto distributions and Zipf's law"

Albert & Barabási: "Statistical mechanics of complex networks".

Barabási & Albert: "Emergence of scaling in random networks".

Dúvidas

Dúvidas?