

Redes Complexas

Aula 9

Roteiro

- Aplicando o $G(n,p)$
- Avaliando o modelo
- Preferential attachment
- Modelo BA
- Propriedades

Aplicando o $G(n,p)$

- Como aplicar modelo $G(n,p)$ a uma rede real?
- Determinar parâmetros do modelo de acordo com a rede real
 - parâmetros: n e p
- Ex. AS Graph, 10697 vértices e 31992 arestas



Quanto vale n e p neste caso?

Aplicando o $G(n,p)$

- Determinar p de forma a preservar grau médio da rede real

- Grau médio da rede real com n vértices e m arestas $\longrightarrow \bar{d} = \frac{2m}{n}$

- Grau médio do $G(n,p)$ $\longrightarrow \bar{d}_G = (n-1)p$

- Igualando e resolvendo para p , temos

$$p = \frac{2m}{n(n-1)}$$

Exemplos

- AS Graph, 10697 vértices e 31992 arestas
 - $d = 5.98$
 - $p = 0.00056$
- Rede de atores, 449913 vértices e 25516482 arestas
 - $d = 113.4$
 - $p = 0.00025$
- Rede metabólica, 765 vértices 3686 arestas
 - $d = 9.67$
 - $p = 0.0126$

Observação?

- p geralmente é muito pequeno (redes reais são muito esparsas)

Utilidade do Modelo $G(n,p)$

- O quão bom é o modelo $G(n,p)$ para representar redes reais?

Captura fielmente número de vértices e grau médio!

- Isso é suficiente?
- Depende do propósito do modelo

Captura outros aspectos estruturais de redes reais?

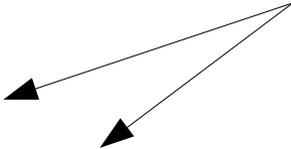
Aspectos Estruturais

- Presente em muitas redes reais
 - distâncias pequenas
 - alta clusterização
 - distribuição do grau em cauda pesada

■ Modelo $G(n,p)$

- Distâncias: $O(\log n / \log z)$
- Clusterização (média): p
- Distribuição do grau: Binomial($n-1, p$)

Muito diferentes!



Aspectos fundamentais para muitas aplicações

Exemplo

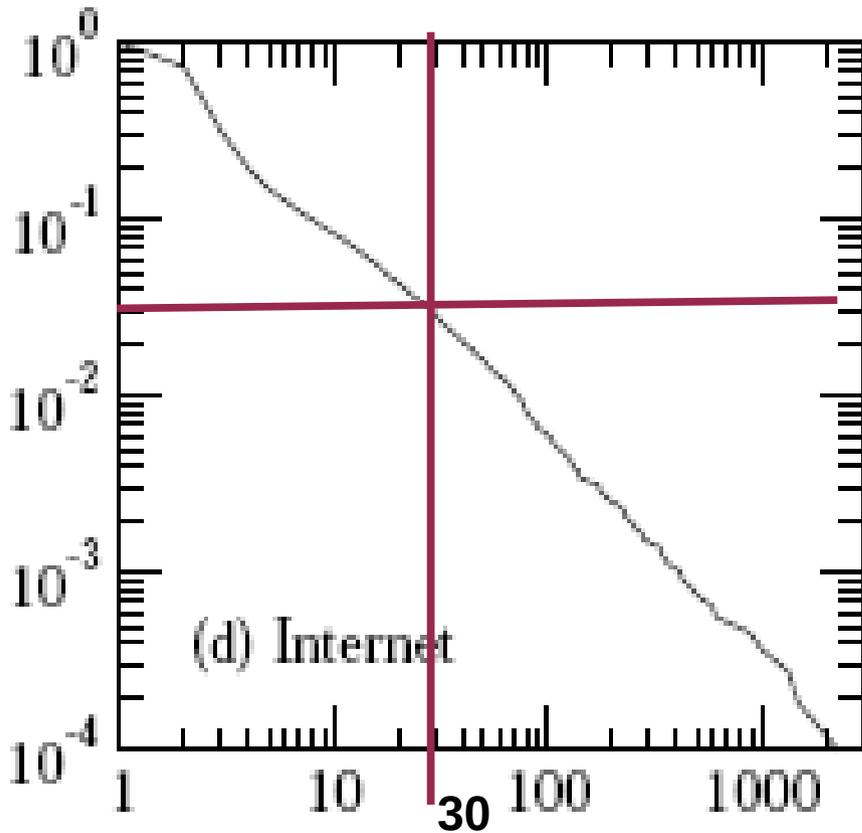
- AS Graph, 10697 vértices e 31992 arestas
- Distância média
 - rede real: 3.31
 - modelo: $O(\log n / \log z) = 5.18$
- Clusterização
 - rede real: 0.39
 - modelo: $p = 0.00056$

~1000 vezes menor!

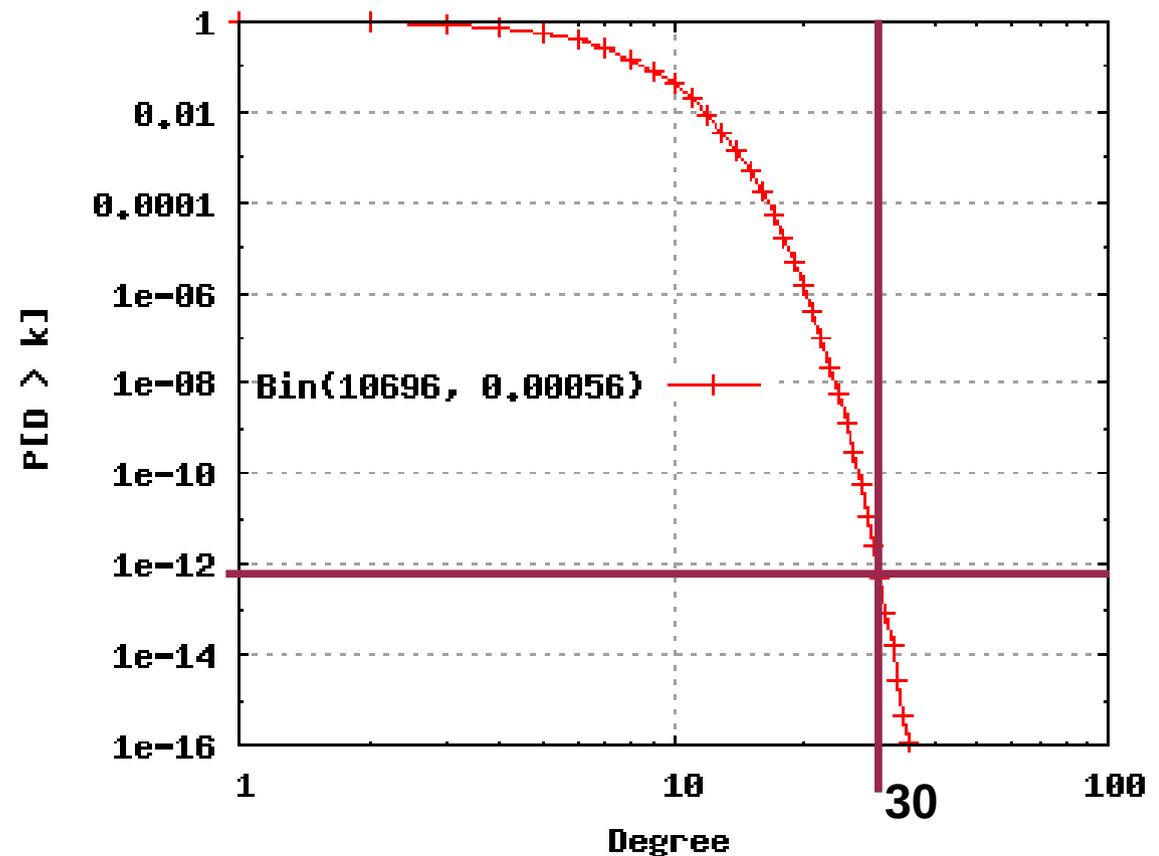
Exemplo

- AS Graph, 10697 vértices e 31992 arestas
- Distribuição do grau: $P[D > k]$

Rede real



Modelo $G(n,p)$, parametrizado



Fundamentalmente diferente!

Modelos de Rede

- **Objetivo:** modelos matemáticos que capturem aspectos importantes de redes reais
- Modelos diretos
 - Modelo define diretamente estrutura do grafo, número de vértices é fixo
 - ex. $G(n,p)$, modelo com sequência de graus
- Modelos de crescimento (*growth*)
 - baseados em algum processo incremental ou evolucionário, número de vértices cresce
 - Processo iterativo constroi a rede

Modelos de Crescimento

- Vértices e arestas são adicionados (ou removidos) incrementalmente
 - refletem o “crescimento” da rede
- Processo gerador tenta capturar a realidade
 - estrutura observada na rede é consequência do processo gerador
- Explicam o surgimento das propriedades estruturais nas redes
 - a partir de um processo gerador simples

Muitos modelos propostos são desta classe

Preferential Attachment (PA)

- Regra fundamental (e antiga) de formação
 - aka. cumulative advantage, rich-gets-richer, Matthew effect
- **Idéia:** objetos têm preferência em se *relacionar* com objetos mais populares
 - popularidade é função dos relacionamentos
 - objetos mais populares “atraem” novos objetos, aumentando sua popularidade
- Ideia que possui muitas aplicações, observada em diversos contextos

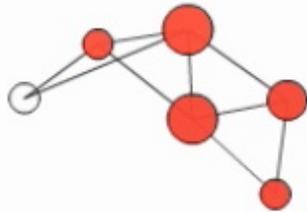
Exemplos (PA)

- Criação de links na web (Barabasi et al 99)
 - Páginas novas tendem a criar links para páginas com mais links
- Aglomeração de pessoas em cidades (Zipf 32)
 - Cidades maiores tendem a atrair mais pessoas
- Espécies de plantas de um mesmo gênero (Yule 25)
 - Mutações são mais prováveis em gênero com mais espécies, levando a outras espécies
- Número de artigos publicados (Simon 55)
 - Novos artigos tendem a ter co-autores que publicaram mais artigos

Modelo BA (Barabási-Albert)

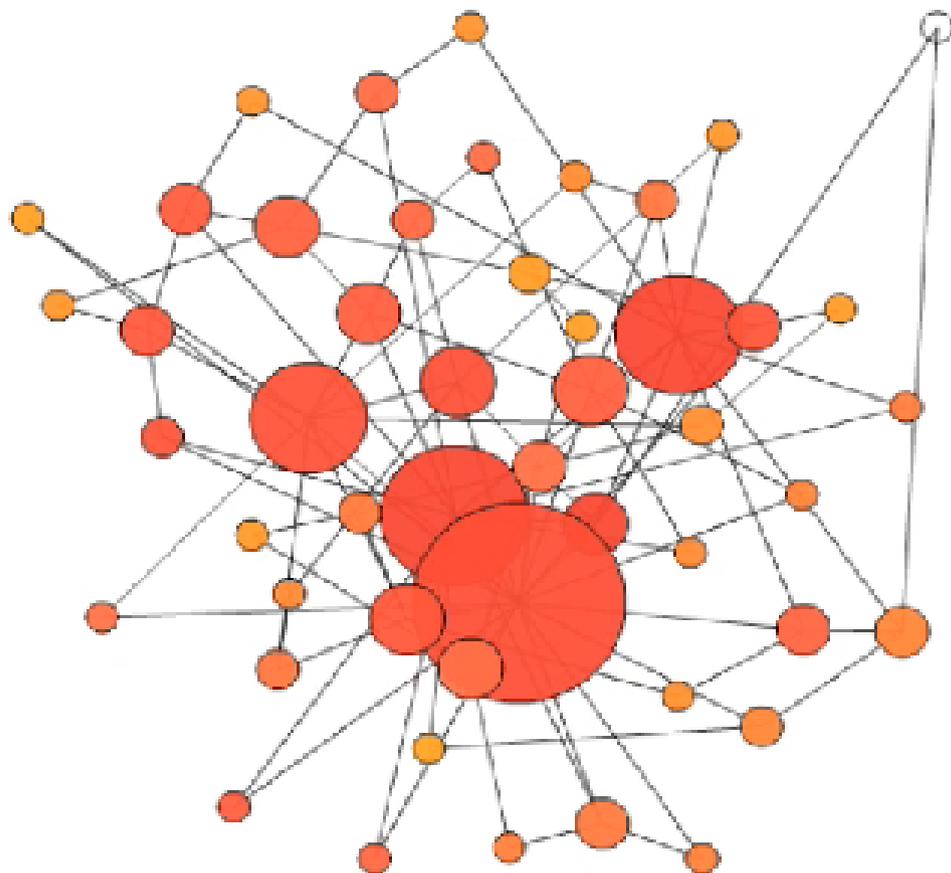
- Modelo aleatório para redes baseado em PA (Barabasi, Albert 99, +40K citações)
 - objetos são vértices e popularidade é dada pelo grau do vértice
- Processo de formação
 - inicialmente rede é um pequeno clique
 - a cada passo, adicionar 1 vértice com grau m
 - vértices adjacentes escolhidos aleatoriamente, com prob. proporcional ao seu grau
- m é parâmetro do modelo
 - clique inicial não é (muito) importante

PA Exemplo



- Clique inicial com 3 vértices, $m = 2$
- Tamanho do vértice proporcional ao grau

PA Exemplo



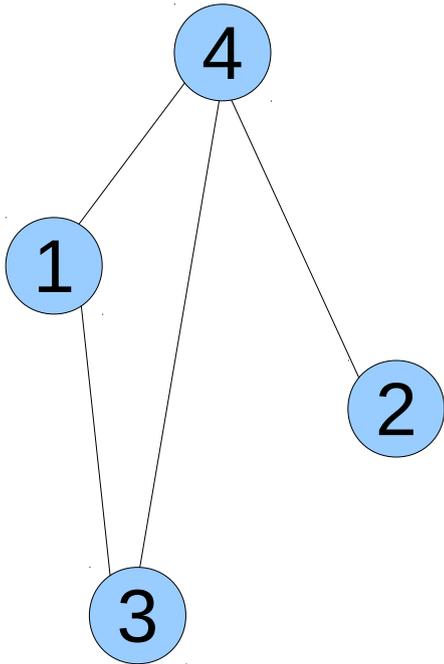
- “*Rich gets richer*” em ação!
- Muitos vértices com grau 2 (mínimo), alguns com grau muito maior

Definindo Preferências

- Dado rede $G_t = (V_t, E_t)$ no instante $t = 1, \dots$
- $d_u(t)$: grau do vértice u no instante t
- Considere a chegada de um novo vértice no instante t
 - vértice traz m arestas
- $p_u(t)$: prob. do vértice u ser incidente a uma nova aresta no instante t

$$p_u(t) = \frac{d_u(t)}{\sum_{v \in V} d_v(t)} = \frac{d_u(t)}{2mt} \leftarrow m \cdot t = \text{número de arestas adicionadas até } t$$

PA Exemplo



$$p_u(t) = \frac{d_u(t)}{\sum_{v \in V} d_v(t)}$$

$$\sum_{v \in V} d_v = 2 + 1 + 2 + 3 = 8$$

$$p_1 = 2/8 = 1/4$$

$$p_2 = 1/8$$

$$p_3 = 2/8 = 1/4$$

$$p_4 = 3/8$$

Evolução do Grau do Vértice

- Quanto vale $d_u(t)$?
 - $d_u(t)$ é uma v.a. discreta
- Iremos fazer uma aproximação
 - Assumir $d_u(t)$ é determinístico (valor esperado)
 - Logo $d_u(t)$ é contínuo
- Quanto vale variação de $d_u(t)$ no instante t ?
 - grau de u só pode crescer, taxa depende de m e sua probabilidade de receber aresta

$$\frac{dd_u}{d_t} = m p_u(t) \longrightarrow \frac{dd_u}{d_t} = \frac{d_u}{2t}$$

Evolução do Grau do Vértice

- Condições iniciais

- assumir que vértice u entra no instante t_u

- ao entrar, vértice tem m arestas: $d_u(t_u) = m$

- Resolver diferencial, com condição inicial

$$\frac{dd_u}{d_t} = \frac{d_u}{2t} \longrightarrow d_u(t) = m \left(\frac{t}{t_u} \right)^{1/2}$$

- Dependência temporal no grau

- grau cresce como \sqrt{t}

- mais antigos (menor t_u) tem grau maiores

Distribuição do Grau

- Distribuição de grau da rede no tempo t
 - distribuição depende do tempo?

$$P[d_u(t) < k]$$

- Assumir que instante de entrada é uniforme
 - vértice u entra entre $[1, t]$, t_u é v.a.
 - substituindo $d_u(t)$

$$P[d_u(t) < k] = P[t_u > m^2 t / k^2] = 1 - \frac{m^2}{k^2}$$

$m^2/k^2 < 1$, e definem uma fração do intervalo $[1, t]$

Distribuição do Grau

- Transformando a CDF em PDF
 - diferenciando a CDF, pois é uma aproximação contínua

$$P[d_u(t) = k] \approx \frac{2m^2}{k^3}$$

Lei de potência na distribuição do grau com expoente 3

Observações sobre PA

- Fenômenos observados empiricamente podem ser modelados por PA
- Modelo leva ao surgimento de *lei de potência* na popularidade dos objetos
- PA como *explicação* para leis de potência observadas empiricamente
 - lei de potência surge, pois processo de formação segue o paradigma de PA
- **Cuidado:** nem toda lei de potência é explicada por PA

Limitação do Modelo BA

- Lei de potência com expoente 3
 - muitas redes tem outros expoentes
- Vértices mais antigos tem grau maiores
 - vértices novos não tem chance de se tornarem mais populares
- Vértices criam arestas somente no instante de chegada
 - não há dinâmica entre vértices já na rede
- Rede gerada tem baixa clusterização

Todas as limitações acima superadas em modelos subsequentes!