

Redes Complexas

Aula 11

Roteiro

- Configuration Model
- Propriedades
- Stochastic Block Model (SBM)
- Propriedades

Aula passada

- Experimento de Milgram
- Mundo pequeno
- Modelo de Watts-Strogatz
- Propriedades estruturais



Configuration Model

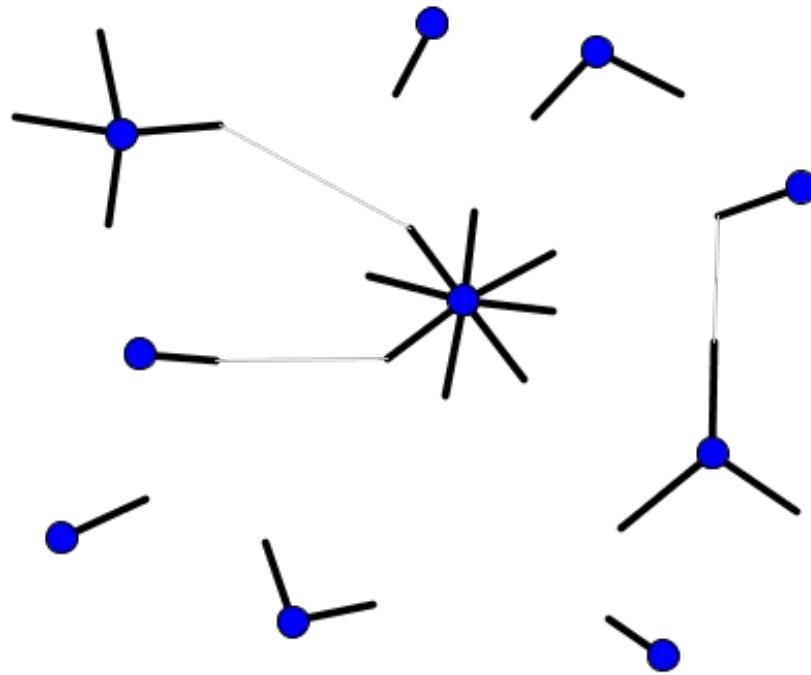
- Como construir modelo com distribuição de grau específica?

Graus como parâmetro do modelo!

- Dado n vértices e uma sequência de graus: d_1, d_2, \dots, d_n
 - soma de d_i deve ser par
- Criar cada nó com d_i pontas de arestas (*stubs*)
 - 1) Escolher duas pontas livres uniformemente
 - 2) Conectar as pontas. Voltar para (1)

Configuration Model

- Exemplo com $n=10$



- Escolher duas pontas livres para conectar

- Múltiplas arestas e loops (auto-aresta) podem acontecer

- valor esperado é constante, fração vai a zero com n crescente

Probabilidade de Aresta

- Qual probabilidade da aresta (i, j) ?
- Grau de i, j é dado por d_i, d_j

$$2m = \sum_{i=1}^n d_i \quad \text{Soma dos graus}$$

- Cada stub de i tem chance (aprox) $d_j / 2m$ de ser incidente a j
- Como temos d_i stubs em i ,

$$p_{ij} \approx \frac{d_i d_j}{2m}$$

probabilidade proporcional ao produto dos graus de i e j

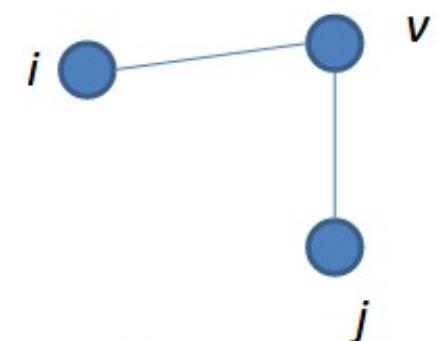
Clusterização

■ Probabilidade de restante de grau

- seguir uma aresta ao acaso
- q_k = prob. de restante de grau ser k
- p_k = fração de vértices com grau k

$$q_k = \frac{(k+1)p_{k+1}}{\bar{d}}$$

- Dado v conectado a dois vértices, i e j
- Chance de fechar o triângulo?
- Condicionar no restante de grau de i e j



$$C = \sum_{k_i, k_j=0}^{\infty} q_{k_i} q_{k_j} \frac{k_i k_j}{2m} = \frac{1}{2m} \left(\sum_{k=0}^{\infty} k q_k \right)^2 = \dots = \frac{1}{n} \frac{(\langle k \rangle^2 - \langle k \rangle)^2}{\langle k \rangle^3}$$

Obs: k_i é o restante de grau do vértice i , $\langle X \rangle = E[X]$ valor esperado

Componente Gigante

- GCC : componente conexa gigante
 - maior componente conexa tem εn vértices, para algum $\varepsilon > 0$, a.a.s.
- Condição para termos uma GCC em um grafo aleatório qualquer

$$E[d_i|i-j] > 2 \quad , \text{ para todos vértices } i, j \text{ dentro da GCC}$$

- Condição pode ser reescrita em função do primeiro e segundo momento da distribuição de grau

$$E[d_i|i-j] > 2 \quad \longleftrightarrow \quad \kappa = \frac{E[d^2]}{E[d]} > 2$$

- Dado d_1, d_2, \dots, d_n podemos determinar se teremos uma GCC comparando os dois primeiros momentos

Chung-Lu Model

- Variação do Configuration Model
 - parecida com $G(n,p)$, preferida pelos matemáticos
- Vetor de pesos w_1, w_2, \dots, w_n associado aos vértices (possivelmente reais)
- Conectar cada par de vértices com probabilidade (permitindo loops)

$$p_{ij} = \frac{w_i w_j}{\sum_k w_k}$$

- valor esperado de grau do vértice i :
$$\bar{d}_i = \sum_j p_{ij} = \sum_j \frac{w_i w_j}{\sum_k w_k} = w_i$$
- w_i é o grau médio do vértice i , grau é alteatório



Modelos com Homofilia

- Como criar modelos com padrões de conexão distintos (ex. homofilia)?

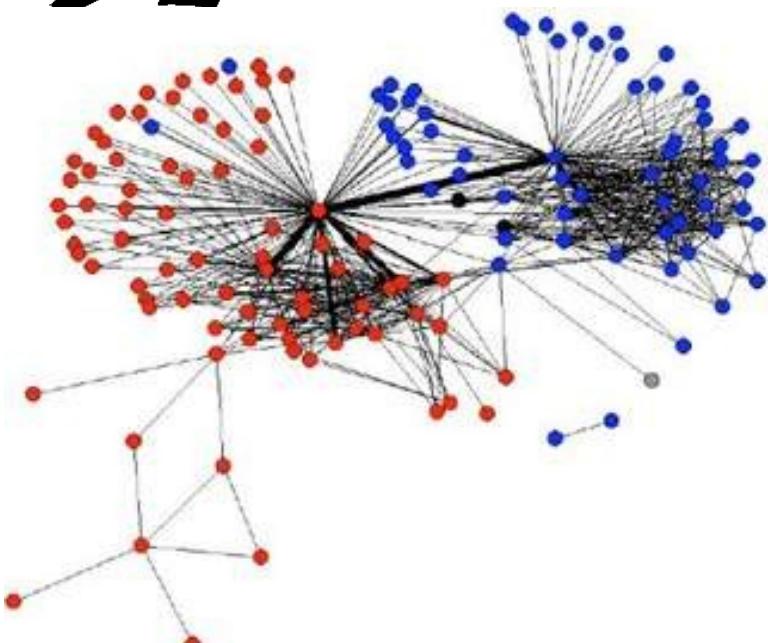
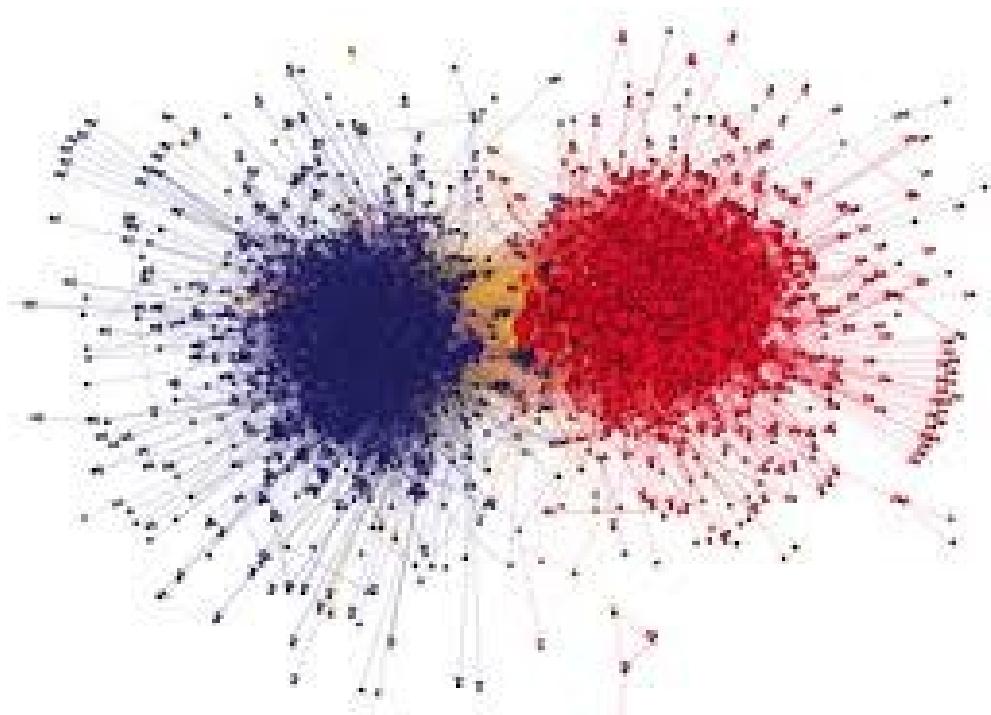


Figure 5: Strong ties in the candidate network

Políticos nos EUA (DEM x REP)



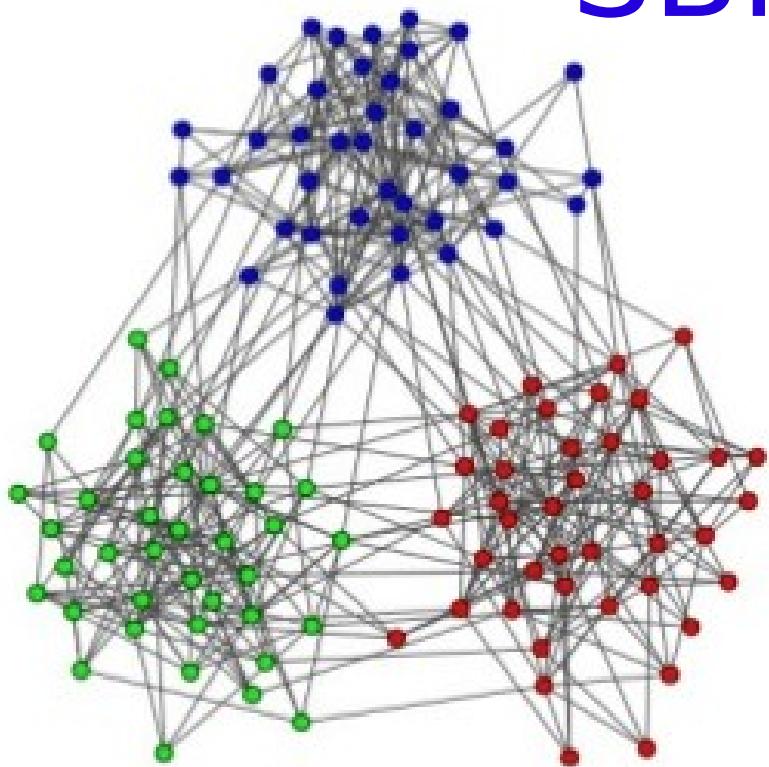
Blogs nos EUA (DEM x REP)

Embutir homofilia no modelo!

Stochastic Block Model (SBM)

- Generalização do $G(n,p)$ para grupos
- Modela homofilia entre grupos (blocos)
 - proposto em 1983, na revista *Social Networks*
- k blocos, V_i conjunto de vértices no bloco i ,
 $n_i = |V_i|$, tamanho do bloco i
- B = matriz (simétrica) de probabilidade de aresta entre vértices dos blocos
 - $b_{ij} = \text{prob. de aresta } (u,v), u \text{ em } V_i, v \text{ em } V_j$
- Em geral, b_{ii} maior que b_{ij} para $i \neq j$
 - maior chance de arestas dentro do grupo

SBM Exemplo



- 3 blocos ($k=3$)
- $n_k = 40$ para todo k

$$B = \begin{pmatrix} 0.2 & 0.01 & 0.01 \\ 0.01 & 0.2 & 0.01 \\ 0.01 & 0.01 & 0.2 \end{pmatrix}$$

- Valor esperado do grau do vértice v ?
- Condicionar no bloco (v do bloco i)
$$E[d_v | v \in V_i] = \sum_{j=1}^k n_j b_{ij}$$
- No exemplo acima: $E[d_v] = 40 \cdot 2 + 40 \cdot 0.01 + 40 \cdot 0.01 = 8.8$
 - para v em qualquer bloco i

Aplicações de SBM

- Modelo canônico para redes com comunidades (grupos)
 - em muitos casos $k=2$, $n_k = n$
- Importante no estudo de algoritmos de detecção de comunidades
 - como saber que algoritmo detectou corretamente?
- Importante no estudo de espalhamento de informação e formação de opinião
 - impacto da homofilia estrutural, incluindo a polarização