Estatística de Redes Sociais

Antonio Galves

Módulo 2

4a. aula

Grafos aleatórios e redes sociais. O modelo de Erdős-Rényi.

Recapitulando: Grafos

- Um grafo é um par ordenado $G = (V, E)$,
- tendo como primeiro elemento um conjunto finito de vértices
- e tendo como segundo elemento um conjunto de arestas ligando pares (não orientados) de vértices.

Recapitulando: Grafos dirigidos

- Um **grafo** $G = (V, E)$ é dirigido, se cada aresta tem uma direção.
- Ou seja, um grafo dirigido é um par formado por um conjunto de vértices e por um conjunto de flechas ligando vértices.

O grafo dirigido em nosso modelo básico de rede social

- O conjunto de vértices é o conjunto de atores;
- e o conjunto de flechas indo de um ator a outro é deduzido do conjunto de influenciadores:
- dado um par de atores a, b, existe uma flecha indo de a para b, se a pertencer ao conjunto de influenciadores de b.
- Ou seja, (a, b) é uma flecha do grafo dirigido, se $a \in \mathcal{V}_{\cdot \to b}$.

A matriz de adjacência de um grafo

- A matriz de adjacência é uma forma computacional de representar o grafo $G = (V, E)$.
- Se o grafo for dirigido, para todo par ordenado de vértices (v, v'), com $v \neq v'$, definimos

  $$M(v, v') = 1$$

  se houver uma flecha (uma aresta dirigida), indo de v a v', e $M(v, v') = 0$ se não houver uma flecha indo de v a v'.
- Se o grafo não for dirigido, a ordem (v, v') ou (v', v) não importa e $M(v, v') = M(v', v)$.

Grafos conectados

- Dados dois vértices a e b, existe um número inteiro $k \geq 1$, e uma sequência de vértices $v_0, \ldots, v_k$, tais que
  1. $v_0 = a$ e $v_k = b$;
  2. para todo $j = 0, \ldots, k-1$,

$$M(v_j, v_{j+1}) = 1.$$

Redes sociais são modeladas por grafos aleatórios

- Na vida real as redes de relacionamento são quase que inevitavelmente aleatórias.
- Isso quer dizer que relações de amizade, ou de influência entre atores se constituem em parte por obra do acaso.
- Isso nos conduz à seguinte pergunta:
- Como garantir que a grande maioria dos atores de uma rede construida aleatoriamente esteja conectada?

O grafo de Erdős-Rényi

- O grafo de Erdős-Rényi foi introduzido como um grafo não dirigido.
- Seu conjunto de vértices é $V = \{1, \ldots, N\}$.
- Entre dois vértices v e v', $v \neq v'$ quaisquer, colocamos uma aresta com probabilidade p, onde $p \in [0, 1]$. Essa decisão é tomada independentemente para cada par (não ordenado) de vértices.
- Essa classe de grafos será designada com a notação $G(N, p)$.

Simulando $G(N, p)$

- O conjunto de vértices é $V = \{1, \ldots, N\}$. Temos que atribuir o valor 0 ou 1 a cada par (v, v').
- Como se trata de um grafo não dirigido, basta sortear os valores de $M(v, v')$, para cada $v = 1, \ldots, N-1$ e cada $v' = v+1, \ldots, N$.
- Os valores de $M(v, v')$, $v < v'$, são variáveis aleatórias independentes identicamente distribuidas (i.i.d.) com

$$\mathbb{P}\{M(v, v') = 1\} = p \text{ e } \mathbb{P}\{M(v, v') = 0\} = 1 - p.$$

- Lembre que, o grafo sendo não dirigido, $M(v, v') = M(v', v)$.

QUIZ

- Quantos sorteios teremos que fazer para encontrar todos os valores da matriz de adjacência M caracterizando o grafo $G(N,p)$?
- Em outras palavras de quantas maneiras podemos escolher subconjuntos com exatamente dois elementos em $V = \{1,\ldots,N\}$?
- Combinações de N elementos 2 a 2:

$$\binom{N}{2} = \frac{N!}{2(N-2)!} = \frac{N(N-1)}{2}$$

Simulando $G(N, p)$

1. Para cada um dos $\binom{N}{2}$ pares não ordenados de vértices (v, v') sorteamos $U(v, v')$ uniformemente no intervalo $[0, 1]$.
2. Esse sorteio é feito independentemente para todos os pares de vértices.
3. $M(v, v') = 1$, se $U(v, v') \leq p$
4. $M(v, v') = 0$, se $U(v, v') > p$.

O grau dos vértices de $G(N, p)$

- Em um grafo não dirigido, tendo $V = \{1, \ldots, N\}$ como conjunto de vértices, o grau de um vértice v é o número de arestas que incidem em v.
- Formalmente,

$$deg(v) = \sum_{v' \in V : v' \neq v} M(v, v')$$

- Se o grafo for dirigido, teremos que distinguir entre arestas que chegam em v (grau de entrada de v) e arestas que saem de v (grau de saída).

QUIZ

- Que valores $deg(v)$ pode assumir, onde v é um elemento qualquer do conjunto de vértices $V = \{1, \ldots, N\}$?
- Resposta: $deg(v) \in \{0, \ldots, N-1\}$

Cálculos básicos

- Em um grafo $G(N, p)$, calcule o valor médio (a esperança) de $deg(v)$

$$\mathbb{E}[deg(v)] = \mathbb{E}\left[\sum_{v' \in V : v' \neq v} M(v, v')\right]$$

- Calcule também a distribuição de probabilidades de $deg(v)$, isto é, calcule a probabilidade $\mathbb{P}\{deg(v) = k\}$ para todos os valores possíveis de $k \in \{0, ..., N-1\}$.

Respostas

$$\mathbb{E}[deg(v)] = \mathbb{E}\left[\sum_{v' \in V : v' \neq v} M(v, v')\right] = \sum_{v' \in V : v' \neq v} \mathbb{E}[M(v, v')].$$

$$\mathbb{E}[M(v, v')] = 1.p + 0(1 - p) = p$$

Logo

$$\mathbb{E}[deg(v)] = (N - 1)p\,.$$

Respostas

Para todo $k = 0, \dots, N-1$

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

Queremos estudar o caso em que $N \to +\infty$

- Vamos supor que $p = p_N = \lambda/N$, onde λ é uma constante positiva.
- **Teorema** Para todo inteiro $k \geq 0$

$$\lim_{N \to +\infty} \binom{N-1}{k} p_N^k (1 - p_N)^{(N-1-k)} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Demonstração

- $$\binom{N-1}{k} p_N^k (1-p_N)^{(N-1-k)} = \frac{(N-1)!}{k!(N-1-k)!} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^{(N-1-k)}$$

$$= \frac{\lambda^k}{k!} \frac{(N-1)...(N-k)}{N^k} \left(1 - \frac{\lambda}{N}\right)^{(N-1-k)}$$

- Concluímos, observando que

$$\frac{(N-1)}{N} ... \frac{(N-k)}{N} = \left(1 - \frac{1}{N}\right) ... \left(1 - \frac{k}{N}\right) \to 1,$$

$$\left(1 - \frac{\lambda}{N}\right)^{(N-1-k)} \to e^{-\lambda}, \text{ quando } N \to \infty.$$

Exercícios

1. Verifique a seguinte identidade:

$$\sum_{k=0}^{K} \binom{K}{k} p^k (1-p)^{(K-k)} = Kp.$$

2. Transforme o pseudo-código para a construção de $G(N, p_N)$ dado no curso em um código efetivo para simular um grafo desta classe, com $p_N = \lambda/N$ para diferentes valores de $\lambda > 0$.
3. Observe as diferenças do grafo $G(N, p_N)$ com $N = 100$, para valores de $\lambda < 1$, e valores de $\lambda > 1$.
4. Considere o modelo do votante, tendo como conjunto de atores $A = \{1, \ldots, 100\}$ e tendo relações de influência dadas pelo grafo $G(100, \lambda/100)$ da seguinte maneira. O ator a e b se influenciam mutuamente se $M(a, b) = 1$. Observe a evolução dessa rede quando $\lambda = 1/2$ e $\lambda = 5$.