Estatística de Redes Sociais

Antonio Galves

Módulo 2

9a. aula

Grafos aleatórios e redes sociais.

Módulos 1 e 2: complementos e revisão.

QUIZ

- Seja G um grafo gerado aleatoriamente na classe $G(N, p)$ com $p \in (0, 1)$.
- Sabendo que G está na classe $G(N, p)$, calculamos a sua verossimilhança
$$p^{N_G(1)}(1-p)^{N_G(0)}.$$
- É ou não é verdade que sempre vale
$$p^{N_G(1)}(1-p)^{N_G(0)} \geq q^{N_G(1)}(1-q)^{N_G(0)}$$
para todo $q \in [0, 1]$.

RESPOSTA

- A resposta é **NÃO**!
- De fato

$$p^{N_G(1)}(1-p)^{N_G(0)} < \hat{p}^{N_G(1)}(1-\hat{p})^{N_G(0)}$$

onde

$$\hat{p} = \frac{\mathcal{N}_G(1)}{\mathcal{N}_G(0) + \mathcal{N}_G(1)}.$$

- O máximo só é atingido por $p^{N_G(1)}(1-p)^{N_G(0)}$, quando por acaso $\hat{p} = p$, o que é pouco provável quando $|V|$ é grande.
- Por exemplo, se $p = 1/2$ e $N = 5$, é preciso que exatamente 5 das 10 possíveis arestas esteja presente, o que acontece com probabilidade

$$\frac{1}{2^{10}}\binom{10}{5} = \frac{252}{1024} \approx 0.25.$$

Lembrando a noção de distância entre dois vértices

- Dado um grafo tendo M como matriz de adjacência, definimos a distância entre dois vértices v e v' como sendo o menor $k \geq 1$ tal que existem vértices $v_0, v_1, ..., v_k$, com $v = v_0$ e $v_k = v'$, tais que $M(v_i, v_{i+1}) = 1$, para $i = 0, ..., k-1$.
- $dist(v, v') =$

$$\min\{k : \exists (v_j)_{j=0,\ldots,k}, v = v_0, v_k = v', M(v_i, v_{i+1}) = 1, i = 0, ..., k-1\}$$

Recapitulando o modelo de Watts e Strogatz

- Começamos com um grafo regular, tendo $V = \{1, \ldots, N\}$ e com matriz de adjacência M satisfazendo: $M(v, v') = 1$, se $1 \leq |v - v'| \leq k$, e $M(v, v') = 0$ caso contrário.
- Aqui, $k \geq 1$ é um inteiro muito menor do que N. Usamos a convenção $N + 1 = 1$.
- Aleatorizamos esse grafo regular, decidindo alterar ou manter, com probabilidade $p \in [0, 1]$ ou $1 - p$ respectivamente, cada aresta, independentemente uma das outras.
- Se decidirmos alterar a aresta, uma de suas duas pontas escolhida com probabilidade $1/2$ é reconectada em um outro vértice escolhido uniformemente dentre todos os vértices, com exceção da extremidade que decidimos não alterar na aresta considerada.
- Se no final do processo houver arestas duplicadas, essas arestas serão unificadas em uma só.

D E S A F I O

- Watts e Strogatz aleatorizaram o grafo regular para diminuir a distância típica entre dois vértices.
- Será que a aleatorização tem realmente esse efeito?
- Considere o modelo de Watts e Strogatz com 100 vértices, $k = 10$ e $p \in (0, 1)$. Calcule $\mathbb{P}(dist(1, 50) = 1)$.

RESPOSTA

- Seja $B(1, v)$, com $1 \leq |1 - v| \leq 10$, o evento que consistem em:
- escolher mudar a aresta $\{1, v\}$, com probabilidade p.
- Dado que escolhemos alterar a aresta $\{1, v\}$, escolhemos alterar a ponta que inicialmente estava em v, com probabilidade $1/2$.
- Finalmente, dado que escolhemos alterar a extremidade v da aresta $\{1, v\}$, escolhemos como sua nova ponta o vértice 50, com probabilidade $1/98$.
- Como essas três escolhas são independentes entre si,

$$\mathbb{P}(B(1, v)) = p \times \frac{1}{2} \times \frac{1}{98} = \frac{p}{196}$$

- Definimos de maneira equivalente o evento $B(50, v)$, trocando o vértice 1 pelo vértice 50.
- Da mesma maneira,

$$\mathbb{P}(B(50, v)) = \frac{p}{196}.$$

- A probabilidade $\mathbb{P}(dist(1,50)=1)$ é igual a

$$\mathbb{P}\left(\left\{\bigcup_{1\leq|1-v|\leq 10} B(1,v)\right\}\cup\left\{\bigcup_{1\leq|50-v|\leq 10} B(50,v)\right\}\right).$$

- Os eventos $(B(1,v):1\leq|1-v|\leq 10)$ e $B(50,v:1\leq|50-v|\leq 10)$ são independentes mas não são disjuntos.
- Assim sendo, a probabilidade da união desses eventos é menor do que a soma das probabilidades dos eventos.
- Como calcular a probabilidade dessa união?
- Calculando a probabilidade do evento complementar e aproveitando a independência dos eventos complementares.

- $$\mathbb{P}\left(\left\{\bigcup_{1\leq|1-v|\leq 10} B(1,v)\right\}\cup\left\{\bigcup_{1\leq|50-v|\leq 10} B(50,v)\right\}\right)$$
  $$1-\mathbb{P}\left(\left\{\bigcap_{1\leq|1-v|\leq 10} (B(1,v))^c\right\}\cap\left\{\bigcap_{1\leq|50-v|\leq 10} (B(50,v))^c\right\}\right)$$
- Os eventos $((B(1,v))^c : 1 \leq |1-v| \leq 10)$ e $((B(50,v))^c : 1 \leq |50-v| \leq 10)$ são independentes entre si e, por isso, a probabilidade da intersecção é igual ao produto das probabilidades.
- $$= 1 - \prod_{1\leq|1-v|\leq 10} \mathbb{P}\{(B(1,v))^c\} \prod_{1\leq|50-v|\leq 10} \mathbb{P}\{(B(50,v))^c\}$$

$$1 - \prod_{1 \leq |1-v| \leq 10} \mathbb{P}\{(B(1,v))^c\} \prod_{1 \leq |50-v| \leq 10} \mathbb{P}\{(B(50,v))^c\} =$$

$$1 - \prod_{1 \leq |1-v| \leq 10} \left\{1 - \mathbb{P}\{B(1,v)\}\right\} \prod_{1 \leq |50-v| \leq 10} \left\{1 - \mathbb{P}\{B(50,v)\}\right\} =$$

$$1 - \left(1 - \frac{p}{196}\right)^{40}.$$

Lembrando o modelo *rico fica mais rico*

Queremos definir uma sequência de grafos $G_t = (V_t, E_t)$, $t \geq 0$.

1. Inicialização: Definimos $V_0 = \{1, \ldots, N_0\}$ com $N_0 \geq 2$ qualquer fixado e um conjunto de arestas E_0 de modo que o grau $D_0(v) \geq 1$ para todo $v \in V_0$.
2. Para todo $t \geq 1$:
   2.1 $V_t = V_{t-1} \cup \{|V_{t-1}| + 1\}$.
   2.2 $E_t = E_{t-1} \cup \{|V_{t-1}| + 1, \xi_t\}$, onde ξ_t é um elemento de V_{t-1} escolhido aleatóriamente com a distribuição

$$\mathbb{P}(\xi_t = v) = \frac{D_{t-1}(v)}{\sum_{v' \in V_{t-1}} D_{t-1}(v')}.$$

QUIZ

- Construa uma sequência $(G_t : t \geq 0)$ de grafos *rico fica mais rico*, com $V_0 = \{1, ..., 10\}$, e com M_0, tal que para todo par de vértices $v < v'$, $M_0(v, v') = 1$, se e somente se, v e v' pertencerem ambos a $V_0^{(1)} = \{1, ..., 5\}$, ou ambos a $V_0^{(2)} = \{6, ..., 10\}$.
- O que essa condição inicial diz sobre G_0?
- O que essa condição inicial implica sobre $(G_t : t \geq 0)$?

R E S P O S T A

- **Pergunta:** O que essa condição inicial diz sobre G_0?
- Ela diz que G_0 é desconectado, pois não há nenhuma aresta tendo uma extremidade no conjunto $\{1, ..., 5\}$ e a outra no conjunto $\{6, ..., 10\}$.

- **Pergunta:** O que essa condição inicial implica sobre $(G_t : t \geq 0)$?
- Vejamos o que acontece no instante $t = 1$. $V_0 = \{1, \dots, 10\}$. O novo vértice, representado pelo ponto 11 vai escolher um vértice ξ_1 de V_0 para se ligar.
- Se ξ_1 for escolhido em V_0^1, vamos chamar de V_1^1 o conjunto $V_0^1 \cup \{11\}$ e $V_1^2 = V_0^2$.
- Se ξ_1 for escolhido em V_0^2, vamos chamar de V_1^2 o conjunto $V_0^2 \cup \{11\}$ e $V_1^1 = V_0^1$.
- Em ambos os casos, $M_1(v, v') = 0$ para todo $v \in V_1^1$ e $v' \in V_1^2$.
- Ou seja, o grafo $G_1 = (V_1, E_1)$ é desconectado, assim como G_0.
- Por indução, essa propriedade vale para todo $t \geq 1$.

D E S A F I O

- Seja V um conjunto finito.
- Seja $\mathcal{G}(V)$ o conjunto de todos os grafos não dirigidos tendo V como conjunto de vértices.
- Quanto vale $|\mathcal{G}(V)|$?
- Em outras palavras, quantos grafos não dirigidos diferentes podemos construir tendo V como conjunto de vértices?

RESPOSTA

- Seja $\mathcal{P}_2(V)$ o conjunto de todos os subconjuntos de V tendo exatamente dois elementos.
- Um grafo $G \in \mathcal{G}(V)$ pode ser visto como uma função de $\mathcal{P}_2(V)$ no conjunto $\{0, 1\}$.
- De fato, um grafo $G \in \mathcal{G}(V)$ é totalmente caracterizado pela função que a todo par não orientado $\{u, v\} \in \mathcal{P}_2(V)$ associa o valor $M(u, v) = 1$, se $\{u, v\}$ for uma aresta de $\mathcal{G}(V)$ e associa valor $M(u, v) = 0$, se $\{u, v\}$ não for uma aresta.
- Ou seja, $\mathcal{G}(V) = \{0, 1\}^{\mathcal{P}_2(V)}$.
- Em conclusão,

$$|\mathcal{G}(V)| = 2^{|\mathcal{P}_2(V)|} = 2^{\frac{|V||V-1|}{2}}.$$

Tópicos que podem ser discutidos na prova

1. Evolução do modelo do votante a partir de uma certa configuração
2. Origem das opiniões de dois atores num certo instante no modelo do votante
3. A ida para a unanimidade no modelo do votante e casos em que isso não ocorre
4. O grau de um vértice $D_N(v)$ e o número de arestas $\mathcal{N}_G(1)$ em um grafo G na classe $G(N,p)$: distribuição binomial
5. $D_N(v)$ e $\mathcal{N}_G(1)$ em $G(N,p)$: lei dos grandes números
6. $D_N(v)$ e $\mathcal{N}_G(1)$ em $G(N,p)$: aproximação normal
7. $N_{i,j}(0) + N_{i,j}(1)$ em $G(V^{(1)}, ..., V^{(k)}, \mathbf{p} = (p_{i,j} : i,j = 1, ..., k, i \leq j))$.

Tópicos que podem ser discutidos na prova

8. Dada a matriz de adjacência M com N vértices, calcular o coeficiente de aglomeração do vértice 1, ou o coeficiente médio de aglomeração.
9. Dada a matriz de adjacência M com N vértices, calcular o diâmetro do grafo.
10. Na família de grafos rico fica mais rico, como as propriedade de conexão de G_0 afetam os sucessivos grafos $(G_t^t \geq 0)$.
11. Calcular o estimador de máxima verossimilhança de p, dada uma matriz de adjacência M gerada por $G(V^{(1)}, V^{(2)}, p)$.
12. Estudar o diâmetro do grafo de Watts-Strogatz quando $k = \log(N)$ com N grande e $p \in (0, 1)$.