

Desvios moderados do número de triângulos no grafo aleatório $G(n, m)$

Leonardo Gonçalves de Oliveira

PUC-RIO

20 de maio de 2021

Trabalho conjunto com José D. Alvarado e Simon Griffiths

Grafos aleatórios

Modelo $G(n, p)$: grafo com n vértices em que cada aresta é incluída independentemente com probabilidade p . Para um grafo G de n vértices e m arestas, temos $\mathbb{P}(G(n, p) = G) = p^m(1 - p)^{n-m}$.

Modelo $G(n, m)$: grafo escolhido aleatoriamente e uniformemente dentre todos os grafos com n vértices e m arestas. Para qualquer G nessa família de grafos, temos $\mathbb{P}(G(n, m) = G) = 1/\binom{N}{m}$, onde $N = \binom{n}{2}$.

Contando triângulos em $G(n, p)$

Número esperado de triângulos: $p^3 \binom{n}{3}$.

Desvio padrão: $\Theta(n^2)$.

Barber, Karoński, Ruciński (1989): distribuição assintoticamente normal.

Grandes desvios no número de triângulos em $G(n, p)$

Representamos o número de cópias isomórficas de um grafo F em um grafo G por $N_F(G)$.

Problema: determinar o comportamento de

$$r(\delta, p, n) := -\log \mathbb{P} (N_{\Delta}(G_p) > (1 + \delta)p^3(n)_3) .$$

Vu(2001), Janson e Ruciński (2002):

$$c(\delta)p^2n^2 \leq r(\delta, p, n) \leq C(\delta)p^2n^2 \log(1/p) \text{ para constantes } c(\delta), C(\delta).$$

DeMarco, Khan (2011):

$$c(\delta)p^2n^2 \log(1/p) \leq r(\delta, p, n) \leq C(\delta)p^2n^2 \log(1/p).$$

Grandes desvios no número de triângulos em $G(n, p)$

Lubetzky, Zhao (2016): Se $n^{-1/42} \log n \ll p \ll 1$ então

$$r(\delta, p, n) = (1 + o(1)) \min \left\{ \frac{\delta^{2/3}}{2}, \frac{\delta}{3} \right\} p^2 n^2 \log(1/p)$$

Harel, Mousset, Samotij (2019): A expressão acima vale para $p \gg n^{-1/2}$. Além disso, para $n^{-1} \ll p \ll n^{-1/2}$,

$$r(\delta, p, n) = (1 + o(1)) \frac{\delta^{2/3}}{2} p^2 n^2 \log(1/p).$$

Também encontraram o valor assintótico de $r(\delta, p, n)$ no caso em que $p^2 n \rightarrow c \in \mathbb{R}$.

Desvios moderados no número de triângulos em $G(n, p)$

Döring, Eichelsbacher (2009): Se $p^{-1/2}n^{-1} \ll \delta_n \ll p^7$ então

$$r(\delta, p, n) = \frac{-\delta_n p n^2}{36(1-p)} + o(\delta_n^2 p n^2)$$

No caso denso em que $p \in (0, 1)$, Féray, Méliot e Nikeghbali (2013) encontraram a probabilidade assintótica de desvio.

Contando triângulos em $G(n, m)$

Consideramos $N_{\Delta}(G_m)$, cuja esperança é dada por

$$L_{\Delta}(m) = \frac{(n)(n-1)(n-2)m(m-1)(m-2)}{N(N-1)(N-2)}.$$

e desvio padrão é da ordem $\Theta(n^{3/2})$.

Para desvios pequenos, de ordem até a do desvio padrão, Janson (1994) provou um teorema central do limite que garante que a distribuição de $N_{\Delta}(G_m)$ é assintoticamente normal.

Desvios moderados no número de triângulos em $G(n, m)$

Seja $t = m/N$ a densidade de arestas em G_m . O caso denso (em que t é constante) foi tratado em trabalho de Goldschmidt, Griffiths e Scott (2020).

Theorem (Goldschmidt, Griffiths e Scott)

Seja α_n uma sequência tal que $1 \ll \alpha_n \ll n^{1/2}$. Seja $t \in (0, 1)$ constante e $m = \lfloor tN \rfloor$. Então

$$\mathbb{P} \left(D_{\Delta}(G_m) > \alpha_n n^{3/2} \right) = \exp \left(- \frac{\alpha_n^2 (1 + o(1))}{12t^3(1-t)^2(1+2t)} \right).$$

Desvios moderados no número de triângulos em $G(n, m)$

Consideramos os casos em que G_m é esparso. Mais precisamente, gostaríamos de determinar o comportamento de

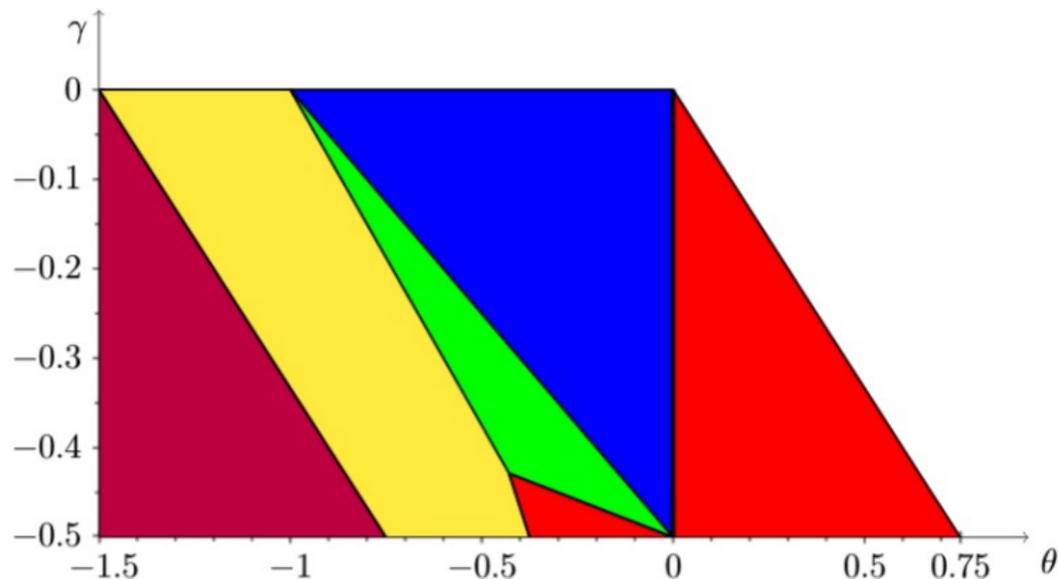
$$r(\delta, t, n) = -\log \mathbb{P} (N_{\Delta}(G_m) > (1 + \delta)t^3 n^3)$$

quando $n^{-1/2}(\log n)^{1/2} \ll t \ll 1$ e $t^{-3/2}n^{-3/2} < \delta < t^{-3/2}$.

Dizemos que ocorre localização se o desvio acontece devido à presença de um certo objeto no grafo, ao invés de simplesmente ser causado pela presença de mais arestas.

Causas dos desvios

Considere $t = n^\gamma$ e $\delta = n^\theta$. A figura a seguir mostra diferentes regiões separadas pelas suas causas de desvio.



Causas dos desvios

Tomando $b = b(n)$ uma sequência pensamos em quão grande pode ser um desvio com probabilidade e^{-b} de ocorrer. Definimos

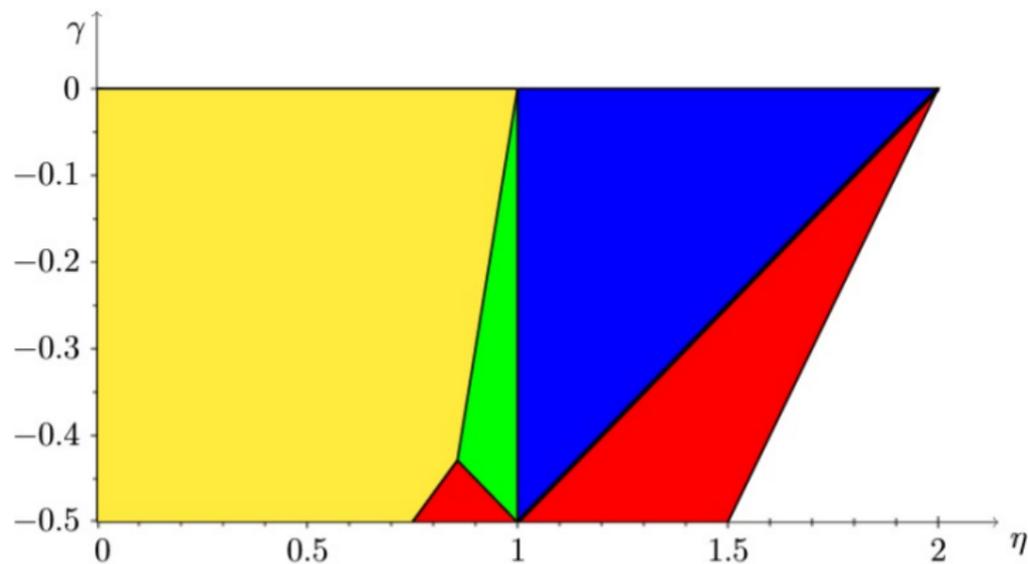
$$DEV_{\Delta}(b, t) = \min\{x : \mathbb{P}(D_{\Delta}(m) > x) \leq \exp(-b)\}.$$

Seja $\ell = \log(1/t)$. Dependendo de b e t , o desvio no número de triângulos pode ter diferentes causas:

- Regime normal: $N(b, t) := b^{1/2} t^{3/2} n^{3/2}$.
- Estrela: $S(b, t) := \frac{b^2 t}{\ell^2} \mathbf{1}_{b \leq n\ell}$
- Bipartido: $H(b, t) := \frac{b t n}{\ell} \mathbf{1}_{b \geq n\ell}$
- Clique: $C(b, t) := \frac{b^{3/2}}{\ell^{3/2}}$

Causas dos desvios

Considere $t = n^\gamma$ e $b = n^\eta$. Obtemos a figura correspondente para esse caso.



Resultado principal

Theorem (Alvarado, Griffiths, O.)

Existem constantes absolutas c, C tais que vale a seguinte afirmação. Para todo $t \geq Cn^{-1/2}(\log n)^{1/2}$ e $3 \log n \leq b \leq tn^2\ell$ vale

$$cM(b, t) \leq DEV_{\Delta}(b, t) \leq CM(b, t)$$

onde $M(b, t) = \max\{N(b, t), S(b, t), H(b, t), C(b, t)\}$.

Hoeffding-Azuma

Seja $(S_i)_{i=1}^m$ um martingal com incrementos $(X_i)_{i=1}^m$. Então

$$\mathbb{P}(S_m - S_0 > a) \leq \exp\left(-\frac{a^2}{2 \sum_{i=1}^m c_i^2}\right)$$

para todo $a > 0$, onde $c_i = \|X_i\|_\infty$.

Freedman

Melhores cotas do que Hoeffding-Azuma no caso em que os incrementos X_i são tipicamente muito menores do que seus respectivos supremos.

Para um martingal $(S_i)_{i=0}^m$ com incrementos $(X_i)_{i=1}^m$ definimos

$$V_j = \sum_{i=1}^j \mathbb{E} [X_i^2 | \mathcal{F}_{i-1}].$$

Para $\alpha > 0$, definimos $\tau_\alpha = \min\{1 \leq j \leq m : S_j - S_0 \geq \alpha\}$ com $\tau_\alpha = m$ caso esse conjunto seja vazio.

Definimos $T_\alpha = V_{\tau_\alpha}$, que representa a variância condicional acumulada até o momento que o martingal ultrapassa α , caso isso ocorra.

Theorem (Freedman (1975))

Seja $(S_i)_{i=0}^m$ um supermartingal com incrementos $(X_i)_{i=1}^m$. Seja $R > 0$ tal que $|X_i| \leq R$ q.c. para todo i . Então, para todos $\alpha, \beta > 0$ vale

$$\mathbb{P}(T_\alpha \leq \beta) \leq \exp\left(-\frac{\alpha^2}{2(\beta + R\alpha)}\right).$$

Caso $aR \ll b \ll a^2$ também vale que

$$\mathbb{P}(T_\alpha \leq \beta) = \exp\left(-\frac{\alpha^2(1 + o(1))}{2\beta}\right).$$

Cota superior

Nosso resultado principal utiliza fortemente as desigualdades de martingais.

Vejamos uma ideia de como provar a cota superior, ou seja, que existe uma constante C tal que

$$\mathbb{P}(D_{\Delta}(G_m) \geq CM(b, t)) \leq \exp(-b)$$

para todo $t \geq Cn^{-1/2}(\log n)^{1/2}$ e $3 \log n \leq b \leq tn^2\ell$.

Processo de Erdős-Rényi

O processo de Erdős-Rényi é uma sequência de grafos aleatórios $(G_i)_{i=0}^N$ que começa com G_0 sendo o grafo vazio com n vértices. A cada passo, adicionamos uma aresta escolhida uniformemente dentre todas as possíveis.

De maneira equivalente, podemos definir G_i como o grafo com arestas $\{e_1, \dots, e_i\}$ onde e_1, e_2, \dots, e_N é uma permutação aleatória das arestas de K_n .

Em ambos os casos, $G_m \sim G(n, m)$.

Representação martingal

Agora os martingais entram em cena. Considere

$A_F(G_i) = N_F(G_i) - N_F(G_{i-1})$ como o número de cópias isomórficas de F criadas com a adição da i -ésima aresta.

Definimos $X_F(G_i) = A_F(G_i) - \mathbb{E}[A_F(G_i)|G_{i-1}]$.

As variáveis $X_F(G_i)$ funcionam como incrementos de um martingal, pois $\mathbb{E}[X_F(G_i)|G_{i-1}] = 0$.

Goldschmidt, Griffiths e Scott mostraram a seguinte representação martingal para o desvio no número de triângulos em G_m :

$$D_{\Delta}(G_m) = \sum_{i=1}^m \left[\frac{(N-m)_2(m-i)}{(N-i)_3} X_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3} X_{\Delta}(G_i) \right].$$

Truncamento

Temos um martingal para $D_{\Delta}(G_m)$! Basta controlar as variâncias condicionais dos incrementos, aplicar Freedman e obter o resultado...errado!

Problema: os incrementos podem ser muito grandes, prejudicando as cotas obtidas.

Como resolver? Fazemos um truncamento e analisamos de modo separado os incrementos grandes.

Truncamento

Definimos

$$K_{\wedge}(b, t) := \frac{M(b, t)}{bt} \text{ e } K_{\Delta}(b, t) := \frac{M(b, t)}{b}.$$

Fazemos então

$$X'_{\wedge}(G_i) = X_{\wedge}(G_i)1_{X_{\wedge}(G_i) \leq K_{\wedge}} \text{ e } X'_{\Delta}(G_i) = X_{\Delta}(G_i)1_{X_{\Delta}(G_i) \leq K_{\Delta}}.$$

Definimos também

$$Z_{\wedge}(G_i) = X_{\wedge}(G_i)1_{X_{\wedge}(G_i) > K_{\wedge}} \text{ e } Z_{\Delta}(G_i) = X_{\Delta}(G_i)1_{X_{\Delta}(G_i) > K_{\Delta}}.$$

Truncamento

Finalmente, definimos

$$D'_{\Delta}(G_m) = \sum_{i=1}^m \left[\frac{(N-m)_2(m-i)}{(N-i)_3} X'_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3} X'_{\Delta}(G_i) \right]$$

e

$$N_{\Delta}^*(G_m) = \sum_{i=1}^m \left[\frac{(N-m)_2(m-i)}{(N-i)_3} Z_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3} Z_{\Delta}(G_i) \right].$$

Assim, $D_{\Delta}(G_m) = D'_{\Delta}(G_m) + N_{\Delta}^*(G_m)$. Controlamos esses termos separadamente.

Controlando $D'_\Delta(G_m)$

Como $D_\Delta(G_m)$ é um supermartingal, controlamos usando Freedman. Para isso, precisamos encontrar cotas adequadas para os incrementos e para a soma das variâncias.

Escrevemos \mathbb{X}'_i para os incrementos de $D'_\Delta(G_m)$. Notamos que os coeficientes que multiplicam $X'_\wedge(G_i)$ e $X'_\Delta(G_i)$ são menores que t e 1 , respectivamente. Considerando o truncamento, obtemos

$$|\mathbb{X}'_i| \leq \frac{2M(b, t)}{b} \text{ e } |\mathbb{X}'_i|^2 \leq 2t^2|X'_\wedge(G_i)|^2 + 2|X'_\Delta(G_i)|^2.$$

Variância condicional dos incrementos

Lemma

Existe uma constante C tal que a seguinte afirmação é verdadeira. Sejam $Cn^{-1/2}(\log n)^{1/2} \leq t \leq 1/2$ e $b \geq n$. Então

$$\mathbb{P} \left(\sum_{i=1}^m \mathbb{E} [(\mathbb{X}'_i)^2 | G_{i-1}] \geq C(bt^3n^2 + b^2t) \right) \leq \exp(-b).$$

Variância condicional dos incrementos

Temos $\mathbb{E} [X_{\wedge}(G_i)^2 | G_{i-1}] = \text{Var}(A_{\wedge}(G_i) | G_{i-1})$. Se $e_i = uw$, então $A_{\wedge}(G_i) = 2d_u(G_{i-1}) + 2d_w(G_{i-1})$. Como e_i é selecionado uniformemente entre as $N - i + 1$ arestas de $K_n \setminus G_{i-1}$ é possível mostrar que

$$\mathbb{E} [X_{\wedge}(G_i)^2 | G_{i-1}] \leq \frac{32}{n} \sum_u D_u(G_{i-1})^2.$$

De maneira análoga,

$$\mathbb{E} [X_{\Delta}(G_i)^2 | G_{i-1}] \leq \frac{32}{n^2} \sum_{uw} D_{uw}(G_{i-1})^2.$$

Desvios dos graus

Proposição

Existe uma constante $C > 0$ tal que a seguinte afirmação é verdadeira. Sejam $t \geq 2n^{-1} \log n$ e $b \geq tn$. Para todo $i \leq m$, vale

$$\mathbb{P} \left(\sum_u D_u(G_i)^2 \geq C\kappa(b, t) \right) \leq \exp(-b)$$

onde

$$\kappa(b, t) = \begin{cases} tn^2 & tn \leq b < t^{1/2}nl \\ b^2/\ell^2 & t^{1/2}nl \leq b < nl \\ bn/\ell & nl \leq b \leq tn^2\ell. \end{cases}$$

Desvios dos cograus

Proposição

Existe uma constante $C > 0$ tal que a seguinte afirmação é verdadeira. Sejam $t \geq Cn^{-1/2}(\log n)^{1/2}$. Para todos $n \leq b \leq t^2 n^2$ e $i \leq m$, vale

$$\mathbb{P} \left(\sum_{uw} D_{uw}(G_i)^2 \geq Cbt^2 n^2 \right) \leq \exp(-b)$$

e para todo $b > t^2 n^2$ temos

$$\mathbb{P} \left(\sum_{uw} D_{uw}(G_i)^2 \geq Cb^2 \right) \leq \exp(-b).$$

Variância condicional dos incrementos

Assim, existe uma constante $C > 0$ tal que com probabilidade $1 - \exp(-b)$ temos

$$|\mathbb{X}'_i|^2 \leq 2t^2 |X'_\wedge(G_i)|^2 + 2|X'_\Delta(G_i)|^2 \leq \frac{C\kappa(b, t)t^2}{n} + \frac{\max\{bt^2n^2, b^2\}}{n^2}.$$

O restante da prova do lema segue observando que $\kappa(b, t)t^2 \leq \max\{bt^2n^2, b^2\}$ e aplicando a cota da união.

Controlando $N_{\Delta}^*(G_m)$

Essa é a contribuição dada pelos incrementos que possuem valor alto. Nesse caso, observamos que

$$N_{\Delta}^*(G_m) \leq \sum_{i=1}^m tZ_{\wedge}(G_i) + Z_{\Delta}(G_i).$$

Utilizando a definição, é possível mostrar que $\sum_i Z_{\wedge}(G_i) \leq \sum_v D_v(G_m)^2$.

Segue que existe uma constante $C > 0$ tal que $\sum_i tZ_{\wedge}(G_i) \leq Ct\kappa(b, t)$ com probabilidade mínima de $1 - \exp(-b)$. Como $t\kappa(b, t) \leq M(b, t)$, esse termo está controlado.

Controlando $N_{\Delta}^*(G_m)$

Note que ocorre um aumento em $Z_{\Delta}(G_i)$ quando, ao adicionarmos a aresta e_i criamos muitos triângulos. Isso acontece quando o cograu de e_i em G_{i-1} é muito alto. De fato,

$$\sum_{i=1}^m Z_{\Delta}(G_i) \leq 6 \sum_{i=1}^m d_{e_i}(G_{i-1}) 1_{\{d_{e_i}(G_{i-1}) > K_{\Delta}/6\}}.$$

Dizemos que e_i é ruim se

$$|D_{e_i}(G_{i-1})| \geq 100t(|D_u(G_{i-1})| + |D_w(G_{i-1})|).$$

Finalmente, controlamos a soma de maneira separada para arestas ruins e não ruins.

Cota inferior

Também provamos que existe uma constante c tal que

$$\mathbb{P}(D_{\Delta}(G_m) \geq cM(b, t)) \geq \exp(-b)$$

para todo $t \geq cn^{-1/2}(\log n)^{1/2}$ e $3 \log n \leq b \leq tn^2\ell$.

No caso do regime normal, usamos a recíproca da desigualdade de Freedman. Fazemos uma leve alteração no supermartingal truncado $D'_{\Delta}(G_m)$ para que tenhamos novamente um martingal.

Nos outros regimes, a prova segue construindo uma estrutura específica que é a responsável pelo desvio naquela região.

De volta ao modelo $G(n, p)$

Podemos deduzir resultados para o modelo $G(n, p)$ utilizando o fato que $G(n, p)$ é uma média dos modelos $G(n, m)$:

$$\mathbb{P}_p(E) = \sum_{m=0}^N \mathbb{P}(\text{Bin}(N, p) = m) \mathbb{P}_m(E)$$

Estendemos os resultados de Döring e Eichelsbacher (2009):

Theorem

Let $n^{-1/2} \log n \ll p \ll 1$ and let δ_n be a sequence satisfying

$$p^{-1/2} n^{-1} \ll \delta_n \ll p^{3/4} (\log n)^{3/4}, n^{-1/3} (\log n)^{2/3} + p \log(1/p).$$

Then

$$r(\delta_n, p, n) = (1 + o(1)) \frac{\delta_n^2 p n^2}{36}.$$

Sobre cerejas

Usando as mesmas ferramentas, obtemos resultados para cerejas (caminhos de tamanho 2). Definimos

$$DEV_{\wedge}(b, t) = \min\{x : \mathbb{P}(D_{\wedge}(m) > x) \leq \exp(-b)\}.$$

Nesse caso, temos apenas três possíveis causas para o desvio:

- Regime normal: $N(b, t) := b^{1/2}tn^{3/2}$.
- Estrela: $S(b, t) := \frac{b^2}{\ell^2}1_{b \leq n\ell}$
- Bipartido: $H(b, t) := \frac{bn}{\ell}1_{b \geq n\ell}$

Theorem (Alvarado, Griffiths, O.)

Existem constantes absolutas c, C tais que vale a seguinte afirmação. Para todo $t \geq Cn^{-1/2}(\log n)^{1/2}$ e $3 \log n \leq b \leq tn^2\ell$ vale

$$cM(b, t) \leq DEV_{\wedge}(b, t) \leq CM(b, t)$$

onde $M(b, t) = \max\{N(b, t), S(b, t), H(b, t)\}$.

E agora?

Algumas possibilidades de trabalho futuro:

- Encontrar a versão assintótica de $r(\delta, t, n)$.
- Estender os resultados para qualquer grafo H , como no caso denso.
- Estudar progressões aritméticas em subconjunto esparsos $B_m \subseteq \mathbb{Z}_n$ (ver Griffiths, Koch, Secco (2020)).

Fim

Obrigado!