Zufällige Graphen

Christoph Schumacher

17. Mai 2020

RGCN I, Kap. 4.1, 4.2, 4.3:
Der subkritische Erdős-Rényi-Graph

Zusammenfassung

Einführung zum Erdős-Rényi-Graph, Verhalten der größten Zusammenhangskomponente im subkritischen Fall $\lambda < 1$

Inhaltsverzeichnis

$f(n) = \Theta(n^2)$ bekannt?

$$\exists \gamma > 0\colon \gamma^{-1} \leq \frac{f(n)}{n^2} \leq \gamma \tag{1}$$

bzw.

$$0 < \liminf \frac{f(n)}{n^2} \leq \limsup \frac{f(n)}{n^2} < \infty \tag{2}$$

Link zu Wikipedia

Fragen

1. —

1 Erforschung der Zusammenhangskomponenten

Was soll das eigenlich bedeuten: „Bedingt auf S_{t-1} ist

$$X_t \sim \mathrm{Bin}(n-(t-1)-S_{t-1}, p).\text{“?} \tag{3}$$

Um das umzusetzen, kann man für jedes $s \in \{0, 1, \ldots, n-(t-1)\}$ eine von allem anderen unabhängige ZV

$$X_t^{(s)} \sim \mathrm{Bin}(n-(t-1)-s, p) \tag{4}$$

vorrätig halten und damit

$$X_t := X_t^{(S_{t-1})} \tag{5}$$

definieren. Dann erhalten wir

$$\begin{aligned}
\mathbb{P}(X_t = k \mid S_{t-1} = s) &= \mathbb{P}(X_t^{(S_{t-1})} = k \mid S_{t-1} = s) && (6)\\
&= \mathbb{P}(X_t^{(s)} = k \mid S_{t-1} = s) && (7)\\
&= \mathbb{P}(X_t^{(s)} = k) && (8)\\
&= \binom{n-(t-1)-s}{k} p^k (1-p)^{n-(t-1)-s-k}, && (9)
\end{aligned}$$

also X_t ist, bedingt auf $\{S_{t-1} = s\}$, $\mathrm{Bin}(n-(t-1)-s, p)$-verteilt, oder, etwas kürzer: X_t ist, bedingt auf S_{t-1}, $\mathrm{Bin}(n-(t-1)-S_{t-1}, p)$-verteilt.

Wie rechnet man damit? — Testrechnung mit $q := 1-p$:

$$\begin{aligned}
\mathbb{P}(X_2 = k) &= \sum_{j=0}^{\infty} \mathbb{P}(\underbrace{S_1}_{=X_1 \sim \mathrm{Bin}(n-1,p))} = j)\, \mathbb{P}(X_2 = k \mid S_1 = j) && (10)\\
&= \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \binom{n-1-j}{k} p^k q^{n-1-j-k} && (11)
\end{aligned}$$

Damit (4.1.4) überhaupt sinnvoll ist, braucht man $N_t := n-t-S_t \geq 0$. Warum stimmt das? — Beweis per Induktion über t:

IA: Für $t = 0$ ist $N_0 = n - 0 - S_0 = n - 1 \geq 0$, denn der die Zshgskomponente hat mindestens einen Knoten, also ist $n \geq 1$.

IV: Für ein $t \in \mathbb{N}$ gelte $N_{t-1} \geq 0$.

IS: Nach der Induktionsannahme können wir $X_t \sim \mathrm{Bin}(N_{t-1}, p)$ bilden. Damit ist dann $N_t = n - t - S_t = n - t - (S_{t-1} + X_t - 1) = N_{t-1} - X_t \geq 0$ fast sicher, denn wegen der Verteilung von X_t ist $X_t \leq N_{t-1}$ fast sicher.

Also können wir tatsächlich die Folgen $(X_t)_{t\in\mathbb{N}_0}$, $(S_t)_{t\in\mathbb{N}_0}$ und $(N_t)_{t\in\mathbb{N}_0}$ bilden. Dabei ist $S_t \leq n - t$ und $N_t \geq 0$, und (N_t) fällt monoton und konvergiert fast sicher gegen 0.

$$X \sim \mathrm{Bin}(0, p) \implies \mathbb{P}(X = 0) = \binom{0}{0} p^0 (1-p)^0 = 1 \tag{12}$$

2 Monotonie in der Kantenwahrscheinlichkeit

Alternative Formulierung von Definition 4.1: Jede Konfiguration $\omega \in \Omega$ bestimmt einen Graphen G_ω auf $[n]$. Eine ZV $X : \Omega \to \mathbb{R}$ is *wachsend*, wenn

$$\forall \omega, \omega' \in \Omega \colon G_\omega \subseteq G_{\omega'} \implies X(\omega) \leq X(\omega'). \tag{13}$$

Ein Ereignis $A \subseteq \Omega$ ist *wachsend*, wenn $\mathbf{1}_A$ eine wachsende Zufallsgröße ist.

Erklärung: Wenn $\{X \geq x\}$ für alle $x \in \mathbb{R}$ ein wachsendes Ereignis ist, dann ist für ω, ω' mit $G_\omega \subseteq G_{\omega'}$:

$$x := X(\omega) \implies \omega \in \{X \geq x\} \implies \omega' \in \{X \geq x\} \implies X(\omega') \geq x = X(\omega), \tag{14}$$

also ist X wachsend im Sinne von (13). Wenn X wachsend im Sinne von (13) ist, dann ist für alle $x \in \mathbb{R}$ und $\omega, \omega' \in \Omega$ mit $G_\omega \subseteq G_{\omega'}$:

$$\omega \in \{X \geq x\} \implies X(\omega) \geq x \overset{X(\omega)\leq X(\omega')}{\implies} X(\omega') \geq x \implies \omega' \in \{X \geq x\}. \tag{15}$$

Für Ereignisse: Wenn $A \subseteq \Omega$ ein wachsendes Ereignis ist, dann ist $\mathbf{1}_A$ wachsend: Seien $\omega, \omega' \in \Omega$ mit $G_\omega \subseteq G_{\omega'}$. Wenn $\omega \notin A$, dann ist $\mathbf{1}_A(\omega) = 0 \leq \mathbf{1}_A(\omega')$. Wenn $\omega \in A$, dann ist auch $\omega' \in A$, also gilt in beiden Fällen

$$\mathbf{1}_A(\omega) \leq \mathbf{1}_A(\omega'). \tag{16}$$

Umkehrung: siehe oben mit $A = \{\mathbf{1}_A \geq 1\}$.

3 Zshgskomponente $\preceq$ Nachkommenschaft

Warum ist $(X_j^{\geq})_j$ i. i. d.? — Um die bedingte Verteilung auszurechnen, bedingt man:

$$\mathbb{P}(X_i^{\geq} = k) = \sum\nolimits_t \mathbb{P}(N_{i-1} = t)\, \mathbb{P}(X_i^{\geq} = k \mid N_{i-1} = t) \tag{17}$$

$$= \sum\nolimits_t \mathbb{P}(N_{i-1} = t)\, \mathbb{P}(X_i^{(t)} + Y_i^{(t)} = k \mid N_{i-1} = t) \tag{18}$$

$$= \sum_t \mathbb{P}(N_{i-1} = t) \binom{n}{k} p^k (1-p)^{n-k} \tag{19}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}, \tag{20}$$

also ist $X_j^{\geq} \sim \mathrm{Bin}(n,p)$ ohne Bedingung, und damit ist $(X_j^{\geq})_j$ identisch verteilt. Zur Unabhängigkeit von $(X_j^{\geq})_j$: Für alle x_j gilt

$$\mathbb{P}\big((X_j^{\geq})_{j=1}^i = (x_j)_{j=1}^i\big) \tag{21}$$

$$= \sum_t \mathbb{P}\big((N_j)_{j=1}^i = t\big)\, \mathbb{P}\big((X_j^{\geq})_{j=1}^i = (x_j)_{j=1}^i \mid (N_j)_{j=1}^i = t\big) \tag{22}$$

$$= \sum_t \mathbb{P}\big((N_j)_{j=1}^i = t\big)\, \mathbb{P}\big((X_j^{(t_j)} + Y_j^{(t_j)})_{j=1}^i = (x_j)_{j=1}^i \mid (N_j)_{j=1}^i = t\big) \tag{23}$$

$$= \sum_t \mathbb{P}\big((N_j)_{j=1}^i = t\big) \prod_{j=1}^i \binom{n}{x_j} p^{x_j} (1-p)^{n-x_j} \tag{24}$$

$$= \prod_{j=1}^i \mathbb{P}(X_j^{\geq} = x_j) \tag{25}$$

Warum wird im Buch auf $(X_j)_{j=1}^{i-1}$ bedingt? — Die ZV N_{i-1} ist $\sigma(X_1, \ldots, X_{i-1})$-messbar. Wenn man auf $(X_j)_{j=1}^{i-1}$ bedingt, dann kennt man auch N_{i-1} und damit auch die Verteilung von $X_i^{\geq}$.

4 Untere Schranke an Clustergröße

Warum ist $T^{\leq}$ keine stochastische untere Schranke an $|\mathcal{C}(1)|$? — Die Verteilung von $T^{\leq}$ ist für jedes k eine andere, es handelt sich also gar nicht um eine einzige ZV.

Im Beweis von Theorem 4.3 ist

$$\mathcal{T}_k := \min\{t \in \mathbb{N}_0 \colon N_t \leq n-k\} = \min\{t \in \mathbb{N}_0 \colon t + S_t \geq k\} \tag{4.2.6}$$

ist der erste Zeitpunkt, bei der Erforschung von $\mathcal{C}(1)$ mindestens k Knoten gefunden wurden, und zwar als inaktive oder aktive Knoten, wenn $\mathcal{C}(1)$ nicht schon vorher erschöpft wurde, was man mit $S_t > 0$ für alle $t \leq \mathcal{T}_k$ überprüfen kann.

In jedem Fall ist $\mathcal{T}_k$ für alle $k \in \{0, \ldots, n\}$ wohldefiniert wegen $N_t \xrightarrow{t\to\infty} 0$ fast sicher, und die Folge $(\mathcal{T}_k)_k$ monoton wachsend.

Zu (4.2.7): Wenn $|\mathcal{C}(1)| \geq k$, dann muss man mindestens k Knoten aus $\mathcal{C}(1)$ finden, bevor man $\mathcal{C}(1)$ ausgeschöpft hat, also ist $\mathcal{T}_k \leq k-1$. Formal: Aus $|\mathcal{C}(1)| \geq k$

folgt $S_t \geq 1$ für alle $t \in \{0, \dots, k-1\}$, also

$$
\begin{aligned}
N_{k-1} &= n - (k-1) - S_{k-1} && (26)\\
&\leq n - (k-1) - 1 = n - k, && (27)
\end{aligned}
$$

also ist $\mathcal{T}_k \leq k - 1 < |\mathcal{C}(1)|$. Damit haben wir

$$
\begin{aligned}
\{|\mathcal{C}(1)| \geq k\} &\subseteq \{\mathcal{T}_k < |\mathcal{C}(1)|, \forall t < |\mathcal{C}(1)| \colon S_t > 0\} && (28)\\
&\subseteq \{\forall t \leq \mathcal{T}_k \colon S_t > 0\}. && (29)
\end{aligned}
$$

Für die andere Richtung: Für alle $t \leq |\mathcal{C}(1)|$ ist S_t die Anzahl der aktiven Knoten, und t ist die Anzahl der inaktiven Knoten. Sowohl aktive als auch inaktive Knoten sind in $\mathcal{C}(1)$, also ist für alle $t \leq |\mathcal{C}(1)|$:

$$
|\mathcal{C}(1)| \geq t + S_t = n - N_t. \tag{30}
$$

Ist $\mathcal{T}_k \leq |\mathcal{C}(1)|$, dann können wir $t = \mathcal{T}_k$ einsetzen und bekommen

$$
|\mathcal{C}(1)| \geq n - N_{\mathcal{T}_k} \geq n - (n-k) = k. \tag{31}
$$

Damit können wir zeigen:

$$
\begin{aligned}
\{\forall t \leq \mathcal{T}_k \colon S_t > 0\} &\subseteq \{\mathcal{T}_k < \inf\{t \colon S_t = 0\}\} && (32)\\
&= \{\mathcal{T}_k < |\mathcal{C}(1)|\} && (33)\\
&\subseteq \{|\mathcal{C}(1)| \geq k\}. && (34)
\end{aligned}
$$

5 Maximale Zusammenhangskomponente subkritisch

Zu Gleichung (4.3.5) bzw. **Aufgabe 4.14**: Es gilt

$$
\{|\mathcal{C}_{\max}| < k\} = \{Z_{\geq k} = 0\}, \tag{35}
$$

denn genau dann, wenn die größte Zusammenhangskomponente kleiner als k ist, gibt es keinen Knoten in einer Zusammenhangskomponente mit Mindestgröße k. Wir bilden die Komplemente und erhalten

$$
\{|\mathcal{C}_{\max}| \geq k\} = \{Z_{\geq k} > 0\}. \tag{36}
$$

Wenn es aber mindestens einen Punkt in einer Zusammenhangskomponente der Mindestgröße k gibt, so hat man automatisch mindestens k solche Punkte:

$$
\{Z_{\geq k} > 0\} = \{Z_{\geq k} \geq k\}. \tag{37}
$$

Damit können wir umschreiben:

$$|\mathcal{C}_{\max}| = \max\{k\colon k \leq |\mathcal{C}_{\max}|\} = \max\{k\colon Z_{\geq k} > 0\} = \max\{k\colon Z_{\geq k} \geq k\} \tag{38}$$

Zu (4.3.7):

$$\begin{aligned}
\mathbb{P}_\lambda(|\mathcal{C}_{\max}| \geq k) &= \mathbb{P}_\lambda\Big(\bigcup_{v\in[n]} \{|\mathcal{C}(v)| \geq k\}\Big) && (39)\\
&\leq \sum_{v\in[n]} \mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k) && (40)\\
&= n\,\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k) && (41)
\end{aligned}$$

Der Erwartungswert von $Z_{\geq k}$ ist

$$\mathbb{E}[Z_{\geq k}] = \sum_{v\in[n]} \mathbb{E}[\mathbf{1}_{\{|\mathcal{C}(v)|\geq k\}}] = \sum_{v\in[n]} \mathbb{P}(|\mathcal{C}(v)| \geq k) = n\,\mathbb{P}(|\mathcal{C}(1)| \geq k). \tag{42}$$

Mit der first moment method läuft der Beweis so:

$$\begin{aligned}
\mathbb{P}(|\mathcal{C}_{\max}| \geq k) &= \mathbb{P}(Z_{\geq k} \geq k) && (43)\\
&\leq k^{-1}\,\mathbb{E}[Z_{\geq k}] && (44)\\
&= nk^{-1}\,\mathbb{P}(|\mathcal{C}(1)| \geq k) && (45)\\
&\leq \dots && (46)
\end{aligned}$$

Aufgabe 4.16:

$$\begin{aligned}
&\mathbb{P}_\lambda(1 \not\longleftrightarrow 2 \mid |\mathcal{C}(1)| = l) && (47)\\
&= \sum_{A\in\binom{[n]\setminus\{1\}}{l}} \mathbb{P}_\lambda(1 \not\longleftrightarrow 2 \mid \mathcal{C}(1) = A, |\mathcal{C}(1)| = l)\,\mathbb{P}_\lambda(\mathcal{C}(1) = A \mid |\mathcal{C}(1)| = l) && (48)\\
&= \sum_{A\in\binom{[n]\setminus\{1\}}{l}} \mathbf{1}_{[n]\setminus A}(2)\frac{1}{\binom{n-1}{l-1}} && (49)\\
&= \sum_{A\in\binom{[n]\setminus\{1,2\}}{l}} \frac{1}{\binom{n-1}{l-1}} && (50)\\
&= \frac{\binom{n-2}{l-1}}{\binom{n-1}{l-1}} = \dots = \frac{n-l}{n-1} = 1 - \frac{l-1}{n-1} && (51)
\end{aligned}$$