Slučajni grafi in verjetnostna metoda

Borut Lužar

1 Uvod

Slučajni graf (ang. *random graph*) je graf, ki mu na dani množici vozlišč naključno določimo povezave. Predpostavljali bomo, da povezave določamo neodvisno, torej, da je verjetnost izbora neke povezave, enaka verjetnosti izbora katerekoli druge. Slučajne grafe bomo uporabljali pri dokazovanju trditev s pomočjo *verjetnostne metode* (ang. *probabilistic method*). Razvil jo je madžarski matematik Paul Erdős leta 1959 za dokaz izreka o povezanosti kromatičnega števila in notranjega obsega grafa (ang. *girth*), kjer je *notranji obseg* grafa $g(G)$ dolžina najkrajšega cikla v grafu G.

Izrek pravi, da za vsako naravno število k obstaja graf G, za katerega veljata neenačbi: $g(G) > k$ in $\chi(G) > k$. Standarden pristop k dokazovanju takšnega izreka, bi bila konstrukcija grafa z željeno lastnostjo za neki manjši k in nato s pomočjo indukcije prehod na grafe z lastnostjo za večje k. V našem primeru takšen način ni uporaben, saj, če ustrežemo lastnosti z velikim notranjim obsegom, dobimo graf, ki je lokalno 2-obarvljiv. Erdős je k stvari pristopil s popolnoma druge strani. Za vsak n je definiral verjetnostni prostor na množici grafov z n vozlišči in pokazal, da za neke pazljivo izbrane verjetnostne mere, obstaja pozitivna verjetnost, da obstaja graf z n vozlišči, ki zadošča obema lastnostima za vse dovolj velike n.

2 Pojem slučajnega grafa

Naj bo V množica z n elementi, recimo $V = \{0, \ldots, n-1\}$. Naš cilj je spremeniti množico G vseh grafov na V v verjetnostni prostor in potem poskusiti odgovoriti na najbolj tipično zastavljeno vprašanje o slučajnih objektih, kakšna je verjetnost, da ima nek graf $G \in \mathcal{G}(n,p)$ to ali ono lastnost.

Lahko je videti, da graf G zgradimo naključno tako, da se za vsako povezavo $e \in [V]^2$ odločimo naključno ali je, oziroma ni vsebovana v grafu G. Te odločitve so izvedene neodvisno, verjetnost dogodka, da je e element grafa G, pa je enaka $p \in [0,1]$.

Naj bo G_0 nek graf na množici vozlišč V in z m povezavami. Verjetnost, da ga bomo izbrali, bo enaka

$$p^m(1-p)^{\binom{n}{2}-m}.$$

Torej bo za vsak izbran graf G, verjetnost, da je enak G_0, enaka zgornji. Verjetnost, da je G izomorfen G_0, bo seveda precej večja. Če so verjetnosti vseh elementarnih dogodkov tako določene, je tako določena tudi mera željenega prostora $\mathcal{G}(n,p)$. Preveriti je potrebno, da verjetnostna mera na $\mathcal{G}(n,p)$, ki se pojavi za vsako povezavo neodvisno z verjetnostjo p, obstaja.

Naj bo $\Omega_e = \{0_e, 1_e\}$ verjetnostni prostor za vsako povezavo $e \in [V]^2$. Definirajmo $P_e(\{1_e\}) = p$ in $P_e(\{0_e\}) = q$, verjetnosti, da povezava je oziroma ni v grafu (verjetnosti elementarih dogodkov prostora Ω_e). Kot verjetnostni prostor $\mathcal{G}(n,p)$ vzemimo produktni prostor

$$\Omega = \prod_{e\in[V]^2} \Omega_e.$$

Element Ω bo preslikava ω, ki vsaki povezavi $e \in [V]^2$ priredi 0_e ali 1_e, verjetnostna mera P na Ω, pa bo produktna mera vseh P_e. V praksi to pomeni, da ω predstavlja graf G na n vozliščih V s povezavami

$$E(G) = \{e \mid \omega(e) = 1_e\}.$$

Grafu G pravimo *slučajni graf* na V z verjetnostjo povezave p.

Vsako množico grafov na V si v prostoru $\mathcal{G}(n,p)$ lahko predstavljamo kot dogodek. Ali drugače, za vsako povezavo $e \in [V]^2$ je množica

$$A_e = \{\omega \mid \omega(e) = 1_e\}$$

vseh grafov G na V, kjer je $e \in E(G)$, dogodek, da je e povezava G. S pomočjo teh dogodkov pa lahko dokažemo prvo trditev.

Trditev 1 *Dogodki A_e so neodvisni in se pojavljajo z verjetnostjo p.*

Dokaz. Po definiciji je

$$A_e = \{1_e\} \times \prod_{e'\neq e} \Omega_{e'}.$$

Ker je P produktna mera vseh mer P_e, dobimo

$$P(A_e) = p \prod_{e'\in e} 1 = p$$

Podobno, če je $\{e_1, \dots, e_k\}$ poljubna podmnožica $[V]^2$, velja

$$\begin{aligned} P(A_{e_1} \cap \cdots \cap A_{e_k}) &= P(\{1_{e_1}\} \times \dots \{1_{e_k}\} \times \prod_{e\notin\{e_1,\dots,e_k\}} \Omega_e) \\ &= p^k \\ &= P(A_{e_1}) \cdots P(A_{e_k}). \end{aligned}$$

■

P je torej enolično določen s p in našo predpostavko, da so dogodki A_e neodvisni. Da izračunamo verjetnosti v prostoru $\mathcal{G}(n,p)$, zadošča delati s tema predpostavkama in uporaba prostora ni več potrebna.

Lema 2 *Za vsa naravna števila n in k, kjer velja $n \geq k \geq 2$, je verjetnost, da ima $G \in \mathcal{G}(n,p)$ množico k neodvisnih vozlišč največ*

$$P[\alpha(G) \geq k] \leq \binom{n}{k} q^{\binom{k}{2}}.$$

Dokaz. Verjetnost, da je neka določena množica $U \subseteq V$ z močjo k neodvisna v grafu G, je enaka $q^{\binom{k}{2}}$. Trditev nato sledi iz dejstva, da je natanko $\binom{n}{k}$ takih množic U. ∎

V kontekstu slučajnih grafov lahko vsako izmed invariant grafa (npr. povprečno stopnjo, povezanost, notranji obseg, kromatično število, itd) interpretiramo kot nenegativno **naključno spremenljivko** na $\mathcal{G}(n,p)$, funkcijo

$$X : \mathcal{G}(n,p) \to [0,\infty).$$

Pričakovana vrednost X pa je število

$$E(X) := \sum_{G \in \mathcal{G}(n,p)} P(G) \cdot X(G).$$

Izračun povprečne vrednosti naključne spremenljivke X je preprost in učinkovit način za dokaz obstoja grafa G z $X(G) < a$ za nek določen $a > 0$ in da ima G hkrati neko lastnost L. Če je pričakovana vrednost X majhna, $X(G)$ ne more biti velik za več kot le nekaj grafov iz $\mathcal{G}(n,p)$, ker je $X(G) \geq 0$ za vse $G \in \mathcal{G}(n,p)$. Torej mora biti X majhen za večino grafov iz $\mathcal{G}(n,p)$ in lahko pričakujemo, da bomo med njimi našli nekega z željeno lastnostjo L.

Ta preprosta ideja je jedro številnih nekonstruktivnih dokazov obstoja z slučajnimi grafi vključno z Erdősevim izrekom, ki ga bomo dokazali v naslednjem razdelku.

Lema 3 (Markova neenakost) *Naj bo $X > 0$ naključna spremenljivka v prostoru $\mathcal{G}(n,p)$ in $a > 0$. Potem velja*

$$P[X \geq a] \leq \frac{E(X)}{a}.$$

Dokaz.

$$\begin{aligned} E(X) &= \sum_{G \in \mathcal{G}(n,p)} P(G) \cdot X(G) \\ &\geq \sum_{\substack{G \in \mathcal{G}(n,p) \\ X(G) \geq a}} P(G) \cdot X(G) \\ &\geq \sum_{\substack{G \in \mathcal{G}(n,p) \\ X(G) \geq a}} P(G) \cdot a \\ &= P[X \geq a] \cdot a. \end{aligned}$$

∎

Ker so naši verjetnosti prostori končni, lahko pričakovano vrednost pogosto izračunamo z metodo dvojnega štetja (ang. *double counting*). Če je, na primer, X naključna spremenljivka na prostoru $\mathcal{G}(n,p)$, ki šteje podgrafe grafa G v neki določeni množici $\mathcal{H}$ grafov na V, potem $E(X)$ po definiciji šteje pare (G,H) tako, da je $H \subseteq G$. Vsak par je utežen z verjetnostjo $P(G)$. Algoritem za izračun $E(X)$ bo približno takšen: Za vsak graf $G \in \mathcal{G}(n,p)$ pogledamo koliko podgrafov vsebuje množica $\mathcal{H}$ in tako izračunamo $P(G)$.

Izračunajmo število pričakovanih ciklov neke dane dolžine $k \geq 3$ v naključnem grafu $G \in \mathcal{G}(n,p)$. Naj bo torej $X : \mathcal{G}(n,p) \to \mathbb{N}$ naključna spremenljivka, ki vsakemu izmed slučajnih grafov G priredi število k-ciklov v njem, to je število podgrafov izomorfnih C^k. Definirajmo

$$(n)_k := n(n-1)(n-2)\cdots(n-k+1),$$

število zaporedij k različnih elementov na dani množici z n elementi.

Lema 4 *Pričakovano število k-ciklov v $G \in \mathcal{G}(n,p)$ je*

$$E(X) = \frac{(n)_k}{2k}\,p^k.$$

Dokaz. Za vsak k-cikel C z vozlišči v $V = 0, \ldots, n-1$, množici vozlišč grafov iz $\mathcal{G}$, naj $X_C : \mathcal{G}(n,p) \to \{0,1\}$ predstavlja *indikacijsko naključno spremenljivko* za C:

$$X_C : G \mapsto \begin{cases} 1 & C \subseteq G; \\ 0 & \text{sicer.} \end{cases}$$

Ker X_C dobi samo 1 za pozitivno vrednost, je pričakovana vrednost $E(X_C)$ enaka meri $P[X_C = 1]$ množice vseh grafov v $\mathcal{G}(n,p)$, ki vsebujejo C. Ampak to je natanko verjetnost, da je $C \subseteq G$:

$$E(X_C) = P[C \subseteq G] = p^k. \tag{1}$$

Koliko takšnih ciklov $C = v_0 \ldots v_{k-1} v_0$ obstaja? Takšnih je $(n)_k$ zaporedij $v_0 \ldots v_{k-1}$ različnih vozliščv V in vsak izmed ciklov je opisan z $2k$ takšnih zaporedij. Torej je natanko $(n)_k/2k$ različnih ciklov.

Naša naključna spremenljivka X priredi vsakemu grafu G njegovo število k-ciklov. Torej je X vsota vseh vrednosti $X_C(G)$:

$$X = \sum_C X_C.$$

Pričakovana vrednost je linearna, in zato iz (1) sledi:

$$E(X) = E\big(\sum_C X_C\big) = \sum_C E(X_C) = \frac{(n_k)}{2k} p^k.$$

■

3 Verjetnostna metoda

Verjetnostna metoda se je v diskretni matematiki razvila iz naslednje ideje. Da bi dokazali obstoj nekega objekta z željeno lastnostjo, definiramo verjetnostni prostor na nekem večjem razredu objektov in nato pokažemo, da ima nek element tega razreda željeno lastnost s pozitivno verjetnostjo. Objekti vsebovani v tem prostoru so lahko kakršnikoli. Grafi, preslikave, particije ali ureditve vozlišč ipd. V tem razdelku bomo prikazali uporabo verjetnostne metode pri dokazu Erdősevega izreka.

Kot smo povedali že zgoraj, izrek govori o obstoju grafa z notranjim obsegov in kromatičnim številom večjim od k, kjer je k poljubno naravno število. Recimo ciklom dolžine največ k *kratki* in množicam vozlišč z močjo večjo ali enako $\frac{|V(G)|}{k}$ *velike.* Za dokaz Erdősevega izreka bo torej zadostovalo najti graf G brez kratkih ciklov in brez velikih neodvisnih množic vozlišč. Kajti v tem primeru bodo barvni razredi vsakega barvanja vozlišč G majhni (ne veliki) in zato bomo potrebovali več kot k barv, da pobarvamo G.

Kako takšen graf najti? Če izberemo dovolj majhen p, ni verjetno, da bo slučajni graf iz prostora $\mathcal{G}(n,p)$ vseboval kakšen (kratek) cikel. Če izberemo dovolj velik p, ni verjetno, da bo G imel velike neodvisne množice vozlišč. Vprašanje je torej ali se ti območji p prekrivata, torej ali lahko izberemo takšen p za nek n, da bo oboje: dovolj majhen, da bo $P[g \leq k] < \frac{1}{2}$ in dovolj velik za $P[\alpha \geq \frac{n}{k}] < \frac{1}{2}$, kjer je α velikost neodvisne množice vozlišč. Če bi to bilo res, bi $\mathcal{G}(n,p)$ vseboval vsaj en graf brez kratkih ciklov ali velikih neodvisnih množic vozlišč.

Na žalost je takšna izbira p nemogoča. Območji se ne prekrivata! Kot bo pokazano kasneje, moramo obdržati p pod n^{-1}, da ni preveč verjetno, da se pojavijo kratki cikli, toda za vsak tak p, sploh ni verjetno, da se cikli pojavijo! Vse pa še ni izgubljeno. V želji, da obdržimo pojavljanje velikih neodvisnih množic malo verjetno, postavimo p nad n^{-1}, na $n^{\epsilon-1}$ za nek $\epsilon > 0$. To bo sicer povzročilo pojavitev nekaj kratkih ciklov v G, neodvisno od n tako, da bomo vedno vozlišče v vsakem izmed takšnih ciklov lahko odstranili. Dobljenemu grafu recimo H. Graf H bo brez kratkih ciklov, njegovo neodvisnostno število pa bo kvečjemu enako neodvisnostnemu številu G. Ker H ni veliko manjši od grafa G, bo njegovo kromatično število še vedno veliko in tako smo našli graf z velikim notranjim obsegom in velikim kromatičnim številom.

Za pripravo na formalni dokaz Erdősevega izreka, pokažimo najprej da je verjetnost pojavitve povezave $p = n^{\epsilon-1}$ v resnici vedno dovolj velika, da zagotovi, da $G \in \mathcal{G}(n,p)$ 'skoraj gotovo' nima velikih neodvisnih množic vozlišč.

Lema 5 *Naj bo $k > 0$ naravno število in naj bo $p = p(n)$ funkcija n, takšna da velja $p \geq \frac{6k \ln n}{n}$ za velike n. Potem velja*

$$\lim_{n \to \infty} P[\alpha \geq \frac{n}{2k}] = 0.$$

Dokaz. Za vsa naravna števila n in r z lastnostjo $n \geq r \geq 2$ in vse $G \in \mathcal{G}(n,p)$, Lema 2 implicira

$$\begin{aligned} P[\alpha \geq r] &\leq \binom{n}{k} q^{\binom{r}{2}} \\ &\leq n^r q \binom{r}{2} \\ &= \left(n q^{(r-1)/2}\right)^r \\ &\leq \left(n e^{-p(r-1)/2}\right)^r. \end{aligned}$$

Zadnja neenakost sledi iz dejstva, da je $1 - p \leq e^{-p}$ za vse p. Če je $p \geq (6k \ln n)n^{-1}$ in $r \geq \frac{n}{2k}$, za osnovo potence v zgornjem izrazu velja

$$\begin{aligned} n e^{-p(r-1)/2} &= n e^{-pr/2 + p/2} \\ &\leq n e^{-(3/2)\ln n + p/2} \\ &\leq n n^{-3/2} e^{1/2} \\ &= \sqrt{e}/\sqrt{n} \longrightarrow_n \to \infty 0. \end{aligned}$$

Ker je $p \geq (6k \ln n)n^{-1}$ za velike n vzamemo $r = \lceil \frac{n}{2k} \rceil$

$$\lim_{n \to \infty} P[\alpha \geq \frac{n}{2k}] = \lim_{n \to \infty} P[\alpha \geq r] = 0,$$

kar dokaže lemo. ∎

Izrek 6 (Erdős, 1959) *Za vsako naravno število k obstaja graf H z notranjim obsegom $g(H) > k$ in kromatičnim številom $\chi(H) > k$.*

Dokaz. Naj bo $k \geq 3$, ϵ določen in naj velja $0 < \epsilon < \frac{1}{k}$ ter $p = n^{\epsilon-1}$. Naj bo $X(G)$ število kratkih ciklov v naključnem grafu $G \in \mathcal{G}(n,p)$, to je število ciklov dolžine največ k. Po Lemi 4 imamo

$$E(X) = \sum_{i=3}^{k} \frac{(n)_i}{2i} p^i \leq \frac{1}{2} \sum_{i=3}^{k} n^i p^i \leq \frac{1}{2}(k-2)n^k p^k.$$

Velja $(np)^i \leq (np)^k$, saj je $np = n^\epsilon \geq 1$. Po Lemi 3 velja

$$\begin{aligned} P[X \geq n/2] &\leq E(X)/(n/2) \\ &\leq (k-2)n^{k-1}p^k \\ &= (k-2)n^{k-1}n^{(\epsilon-1)k} \\ &= (k-2)n^{k\epsilon-1}. \end{aligned}$$

Ker smo izbrali ϵ tako, da velja $k\epsilon - 1 < 0$ velja tudi

$$\lim_{n\to\infty} P[X \geq n/2] = 0.$$

Naj bo n dovolj velik, da bo $P[X \geq \frac{n}{2}] < \frac{1}{2}$ in $P[\alpha \geq \frac{n}{2k}] < \frac{1}{2}$, zadnje je mogoče zaradi naše izbire p in Leme 5. Potem obstaja graf $G \in \mathcal{G}(n,p)$ z manj kot $n/2$ kratkimi cikli in $\alpha(G) < \frac{n}{2k}$. Iz vsakega izmed kratkih ciklov izbrišemo vozlišče. Naj bo H dobljeni graf. Potem je $|H| \geq n/2$ in H ne vsebuje kratkih ciklov, torej je notranji obseg $g(H) > k$. Po definiciji G pa velja še

$$\chi(H) \geq \frac{|H|}{\alpha(H)} \geq \frac{n/2}{\alpha(G)} > k.$$

∎

Literatura

[1] R. Diestel, Random graphs, *Graph theory*, Springer-Verlag, Berlin, 2000.