

Grafs aleatoris amb una seqüència de graus donada

GUILLEM PERARNAU

Resum: Des de l'aparició de la noció de xarxes complexes, els grafs aleatoris han resultat una eina fonamental per modelar-les i estudiar-les. En aquest article tractarem l'estudi dels grafs aleatoris amb una seqüència de graus donada, en què el grau de cada vèrtex és fixat *a priori* per després triar una instància aleatòria que respecti aquestes restriccions. El nostre objectiu és introduir en aquest tema el lector que no hi està familiaritzat, enunciant els resultats més rellevants en l'àrea i explicant perquè són certs, sense donar-ne proves rigoroses.

Paraules clau: grafs aleatoris, seqüència de graus, models per a xarxes complexes, components connexes, distàncies en grafs, enumeració de grafs, mostreig de grafs aleatoris.

Classificació MSC2020: 05C80, 05C82.

1 Introducció

Els grafs aleatoris són estructures combinatòries discretes que des del seu naixement han trobat nombroses aplicacions en una àmplia gamma de disciplines científiques, des de la informàtica i les matemàtiques fins a la biologia i les ciències socials. Originats en el treball seminal d'Erdős i Rényi a la dècada de 1960, la teoria dels grafs aleatoris ha evolucionat, fins a transformar-se en un camp ric, divers i en constant creixement i esdevenir una eina fonamental per estudiar diversos fenòmens en la ciència moderna.

Recentment, l'estudi dels grafs aleatoris s'ha lligat de manera més estreta a l'anàlisi de les xarxes complexes. Les *xarxes complexes* modelen sistemes caracteritzats per elements interconnectats (nodes connectats a través d'arestes), i apareixen en dominis diversos com ara les xarxes socials, les xarxes biològiques, les xarxes neuronals, les xarxes de transport o les xarxes de la informació (Internet, 5G, etc.). Els desenvolupaments tecnològics de les últimes dècades ens han proporcionat la capacitat computacional per processar aquestes xarxes de mida ingent, i han fet palesa la necessitat creixent de comprendre-les des del punt de vista teòric. Això ha plantejat l'ambiciós repte de desenvolupar i analit-

zar models de xarxes que encaixin amb el que s'observa a la realitat. La teoria dels grafs aleatoris ofereix un marc idoni per a aquest objectiu, permetent-nos extreure informació sobre les seves propietats estructurals, estudiar-ne l'evolució dinàmica o mesurar la robustesa davant de pertorbacions.

Objectes de definició senzilla, com ara el graf aleatori binomial, proporcionen un model bàsic inicial per a les xarxes complexes. Tot i la seva simplicitat, aquest model mostra una riquesa estructural extraordinària, capturant, per exemple, dos fenòmens com són l'existència de transicions de fase en l'estructura de les components connexes, o la propietat petit-món (*small-world* en anglès), que assegura que qualsevol parell de vèrtexs es poden connectar a través de la xarxa amb pocs passos. No obstant això, aquest model no aconsegueix capturar algunes de les altres propietats essencials que apareixen en les xarxes del món real, com ara l'heterogeneïtat o les correlacions entre arestes (*clustering*). De fet, les xarxes que trobem al món real exhibeixen una topologia tan rica i diversa que té poc sentit intentar construir un sol objecte que les modeli a totes. Per exemple, les propietats d'una xarxa social com Instagram tenen poques coses en comú amb les de les xarxes elèctriques, fortament marcades per les restriccions geogràfiques. Això justifica la necessitat d'introduir nous models, més sofisticats que el del graf aleatori binomial, que donen lloc a una àrea de recerca nova i dinàmica en la intersecció de la física estadística, la informàtica teòrica i les matemàtiques.

En aquest article ens centrarem en l'objectiu d'introduir heterogeneïtat als models de xarxes, i ho farem a través de la noció de grau. El grau d'un node en una xarxa és el nombre de nodes als quals està connectat. Per exemple, en una xarxa social el grau d'una persona (representada per un node) és el nombre d'amics d'aquesta persona en la xarxa. És ben sabut que la majoria de xarxes exhibeixen seqüències de graus particulars o amb fortes propietats estructurals. Un exemple és l'existència de *nodes concentradors* (*hubs* en anglès), nodes de la xarxa que tenen un grau molt superior a la resta; per exemple, Frankfurt és un hub a la xarxa de transport aeri. D'altra banda, hi ha xarxes en què la seqüència de graus segueix una certa llei de probabilitat; per exemple, a les xarxes anomenades *invariants d'escala* (*scale-free* en anglès), els graus segueixen una distribució de llei de potències.

Aquests exemples motiven la creació d'un model parametritzat pels graus dels seus nodes. És a dir, fixant els graus *a priori*, volem estudiar un graf aleatori cada vèrtex del qual tingui el grau que se li ha assignat. A aquest efecte, si volem generar un graf amb n vèrtexs, podem prendre una seqüència de graus \mathbf{d}_n de longitud n i després prendre $\mathbb{G}_n = \mathbb{G}_{n, \mathbf{d}_n}$ que sigui un graf triat uniformement a l'atzar de $\mathbb{G}_{n, \mathbf{d}_n}$, el conjunt de tots els grafs amb n vèrtexs i seqüència de graus \mathbf{d}_n . Aquest objecte combinatori s'anomena *graf aleatori amb una seqüència de graus donada* i serà el nostre principal objecte d'estudi.

Cal fer notar que aquest model no serà el més adequat per a xarxes que exhibeixen *clustering*, el fenomen que es dona quan els veïns d'un vèrtex tenen tendència a estar connectats. En tal cas, hi ha altres models aleatoris de xarxes, com els grafs geomètrics aleatoris o els grafs hiperbòlics aleatoris, que capturen aquest fenomen.

Per al lector interessat en l'evolució de la teoria de grafes i la seva relació amb els grafes aleatoris, referim a l'article de Rué [47]. D'altra banda, cal destacar que també hi ha models deterministes de xarxes complexes; vegeu Comellas [9].

1.1 Una primera mirada als grafes aleatoris

Tot i que aquest model dona una representació més realista de les xarxes complexes, la seva descripció en dificulta l'anàlisi teòrica. Abans de plantejar-nos com el podem estudiar des del formulisme matemàtic, discutirem els principals obstacles que presenta la definició del model i com superar-los.

Una primera opció per entendre aquest model és donar-hi un enfocament algorímic i/o probabilístic: trobar un procediment que generi \mathbb{G}_n i que sigui prou simple per ser analitzat. Per exemple, el graf aleatori binomial es pot generar seleccionant independentment cada aresta amb probabilitat p , com veurem més endavant.

És temptador voler generar \mathbb{G}_n de la manera següent: començant amb el graf buit amb n vèrtexs, a cada pas, seleccionem «aleatòriament» dos vèrtexs diferents i no adjacents u, v tals que els seus graus actuals siguin més petits que els seus graus finals d_u, d_v i hi afegim una aresta que els connecti. Si un prova de fer-ho, ràpidament s'adonarà que aquest procediment podria quedar bloquejat abans de generar un graf amb la seqüència de graus desitjada, ja que totes les parelles de vèrtexs susceptibles de ser triades podrien ja estar unides per una aresta. I encara que el procés tingués èxit, podria ser que no es generessin tots els grafes de \mathcal{G}_{n,d_n} amb la mateixa probabilitat.

Podem modificar el procediment anterior seleccionant qualsevol u, v , sempre que els seus graus actuals siguin més petits que els seus graus finals, permetent llaços i arestes múltiples. Això es coneix com a *model de configuració* i s'introduirà en detall més endavant. Un algorisme senzill, conegut com a *mostreig per rebuig*, consisteix a executar repetidament aquest procediment fins que s'obté un graf simple (sense llaços ni arestes múltiples). Es pot provar que aquest algorisme de mostreig genera un graf final amb la distribució uniforme a \mathcal{G}_{n,d_n} . Addicionalment, si la seqüència de graus té determinades bones propietats, el procediment aleatori és computacionalment eficient.

Una segona opció és donar-hi un enfocament combinatori. Com que estem considerant un model uniforme, per a qualsevol propietat del graf, la probabilitat que \mathbb{G}_n tingui aquesta propietat és la raó entre el nombre de grafes a \mathcal{G}_{n,d_n} que satisfan aquesta propietat i el nombre total de grafes a \mathcal{G}_{n,d_n} . Aquesta idea és força útil en models com el graf aleatori d'Erdős-Rényi $\mathbb{G}_{n,m}$, on el nombre d'arestes m es fixa com a paràmetre del model. Estimar el nombre de grafes amb una seqüència de graus donada que satisfacin una determinada propietat no és una tasca fàcil. De fet, determinar $|\mathcal{G}_{n,d_n}|$ és un problema famós en enumeració que encara no està completament resolt i que tractarem a la part final d'aquest article. Tot i això, resulta que es pot aproximar la raó entre aquests dos conjunts utilitzant un senzill truc de doble recompte.

Per tant, un dels principals obstacles en l'estudi de \mathbb{G}_n és la manca d'un bon model. Tot i que els enfocaments algorímic, probabilístic i combinatori

més simples tenen moltes limitacions, veurem com es poden modificar per permetre l'anàlisi d'aquests grafs aleatoris.

1.2 Contingut i organització

En aquest article donarem una visió general sobre els grafs aleatoris amb una seqüència de graus donada, que d'ara endavant anomenarem simplement *grafs aleatoris*.

L'article s'estructura de la manera següent. Dediquem primer la secció 2 a introduir la notació bàsica en teoria de grafs, probabilitat, asimptòtica i grafs aleatoris que farem servir al llarg de l'article. La secció 3 introdueix el model de configuració, un model molt útil per estudiar grafs aleatoris amb poques arestes. En aquesta secció parlarem de les condicions suficients sobre la seqüència de graus per tal que el graf generat sigui simple amb una probabilitat positiva. Continuem a la secció 4 establint la connexió entre els veïnatges locals en un graf aleatori i els processos de ramificació. A la secció 5 discutim l'estructura de les components connexes del graf aleatori, introduint el criteri de Molloy-Reed i cobrint temes com la percolació i la connectivitat. La secció 6 està dedicada a l'estudi de les distàncies en grafs aleatoris. Finalment, la secció 7 tracta sobre grafs aleatoris amb moltes arestes, introduint el mètode de commutació (*switching* en anglès) i discutint-ne les aplicacions a l'enumeració i al mostreig.

El nostre objectiu és proporcionar una bona intuïció sobre els resultats més rellevants en la literatura. En aquest article no donarem proves d'aquests resultats, però hi ha diverses revisions i llibres que inclouen demostracions del contingut que es presenta aquí i resultats addicionals. Consulteu, per exemple, les referències següents: model de configuració [21, 26], estructura de components connexes i distàncies [27], subgrafs [40], enumeració [52], mostreig [23].

D'altra banda, aquests grafs aleatoris han atret l'atenció de bona part de la comunitat científica per la seva versatilitat a l'hora de modelar xarxes. Actualment es fan servir per a tot tipus de xarxes en què la informació sobre els graus es pot extreure empíricament d'unes dades. La *teoria de xarxes complexes* és una branca de la física estadística que es dedica a interpretar aquestes xarxes des d'un punt de vista físic i es basa en anàlisis heurístiques i simulacions numèriques sobre models de grafs aleatoris com el que estudiarem en aquest article. Podeu consultar els articles d'Albert i Barabási [1] o de Newman [45] per entendre els grafs aleatoris des d'aquest punt de vista, així com per trobar-hi nombrosos exemples d'aplicacions concretes.

Finalment, en aquest article no ens centrarem en famílies específiques de seqüències de graus, sinó que tractarem les seqüències al nivell més general possible. Per tal de fer-ho, identificarem un conjunt comú d'hipòtesis (vegeu la hipòtesi A) sota les quals es poden enunciar la majoria dels resultats, amb l'excepció dels resultats de la secció 7, i destacarem i justificarem l'ús de condicions addicionals sempre que sigui necessari. Els lectors interessats en famílies particulars de seqüències de graus, els remetem a l'article de Wormald [51] per a seqüències regulars, i als llibres de van der Hofstad [26, 27] per a seqüències de graus que segueixen una llei de potència.

2 Notació

Abans de començar discutint els resultats sobre grafes aleatoris, introduïm la notació bàsica que farem servir en aquest article.

2.1 Grafes

Denotem amb $G = (V, E)$ un graf amb conjunt de vèrtexs $V = V(G)$ i conjunt d'arestes $E = E(G) \subseteq \binom{V}{2}$, un conjunt de parelles no ordenades d'elements de V . L'ordre del graf és $v(G) = |V(G)|$ i la seva mida, $e(G) = |E(G)|$. Al llarg de l'article fem servir «:=» per denotar una definició usant una equació. Per a $n \in \mathbb{N}$, denotem $[n] := \{1, \dots, n\}$. Tots els grafes d'aquest article tenen conjunt de vèrtexs $V(G) = [n]$ i parlem de *grafes etiquetats*, en els quals l'etiqueta d'un vèrtex és el valor $j \in [n]$ que se li assigna. En particular, $v(G) = n$ sempre que no es digui el contrari. Un *subgraf* de G és un graf $H = (V_H, E_H)$ tal que $V_H \subseteq [n]$ i $E_H \subseteq \binom{V_H}{2} \cap E$ i ho denotem amb $H \subseteq G$. El *grau* d'un vèrtex $v \in [n]$, $d(v)$, és el nombre d'arestes incidents a v en G .

Podem definir una *distància* sobre els vèrtexs de G . Donat $v, w \in [n]$, denotem amb $\text{dist}(v, w)$ la longitud del camí més curt que uneix v i w en cas que existeixi, altrament $\text{dist}(v, w) = \infty$. El *diàmetre* d'un graf, $\text{diam}(G)$, es defineix com el màxim de les distàncies entre vèrtexs de G . Un graf és *connex* si per a tot $v, w \in [n]$, $\text{dist}(v, w) < \infty$. Si un graf no és connex, aleshores té sentit parlar de les seves components connexes. La *component connexa* d'un vèrtex v és el subgraf maximal que conté tots els vèrtexs w tals que $\text{dist}(v, w) < \infty$. Denotem la component connexa més gran (en ordre) de G amb $C_1(G)$; en cas d'haver-n'hi més d'una, escollim la que conté el vèrtex amb l'etiqueta més petita.

Per a $h \in \mathbb{N}$, definim el *h-veïnatge* i el *(≤ h)-veïnatge* de v , respectivament, com

$$N_h(v) := \{w \in [n] : \text{dist}(v, w) = h\},$$

$$N_{\leq h}(v) := \{w \in [n] : \text{dist}(v, w) \leq h\}.$$

Denotarem amb $\vec{G} = (V, \vec{E})$ un *graf dirigit* (o *digraf*) amb conjunt de vèrtexs $V = [n]$ i conjunt d'arcs $\vec{E} = \vec{E}(G)$, un conjunt de parelles *ordenades* d'elements de V . El *grau d'entrada* de v , $d^-(v)$, és el nombre d'arcs $(w, v) \in \vec{E}$ amb $w \in [n]$, i el *grau de sortida* de v , $d^+(v)$, és el nombre d'arcs $(v, w) \in \vec{E}$ amb $w \in [n]$. Podem definir una distància no simètrica a \vec{G} : $\text{dist}(u, v)$ és la longitud del camí dirigit més curt que va des de v fins a w en cas que existeixi, altrament $\text{dist}(v, w) = \infty$. Definim la *component fortament connexa* de $v \in [n]$ com el subdigraf maximal que conté els vèrtexs $w \in [n]$ tal que $\text{dist}(v, w) < \infty$ i $\text{dist}(w, v) < \infty$. Denotem la component fortament connexa més gran (en ordre) de \vec{G} amb $C_1(\vec{G})$; en cas d'haver-n'hi més d'una, escollim la que conté el vèrtex amb l'etiqueta més petita.

Parlem de *multigrafes* $G = (V, E)$ per denotar un graf en què E és un multiconjunt de parelles d'elements de V no necessàriament diferents. Si tenim una

aresta $\{v, v\} \in E$, aleshores diem que aquesta aresta és un *llaç* a v . Si tenim dues, tres o quatre arestes $\{v, w\}$ a E , aleshores parlem d'una *aresta doble*, *triple* o *quàdruple*, respectivament; en general diem que és una *aresta múltiple*.

2.2 Probabilitat

Denotarem un *espai de probabilitat finita* amb $(\Omega, \mathcal{A}, \mathbb{P})$, on Ω és un conjunt finit, \mathcal{A} és una σ -àlgebra sobre Ω i $\mathbb{P}: \mathcal{A} \rightarrow \mathbb{R}$ és una funció de probabilitat. Farem servir X per denotar una *variable aleatòria* en $(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{E}[X]$ per a la seva *esperança* i $\text{Var}(X)$ per a la seva *variància*. Farem referència a alguna de les distribucions de probabilitat discreta més habituals, com la *distribució binomial* $\text{Bin}(n, p)$, on $n \in \mathbb{N}$ i $p \in [0, 1]$, i la *Poisson* $\text{Po}(\lambda)$, on $\lambda > 0$. Usarem $(X_t)_{t \geq 0}$ per denotar un *procés estocàstic* en $(\Omega, \mathcal{A}, \mathbb{P})$.

Ens caldrà considerar seqüències d'espais de probabilitat finits $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)_{n \in \mathbb{N}}$, on $|\Omega_n| \rightarrow \infty$ quan $n \rightarrow \infty$. Per tal d'alleugerir la notació és estàndard usar $\mathbb{P} = \mathbb{P}_n$ i així ho farem. Donada una seqüència d'esdeveniments $(A_n)_{n \in \mathbb{N}}$ tal que $A_n \in \mathcal{A}_n$, diem que $(A_n)_{n \in \mathbb{N}}$ se satisfà *amb alta probabilitat (aap)* si

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(A_n) = 1.$$

Donada una seqüència $(X_n)_{n \in \mathbb{N}}$, on X_n és una variable aleatòria definida a $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ i X és una variable aleatòria definida a $(\Omega, \mathcal{A}, \mathbb{P})$, diem que X_n *convergeix a X en distribució* (o $X_n \xrightarrow{d} X$) si

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(X_n \leq x) = \mathbb{P}(X \leq x), \quad \text{per a tot } x \in \mathbb{R}.$$

D'altra banda, donada una constant $c \in \mathbb{R}$, diem que X_n *convergeix a c en probabilitat* (o $X_n \xrightarrow{p} c$) si per a tot $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(|X_n - c| \geq \epsilon) = 0.$$

2.3 Notació asimptòtica

Per a $f, g: \mathbb{N} \rightarrow \mathbb{N}$, diem que $f(n) = O(g(n))$ si

$$\limsup_{n \rightarrow \infty} \frac{f(n)}{g(n)} < \infty,$$

i que $f(n) = \Theta(g(n))$ si $f(n) = O(g(n))$ i $g(n) = O(f(n))$.

Diem que $f(n) = o(g(n))$ si

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0,$$

en particular $f(n) = o(1)$ denota que $\lim_{n \rightarrow \infty} f(n) = 0$.

Finalment, diem que $f(n) \sim g(n)$ si

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1.$$

Quan parlem de resultats asimptòtics sobre seqüències d'espais de probabilitat, farem servir expressions de l'estil « $f(n) = O(g(n))$ aap»; formalment això cal entendre-ho com: existeix una constant $c > 0$ tal que

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(f(n) \leq cg(n)) = 1,$$

i similarment per a l'altra notació asimptòtica.

2.4 Grafes aleatoris

Una *seqüència de graus* de longitud n és una seqüència $\mathbf{d}_n = (d_i)_{i \in [n]}$, on d_i és un número enter positiu¹ amb $d_i \leq n - 1$ per a tot $i \in [n]$ i $m_n := \sum_{i \in [n]} d_i$ és parell. Definim $\Delta_n := \max_{i \in [n]} d_i$.

Redefinim $\mathcal{G}_{n, \mathbf{d}_n}$ com el conjunt de grafes amb conjunt de vèrtexs $[n]$ i on el vèrtex i té grau $d(i) = d_i$. Una seqüència de graus \mathbf{d}_n és *factible* si $\mathcal{G}_{n, \mathbf{d}_n} \neq \emptyset$. Donada una seqüència de graus factible \mathbf{d}_n , sigui $\mathbb{G}_{n, \mathbf{d}_n}$ un graf triat uniformement a l'atzar de $\mathcal{G}_{n, \mathbf{d}_n}$. Anomenem $\mathbb{G}_{n, \mathbf{d}_n}$ un *graf aleatori amb seqüència de graus \mathbf{d}_n* . Quan la seqüència de graus a la qual fem referència sigui clara pel context, escriurem simplement $\mathbb{G}_n = \mathbb{G}_{n, \mathbf{d}_n}$ i ens hi referirem com a *graf aleatori*. Donat $d = d(n)$ amb $1 \leq d \leq n - 1$, un *graf aleatori d -regular*, denotat per $\mathbb{G}_{n, d}$, és el graf aleatori \mathbb{G}_n on $d_i = d$ per a tot $i \in [n]$.

Una *seqüència de seqüències de graus* és una seqüència $\mathbf{d} = (\mathbf{d}_n)_{n \in \mathbb{N}}$, on per a cada $n \in \mathbb{N}$, $\mathbf{d}_n = (d_{n,i})_{i \in [n]}$ és una seqüència de graus de longitud n . Definim $n_k = n_{n,k} := |\{i \in [n] \mid d_{n,i} = k\}|$ com el nombre de vèrtexs de grau k en \mathbf{d}_n . Donada \mathbf{d} , definim la *seqüència de grafes aleatoris* $\mathbb{G}_{\mathbf{d}} = (\mathbb{G}_{n, \mathbf{d}_n})_{n \in \mathbb{N}}$.

Al llarg de l'article, compararem el comportament del graf aleatori \mathbb{G}_n amb el del *graf aleatori binomial*, $\mathbb{G}_{n,p}$, el qual té el conjunt de vèrtexs $[n]$ i en què cada aresta s'afegeix independentment amb probabilitat $p \in [0, 1]$. Sovint considerarem probabilitats $p = p_n \rightarrow 0$ quan $n \rightarrow \infty$. Finalment, per a $m \in \{0, \dots, \binom{m}{2}\}$, introduïm el *graf aleatori d'Erdős-Rényi*, $\mathbb{G}_{n,m}$, que és un graf triat uniformement de tots els grafes amb conjunt de vèrtexs $[n]$ i m arestes. A la literatura, sovint s'anomena $\mathbb{G}_{n,p}$ el graf aleatori d'Erdős-Rényi, fet que pot comportar una lleu confusió. Tot i que les definicions són diferents, si $m = \binom{n}{2}p$, $\mathbb{G}_{n,p}$ i $\mathbb{G}_{n,m}$ es comporten de manera molt similar, d'aquí l'abús de notació.

¹ En general, les seqüències de graus poden contenir vèrtexs de grau zero. Tot i això, no tenen cap paper en l'estructura del graf aleatori i es poden eliminar de la seqüència de graus reduint a la vegada n , d'aquí l'assumpció.

3 El model de configuració

A la subsecció 1.1 hem proposat el següent algorisme per generar grafs amb la seqüència de graus donada: començant amb el graf buit, escollim dos vèrtexs a l'atzar que no tinguin el seu grau saturat i afegim una aresta entre ells. Aquest procediment sempre acaba, tot i que pot generar grafs que no són simples. Tot i això, la idea d'aparellar vèrtexs de dos en dos és molt útil, i és la base principal del model de configuració.

El *model de configuració* (també conegut com a *model d'aparellament*) amb seqüència de graus \mathbf{d}_n , denotat per $\mathbb{CM}_{n,\mathbf{d}_n}$ (o simplement \mathbb{CM}_n si la seqüència de graus és clara pel context) i introduït per Bollobás a [5] (vegeu també [3, 4, 49, 50]), és el multigraf aleatori amb conjunt de vèrtexs $[n]$ que s'obté assignant d_i *semiaarestes* a cada vèrtex $i \in [n]$, i després aparellant-les totes uniformement a l'atzar per crear les arestes (vegeu la figura 1). El nombre d'aparellaments que es poden realitzar és $m_n!! := m_n(m_n - 2)(m_n - 4) \dots 2$.

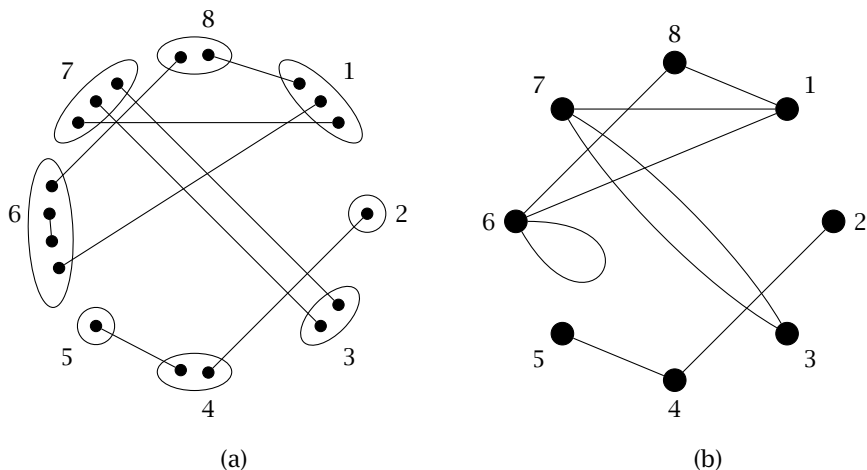


FIGURA 1: Exemple del model de configuració amb $\mathbf{d}_n = (3, 1, 2, 2, 1, 4, 3, 2)$: (a) aparellament de les semiaarestes, representades com a punts; (b) multigraf resultant.

Observeu que una semiaresta incident a v es pot aparellar amb una altra semiaresta també incident a v , i produir així un llaç en aquest vèrtex. De manera similar, més d'una semiaresta incident a u es pot aparellar amb semiarestes incidents a v , i produir una aresta múltiple. A la figura 1, aquest és el cas per al vèrtex 6 (llaç) i entre els vèrtexs 3 i 7 (aresta doble). Per aquesta raó, en general, el model de configuració no dona lloc a un graf simple, sinó a un multigraf. No obstant això, proporciona un model fàcilment analitzable que genera exemples amb la seqüència de graus correcta.

Donat un multigraf G amb conjunt de vèrtexs $[n]$ i seqüència de graus \mathbf{d}_n , hi ha molts aparellaments que el generen. Concretament,

$$\mathbb{P}(\mathbb{CM}_n = G) = \frac{\prod_{i \in [n]} d_i!}{(m_n - 1)!! \prod_{i \in [n]} 2^{a_i} a_i! \prod_{1 \leq i < j \leq n} b_{i,j}!}, \quad (1)$$

on a_i és el nombre de llaços incidents al vèrtex i , i $b_{i,j}$ és el nombre d'arestes entre els vèrtexs i i j . Fem, doncs, el càlcul del nombre d'aparellaments que donen lloc a G . Fixem un ordre arbitrari de les arestes de G . Llavors podem ordenar les semiaarestes incidents al vèrtex i de $d_i!$ maneres i aparellar-les segons aquest ordre i l'ordre de les arestes. A causa dels llaços i de les arestes múltiples, un aparellament particular apareixerà $\prod_{i \in [n]} 2^{a_i} a_i! \prod_{1 \leq i < j \leq n} b_{i,j}!$ vegades. Finalment, la probabilitat d'aparellar les dues semiaarestes correctes al pas t és $\frac{1}{m_n - 2t + 1}$ i d'aquí s'obté la fórmula (1).

Observeu que, donada \mathbf{d}_n , la probabilitat en (1) només depèn del nombre de llaços i arestes múltiples de G , de manera que la probabilitat de qualsevol graf simple és la mateixa. Així doncs, \mathbb{CM}_n condicionat a l'esdeveniment de ser simple i \mathbb{G}_n tenen la mateixa llei de probabilitat, és a dir,

$$\mathbb{P}(\mathbb{CM}_n = G \mid \mathbb{CM}_n \text{ és simple}) = \mathbb{P}(\mathbb{G}_n = G). \tag{2}$$

Aquesta és l'observació clau que ens permet estudiar grafs aleatoris utilitzant el model de configuració. Per a qualsevol propietat \mathcal{P} de grafs simples,

$$\mathbb{P}(\mathbb{G}_n \in \mathcal{P}) = \mathbb{P}(\mathbb{CM}_n \in \mathcal{P} \mid \mathbb{CM}_n \text{ és simple}) = \frac{\mathbb{P}(\mathbb{CM}_n \in \mathcal{P})}{\mathbb{P}(\mathbb{CM}_n \text{ és simple})} \tag{3}$$

i, entenen la probabilitat que el model de configuració produeixi un graf simple, podem transferir les probabilitats del model de configuració als grafs aleatoris.

Abans de donar una expressió per a la probabilitat que \mathbb{G}_n sigui simple, parlem breument de la distribució de llaços i arestes múltiples. Siguin L_n i M_n variables aleatòries que compten, respectivament, el nombre de llaços i de parelles d'arestes dobles que no són llaços a \mathbb{CM}_n . Podem escriure $L_n := \sum_{i \in [n]} \sum_{1 \leq r < s \leq d_i} X_{i,r,s}$, on $X_{i,r,s}$ és la variable aleatòria indicadora que la r -èsima i la s -èsima semiaarestes incidents al vèrtex i s'han aparellat, d'on es desprèn que $\mathbb{E}[X_{i,r,s}] = \frac{1}{m_n - 1}$. Es pot obtenir una expressió similar per a M_n . Per linealitat de l'esperança,

$$\begin{aligned} \mathbb{E}[L_n] &= \frac{1}{2(m_n - 1)} \sum_{i \in [n]} d_i(d_i - 1) = \frac{\nu_n}{2} (1 + O(1/m_n)), \\ \mathbb{E}[M_n] &= \frac{1}{4(m_n - 1)(m_n - 3)} \sum_{i,j \in [n], i \neq j} d_i(d_i - 1)d_j(d_j - 1) \\ &= \frac{1}{4} (\nu_n^2 - \kappa_n) (1 + O(1/m_n)), \end{aligned} \tag{4}$$

on

$$\begin{aligned} \nu_n &:= \frac{1}{m_n} \sum_{i \in [n]} d_i(d_i - 1), \\ \kappa_n &:= \frac{1}{m_n^2} \sum_{i \in [n]} d_i^2(d_i - 1)^2. \end{aligned}$$

Els paràmetres ν_n i κ_n es poden expressar com a combinació lineal de moments de la seqüència de graus. Com veurem més endavant, ν_n és un paràmetre clau, no tan sols per controlar el nombre de llaços i arestes dobles, sinó també per entendre l'estructura de components i les distàncies dins del graf aleatori. D'altra banda, κ_n juga un paper més secundari i, per a la gran majoria de seqüències de graus, es té que $\kappa_n = o(\nu_n)$. Arribats a aquest punt, és útil establir algunes condicions sobre la seqüència de graus considerada. Donada \mathbf{d}_n , definim la *distribució de graus* $D_n = D_n(\mathbf{d}_n)$ com la variable aleatòria sobre \mathbb{N} que indica el grau d'un vèrtex triat uniformement a l'atzar, és a dir,

$$\mathbb{P}(D_n = k) = \lambda_{n,k} := \frac{n_{n,k}}{n}, \quad \text{per a tot } k \geq 1. \quad (5)$$

Enunciem ara, doncs, les nostres hipòtesis sobre la seqüència de graus que suposarem per a la majoria dels resultats presentats en aquest article.

HIPÒTESI A. *Existeix una distribució de probabilitat $D = D(\mathbf{d})$ sobre \mathbb{N} amb $\lambda_k := \mathbb{P}(D = k)$ per a tot $k \geq 1$ i*

1. $D_n \xrightarrow{d} D$, és a dir, $\lim_{n \rightarrow \infty} \lambda_{n,k} = \lambda_k$ per a tot $k \geq 1$;
2. $\lim_{n \rightarrow \infty} \mathbb{E}[D_n]$ existeix, és finit i és igual a $\lambda := \mathbb{E}[D]$;
3. $\lim_{n \rightarrow \infty} \mathbb{E}[D_n^2]$ existeix, és finit i és igual a $\mathbb{E}[D^2]$.

OBSERVACIÓ 1 (DISCUSSIÓ SOBRE LA HIPÒTESI A). La primera hipòtesi ens permet descriure \mathbb{CM}_n en termes dels paràmetres límit. La segona hipòtesi restringeix el nombre d'arestes, en particular, $m_n = \Theta(n)$. Finalment, la tercera hipòtesi controla la variància de la distribució dels graus. Definim

$$\nu := \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} = \frac{1}{\lambda} \sum_{k \geq 1} k(k-1)\lambda_k.$$

La hipòtesi implica que

$$\lim_{n \rightarrow \infty} \nu_n = \nu \in [0, \infty), \quad (6)$$

i que el grau màxim compleix

$$\Delta_n = o(\sqrt{n}).$$

OBSERVACIÓ 2 (HIPÒTESIS A LA LITERATURA). Cada resultat sobre els grafs aleatoris a la literatura utilitza el seu propi conjunt d'hipòtesis sobre la seqüència de graus donada, sovint sobre característiques com la suavitat, la convergència (uniforme) o l'acotació de certs paràmetres com el grau màxim i el grau mínim. Aquestes hipòtesis acostumen a ser tècniques i complicades d'enunciar. Tanmateix, una gran part dels resultats se satisfan sota hipòtesis molt similars a la hipòtesi A. De fet, les nostres condicions són essencialment les condicions més febles sobre \mathbf{d} sota les quals $\mathbb{CM}_{\mathbf{d}}$ té una descripció local simple (vegeu la secció 4). Per tant, escollim la hipòtesi A com un marc comú per a

tots els resultats donats en aquest article. Eventualment, alguns dels resultats requeriran condicions addicionals; en tal cas les destacarem i en justificarem la necessitat. En altres casos, adaptarem alguns resultats a seqüències de graus més enllà de les cobertes per les nostres hipòtesis principals. En particular, la secció 7 estarà dedicada a seqüències de graus sense restriccions en el segon moment.

Tornant a la probabilitat que el model de configuració sigui simple, sota la hipòtesi A, el nombre esperat de llaços i arestes múltiples és constant: de (4) en deduïm

$$\mathbb{E}[L_n] \sim \nu/2 \quad \text{i} \quad \mathbb{E}[M_n] \sim \nu^2/4,$$

on hem utilitzat (6), i que $\kappa_n = o(1)$, ja que $\nu_n = O(1)$. Per a n prou gran, els llaços i les arestes dobles probablement apareixeran lluny els uns dels altres, mostrant correlacions febles que s'esvaniran en el límit. Les variables aleatòries que compten el nombre d'aparicions independents d'esdeveniments amb poca probabilitat segueixen distribucions de Poisson. Així doncs, en el límit, L_n i M_n esdevenen variables aleatòries de Poisson independents amb paràmetres $\nu/2$ i $\nu^2/4$, respectivament.

Un llaç es pot entendre com un cicle de longitud 1, i una aresta doble, com un cicle de longitud 2. Donem, doncs, el següent resultat més general per a la distribució de cicles curts en el model de configuració (vegeu [26] per una demostració usant el mètode dels moments).

TEOREMA 3. *Suposem que \mathbf{d} satisfà la hipòtesi A. Per a $k \in \mathbb{N}$, sigui $X_{n,k}$ el nombre de cicles de longitud k en \mathbb{CM}_n . Aleshores, per a qualsevol seqüència finita k_1, \dots, k_l d'enters no negatius, el vector aleatori $(X_{n,k_1}, \dots, X_{n,k_l})$ convergeix en distribució a un vector de variables de Poisson independents amb paràmetres $\xi_{k_i} := \nu^{k_i}/2k_i$, respectivament.*

Com a conseqüència del teorema anterior i (3), obtenim el resultat següent.

TEOREMA 4. *Suposem que \mathbf{d} satisfà la hipòtesi A. Llavors,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{CM}_n \text{ és simple}) = e^{-\frac{\nu}{2} - \frac{\nu^2}{4}} > 0.$$

En particular, qualsevol esdeveniment que es compleixi amb alta probabilitat per a \mathbb{CM}_n també es complirà amb alta probabilitat per a \mathbb{G}_n .

La demostració original d'aquest resultat és de Bender i Canfield [4] per a $\Delta_n = O(1)$ i de Bollobás [5] per a $\Delta_n \leq \sqrt{2 \log n} - 1$. La demostració del teorema sota les nostres hipòtesis és de Janson [31].

Si X és una variable de Poisson amb paràmetre $\lambda > 0$, la probabilitat que $X = 0$ és $e^{-\lambda}$. La probabilitat que \mathbb{CM}_n sigui simple és equivalent a $L_n = 0$ i $M_n = 0$. La primera part del teorema encaixa amb el fet que, en el límit, L_n i M_n són variables aleatòries de Poisson independents amb paràmetres $\nu/2$ i $\nu^2/4$, respectivament. La segona part justifica l'ús del model de configuració per a estudiar les propietats típiques dels grafs aleatoris.

Al llarg d'aquest article, ens centrarem en resultats sobre el graf aleatori \mathbb{G}_n que passen amb alta probabilitat. La gran majoria són demostrats primer pel model de configuració \mathbb{CM}_n i després transferits a \mathbb{G}_n utilitzant el teorema 4.

OBSERVACIÓ 5 (GRAFS ALEATORIS BINOMIALS). El grau d'un vèrtex en $\mathbb{G}_{n,p}$ es distribueix com una distribució binomial $\text{Bin}(n-1, p)$. Si $p = p_n = \lambda/n$ per a $\lambda \in (0, \infty)$, llavors el grau convergeix en distribució a una Poisson $\text{Po}(\lambda)$. Observeu que la seqüència de graus de $\mathbb{G}_{n,p}$ és una seqüència aleatòria i, per tant, $\mathbb{G}_{n,p}$ no és un cas particular de $\mathbb{G}_{n,\mathbf{d}_n}$, en què la seqüència de graus \mathbf{d}_n està fixada *a priori*. Tot i això, la seqüència de seqüències de graus de $\mathbb{G}_{n,p}$ satisfà la hipòtesi A amb alta probabilitat. Per tant, qualsevol esdeveniment que es compleixi amb alta probabilitat en \mathbb{CM}_n també es complirà amb alta probabilitat en $\mathbb{G}_{n,p}$. Aquest argument es pot estendre als *grafs aleatoris generalitzats* (vegeu, per exemple, [26, secció 7]).

4 Aproximació per processos de ramificació

Assumint la hipòtesi A i segons el teorema 3, esperem un nombre constant de cicles de longitud acotada. Així doncs, amb alta probabilitat, el veïnatge d'un vèrtex aleatori d'una profunditat constant és un arbre construït de manera aleatòria. Aquesta és la principal motivació per utilitzar processos de ramificació. Aquesta secció conté una breu introducció als processos de ramificació i s'explica com es poden acoblar amb l'exploració del veïnatge d'un vèrtex a \mathbb{CM}_n .

4.1 Processos de ramificació

Considerem una variable aleatòria ξ sobre els enters no negatius i siguin $(\xi_{i,t})_{i \geq 1, t \geq 0}$ còpies independents i idènticament distribuïdes de ξ . El *procés de ramificació* $(X_t)_{t \geq 0}$ amb distribució de descendència ξ (també conegut com a *arbre de Bienaymé-Galton-Watson*) es defineix com $X_0 = 1$ i, si $t \geq 0$,

$$X_{t+1} := \sum_{i=1}^{X_t} \xi_{i,t}.$$

Els processos de ramificació solen representar-se com a arbres arrelats. Per a qualsevol $t \geq 0$ i qualsevol element i comptat per X_t , li assignem $\xi_{i,t}$ fills, que representen els elements comptats per X_{t+1} . Aleshores, la t -èsima generació del procés és el conjunt d'elements a distància t de l'arrel, que té mida X_t . En algunes de les aplicacions d'aquest article, l'arrel del procés tindrà una distribució de descendència diferent de la resta d'elements del procés; en tal cas ho destacarem.

Definim l'*esdeveniment d'extinció* i la *probabilitat d'extinció*, respectivament, com

$$\mathcal{E} := \{\exists t \geq 0 : X_t = 0\} \quad \text{i} \quad \eta := \mathbb{P}(\mathcal{E}). \quad (7)$$

Observeu que si $\eta = 1$, llavors el procés s'extingeix quasi segurament (amb probabilitat 1), mentre que si $\eta < 1$, hi ha una probabilitat positiva que el procés sobrevisqui. Un dels resultats més notoris sobre els processos de ramificació en determina la probabilitat d'extinció.

TEOREMA 6. *Sigui $(X_t)_{t \geq 0}$ un procés de ramificació que té distribució de descendència ξ . Aleshores η és la solució no negativa més petita de $z = \sum_{k \geq 0} \mathbb{P}(\xi = k)z^k$. En particular:*

- (i) si $\mathbb{E}[\xi] \leq 1$ i $\mathbb{P}(\xi = 1) < 1$, llavors $\eta = 1$;
- (ii) si $\mathbb{E}[\xi] > 1$ o $\mathbb{P}(\xi = 1) = 1$, llavors $\eta < 1$.

Descartant el cas degenerat $\mathbb{P}(\xi = 1) = 1$, el teorema estableix que, si el nombre esperat de descendents de cada vèrtex és com a màxim 1, llavors el procés s'extingeix quasi segurament.

La *distribució conjugada* de ξ , denotada per $\hat{\xi}$, es defineix com

$$\mathbb{P}(\hat{\xi} = k) := \eta^{k-1} \mathbb{P}(\xi = k), \quad \text{per a tot } k \geq 0. \quad (8)$$

Les distribucions conjugades s'utilitzen per descriure els processos de ramificació condicionats a l'extinció.

LEMA 7. *Sigui $(X_t)_{t \geq 0}$ un procés de ramificació amb distribució de descendència ξ que compleix $\eta > 0$. Condicionat a \mathcal{E} , $(X_t)_{t \geq 0}$ té la mateixa distribució que un procés de ramificació amb distribució de descendència $\hat{\xi}$.*

Aquest lema es pot interpretar de la següent manera: com que la probabilitat d'extinció η és positiva, pot ser que el procés s'extingeixi. Aquest esdeveniment es defineix de manera global i, per tant, *a priori*, el procés condicionat a l'extinció podria tenir una llei de probabilitat complicada. El resultat diu que aquest no és el cas; la seva descripció és molt senzilla ja que no deixa de ser un procés de ramificació amb una altra distribució de descendència. En particular, manté totes les propietats d'independència intrínseques d'aquests processos.

Les demostracions dels resultats d'aquesta part es poden trobar, per exemple, a [26, secció 3].

4.2 Acoblament amb el veïnatge d'un vèrtex

El veïnatge local d'un vèrtex en el model de configuració pot ser acoblat amb un procés de ramificació amb distribució provinent de la seqüència de graus, de manera que l'acoblament tingui èxit amb alta probabilitat.

Recordeu la definició de veïnatge a la secció 2. El nostre objectiu és estudiar el procés estocàstic $(|N_t(v)|)_{0 \leq t \leq h}$, per al qual necessitem introduir un procés d'exploració de $\mathbb{C}\mathbb{M}_n$. Això ens permetrà estudiar aquest model a través de l'estudi dels processos de ramificació.

Donada una regla de prioritat \mathcal{R} sobre les semiarestes, considerem el següent procés que genera $\mathbb{C}\mathbb{M}_n$:

- (i) triem una semiaresta sense emparellar e (si n'hi ha alguna) segons \mathcal{R} ;
- (ii) triem una semiaresta sense emparellar $f \neq e$ uniformement a l'atzar;
- (iii) emparellem e i f , i prosseguim.

Tot i que aquest procediment genera el model de punt, és senzill d'analitzar, ja que el conjunt de semiarestes sense emparellar en un cert moment és aleatori i depèn de les eleccions fetes anteriorment. En lloc d'aquest procés inicial, considerem l'objecte \mathbb{CM}_n^* generat pel procediment anterior substituint (ii) pel pas següent:

- (ii*) triem una semiaresta f uniformement a l'atzar entre *totes* les semiarestes.

Cal observar que en \mathbb{CM}_n^* les semiarestes es poden emparellar més d'un cop, de manera que l'objecte resultant és un conjunt de parelles de semiarestes (no necessàriament diferents). En aquest objecte hi podem definir una distància,² i té sentit considerar els h - i ($\leq h$)-veïnatges de v en \mathbb{CM}_n^* .

Per determinar $N_{\leq h}(v)$, executem el procés d'exploració començant a v i amb la regla de prioritat donada per l'ordre *BFS* (*breadth first search*): en cada moment triem una de les semiarestes e més properes a v que encara no hagi estat emparellada, i si totes les semiarestes estan a distància infinita de v , triem llavors la que tingui l'etiqueta més petita. Com que només necessitem generar el veïnatge de v fins a una profunditat h , aturarem el procés quan totes les semiarestes sense emparellar estiguin a una distància més gran o igual que h de v .

Definim $\mathcal{A}_h(v)$ com l'esdeveniment en \mathbb{CM}_n^* que, fins al pas que determina completament $N_{\leq h}(v)$, cap semiaresta emparellada ni la semiaresta e seleccionada al pas (i) no s'han seleccionat al pas (ii*). Condicionats a $\mathcal{A}_h(v)$, \mathbb{CM}_n i \mathbb{CM}_n^* tenen la mateixa distribució i podem considerar l'acoblament natural entre els dos.

Definim $\mathcal{B}_h(v)$ com l'esdeveniment en \mathbb{CM}_n (o en \mathbb{CM}_n^*) en què no s'ha creat cap cicle fins al pas que determina $N_{\leq h}(v)$. Si $\mathcal{B}_h(v)$ es compleix, $N_{\leq h}(v)$ indueix un arbre a \mathbb{CM}_n (o a \mathbb{CM}_n^*).

En aquest punt podem relacionar l'exploració del veïnatge local de v a \mathbb{CM}_n amb el procés de ramificació corresponent. Sigui U el vèrtex (aleatori) incident a f , la semiaresta que s'ha triat al pas (ii*). Com que triem f del conjunt de *totes* les semiarestes, per a tot $u \in [n]$, la probabilitat que $U = u$ és d_u/m_n . En particular, la probabilitat que U tingui grau k és

$$\mathbb{P}(d_U = k) = \sum_{u:d_u=k} \frac{d_u}{m} = k \cdot \frac{n_{n,k}}{m} =: \mathbb{P}(D_n^* = k),$$

on D_n^* és coneguda com la *distribució esbiaixada per la mida* de D_n , i D_n es defineix com a (5).

Com que v està fixat, $|N_1(v)| = d_v$. Sigui $2 \leq t < h$ i suposem que som al pas just després que $N_{t-1}(v)$ hagi estat determinat. El nombre de semiarestes

² La distància entre v i w en \mathbb{CM}^* es defineix com la longitud de la seqüència més curta de parelles de semiarestes tal que cada dues parelles consecutives tenen dues semiarestes incidents a un mateix vèrtex, la primera parella és incident a v i l'última és incident a w .

incidents a aquest conjunt i encara no aparellades és mesurable respecte a la història actual del procés d'exploració, mentre que $N_t(v)$ encara és aleatori. Si es compleix $\mathcal{A}_h(v) \cap \mathcal{B}_h(v)$, hi ha $|N_t(v)|$ semiarestes incidents a $N_{t-1}(v)$ (ja que no hi ha cicles curts) i cadascuna d'elles s'aparellarà a una semiaresta aleatòria, incident a un vèrtex el grau del qual es distribueix segons D_n^* . Com que no hi ha cicles, tots aquests vèrtexs són diferents i

$$|N_{t+1}(v)| = \sum_{i=1}^{|N_t(v)|} \xi_{i,t},$$

on $\xi_{i,t}$ són variables aleatòries independents amb distribució $D_n^* - 1$. Aquí, el -1 descompta la semiaresta adjacent a cada vèrtex de $N_t(v)$ que connecta amb $N_{t-1}(v)$.

Per tant, si es compleix $\mathcal{A}_h(v) \cap \mathcal{B}_h(v)$, $(|N_t(v)|)_{0 \leq t \leq h}$ es pot acoblar amb el procés $(X_t)_{0 \leq t \leq h}$, que compleix $X_1 = d_v$.

Resta calcular la probabilitat que qualsevol dels esdeveniments falli. Com que la probabilitat de triar f incident amb u en un pas donat és d_u/m_n i això produeix un increment de $(d_u - 1)$ semiarestes noves que podrien fer que l'acoblament fallés, el nombre esperat de semiarestes conflictives³ afegides a cada pas és

$$\mathbb{E}[D_n^* - 1] = \sum_{i=1}^n \frac{d_i(d_i - 1)}{m_n} = v_n,$$

que està acotat per la hipòtesi A. Per tant, en el nombre de passos que es requereixen per revelar el veïnatge de v fins a una profunditat constant, el nombre esperat de semiarestes conflictives és constant i és improbable que aquestes siguin triades com a semiarestes f durant el procés. Aquestes són les idees principals que hi ha darrere del resultat següent, que determina asimptòticament la llei de la mida del veïnatge local d'un vèrtex.

LEMA 8 (PRIMER LEMA D'ACOBLEMENT). *Suposem que \mathbf{d} satisfà la hipòtesi A. Sigui $v \in [n]$. Considerem $(X_t)_{t \geq 0}$ un procés de ramificació amb descendència $D_n^* - 1$ que compleix $X_1 = d_v$. Per a qualsevol $h \in \mathbb{N}$, hi ha un acoblament entre el procés d'exploració de \mathbb{CM}_n començant a v i $(X_t)_{t \geq 0}$, de manera que*

$$\mathbb{P}((|N_t(v)|)_{0 \leq t \leq h} = (X_t)_{0 \leq t \leq h}) = 1 - O(1/n). \tag{9}$$

Cal observar que la probabilitat que l'acoblament falli és massa gran per demostrar que tots els veïnatges de profunditat acotada h són arbres. Això encaixa bé amb el teorema 3, que implica que el nombre de cicles curts és positiu amb probabilitat positiva.

El lema anterior descriu el veïnatge d'un vèrtex v fixat *a priori*, però es pot adaptar al veïnatge d'un vèrtex V_n triat uniformement a l'atzar de $[n]$. En aquest últim cas, X_1 segueix la distribució D_n , i, per tant, definim els processos següents.

³ Les semiarestes conflictives són aquelles incidents a vèrtexs que tenen una semiaresta incident ja emparellada, ja que aquestes podrien donar lloc a cicles o ser reaparellades a \mathbb{CM}_n^* .

DEFINICIÓ 9. Sigui $(Y_{n,t})_{t \geq 0}$ el procés de ramificació amb distribució de descendència $D_n^* - 1$, on l'arrel té distribució de descendència D_n . Sigui $(Y_t)_{t \geq 0}$ el procés de ramificació amb distribució de descendència $D^* - 1$, on l'arrel té distribució de descendència D . Aquí D^* denota la distribució esbiaixada per la mida de D , és a dir,

$$\mathbb{P}(D^* = k) := \frac{k\lambda_k}{\lambda}, \quad \text{per a tot } k \geq 1,$$

que en particular satisfà

$$\mathbb{E}[D^* - 1] = \nu. \quad (10)$$

LEMA 10 (SEGON LEMA D'ACOBLEMENT). *Suposem que \mathbf{d} satisfà la hipòtesi A i sigui $(Y_{n,t})_{t \geq 0}$ com en la definició 9. Per a qualsevol $h \in \mathbb{N}$, hi ha un acoblament entre el procés d'exploració de $\mathbb{C}\mathbb{M}_n$ que comença en un vèrtex V_n escollit uniformement a l'atzar i $(Y_{n,t})_{t \geq 0}$, de manera que*

$$\mathbb{P}((|N_t(V_n)|)_{0 \leq t \leq h} = (Y_{n,t})_{0 \leq t \leq h}) = 1 - O(1/n). \quad (11)$$

OBSERVACIÓ 11. Al lema 10 podem substituir $(Y_{n,t})_{t \geq 0}$ per $(Y_t)_{t \geq 0}$ a costa de reemplaçar $1 - O(1/n)$ per $1 - o(1)$, fent més feble la conclusió. De fet, la probabilitat que l'acoblament falli depèn de la velocitat de convergència en les condicions de la hipòtesi A.

Els lemes d'acoblament anteriors es poden estendre si $h = h_n \rightarrow \infty$ quan $n \rightarrow \infty$, sempre que el $(\leq h_n)$ -veïnatge tingui mida $o(\sqrt{n})$; en aquest cas, la probabilitat d'èxit a (9) i (11) s'ha de reemplaçar per $1 - o(1)$.

5 Estructura de les components connexes

Un dels problemes fonamentals en grafs aleatoris és descriure la seva estructura de components connexes en funció dels valors dels seus paràmetres. De fet, una de les aplicacions principals dels grafs aleatoris és la modelització de fenòmens estudiats per la física estadística, especialment en relació amb les transicions de fase. Aquest fenomen té lloc quan un sistema físic experimenta un canvi abrupte i qualitatiu en resposta a petites variacions en un paràmetre extern, com ara la temperatura o la pressió. Un exemple d'aquesta transició de fase es pot observar en el pas entre els diferents estats de l'aigua, els quals es poden modelar mitjançant grafs aleatoris diferents; vegeu, per exemple, la revisió de Duminil-Copin [15]. En aquests grafs, els paràmetres externs es representen a través dels paràmetres del model.

En aquesta secció ens endinsem en l'estudi de la transició de fase per la propietat de l'existència d'una component connexa macroscòpica en el graf, en contrast amb un graf altament fragmentat. El resultat fundacional i més important en la teoria de grafs aleatoris va ser provat per Erdős i Rényi [16]: essent $C_1(\mathbb{G}_{n,p})$ la component més gran del graf aleatori binomial $\mathbb{G}_{n,p}$, si $p = p(n) = \lambda/n$ per a $\lambda > 0$, tenim el següent:

- si $\lambda < 1$, llavors $\nu(C_1(\mathbb{G}_{n,p})) = \Theta(\log n)$ aap;
- si $\lambda = 1$, llavors $\nu(C_1(\mathbb{G}_{n,p})) = \Theta(n^{2/3})$ aap;
- si $\lambda > 1$, llavors $\nu(C_1(\mathbb{G}_{n,p})) = \Theta(n)$ aap.

Aquestes tres fases són conegudes, respectivament, com els règims *subcrític*, *crític* i *supercrític*. En el règim supercrític, la component connexa d'ordre lineal és coneguda com la *component gegant* i totes les altres components tenen ordre $O(\log n)$. Aquest resultat es coneix informalment com el *doblet salt*, degut a l'evolució de l'ordre de la component més gran al voltant de $p = 1/n$.

Un fenomen similar emergeix en els grafs aleatoris amb una seqüència de graus donada. En el nostre cas, el paràmetre del model $\mathbb{G}_{n,\mathbf{d}_n}$ és la seqüència de graus \mathbf{d}_n . Així doncs, l'objectiu principal d'aquesta secció és determinar les seqüències de graus \mathbf{d}_n per a les quals $\nu(C_1(\mathbb{G}_n)) = o(n)$ aap, i aquelles per a les quals $\nu(C_1(\mathbb{G}_n)) = \Theta(n)$ aap.

5.1 El criteri de Molloy-Reed

En aquesta secció presentem un criteri per determinar l'existència d'una component gegant que és vàlid sota la hipòtesi A. Abans de presentar-lo, donem una intuïció per entendre'l.

Segui V_n un vèrtex triat uniformement a l'atzar de $[n]$. Pel lema 10 i l'observació 11, aap el veïnatge local de V_n es comporta com les primeres generacions de $(Y_t)_{t \geq 0}$. La sub/supercriticalitat d'aquest procés no depèn de la distribució de descendència de l'arrel, que és D , sinó només de la distribució de descendència dels altres elements, que és $D^* - 1$ i que té esperança ν segons l'equació (10). Pel teorema 6, el valor de ν determina si el procés s'extingeix quasi segurament o si té probabilitat positiva de sobreviure. Aquesta és la intuïció que hi ha darrere del resultat següent.

TEOREMA 12 (CRITERI DE MOLLOY-REED). *Suposem que \mathbf{d} satisfà la hipòtesi A. Aleshores,*

- (i) *si $\nu < 1$, llavors $\nu(C_1(\mathbb{G}_n)) = o(n)$ aap;*
- (ii) *si $\nu > 1$, llavors $\nu(C_1(\mathbb{G}_n)) = \Theta(n)$ aap.*

La primera demostració d'aquest teorema és de Molloy i Reed [43], però requereix condicions més fortes sobre la seqüència de graus. El resultat com es presenta aquí és conseqüència d'un resultat de Janson i Luczak [32] que discutirem amb detall més endavant.

OBSERVACIÓ 13. A la literatura també s'utilitzen els paràmetres

$$Q_n := \frac{1}{m} \sum_{i \in [n]} d_i(d_i - 2) = \nu_n - 1, \quad Q := \frac{1}{\lambda} \sum_{k \geq 1} k(k - 2)\lambda_k = \nu - 1.$$

Mentre que ν_n és aproximadament el nombre esperat de nous veïns de cada vèrtex en les primeres etapes del procés d'exploració, Q_n es pot interpretar

de la manera següent. Sigui E_t el nombre d'arestes entre els conjunts de vèrtexs explorats i de no explorats en el pas t del procés. Si t és prou petit, $Q_n \sim \mathbb{E}_t(E_t - E_{t-1})$, on \mathbb{E}_t és l'esperança condicionada, on condicionem a la història generada pels primers $t - 1$ passos del procés.

5.2 El règim subcrític

El cas $\nu < 1$ (o $Q < 0$) es coneix com a *règim subcrític*. D'una banda, si $D_n \sim \text{Po}(\lambda)$ amb $\lambda < 1$, el graf aleatori s'assembla a un $\mathbb{G}_{n,p}$ amb $p = \lambda/n$, que satisfà $\nu(C_1(\mathbb{G}_{n,p})) = \Theta(\log n)$ aap. D'altra banda, de manera determinista, $\nu(C_1(\mathbb{G}_n)) \geq \Delta_n + 1$, que pot créixer més ràpidament que de manera logarítmica.

Per determinar l'ordre de la component més gran, ens centrarem en l'estudi de la component d'un vèrtex triat uniformement a l'atzar. Sigui $(Y_{n,t})_{t \geq 0}$ com en la definició 9. Per la desigualtat de Markov, la probabilitat que el procés de ramificació sobrevisqui fins a la generació $t \geq 1$ és

$$\mathbb{P}(Y_{n,t} \geq 1) \leq \mathbb{E}[Y_{n,t}] = \mathbb{E}[D_n] \mathbb{E}[D_n^* - 1]^{t-1} = \frac{m_n}{n} \nu_n^{t-1},$$

que té un decreixement exponencial en t . Si t és més gran que $\frac{\log n}{|\log \nu_n|} \sim \frac{\log n}{|\log \nu|}$, aleshores obtenim que la probabilitat que algun vèrtex a \mathbb{CM}_n tingui vèrtexs a distància t és a com a màxim

$$\sum_{v \in [n]} d_v \nu_n^{t-1} = m_n \nu_n^{t-1} = o(1).$$

Tot i que això indica que el diàmetre de cada component és logarítmic, bé podria ser el cas que el nombre de vèrtexs en aquestes components fos molt més gran. El resultat següent demostra que no és així.

TEOREMA 14. *Suposem que \mathbf{d} satisfà la hipòtesi A. Si $\nu < 1$, llavors aap*

$$\nu(C_1(\mathbb{G}_n)) = O(\Delta_n \log n).$$

Molloy i Reed [43] anteriorment havien demostrat que $\nu(C_1(\mathbb{G}_n)) = O(\Delta_n^2 \log n)$ i aquesta millora en la cota superior ha estat provada recentment per Coulson i l'autor [12].

5.3 El règim supercrític

El cas $\nu > 1$ (o $Q > 0$) es coneix com a *règim supercrític*. El teorema 12 garanteix l'existència d'una component gegant, però quin és el seu ordre i la seva mida? Per determinar-ho, estudiarem la probabilitat que un vèrtex triat uniformement V_n pertanyi a la component gegant.

Per obtenir una intuïció podem usar l'acoblament amb processos de ramificació descrit en la secció 4. Sigui $(Y_t)_{t \geq 0}$ com a la definició 9. Com que $\nu > 1$,

hi ha una probabilitat positiva que el procés sobrevisqui indefinidament. D'una banda, si el procés s'extingeix, ho farà típicament en un nombre constant de passos, i V_n pertanyerà a una component petita. D'altra banda, si el procés sobreviu, la component de V_n creixerà fins a un punt en què ja no es podrà acoblar amb $(Y_t)_{t \geq 0}$. En aquest punt, la component exposada serà suficientment gran i aap formarà part de la component gegant.

Tot i que el comportament sub/supercrític de $(Y_t)_{t \geq 0}$ no depèn de la distribució de descendència de l'arrel, la probabilitat d'extinció sí que en depèn. Sigui η la probabilitat d'extinció d'un procés de ramificació amb descendència $D^* - 1$. Condicionat a $Y_1 = k$, $(Y_t)_{t \geq 0}$ només s'extingirà si tots els k subprocessos independents que comencen a la primera generació s'extingeixen, cosa que passa amb probabilitat η^k . Com que $\mathbb{P}(Y_1 = k) = \lambda_k$, tenim que la probabilitat de supervivència de $(Y_t)_{t \geq 0}$ és

$$\sum_{k \geq 1} (1 - \eta^k) \lambda_k = 1 - \sum_{k \geq 1} \eta^k \lambda_k.$$

Pel que fa al nombre d'arestes contingudes a la component gegant, la situació és diferent. Sigui E_n una aresta triada uniformement a l'atzar, i U_n un dels extrems de E_n . Llavors, per a qualsevol $u \in [n]$, la probabilitat que $U_n = u$ és proporcional al nombre de semiaarestes incidents a u . Així, el grau d_{U_n} es distribueix segons D_n^* , que s'aproxima per D^* . El veïnatge de E_n és la unió dels veïnatges dels seus dos extrems. En contrast amb els veïnatges de vèrtexs, aquí l'extrem U_n dona lloc a $d_{U_n} - 1$ arestes noves en el primer pas (ja que E_n no es compta). Per tant, el veïnatge de E_n es pot acoblar amb un procés de ramificació $(X_t)_{t \geq 0}$ amb distribució de descendència $D^* - 1$ i $X_0 = 2$. D'aquí es dedueix que la probabilitat que una aresta E_n pertanyi a la component gegant és $1 - \eta^2$.

La discussió anterior proporciona la intuïció que hi ha darrere del teorema següent.

TEOREMA 15. *Suposem que \mathbf{d} satisfà la hipòtesi A. Si $\nu > 1$, llavors*

$$\lim_{n \rightarrow \infty} \frac{v(C_1(\mathbb{G}_n))}{n} = 1 - \sum_{k \geq 1} \eta^k \lambda_k, \quad \text{en probabilitat,} \tag{12}$$

$$\lim_{n \rightarrow \infty} \frac{e(C_1(\mathbb{G}_n))}{n} = \frac{\lambda}{2} (1 - \eta^2), \quad \text{en probabilitat,}$$

on η és la probabilitat d'extinció d'un procés de ramificació amb distribució de descendència $D^* - 1$.

A més a més, per a qualsevol altra component C de \mathbb{G}_n ,

$$\lim_{n \rightarrow \infty} \frac{v(C)}{n} = 0, \quad \text{en probabilitat.}$$

El teorema 15 va ser demostrat inicialment per Molloy i Reed [44] sota condicions més fortes sobre la seqüència de graus; el resultat com s'indica aquí és de Janson i Luczak [32]. De fet, (12) també es compleix si eliminem la hipòtesi A.3 i demanem que $\mathbb{P}(D \geq 3) > 0$, com van demostrar Bollobás i Riordan [6] (vegeu també van der Hofstad [27, secció 4]).

5.4 El règim crític

El cas $\nu = 1$ (o $Q = 0$), conegut com a *règim crític*, ha generat molt d'interès en la literatura. Aquest règim es divideix en tres subrègims: el *quasisubcrític*, la *finestra crítica* i el *quasisupercrític*. En paraules, la finestra crítica és el conjunt de paràmetres que fan que el graf aleatori es comporti com un objecte crític. Aleshores, el subrègim quasisubcrític (o quasisupercrític) és el subrègim entre el règim subcrític (o supercrític) i la finestra crítica. Per distingir entre els tres règims, cal considerar la velocitat a la qual Q_n convergeix a Q quan $n \rightarrow \infty$ i utilitzar informació addicional sobre el procés.

Definim

$$R_n := \frac{1}{m_n} \sum_{i \in [n]} d_i(d_i - 2)^2.$$

Recordeu la definició de E_t a l'observació 13. Per a t prou petit, $Q_n \sim \mathbb{E}_t(E_t - E_{t-1})$ i $R_n \sim \mathbb{E}_t((E_t - E_{t-1})^2)$. Si $Q_n \rightarrow 0$ prou ràpidament, llavors el temps que triga E_t a ser zero (moment en què s'acaba l'exploració d'una component connexa) només depèn de les fluctuacions del procés, donades per R_n i no per Q_n . Hatami i Molloy ([25]) van trobar la condició necessària i suficient perquè això passi, establint l'amplada de la finestra crítica sota la hipòtesi A (i un parell de condicions addicionals més): per a qualsevol $\lambda, \epsilon > 0$ existeix $C > 1$ tal que si $Q_n \rightarrow 0$ i $|Q_n|^3 n / R_n^2 \leq \lambda$, llavors

$$\mathbb{P} \left(\frac{\nu(C_1(\mathbb{G}_n))}{R_n^{-1/2} n^{2/3}} \notin (C^{-1}, C) \right) \leq \epsilon.$$

Observeu que si $R_n = O(1)$, llavors la finestra crítica es comporta com la del graf aleatori binomial $\mathbb{G}_{n,p}$, quan $p = \frac{1}{n} + O(n^{-4/3})$, fet que dona lloc a components de mida $\Theta(n^{2/3})$.

La distribució límit dels ordres de les components més grans, adequadament normalitzats, ha estat estudiada en [13, 14, 35]. Riordan ([46]) va donar la descripció completa dels dos subrègims per a seqüències de graus acotats. Aquests resultats s'han estès recentment per seqüències de graus no acotats per al subrègim quasisubcrític [12] i el quasisupercrític [28, 32].

5.5 Un criteri universal

El criteri de Molloy-Reed no es compleix de manera general per a seqüències de graus més enllà de les que compleixen la hipòtesi A. Com es mostra a [6, 27], la hipòtesi A.3 es pot eliminar de les condicions, però això no passa amb

les altres suposicions. Joos, Rautenbach, Reed i l'autor ([34]) van proposar el criteri següent, que funciona per a totes les seqüències de graus \mathbf{d}_n . Fixem \mathbf{d}_n i assumim que $d_1 \leq \dots \leq d_n$ (altrament, podem reetiquetar els vèrtexs perquè així sia) i definim els paràmetres següents:

$$\hat{m}_n := \sum_{d_i \neq 2} d_i,$$

$$j_n^* := \inf \left\{ j \in [n] : \sum_{i=1}^j d_i(d_i - 2) > 0 \right\} \wedge n,$$

$$S_n := \sum_{j=j_n^*}^n d_j.$$

En termes d'aquests paràmetres, es pot donar la següent classificació de seqüències per a l'existència d'una component gegant.

TEOREMA 16 (CRITERI UNIVERSAL PER A LA COMPONENT GEGANT). *Per a qual-sevol seqüència de graus \mathbf{d}_n ,*

- (i) *si $\hat{m}_n \rightarrow \infty$ i $S_n = o(n)$, llavors $v(C_1(\mathbb{G}_n)) = o(n)$ aap;*
- (ii) *si $\hat{m}_n \rightarrow \infty$ i $S_n = \Theta(n)$, llavors $v(C_1(\mathbb{G}_n)) = \Theta(n)$ aap;*
- (iii) *si $\hat{m}_n = O(1)$, llavors, per a cada $\epsilon > 0$, existeix $\delta > 0$ tal que*

$$\mathbb{P} \left(\frac{v(C_1(\mathbb{G}_n))}{n} \leq \epsilon \right) \geq \delta, \quad \mathbb{P} \left(\frac{v(C_1(\mathbb{G}_n))}{n} \geq 1 - \epsilon \right) \geq \delta.$$

5.6 Grafs aleatoris dirigits

Moltes de les xarxes que ens trobem en el món real són dirigides, és a dir, el fet que un vèrtex es connecti amb un altre no necessàriament implica que el segon estigui connectat amb el primer. Alguns exemples són la *world wide web* (xarxa d'Internet on les arestes són hiperenllaços d'una pàgina a una altra), la xarxa de citacions en articles científics, algunes xarxes socials (Instagram, Twitter, etc.) o la xarxa de carreteres d'una ciutat. Els grafs dirigits tenen similituds amb els no dirigits, però algunes nocions, així com els resultats que les estudien, canvien dràsticament. En aquesta secció discutim breument l'estructura de les components fortament connexes dels grafs aleatoris dirigits, però primer ens cal definir-los.

Una *seqüència de digraus* de longitud n és una seqüència $\vec{\mathbf{d}}_n = ((d_i^-, d_i^+))_{i \in [n]}$, on d_i^- i d_i^+ són números enters no negatius amb $d_i^-, d_i^+ \leq n-1$, per a tot $i \in [n]$. Totes les definicions i notacions bàsiques per al cas no dirigit, s'adapten de forma natural a grafs (aleatoris) dirigits. Donada una seqüència de digraus factible $\vec{\mathbf{d}}_n$, $\vec{\mathbb{G}}_n = \vec{\mathbb{G}}_{n, \vec{\mathbf{d}}_n}$ és un graf dirigit triat uniformement aleatòriament entre tots els grafs dirigits amb conjunt de vèrtexs $[n]$ i amb seqüència de digraus $\vec{\mathbf{d}}_n$.

Sigui $\vec{D}_n = (D_n^-, D_n^+)$ el grau d'un vèrtex triat uniformement a l'atzar a $[n]$. Podem establir hipòtesis anàlogues a la hipòtesi A per a seqüències de digraus: existeix una distribució de probabilitat $\vec{D} = (D^-, D^+)$ a $\mathbb{Z}_{\geq 0}^2$ tal que $\vec{D}_n \xrightarrow{d} \vec{D}$ i per a tots els números enters no negatius a, b amb $a + b \leq 2$,

$$\lim_{n \rightarrow \infty} \mathbb{E}[(D_n^-)^a (D_n^+)^b] = \mathbb{E}[(D^-)^a (D^+)^b] < \infty.$$

Definim

$$\vec{\nu} := \frac{\mathbb{E}[D^- D^+]}{\mathbb{E}[D^-]}. \quad (13)$$

Llavors $\vec{\nu}$ controla si la component fortament connexa més gran té ordre lineal (vegeu [11, 7]): si $\vec{\nu} > 1$,

$$\lim_{n \rightarrow \infty} \frac{\nu(C_1(\vec{\mathbb{G}}_n))}{n} = \zeta, \quad \text{en probabilitat,} \quad (14)$$

per a alguna constant $\zeta > 0$; i si $\vec{\nu} < 1$, llavors $\nu(C_1(\vec{\mathbb{G}}_n)) = O(1)$.

5.7 Percolació

Donada una propietat d'una xarxa, és natural plantejar-nos com és de robusta davant de determinats canvis en l'ambient. En el nostre cas, ho mesurarem a través del concepte de *percolació*: si cada aresta és eliminada amb una certa probabilitat, quina és la probabilitat que la propietat es perdi? Ens focalitzarem en la propietat de contenir una component gegant; per tant, assumirem que $\nu > 1$ per tal que el graf aleatori inicialment satisfaci aquesta propietat aap. Aleshores estudiarem quina és la mínima probabilitat d'eliminació que fa que la percolació canvia l'estructura de components de manera substancial. Aquesta noció té múltiples aplicacions, per exemple, en l'estabilitat dels sistemes ecològics o en la resistència d'una xarxa de comunicació a atacs aleatoris.

Donat un graf G i $p \in [0, 1]$, el *graf p -percolat* G^p és un graf aleatori obtingut a partir de G mantenint cada aresta independentment amb probabilitat p . Per exemple, si $G = K_n$ és el graf complet amb n vèrtexs, llavors $G^p = \mathbb{G}_{n,p}$ és el graf aleatori binomial.

Quan considerem \mathbb{G}_n^p , primer cal observar que hi ha dos nivells d'aleatorietat: primer triem un graf aleatori \mathbb{G}_n i després percolem cada aresta independentment amb probabilitat p . No obstant això, l'heurística següent és vàlida sempre que la seqüència de graus satisfaci la hipòtesi A. Construïm una seqüència (aleatòria) de graus percolats \mathbf{d}_n^p mantenint cada semiaresta de la seqüència de graus original \mathbf{d}_n independentment amb probabilitat p . Suposant que la suma dels graus percolats sigui parell (condició necessària per ser seqüència de graus), podem considerar el model de configuració $\mathbb{CM}_{n, \mathbf{d}_n^p}$. Llavors, $\mathbb{CM}_{n, \mathbf{d}_n^p}$ té la mateixa llei de probabilitat que $(\mathbb{CM}_{n, \mathbf{d}_n})^p$.

Aleshores, la *probabilitat crítica de percolació* p_c es defineix com l'ímfim de les probabilitats p que fan que el graf aleatori percolat \mathbb{G}_n^p tingui una

component gegant aap. El grau en \mathbf{d}_n^p d'un vèrtex de grau ℓ en \mathbf{d}_n està distribuït com a $\text{Bin}(\ell, p)$. Per tant, el nombre esperat de vèrtexs de grau k en \mathbf{d}_n^p és

$$\lambda_k^p := \sum_{\ell \geq k} \lambda_\ell \binom{\ell}{k} p^k (1-p)^{\ell-k},$$

i podem calcular el grau mitjà esperat de \mathbf{d}_n^p

$$\begin{aligned} \sum_{k \geq 1} k \lambda_k^p &= \sum_{k \geq 1} k \sum_{\ell \geq k} \lambda_\ell \binom{\ell}{k} p^k (1-p)^{\ell-k} \\ &= \sum_{\ell \geq 1} \lambda_\ell \sum_{k \leq \ell} k \binom{\ell}{k} p^k (1-p)^{\ell-k} \\ &= \sum_{\ell \geq 1} p \ell \lambda_\ell = p \lambda, \end{aligned}$$

on hem utilitzat que $\mathbb{E}[\text{Bin}(\ell, p)] = \ell p$. De manera similar podem fer servir que $\text{Var}(\text{Bin}(\ell, p)) = \ell p(1-p)$ per obtenir

$$\sum_{k \geq 1} k(k-1) \lambda_k^p = p^2 \sum_{k \geq 1} k(k-1) \lambda_k.$$

Definim

$$\nu^p := \frac{\sum_{k \geq 1} k(k-1) \lambda_k^p}{\sum_{k \geq 1} k \lambda_k^p} = p \nu.$$

Segons l'heurística introduïda en les seccions anteriors, la condició per a la supercriticalitat del graf aleatori percolat és $\nu^p > 1$, que és equivalent a $p > 1/\nu$, com estableix el teorema següent.

TEOREMA 17. *Suposem que \mathbf{d} satisfà la hipòtesi A amb $\nu > 1$. Sigui $p_c := 1/\nu \in (0, 1)$. Llavors,*

- (i) *si $p < p_c$, llavors $\nu(C_1(\mathbb{G}_n^p)) = o(n)$ aap;*
- (ii) *si $p > p_c$, llavors $\nu(C_1(\mathbb{G}_n^p)) = \Theta(n)$ aap.*

Aquest resultat va ser demostrat per Fountoulakis [20] sota condicions més fortes en la seqüència de grau, i l'enunciat donat aquí és de Janson [30], que també va calcular l'orde asimptòtic de $C_1(\mathbb{G}_n^p)$ quan $p > p_c$.

5.8 Connectivitat

Més enllà del llindar per a l'existència d'una component gegant, arribem al llindar per a la connectivitat del graf. Recordem que un graf és connex si tots els vèrtexs estan a distància finita els uns dels altres. La connectivitat és una propietat desitjable en moltes xarxes. Per exemple, en una xarxa de telecomunicacions, com la xarxa 5G, la connectivitat certifica que dos nodes qualssevol es poden comunicar a través d'ella.

Luczak ([37]) va demostrar que si $d_i \geq 3$ per a tot $i \in [n]$ (i el grau màxim no és massa gran), llavors \mathbb{G}_n és connex aap. Des del punt de vista intuïtiu, si cada vèrtex està connectat com a mínim amb tres més, això ens garanteix que típicament tots els veïnatges creixen a una velocitat exponencial, i acaben per formar part de la mateixa component. Així doncs, és natural preguntar-nos quants vèrtexs de grau 1 i grau 2 podem permetre en la seqüència de graus abans que el graf aleatori deixi de ser connex aap.

Recordem que la nostra hipòtesi implica que $m_n = \Theta(n)$. D'una banda, suposem que hi ha k vèrtexs de grau 1. La probabilitat que cap de les semiarestes incidents a vèrtexs de grau 1 s'aparelli entre elles, formant una aresta aïllada, és

$$\prod_{i=1}^k \left(1 - \frac{k-i}{m_n - 2i + 1}\right) \sim e^{-k^2/2m_n},$$

sempre que k no sigui massa gran. Així doncs, si k és d'ordre més gran o igual que \sqrt{n} , la probabilitat que existeixi una aresta aïllada, i que, per tant, el graf no sigui connex, és positiva. Això es pot veure com un exemple del problema clàssic en probabilitat discreta conegut com a *paradoxa dels aniversaris*; vegeu [18, p. 33].

D'altra banda, recordeu que λ_2 és la proporció de vèrtexs de grau 2. Per a $k \geq 3$ fixat, la probabilitat que un vèrtex de grau 2 concret pertanyi a un cicle de longitud k aïllat és

$$\frac{1}{m_n - 2k + 1} \prod_{i=1}^k \frac{2(n_{2,n} - i + 1)}{m_n - 2i + 1} \sim \frac{(2\lambda_2)^{k-1}}{\lambda^k} \cdot \frac{1}{n}.$$

Per tant, si $\lambda_2 > 0$, el nombre esperat de cicles curts aïllats de qualsevol longitud constant és positiu. En cas que algun d'ells existeixi, \mathbb{G}_n no seria connex.

Si el límit existeix, definim

$$\rho_1 := \lim_{n \rightarrow \infty} \frac{n_{1,n}}{\sqrt{n}} \in [0, \infty].$$

TEOREMA 18. *Suposem que \mathbf{d} satisfà la hipòtesi A i que ρ_1 existeix. Llavors,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}_n \text{ és connex}) = \sqrt{\frac{\lambda - 2\lambda_2}{\lambda}} \cdot e^{-\frac{\rho_1^2}{2(\lambda - 2\lambda_2)} + \frac{\lambda_2^2 + \lambda\lambda_2}{\lambda^2}},$$

on λ es defineix al punt 2 de la hipòtesi A. En particular,

- (i) si $\rho_1 = \infty$ o $\lambda_2 = 1$, llavors \mathbb{G}_n no és connex aap;
- (ii) si $\rho_1 = 0$ i $\lambda_2 = 0$, llavors \mathbb{G}_n és connex aap.

Aquest resultat va ser demostrat per Federico i van der Hofstad [17].

6 Distàncies

Una de les característiques més importants de les xarxes complexes és el fenomen petit-món (*small-world*): és possible connectar dos nodes qualssevol de la xarxa a través d'un camí sorprenentment curt. És ben conegut l'experiment de Milgram, que va estudiar la distància mitjana de la població dels Estats Units d'Amèrica a través de la seva xarxa de coneguts. D'aquí en va sortir el concepte dels *sis graus de separació*: entre dues persones qualssevol sempre hi ha una connexió a través de sis o menys persones conegudes entre elles. Per tant, és important que els grafes aleatoris que utilitzem per modelar aquestes xarxes tinguin també aquesta característica. Com veurem en aquesta secció, la distància típica entre dos vèrtexs aleatoris, així com el diàmetre del graf \mathbb{G}_n , són d'ordre logarítmic en n , que és molt menor que l'ordre del graf.

6.1 Distància típica

Siguin U_n i V_n triats independentment i de manera uniforme de $[n]$. El nostre objectiu és estudiar la variable aleatòria $\text{dist}(U_n, V_n)$, coneguda com a *distància típica*, on la distància es mesura a \mathbb{G}_n com s'ha descrit a l'inici de la secció 2. Si $\nu < 1$, la gran majoria de parelles no estan connectades, així que ens limitarem al cas supercrític $\nu > 1$. Com que el nostre graf aleatori pot no ser connex, té sentit condicionar U_n i V_n a pertànyer a la mateixa component connexa, la gegant.

Donarem un argument heurístic per calcular la distància típica. Segons el lema 8, els veïnatges de U_n i V_n creixen a la mateixa velocitat que un procés de ramificació amb distribució de descendència $D^* - 1$ (la descendència de l'arrel no és rellevant per a aquesta anàlisi), on D^* s'ha introduït a la definició 9, de manera que la mida esperada del h -veïnatge d'un vèrtex és de l'ordre de ν^h . La mida dels veïnatges no està concentrada al voltant de la seva esperança; de fet, hi ha molts vèrtexs amb veïnatges relativament petits. No obstant això, condicionat al fet que U_n i V_n pertanyin a la component gegant, la mida dels dos veïnatges sí que està concentrada al voltant de ν^h . Fixant $h = \frac{\log n}{2 \log \nu}$, cadascun dels dos h -veïnatges tenen una mida aproximada de \sqrt{n} . A més a més, donats dos conjunts fixats de mida \sqrt{n} , hi ha una probabilitat constant de tenir una aresta entre ells, connectant aquests dos veïnatges i creant un camí entre U_n i V_n . Així doncs, la distància típica hauria de ser aproximadament $2h$.

El resultat següent determina la distància típica en grafes aleatoris en el règim supercrític. La seva prova formalitza l'heurística anterior i es pot trobar a [27, secció 6].

TEOREMA 19. *Suposem que \mathbf{d} satisfà la hipòtesi A i $\nu > 1$. Si U_n i V_n es trien independentment i de manera aleatòria de $[n]$, condicionats al fet que U_n i V_n estiguin a la component gegant de \mathbb{G}_n , aleshores*

$$\text{dist}(U_n, V_n) \sim \frac{\log n}{\log \nu}.$$

Si la hipòtesi A.3 falla, en particular si $\mathbb{E}[D^2] = \infty$, el resultat encara és vàlid i $\text{dist}(U_n, V_n) = o(\log n)$.

6.2 Diàmetre

Recordeu que el diàmetre d'una component connexa és la distància més gran entre dos dels seus vèrtexs. En el règim subcrític $\nu < 1$, l'argument usat a la subsecció 5.2 dona una idea del fet que el diàmetre més gran d'una component és asimptòticament igual a $\frac{\log n}{|\log \nu|}$. En el règim supercrític $\nu > 1$, tot i que la component gegant té ordre lineal, el seu diàmetre és d'ordre logarítmic, com la distància típica. No obstant això, el diàmetre i la distància típica difereixen per una constant multiplicativa. Descartant el cas degenerat $\lambda_1 = 0$, on λ_1 està definit a la hipòtesi A, donem una discussió informal sobre aquesta constant abans d'enunciar el resultat.

Al règim supercrític, típicament, el veïnatge d'un vèrtex o bé creix ràpidament o bé s'extingeix en un nombre petit de passos. No obstant això, hi ha veïnatsges atípics que ni s'expandiran ni s'extingiran durant molts passos. Per donar una intuïció sobre aquest fenomen, considerem un procés de ramificació supercrític $(X_t)_{t \geq 0}$ amb distribució de descendència ξ . Sigui $\hat{\xi}$ la distribució conjugada de ξ , definida com a (8); podem escriure

$$\hat{\nu} := \mathbb{E}[\hat{\xi}] = \frac{1}{\lambda} \sum_{k \geq 1} k(k-1)\eta^{k-2}\lambda_k \in (0, 1), \quad (15)$$

que indica el creixement esperat del procés $(X_t)_{t \geq 0}$ condicionat a la seva extinció. En particular, com que $\hat{\nu} < 1$, el procés condicionat és subcrític.

Diem que un element del procés té *descendència infinita* si el subarbre que en penja no s'extingeix. Sigui $(X'_t)_{t \geq 0}$ el subprocés dels elements amb descendència infinita. Podem calcular la probabilitat que, condicionat a la seva supervivència, el procés només tingui un element amb descendència infinita a la generació t , és a dir, $X'_t = 1$. Recordeu que \mathcal{E} és l'esdeveniment d'extinció i η és la seva probabilitat; vegeu (7). Tenim

$$\mathbb{P}(X'_1 = 1 \mid \overline{\mathcal{E}}) = \frac{\mathbb{P}(X'_1 = 1)}{\mathbb{P}(\overline{\mathcal{E}})} = \frac{\sum_{k \geq 1} \mathbb{P}(\xi = k)k(1-\eta)\eta^{k-1}}{1-\eta} = \hat{\nu}.$$

Aquesta fórmula té l'explicació següent: l'arrel pot tenir qualsevol nombre positiu k de fills, distribuïts d'acord amb la llei de ξ . Llavors hi ha k opcions per seleccionar el fill que tindrà descendència infinita, cosa que passa amb probabilitat $(1-\eta)$. D'altra banda, cadascun dels altres $(k-1)$ fills té probabilitat η que s'extingeixi la seva descendència. Com que tots els esdeveniments anteriors són mútuament independents, se n'obté la fórmula.

Noteu que $(X'_t)_{t \geq 0}$ és un procés no decreixent. Per tant, si $X'_t = 1$, llavors $X'_s = 1$ per a tot $s \in [t]$. Com que $\{X'_0 = 1\} = \overline{\mathcal{E}}$, usant la propietat markoviana dels processos de ramificació, tenim que:

$$\mathbb{P}(X'_t = 1 \mid \overline{\mathcal{E}}) = \prod_{s=1}^t \mathbb{P}(X'_s = 1 \mid X'_{s-1} = 1) = (\mathbb{P}(X'_1 = 1 \mid \overline{\mathcal{E}}))^t = \hat{\nu}^t. \quad (16)$$

L'equació (16) resulta ser una bona aproximació de la probabilitat que el veïnatge d'un vèrtex ni s'expandeixi ni s'extingeixi durant t passos.

Com que els veïnatges es poden acoblar amb processos de ramificació amb distribució de descendència $D^* - 1$ (la descendència de l'arrel no és rellevant per a aquesta anàlisi) i $\lambda_1 > 0$ implica $\eta > 0$, tenim que la probabilitat d'un veïnatge atípicament «prim» de longitud t és asimptòticament $\hat{\nu}^t$, on

$$\hat{\nu} := \mathbb{E}[\hat{\xi}] = \frac{1}{\lambda} \sum_{k \geq 1} k(k-1)\eta^{k-2}\lambda_k \in (0, 1).$$

Com que hi ha un nombre lineal de vèrtexs a la component gegant, aap tindrem vèrtexs amb veïnatges «prims» de longitud $t = \frac{\log n}{|\log \hat{\nu}|}$. Els vèrtexs al final d'aquests veïnatges «prims» es comportaran com a vèrtexs triats uniformement a l'atzar, i la seva distància vindrà donada pel teorema 19.

Fernholz i Ramachandran ([19]) van demostrar que aquesta intuïció és correcta i el diàmetre és d'ordre logarítmic amb la constant principal donada per les contribucions anteriors.

TEOREMA 20. *Suposem que \mathbf{d} satisfà la hipòtesi A. Si $\nu < 1$,*

$$\text{diam}(C_1(\mathbb{G}_n)) \sim \frac{\log n}{|\log \nu|}.$$

Si $\nu > 1$ i $\lambda_1 > 0$,

$$\text{diam}(C_1(\mathbb{G}_n)) \sim \left(\frac{1}{\log \nu} + \frac{2}{|\log \hat{\nu}|} \right) \log n.$$

6.3 Grafs aleatoris dirigits

Sota les hipòtesis descrites a la subsecció 5.6, podem estudiar les distàncies típiques [29] i el diàmetre [8] dels grafs aleatoris dirigits en el règim supercrític, donat per $\bar{\nu} > 1$; vegeu (13)–(14).

Siguin U_n i V_n triats independentment i de manera uniforme de $[n]$. Condiçionat al fet que U_n i V_n estiguin a la mateixa component fortament connexa de $\bar{\mathbb{G}}_n$, aap

$$\text{dist}(U_n, V_n) \sim \frac{\log n}{\log \bar{\nu}},$$

i

$$\text{diam}(C_1(\bar{\mathbb{G}}_n)) \sim \left(\frac{1}{\log \bar{\nu}} + \frac{1}{|\log \hat{\nu}^-|} + \frac{1}{|\log \hat{\nu}^+|} \right) \log n,$$

on $\hat{\nu}^-$ i $\hat{\nu}^+$ són constants que s'obtenen com a esperances de certes distribucions de descendència relacionades amb la seqüència de digraus, esbiaixada per la mida (vegeu [8] per a una definició precisa).

7 Grafs aleatoris densos

La gran majoria dels resultats presentats en les seccions anteriors són per a seqüències de graus que satisfan la hipòtesi A, la qual cosa en particular implica que el graf té poques arestes. Tot i que les xarxes que trobem al món real acostumen a tenir relativament poques connexions comparat amb el nombre d'elements, des del punt de vista combinatori, és interessant entendre aquells casos en què el nombre mitjà de connexions tendeix cap a infinit quan n també ho fa. En aquesta secció discutirem alguns resultats per a grafs aleatoris densos, les seqüències de graus dels quals poden tenir variància, o fins i tot esperança, no acotada. El principal entrebanc a l'hora d'estudiar aquestes seqüències és que la probabilitat que el model de configuració sigui simple tendeix ràpidament a zero (vegeu el teorema 4). Explicarem com superar aquest obstacle introduint el mètode de commutació i discutirem com s'aplica aquesta tècnica en dos contextos: l'enumeració i el mostreig de grafs amb una seqüència de graus donada.

7.1 El mètode de commutació (*switching*)

El mètode de commutació per a grafs aleatoris va ser introduït per McKay [38]. Es tracta d'una forma particular de l'argument clàssic de doble recompte i, en el context dels grafs aleatoris, es pot descriure com segueix. Sigui $\mathcal{P}_n = \mathcal{P}_{n, \mathbf{d}_n}$ el conjunt de grafs de $\mathcal{G}_{n, \mathbf{d}_n}$ que satisfan certa propietat. El nostre objectiu és calcular la probabilitat que un graf aleatori sigui a \mathcal{P}_n , és a dir,

$$\mathbb{P}(\mathbb{G}_n \in \mathcal{P}_n) = \frac{|\mathcal{P}_{n, \mathbf{d}_n}|}{|\mathcal{G}_{n, \mathbf{d}_n}|} = \frac{|\mathcal{P}_n|}{|\mathcal{P}_n| + |\mathcal{P}_n^c|} = \frac{\alpha}{\alpha + 1}, \quad (17)$$

on $\alpha = |\mathcal{P}_n|/|\mathcal{P}_n^c|$ és la raó entre el nombre de grafs que tenen la propietat i els que no la tenen.

El mètode de commutació ens permet estimar α mitjançant un argument basat en una operació simple anomenada commutació (*switch*). Per a $G \in \mathcal{P}_n$ i $H \in \mathcal{P}_n^c$, sigui $d(G, H)$ el nombre de maneres en què H es pot obtenir a partir de G mitjançant una commutació, i definim $\bar{d}(G) = \sum_{H \in \mathcal{P}_n^c} d(G, H)$. De manera similar, podem definir $\bar{d}(H)$ com el nombre de maneres en què H es pot obtenir a partir d'un graf de \mathcal{P}_n . Llavors,

$$\bar{d}(\mathcal{P}_n)|\mathcal{P}_n| = \sum_{G \in \mathcal{P}_n} \bar{d}(G) = \sum_{H \in \mathcal{P}_n^c} \bar{d}(H) = \bar{d}(\mathcal{P}_n^c)|\mathcal{P}_n^c|,$$

on $\bar{d}(\mathcal{A})$ és el grau mitjà al conjunt \mathcal{A} . Aleshores $\alpha = \bar{d}(\mathcal{P}_n^c)/\bar{d}(\mathcal{P}_n)$ i, estimant aquestes dues quantitats, podem obtenir una aproximació de α .

Com a operació, una commutació és una modificació local del graf. La commutació més usada és el del cicle de longitud 4: donat $G \in \mathcal{P}_n$ amb $vw, xy \in E(G)$ i $vx, wy \notin E(G)$, obtenim el graf $H \in \mathcal{P}_n^c$ eliminant les arestes vw, xy i afegint les arestes vx, wy (vegeu la figura 2(a)). Una altra possible operació és la commutació sobre un cicle de longitud 6: donat $G \in \mathcal{P}_n$ amb

$vw, x_1y_1, x_2y_2 \in E(G)$ i $vx_1, wy_2, x_2y_1 \notin E(G)$, obtenim el graf $H \in \mathcal{P}_n^c$ eliminant les arestes vw, x_1y_1, x_2y_2 i afegint les arestes vx_1, wy_2, x_2y_1 (vegeu la figura 2(b)). Aquests dos són exemples de commutacions senzilles; a la literatura s'han utilitzat commutacions més complexes per superar certes dificultats tècniques i estendre l'aplicabilitat del mètode.

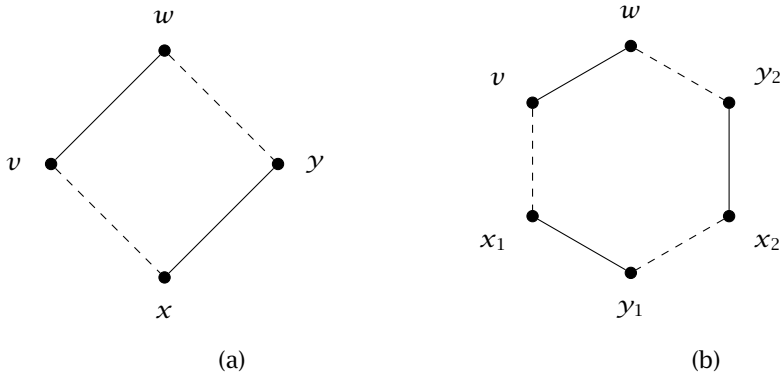


FIGURA 2: L'operació de commutació; les arestes representades amb línies contínues són reemplaçades per les representades amb línies discontinües: (a) commutació sobre un cicle de longitud 4; (b) commutació sobre un cicle de longitud 6.

Per il·lustrar el mètode, ara en donem un exemple. Donada \mathbf{d}_n i $v, w \in [n]$ diferents, sigui \mathcal{P}_n el conjunt de grafs que contenen l'aresta vw . Farem commutacions sobre cicles de longitud 6 per estimar la probabilitat que $vw \in E(\mathbb{G}_n)$. Per una banda, donat un graf $G \in \mathcal{P}_n$, hem de trobar $x_1y_1, x_2y_2 \in E(G)$ tals que $vx_1, wy_2, x_2y_1 \notin E(G)$. Hi ha m_n^2 opcions per a $x_1y_1, x_2y_2 \in E(G)$ qualssevol, d'on

- a com a màxim $d_v \Delta_n m_n$ d'elles, x_1 és incident a v ;
- a com a màxim $d_w \Delta_n m_n$ d'elles, y_2 és incident a w ;
- a com a màxim $\Delta_n^2 m_n$ d'elles, y_1 és incident a x_2 .

Així doncs, el nombre total de commutacions és $m_n^2(1 + O(\Delta_n^2/m_n))$ i $\bar{d}(\mathcal{P}_n) = m_n^2(1 + O(\Delta_n^2/m_n))$.

D'altra banda, donat un graf $H \in \mathcal{P}_n^c$, hem de trobar $vx_1, wy_2, x_2y_1 \in E(H)$ tals que $x_1y_1, x_2y_2 \notin E(H)$. Hi ha $d_v d_w m_n$ opcions per a $vx_1, wy_2, x_2y_1 \in E(H)$ qualssevol, d'on

- a com a màxim $d_v d_w \Delta_n^2$ d'elles, y_1 és incident a x_1 ;
- a com a màxim $d_v d_w \Delta_n^2$ d'elles, x_2 és incident a y_2 .

Així doncs, el nombre total de commutacions és $d_v d_w m_n(1 + O(\Delta_n^2/m_n))$ i $\bar{d}(\mathcal{P}_n^c) = d_v d_w m_n(1 + O(\Delta_n^2/m_n))$. Per (17), obtenim

$$\mathbb{P}(vw \in E(\mathbb{G}_n)) \sim \frac{d_v d_w}{m_n}, \tag{18}$$

sempre que $\Delta_n^2 = o(m_n)$. Animem el lector a comprovar la limitació de les commutacions sobre cicles de longitud 4 per calcular la probabilitat anterior.

Veient el resultat (18), es pot deduir que el mètode de commutació permetrà estudiar seqüències de graus sempre que el seu grau màxim no sigui massa gran respecte al nombre d'arestes. En les seccions següents apareixeran condicions similars a $\Delta_n^2 = o(m_n)$.

7.2 Enumeració

Un dels temes fonamentals en matemàtica discreta és l'enumeració d'objectes combinatoris. En el nostre cas, ens agradaria comptar quants grafs hi ha amb conjunt de vèrtexs $[n]$ i seqüència de graus \mathbf{d}_n , és a dir, calcular $|\mathcal{G}_{n,\mathbf{d}_n}|$. En aquesta secció farem una introducció al tema. Remetem el lector interessat a l'article de Wormald [52] per a una anàlisi més detallada.

Una possible manera d'enfocar aquest problema és fer servir el model de configuració $\mathbb{CM}_n = \mathbb{CM}_{n,\mathbf{d}_n}$ en combinació amb les estimacions sobre la probabilitat que el graf generat sigui simple. El nombre total d'aparellaments de les m_n semiaarestes és $(m_n - 1)!!$, però hi ha més d'un aparellament que dona lloc al mateix graf etiquetat; vegeu (1). Tot i això, com que el nombre d'aparellaments que donen lloc al mateix graf només depèn del nombre de llaços i arestes múltiples, és el mateix per a tots els grafs simples i obtenim el resultat següent:

$$|\mathcal{G}_{n,\mathbf{d}_n}| = \frac{(m_n - 1)!!}{\prod_{i \in [n]} d_i!} \cdot \mathbb{P}(\mathbb{CM}_n \text{ és simple}). \quad (19)$$

Com hem exposat a la secció 3, assumint la hipòtesi A, el nombre de llaços i d'arestes dobles es pot aproximar per distribucions de Poisson independents amb paràmetres $\nu_n/2$ i $\nu_n^2/4$, i obtenir així el teorema 4, que reformulem ara en termes de ν_n :

$$\mathbb{P}(\mathbb{CM}_n \text{ és simple}) \sim \exp\left(-\frac{\nu_n}{2} - \frac{\nu_n^2}{4}\right) > 0. \quad (20)$$

Combinat amb (19), ens dona el nombre asimptòtic de grafs amb una seqüència de graus donada sempre que se satisfaci la nostra hipòtesi.

Una línia de recerca central en grafs aleatoris és entendre com aquesta fórmula s'estén a altres seqüències de graus per a les quals $\nu_n \rightarrow \infty$. En la resta d'aquesta secció, discutirem la probabilitat de ser simple en termes de la relació entre Δ_n i m_n .

Definim els següents *moments factorials* de D_n :

$$\gamma_n := \frac{1}{m_n} \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2),$$

$$\mu_n := \frac{1}{m_n} \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2)(d_i - 3).$$

Per obtenir una intuïció sobre quan l'aparició de llaços i arestes dobles perd el seu comportament poissonià, mirarem el nombre de llaços dobles DL_n , arestes triples T_n i arestes quàdruples Q_n , el nombre esperat de les quals es pot calcular seguint els mateixos passos que en (4). Si assumim que $\sum_{i \in [n]} d_i^k = o((\sum_{i \in [n]} d_i)^k)$ per a $k \in \{2, 3, 4\}$, llavors tenim:

$$\mathbb{E}[DL_n] \sim \frac{\mu_n}{8m_n}, \quad \mathbb{E}[T_n] \sim \frac{\gamma_n^2}{12m_n}, \quad \mathbb{E}[Q_n] \sim \frac{\mu_n^2}{48m_n^2},$$

que implica

$$\mathbb{E}[DL_n] = O\left(\frac{\Delta_n^3}{m_n}\right), \quad \mathbb{E}[T_n] = O\left(\frac{\Delta_n^4}{m_n}\right), \quad \mathbb{E}[Q_n] = O\left(\frac{\Delta_n^6}{m_n^2}\right).$$

Per tant, si $\Delta_n^4 = o(m_n)$, no esperem ni llaços dobles ni arestes triples, de manera que els llaços i les arestes dobles apareixen de manera disjunta i les correlacions entre ells són febles. Per tant, l'aproximació de Poisson donada per (20) continua sent vàlida, com va demostrar McKay [39].

Si $\Delta_n^3 = o(m_n)$, no esperem ni llaços dobles ni arestes quàdruples, però el nombre d'arestes triples no és negligible, de manera que el nombre d'arestes dobles ja no es comportarà com una Poisson.

TEOREMA 21. *Suposem que \mathbf{d}_n satisfà $\Delta_n^3 = o(m_n)$. Llavors,*

$$\mathbb{P}(\text{CM}_n \text{ és simple}) \sim \exp\left(-\frac{\nu_n}{2} - \frac{\nu_n^2}{4} - \frac{1}{12m_n}(6\nu_n^2\gamma_n - 3\nu_n^4 - 2\gamma_n^2)\right).$$

Aquest resultat, provat per McKay i Wormald [42], és l'estimació més general coneguda per a aquest problema.

La prova del teorema 21 es basa en el mètode de commutació. La idea principal és obtenir un valor asimptòtic per a $|\mathcal{P}_n(\ell, d, t)|$, el nombre de configuracions amb ℓ llaços, d arestes dobles i t arestes triples, estimant les raons següents:

$$\frac{|\mathcal{P}_n(\ell, d, t)|}{|\mathcal{P}_n(\ell, d, t-1)|}, \quad \frac{|\mathcal{P}_n(\ell, d, 0)|}{|\mathcal{P}_n(\ell-1, d, 0)|} \quad \text{i} \quad \frac{|\mathcal{P}_n(0, d, 0)|}{|\mathcal{P}_n(0, d-1, 0)|}.$$

Per tal de fer-ho s'utilitzen diferents tipus de commutacions per crear o destruir llaços, arestes dobles o arestes triples.

7.3 Mostreig

Un dels reptes centrals en la ciència de la computació teòrica és desenvolupar algorismes aleatoris eficients per generar objectes combinatoris d'acord amb una distribució específica. Un algorisme es considera eficient si el seu temps (o temps esperat) d'execució és un polinomi on la variable és la mida de l'entrada. Aquí estudiarem la generació de grafs de $\mathcal{G}_{n, \mathbf{d}_n}$ en temps (o temps esperat) polinomial en n . Parlarem de dues famílies d'algorismes de mostreig:

mostreig uniforme, en què es generen grafs aleatoris seguint la distribució uniforme, i *mostreig quasiuniforme*, en què es generen grafs aleatoris d'una distribució de probabilitat molt propera a la uniforme. En aquesta secció presentem breument els mètodes més prominents i remetem el lector interessat a l'article de Greenhill [23].

Un algorisme senzill per mostrejar aquests grafs és generar un exemple utilitzant el model de configuració. Si el parell resultant és simple, per (2), serà un graf aleatori uniforme. En cas contrari, el descartem i repetim l'experiment fins que aquest tingui èxit. Aquest algorisme es coneix com a *mostreig per rebuig* i el seu temps d'execució esperat està determinat per la probabilitat que el model de configuració produeixi un graf simple. Usant el teorema 4, podem deduir que l'algorisme és eficient per a seqüències de graus que satisfan la hipòtesi A. De fet, permet mostrejar en temps polinomial qualsevol seqüència de graus que compleixi $\Delta_n = O(\sqrt{\log n})$.

McKay i Wormald ([41]) van proposar un algorisme basat en commutacions que realitza commutacions aleatòries per eliminar llaços i arestes dobles. Per tal d'obtenir la distribució uniforme, cal afegir una *probabilitat de rebuig* a cada pas. En el context més general de mostrejar aparellaments perfectes de grafs, Jerrum i Sinclair ([33]) van introduir un algorisme que funciona per a seqüències de graus *P-estables*, és a dir, seqüències \mathbf{d}_n en què el nombre de grafs amb aquesta seqüència de graus és estable sota petites pertorbacions de la seqüència. D'altra banda, Kannan, Tetali i Vempala ([36]) van introduir la *cadena de commutacions* (*switching chain* en anglès), una cadena de Markov amb espai d'estats $\mathcal{G}_{n, \mathbf{d}_n}$ i transicions obtingudes mitjançant la realització d'una commutació triada uniformement a l'atzar. Es conjectura que la cadena és un mostrejador quasiuniforme eficient per a qualsevol seqüència de graus. Cooper, Dyer i Greenhill ([10]) van demostrar la conjectura per a grafs d -regulars, per a qualsevol $1 \leq d \leq n - 1$. Greenhill ([22]) ha estès aquest resultat per a seqüències de graus donades que satisfan $\Delta_n^2 = O(m_n)$, i, juntament amb Sfragara [24], a grafs dirigits.

Finalment, Steger i Wormald ([48]) van proposar un algorisme similar al model de configuració per generar grafs aleatoris quasiuniformes d -regulars, que ha estat estès a grafs aleatoris amb una seqüència de graus donada per Bayati, Kim i Saberi [2].

Agraïments

L'autor dona les gràcies a Oriol Serra pels comentaris que han ajudat a millorar aquest article. L'autor d'aquest article ha rebut suport de l'Agència Estatal de Investigació a través del projecte PID2020-113082GB-I00, del projecte RED2022-134947-T, i del programa Severo Ochoa i María de Maeztu per a centres i unitats d'excel·lència en R + D (CEX2020-001084-M).

L'autor expressa el seu agraïment pels comentaris d'un revisor anònim, així com als correctors d'estil, a la tècnica d'edició i composició i als editors de la revista, que han contribuït significativament a millorar aquest article.

Referències

- [1] ALBERT, R.; BARABÁSI, A.-L. «Statistical mechanics of complex networks». *Rev. Modern Phys.*, 74 (1) (2002), 47–97.
- [2] BAYATI, M.; KIM, J. H.; SABERI, A. «A sequential algorithm for generating random graphs». *Algorithmica*, 58 (4) (2010), 860–910.
- [3] BÉKÉSSY, A.; BÉKÉSSY, P.; KOMLÓS, J. «Asymptotic enumeration of regular matrices». *Studia Sci. Math. Hungar.*, 7 (1972), 343–353.
- [4] BENDER, E. A.; CANFIELD, E. R. «The asymptotic number of labeled graphs with given degree sequences». *J. Combinatorial Theory Ser. A*, 24 (3) (1978), 296–307.
- [5] BOLLOBÁS, B. «A probabilistic proof of an asymptotic formula for the number of labelled regular graphs». *European J. Combin.*, 1 (4) (1980), 311–316.
- [6] BOLLOBÁS, B.; RIORDAN, O. «An old approach to the giant component problem». *J. Combin. Theory Ser. B*, 113 (2015), 236–260.
- [7] CAI, X. S.; PERARNAU, G. «The giant component of the directed configuration model revisited». *ALEA Lat. Am. J. Probab. Math. Stat.*, 18 (2) (2021), 1517–1528.
- [8] CAI, X. S.; PERARNAU, G. «The diameter of the directed configuration model». *Ann. Inst. Henri Poincaré Probab. Stat.*, 59 (1) (2023), 244–270.
- [9] COMELLAS, F. «Models deterministes de xarxes complexes». *Butlletí de la Societat Catalana de Matemàtiques*, 22 (1) (2007), 23–43.
- [10] COOPER, C.; DYER, M.; GREENHILL, C. «Sampling regular graphs and a peer-to-peer network». *Combin. Probab. Comput.*, 16 (4) (2007), 557–593.
- [11] COOPER, C.; FRIEZE, A. «The size of the largest strongly connected component of a random digraph with a given degree sequence». *Combin. Probab. Comput.*, 13 (3) (2004), 319–337.
- [12] COULSON, M.; PERARNAU, G. «Largest component of subcritical random graphs with given degree sequence». *Electron. J. Probab.*, 28 (2023), article núm. 34, 28 p.
- [13] DHARA, S.; VAN DER HOFSTAD, R.; VAN LEEUWAARDEN, J. S. H.; SEN, S. «Critical window for the configuration model: finite third moment degrees». *Electron. J. Probab.*, 22 (2017), article núm. 16, 33 p.
- [14] DHARA, S.; VAN DER HOFSTAD, R.; VAN LEEUWAARDEN, J. S. H.; SEN, S. «Heavy-tailed configuration models at criticality». *Ann. Inst. Henri Poincaré Probab. Stat.*, 56 (3) (2020), 1515–1558.
- [15] DUMINIL-COPIN, H. «Sharp threshold phenomena in statistical physics». *Jpn. J. Math.*, 14 (1) (2019), 1–25.
- [16] ERDŐS, P.; RÉNYI, A. «On the evolution of random graphs». *Bull. Inst. Internat. Statist.*, 38 (1961), 343–347.

- [17] FEDERICO, L.; VAN DER HOFSTAD, R. «Critical window for connectivity in the configuration model». *Combin. Probab. Comput.*, 26 (5) (2017), 660–680.
- [18] FELLER, W. *An Introduction to Probability Theory and Its Applications*. Nova York: John Wiley, 1968.
- [19] FERNHOLZ, D.; RAMACHANDRAN, V. «The diameter of sparse random graphs». *Random Structures Algorithms*, 31 (4) (2007), 482–516.
- [20] FOUNTOULAKIS, N. «Percolation on sparse random graphs with given degree sequence». *Internet Math.*, 4 (4) (2007), 329–356.
- [21] FRIEZE, A.; KAROŃSKI, M. *Introduction to Random Graphs*. Cambridge: Cambridge University Press, 2016.
- [22] GREENHILL, C. «The switch Markov chain for sampling irregular graphs (extended abstract)». A: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. Filadèlfia, PA: Society for Industrial and Applied Mathematics (SIAM), 2015, 1564–1572.
- [23] GREENHILL, C. «Generating graphs randomly». A: *Surveys in Combinatorics 2021*. Cambridge: Cambridge University Press, 2021, 133–186. (London Math. Soc. Lecture Note Ser.; 470)
- [24] GREENHILL, C.; SFRAGARA, M. «The switch Markov chain for sampling irregular graphs and digraphs». *Theoret. Comput. Sci.*, 719 (2018), 1–20.
- [25] HATAMI, H.; MOLLOY, M. «The scaling window for a random graph with a given degree sequence». *Random Structures Algorithms*, 41 (1) (2012), 99–123.
- [26] VAN DER HOFSTAD, R. *Random Graphs and Complex Networks*. Vol. 1. Cambridge: Cambridge University Press, 2017. (Camb. Ser. Stat. Probab. Math.; 43)
- [27] VAN DER HOFSTAD, R. *Random Graphs and Complex Networks*. Vol. 2. Cambridge: Cambridge University Press, 2024. (Camb. Ser. Stat. Probab. Math.)
- [28] VAN DER HOFSTAD, R.; JANSON, S.; LUCZAK, M. «Component structure of the configuration model: barely supercritical case». *Random Structures Algorithms*, 55 (1) (2019), 3–55.
- [29] VAN DER HOORN, P.; OLVERA-CRAVIOTO, M. «Typical distances in the directed configuration model». *Ann. Appl. Probab.*, 28 (3) (2018), 1739–1792.
- [30] JANSON, S. «On percolation in random graphs with given vertex degrees». *Electron. J. Probab.*, 14 (5) (2009), 87–118.
- [31] JANSON, S. «The probability that a random multigraph is simple». *Combin. Probab. Comput.*, 18 (1-2) (2009), 205–225.
- [32] JANSON, S.; LUCZAK, M. «A new approach to the giant component problem». *Random Structures Algorithms*, 34 (2) (2009), 197–216.
- [33] JERRUM, M.; SINCLAIR, A. «Fast uniform generation of regular graphs». *Theoret. Comput. Sci.*, 73 (1) (1990), 91–100.

- [34] JOOS, F.; PERARNAU, G.; RAUTENBACH, D.; REED, B. «How to determine if a random graph with a fixed degree sequence has a giant component». *Probab. Theory Related Fields*, 170 (1-2) (2018), 263-310.
- [35] JOSEPH, A. «The component sizes of a critical random graph with given degree sequence». *Ann. Appl. Probab.*, 24 (6) (2014), 2560-2594.
- [36] KANNAN, R.; TETALI, P.; VEMPALA, S. «Simple Markov-chain algorithms for generating bipartite graphs and tournaments». *Random Structures Algorithms*, 14 (4) (1999), 293-308.
- [37] ŁUCZAK, T. «Sparse random graphs with a given degree sequence». A: *Random Graphs*. Vol. 2 (Poznań, 1989). Nova York: John Wiley & Sons, Inc., 1992, 165-182. (Wiley-Intersci. Publ.)
- [38] MCKAY, B. D. «Subgraphs of random graphs with specified degrees». *Congr. Numer.*, 33 (1981), 213-223.
- [39] MCKAY, B. D. «Asymptotics for symmetric 0-1 matrices with prescribed row sums». *Ars Combin.*, 19 (1985), 15-25.
- [40] MCKAY, B. D. «Subgraphs of random graphs with specified degrees». A: *Proceedings of the International Congress of Mathematicians. Volume IV*. Nova Delhi: Hindustan Book Agency, 2010, 2489-2501.
- [41] MCKAY, B. D.; WORMALD, N. C. «Uniform generation of random regular graphs of moderate degree». *J. Algorithms*, 11 (1) (1990), 52-67.
- [42] MCKAY, B. D.; WORMALD, N. C. «Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$ ». *Combinatorica*, 11 (4) (1991), 369-382.
- [43] MOLLOY, M.; REED, B. «A critical point for random graphs with a given degree sequence». *Random Structures Algorithms*, 6 (2-3) (1995), 161-179.
- [44] MOLLOY, M.; REED, B. «The size of the giant component of a random graph with a given degree sequence». *Combin. Probab. Comput.*, 7 (3) (1998), 295-305.
- [45] NEWMAN, M. E. J. «The structure and function of complex networks». *SIAM Rev.*, 45 (2) (2003), 167-256.
- [46] RIORDAN, O. «The phase transition in the configuration model». *Combin. Probab. Comput.*, 21 (1-2) (2012), 265-299.
- [47] RUÉ, J. «De la teoria de grafs clàssica a l'anàlisi de les grans xarxes». *Butlletí de la Societat Catalana de Matemàtiques*, 38 (1) (2023), 67-105.
- [48] STEGER, A.; WORMALD, N. C. «Generating random regular graphs quickly». *Combin. Probab. Comput.*, 8 (4) (1999), 377-396.
- [49] WORMALD, N. C. «Some problems in the enumeration of labelled graphs». Tesi doctoral. Newcastle: University of Newcastle, 1978.
- [50] WORMALD, N. C. «The asymptotic distribution of short cycles in random regular graphs». *J. Combin. Theory Ser. B*, 31 (2) (1981), 168-182.

- [51] WORMALD, N. C. «Models of random regular graphs». A: *Surveys in Combinatorics, 1999*. Cambridge: Cambridge University Press, 1999, 239-298. (London Math. Soc. Lecture Note Ser.; 267)
- [52] WORMALD, N. C. «Asymptotic enumeration of graphs with given degree sequence». A: *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*. Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd., 2018, 3245-3264.

DEPARTAMENT DE MATEMÀTIQUES
UNIVERSITAT POLITÈCNICA DE CATALUNYA
CAMPUS DIAGONAL SUD, EDIFICI U, C. DE PAU GARGALLO, 14,
08028 BARCELONA

I CENTRE DE RECERCA MATEMÀTICA
EDIFICI C, CAMPUS DE LA UNIVERSITAT AUTÒNOMA DE BARCELONA,
08193 BELLATERRA (CERDANYOLA DEL VALLÈS)

guillem.perarnau@upc.edu