

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
MATEMATIKAI INTÉZET



Ph.D. tézisfüzet

Nagy gráfok mintavételezése és lokális algoritmusok

Csóka Endre

Matematika Doktori Iskola

Igazgató: Laczkovich Miklós, a Magyar Tudományos Akadémia rendes tagja

Elméleti matematikus doktori program

Igazgató: Szűcs András, a Magyar Tudományos Akadémia rendes tagja

Témavezető:

Lovász László, a Magyar Tudományos Akadémia rendes tagja

Eötvös Loránd Tudományegyetem, Számítógéptudományi Tanszék
2012 december

A nagyon nagyméretű gráfok az élet számos területén előfordulnak. Ilyenek a biológiai hálózatok, mint például az idegrendszer, előjönnek fizikából, például a szilárd testek részecskéinek kölcsönhatásgráfjaként, de ide tartozik az internet, a világ közlekedési hálózata, az elektromos hálózat, és az emberek ismeretségi gráfja is. Ezek közül számos esetben nemcsak nagy méretű a gráf, hanem nehezen hozzáférhető is, ezért reménytelen a gráf pontos megismerése. Ez azonban nem zárja ki, hogy megértsük a gráf számos fontos tulajdonságát.

Először más tudományterületeken, főleg a fizikusok körében vált népszerűvé a nagy gráfok statisztikai típusú elemzése. Kutatásaik során ők a gráfoknak leginkább csak a fokszámeloszlását, esetleg a szomszédos csúcsok fokszámainak korrelációját és közös szomszédainak számát vizsgálják, és kizárólag ezen információk alapján adnak jellemzést a gráfokra. Gyakran ezt olyan formában teszik, hogy egy nehezen megismerhető gráfnak statisztikusan kimérik az említett paramétereit, generálnak ilyen paraméterekkel rendelkező gráfokat, gyakran nem is egyetlen véletlennel, hanem csak valamilyen heurisztikus módszerrel, és ezen gráfok tulajdonságaiból következtetnek az eredeti gráf tulajdonságaira. Ez a módszer meglepően jól működik a gyakorlatban, legalábbis az érintett területeken mindezt bevált módszerként tartják számon. A nagyon nagy gráfok elméletének megalkotását elsősorban az ezen jelenségek mögött rejlő matematikai háttér megismerése motiválta.

A kérdés matematikai megfogalmazása úgy szól, hogy a gráfoknak mik azok a tulajdonságai és paramétereit, amiket konstans méretű mintavételezéssel is jól meg lehet becsülni. Egy gráfparamétert akkor nevezünk becsülhetőnek, ha minden $\varepsilon > 0$ -ra létezik olyan konstans idejű mintavételező algoritmus, ami minden gráfhoz hozzárendel egy olyan számot ami várhatóan legfeljebb ε -nal tér el a gráf paraméterértékétől. Arra, hogy mit értünk mintavételezésen, kétféle modell is született.

Oded Goldreich, Shafi Goldwasser és Dana Ron [17] úgy definiálta a mintavételezést, hogy egyetlen véletlennel veszünk konstans sok pontot (akár ugyanazt többször is), és tekintjük az általuk feszített részgráfot. Ez annak a szemléletesebb módszernek az egyszerűbb, de ekvivalens erejű alakja, hogy konstans sokszor használhatjuk azt a két eszközt, hogy választunk egy véletlen pontot, illetve hogy két már kiválasztott pont között megnézzük, van-e él. Az erre a mintavételezésre épülő gráflimesz-elmélet Lovász Lászlónak Szegedy Balázssal [26], illetve Borgs, Chayes, T. Sós és Vesztergombival [6], [5] közös cikkein alapul. Ennek az elméletnek sikerült kimerítő választ adnia azokra a fő kérdésekre, hogy mik egy gráf tesztelhető tulajdonságai és becsülhető paramétereit. Ezáltal ez a kérdéskör, melyet a sűrű gráfok elméletének nevezünk, bizonyos szempontból megoldott és lezárt.

Ugyanakkor az elmélet új, érdekes összefüggésekre is rávilágított, ezért ezekben az irányokban továbbra is folyik a kutatás. A gráfokat általánosító grafingnak, és a gráfhomomorfizmusnak már a fogalma önmagában is hasznosnak bizonyult, új, egyszerűbb nyelvezetet és szemléletet biztosít számos már korábban is létező problémához. Ezekre példa a 8. fejezetben bemutatott sejtés, és az azt alátámasztó részeredmények. Ami azonban az eredeti motivációt illeti, ezzel a módszerrel csak a sűrű gráfokat lehet jellemezni, tehát azokat, amiknek $\Theta(n^2)$ éle van; sajnos a gyakorlatban előforduló gráfokban azonban korántsincs ennyi él, ezért ezekből a mintavétel nagy valószínűséggel üres gráfot ad.

Ezzel szemben Oded Goldreich és Dana Ron [18] modellje azt feltételezi, hogy a fokszámok korlátosak, illetve néha kevesebbet vár el, de azt mindig megköveteli, hogy a gráfnak $O(n)$ éle legyen. Ez sokkal alkalmasabbnak mutatkozik a való életbeli gráfok megértéséhez, cserébe ennek a matematikai leírása sokkal nehezebbnek bizonyul. Bár már számos fontos eredmény birtokában vagyunk, de ellentétben a sűrű gráfok esetével, ez a kérdéskör semmilyen értelemben sincsen még lezárva. Sőt, algoritmikusan eldönthetetlen kérdéseket is érint, mint azt a 4. fejezetben megmutatjuk. Disszertációm nagyobbik része ezzel a kérdéskörrel foglalkozik, melyet a ritka gráfok elméletének nevezünk.

Ebben a modellben a mintavételezés a következő. Egyetlen véletlennel kiválasztunk kons-

tans sok pontot, és mindnek tekintjük a konstans sugarú környezetét. Ez annak a szemléletesebb módszernek az egyszerűbb, de ekvivalens erejű alakja, hogy konstans sokszor használhatjuk azt a két eszközt, hogy választunk egy véletlen pontot, illetve hogy megnézzük egy már megtalált pont szomszédainak listáját.

Másféle struktúrák mintavételezéséről is született már elmélet, például Kohayakawa [20] elmélete permutációkról, Janson [22] elmélete részben rendezett halmazokról, Szegedy [32] elmélete Abel-csoportokról, illetve Gromov [19] és Elek [15] elmélete metrikus terekről. Mivel a sűrű gráfok elmélete az első, és az egyetlen teljes elmélet, ezért ennek tapasztalatai hasznos útmutatásokkal szolgálnak a többi elmülethez is. Ezek a fő eredmények nagyrészt az alábbiakban foglalhatóak össze.

A homomorfizmuszámoknak a sűrű és a ritka gráfok elméletében is fontos szerepük van. Egy F és egy G gráfra az éltartó $h: V(F) \rightarrow V(G)$ függvényeket homomorfizmusnak nevezzük, és $hom(F, G)$ jelöli ezeknek a számát. Formálisan

$$hom(F, G) = \left| \left\{ h: V(F) \rightarrow V(G) \mid \forall (x, y) \in E(F): (h(x), h(y)) \in E(G) \right\} \right|.$$

A sűrű és a ritka gráfok elméletében is egy G gráf mintavételezése egyenértékű azzal, hogy konstans sok F gráfra kapunk a közelítő értéket a sűrű esetben $t(F, G) = hom(F, G)/|V(G)|^{|V(F)|}$ -ről, a ritka esetben pedig $hom(F, G)/|V(G)|$ -ről. Tehát itt a normalizálás adja az egyetlen különbséget.

Jelölje \mathcal{G} a gráfok halmazát (izomorfia erejéig megkülönböztetve). Az alábbi \mathcal{T} topológiával azt próbáljuk meg kifejezni, hogy különböző gráfok mennyire különböztethetőek meg mintavételezéssel. Legyen \mathcal{T} az a legszűkebb topológia, melyre nézve a $t(F, \cdot)$ homomorfizmuszúség minden F esetén folytonos függvény. Avagy egy G_1, G_2, \dots gráfsorozat pontosan akkor konvergens a \mathcal{T} topológiában, hogyha minden F gráfra a $t(F, G_n)$ sorozat konvergens.

Könnyen látszik, hogy ha a gráf minden csúcsát ugyanannyiszor megtöbbszörözzük, akkor a topológia szerint ekvivalens gráfot kapunk, ezért az ilyen gráfokat nem is különböztetjük meg. Másrészt ha két gráf nem áll elő ugyanannak a gráfnak két megtöbbszörözéseként, akkor viszont a topológia szerint sem esnek egybe.

*Grafonok*nak hívjuk a mérhető, szimmetrikus $[0, 1]^2 \rightarrow [0, 1]$ függvényeket. A $\{0, 1, \dots, n-1\}$ csúcshalmaz feletti gráfoknak megfeleltetjük az alábbi grafont.

$$w(x, y) = \begin{cases} 1, & \text{ha van él } [nx] \text{ és } [ny] \text{ között;} \\ 0 & \text{egyébként} \end{cases}$$

Egy grafonból való mintavételezés azt jelenti, hogy veszünk konstans sok egyenletes véletlen számot $[0, 1]$ -ből, tekintjük az ilyen koordinátájú pontok alkotta részmatrixot, mint szomszéd-sági matrixot, és az így kapott gráf a mintevétel eredménye. Könnyen látszik, hogy egy gráfból, vagy az annak megfelelő grafonból való mintavételezés egyenértékű.

A $w_1, w_2: [0, 1]^2 \rightarrow [0, 1]$ grafonokat gyengén izomorfoknak nevezzük, ha léteznek $\sigma_1, \sigma_2: [0, 1] \rightarrow [0, 1]$ mértéktartó transzformációk, melyekre majdnem minden $(x, y) \in [0, 1]^2$ esetén $w_1(\sigma_1 x, \sigma_1 y) = w_2(\sigma_2 x, \sigma_2 y)$. A gyengén izomorf grafonokból vett mintavétel eredményének eloszlása is azonos.

Lovász és Szegedy [26] bizonyította be, hogy a $(\mathcal{G}, \mathcal{T})$ topologikus tér lezártja megegyezik a grafonok terének a gyenge izomorfiaival való faktorizáltjával. Azt is megmutatták, hogy ez a tér kompakt, melynek számos fontos következménye adódik, például az extrémális kombinatorika terén.

Az alábbiak szerint definiáljuk a vágástávolságot a grafonok terén.

$$\delta_{\square}(w_1, w_2) = \inf_{\sigma_1, \sigma_2} \sup_{S, T} \int_{x \in S} \int_{y \in T} w_1(\sigma_1 x, \sigma_1 y) - w_2(\sigma_2 x, \sigma_2 y) dx dy,$$

ahol $\sigma_1, \sigma_2: [0, 1] \rightarrow [0, 1]$ mértéktartó transzformációk, S és T pedig $[0, 1]$ -nek mérhető részhalmazai. Lovász és Szegedy [26] írta le az alábbi egyenlőtlenséget.

$$\forall F \in \mathcal{G}, w_1, w_2 \in W : \quad t(F, w_1) - t(F, w_2) \leq |E(F)| \cdot \delta_{\square}(w_1, w_2).$$

A másik irányhoz is van egy ennél sokkal bonyolultabb egyenlőtlenség. Ezekből együtt pedig már következik, hogy a δ_{\square} vágástávolság által indukált topológia megegyezik \mathcal{T} -vel. Más szóval, a G_n gráfsorozat pontosan akkor konvergens, ha a δ_{\square} metrika szerint Cauchy-sorozatot alkot.

Összefoglalva, a gráfok terét beágyasztuk egy szép és jól kezelhető kompakt metrikus térbe, amely kifejezi a gráfok távolságát a mintavételezésre nézve. Ezzel pedig azt kapjuk, hogy egy gráfparaméter pontosan akkor becsülhető, ha folytonosan kiterjeszthető a grafonok terére. Mivel ez a tér csak gráfokból és gráfsorozatok határértékeiből áll, ezért ha létezik folytonos kiterjesztés, akkor az egyértelmű is.

Az elmélet erejét jól érzékelteti az alábbi állítás. Egy gráf pontosan akkor kvázivéletlen, ha kicsi a vágás távolsága egy konstans grafontól. Más szóval, egy gráf pontosan akkor van közel a konstans p grafonhoz, ha a mintaeloszlása közel van a p paraméterű Erdős–Rényi véletlen gráféhoz.

Tekintsük most a ritka gráfok elméletét. Elég nagy gráfot véve, a mintavételező algoritmus 1-hez tartó valószínűséggel diszjunkt környezeteket választ ki. Ezért határértékben nézve egy mintavétel az alábbi még egyszerűbb alakra hozható. Valamilyen r és n konstansokra tekintjük az egyenletes véletlen pont r sugarú környezeteknek eloszlását, és ebből veszünk egy n elemű mintát.

Korlátos fokú gráfokhoz kétféle határobjektumot is rendelhetünk. Elek [13], illetve Aldous és Lyons [1] vezette be a grafingot, ami egy mértéktér feletti véges sok mértéktartó bijekciót jelent. A másik pedig a véletlen gyökeres gráf. Mindkettőnek megvan a maga előnye a másikkal szemben, de a mi céljainknak ez utóbbi fog megfelelni.

Egy központi kérdés, hogy a gráfnak mely valós paramétereit tudjuk mintavételezéssel megbecsülni. Ilyen paraméter lehet például, hogy a legnagyobb független halmaz a pontok mekkora arányát tartalmazza, vagy ugyanez lefoglaló pontokra, vagy párosításra; vagy az élek mekkora arányát kell elhagyni, hogy a gráf legfeljebb feleakkora komponensekre essen szét, vagy hogy síkgráf legyen. Formálisan, gráfparaméternek nevezzük a $p: \mathcal{G} \rightarrow \mathbb{R}$ függvényeket. Paraméterbecslő függvénynek pedig olyan függvényt nevezünk, amely a gráf konstans sok véletlen pontjának konstans sugarú környezetéhez hozzárendel egy valós számot. Egy p gráfparamétert pedig akkor nevezünk becsülhetőnek, ha minden $\varepsilon > 0$ -hoz létezik olyan paraméterbecslő függvény, mely minden G gráfból vett véletlen mintához legfeljebb ε várható hibával $p(G)$ -t rendeli.

Egy paraméter becsülhetősége olyasmint fejez ki, hogy a paraméter értékét meghatározza a gráf pontjai környezetének eloszlása. Ezt az alábbi módon lehet pontosabban megfogalmazni. Tekintsük a gyökeres, összefüggő, korlátos fokú véges és végtelen gráfoknak az eloszlásait, a konstans sugarú környezetek eloszlásai által generált szigma-algebrával. Nevezzük ezeket *véletlen gráfoknak*. Minden véges G gráfhoz hozzátartozik az a $H(G)$ véletlen gráf, hogy vesszük egy egyenletes véletlen pontnak, mint gyökérnek az összefüggőségi komponensét. Véletlen gráfok egy sorozatát akkor tekintjük konvergensnek, ha minden r -re az r sugarú környezetek eloszlása (ami egy véges dimenziós tér egy pontja) konvergens. Az így kapott topológikus teret X -szel jelöljük. Belátható, hogy egy p paraméter pontosan akkor becsülhető, ha létezik olyan folytonos valós $\hat{p}: X \rightarrow \mathbb{R}$ függvény a topológikus téren, ami minden gráfhoz tartozó véletlen gráfhoz a gráf paraméterét rendeli, vagyis

$$\forall G \in \mathcal{G}: p(G) = \hat{p}(H(G)). \quad (1)$$

Mindezek alapján már meg lehet mondani néhány paraméterről, hogy nem becsülhetőek. Vegyük például azt a kérdést, hogy az élek mekkora arányát kell elhagyni, hogy a gráf legfeljebb feleakkora komponensekre essen szét. d -reguláris véletlen gráfok növekvő csúcsszámú

sorozatán ez az arány pozitív számhoz tart. Ellenben, ha egy ugyanilyen gráf két példányának a diszjunkt unióját vesszük, ott 0 az arány, hiszen él elhagyása nélkül is két feleakkora komponensekre van bontva. A két gráf környezetének eloszlása ugyanakkor ugyanoda tart: a d -reguláris végtelen fához (mint koncentrált eloszlású véletlen gráfhoz). Ebből az adódik, hogy a becslhetőséghez \tilde{p} -nek a d -reguláris végtelen fán egyszerre kellene 0-nak és a fenti pozitív értéknek lennie, ami ellentmondás.

Egy sokkal kevésbé nyilvánvaló probléma a függetlenségi arány. A d -reguláris véletlen páros gráfon ez az arány $1/2$, míg a d -reguláris véletlen gráfon Bollobás Béla [4] eredménye szerint ez az arány $< 6/13$. Mivel ez a két gráfsorozat is a d -reguláris végtelen fához tart, ezért a függetlenségi arány sem lehet becslhető. Másrészt sok minden más, mint például a maximális párosítás relatív nagysága, becslhető. De a függetlenségi arány is becslhető bizonyos gráfosztályokon, például a síkgráfokon.

Valójában \hat{p} -t csak a véges gráfokhoz tartozó eloszlások lezártjára, vagyis $cl\{H(G): G \in \mathcal{G}\}$ -re kell kiterjeszteni, úgy is igaz lesz (1). Ez pedig egy lényegesen szűkebb tér. Ha például a gyöker 1 valószínűséggel harmadfokú, de minden szomszédja 1 valószínűséggel negyedfokú, akkor ezt nyilván nem kaphattuk, vagy közelíthettük meg egy gráf környezetének eloszlásaként, vagyis ez nincs $cl\{H(G): G \in \mathcal{G}\}$ -ben, ezért itt felesleges megadni \hat{p} -t.

Jelöljük a fokszámkorlátot d -vel. Vegyünk egy véletlen gráfot, és tegyük át a gyökeret valamelyik szomszédjába egyenként $1/d$ eséllyel, a maradék eséllyel pedig hagyjuk helyben a gyökeret. Ezáltal egy másik véletlen gráfot kaptunk. Ha a kettő eloszlása megegyezik, akkor a véletlen gráfot unimodulárisnak nevezünk. Az unimoduláris véletlen gráfok terét X_u -val jelöljük.

The random rooted graphs assigned to finite graphs are unimodular. By the conjecture of David Aldous and Russell Lyons, the other direction is also true for the closure, that is, $X_u = cl\{H(G): G \in \mathcal{G}\}$. Or equivalently, for all unimodular random rooted graphs U , there exists a sequence of graphs G_n such that the corresponding random rooted graphs $H(G_n)$ tend to U . This conjecture is already known in some special cases, e.g. when the distribution is concentrated on trees. [7] [14]

Könnyen látszik, hogy a gráfokhoz tartozó véletlen gráfok unimodulárisak. David Aldous és Russell Lyons sejtése szerint a másik irány is igaz, vagyis $X_u = cl\{H(G): G \in \mathcal{G}\}$. Avagy ekvivalens megfogalmazásban, minden U unimoduláris véletlen gráfhoz létezik olyan G_n gráfsorozat, hogy az azokhoz tartozó $H(G_n)$ véletlen gráfok U -hoz konvergálnak. Ezt eddig csak speciális esetekre sikerült bebizonyítani, például fákra koncentrált eloszlásokra. [7] [14]

Az Aldous-Lyons sejtésről azóta kiderült, hogy más témákkal is fontos kapcsolatban áll. Van például számos csoportelméleti sejtés, amit minden megszámlálható diszkrét csoportra igaznak sejtene, de csak az ún. szofikus csoportokra sikerült belátni. Mikhail Gromov tette fel a kérdést, hogy szofikus-e minden megszámlálható diszkrét csoport. Az volt az általános sejtés, hogy ez nem igaz, de ellenpéldát sem sikerült találni. Mint azt Elek Gábor a csoport Cayley-gráfjának vizsgálatával megmutatta, az Aldous-Lyons sejtés egy változatából következne, hogy Gromov kérdésére igen a válasz. Ezáltal pedig ezek az említett csoportelméleti sejtések is bizonyítást nyernének.

Emiatt és ettől függetlenül is természetesen felvetődik, hogy próbáljuk meg leírni az unimoduláris véletlen gráfok X_u terét, illetve a gráfokhoz tartozó véletlen gráfok terének $cl\{H(G): G \in \mathcal{G}\}$ lezártját. Sajnos azonban, mint azt a 4. fejezetben megmutatjuk, ezek nem lehetnek szép alakú halmazok, mivel az alakjukra vonatkozó bizonyos természetes kérdések algoritmikusan eldönthetetlenek. Megemlítendő, hogy ezen kérdések mindegyikére bizonyítottan ugyanaz a válasz a két halmazon, ami némi megerősítést is adja a sejtésnek.

A 2. fejezetben belátjuk, hogy ha az Aldous-Lyons sejtés nem igaz, akkor van olyan unimoduláris véletlen gráf, amit már egyetlen véletlen pont konstans sugarú környezetét véve is nagy biztonsággal meg tudunk különböztetni a véges gráfoktól. Ehhez eszközként megmutatjuk,

hogy egy korlátos fokú hálózatban a közelítőleg maximális folyam megkonstruálható determinisztikus lokális algoritmussal, amit a következőkben definiálunk, és ami a paraméterbecsléssel egyébként is szorosan összefüggő fogalom.

Lokális algoritmusok

A korlátos fokú gráfokon vett megosztott algoritmus a következőt jelenti. A bemenetként kapott gráf minden csúcsába elhelyezünk egy processzort, és csak azok tudnak közvetlenül kommunikálni, melyek szomszédos csúcsokban vannak elhelyezve. A végén minden processzor hoz egy döntést, és ez lesz az algoritmus kimenete. Ha például szeretnénk találni egy nagyméretű párosítást, akkor minden csúcsnak döntenie kell, hogy melyik szomszédjával legyen összepárosítva, vagy maradjon párosítatlanul. Ezeknek a döntéseknek természetesen konzisztensnek kell lenniük. A megosztott algoritmusokat többféle, nem egyenértékű módon is lehet definiálni, de esetünkben ezek a különbözőségek nem fognak jelentkezni.

Azokat a megosztott algoritmusokat nevezzük lokális algoritmusnak, melyek csak konstans számú egyidejű kommunikációs lépést használnak. Ennek egy ekvivalens megfogalmazása szerint egy algoritmus akkor lokális, ha minden csúcsához tartozó kimenet csak a csúcs konstans sugarú környezetétől függ (izomorfia erejéig).

A lokális algoritmusoknak Angluin [3], Linial [24], illetve Naor és Stockmeyer [29] voltak az úttörői. Auglin [3] mutatott rá a módszer korlátaira teljesen címkézetlen gráfok esetén. Linial arra az esetre is mutatott negatív eredményeket, ha minden csúcsnak egyedi címkéje van. Naor és Stockmeyer [29] bizonyította az első nem nyilvánvaló pozitív eredményeket. A lokális algoritmusok témájáról bővebb áttekintés Suomela-tól [31] olvasható.

A megosztott algoritmusoknak egy fontos kiterjesztését adja a véletlen megengedése, ami elsősorban a szimmetria megtörése érdekében szükséges. [2, 21, 27] Például tranzitív gráfok esetén egy lokális algoritmusnak minden csúcsban ugyanazt a döntést kellene meghoznia, ezért ezzel nem kaphatunk meg nem üres független halmazt. Ha azonban a véletlent is megengedjük, akkor már rögtön más a helyzet. A véletlen lokális algoritmusoknak egy ekvivalens megfogalmazása az alábbi. A csúcsokhoz független véletlen számokat rendelünk, és a csúcsbeli döntések függhetnek a konstans sugarú környezetben belüli véletlen számoktól is. Ha például azokat a csúcsokat választjuk ki, amelyekhez nagyobb számot sorsoltunk, mint minden szomszédjához, akkor ezzel már kapunk is egy a csúcsoknak várhatóan legalább $1/(d+1)$ arányát tartalmazó független halmazt.

A lokális algoritmusoktól általában nem optimális, hanem csak közelítő megoldást várunk. Azt mondjuk például, hogy lokális algoritmussal készíthető közel maximális méretű független halmaz, ha minden $\varepsilon > 0$ -ra létezik olyan lokális algoritmus, ami minden G gráfra mindig független halmazt ad, és ennek várható mérete legfeljebb εn -nel kisebb a G -beli maximális független halmaz méreténél.

A lokális algoritmusok szoros kapcsolatban állnak a paraméterbecsléssel, hiszen a vizsgált paraméterek között nagyon sok olyan van, ami maximalizációs problémából származik. Ilyen például a maximális párosítás, a maximális független halmaz, és a minimális vágás relatív mérete, a csúcsszámmal normalizálva. A kapcsolatot az adja, hogy ha létezik véletlen lokális algoritmus, ami közel optimális struktúrát ad, például egy közel maximális párosítást, akkor a maximális párosítás relatív mérete becsülhető. A paraméterbecslő függvény a következő. Vesszük konstans sok véletlen pontnak az ugyanakkora sugarú környezetét, mint amit a lokális algoritmus használ. Mindegyik környezetben a csúcsokra sorsolunk véletlen számokat, és megnézzük, hogy a lokális algoritmus párosítaná-e a gyökeret. A párosított csúcsok arányában fele pedig jó becslést ad a maximális párosítás relatív méretére. Huy Ngoc Nguyen és Krzyzstof Onak [30] bizonyította be számos gráfparaméter, köztük a maximális párosítás relatív méretének becsülhetőségét ezzel a módszerrel.

A lokális algoritmusoknak vehetjük azt az előfeldolgozós kiterjesztését is, hogy minden csúcsbeli döntés függhet magától a gráftól is, izomorfia erejéig. Tehát minden csúcs kap egyfajta központi információt, ami függhet a gráf egészétől (izomorfia erejéig), utána végrehajtják a konstans sok kommunikáció lépést, és végül meghozzák a döntéseiket. Mindezt egy szolgáltatásként is leírhatjuk. Van egy központ, ami rendelkezik valamilyen információval a gráf egészéről. A független halmaz példáját véve, minden csúcs „névtelenül” megkérdezheti a központtól, hogy ő benne van-e a független halmazban, és a központnak erre csak a gráfról szerzett információja, és a csúcs konstans sugarú környezete alapján kell válaszolnia. Ezeknek a válaszoknak pedig konzisztenseknek kell lenniük, tehát semelyik két szomszédos csúcs sem kaphat „igen” választ, miközben az „igen” választ kapó csúcsok arányának meg kell közelítenie a függetlenségi arányt.

Elek Gábor [13] a szubexponenciális növekedésű gráfok körében mutatta meg számos paraméter becslhetőségét az alábbi előfeldolgozást használva. Vette konstans sok véletlen pont konstans sugarú környezetét, és ezt a statisztikát kapta meg minden csúcs. Vagyis minden csúcsban a saját konstans sugarú környezete, a benne szereplő véletlen számok, és ezen közös környezetstatisztika alapján kell meghozni a döntést. Ennek az az értelme, hogy ha létezik ilyen típusú, közel optimális lokális algoritmus, még abból is következik a megfelelő gráfparaméter becslhetősége. A becsléshez vesszük konstans sok véletlen pont konstans sugarú környezetét a környezetstatisztikához, utána veszünk még egy pontot és környezetét, odaadjuk neki a környezetstatisztikát, és kiszámoljuk a döntést ebben a pontban. Ezt az egészet konstans sokszor megismételjük, és az eredményekből nyerhetünk egy becslést a paraméterre. Ezt az észrevételt Borel-gráfokról szóló tételek lokális algoritmusokra való lefordítására is használták. [16]

In Chapter 3, we compare the strengths of the different generalizations of local algorithms. In particular, we show that preprocessing is useless. More precisely, if there exists a local algorithm using preprocessing, then there exists another local algorithm with the same radius, in which the only „preprocessing” is a random variable with a continuous distribution, and this provides an output with at most the same error from the optimum, in expectation. The use of this random variable can also be interpreted as follows. We draw a local algorithm from a given probability distribution of local algorithms, and we use this at each node.

A 3. fejezetben összevetjük, hogy a lokális algoritmusoknak ezen általánosításai mennyiben is növelik a módszer erejét. Belátjuk azt a meglepő állítást, hogy az előfeldolgozás semennyit sem segít. Pontosabban, minden olyan lokális algoritmushoz, ami használ előfeldolgozást, van olyan lokális algoritmus is, ami „előfeldolgozásként” csak egy globális véletlen számot használ, és mégis ugyanakkora sugárt használva előállít egy várhatóan legalább ugyanakkora értékű struktúrát. Ennek a globális véletlen változónak a használata megfelel annak, hogy lokális algoritmusok egy valószínűségi eloszlásából kisorsolunk egyet, és ezt használjuk minden csúcsban.

Visszatérve a téma eredeti motivációjára, a fizikusok körében él egy olyan tapasztalat, hogy a valóságban előforduló gráfokban is, és a véletlenül generált gráfokban is minden lokális. Bár ez az állítás matematikailag nagyon nem precíz, mégis bármikor felteszik, amikor csak szükségük van rá, és elégedettek szoktak lenni a kapott eredményekkel. Egy ilyen eredményre mutatunk egy példát a 7. fejezetben. Itt egy „fázisátalakulás” jelenségét írjuk le, és a matematikai elemzés során fel kell tennünk valaminek a lokalitását. Ennek megfelelően a cikkben megkapott elméleti eredmények egyeznek is a véletlen gráfokon számolt szimulációs eredményekkel.

We try to find a true mathematical statement expressing the experience that „everything is local” on random graphs. In addition, this could open the door to describe the terms „typical” or „real-life” graphs better: a graph is as much „typical” as true that „everything is local” on it. There are many statements about graphs which are not true for all graphs, but which are true for the typical graphs. About such statements, all what we can do is proving that this is true for uniform random graphs on a large vertex set, with probability tending to 1. The theoretical imperfectness of this technique is pointed out by computer programs that are able to distinguish between uniform random graphs and different kind of real-life graphs, with high

probability. Therefore, something to be true for almost all graphs does not mean that it is true for the typical graphs. The curiosity of the result in Chapter 7 is it proves something for typical graphs, but in an unusual sense.

Ezt a tapasztalatot megpróbáljuk általános matematikai állításként is megfogalmazni. Olyan jellegű állítást várunk, hogy véletlen gráfokon „minden lokális”. Ezzel egyszersmind lehetőségünk nyílna arra is, hogy jobban megfogjuk a „tipikus”, a „gyakorlatban előforduló” gráf fogalmát: egy gráf annyira „tipikus”, amennyire igaz rá, hogy rajta „minden lokális”. Ez azért lenne nagyon hasznos, mert számos olyan állítást ismerünk, ami nem igaz minden gráfra, de igaz a tipikus gráfokra. Ilyen esetben jelenleg általában csak annyit tudunk tenni, hogy belátjuk, hogy egy egyenletes véletlennel választott gráfra nagy eséllyel igaz. Ennek a módszernek a tökéletlenségét jól szemlélteti, hogy vannak programok, amik nagyon nagy pontossággal meg tudják különböztetni a gyakorlatból jövő gráfokat az egyenletes véletlennel választott gráfoktól. Ezért az, hogy valami igaz a legtöbb gráfra, még nem kell, hogy azt jelentse, hogy igaz lenne a gyakorlatban előforduló gráfok többségére is. A 7. fejezetbeli eredménynek az az érdekessége, hogy itt a tipikus gráfokra bizonyítunk valamit, a tipikusságot egy szokatlan értelemben véve.

Ami azonban a tipikusság remélt definícióját illeti, mindeddig sajnos csak sejtések megfogalmazásáig jutottunk, és azt sem tudhatjuk, hogy megtaláltuk-e már a megfelelő állítást. A jó sejtés megtalálásához érdemes a legegyszerűbb speciális esetből indulni.

Talán a legegyszerűbb kérdés a függetlenségi arány 3-reguláris nagykörű gráfokon. Egyrészt azért érdemes ezt vizsgálni, mivel itt a környezetstatisztika a 3-reguláris végtelen gráfra koncentrálódik. Másrészt a függetlenségi arányt nem határozza meg a környezetstatisztika. Harmadrészt, ez esetben világos, hogy mit értünk véletlen gráfon, hiszen az egyenletes véletlen 3-reguláris gráfok növekvő sorozata ugyanehhez a környezetstatisztikához tart.

Egy lokális algoritmus által generált függetlenségi arány várható értéke csak a gráf környezeteinek eloszlásától függ. Továbbá, adott környezetstatisztika mellett a lokális algoritmussal elérhető arányok szuprénuma alsó becslést ad az ilyen környezetstatisztikájú gráfok függetlenségi arányára is. A fenti tapasztalatok pedig azt sugallják, hogy véletlen gráfokon lokális algoritmussal megkonstruálható egy közel maximális független halmaz. Ezt a sejtést két részre bonthatjuk. Az egyik szerint adott környezetstatisztika mellett a véletlen gráfon a legkisebb a függetlenségi arány. A másik szerint ez a legkisebb arány lokális algoritmussal el is érhető. Mindkét állítás némileg meglepően hangzik, ezért ezen sejtések helyességének eldöntése megerősítheti, vagy cáfolhatja, hogy ez-e a jó út a jelenség megértéséhez. A 5. fejezetben mutatunk egy lokális algoritmust, ami a 3-reguláris nagykörű gráfokon az eddig ismert legnagyobb függetlenségi arányt éri el.

A lokális algoritmusoknak egy természetes limeszét adják a factor of iid folyamatok. Ez azt jelenti, hogy itt is hozzárendelünk a csúcsokhoz független azonos eloszlású véletlen számokat, és a gyökeres végtelen környezet egy természetes értelemben vett mérhető függvényének kell lennie a csúcshelyi döntésnek. Ezzel a nyelvezettel úgy fogalmazható át az előző kérdés, hogy mi a legnagyobb elérhető függetlenségi arány a 3-reguláris végtelen fán factor of iid folyamattal.

Bár a közelítőleg maximális folyam elérhető lokális algoritmussal, az a kérdés még nyitva áll, hogy mikor létezik factor of iid maximális párosítás. Ez csoportelméleti kérdésekkel is kapcsolatban áll, azok Cayley-gráfján keresztül, és még sok más kérdéssel is. Például Laczkovich Miklós konstrukciójában [23] Tarski körnégszögesítésére, a fő ötlet egy teljes párosítás keresése volt egy véges sok eltolás által generált grafingban. Az azonban még nyitott, hogy létezik-e a körnek Lebesgue-mérhető átdarabolása is a négyzetbe. Egy megfelelő factor of iid teljes párosítás pedig alkalmas módszer lehetne a mérhető átdarabolásra.

Russell Lyons és Fedor Nazarov [28] bebizonyították, hogy minden nemamenábilis csoport páros Cayley-gráfján létezik factor of iid teljes párosítás. A nem páros eset ugyanakkor lényegesen bonyolultabb, de a 6. fejezetben belátjuk, hogy létezik factor of iid teljes párosítás minden nemamenábilis Cayley-gráfon, sőt, minden nemamenábilis csúcstranzitív unimoduláris véletlen

gráfon is.

A 2., 3. és 4. fejezet saját egyszerzős cikkeimen alapul [9–11]. Az 5. fejezet közös munka Gerencsér Balázssal, Harangi Viktorral és Virág Bálinttal, a 6. fejezet közös munka Lippner Gáborral [12], a 7. fejezet közös munka Pósfai Mártonnal és Yang-Yu Liuval [25], a 8. fejezet pedig közös munka elsősorban Hubai Tamással és Lovász Lászlóval, továbbá Omar Antolín Camarena-val és Lippner Gáborral [8].

Hivatkozások

- [1] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, 12:no. 54, 1454–1508, 2007.
- [2] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of algorithms*, 7(4):567–583, 1986.
- [3] D. Angluin. Local and global properties in networks of processors. In *Proceedings of the twelfth annual ACM symposium on Theory of computing*, pages 82–93. ACM, 1980.
- [4] B. Bollobás. The independence ratio of regular graphs. *Proc. Amer. Math. Soc.*, 83(2):433–436, 1981.
- [5] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, B. Szegedy, and K. Vesztergombi. Graph limits and parameter testing. In *Annual ACM Symposium on Theory of Computing: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, volume 21, pages 261–270, 2006.
- [6] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. Counting graph homomorphisms. *Topics in discrete mathematics*, pages 315–371, 2006.
- [7] L. Bowen. Couplings of uniform spanning forests. *Proceedings of the American Mathematical Society*, pages 2151–2158, 2004.
- [8] O.A. Camarena, E. Csóka, T. Hubai, G. Lippner, and L. Lovász. Positive graphs. *arXiv preprint arXiv:1205.6510*, 2012.
- [9] E. Csóka. Maximum flow is approximable by deterministic constant-time algorithm in sparse networks. *arXiv preprint arXiv:1005.0513*, 2010.
- [10] E. Csóka. Local algorithms with public randomisation on sparse graphs. *arXiv preprint arXiv:1202.1565*, 2012.
- [11] E. Csóka. An undecidability result on limits of sparse graphs. *The Electronic Journal of Combinatorics*, 19(2):P21, 2012.
- [12] E. Csóka and G. Lippner. Invariant random matchings in cayley graphs. *arXiv preprint arXiv:1211.2374*, 2012.
- [13] G. Elek. Note on limits of finite graphs. *Combinatorica*, 27(4):503–507, 2007.
- [14] G. Elek. On the limit of large girth graph sequences. *Combinatorica*, 30(5):553–563, 2010.
- [15] G. Elek. Samplings and observables. convergence and limits of metric measure spaces. *arXiv preprint arXiv:1205.6936*, 2012.

- [16] G. Elek and G. Lippner. Borel oracles. an analytical approach to constant-time algorithms. In *Proc. Amer. Math. Soc.*, volume 138, pages 2939–2947, 2010.
- [17] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 339–348. IEEE, 1996.
- [18] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2):302–343, 2002.
- [19] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Birkhäuser Boston, 2006.
- [20] C. Hoppen, Y. Kohayakawa, C.G. Moreira, B. Rath, and R. Menezes Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory, Series B*, 2012.
- [21] A. Israeli and A. Itai. A fast and simple randomized parallel algorithm for maximal matching. *Information Processing Letters*, 22(2):77–80, 1986.
- [22] S. Janson. Poset limits and exchangeable random posets. *Combinatorica*, pages 1–35, 2009.
- [23] M. Laczkovich. Equidecomposability and discrepancy; a solution of tarski’s circle squaring problem. *J. Reine Angew. Math.*, 404:77–117, 1990.
- [24] N. Linial. Locality in distributed graph algorithms. *SIAM Journal on Computing*, 21:193, 1992.
- [25] Yang-Yu Liu, Endre Csóka, Haijun Zhou, and Márton Pósfai. Core percolation on complex networks. *Phys. Rev. Lett.*, 109:205703, Nov 2012.
- [26] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.
- [27] M. Luby. A simple parallel algorithm for the maximal independent set problem. In *Proceedings of the seventeenth annual ACM symposium on Theory of computing*, pages 1–10. ACM, 1985.
- [28] R. Lyons and F. Nazarov. Perfect matchings as iid factors on non-amenable groups. *European Journal of Combinatorics*, 32(7):1115–1125, 2011.
- [29] M. Naor and L. Stockmeyer. What can be computed locally? In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 184–193. ACM, 1993.
- [30] H.N. Nguyen and K. Onak. Constant-time approximation algorithms via local improvements. In *Foundations of Computer Science, 2008. FOCS’08. IEEE 49th Annual IEEE Symposium on*, pages 327–336. IEEE, 2008.
- [31] J. Suomela. Survey of local algorithms, 2009.
- [32] B. Szegedy. Gowers norms, regularization and limits of functions on abelian groups. *arXiv preprint arXiv:1010.6211*, 2010.