

Hatékony tulajdonságtesztelés sűrű gráfokban

Fekete Panna Tímea 

HUN-REN Rényi Alfréd Matematikai Kutatóintézet, Budapest, Magyarország
Budapesti Műszaki és Gazdaságtudományi Egyetem, Villamosmérnöki és Informatikai Kar,
Számítástudományi és Információelméleti Tanszék, Budapest, Magyarország
Levelező szerző, e-mail: feketepanna.timea@renyi.hu

Beérkezett: 2025. július 31.; elfogadva: 2026. február 3.

Összefoglalás

Nagy, sűrű hálózatok számos tudományterületen jelen vannak, így általános vizsgálatuk kiemelt jelentőséggel bír. Méretüknél fogva teljes egészében nehezen vizsgálhatók, emiatt mintavételezéssel kapott kisebb hálózatok tulajdonságaiból következtethetünk az eredeti hálózat struktúrájára. Jelen dolgozatban a hálózatokat a gráfelmélet segítségével formalizáljuk. Bizonyítjuk, hogy részbenrendezett halmazok monoton tulajdonságai könnyen tesztelhetők kis minták alapján. Továbbá kiterjesztünk két mintavételezési lemmát nemkorlátos magfüggvényekre, igazolva egy majdnem biztos konvergenciát a vágástávolságra. Összességében jelen dolgozat hozzájárul újabb komplex hálózatcsaládok globális tulajdonságainak feltárásához lokális minták segítségével.

Kulcsszavak: polinomiális törlési lemma, részbenrendezett halmaz, mintavételezési lemma, magfüggvény

Efficient property testing for dense graphs

Panna Tímea Fekete

HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary
Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics,
Department of Computer Science and Information Theory, Budapest, Hungary

Summary

Large-scale networks are ubiquitous in nature and technology, spanning domains such as neuroscience, epidemiology, transportation, and social sciences. Their analysis often requires the study of their graph counterparts, whose complex global behavior is often computationally prohibitively expensive to study directly. A common strategy is to analyze smaller substructures or simplified models of these graphs. However, this raises a fundamental question: to what extent do these smaller samples retain the key properties of the original network?

This current research focuses on dense graphs, where the number of edges grows quadratically with the number of vertices. We further develop efficient tools for understanding their structure through sampling and removal lemmas. Building on Szemerédi's regularity lemma—a cornerstone in the field of graph theory—we address its limitations by constructing more practical approaches that require significantly fewer computational resources.

We investigate monotone classes of finite posets (that is, partially ordered sets closed under taking subposets) and prove that these are easily testable. This means that their structural properties can be reliably inferred by sampling only a number of vertices polynomial in the error parameter, a substantial improvement over earlier approaches relying on tower-type bounds. We prove this via a polynomial removal lemma for posets. Furthermore, we show that for every monotone class of posets, there exists an h such that the class is indistinguishable with strong testing from the class of C_h -free posets, where C_h is a chain of length h . Using this, we also obtain a two-sided test, with sampling size $O(\varepsilon^{-1})$. Moreover, the test is essentially one-sided.

We examine the sampling method from a graph limit point of view. The Counting Lemma and the Inverse Counting Lemma establish a connection for comparing two dense large graphs by relating homomorphism densities and cut distance. It has been shown that while the Counting Lemma holds for unbounded kernels (i.e., limits of multi-graphs), the Inverse Counting Lemma does not. The Inverse Counting Lemma is built on the First and Second

Sampling Lemmas, which we demonstrate also hold for unbounded kernels. The First Sampling Lemma gives bounds on the difference between the cut norms of the original and the sampled kernel. The main feature of the cut norm is that it emphasizes structure over randomness, and “filters out noise”. The Second Sampling Lemma gives an upper bound on the so-called ‘cut-metric’ of the original and the sampled kernel, i.e., the cut norm of the difference of the original and the sampled kernel where one can use any measure preserving transformation on the sample. We also derive that for any kernel in L^p ($p > 4$), the samples converge almost surely in the cut metric to the original kernel.

Overall, the presented methods and theorems contribute to the more efficient and accurate study of the structure of large networks.

Keywords: polynomial removal lemma, poset, sampling lemma, kernel

Bevezetés

Nagy struktúrák szinte minden tudományterületen jelen vannak, a humán és társadalomtudományok területén éppúgy, mint az élettelen vagy élő természettudományokban. A humán tudományokban segítséget nyújthatnak társadalmi hálózatok szerkezetének és dinamikájának megértésében, az információ terjedésének és torzulásának tanulmányozásában vagy gazdasági rendszerek működésének feltérképezésében. Műszaki tudományokban többek között elektromos hálózatok hibatűrésének vizsgálatára, közlekedési rendszerek robusztusságának modellezésére vagy anyagi rendszerek (halmaz)állapotváltozásainak megértésére használják őket. Az élettudományokban több hierarchikus szinten megtalálhatók: molekuláris szinten (fehérje–fehérje kölcsönhatási hálózatok vagy génhálózatok vizsgálata), az egyed szintjén (az emberi agyi struktúrák közötti kapcsolat vagy az *ecetmuslica* – *Drosophila melanogaster* – teljes agyi konnektóm, azaz ideghálózati térképének modellezése), valamint az egyének feletti szinteken is (fertőző betegségek, például COVID–19 terjedésének modellezése). A felsorolás természetesen nem teljes; annak bemutatására szolgál, hogy a hálózatokkal valamilyen absztrakciós szinten minden tudományterület foglalkozik. E hálózatok vizsgálata sokszor nemcsak a modellezni kívánt rendszerek pusztá méretéből adódóan okoz nehézséget, hanem a lokális és globális szerkezet áttekintése és megértése is komoly kihívás jelent.

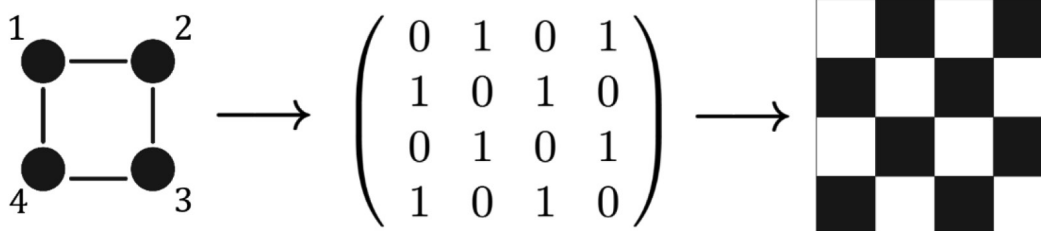
A hálózatokat gráfokkal, azaz csúcsokkal és azokat összekötő élekkel modellezhetjük. A különböző tudományterületeken fellelhető komplex rendszerek mögött megbúvó gráfstruktúrák lehetővé teszik, hogy hasonló matematikai eszközkészlet használatával közelebb kerüljünk a megértésükhöz. Az így képzett gráfok sok esetben

hatalmas méretűek, emiatt teljes feltérképezésük (szinte) lehetetlen. Felmerül a kérdés, hogy mégis hogyan ismerhetjük meg ezeknek a rendszereknek a szerkezetét. Tudunk-e az egész hálózat struktúrájára következtetni bizonyos jól megválasztott részstruktúrák vizsgálata révén? Milyen információt veszítünk el eközben? Jelen tanulmányban ezekre és ehhez hasonló kérdésekre keressük a választ.

Sűrű gráfok, grafonok, magfüggvények

Egy irányított gráf alatt csúcsok és csúcspárok (irányított élek) halmazát értjük. Gráfokat megadhatunk az úgynevezett szomszédsági mátrixuk segítségével is, ahol az i . sor j . eleme az (i, j) élek száma (egyszerű gráfok esetén 0 vagy 1). Irányítatlan gráfok szomszédsági mátrixa szimmetrikus. Az irányítatlan gráfokra tekinthetünk úgy, mint irányított gráfokra: az irányítatlan gráf minden éle két irányított éllel helyettesíthető. Az 1. ábrán egy $\{1, 2, 3, 4\}$ csúcs-halmazú, $\{(1, 2), (2, 3), (3, 4), (1, 4)\}$ élhalmazú irányítatlan gráf szemléltetését láthatjuk, illetve a hozzá tartozó szomszédsági mátrixot. A szomszédsági mátrixot pixeles ábrán is ábrázolhatjuk; erre is láthatunk példát az 1. ábrán, illetve a későbbiekben még visszatérünk rá.

Nagy gráfok az élek aszimptotikus száma szerint osztályozhatók. Sűrűnek nevezünk egy n csúcsú gráfot, ha az éleinek száma n^2 -tel arányos; ritkának nevezzük, ha az éleinek száma n -nel arányos, ahogy $n \rightarrow \infty$. Előfordulhat, hogy egy gráf se nem ritka, se nem sűrű. A ritka és a sűrű gráfok viselkedése jelentősen eltér, így elemzésük különböző módszereket igényel. A tanulmány további részében kizárólag sűrű (irányított és irányítatlan) gráfokkal foglalkozunk.



1. ábra | Egy $\{1, 2, 3, 4\}$ csúcs-halmazú, $\{(1, 2), (2, 3), (3, 4), (1, 4)\}$ élhalmazú gráf szemléltetése, a hozzá tartozó szomszédsági mátrix, illetve grafon
 Forrás: Glasscock 2015

A gráflimeszelméletben a gráfokat (izomorfizmus, azaz „csúcsok átnevezése” erejéig) a szomszédsági mátrixukkal azonosítjuk. A homomorfizmus-sűrűség alapú konvergencia segítségével a gráfok határértékei szimmetrikus, mérhető $[0,1]^2 \rightarrow [0,1]$ függvényekkel írhatók le, ezeket grafonoknak nevezzük. Ezekre a függvényekre tekinthetünk úgy, mint a szomszédsági mátrixok folytonos analógiája. A nagy gráfok sorozatának az úgynevezett vágásnorma szerinti limeszobjektumát Borgs, Chayes, Lovász, T. Sós, Szegedy és Vesztergombi vezették be és dolgozták ki (Borgs et al. 2008). A grafonokra tekinthetünk úgy, mint „végtelen, (súlyozott) gráfokra”, amelyek csúcshalmaza a $[0,1]$ intervallum, a függvényérték pedig az élekre adott súlyokat adja meg bármely két csúc között. A véges gráfok is leírhatók grafonokkal; erre láthatunk példát az 1. ábrán, a pixeles ábrázolás ezt szemlélteti. Egy n csúcsú véges gráf minden csúcsa megfeleltethető egy $1/n$ hosszú intervallumnak; a függvényérték az i . intervallum és j . intervallum tetszőleges pontjai között 0 („fehér”), ha (ij) nem él, illetve 1 („fekete”), ha él. A 2. ábrán teljes páros gráfokból készített grafonok ábrázolását láthatjuk. Vegyük észre, hogy a csúcsok elnevezése befolyásolja a grafont, azonban a limeszobjektumot úgy definiálták, hogy ne befolyásolja őket. Grafonokról röviden Glasscock írt (Glasscock 2015), magyar nyelvű fordítása is elérhető (Tóth 2024). Bővebben pedig Lovász 2012-es könyvében olvashatunk róluk (Lovász 2012).

Grafonok általánosításaként kaphatjuk meg az úgynevezett kerneleket, azaz magfüggvényeket, amelyek szimmetrikus, korlátos $[0,1] \times [0,1] \rightarrow \mathbb{R}$ függvények. Amennyiben az egyszerű gráfok világából a súlyozott multigráfok világába lépünk, akkor limeszobjektumként nemkorlátos L^p kerneleket kapunk, valamilyen véges p -re (Borgs et al. 2018, 2019).

Kutatásaim során sűrű gráfok különféle tulajdonságait vizsgáltam két különböző megközelítéssel. Az egyik

esetben tulajdonságokat vizsgálunk: megnézzük, hogy mely tulajdonságok esetén tudunk következtetni egy bizonyos gráfosztály minden gráfa esetében a gráfból véletlenszerűen választott (mintavételezett) kisebb, úgynevezett részgráf alapján az eredeti gráf tulajdonságára. A másik esetben pedig azt vizsgáljuk, hogy egy gráf milyen messze kerülhet a mintavételezett gráftól az úgynevezett vágásnorma szerint, amely a fehér zaj kiszűrésére alkalmas norma. Mindkét probléma szerves részét képezi doktori disszertációmnak, így a tanulmány doktori tézisfüzetemre, tézisemre épül (annál azonban bővebb).

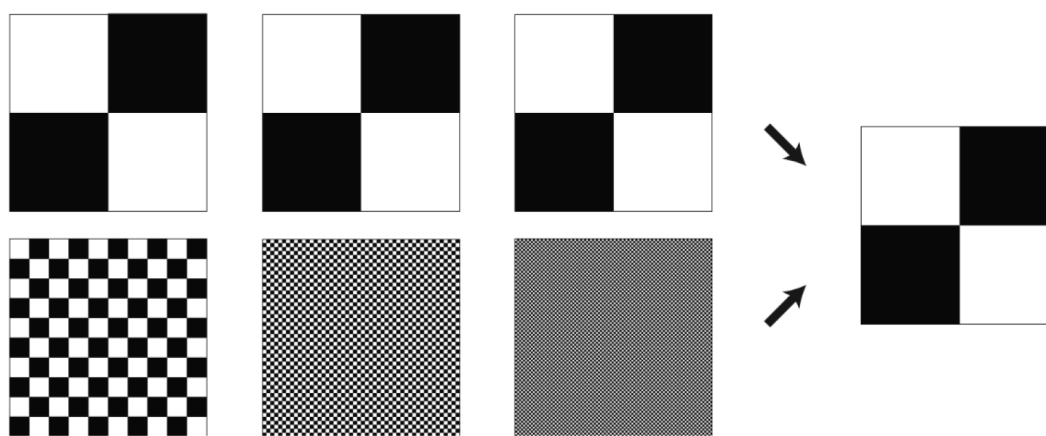
Mindegyik, alább ismertetett probléma esetében alkalmazható lenne a Szemerédi-féle regularitási lemma, azonban ez minden esetben jóval gyengébb eredményhez vezetne. Jelen esetben a hatékony becslés a cél, ami azt jelenti, hogy a mintavételezés méretére nagy valószínűséggel polinomiális korlátot tudunk adni.

Tulajdonságtesztelés mintavételezéssel

Nagy irányított gráfokat vizsgálunk kis méretű mintavételezéssel. Egy irányított gráf részgráfa egy olyan irányított gráf, amelynek csúcsai, illetve élei részalmazai az eredeti gráf csúcsainak, illetve éleinek. Egy gráf feszített részgráfa alatt egy olyan gráfot értünk, amelynek csúcshalmaza részalmaz a gráf csúcshalmazának, élhalmaza pedig a gráfban ezen csúcsok által feszített élek. A fejezet további részében gráf alatt irányított gráfot értünk.

Tulajdonság

Gráfok tulajdonságán véges gráfok egy izomorfizmusra zárt osztályát értjük. Ez azt jelenti, hogy a teljes gráf ismeretében minden gráfról egyértelműen el lehet dönteni, hogy rendelkezik-e az adott tulajdonsággal, vagy sem; továbbá, ha a gráf csúcsait „átnevezzük”, az nem változtat azon, hogy rendelkezik-e a tulajdonsággal. Egy



2. ábra

Az első, második és harmadik oszlopokban rendre 10, 50 és 100 csúcsú teljes páros gráfokból készített grafonok ábrázolását láthatjuk. A páros gráfok limeszobjektumának ábrázolása a negyedik oszlopban látható.

Forrás: Glasscock 2015

tulajdonságot monoton tulajdonságnak hívunk, ha él- és csúcshagyásra zárt gráfosztály, vagyis ha egy gráf rendelkezik a tulajdonsággal, akkor minden (nem feltétlenül feszített) részgráfja is.

Irányítatlan gráfok egy monoton tulajdonsága például a háromszög-mentesség: háromszög-mentesnek hívunk egy irányítatlan gráfot, ha nem tartalmaz részgráfként 3 csúcshalmaz teljes irányítatlan gráfot. Az 1. ábrán látható gráf például háromszög-mentes, illetve minden részgráfja is háromszög-mentes. Azonban ha az $(1,4)$ él is az élhalmaz része lenne, akkor már nem lenne háromszög-mentes.

Tulajdonságtól való távolság

Egy $V(G)$ csúcshalmazú G gráf ε -távol van egy tulajdonságtól, ha semelyik $V(G)$ csúcshalmazú gráf, amelynek élhalmaza legfeljebb $\varepsilon V(G)^2$ élben tér el G élhalmazától, nem rendelkezik a tulajdonsággal.

Monoton gráftulajdonságok vizsgálata esetén a definíció alapján a tulajdonságtól való távolsághoz elegendő az élek törlését vizsgálni, élek hozzáadása esetén a kapott gráf ugyanis biztosan nem rendelkezhet a tulajdonsággal.

Célunk, hogy ha egy gráfról tudjuk, hogy vagy rendelkezik egy adott tulajdonsággal, vagy távol van tőle, akkor mintavételezéssel el tudjuk dönteni, hogy nagy valószínűséggel melyik igaz rá.

Tesztelhetőség

Egy tulajdonságot erősen tesztelhetőnek nevezünk, ha tetszőleges pozitív valós ε esetén létezik $f(\varepsilon)$ egész szám, amelyre teljesülnek a következők:

- ha G egy gráf, amelyre teljesül a tulajdonság, akkor tetszőleges $f(\varepsilon)$ csúcsa által feszített részgráfjára is teljesül;
- továbbá, ha egy G gráf ε -távol van a tulajdonságtól, akkor legalább $1/2$ valószínűséggel egy tetszőleges $f(\varepsilon)$ csúcsa által feszített részgráfjára sem teljesül a tulajdonság.

A kis elemszámú mintavétellel történő tulajdonságtesztelés egyik klasszikus eredménye a Ruzsa és Szemerédi által bizonyított háromszög-törlési lemma (Ruzsa–Szemerédi 1978). A lemma szerint tetszőleges $\varepsilon > 0$ esetén, ha egy $V(G)$ csúcsoztályú G irányítatlan gráfban legfeljebb $\delta V(G)^3$ háromszög található (ahol δ csak ε -tól függ), akkor a gráfnak létezik egy háromszög-mentes részgráfja, amelyet G -ből legfeljebb $\varepsilon V(G)^2$ él törlésével kapunk.

Alon és Shapira bebizonyították, hogy az irányítatlan gráfok minden monoton tulajdonsága erősen tesztelhető (Alon–Shapira 2005). Az ε -tól való függés sajnos már irányítatlan gráfok esetében is nagyon rossz lehet: az Alon–Shapira-tételben ismert felső korlátok a Szemerédi-

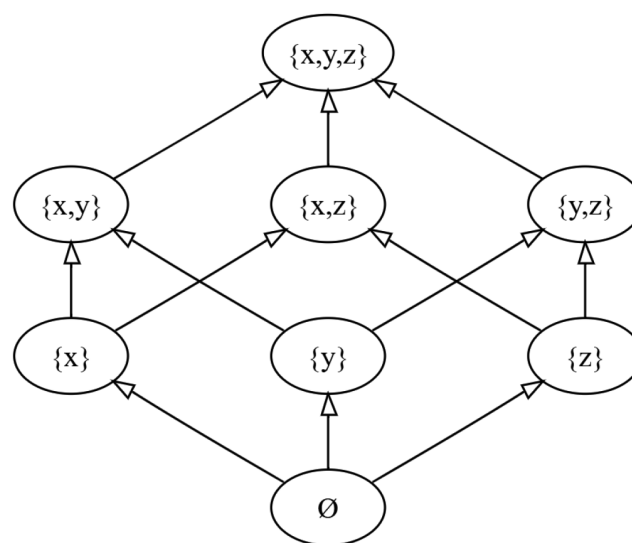
féle regularitási lemma iterált használata miatt toronyfüggvényeket tartalmaznak.

Alon és Fox nyomán (Alon–Fox 2015) egy tulajdonságot könnyen tesztelhetőnek nevezünk, ha erősen tesztelhető, és $f(\varepsilon)$ felülről becsülhető $1/\varepsilon$ egy polinomjával; ellenkező esetben a tulajdonság nehezen tesztelhető. Ismert, hogy mind a perfekt gráfok (Alon–Fox 2011), mind az összehasonlíthatósági gráfok tesztelése nehéz (Alon–Fox 2015); kevés tulajdonság bizonyult könnyen tesztelhetőnek. Alon és Fox bebizonyították, hogy irányítatlan gráfok esetén tetszőleges H gráfra a H -mentesség pontosan akkor könnyen tesztelhető, ha H páros (Alon–Fox 2015). A könnyű és nehéz tulajdonságok megkülönböztetése még sejtés szintjén is elérhetetlennek tűnik.

Poset, részbenrendezett halmaz

Poseten (részbenrendezett halmazon) egy olyan halmazt értünk, amely rendelkezik egy antireflexív és tranzitív $<$ relációval. A halmaz elemeire tekinthetünk úgy, mint egy irányított gráf csúcsaira, és a halmaz ij elemeire (ij) pontosan akkor él, ha $i < j$. A 3. ábrán az $\{\{x,y,z\}, \{x,y\}, \{x,z\}, \{y,z\}, \{x\}, \{y\}, \{z\}, \emptyset\}$ halmazon a részhalmaz relációval kapott poset Hasse-diagramja látható. Ezen az ábrán – az átláthatóság kedvéért – nem szerepelnek azok az élek, amelyek egyértelműen következnek a tranzitivitásból; például $(\{x\}, \{x,y,z\})$ és $(\emptyset, \{x,y,z\})$.

A poset rendezett részhalmazait láncnak hívjuk – azaz $\{i_1, i_2, \dots, i_k\}$ csúcshalmaz egy lánc, ha $i_1 < i_2 < \dots < i_k$, illetve az $\{i_1, i_2, \dots, i_k\}$ csúcshalmaz antiláncot alkot, ha semelyik két elem nem hasonlítható össze. A 2. ábrán látható poset egy leghosszabb lánc például az $\{\{x,y,z\}, \{x,y\}, \{y\}, \emptyset\}$ halmaz, míg egy leghosszabb anti-



3. ábra Az $\{\{x,y,z\}, \{x,y\}, \{x,z\}, \{y,z\}, \{x\}, \{y\}, \{z\}, \emptyset\}$ halmazon a részhalmaz relációval kapott poset Hasse-diagramja
 Forrás: Wikipedia, Partially ordered set (letöltve: 2025. 07. 18.)

lánca például $\{\{x\},\{y\},\{z\}\}$. Jelölje egy P poset leghosszabb láncának hosszát $h(P)$, leghosszabb antiláncának hosszát $w(P)$. A 3. ábrán szemléltetett poset esetén $h=4$, $w=3$. A h elemű láncot C_h -val jelöljük.

Posetek egy \mathcal{P} (nem feltétlenül véges) halmazára legyen $h(\mathcal{P}) = \min_{P \in \mathcal{P}} h(P)$, illetve $w(\mathcal{P}) = \min_{P \in \mathcal{P}: h(P)=h(\mathcal{P})} w(P)$.

Kun Gáborral bebizonyítottuk (Fekete–Kun 2025), hogy posetek monoton (részhalmazok vételére zárt) osztályai könnyen tesztelhetők. Ezen eredmény ekvivalens egy törlési lemmával polinomiális korlátokkal.

Homomorfizmus

Legyen P és Q két poset. Ha az $f: Q \rightarrow P$ leképezés rendezéstartó, akkor homomorfizmusnak nevezzük. A $t(Q,P)$ homomorfizmus-sűrűség annak a valószínűsége, hogy egy véletlenszerűen választott $f: Q \rightarrow P$ leképezés homomorfizmus. Egy P poset Q -mentes, ha nem tartalmazza Q -t (nem feltétlenül indukált) részposetként.

Polinomiális törlési lemma posetekre (Fekete–Kun 2025)

Legyen Q véges poset, amire $h(Q) \geq 2$. Bármely $\varepsilon > 0$ és P véges poset esetén, ha $t(Q,P) < (\varepsilon/2)^{h(Q)w(Q)^2}$, akkor létezik P -nek egy Q -mentes (sőt $C_{h(Q)}$ -mentes) részposete, amelyet legfeljebb $\varepsilon|P|^2$ él eltávolításával kapunk.

Ez tulajdonképpen azt jelenti, hogy ha Q -nak kevés példánya van P -ben, akkor P kevés él eltávolításával Q -mentessé tehető, azaz nincs ε -távol a Q -mentességtől. Tehát mivel P vagy távol van a Q -mentességtől, vagy Q -mentes, ezért vagy sok Q van benne, vagy egyáltalán nem tartalmaz Q -t.

Posetek monoton osztályai könnyen tesztelhetők (Fekete–Kun 2025)

Legyen \mathcal{P} véges posetek halmaza, amire $h(P) \geq 2$. Legyen $Q \in \mathcal{P}$ olyan, hogy $h(Q) = h(\mathcal{P})$ és $w(Q) = w(\mathcal{P})$. Bármely $\varepsilon > 0$ és P véges poset esetén, ha $t(Q,P) < (\varepsilon/2)^{h(P)w(P)^2}$, akkor létezik P -nek egy \mathcal{P} -mentes (sőt $C_{h(\mathcal{P})}$ -mentes) részposete, amelyet legfeljebb $\varepsilon|P|^2$ él eltávolításával kapunk.

Megkülönböztethetetlen tulajdonságok

A láncok fontos szerepet játszanak monoton poset osztályok hatékony tesztelésében: megadjuk ezen osztályok egy egyszerű osztályozását a tesztelhetőség szempontjából.

A Φ_1 és Φ_2 tulajdonságokat megkülönböztethetetlenek nevezzük, ha tetszőleges $\varepsilon > 0$ -ra és $i=1,2$ -re létezik N , hogy egy legalább N elemű halmazon vett, tetszőleges $P \in \Phi_i$ posethez létezik $P' \in \Phi_{3-i}$ poset ugyanazon a halmazon, amelyet P -ből legfeljebb $\varepsilon|P|^2$ él változtatásával kapunk meg.

Mivel monoton tulajdonságokat vizsgálunk, elegendő éleket törölni.

Megkülönböztethetlenség (Fekete–Kun 2025)

Legyen \mathcal{P} véges posetek egy halmaza, legyen $h=h(\mathcal{P}) \geq 2$ és $w=w(\mathcal{P})$. Ekkor a \mathcal{P} -mentes posetek osztálya és a C_h -mentes posetek osztálya megkülönböztethetetlenek.

Nevezetesen, minden C_h -mentes poset \mathcal{P} -mentes is; és ha egy P poset \mathcal{P} -mentes, akkor létezik egy C_h -mentes

részposete, amelyet legfeljebb $2 \left(\frac{h^2 w^2}{|P|}\right)^{\frac{1}{hw^2}} |P|^2$ él eltávolításával kapunk.

Ebből azonnal adódik, hogy minden legalább $N = h^2 w^2 (\varepsilon/2)^{-hw^2}$ elemű \mathcal{P} -mentes P posetnek létezik egy C_h -mentes (nem feltétlenül indukált) P' részposete, amelyet legfeljebb $\varepsilon|P|^2$ él törlésével kapunk. Azonban sokkal hatékonyabb tesztet kaphatunk, ha a konstans elemszámot mintavételező teszt iterálása helyett egy nagyobb részposetet mintavételezünk.

Poset C_h -mentességének részposet tesztelése s méretű mintavételezéssel

Bemenet: P poset
 s elemű véletlen indukált részposet R
 Ha $R = C_h$, akkor elutasítjuk P -t,
 egyébként elfogadjuk P -t.

A megkülönböztethetlenség alapján ez a teszt a \mathcal{P} -mentes posetek tesztelésére is alkalmas: legalább $1/2$ valószínűséggel elutasítja azokat a poseteket, amelyek ε -távol vannak a \mathcal{P} -mentességtől, azonban elhanyagolható valószínűséggel megengedi, hogy tévesen elutasítson \mathcal{P} -mentes posetet is.

Részposet teszt (Fekete–Kun 2025)

Legyen $h \geq 2$ egész szám, $\varepsilon > 0$, $c > 0$ és P egy véges poset. Ha a P poset ε -távol van a C_h -mentességtől, akkor

$\left\lceil \frac{4 \log(h) + 5}{2\varepsilon} \right\rceil$ véletlenszerűen választott elem legalább $1/2$ valószínűséggel tartalmazza C_h egy példányát.

Tetszőleges rögzített h -ra a C_h -mentesség egyoldalú hibával való tesztelésére a fenti korlát pontos nagyságrendet ad a szükséges mintavételezési csúcscsúszámot tekintve ε -ban.

Összehasonlíthatósági gráfok monoton tulajdonságai (Fekete–Kun 2025)

Egy P posethez tartozó összehasonlíthatósági gráf csúcshalmaza P , élhalmaza $\{(x,y): x < y \text{ vagy } y < x\}$.

Alon és Fox megmutatták, hogy az összehasonlíthatósági gráfok osztálya nehezen tesztelhető (Alon–Fox 2015). Azonban azzal az ígérettel, hogy a bemeneti gráf

egy összehasonlíthatósági gráf (promise problem), az összehasonlíthatósági gráfok monoton osztályai tesztelhetővé válnak – annak ellenére, hogy a hozzájuk tartozó poseteket nem ismerjük. Eredményeink teljesülnek monoton összehasonlíthatósági gráfosztályok tesztelésére is.

Kernel és mintavételezett gráf távolsága

Vágásnorma

A vágásnormát Frieze és Kannan definiálták mátrixokra (Frieze–Kannan 1999). A vágásnorma legfontosabb jellemzője, hogy nagyobb hangsúlyt fektet a struktúrára a „véletlen ingadozással” szemben, és segítségével „kiszűrhető a zaj”: amennyiben két mátrix csak fehér zajban tér el egymástól, úgy a különbségük vágásnormája nagyon kicsi, de amint a különbség „strukturáltabbá” válik, a norma is nő. A vágásnorma erős kombinatorikai jellege miatt kulcsszerepet játszott a (sűrű) gráflimeszelmélet kidolgozásában.

Egy $W \in L^1([0,1]^2)$ függvény vágásnormája

$$\|W\|_{\square} = \sup_{S,T \subseteq [0,1]} \left| \iint_{S \times T} W(x,y) dx dy \right|.$$

Mintavételezés

Legyenek X_0, X_1, \dots független, a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változók, és legyen k egy pozitív egész paraméter. Jelöljön $U: [0,1]^2 \rightarrow \mathbb{R}$ egy szimmetrikus, valós értékű mérhető függvényt. Jelölje $X = (X_1, X_2, \dots, X_k)$ a k -hosszú valószínűségi változó vektort, és jelölje M_k a $[0,1]^2$ -en azt a lépcsős függvényt, amelyben a szomszédos szakadáspontok közötti távolság $1/k$ mindkét változóban, a függvényérték pedig $(U(X_i, X_j))_{i,j \in [k]}$, és 0 a főátlón. Tehát M_k a mintavételezett k csúcsú gráfhoz tartozó kernel.

Első mintavételezési lemma

Az Alon és szerzőtársai által bizonyított első mintavételezési lemma két mátrix vágásnormabeli távolságának mintavételezéssel történő becsléséről szól (Alon et al. 2003). Ezt az eredményt Borgs és szerzőtársai általánosították és javították korlátos kernelekre (szimmetrikus $L^\infty([0,1]^2)$ függvényekre): k méretű mintavételezés esetén a hiba nagy valószínűséggel $O(k^{-1/4})$ nagyságrendű (Borgs et al. 2008).

Első mintavételezési lemma nemkorlátos kernelekre (Fekete–Kunszenti–Kovács 2025)

Legyen $k \geq 2$ egész, $p > 2$, és $U \in L^p_{sym}([0,1]^2)$. Ekkor

- tetszőleges $\phi > 0$ esetén teljesül, hogy

$$\begin{aligned} P\left(\|M_k\|_{\square} - \|U\|_{\square} \leq \left(C_1 + C_2 \frac{\sqrt{\ln k}}{k^{\frac{p-3}{4p}-\phi}}\right) k^{-\frac{1}{4} + \frac{1}{4p}}\right) \\ \geq 1 - 4k^{-\phi p}, \end{aligned}$$

ahol C_1, C_2 konstansok csak $p, \|U\|_p, \|U\|_1$ és ϕ értékektől függenek.

- tetszőleges $\gamma \in (\frac{1}{p}, \frac{1}{2})$ esetén teljesül, hogy

$$P(-C_3 k^{-1/2 + \gamma} \sqrt{\ln k} \leq \|M_k\|_{\square} - \|U\|_{\square}) \geq 1 - 4k^{1-\gamma p},$$

ahol C_3 konstans csak $p, \|U\|_{\square}, \|U\|_p, \|U\|_1$ és γ értékektől függ.

Ahogy az a tételből is látszik, az alsó és a felső korlát nagyságrendje eltér. Nemcsak a nagyságrend, hanem a bizonyítás és a bizonyítás nehézsége is nagyban eltér a két esetben.

Vágástávolság

Amennyiben két kernelt szeretnénk összehasonlítani, akkor egyrészt vizsgálhatjuk vágásnormáik különbségét, ahogyan azt fentebb tettük. Másrészt elemezhetjük a két kernel különbségének vágásnormáját. A 2. ábrán teljes páros gráfok két különböző csúcselnevezéshez tartozó grafonjait láthatjuk; ezek különbsége nem 0, holott egyazon gráfhoz tartoznak. Ennek kiküszöbölésére szeretnénk megengedni a kernelek különbségvétele előtt a „csúcs átnevezést”, vagyis a $[0,1]$ intervallumon egy mértéktartó bijekció alkalmazását.

Jelölje a mértéktartó $[0,1] \rightarrow [0,1]$ bijekciók halmazát Ψ . Az $U, W \in L^1([0,1]^2)$ függvények vágástávolsága $\delta_{\square}(U, W) := \inf_{\psi \in \Psi} \|U - W_{(\psi(x), \psi(y))}\|_{\square}$.

Második mintavételezési lemma

Szeretnénk becslést adni az M_k minták U körüli koncentrációjáról, azaz korlátokat szeretnénk adni a $\delta_{\square}(M_k, U)$ vágástávolságra nagy valószínűséggel.

Borgs és szerzőtársai megmutatták, hogy korlátos kernelekre a mintavételezett és az eredeti kernel vágástávolsága nagy valószínűséggel $O((\ln k)^{-1/2})$ nagyságrendű (Borgs et al. 2008). Az első mintavételezési lemmához képest szignifikánsan gyengébb nagyságrend a Szemerédi-féle regularitási lemma használatának köszönhető.

Második mintavételezési lemma nemkorlátos kernelekre (Fekete–Kunszenti-Kovács 2025)

Legyen $k \geq 2$ egész, $p > 2$, és $U \in L_{sym}^p([0,1]^2)$. Tetszőleges

$\phi \in \left(0, \frac{1}{2} - \frac{1}{p}\right)$ esetén teljesül, hogy

$$P\left(\delta_{\square}(M_k, U) \leq C_4(\ln k)^{\frac{1}{2} + \frac{1}{2p}}\right) \geq 1 - e^{-\frac{k^2}{2 \log_2(k)}} - 4k^{-\phi p}$$

ahol C_4 csak $p, \|U\|_p, \|U\|_1$ és ϕ értékektől függ.

Kapcsolat metrikák között

A sűrű gráfok témakörén belül ezek a mintavételezési lemmák kulcsszerepet játszanak annak bizonyításában, hogy a korlátos kernelek vágásnormája és a bal-konvergencia (más néven sűrűség-konvergencia) által indukált metrika megegyezik. Kunszenti-Kovács megmutatta, hogy ez az állítás már nem igaz nemkorlátos kernelekre (Kunszenti-Kovács 2019). Ismert, hogy nemkorlátos U kernelekre a véletlen k -méretű minták 1 valószínűséggel bal-konvergálnak U -hoz, ahogy $k \rightarrow \infty$ (Kunszenti-Kovács–Lovász–Szegedy 2022), de ez a kérdés ez idáig nyitott volt a vágástávolság esetén. Ez alapján a második mintavételezési lemma nemkorlátos kernelekre való kiterjeszhetősége nem volt feltétlenül sejthető.

Majdnem biztos konvergencia (Fekete–Kunszenti-Kovács 2025)

Legyen $p > 4$, továbbá $U \in L_{sym}^p([0,1]^2)$ kernel, és legyen $(M_k)_{k \geq 2}$ egy k elemű véletlen részgráf sorozata U -nak. Ekkor a következő teljesül: $P(\delta_{\square}(U, M_k) \rightarrow 0 \text{ ahogy } k \rightarrow \infty) = 1$.

Konklúzió

Kutatásaim során mintavételezéssel vizsgáltam nagy gráfokat. Eredményeim hozzájárultak ahhoz, hogy sűrű gráfok globális tulajdonságai kis minták alapján is megbízhatóan vizsgálhatók legyenek. A posetekre vonatkozó polinomiális törlési lemma bizonyítása lehetővé tette további, könnyen tesztelhető gráfosztályok azonosítását. Emellett a mintavételezési lemmák kiterjesztése a nemkorlátos kernelekre igazolta, hogy a minták konvergenciája a vágásmetrikában általánosabb körülmények között is fennáll. Ezek az eredmények nemcsak elméleti szempontból fontosak, hanem gyakorlati alkalmazásokat is előrevetítenek a biológiai hálózatok, a társadalmi kapcsolatrendszerek és a technológiai infrastruktúrák modellezésében. Összességében a bemutatott módszerek és tételek hozzájárulnak ahhoz, hogy a nagy hálózatok szerkezete hatékonyabban és pontosabban tanulmányozható legyen.

Köszönetnyilvánítás

Az 1016492. számú projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási, Fejlesztési és Innovációs Alapból nyújtott támogatásával, a KDP-2020 pályázati program, illetve az ERC Synergy Grant No. 810115 finanszírozásában valósult meg.



AZ NKFI ALAPBÓL MEGVALÓSULÓ PROGRAM

Irodalomjegyzék

- Alon, N., Fernandez de la Vega, W., Kannan, R. & Karpinski, M. (2003) Random sampling and approximation of MAX-CSPs. *Journal of Computer and System Sciences*, Vol. 67. No. 2. pp. 212–243. Special issue on STOC2002 (Montreal, QC). [https://doi.org/10.1016/S0022-0000\(03\)00008-4](https://doi.org/10.1016/S0022-0000(03)00008-4)
- Alon, N. & Fox, J. (2011) Testing perfection is hard. <https://arxiv.org/abs/1110.2828> [Letöltve: 2025. 07. 18.]
- Alon, N. & Fox, J. (2015) Easily testable graph properties. *Combinatorics, Probability and Computing*, Vol. 24. No. 4. pp. 646–657. <https://doi.org/10.1017/S0963548314000765>
- Alon, N. & Shapira, A. (2005) Every monotone graph property is testable. *STOC'05: Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, New York. pp. 128–137. <https://doi.org/10.1145/1060590.1060611>
- Borgs, C., Chayes, J. T., Cohn, H. & Zhao, Y. (2018) An L_p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *The Annals of Probability*, Vol. 46. No. 1. pp. 337–396. <https://doi.org/10.1214/17-AOP1187>
- Borgs, C., Chayes, J. T., Cohn, H. & Zhao, Y. (2019) An L_p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, Vol. 372. No. 5. pp. 3019–3062. <https://doi.org/10.1090/tran/7543>
- Borgs, C., Chayes, J. T., Lovász, L., T. Sós, V. & Vesztegombi, K. (2008) Convergent sequences of dense graphs I. Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, Vol. 219. No. 6. pp. 1801–1851. <https://doi.org/10.1016/j.aim.2008.07.008>
- Fekete, P. T. & Kun, G. (2025) Posets are easily testable. *European Journal of Combinatorics*, Vol. 129. No. 4. Article No. 104044. <https://doi.org/10.1016/j.ejc.2024.104044>
- Fekete, P. T. & Kunszenti-Kovács, D. (2025) The cut norm and Sampling Lemmas for unbounded kernels. *Analysis Mathematica*, Vol. 51. No. 2. pp. 477–514. <https://doi.org/10.1007/s10476-025-00090-9>

- Frieze, A. & Kannan, R. (1999) Quick approximation to matrices and applications. *Combinatorica*, Vol. 19. No. 2. pp. 175–220. <https://doi.org/10.1007/s004930050052>
- Glasscock, D. (2015) What is a graphon? *Notices of the American Mathematical Society*, Vol. 62. No. 1. pp. 46–48. <https://doi.org/10.48550/arXiv.1611.00718>
- Kunszenti-Kovács, D. (2019) Uniqueness of Banach space valued graphons. *Journal of Mathematical Analysis and Applications*, Vol. 474. No. 1. pp. 413–440. <https://doi.org/10.1016/j.jmaa.2019.01.052>
- Kunszenti-Kovács, D., Lovász, L. & Szegedy, B. (2022) Multigraph limits, unbounded kernels, and Banach space decorated graphs. *Journal of Functional Analysis*, Vol. 282. No. 2. Article No. 109284. <https://doi.org/10.1016/j.jfa.2021.109284>
- Lovász, L. (2012) *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications 60. American Mathematical Society, Providence, RI. <https://doi.org/10.1090/coll/060>
- Ruzsa, I. Z. & Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. *Combinatorics (Proceedings of the Fifth Hungarian Colloquium of the Mathematical Society János Bolyai, Keszthely, 1976)*, Vol. 2. *Colloquia Mathematica Societatis János Bolyai* 18. pp. 939–945.
- Tóth L. M. (2024) Mi is... a grafon? *Érintő – Elektronikus Matematikai Lapok*, Vol. 34. <https://ematlap.hu/tudomany-tortenet-2024-10/1460-mi-is-az-a-grafon> [Letöltve: 2025. 07. 18.]
- Wikipedia. Partially ordered set. https://en.wikipedia.org/wiki/Partially_ordered_set [Letöltve: 2025. 07. 18.]