Válasz Telcs András opponensi bírálatára

Egyenként válaszolok a felmerült kritikákra, kérdésekre. Először a szabad formátumban megfogalmazott észrevételekre reagálok, majd utána tételesen az opponensi kérdésekre.

A 115. tétel [...] meglepő a laikus olvasó számára
Válasz: Ez az egyik kedvenc tételem (a nem saját eredmények között), így nem tudom megállni, hogy ne válaszoljak erre. A bizonyítás azon alapszik, hogy ha egy fekete-fehér gráfnak elkészítjük azt a módosított A szomszédsági mátrixát, amely a főátlón kívül megegyezik a szomszédsági mátrixszal, a főátóban pedig minden i-re $a_{i,i}$ 0 ha az i-edik pont fehér, és 1, ha az i-edik pont fekete, akkor minden sikeres megnyomási szekvencia A-nak a $\mathbb{F}_2$ kételemű test feletti rangja. Ha valaki már tudja ezt a tényt, akkor a bizonyítás már nem annyira nehéz: azt kell belátni, hogy minden megnyomás hatása A-ra egy Gauss eliminációs lépésnek felel meg, és könnyű belátni, hogy ez a Gauss elminációs lépés minden esetben eggyel csökkenti a rangot. Ami számomra nagyon meglepő, hogy a 115. tételnek nem ismert elemi (azaz kételemű test feletti lineáris algebrát nem használó) bizonyítása.

Az értekezésben a lineáris algebrai kapcsolat terjedelmi okok miatt nem került bemutatásra: a 9. fejezetben elért eredményekhez nem kellett lineáris algebrai módszereket használni. Azonban az elmúlt két évben nagyon sokat foglalkoztunk ezzel a problémával BSM-es diákokkal, és úgy néz ki, hogy jelentős eredményeket tudtunk elérni a 4 reverzió sejtésben, amihez erőteljesen kellett lineáris algebrai eszközöket használni. Jelenleg dolgozunk a kéziraton, amit most nincs olyan állapotan, hogy megosztható lenne, de bízom benne, hogy a védésre be fogom tudni mutatni ezt a kéziratot.

A dolgozatban szereplő cikkek copy-paste beillesztése csöppet sem kárhoztatható, a bíráló kifejezetten fölöslegesnek és időrablónak gondolja a teljes dolgozat elvárását a rövid dolgozat és tézisek benyújtása helyett, de a beillesztés közben három helyen észrevettük az „in this paper” utalást
Válasz: Igyekeztem az „in this paper” utalásokat „in this thesis”-re átírni, sajnos ez nem sikerült mindenhol. Szeretnék válaszolni a rövid vs. teljes dolgozat kérdésére is. A III. osztály támogatja a rövid dolgozat benyújtását, a teljes dolgozat írását két okból választottam. Az első ok az, hogy a doktori értekezésem a szokásosnál több cikken alapszik, amelyek együttesen adják ki az általam bemutatni kívánt eredményeket. Igyekeztem az egyes részek közötti kapcsolatokat kiemelni, és bemutatni az értekezésben. A másik ok az, hogy az értekezés alapját nyújtó publikációk döntő többsége társszerzőkkel együtt készült. Ezekből a cikkekből azokat a részeket emelem ki, amelyek lényegében az én eredményeim, és a társszerzők eredményeit igyekeztem ettől külön választani. Különösen fontosnak tartom két fiatal kutatónk, Soltész Dániel és Mezei Tamás esetén kiemelni azokat a részeket, amelyek az ő eredményeik, így azok az ő tudományos fokozataik elérésében beszámíthatóak lesznek.

p 42 l 8: valami hiányzik [...]; p64 l18 fölösleges „in”; p80 és a későbbiekben számos helyen és a bibliográfiában is Erdős P.L. helyett hibásan Erdős E.P. van; tézisekben p 23 l 13 „miden”
Válasz: Köszönöm ezeknek a typoknak az észrevételét. Különösen sajnálom a társszerzőm, Erdős Péter monogrammjának az elírását.

p84. a P-stabilitás központi szerepet játszik, jó lenne egy kis heurisztika arra, mit is jelent. Ugyanitt talán jó lett volna a multicommodity flow-ról pár szót említeni (előtörténet, jelentős

eredmények, hasznosság)
Válasz: A P-stabilitás heurisztikusan azt jelenti, hogy ha egy P-stabil fokszámsorozatot kismértékben megváltoztatunk (L_1 távolágban kettővel), akkor a realizációinak a száma legfeljebb a fokszámsorozat hosszának polinomszorosára változik. Egy nemrég írt kéziratunkban [1] azt mutatjuk meg, hogy korlátokkal megadott fokszámsorozat-osztályokra a P-stabil osztályok lényegében megegyeznek azokkal a fokszámsorozat-osztályokkal, amelyek mindig grafikusak (fully graphic, always graphic). Ez már önmagában érdekes, de kibontakozni látszik egy még érdekesebb téma a jelenlegi kutatásaink mentén. 3-uniform hipergráfok korlátokkal megadott fokszámsorozataira tudunk bizonyítani egy dichotómia tételt: a korlátoktól függően vagy egy ilyen osztályban véges sok kivétellel minden fokszámsorozat, ami teljesíti a kézfogási lemmát (3-uniform hipergráfok esetén ez az, hogy a fokszámok összege 3-mal osztható) grafikus, vagy egy ilyen osztályra NP-teljes eldönteni azt, hogy egy fokszámsorozat grafikus. 3-uniform hipergráfok esetén még nem sikerült P-stabilitásról eredményeket elérni (ugye itt most L_1 távolságban nem 2-vel, hanem 3-mal kellene egy fokszámsorozatot megváltoztatni). Egyszerű gráfok esetén meg nem világos, hogy mi lenne a dichotómia tétel, hiszen egyszerű gráfokra bármilyen fokszámsorozat grafikusságának az eldöntése P-ben van. A kéziratunkban ([1]) megfogalamzunk egy sejtést, hogy fokszámsorozatok kis perturbációira a realizációk számának a változása vagy egy fix fokú polinom felső korlátú vagy szuperpolinomiális. A kéziratban nem szerepel, de nekem van egy ennél merészebb sejtésem is: A polinom felső korlát fokára több fázisátmenet is van, ahogy a fokszámok alsó és felső határát változtatjuk: *i)* a legkisebb korlát egyszerű gráfok esetén a majdnem reguláris fokszámsorozatokra, páros gráfok esetén a félig majdnem reguláris fokszámsorozatokra, *ii)* majd az alsó és felső korlátokat nyitva a polinom foka nagyobb lesz; a határokat definiáló egyenlőtlenség egyszerű gráfokra nem ismert, páros gráfokra az értekezés 4.9-es egyenlőtlensége adja meg, *iii)* majd a legnagyobb korlátokat a Jerrum-McKay-Sinclair egyenlőtlenség [2]:

$$(c_1 - c_2 + 1)^2 \leq 4c_2(n - c_1 - 1)$$

adja meg, ahol c_1 a maximális fok, c_2 a minimális fok egy n hosszú fokszámsorozatban (és ennek az analógja meg páros gráfokra nem ismert), és erre a polinom felső korlát foka még nagyobb. És innentől nincs még nagyobb fokú polinom, mint felső korlát, hanem szuperpolinomiális felső korlát van csak. Személy szerint én nagyon szeretnék még ennél mélyebb kapcsolatot is találni a P-stabilitás és gráfalgoritmusok bonyolultságelmélete között, de egyelőre homályban tapogatózunk.

A multicommodity flow az értekezés bevezetőjében már említve van. Talán ott lehetett volna bővebben írni róla, de mivel ez nem saját eredmény, terjedelmi okok miatt nagyon korlátozottan írtam róla.

p 128 l -13 „markers” Valahol definiálva van? A genomikában járatlan számára szükséges volna definiálni.
Válasz: Az 1. oldal lábjegyzetében érintőlegesen meg van említve, de valóban ezt célszerű lett volna pontosaban definiálni, sőt, akár utána a 128. oldalon visszahivatkozni a definícióra/leírásra.

p 180 e(10.2) szerencsés lenne megmagyarázni, miért tekintjük ezt energiának
Válasz: A definíció előtt próbálkoztam ezt elmagyarázni: ha a DCJ rendező szcenárió csak reverziókat tartalmaz, akkor nincsenek a rendező szcenárión cirkuláris kromoszómák. Egy

rendezési szcenáriót annál távolabbinak tekintünk tisztán reverziókat tartalmazó szcenáriónál, minél több cirkuláris kromoszóma van benne. A (10.2) képlet ezt fejezi ki. Ha így definiáljuk a hipotetikus energiát, akkor az összes DCJ rendező szcenárió terében a legalacsonyabb (= 0) energiájú rendező szcenáriók a csak reverziókat tartalmazó szcenáriók lesznek.

Opponensi kérdés

1. Mi okozhatja, hogy a 13. Tézisben ismertetett Markov lánc jó keverési tulajdonságát nehéz igazolni?

   Válasz: Woodard, Schmidler és Huber adott meg 2009-ben szükséges és elégséges feltételeket ahhoz, hogy egy Parallel Tempering Markov lánc gyorsan keveredjen [3, 4]. Ezen feltételeket nehéz ellenőrizni a megadott Markov láncra. Szemléletesen a következőről van szó. Legyen G_1 és G_2 két unikromoszómális, megegyező telomerű, akadálymentes (hudle-free, értekezés 34. oldal) genom azonos génekkel. Jelölje $\mathcal{D}(G_1, G_2)$ a G_1-ből G_2-be menő legtakarékosabb DCJ rendezési szcenáriók halmazát, és jelölje $\mathcal{R}(G_1, G_2)$ a G_1-ből G_2-be menő legtakarékosabb reverziós rendezési szcenáriók halmazát. Minden $R \in \mathcal{R}(G_1, G_2)$ legtakarékosabb reverziós rendezési szcenárióhoz definiálhatjuk a vonzási tartományát (attraktorát), $\mathcal{A}(R)$-t, amely azon $D \in \mathcal{D}(G_1, G_2)$ legtakarákosabb DCJ rendezési szcenáriókból áll, amelyekből eljuthatunk R-be a 124. tételben megadott módon. Továbbá minden $\mathcal{X} \subseteq \mathcal{R}(G_1, G_2)$ halmazra legyen a $\mathcal{A}(\mathcal{X})$ az $\mathcal{X}$ szcenáriói attraktorainak az úniója. Ha tudunk genompárok végtelen sorozatát megadni, úgy, hogy azokban vannak olyan $\mathcal{X}_1$ és $\mathcal{X}_2$ halmazok melyre

   $$\frac{|\mathcal{R}(G_1, G_2)|}{|\mathcal{X}_1|} = O(poly(n)) \quad \text{valamint} \quad \frac{|\mathcal{R}(G_1, G_2)|}{|\mathcal{X}_2|} = O(poly(n))$$

   ahol n a genomok mérete, ugyanakkor

   $$\frac{|\mathcal{A}(\mathcal{X}_1)| \times |\mathcal{X}_2|}{|\mathcal{A}(\mathcal{X}_2)| \times |\mathcal{X}_1|} = \Omega(c^n)$$

   valami $c > 1$-re, akkor a Parallel Tempering lassan kever (további, itt nem részletezett technikai feltételek kellenek, ami azt biztosítja, hogy alacsonyabb hőmérsékleten ne lehessen nagy valószínűséggel $\mathcal{X}_1$-ből, illetve annak adott hőmérsékleten vett nagy valószínűségű vonzástartományából $\mathcal{X}_2$-be, illetve annak vonzástartományába jutni). Valóban, ekkor hiába kever gyorsan a Parallel Tempering a legtakarékosabb DCJ rendező szcenáriókon, elhanyagolhatóan kevés időt fog $\mathcal{X}_2$ vonzástartományában tölteni, így lehűtéssel elhanyagolhatóan kis valószínűséggel fogunk $\mathcal{X}_2$-be jutni. Azt kellene belátni, hogy nem létezik az attraktorok mérete közötti ilyen nagy különbség.

Budapest, 2024. október 16.

Miklós István

Hivatkozások

[1] Erdős, P.L., Miklós, I., Soukup, L. (2024) Fully graphic degree sequences and P-stable degree sequences. `https://arxiv.org/abs/2405.12013`

[2] Jerrum, M.R., McKay, B.D., Sinclair, A. (1992) When is a Graphical Sequence Stable? Chapter 8 in Random Graphs: Volume 2 Eds.: A. Frieze and T. Luczak, Wiley, pp. 304.

[3] Woodard, D., Schmidler, S.C., Huber, M. (2009) Conditions for rapid mixing of parallel and simulated tempering of multimodal distributions. Annals of Applied Probability, 19(2): 617–640.

[4] Woodard, D., Schmidler, S.C., Huber, M. (2009) Sufficient Conditions for Torpid Mixing of Parallel and Simulated Tempering. Electronic Journal of Probability, 14(29):780–804.