Bírálat

Bolla Marianna

Clustering graphs and contingency tables with spectral methods

című akadémiai doktori értekezéséről

Bolla Marianna értekezésének fő témája nagy gráfok szerkezetének és spektrális tulajdonságainak kapcsolatának megértése, valamint spektrális módszereket használó algoritmusok építése gráfokkal kapcsolatos feladatok megoldására. A nagyméretű hálózatok témaköre az utóbbi tizenöt-húsz évben az alkalmazott matematika egyik sokat kutatott területévé vált, amit részben az informatikai hálózatok (internet, közösségi hálózatok) kiépülése motivált. Ezeknek a gráfoknak a globális szerkezetének feltárásában az egyik fontos tényező a sűrűbben összekötött részek, klaszterek megtalálása – mind elméleti, mind gyakorlati, algoritmikus szempontból. A dolgozat ilyen jellegű feladatokat old meg a gráf Laplace-mátrixának, illetve modularitásmátrixának spektrumát vizsgálva.

Az első fejezet bevezeti az értekezésben használt alapvető fogalmakat, valamint bemutatja a szerző dimenziócsökkentéssel kapcsolatos eredményeit. Itt tehát az a cél, hogy egy n csúcsú gráfot (melynek adjacenciamátrixa $n \times n$ méretű), n darab, de n-nél jóval kisebb dimenziós vektorral reprezentáljunk úgy, hogy az egymással összekötött csúcsokhoz tartozó vektorok minél közelebb kerüljenek egymáshoz az alacsonyabb dimenziós térben. Súlyozott gráf esetén pedig minél nagyobb súlya van a két csúcs közötti élnek, annál közelebb kell, hogy kerüljenek, a megoldandó minimalizálási feladatban. A szerző tárgyalja a súlyozott négyzetes mátrixok, illetve a kontingencia-táblázatok esetét, a főkomponens-analízishez hasonlóan megmutatva, hogy különféle, a klasszikustól eltérő (pl. a modularitásmátrixszal megfogalmazott) feladatokban hogyan kapható meg az optimális reprezentáció a sajátvektorokból. Ezek az általánosítások az alkalmazások szempontjából is figyelemre méltók lehetnek, hiszen például biológiai, genetikai feladatokban előfordulhat, hogy két különböző halmaz közötti kapcsolatokat kell modellezni, illetve hogy nem csak a kapcsolat megléte, hanem annak erőssége is fontos jellemzője a vizsgált rendszernek. Ezen kívül a szerző a modularitás-mátrixnak a gráf szerkezetével való kapcsolatát (pl. független halmazok) is vizsgálja, többek között a Hoffmann-korláttal rokon állításokat bizonyítva. A fő eredményeket bizonyos Hilbert-tér értékű operátorokra is általánosítja, megkeresve a lineáris algebrával, optimalitási tételekkel megfogalmazott eredmények mélyebb, funkcionálanalízis-beli hátterét. A fejezet főbb eredményei tehát a klasszikus eredmények modularitás-mátrixokra vonatkozó analógiái köré csoportosulnak.

A második fejezetben a szerző bizonyos véletlen mátrixok, a blokk-modell spektrumával kapcsolatos eredményeit ismerteti. Ez a véletlen gráf az Erdős–Rényi-gráfnak egy inhomogén általánosítása: a csúcsokat csoportokra osztjuk, majd a csúcspárokat véletlenül, függetlenül kötjük össze, úgy, hogy az egyes élek behúzásának valószínűsége csak attól függ, hogy a végpontok melyik csoportba tartoznak. Mivel erre úgy is gondolhatunk, mint egy fix mátrix perturbációja egy véletlen Wigner-mátrixszal, a (nagyobb) sajátérték elhelyezkedésére, távolságára vonatkozó eredmények (13–17. tételek) a véletlen mátrixok elméletének szempontjából is érdekesek lehetnek. A szerző ebben az esetben sem csupán a négyzetes mátrixokat, hanem a kontingenciatáblázatokat is vizsgálja, ahol a két dimenzió akár jelentősen is eltérhet. Az általánosítással véletlen gráfoknak olyan tág, rugalmas családját kapjuk, melyek alkalmasak lehetnek többféle valós hálózat modellezésére is. A szerző több egyenlőtlenséget bizonyít a modularitásmátrix sajátértékei és egy, a kontingenciatáblázat diszkrepanciáját mérő mennyiség között, amivel a spektrum és a kombinatorikai struktúrák szerkezete között fennálló kapcsolatok jobb megértését segíti (ahogyan a Cheeger-egyenlőtlenség a klasszikus esetben). Az ilyen jellegű állítások szintén hasznosak lehetnek valós hálózatok modellezése során is.

A harmadik fejezetben először a modularitási mátrix spektrumának és a homomorfizmus-sűrűségen alapuló gráflimesz-fogalomnak a viszonyát vizsgálja a szerző, többek között egy folytonossági tételt bizonyítva. Ezután ugyenebben a témakörben egy sejtést fogalmaz meg, mely a gráfkonvergenciát karakterizálná többek között a modularitási mátrix spektrumának, illetve az előző fejezetben szereplő diszkrepancia jellegű mennyiségeknek a segítségével (a sejtést a szerző egy, a disszertáció megszületése óta készült munkájában megcáfolta). Végül egy EM-típusú algoritmust javasol egy speciális véletlengráf-modell paramétereinek becslésére, azaz egy statisztikai feladatot old meg egy sokparaméteres, sűrű gráfban.

A dolgozat témaválasztása időszerű, a nagy hálózatok klaszterezettségével és spektrumával kapcsolatos kérdéseket sokan vizsgálják napjainkban. A dolgozatban a szerző (és néhol társszerzők) nevével jelzett tételeket új eredménynek fogadom el. Az értekezés jól felépített, és egységes képet ad a gráfok és kontingenciatáblázatok spektrumának elméletéről, illetve a lehetséges alkalmazásokról. A dolgozatban bemutatott számos eredmény közül sok olyan van (különösen a második fejezetben), amely önmagában is érdekes lehet a gráfok spektrálelméletében. Ezzel párhuzamosan a szerző folyamatosan szem előtt tartja az alkalmazások szempontjait, az eredmények többségét jól körülírható, modellezési feladatokban előforduló eredmények motiválják. A bizonyítások során alkalmazott módszerek helyenként inkább klasszikusak, esetleg már ismert tételekre épülnek, máshol azonban több ötletet, mélyebb módszereket, kifinomultabb technikákat is igényelnek. A bőséges irodalomjegyzék és rengeteg megfelelő hivatkozás ellenére a dolgozatban nem találtam a szorosan kapcsolódó "community detection" témakör véletlen blokk-modellre vonatkozó fontos új eredményeire vonatkozó említést (például [4]). A második fejezetben érdemes lett volna kitérni arra, hogy az értekezésben ismertetett eredmények hogyan kapcsolódnak ehhez a friss kutatási irányhoz. A harmadik fejezetben javasolt becslési eljárással kapcsolatban jó lett volna részletesebben látni a valós adatokon, valós hálózatokon való eredményeket (akár azt is, hogy milyen méretű, milyen jellemzőkkel rendelkező hálózatokon történt az ellenőrzés), hiszen ezek lényeges információkat tartalmazhatnak az algoritmus alkalmazhatóságáról – ami ebben az esetben a cél lenne.

Összességében az értekezés eredményeit az akadémiai doktori fokozat követelményeinek megfelelőnek tartom, és javaslom a nyilvános védés kitűzését.

Kérdések.

1. Az első fejezetben javasolt eljárások alkalmazását az értekezésben bemutatott képszegmentációs feladaton kívül sikerült-e más feladatokban is megvalósítani? Ha igen, mik a megközelítés előnyei az adott feladatban, és milyen nehézségek

adódnak a megvalósítás során?

2. A 2.3. szakaszban szereplő „multiway discrepancy” mennyiségnek (25. definíció) van-e valamilyen kapcsolata a gráfok vágás-normájával, illetve a jumble normával (lásd [2])?

3. Van-e valamilyen kapcsolat az értekezés második fejezetében szereplő, illetve az [1] survey cikk 3.4., illetve 4.6. szakaszában ismertetett eredmények között? Az értekezésben szereplő eredményekből esetleg lehet-e következtetést levonni a rekonstruálhatóság kritikus értékére?

4. A harmadik fejezet harmadik részében szereplő α-modellhez hasonlóan tekinthetnénk egy olyan véletlen gráfot, melyet egy speciális grafonból, egy lépcsős függvényből sorsolunk (a [3] cikkben definiált értelemben). Módosítható-e a dolgozatban szereplő EM-algoritmus úgy, hogy ennek a véletlen gráfnak a paramétereinek becslésére adjon egy jó eljárást? Ha nem, mi lenne ennek a fő akadálya?

Hivatkozások

[1] E. Abbe. Community detection and stochastic block models: recent developments. Kézirat. arXiv:1703.10146.

[2] D. Kunszenti-Kovács, L. Lovász and B. Szegedy. Multigraph limits, unbounded kernels,and Banach space decorated graphs. Kézirat. arXiv:1406.7846.

[3] L. Lovász and B. Szegedy. Limits of dense graph sequences, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957.

[4] E. Mossel, J. Neeman and A. Sly. Reconstruction and estimation in the planted partition model, Probab. Theory Related Fields **162** (2015), no. 3-4, 431–461.

Budapest, 2017. április 18.

Backhausz Ágnes
Eötvös Loránd Tudományegyetem