

Válasz Szolnoki Attila opponensi véleményére

Köszönöm szépen a disszertáció gondos átolvasását és a nagyon támogató opponensi véleményt. Válaszaim a feltett kérdésekre a következők:

1) A 3.3.3-as a „csoportok hálózata” alfejezet felveti azt a kérdést, hogy milyen eredményt adna az, ha a klikkperkoláción alapuló csoportkeresési módszert a csoportok hálózatára is alkalmaznánk (esetleges kisebb k -val, mint az eredeti klikkméret)? A fizikában ismert fázisátalakulásoknál használt technikák mintájára esetleg lehet-e ilyen módon egyfajta „renormálást” csinálni és ezáltal bizonyos csoport-agglomerátumokat azonosítani? Esetleg volt-e a szerzőnek ilyen jellegű próbálkozása, vagy ha nem, akkor milyen praktikus ok indokolja azt, hogy kár ebbe az irányba erőfeszítéseket tenni?

A vázolt „renormálás” egy nagyon természetes gondolat, mely bennünk is felmerült a hivatkozott ábrán szereplő hálózatok csoportszerkezetének elemzése során. Azonban részletesebb tanulmányokat nem folytattunk ezen a téren az alábbi okok miatt. Egyfelől a csoportok hálózata és az alap hálózat közt van néhány fontos eltérés, ami például a fokszámoszlás alakjából is látszik. Ennek révén nem mondhatjuk azt, hogy a renormálás „kiátlagolja a lokális fluktuációkat” és utána önhasonló módon ugyanazt a rendszert látjuk, csak nagyobb skálán. Emellett ha mégis nekilátnánk „renormálni”, akkor több eltérő lehetőség is felmerül ennek véghezvitelére. A k paraméter csökkentése is hasonló dolgokat eredményezhet mint a csoportok hálózatára való áttérés, hiszen előfordulhatnak olyan csoportthalmazok, melyek egy nagyobb k -nál még elkülönülnek egymástól, kisebb k -ra való áttérésnél viszont már összeolvadnak egy csoporttá. Ezek alapján legalább 3 különböző alternatíva is felmerül a „renormálás” konkrét megvalósítására:

- a) A csoportok hálózatán ugyanazon k paraméter mellett keresünk csoport-agglomerátumokat mint amilyen k paraméter mellett az alap hálózaton.
- b) Megengedjük, hogy a csoportok hálózatán más k paraméterrel keressük meg a csoport-agglomerátumokat (és minden szinten az optimális k értéket használjuk).
- c) Nem a csoportok hálózata alapján állítjuk elő a csoport-agglomerátumokat, hanem egyszerűen csak csökkentjük a k paramétert.

Továbbá az is joggal merül fel, hogy a csoportok hálózatát ne az átfedések alapján definiáljuk, hanem a csoportok között húzódó élek alapján, ezzel már akár 6 különböző módon is vizsgálhatjuk a „renormálás” hatásait, és nem világos, hogy ezek közül melyik eljárás a „kanonikus”. Végezetül az általunk tanulmányozott hálózatok többségénél maximum néhány lépésig lehetne „renormálni”, akármelyik módszert is választjuk a fentiek közül.

2) *Egy számomra nem egyértelmű apróság a 4.10-es ábra kapcsán az, hogy milyen módon definiált csoportméret függvényében van az átlagos élettartam ábrázolva? Mint ahogy a 4.11d ábra is szépen illusztrálja, egy csoport mérete dinamikus változhat, így az „s” érték se állandó. A csoport induló értékét, vagy valamilyen időátlagot tekintsük-e a 4.10-es ábrán szereplőnek?*

A 4.10-es ábra és a hozzá tartozó szövegrészek valóban nem adnak felvilágosítást erről. Az s-el jelölt tengelyen a csoportméret időátlaga szerepel.

3) *Ugyanehhez a részhez kapcsolódik a tárgyalt csoportdinamikai modell alkalmazhatósága. Nyilvánvalóan ez egy nagyon fontos, széles körben (pl. vállalatok fejlődésére vagy más biológiai rendszerekben) érvényes modell-család. De a bemutatott lényegében topológiai analízis nem veszi figyelembe azt, hogy a csoport élettartama nem feltétlen elsődleges fitnessze a csoportnak, hanem inkább annak egy következménye. (Pl. valamilyen szempontból sikeres csoport inkább együtt marad, míg a sikertelen könnyebben átalakul.) Természetesen a „valódi” fitnesszt a szereplők konkrét kölcsönhatásából lehetne származtatni, amit egy ilyen általános modelltől nem várhatunk el. Éppen ezért meglepő az, hogy mégis ilyen szemléletes és a mindennapi tapasztalatainkkal összhangban álló jóslatokra képes még ez általános modell is. Van-e esetleg valamilyen elképzelése a szerzőnek arra vonatkozóan, hogy mi az az általános tulajdonság, ami ennek a robusztusságnak a hátterében áll?*

Maximálisan egyetértek azzal, hogy ha ismerjük részleteiben a vizsgált rendszert, akkor a fitnesszt valószínűleg érdekesebb ezen további paraméterek alapján származtatni, és az élethossz csak az egyik lehetséges indikátora a fitnessznek. A tapasztalt robusztusságra nincs mély, jól megalapozott magyarázat. Véleményem szerint az lehet talán egy spekulatív magyarázat, hogy mindkét hálózat emberek közti kapcsolatokon alapszik, és az emberek gyakran szeretik megtartani azt, ami működik (fitt), és eldobni azt ami nem. Emiatt a jól működő csoportok hosszasan megmaradnak, és a „jól működő emberi csoport” összetételének változékonysága ugyanolyan, akár egy telefonhívási hálózatot vizsgálunk, akár egy tudományos társszerzőségi hálózatot, hiszen a tudósok is valahol emberek.

4) A dolgozat utolsó fejezetéhez érve az olvasóban önkéntelenül is felmerül az, hogy az előző fejezetekben ismerttetett csoportazonosítási módszert milyen eredménnyel lehetne alkalmazni a javasolt multifraktálok alapuló generált gráfokra? Esetleg ilyen módon kapcsolatot lehetne teremteni a csoportokat leíró statisztikai tulajdonságok és a generálás alapjául szolgáló $L(x, y)$ függvény bizonyos tulajdonsága között. A szerző véleménye szerint milyen csoportkeresési módszerek jöhetnének itt szóba?

Volt egy olyan projektünk, ahol a generálás alapjául szolgáló függvényt maximum likelihood technika alkalmazásával próbáltuk „visszafejteni” egy célul megjelölt hálózathoz, sajnos a diák aki diplomamunkaként foglalkozott vele nem csinálta végig, abbahagyta tanulmányait. A csoportok- és csoportstatisztikák összekapcsolása a multifraktál generálással nekünk nem jutott eszünkbe, de nagyon jó ötletnek tűnik; ugyanakkor a csoportkereső módszerek közül a sztochasztikus blokkmodell ötleten alapuló módszerek [1, 2, 3, 4] szerintem jobban passzolnak a multifraktál generáló koncepciójához mint a klikkperkoláció. A sztochasztikus blokkmodell alap koncepciója az, hogy megpróbáljuk a csúcsokat rögzített számú blokkokba (csoportokba) beosztani úgy, hogy minden B blokkon belül megadunk egy, a blokkra jellemző p_B élvalószínűséget, illetve minden A, B blokkpáros között is megadunk egy, az adott párosra jellemző p_{AB} élvalószínűséget, és az a cél, hogy a megfigyelt élszámok és a modell által jósolt élszámok minél jobban passzoljanak. Amikor a multifraktál generálót illesztjük egy hálózathoz, akkor majdnem ugyanezt kell csinálni, csak itt a blokkok egy multifraktálba rendeződnek. Amit nyerünk a multifraktállal az az, hogy kevesebb paramétert kell illeszteni, mintha az összes blokk és összes blokkpáros paraméterét akarnánk meghatározni.

2017. november 17.

Palla Gergely

Hivatkozások

- [1] A. Decelle – F. Krzakala – C. Moore – L. Zdeborova: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84. évf. (2011), 066106. p.
- [2] S. E. Fienberg – S. S. Wasserman: Categorical data analysis of single sociometric relations. *Sociological Methodology*, 12. évf. (1981), 156–192. p.
- [3] P. W. Holland – K. B. Laskey – S. Leinhardt: Stochastic blockmodels: First steps. *Social Networks*, 5. évf. (1983), 109–137. p.
- [4] T. A. B. Snijders – K. Nowicki: Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14. évf. (1997), 75–100. p.