

Palla Gergely “Komplex hálózatok
szerkezetének es dinamikájának feltárása
és modellezése statisztikus fizikai
módszerekkel” című MTA doktori
értekezésének bírálata

July 10, 2017

1 A tudományterületről

Az értekezés a hálózattudomány témakörébe tartozik. Ez egy meglehetősen új tudományterület, amelynek fogalmi struktúrája a kétezres évek elején alakult ki. A hálózattudomány egyik főágának - a dolgozat bevezetőjében is említett - alapvető paradigmája az, hogy 1. a valóságban is létező komplex hálózatok rendelkeznek a “scale-free” illetve “small world” tulajdonságokkal és ezért 2. (lényegesnek tartom, hogy ez egy implikáció) kiemelkedő jelentőségű a matematikailag egyszerűen leírható “scale-free” (skálafüggetlen) és “small world” tulajdonságokkal rendelkező hálózatok (valójában gráfsorozatok) tanulmányozása. Az alapvető modell a Barabási-Albert László által bevezetett preferential attachment hálózatkonstrukció. Ezekről a gráfsorozatokról Strogatz es Watts mutatta ki, hogy rendelkeznek a “small world” tulajdonsággal, illetve Barabási es Albert Réka, minden idők egyik legidézettebb tudományos publikációjában, hogy rendelkeznek a “scale free” tulajdonsággal. Hálózattudományi cikkek százai (talán ezrei) foglalkoznak azzal, hogy statisztikus fizikai módszerekkel tanulmányozzák ezeket a gráfokat (különös tekintettel a preferential attachment gráfokra illetve az Erdős-Rényi véletlen gráfokra). Nem kerülhető meg a kérdés, hogy valóban igaz-e az, hogy a valóságos hálózatok skálafüggetlenek. Erre a válasz, inkább nemleges. A Faloutsos testvérek 1999-ben publikált cikkükben azt állították, hogy az Internet router hálózata skálafüggetlen jellegű. Ez a cikk volt hosszú ideig a hálózattudomány legjobban kommunikálható eredménye, amelynek Google Scholar idézettsége hatezer felett jár. Erről a cikkről ma már egyértelműen tudható, hogy teljesen hibás, egy olyan módszerre épül, amelyik szinte minden hálózatról azt mutatná ki, hogy skálafüggetlen tulajdonságokkal rendelkezik. Ezen cikk hibájának feltárása után jelent meg a hálózatelmélet első jelentős kritikája Willinger, Alderson es Doyle cikke [4] az AMS Notices-ben. Később

Clauset, Shalizi es Newman [3] mutatta meg, hogy számos hálózat, amelyről a közvélekedés az volt, hogy skálafüggetlen, valójában nem az. Számos világhírű cikk esetében súlyos kétségek merültek fel. Ezek közül a valószínűleg Christakis es Fowler cikke a kövérség járványos voltáról, illetve Barzel es Barabási Nature Biotechnology cikke. Az előbbiről Russell Lyons matematikus írt bírálatot, az utóbbiról Lior Pachter a Berkeley computational biology professzora írt egy azóta világhírűvé vált, rendkívül kritikus blogbejegyzést, ami a hálózatelmélet globális kritikájának egyik fontos eleme.

Meg kell jegyeznünk, hogy a preferential attachment sorozatok egy valószínűséggel expandert alkotnak, és semmi nem mutat arra, hogy a valóságos hálózatok is expanderek, függetlenül attól, hogy esetleg logaritmikus az átmérőjük a méretük függvényében. Ugyancsak fontos hozzátenni, hogy a skálafüggetlenség a fokszámeloszlás tulajdonsága, tehát önmagában a skálafüggetlenségből a gráfok mélyebb struktúrális tulajdonságára nem lehet következtetni, magyarul, abból, hogy egy konkrét scale-free/small world gráfsorozat rendelkezik egy bizonyos tulajdonsággal nem következtethetünk arra, hogy más tulajdonságai is hasonlóak lehetnek, az ilyen matematikai hálózatok vizsgálata nem feltétlenül segíti a konkrét komplex hálózatok megértését (egyes kutatók szerint egyáltalán nem segíti, de itt a viták oly mértékben mérgesedtek el, hogy ezt egy konkrét kutató doktori értekezésének vizsgálatakor nem lenne fair részletesebben tárgyalni). Mindazonáltal a fentiek alapján kimondhatjuk, hogy a hálózatelmélet matematikai hálózatok vizsgálatára épülő ágának valóságos tudományos jelentősége, mondjuk úgy, nem teljesen elfogadott a tudományos közéletben.

2 Az értekezés módszereiről

A dolgozat matematikai fogalmak statisztikus fizikai jellegű tanulmányozásával foglalkozik. Matematikusként el tudom fogadni azt, hogy a statisztikus fizika vizsgálatának tárgya esetleg egy absztrakt matematikai struktúra, nem állítom, hogy nincs létjogosultsága a gráfokra vonatkozó heurisztikus jellegű vizsgálatoknak, de matematikai szempontból csak akkor tudok értékelni egy ilyen eredményt, ha az egy logikailag tökéletes levezetésen alapul. Az előző részben említett preferential attachment gráfsorozat egy tökéletesen definiált es matematikai szempontból érdekes fogalom. Tehát, amikor Strogatz es Watts azt állítja, hogy a preferential attachment gráfokra igaz az, hogy az átmérőjük asszimptotikusan $\log(n)$ nagyságrendű, akkor az egy valóságos matematikai tétel. Itt érkeztem el a bírálat számomra legnehezebb részéhez. A hálózatelmélészek számára Strogatz es Watts bizonyítása "heurisztikus", egy matematikus számára pedig hibás. Ez nem egy bizonyítása volt egy matematikai tételnek, hanem egy érvelés amellett, hogy a tétel talán igaz. Utóbb kiderült, hogy egyetlen új csúcs megjelenése esetén a valóságos asszimptotika $\frac{\log(n)}{\log \log(n)}$ (ezt Bollobás Béla es Oliver Riordan [1] bizonyította be). Strogatz es Watts cikkének idézettsége ennek ellenére nagyságrenddel nagyobb, mint Bollobáséké. A skálafüggetlenséget

ugyanezen gráfsorozatokra Barabási és Albert Réka mutatta meg, de ez sem bizonyítás, hanem inkább csak egy erős érvelés, ami nem tehető pontossá. Itt több nagyságrend is van a hibátlan bizonyítás (Bollobás, Riordan, Spencer, Tusnády) [2] és az eredeti cikk idézettsége között, az utóbbi javára. Hálózatzelméleti szempontból az “ n ” az egészen biztosan kisebb, mint 10^{100} , nincs valódi jelentősége annak, hogy valahol van egy $\log \log(n)$ -es tag, és általában is igaz, hogy a statisztikus fizikai heurisztikákat megtámogató számítógépes modellek is elégségesek lehetnek, ha valakit a valóban fellépő hálózati méretek érdekelnek. Azt azonban ki kell mondanom, hogy a heurisztikus eredményekből pont a legfontosabb dolog hiányzik egy matematikus számára.

Az értekezés egyik fontos eredménye a k -klikk perkolációs küszöb meghatározása. A Palla által megadott formula korrekt, ahogy ezt később Bollobás és Riordan, illetve Ráth és Tóth kimutatta, de Palla eredménye matematikai szempontból nehezen értékelhető. Ugyanez mondható el a dolgozat első részében tárgyalt topológikus fázisátalakulásról szóló eredményekről. A Lovász Lászlóval és Vicsek Tamással közösen írt cikk egészen közel van a klasszikus gráfimeszelmülethez. Ebben a cikkben (PNAS) a Lovász-Szegedy féle graphon modell segítségével próbálnak skálafüggetlen gráfsorozatokat konstruálni. Maga a gondolat nagyon érdekes, hiszen az eredeti megközelítés az, hogy egy $W : [0, 1] \times [0, 1] \rightarrow [0, 1]$ függvény, azaz egy abszolút folytonos valószínűségi mérték Radon-Nykodym deriváltja segítségével konstruálnak konvergens sorozatokat, itt pedig egy szinguláris mérték van a háttérben. A probléma az, hogy ezzel a módszerrel valószínűleg nem lehet gráfsorozatokat nyerni. Nagy n -ekre a keletkező gráfok csúcsainak jelentős részéből egyáltalán nem indul ki él, ez egyébként az eredeti cikkhez csatolt pdf-fileban is világosan le van írva. Bizonyos kisebb n -ekre valóban működik az ötlet, ezen n -ek a valóságos hálózatok mérettartományában vannak, tehát megint arról van szó, hogy hálózatzelméleti szempontból semmiféle probléma sem lép fel, eltekintve attól, hogy vajon miért is fontos egy biológus vagy egy szociológus számára egy ilyen különleges matematikai konstrukció, csak azért mert az ő hálózata is skálafüggetlen, ha megfelelő szögben nézzük. A “megfelelő szög” itt azt jelenti, hogy log-log skálán általában skálafüggetlennek, vagy legalábbis hasonlóknak néznek ki a hálózatok, nagyjából mindegyik, amelynek “heavy tail”-je van (lásd az értekezés 52. oldalán lévő ábrát).

3 A módszerek hasznosságáról

A dolgozat harmadik fejezetében a szerző megmutatja, hogyan használható a k -klikkperkoláció csoportszerkezetek feltárására. Matematikai értelemben nem egészen pontosan definiált a hálózati csoportok fogalma és számítástudományi/matematikai szempontból ezek az algoritmusok nehezen értékelhetők, de a dolgozatban szereplő módszer alapján kifejlesztett CFinder algoritmus sikeressége valódi hálózatok vizsgálatára számomra (a hivatkozások vizsgálata után) meggyőzőnek tűnik.

4 Értékelés

Palla Gergely értekezése nyolc 2004 és 2010 között publikált cikkekre épül, mely cikkek közül kettő a Nature-ben, egy pedig a Proc. Natl. Acad. Sci. USA-ben jelent meg. Ez a két lap a világ legjelentősebb tudományos folyóiratai közé tartozik. Ezen cikkekben Palla első szerzőként szerepel, amely tény ezen a területen nem hagyható figyelmen kívül. Az öt másik eredmény is világszínvonalú folyóiratokban lett publikálva. A két Nature cikk Scholar Google idézettsége 4124, illetve 1335, Palla Gergely összesített idézettsége a bíráló megírása idején 7588. Összefoglalva a fentieket: matematikusként vannak bizonyos fenntartásaim a heurisztikus hálózatelmélettel, mint tudományággal kapcsolatban, ezek a fenntartások azonban nem a szerzőre vonatkoznak, hanem egy rendkívül fiatal tudományterületnek a természettudományokon belüli pozíciójára. Az nem kétséges számomra, hogy valóságos tudományterületről van szó, amely talán több figyelmet kapott, mint amennyit megérdemel. Mindazonáltal úgy gondolom, hogy értékes eredmények irányába mutat a hálózatelmélet fejlődése. Palla Gergely tudományos beágyazottsága a területen vitathatatlan, különös tekintettel kiemelkedő idézettségére, és ezért feltétlenül javaslom a nyilvános védelem kitűzését és az MTA doktora cím odaítélését.

References

- [1] B. BOLLOBÁS, O. RIORDAN, The Diameter of a Scale-Free Random Graph. *Combinatorica* **24** (2004) 5–34.
- [2] B. BOLLOBÁS, O. RIORDAN, J. SPENCER AND G. TUSNÁDY, The degree sequence of a scale-free random graph process. *Random Structures and Algorithms* **18** (2001) 279–290.
- [3] A. CLAUSET, R. S. SHALIZI AND M. E. J. NEWMAN, Power law distribution in empirical data, *SIAM Reviews* **51** (2009) 661–703.
- [4] W. WILLINGER, D. ALDERSON AND JOHN C. DOYLE, Mathematics and the Internet: A Source of Enormous Confusion and Great Potential, *Notices of the AMS* **56** (2009) 586–599.

Kérdés: Melyek azok a dolgozatban érintett problémák, ahol a statisztikus fizikai módszerek esetleg matematikailag precízzé tehetők?