

Az exponenciális random gráf modellek bemutatása:

Egy iskolai osztály baráti hálózatának modellezése¹

Kisfalusi Dorottya

kisfalusi.dorottya@tk.mta.hu

Beérkezés: 2017. 12. 11.

Átdolgozott változat beérkezése: 2018. 03. 12.

Elfogadás: 2018. 03. 12.

Összefoglaló: Az exponenciális random gráf modellek (ERGM) olyan statisztikai modellek, melyek társadalmi kapcsolathálók elemzésére alkalmasak. A széles körben ismert regressziós technikák (pl. logisztikus regresszió) nem alkalmasak a kapcsolatok egymástól való függőségének figyelembevételére, mivel a megfigyelési egységek között függetlenséget tételeznek fel. A társadalmi kapcsolathálók jelölései azonban nem függetlenek egymástól, és az ERG-modellek képesek a jelölések közötti függőségi viszonyok modellezésére. Az ún. lokális hálózati konfigurációk hatásának becslésével egyrészt következtetéseket vonhatunk le arról, hogy az adott kapcsolatháló milyen mechanizmusok hozták létre, másrészt egyéb, számunkra fontos magyarázó változók paramétereit (pl. a hálózati szereplők neme, etnikai hovatartozása) úgy tudjuk megbecsülni, hogy közben kontrollálunk az endogén hálózati mechanizmusokra. A tanulmányban az ERG-modellek ismertetése után egy empirikus példával, egy hatodikos általános iskolai osztály baráti hálózatának elemzésével is illusztrálom e statisztikai modellek használatát.

Kulcsszavak: ERGM, kapcsolatháló-elemzés, statisztikai modellezés

Bevezetés

A széles körben ismert regressziós modellek (pl. logisztikus regresszió) a megfigyelési egységek között függetlenséget tételeznek fel, ezért nem alkalmasak a kapcsolatok egymástól való függőségének figyelembevételére és modellezésére. Ehhez sajátos statisztikai modellekre van szükség, és az elmúlt évtizedekben több statisztikai modellcsalád is született a kapcsolathálózati adatok elemzésére.

A társadalmi kapcsolathálókból számos olyan mechanizmus ismert, melyekben a kapcsolati jelölések egymástól függnék. Az egyik ilyen mechanizmus például a reciprocitás: sokfajta olyan kapcsolattípust ismerünk (pl. a baráti jelölések), melyekben kölcsönös jelölések nagy arányban fordulnak elő. Ha *ego* barátjának jelöl egy másik embert, *altert*, akkor nagy a valószínűsége annak, hogy *alter* is barátjának tartja *egót*, az *alter* $\rightarrow$ *ego*

¹ A tanulmányban felhasznált adatokat az MTA TK „Lendület” RECENS Kutatócsoport gyűjtötte az MTA Lendület programja által támogatott „Verseny és negatív hálózatok” kutatás keretében. Köszönettel tartozom a két anonim bírálónak a kézirat korábbi változatához fűzött értékes hozzászólásaikért.

jelölés megléte tehát függ attól, hogy van-e jelölés $ego \rightarrow alter$ között. Egy másik gyakran megfigyelt mechanizmus a tranzitivitás, melyet „a barátom barátja az én barátom is” jelenséggel illusztrálhatunk: ha ego barátjának jelöli $alter$ 1-et és $alter$ 1 barátjának jelöli $alter$ 2-t, akkor nagy a valószínűsége annak, hogy ego barátjának jelöli $alter$ 2-t is. Az $ego \rightarrow alter$ 2 jelölés megléte tehát függ az $ego \rightarrow alter$ 1 és az $alter$ 1 $\rightarrow alter$ 2 jelölések meglététől. E bemutatott két mechanizmuson kívül nagyon sok egyéb olyan mechanizmus létezik, melyek a kapcsolati jelölések közötti függőségi viszonyt írják le.

Ebben a tanulmányban az exponenciális random gráf modellek (*exponential random graph models*, ERGM, más néven p^* modellek – Lusher et al. 2013; Robins et al. 2007) használatát mutatom be, mely modellek képesek a kapcsolathálózati jelölések közötti függőségi viszonyok modellezésére. Az ERG-modelleket elsősorban keresztmetszeti adatok elemzésére használják, bár léteznek a modellcsaládnak longitudinális adatok elemzésére használható változatai is (Desmarais–Cranmer 2012; Hanneke et al. 2009; Krivitsky–Handcock 2014; Lusher et al. 2013a). Longitudinális kapcsolathálózati adatok elemzésére azonban jóval szélesebb körben használják az ún. sztochasztikus aktororientált modelleket (*stochastic actor-oriented models* – Snijders et al. 2010; Steglich et al. 2010; Block et al. 2018), ezért ebben a tanulmányban a keresztmetszeti ERG-modellek bemutatására koncentrálok.

A statisztikai modell ismertetése után egy empirikus példával, egy általános iskolai osztály baráti hálózatának elemzésével is illusztrálom az ERG-modellek használatát. E példa segítségével azt mutatom be, miért fontos a kapcsolatok közötti függőségi viszonyok figyelembevétele olyan esetekben is, amikor nem ezek a függőségi viszonyok állnak kutatásunk középpontjában. Bizonyos hálózati struktúrák (pl. etnikailag szegregált baráti hálózatok) létrejöhetnek és erősödhetnek ugyanis hálózati (pl. tranzitivitás) és nem hálózati (pl. etnikai homofíliára vonatkozó preferencia) mechanizmusok hatására is. Az ERG-modellek alkalmasak e mechanizmusok szétválasztására.

Az exponenciális random gráf modellek kialakulása

Az 1970-es, 1980-as években a kapcsolathálózati elemzések alkalmazásakor még gyakori volt a logisztikus regressziós modellek, illetve a lineáris valószínűségi modellek használata (lásd például Hallinan–Williams 1987, 1989). Ezeknek az elemzéseknek a függő változója egy kétértékű változó volt, és a regressziós modellekkel azt kívánták magyarázni, milyen változók hatására alakul ki (1 – van kapcsolat, 0 – nincs kapcsolat) vagy marad fenn időben (1 – fennmarad, 0 – megszűnik) két egyén között a vizsgálni kívánt kapcsolattípus. Magyarázó változóként egyéni szintű (pl. nem, etnikum stb.), diadikus (pl. azonos nem, azonos etnikum stb.), és kontextuális változókat (pl. osztálylétszám, kisebbségi tanulók aránya az osztályban stb.) vontak be az elemzésbe.

A hagyományos regressziós modellek azonban függetlenséget tételeznek fel a megfigyelési egységek között, miközben egy közösség kapcsolatai nem függetlenek egymástól. Ennek is köszönhető, hogy a megfigyelt társadalmi kapcsolathálókat eltér-

nek a véletlen hálózatoktól (Rapoport 1957). Felmerült tehát az igény olyan statisztikai modellek megalkotására, melyek figyelembe tudják venni, hogy egy hálózat két szereplője közötti kapcsolat kialakulása és fennmaradása függ a környezetükben meglévő többi kapcsolattól is.

Ezek a statisztikai modellek arra alkalmasak, hogy modellezzük, milyen mechanizmusok hozták létre az empirikusan megfigyelt kapcsolathálót. A függő változó tehát maga a kapcsolatháló, magyarázó változóként pedig nem csak a szereplők egyéni és diádszintű jellemzőit tudjuk bevonni az elemzésbe, hanem hálózati hatásokat is. Holland és Leinhardt (1981) ún. p1 modellje volt az első olyan modell, melyben bizonyos hálózati hatások (pl. reciprocitás, a szereplők kimenő és bejövő jelölései) paraméterbecslése lehetséges volt.² Ez a modell azonban a diádok között függetlenséget tételezett fel.

Frank és Strauss (1986) tanulmánya vezette be a kapcsolatháló-elemzésbe a Markov-függőség fogalmát, mely szerint egy hálózatban a kapcsolatok egymástól függetlenek, kivéve, ha egy pontjuk közös. E modell exponenciális függvényként való általánosításával született meg az ún. p* modell (Wasserman–Pattison 1996). A későbbi ERG-modellek lazítottak a Markov-függőség szigorú feltételezésén annyiban, hogy bizonyos esetekben két kapcsolat között akkor is állhat fenn függőségi viszony, ha nincsen közös pontjuk (Pattison–Robins 2002; Snijders et al. 2006). E modellek kialakulásával lehetővé vált annak modellezése, milyen endogén hálózati folyamatok jellemzik a kapcsolathálókat.

Az exponenciális random gráf modellek mint statisztikai modellek

Az ERG-modellek abból a feltevésből indulnak ki, hogy egy hálózat kapcsolatai nem függetlenek egymástól, a hálózati szereplők közötti kapcsolatok létrejöttét a szereplők lokális környezetében meglévő más kapcsolatok is befolyásolják. Ezek a nem független viszonyok különböző lokális konfigurációkat alkotnak (pl. reciprok viszonyok, tranzitív viszonyok stb.), és a modell becslése során ezekhez a hálózati konfigurációkhoz rendelünk valószínűségeket. A modellek segítségével következtetni tudunk arra, hogy az egyes hálózati konfigurációk szignifikánsan gyakrabban vagy ritkábban fordulnak-e elő, mint ahogy azt a véletlen alapján várnánk, kontrollálva a modell összes többi paraméterére (Lusher et al. 2013b: 1).

Az ERG-modellek elméletileg megalapozott modellek: a modell becslése során arra vagyunk kíváncsiak, hogy az adott hálózatban kialakult struktúra milyen mechanizmusoknak köszönhetően jött létre. Vizsgálhatjuk például, hogy egy adott is-

2 Azokban a kapcsolatháló-elemzésekben, melyek során főként a hálózati szereplők valamely attribútumának hatása áll a kutatás középpontjában, gyakran használják a 2000-es évek során megalkotott ún. p2 modellt is (van Duijn et al. 2004), mely multilevel elemzésre is alkalmas (lásd pl. Veenstra et al. 2007; Tolsma et al. 2013). Továbbá azokban az esetekben, amikor a hálózati mechanizmusok modellezése nem célja az elemzésnek, de a hálózati struktúrára kontrollálni szeretnénk, az ún. QAP- (quadratic assignment procedure) modellek is használhatók (Krackhardt 1988; Dekker et al. 2007).

kolai osztály baráti kapcsolathálója³ által jött-e létre, hogy a diákok (a) hajlamosak a hozzájuk hasonlókkal barátkozni (homofília), (b) hajlamosak a baráti jelöléseket viszonzni (reciprocitás), (c) hajlamosak barátaik barátaival összebarátkozni (transzitivitás), vagy esetleg (d) e három tényező együttesen játszott szerepet a kapcsolatok alakulásában. Ugyan keresztmetszeti adataink vannak, de e tényezők együttes bevonásával a modellünkbe következtethetünk arra, hogy melyik tényező milyen mértékben járult hozzá a hálózati struktúra kialakulásához.

Az ERG-modellekben a függő változót a hálózat kötési alkotják. Azokat a diádokat, ahol *ego* megjelöli *altert* a vizsgált kapcsolattípusban (pl. Kik a barátaid?), 1-es számmal kódoljuk, azokat a diádokat, ahol nincsen jelölés *ego* és *alter* között, 0-val kódoljuk. A lehetséges magyarázó változók egyéni és diádszintű változók (pl. *ego* neve, *alter* neve, azonos nem), valamint az endogén strukturális folyamatokat (pl. reciprocitás, transzitivitás) megjelenítő konfigurációk lehetnek.

Minden hálózat leírható különböző összefoglaló mérőszámokkal (statisztikákkal), mint például a kapcsolatok száma, a kölcsönös jelölések száma, a különböző triádkonfigurációk száma, stb. Az ERG-modellek ezeket a statisztikákat használják fel ahhoz, hogy a hálózat valószínűségét megbecsüljék. A modell az alábbi képlettel írható le:

$$P_{\theta}(x) = \frac{1}{\kappa} \exp\left\{\sum_k \theta_k z_k(x)\right\} \quad ,$$

ahol $P_{\theta}(x)$ az adott hálózat valószínűsége, z_k a különböző statisztikák vektora, θ_k a paraméterek vektora, κ pedig egy konstans (Koskinen–Daraganova 2013: 55). Maguk a statisztikák a hálózati konfigurációk számát (vagy ennek valamilyen függvényét) jelenítik meg az x hálózatban. A hálózat valószínűsége tehát attól függ, hogy hány ilyen konfiguráció van jelen a megfigyelt hálózatban, és a paraméterek azt mutatják meg, milyen szerepet játszanak az egyes konfigurációk a hálózatban (Robins–Lusher 2013a: 9).

A becslési eljárás

A becslési eljárás során azt becsüljük meg, hogy a különböző paraméterek milyen hatást gyakorolnak annak a valószínűségére, hogy *ego* és *alter* között van jelölés. A modell specifikálásakor kiválasztjuk azokat a konfigurációkat, melyeket a hálózat létrejötte szempontjából elméletileg fontosnak tartunk. A modell paramétereit ezek a konfigurációk alkotják, a becslés során azok paraméterértékeit becsüljük meg egy szimulációs eljárás segítségével. A szimulációs eljárással azt a gráfeloszlást keressük meg, melyben a megfigyelt hálózat központi helyet foglal el, tehát aminek a jellemzői kellőképpen hasonlítanak a megfigyelt hálózathoz. Az eljárás során megbecsüljük azokat a paraméterértékeket, amik a legnagyobb valószínűséggel vettek részt a megfigyelt hálózat kialakításában (Koskinen–Snijders 2013: 141).

3 Kapcsolathálózati kutatásokban a baráti jelölések *ego* *percepcióját* fejezik ki arra vonatkozóan, hogy *ego* és *alter* között van-e baráti viszony. Mivel *ego* és *alter* *percepciója* eltérhet arról, hogy van-e köztük baráti viszony, a baráti jelölések nem minden esetben kölcsönösek. E tanulmányban az irányított baráti jelölések hálózatát elemzem, de az ERG-modellek lehetőséget adnak irányítatlan kapcsolatok elemzésére is.

A becslési eljárásra két megközelítés ismert, az alábbiakban a Snijders (2002) által javasolt módszert ismertetem.⁴ A becslési eljárás akkor konvergált, ha a szimulált hálózatok statisztikáinak várható értékei kellően hasonlóak a megfigyelt hálózatunk statisztikáinak értékeihez. A hasonlóság mértékét minden paraméter esetében egy ún. t -hányados segítségével értékeljük, mely a statisztika szimulált hálózatokból vett mintában számolt átlaga és a megfigyelt hálózatban mért értéke különbségének, valamint a mintabeli szórásnak a hányadosa. Ha a t -hányados abszolút értéke minden paraméter esetében kisebb mint 0,1, a modellünk konvergált. Ha a t -hányados értéke valamelyik vagy több paraméter esetében abszolút értékben nagyobb mint 0,1, lehetőség van a becslési eljárás megismétlésére oly módon, hogy a korábban kapott paraméterértékeket állítjuk be a paraméterbecslés kezdő értékeként. Ezt az eljárást addig folytathatjuk, míg a modell konvergál (Koskinen–Snijders 2013: 153–154).⁵

A becslési eljárás során minden paraméter esetében kiszámoljuk a paraméter becsült értékét és a standard hibát. Az együtthatók szignifikanciája Wald-teszt segítségével tesztelhető: ha az együttható és a standard hiba hányadosának abszolút értéke nagyobb 2-nél, a paraméter hatását (a konvencionálisan gyakran használt 5 százalékos szinten) statisztikailag szignifikánsnak tekintjük (Koskinen–Snijders 2013: 157).

A modell illeszkedésének vizsgálata

A modell illeszkedésének (*goodness of fit*, GOF) vizsgálatakor arra vagyunk kíváncsiak, mennyire jól reprezentálja becsült modellünk a hálózat azon jellemzőit, melyeket a paraméterbecslés során expliciten nem modelleztünk. A modell illeszkedését egy szimulációs folyamat segítségével tudjuk értékelni. A becslési eljárás során becsült paraméterértékeket felhasználva szimulálunk számos hálózatot, majd megvizsgáljuk, e hálózatok eloszlásában hol helyezkedik el az eredetileg megfigyelt hálózatunk bizonyos, a modellünkben nem feltétlenül becsült konfigurációk tekintetében (Koskinen–Snijders 2013: 165). A paraméterbecslésünk ugyanis akkor megfelelő, ha a becsült paramétereink alapján szimulált hálózatok minél több szempontból minél jobban hasonlítanak a ténylegesen megfigyelt hálózathoz.

Egy konfiguráció illeszkedése akkor elfogadható, ha az empirikusan megfigyelt érték és a szimulált hálózatokban mért átlag közötti különbség, valamint a szórás hányadosa (GOF t -hányados) abszolút értékben nem haladja meg a 2-t azoknak a konfigurációknak az esetében, amiket nem becsültünk a modellben, és nem haladja meg a 0,1-et azok esetében, amiket becsültünk (Koskinen–Snijders 2013: 166; Robins–Lusher 2013b: 181). Ha valamelyik vagy több általunk fontosnak tartott konfiguráció tekintetében nem megfelelő a modellünk illeszkedése, újakezdhetjük a becslési eljárást oly módon, hogy újabb paraméter vagy paraméterek bevonásával a modellbe megpróbálunk e konfigurációk esetében is jobb illeszkedést elérni.

⁴ Mindkét eljárásról részletes matematikai leírást nyújtanak Koskinen és Snijders (2013: 149–154).

⁵ Ha az új modellt oly módon becsüljük, hogy a korábban kapott paraméterértékeket állítjuk be a paraméterbecslés kezdő értékeként, akkor oda kell figyelnünk az ún. minta-autokorrelációs együttható értékére is. Ennek értéke ideális esetben 0 körül van, de abszolút értékben 0,4-es értékig tolerálható (Koskinen–Snijders 2013: 147).

Az ERG-modellek szociológiai relevanciája

Az exponenciális random gráf modellek alkalmasak arra, hogy endogén hálózati hatásokat modellezzünk, ezáltal vizsgálni tudjuk, milyen mechanizmusok játszhattak szerepet a hálózat létrejöttében. Ez egyben azt is jelenti, hogy egyéb, számunkra fontos magyarázó változók paramétereit úgy tudjuk megbecsülni, hogy közben kontrollálunk ezekre a mechanizmusokra. Tegyük fel például, hogy kutatási kérdésként az érdekel minket, kialakulnak-e pozitív kapcsolatok az integrált oktatást nyújtó iskolákban a különböző etnikai háttérrel rendelkező tanulók között, vagy a diákok kapcsolathálója továbbra is etnikailag szegregált marad. A regressziós modellben azt szeretnénk vizsgálni tehát, hogy tapasztalhatunk-e a baráti jelölések tekintetében etnikai homofíliát (Kandel 1978; Lazarsfeld–Merton 1954; McPherson et al. 2001), azaz hajlamosabbak-e a tanulók saját etnikai csoportjuk tagjaival barátkozni, mint a más etnikai háttérrel rendelkező osztálytársaikkal. Tudjuk továbbá, hogy társadalmi kapcsolathálóokban a baráti jelölések általában viszonzottak, illetve gyakran előfordul, hogy összebarátkozunk a barátaink barátaival. Ahhoz tehát, hogy az etnikai homofília mértékét torzítatlanul tudjuk becsülni, kontrollálnunk kell arra, hogy az egyének hajlamosak reciprok, valamint tranzitív viszonyok kialakítására.

Példaként képzeljünk el egy az általános iskola első évfolyamát most kezdő nem roma tanulót, Sárát. Sára családja most költözött a környékre, ezért a lány nem ismer senkit új osztálytársai közül. Szeretne barátokat szerezni az osztályból, és egyáltalán nem fontos neki, hogy a barátai a romák vagy a nem romák közül kerülnek-e ki. Az első héten feltűnik neki, hogy egyik osztálytársa, Judit, nagyon gyakran kezdeményez beszélgetést. Sárának szimpatikus Judit, ezért viszonozza a baráti közeledést. Mivel Judit nem roma, Sárának máris lett egy nem roma barátja. Judit még az óvodában összebarátkozott Lucával, s a születésnapjára elhívja a barátait, így Sárát és Lucát is, akik így szintén barátok lesznek. Sárának már két nem roma barátja van. Szimpatikus neki az osztályból egy roma fiú, István, de azt tapasztalja, hogy fiúk és lányok mindig egymástól elkülönülve játszanak. Így aztán úgy dönt, hogy István helyett inkább a lányokkal barátkozik, és összebarátkozik Melindával, aki nem roma. Sárának tehát – annak ellenére, hogy nem tartja fontosnak barátai etnikai hovatartozását – nem roma barátai lettek, azáltal, hogy viszonozta mások baráti közeledését (reciprocitás), összebarátkozott a barátai barátaival (tranzitivitás), illetve szívesebben barátkozott lányokkal, mint fiúkkal (nemi homofília). Ezek az endogén strukturális mechanizmusok tehát képesek arra, hogy makroszinten (a teljes osztály kapcsolathálóját figyelembe véve) erősítsék az etnikai elkülönülést. Annak érdekében, hogy az etnikai preferenciákra következtetni tudjunk, kontrollálnunk kell ezekre a mechanizmusokra.

Empirikus példa: egy iskolai osztály baráti hálózatának modellezése

Az exponenciális random gráf modellek használatát egy iskolai osztály baráti hálózatának példáján mutatom be. A kutatási kérdés, amire választ kívánunk kapni, az, hogy a tanulók hajlamosabbak-e a saját etnikai csoportjukból barátokat választani, mint a más etnikai háttérrel rendelkező osztálytársaik közül. A példaként szolgáló

osztály egy Nógrád megyei általános iskola 6. évfolyamának egyetlen osztálya, 23 tanulóval. Az osztályba 8 fiú és 15 lány jár, 8-an tartják magukat romának vagy a magyar mellett romának is, és 15-en tartják magukat kizárólag magyarnak. Az ERG-modellek becslésére az MPNet programot használtam⁶ (Wang et al. 2014).

A kutatási kérdésre három paraméter segítségével adok választ. A *Roma ego* paraméter azt mutatja meg, hogy a roma tanulók hajlamosabbak-e másokat barátnak jelölni, mint a nem romák. A *Roma alter* paraméter azt mutatja meg, hogy a roma tanulók hajlamosabbak-e baráti jelöléseket kapni, mint a nem romák. A *Roma ego*alter* paraméter egy interakciós tag, melynek a segítségével következtetni tudunk a romák csoportján belül a homofília mértékére. E három paraméter felhasználásával feltételes esélyhányadosokat tudunk számolni, melyek megmutatják, hogy a nem roma-nem roma referenciakategóriához viszonyítva milyen a roma-nem roma [$\exp(\beta_{\text{Roma ego}})$], a nem roma-roma [$\exp(\beta_{\text{Roma alter}})$] és a roma-roma [$\exp(\beta_{\text{Roma ego}} + \beta_{\text{Roma alter}} + \beta_{\text{Roma ego*alter}})$] baráti jelölések esélye (Robins–Daraganova 2013: 98).

Az elemzés során különböző strukturális mechanizmusokra kontrollálok, ezek leírása és grafikus ábrázolása az 1. táblázatban található. A reciprocitás paramétere azt mutatja meg, hogy a baráti jelöléseket hajlamosak-e a tanulók viszonzni. A 2-út (2-path) paraméter azt modellezi, hogy akik több jelölést küldenek, azok hajlamosak-e több jelölést is kapni. A többes 2-út (multiple 2-paths) paraméter azt mutatja, hogy milyen mértékű lokális összeköttetésben állnak egymással a tanulók. A tranzitivitás paramétere modellezi, hogy a hálózatban milyen eséllyel alakulnak ki tranzitív triadikus kapcsolatok a többes 2-utakból, míg a ciklicitás paramétere azt mutatja meg, hogy a hálózatban milyen eséllyel alakulnak ki ciklikus triadikus kapcsolatok a többes 2-utakból⁷ (Lusher–Robins 2013a: 44). A befok-csillag (alternating in-star vagy popularity spread) paraméter a bejövő jelölések eloszlásának szóródását modellezi, tehát azt, hogy egyes tanulóknál sok bejövő jelölés koncentrálódik. A kifok-csillag (alternating out-star vagy activity spread) paraméter a kimenő jelölések eloszlásának szóródását modellezi, tehát azt, hogy egyes tanulók másoknál hajlamosabbak sok baráti jelölést küldeni⁸ (Koskinen–Daraganova 2013: 69).

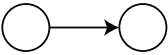
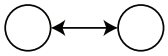
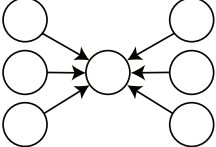
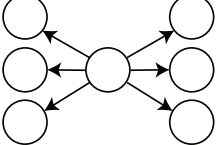
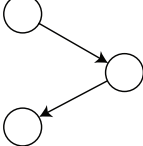
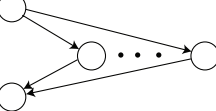
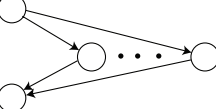
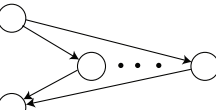
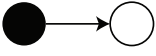
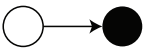
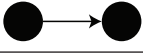
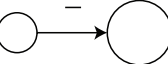
A modellben kontrollálok továbbá a nemi homofíliára (fiú ego, fiú alter, fiú ego*alter), valamint arra, hogy a magasabb tanulmányi átlaggal rendelkező tanulók hajlamosabbak-e baráti jelöléseket küldeni (tanulmányi átlag ego) és kapni (tanulmányi átlag alter), mint az alacsonyabb átlaggal rendelkezők, illetve a hasonló tanulmányi átlaggal rendelkezők hajlamosabbak-e barátok lenni, mint a különböző átlaggal rendelkezők (tanulmányi átlag különbsége).

6 A program a www.sna.unimelb.edu.au/PNet weboldaltól tölthető le. Az ERG-modellek az R programon belül is becsülhetők a statnet elnevezésű csomag használatával (Handcock et al. 2008).

7 Az ERG-modellek konfigurációi gyakran egymásba ágyazottak, ami azt jelenti, hogy egy-egy konfiguráció magában foglal más, egyszerűbb konfigurációkat is. Az 1. táblázat 7. sorában ábrázolt tranzitivitás konfiguráció például tartalmazza a 6. sorban ábrázolt többes 2-út konfigurációt is. Ez azt jelenti, hogy ha a hálózatban a többes 2-út a véletlennél gyakrabban fordul elő, akkor ennek következményeként tranzitív triádból is több várható a hálózatban. Annak érdekében, hogy meg tudjuk mondani, a tranzitív triádok magasabb száma abból adódik, hogy a hálózatban a véletlennél gyakrabban fordulnak elő a többes 2-utak, vagy abból, hogy a többes 2-utakból a véletlennél gyakrabban alakul ki tranzitív triád, a modellben kontrollálnunk kell az alacsonyabb szintű konfigurációkra (Lusher–Robins 2013b: 22–23).

8 A modellben becslést befok-csillag és a kifok-csillag paraméterek az összes fajta befok-, illetve kifok-csillag alakzat (1-csillag, 2-csillag, 3-csillag stb.) súlyozott összegéből tevődnek össze (Koskinen–Daraganova 2013: 65).

1. táblázat: Az exponenciális random gráf modellekben becsült konfigurációk leírása és grafikus ábrázolása

Paraméter	MPNet név	Paraméter leírása	Paraméter grafikus ábrázolása
Jelölés	<i>arc</i>	Általános tendencia a jelölésre	
Reciprocitás	<i>reciprocity</i>	Tendencia a jelölések viszonzására	
Befok-csillag	<i>A-in-S</i>	A bejövő jelölések szóródása	
Kifok-csillag	<i>A-out-S</i>	A kimenő jelölések szóródása	
2-út	<i>TwoPath</i>	A több jelölést küldők hajlamosabbak-e több jelölést is kapni	
Többes 2-út	<i>A2P-T</i>	Tendencia a lokális összeköttetésre több 2-úton keresztül	
Tranzitivitás	<i>AT-T</i>	Tendencia a többes 2-utak tranzitív triadikus kapcsolattá alakulására	
Ciklicitás	<i>AT-C</i>	Tendencia a többes 2-utak ciklikus triadikus kapcsolattá alakulására	
Attribútum ego	<i>Sender</i>	A vizsgált attribútum esetében magasabb értékkel rendelkező személyek hajlamosabbak-e jelölést küldeni	
Attribútum alter	<i>Receiver</i>	A vizsgált attribútum esetében magasabb értékkel rendelkező személyek hajlamosabbak-e jelölést kapni	
Attribútum ego*alter	<i>Interaction</i>	Az attribútum ego és alter paraméterek interakciója	
Attribútum különbség	<i>Difference</i>	Hajlamosabbak-e a szereplők jelölést küldeni, ha nagyobb a különbség köztük és az alter között az attribútum tekintetében	

Megjegyzés: az attribútum ego*alter interakció kategoriális, míg az attribútum különbség paraméter folytonos változók esetén alkalmazható.

Eredmények

A 2. táblázat 1. modellje csak a roma ego, a roma alter és a roma ego*alter paraméterek becslését tartalmazza, míg a 3. táblázat első fele az e modell paraméterbecslései által számolt esélyhányadosokat tartalmazza. Ez a modell azt mutatja meg, milyen esélye van az interetnikus barátságoknak az osztályban abban az esetben, ha nem kontrollálunk a strukturális paraméterekre, valamint a nemi és tanulmányi átlagbeli homofíliára. A roma ego és a roma alter paraméter értéke negatív, ami annyit jelent, hogy a roma tanulók az osztályban kevésbé hajlamosak baráti jelöléseket küldeni és kapni, mint a nem roma tanulók, bár a paraméterbecslések csak a roma alter és a roma ego*alter interakció esetében állnak közel ahhoz, hogy statisztikailag szignifikánsak legyenek (e két paraméter esetében a t -érték abszolút értéke megközelíti a 2-t). A feltételes esélyhányadosok alapján az látszik, hogy a nem roma-nem roma referenciakategóriához viszonyítva a roma-roma baráti jelölések aránya közel azonos (1,02), míg mind a roma-nem roma, mind a nem roma-roma baráti jelölések esélye alacsonyabb (0,68 és 0,57)⁹. Önmagában a tanulók etnikai hovatartozását vizsgálva tehát azt látjuk, hogy a tanulók hajlamosabbak intraetnikus, mint interetnikus barátságok kialakítására és fenntartására.

A 2. modellben kontrollálunk a nemi és tanulmányi átlagbeli homofíliára, viszont nem kontrollálunk a strukturális mechanizmusokra. Más nem lineáris regressziós modellekhez hasonlóan a paraméterbecslések és esélyhányadosok a modellek között nem közvetlenül összehasonlíthatóak (Mood 2010), de annyi megállapítható, hogy a roma ego és roma alter paraméterek értéke pozitívvá válik, és statisztikailag egyik sem szignifikáns. Az esélyhányadosok azt mutatják, hogy az interetnikus baráti jelölések esélye magasabb, mint a nem roma-nem roma jelölésé (roma-nem roma: 1,24, nem roma-roma: 1,96), és a legnagyobb esélye a roma-roma jelöléseknek van (4,10). A nemi homofíliát vizsgálva a lány-lány baráti jelöléshez viszonyítva a fiú-lány és a lány-fiú jelölések esélye jóval kisebb (0,20 és 0,35), míg a fiú-fiú jelölések esélye jóval magasabb (4,33). Az alter tanulmányi átlagára vonatkozó pozitív és szignifikáns paraméterbecslés (tanulmányi átlag alter) azt mutatja, hogy a magasabb tanulmányi átlaggal rendelkező tanulók több baráti jelölést kapnak, míg a negatív és szignifikáns tanulmányi átlag különbség paraméter azt mutatja, hogy minél nagyobb különbség van két tanuló tanulmányi átlaga között, annál kisebb az esély arra, hogy barátok lesznek.

A 3. modell az eddigi paraméterek mellett a strukturális paraméterek becsléseit is tartalmazza. A jelölés paraméter a tanulók általános hajlandóságát mutatja arra vonatkozóan, hogy jelölést küldjenek. A pozitív reciprocitás paraméter azt mutatja, hogy a tanulók hajlamosak a baráti jelöléseket viszonzni. A negatív többes 2-út paraméter alapján a tanulók közötti, tranzitív triád nem képező lokális összeköttetés alacsonyabb mértékű, mint ahogy azt a véletlen alapján várnánk. A pozitív tranzitivitás paraméter alapján a többes 2-út konfigurációkból nagy eséllyel tranzitív triadikus kapcsolatok alakulnak ki.

9 Azt, hogy a különböző feltételes esélyhányadosok közötti különbség statisztikailag szignifikáns-e, további Wald-tesztekkel vizsgálhatjuk.

A strukturális paraméterek bevonásával a hálózati szereplők tulajdonságainak hatását modellező paraméterbecslések alacsonyabbak lettek, mint a 2. modellben.¹⁰ A nem roma-nem roma jelölésekhez viszonyítva a roma-nem roma jelölések esélye közel azonos (1,05), a nem roma-roma jelölések esélye 1,47-szeres (de a különbség statisztikailag nem szignifikáns), míg a roma-roma jelölések esélye 2,64-szeres. A lány-lány jelölésekhez viszonyítva a fiú-lány jelölések esélye 0,56-szoros, a lány-fiú jelölések esélye 0,75-szoros (de a különbség statisztikailag nem szignifikáns), a fiú-fiú jelölések esélye több mint kétszeres. A jobb tanulmányi átlaggal rendelkező tanulók továbbra is hajlamosabbak baráti jelöléseket kapni, mint a rosszabb tanulmányi átlaggal rendelkezők, de a tanulmányi átlag különbségének hatását mérő változó már nem szignifikáns a strukturális paraméterek bevonása után. Látható tehát, hogy a strukturális paraméterek bevonása után a hálózati szereplők jellemzőinek (pl. nem, etnikai hovatartozás, tanulmányi átlag) hatását mérő paraméterbecslések értéke jelentősen megváltozott, a strukturális paraméterek bevonása nélkül e változók hatását felülbecsülnénk.

A 3. modell paraméterbecslése az MPNet programba bevonható összes paraméter tekintetében, és a hálózat egyéb jellemzőit tekintve (pl. bejövő és kimenő jelölések szórása és eloszlásának ferdesége, klaszteresedés) is jó illeszkedést mutat, tehát a modell jól reprezentálja a hálózatnak azokat a jellemzőit is, melyeket a becslési eljárás során nem modelleztünk.

2. táblázat: Az exponenciális random gráf modellek paraméterbecslései

	1. modell			2. modell			3. modell		
Paraméter	Paraméterbecslés	Standard hiba		Paraméterbecslés	Standard hiba		Paraméterbecslés	Standard hiba	
Jelölés	-1,111	0,159	*	-3,654	1,166	*	-3,878	1,218	*
Reciprocitás							1,876	0,488	*
2-út							-0,021	0,059	
Befok-csillag							0,059	0,459	
Kifok-csillag							0,344	0,416	
Tranzitivitás							0,638	0,224	*
Ciklicitás							-0,180	0,157	
Többes 2-út							-0,207	0,071	*
Roma ego	-0,385	0,289		0,217	0,405		0,046	0,320	
Roma alter	-0,556	0,289		0,673	0,401		0,385	0,363	
Roma ego*alter	0,964	0,497		0,523	0,623		0,541	0,514	
Fiú ego				-1,592	0,408	*	-0,586	0,291	*
Fiú alter				-1,051	0,383	*	-0,294	0,291	
Fiú ego*alter				4,107	0,572	*	1,657	0,481	*
Tanulmányi átlag ego				0,122	0,178		0,055	0,152	
Tanulmányi átlag alter				0,624	0,172	*	0,385	0,168	*
Tanulmányi átlag különbség				-0,436	0,169	*	-0,179	0,119	

10 Mivel nem lineáris regressziós modellekben újabb magyarázó változók bevonásával a korábbi paraméterbecslések értéke növekedhet, a paraméterbecslések csökkenésének mértékét valószínűleg még alul is becsültük (Mood 2010).

3. táblázat: Az etnicitás hatását mutató feltételes esélyhányadosok a különböző modellekben

	1. Modell		2. Modell		3. Modell	
	nem roma	roma	nem roma	roma	nem roma	roma
nem roma	1,00	0,57	1,00	1,96	1,00	1,47
roma	0,68	1,02	1,24	4,10	1,05	2,18

Megjegyzés: a referenciakategória a nem roma-nem roma jelölés.

A bemutatott modellben egyetlen osztály esetében vizsgáltuk a kapcsolathálót kialakító mechanizmusokat. Lehetséges azonban több hálózat (pl. több iskolai osztály) alapján metaelemzést végezni oly módon, hogy minden hálózat esetében külön-külön megbecsüljük a paraméterértékeket és a standard hibákat, majd ezeket felhasználva metaelemzés segítségével vizsgáljuk, hogy látunk-e valamilyen általános tendenciát a paraméterbecslésekben (Snijders–Baerveldt 2003).

Összefoglalás

A tanulmányban az exponenciális random gráf modellek használatát mutattam be. Az ERG-modellek olyan statisztikai modellek, melyek képesek figyelembe venni és modellezni a kapcsolatok közötti függőségi viszonyokat, ezáltal alkalmasak arra, hogy feltárják a társadalmi kapcsolatháló alakulásában szerepet játszó mechanizmusokat még akkor is, ha keresztmetszeti adataink vannak. Az ERG-modellek használata akkor is hasznos, ha nem e mechanizmusok állnak kutatásunk középpontjában, mivel e mechanizmusokra kontrollálva más magyarázó változók hatását (pl. nem, etnikai hovatartozás) torzítatlanul tudjuk becsülni.

Példaként egy 6. évfolyamos általános iskolai osztály baráti hálózatát elemeztem. Az elemzés során arra voltam kíváncsi, tapasztalható-e etnikai homofília a tanulók baráti jelöléseiben, és ha igen, azt milyen mechanizmusok hozhatták létre. Az elemzés során láttuk, hogy a strukturális mechanizmusok bevonása nélkül felülbecsültük a tanulók tulajdonságainak (nem, etnikai hovatartozás, tanulmányi átlag) baráti kapcsolatokra gyakorolt hatását. Az endogén hálózati mechanizmusok tehát hozzájárulnak ahhoz, hogy a tanulók bizonyos jellemzői szerint szegregált kapcsolatháló jönjenek létre.

E tanulmányban bemutattam, hogyan tudunk ERG-modelleket becsülni, ha egyetlen keresztmetszeti hálózat áll kutatásunk középpontjában. ERG-modellek segítségével azonban két hálózat egymással való összefüggése is vizsgálható, valamint többszintű hálózatokat (Wang 2012) és longitudinális adatokat (Snijders–Koskinen 2012) is modellezni tudunk. Eme eljárások részletes ismertetése nem témája tanulmányomnak, leírásukat az idézett szakirodalmak tartalmazzák.

Abstract: Exponential random graph models (ERGMs) are statistical models for social networks. Social networks cannot be adequately analysed with standard statistical models such as logistic regression,

because these techniques assume independence among observations. In social networks, the presence or absence of ties in the network affect the presence or absence of other ties. ERGMs explicitly model these dependence among ties. By estimating the effects of several local network configurations, inferences about the underlying social mechanisms can be made. Moreover, the effects of attribute-based parameters (e. g., ethnicity, gender) on network ties might be overestimated if endogenous structural network processes are not controlled for. In this study, I give an overview of ERGMs and illustrate their use with an empirical example analysing the friendship network of a sixth-grade school class.

Keywords: ERGM, social network analysis, statistical modelling

Irodalom

- Block, P. – Koskinen, J. – Hollway, J. – Steglich, C. – Stadtfeld, C. (2018): Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, 52(1): 180–191.
- Dekker, D. – Krackhardt, D. – Snijders, T. A. B. (2007): Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, 72(4): 563–581.
- Desmarais, B. A. – Cranmer, S. J. (2012): Micro-level interpretation of exponential random graph models with application to estuary networks. *Policy Studies Journal*, 40(3): 402–434.
- Frank, O. – Strauss, D. (1986): Markov random graphs. *Journal of the American Statistical Association*, 81: 832–842.
- Hallinan, M. T. – Williams, R. A. (1987): The stability of students' interracial friendships. *American Sociological Review*, 52(2): 653–664.
- Hallinan, M. T. – Williams, R. A. (1989): Interracial friendship choices in secondary schools. *American Sociological Review*, 54(1): 67–78.
- Handcock, M. S. – Hunter, D. R. – Butts, C. T. – Goodreau, S. M. – Morris, M. (2008): statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1): 1–11. <http://www.jstatsoft.org/v24/i01>.
- Hanneke, S. – Fu, W. – Xing, E. (2009): Discrete Temporal models of social networks. *arXiv:0908.1258*, <http://arxiv.org/abs/0908.1258>.
- Holland, P. W. – Leinhardt, S. (1981): An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373): 33–50.
- Kandel, D. B. (1978): Homophily, selection and socialization in adolescent friendships. *American Journal of Sociology*, 84(2): 427–436.
- Koskinen, J. – Daraganova, G. (2013): Exponential random graph model fundamentals. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 49–76.

- Koskinen, J. – Snijders, T. A. B. (2013): Simulation, estimation, and goodness of fit. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 141–166.
- Krackhardt, D. (1988): Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10(4): 359–381.
- Krivitsky, P. N. – Handcock, M. S. (2014): A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1): 29–46.
- Lazarsfeld, P. – Merton, R. (1954): Friendship as a social process: A substantive and methodological analysis. In Berger, M. – Abel, T. – Pege, C. (eds.): *Freedom and Control in modern Society*. New York: Van Norstrand, 18–66.
- Lusher, D. – Koskinen, J. – Robins, G. (eds.). (2013a): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press.
- Lusher, D. – Koskinen, J. – Robins, G. (2013b): Introduction. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 1–6.
- Lusher, D. – Robins, G. (2013a): Example exponential random graph model analysis. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 37–46.
- Lusher, D. – Robins, G. (2013b): Formation of social network structure. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 16–28.
- McPherson, M. – Smith-Lovin, L. – Cook, J. M. (2001): Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27: 415–444.
- Mood, C. (2010): Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1): 67–82.
- Pattison, P. – Robins, G. (2002): Neighborhood-based models for social networks. *Sociological Methodology*, 32(1): 301–337.
- Rapoport, A. (1957): Contribution to the theory of random and biased nets. *The Bulletin of Mathematical Biophysics*, 19(4): 257–277.
- Robins, G. – Daraganova, G. (2013): Social selection, dyadic covariates, and geospatial effects. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 91–101.
- Robins, G. – Lusher, D. (2013a): What are exponential random graph models? In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 9–15.

- Robins, G. – Lusher, D. (2013b): Illustrations: Simulation, Estimation, and goodness of fit. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 167–85.
- Robins, G. – Pattison, P. – Kalish, Y. – Lusher, D. (2007): An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, 29(2): 173–191.
- Snijders, T. A. B. – van de Bunt, G. G. – Steglich, C. E. G. (2010): Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1): 44–60.
- Snijders, T. A. B. – Koskinen, J. (2012): Longitudinal models. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 130–40.
- Snijders, T. A. B. – Pattison, P. E. – Robins, G. L. – Handcock, M. S. (2006): New specifications for exponential random graph models. *Sociological Methodology*, 36(1): 99–153.
- Snijders, T. A. B. (2002): Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2): <http://www.cmu.edu/joss/content/articles/volume3/Snijders.pdf>.
- Snijders, T. A. B. – Baerveldt, C. (2003): A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology*, 27(2): 123–151.
- Steglich, C. – Snijders, T. A. B. – Pearson, M. (2010): Dynamic networks and behavior: separating selection from influence. *Sociological Methodology*, 40(1): 329–39.
- Tolsma, J. – van Deurzen, I. – Stark, T. H. – Veenstra, R. (2013): Who is bullying whom in ethnically diverse primary schools? Exploring links between bullying, ethnicity, and ethnic diversity in Dutch primary schools. *Social Networks*, 35(1): 51–61.
- Van Duijn, M. A. J. – Snijders, T. A. B. – Zijlstra, B. J. H. (2004): p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2): 234–254.
- Veenstra, R. – Lindenberg, S. – De Winter, A. F. – Zijlstra, B. J. H. – Verhulst, F. C. – Ormel, J. (2007): The dyadic nature of bullying and victimization: Testing a dual-perspective theory. *Child Development*, 78(6): 1843–1854.
- Wang, P. (2012): Exponential random graph model extensions: Models for multiple networks and bipartite networks. In Lusher, D. – Koskinen, J. – Robins, G. (eds.): *Exponential Random Graph Models for Social Networks: Theories, Methods, and Applications*. Cambridge, New York: Cambridge University Press, 115–129.
- Wang, P. – Robins, G. – Pattison, P. – Koskinen, J. (2014): *MPNet, Program for the Simulation and Estimation of (p^*) Exponential Random Graph Models for Multilevel Networks*. Melbourne, Australia: Melbourne School of Psychological Sciences, The University of Melbourne.
- Wasserman, S. – Pattison, P. (1996): Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p. *Psychometrika*, 61(3): 401–425.