

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Mázsár Noémi

A járványterjedés modellezése véletlen gráfokon

BSc Szakdolgozat

Témavezető:

Backhausz Ágnes

Valószínűségelméleti és Statisztika Tanszék



Budapest, 2013

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek, Backhausz Ágnesnek, hogy figyelemmel kísérte a szakdolgozatom készülését, segített a felmerülő problémákat megoldani, és mindig időt szakított rám a félév során. Külön köszönetet szeretnék mondani a \LaTeX használatával kapcsolatos tanácsokért.

Köszönöm minden tanáromnak, akik az elmúlt három év során segítettek a szakmai fejlődésemet, köszönöm a gimnáziumi matematika tanáromnak, hogy elindított ezen az úton, és köszönettel tartozom családomnak, páromnak, szaktársaimnak a bátorításukért és támogatásukért.

Budapest, 2013. május 27.

Mázsár Noémi

Tartalomjegyzék

Bevezetés	4
1. Véletlen gráfmodellek	5
1.1. Erdős–Rényi véletlen gráf	6
1.2. Skálafüggetlen véletlen gráf általános modellje	7
1.3. Konfigurációs modell	8
1.4. „Preferential attachment” modell	9
1.4.1. Barabási–Albert-modell	12
1.4.2. Barabási–Albert-fa	12
1.5. Egyéb gráfmodellek	13
2. A járványterjedés modellezése	14
2.1. Modellezés az egy- és kétkomponensű Erdős–Rényi-gráfon	14
2.2. Járványterjedés a „preferential attachment” gráfmodellen	22
2.3. Az influenzavírus terjedése	28
2.4. A 2008 – 2009-es influenza terjedése megfigyelések alapján	31
3. Az internetes vírusok terjedése	34
3.1. Vírusok terjedése mobiltelefonok között	36
Irodalomjegyzék	38

Bevezetés

Képzeljük el a következő szituációt: egy repülőtéren a várakozó utasoknak 0,1%-a egy adott vírussal fertőzött. Ez a vírus veszélyes, könnyen terjed, és tegyük fel, hogy máshol még nincsenek fertőzött személyek, csak itt. Ám a repülőgépeket nem sikerül megállítani, így a fertőzést hordozó utasok elrepülnek különböző országokba, akár más-más földrészekre. Nem tudjuk, hogy személy szerint kik hordozzák a betegséget, így nem tudjuk, mely városokba kerülhet be a vírus. Vajon hogy tudnánk ilyen kevés információ mellett megakadályozni egy járvány kitörését? Ebben segít a gráfmodellek használata.

Természetesen ebben a szituációban kifejezetten nehéz dolgunk van, mivel csak azután tudjuk elkezdni a vírus elleni védekezést, miután már megjelent az egyes városokban. Ezután sok megoldás jöhet szóba: a vírus kellő ismeretében tudhatjuk, hogy mi lehet ellene a hatásos védekezési módszer, milyen módon terjed a fertőzés, illetve mennyi ideig tudja egy ember továbbadni a vírust egy másiknak. Ezen szempontok alapján különböző kezeléseket alkalmazhatunk.

A gráfmodellek rendkívül sokfélék lehetnek, ezek közül ismertetünk néhány típust az 1. fejezetben. Vannak, melyek alkalmasak a valós hálózatok modellezésére, vannak, amik ehhez túl szabályosak, rendszerezettek, ezért kevésbé életszerűek. A két legrészletesebben vizsgált modell az Erdős–Rényi-gráf és a „preferential attachment” típusú gráfmodell. Ezekon a gráfokon nézzük a járványterjedést, ennek legfőbb tulajdonságait, és a modellek közti különbségeket. Szót ejtünk az influenza terjedéséről, és megemlítünk pár konkrét mérési adatot is ezzel kapcsolatban, melyek mutatják néhány elméleti eredményünk helyességét, mindemellett hangsúlyozzák, hogy a modellezés nem minden esetben tökéletesen valóságos, és a konkrét esetekben a modellek jóval komplexebbek, és bonyolultabbak, mint azt gondolnánk. Végül, de nem utolsósorban megemlítjük, hogy az emberek közti járványterjedéshez nagy mértékben hasonlít az internetes vírusok terjedése, vagy akár a mobiltelefonos vírusok terjedése is.

1. fejezet

Véletlen gráfmodellek

A véletlen gráfmodellek tanulmányozása egy viszonylag új ága a matematikának, ami ma is látványosan fejlődik, napjainkban is születnek újabb és újabb eredmények, módosítások. Rengeteg véletlen gráfot ismerünk, melyek közül néhány igen csak mesterségesnek tűnik, de vannak, amikkel valós problémákat is jól modellezhetünk. Mindkét esetre láthatunk majd példát, de természetesen a teljesség igénye nélkül.

Először is nézzünk néhány érdekes tulajdonságot, ami a valós hálózatokra jellemző. Ezek egyike a kis világ jelenség, ami szerint bármely két ember a világból hat személyes ismeretségen keresztül ismeri egymást. 1967-ben Stanley Milgram, amerikai pszichológus végzett kísérletet, azzal a céllal, hogy a hat lépés távolságot igazolja. Leveleket küldött több embernek, hogy juttassák el egy adott személynek, de csak a személyes ismeretségeiken keresztül. Az első kísérlet még nem járt nagy sikerrel, de a későbbiek – ahol Milgram kisebb trükkökkel fontosnak tüntette fel, hogy a levél eljusson a címzetthez – már több sikerrel értek célba. Napjainkban a technika hihetetlen gyorsaságú fejlődése mellett a hat lépés távolság bármely két ember között a Földön már nem is olyan elképzelhetetlen: gondoljuk csak az internetre, mint összekötő hálózatra. Persze felmerülhet bennünk a kérdés: mit értünk az alatt, hogy két ember ismeri egymást? Lehet, hogy csak elektronikusan kommunikáltak, épp csak tudják egymás nevét, vagy személyesen találkoznak nap, mint nap, ezen eseteket most nem vizsgáljuk részletesebben.

A hat lépésnyi távolság modellezésére, mind a matematikusok, mind a színészek körében készült ismert példa. A matematikusok körében az Erdős Pállal való ismeretséget figyeljük meg. Megfeleltetjük a matematikusokat egy gráf csúcsainak, és két csúcs között akkor vezet él, ha a két csúcsnak megfelelő matematikusok publikáltak közös cikket. Erdős Páltól vett távolságot az adott csúcsnak megfelelő matematikus Erdős-számának nevezzük. Az átlagos-Erdős szám a [16][10. oldal, 1.3 táblázat] alapján 4,646. Ugyanezt a modellt a színészek között is létrehozták, Erdős Pál helyett Kevin Bacon szereplésével. Itt két színésznek megfeleltetett csúcsok akkor vannak összekötve, ha szerepeltek közös filmben.

Az eddigi néhány példában különböző nagyságú és összetettségű gráfmodellt láthattunk. A különbözőségük miatt érdemes véletlen gráfokat használni a modellezéshez, azonban a mélyebb matematikai megértéshez előbb meg kell ismernünk a különböző csoportosítási szempontokat, és a gráfmodellek alapvető tulajdonságait.

A véletlen gráfokat sok tulajdonság alapján csoportosíthatjuk. Az egyik legfontosabb szempont, hogy a gráf képes-e a növekedésre, változásra, vagy sem. Ez alapján vannak a statikus gráfmodellek, melyeknek előre adott a méretük, és a dinamikus modellek, azaz amik növekedhetnek, ezáltal élethűbben modellezve a valós hálózatokat. Nézzük először a statikus modelleket.

1.1. Erdős–Rényi véletlen gráf

Ez talán az egyik legegyszerűbb gráfmodell, melynek több változatát is ismerjük, de az eltéréseknek nem szükséges túl nagy jelentőséget tulajdonítanunk. Az eredeti Erdős Pál és Rényi Alfréd által 1959-ben leírt modellben [11] adott n csúcs és M él, és az összes ilyen gráf közül egyenlő valószínűséggel választjuk az egyiket. Itt ugye látjuk, hogy ha egy gráfmodellt tekintünk, akkor abban az egyik él behúzásának valószínűsége valamilyen értelemben függ a másiktól, mivel az élek száma előre meghatározott. A későbbiekben majd láthatjuk, hogy ennek itt nem lesz fontos szerepe. A másik megközelítés, amit napjainkban szintén Erdős–Rényi-modell néven ismerünk, először az Edgar Gilbert által 1960-ban kiadott cikkben [12] jelent meg, miszerint előre adott n , a gráf csúcsainak száma, és ezek között minden élt a többitől függetlenül p valószínűséggel húzunk be. Ezt a p valószínűséget gyakran élvalószínűségnek nevezik. Ebben az esetben az élek száma nem előre meghatározott.

Vizsgáljuk meg az Erdős–Rényi-gráf néhány fontos tulajdonságát. Az élek behúzási valószínűsége két csúcs között p , amit tetszőlegesen megválaszthatunk a $[0, 1]$ intervallumból. A p -t gyakran $p = \frac{\lambda}{n}$ alakban adjuk meg, ez a jelölés fontos lesz a későbbiekben. Ettől a választástól függően a gráf különböző alakú lehet: ha $p = \frac{\lambda}{n}$ és $\lambda < 1$, akkor a gráf több kis $\Theta(\log n)$ nagyságú komponensből fog állni, ha $\lambda = 1$ akkor a gráf legnagyobb komponense nagy valószínűséggel $n^{\frac{2}{3}}$ nagyságrendű lesz, ha pedig $p = \frac{\lambda}{n}$ olyan, hogy $\lambda > 1$, akkor a gráf egy nagy $\Theta(n)$ méretű és több kisebb $\Theta(\log n)$ méretű komponensből fog állni. Azaz, ha $\lambda > 1$, akkor $\exists c > 0$ adott, amire

$$P(\text{az } n \text{ csúcsú gráfban a legnagyobb komponens } \geq cn) \rightarrow 1,$$

ha $n \rightarrow \infty$.

1.1.1. Megjegyzés. A Θ jelentése általánosan a következő: $f(n) = \Theta(g(n))$, ha $\exists c_1, c_2$ pozitív számok, hogy $c_1 g(n) \leq f(n) \leq c_2 g(n)$ teljesül $\forall n$ -re.

Végül nézzük meg az Erdős–Rényi-gráf foksámeloszlását. Feltehetjük, hogy $\lambda \in (0, 1]$, hiszen, ha $\lambda = 0$ lenne, akkor $p = 0$, ekkor nincsenek élek a gráfban. Ekkor az élek számának eloszlása

binomiális p paraméterrel. A k fokszámú csúcsok aránya közel van a $P(\text{Bin}(n-1, \frac{\lambda}{n}) = k)$ valószínűséghez. Tudjuk, hogy ha n nagy, akkor a megfelelő binomiális eloszlás közel van a λ paraméterű Poisson-eloszláshoz. Ehhez ismernünk kell a Poisson-eloszlást, ami a következő:

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!},$$

ahol $k \geq 0$. Jelölje D_i az i csúcs fokszámát, ekkor a fokszámeloszlás legyen

$$P_k^{(n)} = \frac{1}{n} \sum_{i=1}^n I_{\{D_i=k\}},$$

tehát $P_k^{(n)}$ jelöli a k fokú csúcsok arányát n lépés után.

1.1.2. Tétel (Az Erdős–Rényi-gráf fokszámeloszlása). *Rögzítsünk egy $\lambda \in (0, 1]$ számot. $\forall \varepsilon_n$ -re, ha $\sqrt{n}\varepsilon_n \rightarrow \infty$, akkor*

$$P_\lambda(\max_k |P_k^{(n)} - p_k| \geq \varepsilon_n) \rightarrow 0.$$

1.1.3. Megjegyzés. *Tehát az előző tételben p_k a k fokú csúcsok arányának a limeszét jelöli.*

Ez a modell nem igazán hasonlít a valós hálózatokhoz, hiszen gondoljunk csak bele: igaz lenne az, hogy bármely ember egyforma valószínűséggel kap el egy betegséget? Vagy mindegy, hogy egy fontos hírportálról vagy egy kevésbé látogatott weboldalról fertőződött meg egy internetes vírussal, vajon egyforma eséllyel lesz vírusos a gépünk? Érezzük, hogy a helyes válasz mindkét kérdésre a nem. Szerencsére ezek a kérdések nem csak bennünk merültek fel, így már vannak olyan modellek, amik jobban ábrázolják a valódi hálózatokat.

1.2. Skálafüggetlen véletlen gráf általános modellje

Itt adott n csúcs, és minden csúcsnak van egy adott vagy egy véletlenül sorsolt súlya, W_i ($i = 1, \dots, n$). Az éleket egymástól függetlenül húzzuk be, de egy él szereplésének esélyét befolyásolják a csúcsok súlyai, amik közt be akarjuk húzni. Az adott él valószínűség i és j csúcsok közt a következő:

$$p_{ij} = \frac{W_i W_j}{L_n + W_i W_j} \quad (1.1)$$

ahol $\{W_i\}_{i=1}^n$ a csúcsok súlya, $L_n = \sum_{i=1}^n W_i$, az összes csúcs súlyának összege. Az eddigi gráfmodelljeink nem voltak skálafüggetlenek, de itt tudjuk úgy választani a súlyokat, hogy skálafüggetlen modellt kapjunk. A valós hálózatokra ez gyakran jellemző tulajdonság, ezért érdemes pontosan definálnunk.

1.2.1. Definíció. Egy adott G_n ($n = 1, 2, \dots$) gráfsorozat skálafüggetlen, ha a gráfban a k fokú csúcsok aránya 1 valószínűséggel konvergál valamely p_k számhoz $\forall k$ esetén, ha $n \rightarrow \infty$, és valamilyen $\gamma > 0$ -ra és $c > 0$ -ra $p_k k^\gamma \rightarrow c$, ha $k \rightarrow \infty$. Azaz $p_k \approx ck^{-\gamma}$.

Következő fontos tulajdonság a fokszámoszlás, térjünk ki erre ennél a modellenél is. Vezessük be itt is a következő jelölést a k fokú csúcsok arányára n lépés után:

$$P_k^{(n)} = \frac{1}{n} \sum_{i=1}^n I_{\{D_i=k\}}.$$

A fokszámoszlás a Poisson eloszláshoz hasonlít. Ha $k \geq 0$, akkor

$$p_k = E\left(e^{-W} \frac{W^k}{k!}\right).$$

Ha itt a W eloszlását megfelelően választjuk, akkor elérhetjük a skálafüggetlenséget.

1.2.2. Megjegyzés. Ha W_i súlyok független azonos eloszlásúak, akkor W is ugyanilyen eloszlású. Ha pedig a W_i -k nem ilyenek, akkor nem tudunk egyetlen közös W -t találni.

1.2.3. Tétel (A skálafüggetlen gráf fokszámoszlása [16] 6.9 tétel). Két feltételre van szükségünk a tétel teljesüléséhez:

(a) Gyenge konvergencia a csúcsok súlyaira: $W_n \xrightarrow{d} W$, ha $n \rightarrow \infty$.

(b) Konvergencia az átlagos súlyú csúcsok súlyaira: ha $E(W) > 0$, $\lim_{n \rightarrow \infty} E(W_n) = E(W)$.

Ekkor $\forall \varepsilon > 0$ esetén $P\left(\sum_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon\right) \rightarrow 0$.

Azaz a tétel azt mondja ki, hogy a k fokú csúcsok aránya n lépés után tart p_k -hoz, sztochasztikus konvergenciával, tehát, ha $n \rightarrow \infty$.

1.3. Konfigurációs modell

Ennél a modellenél a csúcsok fokszáma előre adott, és természetesen a csúcsok száma is, legyen ez n . Jelölje D_i az i csúcs foksámát, és $L_n = \sum_{i=1}^n D_i$ a csúcsok foksámainak összegét. Ismerjük tehát az összes csúcs foksámát, és ez alapján húzunk be éleket a gráfban véletlenszerűen (természetesen ez alapján többféle gráfot is készíthetünk). A modell készítését úgy képzelhetjük el, hogy van n darab csúcsunk, és minden csúcshoz csatlakozik éppen annyi él, amennyi az adott csúcs fokszáma. Az él másik vége szabad, ezt egyelőre nem kötöttük sehova. Ezután ha véletlenszerűen választunk két szabad élvéget, és ezeket összekötjük, akkor megkapjuk a gráf egy élet. Ezt többféleképp is megtehetjük, azt a célt tartva szem előtt, hogy a gráfunk egyszerű legyen, azaz ne legyenek benne többszörös élek, illetve hurokélek, ami persze nem minden esetben megvalósítható.

Tehát adott egy konfigurációs gráfmodell, és mi ezt egyszerűvé szeretnénk alakítani. Erre az egyik lehetőség az ismétléses konfigurációs modell, ahol először elkezdjük véletlenszerűen összekötni a csúcsokat. Ezt addig csináljuk, míg a gráf egyszerűsége meg nem szűnik, ekkor minden eddigi összekötésről elfeledkezünk az ismétléses modell szerint, majd újrakezdjük a próbálgatást, egészen addig míg nem sikerül az egyszerű gráfmodellünket létrehozni. Ez alapján a megfelelőek, azaz az egyszerű gráfok egyformán valószínűek lesznek.

A másik lehetőség a törléses konfigurációs modell, ahol először fixáljuk a fokszámot, legyen ez d . Majd az n csúcsú, már adott konfigurációs gráfmodellünkből, ahol többszörös éleket is megengedünk, és minden csúcs foka d , elhagyjuk a hurokéleket és a többszörös élek közül kitörölünk annyit, hogy csak egy él maradjon ott is, ahol eddig több volt. Így eltérünk ugyan az eredeti gráftól, nem kapunk pontos eredményt, hiszen a fokszámok csökkenhetnek, de aszimptotikusan ugyanaz marad a gráf, azaz egy viszonylag nagy modell esetén nem történik számottevő változás.

Az eddigi gráfmodellek nem mutatják a gráf létrejöttét, változását, esetleg jövőbeli növekedését, így ahhoz, hogy ezeket is megismerhessük, áttérünk a dinamikus modellekre.

1.4. „Preferential attachment” modell

Ez egy növekedő gráfmodell, azaz folyamatosan új csúcsokkal, és élekkel bővül. Yule volt az első, aki a növekvő gráfokkal foglalkozott 1925-ben, majd Barabási Albert-László és Albert Réka jelentős eredményeket értek el ezen a téren, az ő modelleikről még lesz szó a későbbiekben. Pontosan ezzel a modellel Bollobás Béla foglalkozott először [6].

A „preferential attachment” modell alkalmas a valós hálózatok növekedésének modellezésére, hiszen itt nem minden csúcshoz egyforma valószínűséggel kapcsolódnak az új élek, hanem a nagyobb fokszámú csúcsokhoz nagyobb valószínűséggel kapcsolódnak, mint a kisebbekhez. A valóságban gyakran megjelenik ilyen típusú hálózat, mely növekszik, de nem mindenhol egyforma mértékben.

A növekedés miatt ez egy gráfsorozatot modellez, melyet ezután jelöljünk a következőképpen: $\{PA_t(m, \delta)\}_{t=1}^{\infty}$. Ez egy adott t -re egy t csúcsú gráfot ad, amelyben az élszám mt , m pozitív egész szám, mely azt jelöli, hogy egyszerre hány darab éllel fog kapcsolódnani az új csúcs. A δ szintén egy általunk választott paraméter, amely megfelelő megadásáról kicsit később lesz szó. Vizsgáljuk először az $m = 1$ esetet. $PA_1(1, \delta)$ egy izolált pontból és egy hurokélből áll. Ezután tegyük fel, hogy már van egy t csúcsból álló gráfunk, és nézzük meg, hogy zajlik a $(t + 1)$ -edik csúcs hozzávétele. Az új csúcs egy éllel kapcsolódik valamely ponthoz a fokszámmal arányos valószínűség szerint. Természetesen önmagához is kapcsolódhat, de ugye nem túl nagy eséllyel.

Tehát annak a valószínűsége, hogy a $t + 1$ -edik csúcs hova kapcsolódik, a következő:

$$P(v_{t+1}^{(1)} \rightarrow v_i^{(1)} \mid PA_t(1, \delta)) = \begin{cases} \frac{1+\delta}{t(2+\delta)+(1+\delta)} & \text{ha } i = t + 1 \\ \frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)} & \text{ha } i \in \{1, 2, \dots, t\} \end{cases} \quad (1.2)$$

A képletben $D_i(t)$ jelöli a $v_i(t) \in PA_t(1, \delta)$ fokszámát.

A definíciót felhasználva, gondoljuk át a δ szerepét és lehetséges értékeit. Látjuk, hogy szükséges feltétel a $\delta \geq -1$, hiszen különben a definícióban a felső ágon negatív érték szerepelne, ami nem lehet valószínűség. Ha δ -t elég nagyra választjuk, akkor a fokszámok nem igazán számítanak, egy egyenletes valószínűségeloszlást kapunk. Ha pedig δ kicsi, vagy esetleg nulla, akkor a fokszámokkal arányosan oszlanak meg a valószínűségek.

Nézzük az $m > 1$ esetet. Kezdetben legyen adott $PA_{mt}(1, \frac{\delta}{m})$ gráfsorozat, aminek csúcsai $v_1^{(1)}, \dots, v_{mt}^{(1)}$. $PA_{mt}(1, \frac{\delta}{m})$ -beli $v_1^{(1)}, \dots, v_m^{(1)}$ csúcsokat vonjuk össze a $PA_t(m, \delta)$ gráfban a $v_1^{(m)}$ csúccsá. Általánosan leírva a $PA_{mt}(1, \frac{\delta}{m})$ gráfbeli $v_{(j-1)m+1}^{(1)}, \dots, v_{jm}^{(1)}$ csúcsokból keletkezik a $v_j^{(m)}$ csúcs a $PA_t(m, \delta)$ gráfban. Tehát az eredetiből vett m darab csúcsonként a csúcsok összevonásával keletkezik az új gráfban 1 csúcs. Az m darab csúcsot a régi gráfban az új csúcs őseinek nevezzük. Két csúcs össze van kötve az $PA_t(m, \delta)$ gráfban, ha az adott két csúcsnak bármely két őse össze van kötve a $PA_{mt}(1, \frac{\delta}{m})$ gráfban. Az így keletkezett $PA_t(m, \delta)$ egy t csúcsú, mt élű, $2mt$ összfokszámú multigráf, azaz lehet benne hurokél vagy többszörös él. A modell egyszerűsítéséhez természetesen szükséges meghatározni azokat az együttes valószínűségeket, hogy ha egy új csúcs m éllel kapcsolódik az eddigi gráfhoz, akkor tetszőleges m tagból álló csoportra mennyi a valószínűsége, hogy épp ehhez fog kapcsolódni. Ezt persze nem csak ennél a modellnél adhatjuk meg, hanem minden olyanánál, ahol az új csúcs egyszerre nem csak egy éllel kapcsolódik.

A következő tételben a „preferential attachment” modell maximális fokszámát vizsgáljuk, amire vezessük be a következő jelölést: $M_t = \max_{i=1, \dots, t} D_i(t)$, ahol $D_i(t)$ a $v_i \in PA_t(m, \delta)$ fokszámát jelöli.

1.4.1. Tétel. *A „preferential attachment” modell esetén $PA_t(m, \delta)$ -ban, rögzítsük $m \geq 1$ -et és $\delta \geq -m$ -et. Ekkor*

$$M_t t^{-\frac{1}{\tau-1}} \rightarrow \mu,$$

ahol $t \rightarrow \infty$ és $P(\mu = 0) = 0$.

1.4.2. Megjegyzés. *Definiáljuk a τ -t az előbbi tétel megértéséhez: $\tau = 3 + \frac{\delta}{m}$. Egy triviális becslést ismerünk τ -ra, miszerint $\tau = 3 + \frac{\delta}{m} > 2$, hiszen $\frac{\delta}{m} > -1$.*

A valós hálózatokban is $\tau > 2$ az általános.

Következő célunk, hogy megvizsgáljuk, hogy a modell skálafüggetlen, vagy sem. Ehhez határozzuk meg a k fokú csúcsok arányát a gráfban, hiszen ez szorosan összefügg a skálafüggetlenséggel. Ez egy viszonylag könnyen mérhető mennyiség az irányított gráfok esetén, mert egyszerre

csak egy csúcsot kell vizsgálnunk. $P_k(t) = \frac{1}{t} \sum_{i=1}^t I_{\{D_i(t)=k\}}$ legyen annak a jelölése, hogy t lépés után a k fokszámmal rendelkező csúcsok aránya mennyi a gráfban. Ha $m > 1$ és $\delta \geq -m$, akkor tudunk definiálni egy olyan $\{p_k\}_{k=0}^{\infty}$ eloszlást, amit a következő konstansok határoznak meg: $p_k = 0$, ha $k = 0, \dots, m-1$, és

$$p_k = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(k+\delta)\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(k+3+\delta+\frac{\delta}{m})}, \quad \text{ha } k \geq m.$$

Ha $m = 1$, akkor a képlet a következő egyszerűbb alakban írható fel:

$$p_k = (2+\delta) \frac{\Gamma(k+\delta)\Gamma(3+2\delta)}{\Gamma(\delta)\Gamma(k+3+2\delta)},$$

ha pedig $\delta = 0$ és $k \geq m$, akkor:

$$p_k = \frac{2\Gamma(k)\Gamma(m+2)}{\Gamma(m)\Gamma(k+3)} = \frac{2m(m+1)}{k(k+1)(k+2)}$$

alakban írhatjuk.

Ahhoz, hogy az előbbi összefüggéseket értelmezni tudjuk, ismernünk kell a Γ jelentését.

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Nekünk most csak a Γ azon tulajdonságára van szükségünk, miszerint $\Gamma(n) = (n-1)!$ és $\Gamma(a) = (a-1)\Gamma(a-1)$. Kifejtve: $\Gamma(k+\delta) = (k+\delta-1)(k+\delta-2)\dots(m+\delta)\Gamma(m+\delta)$, ahol $k > m$. Éppen a Γ függvény ezen kifejtése miatt le lehet egyszerűsíteni, így kapjuk, hogy a p_k polinomiális.

A következő tétel azt mondja ki, hogy ha $p_k \geq 0$, $\{p_k\}_{k=1}^{\infty}$ valóban eloszlás, mégpedig a $PA_t(m, \delta)$ fokszámeloszlása.

1.4.3. Tétel (Fokszámsorozat a „preferential attachment” modellben). *Rögzítsük $m > 1$ -et és $\delta \geq -m$ -et. Ekkor létezik $C = C(\delta, m) > 0$ δ -tól és m -től függő konstans, amire teljesül, hogy ha $t \rightarrow \infty$, akkor $P\left(\max_k |P_k(t) - p_k| \geq C\sqrt{\frac{\log(t)}{t}}\right) = o(1)$.*

1.4.4. Megjegyzés. *Az $o(1)$ kifejezésnek az előbbi tételben annyi a jelentése, hogy, ha $t \rightarrow \infty$, akkor a megadott valószínűség tart a nullához.*

Az eddigiek alapján már tétel formában is kimondhatjuk a következő állítást.

1.4.5. Tétel. *A „preferential attachment” modell skálafüggetlen.*

A „preferential attachment” modell tehát skálafüggetlen, és fokszámeloszlása polinomiális, ezért is jellemzi jól a valós hálózatokat. Mint láttuk, ez nem minden modellre igaz, példaképp

megemlíthető az Erdős–Rényi modell, aminél a fokszámeloszlás a Poisson-eloszlással jellemezhető, melynek lecsengése exponenciálisnál is gyorsabb.

A polinomiális fokszámeloszlás leegyszerűsítve azt jelenti, hogy a kétszeres fokszám esetén, azaz, ha a k fokú csúcsok aránya helyett a $2k$ fokúak arányát nézzük, akkor a p_k ugyanannyiadrészére csökken. A k fokú csúcsok arányát körülbelül $ck^{-\gamma}$ jellemzi, láthatjuk, hogy ezt $2k$ -ra alkalmazva csak konstanssal, pontosabban $2^{-\gamma}$ -val változik az eredmény. Ezzel szemben az exponenciális eloszlásnál, ahol a k fokú csúcsok arányát λ^k jellemzi, ha a $2k$ fokú csúcsok arányát nézzük, akkor ez a k fokú csúcsok arányának gyöke lesz, mivel $\lambda^{2k} = (\lambda^k)^2$. Az Erdős–Rényi-modellt már ismert, hogy a Poisson-eloszlás jellemzi, tehát itt a fokszámeloszlás még gyorsabban cseng le, mivel itt $p_k = \frac{\lambda^k}{k!}e^{-\lambda}$, azaz még k faktoriálissal is leosztunk.

A valós hálózatok általában polinomiális fokszámeloszlásúak, ezért is alkalmazható jobban a gyakorlatban „preferential attachment” modell, mint az Erdős–Rényi-modell.

1.4.1. Barabási–Albert-modell

A modellt Barabási Albert-László és tanítványa, Albert Réka dolgozta ki 1999-ben. A Barabási–Albert-modell a „preferential attachment” modell egy olyan altípusa, ahol $\delta = 0$, és m , az egy lépésben a gráfhoz adott élek száma tetszőleges pozitív egész szám. Tehát itt a fokszámok a meghatározóak, és az új csúcs mindig m éllel kapcsolódik az eddigi hálózatunkhoz. Itt is teljesülnek a „preferential attachment” modell tulajdonságai: ez is dinamikus, növekedő modell, és érvényesül a „rich get richer” kifejezéssel is leírható sajátosság, azaz a nagyobb fokszámú csúcsokhoz nagyobb valószínűséggel fog kapcsolódni az új csúcs, mint a kisebb fokszámúakhoz. Az aszimptotikus fokszámeloszlás itt is polinomiálisan cseng le, negatív kitevőjű hatvány szerint.

1.4.2. Barabási–Albert-fa

Itt további megkötéseket teszünk a Barabási–Albert-modellhez képest. Az $m = 1$, azaz egyszerre csak egy él kapcsolódik, a $\delta = 0$ feltétel pedig továbbra is fennáll. Ezenkívül kizárjuk a hurok-élek lehetséges létezését a gráfban. A modell kialakulásakor kezdetben egyetlen csúcsunk van. A következő csúcsot ugye csak ehhez tudjuk kapcsolni egy éllel, ez az eredeti csúcs gyereke lesz. A harmadiknál már két lehetőségünk van: kapcsolhatjuk az első vagy a második csúcshoz, de mindig szigorúan csak egyhez, hiszen fáról van szó. Így minden csúcsnál annak a valószínűsége, hogy épp ahhoz kötjük az újat az adott csúcs fokszámának és az összes csúcs fokszámösszegének a hányadosa. A Barabási–Albert-fa fokszámeloszlása ugyanaz, mint a „preferential attachment” modell fokszámeloszlása, csak itt jelentősen egyszerűsödik a helyzet, hiszen felhasználhatjuk a $\delta = 0$ és $m = 1$ kikötéseket. Annak az esélye, hogy az $(n + 1)$ -edik csúcsot egy k fokú csúcshoz kötjük, $\frac{k}{2n-2}$. Hiszen azt, hogy az új csúcs hová fog kapcsolódni a fokszámokkal arányos való-

színűséggel választjuk ki. Ha $\delta = 0$ és $k \geq m$ vagy $m = 1$, ezek már szerepeltek a „preferential attachment” modellnél, most pedig ezt a kettőt felhasználva kapjuk, hogy ha $\delta = 0$ és $m = 1$, akkor

$$p_k = \frac{2\Gamma(k)\Gamma(3)}{\Gamma(k+3)\Gamma(1)} = \frac{4}{(k+2)(k+1)k}.$$

Itt a Γ tulajdonságait használjuk, pontosabban azt, hogy $\Gamma(k+1) = k\Gamma(k)$, azaz $\Gamma(k) = (k-1)!$. Vagyis a Barabási–Albert-fa fokszámeloszlása $\frac{4}{(k+2)(k+1)k} \approx \frac{4}{k^3}$ -höz tart, azaz ennyi a valószínűsége, hogy épp egy k fokszámú csúcsot választunk ki, ha az eloszlás egyenletes és ha $k \rightarrow \infty$.

Ebből a modellből alakult ki az általánosított Barabási–Albert-fa, ahol az $m = 1$ kikötés megmarad, de $\delta > -1$ értéke tetszőleges. Erre a modellre most nem térünk ki külön, de a felépítése természetesen hasonlít a Barabási–Albert-fához.

1.5. Egyéb gráfmodellek

Sok különböző gráfmodell van, melyeknek szintén érdekes és hasznos alkalmazásai vannak az élet számos területén. Példaképp megemlíteném a fitness modelleket, amik szintén dinamikus modellek, a „preferential attachment” modell továbbfejlesztései. Itt a gráfban minden csúcsonk adott egy sorszáma, amilyen sorrendben bevettük őket a gráfba. Minden új csúcs csatlakozásakor, még mielőtt az éllel kapcsolódna, az összes csúcsonk kiosztunk egy fitnesset, azaz egy véletlen súlyt. Az úgynevezett „véletlen” tetszőleges eloszlással megadható, ezt mi határozzuk meg, majd a sorsolás független azonos eloszlás szerint történik. Így annak a valószínűsége, hogy egy újabb csúcs majd hová csatlakozik, nem csak a fokszámoktól függ, mint a „preferential attachment” modellnél, hanem a véletlenszerűen kiosztott súlytól is. Ezáltal a gráfban még egy tényező fellép, ami az eddigi inhomogenitást tovább növelheti, vagy akár bizonyos mértékben ki is egyenlítheti azt. A modell egy fontos tulajdonsága, hogy ahogy egyre növekszik a gráf, a kisebb fokú csúcsok is naggyá változhatnak, és ezt a véletlen súlyok teszik lehetővé.

Most pedig térjünk rá az eredeti témánkra, azaz a járványterjedésre, amely modellezéséhez az eddig megismert véletlen gráftípusokra és hozzájuk kapcsolódó fogalmakra nagy szükségünk lesz.

2. fejezet

A járványterjedés modellezése

Az eddig megismert gráfmodellek közül az Erdős–Rényi-gráfon, és „preferential attachment” típusú gráfmodelleken vizsgáljuk a járványterjedést. Az Erdős–Rényi-gráfmodell esetén egy és két komponensből állót is vizsgálunk, és megadjuk, hogy az utóbbinál bizonyos körülmények között a vírus mikor tud átterjedni egyik komponensből a másikba. A gyakorlatban a két komponens értelmezhetjük például két városként. Majd megnézzük, hogy egy adott Erdős–Rényi-gráfban, azon feltételek mellett, hogy a gráf viszonylag nagy mértékben összefüggő, van-e olyan város, ami megállítja a fertőzést, azaz a határa kellőképp kicsi, olyan kevés él megy ki belőle, hogy innen nem terjed tovább a vírus.

Majd a „preferential attachment” modellen nézzük a járványok terjedését, hiszen ez jobban hasonlít az emberi hálózatokhoz, erről már esett szó az első fejezet során. Alsó és felső korlátokat adunk a járvány túlélési valószínűségére bizonyos feltételek mellett.

A fejezet végén az influenza modellezésére mutatunk néhány alapvető példát, és néhány valós adatot a vírus terjedéséről. Ezek során újabb és újabb szempontok merülhetnek fel, ami alapján érdemes a modelleket létrehozni. Nem mindegy a szempontok preferenciája, sem az, hogy hány szempontot veszünk figyelembe. Azt mindenesetre láthatjuk, hogy a valóságban a járványok terjedése jóval komplexebb és bonyolultabb, mintsem hogy tökéletes modellt készíthessünk. Ráadásul a vírusok változékonyak, a modelleken is mindig módosítani, javítani kell, mindezek ellenére sok, a gyakorlatban jól használható modellt ismerünk.

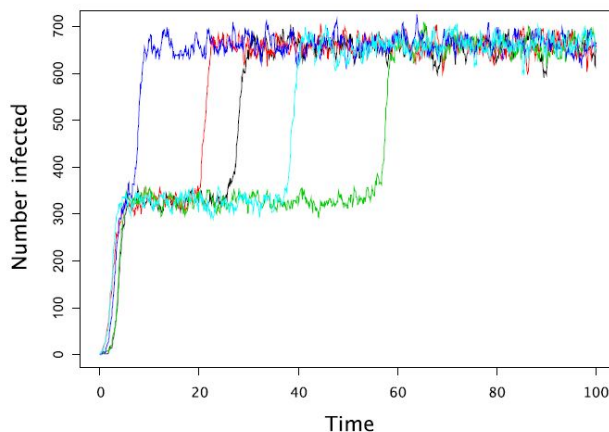
2.1. Modellezés az egy- és kétkomponensű Erdős–Rényi-gráfon

Ebben az alfejezetben a [8] cikk információit és adatait használjuk fel.

Legyen $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ különálló Erdős–Rényi-gráfok, mindkettőben n csúcs, és p valószínűséggel húzunk be élt két csúcs között, melyek ugyanabban a komponensben vannak. Az egész gráf $G = (V, E)$, ahol $V = V_1 \cup V_2$, és $E = E_1 \cup E_2 \cup B$, amelyben $B \subset V_1 \times V_2$ az

élek által alkotott hídát jelöli G_1 és G_2 között. Általában B az éleknek nagyon kis hányada, így látszik jól, hogy két külön, jobban összefüggő komponens van a gráfban, melyeket B élhalmaz mint híd köt össze.

Nézzünk egy ehhez a modellhez készített szimulációt. A következő ábrához tartozó modellben a két komponens 500 – 500 csúcsból áll, ahol az átlagos fokszám $np = 50$, azaz $p = 0,1$, és $|B| = 2$. n a csúcsok száma, p az élbehúzási valószínűség két csúcs között. Bevezetjük még a λ változót, mely azt jelöli, hogy egy fertőzött csúcs mekkora rátával fertőzi meg a szomszédos csúcsokat. A ráta jelentéséről hamarosan lesz szó, addig képzelhetjük azt, mintha a valószínűségről lenne szó. Ebben a példában $\lambda = 0,06$. Az adatokból kiszámolhatjuk, hogy $np\lambda = b = 3$ ebben az esetben. Az ábrán a különböző színű görbék egymástól független kísérleteket jelenítenek meg. Kiinduláskor az első komponensben adott két fertőzött csúcs, a többi csúcs egészséges. A fertőzött csúcsok 1 rátával gyógyulnak meg. Az összes szimulációban a vírus gyorsan elérte a látszólagos egyensúlyt a V_1 -beli csúcsokon, majd véletlen időpontban áttért a fertőzés a V_2 -beli csúcsokra, a második komponensbe. Felmerül bennünk a kérdés, hogy pontosan mikor kerül át a vírus a második komponensbe? Erre még visszatérünk a későbbiekben.



2.1. ábra. Két komponens közti vírusterjedés

[8]

A folyamat, ami a G gráfban végbemegy, egy Markov-folyamat, ahol $\xi_t \subset V$ azon csúcsok halmaza, akik t időpillanatban fertőzöttek.

2.1.1. Megjegyzés. *A Markov-folyamat egy olyan típusú véletlen folyamat, ahol a jövőbeli változás csak a pillanatnyi állapottól függ, a múlttól nem.*

Ha $\lambda = \lambda(n, p) > 0$ a vírus terjedésének sebessége, akkor egy fertőzött csúcs λ paraméterű Poisson-folyamattal adja tovább a fertőzést a szomszédjainak. A gyógyulás pedig az 1 paramé-

terű Poisson-folyamat szerint zajlik. Valójában egy egészséges csúcs, azaz $v \notin \xi_t$, $\lambda|N(v) \cap \xi_t|$ rátával fertőződhet meg, ahol $N(v)$ jelöli a v szomszédait G -ben.

A Poisson-folyamat modellezését a következőképpen képzelhetjük el. Mindenkinnek, azaz minden csúcsnak van egy órája, mely véletlen időközönként cseng. Miután csengett, akkor kisorsolunk egy új időtartamot, amennyi idő múlva újból cseng majd az óra, ami időtartamonként egymástól függetlenül, λ paraméterű exponenciális eloszlás szerint történik. Ha az időtartamok összességét vesszük, akkor egy λ rátájú Poisson-folyamatról van szó. Tehát a ráta a Poisson-folyamat paramétere, ez szerepelt az előbbi példában is.

Azaz, ha a folyamatban szereplő egymástól független időtartamok exponenciális eloszlásúak, λ paraméterrel, akkor ezek összesége a λ rátájú Poisson-folyamat. Ekkor, ha egy fertőzött csúcsot tekintünk, akkor ha ez nem gyógyul meg előbb, minthogy cseng az órája, akkor megfertőzi az összes hozzá közvetlen éllel csatlakozó csúcsot. Erre úgy is gondolhatunk, mintha két órája lenne: ha az egyik cseng, akkor fertőz, ha a másik, akkor meggyógyul. Természetesen fertőzni csak akkor fog, ha még nem gyógyult meg előtte.

Ha a λ ráta nagy, akkor az időtartamok várható értéke kicsi, hiszen ezek exponenciális eloszlásúak, és ismert, hogy a λ paraméterű exponenciális eloszlás várható értéke $\frac{1}{\lambda}$. Ekkor sok rövidebb időtartam van, tehát az óra gyakrabban cseng. Emiatt a λ -t másképp intenzitásnak is nevezhetjük. Egy csúcs megfertőződésének időpontja nem csak a λ nagyságától, hanem a szomszédainak számától is függ. Hiszen akkor kapja el a fertőzést, mikor a szomszédai közül az első fertőzni fog. Tudjuk, hogy mind exponenciális eloszlásúak, és exponenciálisak minimuma is exponenciális, ebből adódik a következő állítás: Ha X_1, X_2, \dots, X_n független λ paraméterű exponenciális eloszlásúak, akkor $\min(X_1, \dots, X_n) \sim \exp(n\lambda)$. Ez a magyarázata annak, hogy a fertőzés rátája a rátának és a fertőzött szomszédok számának szorzata.

Ahhoz, hogy a járvány elterjedését biztosítsuk, feltesszük, hogy $np = n^a$, ahol $a \in (0, 1]$. Ez egy nagyságrendi becslés, amivel könnyen lehet számolni. Az előbbit átrendezve azt kapjuk, hogy $p = n^{a-1}$ az élvalószínűség. Azért van szükség a viszonylag nagy élvalószínűségekre, vagyis nagy fokszámra, hogy a fertőzés továbbadása ne legyen túl ritka. Kisebb fokszámok esetén más plusz feltételekkel kell garantálni, hogy a vírus gyorsan el tudjon terjedni, és ne haljon ki. Ha az előző képletbe $a = 0$ -t helyettesítünk, akkor $p = \frac{1}{n}$, és az első fejezetből Erdős–Rényi-gráfmodellnél kiderül, hogy épp itt változik meg a gráf összefüggősége. A $p = \frac{1}{n}$ még nem elég, de $p = n^{a-1}$ pozitív a esetén már jó. Az $np = \frac{\log n}{n}$ is elég lenne, csak most nem ezt nézzük, mert könnyebb számolni n^a -val. Tehát ahhoz, hogy a gráf összefüggő legyen, az is elég, ha az átlagos fokszám legalább $np \geq c \log n$ ($c > 1$). A továbbiakban az egyszerűség kedvéért feltesszük, hogy $np = n^a$, és erre a speciális esetre vonatkozó eredményeket vizsgálunk majd.

A következő két tétel értelmezéséhez be kell vezetnünk néhány jelölést. Legyen $P[\cdot]$ a $G = G_1 \cup G_2 \cup B$ független Erdős–Rényi-féle véletlen gráf kiválasztásának valószínűsége, ahol $G_1, G_2 \sim$

$G(n,p)$ különálló véletlen gráfok, és B ezektől független élhalmaz, ami csak olyan éleket tartalmaz, amelyeknek egyik végpontja G_1 -ben, a másik végpontja G_2 -ben van. Miután véletlenszerűen választottunk egy gráfot, ezt lerögzítjük, ezután is marad a véletlennek szerepe a gráfban, pontosabban a járvány terjedésében. Ezt több dolog határozza meg: az egyik, hogy kezdetben melyik csúcsok voltak megfertőzve. P_v jelöli azt, ha először csak a v csúcs volt fertőzött. Természetesen a v csúcs helyett a csúcsok bármely részhalmaza szerepelhet. A másik befolyásoló tényező a vírus terjedésének sebessége, mely szintén véletlen. Ez határozza meg, hogy a járvány mennyi idő után hal ki.

Azt is érdemes megvizsgálni, hogyan terjed a vírus, ha csak az egyik komponensen nézzük a terjedést, azaz ha a B élekről egy időre megfeledkezünk. Erről fog szólni az első tétel. Legyen $\xi_t = \xi_t \cap V_1 \quad \forall t$ -re, ahol ξ_t jelöli a fertőzött csúcsok halmazát t időpillanatban. Legyen $\xi_0 = A \subset V_1$, a kezdetben fertőzött csúcsok halmaza. Jelöljük ennek az eloszlását $P_A^{G_1}$ -el.

Ezenkívül használni fogjuk még az $a_n = O(b_n)$ jelölést, ami azt jelenti, hogy $a_n \leq Cb_n$ valamely $C > 0$ -ra, $a_n = o(b_n)$ -t, ami akkor teljesül, ha $\frac{a_n}{b_n} \rightarrow 0$, és $a_n = \Omega(b_n)$ jelölést, miszerint $a_n \geq cb_n$ valamely $c > 0$ -ra.

Valahogy biztosítanunk kell, hogy a vírus ne haljon ki, mielőtt átterjed az első komponensből a másodikba. Ha $np\lambda = b < 1$, akkor egy adott csúcs várhatóan egynél kevesebb másikat fertőz meg, ezért gyorsan kihal a fertőzés. Ezt nevezik szubkritikus elágazó folyamatnak. (Ez olyan, mintha minden embernek egynél kevesebb gyereke lenne, így előbb-utóbb az emberek kihálnának.) Ha $np\lambda = b = 1$, ekkor szintén egy valószínűséggel kihal a vírus, csak ebben az esetben ezt nehezebb megmutatni, bonyolultabb számolásokra van szükség. Ha pedig $np\lambda = b > 1$, ez a szuperkritikus elágazó folyamat, ekkor tud elterjedni a vírus.

Tehát feltehetjük, hogy $np\lambda = b > 1$, mivel csak ebben az esetben érdemes vizsgálni, hogy mikor terjed át a vírus a második komponensre. Ha ez a feltétel nem teljesül, akkor a vírus már az első komponensben kihal, mielőtt átterjedhetne a másodikra is. Ebben az esetben teljesül a következő tétel, ami azt mondja ki, hogy szuperkritikus elágazó folyamat esetén a vírus exponenciálisan hosszú ideig él, ha fennmaradt $\Omega(\log \log n)$ ideig. Azt pedig feltehetjük, hogy életben marad legalább $\Omega(\log \log n)$ ideig, mivel ennyi idő alatt a vírus vagy kihal körülbelül $\frac{1}{b}$ valószínűséggel, vagy megfertőz nagyságrendben $\log n$ csúcsot.

Tehát ahhoz, hogy a kétkomponensű modellen nézhessük a terjedést, előbb azt az esetet vizsgáljuk, mikor csak egy komponens van. Ez segít majd meghatározni, hogy mi szükséges ahhoz, hogy két komponens esetében a másodikra is átterjedjen a fertőzés.

2.1.2. Tétel. *Adott egy Erdős–Rényi-féle véletlen gráfunk, $G_1 \sim G(n,p)$, ahol $np = n^a$, $np\lambda = b > 1$. Jelöljük ξ_t -vel azon csúcsok halmazát, akik t időpillanatban fertőzött állapotban vannak. Ez*

a kezdeti állapotban jelenleg csak egyetlen csúcs, azaz $\xi_0 = \{v\} \subset V_1$. Legyen $r = \frac{2}{b-1} \log \log n$. Ekkor léteznek $\eta_3, \varepsilon, c > 0$ b -től függő konstansok, hogy $\forall \delta > 0$ -ra

$$P \left[P_v^{G_1} \left(\min_{t \in [\eta_3 \log n, e^{cn}]} |\xi_t| \leq \varepsilon n \mid |\xi_r| > 0 \right) > \delta \right] \rightarrow 0, \quad (2.1)$$

ha $n \rightarrow \infty$.

Fogalmazzuk meg, mit is mond ki az előbbi tétel. Ha az r időpillanatig nem halt ki a fertőzés, akkor a G_1 gráfban a v pontból elterjedő vírusnál annak a valószínűsége, hogy valamely $t \in [\eta_3 \log n, e^{cn}]$ időpillanatban viszonylag kevés fertőzött csúcs van, maximum a populáció ε -szorososa, és ez nagyobb mint δ , ennek a valószínűsége tart a nullához. Tehát az, hogy a megadott hosszú időintervallumban van egy olyan időpont, mikor kevés a fertőzött csúcs, ennek a valószínűsége nagyobb, mint δ , feltéve, hogy az r -edik időpillanatig nem hal ki a járvány. Ha ennek a valószínűségét nézzük, az tart nullához.

Az előbbi tételben a konstansok választására a legmegfelelőbb értékek $\eta_3 = \frac{4}{(1-4\varepsilon)b-1}$, és $\varepsilon = \frac{1}{4} \left(1 - \frac{1}{b^{\frac{1}{3}}}\right)$. Arra, hogy ezeket miért érdemes így választani, most nem térünk ki részletesen, de ezen értékek mellett elég pontos becslést kapunk. A becslés ugyanis nem minden η_3, ε -ra igaz, de ezek a legkisebb értékek, amire bebizonyítható, hogy a becslés teljesül.

Visszatérve a kétkomponensű modellhez, az egyik fontos kérdés, hogy ha a járvány egyetlen $v_0 \in V_1$ csúcsból kezd el terjedni, akkor mennyi idő szükséges ahhoz, hogy a V_2 -beli csúcsok egy pozitív hányada is fertőzött legyen. Másképp fogalmazva, mikor jut át a járvány a B élek által alkotott hídon, mikor kezd el terjedni a második komponensben is? Erre a kérdésre ad választ a következő tétel.

2.1.3. Tétel. Rögzítsünk $a \in (0, 1]$ -et. Tegyük fel, hogy $np = n^a$, $np\lambda = b > 1$, és legyen $|B| = o\left(\frac{n^a}{\log n \log \log n}\right)$ az átkötő élek száma V_1 és V_2 között. Válasszunk egy tetszőleges $v_0 \in V_1$ -et. Ekkor $\exists \varepsilon > 0$, hogy ha $\tau := \inf\{t > 0 : |\xi_t \cap V_2| > \varepsilon n\}$, akkor $\forall x \in [0, \infty)$ és $\forall \delta > 0$ -ra

$$P \left[\left| P_{v_0} \left(\frac{\tau}{|B|} \leq x \right) - \left(1 - \frac{1}{b}\right) \left[1 - \exp\left(-\frac{(b-1)^2}{b}x\right)\right] \right| > \delta \right] \rightarrow 0, \quad (2.2)$$

ha $n \rightarrow \infty$.

2.1.4. Megjegyzés. A külső $P[\cdot]$ a véletlen gráf megválasztásának valószínűségét jelzi. A belső P_{v_0} valószínűségnél nem írjuk ki a G_1, G_2 gráfokat, de ennek ellenére ezek itt már rögzítettek. A tétel maga pedig a τ eloszlásfüggvényéről szól.

Bizonyítás. Mivel a precíz bizonyítás túl hosszú lenne, csak vázlatos bizonyítást adunk az előző tételre. Az egyszerűség kedvéért tegyük fel, hogy $B = \{(u_b, v_b)\}$ egyetlen élből álló halmaz,

ahol $u_b \in V_1, v_b \in V_2$. Először is, amikor a fertőzés egyetlen $\xi_0 = \{v_0\} \subset V_1$ csúcsból indul ki, viszonylag kicsi $r = O(\log \log n)$ idő alatt $|\xi_t|$ jól közelíthető egy adott folyamattal, amelynél a túlélési valószínűség $(1 - \frac{1}{b})$, ahol $b = np\lambda$. Ekkor r idő alatt a fertőzés vagy kihal, vagy exponenciálisan hosszú ideig túlél pozitív sűrűséggel, mely már szerepelt a 2.1.2. tételben. Ezt az akadályt úgy győzhetjük le, hogy az izoperimetrikus egyenlőtlenséget alkalmazzuk, melyről a 2.1.6. tételben lesz szó.

Második lépésben a fertőzés duális folyamatát alkalmazzuk, és azt, hogy ez ugyanolyan eloszlású, mint az eredeti fertőzés. A duális folyamatnál azt figyeljük meg, hogy kit ki fertőzött meg. Tehát mintha visszafelé haladna az idő, a már fertőzöttektől haladunk arra, ahonnan a vírus elkezdett terjedni. A duális folyamat $t > 0$ időpontban kezdődik, és $\{\zeta_s^t\}_{s \in [0, t]}$ -vel jelöljük. Tehát, ha a folyamat egy $\zeta_0^t = v \subset V$ csúcsból indul ki, akkor $\zeta_s^t \cap \xi_{t-s} \neq \emptyset$ valamely $s \in (0, t)$ -re akkor és csak akkor, ha $v \in \xi_t$. Az eredeti folyamatban λ rátával terjed a fertőzés az éleken keresztül, és ez a λ a ξ_t -től független. Ezért azt figyeljük meg, mikor a duális folyamat u_b -ből indul, azaz először csak ezt a részt nézzük. A duális folyamatban a fertőzött csúcsok halmaza az r időpillanatra eléri az $\Omega(\log n)$ méretet körülbelül $1 - \frac{1}{b}$ valószínűséggel. Itt a nehézség valójában a duális folyamat modellezésében rejlik, mert a modell az időben visszafelé haladva növekszik, és a csúcsok sokkal messzebb lehetnek egymástól, mint az eredeti folyamatban. Végül, mikor a fertőzés átterjed az (u_b, v_b) élen, akkor egy új, független duális folyamatot indítunk. Minden duális folyamatnak, hogy túléljen r ideig, kereszteznie kell ξ_t -vel. Ennek eredményeképp a vírus átterjed v_b csúcsra. Innen pedig elterjed a fertőzés V_2 -ben körülbelül $1 - \frac{1}{b}$ valószínűséggel, ez már a bizonyításvázlat elején is előkerült. Szóval abban az esetben, ha $\xi_r \neq \emptyset$, akkor a szükséges idő, hogy a fertőzés átterjedjen V_2 -be a $\lambda(1 - \frac{1}{b})^2$ paraméterű exponenciális eloszlás szerint zajlik, és annak a valószínűsége, hogy $\xi_r = \emptyset$, $1 - \frac{1}{b}$ -vel egyenlő. Éppen ezt állította a 2.1.3. tétel. \square

A tételben τ az első olyan időpontot jelzi, mikor a második komponensben az emberek egy pozitív hányada, azaz jelenleg az emberek ε -szorososa beteg. Ha a vírus elég sokáig fennmarad (ami $\approx (1 - \frac{1}{b})$ valószínűséggel következik be), akkor τ eloszlása megközelítőleg exponenciális $|B| \frac{(b-1)^2}{bn^a}$ paraméterrel. A felső határa az élszámnak a két komponens között $|B|\lambda = \frac{b|B|}{n^a}$. Ezt a megkötést az indokolja, hogy legyen értelme a két komponensnek, ne rögtön az egész gráfban terjedjen el a vírus, hanem a második komponensben csak bizonyos idő után. Ha a hídon túl nagy valószínűséggel húzunk be éleket, akkor a két komponens nem különül el, szinte olyan, mintha egy homogén Erdős–Rényi-gráfmodellt vizsgálnánk. A maximális megengedett $|B|$ élszám, ami mellett még teljesül a két komponens elkülönülése, $O(n^a)$. Kis t esetén a fertőzött csúcsok száma V_1 -ben t időpillanatban körülbelül b^t , és minden vírusos csúcs fokszáma körülbelül np , és a fertőzött csúcsok minden szomszédjukat λ valószínűséggel fertőzik meg. Tehát a fertőzött csúcsok száma eléri az εn -t $s = \frac{\log(\varepsilon n)}{\log b}$ időben. Előfordulhat, hogy a vírus nem csak egyszer, hanem többször is átterjed az egyik komponensből a másikba. Nézzük meg, hogy mennyi az átterjedések

számának várható értéke. Ehhez felhasználjuk a következő összefüggést: $\frac{b^t \lambda |B|}{n} = \frac{b^{t+1} |B|}{n^{1+a}}$. Tehát az áttérjedések számának várható értéke a következő:

$$\int_0^s \frac{b^{t+1} |B|}{n^{1+a}} dt = \frac{b^{s+1} |B| \log b}{n^{1+a}} \approx \frac{|B| \varepsilon b \log b}{n^a}.$$

Hogyha $|B| \gg n^a$, azaz B nagyságrendje n^a nagyságrendjénél nagyobb, azaz ezzel osztva végtelenhez tart, akkor a járvány gyorsabban áttérjed V_2 -be, minthogy elérné V_1 -ben az εn méretet. A 2.1.3. tétel alapján alapján hasonlóan felírhatjuk a problémát m komponensre is.

A most következő részben az izoperimetrikus egyenlőtlenségről lesz szó, mely lényegét a következő példa szemlélteti: vegyünk egy tetszőleges országot, és nézzük ezen belül a járványterjedést. Azt vizsgáljuk, hogy van-e olyan város, melynek a határa önmagához viszonyítva olyan kicsi, hogy ez meggátolja a járvány elterjedését. Azaz, attól, hogy ebben a városban kitör a járvány, ez nem fog innen továbbterjedni, mert annyira kicsi az innen kivezető élek száma. Természetesen a központi városokat nem érdemes nézni, csak a kisebb városokat, ahonnan nem vezet ki sok él, azaz ahol kicsi a forgalom.

Ezenkívül bebizonyítunk egy tételt, ami korlátot ad az ε -izoperimetrikus számra egy Erdős–Rényi véletlen gráf segítségével, aminek nagy az átlagos fokszáma.

2.1.5. Definíció. *Legyen adott egy $G = (V, E)$ gráf. Ekkor a G gráfhoz tartozó ε -izoperimetrikus számot jelöljük $i_\varepsilon(G)$ -vel, és a következőképpen definiáljuk:*

$$i_\varepsilon(G) = \min \left\{ \frac{|\partial U|}{|U|} \mid U \subset V, |U| \leq \varepsilon |V| \right\},$$

ahol $\partial U \subset E$ élek egy olyan halmaza, melynek pontosan az egyik végpontja az U csúcshalmazban van.

2.1.6. Tétel. *Ha $G \sim G(n, p)$ egy Erdős–Rényi véletlen gráf, ahol a csúcsok száma n , két csúcs közt az él behúzási valószínűsége p , és $np \geq 28(\log n)^3$, akkor $\forall \varepsilon > 0$ -ra*

$$i_\varepsilon(G) \geq (1 - \varepsilon)np - (np)^{\frac{2}{3}} \quad (2.3)$$

$1 - o(1)$ valószínűséggel, ha $n \rightarrow \infty$.

Érdemes megemlíteni, hogy ez a korlát lényegében nagyon éles, kevéssé tér el attól, hogy ha U εn -edrészre a csúcsoknak, akkor $|\partial U| \sim \text{Binomiális}(|U|(1 - \varepsilon)n, p)$, ezért nagy valószínűséggel $|\partial U| = |U|(1 - \varepsilon)np(1 + o(1))$.

A viszonylag nagy foksámra pedig azért van szükség, hogy ne legyenek elszigetelt városok, mivel

ha kevés él lenne, akkor nyilván előfordulna ilyen, és egy elszigetelt városból nem tud elterjedni a vírus.

A tétel bizonyításához felhasználjuk a következő két lemmát.

2.1.7. Lemma. *Markov-egyenlőtlenség: Ha $X \geq 0$ valószínűségi változó, és létezik az $E(X) > 0$ várható érték, akkor $\forall a > 0$ valós szám esetén $P(X \geq a) \leq \frac{E(X)}{a}$.*

2.1.8. Lemma. *Bernstein-egyenlőtlenség: Legyenek az X_1, X_2, \dots valószínűségi változók függetlenek, azonos eloszlásúak, és tegyük fel, hogy létezik a várható értékük. Ha $x \geq E(X_1)$, akkor*

$$P\left(\frac{X_1 + X_2 + \dots + X_n}{n} \geq x\right) \leq \Psi^n(x),$$

míg, ha $x \leq E(X_1)$, akkor

$$P\left(\frac{X_1 + X_2 + \dots + X_n}{n} \leq x\right) \leq \Psi^n(x),$$

ahol $\Psi(x) = \inf_t e^{-tx} R(t)$.

Az előbbi $R(t) = E(e^{tX_1})$ függvényt az X_1 valószínűségi változó momentum generáló függvényének nevezzük. Akkor mondjuk, hogy $R(t)$ létezik, ha az origó az $R(t)$ értelmezési tartományának belső pontja.

Bizonyítás. Jelölje $P(u, m)$ annak a valószínűségét, hogy a G gráf tartalmazza a csúcsok egy $U \subset V$ halmazát, ahol $|U| = u$, és $|\partial U| \leq m$. Ekkor csak azt kell belátnunk, hogy

$$\sum_{u=1}^{\varepsilon n} P(u, m(u)) = o(1),$$

ahol $m(u) := u[(1 - \varepsilon)np - (np)^{\frac{2}{3}}]$, és ezzel bebizonyítottuk a tételt. A Markov-egyenlőtlenséget felhasználva kapjuk, hogy

$$P(u, m(u)) \leq \binom{n}{u} \sum_{s=0}^m m(u) \binom{u(n-u)}{s} p^s (1-p)^{u(n-u)-s} = \left(\binom{n}{u} P(X_u \leq m(u)) \right),$$

ahol $X_u \sim \text{Binomiális}(u(n-u), p)$. Ezután a Bernstein-egyenlőtlenséget alkalmazzuk. Ennek eredményeképp azt kapjuk, hogy

$$P(|X_u - E(X_u)| \geq t) \leq 2 \exp \left\{ \frac{-t^2}{2(D(X_u) + \frac{t}{3})} \right\}.$$

Legyen $t = E(X_u) - m(u) = u(\varepsilon np - up + (np)^{\frac{2}{3}})$. Megjegyezzük, hogy, ha $t \geq u(np)^{\frac{2}{3}}$, ahol

$u \leq \varepsilon n$, ott

$$\begin{aligned}
P(X_u \leq m(u)) &\leq P(|X_u - E(X_u)| \geq u(np)^{\frac{2}{3}}) \\
&\leq 2 \exp \left\{ \frac{-u^2(np)^{\frac{4}{3}}}{2[u(n-u)p(1-p) + \frac{1}{3}u(np)^{\frac{2}{3}}]} \right\} \\
&\leq 2 \exp \left\{ \frac{-u(np)^{\frac{2}{3}}}{2(np)^{\frac{1}{3}} + \frac{2}{3}} \right\} \\
&\leq 2 \exp \left\{ -\frac{1}{3}u(np)^{\frac{1}{3}} \right\},
\end{aligned}$$

feltéve, hogy $np > \frac{1}{3}$. Ezután felhasználjuk azt a feltételt, miszerint $np \geq 28(\log n)^3$.

$$\sum_{u=1}^{\varepsilon n} P(u, m(u)) \leq 2 \sum_{u=1}^{\infty} [ne^{-\frac{(np)^{\frac{1}{3}}}{3}}]^u = 2ne^{-\frac{(np)^{\frac{1}{3}}}{3}} (1 - ne^{-\frac{(np)^{\frac{1}{3}}}{3}})^{-1} = o(1).$$

Ezzel az eredeti állítást beláttuk. \square

2.2. Járványterjedés a „preferential attachment” gráfmodellen

Ebben az alfejezetben főként a [4] cikk tartalmát használjuk fel. Mint azt láthattuk az első fejezetben, a „preferential attachment” modell sokban hasonlít a valós hálózatokhoz. Így a valódi fertőzések terjedését az emberek közt, illetve az internetes vírusokét is jobban modellezi ez a gráf az Erdős–Rényi-modellnél. Belátjuk, hogy bármely vírus, ami pozitív valószínűséggel terjed át a szomszédaira, nem elhanyagolható eséllyel alakulhat járvánnyá. Ezenkívül megmutatjuk, hogy a fertőzés terjedésének folyamata erősen függ a gráfban előforduló fokszámoktól. Megmutatjuk, hogy, ha a gráfban minden csúcs fokszáma jelentősen kisebb, mint $\frac{1}{\lambda}$, akkor a járvány gyorsan kihal. Ha pedig a vírus elér egy olyan csúcsot, melynek fokszáma jelentősen nagyobb, mint $\frac{1}{\lambda^2}$, akkor a járvány nagy valószínűséggel hosszú ideig túlél ennek a csúcsnak a szomszédai között. Itt a λ a fertőzés terjedésének rátáját jelöli. Ebben a részben feltesszük, hogy $\lambda \in [0, 1]$.

Az általános modell, amit a vírusos fertőzések vizsgálatára használunk, a „susceptible - infected - susceptible”, azaz a SIS modell. Itt minden csúcs vagy fertőzött, vagy egészséges, de a fertőzésre fogékony. Egy fertőzött csúcs 1 rátával egészségesé válhat, függetlenül a szomszédai állapotától. Egy egészséges csúcs pedig $c\lambda$ rátával fertőződik meg, ahol c az adott csúcs fertőzött szomszédainak számát jelöli. Tehát ez ugyanúgy zajlik, mint az Erdős–Rényi-modellnél a 2.1. szakaszban: szintén Poisson-folyamatról van szó, melyet a ráta jellemez. Itt a λ megegyezik a fertőzés terjedésének rátájával.

Ez a modell természetesen nem minden körülményt vesz figyelembe. Például, ha a körülöttünk élők közül mindenki beteg, akkor mi is nehezebben gyógyulunk meg a valóságban, esetleg elkapjuk az adott vírus egy mutálódott változatát, vagy egyéb vírusfertőzés is terjedhet ugyanakkor,

más valószínűséggel, és a kettő találkozásakor kialakulhatnak bizonyos szövődmények. Láthatjuk, hogy a valóság modellezése nem olyan könnyű, de néhány körülményt kiemelve, jól használható modelleket alkothatunk, melyek a gyakorlatban hasznunkra válnak.

Az egyik legfontosabb eredmény ebben a témakörben a járványterjedési kritikus érték létezése, melyre a továbbiakban küszöbérték néven hivatkozunk. Végtelen gráfok esetén két küszöbérték, λ_1 és λ_2 létezik a járványterjedésre, és teljesül rájuk, hogy $\lambda_1 \leq \lambda_2$. Ha $\lambda > \lambda_2$, akkor pozitív valószínűséggel elterjedhet a vírus a gráf bármely pontjára. A $\lambda_1 < \lambda < \lambda_2$ esetben a vírus pozitív valószínűséggel túlél, de végül szinte biztos, hogy minden csúcs meggyógyul. Ha $\lambda < \lambda_1$, akkor a vírus nagy eséllyel kihál. Ezenkívül ismert más alaptípusú gráfokra is hasonló témájú eredmény. Például, hogy a Z_d , azaz d dimenziós kockarács esetén $\lambda_1 = \lambda_2$, a reguláris fáknál pedig, azaz az olyan fáknál, ahol minden csúcs fokszáma megegyezik, $\lambda_1 < \lambda_2$.

Véges gráfoknál a fertőzés 1 valószínűséggel kihál. Mivel, ha a csúcsok pozitív valószínűséggel gyógyulnak, akkor kis gyógyulási valószínűség esetén is lesz viszonylag sok idő múlva egy valószínűséggel olyan rövid időtartam, amikor minden csúcs egyszerre meggyógyul. Tehát a csúcsok pozitív valószínűséggel gyógyulnak meg egy adott időtartam alatt.

Egy természetes definíció a járványokra véges esetben, ha egy jól definiált végtelen gráf egy véges részhalmazát nézzük, például Z^d -t. Tehát az $[-n, n]^d$ d dimenziós kockára van egy λ_c pozitív szám, amire ha $\lambda > \lambda_c$, akkor a valószínűség ugrásszerűen eltávolodik a nullától, a vírus túlélési ideje exponenciális lesz n^d -ben. Ha pedig $\lambda < \lambda_c$, akkor a fertőzés $\log n$ idő alatt kihál ($1 - o(n)$) valószínűséggel.

Ráadásul ez a λ_c szám megegyezik a küszöbértékkel Z^d esetben. Innen adódik a következő állítás, miszerint a vírus járvánnyá alakul, ha a kihalásához szükséges idő polinomiálisnál gyorsabban változik a gráf csúcscsámának függvényében.

2.2.1. Megjegyzés. *A polinomiálisnál gyorsabb növekedés alatt azt értjük, hogy, ha n gráf csúcscsáma, $f(n)$ a kihaláshoz szükséges idő n csúcs esetén, akkor $\frac{f(n)}{n^k} \rightarrow \infty \quad \forall k$ -ra.*

2.2.2. Megjegyzés. *A túlélés alatt azt értjük, hogy tetszőlegesen sok idő után is van fertőzött csúcs. A járvány kihalása alatt pedig azt, hogy a fertőzött csúcsok halmaza üres, és üres is marad.*

Vajon mennyi az értéke a λ_c küszöbértéknek skálafüggetlen gráfok esetén? Ebben a témában korábban is születtek cikkek, melyekben az szerepelt, hogy a λ_c küszöbszám a skálafüggetlen gráfokban 0. Ezt aktuális adatokkal és kutatásokkal alá is támasztották, de nem volt ismert a precíz bizonyítás. A jelenleg használt cikk [4] az, ahol az első szigorú bizonyítás jelent meg erről skálafüggetlen gráfokra, konkrétan a „preferential attachment” modellre. A bizonyításnál a „preferential attachment” modell egy másfajta ábrázolása, a Pólya-urnamodell alkalmazható. Az alap gondolat a következő: legyen adott egy rögzített csúcs, ez legyen például piros, a többi

pedig kék. Megnézzük, hogy az új csúcs hová fog kapcsolódni. Ha a piros csúcsához, akkor ennek a fokszáma nő eggyel, ha valamelyik kékhez, akkor a kék csúcsok össz-fokszáma nő eggyel.

A küszöbérték, vagy másnéven kritikus érték, valóban 0 a skálafüggetlen gráfokban. Ez azért olyan fontos a járványterjedés szempontjából, mert azt jelenti, hogy azok a vírusok, melyek lassan szaporodnak, szintén pozitív eséllyel válhatnak járvánnyá. A továbbiakban nem csak erről lesz szó, hogy nulla vagy pozitív ez a valószínűség, hanem egy alsó- és felső korlátot is adunk erre a valószínűsége, a λ függvényében. Ezt két különböző esetben vizsgáljuk: mikor a fertőzés egy előre kijelölt konkrét csúcsból kezd terjedni, vagy pedig egy teljesen véletlenszerűen választott csúcsból. Kiderül, hogy a két lehetőség nagyban különbözik egymástól, ezekről lesz szó a következő két tételben.

A skálafüggetlen gráf modelljének hatására alakult ki a Barabási–Albert-modell, melyről már volt szó a az 1.4.1. részben. Elevevítsük fel ezt a gráfmodellt újból. Rögzítsünk egy $m \geq 2$ egész számot, és egy $0 \leq \alpha < 1$ valós számot. Jelölje $\{v_i\}$ a csúcsok halmazát, és G_i a gráfot az i . időpillanatban. G_1 csak a v_1 csúcsot tartalmazza, G_2 a v_1, v_2 csúcsokat, és m darab öket összekötő élt. Ebből kiindulva a következő rekurzió szerint építjük fel a gráfot: ha adott G_{n-1} , akkor G_n -t úgy kaphatjuk meg, hogy hozzáveszünk egy v_n csúcsot, és kiválasztunk m csúcsot G_{n-1} -ből, melyek közt egy csúcs akár többször is előfordulhat. A kiválasztott m csúcsot jelöljük w_1, \dots, w_m -mel. A w_1 csúcsot α valószínűséggel egyenletesen, $1 - \alpha$ valószínűséggel pedig a „preferential attachment” modell szabályai szerint választjuk, azaz ekkor $w_1 = v_i$ ($i = 1, \dots, n - 1$),

$$\frac{\deg_{n-1}(v_i)}{\sum_{i=1}^{n-1} (\deg_{n-1}(v_i))} = \frac{\deg_{n-1}(v_i)}{2m(n-2)}$$

valószínűséggel. A képletben $\deg_{n-1}(v_i)$ a v_i csúcs fokszámát jelöli az $(n - 1)$. időpillanatban. Majd ugyanígy folytatjuk, csak w_k meghatározásakor $\deg_{n-1}(v_i)$ helyett $\deg'_{n-1}(v_i) = \deg_{n-1}(v_i) + |\{1 \leq j \leq k - 1 | w_j = v_i\}|$ -t használjuk. Ezután húzzuk be az új éleket oly módon, hogy v_n -t összekötjük w_1, \dots, w_m -nel. Amiatt, hogy megengedtük egy csúcs többszöri beválasztását, a gráfban többszörös élek is előfordulhatnak.

Ha $\alpha = 0$, akkor a „preferential attachment” modelltől ez mindössze annyiban tér el, hogy ott megengedettek voltak a hurokélek, míg itt nem. Ezzel ellentétben, a többszörös élek mindkét modellben engedélyezettek. Bonyolultabb eset, mikor többszörös éleket sem engedünk meg, ezzel most nem foglalkozunk. Most azt az esetet vizsgáljuk, ahol többszörös éleket megengedünk, de hurokéleket nem.

2.2.3. Tétel. $\forall \lambda > 0$ -ra létezik N , hogy ha a gráf mérete $n > N$, azaz a gráf elég nagy, akkor teljesül a következő állítás. A gráf minden csúcsa közül azonos valószínűséggel választunk egy v csúcsot, ez lesz a vírus kiindulópontja. Jelöljük $Q(v)$ -vel, hogy mennyi annak a valószínűsége,

hogy v -ből indítva túlél a járvány. Ekkor $1 - O(\lambda^2)$ valószínűséggel v egy olyan csúcs, melyre

$$\lambda^{C_1 \frac{\log \frac{1}{\lambda}}{\log \log \frac{1}{\lambda}}} \leq Q(v) \leq \lambda^{C_2 \frac{\log \frac{1}{\lambda}}{\log \log \frac{1}{\lambda}}},$$

ahol C_1, C_2 konstansok, melyek nem függenek λ -tól, sem n -től. Ezzel a túlélési valószínűségre adtunk egy alsó és felső korlátot is.

2.2.4. Megjegyzés. Továbbra se feledkezzünk meg arról a feltételről, hogy $\lambda \in [0, 1]$, erre a tételben is szükség van.

A maradék $O(\lambda^2 n)$ csúcs, amit kihagytunk az előző tételnél, erősen befolyásolja az átlagos túlélési valószínűséget, ezt mutatja be a következő tétel. Ez a kimaradt csúcshalmaz olyan erős befolyással bír, hogy az összes csúcsot tekintve, ez úgy fog viselkedni, mint az a kevés csúcs, amit most hozzávettünk.

2.2.5. Megjegyzés. Azért $O(\lambda^2 n)$, mivel n db csúcs van összesen, és az egyenletes eloszlás miatt mindnek $\frac{1}{n}$ a valószínűsége.

Tehát most az összes csúcsot nézzük, azokat is, amiket az előző tételben figyelmen kívül hagytunk.

2.2.6. Tétel. $\forall \lambda > 0$ -ra létezik N , amire igaz, hogy ha a gráf mérete $n > N$, ekkor teljesül a tételünk állítása. A gráf minden csúcsa közül azonos valószínűséggel, egyenletes eloszlás alapján választunk egy v csúcsot, ebből a csúcsból kezd el terjedni a fertőzés. Jelöljük itt is $Q(v)$ -vel, hogy mennyi annak a valószínűsége, hogy v -ből indítva túlél a járvány. Ekkor a túlélési valószínűség alsó és felső korlátja a következő: $\lambda^{C_3} \leq \frac{1}{n} \sum_{i=1}^n Q(v) \leq \lambda^{C_4}$, ahol C_3, C_4 abszolút konstansok, azaz nem függenek λ -tól, sem n -től.

Ebben a tételben az összes csúcs közül választottunk egyenletes valószínűséggel, azaz a túlélési valószínűség $\frac{1}{n} \sum_{i=1}^n Q(v)$.

Míg a 2.2.3. tételben $n(1 - O(\lambda^2))$ csúcsból választottunk, és így egy más nagyságrendű becslést kaptunk, mint a mostani esetben. Ezt úgy képzelhetjük el, hogy ha a gráfban található viszonylag kevés nagy fokú csúcsot elhagyjuk, akkor a járvány nem terjed olyan mértékben, mintha az elhagyott nagy fokszámmal rendelkező csúcsok is szereplnének. Mivel a gráfban a csúcsok nagy része olyan, hogy onnan indítva a járvány lassan terjed.

A következő definíció pontosítja a gyógyulás, illetve a fertőzés valószínűségét, illetve ezek idejét. Ebben a modellben egy csúcs kétféle állapotban lehet: egészséges vagy fertőzött. Emiatt elég csak a gyógyulást és a fertőzést vizsgálnunk, mivel ez a két folyamat mehet végbe a gráfban. Egy fertőzött csúcs exponenciális idő alatt egészségesé válik, függetlenül a szomszédai állapotától. Egy egészséges csúcs pedig fertőzötté válik, a fertőzött szomszédainak számának arányos rátával.

2.2.7. Definíció. Egy $G(V, E)$ gráfon adott egy λ paraméterű járványterjedés, ami egy folytonos idejű η_t Markov-folyamat. Ez bármely t időpillanatban meghatározható egy $A = \{v \in V : \eta_t(v) = 1\}$ csúcshalmaz által. Ebben az A halmazban vannak a fertőzött csúcsok, a többi csúcs egészséges. A terjedés rátája pedig a következő η_t -re:

$$A \rightarrow A \setminus \{v\}, \quad 1 \text{ rátával, ha } v \in A \text{ és}$$

$$A \rightarrow A \cup \{v\}, \quad \lambda |\{u \in A : \{u, v\} \in E\}| \text{ rátával, ha } v \notin A.$$

Azaz a definíció szerint egy adott csúcs egy rátával gyógyul meg, a fertőzés rátája pedig a szomszédai számának λ -szorososa.

Feltesszük, hogy a $t = 0$ időpontban a gráf egyetlen csúcsa fertőzött. Egy végtelen gráfban a fertőzés akár végtelen ideig is fennmaradhat. Ezzel szemben egy véges gráfban egy idő után a fertőzés kihal, azaz az A csúcshalmaz üres lesz, és üres is marad. Véges gráfok esetében vizsgáljuk, hogy mennyi idő szükséges ahhoz, hogy minden csúcs egészségesse váljon. Általánosságban azt mondhatjuk, hogy egy fertőzés akkor alakul járvánnyá, ha a kihalásához szükséges idő exponenciális, a gráf csúcscsámának függvényében. Megmutatjuk, hogy egy n csúcsú skálafüggetlen gráf esetében létezik λ_n küszöbérték, amire igaz, hogy, ha a fertőzés λ rátájára teljesül, hogy $\lambda > \lambda_n$ akkor ez egy adott konstans valószínűséggel alakul járvánnyá, és $\lambda_n \rightarrow 0$, ha $n \rightarrow \infty$. Ezzel ellentétben állnak a korlátos fokú gráfok. Ezek jellemző tulajdonsága, hogy minden csúcs foka legfeljebb d lehet. Itt a fertőzés nagy valószínűséggel exponenciálisan gyorsan kihal, ha $\lambda < \frac{1}{2d}$. A következő tétel nem mondja ki az exponenciális időben történő kihalást, csak egy felső korlátot ad annak a valószínűségére, hogy egyszerre sok fertőzött csúcs volt.

2.2.8. Tétel. Legyen adott egy G gráf, amelyben a maximális fokszám d . Legyen S azon csúcsok halmaza a G gráfban, melyek valaha is fertőzöttek voltak. Ekkor $P(|S| > k) < (2d\lambda)^k$ teljesül minden k -ra.

Bizonyítás. Feltesszük, hogy $\lambda d > 1$, ez nem sokban fog különbözni az általános esettől. Legyen az A halmaz számossága X . Pontosabban, legyen az A halmaz számossága a t időpontban X_t , tehát ez végig függ a t -től, de ezt az egyszerűség kedvéért most nem jelöljük. Hogy egy egészséges csúcs fertőzötté váljon, és fordítva, ezen két esemény egyszerre megtörténésének valószínűsége nulla. Ezért a fertőzés terjedési rátáját X -re a következőképpen határozzuk meg:

$$X \rightarrow X - 1 \quad X \text{ rátával, és}$$

$$X \rightarrow X + 1 \quad \lambda |c(A, \bar{A})| \text{ rátával,}$$

$$\text{ahol } c(A, \bar{A}) = \{\{u, v\} \in E : u \in A, v \in \bar{A}\}.$$

Tehát $|c(A, \bar{A})| \leq Xd$, mivel X csúcs van, és mindegyik foka legfeljebb d . Ezért az X_t növekedésének valószínűsége legfeljebb

$$\frac{\lambda X d}{X + \lambda X_t d} = \frac{\lambda d}{1 + \lambda d} < \frac{1}{2} \lambda d,$$

azon feltevés mellett, hogy $\lambda d < 1$, és X csökkenésének a valószínűsége pedig legalább

$$\frac{1}{1 + \lambda d} > 1 - \frac{1}{2}\lambda d.$$

Itt a növekedés és a csökkenés a fertőzött csúcsok számára vonatkozik. Ahhoz, hogy elérjük $X = k + 1$ egyenlőséget, szükség van legalább k növekedésre az első $2k$ esemény között. Ennek a valószínűsége legfeljebb $2^{2k}(\frac{\lambda d}{2})^k = (2d\lambda)^k$. Ezzel megadtuk a felső korlátot, épp ez volt a célunk. \square

2.2.9. Megjegyzés. *Egy csúcs megfertőződése abban az időpillanatban történik, mikor szomszédjai között időben az első fertőz. Az, hogy ezután egy másik szomszédja is terjeszti a fertőzést, már nem számít, hiszen ekkor már úgymint fertőzött a vizsgált csúcs. (Természetesen, ha közben meggyógyult, akkor újból számít.) Ehhez kapcsolódóan használjuk fel, hogy exponenciális eloszlások minimuma exponenciális, ezért lesz a megfertőződés rátája $\lambda|c(A, \bar{A})|$.*

2.2.10. Következmény. *Legyen G egy gráf, és egy adott $v \in G$ csúcs. Legyen l egy rögzített pozitív egész szám. Tegyük fel, hogy v -nek az l sugarú gömbi környezetében minden csúcs fokszáma legfeljebb d . Tegyük fel, hogy a járványterjedés a $\{v\}$ csúcsból indul ki $\lambda < \frac{1}{2d}$ paraméterrel. Ha $T > 0$, akkor $S(T)$ legyen az az esemény, mikor $A_T \neq 0$, és legyen $B(l)$ azon esemény, mikor a fertőzés sosem kerül ki a v csúcs l sugarú környezetéből. Ekkor minden T -re*

$$P(S(T)|B(l)) < (2\lambda d)^T.$$

Tehát, ha nem fertőződik meg egy csúcs sem az adott környezeten kívül, akkor a járvány nagy valószínűséggel kihal.

A következő tételben egy csillag alakú gráfon vizsgáljuk a fertőzés terjedését. Ez a 2.2.3. tétel bizonyításához használható fel. Megmutatjuk, hogy nagy valószínűséggel a csillag formájú gráfban a fertőzés exponenciálisan hosszú ideig túléli a csúcscsillag függvényében. A bizonyítás ötlete a következő: amikor a csillag közepe megfertőződik, akkor a levelek is magas rátával fertőződnek. A fertőzött levelek száma elég magas ahhoz, hogy biztosítsa a fertőzés túlélését a gráfban, amíg a középső csúcs újra meg nem fertőződik. Itt a középső csúcs fontos tulajdonsága, mely a bizonyításhoz is szükséges, hogy viszonylag nagy a fokszáma.

2.2.11. Tétel. *Legyen G egy csillag alakú gráf, x középponttal, és y_1, \dots, y_k levelekkel. Legyen A_t azon csúcsok halmaza, amelyek a t időpontban fertőzöttek. Ekkor létezik C valós szám, hogy ha $A_0 = \{x\}$, akkor $P(A_{\exp(Ck\lambda^2)} \neq 0) = 1 - o(k)$.*

Azaz a tétel azt mondja ki, hogy a fertőzés nagy valószínűséggel exponenciális idő múlva is életben lesz a csillag alakú gráfon.

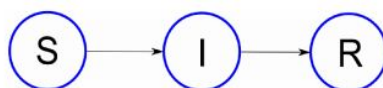
2.2.12. Megjegyzés. *A levelek azon csúcsok, amikből nem megy ki él.*

A következő részben egy kicsit gyakorlatiasabb szemszögből nézzük a járványterjedést, ami talán a modellezés bonyolultságát is előtérbe helyezi.

2.3. Az influenzavírus terjedése

Ebben a részben a [14] cikk ötleteit és ábráit használjuk fel, és az influenza terjedését vizsgáljuk részletesebben. Ez egy olyan probléma, amivel minden évben újra és újra szembesülünk világszerte. Ezért fontos minél jobb és hatékonyabb modellt találni az influenza megelőzésére. Célunk, hogy matematikai modellek segítségével az adott járvány kitörése esetén csökkentjük minimálisan a betegség továbbterjedését, és az esetleges halálozást. Ehhez hasznos, ha ismerjük a vírus fontosabb jellemzőit. A gráfmodellünkben a csúcsok az embereket jelképezik, két csúcs közt futó él pedig azt jelenti, hogy a két ember ismeri egymást, valamilyen kapcsolat van köztük. A vírus adott valószínűséggel terjed a gráfban az éleken keresztül.

A fertőző betegségek modellezésének alapötlete, hogy felosztjuk a populációt független csoportokra néhány jellegzetes tulajdonság szerint, ami fontos a járvány megelőzése szempontjából. Először az egyik legegyszerűbb járványterjedési modellt, a SIR modellt használjuk. Itt 3 csoportba osztjuk az embereket: S , I és R , ahol az S csoportba tartoznak a fertőzésre fogékony személyek, azaz, akik elkaphatják a betegséget. I jelöli a fertőzöttek csoportját, ők azok, akik tovább tudják terjeszteni a betegséget. R pedig a már felépült személyek csoportja, akik már immunisak az adott influenzatípus ellen, azaz őket már nem lehet megfertőzni.



2.2. ábra. Az alap SIR modell

Persze a kategóriákat más szempontok alapján is választhatnánk. Például külön csoportba sorolhatnánk azokat, akik be vannak oltva az influenza ellen, és akik nem, vagy kialakíthatnánk a csoportokat korosztályok szerint is. Sőt, akár az előbbieket kombinálva egy még bonyolultabb modellt is felírhatnánk. Természetesen a dinamikus modelleknél a csoportok közt van átjárás, az egyének átkerülhetnek egyik csoportból a másikba.

Egy fontos fogalom, amit be kell vezetnünk, a „basic reproduction number”, amit R_0 -al jelölünk. R_0 az egy lépésben újonnan megfertőződött csúcsok átlagos száma. Kezdetben egy fertőzött csúcs van, és innen kezd el terjedni a vírus az S csoportban. Általában, ha egy vírus megjelenik

egy közösségben, akkor alakul járvánnyá, ha $R_0 > 1$, ha pedig $R_0 < 1$, akkor a vírus gyorsan kihal, mielőtt még jelentősen elterjedne. Szintén fontos mennyiség, amivel jellemezhetjük a vírus erősségét, a támadási ráta. Ez azt fejezi ki, hogy a lakosság mekkora hányadának kell átesni a betegségen, mielőtt a vírus kihal.

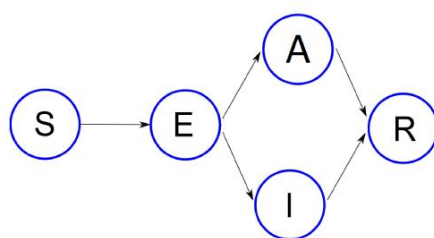
Az influenza egy szezonális vírus, azaz mindig az év egy bizonyos időszakában jelenik meg. Az egyik legjobb védekezési módszer ellene a védőoltás, ám az oltóanyag gyártása időigényes folyamat, ezért már jóval a vírus kitörése előtt megkezdik a megfelelő mennyiség legyártását. Magyarországon általában január és március között zajlik az influenzaszezon, az oltóanyagok pedig már október körül megfelelő mennyiségben készen vannak. Az oltást beadatni körülbelül novemberben érdemes. De az influenza sajnos gyakran mutálódik, és ekkor a tavaly kifejlesztett oltóanyag már nem nyújt védelmet az új változattal szemben. Ilyenkor, hirtelen kitörő új változat esetén előfordulhat, hogy a vírus terjedése és az oltóanyag gyártása párhuzamosan zajlik, gyakorlatilag verseny folyik a kettő között, hogy melyik a gyorsabb. Erre egy példa a 2009 tavaszán megjelent új influenzatörzs, H1N1 típusú járvány volt, melyről lesz szó a 2.4. részben. Több országban nem voltak felkészülve egy ilyen méretű járványra, és nem volt megfelelő mennyiségű oltóanyag sem, a vírus pedig világszerte elkezdett terjedni. Azonban egy ilyen esetet modellezni jóval bonyolultabb, mint mikor az oltóanyag már megfelelő mennyiségben rendelkezésre áll a vírus kitörésekor. Ezért most csak az egyszerűbb esettel foglalkozunk.

Ha az oltások a megfelelő időben beadásra kerültek, akkor ezt a modellben úgy jeleníthetjük meg, hogy a kezdeti állapotban az immunis csoportba kerülnek mindazok, akik megkapták az oltást. Akik elkapják az influenzát, azok miután átestek a betegségen, már nem kapják el újból a vírust az adott szezonban, ellenállóak lesznek ezzel a betegséggel szemben. Az SIR modellnél feltesszük, hogy mindenki egyértelműen besorolható az egyik kategóriába, tehát a csoportok számosságának összege kiadja az egész lakosság számosságát. Az egyes csoportokban lévő személyek száma a következő esetekben változhat: ha egy fertőzött és egy fertőzésre fogékony személy találkozik, vagy amikor valaki felépül a betegségből. Ennél a modellnél az egyszerűség kedvéért azt is feltesszük, hogy nincs be- vagy kilépés a közösségbe. Figyelmen kívül hagyjuk a születést, a természetes halált és a migrációt. Ezenkívül feltesszük azt is, hogy a populáció homogén, azaz minden egyén ugyanazzal a tulajdonságokkal rendelkezik, egyforma valószínűséggel kapják el a vírust, pedig ez a valóságban nem igaz. Az életben ez függ a kortól, környezettől, a saját immunrendszer állapotától, és sok egyéb tényezőtől is.

Az influenza modellezésénél nehézséget jelent, hogy nem minden fertőzött személy produkál tüneteket. Egy jelentős hányaduk tünetmentes, de ugyanúgy hordozza a vírust, képes annak továbbadására. Tehát érdemes bevezetnünk még egy csoportot a modellbe, melybe a tünetmentes, de fertőzést hordozó egyének kerülnek.

Vizsgáljunk meg tehát egy kicsit bonyolultabb modellt, mely a SIR modell továbbfejlesztése.

Két, tipikusan az influenzára jellemző új csoporttal bővítjük az eddigi modellünket, így kapjuk a SEAIR modellt. Az E -vel jelölt új csoportba azok kerülnek, akik épp elkapták a betegséget, de még nem fertőznek. Ezt lappangási időszaknak nevezzük, ami általában viszonylag rövid ideig tart. A másik újonnan bevezetett A csoportba pedig a már említett tünetmentes személyek tartoznak, akik hordozzák a vírust, de nincsenek tüneteik. Róluk nagy valószínűséggel nem derül ki, hogy elkapták a betegséget, de ők is ugyanúgy terjeszthetik tovább, ami nehezebbé teszi a valóság-hű modell létrehozását.



2.3. ábra. A SEAIR modell

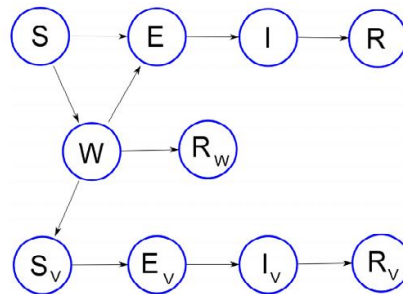
Mint már korábban említettük, az influenza megelőzésére a legjobb módszer a védőoltás. Jelen esetben olyan modelleket vizsgálunk, ahol az oltóanyag gyártása megelőzi az influenzaszezont, szóval, ha egy beoltott közösségben modellezzük az influenza terjedését, akkor kezdetben is vannak immunis személyek. Ekkor a „reproduction number” módosul, ez lesz a „control reproduction number”, amit R_c -vel jelölünk. Ahhoz, hogy elkerüljük a járvány kitörését, $R_c < 1$ szükséges. Néhány esetben előfordulhat, hogy a védőoltás nem nyújt teljes védelmet, vagy az is, hogy nem oltottak be elég embert, így a járvány az oltások ellenére is elterjedhet.

Egy másik lehetséges stratégia az influenza megelőzésére az antivírusos kezelés. Ezt általában akkor alkalmazzák, ha nem áll rendelkezésre megfelelő mennyiségű oltóanyag. Ez a kezelés csökkenti vírus lefolyási idejét, és a fertőzőképességét a beteg embereknek.

Természetesen vannak más megelőzési módszerek is, például a kapcsolatok számának csökkentése. Ez megvalósítható esetleg az iskolák bezárásával, vagy a zsúfolt helyek, nagyobb csomópontok lezárásával. (Ilyen helyek például a bevásárlóközpontok, konferenciák, nagyobb közlekedési csomópontok.) Vagy a vírus terjedésének kezdetekor elzárhatjuk a már fertőzött személyeket az egészségesektől, hogy ne fertőzzenek meg másokat. Persze itt is felmerülnek problémák, példaképp, hogy vannak, akik tünetmentesek, de mégis képesek megfertőzni másokat.

Végül nézzünk meg egy még bonyolultabb modellt, ahol az influenza járvány néhány további jellegzetességét használjuk fel. Ez a SEIR modellen alapul, ahol az eddig használt S, E, I, R jelentése továbbra is ugyanaz. Az első szempont, amit az új modellnél figyelembe veszünk az

emberek kora. A korosztály szerinti besorolás több szempontból is fontos, hiszen sok paraméter korfüggő, ami befolyásolja, hogy mekkora valószínűséggel kaphatja el valaki a vírust, ráadásul a különböző korú személyek más-más közösségekben mozognak. Ezenkívül figyelembe vesszük, a vírus lefolyásához szükséges időt is. Tanulmányok szerint az emberi testnek összesen 14 nap szükséges a védőoltás után annyi antitest előállításához, ami már biztosítja a kellő védettséget. A köztes periódusban sajnos elkapható a vírus. Ez a késleltetés akkor lehet fontos, ha az oltóanyag csak a vírus kirobbanása után lesz elérhető. A harmadik tulajdonságként pedig azt vizsgáljuk, hogy melyik korcsoportot célszerű előbb beoltani. Ehhez össze kell hasonlítani különböző oltási stratégiákat. Cél a 60%-os beoltottság a három hónapos oltási periódus végére.



2.4. ábra. A SEIR modell egy továbbfejlesztése

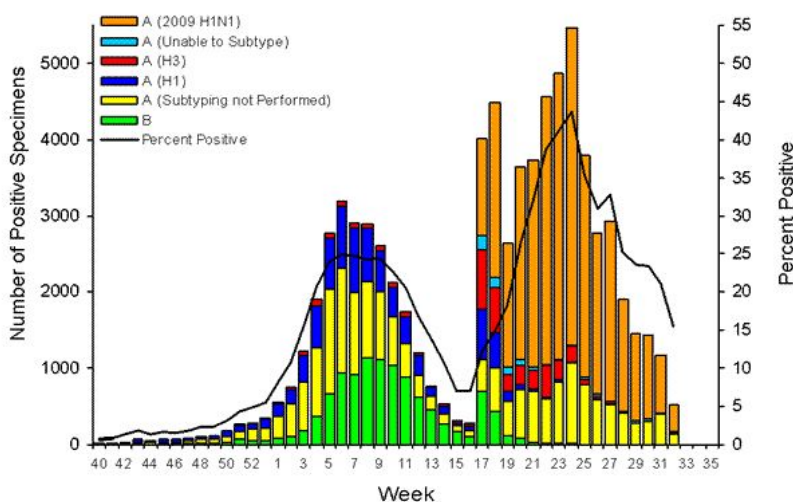
Mivel a lappangási időszak egészen rövid, mindössze körülbelül 1,25 nap, ezért figyelmen kívül hagyjuk azt a kis valószínűséget, hogy valaki akkor kapja meg az oltóanyagot, mikor épp az E csoportban van, azaz mikor már elkapta a betegséget, de még nem jelentkeztek a tünetek. Az S csoportbeli személyeket oltják be, míg el nem éri a 60%-os beoltottságot. W osztályba pedig azok kerülnek, akik már megkapták az oltást, de még fertőzhetőek (ez a periódus körülbelül 14 nap), vagy azok, akiknél az oltás hatástalan. Az oltás beadása után 14 nappal vagy immunissá válnak q valószínűséggel, és az R_W csoportba kerülnek, vagy, ha az oltás hatástalan volt, vagy elkapták a vírust ebben a köztes periódusban, akkor továbbra is fogékonyak a fertőzésre, ekkor az S_V csoportba kerülnek. Itt ugyanúgy történik a fertőzés terjedése, mint az S csoportban. Tehát például a beoltott, de ennek ellenére fertőzött személyek csoportja az I_V .

2.4. A 2008 – 2009-es influenza terjedése megfigyelések alapján

Ebben a fejezetben valós statisztikai adatok által [18] vizsgáljuk a járványterjedés néhány tulajdonságát. Az influenzát nézzük példaképp, melyről már esett szó a 2.3. fejezetben. Ennek a vírusnak is több típusa van, az évek során különböző változatok jelentek meg, mely rendkívül

megnehezítette a kezelést. Most a 2008–2009-es szezonban elterjedt H1N1 típusú vírus terjedését vizsgáljuk részletesebben. A felhasznált adatokat az Amerikai Egyesült Államokban rögzítették, a 2009-es influenzaszezon végével bezárólag.

Influenza Positive Tests Reported to CDC by U.S. WHO/NREVSS Collaborating Laboratories, National Summary, 2008-09

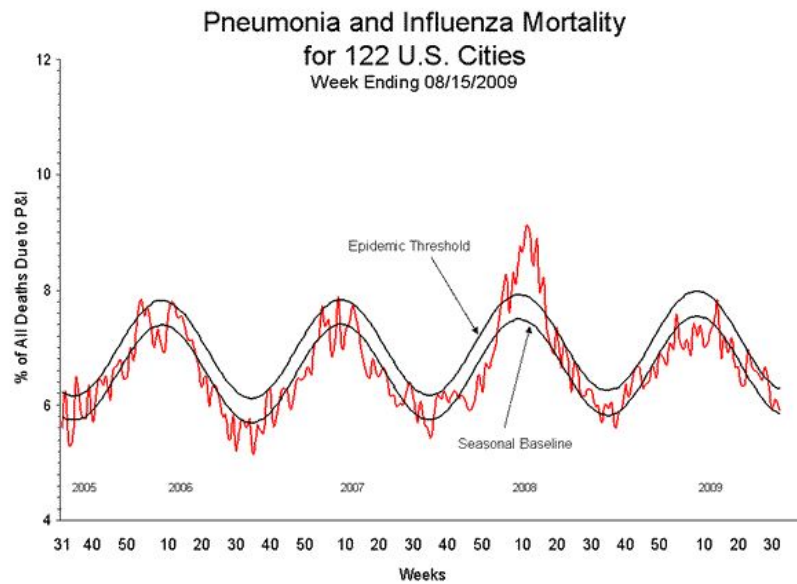


A előző ábrán láthatjuk, hogy a H1N1 vírus megjelenésekor a megbetegedések száma ugrásszerűen megnőtt. Ekkor a legtöbb embert a H1N1 típusú influenza érintett, a többitől diagnosztizált fertőzés láthatóan kevesebb volt.

Ezt úgy képzelhetjük el, mint mikor a vírus átterjed egy komponensből egy másikba. Ilyet az Erdős–Rényi-típusú gráfnál láttunk a 2.1. részben. Ezt hasonlóképpen elképzelhetjük „preferential attachment” modell esetén. Azért ezt nézzük, és nem az Erdős–Rényi-gráfot, mivel a valós hálózatok, így az emberek közti kapcsolathálózat is a „preferential attachment” modellel írható le jobban. Legyen a példánkban a két nagyobb komponens Ázsia és Amerika. Ezen területek közt viszonylag kevés a heti szintű mozgás, azaz a két komponens közt viszonylag kevés él fut. Az ábra szerint a 17. héten jelent meg a vírus Amerikában, és rögtön sok személyt elért a fertőzés. Volt egy kezdeti hullám, ami után kicsit csökkent a fertőzöttek száma, majd ezután, a 24. hét környékén lett a legmagasabb a megbetegedések száma.

Szinte biztos, hogy az új járványtípus miatt volt kiemelkedően magas az influenza és a tüdőgyulladásos megbetegedések által okozott halálozások száma. Itt a következő ábrán az együttes fertőzést jelző piros görbe a 2008-as évben jóval magasabb értékeket mutat, mint a kritikus érték. Itt is megfigyelhető az a jelenség, hogy miután a halálozások száma átlépte a kritikus értéket, először egy kisebb hirtelen növekedés, majd kis visszaesés után egy még nagyobb növekedés tör-

tént. Az ábrán az alsó fekete görbe az átlagos értéket jelöli, míg a felső a kritikus értéket: ha nincs nagy járvány, akkor a halálozások száma ez alatt az érték alatt marad.



Természetesen más szempontok alapján is készíthetők statisztikák, például korcsoportok szerinti felbontásban, vagy földrajzi elhelyezkedés szerint, és így még csak a legtriviálisabb szempontokat vettük sorra. Láthatjuk, hogy a valóság modellezése igen nehéz, és bonyolult feladat. Rengeteg újabb modellt hozhatunk létre, melyek az élet egy bizonyos területéhez nagyban hasonlítanak, de a valóság pontos másának elkészítése gyakran nem megvalósítható, csak tetszőlegesen közelíthető.

3. fejezet

Az internetes vírusok terjedése

A járványterjedés kapcsán használt modelleket az internetes vírusok modellezésére is alkalmazhatjuk. Ezt a modellt kétféleképpen képzelhetjük el. Az egyik lehetőség, hogy a gráf csúcsai a számítógépek lesznek, és azt vizsgáljuk hogyan terjedhet át a vírus az egyik gépről a másikra. A másik lehetőség, hogy az internetes weboldalakat tekintjük csúcsoknak, és egy irányított gráfot kapunk oly módon, hogy amelyik oldalon van a másikra hivatkozás, onnan megy egy irányított él oda, amire hivatkozik. Ez utóbbit Word Wide Web néven ismerjük, most ennek a felépítést nézzük meg részletesebben.

Tehát a Word Wide Web hálózatát egy nagy irányított gráfként képzelhetjük el, ahol a csúcsok a weboldalak, másnéven HTML dokumentumok, az irányított élek pedig a hivatkozások az oldalak közt. A nagy méret, a rengeteg kapcsolat, és ezek folyamatos változása miatt nehéz, sőt szinte lehetetlen a pontos modell megalkotása. Valószínűleg e pillanatban is újabb weboldalak jönnek létre, melyek élekkel kapcsolódnak be a hálózatba. Ezt a gyors változást pedig szinte lehetetlen követni.

Ezekre a modellekre alkalmazhatóak az eddig megismertek: ha csak egészséges és fertőzött csúcsok vannak, akkor ez leírható a SIS modellel, ha vannak védettek az adott számítógépes vírussal szemben, akkor használható a 2.3. alfejezetben megismert SIR modell.

Egy fontos tulajdonság a gráf fokszámeloszlása, melynek ismerete elősegíti a hálózat általános jellemzőinek feltárását. Ha $P(k)$ annak a valószínűsége, hogy egy adott csúcs a gráfban k másik csúcsához csatlakozik, azaz a fokszáma éppen k , akkor $P(k) \sim ck^{-\gamma}$, ahol γ egy 2-nél valamivel nagyobb szám. Itt a csúcsok fokszáma általában kicsivel az átlag alatt van, és van néhány nagyon nagy fokú csúcs, míg az Erdős–Rényi-modellnél a fokszámok minden csúcs esetében közel azonosak voltak. Ezért itt nem is tudunk kiválasztani egy tipikus csúcsot, hiszen a jóval nagyobb és a kisebb fokú csúcsok együttesen alkotják a hálózatot. Valamilyen szinten ez a hálózat skálafüggetlenségét mutatja, mely egy nagyon fontos sajátosság. Az eddig leírt tulajdonságok az általunk már az 1.4. részben említett „preferential attachment” modellel jellemzőek. Valójában a Word

Wide Web felépítését is ezzel a modellel írhatjuk le a legpontosabban.

Másik jellemző a hálózatunkra a kis világ tulajdonság, miszerint bármely két csúcs viszonylag rövid úttal össze van kötve a gráfban. A Web esetén ez a viszonylag rövid út körülbelül 19 hosszúságú, [1] cikk szerint.

A hálózat növekedése a „preferential attachment” modell szabályai szerint történik. Az új weblapok nagyobb eséllyel kapcsolódnak a nagy fokú csúcsokhoz. Azt gondolnánk, hogy azok a csúcsok, melyek időben korábban kerültek be a gráfba, nagyobb eséllyel válhatnak nagy fokú csomóponttá a hálózatban, hiszen több esélyük volt, hogy épp hozzájuk kapcsolódnak az új csúcsok. De láthatjuk, hogy az időben később is kialakultak nagy fokú csúcsok, például a Google és a Facebook. Ezen viszonylag gyorsan létrejött nagy fokú csúcsok létezésének megmagyarázásához Barabási-Albert László bevezette a fitness fogalmát, melyet már említettünk az 1.5. részben. Ez egy adott véletlen szám, amit kisorsolunk minden csúcshoz, és ez is nagymértékben befolyásolja, hogy hány él fog kapcsolódni az adott csúcshoz. Egy csúcs, melyhez egy nagy fitness tartozik, könnyebben válhat nagy fokszámú csomóponttá. Természetesen ezen kívül a fokszámok is befolyásolják, annak valószínűségét, hogy az új csúcsok hová kapcsolódnak.

Az internetet egyre többen használják információcsere céljából, mégis ez talán a legkevésbé ellenőrzött kommunikációs hálózat. Épp emiatt terjedhetnek könnyedén a vírusok, például adott dokumentumok letöltése, vagy adatszere során. A számítógépes vírusok nagyon hasonlóak a biológiai vírusokhoz. De a biológiai vírusokkal szemben a számítógépes vírusokat szinte kivétel nélkül egy-egy személy készítette, előfordul tehát olyan eset, mikor a fertőzés szándékos.

A vírusok károsíthatják a számítógépet, megjelenhetnek szöveggként, videóüzenetként, vagy egyéb formában. Sok kárt okozhatnak, például módosíthatják a memória tartalmát, törölhetnek különböző fájlokat, vagy akár törölhetik az egész merevlemez tartalmát is. Az is előfordulhat, hogy nem vesszük észre, hogy vírusos a gépünk, csak azt hogy minden téren lassabban működik az eddigénél. Ez szintén károsodásra, vagy adathiányra utalhat, amit valamilyen vírus is okozhatott. A vírusnak szüksége van egy fertőzött alkalmazásra, vagy dokumentumra, ahhoz, hogy a fertőzést elterjeszthesse a gépen. Nem minden vírus viselkedik ugyanúgy, sokféle változat létezik, ezért is nehéz védekezni az összes típus ellen. Néhány vírus csak addig aktív, míg az adott fertőzött alkalmazást futtatjuk, de van, amelyik a számítógép kikapcsolása után is aktív marad. Mivel mikor kikapcsoljuk a gépet, akkor általában a vírus csak a memóriából törlődik, de az eleve fertőzött fájlból nem. Ezért mikor újból bekapcsoljuk a gépet, a fertőzés megint aktív lehet.

Fontos kérdés, hogy mi védheti meg a számítógépeket az interneten tejedő vírusokkal szemben. Sokféle vírusirtó, tűzfal, és egyéb védelmi rendszer ismert, vegyünk példaképp egy általános vírusirtót. Azok a számítógépek sem teljesen védettek a vírusoktól, amikre vírusirtót telepítettek, de ezeket bizonyos időközönként átvizsgálja a program, és jelzi, ha talál valamilyen vírust. A számítógép ugyanazzal a vírussal akár többször is megfertőződhet, és minden esetben fertőzött

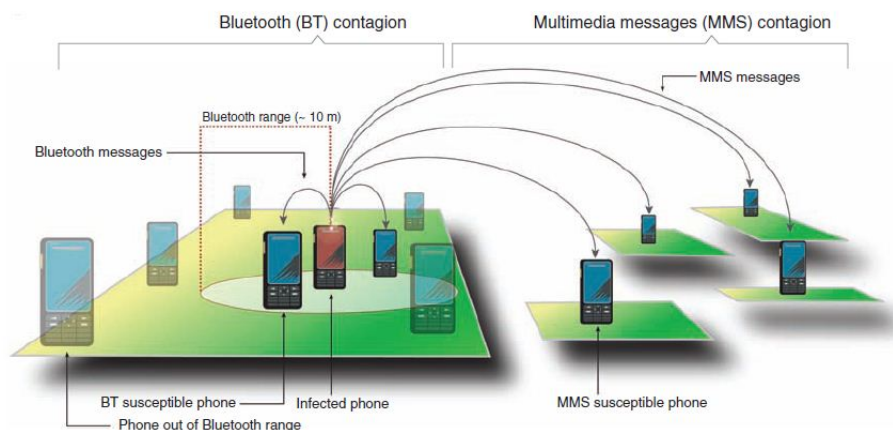
is marad, míg a vírusirtó program meg nem találja újból a vírust. A gyakran frissített és futatott antivírus szoftverek bizonyos védelemmel jelenthetnek a számítógépnek, de az internetes vírusok is képesek a mutációra. A vírusirtó sajnos csak az eddig ismert vírusok ellen nyújt védelmet, az új, mutálódott, eddig még nem ismert változatokkal szemben nem.

Sajnos a hálózat nagy mérete, illetve a gyors változása miatt rengeteg olyan internetes vírus is létezik, melyet nem ismerünk, így nem tudunk ellene megfelelő védelmet sem kialakítani. A vírus felismerése után van lehetőség az adott védelem fejlesztésére, majd ennek beépítésére a weblapokra, hogy a vírus máskor se támadhassa meg az adott oldalakat.

3.1. Vírusok terjedése mobiltelefonok között

Napjainkban egyre gyakoribbak az okostelefonok, melyek közül néhány lassan már a számítógépek teljesítményével is felveszi a versenyt. Mobiltelefonon keresztül is tudunk internetezni, illetve többféleképp is bonyolíthatunk adatforgalmat, például SMS-ben, MMS-ben, vagy Bluetooth-on. A mobiltelefonos vírusok is többféleképp terjedhetnek, most kétféle megközelítést vizsgálunk részletesebben, melyhez a [17] cikk információit használtuk fel.

Az első a Bluetooth kapcsolatokon keresztül történő terjedés. Ez természetesen csak az aktív Bluetooth-os készülékeket veszélyezteti, körülbelül 10 – 30 m hatótávolságon belül. A másik az MMS-en keresztül történő terjedés, mely úgy zajlik, hogy, ha egy adott telefon megfertőződik egy MMS által, akkor ez a vírusos fájl továbbterjed az adott mobil névjegyzékében található összes telefonszámra. Ezt a típusú terjedést nem befolyásolja a mobiltelefonok távolsága, akármilyen távolra eljuthat a vírus ennek segítségével. Itt a távolság alatt természetesen a fizikai távolságra gondolunk. Az előző két lehetséges terjedést mutatja be a következő ábra.



Mindkét lehetséges terjedést jellemezhetjük a SIS modellel, mivel nem ismert nagy hatékonyságú

védelem, mellyel a vírus elkerülhető. De a Bluetooth vírusra használható akár az SIR modell is, ha védetteknek tekintjük azokat a telefonokat, melyeken nincs éppen bekapcsolt állapotban a Bluetooth. Az egyik legszembetűnőbb különbség a két típus között a terjedéshez szükséges idő. Ahhoz, hogy Bluetooth-on keresztül megfertőzzük az összes hatótávolságon belül lévő telefont, néhány napra lenne szükségünk. Viszont egy átlagos MMS vírus esetén a fertőzött tartalom körülbelül 2 perc alatt kerül át egy másik készülékre, így néhány óra alatt sok telefonra terjedhet át a vírus.

Kísérletképp egy adott telefonról indítva követték az MMS és a Bluetooth vírus terjedését, hogy ezzel illusztrálják a terjedésük különbözőségét. A Bluetooth vírus hullámszerűen terjedt, túlnyomórészt a kiindulási pont szomszédságában, míg az MMS vírus szétszórtabb mintát hoz létre, mivel a névjegyekben gyakran távoli felhasználók szerepelnek.

A Bluetooth és az MMS vírusnak is megvannak a maga korlátai. A Bluetooth vírusnál ez a korlát az emberek folyamatos mozgása, helyzetváltoztatása, emiatt a vírus terjedése viszonylag lassú. Az MMS vírus pedig az embereknek csak egy nagyon kis hányadához tud eljutni, mert egy telefonban nincs benne olyan sok telefonszám, ahová a vírus továbbjuthat. Ezt úgy képzelhetjük el, hogy létrehozunk egy irányított gráfot, melyben a mobiltelefonok a csúcsok, és az egyikből megy irányított él a másikba, ha az elsőben benne van a második készülékhez tartozó telefonszám. A Bluetooth vírussal tehát több ember érhető el, viszont lassabban, míg az MMS vírussal kevesebb, de ez gyorsabban is megtörténhet, akár néhány óra alatt.

Láthattuk, hogy a fertőzések terjedésére sok vonatkozásban, többféle modell is felírható. Ezek nem tökéletesek, de viszonylag jól ábrázolják a nagy hálózatokat is, és a különböző vírusok terjedése közt is segítenek felfedezni mind a hasonlóságokat, mind az eltéréseket is. Ezen eredményeknek pedig nagy hasznuk van a gyakorlatban: megfelelő ismeretek mellett megakadályozható egy világméretű járvány kitörése, a minimálisra mérsékelhetjük egy fertőzés elterjedését, akár egy valós betegségről, akár egy internetes vírusról van szó.

Irodalomjegyzék

- [1] BARABÁSI ALBERT-LÁSZLÓ, *Network Science*, Phil. Trans. R. Soc. A, **371** 2013.
- [2] BARABÁSI ALBERT-LÁSZLÓ ÉS ALBERT RÉKA, *Emergence of scaling in random networks. Science*, **286** 509–512, 1999.
- [3] BARABÁSI ALBERT-LÁSZLÓ , ALBERT RÉKA, HAWOONG JEONG, *Scale-free characteristics of random networks: the topology of the world-wide web*, Physica A: Statistical Mechanics and its Applications, **281** 69–77, 2000.
- [4] BERGER, N., BORGS, C., CHAYES, J. AND SABERI, A., *On the spread of viruses on the internet*. In Proc. Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms 301–310. SIAM, Philadelphia., 2005.
- [5] BOLLOBAS B., *A probabilistic proof of an asymptotic formula for the a number of labelled regular graphs*. European J. Combin., 1(4):311–316, 1980.
- [6] BOLLOBÁS B., RIORDAN O., SPENCER J., AND TUSNÁDY G., *The degree sequence of a scale-free random graph process*. Random Structures Algorithms, 18(3):279–290, 2001.
- [7] CSISZÁR VILLŐ, *Diszkrét és folytonos paraméterű Markov láncok*. Egyetemi jegyzet. <http://www.cs.elte.hu/villo/ml/ML.pdf>
- [8] SIVAKOFF D., *Contact process on a graph with communities*. <http://arxiv.org/abs/1210.3434>
- [9] DEREICH, STEFFEN, ORTGIESE, MARCEL, *Robust analysis of preferential attachment models with fitness*. <http://arxiv.org/abs/1302.3385v1>
- [10] DURRETT, RICK., *Random graph dynamics*. Cambridge University Press, 2007.
- [11] ERDŐS PÁL ÉS RÉNYI ALFRÉD, *On random graphs. I*. Publ. Math. Debrecen, **6** 290–297, 1959.

- [12] GILBERT, *E. N. Random graphs*. Ann. Math. Statist., **30** 1141–1144, 1959.
- [13] KATONA ZSOLT, *Véletlen gráfmodellek*. Doktori értekezés, 2006.
- [14] KNIPL DH., RÖST G., *Influenza models with Wolfram Mathematica*, pp 1–24, Chapter 4 in: Interesting Mathematical Problems in Sciences and Everyday Life -2011 (eds: J Karsai, R Vajda), Szeged, 2011.
- [15] KNIPL DH., RÖST G., *Modelling the strategies for age specific vaccination scheduling during influenza pandemic outbreaks*, Math. Biosci. Eng. 8(1), 123-139 2011.
- [16] VAN DER HOFSTAD, REMCO., *Random Graphs and Complex Networks*. Előkészületben.
<http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>
- [17] WANG P., GONZÁLEZ C., HIDALGO A., BARABÁSI ALBERT-LÁSZLÓ, *Understanding the Spreading Patterns of Mobile Phone Viruses*, Science, **324** 1071–1076, 2009.
- [18] <http://www.cdc.gov/flu/weekly/weeklyarchives2008-2009/weekly32.htm>
- [19] <https://www.wiredsafety.org/internet101/viruses.html>