Eötvös Loránd Tudományegyetem

Természettudományi Kar

Hálózati folyamatok modellezése differenciálegyenletekkel

MSc szakdolgozat

Írta: Szabó András

Alkalmazott Matematikus MSc szak

Témavezető:

Simon Péter, egyetemi docens

Alkalmazott Analízis és Számításmatematikai Tanszék

Eötvös Loránd Tudományegyetem, Természettudományi Kar

Budapest, 2011

Tartalomjegyzék

1. Bevezetés **1**

1.1. Rövid áttekintés 1

2. Elméleti alapok **1**

2.1. Az SIS modell 1

2.1.1. További modellek 2

2.2. Kolmogorov egyenletek 2

2.3. Várható érték egyenletek 5

2.4. Az egyenletrendszer lezárása 7

2.4.1. Harmadrendű lezárás 9

2.5. Összevonás teljes gráf esetén 10

3. Közelítések és szimulációk **13**

3.1. Monte Carlo módszer és a szimulációs algoritmus 13

3.2. Teljes gráf vizsgálata 15

4. Vizsgálatok és eredmények **22**

4.1. Módosított körgráf vizsgálata 22

4.2. Új módszerek 26

4.2.1. A *pair closure* javítása 26

4.2.2. Új közelítő differenciálegyenlet 28

4.3. Összefoglalás 35

Köszönetnyilvánítás **36**

Irodalomjegyzék **37**

1. fejezet

Bevezetés

Jelen dolgozat célja, hogy bevezesse az Olvasót a hálózati folyamatok témakörének egyik mai napig sokat vizsgált ágazatába. Az egyik leggyakrabban vizsgált folyamat az úgynevezett SIS járványterjedési modell, melynek számos gyakorlati alkalmazása van. Ezen folyamatokat dinamikai szempontból sok esetben csak különböző közelítések kidolgozásával lehet megérteni. A fő célunk ilyen közelítések kidolgozása, azonban ez általában nehéz és hosszú kutató munkát igényel. A dolgozatban egy speciális hálózat megértését tűztük ki elsődleges feladatunknak és reményeink szerint ezen eredmények illetve módszerek a későbbiekben még további problémák esetén is jól hasznosíthatóak lesznek. A felépítést illetően egy rövid betekintést nyújtunk a témakörbe, majd az elméleti alapok tárgyalása után a vizsgálati módszerekhez kalauzoljuk az Olvasót, végül a dolgozat zárásaként tömören bemutatjuk az új módszereket melyeket kidolgoztunk.

1.1. Rövid áttekintés

A hálózatok témaköre a mai modern alkalmazott matematika talán leginkább vizsgált területe. Számos cikk és tudományos értekezés jelenik meg a témában, érintve a matematika számos területét. Az alkalmazások száma egyre csak nő, hálózatokkal kapcsolatos kérdések felvetődnek olyan tudományágakban mint például az ökológia, epidemiológia, vagy akár a neurológiától kezdve a bioinformatikán keresztül egészen az számítástudományig. A kutatások egyik fő célja, a hálózatok szerkeze-

tének megértése. Olyan kérdések megválaszolása mint például, hogyan fejlődik egy adott hálózat, vagy milyen szabályszerűségek fedezhetőek fel benne. A másik fontos kérdéskör, hogy a hálózat struktúra, hogyan befolyásolja azon folyamatokat, melyek a hálózaton mennek végbe. Ilyen folyamatokra jó példa a járványterjedés vagy a neurális hálózatokban az ingerületvezetés. Ezen folyamatoknál felmerülő kérdéseket sokszor szimulációkból kapott eredményekkel próbálják megválaszolni. Ezek sokszor könnyen megvalósítható algoritmusok, de a hálózat szerkezetének dinamikára gyakorolt hatását nehéz megérteni ezen az úton. Szükség van tehát, általánosabb elméleti eredményekre, akár a sztochasztikus folyamatok vagy akár ezen folyamatokat közelítő differenciálegyenletek témakörében.

A sztochasztikus megközelítések elméletében úttörő szerepet játszott Kurtz [4, 5], aki elsősorban az ugró Markov-folyamatok, illetve a martingálok felől közelítette meg a problémát. Őt követték Keeling és Rand, akik a későbbiekben tárgyalt lezárásokban alkalmazott közelítéseket kidolgozták, ezzel megalapozva a differenciálegyenletes közelítő egyenletek alkalmazását. A mi célunk, hogy elsősorban a dinamikai rendszerek illetve a differenciálegyenletek elméletét felhasználva érjünk el eredményeket, azonban megjegyezzük, hogy a kutató munkánk jártasságot kíván a sztochasztikus folyamatok illetve szimulációs algoritmusok terén is. A hálózati folyamatok népszerűségét nagyban erősíti, hogy a matematika sokszor távolinak tűnő ágazatait összekapcsolja, nem ritka, hogy találkozik a diszkrét matematika, valószínűség számítás és az alkalmazott analízis.

2. fejezet

Elméleti alapok

A bevezető után jelen fejezetben áttérünk az elméleti alapok tárgyalására és bemutatjuk az úgynevezett SIS modellt. A dolgozatban mindvégig ezt a modellt fogjuk használni a vizsgálatainkhoz, azonban megemlítünk néhány további modellt is megjegyezve, hogy általában ezek az SIS modell általánosításai.

2.1. Az SIS modell

Tegyük fel, hogy adott egy N csúcsú gráf, melyet hálózatnak hívunk. Az egyszerűség kedvéért ezt a gráfot az adjacencia mátrixával fogjuk reprezentálni, azaz legyen $G = (g_{ij})_{i,j=1,2,\ldots,N} \in \{0,1\}^{N^2}$ és $g_{ij} = 1$ ha az i és j csúcsok közt él fut és $g_{ij} = 0$ ha nincsenek össze kötve. A hálózatunk irányítatlan és hurokélmentes, tehát $G^T = G$ és $g_{ii} = 0$ minden i-re. A gráfban minden csúcsnak két lehetséges állapota van, S vagyis fertőzhető, illetve I azaz fertőzött. Az iménti jelölések az angol susceptible és infected szóból erednek, megjegyezzük továbbá, hogy minden fertőzött fertőz is egyben. Egy egészséges csúcsot csak a fertőző szomszédai fertőzhetnek egyenként τ rátával, a fertőzött csúcsok pedig γ rátával gyógyulnak. A modellben bizonyos valószínűséggel történik fertőzés illetve gyógyulás és az imént említett ráták az ezekhez tartozó eloszlások paraméterei. Ha egy csúcs meggyógyult utána újrafertőződhet, innen adódik az elnevezés ($S \to I \to S$). A gyógyulás és a fertőzés független Markov folyamatok, hiszen a múlt nem befolyásolja a jelent. Ha tekintünk egy kicsi δt időintervallumot, akkor fertőzés valószínűsége $1 - \exp(-N_I \tau \delta t)$, ahol N_I jelöli a

fertőző szomszédok számát. Hasonlóan a gyógyulás is exponenciális eloszlású, azaz a gyógyulás valószínűsége $1 - \exp(-\gamma\delta t)$.

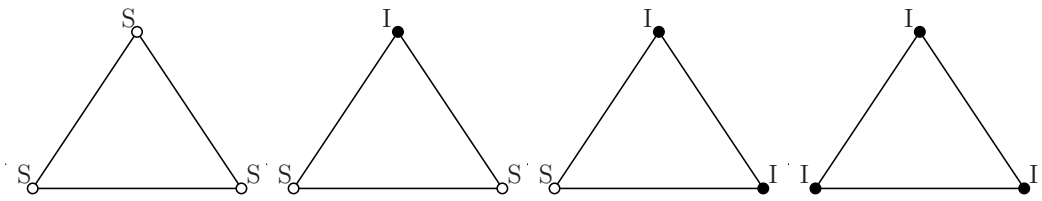
2.1.1. További modellek

A továbbiakban említés szintjén ismertetünk néhány fontos modellt. Az egyik legfontosabb az úgynevezett Kermack-McKendrick féle SIR modell. Az S és az I az előzőekhez hasonló, azonban van egy harmadik csoport amit R-el jelölünk. Ezen csoportba tartoznak azok akik megfertőződtek és utána már immunisak a betegségre, tehát nem lehet újra fertőzni őket. Ebből adódik az elnevezés amit szemléletesen így szoktak jelölni, $S \to I \to R$. Megjegyezzük, hogy általában az SIS modellt is ebből szokták származtatni, méghozzá úgy, hogy nem alakulhat ki immunitás. Még egy fontos modellt említünk ez pedig nem más mint az $SEIR$ modell ami abban különbözik, az SIR modelltől, hogy ha valaki megkapja a fertőzést akkor még nem válik fertőzővé hanem először az E csoportba kerül, majd egy bizonyos lappangási időt követően lesz fertőző.

2.2. Kolmogorov egyenletek

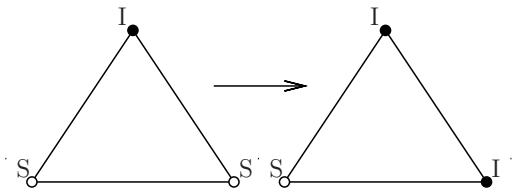
Az modellünk szerint a terjedés tehát nem más mint egy sztochasztikus folyamat melynek állapotai N csúcs gráfok, összesen 2^N darab. A továbbiakban szeretnénk felírni az úgynevezett Kolmogorov egyenleteket. Ezen egyenletek segítségével megkaphatjuk, hogy melyik állapotnak mi a valószínűsége, ez nyilván t függvénye. Kezdetben mesterségesen meg kell fertőznünk legalább egy csúcsot, hiszen ha mindenki egészséges akkor nincs terjedés, ezzel bővebben a szimulációs algoritmus tárgyalásánál foglalkozunk. A további időpillanatokban bármely csúcs bármely állapotú lehet, ezért érdemes egy N hosszú vektorral reprezentálni a rendszer állapotát, ilyen vektorokból összesen 2^N darab van. A lehetséges állapotokat osszuk fel $N+1$ részhalmazra a következő módon. Legyen $\mathcal{S}^0$ azon állapot halmaza amelyben minden csúcs egészséges, azaz $\mathcal{S}^0 = (S, S, \ldots, S)$, általában pedig $\mathcal{S}^k$ jelölje azon állapotokat melyekben k darab fertőzött van. Az $\mathcal{S}^k$ halmaz elemei legyenek $\mathcal{S}_1^k, \mathcal{S}_2^k, \ldots, \mathcal{S}_{c_k}^k$, ahol $c_k = \binom{N}{k}$ és egy l csúcs állapotára a következőképpen fogunk hivatkozni, $\mathcal{S}_j^k(l) = S$

ha egészséges, illetve $\mathcal{S}_j^k(l) = I$ ha fertőzött.



2.1. ábra. $\mathcal{S}^k$ $(k = 0, 1, 2, 3)$ halmaz néhány eleme 3 csúcsú teljes gráf esetén

Fertőzés esetén $\mathcal{S}_j^k \to \mathcal{S}_i^{k+1}$ típusú átmenetről van szó, ahol i és j olyanok, hogy létezik l, melyekre $\mathcal{S}_j^k(l) = S$ és $\mathcal{S}_i^{k+1} = I$, valamint $S_j^k(m) = S_i^{k+1}(m)$ is teljesül minden $m \neq l$ esetén (2.2 ábra). Ekkor nyilván létezik egy r csúcs, melyre $S_j^k(r) = I$ és $g_{lr} = 1$. Hasonlóan az előbbihez, gyógyulásnál egy $\mathcal{S}_j^k \to \mathcal{S}_i^{k-1}$ típusú változás történik, ahol i-re és j-re teljesül, hogy létezik l amelyre $\mathcal{S}_j^k(l) = I$ és $\mathcal{S}_i^{k-1}(l) = S$, továbbá $\mathcal{S}_j^k(m) = \mathcal{S}_i^{k-1}(m)$ minden $m \neq l$-re.



2.2. ábra. Fertőzés három csúcsú teljes gráfon

Jelöljük $X_j^k(t)$-vel annak a valószínűségét, hogy a t időpillanatban az $\mathcal{S}_j^k$ állapotban van és legyen $X^k(t) = \left(X_1^k(t), X_2^k(t), \ldots, X_{c_k}^k(t)\right)$, $k = 0, 1, \ldots, N$. Az $X_j^k(t)$ felírjuk a Kolmogorov egyenleteket:

$$\dot{X}^k(t) = A^k X^{k-1}(t) + B^k X^k(t) + C^k X^{k+1}(t) \qquad k = 0, 1, \ldots, N \tag{2.1}$$

ahol az A^k, B^k, C^k mátrixokat az adott gráfstruktúra határozza meg. Az előbbi egyenletrendszert mátrix alakban is felírhatjuk, azaz

$$\dot{X}(t) = PX(t)$$

ahol

$$P = \begin{pmatrix} B^0 & C^0 & 0 & 0 & 0 & 0 \\ A^1 & B^1 & C^1 & 0 & 0 & 0 \\ 0 & A^2 & B^2 & C^2 & 0 & 0 \\ 0 & 0 & A^3 & B^3 & C^3 & 0 \\ 0 & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & A^N & B^N \end{pmatrix}$$

Általános elmondható, hogy az A^k mátrix mindig a fertőzésért felelős tag és C^k felel a gyógyulásért, továbbá A^0, C^N nullmátrixok. Jelöljük az A^k mátrix (i,j) elemét $A^k_{i,j}$-val, továbbá a C^k mátrixét $C^k_{i,j}$-val. A 2.1 egyenletből következik, hogy az $A^k_{i,j}$ elem adja meg az $\mathcal{S}^{k-1}_j \to \mathcal{S}^k_i$ fertőzés rátáját. Legyen az $\mathcal{S}^{k-1}$ halmaz elemszáma c_{k-1}, illetve az $\mathcal{S}^k$ halmazé c_k, ekkor az A^k blokk $c_{k-1} \times c_k$ méretű lesz. Tekintsünk egy l csúcsot amely az $\mathcal{S}^{k-1}_j$ állapotban S típusú volt és legyen $\mathcal{S}^k_i$ az az állapot, melyre $\mathcal{S}^k_i(l) = I$ és a többi csúcs típusa azonos. Az $\mathcal{S}^{k-1}_j$ állapotban az l csúcs és fertőzött szomszédjai közt futó élek számát jelöljük q-val. Ekkor mivel minden fertőzött külön-külön τ rátával fertőz, ezért $A^k_{i,j} = \tau q$. A továbbiakban (S, I) típusú élnek fogjuk nevezni azon éleket, melyeknél az egyik csúcs egészséges, a másik csúcs pedig fertőzött. Ha az előbbi számolást elvégezzük az összes l csúcsra, amelyre $\mathcal{S}^{k-1}_j(l) = S$, akkor megkapható az A^k mátrix j-edik oszlopösszege, tehát

$$\sum_{i=1}^{c_k} A^k_{i,j} = \tau N_{SI}(\mathcal{S}^{k-1}_j) \tag{2.2}$$

minden $j = 1, 2, \dots, c_{k-1}$-re ahol $N_{SI}(S^{k-1}_j)$ jelöli az (S, I) élek számát.

Röviden tárgyaljuk a C^k mátrix esetét, az előzőekhez hasonlóan. Mivel C^k az $\mathcal{S}^{k+1}_j \to \mathcal{S}^k_i$ átmenet rátáját tartalmazza, ezért a blokk mérete $c_{k+1} \times c_k$-es. Ha valamely l csúcsra $\mathcal{S}^{k+1}_j = I$ és $\mathcal{S}^k_i(l) = S$, valamint a többi csúcsra megegyezőek az állapotok, akkor $C^k_{i,j} = \gamma$. Az $\mathcal{S}^{k+1}_k$ állapotban $k+1$ darab fertőző csúcs van, ezért a C^k mátrix j-edik oszlopában $k+1$-szer szerepel a γ ráta, tehát az alábbi egyenlet adódik

$$\sum_{i=1}^{c_k} C^k_{i,j} = \gamma(k+1) \tag{2.3}$$

minden $j = 1, 2, \dots, c_{k+1}$ esetén.

A B^k mátrix az $\mathcal{S}_i^k \to \mathcal{S}_j^k$ átmenetek rátáját adja meg, azonban ez $i \neq j$ esetén nulla. Tehát B^k egy $c_k \times c_k$ méretű diagonális mátrix. Mivel a P mátrix minden oszlopösszege nulla ezért az előző két egyenlet ismeretében az alábbi feltétel adódik

$$B_{i,i}^k = -\sum_{j=1}^{c_{k+1}} A_{j,i}^{k+1} - \sum_{j=1}^{c_{k-1}} C_{j,i}^{k-1} \tag{2.4}$$

2.3. Várható érték egyenletek

Az előbb tárgyalt P mátrixot sokszor nem könnyű felírni, ez ugye mindig a gráfstruktúrától függ. Azonban ha sikerül is a rendszer mérete 2^N, ahol N a csúcsszám. Valós problémák esetén ez reménytelenül sok egyenletet jelent, még numerikus módszerekkel sem tudjuk előállítani a megoldást. A problémát úgy orvosoljuk, hogy csak a fertőzöttek számának várható értékét szeretnénk kiszámítani, általában ez elég információt szolgáltat a terjedés viselkedéséről. Nyilvánvalóan a várható értéket az előbbi differenciálegyenlet rendszer megoldásaiból is ki tudnánk számolni, azonban akad más megoldás is. Vezessük be az $[I](t)$ jelölést a fertőzöttek, míg az $[S](t)$ jelölést az egészségesek számának várható értékére az alábbi módon

$$[I](t) = \sum_{k=0}^{N} k \sum_{j=1}^{c_k} X_j^k(t), \qquad [S](t) = \sum_{k=0}^{N} (N-k) \sum_{j=1}^{c_k} X_j^k(t) \tag{2.5}$$

Az a célunk, hogy ezen mennyiségekre felírjunk valamilyen összefüggést, előbb azonban bevezetjük az (S, I) típusú élek számának várható értékét a következőképpen

$$[SI](t) = \sum_{k=0}^{N} \sum_{j=1}^{c_k} N_{SI}(\mathcal{S}_j^k) X_j^k(t)$$

ahol $N_{SI}(\mathcal{S}_j^k)$ jelöli az $\mathcal{S}_j^k$ állapotbeli SI élek számát.

A továbbiakban megfogalmazunk és igazolunk egy összefüggést a várható értékekkel kapcsolatban, amely alapvető szerepet tölt be a későbbiekben. Nem túlzás azt állítani, hogy az eredményeink nagyrésze az alábbi lemmára épül. A továbbiakban erősen támaszkodunk a [6] cikk tartalmára.

2.3.1. Lemma. *Az $[S]$ illetve $[I]$ várható értékek kielégítik az alábbi egyenletrendszert*

$$[\dot{S}] = \gamma[I] - \tau[SI] \tag{2.6}$$

$$[\dot{I}] = \tau[SI] - \gamma[I] \tag{2.7}$$

Bizonyítás: Legyen $e_k = (1, 1, \ldots, 1)$ csupa 1-esekből álló sorvektor, melynek c_k eleme van. Ezt a jelölést felhasználva $\sum_{j=1}^{c_k} X_j^k = e_k X^k$ adódik. Az (2.5) egyenletet az alábbi alakba írhatjuk

$$[I](t) = \sum_{k=0}^{N} k e_k X^k, \qquad [S](t) = \sum_{k=0}^{N} (N-k) e_k X^k \tag{2.8}$$

valamint az (2.4) összefüggés a következő alakot ölti

$$B_{i,i}^k = -(e_{k+1}A^{k+1})_i - (e_{k-1}C^{k-1})_i \tag{2.9}$$

Mivel B^k diagonális mátrix, ezért fenn áll, hogy $B_{i,i}^k = (e_k B^k)_i$. Ezt felhasználva a következőt kapjuk

$$(e_k B^k)_i = -(e_{k+1}A^{k+1})_i - (e_{k-1}C^{k-1})_i, \quad \forall i = 1, \ldots, c_k \tag{2.10}$$

Ebből adódik, hogy minden $k = 0, 1, \ldots, N$ esetén teljesül, hogy

$$e_{k+1}A^{k+1} + e_k B^k + e_{k-1}C^{k-1} = 0 \tag{2.11}$$

Deriválva az $[I](t)$ függvényt és felhasználva az (2.1) összefüggést az alábbi egyenlet adódik

$$[\dot{I}] = \sum_{k=0}^{N} k e_k \dot{X}^k = \sum_{k=0}^{N} k e_k (A^k X^{k-1} + B^k X^k + C^k X^{k+1}) =$$

$$\sum_{k=1}^{N} k e_k A^k X^{k-1} + \sum_{k=0}^{N} k e_k B^k X^k + \sum_{k=0}^{N-1} k e_k C^k X^{k+1} =$$

$$\sum_{k=0}^{N-1} (k+1) e_{k+1} A^{k+1} X^k + \sum_{k=0}^{N} k e_k B^k X^k + \sum_{k=1}^{N} (k-1) e_{k-1} C^{k-1} X^k =$$

$$\sum_{k=0}^{N} \Big((k+1) e_{k+1} A^{k+1} + k e_k B^k + (k-1) e_{k-1} C^{k-1} \Big) X^k =$$

$$\sum_{k=0}^{N} \Big(\underbrace{k e_{k+1} A^{k+1} + k e_k B^k + k e_{k-1} C^{k-1}}_{=0} + e_{k+1} A^{k+1} - e_{k-1} C^{k-1} \Big) X^k.$$

Tehát az alábbi egyenletet kaptuk

$$[\dot{I}] = \sum_{k=0}^{N} \Big(e_{k+1} A^{k+1} - e_{k-1} C^{k-1} \Big) X^k = \sum_{k=0}^{N} e_{k+1} A^{k+1} X^k - \sum_{k=0}^{N} e_{k-1} C^{k-1} X^k \quad (2.12)$$

A (2.2) összefüggésből $(e_{k+1}A^{k+1})_j = \sum_{i=1}^{c_{k+1}} A_{i,j}^{k+1} = \tau N_{SI}(\mathcal{S}_j^k)$ adódik. Ezt behelyettesítve a fenti összeg első tagjába kapjuk, hogy

$$\sum_{k=0}^{N} e_{k+1} A^{k+1} X^k = \sum_{k=0}^{N} \sum_{j=1}^{c_k} (e_{k+1} A^{k+1})_j X_j^k = \tau \sum_{k=0}^{N} \sum_{j=1}^{c_k} N_{SI}(\mathcal{S}_j^k) X_j^k(t) = \tau [SI]$$

Hasonlóan (2.3) egyenletből $(e_{k-1}C^{k-1})_j = \sum_{i=1}^{c_{k-1}} C_{i,j}^{k-1} = \gamma k$. Ezt összevetve a (2.12) egyenlet második tagjával

$$\sum_{k=0}^{N} e_{k-1} C^{k-1} X^k = \sum_{k=0}^{N} \sum_{j=1}^{c_k} (e_{k-1} C^{k-1})_j X_j^k = \gamma \sum_{k=0}^{N} k \sum_{j=1}^{c_k} X_j^k = \gamma [I]$$

Vagyis igazoltuk a lemmát az $[I](t)$ mennyiségre, az $[S](t)$-re vonatkozó egyenlet esetében a bizonyítás hasonló $\square$.

2.4. Az egyenletrendszer lezárása

Az imént tárgyalt lemma elvileg lehetőséget biztosít, hogy egyetlen egy egyenlet megoldásával megkapjuk a fertőzöttek számának várható értékét, azonban a jobboldal tartalmazza az $[SI]$ mennyiséget amit nem ismerünk. Az a tervünk, hogy valamilyen közelítést alkalmazunk ami azonban már csak $[I]$ illetve $[S]$ ismeretleneket tartalmazza, és így már megoldhatjuk az egyenletrendszert. Ezt az eljárást hívják az egyenletrendszer lezárásának (angolul *closure*), azaz amikor a plusz ismeretlent a meglévőek valamely kifejezésével helyettesítjük. Pontos formulát nehéz lenne találni általános esetben, így csak valamilyen közelítését fogjuk használni az $[SI]$ várható értéknek, ez bizonyos esetekben jó közelítést ad máskor kevésbé. Megjegyezzük, hogy ezen ötlet Keeling-től ered [3].

Most bemutatjuk a legelterjedtebb közelítést amit a szakirodalomban *pair closure* néven említenek. Tegyük fel, hogy a G gráf olyan, hogy minden csúcs foka azonos, mégpedig legyen ez a fokszám n, tovább a csúcsszámot jelöljük N-el. Ekkor a közelítésünk az alábbi

$$[SI] \approx [S][I] \frac{n}{N-1}$$

A fenti összefüggést röviden, heurisztikus magyarázattal indokoljuk. Tekintsünk egy S típusú csúcsot, mivel $[I]$ darab fertőzött van ezért átlagosan $\frac{[I]}{N-1}$ eséllyel lesz egy szomszédja fertőzött, de mivel n szomszédja van ezzel meg kell szorozzuk az előbbi kifejezést végül pedig szorzunk $[S]$-el mivel ennyi darab S, azaz egészséges típusú csúcs van. A közelítés előnye, hogy igen egyszerű, azonban ahogy az a heurisztikus magyarázatból is kiderül, csak akkor jó, ha fertőzöttek illetve egészséges száma nagyjából független.

Legyen a továbbiakban a hálózatunk az N csúcs teljes gráf. Alkalmazzuk a fenti közelítő egyenleteket és jelöljük a kapott megoldásokat $\tilde{I}$ illetve $\tilde{S}$-al. Mivel ekkor a gráfunk fokszáma $n = N - 1$ minden csúcs esetén, ezért a fenti közelítés $[SI] \approx [S][I]\frac{n}{N-1} = [S][I]$ alakot ölti, azaz az egyenletrendszerünk az alábbi,

$$\dot{\tilde{S}} = \gamma\tilde{I} - \tau\tilde{S}\tilde{I} \tag{2.13}$$

$$\dot{\tilde{I}} = \tau\tilde{S}\tilde{I} - \gamma\tilde{I} \tag{2.14}$$

A keresett ismeretlen mennyiségeket az egyszerűség kedvéért normáljuk a $[0, 1]$ intervallumba. Ehhez vezessük be a következő jelöléseket, $\tau = \frac{\beta}{N}$ valamely rögzített β szám mellett, továbbá $\tilde{s} = \frac{\tilde{S}}{N}$, $\tilde{i} = \frac{\tilde{I}}{N}$, $[s] = \frac{[S]}{N}$, $[i] = \frac{[I]}{N}$. Ekkor az egyenletrendszer az alábbi alakot ölti

$$\dot{\tilde{s}} = \gamma\tilde{i} - \beta\tilde{s}\tilde{i} \tag{2.15}$$

$$\dot{\tilde{i}} = \beta\tilde{s}\tilde{i} - \gamma\tilde{i} \tag{2.16}$$

Felmerül a kérdés, hogy ezen típusú közelítés mennyire ad jó eredményt. Ezt általában nehéz megválaszolni és sokszor csak szimulációs eredményekkel összevetve kaphatunk képet a hibáról, azonban teljes gráf esetén az alábbi tételt mondhatjuk, melyről bővebben [6] dolgozatban.

1. Tétel. *Legyen G egy teljes gráf és $\tau = \frac{\beta}{N}$ valamely rögzített β-ra. Jelölje az* (2.1) *egyenlet megoldását X és legyenek $(\tilde{s}, \tilde{i})$ az* (2.15) *egyenlet megoldásai a következő kezdeti feltételekkel $\tilde{s}(0) = [s](0)$ és $\tilde{t}(0) = [t](0)$. Ekkor bármely $t \geq 0$ esetén*

$$\lim_{N\to\infty} |\tilde{s}(t) - [s](t)| = 0 \tag{2.17}$$

$$\lim_{N\to\infty} |\tilde{i}(t) - [i](t)| = 0 \tag{2.18}$$

A teljes gráf mint speciális hálózat még a későbbi fejezetekben is fontos szerepet fog játszani, azonban most egy újabb fajta közelítést tárgyalunk

2.4.1. Harmadrendű lezárás

Az előző lemma ereje abban rejlett, hogy a várható értéket egyetlen egy egyenletből megkaphatjuk, mivel az $[S] + [I] = N$ összefüggés fenn áll. Most bemutatunk egy módszert amely ugyancsak konstans számú egyenletrendszert eredményez és általában jobb közelítést kapunk mint ha a korábbi módszert használnánk.

Vezessük be az $[II], [SS]$ jelöléseket az (I, I) illetve az (S, S) típusú élek számának várható értékére. Az alapötlet az, hogy az $[SI]$ értéket további új mennyiségekkel kifejezzük, majd azokra ugyancsak közelítést adunk. Tegyük fel, hogy egy A típusú csúcs össze van kötve egy B típusúval és az egy C típusúval ($A, B, C \in \{S, I\}$), így egy élpárt kapunk ami egy 2 hosszú út A és C között. Az ilyen élpárok számának várható értékét a továbbiakban $[ABC]$-vel jelöljük. Ezen mennyiségek bevezetése után megfogalmazhatunk egy lemmát az előző szakaszban tárgyalthoz hasonlóan.

2.4.1. Lemma. *Az $[S], [I], [SI], [II], [SS]$ mennyiségekre fenn áll az alábbi egyenletrendszer.*

$$\begin{aligned}
[\dot{S}] &= \gamma[I] - \tau[SI] \\
[\dot{I}] &= \tau[SI] - \gamma[I] \\
[\dot{SI}] &= \gamma([II] - [SI]) + \tau([SSI] + [ISI] - [SI]) \\
[\dot{II}] &= -2\gamma[II] + 2\tau([ISI] - [SI]) \\
[\dot{SS}] &= 2\gamma[SI] - 2\tau[SSI]
\end{aligned}$$

Az [ISI] illetve az [SSI] várható értékekre közelítést kell alkalmazzunk, ezt a módszert a szakirodalomban *triple closure* néven találjuk, vagyis háromszoros lezárás. Gyakorlatilag az előző rendszert bővítettük és hasonlóan bővíthetnénk tovább akár végtelen dimenziós rendszerré is. A lemmában szereplő ismeretlen mennyiségek közelítésére Keeling és Rand formuláját [3] használjuk, amely az alábbi

$$[ABC] \approx \frac{n-1}{n} \frac{[AB][BC]}{[B]}$$

A fenti képlet heurisztikus magyarázata hasonló a korábbi közelítéséhez, részletesebben a már említett [3] cikkben olvashatunk róla.

A korábban tárgyalt lezárás esetében egy tétel garantálta, hogy a csúcsszám növelésével a közelítésünk is egyre jobb lesz. Jelen esetben nincs ilyen erős tételünk, ezért általában a korábbi *pair closure* közelítést vesszük az összehasonlítás alapjául.

2.5. Összevonás teljes gráf esetén

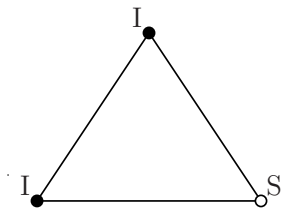
Korábban már említettük, hogy a Kolmogorov egyenletek egy 2^N méretű differenciálegyenlet rendszert alkotnak, ahol N a gráfunk csúcsszáma. Ez kezelhetetlenül nagy, de mivel úgyis csak a várható értékre vagyunk kiváncsiak ezért felírtunk a várható értékre egy alacsonyabb dimenziós differenciálegyenlet rendszert. Ezeknél azonban valamilyen közelítést kellett használnunk, tehát a pontos megoldás nem volt a kezünkben. Most bemutatunk egy módszert amely segítségével csökken az egyenletek száma és a kapott várható érték a pontos Kolmogorov rendszerből kapott várható értékkel egyezik meg. A módszer hátránya, hogy a csökkenés mértéke nagyban függ a gráfstruktúrától, azonban bizonyos speciális gráfok mint például az N csúcsú teljes gráf esetén nagy hatékonysággal alkalmazható.

Az módszer alapötlete az, hogy új változók bevezetésével csökkentjük az ismeretlenek számát, ezt hívják általánosan összevonásnak (angolul *lumping*). Pontosabban a régi változók valamely lineáris kombinációját fogjuk új változókkal leírni. A korábban tárgyalt $X^k = \left(X_1^k, X_2^k, \ldots, X_{c_k}^k\right)$ vektor elemeit helyettesítsük egyetlen x^k változóval a következőképpen

$$x^k = X_1^k + X_1^k + \ldots + X_{c_k}^k \quad \forall k = 1, 2, \ldots, N$$

ekkor x^k tulajdonképpen annak a valószínűségét jelenti, hogy k darab fertőzött van. Megjegyezzük, hogy ez a konkrét összevonás teljes gráf esetén lesz megfelelő, általános esetben a gráf automorfizmus csoportja adja meg a megfelelő változókat, bővebben [6]. A továbbiakban rögzítjük a csúcsszámot $N = 3$-ra és felírjuk a Kolmogorov illetve az összevonások után kapott egyenletrendszert. Számozzuk meg a csúcsokat 1-től 3-ig és legyen X_{ABC} annak az állapotnak a valószínűsége amikor

az első csúcs A, a második B illetve a harmadik C típusú, ahol $A, B, C \in \{S, I\}$. Például X_{ISI} az alábbi állapot valószínűségét jelenti 2.3 ábra.



2.3. ábra. ISI állapot, melynek valószínűsége X_{ISI}

Ekkor a Kolmogorov egyenletek

$$\begin{aligned}
\dot{X}_{SSS} &= \gamma(X_{SSI} + X_{SIS} + X_{ISS}) \\
\dot{X}_{SSI} &= \gamma(X_{SII} + X_{ISI}) - (2\tau + \gamma)X_{SSI} \\
\dot{X}_{SIS} &= \gamma(X_{SII} + X_{IIS}) - (2\tau + \gamma)X_{SIS} \\
\dot{X}_{ISS} &= \gamma(X_{ISI} + X_{IIS}) - (2\tau + \gamma)X_{ISS} \\
\dot{X}_{SII} &= \gamma X_{III} + \tau(X_{SSI} + X_{SIS}) - 2(\tau + \gamma)X_{SII} \\
\dot{X}_{ISI} &= \gamma X_{III} + \tau(X_{SSI} + X_{ISS}) - 2(\tau + \gamma)X_{ISI} \\
\dot{X}_{IIS} &= \gamma X_{III} + \tau(X_{SIS} + X_{ISS}) - 2(\tau + \gamma)X_{IIS} \\
\dot{X}_{III} &= -3\gamma X_{III} + 2\tau(X_{SII} + X_{ISI} + X_{IIS})
\end{aligned}$$

A fentebb tárgyalt összevonásokkal a fenti egyenletrendszert egyszerűsítsük, ekkor az alábbit kapjuk

$$\begin{aligned}
\dot{x}^0 &= \gamma x^1 \\
\dot{x}^1 &= -2(\tau + \gamma)x^1 + 2\gamma x^2 \\
\dot{x}^2 &= 2\tau x^2 - 2(\tau + \gamma)x^2 + 3\gamma x^3 \\
\dot{x}^3 &= 2\tau x^2 - 3\gamma x^3
\end{aligned}$$

ahol $x^0 = X^0$, $x^1 = X_1^1 + X_2^1 + X_3^1$, $x^2 = X_1^2 + X_2^2 + X_3^2$ és $x^3 = X^3$.

Általános csúcsszám esetén hasonló rendszert kapunk, erről szól a következő tétel.

2. Tétel. *Legyen G egy teljes gráf, ekkor az x^k függvények kielégítik az alábbi differenciálegyenleteket*

$$\begin{aligned}
\dot{x}^0 &= \gamma x^1 \\
\dot{x}^k &= (k-1)(N-k+1)\tau x^{k-1} - (k(N-k)\tau + k\gamma)x^k + (k+1)\gamma x^{k+1} \\
\dot{x}^N &= (N-1)\tau x^{k-1} - N\gamma x^N
\end{aligned}$$

$\forall k = 1, 2, \ldots, N-1$

Tehát a kezdetben 2^N méretű rendszer helyett elég megoldanunk egy $N+1$ méretűt. A keresett várható értéket ekkor könnyen megkaphatjuk, mégpedig a

$$[I](t) = \sum_{k=0}^{N} k x^k(t)$$

képletből adódik.

3. fejezet

Közelítések és szimulációk

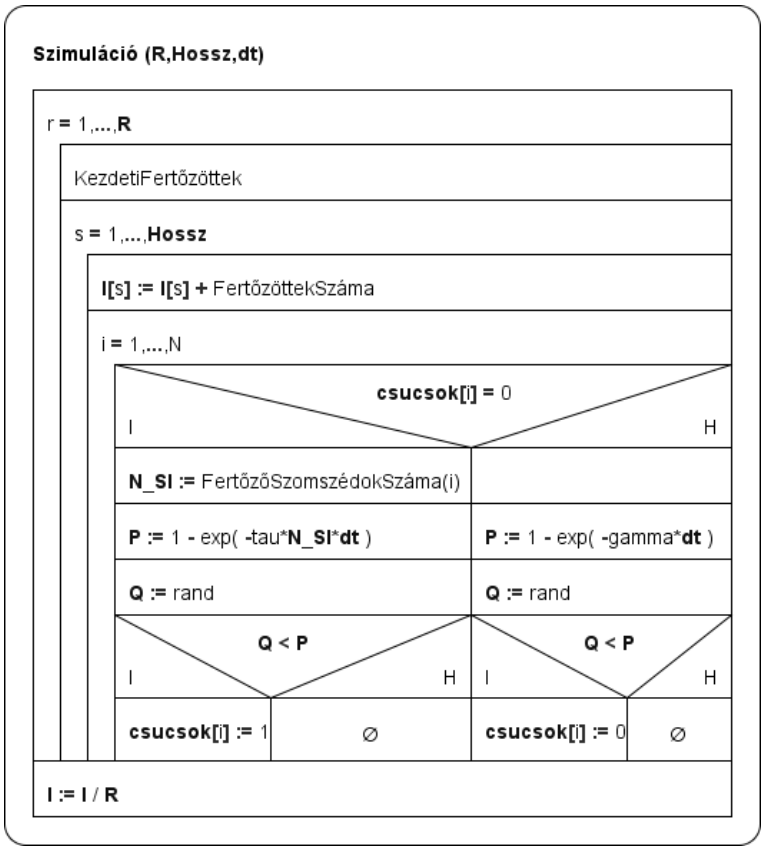
Ebben a fejezetben az elméleti alapok tárgyalása után a szimulációs módszert, illetve annak tesztelését fogjuk bemutatni. Először ismertetjük a szimuláció során használt Monte Carlo módszert és érintjük a konkrét szimulációs algoritmust. A kutatásunk alapjait képező szimulációs eredményekre csak a következő fejezetben térünk át, most kizárólag a teljes gráf esetére szorítkozunk amely megfelelő bevezetést nyújt a további fejezethez. Megjegyezzük, hogy elsősorban egy rövid és tömör áttekintést szeretnénk adni, ezért néhol a precíz megfogalmazások helyett inkább a szemléletes tárgyalást részesítjük előnyben.

3.1. Monte Carlo módszer és a szimulációs algoritmus

A Monte Carlo módszer egy szimulációs módszer, melyet régóta sok területen alkalmaznak véletlen kísérletek eredményeinek előállítására. Lényege, hogy egy véletlenszám generátor segítségével lejátszuk a kísérletben szereplő folyamatot, majd ezt sokszor megismételve az eredményeket átlagoljuk. A Monte-Carlo elnevezés arra utal, hogy a módszerben szereplő véletlenszámot akár egy kaszinó játék eredményeiből is vehetnénk. Az algoritmusunkban mindig a MATLAB beépített véletlenszám generátorát használtuk, tehát a `rand` parancsot amely gyakorlatilag a $[0, 1]$ intervallumon egyenletes eloszlású pszeudovéletlen szám. A pszeudovéletlen azt jelenti, hogy valamilyen determinisztikus algoritmusból származik, amely azonban közelít a

véletlenhez.

Most röviden tárgyaljuk a szimulációhoz használt algoritmust. Tekintsük a $[0, T]$ intervallum egy felosztását δt osztásközzel és jelöljük M-el az osztásközök számát. Egy kísérlet abból áll, hogy M-szer végig megyünk az egész gráfon, minden csúcsot megvizsgálunk és bizonyos valószínűséggel S állapotból I-re változtatjuk vagyis fertőzünk, illetve I-ről S-re vagyis gyógyítunk. Ez a következőt jelenti, vegyük először a fertőzés esetét. Kiszámítjuk egy adott csúcsra a $P = 1 - exp(-N_{SI}\tau\delta t)$ értéket, ahol N_{SI} jelöli az adott egészséges csúcshoz tartozó (S, I) élek számát, majd tekintsük a `rand` parancs segítségével kapott Q véletlen számot. Ha $Q < P$ akkor fertőzünk, vagyis I-re változtatjuk az állapotot, ellenkező esetben egészséges marad a csúcs. Gyógyulásnál hasonlóan járunk el, ekkor a $P' = 1 - exp(-\gamma\delta t)$ számot hasonlítjuk egy másik Q' véletlenszámhoz. Ha $Q' < P'$ akkor a csúcs meggyógyul, ellenkező esetben fertőzött marad. A fenti eljárást megismételjük összesen R-szer és végül átlagot veszünk. A konkrét algoritmus vázlata az alábbi, bemeneti paraméterek az R ismlétlési szám, a δt felosztás és az intervallum hossza, amit $Hossz$-al jelölünk.



3.2. Teljes gráf vizsgálata

Ebben a szekcióban végig rögzítjük a hálózat struktúráját, mindig az N csúcsú teljes gráfot fogjuk használni, melyet C_N-el fogunk jelölni. Ez a hálózat kutatási szempontból nem annyira érdekes, kizárólag csak a szimulációs algoritmus tesztelése céljából használjuk. A továbbiakban konkrét szimulációs eredményeket fogunk összehasonlítani, az előző fejezetben tárgyalt közelítésekkel, illetve a pontos megoldással. A pontos Kolmogorov megoldásból számított várható értéket azért tudjuk számolni, mivel a változó összevonásos módszer $N+1$-re csökkentette a differenciálegyenlet rendszer méretét, ez pedig már numerikusan kezelhető $100-200$ nagyságrendű csúcsszámok esetén.

Először felírjuk konkrétan a két közelítésből kapott rendszert, külön-külön. A *pair closure* típusú közelítés esetében a rendszer

$$[\dot{S}] = \gamma[I] - \tau[S][I] \tag{3.1}$$

$$[\dot{I}] = \tau[S][I] - \gamma[I] \tag{3.2}$$

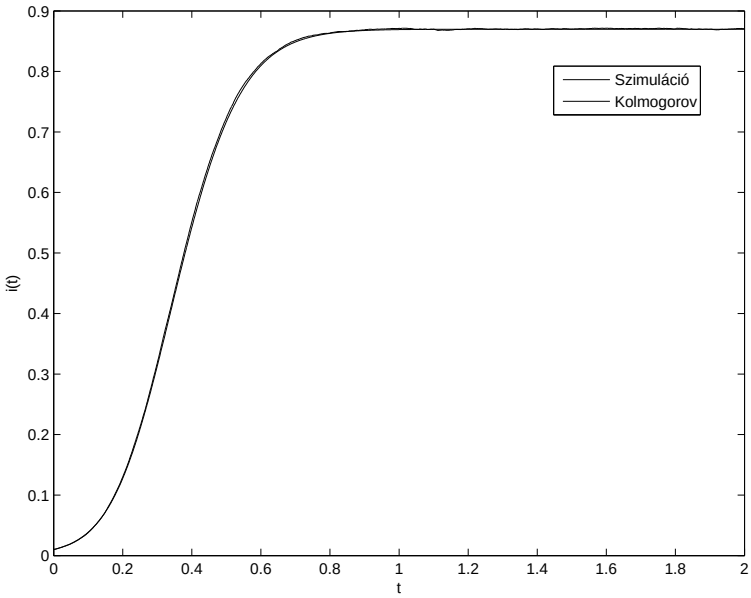
hiszen a fokszám $n = N-1$, így $\frac{n}{N-1} = \frac{N-1}{N-1} = 1$.

A *triple closure* közelítésben hasonlóan kiesik a fokszámot tartalmazó együttható a rendszer tehát

$$\begin{aligned}
[\dot{S}] &= \gamma[I] - \tau[SI] \\
[\dot{I}] &= \tau[SI] - \gamma[I] \\
[\dot{SI}] &= \gamma([II] - [SI]) + \tau\Big(\frac{n-1}{n}\frac{[SS][SI]}{[S]} + \frac{n-1}{n}\frac{[IS][SI]}{[S]} - [SI]\Big) \\
[\dot{II}] &= -2\gamma[II] + 2\tau\Big(\frac{n-1}{n}\frac{[IS][SI]}{[S]} - [SI]\Big) \\
[\dot{SS}] &= 2\gamma[SI] - 2\tau\frac{n-1}{n}\frac{[SS][SI]}{[S]}
\end{aligned}$$

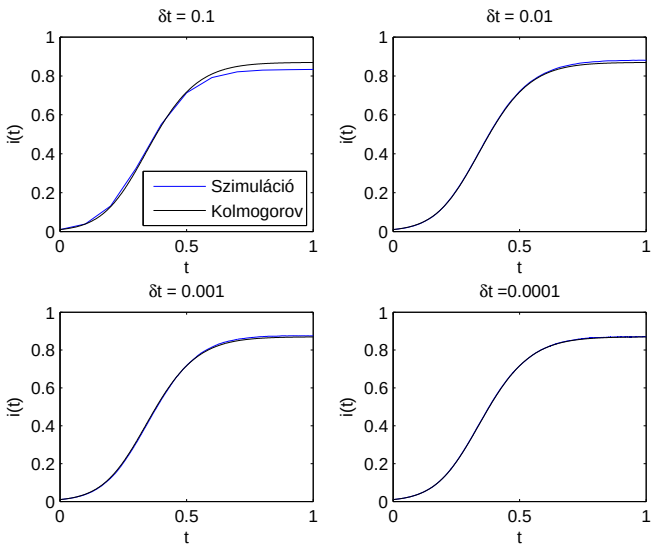
Ahhoz, hogy érdemben vírusterjedésről beszélhessünk kezdetben mesterségesen meg kell fertőznünk néhány csúcsot, hiszen ha mind S típusú lenne akkor ebben az állapotban maradna a folyamat. Kérdés, hogy mekkora legyen a kezdeti fertőzöttek száma, mint azt a későbbiekben látni fogjuk ez meghatározó szerepet játszik. A reális kezdő érték 1 darab fertőzöttől kezdve kb. a csúcsszám 5%-áig terjednek, az ennél nagyobb értékek nem túl reálisak különösen nagy csúcsszám esetén.

Most nézzünk, egy konkrét szimulációs eredményt. A csúcsszám legyen $N = 100$, a paraméterek $\beta = 15$ és $\gamma = 1$, a szimulációs paramétereket illetően legyen az ismétlési szám $R = 1000$, a vizsgált intervallum $[0, 2]$ és az osztásköz $\delta t = 0.001$ (3.1. ábra).



3.1. ábra. Fertőzöttek számának várható értéke, $N = 100$, $\beta = 15$ és $\gamma = 1$ esetén.

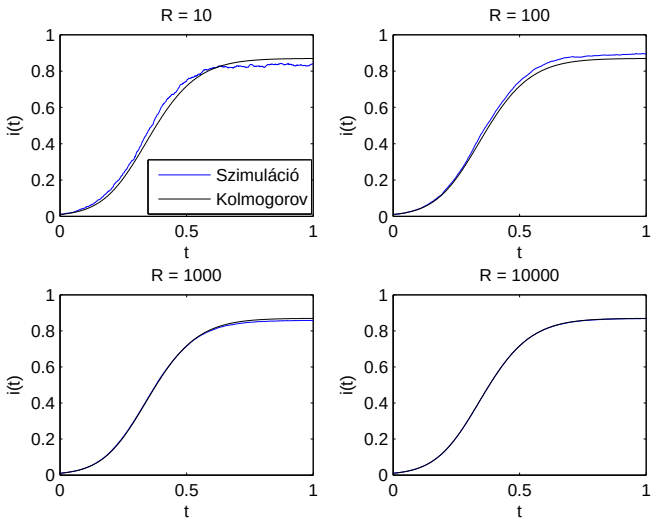
Azért választottuk ezen paramétereket, mert az így kapott görbe tipikus esete a terjedésnek. Látható, hogy van egy úgynevezett felfutási szakasz amely után beáll egy egyensúlyi állapot és gyakorlatilag konstans a függvény. A további vizsgálatokban az intervallumot mindig úgy fogjuk megválasztani, hogy ezt a felfutási szakaszt tartalmazza, utána a függvény úgyis konstans lenne. Megjegyezzük, hogy az SIS modellben a stacionárius állapot minden esetben az, hogy a vírus kihal vagyis mindenki egészséges, de ez általában csak nagyon hosszú idő után következik be, tehát a vizsgálatokban ezzel a szakasszal nem fogunk foglalkozni. Láthatjuk az ábrán, hogy a szimuláció teljesen visszaadta a Kolmogorov megoldást. Ez annak köszönhető, hogy mind az intervallum felosztását, mind az ismétlési számot jól választottuk. Nyilvánvaló, hogy minél nagyobb az ismétlési szám annál simább görbét kapunk, illetve minél kisebb a felosztás annál jobban közelítjük a folytonos Kolmogorov megoldást. Most nézzük meg mi történne alacsonyabb felosztás esetén, minden más paramétert változatlanul hagyva (3.2. ábra).



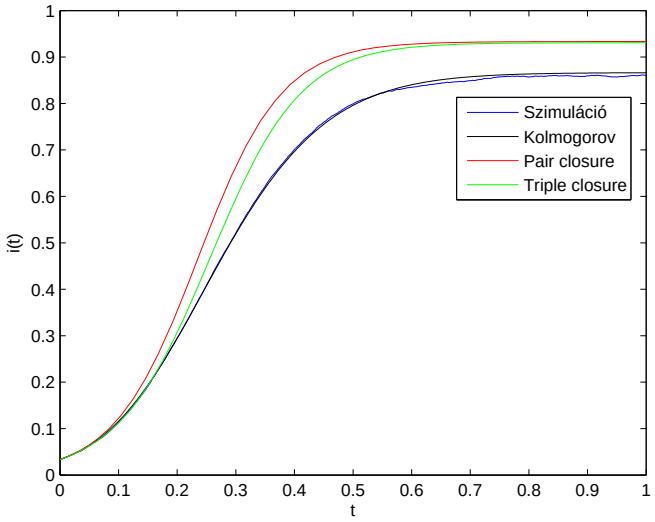
3.2. ábra. Fertőzöttek számának várható értéke, különböző δt felosztások esetén ($N = 100$, $\beta = 15$, $\gamma = 1$)

Azt látjuk, hogy $\delta t = 0.001$ illetve $\delta t = 0.0001$ esetén már nagyon közel van a két görbe. Mejegyezzük, hogy a túl nagy δt érték használata nem célszerű, mert ekkor a szimulációban egy adott lépésben több fertőzés, illetve gyógyulás történik mint amennyi reális lenne. A továbbiakban mindig a $\delta t = 0.001$ értéket fogjuk használni, így a futási idők 10-szeresen csökkennek a finomabb felosztáshoz képest. A másik fontos szimulációs paraméter a ismétlési szám. Most az előző vizsgálathoz hasonlóan rögzítünk minden paramétert és az ismétlési számot változtatjuk (3.3. ábra).

Nyilvánvalóan magasabb ismétlési számnál egyre jobbak a közelítések. Általánosságban elmondható, hogy az $R = 1000$ körüli ismétlési szám már magasnak számít, ezért mi is inkább ezt fogjuk használni. Megjegyezzük, hogy a szimulációs algoritmust módosíthatnánk úgy, hogy nem mi adjuk meg az R értékét, hanem egy úgynevezett tolerancia feltételt használunk a külső ciklusból való kilépésre. Ez azt jelenti, hogy ha a kiátlagolt görbe már alig változik, akkor leállítjuk a szimulációt, hiszen így már további ismétlésekkel sem tudunk jobb közelítést kapni. A továbbiakban tehát mindig az $R = 1000$ és $\delta t = 0.001$ paramétereket fogjuk használni. Most nézzük meg, hogy az előző szakaszban tárgyalt közelítések, hogyan viselkednek. A csúcsszám legyen $N = 30$, és a paraméterek $\tau = 15$ valamint $\gamma = 1$ (3.4.ábra).



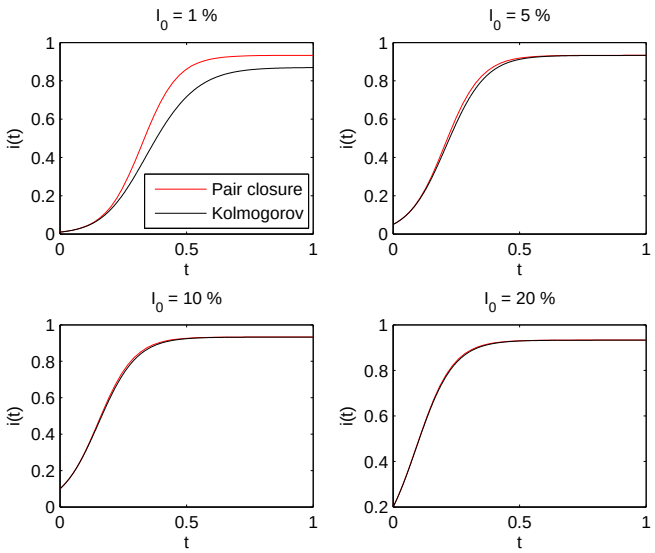
3.3. ábra. Fertőzöttek számának várható értéke, különböző R ismétlési számok esetén ($N = 100$, $\beta = 15$, $\gamma = 1$)



3.4. ábra. Közelítések összehasonlítása, $N = 30$, $\beta = 15$ és $\gamma = 1$ esetén.

A csúcsszámot azért csökkentettük, mert így látható az ábrán a különbség a két közelítés között, továbbá látható, hogy a *triple closure* van közelebb. Elmondható azonban, hogy a közelítések nem jók, akár a felfutási szakaszt, akár az egyensúlyi értéket nézzük. Korábban már említettük, hogy a kezdeti fertőzöttek száma komoly befolyással bír a közelítésre. Jelöljük I_0-al, hogy kezdetben a csúcsszám hány száza-

léka fertőzött, pontosabban legyen kezdeti fertőzöttek száma $\frac{I_0 N}{100}$ egészre kerekített értéke. Most az egyszerűség kedvéért csak a Kolmogorov megoldást illetve a *pair closure* közelítést vizsgáljuk különböző I_0 értékekre, a paramétereket nem változtatjuk az előzőekhez képest (3.5. ábra).

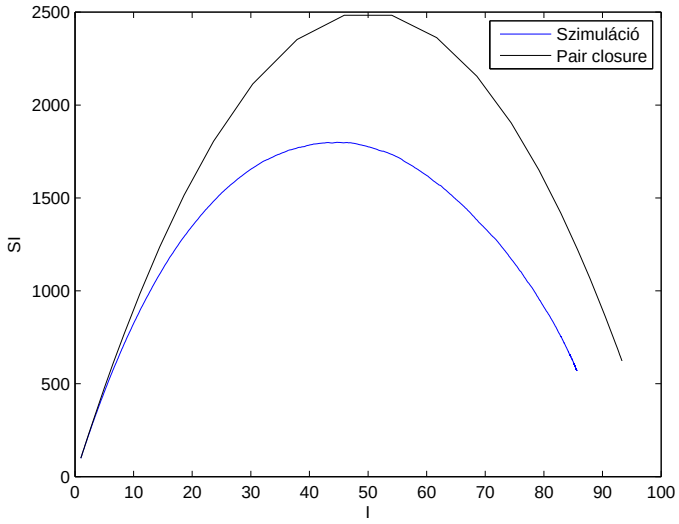


3.5. ábra. Fertőzöttek számának várható értéke, különböző I_0 kezdeti fertőzöttség esetén ($N = 100$, $\beta = 15$, $\gamma = 1$)

Látható, hogy például $I_0 = 10\%$-os értéknél már teljesen jó a közelítés, ennél nagyobb kezdeti fertőzött szám esetén úgyszintén. Megjegyezzük, hogy sokszor azonban az 1%-os vagy annál kisebb I_0 értékek a reálisak ezért a vizsgálatokban is ezeket fogjuk használni. A későbbi fejezetben tárgyaljuk, hogyan lehet ezen közelítéseket javítani úgy, hogy kisebb kezdeti fertőzöttség esetén is jók legyenek, elsősorban az egyensúlyi állapotot illetően. Innentől kezdve ebben a fejezetben végig 1 darab kezdeti fertőzöttel fogunk számolni, ha azt külön nem jelezzük.

A továbbiakban egy másfajta ábrát fogunk vizsgálni amelynek a későbbi fejezetben nagy figyelmet szentelünk. Az a cél, hogy valamennyire megértsük, hogy miért ad rossz eredményt a *pair closure* közelítés, illetve hol "romlik el". Az előzőekkel ellentétben most ne normáljuk a $[0, 1]$ intervallumba a fertőzöttek várható értékét, és jelöljük a *pair closure* közelítéssel kapott függvényt $\tilde{I}(t)$-vel, illetve a szimulációból kapott görbét $I(t)$-vel. Ahhoz, hogy szemléltessük a különbséget ábrázolni fogjuk az

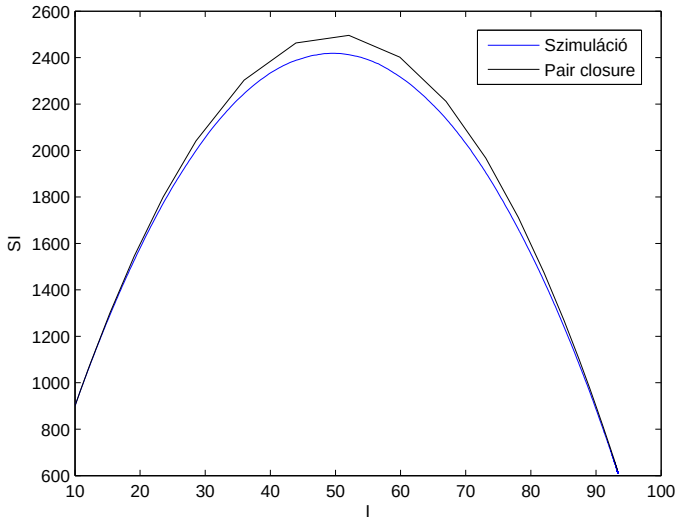
SI élek számát I függvényében, ehhez egy $(I(t), SI(t))$ ábrát készítünk. A közelítés esetében tudjuk, hogy az SI élek számát az $(N - \tilde{I})(\tilde{I})$ görbe írja le, hiszen ezt a közelítést alkalmaztuk, a másik esetben pedig a szimulációs algoritmus végrehajtása közben folyamatosan összeszámoljuk az SI élek számát. Minden paraméter megegyezik a korábbiakkal, a vizsgált intervallum a [0,1] továbbra is, ezt azért említjük meg mivel az ábrán az időfüggés nem látszik (3.6. ábra).



3.6. ábra. $(I(t), SI(t))$ ábra teljes gráf esetén, a paraméterek $N = 100$, $\beta = 15$, $\gamma = 1$, $I_0 = 1\%$

Látható, hogy a közelítés sokkal nagyobb élszámokat használ és ez a magyarázata, hogy miért kapunk nagyobb értékeket. Továbbá az is észrevehető, hogy az egyensúlyi pontban hasonló az SI élek száma, viszont más-más I értéknél. Egy lehetséges stratégia a jó közelítés megtalálására, hogy olyan $SI(t)$ görbéket próbálunk keresni melyek minél közelebb vannak a szimulációból számolthoz. Ezt jól szemlélteti, ha a kezdeti fertőzöttek számát 10%-ra vagyis 10-re növeljük. Láttuk, hogy ekkor a közelítés már nagyon jó, most megnézzük, hogy az $(I(t), SI(t))$ ábra mennyivel változik (3.7. ábra).

Azt kaptuk amire számítottunk, vagyis a görbék sokkal közelebb vannak mint a korábbi ábrán. Ez azért fontos információ mert az SI élek számát meghatározni sokszor, könnyebb akár kombinatorikai módszerek segítségével is.



3.7. ábra. $(I(t), SI(t))$ ábra teljes gráf esetén, a paraméterek $N = 100$, $\beta = 15$, $\gamma = 1$, $I_0 = 10\%$

4. fejezet

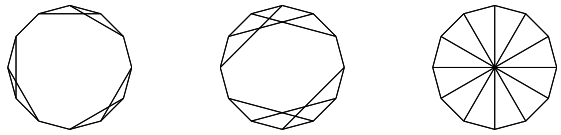
Vizsgálatok és eredmények

Jelen fejezetben tárgyaljuk a dolgozat fő témakörét, olyan új közelítési módszereket melyeket egy speciális gráfon végzett vizsgálatok motiváltak. Ezen speciális gráf az úgynevezett módosított körgráf melyet már a TDK dolgozatomban is vizsgáltam, azonban a szimulációs eredmények pontos magyarázata átfogóbb kutatási munkát igényelt. A célunk, hogy bemutassuk az eddig elért eredményeket megjegyezve, hogy néhány területen még akadnak kérdések, de reményeink szerint ezeket mihamarabb megválaszoljuk és a nyár folyamán témavezetőmmel közösen tudományos cikkünk születhet egy nemzetközileg elismert lapban.

4.1. Módosított körgráf vizsgálata

Ebben a szakaszban a hálózatunk mindvégig a már említett, úgynevezett módosított körgráf lesz. Először is ismertetjük, hogy pontosan mit értünk ezen hálózat alatt.

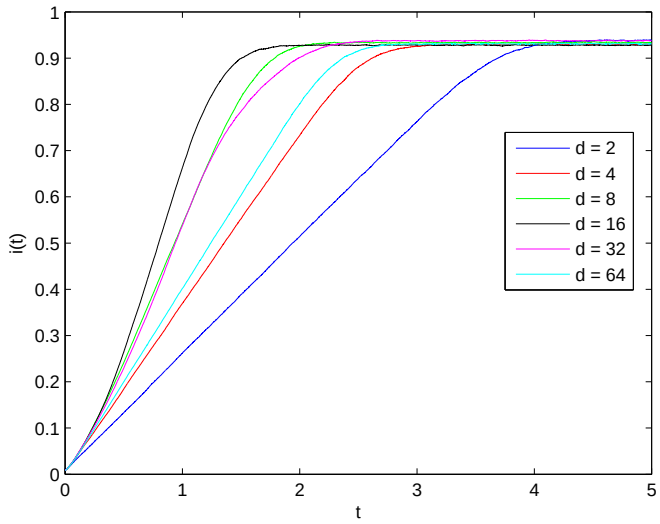
A konstrukció a következő: tekintsünk N csúcsot és kössük össze őket, hogy egy N hosszú kört alkossanak, majd ezek után létrehozunk úgynevezett klasztereket, azaz minden csúcsból a d távolságban lévő egyik szomszédjába élt húzunk úgy, hogy minden csúcs foka 3 legyen, majd az így kapott gráfot jelöljük $G_{N,d}$-vel. Nyilvánvaló, hogy az iménti eljárást csak akkor tudjuk megcsinálni, ha $2d$ osztója a csúcsszámnak. Az alábbi ábrán szemléltetés képpen mutatunk néhány konstrukciót különböző d értékekre, $N = 12$ csúcsszám mellett(4.1. ábra).



4.1. ábra. $G_{N,d}$ gráfok, $N = 12$ és $d = 2, 3, 6$ esetén

Felmerül a kérdés, hogy miért pont ezt a specális gráfot helyeztük vizsgálatunk középpontjába? Az a válasz, hogy ezen gráf hasonlít a $Watts - Strogatz$ véletlen gráfra melyről bővebben [7] dolgozatban olvashatunk. Hasonlóság alatt azt értjük, hogy a $Watt - Strogatz$ kontrukció egy az imént említett módosított körgráfból indul, majd bizonyos valószínűséggel élcserékkel folytatódik. Logikus tehát, hogy először az élcsere mentes hálózatot szeretnénk megérteni, majd később a bonyolultabb kontrukciókat. Amint azt a továbbiakban látni fogjuk már a módosított körgráf megértése sem triviális kérdés.

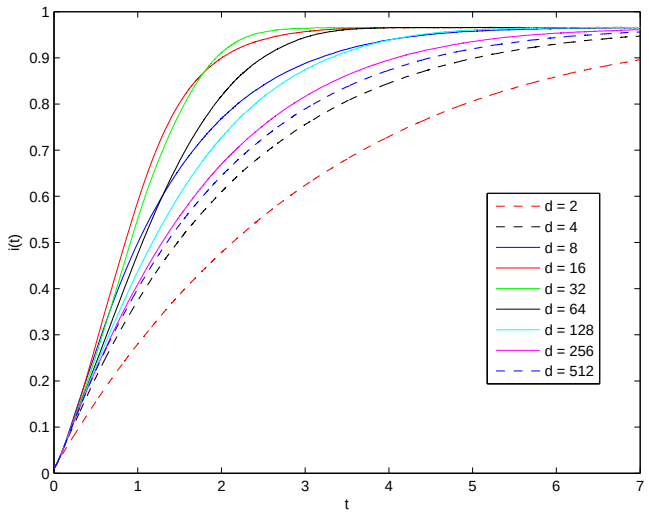
Most megnézünk néhány szimulációs eredményt, amelyek segítségével bemutatjuk a főbb felmerülő kérdéseket. Legyen a kezdetben fertőzött csúcsok száma 1 darab, ahogy a korábbi szakaszban is legtöbbször ezzel számoltunk. Továbbá a csúcsszám $N = 2^k$ valamely pozitív egész k-ra, ezen alak csak azért szükséges, mert ekkor a klaszterozás megvalósítható $d = 1, 2, \ldots, 2^{k-1}$ értékekre. Először futtassuk a szimulációt $N = 128$-ra és a paraméterek legyenek, $\tau = 10$ valamint $\gamma = 1$, ekkor az alábbi eredményeket kapjuk(4.2. ábra).



4.2. ábra. Fertőzöttek számának várható értéke $G_{128,d}$ módosított körgráfon, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ esetén

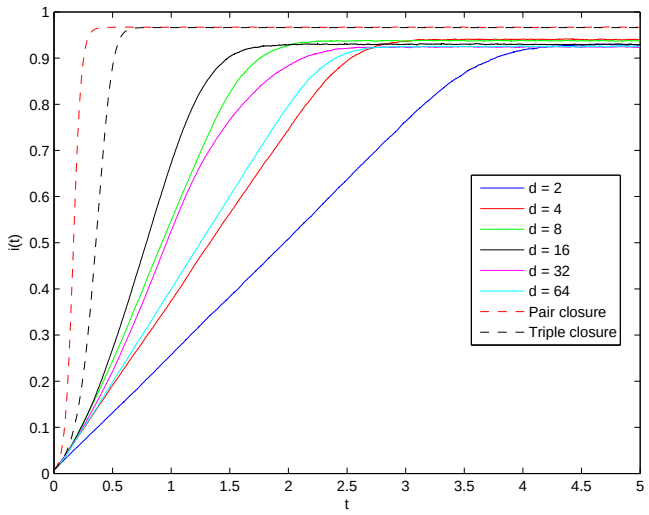
Azért választottuk ezeket a paraméterek, mert a fenti ábra jól mutatja, hogy tipikus esetben milyen a terjedés. Látható, hogy mindegyik görbe ugyanarra az egyensúlyi pontra áll be, viszont a kezdeti szakaszt követően jelentős különbségek vannak. A terjedés meredekségét (sebességét) illetően kézenfekvő lenne először arra gondolni, hogy nagyobb d érték esetén gyorsabban jut el a vírus nagyobb távolságra ezért ott nagyobb a terjedés sebessége is. Azonban a szimulációs tapasztalat ezzel szöges ellentétben áll, látjuk, hogy a 6 lehetséges d érték közül a legnagyobb, azaz a $d = 64$ csak a harmadik a sorban és $d = 16$ esetén a leggyorsabb. Célunk elsősorban ennek a jelenségnek a magyarázata, jelen fejezet nagy részét ennek szenteljük majd a továbbiakban, de először nézzük meg, hogy mi történik magasabb csúcsszám esetén. A következő ábra erejéig legyen a csúcsszám $N = 1024$ a többi paramétert hagyjuk változatlanul (4.3. ábra).

A tendencia hasonló, azaz a legnagyobb $d = 512$ érték itt is a harmadik leglassabb, őt követi a 256 illetve a 128. A fenti ábrát csak motiváció gyanánt néztük meg, most visszacsökentjük a csúcsszámot $N = 128$-ra és megnézzük, hogy a korábban tárgyalt közelítéseket, mennyire bizonyulnak hasznosnak.



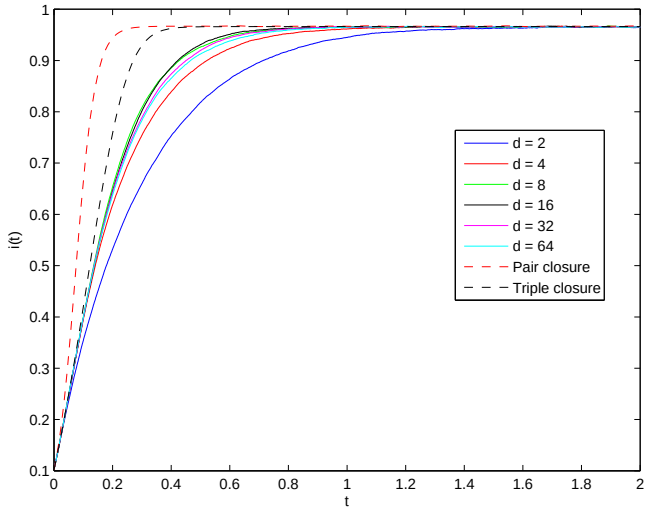
4.3. ábra. Fertőzöttek számának várható értéke $G_{1024,d}$ módosított körgráfon, $d = 2, 4, 8, 16, 32, 64, 128, 256, 512$, $\tau = 10$ és $\gamma = 1$ esetén

Az világos, hogy mivel ezen közelítések nem tartalmazzák a d paramétert, legfeljebb egy görbe felfutási szakaszát közelíthetik csak jól, azonban esetünkben ez sem áll fenn amit a következő ábra jól szemléltet (4.4. ábra).



4.4. ábra. Közelítések és a $G_{128,d}$ módosított körgráfon tekintett terjedések, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ esetén

Látható, hogy sem a *pair closure* sem a *triple closure* közelítés nem ad jó eredményt, az egyensúlyi értéket illetően sem. Az előző szakaszban láttuk, hogy ha növeljük a kezdeti fertőzöttek számát, akkor jelentős javulást tapasztalunk. Legyen most $I_0 = 10\%$, minden más paramétert hagyjunk változatlanul és nézzük meg mit tapasztalunk (4.5. ábra).



4.5. ábra. Közelítések és a $G_{128,d}$ módosított körgráfon tekintett terjedések, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ valamint $I_0 = 10\%$ esetén

Hasonlóan mint teljes gráf esetén, itt is jelentős javulást tapasztalunk, de csak az egyensúlyi pontot illetően, a kezdeti szakasz közelítése továbbra is rossz.

4.2. Új módszerek

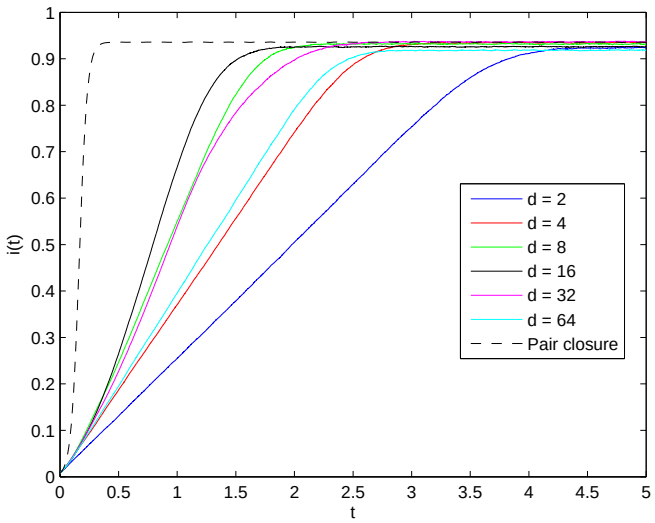
4.2.1. A *pair closure* javítása

Az előbb láttuk, hogy a *pair closure* közelítés mindig a szimulációból kapott görbe felett volt, akár a felfutási szakaszt akár az egyensúlyi állapotot nézzük, mitöbb így volt ez a teljes gráf esetében is. Viszont, ha növeltük a kezdeti fertőzöttek számát, akkor javult a közelítés. Pontosabban a szimulációs eredmények közeledtek a *pair closure* görbéhez, mivel a *pair closure* egyensúlyi állapota független az I_0 értéktől. Az a magyarázat, hogy kis kezdeti fertőzöttség esetén nagyobb valószínű-

séggel előfordulhat, hogy a vírus rövid időn belül kihal és nincs terjedés, viszont ha több kezdetben a fertőzött csúcs akkor ennek a kihalásnak az esélye sokkal kisebb. Tehát röviden az a *pair closure* hibája ebben az esetben, hogy nem veszi figyelembe a gyors kihalásokat. Ez a megfigyelés lehetőséget ad arra, hogy kis I_0 értékek esetén is jó közelítéseket kapjunk, legalábbis az egyensúlyi értéket illetően. Az az ötletünk, hogy a közelítésből kapott görbét egy 1-nél kisebb konstans értékkel megszorozzuk, így közelebb lesz a szimulációs eredményhez. Jelöljük a kezdeti fertőzött (egyetlen) csúcsot i-vel. Ennek a fertőzött csúcsnak 3 egészséges szomszédja van, tehát 3τ rátával adhatja át a fertőzést és γ rátával gyógyulhat. Ez azt jelenti, hogy a vírus kihalásának az esélye $\frac{\gamma}{3\tau+\gamma}$ és a tovább terjedés valószínűsége $\frac{3\tau}{3\tau+\gamma}$, azaz ötletünk, hogy ezekkel a valószínűségekkel súlyozzuk a kihalás illetve tovább terjedés esetét. Jelöljük $I_m(t)$-el a *pair closure* közelítésből kapott görbét. Ezen jelöléssel a fertőzöttek száma $\frac{3\tau}{3\tau+\gamma}I_m + \frac{\gamma}{3\tau+\gamma}0$, hiszen kihalás esetén nulla a fertőzöttek száma. Tehát az új közelítésünk az alábbi

$$\frac{3\tau}{3\tau+\gamma}I_m$$

Nézzük meg, hogy az így kapott görbe mennyit javult, ehhez minden korábbi paramétert változatlanul hagyunk (4.6. ábra).

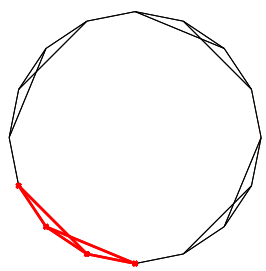


4.6. ábra. A javított *pair closure* és a $G_{128,d}$ módosított körgráfon tekintett terjedések, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ valamint $I_0 = 10\%$ esetén

Látható, hogy az egyensúlyi szakaszt teljesen jól közelíti, azonban a felfutási szakaszt továbbra is rosszul. Megjegyezzük, hogy nagyobb γ gyógyulási ráta esetén előfordulhat, hogy ez sem kompenzálja eléggé a kihalásokat például ha a szimulációból kapott eredményben teljesen kihal a vírus attól függetlenül a fenti közelítés nem fog nullát adni, egyszerűen csak egy kis konstans értékkel fogunk szorozni. Továbbá nagyobb γ esetén érdemes további valószínűségekkel súlyozni, tehát azt is figyelembe venni, hogy mekkora a későbbi kihalások esélye.

4.2.2. Új közelítő differenciálegyenlet

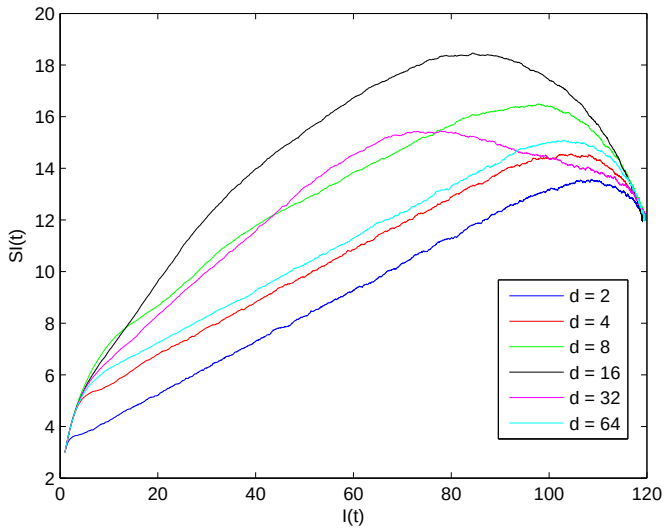
A továbbiakban egy új megközelítést tárgyalunk a gráfstruktúrát illetően, mely közelebb visz minket a vírusterjedések heurisztikus magyarázatához. Az alapgondolat visszavisz minket a módosított körgráf szerkezetéhez. A gráf konstrukciója során a kör elkészítése után elkezdünk további éleket behúzni a d távolságban lévő jobb oldali szomszédokhoz, mégpedig úgy, hogy a gráf fokszáma minden csúcs esetén 3 legyen. Ekkor létrejönnek $2d$ pontból álló csoportok. Gondoljunk mostantól a gráfra úgy, hogy ilyen csoportokat kötünk össze egy-egy éllel és jelölje $M = 2d$ az egy csoportban lévő pontok számát, továbbá $m = \frac{N}{M}$ a csoportok számát, ahol N a csúcsszám. Az előző gondolatmenetet az alábbi ábra jól példázza (4.7. ábra)



4.7. ábra. $G_{16,2}$ módosított körgráf, kiemelve az egyik $M = 4$ méretű csoport

Az nyilvánvaló, hogy a $d = 1$ esetben terjed a vírus a leglassabban, tehát az egyszerű N csúcsú körgráfon, ekkor minden csúcs foka 2. A $d > 1$ esetre gondolhatunk úgy, hogy az előbb említett M méretű csoportoknak egy körgráfja. Ha a d érték kicsi akkor ezek a csoportok hasonlóan viselkednek a terjedés szempontjából mint

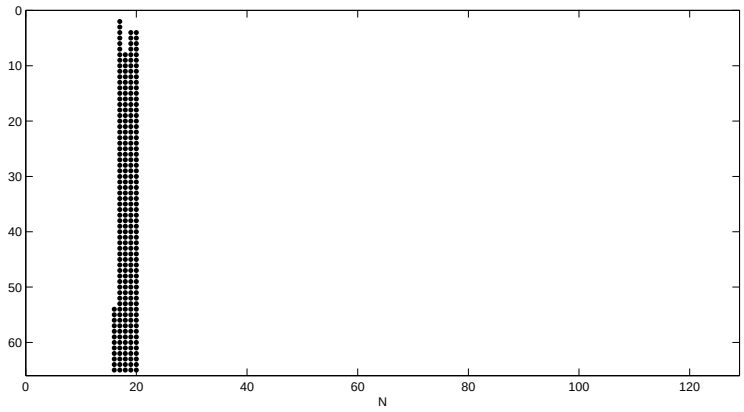
egyetlen csúcs. Mindezt egybevetve magyarázható, hogy miért a $d = 2, 4$ esetben a leglassabb a terjedés. Az igazi kérdés az, hogy miért a legnagyobb d érték követi őket a sorban. Ennek megértéséhez az előző fejezetben használt $(I(t), SI(t))$ ábra hasznos lesz. Nézzük meg mit látunk különböző d értékek esetén, az egyszerűség kedvéért nem normáljuk sem az csúcsokat, sem az élszámokat (4.8. ábra).



4.8. ábra. $(I(t), SI(t))$ görbék $G_{128,d}$ módosított körgráfon, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ esetén

Látható, hogy terjedések sorrendjét teljes mértékben visszaadja a fenti ábra. Mitöbb a probléma visszavezethető, az SI élek számának meghatározására, azaz ha valamely d érték esetén nagyobb az SI élek száma, akkor gyorsabb a terjedés is. Kérdés, hogy ezen élek számát hogyan tudnánk valahogy közelíteni figyelembe véve a d paramétert. Mindenek előtt be kell vezetnünk a frontok fogalmát és egyben megjegyezzük, hogy a front elnevezés nem egységes de rendkívül szemléletes. Képzeljük el, hogy a gráf N csúcsa egy kört alkot, amikor a vírusterjedés elindul egy pontból akkor a vírus átterjedhet a bal vagy jobb oldali szomszédjára illetve a d távolságban lévő szomszédhoz. Ha valamelyik oldalirányú szomszédot fertőzi meg, akkor létrejön egy szakasz melyben mindenki fertőzött és ez a szakasz egyre csak hosszabb lesz. Előfordulhat persze, hogy néhány csúcs meggyógyul és ekkor a szakaszon belül lesz pár csúcs ami egészséges vagy ha szélső csúcs gyógyul akkor rövidebb lesz, de ti-

pikusan ez a körön egy ív mely fertőzöttekből áll néhány egészségestől eltekintve. Egy ilyen ívet fogunk a továbbiakban frontnak nevezni. Az alábbi ábrán egy front kialakulását látjuk (4.9. ábra).

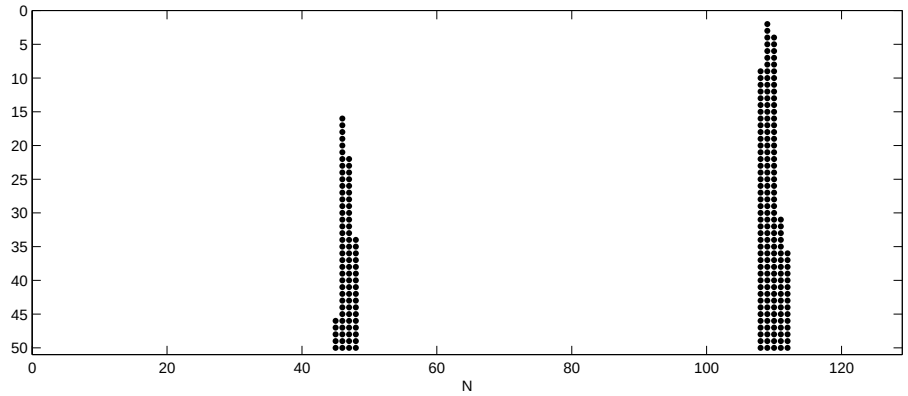


4.9. ábra. Példa egy adott front kialakulására

A vízszintes tengely a csúcsokat a függőleges az időt jelenti és ahol fekete pont látható, az a csúcs fertőzött. A fenti ábra egy konkrét szimuláció eredményét mutatja, a paraméterek $\tau = 20$, $\gamma = 1$ és a vizsgált intervallum a $[0, 1]$ az osztásköz pedig $\delta t = 0.001$.

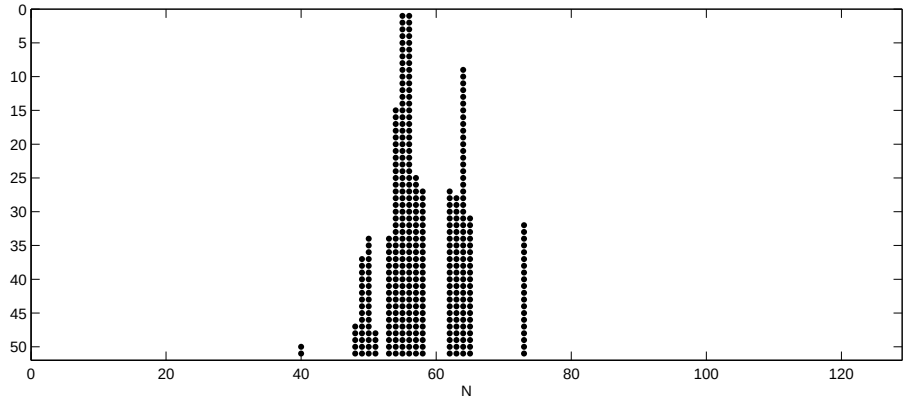
Ha az előző körgráfos gondolatmenetet egybevetjük a frontokkal, akkor megállapítható, hogy kis d értékekre csak egy front indul meg a kör valamely részén. Ez a front fog teljesen szétterjedni a gráfon, hiszen második front kialakulása csak d távolságra lenne lehetséges, de ekkor a két front között keletkező kis hézag gyorsan megtelik fertőzöttekkel. Nézzük meg mi történik $d = 64$ esetén, (4.10. ábra).

Azt látjuk, hogy itt rövid időn belül 2 darab front keletkezik. A magyarázat abban rejlik, hogy a harmadiknak behúzott élek pont a szemközti csúcsokat kötik össze vagyis rögtön a kör szemközti részére átterjed a fertőzés és ott egy másik front kezd kialakulni. A két front közötti távolság a lehető legnagyobb, így sok időbe telik amíg összeérnek. Fontos megállapítás, hogy harmadik front azonban nem alakulhat ki, ugyanis erre csak úgy lenne esély ha a szemközti csúcsokat összekötő élen történne fertőzés, de ekkor ez pont a másik a fronthoz tartozó csúcsot fertőzne. Összefoglalva a legnagyobb d érték esetén 2 darab egymással szemközti front kezd el terjedni.



4.10. ábra. Frontok kialakulása, $G_{128,64}$ módosított körgráf $\tau = 20$ és $\gamma = 1$ esetén

Az eddig vizsgált d-k esetén láttuk, hogy 1 vagy 2 front kialakulására van esély. A középső $d = 8$, illetve $d = 16$ értékeknél azonban sokszor előfordulhat, hogy 3 front alakul ki és elsősorban ez az oka, hogy a terjedés gyorsabb. Nézzük a $d = 8$ esetet, a másik ehhez nagyon hasonló ábrát ad. A helytakarékosság kedvéért csak azt a szakaszt mutatjuk, ahol elkezdenek kialakulni a frontok (4.11. ábra).



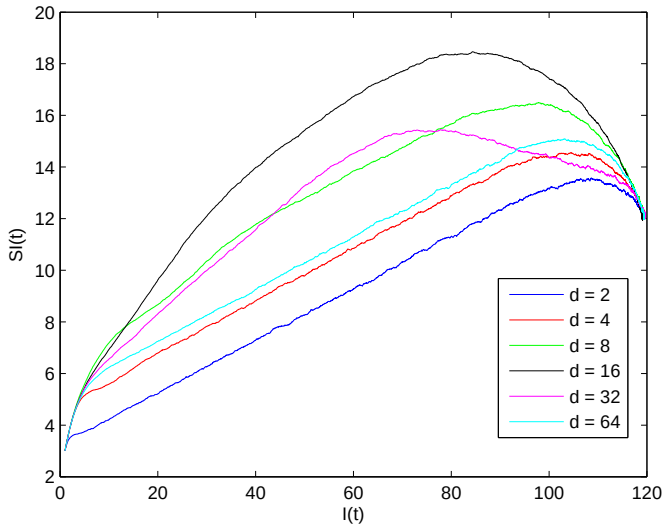
4.11. ábra. Frontok kialakulása, $G_{128,8}$ módosított körgráf $\tau = 20$ és $\gamma = 1$ esetén

A magyarázathoz szét kell választanunk az eseteket aszerint, hogy a kezdeti fertőzött csúcs a csoport szélén, vagy közepén helyezkedik el. Ha a fertőzés eleinte a csoport szélén van, akkor onnan könnyen fertőzi a csoport közepét tehát kialakul középen is egy front, majd innen a másik szélre tevődik át a terjedés ezzel létrehozva a harmadik frontot. Precízebben fogalmazva mindig a csoport első, d-edik illetve $2d$-edik csúcsa körül alakulnak ki frontok. Ha középtájékon helyezkedik el a kezdeti fertőzött akkor ugyanez lesz a helyzet nyilvánvalóan. Ha viszont nem a szélén és nem a közepén hanem a kettő között helyezkedik el a fertőzött csúcs, akkor nem alakul ki 3 front viszont a csoportok közötti fertőzés mindig szélső csúcsok között megy végbe. Tehát ahogy egy második csoportba is megjelenik a fertőzés úgy ott 3 front fog kialakulni és így tovább.

Miért fontos a frontok száma? Az a válasz, hogy alapvetően annál több az SI élek száma, minél több diszjunkt front van. Ezt elsősorban a szimulációs tapasztalatok illetve a heurisztikus érvelés mutatja. Megjegyezzük, hogy a kérdés ennél bonyolultabb ugyanis a frontok hossza és elhelyezkedése befolyásolja ezen élek számát, azonban ennek tisztázása még bizonyos esetekben nem készült el, ezért nem tárgyaljuk. A célunk elsősorban az volt, hogy bemutassuk a szemléletet melynek segítségével magyarázható a konkrét jelenség.

Végül a fejezet zárásaként bemutatunk egy újfajta közelítési módszert, melynek az az alapötlete, hogy az $(I(t), SI(t))$ görbét közelítjük. A korábbiakban már láttuk a 4.12 ábrát és most az a cél, hogy valami hasonlót kapjunk. Numerikus tapasztalatok arra engednek következtetni, hogy a vírusterjedés egy bizonyos kezdeti szakasz után közel lineáris. Ehhez idézzük fel az imént említett 4.12 ábrát.

Az a megfigyelésünk, hogy kezdetben egy vonalba esnek a görbék a d értéktől függetlenül, majd utána elkezdenek sorra leszakadni. Ezt követi egy lineáris szakasz, megjegyezzük, hogy az ábrán az időfüggést nehéz kiolvasni, de mindegyik d értéknél megfigyelhető egy lineáris szakasz csak néhol rövidebbnek tűnnek. Végül sorra elérik az egyensúlyi állapotot és ott találkoznak. Adódik az ötlet, hogy bontsuk 3 részre a közelítő görbét. Az első rész tartalmazza a közös vonalat, ahol együtt megy mindegyik görbe, a második a már említett egyenes szakasz, végül pedig az egyensúlyi ponthoz való visszatérés. Legyenek mindhárom fázisban egyenesek.



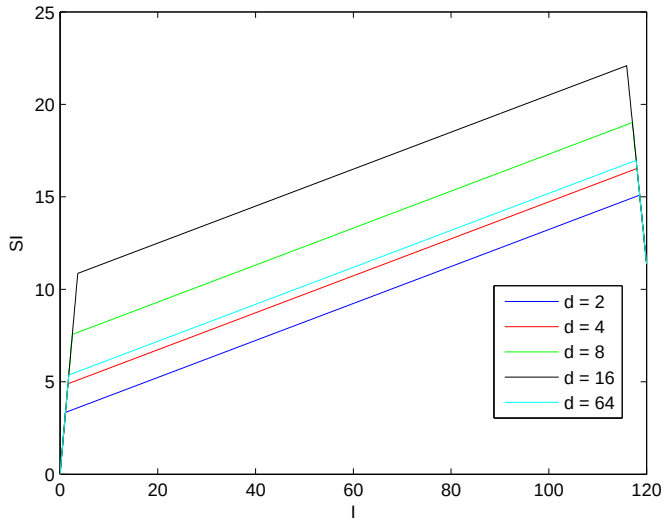
4.12. ábra. $(I(t), SI(t))$ görbék $G_{128,d}$ módosított körgráfon, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ esetén

Az első egyenes egyenlete legyen $SI = 3I$, ennek az a magyarázata, hogy ha kezdetben 1 darab fertőzöttünk van, akkor 3 darab SI élünk lesz, így az egyenes meredeksége adott. A második egyeneshez induljunk ki az $[\dot{I}] = \tau[SI] - \gamma[I]$ egyenletből. Ezt átrendezve

$$[SI] = \frac{[\dot{I}] + \gamma[I]}{\tau}$$

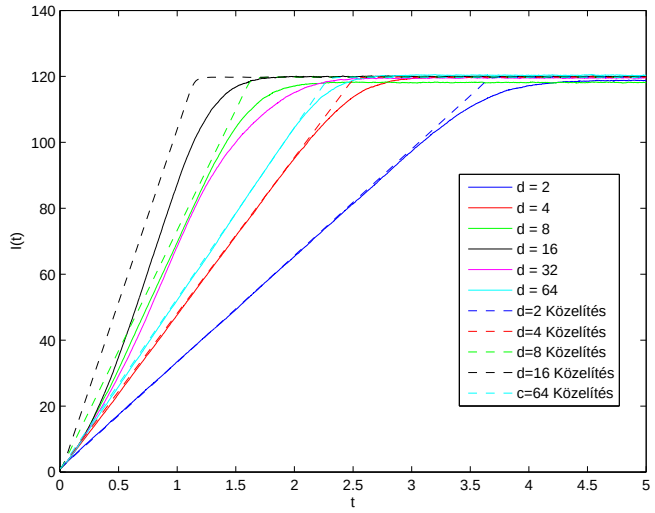
adódik, tehát a meredekség $\frac{\gamma}{\tau}$ és a konstans $\frac{[\dot{I}]}{\tau}$. Ezt a konstans értéket természetesen nem tudjuk, ezért ezt a szimulációs eredményekből közelítjük. Fontos megjegyezni, hogy itt feltételezésünk szerint a meredekség közel konstans. Végül az utolsó egyenes meredeksége egyezzen meg a *pair closure* közelítésből kapott meredekséggel, és természetesen minden egyenes folytonosan illeszkedjen. A következő ábra mutatja az így kapott görbéket (4.13.ábra).

Megjegyezzük, hogy $d = 32$ esetében a deriváltat nehéz egy konstans értékkel közelíteni, ezért nem ábrázoljuk. A tapasztalatok szerint magasabb csúcsszám esetén is megállja a helyét az az érvelés, hogy a középső szakaszon közel lineáris a terjedés. Kutatómunkánk következő fázisában szeretnénk az $[\dot{I}]$ mennyiséget szimulációk nélkül közelíteni a terjedés ezen szakaszán, jelenleg úgy tűnik, hogy ez a frontok módszerével megvalósítható.



4.13. ábra. $(I(t), SI(t))$ görbék közelítése $N = 128$, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ esetén

Legvégül megnézzük, hogy ha az iménti ábrán látható görbéket tekintjük az $[\dot{I}] = \tau[SI] - \gamma[I]$ differenciálegyenletünkben az $[SI]$ mennyiségeknek, akkor milyen közelítést kapunk (4.14. ábra).



4.14. ábra. Újfajta közelítések és a $G_{128,d}$ módosított körgráfon tekintett vírusterjedések összehasonlítása, $d = 2, 4, 8, 16, 32, 64$, $\tau = 10$ és $\gamma = 1$ esetén

Látható, hogy a kisebb d értékekre jobb a közelítés, mint a nagyobbak esetében. Viszont akár a *pair closure*, vagy akár a *triple closure* közelítést vesszük alapul, a javulás jelentős. Ezzel a gondolattal és egy rövid összefoglalással zárjuk a dolgozatot.

4.3. Összefoglalás

Ebben a fejezetben a módosított körgráfnál tapasztalt jelenséget szerettük volna megérteni, miszerint a d paraméter függvényében nem monoton változnak a terjedési sebességek. Láttuk, hogy a szakirodalomban elterjedt közelítések nem adnak jó eredményt, azonban a *pair closure* közelítés esetében konkrét javaslatunk volt annak javításra. Ez azonban elsősorban csak az egyensúlyi érték megállapításában segített a felfutási szakasz közelítését nem javította jelentősen. A front mint szemléletes fogalom bevezetése után, pedig lehetőség nyílt magyarázatot adni a jelenségre, mi több ezen fogalom szoros kapcsolatban áll az SI élek számával ezért a későbbi kutatások során, még fontos szerepet játszhat. Végül egy új ötlet segítségével konkrét közelítési eljárást dolgoztunk ki, melyet kézzel fogható eredményeket szolgáltattunk, megjegyezve, hogy a jövőben még tökéletesíteni szeretnénk.

Köszönetnyilvánítás

Ezúton szeretném megköszönni témavezetőmnek, Simon Péternek, hogy felhívta figyelmemet erre az érdekes témára és, hogy mindig tudott rám időt szakítani, akár külföldön, akár itthon volt. Építő ötleteivel mindig tudott segíteni, és végig támogatta a kutatómunkámat. Végül, de nem utolsó sorban szeretném megköszönni a családomnak, különösen Édesanyámnak, aki egész életemben mindenben támogatott és támogat mai napig.

Irodalomjegyzék

[1] Albert, R., Barabási, A. L. (2002) Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47 – 97.

[2] Keeling, M.J. (1999) The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. Lond. B 266, 859 – 867.

[3] Keeling, M.J., Rand, D.A. & Morris, A.J. (1997) Correlation models for childhood epidemics. Proc. R. Soc. B 264, 1149 – 1156.

[4] Kurtz, T.G. (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. J. Appl. Prob. 7, 49 – 58.

[5] Kurtz, T.G. (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. J. Appl. Prob. 8, 344 – 356.

[6] Simon, P.L., Taylor, M., Kiss., I.Z., (2010) Exact epidemic models on graphs using graph- automorphism driven lumping, J. Math. Biol., 62, 479 – 508.

[7] Watts, D.J., Strogatz, S.H. (1998). Collective dynamics of small-world networks. Nature 393 (6684): 409