

Állatorvostudományi Egyetem

Ökológia tanszék

A szimbiózis megjelenése és dinamikája katalitikus gráfokként modellezett szintróf  
mikrobiális közösségekben

**Készítette:** Máté Jakab

**Témavezetők:** Dr. Boza Gergely, Dr. Zachar István

Ökológiai Kutatóközpont, Evolúciótudományi Intézet

**Belső konzulens:** Dr. Kövér Szilvia

ÁTE, Ökológia Tanszék, tudományos munkatárs

## Tartalomjegyzék

1. Bevezetés.....	3
2. Irodalmi áttekintés.....	4
2.1. Mikrobiális közösségek interakciói.....	4
2.2. Az autokatalitikus halmaz.....	5
2.3. A Jain-Krishna modell bemutatása.....	7
3. Anyag és módszer.....	9
3.1. A Python programnyelv és a felhasznált könyvtárak bemutatása.....	9
3.2. A Jain-Krishna modell működése.....	10
3.3. Ciklusok keresése.....	12
3.4. A Jain - Krishna modell átalakítása.....	15
3.5. A munka menete.....	18
4. Eredmények.....	20
4.1. A Jain - Krishna modell rekonstrukciója.....	20
4.2. A rekonstruált Jain - Krishna modell vizsgálata.....	21
4.3. A bővített modell vizsgálata.....	26
5. Következtetések.....	32
5.1. A JK modell eredményeinek kiértékelése.....	32
5.2. Az bővített modell eredményeinek kiértékelése.....	33
5.3. Tervek és javaslatok.....	34
6. Összefoglaló.....	35
7. English title and summary.....	36
8. Hivatkozások.....	37
9. Melléklet.....	40
10. Köszönetnyilvánítás.....	41
11. Nyilatkozatok.....	42

## 1. Bevezetés

A matematikai modellezés célja a környezetünkben zajló folyamatok megértése. Segítségével reprodukálhatjuk a vizsgált jelenség mögötti alapvető mechanizmust, ami alapján információkat nyerhetünk a zajló és már lezajlott eseményekről, valamint predikciókat tehetünk a jövőre nézvést. A modellezés legnagyobb erénye az egyszerűsítés, hiszen a bonyolultabb rendszerek megértéséhez gyakran kénytelenek vagyunk olyan feltételezéseket tenni, amik bár nem fedik le teljes mértékben a valóságot, de megfelelő mértékben megközelítik azt. Azonban pontosan ezért is kell egy bizonyos fenntartással kezelni a modellek alapján megfogalmazott eredményeket és predikciókat, hiszen a valóság egy elnagyolt másából nyernek információt, és előfordulhat, hogy nem vesznek figyelembe egy-egy releváns körülményt. Mára az életünk számos területén hagyatkozunk matematikai modellek predikcióira, ilyen például a járványkezelés [1] vagy az időjárás-előrejelzés [2].

Az evolúcióbiológia tudományterületén azért bírhat különösen nagy jelentőséggel a modell alapú megközelítés, mert a korai evolúciós eseményekről szinte semmit sem tudunk biztosan [3, 4], viszont rengeteg különböző elmélet fogalmaz meg feltételezéseket az események lezajlására vonatkozóan [5–7], ezeket a feltételezéseket pedig - a fokozatosan előkerülő új bizonyítékok mellett [8] - modellek segítségével is lehet vizsgálni, egymáshoz képest valószínűsíteni. Ez a megközelítés még gyerekcipőben jár, de már kezd kirajzolódni, hogy milyen komplexebb modellekre volna szükség.

Kutatásom során a korai evolúció egyik folyamatát modelleztem: a mitokondrium (és az eukarióta sejtek) kialakulásának szintróf elméletét, Jain és Krishna [9] modelljére építve. A legalább 1.45 milliárd évvel ezelőtt létrejött mitokondrium [10] kialakulásának szintróf elmélete szerint ez a kompartment két prokarióta sejt kölcsönösen előnyös metabolikus kapcsolatán alapuló egyre erősödő függés eredményeként alakulhatott ki [11]. Az elmélet szerint a metabolikus függés a szintrófia jelenségén keresztül valósulhatott meg, ami a prokarióta közösségek reakcióútjainak összekapcsolódását jelenti, ezt a későbbiek során részletesebben kifejtem. Egy ilyen közösségben a mikróbák közötti anyagtranszferen kívül számos más kölcsönhatás is megfigyelhető. Munkám során ezeket a hatásokat összegezve, a szintróf mikrobiális közösségeket katalitikus gráfokként modelleztem. Bár Krause a Jain-Krishna modell adaptációjában [12] említette, hogy a modellben szereplő nóduszokra kémiai és biológiai fajokként is tekinthetünk, kutatásomban a mikrobiális közösségek és a kémiai rendszerek között húzott párhuzam az irodalomban újdonságnak számít.

A dolgozatomban először bemutatom a szintrófia jelenségét, és a prokarióta közösségek egyéb jellemző interakcióit, majd ismertetem a katalitikus rendszerekben szoros függés kialakulásához vezető struktúrákat, a gráfelméleti munkákra támaszkodva. A következő szekcióban ismertetem a munkám alapjául szolgáló modell működését, majd az általam továbbfejlesztett és módosított modellt, és az elvégzett módosítások jelentőségét is. A modell mutatóit különböző paraméterek mellett megvizsgálva értelmezem az eredményeket. Ezek után diszkusszióban ismertetem az eredmények fontosságát és hatását az eukarióta-eredet témakörében és végül javaslatokat fogalmazok meg saját jövőbeni kutatásomra vonatkozólag.

## **2. Irodalmi áttekintés**

### **2.1. Mikrobiális közösségek interakciói**

Egy mikrobiális közösség tagjai számtalan módon lehetnek hatással egymás működésére [13, 14]. A modellünk építése során azt feltételezzük, hogy ezeknek a hatásoknak létezik egy eredője, amivel számosítható, hogy az egyik faj egyedeinek jelenléte összességében semleges, antagonisztikusan vagy katalitikusan hat-e a másik faj egyedeinek növekedésére. Ezen feltételezés alátámasztásaként ebben a szekcióban bemutatom a baktériumpopulációk tipikus kölcsönhatásait, illetve a szintrófia jelenségét, ami a mikrobiális közösségek hálózatos működésének alapját adja.

A szintrófia fogalma az utóbbi években nagy népszerűségnek örvendett a mikrobákkal is foglalkozó tudományterületeken, ennek folyományaként pedig a jelentése kissé elmosódott, így fontos tisztáznunk, hogy mit is értünk alatta. Ebben a dolgozatban a szó egyik hagyományosabb értelmezését használom, mely szerint a szintrófia olyan kapcsolat, amiben a partnerek energetikailag függenek egymástól, és együtt olyan metabolikus folyamatokat bonyolítanak le, amikre egyenként nem volnának képesek [15]. Ez tulajdonképpen azt jelenti, hogy a kapcsolat egyik résztvevője metabolizmusának végterméke a másik félben zajló reakciók kiindulási anyagaként szolgál, így a képződő anyag folyamatosan alacsony koncentrációban van jelen (mivel a partner azonnal felveszi), ami által egyébként energiaigényes folyamatok is energetikailag kedvezővé válhatnak [16]. Egy mikrobiális közösségben jellemzően több ilyen kapcsolatból szerteágazó metabolikus hálózat alakulhat ki. Ennek következményeként kevés erőforrás felhasználásával is változatos anyagcserefunkciók válnak lehetővé, hiszen a termelődő anyagok újabb és újabb mikrobák metabolizmusához biztosítanak (limitáló) forrást. A szintrofikus kapcsolatok energetikailag kedvező hatásával szöges ellentétben áll az exploitatív kompetíció jelensége. Az ilyen kölcsönhatásokban résztvevő felek ugyan azokat a forrásokat hasznosítják működésükhöz, legyen az fény, tapadási felület, szerves

anyag, ezáltal csökkentve a másik fél számára rendelkezésre álló javakat. Azonos reakcióutak és kiindulási anyagok esetén ez a szintrofikus hatással ellentétes kémiai mozgatóerőt fog jelenteni, hiszen a limitáló anyagok alacsony koncentrációban lesznek jelen, ami egyébként energetikailag kedvező reakciókat is veszteségessé tehet, csökken a résztvevő felek metabolizmusának hatékonysága és így a növekedési rátája is.

Az exploitatív kompetíciós kapcsolatban álló fajok között jellemzően egyéb kompetitív adaptációk is megfigyelhetők. Ezeknek az evolúciós háttérben az áll, hogy az adott faj egyedei szelektív előnyhöz jutnak, hogyha a hasonló erőforrásokat felhasználó másik fajok egyedeinek működését valamilyen módon gátolják. Tipikus kompetitív adaptáció lehet az antimikrobiális anyagok szekréciója [17] vagy közvetlen injektálása [18] a környező mikrobákba, illetve a polimerek szekréciója is, mellyel a szekretáló szervezet előnyösebb térbeli elhelyezkedést vagy lokális környezetet biztosíthat magának [19].

Végezetül a véletlen kapcsolatokat is meg kell említenünk, hiszen az adaptív jellegek sokszor másodlagos hatásokat is eredményezhetnek, olyan szervezetekre gyakorolva hatást, amiknek a működésétől a jelleget hordozó élőlény saját fitnesze nem függ [13]. Hogyha a kifejtett másodlagos hatás pozitív, akkor kommenzalizmusról-, ha negatív, akkor ammenzalizmusról beszélhetünk.

## **2.2. Az autokatalitikus halmaz**

Az autokatalitikus halmazok olyan funkcionálisan zárt, önfenntartó reakcióhálózatok, amik képesek önmagukat szaporítani. Ilyen komplex rendszerként tekinthetünk egy-egy élőlényre, a gazdaságra, vagy akár a bioszféra egészére is [20]. Elterjedt feltételezés, hogy a földi élet (illetve az első önfenntartó anyagcserehálózat) is autokatalitikus ciklusok révén jelenhetett meg [21–23]. Kutatásom során mi is autokatalitikus szettek segítségével vizsgáltuk a mikrobiális közösségekben kialakuló szoros függésben létező önfenntartó komponenseket, hiszen hogyha feltételezzük, hogy az ősi mitokondrium kialakulásához két (vagy több) mikroba egyre szorosabb együttműködése vezetett, akkor az általuk alkotott rendszernek is rendelkezni kellett ezekkel a tulajdonságokkal.

Az autokatalitikus szett fogalmát 1971-ben alkotta meg Stuart Kauffman, aki ezalatt molekulák egy olyan csoportját értette, amik az eredetivel közel megegyező molekulacsoport létrejöttét katalizálják, méghozzá olyan módon, hogy egyik molekula sem katalizálja közvetlenül a saját képződését [24]. Később felállította hipotézisét, mely szerint az ilyen struktúrák megjelenése elkerülhetetlen polimerek egy megfelelően komplex csoportjában. A hipotézis tesztelésére egy bináris szekvenciákból álló modellt épített, melyben pár szabadon

elérhető „táplálék” molekulából kiindulva az egyes molekulák lízis, illetve kondenzációs reakciók során alakulhatnak át egymásba, és a reakciók csak katalizátor jelenlétében képesek végbe menni. A katalizátor szerepét a többi szekvencia töltheti be egy bizonyos valószínűséggel, Kaufmann pedig ezen valószínűség függvényében vizsgálta az autokatalitikus szettek létrejöttét [21].

Az autokatalitikus szettek matematikáját és pontosabb definícióját Wim Hordijk és Mike Steel dolgozták ki a 2000-es évek során, bár munkájuk szemléltetéséhez, és az autokatalitikus ciklusokat kereső algoritmusok megalkotásához [25] továbbra is Kaufmann bináris polimer modelljét alkalmazták [22, 26]. A dolgozatom során az ő modell-definíciójukat használtam, mely szerint egy autokatalitikus szett szabadon elérhető táplálékmolekulákból, lehetséges reakciókból, a reakciókhoz tartozó katalizátorokból, valamint a reakciók során képződő molekulákból áll, és két feltételnek kell eleget tennie:

1. Az autokatalitikus szettben található összes reakcióhoz tartozzon legalább egy katalizátor.
2. Az autokatalitikus szettben található reakciókhoz szükséges összes molekula legyen előállítható a szabadon elérhető táplálékmolekulákból, csakis az autokatalitikus szettben szereplő reakciók felhasználásával.

Az fenti definíció egy olyan gráfot feltételez, amelyben a molekulák közötti irányított élek egy része katalizist, egy másik része pedig molekula-transzformációt (hasadást, illetve összeolvadást) jelképez. Azonban én a munkám során kizárólag katalitikus éleket tartalmazó gráfokkal dolgoztam, mivel a mikróbák populációinak egymásra gyakorolt összesített hatását jelenítettem meg a modellemben. Hordijk megmutatta [27], hogy a katalitikus gráfokra is létezik megfeleltetése a definíciójuknak.

Egy különböző molekulákból álló katalitikus gráfban azzal a feltételezéssel kell élnünk, hogy az egyes molekulák szintéziséhez szükséges összes anyag korlátlan mennyiségben biztosítva van. Ekkor az egyik molekulából egy másik felé mutató katalitikus él azt jelenti, hogy az egyik molekula katalizálja a másik molekula szintézisét. Az autokatalitikus szettek szintjén ez úgy fogalmazható meg, hogy minden reakcióhoz biztosítva vannak a szükséges reagensek szabadon elérhető tápanyagmolekulák formájában, illetve hogyha egy molekula beletartozik az autokatalitikus szettbe, akkor az ő szintézisét eredményező reakció is minden esetben bele fog tartozni.

A katalitikus gráfokban értelmezett autokatalitikus szettekben az egyszerű topológia miatt jól elkülöníthető az autokatalitikus mag, és a periféria. Autokatalitikus mag alatt a gráfban található

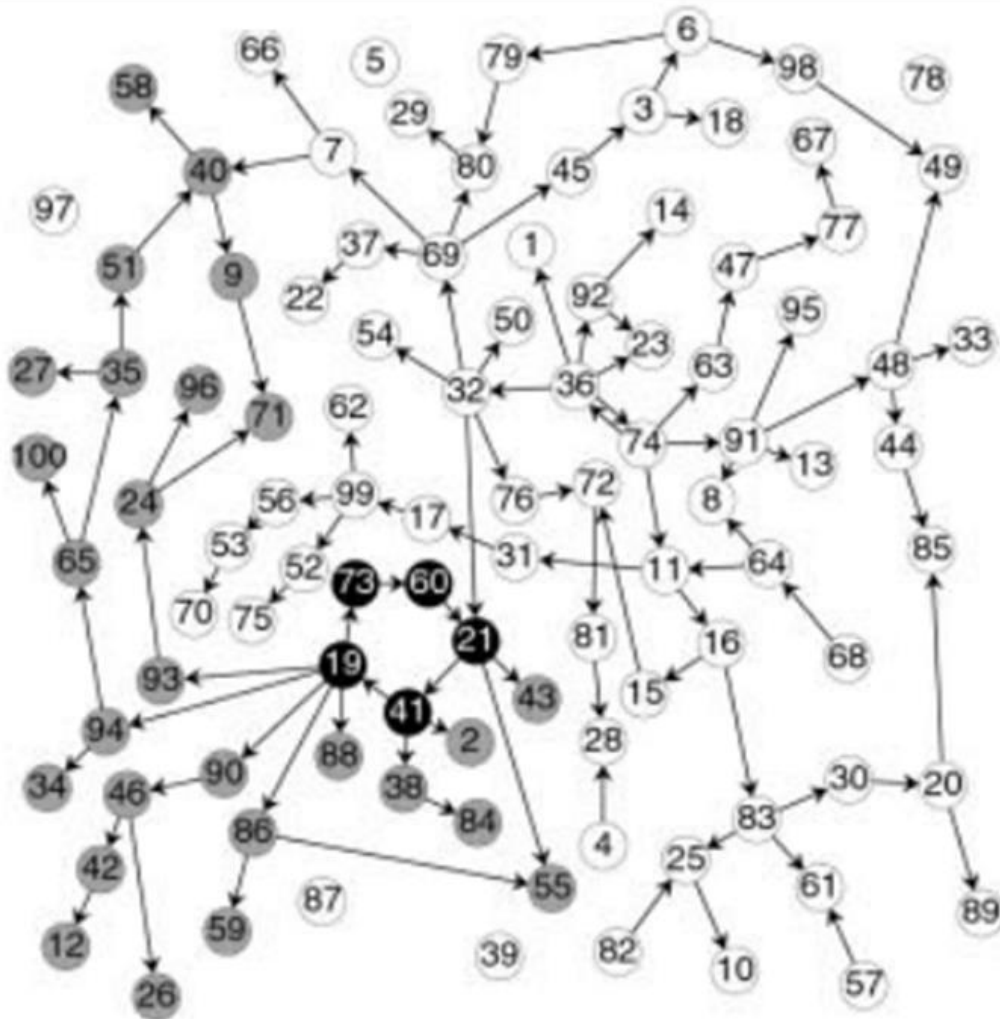
legtöbb nóduszt magába foglaló irányított ciklust értjük, a perifériához pedig minden olyan nódusz tartozik, amihez az autokatalitikus mag valamelyik eleméből katalitikus út vezet. Ez a különbségtétel a későbbiek során fontos lesz, ugyanis a magba és a perifériába tartozó nóduszok eltérő tulajdonságokkal fognak bírni az általunk vizsgált modellekben.

### **2.3.A Jain-Krishna modell bemutatása**

A baktériumközösségek működését szimuláló modellem alapjául Jain és Krishna modellje [9] szolgált. Ebben a fejezetben az ő modelljüket, annak feltételezéseit és eredményit mutatom be. Jain és Krishna (A továbbiakban: JK) modellje egy idealizált kémiai rendszer működését mutatja be, melyben különféle molekulák találhatóak, és az egyes molekulák katalitikusan, illetve inhibitorikusan hathatnak a többi molekula képződésére. Fontos, hogy az inhibitorikus élek megjelenésével az autokatalitikus szettek és magok értelmezésénél továbbra is csak a katalitikus éleket vesszük figyelembe. Azért beszélünk idealizált kémiai rendszerről, mert feltételezzük, hogy a molekulák képződéséhez szükséges összes anyag korlátlan mennyiségben van jelen a rendszerben. Ez az összes molekula mennyiségének korlátlan növekedését eredményezné, azonban feltételezzük azt is, hogy az egyes anyagokat képző reakciók csak katalizátor jelenlétében képesek végbe menni. Ez jól közelíti a természetben lezajló folyamatokat, ugyanis élő szervezetekben valóban ritkán játszódik le reakció katalizátor nélkül. A rendszer molekuláit időről időre evolúciós lépéseknek vetjük alá, melyek során a legalacsonyabb számban jelen lévő molekulák egyike rendre kipusztul, és a helyén egy új molekula jelenik meg. A különböző termelhető molekulák maximális száma tehát nem fog változni az evolúciós lépések hatására, azonban a ténylegesen jelen lévő molekulák száma változhat, hiszen lehetnek olyan molekulák, amelyek egyáltalán nem termelődnek, ekkor a relatív populáció méretük nulla lesz. A modell eredményei szerint a szimuláció kezdeti szakaszában a rendszerben lévő különböző molekulák száma alacsony értékek körül fluktuál (random fázis), azonban amikor megjelenik egy autokatalitikus mag, vagyis záródik egy katalitikus élekből álló kör a gráfban, a különböző molekulák száma hirtelen megnő, és a rendszer növekedésnek indul. Ez azért alakul így, mivel a magban szereplő molekulák - közvetetten, vagy közvetlenül - katalizálják egymás termelődését, így az általuk kifejtett hatás pozitívan csatol vissza a saját növekedésükre. Emellett egy idő után megjelennek olyan molekulák is, amiknek a képződését a mag egyik tagja, vagy a mag egyik tagjából kiinduló katalitikus útban szereplő molekula katalizál, és mivel a mag tagjainak növekedése biztosítva van, a belőlük kiinduló katalitikus leágazások molekulái is rendre növekedésnek indulnak. Ezek

a molekulák képzik a mag köré rendeződő autokatalitikus halmaz periferiáját. Ez a folyamat jellemzően addig fog tartani, amíg az összes rendszerben lévő molekula az autokatalitikus halmaz részévé nem válik. Így a gráf egy kisebb szimbiotikus tagból - a magból -, és egy

### Autokatalitikus mag és periféria egy katalitikus gráfban



1. ábra - Egy katalitikus élekből álló gráf. A feketével jelölt nóduszok az autokatalitikus maghoz-, a szürkével jelölt nóduszok a perifériához tartoznak. A fehérrel jelölt nóduszok nem képezik részét az autokatalitikus szettnek. Forrás: Jain és Krisna, 2002

nagyobb parazitikus részből áll, amit a mag működése táplál. A magot alkotó molekulák száma eközben nőhet, ám a parazitikus jellegű molekulák jellemzően túlsúlyban lesznek.

A növekedési fázis végeztével megkezdődik a rendezett szakasz, melynek során a rendszerben jelenlévő különböző molekulák száma közel maximális. Ebben a fázisban egy újonnan megjelenő molekula rendre azonnal kivesztül a rendszerből, kivéve azt az esetet, amikor képződését katalizálja az autokatalitikus halmaz molekuláinak egyike. Ekkor az újonnan megjelenő molekula populációs mérete nem nulla lesz, így helyette jellemzően a parazitikus periféria egyik molekulája fog kihalni.



Azonban amikor a rendszer kapcsolatait leíró gráfban mindössze egy olyan katalitikus élekből álló kör található, amely mentén a mag összes tagját érinthetjük, a mag tagjai a periférián lévő molekulákhoz hasonló mértékben növekednek, így elképzelhető az is, hogy a mag molekuláinak egyike fog kihalni. Mivel ilyenkor a maghoz tartozó molekulákat egyetlen kör köti össze, bármelyik molekula kihalása a mag széteséséhez vezet, ami a rendszerben jelenlévő különböző molekulák számának drasztikus csökkenésével jár, hiszen így a mag működésétől függő molekulák képződése nincs katalizálva, ezért relatív populációs méretük nullára csökken. A mag szétesése után előfordulhat, hogy megmarad egy más molekulákból álló kisebb ciklus, ami átveszi a mag szerepét, ekkor ismét a növekedési fázisba lép a rendszer, ellenkező esetben pedig a random fázistól kezdődik előlről a folyamat [28].

### 3. Anyag és módszer

#### 3.1.A Python programnyelv és a felhasznált könyvtárak bemutatása

Munkám során a Python programnyelv 3.8-as verzióját, illetve a PyCharm integrált fejlesztési környezet 2020.2.1-es verzióját használtam a modell megépítéséhez és a szimulációk futtatásához. A Python egy magas szintű, általános célú programnyelv. Főbb tulajdonságai:

- **Lefordított**, azaz gépi utasításokra történő fordítás nélkül, általánosan is értelmezhető. A leírt parancsokat nem egy célgép, hanem egy program (interpreter) olvassa be és hajtja végre.
- **Dinamikusan íródó**, azaz az interpreter a változókat típusokba sorolja az aktuális értékük alapján a program futtatása közben.
- **Típus-biztos**, azaz az egyes típusú változókkal csak olyan parancsok hajthatók végre, amelyeket az adott típus támogat.
- **Garbage-collected**, azaz a program automatikusan megpróbálja visszanyerni a memóriának azt a részét, amit korábban a program futtatása lefoglalt, de többé már nincs használatban, így javítva a teljesítményt.

A Python kialakítása felhasználóbarát, a könnyű olvashatóságot és programozhatóságot helyezi előtérbe a gyors futási idővel szemben. Az előbbieket például az operátorok magától értetődő jelölései, és a sorbehúzásra vonatkozó szigorú rendszer által minimálisra redukált zárójelzés segítik elő. [29]

A Pythonba beépített funkciók száma viszonylag limitált, ám könnyen bővíthetjük eszköztárunkat a rengeteg, mindenki számára ingyenesen elérhető könyvtár segítségével. Munkám során a következő könyvtárak funkciói voltak segítségemre:

- **Pip 21.0.1:** A többi könyvtár letöltésére, kezelésére szolgáló bővítmény
- **NumPy 1.19.3** [30]: N - dimenziós sorok (vektorok, mátrixok) létrehozását, indexelését, másolását és egyéb módokon történő manipulálását megkönnyítő funkciókat tartalmazó könyvtár.
- **SciPy 1.5.4** [31]: Matematikai problémák megoldására szolgáló algoritmusokat tartalmazó könyvtár, innen származik az integrátor, amit a modell differenciálegyenleteinek megoldására használtam.
- **Pandas 1.1.4** [32]: Az adatok kiértékelését, manipulálását elősegítő könyvtár, az ábráim létrehozásához szükséges adattáblák megalkotásához használtam.
- **Matplotlib 3.4.1** [33]: Adatvizualizációs csomag
- **Seaborn 0.11.0** [34]: A Matplotlib-en alapuló adatvizualizációs csomag

### 3.2.A Jain-Krishna modell működése

A dolgozatomban használt modell alapjául Jain és Krishna idealizált kémiai rendszerek viselkedését leíró modellje (JK modell a továbbiakban) szolgált [9]. Első lépésben implementáltam a JK modellt, reprodukáltam az ebből kapott eredményeiket, majd vizsgáltam benne az autokatalitikus magok fennmaradását. A modell rekonstruálásához felhasználtam Jain és Krishna egy hasonló modelljének [28] Python-implementációját [35]. Az alábbiakban először bemutatom a rekonstruált modell működését és eredményeit.

Kezdetben adott egy üres gráf,  $s_{max}$  számú nódusszal, amik különböző molekulák populációit reprezentálnak. A nódusok között véletlen módon irányított éleket sorsolunk ki, ezek az élek határozzák meg, hogy milyen típusú molekulák gátolják, illetve katalizálják az egyes típusokba tartozó molekulák képződését. A véletlen kiosztás során  $i$  és  $j$  nódusz között  $1-p$  valószínűséggel nem lesz él  $i$ -ből  $j$ -be. Ekkor  $i$  nódusznak nincs hatása (vagy a hatások eredője semleges)  $j$  nóduszra. Egy adott nódusz  $p$  valószínűséggel van hatással egy másik nóduszra, azaz ekkor van interakciós él a két nódusz között a gráfban. Ekkor a kapcsolathoz véletlen erősséget rendelünk a  $[-1, 1]$  tartományú egyenletes eloszlásból. A negatív súly inhibíciót, a pozitív súly katalízist jelent. Két nódusz között legfeljebb kettő, ellentétes irányú él lehet. Az így létrehozott gráfot (kapcsolatrendszer) és az élek súlyát (kapcsolatok erősségeit) az  $s_{max} \times s_{max}$  méretű  $C$  mátrix írja le. Hogyha egy adott  $j$  nódusz  $c$  hatással van az  $i$  nóduszra, akkor  $C_{ij} = c$ . Azokban az esetekben, amikor  $i = j$ , a kapcsolat erősségét a  $[-1, 0]$  egyenletes eloszlásból sorsoljuk, önmagukban autokatalitikus magokat tehát nem engedünk meg. A kapcsolatrendszer kialakítása után a nóduszokhoz egy-egy random számot rendelünk a  $[0, 1]$

egyenletes eloszlásból, ami a faj denzitását vagy egyedszámát reprezentálja. E számokat olyan módon skálázzuk át, hogy összegük 1 legyen, megkapva így a nóduszokhoz tartozó kezdeti relatív populációméreteket.

A fajok időfejlődése egy egyszerűsített replikátor dinamikát követ. A nóduszokhoz tartozó relatív populációméretek az 1. egyenlet szerint változnak, egészen addig, amíg jó közelítéssel egyensúlyi állapotba nem kerül a rendszer. Az egyensúlyt a karakterisztikus idő elérésevel beálltnak tekintjük.

$$\dot{x}_i = f_i \text{ ha } x_i > 0 \text{ vagy } f_i > 0, \text{ és}$$

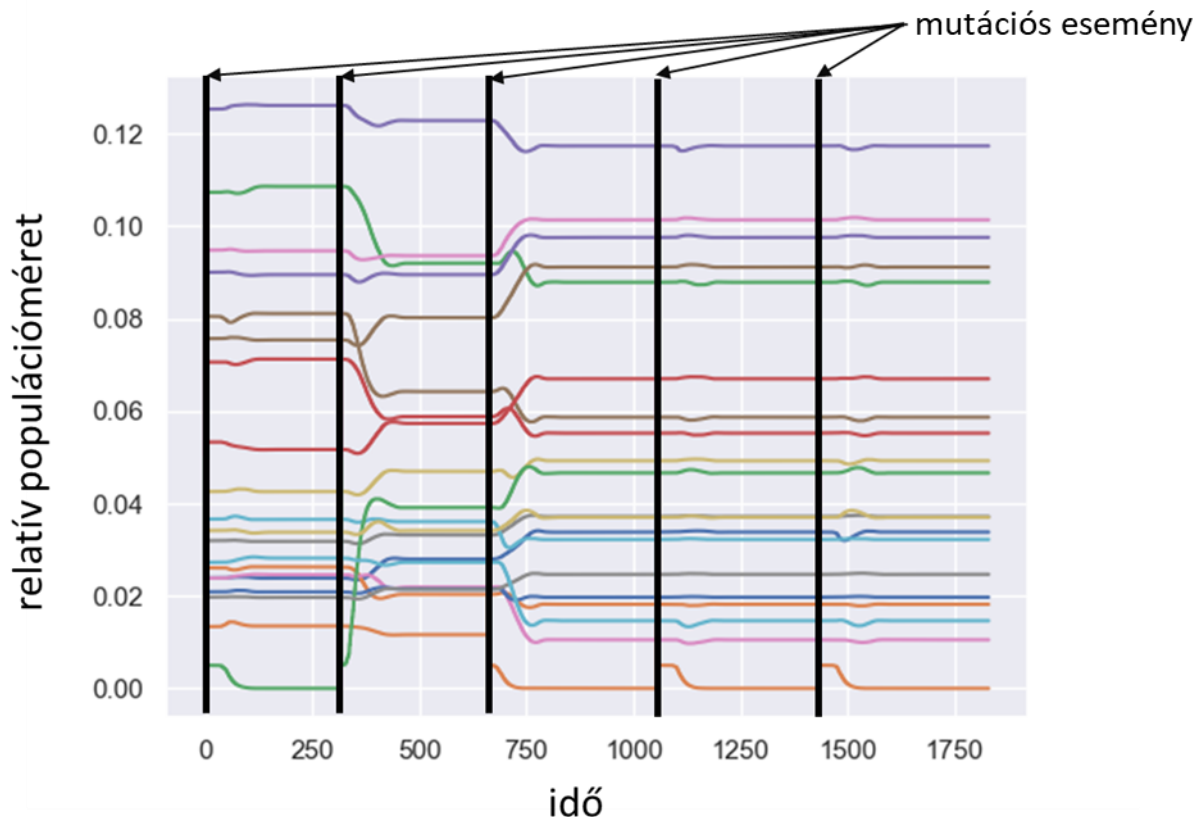
$$\dot{x}_i = 0 \text{ ha } x_i = 0 \text{ és } f_i < 0, \text{ ahol:}$$

$$f_i = \sum_{j=1}^{S_{\max}} c_{ij} x_j - x_i \sum_{k,j=1}^{S_{\max}} c_{kj} x_j$$

**1. egyenlet**

Amikor elértük az egyensúlyi állapotot, beindul a szelekciós mechanizmus. Ennek során meghatározzuk a legkisebb relatív populációmérettel rendelkező nóduszt vagy nóduszokat, melyekből egyet random módon kiválasztunk. A kiválasztott nódusz „kipusztul”, vagyis kivesszük a kapcsolati mátrix rá vonatkozó sorát és oszlopát, illetve lenullázzuk a relatív populációméreteket tartalmazó  $\mathbf{x}$  vektorban is.

## A reprodukált modell működése 5 időlépésen keresztül



2. ábra - A modell működése 5 időlépésen keresztül  $s = 100$  és  $p = 0.005$  paraméterek mellett. Fekete függőleges vonallal a szelekciós események vannak ábrázolva, közöttük a színes vonalak az egyes nóduszokhoz tartozó relatív populációméretet változását jelzik.

A kihalási esemény után azonnal új faj jelenik meg, így mindig  $s_{\max}$  számú nódusz van a gráfban. Az új nódusz kapcsolatrendszerét a gráf kezdeti beállításaihoz hasonló módon, véletlenszerűen generáljuk. Az új faj relatív populációméretét alacsony  $x_0$  értékben állapítjuk meg. A relatív populációméreteket ezután újraszkalázzuk, hogy az összegük továbbra is 1 maradjon. Az így kialakult új kapcsolatrendszer ismét növekedésnek indul, és az új egyensúly elérésekor új szelekciós esemény történik. E lépések további iterációjával modellezzük az idealizált kémiai rendszerekben zajló evolúciós folyamatokat [9].

### 3.3. Ciklusok keresése

Munkám során fontos feladat volt egy cikluskereső algoritmus implementálása, aminek segítségével a szelekciós lépések után megjelenő új faj bevezetése miatt megváltozott kapcsolatrendszerben rendre megállapíthatom a különböző autokatalitikus ciklusok számát és méretét. Erre a célra egy Depth First Search (DFS) algoritmust használtam, ami a gráfban található összes katalitikus utat megvizsgálja visszalépések segítségével. Katalitikus út alatt a

gráfban található irányított éleknek egy olyan, egy irányba mutató, egymáshoz csatlakozó sorozatát értjük, amelyben egyik él vagy nódusz sem szerepel egynél többször. Ehhez Tiernan algoritmusát használtam [36]: a módszer lényege, hogy képletesen „beutazzuk” az egész gráfot. A szemléletes magyarázatot segítő, bemutatom az algoritmus lépéseit.

Adott egy  $N$  nódusú irányított gráf. A gráf nódusaihoz 1-től  $N$ -ig sorszámokat rendelünk. Az algoritmus az 1. számú nódusból indul egy random útvonalon a gráf élei mentén. Utunk során egy vektorban rögzítjük, hogy melyik nóduszokat érintettük már, ezeket többször nem érintjük. Amikor zsákutcába érkezünk, - vagyis az út utolsó nódusából egyetlen olyan nóduszba se mutat él, ami még nem volt érintve - az algoritmus ellenőrzi, hogy a legelső nóduszba vezet-e innen él. Ha vezet, ciklust találtunk, és eggyel megnöveljük a talált ciklusok számát.

Függetlenül attól, hogy találtunk-e ciklust, a zsákutcába érkezés után visszalépünk. A visszalépés során az eddig bejárt utat rögzítő vektor utolsó előtti elemébe haladunk tovább, az utolsó elemet pedig eltávolítjuk a vektorból, felvesszük azonban az utolsó előtti elem előtt zárt útvonalak listájába, amivel garantáljuk, hogy az előbbi út többször ne ismétlődjön. Egészen addig lépkedünk vissza, amíg nem találunk olyan élt, amely mentén az utunk ismét kibővíthetővé válik. Az egyes nóduszok előtt zárt útvonalakat egy  $N \times N$  méretű mátrixban rögzítjük.

A sorozatos kibővítések és visszalépések következtében végül az 1. számú nódusz előtt is zárttá válik az összes útvonal (ha volt ilyen), így tovább lépünk a kettes számú nóduszra. Innen kiindulva az előbbiekhöz hasonló módon járjuk be az összes lehetséges utat, ám ekkorra már az összes olyan ciklust megtaláltuk, amiben szerepel az 1. számú nódusz, éppen ezért a további ciklusok keresésekor már csak a gráf 1-nél nagyobb sorszámmal rendelkező nóduszait kell figyelembe vennünk. Növekvő sorrendben vizsgáljuk tehát az egyes nóduszokat és a belőlük kiinduló utakat, amivel egyrészt sok futási időt spórolunk, hiszen így a gráf vizsgált résztvevőinek száma egyre csökken, valamint azt is garantáljuk, hogy minden ciklus csak egyszer legyen számolva, hiszen egy adott ciklust csak a benne szereplő legkisebb sorszámú nódusból kiinduló utak vizsgálatakor találhatunk meg. Ez az eljárás alkalmas a különböző méretű ciklusok megszámlálására is: létrehozunk egy számlálót az adott ciklusmérethez, amit akkor növelünk meg eggyel amikor ciklust találtunk, és a bejárt út hossza megegyezik a keresett ciklusmérettel.

A számlálók értékét minden időlépés végén egy adattáblában rögzítjük, és minden időlépés elején nullázzuk, hogy csak a kapcsolati mátrix adott konfigurációjában szereplő ciklusokat vegye figyelembe.

A ciklusok száma mellett szintén fontos mutató, hogy az egyes méretű ciklusok átlagosan hány időlépésen keresztül maradnak fenn a rendszerben. Ennek a vizsgálatára is az előbb bemutatott algoritmust használtam fel. A szimulációk indítása előtt megadtam a cikluskereső argumentumaként, hogy milyen méretű ciklusok fennmaradását szeretném vizsgálni, így az első ilyen méretű talált ciklus lesz a választott ciklusunk. A vizsgált ciklus kiválasztása tehát nem teljesen véletlenszerű, és az első talált ciklus mutatói eltérhetnek a később talált ciklusokétól, azonban mivel minden esetben az első talált ciklust fogjuk figyelembe venni, így ez alapján az egyes modellek és ciklusméretek a későbbiekben összehasonlíthatók lesznek. Amikor megtaláljuk az első megadott méretű ciklust, annak az éleit eltároljuk egy listában. Ezeknek az éleknek a meglétét innentől kezdve minden mutációs lépés után ellenőrizzük, az erre a célra létrehozott logikai típusú változó értéke **True** lesz, hogyha továbbra is jelen vannak az élek a gráfban, és **False**, hogyha nem. Az élettartamot úgy kapjuk meg, hogy a ciklus élettartamát számláló mutatót minden mutációs lépés után eggyel növeljük, hogyha a ciklus továbbra is jelen van a gráfban. A fentebb említett logikai változó segítségével a vizsgált ciklus szétválasztása után a szimulációt meg is szakíthatjuk abban az esetben, ha kifejezetten csak a ciklusok élettartamára vagyunk kíváncsiak.

A cikluskereső algoritmus megfelelő működéséhez a vizsgálandó gráf kapcsolatrendszerét egy meghatározott formátumban kell megadnunk, nevezetesen egy  $S_{max}$  számú listát tartalmazó listában, melyben az egyes nóduszok sorszámaikat viselő listák azoknak a nóduszoknak a sorszámaikat tárolják növekvő sorrendben, ahova az adott nódusból mutat él. Ahhoz, hogy ezt az objektumot létrehozzuk, először is másolatot készítünk a kapcsolati mátrix pillanatnyi állapotáról. Ebben a kapcsolatok rögzítése pont fordítva történik (az egyes sorok nem az adott sor sorszámaának megfelelő nódusból kifele mutat, hanem az ide mutató éleket tartalmazzák), így vennünk kell a kapcsolati mátrix transzponáltját. Az így létrehozott mátrixban a negatív és semleges kapcsolatokat jelző helyekre nullákat rendelünk, az összes pozitív kapcsolatot pedig kicseréljük annak az oszlopnak a sorszámaára, amelyikben található, és hozzáadunk 1-et. Mivel a pythonban a vektorok és mátrixok indexelése 0-tól kezdődik, így ezzel elkerülhetjük a 0. sor és oszlop kapcsolatrendszerének elvesztését az utolsó lépés során, amikor a kapott mátrix soraiban az értékeket egyenként növekvő sorba rendezzük, majd eltávolítjuk az elejükről a nullákat. Az így megkapott lista már meg fog felelni az algoritmus bemeneti kritériumainak.

### 3.4.A Jain - Krishna modell átalakítása

A JK modell átalakításához először is számba kellett vennünk, hogy mely aspektusai nem realiztikusak, illetve milyen lényeges tulajdonságok különböztetnek meg egymástól egy kémiai fajokból (molekulákból) álló idealizált rendszert egy baktériumközösségtől.

fontos kérdés a gráf nóduszai között húzódó élek jelentése. A JK modellben egy adott „A” nóduszból egy „B” nóduszba mutató él azt jelenti, hogy az „A” molekula katalizátora (illetve negatív élek esetén inhibitora) annak a reakciónak, melynek során a rendszerben korlátlanul rendelkezésre álló anyagokból a „B” molekula képződik. Ebből az is következik, hogy egy molekula relatív populációméretének nullára csökkenése nem jelenti azt, hogy ezzel ki is halt volna. Későbbi szelekciós események után megjelenő új molekulák katalizálhatják a képződésének reakcióját, és ekkor ismét 0 fölé emelkedhet a relatív populációmérete. Az, hogy a kapcsolatokat kialakító molekulákra katalizátorokként, illetve inhibitorokként tekintünk, azt is magával vonja, hogy a kapcsolatok fenntartásának nincsen költsége, hiszen feltételezzük a katalizátort definíció szerint változatlan formában és mennyiségben kapjuk vissza a reakció végeztével.

Az általam átalakított, bővített modellben (a továbbiakban: bővített modell) a kapcsolatrendszer leíró gráf nóduszaira mikroba fajok populációiként tekintünk. Bár a gráf irányított élei továbbra is katalitikus kapcsolatokat fognak jelezni, ezekre már nem tekinthetünk egy-egy reakció katalizátoraiként (illetve inhibitoraiként), sokkal inkább az egyes fajok populációinak jelenlétéből fakadó, másik fajok populációira gyakorolt összes pozitív (kooperáció) és negatív (inhibíció) hatás összegeként. Mindez magával vonja azt is, hogy az egyes fajok relatív populációméretének nullára csökkenése után nem realiztikus, ha a megfelelő élek megjelenésével ismét növekedésnek indulhatnak. Ha egy faj nincs jelen a rendszerben, akkor az általa hordozott genetikai információ is elvész, nem tarthatja meg tehát a korábbi kapcsolatrendszerét. Éppen ezért a bővített, mikrobiális modellben amikor egy adott  $k$  faj relatív populációmérete nullára csökken ( $x_k = 0$ ), akkor a kapcsolati mátrixban a  $k$ -t reprezentáló nóduszból induló ( $C_{ik}$ ), és az adott nóduszba mutató ( $C_{kj}$ ) éleket egyaránt lenullázzuk.

A következő pont a szelekciós és mutációs mechanizmus módosítása. A JK modellben a kiháló molekulaféleség helyét minden időlépésben egy újonnan megjelenő molekulatípus veszi át, aminek a kapcsolatrendszerét a már ismert random módon kapjuk meg. Ezt az eljárást nyugodtan alkalmazhatnánk baktériumközösségek esetében is, azonban ennél realiztikusabb megközelítés, ha a rendszerben megjelenő új faj - vagy legalábbis kapcsolatainak egy része - a rendszerben már jelenlévő fajok kapcsolatrendszeréből származna, jobban leképezve ezzel a

természetes körülmények között lezajló mutációs folyamatokat. A modellt ennek megfelelően alakítottam át.

Az átalakított szelekciós mechanizmus működtetéséhez az  $m$  és  $i$  paramétereket vezettem be a modellbe. Az újonnan megjelenő faj  $m$  valószínűséggel migráns,  $1-m$  valószínűséggel pedig mutáns lesz. A migránsok esetében a kapcsolatrendszer a már ismert random módon fogjuk megkapni, azonban a mutánsok esetében az egyes kapcsolatok (illetve azok hiánya)  $i$  valószínűséggel a mutációs szülőtől fognak származni. A mutációs szülőt szintén random módon határozzuk meg, mégpedig úgy, hogy minden faj a relatív populációméretének megfelelő eséllyel tölti be ezt a szerepet. A mutáns kapcsolatrendszerének nem-öröklődő részében továbbra is  $p$  és  $1-p$  valószínűséggel alakulnak ki effektív, illetve semleges kapcsolatok.

A következő változtatás nem szerepelt a terveink között, azonban az eredeti modell implementációjánál rá kellett eszmélnem, hogy a szerzők nem biztosítottak elegendő információt a modelljük pontos rekreálásához. Vizsgáljuk meg ismét az 1. egyenletet!

Mint látható, az egyenlet csak azokra az esetekre ad megoldást, amikor a vizsgált faj relatív populációmérete, vagy a relatív populációméret változásának deriváltja nagyobb, mint nulla, illetve amikor a vizsgált faj relatív populációmérete nulla, és deriváltja kisebb, mint nulla. Azokat az eseteket tehát nem értelmezi, amikor  $x_i$  negatív, amit a szerzők azzal magyaráznak, hogy az egyenlet pontosan azért van ilyen formában megadva, hogy a relatív populációméret ne csökkenhessenek nulla alá. Ezzel csak az a probléma, hogy a bemutatott egyenletben semmi sem akadályozza  $x_i$  nulla alá való csökkenését, mint ahogy az a modell rekonstruálása és futtatása során számomra is kiderült. Ezt a szimulációs eredmények vizsgálata nélkül is viszonylag egyszerű belátnunk, hiszen tegyük fel, hogy egy kihalófélben lévő molekulát vizsgálunk, méghozzá abban a pillanatban, amikor relatív populációmérete eléri a nullát. Az ezt megelőző integrációs lépésben  $x_i$  értéke még pozitív volt, így az 1. egyenlet alapján deriváltja  $f_i$  lesz. Ekkor  $x_i$  értéke az adott  $f_i$  érték ismeretében az integrátor lépésközének megfelelő mértékben fog csökkenni, elérve így a nullát. Azt azonban semmi sem fogja garantálni, hogy az  $f_i$  és a lépésköz által meghatározott csökkenés abszolút értéke ne legyen nagyobb  $x_i$ -nél, hiszen az  $f_i$ -ben szereplő két tag közül bármelyik vehet fel negatív értéket, és csak a második függ  $x_i$  méretétől. Mindez arra enged következtetni, hogy az eredeti modellben szerepelt egy küszöbérték, mely alatt a relatív populációméretnek nullának tekintendők. Hogyha ez a küszöbérték megfelelően nagy-, a maximális lépésköz pedig megfelelően kicsi értékre van beállítva, az megoldja az imént vázolt problémát, mivel a relatív populációméretnek már azelőtt lenullázódnak, hogy elérnék a negatív tartományt. Ennek a küszöbértéknek a létezését az is



alátámasztja, hogy a  $\mathbf{p}$  paraméter alacsonyabb értékei mellett általában több olyan nódusz is található a gráfban, amiknek a növekedésére csak a kifolyási tag (lásd 1. egyenlet, az  $\mathbf{f}_i$  értékét megadó egyenlet negatív előjeles tagja) van hatással, mivel semelyik nóduszból nem mutat ide él. Az ilyen molekulák populációiban fokozatos csökkenés következik be, azonban mivel a kifolyási tag a relatív populációméret arányában oszlik meg, értékük jellemzően nem fogja elérni a nullát, ami szintén indokolná egy küszöbérték bevezetését. Ezeknek a körülményeknek a figyelembevételével a Jain-Krishna modell implementációjában a 2. egyenlet szerint számoltuk a relatív populációméreteket, majd az így kapott értékek közül nullára állítjuk be azoknak az értékét, amik  $0.01/s_{\max}$  -nál kisebbek, majd az összes relatív populációméretet újraszámoljuk, hogy összegük továbbra is 1 maradjon.

$$\dot{x}_i = f_i \text{ ha } f_i + x_i \geq 0, \text{ és}$$

$$\dot{x}_i = -x_i \text{ ha } f_i + x_i \leq 0, \text{ ahol:}$$

$$f_i = \sum_{j=1}^{s_{\max}} c_{ij} x_j - x_i \sum_{k,j=1}^{s_{\max}} c_{kj} x_j$$

**2. egyenlet**

Ezeket a módosításokat elvégezve sem működött tökéletesen az eredeti modell rekonstruált változata, ugyanis a kifolyási tagok abszolút értékeinek összege nem egyezett meg a teljes növekménnyel. Ez abból adódott, hogy a kifolyási tagba azok az inhibitorikus élek is bele vannak számolva, amik 0 relatív populációméretű nóduszokba mutattak, így nem fejthették ki hatásukat. Ezt úgy oldottam meg, hogy az ilyen nóduszokba mutató éleket az egyensúly beálltáig átmenetileg eltávolítottuk a kapcsolati mátrixból, ám a következő időlépésben ismét megjelenítettük őket, hogyha az adott molekula nem pusztult ki idő közben.

$$f_i = x_i \sum_{j=1}^{s_{\max}} c_{ij} x_j - x_i \sum_{k,j=1}^{s_{\max}} c_{kj} x_j x_k$$

**3. egyenlet**

Az élő szervezetek populációinak növekedése a populációméretétől is függ. Ennek megfelelően a populációnövekedését leíró differenciálegyenletet is módosítottam. Ezt hivatott megjeleníteni a 3. egyenlet, melyben az 1. egyenlethez hasonlóan a nyereség kifolyása garantálja, hogy a

relatív populációméreték összege 1 maradjon. Itt már mindkét tag függ az adott faj relatív populációméretétől, így a negatív értékekkel nem kell külön foglalkoznunk, azonban ugyanezen ok miatt nullára sem csökkenhet, így itt is indokolt a küszöbérték bevezetése.

### 3.5.A munka menete

A modellépítés első szakasza a tesztelés volt, melynek során a Jain-Krishna modell általam elkészített implementációját igyekeztem úgy alakítani, hogy az eredeti modell eredményeit kapjam vissza. A modell működését  $p = 0.005$  és  $s_{max} = 100$  paraméterek mellett 7000-9000 időlépésen keresztül teszteltem minden jelentősebb változtatás után legalább ötször. Az összehasonlítás alapjául a kapcsolati mátrixban található pozitív (katalitikus) és negatív (inhibitorikus) kapcsolatok-, valamint a pozitív relatív populációs mérettel rendelkező nóduszok számának időbeli alakulása szolgált. A tesztfutások futási ideje a modell különböző konfigurációitól függően 1-5 óra körül mozgott.

A következő lépés a tesztfutások alapján elfogadott modell bővítése volt, melynek során kisebb  $s_{max}$  értékek alkalmazásával a modellt tesztelő szimulációk futtatása 20 - 60 percet vett igénybe. A munka végső szakaszában a létrehozott modellek működését értékeltem ki különböző mutatók segítségével. A reprodukált Jain-Krishna modellben a kevés résztvevőből álló autokatalitikus magok fennmaradásának mutációs lépésekben értelmezett idejét vizsgáltam. A szimulációk egyik felében a  $p$  paramétert  $0.005$ -ön rögzítettem, és egy 12-től 100-ig terjedő skálán nyolc különböző  $s_{max}$  érték mellett 20-20 szimulációt végeztem, melyek a vizsgált mag eltűnéséig, vagy a megszabott maximális 20 000 időlépésig tartottak. Az olyan esetekben, amikor egyáltalán nem jött létre vizsgálható mag, vagy a vizsgált mag a 20 000. időlépésig nem tűnt el a rendszerből, a futást ismét elvégeztem. Ez valamelyest torzíthatja az eredményeket, ám erre a 160 futásból összesen négy esetben került sor, ebből három esetben egyáltalán nem is jött létre vizsgálható mag, egy esetben pedig a 20 000. időlépésben még nem esett szét a vizsgált mag.

A szimulációk második felében az  $s_{max}$  paraméter volt rögzítve 50 értéken, így egy 0.0025-től 0.02-ig terjedő skálán nyolc különböző  $p$  érték mellett szintén 20-20 szimulációt futtattam az előbbiekhöz hasonló módon. Itt 6 futást kellett megismételnem és ebből 4 a 0.0025-ös  $p$  beállítás mellett, ám mivel ebből 3 esetben vizsgálható ciklus hiánya miatt nem kaptam eredményt és csak egy esetben szakadt meg úgy a szimuláció, hogy volt ciklus a rendszerben, így az ismétlések itt sem torzíthatnak sokat az eredményeken.

A bővített modellt is hasonló mutatók alapján terveztem vizsgálni, ám előzetes eredmények alapján kénytelen voltam változtatni az elképzeléseimen, ugyanis a gráf sok esetben túlzottan

bonyolulttá vált, a ciklusok száma exponenciális emelkedésbe kezdett, és amikor nagyságrendileg már 10 millió ciklus volt a kapcsolatrendszerben, eljött az a pont, hogy a számítógép nem bírta kiszámolni a következő időlépéshez tartozó relatív populációs méreteket. Éppen ezért a vizsgált mutató végül ennek a sűrűn kapcsolt állapotnak a kialakulásához szükséges időlépések száma lett. A rendszer állapotát akkor tekintetem sűrűn kapcsoltnak, hogyha a kapcsolatrendszer 100 000-nél több ciklust tartalmazott.

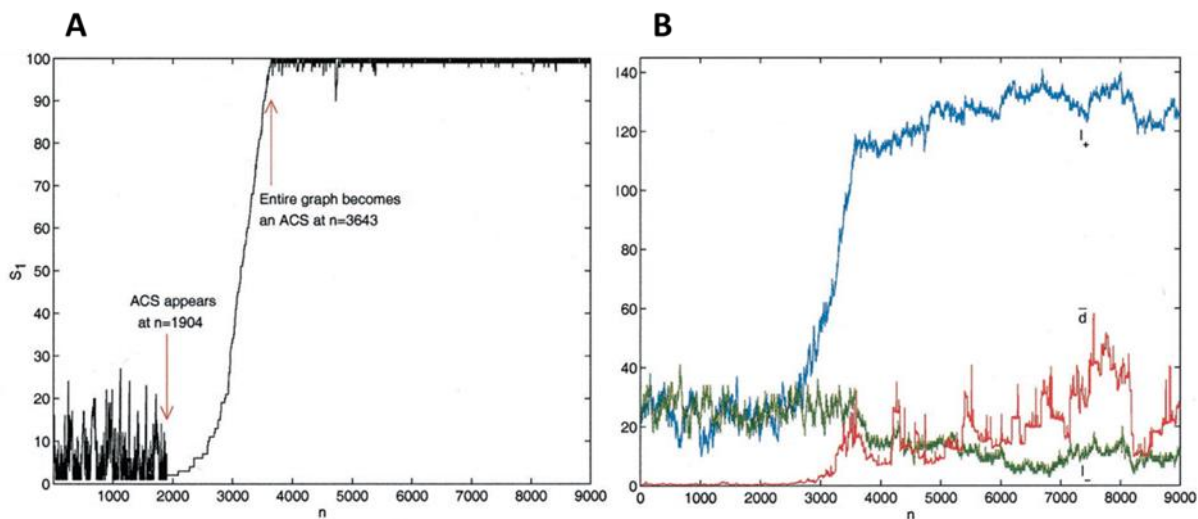
A szimulációk során a  $p$  és  $s_{max}$  paraméterek értékét **0.005**-ben, illetve **100**-ban rögzítettem, emellett az  $m$  (migránsok aránya) és  $i$  (mutáns esetén az öröklődő élek aránya) paraméterek egyikét **0.7**, illetve **0.9** értékeken rögzítettem, míg a másikat egy 0-tól 1-ig terjedő skálán 10 különböző értékre beállítva 30-30 szimulációt futtattam. A szimulációk 10 000 időlépésig, vagy a sűrűn kapcsolt állapot eléréséig tartottak. Az  $m$  paraméter rögzített értékét önkényesen választottam meg, míg az  $i$  paraméter értékét a baktériumokban megfigyelhető alacsony rekombinációs ráta alapján állapítottam meg 0.9-ben [37].

A paraméterek hatását regressziós modellek illesztésével becsültem. Az adatok kiértékelését az R 4.0.2. statisztikai program segítségével, az RStudio 1.2.5001 környezetben végeztem [38].

## 4. Eredmények

### 4.1. A Jain - Krishna modell rekonstrukciója

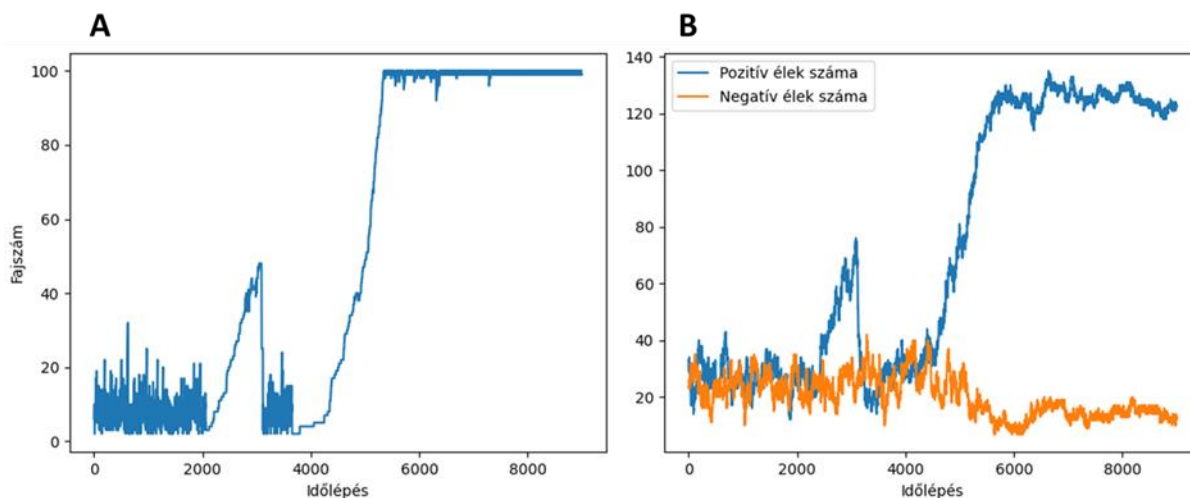
#### Az eredeti Jain-Krishna modell működése



3. ábra - Az eredeti JK modell egy futásának eredményei  $p = 0.005$  és  $s = 100$  paraméterek mellett. Az A ábrán a fajsám, a B ábrán a pozitív élek száma (kék), a negatív élek száma (zöld), és a kölcsönös függőségi mutató (piros) vannak ábrázolva. Forrás: Jain, Krishna 2001

A modellek vizsgálatának első lépéseként a JK modell általam létrehozott rekonstrukciójának működését vettem össze az eredeti modellből kapott eredményekkel. A 3. ábrán az eredeti JK modell egy futásának eredményei láthatók. Az A ábrán a fajsám, a B ábrán a negatív (zöld görbe) és pozitív (kék görbe) élek számának alakulása van ábrázolva 9000 időlépésen keresztül. A B ábrán a piros görbe a JK által kölcsönös függőségre bevezetett mutató, ezt a modell reprodukciójában nem vizsgáltam.

## A reprodukált Jain-Krishna modell működése



4. ábra - A reprodukált JK modell egy futásának eredményei  $s = 100$  és  $p = 0.005$  paraméterek mellett. Az A ábrán a fajszám, a B ábrán a pozitív élek száma (kék) és a negatív élek száma (narancssárga) láthatók.

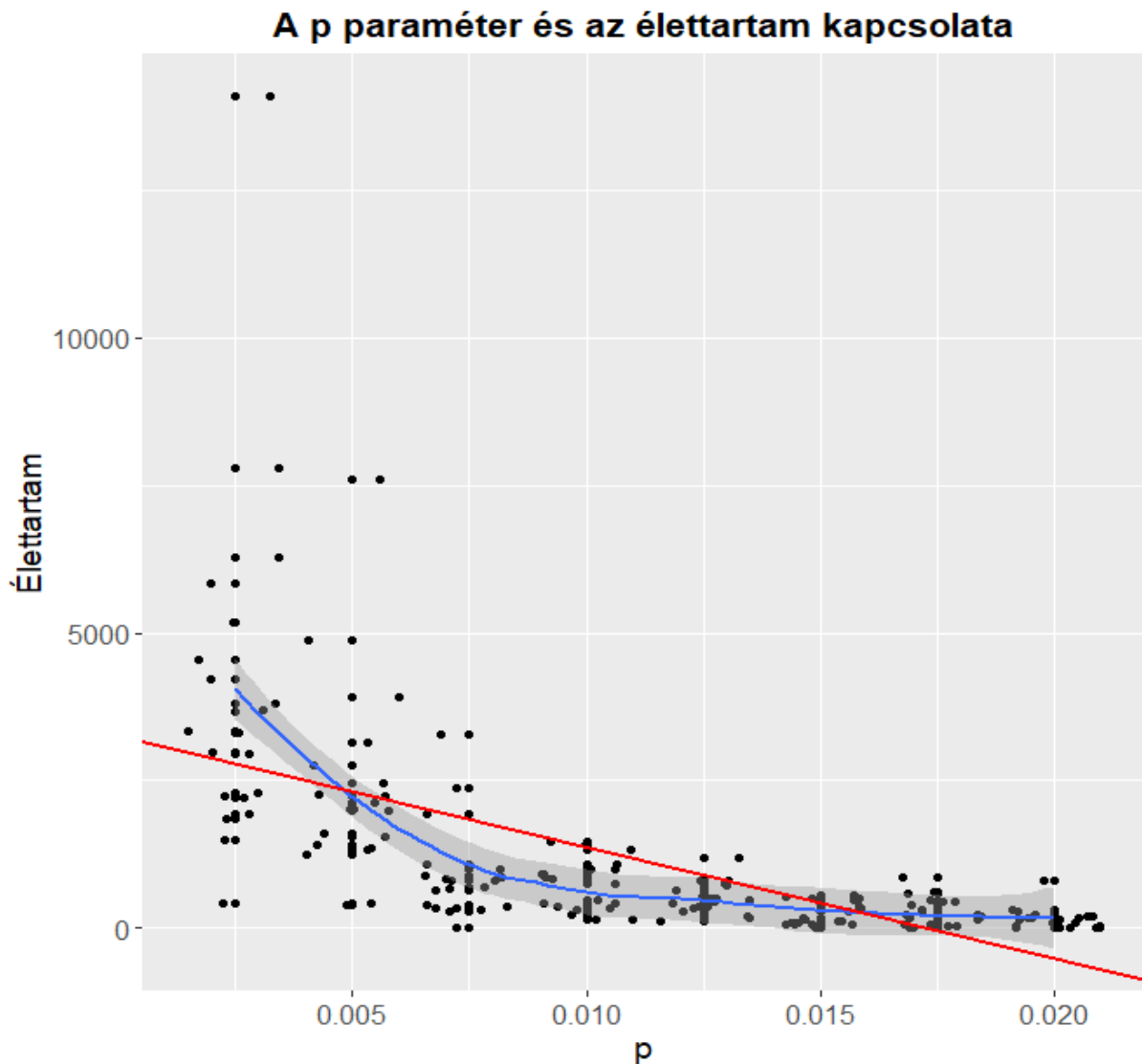
A modell implementációja során létrehoztam a 3. ábra megfelelőjét minden nagyobb változtatás után, ez adta az összehasonlítás alapját. Az 4. ábrán a rekonstruált modell végleges verziójának egy futásából kapott eredmények láthatók. Látható, hogy itt is a 2000. időlépés környékén jelent meg az első autokatalitikus szett, azonban ez rövidesen szétesett, és végül a második autokatalitikus szett bővülésével jutottunk el a rendezett fázisba. A vizsgált mutatók ezt leszámítva azonos értékek körül mozognak, és hasonló görbéket írnak le. Mivel spontán „crash” eseményekről JK is beszámol [28], az 3. és 4. ábrák alapján kijelenthető, hogy a rekonstruált modell és a JK modell működése szerkezetileg azonos.

### 4.2.A rekonstruált Jain - Krishna modell vizsgálata

A rekonstruált JK modellben a kevés elemű autokatalitikus magok fennmaradását vizsgáltam. Az első 80 szimuláció során egyszer sem fordult elő, hogy 3 elemű mag megjelent és szét is esett a megadott maximális időlépés elérése előtt, így feltételeztem, hogy a 4 és 5 elemű magoknál is hasonló eredmények várhatók, és a további elemzések során a 2 elemű magok élettartamának vizsgálatára hagytam.

## A kételemű magok élettartama a $p$ paraméter függvényében

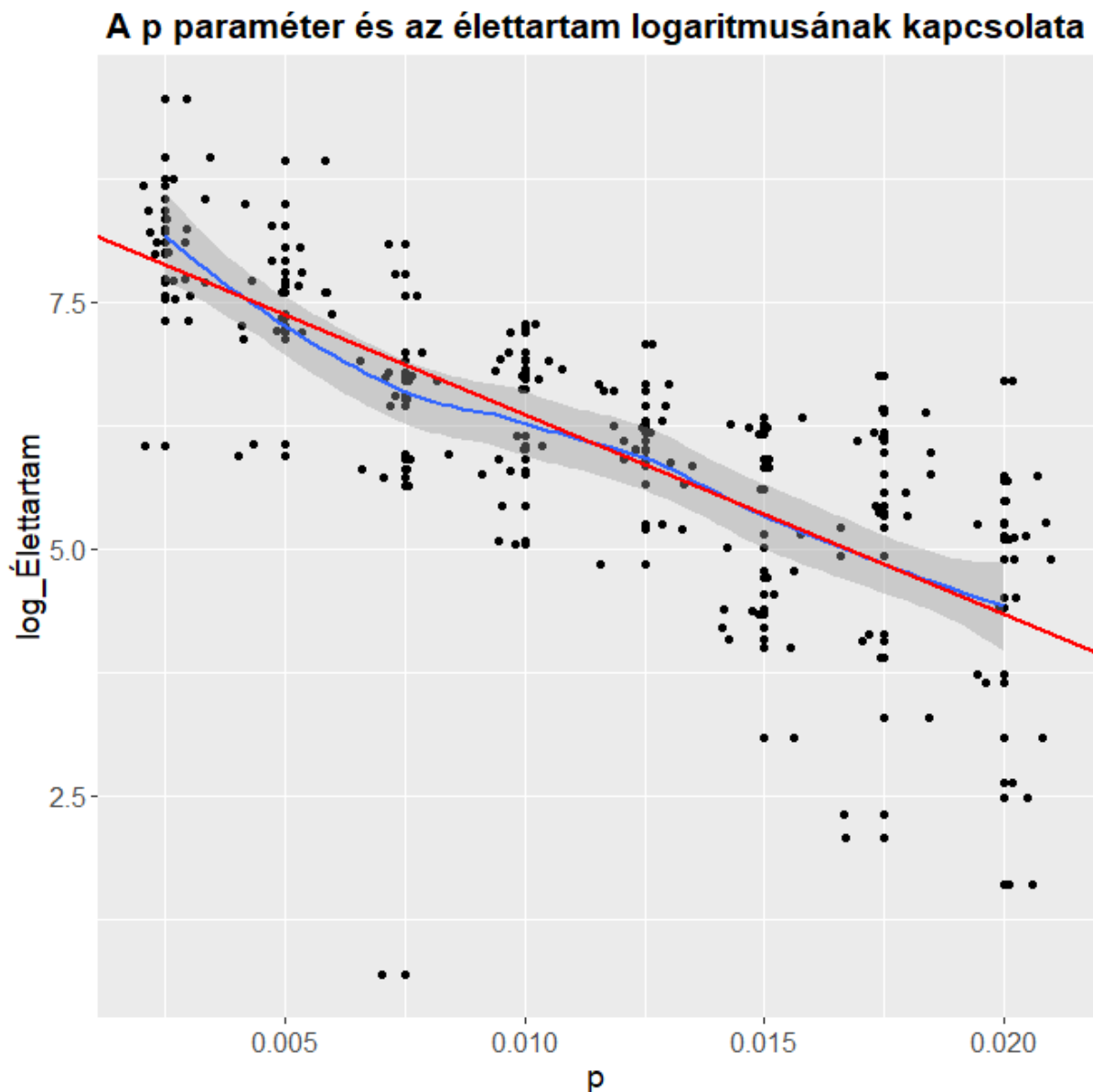
A kapcsolat kialakulásának valószínűsége ( $p$ ) és a kételemű magok élettartama közötti összefüggés feltárásához az  $s_{\max}$  paraméter értékét 100-ban rögzítettem. Először exploratív elemzéseket végeztem. A szórásdiagram (5. ábra) alapján valószínűsíthető, hogy a  $p$  paraméter és az élettartam között exponenciális kapcsolat áll fenn. Az adatpontokra egyszerű lineáris regressziós modellt készítettem, melyben a függő változó a 2 elemű magok élettartama, a magyarázó változó pedig a  $p$  paraméter volt. A modell együtthatóit alkalmazva egyenest (az 5. ábrán piros) illesztettem az adatokra, és bár a regresszió szignifikánsnak bizonyult ( $p < 0.0001$ ), az egyenes láthatólag nem illeszkedik jól a pontokra, ami a korrigált  $R^2$  viszonylag alacsony



5. ábra - A  $p$  paraméter hatása a két elemű magok fennmaradására. Az itt ábrázolt szimulációk során  $s_{\max} = 100$  rögzítve volt. A kék görbét LOESS formulával illesztettem exploratív célzattal, míg a piros egyenes az egyszerű lineáris regressziós modell predikciója. Az adatpontokat az R jitter() funkciójának segítségével megszórtam.

értékében is tükröződik ( $R^2 = 0.3726$ ). A modelldiagnosztikai ábrák vizsgálatával egy torzító pontot is észleltem.

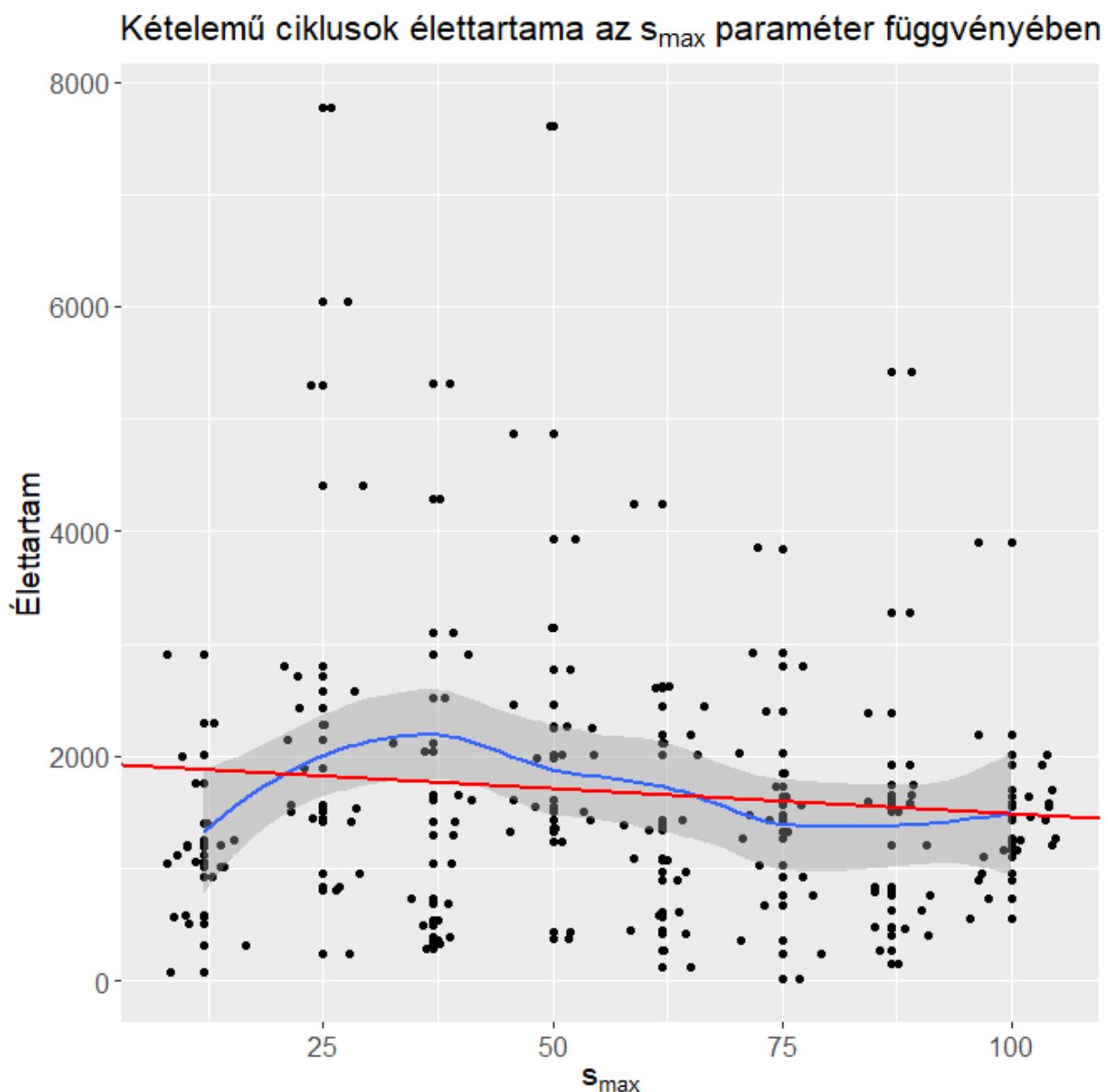
Mivel az exploratív elemzés alapján azt feltételeztem, hogy az élettartam exponenciálisan függ a  $p$  paraméter értékétől, így függő változóként az élettartam természetes alapú logaritmusát használtam a következő regressziós modellemben. A 6. ábrán látható, hogy a logtranszformált élettartam lineárisan függ  $p$ -től. A regressziós modell predikciói (piros egyenes) jól illeszkednek az adatpontjainkra. A regresszió szignifikáns ( $p < 0.0001$ ) és a korrigált  $R^2$  értéke is nagyobb. A modelldiagnosztikai ábrák szerint is megfelelő az illeszkedés, torzító pontok nincsenek. A modell predikciói szerint a  $p$  paraméter 0.01-el történő növelése a 2 elemű magok várható élettartamában bekövetkező  $\sim 0.13$ -szoros csökkenéssel fog járni.



6. ábra - A  $p$  paraméter hatása a két elemű magok fennmaradásának logaritmusára. Az itt ábrázolt szimulációk során  $s_{\max} = 100$  rögzítve volt. A kék görbét LOESS formulával illesztettem exploratív céllal, míg a piros egyenes a lineáris regressziós modell predikciója. Az adatpontokat az R jitter() függvényének segítségével megszórtam.

## A kételemű magok élettartama az $s_{max}$ paraméter függvényében

Az élettartam a maximális fajszámtól ( $s_{max}$ ) való függésének vizsgálatához a  $p$  paraméter értékét 0.005-ben rögzítettem, majd az elemzést ismét egy szórásdiagram (7. ábra) létrehozásával kezdtem. Első ránézésre nem állapítható meg konkrét összefüggés a változók között, az adatpontok eloszlása különböző  $s_{max}$  értékek mellett hasonló. Az adatokra egyszerű lineáris regressziós modellt illesztettem, melyben a függő változó a kételemű magok élettartama, a magyarázó változó pedig az  $s_{max}$  paraméter volt. A regresszió nem bizonyult szignifikánsnak ( $p = 0.2187$ ), valamint a modell magyarázó ereje is elenyészően kicsi ( $R^2 = 0.0033$ ).

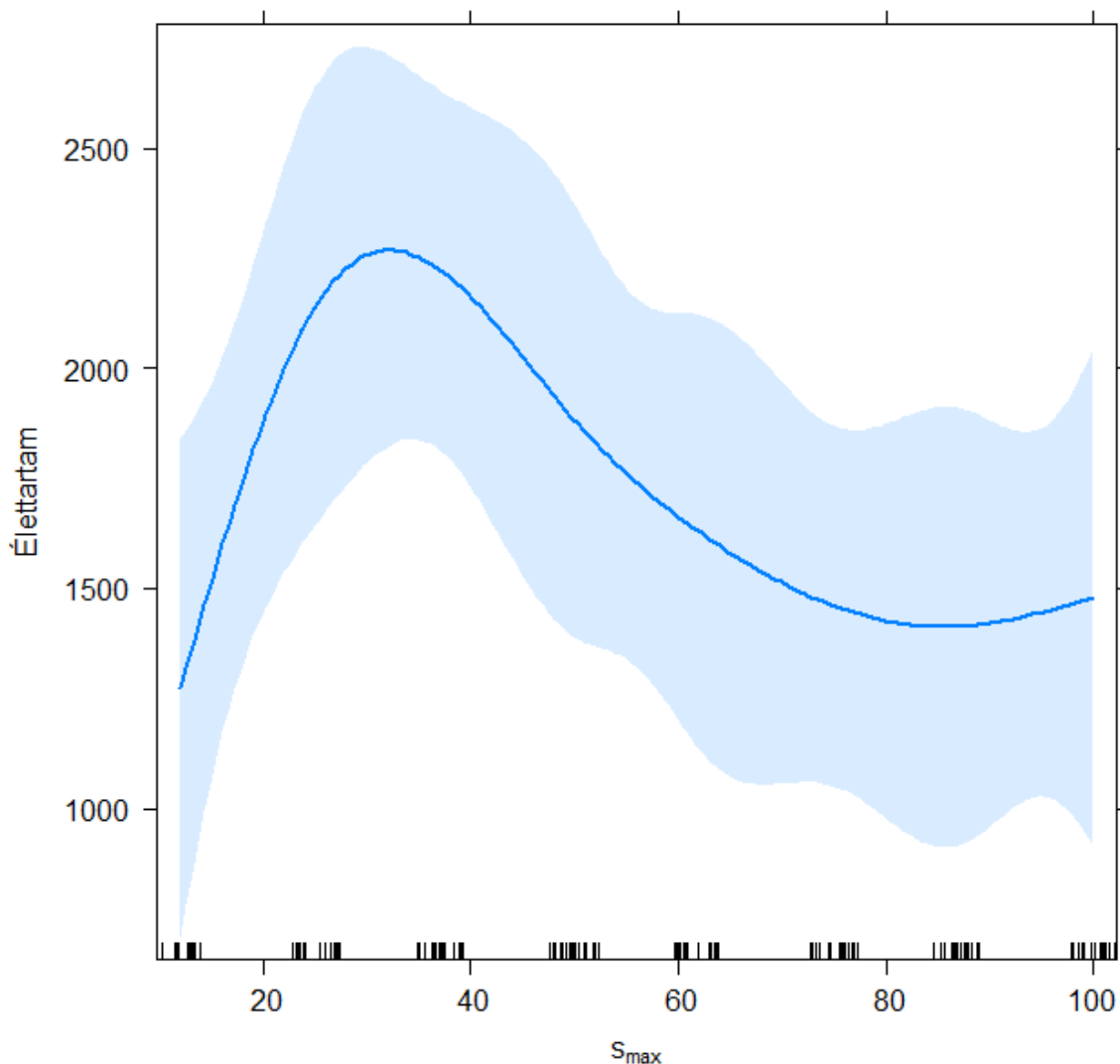


7. ábra - A kételemű magok fennmaradása az  $s_{max}$  paraméter függvényében. A szimulációk során  $p = 0.005$  rögzítve volt. A kék görbét LOESS formulával illeszttem exploratív célzattal, míg a piros egyenes az egyszerű lineáris regressziós modell predikciója. Az adatpontokat az R jitter() funkciójának segítségével megszórtam.



Mivel exploratív elemzésekkel nem sikerült összefüggést feltárnom, az adatokra nemparaméteres spline regressziót végeztem. A spline regresszió feltételezése, hogy a magyarázó változó(k) értékének különböző tartományaiban más és más módokon hat a függő változó értékére, így az egyes tartományokon a regressziós görbe egyenlete más és más paraméterekkel írható le. Mi szabhatjuk meg, hogy hány különböző egyenletet alkalmazunk, amit a szabadsági fok, vagy a csomópontok számának megadásával tehetünk meg. Az 8. ábrán látható a spline regresszió predikciós görbéje 5 szabadsági fok mellett.

### Az élettartam $s_{\max}$ -tól való függésének feltárása spline módszerrel



8. ábra - A kételemű magok élettartama az  $s_{\max}$  paraméter függvényében, rögzített  $p = 0.005$  mellett, spline regresszióval becsülve. A kék görbe a regresszió becslését, a kék sáv pedig a standard hibát ábrázolja.

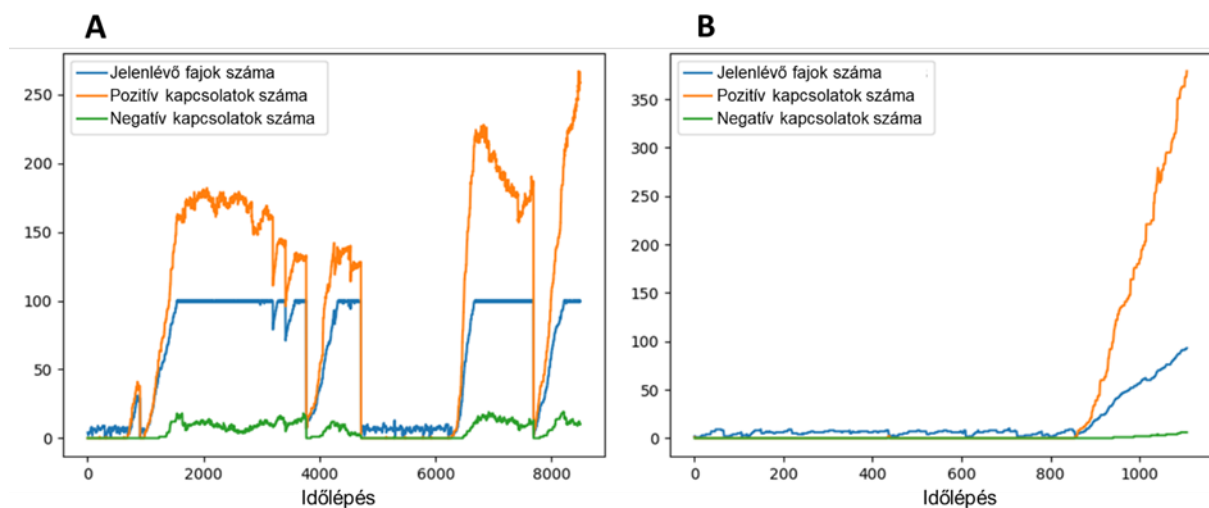
A spline regresszió egyik tartományon sem bizonyult szignifikánsnak, így az ábrázolt összefüggésből semmiképp sem vonhatunk le következtetéseket, bár segítségül szolgálhat a megfelelő paraméteres módszer megtalálásához. A spline görbe alakja alapján Poisson-

regressziós modellt illeszttem az adatokra, ám a modell magyarázó ereje nagyon kicsinek bizonyult, és sok torzító pontot tartalmazott. A kételemű magok élettartama és az  $s_{max}$  paraméter között tehát nem találtam szignifikáns összefüggést.

### 4.3.A bővített modell vizsgálata

A bővített modell különböző  $m$  és  $i$  paraméterek melletti működését a sűrűn kapcsolt állapot eléréséhez szükséges időlépések számának alapján vizsgáltam. A 9. ábrán látható, hogy a sűrűn kapcsolt állapot kialakulása előtt más-más paraméterek esetén különböző struktúrákat figyelhetünk meg. Az **A** ábrán egy alacsonyabb  $i$ -értékhez tartozó szimuláció mutatói láthatók, ekkor még jellemzően megfigyelhetők a JK modellben bemutatott fázisok, a magasabb  $i$ -értékű szimuláció mutatóit ábrázoló **B** ábrán azonban ilyesmit már nem láthatunk. A sűrűn kapcsolt állapot elérését azonban mindkét esetben a katalitikus élek számának meredek növekedése előzi meg.

### A bővített modell működése



9. ábra - Az bővített modell két szimulációjának eredményei. Az A ábrán egy  $m = 0.7$  és  $i = 0.3$  paraméterek melletti futás, míg a B ábrán egy  $m = 0.7$  és  $i = 0.7$  paraméterek melletti futás eredményei láthatók.

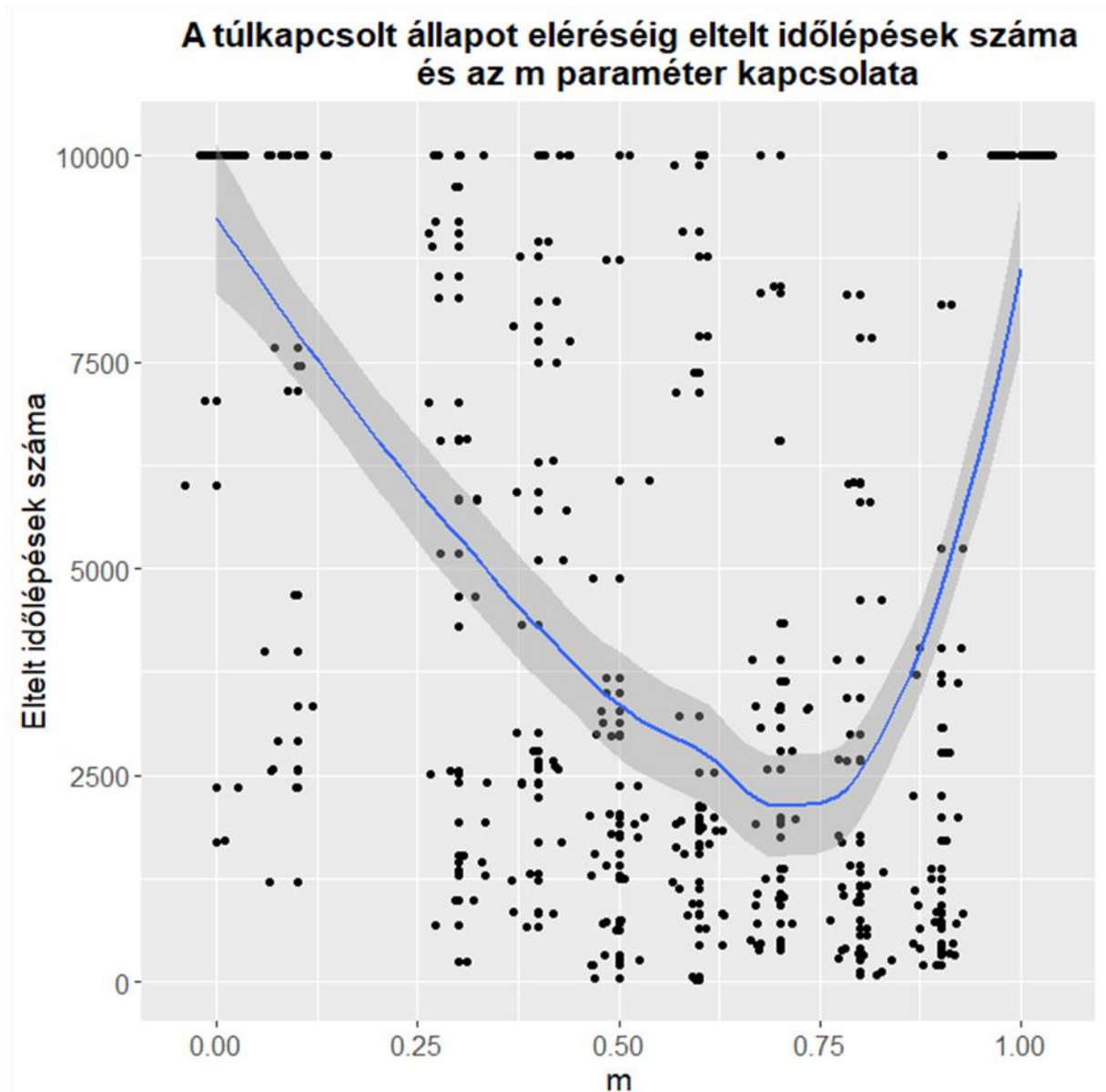
A szimulációkat maximum 10 000 időlépésen keresztül futtattam, rögzített  $p = 0.005$  és  $s_{max} = 100$  paraméterek mellett. Hogyha a 10 000. időlépésig nem értük el a sűrűn kapcsolt állapotot, az adattáblában 10 000 időlépést rögzítettem. Mivel a szimulációk jelentős része nem érte el a sűrűn kapcsolt állapotot a megszabott maximum előtt, a kapott eredmények torzultak, és leszűkült az alkalmazható statisztikai módszerek köre.

### A sűrű kapcsoltság kialakulásának vizsgálata az $m$ paraméter függvényében

A sűrű kapcsoltság eléréséhez szükséges időlépések számának migránsok arányától ( $m$ ) való függését rögzített  $i = 0.9$  paraméter mellett vizsgáltam. A 10. ábrán látható, hogy a sűrűn

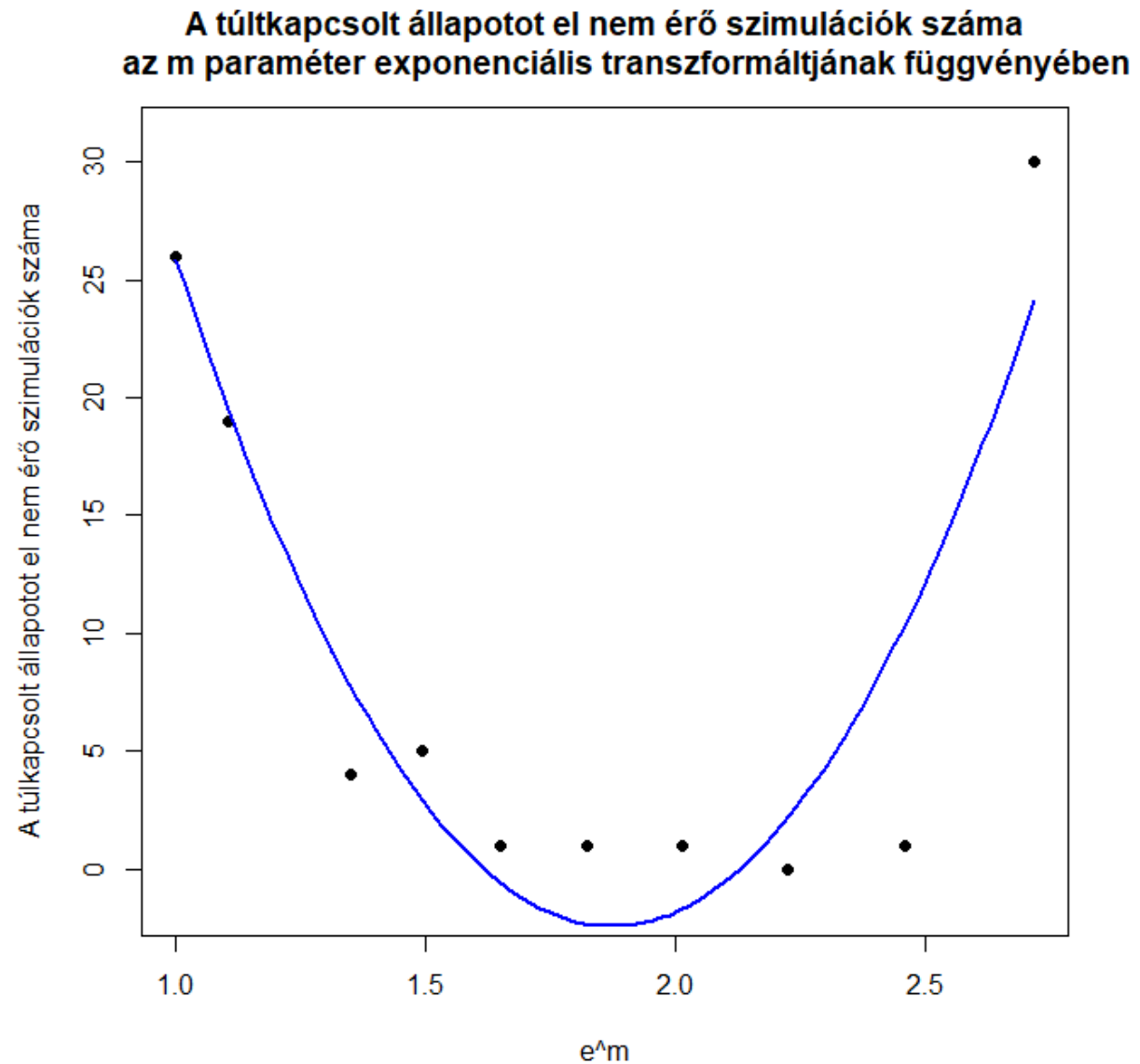
kapcsolt állapotig eltelt idő négyzetesen függ  $m$  értékétől, vagy  $m$  valamilyen transzformáltjától, ugyanis az adatpontokban kirajzolódó hiperbola balra ferde. Az adatokra kvadratikusan illesztettem, ám a modelldiagnosztika során kiderült, hogy a sok 10 000 időlépés-értéket felvevő adatpont miatt nem teljesültek a modell alkalmazhatósági feltételei, így más módszerhez kellett folyamodnom.

Létrehoztam egy új adattáblát, amiben az  $m$  paraméter minden vizsgált értékéhez rögzítettem, hogy az adott paraméterrel futtatott 30 szimuláció közül hány esetben nem alakult ki sűrű



10. ábra - A sűrűn kapcsolt állapot eléréséhez szükséges időlépések száma az  $m$  paraméter függvényében. Az adatpontokat az R `jitter()` funkciójával megszórtam. A kék görbét LOESS formulával illesztettem exploratív céllal.

kapcsoltság. Az így megkapott adatpontokra kvadratikus regressziós modellt illeszttem, és bár a regresszió szignifikáns volt ( $p = 0.0058$ ), és a modell az adatpontokban tapasztalt variancia több mint 70 százalékát magyarázta ( $R^2 = 0.7043$ ), nem voltam elégedett az illeszkedéssel, mivel az adatpontok által kirajzolt parabola továbbra is balra ferde volt, amit ez a modell nem tudott megjeleníteni.



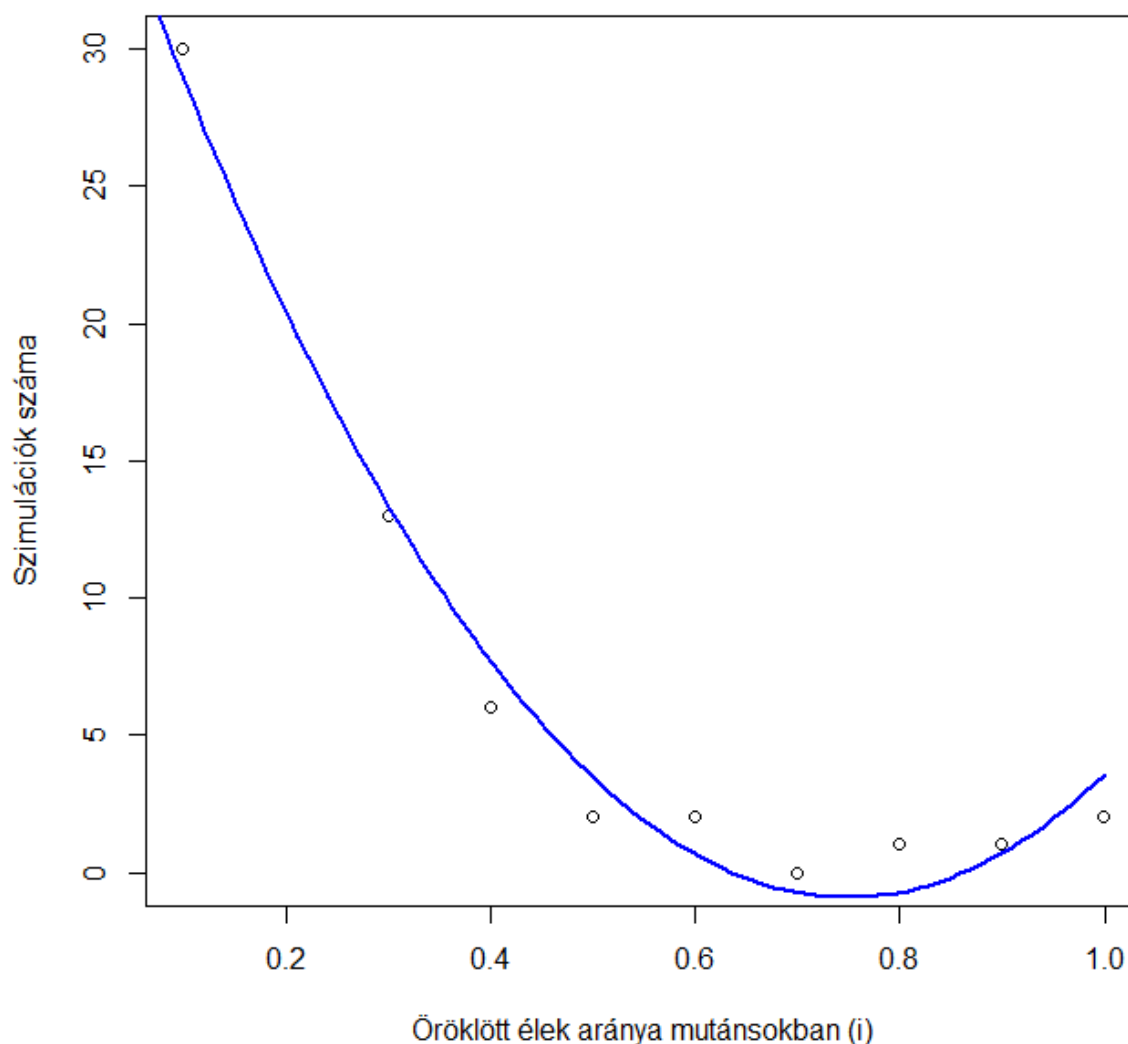
11. ábra - A sűrűn kapcsolt állapotot el nem érő szimulációk száma az  $m$  paraméter exponenciális transzformáltjának függvényében. A kék görbe a kvadratikus regressziós modell predikcióját ábrázolja.

A megfelelő regressziós modell illesztéséhez az **m** paraméter értékeit exponenciálisan transzformáltam, és ezt használtam magyarázó változóként a kvadratikus regressziós modellemben. A 11. ábrán látható, hogy a modell predikciója jól illeszkedik az adatpontokra. A regresszió szignifikánsnak bizonyult ( $p = 0.0010$ ), a modell magyarázó ereje az előbbihez képest nőtt ( $R^2 = 0.8232$ ). A modell predikciója alapján a sűrűn kapcsolt állapot elérése az **m** paraméter 0.62-es értéke mellett fog a legnagyobb eséllyel bekövetkezni a 10 000. időlépés előtt.

### **A sűrű kapcsoltság kialakulásának vizsgálata az **i** paraméter függvényében**

A sűrű kapcsoltság kialakulásának a mutánsokban öröklődő kapcsolatok arányától (**i**) való függését rögzített **m = 0.7** paraméter mellett vizsgáltam. Az exploratív elemzések alapján a függés ismét kvadratikus jellegű, ám az **m** paraméter esetén végzett vizsgálatokhoz hasonlóan a kvadratikus regressziós modell alkalmazási feltételei itt sem teljesültek, így az adatokat ismét átalakítottam úgy, hogy az egyes **i** értékekhez a sűrűn kapcsolt állapotot el nem érő szimulációk száma tartozzon. Az így megkapott adatpontokra kvadratikus regressziós modellt illesztettem. A 12. ábrán látható, hogy a modell predikciója jól illeszkedik az adatokra. A regresszió szignifikánsnak bizonyult ( $p < 0.0001$ ), a modell magyarázó ereje nagy ( $R^2 = 0.9751$ ). A modell predikciója alapján a 0.74 – es **i** érték mellett a legnagyobb a valószínűsége a sűrűn kapcsolt állapot elérésének a 10 000. időlépés előtt.

## A túltkapcsolt állapotot el nem érő szimulációk száma az $i$ paraméter függvényében



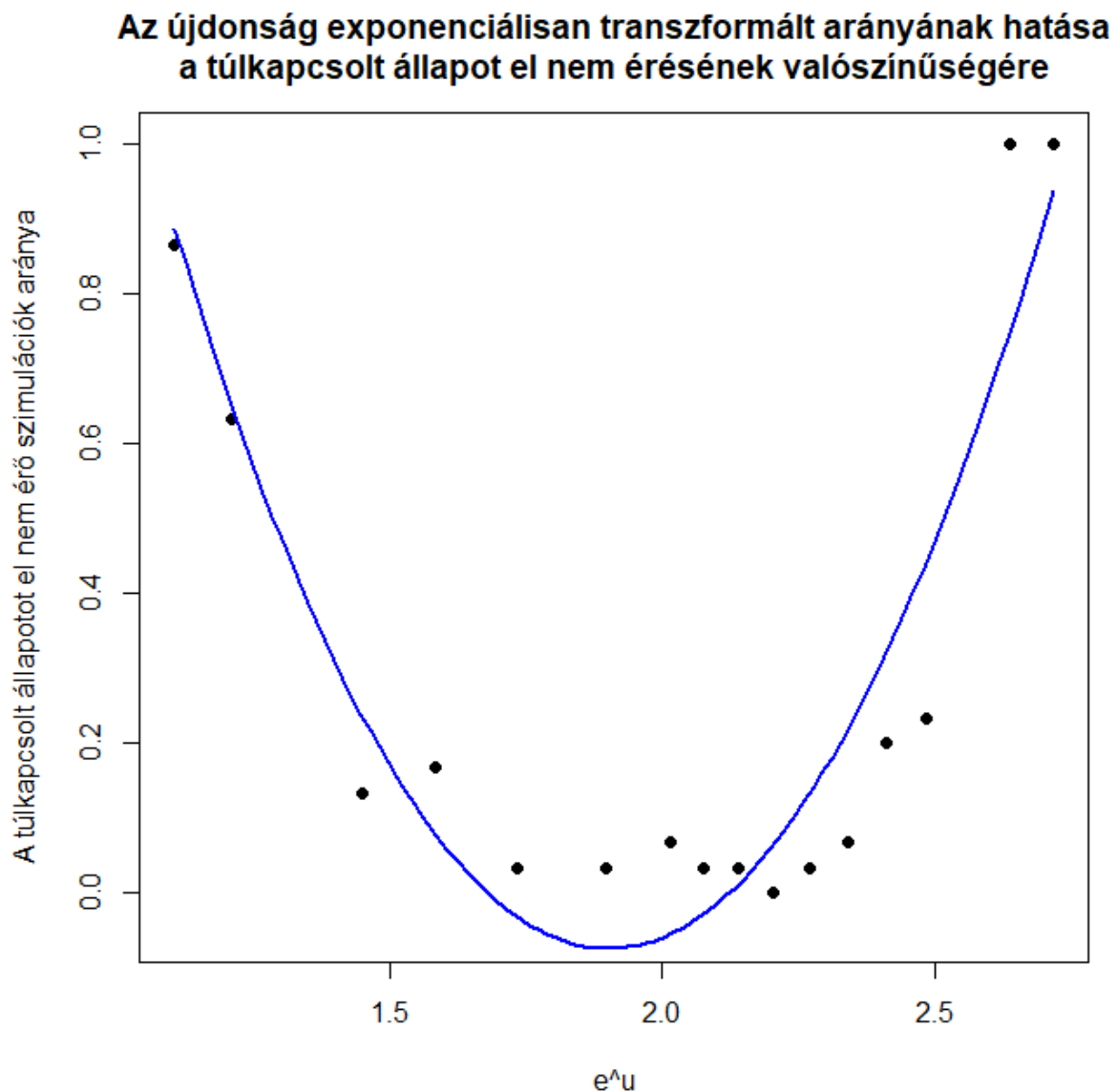
12. ábra - A sűrűn kapcsolt állapotot el nem érő szimulációk száma az  $i$  paraméter függvényében. A kék görbe a kvadratikus regressziós modell predikcióját ábrázolja.

### A sűrű kapcsoltság kialakulásának vizsgálata az összesített adatok alapján

Az  $m$  és  $i$  paraméterek együttes hatását a szimulációkban várható újdonság arányával vizsgáltam. A várható újdonság arányát az  $u = m + (1 - m) * (1 - i)$  képlet alapján számítottam ki, ugyanis a rendszerben megjelenő migránsok teljes kapcsolatrendszere új, míg a mutánsokban csak a nem öröklődő éleket tekintjük újdonságnak. Az  $m$  és  $i$  paraméterek vizsgálata során futtatott összes szimulációra kiszámoltam  $u$  értékét, így összesen 570 szimuláció alapján becsülhettem meg a sűrű kapcsoltság kialakulásának valószínűségére gyakorolt hatását. Mivel az eltérő  $u$  értékekhez különböző számú szimuláció tartozott, az

adatokat lenormáltam az adott érték mellett végzett szimulációk számával, így minden értékhez megkaptam, hogy a szimulációk milyen arányban érték el a sűrűn kapcsolt állapotot. Az így létrehozott 16 megfigyelésből álló adatsort használtam függő változóként regressziós modelleimben.

Az lenormált változó újdonság arányától való függése az exploratív elemzések alapján szintén kvadratikusan összefüggést mutatott. Ennek megfelelően először kvadratikusan regressziós modellt illesztettem az adatokra, ám a nagy magyarázó erő ( $R^2 = 0.7656$ ) és szignifikancia szint ( $p < 0.0001$ ) ellenére nem voltam elégedett az illeszkedéssel. A 13. ábrán már az  $u$  mutató exponenciális transzformáltjával illesztett kvadratikusan regressziós modell predikciója látható.



13. ábra - Annak a valószínűsége, hogy nem alakul ki sűrűn kapcsolt állapot a 10 000. időlépés előtt, az  $u$  mutató exponenciális transzformáltjának függvényében. A kék görbe a kvadratikusan regressziós modell predikcióját ábrázolja.

A modell magyarázó ereje a transzformáció hatására javult ( $R^2 = 0.8771$ ), a regresszió továbbra is szignifikáns ( $p < 0.0001$ ), a predikciós görbe jobban illeszkedik az adatpontokra. A modell becslése szerint az **u** mutató 0.64-es értéke mellett a legnagyobb a valószínűsége, hogy a 10 000. időlépés előtt elérjük a sűrűn kapcsolt állapotot.

## 5. Következtetések

### 5.1. A JK modell eredményeinek kiértékelése

A reprodukált JK modellel végzett szimulációkból a különböző mutatók értékei és a kapcsolatrendszer szerkezetében megfigyelhető fázisok alapján jó közelítéssel visszakaptam az eredeti modell eredményeit. A JK modell számos különböző változatának [12, 39–41] tanulmányozása során nem találtam egyértelmű választ arra, hogy a relatív populációméreték milyen módon vannak a nemnegatív tartományra korlátozva, egyedül Goldstein publikációjában [41] találtam említést „elhanyagolható” populációméretű nóduszokról, ám az elhanyagolás küszöbértéke itt sem derül ki. Az eredmények hasonlósága és Goldstein cikke alapján mindenesetre valószínűsíthető, hogy az eredeti modell is alkalmaz küszöbértéket, és ez nagyságrendileg megegyezik az implementációmban alkalmazott  $0.01/s_{max}$  értékkel.

A reprodukált JK modellben a három elemű magok sokkal ritkábbnak bizonyultak a két elemű magoknál, ami összhangban áll az eredeti modell eredményeivel [9]. A két elemű magok élettartama és az  $s_{max}$  paraméter értékei között nem találtam összefüggést rögzített **p** paraméter mellett, a **p** paraméter értékének növekedésével azonban exponenciálisan csökkent a két elemű magok fennmaradása. Ez elsőre úgy tűnhet, hogy ellentmond Mehrotra JK modellen végzett elemzéseinek [40], aki azt állapította meg, hogy az  $s_{max}$  nagyobb értékei mellett ritkább az autokatalitikus halmaz szétesése, illetve nagyobb **p** értékek mellett az autokatalitikus halmaz ellenállóbbá válik a szétesésekkel szemben. Fontos azonban kiemelni, hogy a JK modellben végzett vizsgálataim során nem a teljes autokatalitikus halmazokat, hanem fajok közötti kétrésztvevős interakciókat. Egy jól konnektált magban az általunk vizsgált kételemű mag egyik tagjának kipusztulása nem feltétlenül fog a teljes mag szétesésével járni, így fordulhat elő, hogy nagyobb **p** értékek esetén a nagyobb konnektáltság miatt ellenállóbbak, a bennük található két elemű magok mégis átlagosan rövidebb ideig maradnak fent, feltehetőleg az inhibitorikus élek gyorsabb cserélődése miatt.

A két elemű ciklusok élettartamát azért lenne fontos baktériumközösségekre kialakított modellekben is vizsgálni, mert ekkor egy autokatalitikus szettben a kételemű magok szétesése a reakciópartnerek cserélődését jelentené, ami alapján lehetne valószínűsíteni, hogy két faj egyedei mennyire lehetnek képesek evolúciós léptékkal is értelmezhető időn keresztül



kölcsönös függésben élni. Hogyha a páros kapcsolatok élettartama evolúciós időskálán elhanyagolhatónak bizonyulna, akkor ez fontos bizonyíték lehetne a mitokondrium kialakulásának szintóf elméletével szemben (ami hosszú ideig stabilan fennálló metabolikus ektoszimbiózist feltételez a fizikai integráció előtt). Hogyha jelentősnek bizonyulna, az az elmélet helyességét valószínűsítene. Mindehhez persze pontos becslésekre volna szükség a szintrof baktériumközösségek mutációs rátájára és egyéb paramétereire vonatkozólag. A jövőben ilyen jellegű adatok gyűjtésével tervezem a modellt realisztikusan paraméterezni.

## 5.2. Az bővített modell eredményeinek kiértékelése

A bővített modell vizsgálata során az első fontos tapasztalat a sűrű kapcsolt állapot gyors megjelenése volt. Ez az állapot akkor alakulhat ki, hogyha a szelekciós események után megjelenő fajokban az újdonság átlagos aránya kisebb mint 1, azaz hogyha az új faj a kapcsolatrendszerének egy részét a rendszerben már jelenlévő nóduszoktól öröklí. Mehrotra és társai hasonló eredményeket kaptak a JK modell egy átalakított változatából, melyben az újonnan megjelenő nóduszok a nagyobb populációméretű nóduszokkal nagyobb eséllyel alakítanak ki kapcsolatot. [40]. A szerzők ellenálló, jól kapcsolt, sokrésztvevős autokatalitikus magok kialakulásáról számolnak be, melyekben az elemi hurkok számát is megállapították. A modellek jobb összehasonlításához az általam bővített modellben is nyomon kellene követni az elemi hurkok számának alakulását, ez a jövőbeli munkám egyik célkitűzése lehet.

A bővített modellben a mutáció/migráció arányára vonatkozó paraméter (**m**) exponenciális transzformáltjától, valamint a mutánsokban öröklődő kapcsolatok arányától (**i**) egyaránt kvadratikusan függött a sűrű kapcsoltság kialakulásának esélye a megszabott 10 000 időlépésen belül. A két paraméterből számolt újdonság arányának exponenciális transzformáltjától (**u**) szintén kvadratikusan függést mutattam ki. Ez arra enged következtetni, hogy a két paraméter szoros interakcióban fejti ki hatását, így a jövőbeli vizsgálatok során a jelenlegi két paraméter helyett érdemes lehet a modellben az **u** mutatót paraméterként meghonosítani.

Az sűrűn kapcsolt állapot **u**-tól való kvadratikusan függése arra enged következtetni, hogy a szoros interakciós hálózatot kialakító baktériumközösségek kialakulása az öröklött és újonnan megjelenő kapcsolatok nagyjából 1:2 aránya mellett a legvalószínűbb. Ez azért alakulhat így, mert az újonnan megjelenő kapcsolatok alacsony aránya mellett jellemzően a domináns baktériumfaj mutánsai fognak megjelenni a rendszerben, amik egymás növekedésére nem hathatnak katalitikusan, hiszen hasonló erőforrásokat hasznosítanak, amit a modell úgy jelenít meg, hogy egymás képződésére csak kicsi [ $(1-i) \times p$ ] eséllyel hathatnak katalitikusan. Az újdonság magas aránya mellett viszont az autokatalitikus magban szereplő fajok mutánsai nem

terjednek el időben, az autokatalitikus mag köré kiterjedt autokatalitikus halmaz rendeződik, melyben a periféria tagjai definíció szerint nem katalizálják a maghoz tartozó fajok növekedését, így a mag szétesésével járó kihalási események hátráltathatják a sűrűn kapcsolt állapot elérését.

### **5.3.Tervek és javaslatok**

A sűrűn kapcsolt állapot vizsgálata, bár érdekes eredményeket produkált, de nem adott egyértelmű választ a kutatás eredeti kérdésére, ami a két résztvevős autokatalitikus magok fennmaradásának idejére vonatkozott. Ennek a vizsgálata akár a modell jelenlegi változatában is lehetséges volna jobb teljesítményű számítógépeken, vagy a ciklusszám meghatározása nélkül, ami feltehetőleg sokat javítana a szimulációk futási idején. Azonban terveim szerint a modelletem tovább fogom optimalizálni, hogy realiztikusabban ábrázolja a szintróf mikrobiális közösségek működését, így az eredmények valóban közelebb vihetnének minket a mitokondrium szintróf elméletének elfogadásához vagy cáfolásához.

A modell realiztikusabbá tételéhez szeretném bevezetni az populációk saját méretétől is függő, 3. egyenlet szerinti növekedést. Ennek a modellnek az eredményeit a megfelelő számú tesztfutás és szimuláció hiányában nem tárgyaltam a dolgozatomban, azonban egyértelmű utat jelöl ki számomra a jövőbeli munka menetére vonatkozólag. Szeretnék a random kapcsolatrendszeren alapuló replikátorok közösségeit leíró modellek [42, 43] körében újat alkotni, a szintróf mikrobiális közösségek több jellemzőjének megjelenítésével.

Egy ilyen jellemző lehet a kompetíció magas aránya a szimbiotikus kapcsolatokhoz képest, becslések szerint ugyanis a mikrobiális közösségek interakcióinak legfeljebb 14 százaléka kooperatív [13]. A kapcsolatok kialakításának Vural modelljéhez [44] hasonlóan lehetne költsége, ugyanis a modellem egyik fő gyengesége, hogy a populációk növekedése nem függ attól, hogy az adott faj hány másik faj növekedésére hat, annak ellenére, hogy mind a toxinok szekréciója, mind a közjó előállítás költséges folyamatok lehetnek [45, 46].

A mitokondrium kialakulásának megfelelő modellezéséhez további újítás lehetne a kétrésztvevős autokatalitikus magok interakcióinak erősödése a mag élettartamának növekedésével. Ezzel a metabolikus partnerek egyre szorosabb függése is meg lenne jelenítve a modellben.

## 6. Összefoglaló

Kutatásom során a mitokondrium kialakulásának szintróf elméletét vizsgáltam egy matematikai modell segítségével. Az elmélet szerint a mitokondrium egy szintróf mikrobiális közösségben alakulhatott ki, két mikroba faj egyre szorosabb metabolikus függése révén.

A szintróf mikrobiális közösségeket katalitikus gráfokként modelleztem, melyekben a nóduszok mikroba fajok populációit, a nóduszok között húzódó irányított élek pedig az egyes populációk egymás növekedésére gyakorolt hatásának eredőit jelentik. A mikroba fajok populációit evolúciós folyamatoknak vettem alá, mely során a populációk növekedése a kapcsolatrendszer alapján történt, és a legkisebb számban jelenlévő mikrobák rendre kipusztultak, helyükön új fajok populációi jelentek meg. A mikrobák kapcsolatrendszerét leíró gráfban a két résztvevős katalitikus élekből álló magok mutációs események számában mért fennmaradását vizsgáltam. Ezek olyan önfenntartó struktúrák, amik szoros függést jelezhetnek számunkra, így ezek segítségével jeleníthetjük meg a modellben a mitokondrium endoszimbionta őseit. A modellem megalkotásához Jain és Krishna idealizált kémiai rendszerek működését bemutató modelljét használtam fel.

A munkám során először reprodukáltam a Jain Krishna modellt, ami az eredmények alapján jól közelítette az eredeti modell működését. A reprodukált modellben a kételemű autokatalitikus magok fennmaradása a katalitikus gráfban az élek kialakulásának valószínűségét megadó paraméter ( $p$ ) növekedésével exponenciálisan csökkent, míg a maximális fajszámtól ( $S_{max}$ ) nem függött.

A reprodukált Jain - Krishna modellt kibővítettem, hogy jobban közelítse a szintróf mikrobiális közösségek működését. Bevezettem a mutáció és a migráció ( $m$ ), illetve a mutánsok által öröklött élek arányát ( $i$ ) meghatározó paramétereket. Az új paramétereket az eredeti modellben szereplő paraméterek rögzített értékei mellett vizsgáltam. Az öröklődő kapcsolatok bevezetésének hatására a mikrobák kapcsolatrendszere a legtöbb szimuláció során sűrűn kapcsolttá vált, ami kezelhetetlenül lassú futási időket eredményezett. Így a sűrűn kapcsolt állapot 10 000. mutációs lépés előtti elérésének valószínűsége alapján vizsgáltam az új paraméterek hatását. Ez a valószínűség mindkét paraméter értékétől kvadratikusan függött, így bevezettem az újdonság arányára vonatkozó mutatót, amit az  $m$  és  $i$  paraméterek értékéből számoltam. Az újdonság arányától szintén kvadratikusan függött a sűrűn kapcsolt állapot kialakulásának valószínűsége, így a jövőbeli munkám során ezt tervezem paraméterként bevezetni a modellbe az  $m$  és  $i$  paraméterek helyett. A dolgozat végén további terveket és javaslatokat fogalmazok meg saját jövőbeli munkámra vonatkozólag.

## 7. English title and summary

### **Emergence and dynamics of symbiosis in syntrophic microbial communities modelled as catalytic graphs**

In this paper, I investigated the syntrophic theory of mitochondrial formation using a mathematical model. The theory proposes that mitochondria may have evolved in a syntrophic microbial community as a result of the increasingly close metabolic dependence of two microbial species.

I used catalytic graphs as a means of modelling syntrophic microbial communities where the nodes represent the populations of microbial species and the directed edges between the nodes are the sums of the effects of populations on each other's growth. I subjected the populations of microbial species to evolutionary processes, during which the growth of the populations was calculated based on the system of their relationships. In each evolutionary step, the smallest species with the smallest population had gone extinct, and a small population of a new species appeared in its place. I examined the survival of autocatalytic cores consisting of two participants measured with the number of survived mutation events. Autocatalytic cores are self-sustaining structures, they may indicate a close dependence, which makes them a useful tool for depicting the ancestors of the mitochondrial endosymbiont in the model. I used Jain and Krishna's 2001 model for idealized chemical reaction systems as the basis of building my model.

I implemented the Jain Krishna model, which, based on my results, appeared to have the same structure as the original model. In the reproduced model, the survival of 2-sized autocatalytic cores in the catalytic graph decreases exponentially with increasing numbers of the parameter indicating the probability of edge formation ( $\mathbf{p}$ ), while it did not depend on the maximum number of species ( $\mathbf{S}_{\max}$ ).

I redesigned the reproduced Jain - Krishna model to capture more accurately the functioning of syntrophic microbial communities. I introduced the parameters determining the mutation and migration ( $\mathbf{m}$ ) and the ratio ( $\mathbf{i}$ ) of the edges inherited by mutants. I examined the new parameters while keeping the parameters in the original model fixed. As a result of the newly introduced inherited connections, the microbe populations became densely connected during most simulations, resulting in unmanageably slow run times. Thus, I examined the effect of the new parameters based on the probability of reaching the densely connected state before the 10,000th mutation step. This probability depended quadratically on the values of both parameters, so I introduced the novelty ratio index ( $\mathbf{u}$ ), which I calculated from the values of the parameters  $\mathbf{m}$  and  $\mathbf{i}$ . The probability of the formation of a densely coupled state also quadratically depended

on the ratio of novelty, so in my future work I plan to introduce this as a parameter in the model instead of the **m** and **i** parameters. At the end of my dissertation, I formulate further plans and suggestions for my own future work.

## 8. Hivatkozások

1. Huppert A, Katriel G (2013) Mathematical modelling and prediction in infectious disease epidemiology. *Clin. Microbiol. Infect.* 19
2. Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. *Nature* 525
3. Gabaldón T (2021) Origin and Early Evolution of the Eukaryotic Cell. *Annu. Rev. Microbiol.* 75
4. Fiore M (2019) The Origin and Early Evolution of Life: Prebiotic Chemistry. *Life* 9:.. <https://doi.org/10.3390/life9030073>
5. Zachar I, Szathmáry E (2017) Breath-giving cooperation: Critical review of origin of mitochondria hypotheses. *Biol. Direct* 12
6. Lake JA, Rivera MC (1994) Was the nucleus the first endosymbiont? *Proc. Natl. Acad. Sci. U. S. A.* 91
7. Martin WF, Garg S, Zimorski V (2015) Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* 370
8. Sousa C (2021) Origin of Life: An Update on New Evidence & Theories. *Am Biol Teach* 83:.. <https://doi.org/10.1525/abt.2021.83.2.76>
9. Jain S, Krishna S (2001) A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proc Natl Acad Sci U S A* 98:.. <https://doi.org/10.1073/pnas.98.2.543>
10. Martin WF (2010) The origin of mitochondria. *Nature education. Nat Educ* 3:
11. Moreira D, López-García P (1998) Symbiosis between methanogenic archaea and  $\delta$ -proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J Mol Evol* 47:.. <https://doi.org/10.1007/PL00006408>
12. Krause A (2004) Predicting crashes in a model of evolving networks. *Complexity* 9:.. <https://doi.org/10.1002/cplx.20017>
13. Mitri S, Richard Foster K (2013) The genotypic view of social interactions in microbial communities. *Annu. Rev. Genet.* 47

14. Nadell CD, Drescher K, Foster KR (2016) Spatial structure, cooperation and competition in biofilms. *Nat. Rev. Microbiol.* 14
15. Schink B (2002) Synergistic interactions in the microbial world. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* 81:.. <https://doi.org/10.1023/A:1020579004534>
16. Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C (2013) Microbial syntrophy: Interaction for the common good. *FEMS Microbiol. Rev.* 37
17. Riley MA, Wertz JE (2002) Bacteriocins: Evolution, ecology, and application. *Annu. Rev. Microbiol.* 56
18. Hood RD, Singh P, Hsu FS, Güvener T, Carl MA, Trinidad RRS, Silverman JM, Ohlson BB, Hicks KG, Plemel RL, Li M, Schwarz S, Wang WY, Merz AJ, Goodlett DR, Mougous JD (2010) A Type VI Secretion System of *Pseudomonas aeruginosa* Targets a Toxin to Bacteria. *Cell Host Microbe* 7:.. <https://doi.org/10.1016/j.chom.2009.12.007>
19. Xavier JB, Foster KR (2007) Cooperation and conflict in microbial biofilms. *Proc Natl Acad Sci U S A* 104:.. <https://doi.org/10.1073/pnas.0607651104>
20. Kaufmann S (2011) Economics and the collectively autocatalytic structure of the real economy. 13.7 Cosmos and Culture. *Natl Public Radio*
21. Kauffman SA (1986) Autocatalytic sets of proteins. *J Theor Biol* 119:.. [https://doi.org/10.1016/S0022-5193\(86\)80047-9](https://doi.org/10.1016/S0022-5193(86)80047-9)
22. Hordijk W, Hein J, Steel M (2010) Autocatalytic sets and the origin of life. *Entropy* 12
23. Wynveen A, Fedorov I, Halley JW (2014) Nonequilibrium steady states in a model for prebiotic evolution. *Phys Rev E - Stat Nonlinear, Soft Matter Phys* 89:.. <https://doi.org/10.1103/PhysRevE.89.022725>
24. Kauffman SA (1971) Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *J Cybern* 1:.. <https://doi.org/10.1080/01969727108545830>
25. Hordijk W, Steel M (2004) Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J Theor Biol* 227:.. <https://doi.org/10.1016/j.jtbi.2003.11.020>
26. Steel M (2000) The emergence of a self-catalysing structure in abstract origin-of-life models. *Appl Math Lett* 13:.. [https://doi.org/10.1016/S0893-9659\(99\)00191-3](https://doi.org/10.1016/S0893-9659(99)00191-3)
27. Hordijk W (2016) Evolution of Autocatalytic Sets in Computational Models of Chemical Reaction Networks. *Orig Life Evol Biosph* 46:.. <https://doi.org/10.1007/s11084-015-9471-0>
28. Jain S, Krishna S (2002) Large extinctions in an evolutionary model: The role of innovation and keystone species. *Proc Natl Acad Sci U S A* 99:..

<https://doi.org/10.1073/pnas.032618499>

29. Van Rossum G (2007) Python Programming language. In: USENIX annual technical conference
30. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. *Nature* 585
31. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli A Pietro, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D V., Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17:. <https://doi.org/10.1038/s41592-019-0686-2>
32. McKinney W (2010) Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference
33. Hunter JD (2007) Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering*. *Comput Sci Eng* 9:
34. Waskom M (2021) seaborn: statistical data visualization. *J Open Source Softw* 6:. <https://doi.org/10.21105/joss.03021>
35. Lison A (2019) resilience-jain-krishna
36. Tiernan JC (1970) An Efficient Search Algorithm to Find the Elementary Circuits of a

- Graph. Commun ACM 13:. <https://doi.org/10.1145/362814.362819>
37. Levin BR, Bergstrom CT (2000) Bacteria are different: Observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 97
  38. R Core Team (2020) (2020) R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria
  39. Zhang RQ, Wei W, Guo BH, Zhang Y, Zheng ZM (2014) Robustness of autocatalytic set in a model of evolving network. In: *Applied Mechanics and Materials*
  40. Mehrotra R, Soni V, Jain S (2009) Diversity sustains an evolving network. *J R Soc Interface* 6:. <https://doi.org/10.1098/rsif.2008.0412>
  41. Goldstein RA (2006) Emergent robustness in competition between autocatalytic chemical networks. *Orig Life Evol Biosph* 36:. <https://doi.org/10.1007/s11084-006-9014-9>
  42. Kärenlampi PP (2014) Symmetry of interactions rules in incompletely connected random replicator ecosystems. *Eur Phys J E* 37:. <https://doi.org/10.1140/epje/i2014-14056-7>
  43. Diederich S, Oppen M (1989) Replicators with random interactions: A solvable model. *Phys Rev A* 39:. <https://doi.org/10.1103/PhysRevA.39.4333>
  44. Vural DC, Isakov A, Mahadevan L (2015) The organization and control of an evolving interdependent population. *J R Soc Interface* 12:. <https://doi.org/10.1098/rsif.2015.0044>
  45. Pintar J, Starmer WT (2003) The costs and benefits of killer toxin production by the yeast *Pichia kluyveri*. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* 83:. <https://doi.org/10.1023/A:1014215200360>
  46. Smith P, Schuster M (2019) Public goods and cheating in microbes. *Curr. Biol.* 29

## 9. Melléklet

Az bővített modell kódja: <https://colab.research.google.com/drive/1BXxqrWS2f2cBkDtLkU2tR6-6iqadozAb?usp=sharing>



## **10. Köszönetnyilvánítás**

Szeretnék köszönetet mondani Témavezetőimnek, Dr. Zachar Istvánnak és Dr. Boza Gergelynek, akik partnerként kezeltek engem a közös munka minden aspektusában, valamint belső konzulensemnek, Dr. Kövér Szilviának, aki vette a fáradságot és megpróbálta megérteni a gyakran túlságosan kusza magyarázataimat, és fontos észrevételekkel segítette a munkámat. Köszönettel tartozom továbbá Rurecht Ádám szaktársamnak, aki kitartóan unszolt szakdolgozatom módszeres alakítására.

## 11. Nyilatkozatok

### HuVetA

#### ELHELYEZÉSI MEGÁLLAPODÁS ÉS SZERZŐI JOGI NYILATKOZAT\*

Név: Máté Jakab

Elérhetőség (e-mail cím): alajos.jakab@gmail.com

A feltöltendő mű címe: A szimbiózis megjelenése és dinamikája katalitikus gráfokként modellezett szintróf mikrobiális közösségekben

A mű megjelenési adatai: 2022. Budapest

Az átadott fájlok száma: 1

---

Jelen megállapodás elfogadásával a szerző, illetve a szerzői jogok tulajdonosa nem kizárólagos jogot biztosít a HuVetA számára, hogy archiválja (a tartalom megváltoztatása nélkül, a megőrzés és a hozzáférhetőség biztosításának érdekében) és másolásvédett PDF formára konvertálja és szolgáltatassa a fenti dokumentumot (beleértve annak kivonatát is).

Beleegyeznek, hogy a HuVetA egynél több (csak a HuVetA adminisztrátorai számára hozzáférhető) másolatot tároljon az Ön által átadott dokumentumból kizárólag biztonsági, visszaállítási és megőrzési célból.

Kijelenti, hogy az átadott dokumentum az Ön műve, és/vagy jogosult biztosítani a megállapodásban foglalt rendelkezéseket arra vonatkozóan. Kijelenti továbbá, hogy a mű eredeti és legjobb tudomása szerint nem sérti vele senki más szerzői jogát. Amennyiben a mű tartalmaz olyan anyagot, melyre nézve nem Ön birtokolja a szerzői jogokat, fel kell tüntetnie, hogy korlátlan engedélyt kapott a szerzői jog tulajdonosától arra, hogy engedélyezhesse a jelen megállapodásban szereplő jogokat, és a harmadik személy által birtokolt anyagrészt mellett egyértelműen fel van tüntetve az eredeti szerző neve a művön belül.

A szerzői jogok tulajdonosa a hozzáférés körét az alábbiakban határozza meg (**egyetlen, a megfelelő négyzetben elhelyezett x jellel**):

- engedélyezi, hogy a HuVetA-ban -ban tárolt művek korlátlanul hozzáférhetővé váljanak a világhálón,
- az Állatorvostudományi Egyetem belső hálózatára (IP címeire) korlátozza a feltöltött dokumentum(ok) elérését,
- a Könyvtárban található, dedikált elérést biztosító számítógépre korlátozza a feltöltött dokumentum(ok) elérését,
- csak a dokumentum bibliográfiai adatainak és tartalmi kivonatának feltöltéséhez járul hozzá (korlátlan hozzáféréssel),

Kérjük, **nyilatkozzon a négyzetben elhelyezett jellel a helyben használatról is:**



Engedélyezem a dokumentum(ok) nyomtatott változatának helyben olvasását a könyvtárban.

Amennyiben a feltöltés alapját olyan mű képezi, melyet valamely cég vagy szervezet támogatott illetve szponzorált, kijelenti, hogy jogosult egyetérteni jelen megállapodással a műre vonatkozóan.

A HuVetA üzemeltetői a szerző, illetve a jogokat gyakorló személyek és szervezetek irányában nem vállalnak semmilyen felelősséget annak jogi orvoslására, ha valamely felhasználó a HuVetA-ban engedéllyel elhelyezett anyaggal törvénytisztító módon visszaélne.

Budapest, 2022 év április hó 28 nap



aláírás

szerző/a szerzői jog tulajdonosa

---

*A HuVetAMagyar Állatorvos-tudományi Archívum – Hungarian Veterinary Archive az Állatorvostudományi Egyetem Hutýra Ferenc Könyvtár, Levéltár és Múzeum által működtetett egyetemi és szakterületi online adattár, melynek célja, hogy a magyar állatorvos-tudomány és -történet dokumentumait, tudásvagyonát elektronikus formában összegyűjtse, rendszerezze, megőrizze, kereshetővé és hozzáférhetővé tegye, szolgáltatassa, a hatályos jogi szabályozások figyelembe vételével.*

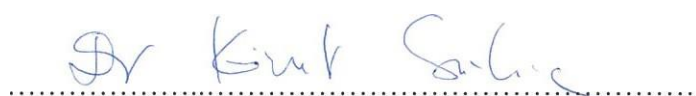
*A HuVetA a korszerű informatikai lehetőségek felhasználásával biztosítja a könnyű, (internetes keresőgépekkel is működő) kereshetőséget és lehetőség szerint a teljes szöveg azonnali elérését. Célja ezek révén*

- *a magyar állatorvos-tudomány hazai és nemzetközi ismertségének növelése;*
- *a magyar állatorvosok publikációira történő hivatkozások számának, és ezen keresztül a hazai állatorvosi folyóiratok impakt faktorának növelése;*
- *az Állatorvostudományi Egyetem és az együttműködő partnerek tudásvagyonának koncentrált megjelenítése révén az intézmények és a hazai állatorvos-tudomány tekintélyének és versenyképességének növelése;*
- *a szakmai kapcsolatok és együttműködés elősegítése,*
- *a nyílt hozzáférés támogatása.*

## KONZULENSI ELLENJEGYZÉS

Alulírott dr. Kövér Szilvia igazolom, hogy Máté Jakab „A szimbiózis megjelenése és dinamikája katalitikus gráfokként modellezett szintróf mikrobiális közösségekben” című diplomamunkát ismerem, azt beadásra és védésre alkalmasnak tartom.

Budapest, 2022.04.28

A handwritten signature in blue ink, reading "Dr. Kövér Szilvia", positioned above a horizontal dotted line.

dr. Kövér Szilvia

Állatorvostudományi Egyetem

Ökológiai tanszék