

SVEUČILIŠTE U ZAGREBU
PMF – Matematički odjel

Alen Lančić

Nino Antulov-Fantulin

**Širenje epidemija na kompleksnim mrežama –
teorijski modeli i računalne simulacije**

Zagreb, 2009.

Ovaj rad izrađen je na zavodu za teorijsku fiziku, Institut Ruđer Bošković i zavodu za elektroničke sustave i obradbu informacija, Fakultet elektrotehnike i računarstva, pod vodstvom dr.sc. Hrvoja Štefančića i dr.sc. Mile Šikića i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2008./2009. Komentor je dr.sc. Zoran Vondraček, Prirodoslovno-matematički fakultet, Matematički odjel.

Sadržaj

1.	Uvod.....	1
2.	Analitički modeli širenja epidemije na grafovima u teoriji vjerojatnosti.....	3
2.1.	Potpuno bipartitni graf i slučajna varijabla infekcije	3
2.2.	Sustav nezavisnih lanaca.....	12
2.3.	Bimodalno ponašanje na sustavu nezavisnih lanaca.....	16
3.	Simulacije širenja epidemije na kompleksnim mrežama.....	22
3.1.	Programski sustav za paralelno programiranje	22
3.2.	Simulacije na kompleksnoj mreži koja odgovara potpunom bipartitnom grafu	23
3.3.	Simulacije na kompleksnim mrežama.....	25
3.3.1.	Simulacije na Erdős-Rényi mrežama:.....	25
3.3.2.	Simulacije na Albert-Barabási mrežama:.....	28
3.3.3.	Simulacije na realnoj mreži ljudskih kontakata kolaboracije znanstvenika:....	29
4.	Zaključak.....	32
5.	Zahvale.....	33
6.	Literatura.....	34
7.	Sažetak	36
8.	Summary	37

1. Uvod

Istraživanje **kompleksnih sustava** i njihovih svojstava zadnjih nekoliko godina zaokuplja pažnju mnogih znanstvenika (fizičara, matematičara, biologa, informatičara, sociologa i dr.).

U kontekstu „teorije grafova” termin „kompleksne mreže” odnosi se na graf s ne trivijalnim topološkim svojstvima koja se ne pojavljuju u jednostavnim grafovima. Struktura im je kompleksnija nego u klasičnim slučajnim mrežama (Erdős–Rényi).

Kompleksne mreže imaju ne trivijalna topološka svojstva kao što su npr. distribucije s debelim repom (engl. *fat-tailed degree distribution*), visoki koeficijent grupiranja, korelacija stupnjeva, postojanje zajednica (engl. *community*) i hijerarhijska struktura.

Statistički opis sustava koji se mogu predstaviti strukturom **kompleksne mreže** daje uvid u bitna svojstva tih sustava, kao što su: njihova otpornost na slučajne greške (kvarove), osjetljivost na namjerne napade (npr. terorističke), efikasnost njihovog pretraživanja ili **podložnost širenju epidemija** (npr. računalnih virusa na Internetu ili SARS-a u ljudskom društvu). U brojnim istraživanjima utvrđeno je da se društveni odnosi mogu prikazati preko kompleksnih mreža. Kompleksna mreža društvenih (socijalnih) kontakata predstavlja strukturu po kojoj se šire epidemije.

Najbitnije i najinteresantnije klase kompleksnih mreža su „**mreže malog svijeta**“ (engl. *small world networks*) gdje prosječan put $\bar{l}(N)$ raste sporije nego ijedna pozitivna potencija od N [1] i „**mreže bez skale**“ (engl. *scale-free networks*) čija je distribucija po zakonu potencija oblika: $P(k) = k^{-\gamma}$ [2]. Albert-Barabashi mreža spada u mreže bez skale koje nastaju preferencijalnim rastom [3].

Postoji **terminološka dualnost** (mreža, čvor i veza) i (graf, vrh i brid). Termine (mreža, čvor i veza) više koriste fizičari dok termine (graf, vrh i brid) pretežno koriste matematičari. Terminološku dualnost smo zadržali zbog konzistentnosti s znanstvenom literaturom. Poštujemo terminologiju po poglavljima pa se u poglavlju 2. analitički modeli širenja epidemije na grafovima u teoriji vjerojatnosti koriste termini (graf, vrh i brid) dok se u poglavlju 3. simulacije širenje epidemija na kompleksnim mrežama koriste termine (mreža, čvor i veza).

Klasični modeli epidemiologije (SIR, SIS, SEIRD...) podrazumijevaju hipotezu homogenog miješanja jedinki populacije. Pretpostavlja se da je kontakt između zaraženih (engl. *infected*) i podložnih zarazi (engl. *susceptible*) konstantan i neovisan o mogućoj heterogenosti u sustavu.

Standardni **SIR model** posjeduje tri odjeljka S podložni, I zaražen, R oporavljeni (engl. *recovered*) u kojem se svaka jedinka populacije može nalaziti. Između odjeljka S i odjeljka I postoji stopa prijelaza (λ) dok između odjeljka I i odjeljka R postoji stopa prijelaza (δ). [4]

Epidemiološki modeli su jako ovisni o modelu koji opisuje populaciju u kojoj se infektivni agenti šire. Dobar model je mreža, gdje svaka jedinka predstavlja čvor, a međusobne veze predstavljaju potencijalni kontakt preko kojeg se zaraza širi. [5]

Interes promatranja epidemioloških modela u "mrežama bez skale" je ogroman, budući da su mreže ljudskih društvenih i seksualnih kontakata *mreže bez skale*. One predstavljaju vrlo zanimljiv slučaj jer imaju distribuciju stupnjeva sljedećeg oblika: $P(k) \approx k^{-\gamma}$.

Satorras, i Vespignani, [6] su 2000. godine pokazali nepostojanje praga epidemije u *mrežama bez skale*.

Razumijevanje samog procesa širenja epidemije nam može pomoći prilikom procjene opasnosti određenog stupnja zaraze, mogu se pronaći bitni čvorovi za širenje epidemije i pandemije i moguće je simulirati **utjecaj protumjera** na proces širenja. [7] [8]

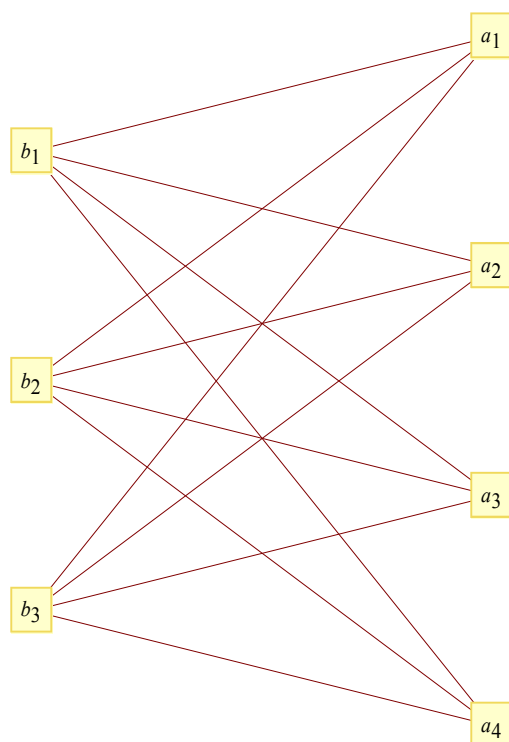
2. Analitički modeli širenja epidemije na grafovima u teoriji vjerojatnosti

2.1. Potpuno bipartitni graf i slučajna varijabla infekcije

Promatramo potpuni bipartitni graf u kojem imamo skup vrhova $\{b_1, \dots, b_s\}$ koji međusobno nisu povezani, ali su svi povezani sa svakim od vrhova iz skupa $\{a_1, \dots, a_n\}$ koji također nisu međusobno povezani. Skup $\{b_1, \dots, b_s\}$ nazovimo skup zaraženih vrhova, a skup $\{a_1, \dots, a_n\}$ nazovimo skupom susceptibilnih vrhova. Promatramo sljedeći slučajni proces:

prvo, za svaki vrh b_j slučajna varijabla $T_j \sim \text{NegBin}(v, q)$ predstavlja broj koraka koji vrh b_j ostaje zaražen, nakon čega postaje imun i više ne prenosi zarazu. Za različite vrhove su duljine oporavka nezavisne slučajne varijable i $v \in \mathbb{N}$, $q \in (0, 1]$.

Drugo, za svaki brid između zaraženog i susceptibilnog vrha nezavisno s vjerojatnošću $p \in (0, 1)$ taj susceptibilni vrh postaje zaražen, a ako niti jedan pokušaj zaraze ne uspije, ostaje susceptibilan. Nakon što odredimo koji vrhovi su ostali susceptibilni, a koji postali zaraženi, korak je gotov i postupak se ponavlja sve dok svi vrhovi nisu ili susceptibilni ili imuni.



Slika 1: Potpuni bipartitni graf, $s = 3, n = 4$

Neka je A_i događaj da se vrh a_i zarazi, za $i \in \{1, \dots, n\}$ i 1_i pripadna indikatorska slučajna varijabla. Imamo

$$P\left(1_i = 0, \sum_{j=1}^s T_j = m\right) = P\left(1_i = 0 \mid \sum_{j=1}^s T_j = m\right) P\left(\sum_{j=1}^s T_j = m\right) = (1-p)^{s+m} \binom{m+vs-1}{m} q^{vs} (1-q)^m$$

jer je $\sum_{j=1}^s T_j$ suma od vs nezavisnih jednako distribuiranih geometrijskih slučajnih varijabli s parametrom q pa ima negativnu binomnu distribuciju s parametrima vs i q . Nadalje,

$$\begin{aligned} P\left(\sum_{j=1}^s T_j = m, \sum_{i=1}^n 1_i = k\right) &= P\left(\sum_{i=1}^n 1_i = k \mid \sum_{j=1}^s T_j = m\right) P\left(\sum_{j=1}^s T_j = m\right) = \\ &= \binom{n}{k} \left(1 - (1-p)^{m+vs}\right)^k \left((1-p)^{m+vs}\right)^{n-k} \binom{m+vs-1}{m} q^{vs} (1-q)^m \end{aligned}$$

jer je $\sum_{i=1}^n 1_i$ suma od n uvjetno nezavisnih jednako distribuiranih Bernoulijevih slučajnih

varijabli s parametrom $1 - (1 - p)^{m+s}$, uz uvjet $\sum_{j=1}^s T_j = m$. Sada imamo

$$\begin{aligned}
P\left(\sum_{i=1}^n 1_i = k\right) &= P\left(\sum_{j=1}^s T_j < \infty, \sum_{i=1}^n 1_i = k\right) = \sum_{m=0}^{\infty} P\left(\sum_{j=1}^s T_j = m, \sum_{i=1}^n 1_i = k\right) = \\
&= q^{\nu s} \binom{n}{k} \sum_{m=0}^{\infty} \left(1 - (1 - p)^{s+m}\right)^k \left((1 - p)^{s+m}\right)^{n-k} \binom{m + \nu s - 1}{m} (1 - q)^m = \\
&= q^{\nu s} \binom{n}{k} \sum_{m=0}^{\infty} \left[\sum_{l=0}^k \binom{k}{l} (-1)^{k-l} \left((1 - p)^{s+m}\right)^{k-l} \right] \left((1 - p)^{s+m}\right)^{n-k} \binom{m + \nu s - 1}{m} (1 - q)^m = \\
&= q^{\nu s} \binom{n}{k} \sum_{l=0}^k \binom{k}{l} (-1)^{k-l} \left((1 - p)^{n-l}\right)^s \sum_{m=0}^{\infty} \binom{m + \nu s - 1}{m} \left((1 - q)(1 - p)^{n-l}\right)^m = \\
&= q^{\nu s} \binom{n}{k} \sum_{l=0}^k \binom{k}{l} (-1)^l \left(\frac{(1 - p)^{n-k+l}}{\left(1 - (1 - q)(1 - p)^{n-k+l}\right)^{\nu}} \right)^s \\
X_n^{(s,\nu)} := \sum_{i=1}^n 1_i &\Rightarrow P\left(X_n^{(s,\nu)} = k\right) = q^{\nu s} \binom{n}{k} \sum_{l=0}^k \binom{k}{l} (-1)^l \left(\frac{(1 - p)^{n-k+l}}{\left(1 - (1 - q)(1 - p)^{n-k+l}\right)^{\nu}} \right)^s
\end{aligned}$$

U daljnjem tekstu ćemo povremeno koristiti oznaku $P\left(X_n^{(s,\nu)} = k\right) = \binom{n}{k} P^*\left(X_n^{(s,\nu)} = k\right)$ zbog

lakšeg pisanja, $P^*\left(X_n^{(s)} = k\right)$ nam govori kolika je vjerojatnost da se od n susjeda zaraženih

vrhova zarazi njih točno k koje smo prethodno odabrali. Također ćemo koristiti i oznaku

$$\beta = P\left(X_1^{(1,\nu)} = 1\right).$$

Definicija 1: Neka su $p \in \langle 0, 1 \rangle$, $q \in \langle 0, 1 \rangle$ i $s, n \in \mathbb{N}$. Za diskretnu slučajnu varijablu X kažemo da ima razdiobu infekcije s parametrima n, s, p, q, ν u oznaci $X \sim I(n, s, p, q, \nu)$, ako joj je distribucija dana sa

$$P(X = k) = q^{sv} \binom{n}{k} \sum_{i=0}^k \binom{k}{i} (-1)^i \left(\frac{(1-p)^{n-k+i}}{(1-(1-q)(1-p)^{n-k+i})^\nu} \right)^s \cdot \mathbf{1}_{\{0, \dots, n\}}(k) \quad (1)$$

Napomena 1: Računamo prvi moment od $X_n^{(s, \nu)} \sim I(n, s, p, q, \nu)$:

$$E[X_n^{(s, \nu)}] = E\left[\sum_{i=1}^n 1_i^{(s, \nu)}\right] = \sum_{i=1}^n E[1_i^{(s, \nu)}] = \sum_{i=1}^n P(X_1^{(s, \nu)} = 1) = nP(X_1^{(s, \nu)} = 1) \quad (2)$$

i drugi moment od $X_n^{(s)}$:

$$\begin{aligned} E\left[\left(X_n^{(s, \nu)}\right)^2\right] &= E\left[\left(\sum_{i=1}^n 1_i^{(s, \nu)}\right)^2\right] = E\left[\sum_{i=1}^n \left(1_i^{(s, \nu)}\right)^2 + 2 \sum_{1 \leq i < j \leq n} 1_i^{(s, \nu)} \cdot 1_j^{(s, \nu)}\right] = \sum_{i=1}^n E\left[\left(1_i^{(s, \nu)}\right)^2\right] + 2 \sum_{1 \leq i < j \leq n} E\left[1_i^{(s, \nu)} \cdot 1_j^{(s, \nu)}\right] = \\ &= \sum_{i=1}^n P(X_1^{(s, \nu)} = 1) + 2 \sum_{1 \leq i < j \leq n} P(1_i^{(s, \nu)} \cdot 1_j^{(s, \nu)} = 1) = nP(X_1^{(s, \nu)} = 1) + 2 \sum_{1 \leq i < j \leq n} P(1_i^{(s, \nu)} = 1, 1_j^{(s, \nu)} = 1) = \\ &= nP(X_1^{(s, \nu)} = 1) + 2 \sum_{1 \leq i < j \leq n} P(X_2^{(s, \nu)} = 2) = nP(X_1^{(s, \nu)} = 1) + n(n-1)P(X_2^{(s, \nu)} = 2) \end{aligned}$$

odakle za varijancu dobivamo:

$$\begin{aligned} \text{Var}\left[X_n^{(s, \nu)}\right] &= E\left[\left(X_n^{(s, \nu)}\right)^2\right] - \left(E\left[X_n^{(s, \nu)}\right]\right)^2 = \\ &= nP(X_1^{(s, \nu)} = 1) + n(n-1)P(X_2^{(s, \nu)} = 2) - n^2P(X_1^{(s, \nu)} = 1)^2 = \\ &= n^2 \left(P(X_2^{(s, \nu)} = 2) - P(X_1^{(s, \nu)} = 1)^2 \right) + n \left(P(X_1^{(s, \nu)} = 1) - P(X_2^{(s, \nu)} = 2) \right) \Rightarrow \\ \sigma\left(X_n^{(s, \nu)}\right) &= n \sqrt{P(X_2^{(s, \nu)} = 2) - P(X_1^{(s, \nu)} = 1)^2 + \frac{1}{n} \left[P(X_1^{(s, \nu)} = 1) - P(X_2^{(s, \nu)} = 2) \right]} \quad (3) \end{aligned}$$

Također, vrijedi $\lim_{s \rightarrow \infty} P(X_n^{(s)} = n) = \lim_{s \rightarrow \infty} \sum_{l=0}^n \binom{n}{l} (-1)^l \left(\frac{q(1-p)^l}{1-(1-q)(1-p)^l} \right)^s =$

$$= 1 + \sum_{l=1}^n \binom{n}{l} (-1)^l \lim_{s \rightarrow \infty} \left(\frac{q(1-p)^l}{1-(1-q)(1-p)^l} \right)^s = 1 \text{ odakle odmah slijedi}$$

$$k < n \Rightarrow \lim_{s \rightarrow \infty} P(X_n^{(s)} = k) = 0.$$

Primjetimo još da $X \sim I(n, 1, p, 1, 1)$ ima istu razdiobu kao $Y \sim B(n, p)$:

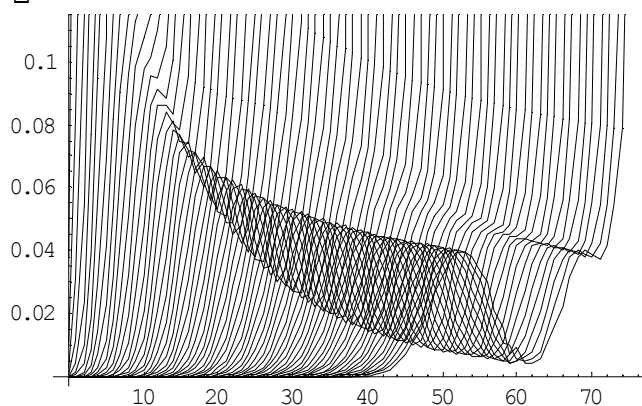
$$\binom{n}{k} (1-p)^{n-k} \left(\sum_{i=0}^k \binom{k}{i} 1^{k-i} (p-1)^i \right) = \binom{n}{k} (1+p-1)^k (1-p)^{n-k} \Rightarrow X = Y.$$

Lema 1: Neka je $X_n \sim I(n, 1, p, q, \nu)$ i $Y_n \sim B(n, p)$. Tada $\lim_{n \rightarrow \infty} \frac{P(X_n = k)}{q \cdot P(Y_n = k)} = 1$.

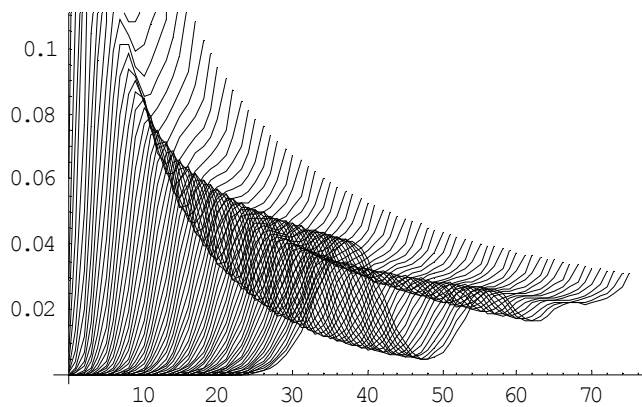
Dokaz:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{P(X_n = k)}{q \cdot P(Y_n = k)} &= \lim_{n \rightarrow \infty} \frac{q \binom{n}{k} \sum_{i=0}^k \binom{k}{i} (-1)^i \frac{(1-p)^{n-k+i}}{(1-(1-q)(1-p)^{n-k+i})^\nu}}{q \binom{n}{k} p^k (1-p)^{n-k}} = \\ &= \lim_{n \rightarrow \infty} \frac{\sum_{i=0}^k \binom{k}{i} (-1)^i \frac{(1-p)^i}{(1-(1-q)(1-p)^{n-k+i})^\nu}}{p^k} = \frac{1}{p^k} \sum_{i=0}^k \binom{k}{i} 1^{k-i} (-1)^i (1-p)^i \lim_{n \rightarrow \infty} \underbrace{\frac{1}{(1-(1-q)(1-p)^{n-k+i})^\nu}}_{=1} = \\ &= \frac{(1-1+p)^k}{p^k} = 1 \end{aligned}$$

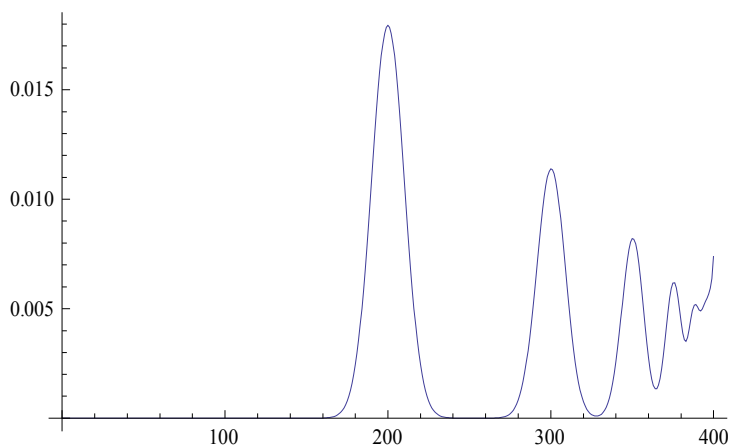
□



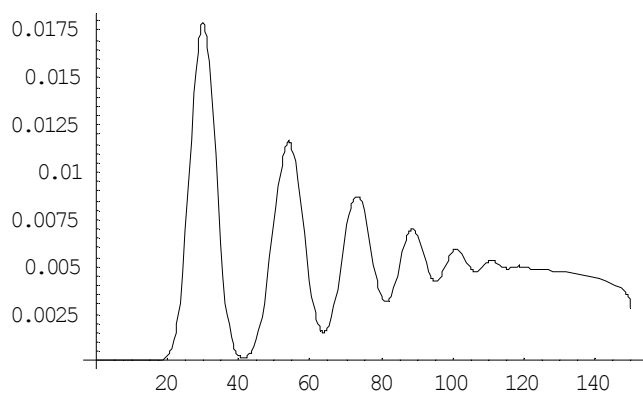
Slika 2: niz grafova funkcija distribucije slučajnih varijabli $X_j \sim I(j, 1, 0.7, 0.4, 1)$ za $j = 1..75$



Slika 3: niz grafova funkcija distribucije slučajnih varijabli $X_j \sim I(j, 1, 0.5, 0.45, 1)$ za $j = 1..75$



Slika 4: graf funkcije distribucije slučajne varijable $X \sim I(400, 1, 0.5, 0.45, 1)$



Slika 5: graf funkcije distribucije slučajne varijable $X \sim I(150, 1, 0.2, 0.22, 1)$

Napomena 2: Ako stavimo $E[X_n^{(s,\nu)}] = \mu$ i $\sigma(X_n^{(s,\nu)}) = \sigma$ iz Čebiševljeve nejednakosti

slijedi da $\forall n, s \in \mathbb{N}$ vrijedi $P(X_n^{(s,\nu)} \in [\mu - 3\sigma, \mu + 3\sigma]) \geq 0.888$ i

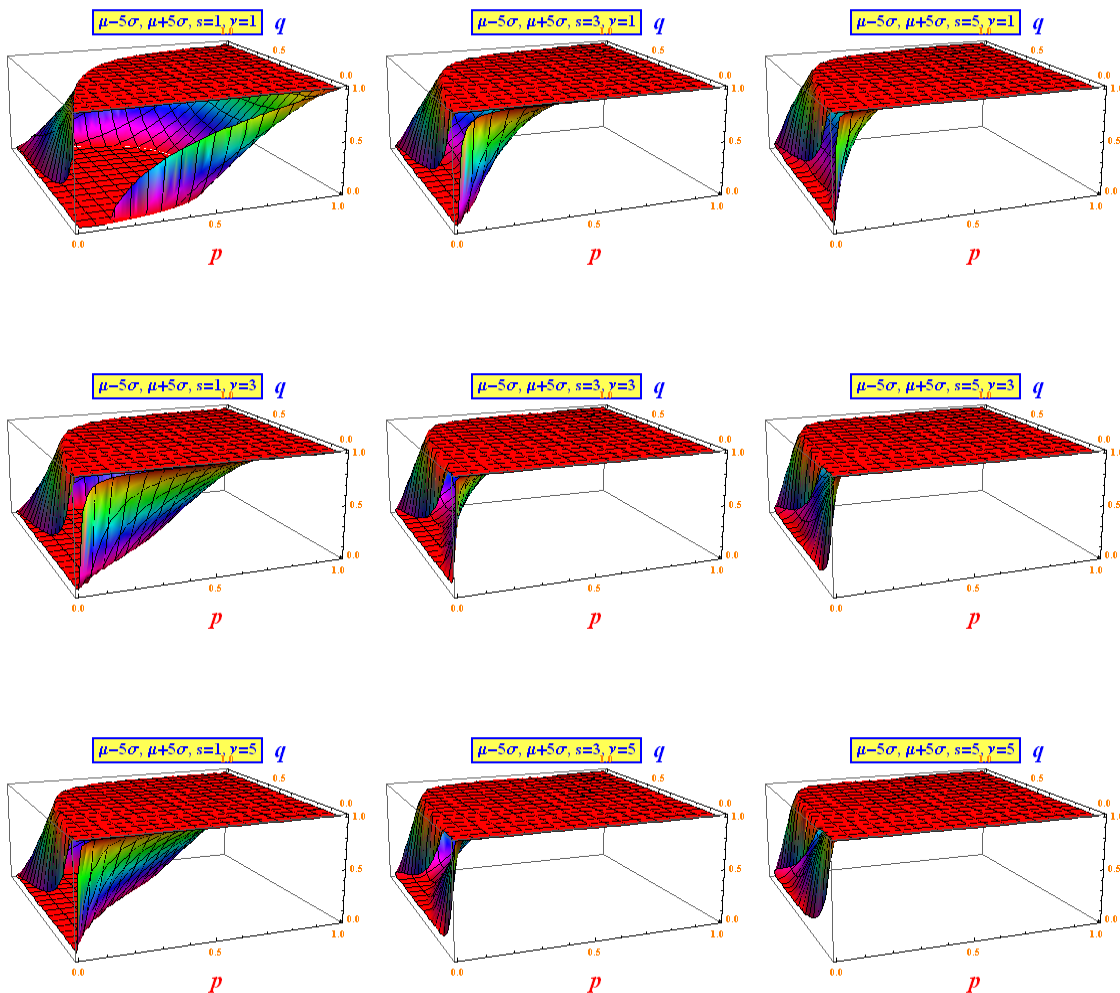
$P(X_n^{(s,\nu)} \in [\mu - 5\sigma, \mu + 5\sigma]) \geq 0.96$. Kako za „dovoljno veliki“ n imamo

$E[X_n^{(s,\nu)}] \pm k\sigma(X_n^{(s,\nu)}) \approx n \left(P(X_1^{(s,\nu)} = 1) \pm k\sqrt{P(X_2^{(s,\nu)} = 2) - P(X_1^{(s,\nu)} = 1)^2} \right)$ vidimo da je

dovoljno pogledati kako izgleda graf od $P(X_1^{(s,\nu)} = 1) \pm k\sqrt{P(X_2^{(s,\nu)} = 2) - P(X_1^{(s,\nu)} = 1)^2}$ jer

znamo da će za „dovoljno veliki“ n devijacija poprimiti gornji izraz i istovremeno će vrijediti

$P(X_n^{(s,\nu)} \in [E[X_n^{(s,\nu)}] - k\sigma(X_n^{(s,\nu)}), E[X_n^{(s,\nu)}] + k\sigma(X_n^{(s,\nu)})]) \geq 1 - k^{-2}$.



Slika 6: graf ploha $E[X_n^{(s,\nu)}] \pm 5\sigma(X_n^{(s,\nu)})$

Napomena 3: Primjetimo još da vrijedi

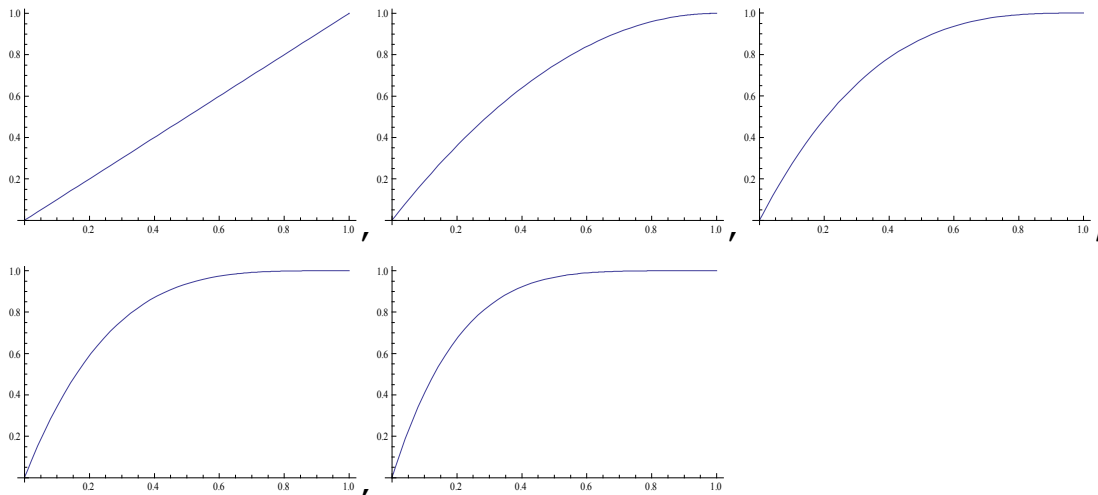
$$P(X_1^{(1,\nu)} = 1) = q^\nu \sum_{i=0}^k (-1)^i \frac{(1-p)^i}{(1-(1-q)(1-p)^i)^\nu} = 1 - \frac{q^\nu (1-p)}{(1-(1-q)(1-p))^\nu}$$

$$P(X_1^{(s,\nu)} = 1) = q^{s\nu} \left(\frac{1}{q^{s\nu}} - \left(\frac{(1-p)}{(1-(1-q)(1-p))^\nu} \right)^s \right) = 1 - \left(\frac{q^\nu (1-p)}{(1-(1-q)(1-p))^\nu} \right)^s =$$

$$= 1 - \left(1 - P(X_1^{(1,\nu)} = 1) \right)^s$$

Ako stavimo $\varphi_s : [0,1] \rightarrow \mathbb{R}, \varphi_s(x) = 1 - (1-x)^s$ imamo $\varphi_s(P(X_1^{(1,\nu)} = 1)) = P(X_1^{(s,\nu)} = 1) \Rightarrow$

$$\varphi_s(E[X_1^{(1,\nu)}]) = E[X_1^{(s,\nu)}] \quad (4)$$



Slika 7: $\varphi_s(x) = 1 - (1-x)^s$ za $s = 1..5$. Vidimo da povećanje broja izvora zaraze povećava očekivani broj zaraženih

b_1
 \nearrow
Sada promatramo graf $a_1 \leftrightarrow b_2$, gdje je a_1 na početku zaraženi vrh, b_i su susceptibilni.
 \nwarrow \vdots
 b_n

Neka je $1_i^{(\nu)}$ indikatorska varijabla događaja da je b_i bio zaražen. U tom slučaju nam je

$$\text{poznato } E\left[\sum_{i=1}^n 1_i^{(\nu)}\right] = E[X_n^{(1,\nu)}] = nP(X_1^{(1,\nu)} = 1).$$

S druge strane promatramo graf

$$\begin{array}{l} a_1 \leftrightarrow b_1 \\ a_2 \leftrightarrow b_2 \\ \vdots \\ a_n \leftrightarrow b_n \end{array}, \text{ gdje su } a_i \text{ zaraženi, } b_i \text{ susceptibilni, } 1_i^{(\nu)}$$

indikatorska varijabla događaja da je b_i bio zaražen, imamo

$$Y_n^{(1,\nu)} = \sum_{i=1}^n 1_i^{(\nu)} \Rightarrow E[Y_n^{(1,\nu)}] = nP(X_1^{(1,\nu)} = 1), \text{ dakle } E[X_n^{(1,\nu)}] = E[Y_n^{(1,\nu)}].$$

Nadalje, $\text{Var}[Y_n^{(1,\nu)}] = nP(X_1^{(1,\nu)} = 1)(1 - P(X_1^{(1,\nu)} = 1))$, dok je

$$\begin{aligned} \text{Var}[X_n^{(s,\nu)}] &= nP(X_1^{(s,\nu)} = 1) + n(n-1)P(X_2^{(s,\nu)} = 2) - n^2P(X_1^{(s,\nu)} = 1)^2 = \\ &= nP(X_1^{(s,\nu)} = 1)(1 - P(X_1^{(s,\nu)} = 1)) + n(n-1)\left(P(X_2^{(s,\nu)} = 2) - P(X_1^{(s,\nu)} = 1)^2\right) = \\ &= \text{Var}[Y_n^{(1,\nu)}] + n(n-1)\underbrace{\left(P(X_2^{(s,\nu)} = 2) - P(X_1^{(s,\nu)} = 1)^2\right)}_{\geq 0} \geq \text{Var}[Y_n^{(1,\nu)}] \end{aligned}$$

Znači vidimo da razdvajanje na nezavisne izvore ne mijenja očekivanje dok smanjuje varijancu, što je dobro jer tada imamo gornju granicu za varijancu.

Pretpostavimo još da imamo graf sa s zaraženih vrhova a_1, \dots, a_s i n susceptibilnih b_1, \dots, b_n .

Neka je b_i spojen na nekih d_i zaraženih vrhova (barem s jednim, inače nije u susjedstvu a_1, \dots, a_s pa graf nije povezan, a taj slučaj nas ne zanima). Neka je $1_i^{(\nu)}$ indikatorska varijabla

događaja da je b_i bio zaražen. Imamo $E\left[\sum_{i=1}^n 1_i^{(\nu)}\right] = \sum_{i=1}^n P(X_1^{(d_i,\nu)} = 1) = \sum_{i=1}^n \varphi_{d_i}\left(P(X_1^{(1,\nu)} = 1)\right)$.

Neka je $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$, $\varphi(d_1, \dots, d_n) = \sum_{i=1}^n \varphi_{d_i}\left(P(X_1^{(1,\nu)} = 1)\right) = n - \sum_{i=1}^n \left(P(X_1^{(1,\nu)} = 0)\right)^{d_i}$ i

$g: \mathbb{R}^n \rightarrow \mathbb{R}$, $g(d_1, \dots, d_n) = \sum_{i=1}^n d_i$. Tražimo kandidate za ekstreme od φ na $g^{-1}(d)$:

$$\nabla g = (1, 1, \dots, 1), \quad \nabla \varphi(d_1, \dots, d_n) = -\ln\left(P(X_1^{(1,\nu)} = 0)\right)\left(\left(P(X_1^{(1,\nu)} = 0)\right)^{d_1}, \dots, \left(P(X_1^{(1,\nu)} = 0)\right)^{d_n}\right),$$

nužno mora vrijediti $\nabla \varphi(d_1, \dots, d_n) = \lambda \nabla g \Rightarrow d_1 = \dots = d_n$, a $g(d_1, \dots, d_n) = d \Rightarrow d_i = \frac{d}{n}$ pa je

$\mathbf{x} = \frac{d}{n}(1, \dots, 1)$ jedini kandidat za ekstrem. Imamo $\varphi(\mathbf{x}) = n \left[1 - \left(P(X_1^{(1, \nu)} = 0) \right)^{\frac{d}{n}} \right]$. Da bismo

vidjeli da je to maksimum, računamo

$$k \in \mathbb{N}_0, x \in [0, 1] \Rightarrow 0 \leq x(1 - (1-x)^k) = (1 - (1-x))(1 - (1-x)^k) \Rightarrow$$

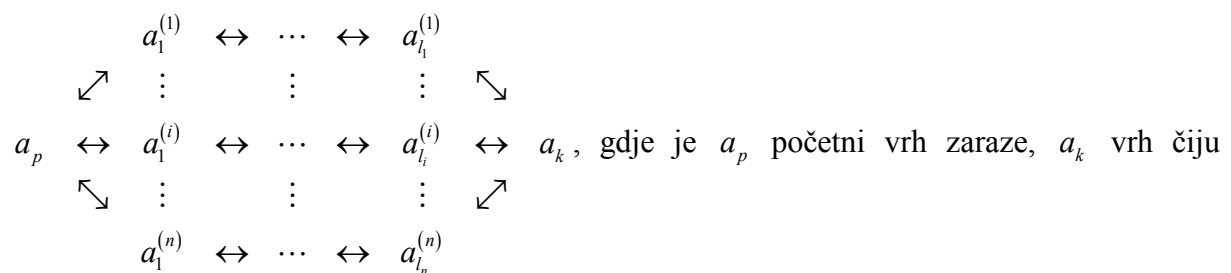
$$\Rightarrow (1-x)(1 - (1-x)^k) \leq 1 - (1-x)^k \Rightarrow (1-x)^k - (1-x)^{k+1} \leq 1 - (1-x) \Rightarrow$$

$$\Rightarrow (1-x)^{s+k} - (1-x)^{s+k+1} \leq (1-x)^s - (1-x)^{s+1} \Rightarrow \varphi_{s+k+1} - \varphi_{s+k}(x) \leq \varphi_{s+1} - \varphi_s(x)$$

Dakle da bismo maksimizirali očekivani broj zaraženih potrebno je ravnomjerno rasporediti veze među izvorima i susceptibilnim vrhovima.

2.2. Sustav nezavisnih lanaca

Napomena 4: Pretpostavimo da imamo graf (sustav nezavisnih lanaca) sa skice



vjerojatnost zaraze tražimo i $l_1, \dots, l_n \in \mathbb{N}$. Neka je $\prod_{i \in \mathcal{I}} a_i := 1$.

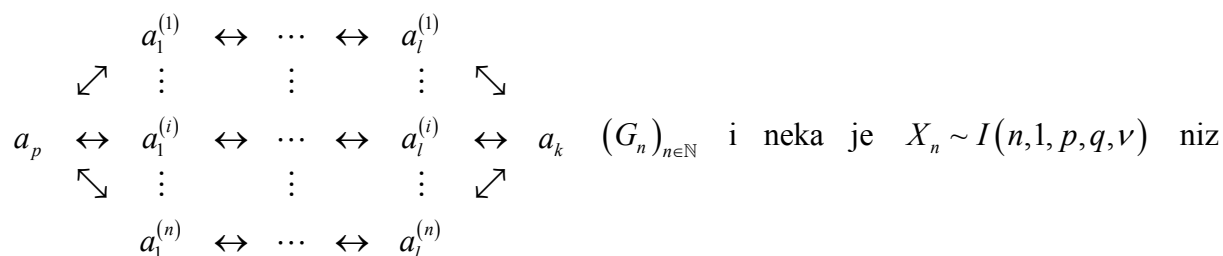
$$\text{Sada imamo } L(l_i) := 1 - \beta^{l_i} \Rightarrow P(A_k) = \sum_{j=1}^n P^*(X_n^{(1, \nu)} = j) \sum_{1 \leq i_1 < \dots < i_j \leq n} \left(1 - \prod_{i \in \{i_1, \dots, i_j\}} L(l_i) \right) =$$

$$\begin{aligned}
&= \sum_{j=1}^n P^*(X_n^{(1,\nu)} = j) \left(\binom{n}{j} - \sum_{1 \leq i_1 < \dots < i_j \leq n} \prod_{i \in \{i_1, \dots, i_j\}} L(l_i) \right) = \sum_{j=1}^n P(X_n^{(1,\nu)} = j) - \sum_{j=1}^n P^*(X_n^{(1,\nu)} = j) \sum_{1 \leq i_1 < \dots < i_j \leq n} \prod_{i \in \{i_1, \dots, i_j\}} L(l_i) = \\
&= 1 - P^*(X_n^{(1,\nu)} = 0) - \sum_{j=1}^n P^*(X_n^{(1,\nu)} = j) \sum_{1 \leq i_1 < \dots < i_j \leq n} \prod_{i \in \{i_1, \dots, i_j\}} L(l_i) = 1 - \sum_{j=0}^n P^*(X_n^{(1,\nu)} = j) \sum_{1 \leq i_1 < \dots < i_j \leq n} \prod_{i \in \{i_1, \dots, i_j\}} L(l_i)
\end{aligned}$$

Poseban slučaj je

$$l_1 = \dots = l_n, L := L(l_i) \Rightarrow P(A_k) = 1 - \sum_{j=0}^n \binom{n}{j} P^*(X_n^{(1,\nu)} = j) L^j = 1 - \sum_{j=0}^n P(X_n^{(1,\nu)} = j) L^j \quad (5)$$

Propozicija 1: Neka je dan niz grafova kao na skici



slučajnih varijabli, $L := 1 - \beta^l$ i a_p početni vrh zaraze. Tada je vjerojatnost zaraze $P(A_k)$

vrha a_k dana izrazom $P(A_k) = 1 - \sum_{j=0}^n P(X_n = j) L^j$ i vrijedi $\lim_{n \rightarrow \infty} P(A_k) = 1$.

Dokaz:

Neka je $f_n : \mathbb{N}_0 \rightarrow \mathbb{R}; f_n(j) = P(X_n = j) L^j$ i μ brojaća mjera na \mathbb{N}_0 .

Sada imamo $\lim_{n \rightarrow \infty} \sum_{j=0}^n P(X_n = j) L^j = \lim_{n \rightarrow \infty} \sum_{j=0}^{\infty} f_n(j) = \lim_{n \rightarrow \infty} \int f_n d\mu$. Definiramo

$g : \mathbb{N}_0 \rightarrow \mathbb{R}; g(j) = L^j$. Primjetimo $\int g d\mu = \sum_{j=0}^{\infty} g(j) = \sum_{j=0}^{\infty} L^j = \frac{1}{1-L} < \infty$ i

$f_n(j) = P(X_n = j) L^j \leq L^j = g(j) \Rightarrow f_n \leq g, \forall n \in \mathbb{N}$. Sada možemo iskoristiti Lebesgueov

teorem o dominiranoj konvergenciji $\lim_{n \rightarrow \infty} \int f_n d\mu \stackrel{LTDK}{=} \int \lim_{n \rightarrow \infty} f_n d\mu$. Računamo

$$0 \leq \lim_{n \rightarrow \infty} f_n(k) = \lim_{n \rightarrow \infty} P(X_n = k) L^k \leq \lim_{n \rightarrow \infty} P(X_n = k) =$$

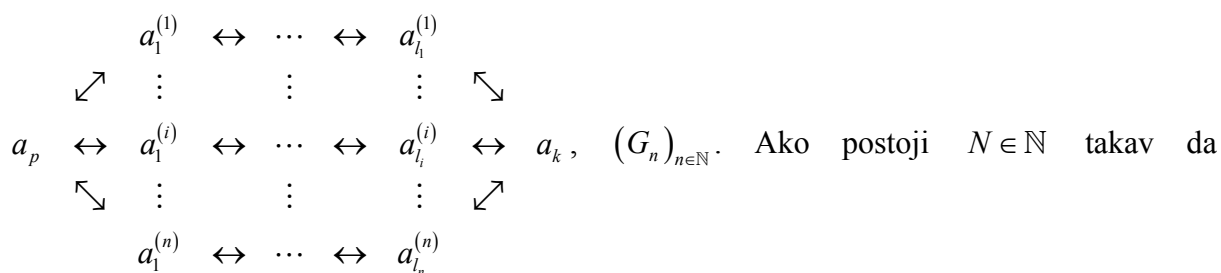
$$\stackrel{\text{Lema 1}}{=} \lim_{n \rightarrow \infty} \underbrace{\frac{P(X_n = k)}{q \binom{n}{k} p^k (1-p)^{n-k}}}_{=1} \cdot q \cdot \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} \leq \lim_{n \rightarrow \infty} \frac{p^k}{k!} (1-p)^{n-k} n^k = 0$$

$$\Rightarrow \lim_{n \rightarrow \infty} f_n = 0 \Rightarrow \lim_{n \rightarrow \infty} \sum_{k=1}^n P(X_n = k) L^k = \int 0 d\mu = 0 \Rightarrow$$

$$\Rightarrow \lim_{n \rightarrow \infty} P(A_k) = 1 - \lim_{n \rightarrow \infty} \sum_{k=1}^n P(X_n = k) L^k = 1$$

□

Korolar 1: Neka je dan niz grafova kao na skici



$\sup_{i \in \mathbb{N}} l_i \leq N$, onda za vjerojatnost zaraze $P(A_k)$ vrha a_k vrijedi $\lim_{n \rightarrow \infty} P(A_k) = 1$.

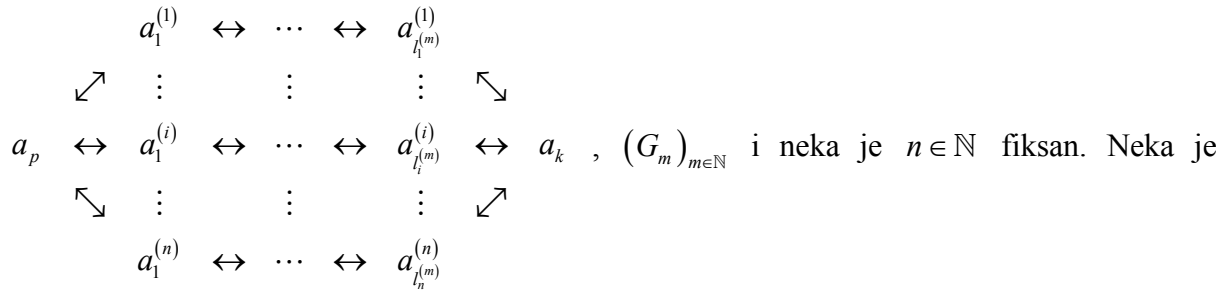
Dokaz: Neka je $L := 1 - \beta^N$ i $X_n \sim I(n, 1, p, q, \nu)$ niz sl. varijabli. Uočimo da

$$1 \geq P(A_k) = 1 - \sum_{j=0}^n P^*(X_n = j) \sum_{1 \leq i_1 < \dots < i_j \leq n} \prod_{i \in \{i_1, \dots, i_j\}} \underbrace{L(l_i)}_{\leq L} \geq 1 - \sum_{j=0}^n P(X_n = j) L^j \Rightarrow$$

$$\Rightarrow 1 \geq \lim_{n \rightarrow \infty} P(A_k) \geq 1 - \lim_{n \rightarrow \infty} \sum_{j=0}^n P(X_n = j) L^j = 1 - 0 = 1 \Rightarrow \lim_{n \rightarrow \infty} P(A_k) = 1$$

□

Propozicija 2: Neka je dan niz grafova kao na skici



$l_m := \min_{i \in \{1..n\}} l_i^{(m)}$. Ako $\lim_{m \rightarrow \infty} l_m = \infty$ onda za vjerojatnost zaraze $P(A_k)$ vrha a_k vrijedi

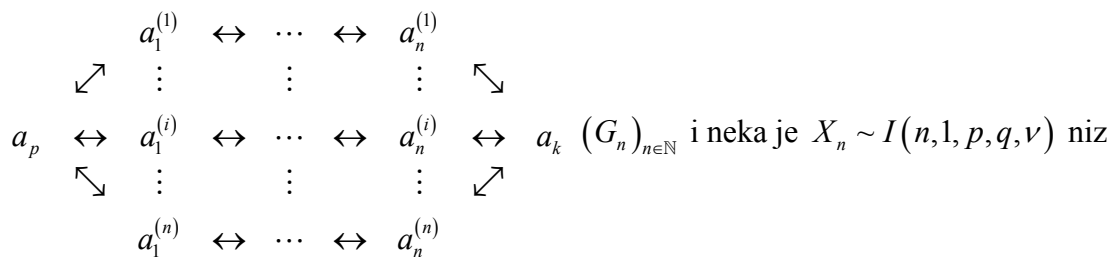
$$\lim_{m \rightarrow \infty} P(A_k) = 0.$$

Dokaz: Neka je $L := 1 - \beta^l \Rightarrow \lim_{m \rightarrow \infty} L = 1$ i $X_n \sim I(n, 1, p, q, \nu)$ slučajna varijabla.

$$\begin{aligned}
 0 \leq P(A_k) &= 1 - \sum_{j=0}^n P^*(X_n = j) \sum_{1 \leq i_1 < \dots < i_j \leq n} \prod_{i \in \{i_1, \dots, i_j\}} \underbrace{L(l_i)}_{\geq L} \leq 1 - \sum_{j=0}^n P(X_n = j) L^j \Rightarrow \\
 \Rightarrow 0 \leq \lim_{m \rightarrow \infty} P(A_k) &\leq 1 - \lim_{m \rightarrow \infty} \sum_{j=0}^n P(X_n = j) L^j = 1 - \sum_{j=0}^n P(X_n = j) \lim_{m \rightarrow \infty} L^j = 1 - \sum_{j=0}^n P(X_n = j) = 0 \Rightarrow \\
 \Rightarrow \lim_{m \rightarrow \infty} P(A_k) &= 0
 \end{aligned}$$

□

Propozicija 3: Neka je dan niz grafova kao na skici



slučajnih varijabli, $L_n := 1 - \beta^n$ i a_p početni vrh zaraze. Tada je vjerojatnost zaraze $P(A_k)$

vrha a_k dana izrazom $P(A_k) = 1 - \sum_{j=0}^n P(X_n = j) L_n^j$ i vrijedi $\lim_{n \rightarrow \infty} P(A_k) = 0$.

Dokaz:

$$\begin{aligned} 1 &\geq \lim_{n \rightarrow \infty} \sum_{k=0}^n P(X_n = k) L_n^k \geq \lim_{n \rightarrow \infty} \sum_{k=0}^n P(X_n = k) L_n^n = \lim_{n \rightarrow \infty} L_n^n \lim_{n \rightarrow \infty} \sum_{k=0}^n P(X_n = k) = \\ &= \lim_{n \rightarrow \infty} (1 - \beta^n)^n \lim_{n \rightarrow \infty} 1 = \lim_{n \rightarrow \infty} e^{\frac{\ln(1 - \beta^n)}{\frac{1}{n}}} \stackrel{L'H}{=} e^{\lim_{n \rightarrow \infty} \frac{n^2 \beta^n \ln \beta}{1 - \beta^n}} = e^0 = 1 \Rightarrow \\ &\Rightarrow \lim_{n \rightarrow \infty} P(A_k) = 1 - \lim_{n \rightarrow \infty} \sum_{k=0}^n P(X_n = k) L_n^k = 1 - 1 = 0 \end{aligned}$$

□

2.3. Bimodalno ponašanje na sustavu nezavisnih lanaca

Propozicija 4 (bimodalnost zaraze na sustavu nezavisnih lanaca): Neka je

$X_n \sim I(n, 1, p, q, 1)$ niz slučajnih varijabli i $L_n := 1 - \beta^{\log_a n}$ za neki $a > 1$. Tada je

$$1 - \lim_{n \rightarrow \infty} \sum_{j=0}^n P(X_n = j) L_n^j = \begin{cases} 0, & \beta \leq a^{-1} \\ 1, & \beta > a^{-1} \end{cases} \quad (6)$$

Dokaz:

$$\begin{aligned} \lim_{n \rightarrow \infty} (1 - \beta^{\log_a n})^n &= e^{\lim_{n \rightarrow \infty} \frac{\ln(1 - \beta^{\log_a n})}{\frac{1}{n}}} \stackrel{L'H}{=} e^{\lim_{n \rightarrow \infty} \frac{n \beta^{\log_a n} \ln \beta}{(1 - \beta^{\log_a n}) \ln a}} = e^{\log_a \beta \lim_{n \rightarrow \infty} n \beta^{\log_a n}} = e^{\log_a \beta \lim_{n \rightarrow \infty} e^{(1 + \log_a \beta) \ln n}} \Rightarrow \\ &\Rightarrow \lim_{n \rightarrow \infty} (1 - \beta^{\log_a n})^n = \begin{cases} 0, & \beta > a^{-1} \\ 1, & \beta \leq a^{-1} \end{cases} \end{aligned}$$

Dakle, za $P(X_1 = 1) \leq a^{-1}$ imamo situaciju kao u prethodnoj propoziciji (dokaz možemo

doslovno kopirati, jedina razlika je definicija L_n) pa je $1 - \lim_{n \rightarrow \infty} \sum_{j=0}^n P(X_n = j) L_n^j = 0$.

U suprotnom, za $\beta > a^{-1}$ radimo sljedeće:

$$\begin{aligned} \sum_{k=0}^n P(X_n = k) L_n^k &= q^\nu \sum_{m=0}^{\infty} (1-q)^m \binom{m+\nu-1}{m} \sum_{k=0}^n \binom{n}{k} \left(L_n \left(1 - (1-p)^{m+1} \right) \right)^k \left((1-p)^{m+1} \right)^{n-k} = \\ &= q^\nu \sum_{m=0}^{\infty} (1-q)^m \binom{m+\nu-1}{m} \left((1-\beta^{\log_a n}) \left(1 - (1-p)^{m+1} \right) + (1-p)^{m+1} \right)^n = \\ &= q^\nu \sum_{m=0}^{\infty} (1-q)^m \binom{m+\nu-1}{m} \left(1 - \beta^{\log_a n} \left(1 - (1-p)^{m+1} \right) \right)^n \end{aligned}$$

Ako sada definiramo $g_n : \mathbb{N}_0 \rightarrow \mathbb{R}, g_n(m) := q^\nu (1-q)^m \binom{m+\nu-1}{m} \left(1 - \beta^{\log_a n} \left(1 - (1-p)^{m+1} \right) \right)^n$,

$g : \mathbb{N}_0 \rightarrow \mathbb{R}, g(m) := q^\nu (1-q)^m \binom{m+\nu-1}{m}$ i μ kao brojaću mjeru na \mathbb{N}_0 , imamo

$$g_n(m) \leq g(m), \forall m \in \mathbb{N} \quad \text{i} \quad \int g d\mu = q^\nu \sum_{m=0}^{\infty} (1-q)^m \binom{m+\nu-1}{m} = 1 < \infty \quad \text{pa je } g \text{ lebesgue-}$$

integrabilna i dominira niz $(g_n)_{n \in \mathbb{N}}$ što znači da opet možemo primjeniti LTDK:

$$\lim_{n \rightarrow \infty} q^\nu \sum_{m=0}^{\infty} (1-q)^m \binom{m+\nu-1}{m} \left(1 - \beta^{\log_a n} \left(1 - (1-p)^{m+1} \right) \right)^n = \lim_{n \rightarrow \infty} \int g_n d\mu \stackrel{LTDK}{=} \int \lim_{n \rightarrow \infty} g_n d\mu. \text{ Još ostaje}$$

pokazati da je $\lim_{n \rightarrow \infty} g_n = 0$ što se svodi na računanje limesa:

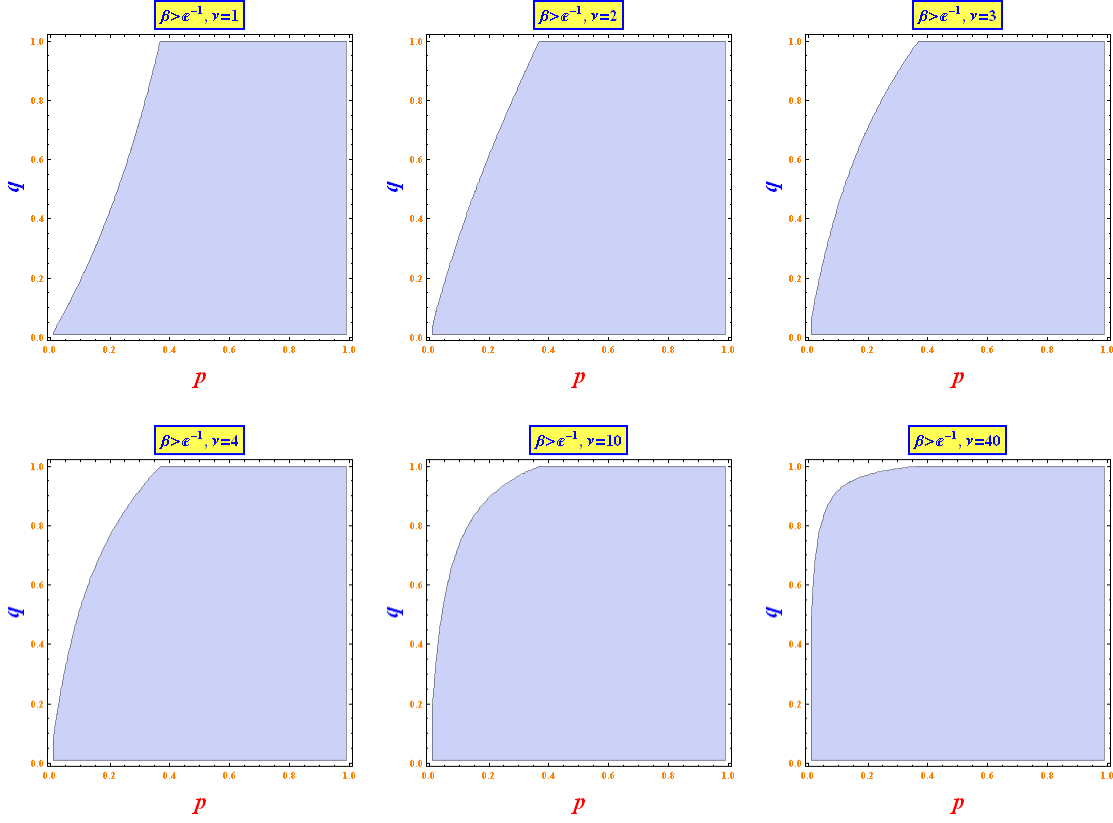
$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \beta^{\log_a n} \left(1 - (1-p)^m \right) \right)^n &= e^{\lim_{n \rightarrow \infty} \frac{\ln \left(1 - \beta^{\log_a n} \left(1 - (1-p)^m \right) \right)}{\frac{1}{n}}} \stackrel{L'H}{=} e^{\lim_{n \rightarrow \infty} \frac{n \left(1 - (1-p)^m \right) \beta^{\log_a n} \log_a \beta}{1 - \beta^{\log_a n} \left(1 - (1-p)^m \right)}} = e^{\left(1 - (1-p)^m \right) \log_a \beta \lim_{n \rightarrow \infty} n \beta^{\log_a n}} = \\ &= e^{\left(1 - (1-p)^m \right) \log_a \beta \cdot \lim_{n \rightarrow \infty} e^{\ln n (1 + \log_a \beta)}} = 0 \text{ odakle slijedi } \lim_{n \rightarrow \infty} g_n = 0 \text{ pa je} \end{aligned}$$

$$\lim_{n \rightarrow \infty} q \sum_{m=0}^{\infty} (1-q)^m \binom{m+\nu-1}{m} \left(1 - \beta^{\log_a n} \left(1 - (1-p)^{m+1} \right) \right)^n \stackrel{LTDK}{=} \int 0 d\mu = 0 \text{ odakle slijedi tvrdnja.}$$

□

Korolar 2: Neka je $X_n \sim I(n, 1, p, q, \nu)$ niz slučajnih varijabli i $L_n := 1 - \beta^{\ln n}$. Tada je

$$1 - \lim_{n \rightarrow \infty} \sum_{j=0}^n P(X_n = j) L_n^j = \begin{cases} 0, & \beta \leq e^{-1} \\ 1, & \beta > e^{-1} \end{cases}.$$



Slika 8: $\beta = P(X_1^{(1,\nu)} = 1) > e^{-1}$. Bimodalnost se očituje u postojanju dva režima; u bijelom području parametarskog prostora zaraza ostaje lokalizirana, a na ostatku se proširi globalno

Napomena 5:

$$\lim_{n \rightarrow \infty} \lim_{s \rightarrow \infty} \left(1 - \sum_{j=0}^n P(X_n^{(s,\nu)} = j) (1 - \beta^{\log_a n})^j \right) = \lim_{n \rightarrow \infty} \left(1 - (1 - \beta^{\log_a n})^n \right) = \begin{cases} 1, & \beta > a^{-1} \\ 0, & \beta \leq a^{-1} \end{cases} \text{ pa vidimo da}$$

rezultat ne ovisi o broju izvora zaraze na lijevom kraju.

Lema 2: Neka je $n, k \in \mathbb{N}_0$ i $X_n^{(s,\nu)} \sim I(n, s, p, q, \nu)$. Neka je

$$P(X_n^{(s,\nu)} = k) = \binom{n}{k} P^*(X_n^{(s,\nu)} = k). \text{ Tada vrijedi}$$

$$P^*(X_k^{(s,\nu)} = k) = \sum_{i=k}^n \binom{n-k}{i-k} P^*(X_n^{(s,\nu)} = i) \quad (7)$$

Dokaz:

$$\begin{aligned}
P^*(X_k^{(s,\nu)} = k) &= q^{s\nu} \sum_{m=0}^{\infty} \left(1 - (1-p)^{s+m}\right)^k \binom{m+s\nu-1}{m} (1-q)^m = \\
&= q^{s\nu} \sum_{m=0}^{\infty} \left(1 - (1-p)^{s+m}\right)^k \binom{m+s\nu-1}{m} (1-q)^m \underbrace{\sum_{i=0}^{n-k} \binom{n-k}{i} \left(1 - (1-p)^{s+m}\right)^i \left((1-p)^{s+m}\right)^{n-k-i}}_{=1} = \\
&= q^{s\nu} \sum_{m=0}^{\infty} \binom{m+s\nu-1}{m} (1-q)^m \sum_{i=k}^n \binom{n-k}{i-k} \left(1 - (1-p)^{s+m}\right)^i \left((1-p)^{s+m}\right)^{n-i} = \\
&= \sum_{i=k}^n \binom{n-k}{i-k} \sum_{m=0}^{\infty} q^{s\nu} \left(1 - (1-p)^{s+m}\right)^i \left((1-p)^{s+m}\right)^{n-i} \binom{m+s\nu-1}{m} (1-q)^m = \sum_{i=k}^n \binom{n-k}{i-k} P^*(X_n^{(s,\nu)} = i)
\end{aligned}$$

□

Propozicija 5: Neka je $X_n^{(s,\nu)} \sim I(n, s, p, q, \nu)$. Tada vrijedi

$$P(X_n^{(s,\nu)} = i) = \binom{n}{i} \sum_{j=i}^n (-1)^{i+j} \binom{n-i}{n-j} P(X_j^{(s,\nu)} = j) \quad (8)$$

Dokaz:

Definiramo $[\mathbf{A}_{(n+1) \times (n+1)}]_{i,j} := \binom{n+1-i}{j-i}$, $[\mathbf{P}^*_{(n+1) \times 1}]_{i,1} := P^*(X_n^{(s,\nu)} = i-1)$ i

$[\mathbf{P}_{(n+1) \times 1}]_{i,1} = P(X_{i-1}^{(s,\nu)} = i-1)$. Sada zbog leme 2 imamo $\mathbf{A}\mathbf{P}^* = \mathbf{P} \Rightarrow \mathbf{P}^* = \mathbf{A}^{-1}\mathbf{P}$ (\mathbf{A} je gornje-

-trokutasta s jedinicama na dijagonali pa je regularna). Lako se vidi da je

$$[\mathbf{A}^{-1}]_{i,j} = (-1)^{i+j} [\mathbf{A}]_{i,j} \Rightarrow P^*(X_n^{(s,\nu)} = i-1) = \sum_{j=1}^{n+1} (-1)^{i+j} \binom{n+1-i}{j-i} P(X_{j-1}^{(s,\nu)} = j-1) \Rightarrow$$

$$P(X_n^{(s,\nu)} = i) = \binom{n}{i} \sum_{j=i}^n (-1)^{i+j} \binom{n-i}{n-j} P(X_j^{(s,\nu)} = j)$$

□

Lema 3: Neka je $X_n \sim I(n, 1, p, q, \nu)$ i $\beta = P(X_1^{(1,\nu)} = 1)$. Definiramo familiju funkcija

$$\mathcal{F} := \left\{ f_{n,p,q,\nu} : \mathbb{R}^+ \rightarrow [0, 1], f_{n,p,q,\nu}(d) = 1 - \sum_{j=0}^n P(X_n^{(1,\nu)} = j) (1 - \beta^d)^j : (n, p, q, \nu) \in \mathbb{N} \times \langle 0, 1 \rangle \times \langle 0, 1 \rangle \times \mathbb{N} \right\}$$

Tada je $f_{n,p,q,\nu}$ strogo padajuća $\forall f_{n,p,q,\nu} \in \mathcal{F}$.

Dokaz:

$$d_1 > d_2 \geq 0 \Rightarrow 1 - \beta^{d_1} > 1 - \beta^{d_2} \Rightarrow \sum_{j=0}^n P(X_n^{(1,\nu)} = j) \left((1 - \beta^{d_2})^j - (1 - \beta^{d_1})^j \right) =$$

$$= f_{n,p,q,\nu}(d_1) - f_{n,p,q,\nu}(d_2) < 0 \Rightarrow f_{n,p,q,\nu}(d_1) < f_{n,p,q,\nu}(d_2)$$

□

Lema 4: Neka je $X_n \sim I(n, 1, p, q, 1)$ i $\beta = P(X_1^{(1,1)} = 1)$. Definiramo familiju funkcija

$$\mathcal{F} := \left\{ f_{n,p,q} : \mathbb{R} \rightarrow \mathbb{R}, f_{n,p,q}(d) = 1 - \sum_{j=0}^n P(X_n^{(1)} = j) (1 - \beta^d)^j : (n, p, q) \in \mathbb{N} \times \langle 0, 1 \rangle \times \langle 0, 1 \rangle \right\}.$$

Tada je (uz $f_0 \equiv 0$):

$$(1) f \in C^\infty(\mathbb{R}), \forall f \in \mathcal{F}$$

$$(2) \frac{d}{dl} f_{n,p,q}(l) = n \ln \beta (f_{n,p,q}(l) - f_{n-1,p,q}(l)), \forall f \in \mathcal{F} \quad (9)$$

$$(3) f_{n,p,q}(l) = \beta^{nl} \left[1 - P(X_n^{(1)} = 0) - \int_0^l f_{n-1,p,q}(z) \beta^{-nz} dz \right], \forall f \in \mathcal{F} \quad (10)$$

Dokaz:

(1) Očito jer je konačna linearna kombinacija funkcija $\in C^\infty(\mathbb{R})$.

(2) Računamo u tri koraka. Prvo imamo

$$f_{n,p,q}(l) = 1 - \sum_{k=0}^n P(X_n = k) L^k \Rightarrow$$

$$\Rightarrow (1 - \beta^l) \frac{d}{dl} f_{n,p,q}(l) = \beta^l \ln \beta \sum_{k=1}^n k P(X_n = k) (1 - \beta^l)^k$$

$$f_{n,p,q}(l) = 1 - q \sum_{m=1}^{\infty} (1 - q)^{m-1} \left(1 - \beta^l (1 - (1 - p)^m) \right)^n \Rightarrow$$

$$\Rightarrow \frac{d}{dl} f_{n,p,q}(l) = qn \beta^l \ln \beta \sum_{m=1}^{\infty} (1 - q)^{m-1} \left(1 - \beta^l (1 - (1 - p)^m) \right)^{n-1} (1 - (1 - p)^m) =$$

$$= n \beta^l \ln \beta \left(q \sum_{m=1}^{\infty} (1 - q)^{m-1} \left(1 - \beta^l (1 - (1 - p)^m) \right)^{n-1} - q \sum_{m=1}^{\infty} (1 - q)^{m-1} \left(1 - \beta^l (1 - (1 - p)^m) \right)^{n-1} (1 - p)^m \right)$$

Zatim nam treba pomoćni rezultat

$$\begin{aligned}
& q(1-p) \sum_{k=0}^{n-1} \binom{n-1}{k} L^k \sum_{m=1}^{\infty} ((1-q)(1-p))^{m-1} \left[\sum_{l=0}^k \binom{k}{l} (-1)^{k-l} ((1-p)^m)^{k-l} \right] ((1-p)^m)^{n-1-k} = \\
& = q \sum_{k=0}^{n-1} \binom{n-1}{k} L^k \sum_{l=0}^k \binom{k}{l} (-1)^{k-l} \sum_{m=1}^{\infty} (1-q)^{m-1} ((1-p)^{n-l})^m = \\
& = \sum_{k=0}^{n-1} \frac{\binom{n-1}{k}}{\binom{n}{k}} L^k q \binom{n}{k} \sum_{l=0}^k \binom{k}{l} (-1)^{k-l} \frac{(1-p)^{n-l}}{1-(1-q)(1-p)^{n-l}} = \sum_{k=0}^n \left(1 - \frac{k}{n}\right) P(X_n = k) L^k
\end{aligned}$$

Na kraju računamo

$$\begin{aligned}
\frac{d}{dl} f_{n,p,q}(l) &= n\beta^l \ln \beta \left(f_{n,p,q}(l) - f_{n-1,p,q}(l) + \frac{1}{n} \sum_{k=1}^n kP(X_n = k)(1-\beta^l)^k \right) = \\
&= n\beta^l \ln \beta (f_{n,p,q}(l) - f_{n-1,p,q}(l)) + \beta^l \ln \beta \sum_{k=1}^n kP(X_n = k)(1-\beta^l)^k = \\
&= n\beta^l \ln \beta (f_{n,p,q}(l) - f_{n-1,p,q}(l)) + (1-\beta^l) \frac{d}{dl} f_{n,p,q}(l) \Rightarrow \\
&\Rightarrow \frac{d}{dl} f_{n,p,q}(l) = n \ln \beta (f_{n,p,q}(l) - f_{n-1,p,q}(l))
\end{aligned}$$

(3) Rješavamo linearnu diferencijalnu jednadžbu prvog reda uz početni uvjet

$f_{n,p,q}(0) = 1 - P(X_n^{(1)} = 0)$ iz (2) pa joj odmah znamo rješenje:

$$\begin{aligned}
f_{n,p,q}(l) &= \left(1 - P(X_n^{(1)} = 0)\right) e^{\int_0^l n \ln \beta dz} - \int_0^l f_{n-1,p,q}(z) e^{\int_z^l n \ln \beta dz} dz = \\
&= \beta^{nl} \left[1 - P(X_n^{(1)} = 0) - \int_0^l f_{n-1,p,q}(z) \beta^{-nz} dz \right]
\end{aligned}$$

□

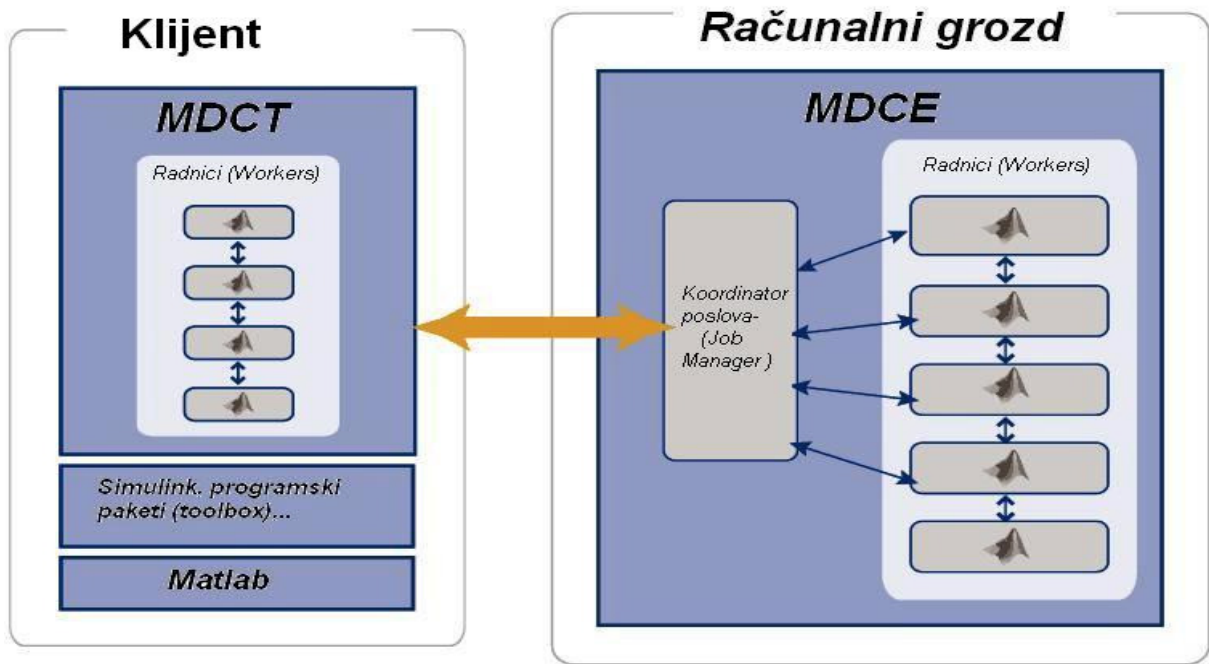
3. Simulacije širenja epidemije na kompleksnim mrežama

3.1. Programski sustav za paralelno programiranje

Zbog velike računalne složenosti simulacija širenja epidemije kroz mreže koristili smo tehnike paralelnog programiranja kako bi ubrzali postupak simuliranja. Korišten je programski sustav Matlab i programski paket "Matlab Distributed Computing Toolbox" [9] (MDCT) koji omogućava distribuirano računanje na računalnom grozdu [sigma.zesoi.fer.hr].

Svaki program koji je vremenski zahtjevan a moguće ga je raspodijeliti na manje podzadatke (engl. *task*) koji će se paralelno izvoditi zove se zadatak (engl. *job*). Sam autor zadatka bira kako će se njegov veliki zadatak podijeliti na manje podzadatke.

Matlab sjednica (engl. *session*) u kojoj su zadatak i njegovi podzadaci definirani zove se klijentska sjednica (engl. *client session*). Matlab Distributed Computing Engine (MDCE) je servis koji omogućava izvođenje zadatka, podzadataka i vraćanje rezultata u klijentsku sjednicu. Koordinator poslova (engl. *job manager*) je dio MDCE sustava koji koordinira izvođenje podzadataka. Koordinator poslova (engl. *job manager*) distribuira podzadatke radnicima (engl. *worker*). [9] Prikaz navedene sheme rada prikazana je na Slici 13.



Slika 9: Prikaz osnovnih modula unutar Matlab Distributed Computing Toolbox-a

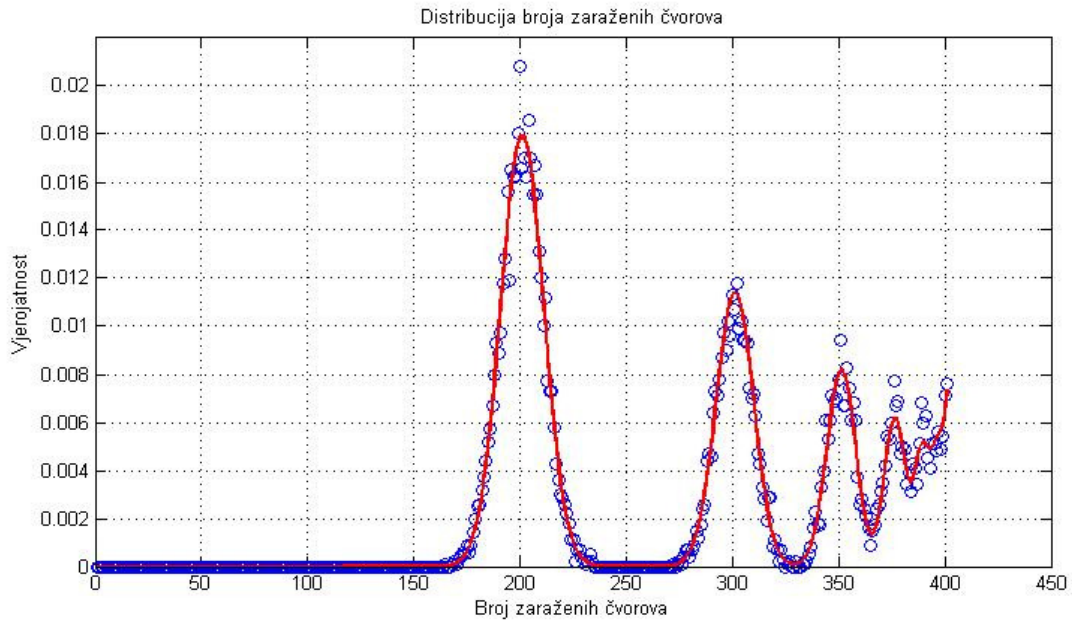
3.2. Simulacije na kompleksnoj mreži koja odgovara potpunom bipartitnom grafu

Generiran je potpuni bipartitni graf sa s čvorova u prvom sloju i n čvorova u drugom sloju. Napravljena su serije od velikog broja simulacija SIR modela širenja epidemija sa parametrima p (vjerojatnost zaraze) i q (vjerojatnost oporavka). Nakon svih simulacija napravljena je distribucija broja zaraženih čvorova u mreži. Dobivena su jako dobra poklapanja sa teorijskim razmatranjima grafa funkcije distribucije slučajne varijable broja zaraženih čvorova $X \sim I(n, s, p, q, v)$ (vidi poglavlje 2.1. potpuno bipartitni graf i slučajna

varijabla infekcije).
$$P(X = k) = q^{sv} \binom{n}{k} \sum_{i=0}^k \binom{k}{i} (-1)^i \left(\frac{(1-p)^{n-k+i}}{(1-(1-q)(1-p)^{n-k+i})^v} \right)^s \cdot \mathbf{1}_{\{0, \dots, n\}}(k)$$

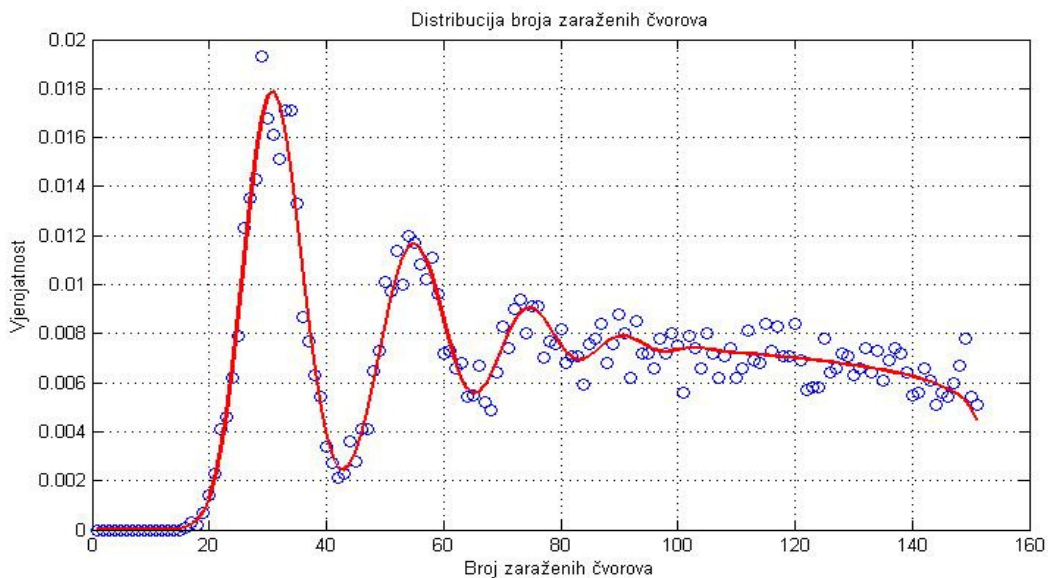
Prikazana su dva primjera:

1.) Generiran je bipartitan graf s parametrima $n = 400$ i $s = 1$. Napravljeno je 10 000 simulacija SIR modela za $p = 0.5$ i $q = 0.45$. Poklapanje distribucije broja zaraženih čvorova sa grafom funkcije distribucije slučajne varijable broja zaraženih čvorova $X \sim I(400, 1, 0.5, 0.45, 1)$ prikazano je na slici 9.



Slika 10: Distribucija broja zaraženih čvorova $p = 0.5$, $q = 0.45$ (plava boja) i funkcija distribucije slučajne varijable broja zaraženih čvorova (crvena boja)

2.) Generiran je bipartitan graf s parametrima $n = 150$ i $s = 1$. Napravljeno je 10 000 simulacija SIR modela za $p = 0.2$ i $q = 0.22$. Poklapanje distribucije broja zaraženih čvorova sa grafom funkcije distribucije slučajne varijable broja zaraženih čvorova $X \sim I(150,1,0.2,0.22,1)$ prikazano je na slici 10. Zbog konačnog broja simulacija prisutne su **fluktuacije** u distribuciji broja zaraženih.



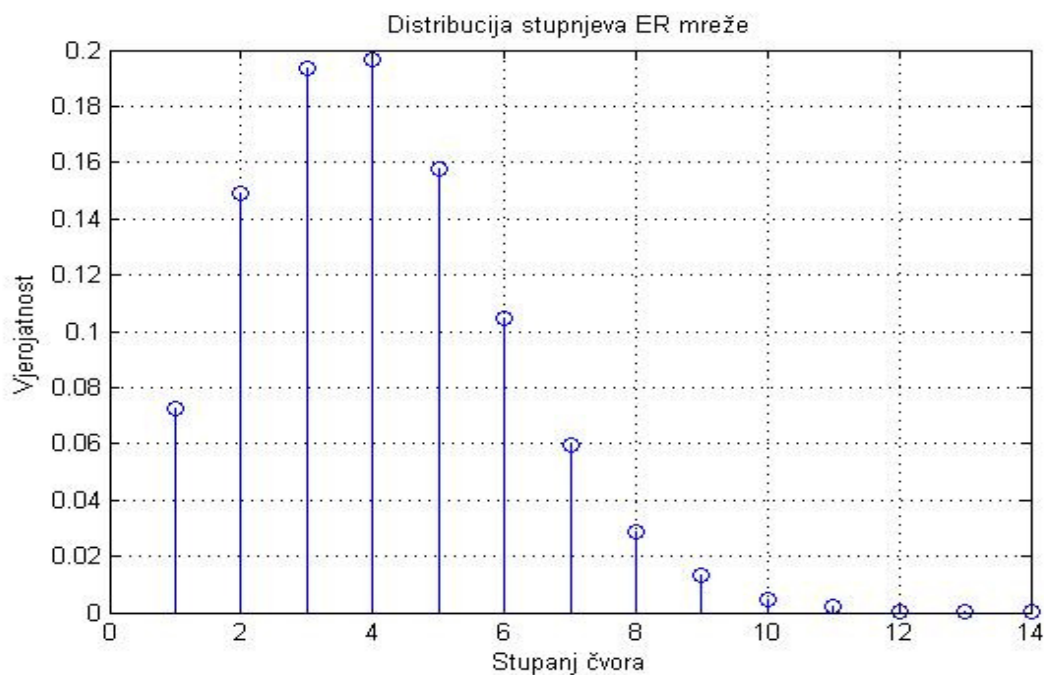
Slika 11: Distribucija broja zaraženih čvorova $p = 0.2$, $q = 0.22$ (plava boja) i funkcija distribucije slučajne varijable broja zaraženih čvorova (crvena boja)

3.3. Simulacije na kompleksnim mrežama

3.3.1. Simulacije na Erdős-Rényi mrežama:

Generira se Erdős-Rényi mreža¹ s N čvorova i s parametrom spajanja slučajno odabrana dva čvora π [10]. Za sve parove čvorova u mreži veza se stvara s vjerojatnošću π . Nad generiranim mrežom za skup vrijednosti p (vjerojatnost zaraze) i q (vjerojatnost oporavka) $\in \langle 0,1 \rangle$ se napravi Z nezavisnih simulacija širenja epidemije iz početnog čvora zaraze s indeksom *poc_id* te se računa očekivani broj zaraženih za svaki promatrani p i q . Distribucija stupnjeva ovakve mreže ravna se prema binomnoj distribuciji. [11][12]

Generirana je Erdős-Rényi mreža sa sljedećim parametrima: $N = 30000$, $\pi = 2.7 \times 10^{-4}$.



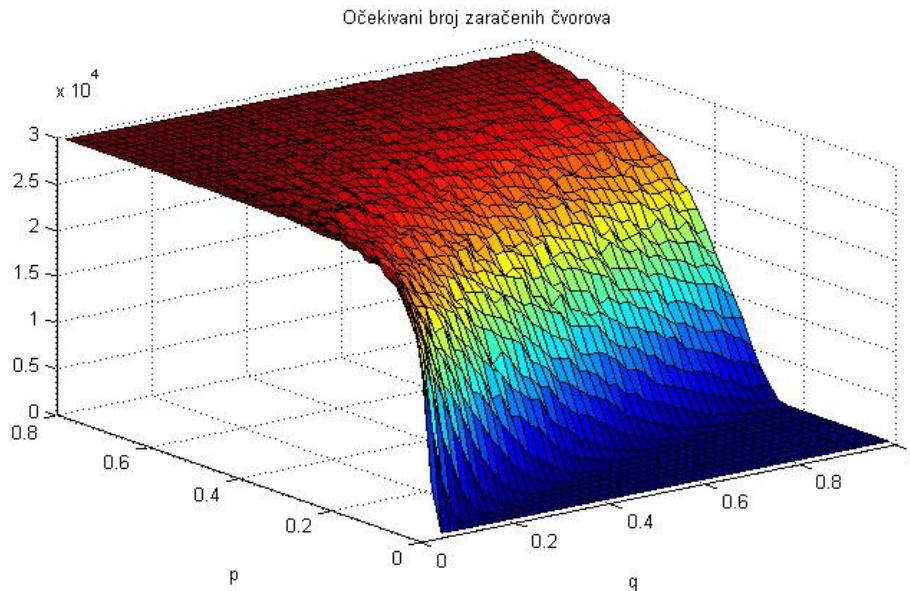
Slika 12: Distribucija stupnjeva promatrane Erdős-Rényi mreže za parametre $N = 30000$,

$$\pi = 2.7 \times 10^{-4}$$

Nad takvom Erdős-Rényi mrežom (N, π) za sve parametre $p \in \{0, 0.02, 0.04, \dots, 0.8\}$ i $q \in \{0, 0.02, 0.04, \dots, 1\}$ s rezolucijom *rez* = 0.02 pokrenuta je simulacija s početnim čvorom

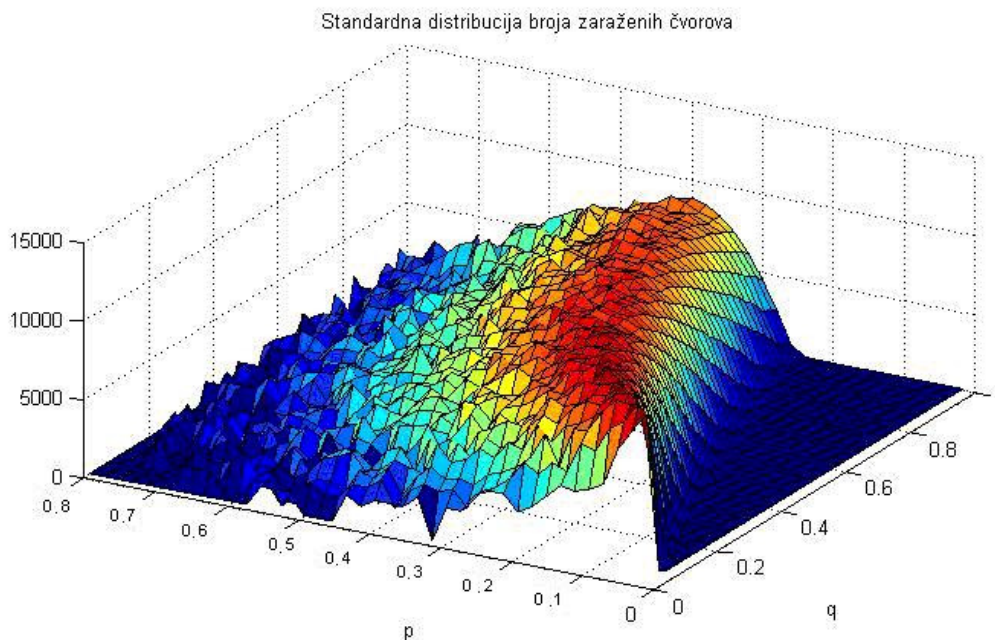
¹ Poznate i kao Erdős-Rényi slučajni grafovi

zaraze $poc_id = 1$. Početni čvor zaraze ima stupanj $d = 4$. Napravljeno je $Z = 500$ nezavisnih simulacija i dobiven je očekivani broj zaraženih čvorova i standardna devijacija broja zaraženih čvorova za svaki promatrani p i q .



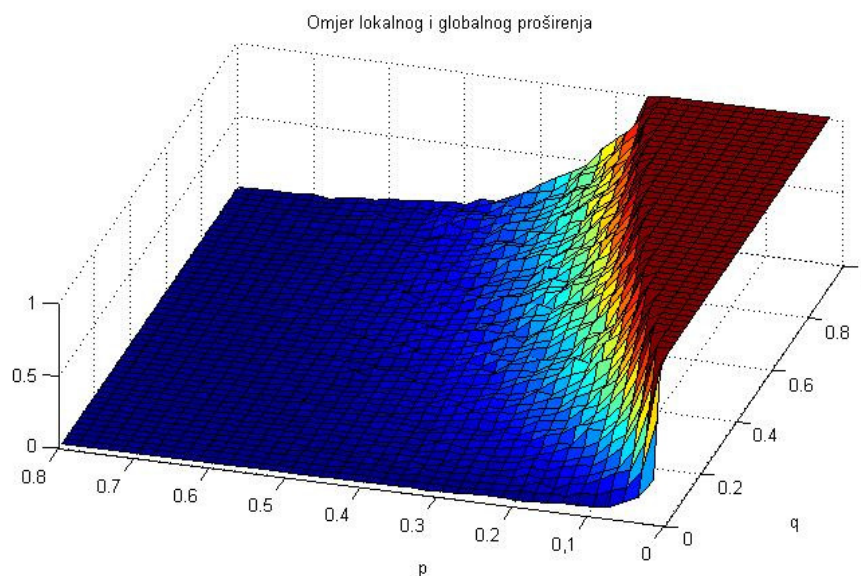
Slika 13: Očekivani broj zaraženih čvorova nad p - q područjem na Erdős-Rényi mreži za parametre

$$N = 30000, \pi = 2.7 \times 10^{-4}, Z = 500 \text{ i } poc_id = 1.$$



Slika 14: Standardna devijacija broj zaraženih čvorova nad p - q područjem na Erdős-Rényi mreži za

$$\text{parametre } N = 30000, \pi = 2.7 \times 10^{-4}, Z = 500 \text{ i } poc_id = 1.$$



Slika 15: Omjer lokalnog i globalnog proširenja epidemije nad p - q područjem na Erdős-Rényi mreži za parametre $N = 30000$, $\pi = 2.7 \times 10^{-4}$, $Z = 500$ i $poc_id = 1$.

Proces širenja zaraze za neke parametre (p, q) u jednoj realizaciji može imati dva moguća ishoda:

- Lokalno proširenje zaraze na malo područje i malo čvorova
- Globalno proširenje zaraze na veliko područje i puno čvorova

Proces širenja zaraze je stohastički proces stoga promatramo ponašanje procesa u velikom broju Z nezavisnih simulacija s istim početnim parametrima.

Ako je tijekom Z nezavisnih simulacija u nekom (p, q) području **dominantan** utjecaj jednog mogućeg ishoda (ili lokalno ili globalno proširenje) onda govorimo da je proces u (p, q) području unimodalalan.

Ako je tijekom Z nezavisnih simulacija u nekom (p, q) području **nije dominantan** utjecaj jednog mogućeg ishoda (ili lokalno ili globalno proširenje) nego je broj realizacija s jednim ishodom sličan broju realizacija suprotnog ishoda proces je u tom (p, q) području je bimodalalan.

Procjena bimodalnosti je vidljiva na grafu standardne devijacije broja zaraženih čvorova u mreži (vidi sliku 13.). Jer ako standardna devijacija za neki (p, q) ne konvergira prema nuli za veliki broj simulacija znači da je to područje (p, q) bimodalno.

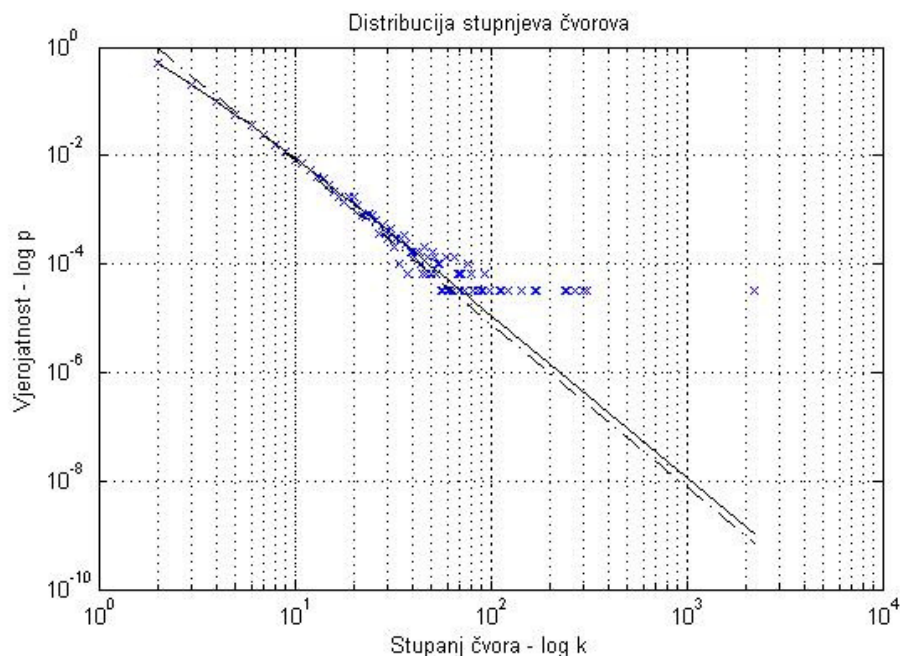
Omjer lokalnog i globalnog proširenja nam omogućava jednostavnu detekciju područja (p, q) u kojemu se zaraza ne proširi ili proširi (vidi sliku 14.).

3.3.2. Simulacije na Albert-Barabási mrežama:

Generira se Albert-Barabási mreža [3] s početnom jezgrom² veličine N_0 . U svakom koraku se dodaje novi čvor koji se s m veza povezuje preferencijalno (veća vjerojatnost spajanja na čvor s većim stupnjem) sa starim čvorovima u mreži. Čvor v_i se odabire sukladno vjerojatnosti $p(v_i) = \frac{k_i}{\sum_j k_j}$, gdje čvor v_i ima stupanj k_i a čvor v_j ima stupanj k_j . Takav

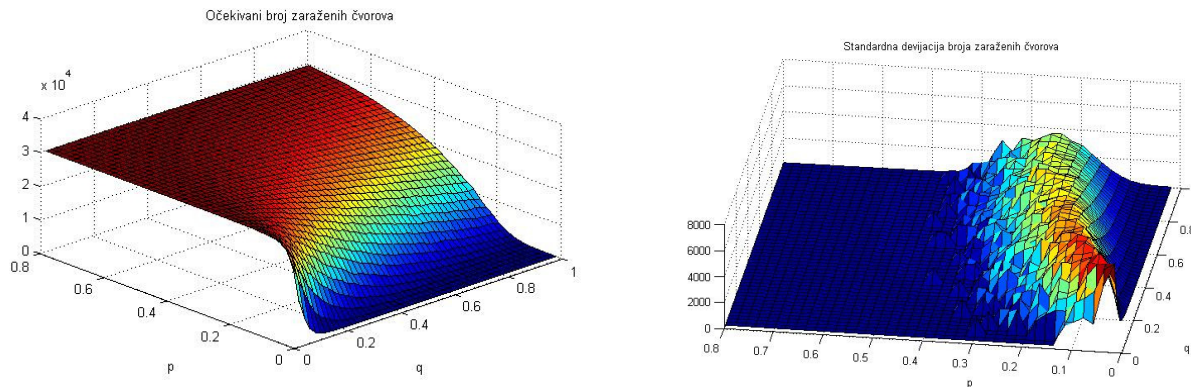
postupak se ponavlja sve dok veličina mreže ne postane $N+N_0$. Distribucija stupnjeva ovako generirane mreže ravna se prema zakonu potencija (engl. *scale-free distribution*) oblika $P(k) = k^{-\gamma}$ i spada u klasu „mreža bez skale“ (engl. *scale-free network*).

Generirana je Albert-Barabási mreža s parametrima: $N = 30000$, $N_0 = 2$ i $m = 2$. Nad tako generiranom mrežom napravljeno (N, N_0, m) za sve parametre $p \in \{0, 0.02, 0.04, \dots, 0.8\}$ i $q \in \{0, 0.02, 0.04, \dots, 1\}$ s rezolucijom $rez = 0.02$ pokrenuta je simulacija s početnim čvorom zaraze $poc_id = 1$. Početni čvor zaraze ima stupanj $d = 2220$. Napravljeno je $Z = 1500$ nezavisnih simulacija i dobiven je očekivani broj zaraženih čvorova i standardna devijacija broja zaraženih čvorova za svaki promatrani p i q .



Slika 16: Distribucija Albert-Barabási mreže, $N_0 = 2$, $N = 30000$ i $m = 2$.

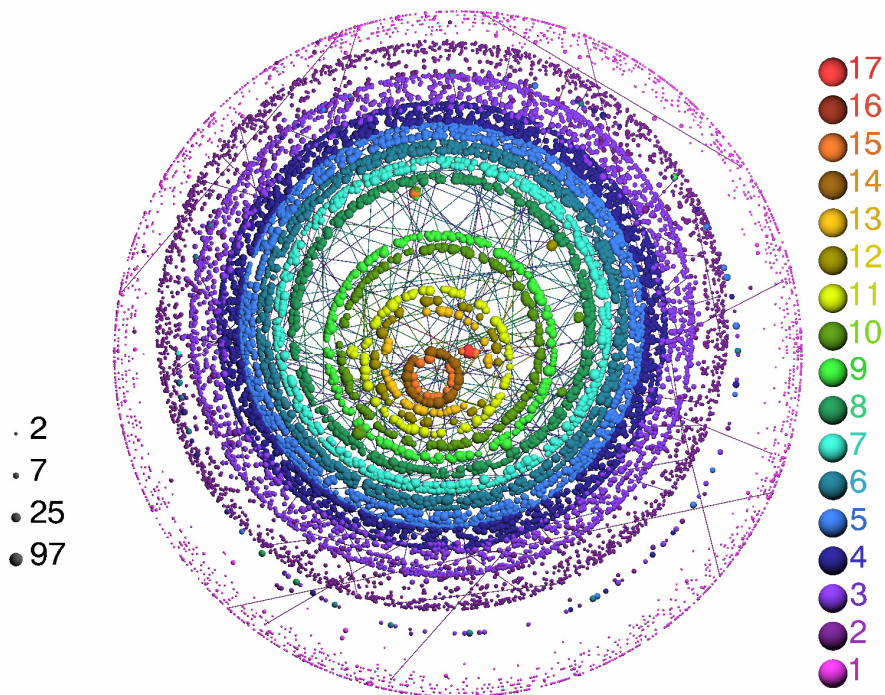
² Jezgra veličine N_0 predstavlja potpuni graf sa N_0 vrhova



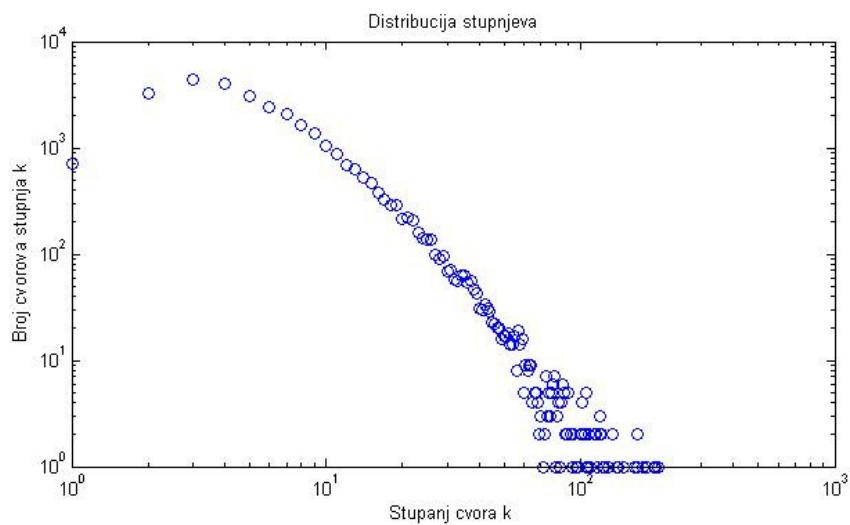
Slika 17: Očekivanje i standardna devijacija broja zaraženih čvorova u Albert-Barabási mreži $N\theta = 2$, $N = 30\,000$, $m = 2$, $poc_id = 1$, $Z = 1500$ simulacija.

3.3.3. Simulacije na realnoj mreži ljudskih kontakata kolaboracije znanstvenika:

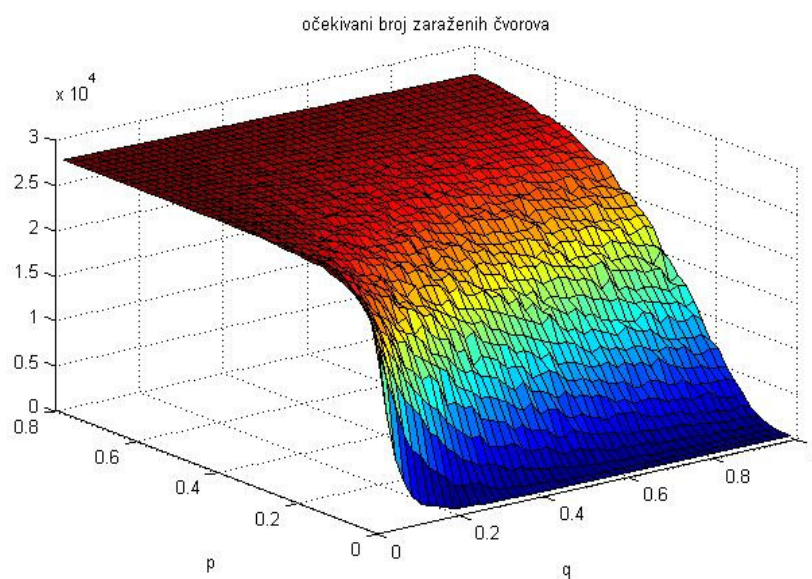
Za potrebe simulacija izabrana je javno dostupna realna mreža kolaboracije znanstvenika "Condensed matter collaborations 2003" [13] [14]. Ta mreža sastoji se od znanstvenika koji su objavljivali radove na "Condensed Matter E-Print Archive" lokaciji između 1. siječnja 1995. godine i 31. prosinca 1999. godine. Mreža je zapisana u GML formatu [15]. Najveća komponenta ove mreže ima 27 519 čvorova.



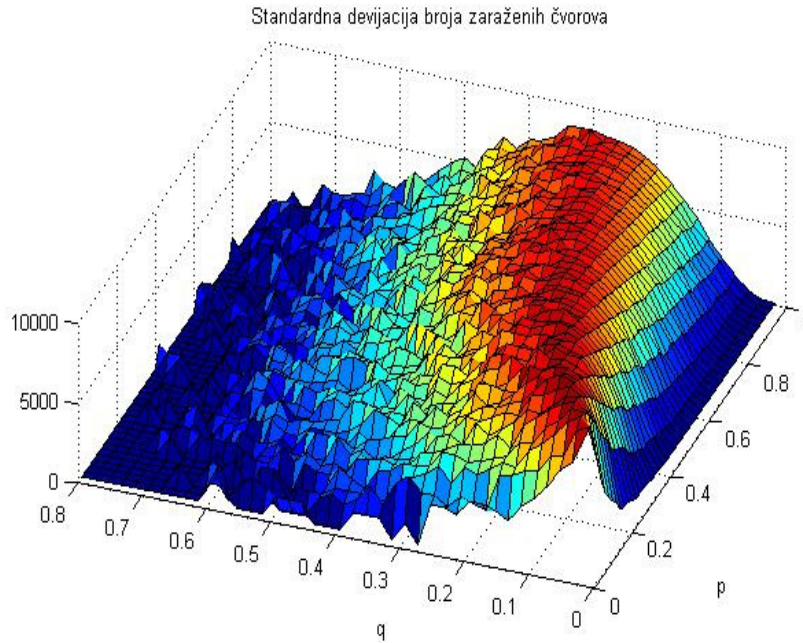
Slika 18: Vizualizacija mreže kolaboracije znanstvenika, svaki čvor ovisno o svom stupnju ima drugačiju boju i veličinu [16]



Slika 19. Distribucija stupnjeva mreže kolaboracije znanstvenika



Slika 20: Očekivanje broja zaraženih čvorova u realnoj mreži kolaboracije znanstvenika



Slika 21 . Standardna devijacija broja zaraženih čvorova u realnoj mreži kolaboracije znanstvenika

Na temelju ovih rezultata simulacija moguće je utvrditi koji dio parametarskog prostora (p,q) uzrokuje globalno a koji lokalno proširenje epidemije. Dok na temelju nagnutosti 3D plohe broja zaraženih čvorova nad (p,q) između globalnog i lokalnog proširenja možemo utvrditi brzinu prijelaza iz lokalnog u globalno. Standardna devijacija broja zaraženih čvorova nam pokazuje u kojem režimu (unimodalnom ili bimodalnom) se proces širenja epidemije nalazi.

4. Zaključak

Standardni modeli epidemiologije kao temeljnu pretpostavku uzimaju hipotezu homogenog miješanja jedinki neke populacije u kojoj se epidemija širi. Takvi modeli ne mogu dobro opisivati heterogene sustave kao što su ljudski kontakti. Dominantan utjecaj heterogenosti mreža ljudskih kontakata najbolje se vidi na distribuciji stupnjeva mreža ljudskih kontakata koji slijede distribuciju po zakonu potencija. Kombiniranjem SIR modela s kompleksom mrežom dobiven je model koji je dobar konsenzus između jednostavnog modela širenja epidemije i kompleksne strukture kao što je kompleksna mreža.

U radu smo pokazali analitički izvod funkcije distribucije, očekivanja i varijance broja zaraženih čvorova u proizvoljnom potpunom bipartitnom grafu. Iduća struktura nad kojom smo proučavali proces širenja epidemije je sustav nezavisnih paralelnih lanaca koji nam je omogućio bolje shvaćanje širenja epidemije i bimodalnosti na jednostavnijoj strukturi nego što je kompleksna mreža. Ako broj lanaca u takvoj strukturi raste eksponencijalno s duljinom lanaca uočavamo dva režima širenja epidemije. Ako je rast sporiji ili brži od eksponencijalnog, očituje se unimodalni režim lokalnog proširenja, odnosno globalnog proširenja.

Poznavanjem početnih parametara SIR modela i izvora zaraze nad nekom mrežom moguće je odrediti konačni ishod epidemije pomoću računalnih simulacija (očekivanje, varijanca broja zaraženih čvorova ili doseg od izvora zaraze). Moguće je odrediti dinamiku širenja epidemije i na temelju tih podataka postupiti preventivno u pokušaju sprječavanja širenja epidemije. Također je moguće simulirati utjecaj intervencije i ocijeniti njezinu uspješnost.

5. Zahvale

Zahvaljujemo se dr.sc. Hrvoju Štefančiću, "Institut Ruđer Bošković" i dr.sc. Mili Šikiću, Fakultet elektrotehnike i računarstva za stručno savjetovanje, uspješnu suradnju i njihovu potporu prilikom rada na ovom istraživanju.

6. Literatura

- [1] D. J. Watts and S. H. Strogatz: Collective dynamics of small-world networks, *Nature* 393, 1998.
- [2] Dorogovtsev,S.N.,Mendes,J.F.F. The shortest path to complex networks. ArXiv:cond-mat/0404593 v4, 24.7.2004.
- [3] A.-L. Barabási, R. Albert and H. Jeong (1999), Mean-field theory for scale-free random networks, *Physica A* 272, 173.
- [4] David Smith and Lang Moore, The SIR Model for Spread of Disease, 22.9.2000, <http://www.math.duke.edu/education/ccp/materials/diffcalc/sir/index.html>, 5.2.2008
- [5] Dorogovtsev,S.N.,Mendes,J.F.F. Evolution of Networks:From Biological Nets to the Inetnet and WWW. Oxford 2003.
- [6] Satorras,R.,Vespignani,A. Epidemics and immunization in scale-free networks. ArXiv: cond-mat/0205260 v1, 14.5.2002.
- [7] Colizza Vittoria , Barrat Alain, Barthelemy Marc, Valleron Alain-Jacques, Vespignani Alessandro. Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *Plos Medicine* | Volume 4 | Issue 1 | e13, sječanj 2007
- [8] Guimera, R., Mossa, S., Turtshi1, A., Amaral1, L. A. N. The world-wide air transportation network: Anomalous centrality, community structure, and cities' global roles. arXiv:cond-mat/0312535 v2,12.7.2005

- [9] Matlab Distributed Computing Toolbox, http://www.mathworks.com/products/distribtb/index.html?s_cid=HP_FP_ML_ParallelComputingToolbox, 13.3.2008.
- [10] Erdős, P.; Rényi, A. (1959). "On Random Graphs. I.". *Publicationes Mathematicae* **6**: 290–297.
- [11] Erdős, P.; Rényi, A. (1960). "The Evolution of Random Graphs". *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**: 17–61.
- [12] Gilbert, E.N. (1959). "Random Graphs". *Annals of Mathematical Statistics* **30**: 1141–1144. doi:10.1214/aoms/11777060
- [13] M. E. J. Newman, Proc. Natl. Acad. Sci. USA 98, 404-409 (2001).
- [14] Network Data, 13.1.2008., Newman, <http://www-personal.umich.edu/~mejn/netdata/>, 15.2.2008.
- [15] GML: A portable Graph File Format, Michael Himsolt, <http://www.infosun.fim.uni-passau.de/Graphlet/GML/gml-tr.html>, 17.4.2008.
- [16] Large Network Visualization tool, <http://xavier.informatics.indiana.edu/lanet-vi/>, 1.6.2008.

7. Sažetak

Alen Lančić, Nino Antulov-Fantulin: Širenje epidemija na kompleksnim mrežama – teorijski modeli i računalne simulacije

U teorijskom dijelu postavili smo nekoliko jednostavnih stohastičkih analitičkih modela širenja epidemije na grafovima. Kao model širenja epidemije u teorijskim razmatranjima koristili smo SIR model koji smo poopćili tro-parametarskim stohastičkim modelom u kojem vrijeme oporavka ima negativnu binomnu distribuciju.

Nad proizvoljnim potpunim bipartitnim grafom izveli smo analitički funkciju distribucije broja zaraženih čvorova te verificirali ispravnost dobivenih rezultata pomoću računalnih stohastičkih simulacija. Ona nam služi kao osnovni okvir za razmatranje složenijih sustava.

Idući precizniji model strukture mreže nad kojom se epidemija širi je sustav nezavisnih paralelnih lanaca. Ovim modelom konstruirali smo primjer strukture koja omogućuje pojavu bimodalno ponašanje procesa širenja epidemije te smo došli korak bliže opisivanju realnih procesa širenja epidemije.

Stohastičkim računalnim simulacijama nad kompleksnim mrežama primijećena su dva moguća režima procesa širenja epidemije koji se očituju u unimodalnom ili bimodalnom režimu. Mjeru bimodalnosti moguće je iskazati standardnom devijacijom ukupnog broja zaraženih čvorova u mreži. Mjera bimodalnosti nam je od velike praktične važnosti za predviđanje konačnog ishoda epidemije. Poznavanjem početnih parametara SIR modela i izvora zaraze nad nekom mrežom moguće je odrediti konačni ishod broja zaraženih čvorova epidemije ili konačni ishod dosega epidemije od izvora zaraze.

Ključne riječi: širenje epidemije, SIR model, kompleksne mreže

8. Summary

Alen Lančić, Nino Antulov-Fantulin: Epidemic spreading in complex networks – theoretical models and computer simulations

In theoretical part of this paper, several analytic stochastic models of epidemic spreading through graphs were constructed. As a model of epidemic spreading in theoretical considerations, basic SIR model was chosen, which was generalized using three-parameter stochastic model with negative binomial time of recovery.

For arbitrary complete bipartite graph, analytic distribution function for number of infected nodes was derived and verified by computer stochastic simulations. It was used as a main framework for analysing of a more complex systems.

System of independent, parallel chains was used as a next, more accurate model of epidemic spreading. Using this model enabled us to construct a network structure which allows emergence of a bimodal behaviour of epidemic spreading that brought us closer to understanding the real process itself.

Using stochastic computer simulations of epidemic spreading in complex networks, unimodal and bimodal behaviour were noticed. As a measure of bimodality, standard deviation of total number of infected nodes in the network can be used. Bimodality measure is of significant practical importance for prediction of an epidemic outcome. Knowledge of starting parameters in the model and epidemic sources in arbitrary network makes it possible to estimate expected number of infected nodes or a distance from the epidemic source.

Keywords: epidemic spreading, SIR model, complex networks