



"Ss. Cyril and Methodius" University in Skopje
**FACULTY OF COMPUTER
SCIENCE AND ENGINEERING**

Моделирање на (COVID -19) вирус со SEIRCD модел

Предмет: Моделирање и симулација на комјутерски мрежи

Ментор:

Д-р Мирослав Мирчев

Изработено:

Јана Николовска 171040

Содржина

Вовед	3
SEICRD модел.....	3
Параметри	5
Имплементација.....	7
Моделот со различни параметри на бета дистрибуцијата	7
Моделот со различни степени на social distancing	8
Моделот со применети мерки после одредено време	8
Резултати.....	9
Заклучок.....	10
Референци	14

Вовед

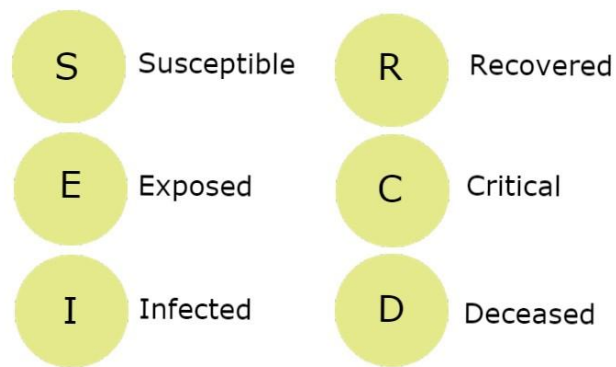
Со појавата на нови епидемии, како на пример моменталната пандемија на вирусот COVID-19, се зголемува потребата од информации како истите да се спречат, на кој начин да се следи нивниот развој и да се контролира, но и кој е најдобриот третман кој може да биде применет. Математичките и статистичките модели се најчесто користени да ги одговорат овие прашања преку проучување на заразеноста на овие преносливи болести. Моделите користат општи претпоставки или собрани статистики заедно со математички формули да пронајдат параметри кои одговараат на испитуваните заразни болести и ги користат тие параметри да направат предвидувања за ширењето на заразата, како и да ги пресметаат ефектите на различен вид на интервенции како на пример масовна вакцинација. Овој пристап помага да се одлучи кои интервенции можат да помогнат и колку можат да помогнат, наспроти оние кои нема да дадат задоволителни резултати.

Во продолжение ќе разгледаме еден ваков модел уште наречен SEIRCD. SEIRCD претставува општ модел кој може да се користи за предвидување на текот на било која заразна болест. Моделот според претходно дефинирани параметри може да определи состојбата на популацијата или во нашиот случај состојба на секој член на дадена популација во временска единица. Параметрите кои ќе ги користиме претставуваат статистички вредности преземени за актуелната пандемија на вирусот COVID-19. Користејќи го овој модел можеме да ги набљудуваме разликите во развојот на вирусот за промена на параметрите и да донесеме заклучок кои вредности на параметрите ќе придонесат во најбавен развој на болеста како можни интервенции во справување на пандемијата.

SEICRD модел

За поедноставување на математичкото моделирање на заразни болести се користат сегментирачки модели (compartmental models). Овие модели името го добиваат поради тоа што популацијата ја поделуваат во повеќе лабелирани сегменти. Класичен сегментирачки модел кој ќе го земеме за пример е SIR моделот кој претставува едноставен модел за моделирање на епидемии. SIR моделот ја дели популацијата на 3 сегменти S за susceptible (потенцијално заболени), I за infectious (заболени) и R за recovered (оздравени). Во овој модел како и моделите слични на него подредувањето на лабелите следи одреден тек на сегментите, па така SIR означува тек од потенцијално заболени до заболени и потоа оздравени.

Со цел да се приближиме поблиску до моделирање на заразни болести со кои се сретнуваме во природата вклучувајќи го и COVID-19 ќе користиме проширен модел на обичниот SIR со 6 сегменти уште наречен SEICRD. Сегментите во овој модел одговараат на Susceptible (потенцијално заразен), Exposed (бил во допир со вирусот), Infected (заразен), Critical (во критична состојба), Recovered (заздравен, со имунитет) и Deceased (починат).



слика 1 Сегменти во моделот SEICRD

Во моделот кој го дефинираме популацијата е претставена со помош на мрежа односно ненасочен граф така што ребро претставува контакт помеѓу два јазли (поединци). Предноста на овој модел во однос на обичниот епидемиолошки модел е што веројатностите за секоја состојба не се исти за секој поединец. Користејќи мрежа се обидуваме подобро да ја измоделираме природата и секојдневните интеракции и динамики во неа, така што секој поединец има различна веројатност да се наоѓа во одредена состојба во зависност од неговите контакти. За да добиеме сумарни статистики за одредена популација потребно е само да ги собереме веројатностите на секој од лицата за даден сегмент.

Секој модел зависи во целост од неговите претпоставки. Дури и со грешни претпоставки моделот ќе даде резултати, но тие ќе се разликуваат од оние во природата. Уште можеме да кажеме дека моделот е онолку добар како и неговите претпоставки. Дел од претпоставките кои ги задаваме за нашиот проблем и кои ги немаме спомнато до сега се:

1. Вирусот може да се пренесе (со одредена веројатност) само доколку си имал контакт со заболено лице
2. Постои период на инкубација по допир со заразен по кој се донесува заклучок дека поединецот е заразен/инфициран
3. Откако поединецот го прележал вирусот се создава имунитет што спречува негово повторно префаќање на истиот вирус
4. Потенцијални текови низ сегментите се SEIR, SEICR, SEICD

Моделот SEICRD, како и останатите сегментирачки модели, е формулиран како Маркова верига. Маркова верига е стохастички модел кој опишува секвенца на можни настани во која веројатноста да се случи еден настан зависи само од состојбата резултат во претходниот настан. Подетално, нашиот модел е формулиран на дискретна Маркова верига поради тоа што промените на настаните се вршат во дискретни временски чекори. Во

нашиот случај временскиот чекор ќе изедначува еден ден, а состојбата (распределбата по сегменти) на секој од поединците ќе зависи од нивната состојба од претходниот ден. Промените од едната состојба во друга се нарекуваат транзиции. Веројатностите асоцирани со промени на состојбата во една временска единица се наречени веројатности на транзиција. Како што веќе споменавме станува збор за стохастички процес бидејќи овие веројатности не се сите предефинирани, туку дел од нив зависат од определена распределба за која ќе споменеме повеќе во делот Параметри.

Параметри

Како што е погоре наведено моделот SEIRCD е формулиран како Маркова верига, па така на крај на секој временски чекор t состојбата на секој поединец i од популацијата ќе зависи само од неговата состојба на почетокот на истиот чекор $t-1$, не и од состојбите во претходните чекори. Уште треба да запомниме дека состојбата претставува распределба на поединецот со соодветни веројатности тој да припаѓа во секој од шестте сегменти – состојба на поединецот i во временска единица t $[S_i(t), E_i(t), I_i(t), C_i(t), R_i(t), D_i(t)]$. Потоа од ново добиените индивидуални состојби можеме да ја пресметаме состојбата на целата популација $[S(t), E(t), I(t), C(t), R(t), D(t)]$.

$$\begin{aligned} S(t) &= \sum_i S_i(t), & E(t) &= \sum_i E_i(t), & I(t) &= \sum_i I_i(t) \\ C(t) &= \sum_i C_i(t), & R(t) &= \sum_i R_i(t), & D(t) &= \sum_i D_i(t) \end{aligned}$$

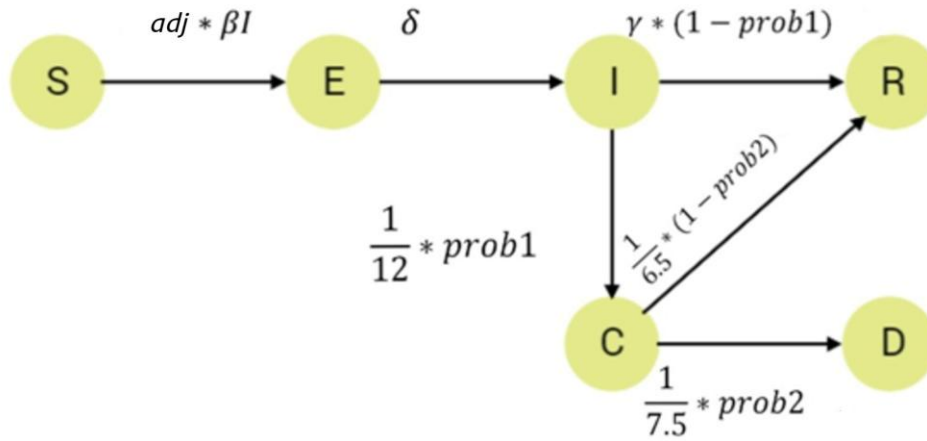
равенка 2 Пресметување на состојба на целата популација замајќи ја во предвид состојбата на секој од поединците i во популацијата

Менувањето на состојбата на еден поединец, односно менување на овие веројатности се вршат со помош на транзиции. Транзициите претставуваат веројатности на премин. На слика 2 можеме да ги забележиме исцртани со помош на стрелки можните текови низ сегментите и соодветно над нив претставени се веројатностите на премин. Во продолжение на слика 2 можеме да видиме ги видиме и детерминистичките равенки со кои се пресметува новата состојба на секој поединец и од кои можеме да видиме како зависи таа од претходната состојба.

Во транзициите се наведени повеќе параметри вклучувајќи ги и:

- adj - вектор со големина на популацијата составен од 1 доколку постои ребро помеѓу двајцата поединци и 0 инаку (различно за секој поединец)
- β - веројатност на превземање на вирусот доколку си во контакт со заолено лице (различно за секој поединец во секоја итерација, се влече од бета распределба)

- δ - веројатност поединецот кој бил во допир со вирусот да покаже симптоми и да може да ја шири болеста. (фиксирано, $=1/5$)
- γ - веројатност поединецот кој бил заболел да оздрави (фиксирано, $=1/10$)
- $prob1$ - веројатност поединецот кој бил заболел да стане критичен (фиксирано, $=1/20$)
- $prob2$ - веројатност поединецот кој бил заболел и во критична состојба да почине (фиксирано, $=1/5$)



слика 2 Состојби и веројатности за транзиција помеѓу состојбите

$$S_i(t) = (1 - (adj * \beta I) * S_i(t - 1)$$

$$E_i(t) = (adj * \beta I) * S_i(t - 1) + (1 - \delta) * E_i(t - 1)$$

$$I_i(t) = \delta * S_i(t - 1) + (1 - (\gamma * (1 - prob1)) - (\frac{1}{12} * prob1)) * I_i(t - 1)$$

$$R_i(t) = (\gamma * (1 - prob1)) * I_i(t - 1) + (\frac{1}{6.5} * (1 - prob2)) * C_i(t - 1) + R_i(t - 1)$$

$$C_i(t) = (\frac{1}{12} * prob1) * I_i(t - 1) + (1 - (\frac{1}{6.5} * (1 - prob2)) - (\frac{1}{7.5} * prob2)) * C_i(t - 1)$$

$$D_i(t) = (\frac{1}{7.5} * prob2) * C_i(t - 1) + D_i(t - 1)$$

равенка 2 Детерминистички равенки со кои се пресметува новата состојба на секој поединец

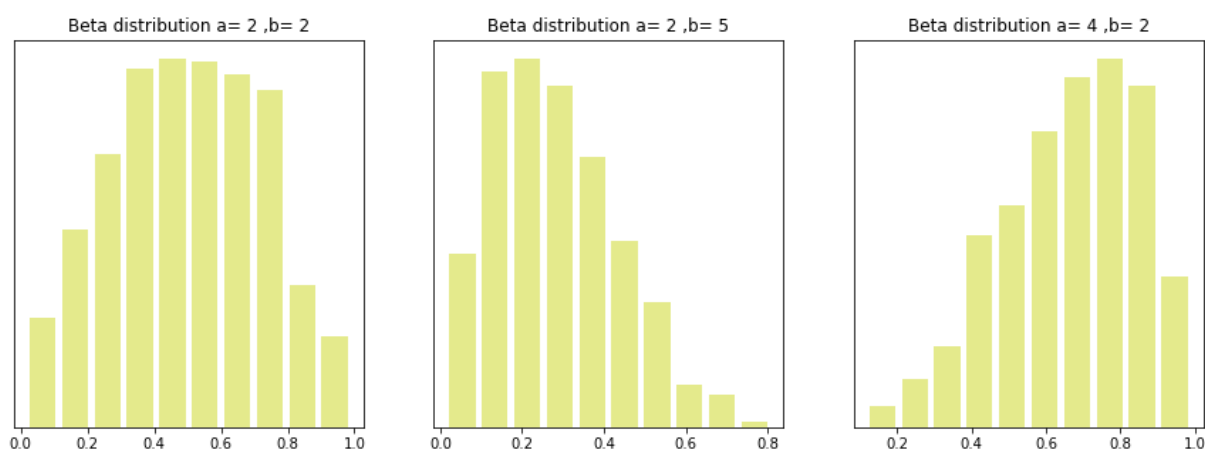
Сегашните вредности на параметрите се превземени од статистиките пресметани за вирусот COVID-19. Истите тие параметри можат лесно да бидат променети.

Имплементација

Имплементацијата на моделот во програмскиот јазик Python може да се погледне во функцијата `model` во Jupyter тетратката со име `SEICRD_model.ipynb`. Функцијата прима како аргументи вредности на параметрите кои се наведени во делот параметри, како и мрежата на контакти на популацијата. Ваквата имплементација ни дозволува да анализираме за различни комбинации на параметри за истата или различна мрежа на контакти како би се одвивало проширувањето на вирусот. Во функцијата се задава и максимален број на временски единици поточно итерации. За секоја од итерациите за секој од поединците во популацијата, односно јазлите во мрежата пресметуваме нова состојба. Дополнително после секоја итерација пресметуваме и заедничка состојба на популацијата која е дополнително визуелизирана на сликите подолу во делот Резултати.

Моделот со различни параметри на бета дистрибуцијата

Бета (β) параметарот претставува веројатноста да се пренесе вирусот од едно заболено лице на здраво лице доколку тие имале контакт. Овој параметар е различен за секој од поединците и неговата дистрибуција претставува бета распределбата. Како што можеме да видиме од граfiците доле, вредностите кои се резултат на бета функцијата се во рангот на 0 и 1, и се разликуваат во однос на нивните хипер-параметри. За различни комбинации на a и b параметрите добиваме дистрибуции чии просеци се поблиску на пример до 0.3 доколку сакаме да кажеме дека постои мала веројатност за пренесување на вирусот, или пак до 0.8 доколку постои поголема веројатност на пренос. Без разлика кои се хипер-параметрите секогаш постои дел од популацијата која има значително поголема или помала вредност од просекот што ја отсликува вистинската ситуација која ја моделираме во природата – некои луѓе повеќе се држат до мерките, за разлика од другите.



слика 3 Хистограми на бета дистрибуцијата за различни вредности на хипер- параметрите

За да направиме споредба на ширење на вирусот со мал број на мерки (поголема веројатност на заразување) и ширење на истиот вирус со поголем број на мерки (помала веројатност на заразување) потребно е истата функција да ја извршиме 2 пати на иста мрежа со различни вредности за хипер-параметрите од кои се пресметува параметарот β .

Моделот со различни степени на social distancing

Доколку манипулирање со β параметарот ни отсликува различни мерки како носење маски, честа дезинфекција итн, за мерката социјално дистанцирање (social distancing) мораме да го намалиме бројот на контакти. Контактите се претставени со ребра помеѓу јазлите во мрежата, па потребно е да отстраниме дел од нив. За најдобро да ја моделираме оваа мерка од надворешниот свет, мораме да забележиме дека контактите кои се намалени се всушност оние контакти кои не ни се секојдневни и оние контакти кои не спаѓаат во нашиот круг на најблиски лица (како блиско семејство, блиски другари, можеби соседи и неколку колеги). Ако ја разгледаме популацијата како мрежа, овие кругови на блиски луѓе дефинитивно би ни претставувале кластери и потребно е да ги отстраниме ребрата кои ги поврзуваат овие кластери. Од теоријата на графовите знаеме дека тие се ребрата кои имаат најголем коефициент на centrality. На сликата подолу може да се забележи хистограмот на пресметаните centrality коефициенти на ребрата на мрежата која ќе ја користиме подолу во одделот Резултати. Во зависност од тоа колку се почитува мерката социјално дистанцирање ја определуваме праг вредноста за тоа кои ребра би ги отстраниле.

За да направиме споредба на ширење на вирусот со нормален број на контакти и ширење на истиот вирус со намален број на контакти потребно е истата функција да ја извршиме 2 пати на цо истите параметри на две различни мрежи, првичната и една со отстранети ребра.

Моделот со применети мерки после одредено време

Функциите `model_restrictions()` и `model_sd_restrictions()` претставуваат проширување на функцијата `model()`. Поради тоа што во ситуациите што ги моделираме, заштитните мерки што се превземаат обично се последица на брзо ширање на вирусот, наместо да споредуваме како би се ширел истиот вирус со различни параметри или различни (скратени) мрежи од самиот почеток, ние ќе претставиме ситуација каде откако ќе се постигне одредена бројка на заболени лица се менуваат и донесуваат нови мерки (промена на параметрите и изгледот на мрежата).

Во функцијата `model_restrictions()` откако ќе се надмине предефинирана бројка на заболени лица ($\approx 30\%$ од популацијата) го намалуваме просекот на дистрибуцијата на β параметарот (ги променуваме `hyper`-параметрите).

Во функцијата `model_sd_restrictions()` откако ќе се надмине предефинирана бројка на заболени лица ($\approx 30\%$ од популацијата) го намалуваме просекот на дистрибуцијата на β параметарот (ги променуваме `hyper`-параметрите) и ги отстрануваме ребрата со висок centrality коефициент (намалување на контакти).

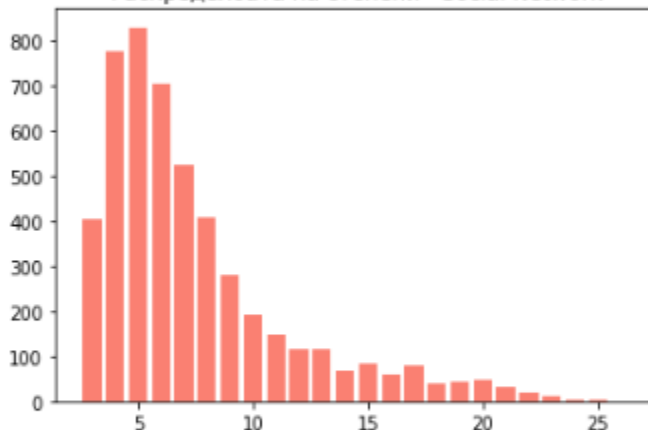
Резултати

Мрежата што ќе ја користиме во нашите модели е добиена со помош на Lancichinetti–Fortunato–Radicchi benchmark алгоритмот и ја нарекуваме “Social Network”. LFR Benchmark алгоритмот се користи за генерирање на вештачки мрежи кои имаат облик на мрежи од вистинскиот свет со јасно дефинирани заедници/кластери (communities). Овие мрежи потоа служат како мерка за проверување на методи за детекција на токму тие заедници. Ние сличноста со мрежите од вистинскиот свет, вклучувајќи ја и power law дистрибуцијата на степени на јазлите, како и поделбата во заедници ќе ја искористиме во решавање и симулирање на нашиот проблем.

Креираме мрежа со 5000 јазли со просечен број на степени по јазел 6 и максимален број на степени по јазел 20. Заедниците кои се формираат би требале да бидат со големина помеѓу 8 и 30. Со оваа мрежа ние претставуваме популација со големина 5000, во која меѓусебен секојдневен контакт едно лице има со околу 6 остана лица, а најмногу со 20. Со ова ги земаме во предвид контактите кои се позначајни и траат подолго време, не и оние кратки секојдневни контакти. Оваа популација е поделена на заедници кои се со големина од 8 до 30 лица.

Transitivity: 0.4873283265013016
Average clustering coefficient: 0.44044534038425237
Average shortest path: 6.702899459891978

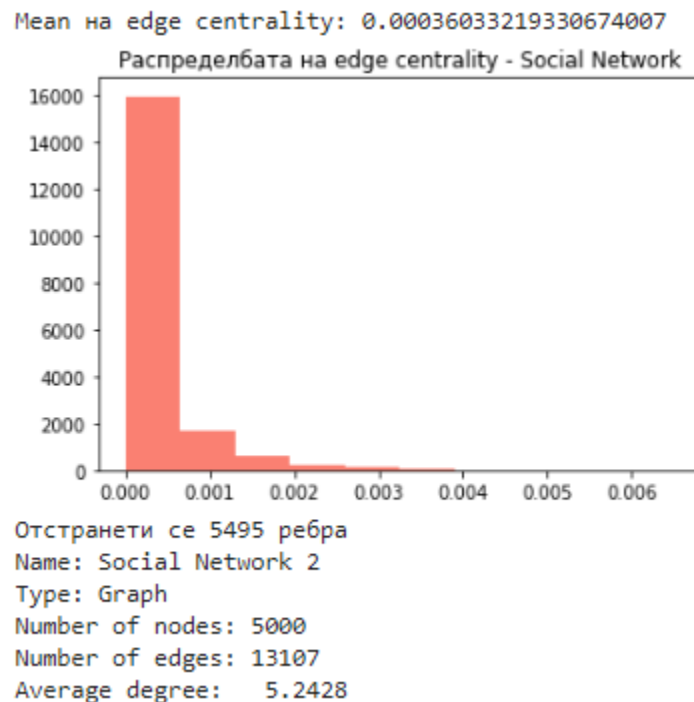
Распределбата на степени - Social Network



Name: Social Network
Type: Graph
Number of nodes: 5000
Number of edges: 18602
Average degree: 7.4408

слика 4 Анализа на мрежата "Social Network"

Дополнително за секој од ребрата е пресметан centrality коефициент и отстранети се ребрата со коефициент поголем од просекот – и новодобиената мрежа се нарекува “Social Network 2”.

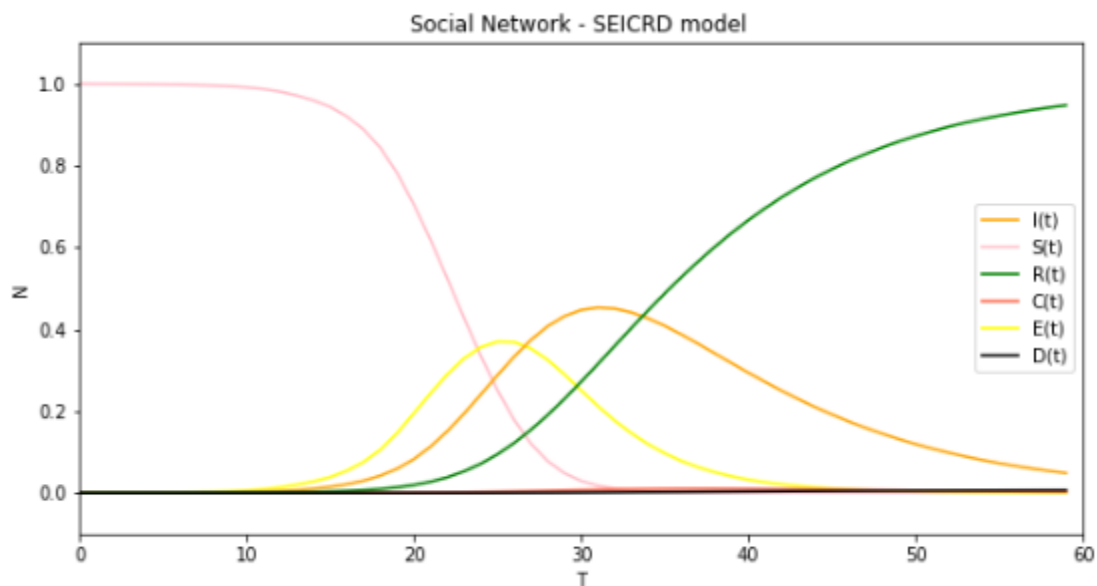


слика 5 Анализа на мрежата "Social Network 2"

Сумарните статистики ги прикажуваме со помош на визуелизација (пример слика 6). Секоја линија ја прикажуваат промената на бројот на лица во соодветниот сегмент низ времето. На x-оската е претставен бројот на временски единици (денови), додека на y-оската е претставена големината на популацијата со помош на проценти.

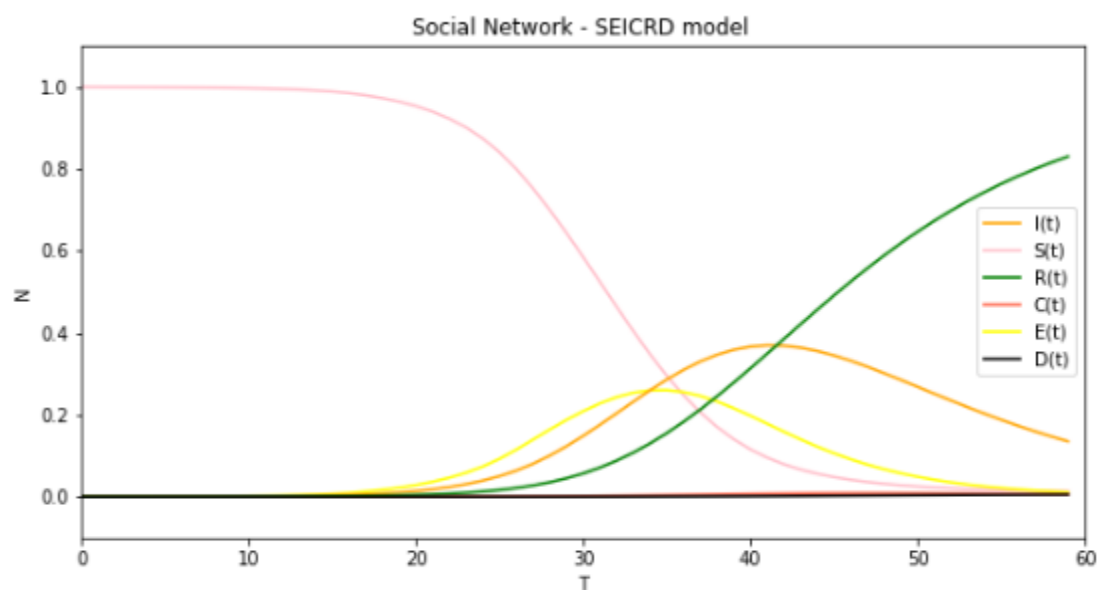
Заклучок

Со цел да се анализираат разликите во ширење на заразата, SEICRD моделот беше извршен пет пати на горе-споментатата мрежа (или нејзини модификации) за различни параметри.



Број на инфицирани со $\beta = (3,2)$: 4999

слика 6 Модел, бета = (3,2)

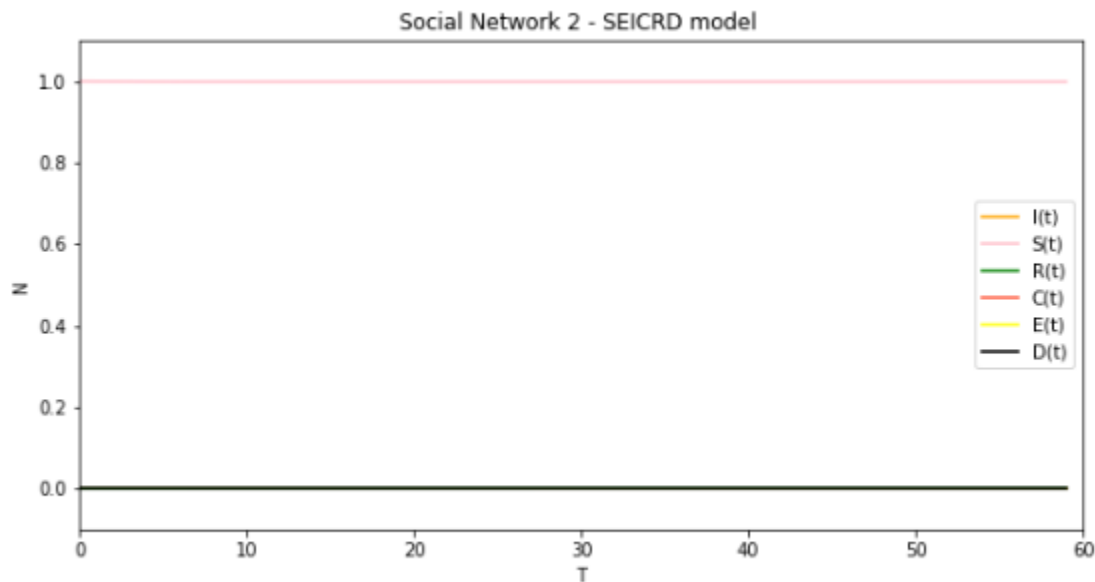


Број на инфицирани со $\beta = (2,5)$: 4933

слика 7 Модел, бета = (2,5)

Во визуелизациите прикажани на сликите 6 и 7 можеме го забележиме ширењето на вирусот со различни вредности на хурег-параметрите на бета распределата за бета параметарот. За вредностите (2,5) намален е број на заболени во однос на вредностите (3,2). Исто така можеме да забележиме побавно, поконтролирано растење на бројот на заболени (Infected)

и бројот на лица во допир со вирусот (Exposed). Овие намалувања можат да бидат позначителни кај поголеми популации.

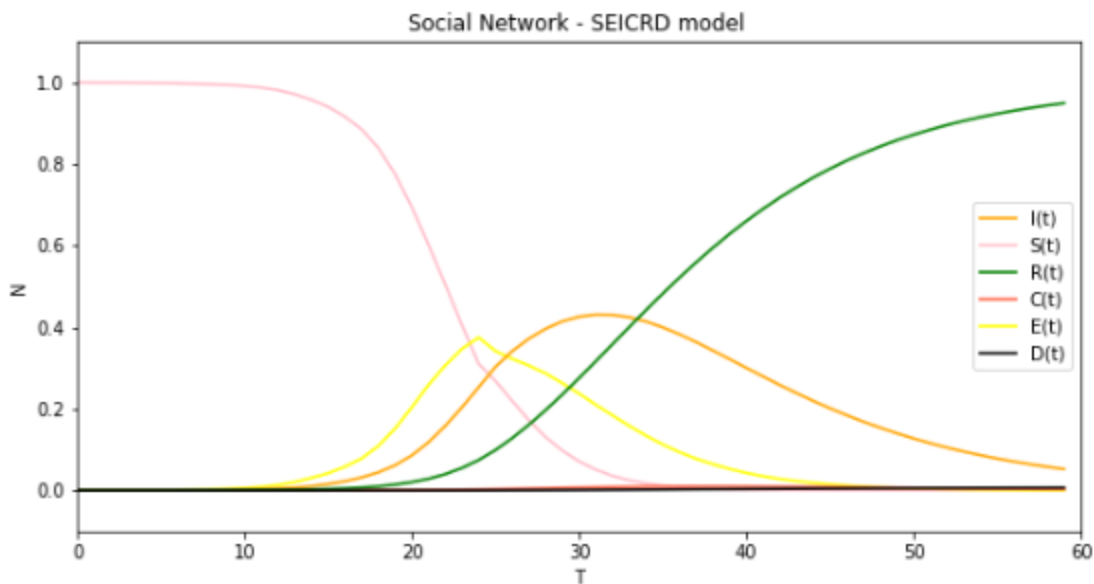


Број на инфицирани со помалце ребра: 7

слика 8 Модел, social distancing, бета = (3,2)

Во визуелизациите прикажани на сликата 8 можеме го забележиме ширењето на вирусот со различни степени на social distancing. Моделот е извршен на оригиналната мрежа “Social Network” и на модифицираната мрежа “Social Network 2”. Можеме да забележиме дека во мрежа со значително намален број на контакти (ребра) вирусот не успева многу да се прошири, и останува само во малата заедница на првобитното заболено лице. Земајќи ги овие статистики во предвид, се заклучува дека волкав степен на social distancing се препорачува бидејќи значително може да ја намали брзината на пренесување на заразата.

Рестрикциите започнуваат $t = 24$

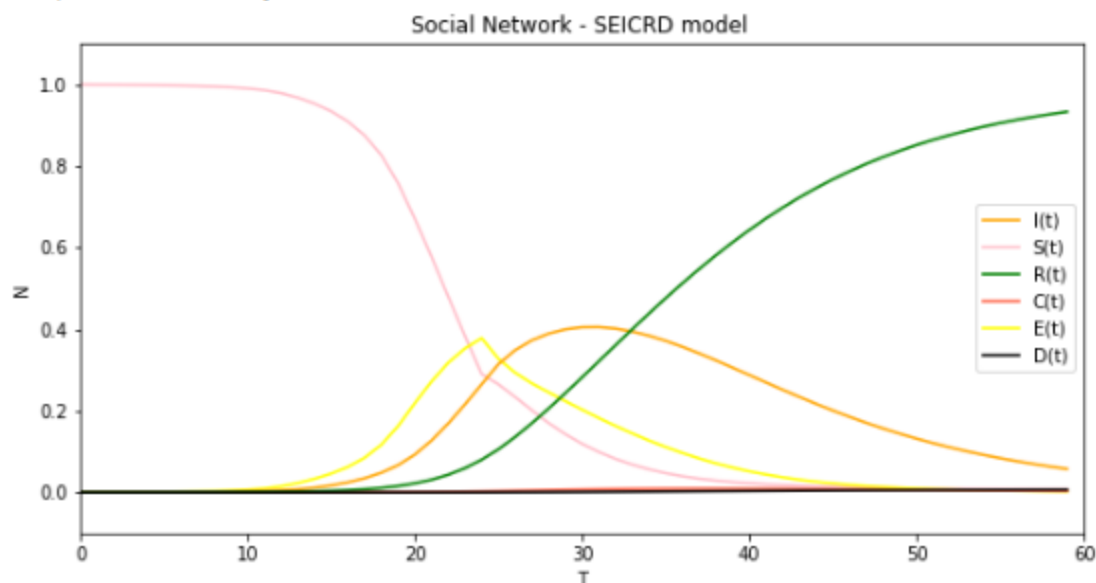


Број на инфицирани со рестрикции: 5000

слика 9 Модел, рестрикции (бета = (3,2) \rightarrow бета = (2,5))

На сликата 9 е прикажана визуелизацијата од SEICRD моделот со рестрикции (`model_restrictions()`) каде што хипер-параметрите на распределбата на бета параметарот се променуваат од (3,2) во (2,5). Можеме да забележиме нагол пад во броевите на заболени (Infected) и броевите на лицата кој биле во допир со вирусот (Exposed) за $t=24$, временската единица од која започнуваат да се применуваат рестрикциите. Оваа визуелизација можеме да ја споредиме со визуелизацијата на сликата 6 за да ги увидиме разликите во ширење на вирусот без применети мерки и после применување на мерки во некоја временска единица.

Рестрикциите започнуваат $t = 24$



Број на инфицирани со social distancing: 4961

слика 10 Модел, рестрикции ($\beta = (3,2) \rightarrow \beta = (2,5)$), social distancing

Сликата 10 претставува визуелизација на SEICRD моделот со рестрикции вклучувајќи и social distancing. За моделирање на рестрикциите ги променуваме hyper-параметрите на распределбата на бета параметарот ($(3,2) \rightarrow (2,5)$), а за моделирање на social distancing ги отстрануваме ребрата со висок centrality коефициент. Слично на претходниот пример, до временската единица $t=24$ ширењето на вирусот наликува на слика 6, но потоа доколку споредуваме продолжуваме да споредуваме со слика 6 забележуваме уште понагол пад на бројот на заболени и лица кои биле во контакт на вирусот. Земајќи ги овие статистики во предвид, се заклучува дека со применување на овие рестрикции се намалува брзината на пренесување на заразата, правејќи го вирусот поконтролиран, намалувајќи ги оптоварувањето на медицинските капацитети, ресурси итн.

Референци

- [1] Froese H., "[Infectious Disease Modelling: Fit Your Model to Coronavirus Data](#)", Towards Data Science, April 2020
- [2] Froese H., "[Build an interactive dashboard to simulate Coronavirus scenarios in Python](#)", Towards Data Science, May 2020
- [3] [NetworkX – Network Analysis in Python](#)