Peer to Peer Systems and Blockchains

Final Project Report

Riccardo Paoletti
Student ID: 532143

Year 2019/2020

Abstract

This report refers about the implementation of a model to simulate the diffusion of a smartphone virus spreading through Bluetooth connection. The diffusion of the virus has been modelled using the SIR model, while the mobility of the peers has been inspired by the one described in [4].

1 Implementation

The implementation exploits adjacency lists.

An HashMap between Peers and a list of Edges keeps tracks of peers in the board and eventual transmissions between them. Hotspots also are modelled as peers but are kept in a separate list and are immutable.

The simulation performs several iterations, through which the moving state and the infection state of peers are muted according to the selected probability model. In each iteration, first the moving state of peers is muted, then all the peers move following the behavior of their new state and finally the next infection state is computed according to the new positions of peers.

The computations have been performed using several thread pools.

1.1 Generation of peers

Initially, each peer is generated providing it an operating system, a direction a position.

The operating system is chosen according to the probability selected by the programmer and this is important because the virus can be propagated only among peers with the same operating system. Before assigning the peer an initial position, its belonging to a hotspot is defined.

If it is generated in a hotspot, the position is chosen using a generation function that is inspired by the Zipf's Law [17]. This function is adapted to this purpose and a bit simplified, in such a way that its more probable to remain close to a hotspot, and the probability to be placed at a certain distance becomes lower and lower with the increase of the distance itself, as illustrated in figure 1.
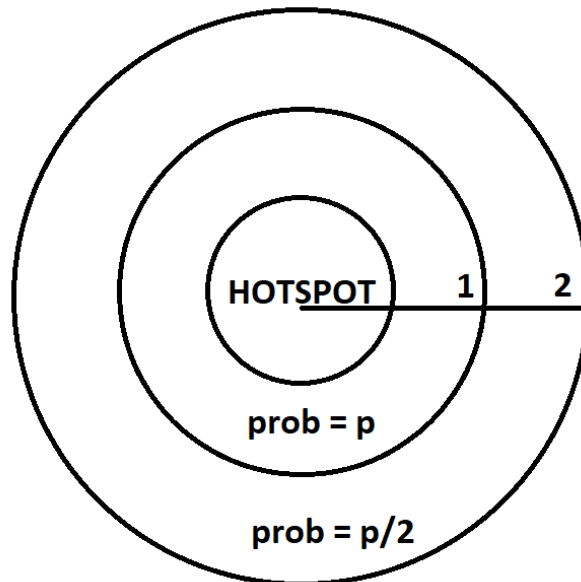

Figure 1: Hotspot distance probability model

In this case, the new position is computed using polar coordinates, generating first the angle in a uniformly random way, and then the distance according to the process described before.

Once all peers have been inserted in the board, the last peer (the infectious one) is generated with the same process and the iterations can start.

1.2 Movement

Every peer has a current position and a current direction (expressed as a position in the board). Every iteration, each peer performs a step toward its direction, only if its state allows it. The length of the step is previously set by the programmer according to its taste.

As described in [4], each peer has a state that describes its movement. Every iteration, each peer can change state according to a probability that depends on the current state and the destination state, and is set before the start of the simulation. Possible states are

- Halting - the peer is not moving this iteration;

- Exploring - the peer is exploring the map;

- Travelling - the peer is moving to a new location in the map.

The change in the movement state of a peer brings to a change of its direction, in such a way that peers are more likely to move around some hotspots in the board, but has also a probability to move outward to hotspots and move randomly.

When a peer has to select a new direction out of a hotspot, it is selected uniformly, and models an exploring behaviour of the peer, while if the new direction has to be selected toward a hotspot, then it is computed with the same process as for the generation of a peer located near a hotspot.

1.3 Infection Spread

Infection spread is modelled with the SIR model [9], dividing the peers in

- Susceptible - can be infected;

- Infectious - can infect susceptible peers;

- Recovered - immune, or dead individuals;

All peers starts being susceptible to the disease, except for one that is infectious. The infection can be propagated only to the peers with the same operating system of the infectious peer.

Every iteration, all peers, either susceptible or infectious, have a probability to install a patch and become recovered. Once recovered a peer can never change state anymore.

When a susceptible peer enters in the infection radius of an infected one then it has a probability to become infectious. This probability is a parameter of the model and is chosen according to which illness is being simulated at the moment, like it is done for the patch probability.

When a peer is infected by another one, a new edge is created in the map, that goes from the infected node to the infecting one, and with a weight that corresponds to the current iteration.

When a peer install the patch and becomes recovered, the current iteration is saved as an attribute of the peer.

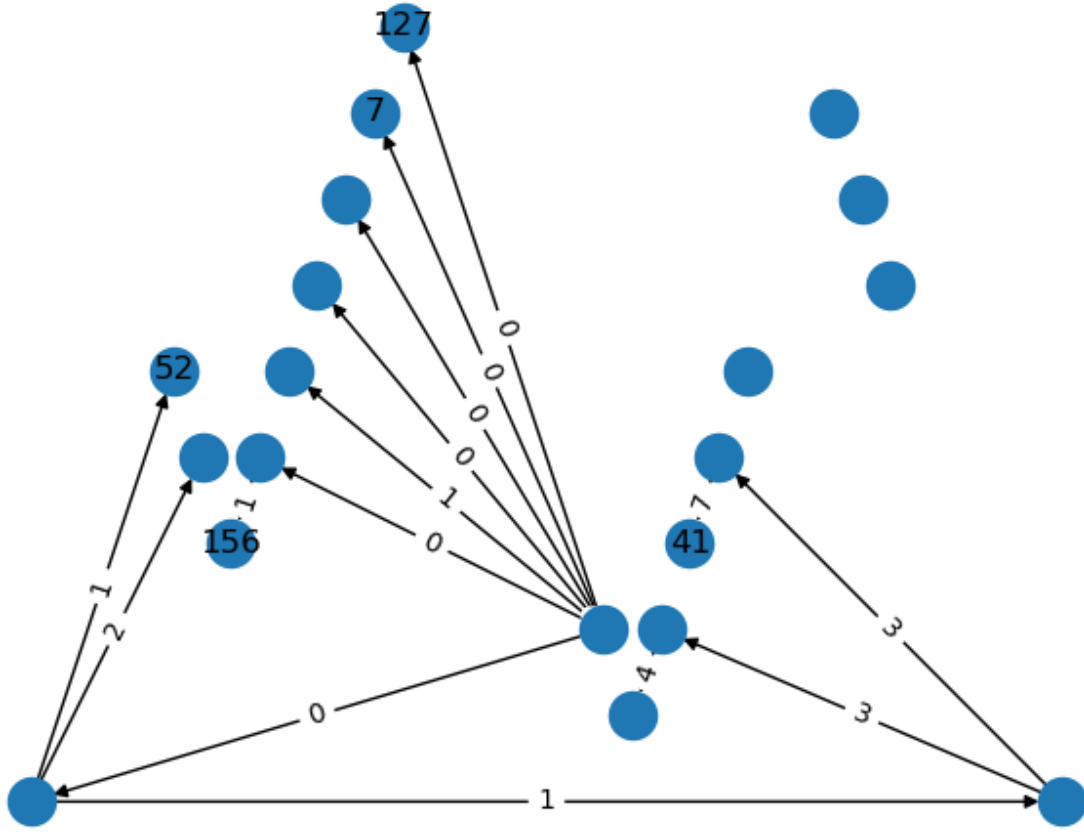An example of a random generated graph is showed in figure 2.

Figure 2: Example of Graph generated using the program. It has 20 peers, with an infection probability of 99% and a path probability of 0,3%. This is the result after 200 iterations. The positions showed in the picture are not the one in the board.

2   Experiments

Several experiments have been conducted trying to relate them with reality.

For such a purpose, in order to guide the selection of parameters related to the number of peers and hostpots, four cities have been selected, among the biggest, situated in different parts of the world: Tokyo, Mexico City, Istanbul and New York.

Then to selected the parameters related to infection spread, three real pandemic have been took as guideline: COVID-19, HIV-AIDS 2005-2012 and SPANISH FLU.

The number of peers and hotspots and the infection and patch probability of each simulation environment has been selected arbitrarily after some researches. The choices in terms of parameters are the following:

- For the cities:

- Tokyo - is the most populated city in the world but it's density is not quite so big, compared with the one of other smallest cities ([15]), so in order to maintain the simulation realistic a population of 60.000 peers has been selected and a number of hotspots equal to 150. Furthermore, the diffusion of mobile phones operating system in Tokyo is: ANDROID - 38,28%; IOS - 61,58%; OTHERS - 0,14%.

- Mexico City - has a density that is the same of Tokyo but with less area ([13]), so the population selected is still 60.000 peers, but the number of hotspot is 100. The operating system diffusion refers: ANDROID - 85,48%; IOS - 14,22%; OTHERS - 0,3%.

- New York - has the biggest density among the selected cities, but with the lesser population or area ([14]), so the number of peers has been put to 108.000 but with a number of hotspots equal to 50. Phone operating system diffusion: ANDROID - 41,65%; IOS - 58,17%; OTHERS - 0,18%.

- Istanbul - has the lesser population density, but with the biggest area ([12]), so a population of 29.000 peers has been selected, with 350 hotspots. Mobile phone operating system diffusion is: ANDROID - 83,91%; IOS - 15,69%; OTHERS - 0,4%.

- For the pandemics:

  - COVID-19 - is characterized by an extreme simplicity in diffusion, a very low fatality and a small radius required to contract the disease ([10], [16]). According to these factors, infection rate has been set to 99%, while path rate has been set to 5%, in such a way to simulate the period of 1-2 weeks that is necessary to manifest symptoms and to heal;

  - HIV-AIDS - has a very low infection rate due to the fact that its transmission is related with blood sharing. It has no chance to heal if infected, but several precautions can be token is order to avoid the contraction ([1], [3], [7], [5], [6]). To model this a patch rate of 5% has been chosen but only for peers that are not already infectious. The other peers cannot heal;

  - SPANISH FLU - is similar to COVID but with less infection rate ([2], [8], [11]), that has been set to 30%, with the same path rate, due to the fact that both has the same way to avoid the diffusion.

All experiments took place in a 100x100 board, with an infection radius of 2, a step size of 1 and with the moving state graph showed in figure 3 with arbitrarily selected probabilities. These parameters also has been arbitrarily selected in order to perform a simulation inspired by the reality.
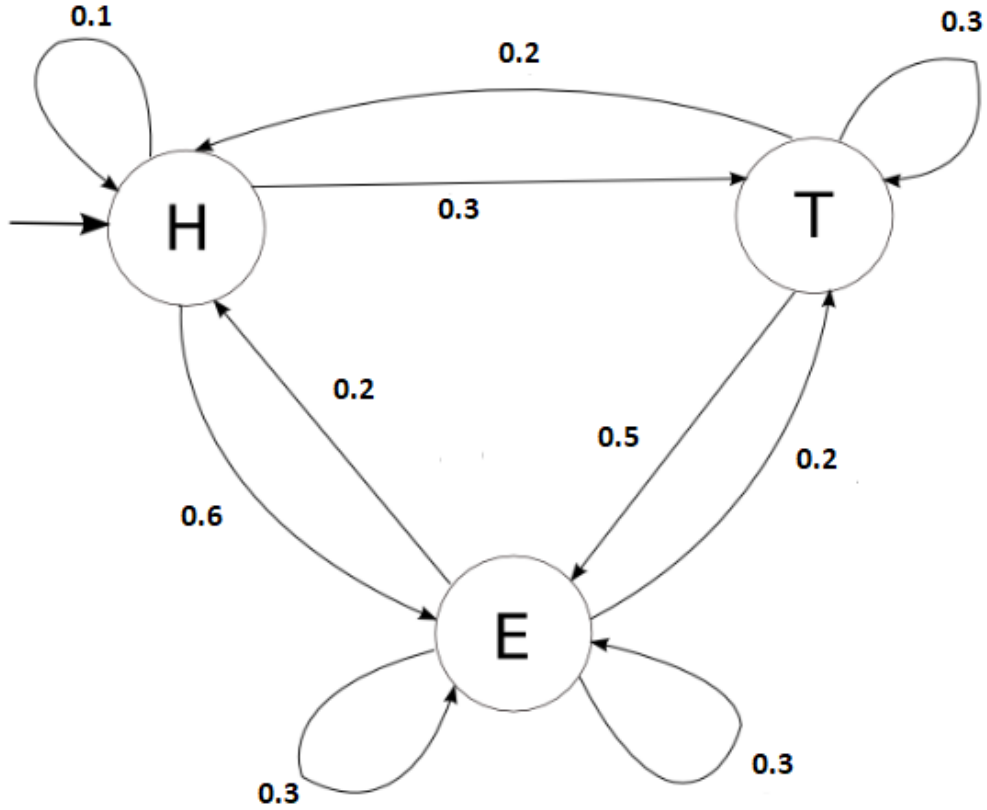
Figure 3: State transition probability graph

The hotspot radius has been set to 6, the probability of remain in a hotspot once inside is 90%, while the percentage of population spawned in hotspots is 70%.

For each experiment 200 iterations have been performed, and 5 tries in order to take the average results.

3   Results

Here are presented the results related to the experiment explained in the section before.

3.1   Number of infected nodes per cycle

3.1.1   COVID-19

Here are presented the number of infected peers per iteration in the COVID-19 simulation environment.
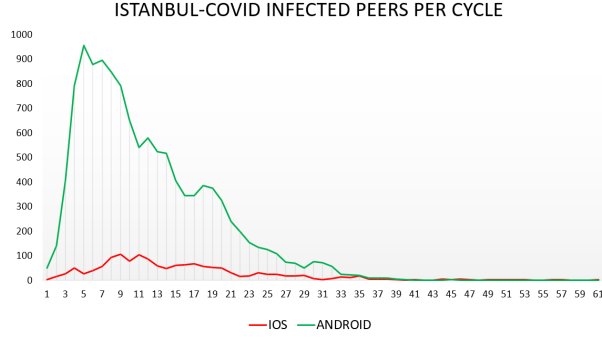
Figure 4: Number of infected peers per cycle in Istanbul simulation
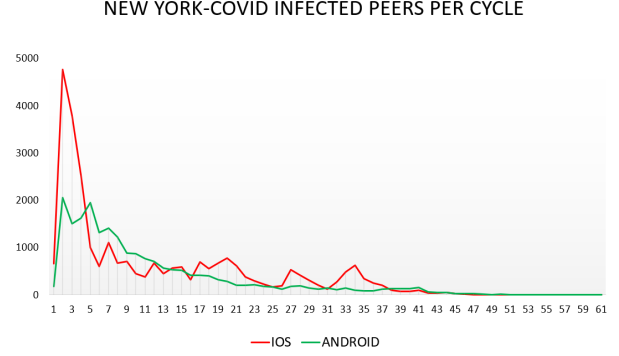
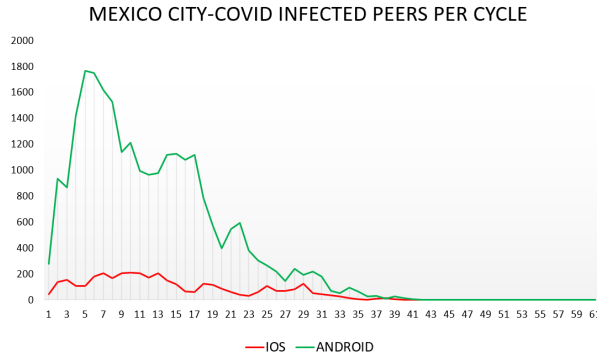Figure 6: Number of infected peers per cycle in New York simulation

Figure 5: Number of infected peers per cycle in Mexico City simulation
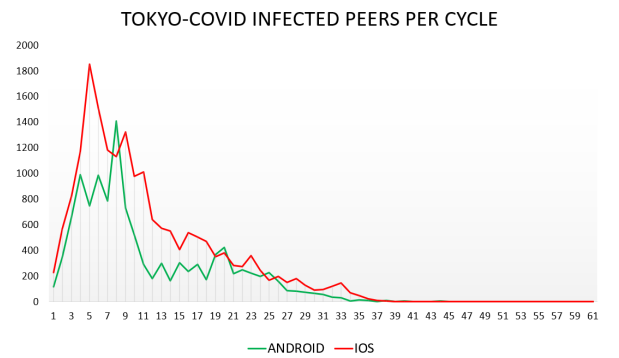

Figure 7: Number of infected peers per cycle in Tokyo simulation

Being this simulation the one with the highest infection probability, I should expect that the peak in number of infected peers per cycle is reached quite early in the execution, and this is in some way always true for each environment.

As we can see from the pictures above, in all the cities the highest value is located always near the third/fifth iteration, with a regular decreasing process that terminates around the thirtieth cycle.

New York simulation represents quite an exception in the sense that, after the peak located around the third cycle, the number of infected peers falls down very quickly and starts to have a sort of saw-tooth behaviour until the thirty-fifth cycle. It is worth to remind that New York simulation is the one with the highest number of peers, but with the lesser number of hotspots.

3.1.2 HIV-AIDS

Here are presented the number of infected peers per iteration in the HIV-AIDS simulation environment.
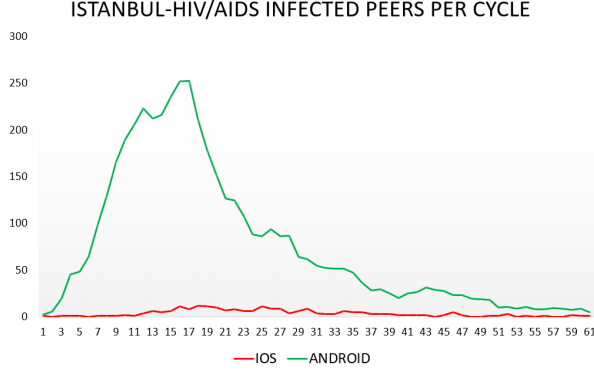
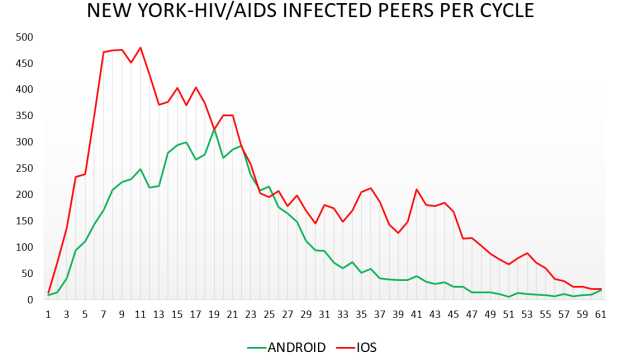Figure 8: Number of infected peers per cycle in Istanbul simulation

Figure 10: Number of infected peers per cycle in New York simulation
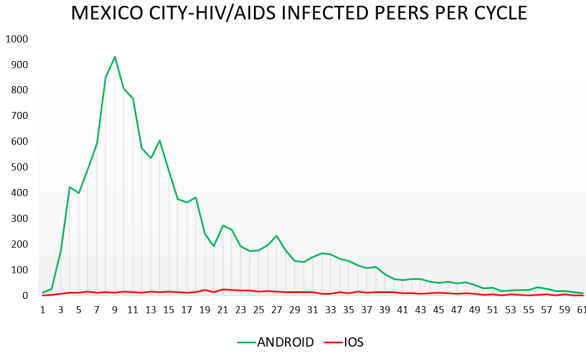

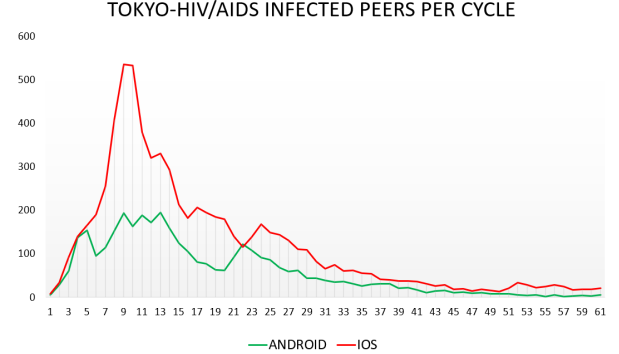Figure 9: Number of infected peers per cycle in Mexico City simulation

Figure 11: Number of infected peers per cycle in Tokyo simulation

This simulation is the one with the lesser infection probability, so we would expect that the number of infected peers per cycle is a bit better distributed in the cycles, respect to what we saw for the COVID-19 simulation.

In some sense our expectations are satisfied because, as we can see from the pictures above, the peak of infected peers this time is located around the ninth cycle, with the only exception of Istanbul simulation, where it is located around the seventeenth iteration.

As it was expected, the number of infected peers is quite lesser than the one saw in COVID-19 simulation for all the environments, and it is worth to notice how for the New York simulation, the number of infected peers per cycle is much more distributed and decreases in a very regular way.

3.1.3 SPANISH FLU

Here are presented the number of infected peers per iteration in the SPANISH FLU simulation environment.
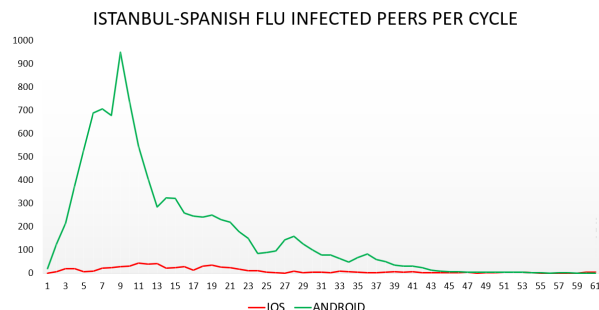
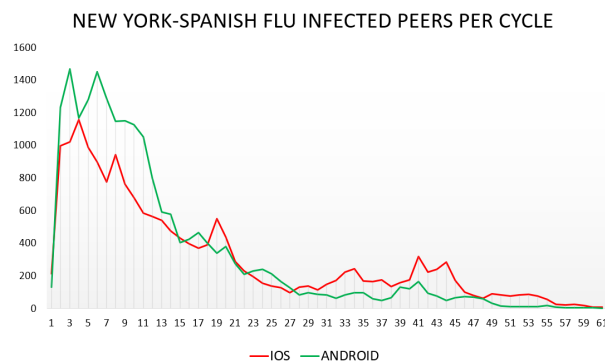Figure 12: Number of infected peers per cycle in Istanbul simulation

Figure 14: Number of infected peers per cycle in New York simulation
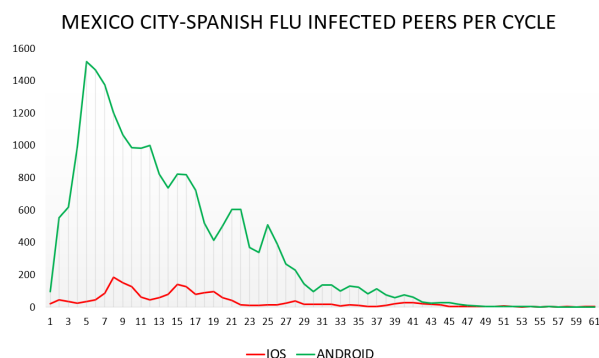

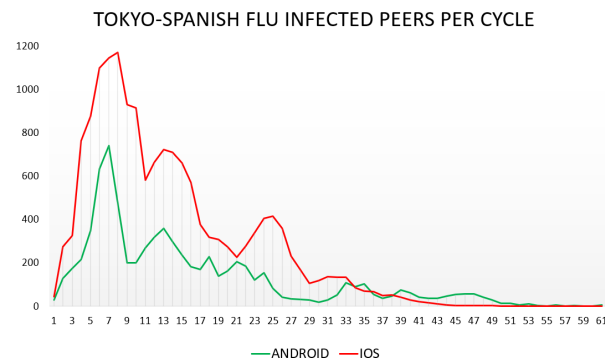Figure 13: Number of infected peers per cycle in Mexico City simulation

Figure 15: Number of infected peers per cycle in Tokyo simulation

This simulation is the one with the infection probability in the middle between the two seen above.

The number of infected peers per cycle is quite the one we should expect, in fact it is lesser than the one for COVID-19 simulation and greater than the one for HIV-AIDS simulation.

As for the simulations showed before, also here the peak is located quite early in the execution, around the fifth/seventh iteration, with a decrease that is regular for all the environments.

The strangest behaviour is, as always, in the New York simulation, where we have two peaks in the second and seventh iterations with a saw-tooth distribution in the decreasing part of the graph.

9

3.2 Number of cycles required to reach an infection percentage

3.2.1 COVID-19

Here are presented the infection percentage per iteration in the COVID-19 simulation environment.
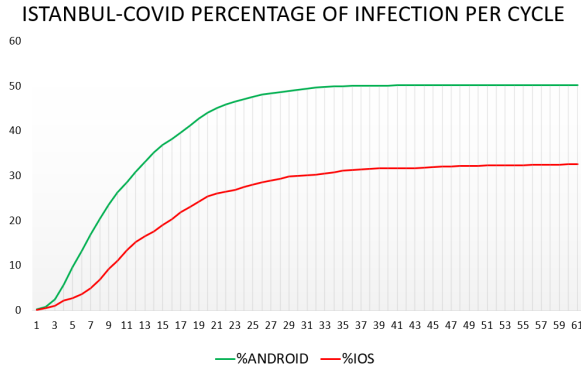

Figure 16: Number of infected peers per cycle in Istanbul simulation
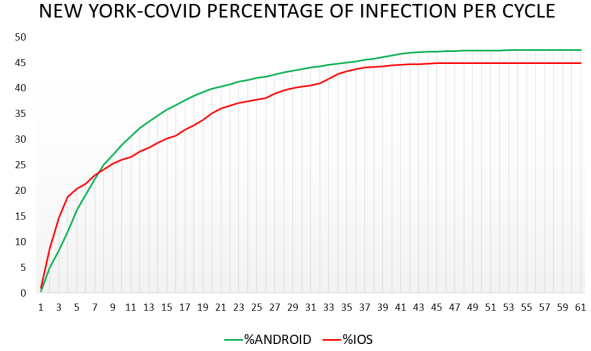

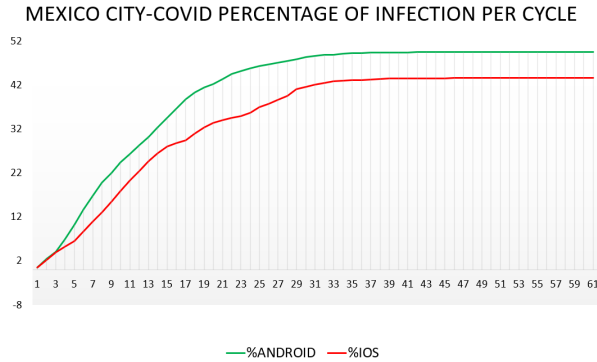Figure 18: Number of infected peers per cycle in New York simulation

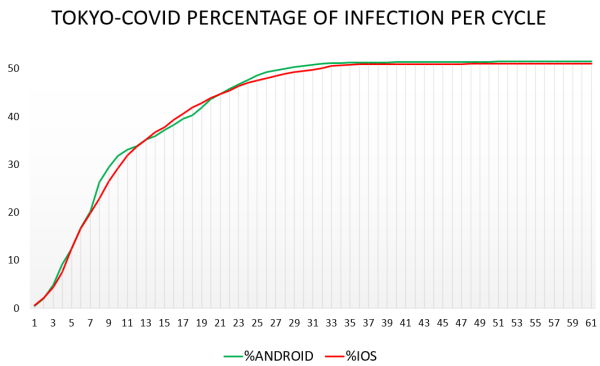Figure 17: Number of infected peers per cycle in Mexico City simulation

Figure 19: Number of infected peers per cycle in Tokyo simulation

In this simulation, as we can see, the percentage of infected peer stabilizes around the 50%. This result is reached around the thirtieth cycle for all the environments, where the behaviour is quite the same.

The only "exception" is New York, where the flattered part of the curve is reached a bit later, around the forty-fifth iteration, probably because of the size of the population.

It is worth to notice that the same behaviour is taken as for the ANDROID population infected, as for the IOS population, even if, for example in Mexico City simulation the IOS population is a much more smaller part, respect to the ANDROID one.

3.2.2 HIV-AIDS

Here are presented the infection percentage per iteration in the HIV-AIDS simulation environment.

ISTANBUL-HIV/AIDS PERCENTAGE OF INFECTION PER CYCLE

Figure 20: Number of infected peers per cycle in Istanbul simulation

NEW YORK-HIV/AIDS PERCENTAGE OF INFECTION PER CYCLE

Figure 22: Number of infected peers per cycle in New York simulation

MEXICO CITY-HIV/AIDS PERCENTAGE OF INFECTION PER CYCLE

Figure 21: Number of infected peers per cycle in Mexico City simulation

TOKYO-HIV/AIDS PERCENTAGE OF INFECTION PER CYCLE

Figure 23: Number of infected peers per cycle in Tokyo simulation

Differently from what we have seen for the COVID-19 simulation, here the maximum percentage reached in the infection is around 20%/25%, due to the much lesser infection probability of this simulation.

The flattered part of the curve is reached around the thirty-fifth/fortieth cycle for quite all the environments.

Here it is worth to notice how the behaviour for the ANDROID population, and the one for the IOS population are quite different, having the second one that reaches a very low percentage of population infected in environments where it represents a minority of the population, like Istanbul or Mexico City.

3.2.3 SPANISH FLU

Here are presented the infection percentage per iteration in the SPANISH FLU simulation environment.
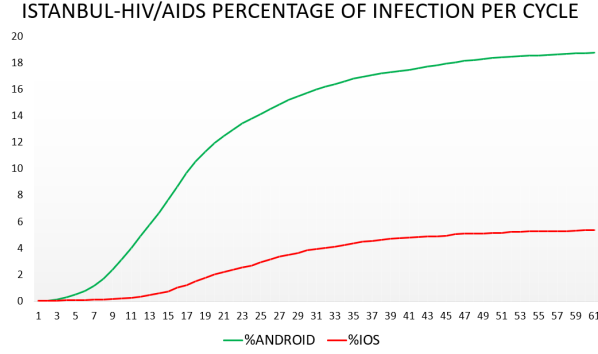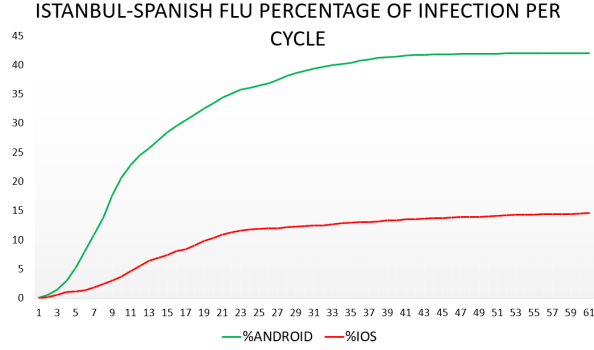

Figure 24: Number of infected peers per cycle in Istanbul simulation
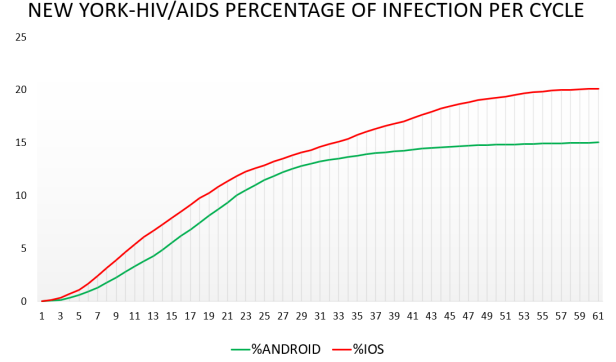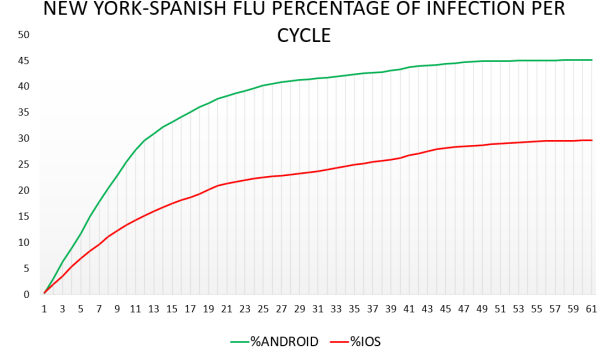

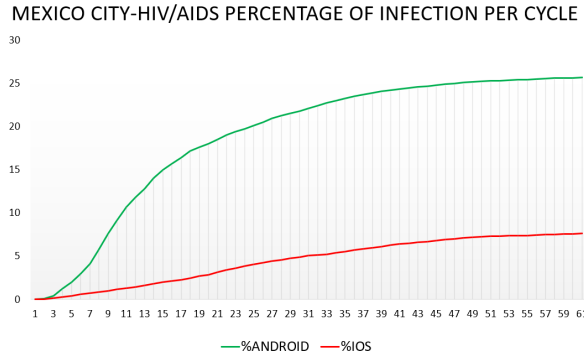Figure 26: Number of infected peers per cycle in New York simulation

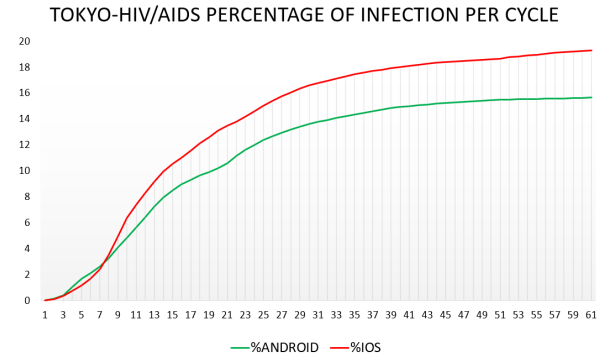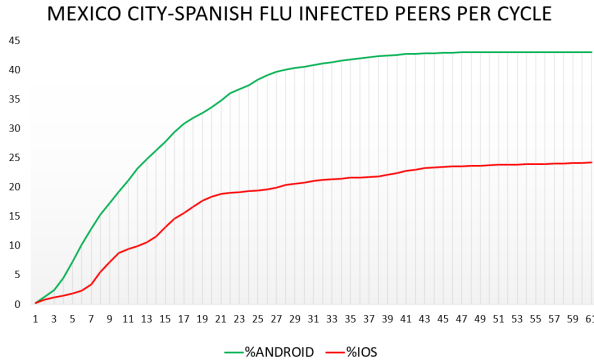Figure 25: Number of infected peers per cycle in Mexico City simulation
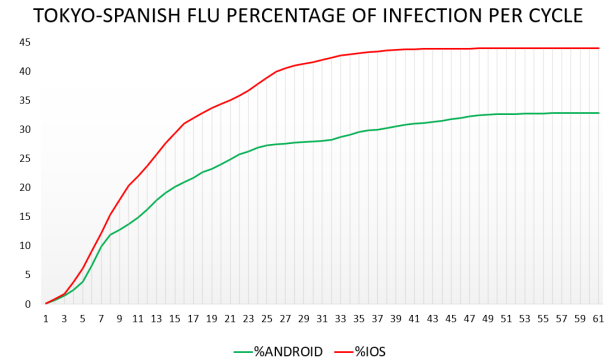
Figure 27: Number of infected peers per cycle in Tokyo simulation

In this simulation, the maximum percentage of infection is located around 45% of all peers. It is reached around the thirtieth cycle for quite all the environments.

We can notice how the percentage is quite similar to the one of COVID-19, even if the infection probability is much smaller. This could lead to think that there is a threshold in the percentage of contagion that, even increasing a lot the infection percentage, cannot be overcome.

3.3   Structural Properties

		AVG node degree	AVG clustering coefficient	AVG diameter	AVG edge density
COVID					
	Istanbul	1,99966	0	52,6	0,0004
	Mexico City	1,99992	0	54,8	0,000076
	New York	2,0096	0	49,4	0,0000756
	Tokyo	1,99992	0	57,6	0,000187
HIV-AIDS					
	Istanbul	2	0	38,2	0,00216
	Mexico City	2	0	41,2	0,00129
	New York	2,0252	0	34,6	0,000222
	Tokyo	2	0	45,6	0,000392
SPANISH FLU					
	Istanbul	1,99982	0	64,4	0,000198
	Mexico City	1,99994	0	61,4	0,000088
	New York	2,0364	0	43,2	0,000102
	Tokyo	1,9999	0	58,6	0,000116

Figure 28: Structural properties of graphs resulting from several simulations

In the table 28 we can see the most important information gathered about the graphs obtained from the simulation described in Experiments section.

First thing we can notice is that the average node degree for quite all the simulation is 2, and this states how every node, in average, has two edges incident, that means that it has been infected by another node and infected himself another node.

We can notice also that the average clustering coefficient is 0 for all the simulations. That could be due to the fact that, having each peer that infected only another peer, they tend to form chains of contagion, without forming groups, and resulting in the low clustering coefficient obtained.

Edge density is very low for all the simulations, but this is a result that is in line with what we have seen for the average node degree, thinking to the number of peers that we have in the graphs and that the number of edges is quite far from the maximum number of possible edges.

The average diameter remains quite the same for each simulation. The biggest exception is New York where it's a bit lower than the other environments, probably due to the fact that it's the environment with more peers of all and so where the infections can propagate better.

4 Conclusions

This report evaluates the development of a simulator for a mobile phone virus spreading via Bluetooth. We have reported the most significant design choices taken and presented some results about several performed experiments.

The implementation exploits adjacency lists, SIR model and a moving system inspired by [4]. Different experiments have been performed, with different populations and different infection rates, trying to simulate situations that could take place in reality, obviously with the limits of the simulation.

In almost all the tested environments, we observed a peak in the infection of peers in the first iterations of the execution. Instead, in the following iterations they decreased quite regularly up to reach a very low value.

Although the behaviours of the different experiments are quite the same but with different scales, for more contagious diseases we observed a greater number of infected peers per cycle, and a greater percentage of infected population, in line with the expectations.

Finally, analyzing the various graphs produced, a set of statistics have been gathered, like average node degree, average clustering coefficient and the average diameter.

The first is the same for all the simulations and it is equal to 2. This shows how every peer in the network has two incident edges on average, and so that every peer is infected by another peer and infects itself only one another peer, forming a chain-like sub-graph of infections, one after the other.

We also noticed that the average clustering coefficient is 0 for all the experiments, that is in line with what we said before about the chains, because the nodes, being connected only with other two nodes, have difficulties to form clusters, even if the number of hotspots is quite high.

References

[1] ars.toscana. *HIV e AIDS: i numeri in Italia, in Europa e nel mondo*. [Online; published in 2012]. 2012. URL: https://www.ars.toscana.it/aree-dintervento/problemi-di-salute/malattie-infettive/hivaids/news/2232-hiv-e-aids-i-numeri-in-italia-in-europa-e-nel-mondo.html.

[2] Centers for Disease Control and Prevention. *1918 Pandemic (H1N1 virus)*. [Online; Page last reviewed: March 20, 2019]. URL: https://www.cdc.gov/flu/pandemic-resources/1918-pandemic-h1n1.html.

[3] Centers for Disease Control and Prevention. *HIV Risk Behaviors*. [Online; Page last reviewed: November 13, 2019]. URL: https://www.cdc.gov/hiv/risk/estimates/riskbehaviors.html.

[4] Sergey Legtchenko, Sébastien Monnet, and Gaël Thomas. "Blue Banana: resilience to avatar mobility in distributed MMOGs". In: *2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN)*. IEEE. 2010, pp. 171–180.

[5] World Health Organization. *AIDS epidemic update: December 2005*. [Online; 2005]. 2005. URL: https://www.who.int/hiv/pub/epidemiology/epiupdate2005/en/.

[6] ourworldindata.org. *HIV / AIDS*. [Online; published in November 2014; last revised in November 2019]. 2019. URL: https://ourworldindata.org/hiv-aids.

[7] Istituto Superiore di Sanità. *Infezione da HIV e AIDS, Aspetti epidemiologici*. [Online; 2018]. URL: https://www.epicentro.iss.it/aids/epidemiologia-mondo.

[8] Istituto Superiore di Sanità. *Le Pandemie Influenzali del Ventesimo Secolo*. [Online; Page last reviewed: November 11, 2010]. URL: https://www.epicentro.iss.it/passi/storiePandemia.

[9] Wikipedia. *Compartmental models in epidemiology*. [Online; checked on 20 July 2020]. 2020. URL: https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology.

[10] Wikipedia. *COVID-19*. [Online; checked on 20 July 2020]. 2020. URL: https://it.wikipedia.org/wiki/COVID-19.

[11] Wikipedia. *Influenza spagnola*. [Online; Page last reviewed: 21 July 2020]. URL: https://it.wikipedia.org/wiki/Influenza_spagnola.

[12] Wikipedia. *Istanbul*. [Online; checked on 20 July 2020]. 2020. URL: https://it.wikipedia.org/wiki/Istanbul.

[13] Wikipedia. *Mexico City*. [Online; checked on 20 July 2020]. 2020. URL: https://en.wikipedia.org/wiki/Mexico_City.

[14] Wikipedia. *New York*. [Online; checked on 20 July 2020]. 2020. URL: https://it.wikipedia.org/wiki/New_York.

[15] Wikipedia. *Tokyo*. [Online; checked on 20 July 2020]. 2020. URL: https://it.wikipedia.org/wiki/Tokyo.

[16] www.statista.com. *Total number of coronavirus-positive (COVID-19) patients and number of deaths in Japan as of July 10, 2020, by prefecture*. [Online; checked on 10 July 2020]. 2020. URL: https://www.statista.com/statistics/1100113/japan-coronavirus-patients-by-prefecture/.

[17] George Kingsley Zipf. *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books, 2016.