UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENTO DE MATEMÁTICA APLICADA

Máster Universitario en Investigación Matemática (INVESTMAT)

Modelización de la difusión y persistencia de datos en redes oportunistas de comunicación móvil

Trabajo fin de Máster
Presentado por:
Carlos Orea Hueso

Dirigido por:
J. Alberto Conejero, Enrique Hernández-Orallo

València, septiembre de 2019

Abstract

In this thesis, we will study how mobile devices can share and keep information and what are the best protocols to achieve different goals. Moreover, we will try to guess which mathematical models best describe our situation so we can do precise forecasts.

The two main traits that distinguish MANETs (Mobile ad hoc Networks) and traditional technologies are the following: First, there is no fixed infrastructure, since the messages are carried and received/sent by the nodes themselves, and second, those nodes are, of course, generally moving. One of the main problems of those networks is to assure that the messages will arrive from the node who generated the message to the one we want to receive it, if necessary, jumping from one to another until reaching to the objective. This can fail for multiple reasons: one of the nodes in between has not enough storage capacity, they are too far apart, etc.

For this reason, we need to simulate different protocols to check which ones are the best, and for this, we need high-quality traces whose behavior assimilate to real-life moving nodes.

The steps that we will follow in this work are the following:

- Obtain information from a real-life scenario ("La plaza de San Vicente Ferrer")
- Feed this data into PEdSim to realistic traces.
- Use these traces into ONE simulator to check the performance of a protocol.
- Understand different mathematical models used to describe these processes.

To obtain the above-mentioned information from this square, we visited "La plaza de San Vicente Ferrer" and one by one, wrote done from where to where people moved and how long it took them. Then, other metrics such as speed can be directly calculated.

After that, we used this data, and altogether with the design of the square, we simulated it to get the traces. We right after simulated them with ONE.

Lastly, different mathematical models can be used in this situation, such as: SIR type models and dynamic graphs, etc. They are generally used in epidemiological disease situations, where there is a group of infected people, susceptible people, and recovered people. Such models admit many different nuances, so it is easily fitted into MANETs experiments. The analogy is clear, since we have nodes without the message (susceptible), and nodes with it (infected) which will infect other nodes if some requirements are met

The first ones are valid as long as the population remaining in the place is large enough. The second ones are more accurate when the size of the population in the

place is small. We will describe the theoretical part of both models and we will see how the second one work with the data obtained.

Contents

Chapter 1

Introduction

We focus on this project on the study of opportunistic networks in the square "Plaza de San Vicente Ferrer" in the Spanish city of València. In a common network, nodes are usually static and the communication paths are set in advance the messages are sent. In opportunistic networks, nodes are mobile devices and the paths are built dynamically, which means that they are not determined in advance, they change as the situation changes.

Nodes are chosen if they bring the message closer to the final node. This is similar to how diseases behave. We say that a node is infected when it creates or receives a message from another node. And we say that a node is susceptible to infection when it has not received the message yet.

1.1 Motivation

Since this kind of networks does not use any Internet connection or any other fixed infrastructure, this technology can be used in places without the above-mentioned technologies.

To further develop the opportunistic networks protocols traces are needed. This project focuses on the generation of realistic traces with which different protocols can be later compared.To have realistic traces, we needed to do real measures in a real place.

As something that probably happened to most of us, in crowded areas such as concerts or music festivals, mobile phones messaging applications stop working properly. This is because since those technologies such as 4G or WiFi do not work properly under these conditions. These are Opportunistic Networks respect to the time since the tool does not need any fixed infrastructure, since messages can be sent from some of the devices to others directly. Therefore, the common technology used here is usually Bluetooth or WiFi direct. As stated here, this is useful in crowded areas, remote areas or areas where a natural disaster has taken down the fixed infrastructure. This technology has already been used multiple times. One of the most famous is in the field of wildlife tracking networks. Also, this is a potentially useful way to communicate for soldiers.

We would like to know how good the performance of our tool is, for this, we need to define some metrics, such as overhead ratio. The performance will depend highly on which algorithm we use (this is, which rule we follow to send the message), and the mobility of the nodes. So different situations might need slightly different approaches. This is a complex task since we have nodes moving in an unknown (or close to unknown) pattern and also the technology carried by the device playing an important role in our study at hand.

1.2 Goals

Our main goal will be to determine how good (or bad) the performance these networks behave since this will help us do improvements later on. We also want to make it as real as possible, and for this reason, we went to the square itself to do real measures. So, even though we used software to obtain the traces, the data fed into it is as real as it can get. Where other works suppose a constant number of people in the area at study, we allow people to go in and out at their will.

These are the main goals in this master final thesis, and those are:

- Obtain a set of realistic traces
- Revise continuous and discrete models for epidemiological spreading.
- Study which mathematical method best describes better each process.
- Analyze what is the performance of the network.

1.3 Methodology

To analyze how the renewal of people in the square affects the performance of the protocol/algorithm. We create a version of our real area in PedSim (which is a software to obtain traces) and pedestrians are generated accordingly to the measurements. With these traces, different protocols will be implemented, and we will explain why different protocols have different performances in this situation. Then the same will be done for a broad range of messages size. Message size is an important variable, since it is related to the time needed for a message to be sent. And since contacts has a finite duration, for big messages, a percentage of them will fail to be sent, because the contact will finish before the message is completely received by the second node.

Therefore, Opportunistic Networks work (nowadays), better with smaller messages than bigger ones.

1.4 Structure

In this thesis, we present the work done as follows. We have 9 different chapters. We move from the more general to the more precise. We introduce the matter at hand, the state of the art of the technologies used and the best mathematical model used to forecast this process. And then we move to our specific case, our scenario simulated, the traces obtained from PedSim and how we did it, the same with ONE and after that, we study how mathematical network models behave. Finally, the results are presented and we draw some conclusions from it.

Chapter 2

MANET

2.1 Introduction

MANET stands for Mobile Ad Hoc Network. Here, there is a group of nodes connected wirelessly with some specific characteristics such as restricted band with or restricted storage capacity. Virtually every electronic device can work as a node: laptops, mobile phones, smartwatches, etc. Therefore, since each of the nodes is mobile, the network itself is mobile too. This technology started, like many other technological advances, with military uses, it was called back then Mobile Packet Radio Network.

Lately, the rise in communication technologies, and instant message applications has made our society highly connected. In MANETs, the structure of the network is not fixed in advance, and each of the nodes works both as a router and a host at the same time.

2.2 Wireless networks

This type of networks allows flexible connection between members in different locations. We have two types of wireless networks:

1. Wireless networks
    - Infrastructure Networks
    - Ad-Hoc Networks

The first type of wireless network (infrastructure networks) depends on an external infrastructure and architecture. This makes possible for different nodes to communicate with each other.

The other type of wireless networks is infrastructure less networks. Commonly known as Mobile ad Hoc Networks (MANETs), the trait of these networks which differentiates them from the others is that there is no centralized administrator. In the before mentioned case, the network relies on the access point and in wired networks it relies on routers. Here the network does not have a fixed topology. Therefore, the source node and destination node can send messages between each other thanks to the other nodes, which will together as a communication chain.

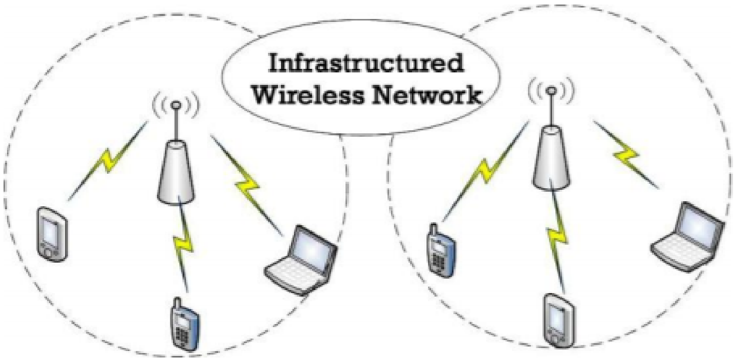


Figure 2.1: Infrastructured Wireless Network

2.3 Mobile Ad Hoc Network

To sum up, a MANET is a group of independent devices (which can move) connected wirelessly. Lately, there has been an amazing evolution in technology, and this makes MANET even more useful since this results in an increase of nodes, which increases the probability of a message to be received by the target node. Despite this being a great invention, there are some issues that are still to be solved, the more urgent one being issues regarding safety, since nodes can join and leave networks freely, this could be a problem.

Moreover, there is a list of characteristic which all MANETs have in common:

- Messages hop
- Self-organization
- Mobility
- Scalability
- Connected wirelessly

As we also said before, MANETs are increasing in importance due to the simultaneous development of greater communication devices and computing power, and the interconnection of these two too. The situations in which those networks are more useful are the following:

- Vehicles network (bikes, buses, cars, etc)
- Emergency response networks
- Remote areas networks

Now we will focus on the issues of MANETs, which we hope that with the fast evolution of technology will soon be solved. Those are mainly the following:

- Network scalability
- Network overhead

- Dynamic topology
- Power management
- Security
- Efficient routing
- Stable routing
- Quality of service

On the other hand, the most notorious advantages of MANET over other communication technologies are: the system can be operated virtually everywhere (low requirements to work), scalability thanks to a better performance with an increasing amount of nodes (which is usually a weakness in other communication technologies).

Now we discuss the main disadvantages of MANETs, not because of technical problems, but because of the very definition of how MANETs are designed. These are:

- Establish a proper security protocol is a hard task to carry out
- If malicious nodes happen to appear, those are hardly detectable within the network
- It is hard to adapt protocols from wired communication systems to MANETs

2.4 Mobile Ad Hoc Routing Protocols

The design of routing protocols is probably the most critical, hard and core part of MANETs, not because of the complexity itself, but because research has to be done to find the best protocols for different situations. Since nodes do not possess any knowledge of how the topology of the network in advance, indeed, it emerges from the micro behavior of each of the nodes. Routing protocols, which dictate how messages should be sent, can be classified into three different categories, regarding the basis of routing information update mechanism.

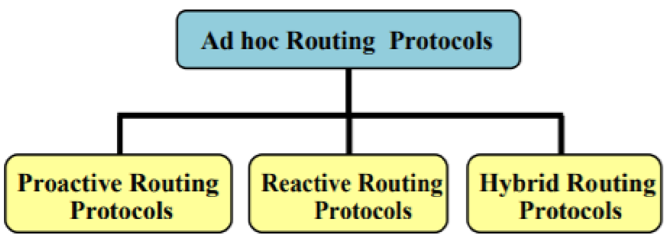


Figure 2.2: Output file given by PedSim

The first one, proactive routing, maintain constantly routes to all nodes in the routing table, even to the ones to which no messages are sent. Those routing protocols are updated dynamically as the topology of the network changes. Therefore, packages (messages) are sent accordingly to the predefined routes (this information is located in the routing table). The main advantage of those protocols is the speed at which messages are sent but the lower hand is the overhead ratio.

The next category is reactive protocols, which are also called on-demand routing protocols. Here, routes are only set under demand, and nodes do not start the routes by themselves (only if they are asked to do so). The main differences with the previous category are:

- Lower overhead
- No maintenance routes without traffic
- No update route information constantly

Lastly, the category of hybrid routing protocols is a mix of the two other categories mentioned above. This protocol benefits from the advantages of the those two, but has a new disadvantage, and that is: the nodes with more topological information has more route information, and therefore, the energy and memory consumption is higher.

There are many more ways in which protocols can be divided regarding different characteristics such as: the amount of information of the message, applications, and others. We can read in this paper [1] how the authors decided to classify the different algorithms depending on the information with which each of them operates. In [2] and [3] we can see other classifications, one of them depending on the existence or not of infrastructure. Finally, we can read an actualized version of the different routing techniques in [4] and [5].

We will now divide the different protocols depending on how many messages are sent forward.

1. Single copy: Single copy protocols are routing techniques that keep only one message "alive" until it reaches the goal node. Since the amount of messages in the network remains constant, and at 1, the network can run smoothly, unluckily, the probability that the destination node receives the message is clearly low, and when it happens it usually takes a large amount of time. The most common ones are:
   - Direct Delivery [6]: This is the simplest way in which we can define a protocol. The node sends the message, if and only if, the node to whom the message is going to be sent, is the destination node. Therefore we can conclude that the message will arrive at its destination only if the seed node and the objective node contact each other.
   - First Contact [7]: When the node with the message contacts another node, the message will be always sent to the node with the one it contacted (assuming the contact time is long enough for the message to be successfully sent). This process will only be stopped when the message arrives at its final destination.
2. Flooding: Flooding protocols follow r-strategy, meaning that they send, depending on the protocol, as many copies as the message as possible, or a fixed

amount of them, hoping that one of the many copies created will arrive the destination. The limitation here is the finite amount of battery and the finite amount of memory, since there is a maximum number of messages that a device can keep.

- Epidemic [8]: This routing technique creates and sends a copy to each of the contacting nodes, so when we are in a crowded area, the number of nodes that have received the message increases exponentially over time (until it usually reaches a maximum, where almost all of the nodes are carrying it). The downside of this protocol is that more often than not the network suffers supporting that many messages. Moreover band with and the capacity of the devices to store messages are two characteristics that constrain this method to be more used. Lately, there has been some modifications to it to benefit from its strong points trying to minimize its weak points. [9,10]
- Spray and Wait [11]: This is a somehow controlled flooding algorithm, there is a fixed amount of copies of the message that can simultaneously exist at each given time. The name of the protocols makes references to how it works, first, the copies are sent to the messages, and then we wait until one of the copies is received by the destination node. Overhead is easily controlled by n (the number of copies of the message).

3. Probabilistic: In this group of algorithms messages are sent to the nodes with a higher probability to contact the destination node.

- PRoPHET [12]: This has the honor to be the first protocol to ever be designed for MANETs. The core of the algorithm relies on the art of identifying which node probably contacts the destination node. If we somehow can know that in advance, therefore we can reduce the amount of copies needed to successfully send the message. Each of the nodes carries the information of which nodes it contacted before, so a simple read in the past contacts of the nodes allows us to make predictions.
- PRoPHET-V2 [13]: This is an upgrade of the above-mentioned protocol. The upgrade relies on the contact time, which improves the predictions.
- MaxProp [14]: Sends a message to the network. If the message arrives at the destination node, another message is sent to allow the nodes with the stored message (which has successfully been sent) to delete it.

4. Other protocols: In this section, messages are selected regarding its sociability.

- BubbleRap [15]: This protocol has its origin and is focused on social network usage, two metrics are defined (centrality and community). The message is sent to a node depending on those two metrics.
- The Friendship and Selfish Forwarding Algorithm [16]: Some metrics are defined and then the message spreading is done regarding those metrics.
- Randomized Rumor Spreading [17]: Floods the network in a random way.

There are plenty of people trying to improve MANETs in multiple ways. In [18] we can read about a novel broadcast-based content delivery system. The goal is to reduce the amount of time needed to deliver a message, for this, the authors first set the optimal size of the message. The conclusions are important since analytical and simulation tools are used to guess the correct size, also it is useful in many scenarios, since the simulations were carried out both in static and mobile devices, and those devices being at different distances away from the transmitter.

The authors at [19] worked in a core problem in our matter at hand, and that is routing when path failure takes place. As we have clearly shown above, all of the protocols do not take into account failures in the delivery system, which is not how it works, when we have real scenarios, we can have multiple problems, being the most notorious ones the followings:

- Buffer overruns
- Errors in the path selection
- Link failures
- Unscheduled delays
- Others

So in those scenarios, the authors propose to develop new strategies, such as splitting, replicating and deleting some of the messages when those failures take place. They also take into account two scenarios, when the message is completely sent or completely lost. And when only a part of the message is sent, this can happen when the time contact is not long enough for the message to be sent, which happen with big enough messages. We will later show how size message affects the success of our protocol.

As we mentioned before, some of the uses of MANETs are in extremely performance challenged environments, where only contact each other from time to time. We can find examples of these scenarios in large areas with low node density, such as wild areas. Also, this effect is similar in areas where nodes are only "active" part of the time, which also makes contact between different nodes a rare thing to happen. Some kind of intelligent flooding is developed in [20], where energy needs to be used wisely, it achieves robustness and reliability.

2.5 MANETs Evaluation

We will now explain how MANETs are evaluated, regarding to which characteristic we focus on, we will we paying attention to some variables.

Firstly, and introduction to the most commons opportunistic networks models will be done, those are the following:

- Analytical models
- Simulation models

The core parts of the methodology that we will use to study those:

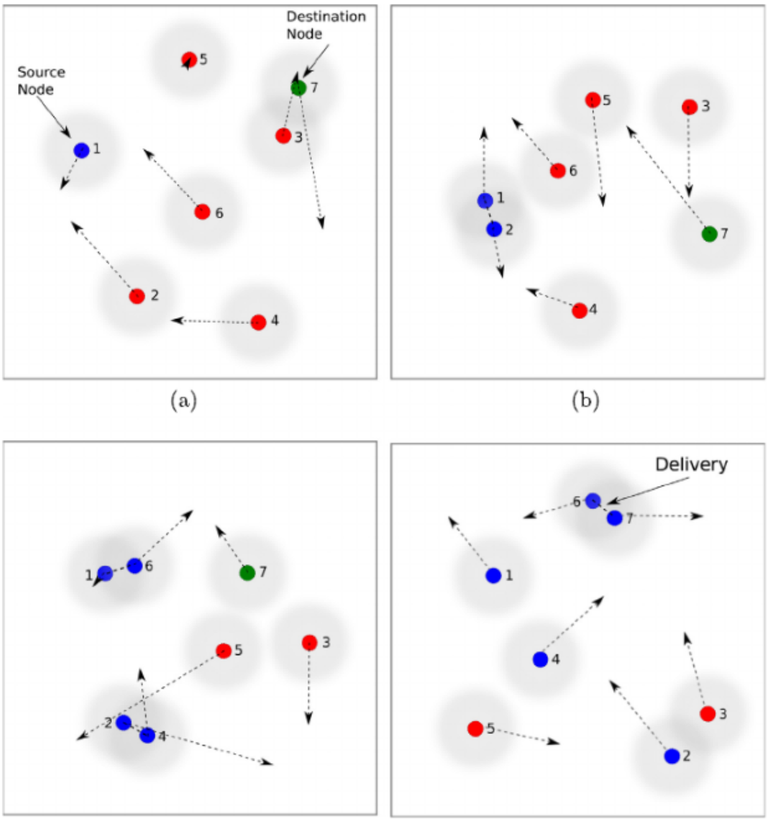
- Opportunistic network
  - Buffer
  - Overhead
  - Transmission range
  - Others
- Works charge: This depends mainly on the protocols used (flooding protocols make the system work hard) and the number of nodes and contacts.
- Metrics
  - Number of contacts per unit of time
  - Probability for a message to be successfully sent
  - Average contact time
  - Delivery time
  - Others

The simulation models use a simplified model of the system, with a lower work charge, which is simulated with software, the one that we will use in our work is named ONE. Usually, these software need to be fed with realistic traces, and due to the complexity of being able to obtain completely real traces, opportunistic network simulation software are used along with traces software. The software to obtain trace here will be PedSim, although those are powerful tools, we should take into account the computation capacity needed to carry out those calculations since some scenarios complexities might exceed the capacity of our computers to carry out the simulation.

Analytical models are mathematical tools that represent our system and our workload. Mathematical models allow us to surpass some of the perks of the simulations, since the workload is usually lower, but on the other hand, the simulation might be less close to reality, this is due to the fact that initial assumptions have to be done, and those can sometimes have some intrinsic errors, which may affect or not the conclusions. There are two main mathematical models used in opportunistic networks: epidemiological and markovian models.

Regarding epidemiological analytical models, there are many papers published about this topic. In [21] the authors proposed a SIR model (susceptible, infected, recovered), and proposed new ideas to reduce deviations on diffusion time depending on different parameters.

In [22] the model is done regarding different social aspects such as: contact time, message size, node density or the amount of people going in and out. The authors try and achieve to obtain analytical expression out of differential equations (which more often than not are not able to be solved and must proceed with numerical methods). Again the mathematical tool used are SIR equations, also used in biological populations, the paper focuses on the equilibrium points and the analytical solution to the equations. The paper also successfully proves that this analysis, commonly used in biology, can also be used in opportunistic networks, and it gives us an insight of the dynamics of the processes.



(a) (b) (c) (d)

Figure 2.3: Message spread evolution over time

In addition to the biological process modeling, some researchers have used markovian chain to model opportunistic networks. The authors in [23] have calculated an approximation to the expected delay of a message to arrived to the destination node. The markovian space represents the distance from the message (the carrier with it) to the destination. Although as they mention, they do not know whether this delay distribution is exponential or not. The researchers at [24] also use a markovian approach to this problem, but this time they use a continuous one.

2.6 Mobility models

When evaluation opportunistic networks and the different protocols applied, we should also take into consideration how the traces used are obtained. There is a trade-off between exactitude and work needed, this means that some traces are easy to obtain but they do not represent the reality as good as traces that need more computation or work to be obtained. The easier to get are the analytical ones where you just input some data into software and after some calculations you obtain them. The two main ones here are markov chain and deterministic models.

We also have simulations, with two different categories, pure simulations and simulations with real mobility traces, in the latter one the amount of real data is bigger than in the former one. Since the mobility creates opportunities for contacts to happen, our study of mobility is crucial. The downside of the more exact methods

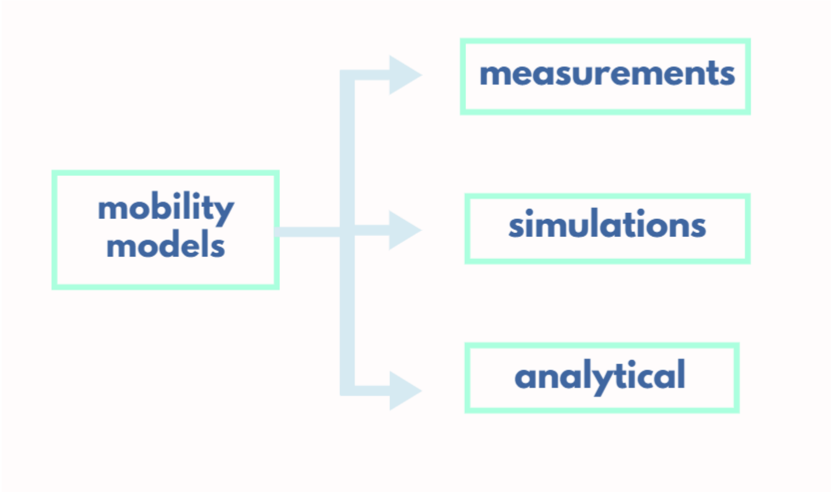


Figure 2.4: Different types of mobility models

is that they are less general, meaning that markovian chains can easily be used in different scenarios varying just some parameters, while the more real ones, we should measure again in the following scenario to get traces as good as the ones we got before. Anyways, even when we face those problems, the experiments carried out by some researchers prove that those methods, while having their downsides, are still useful and a great tool for us. For this reason, we will now introduce the different techniques commonly used, and those are the following:

⟶ **Random mobility models:** Those models mainly use stochastic movement models to move nodes over a given area, here there is a disadvantage of trying to find random numbers, which to out date, are impossible to find, the closer we have are pseudo-random ones. Due to its simplicity, the model more commonly used among all of the random ones is the so-called RWP (Random Way Point). A node moves in a random direction with a random speed. Some more sophisticated models make those "random" directions and speed depend on the previous values. Lately, with new technologies, those models are infrequently used, since randomness (to a certain degree) is not a characteristic of how humans move since we have destinations to which we want to arrive when we displace. So what some researchers are trying to apply is applying the randomness to which places each node goes, therefore defining a clear path.

⟶ **Real traces:** There are some mobility traces which have been obtained with a device that marks the position of the node for each time step, those are the most valuable ones, since they represent closely how real movements happen. Those traces are considered to be higher in quality with a higher number of nodes and/or a longer time duration. There are also contact traces, which tell us when contacts happen, the duration of them and the location where they took place.

One of the most famous traces is [25], and there have been multiple traces studies, one of them can be found in [26].

The biggest benefit of contact traces is its simplicity, experiments can be carried out in this case, but the downside is that some information is lost. On the other hand, location traces need a higher workload but they have the upper hand when it comes to information quality. There is a famous repository [27] called Crawdad, where we can find most of all these traces. So if I were to use some already obtained traces, I would look at this page first.

⟶ **Hybrid mobility models:** In this section, we find models that are a mix of the above mentioned with some additional variables. Some of the most important models used are the following ones, which combine some information on each of the nodes with software simulation. The Lévy Walk in [28] closely describes how some animals move in their habitat or SWIM presented in [29], small world in motion where just some parameters are needed to be set, and the traces have some variables equal to the real ones, SWIM is mainly used to predict forwarding protocols. Another approximation is the one explained in [30], it is important to stress that mobility affects routing protocols performance in MANETs, so the authors focus on the macro-mobility data collected from mobile devices, and they try to benefit from it. They say that can to somehow predict mobile device carriers thanks to previous observation of those users. As they prove on the paper, mobile users present an orbital pattern visiting some places way more often than others.

⟶ **Pedestrian based mobility models:** Mobility models were designed to deal with the complexity of human displacements patterns. Ranging from basic models such as RW or Random Way Point (RWP) in [31], to more complex and realistic ones. Some of those more complex models account for common daily tasks in peoples life such as jobs. Those models were named Working Day and were for example studied in [32]. Also SWIM [33] or SLAW [34]. Altogether with these simulation examples, it is wise to have real traces at hand in case they are needed for a more realistic simulation using other software [35,36].

Chapter 3

Mathematical model: Differential equations

3.1 Introduction

We will try to use mathematics to better understand how these dynamics evolve. A recent survey has shown how important analytical modeling and performance evaluation is to better understand Opportunistic Networks. As we mentioned before, combining real traces with some sort of simulations is one way to evaluate its performance, but sometimes due to lack of resources, we will be forced to use a much less time-consuming approach. Analytical methods offer us a much faster and general tool to be applied to different situations. The two main approaches proposed have been markovian models and models using Ordinary Differential Equations (ODEs).

The one we will use here is the Population Processes approach, in which the goal is to describe the dynamics of population processes. Since then, this approach has been used in at least two situations that if not the same, were similar enough to be mentioned in the present work. In [37] this approach was used in a network in which data carriers were whales, and in [38] it was used to clarify which protocols had better performances. In [39], the authors proposed a SIR model, this is a model used in Epidemic protocols and human epidemic diseases (Susceptible, Infectious and Recovered).

When modeling a real-life situation, we should aim to stay as close to the real situation as possible, and we found that those models miss some important aspects such as: message size, memory capability, battery duration, density of people and many more. Therefore, another type of equations is proposed, Delay Differential Equations (DDEs), which take into account past values of the variable at hand.

After clarifying which tools will be used, we also have to clarify how is the situations that we will try to describe. The contact-based messaging application that we study is short-range communication (with a range of 8 meters), those messages are stored in the different mobile devices that receive it, therefore the information is decentralized over the network.

When two nodes establish a connection, they exchange the messages, each of the nodes storage and forwards messages. Unluckily, as we showed before, depending on the message size, contact time duration and other variables, the message will not be forwarded successfully. For a contact to be defined as effective contact, the contact

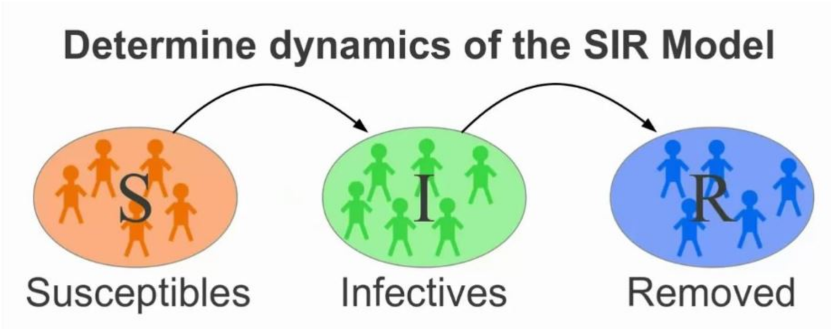


Figure 3.1: Simple dynamics graphically explained

time should allow the message to be completely transferred. If p is the probability of successful transmission is p, and a contact rate of $\overline{\lambda}$, the probability of effective contact rate is

$$\lambda = p\overline{\lambda}$$

Since the contact duration is such an important variable, two situations can take place depending on what we decide to do regarding it:

$\longrightarrow$ When two carriers need to exchange messages, they stop until the transmission is completed, which sometimes it is not likely to happen, thus reducing the effective contact rate (some messages will not be correctly forwarded)

$\longrightarrow$ And on the other hand, we have the situation in which mobile devices carriers do not stop, so message size, band with and contact duration time become extremely important variables, and the message will or will not be forwarded depending on those variables. User speed, communication range also are of high importance.

3.2 Performance

We should not forget that we are aiming for a method to analyze the performance of our network (physical devices, protocols and so on). We should note that people can enter and leave the area freely. In our model, it is comparable to an epidemic spread, in which a seed node is infected, and depending on some probabilities, duration, and location, it spreads the virus to other nodes (in our case, message).

The next assumption, unluckily, does not perfectly describe our scenario, but we think it is close to it. Our entrance and exit rate is not constant, but to obtain analytical solutions we should suppose that they are constant (otherwise we should simulate the dynamics with finite methods). We call β the arrival rate, and γ the exit rate. Therefore, the population in our given area depends directly on the initial number of nodes (N_0) and the entrance and exit rates. We will study three different cases in which differential equations will be applied to obtain interesting results:

BASIC EPIDEMIC MODE: We can see an example of this model in [38], in the most basic example, $N(t)$ remains constant, this is the number of nodes will remain the same. In this case we suppose that the transmission is made instantly (which can be assumed with low transmission times). We name I the infected nodes and S the susceptible ones. The only possible thing to happen is for susceptible nodes to become infected. The equations that describe the situation are the following ones

(ODEs):

$$S'(t) = -\lambda S(t)I(t)$$
$$I'(t) = \lambda S(t)I(t)$$

For an analytical node infected (I(0)=1) the number of infected nodes over time is:

$$I(t) = \frac{N}{1 + (N-1)e^{-\lambda N t}}$$

EPIDEMIC MODEL IN OPEN AREA: We now would like to extend our previous analysis to a more general situation, this is when nodes can enter and leave freely. This is a more realistic situation, since most of real-life situations are described by this model, for example, restaurants, barbershops or malls. We name β the number of nodes entering the area per second (equivalent to the birth rate in the biological process) and γ the number of nodes leaving the area per second (the death rate). Each of the new nodes entering the square are susceptible and nodes leaving it can be susceptible or infected. Here non-infected nodes can die (since a susceptible node can leave the square). Summing up, we have by the previous definitions the following equations:

$$N(t) = N_0 + (\beta - \delta)t$$
$$N(t) = I(t) + N(t)$$

The evolution over time can be derived from the following equations:

$$S'(t) = -\lambda S(t)I(t) + \beta - \delta S(t)N(t)$$
$$I'(t) = \lambda S(t)I(t) - \delta I(t)/N(t)$$
$$N'(t) = \beta - \delta$$

We would like to study the system further at equilibrium, this happens when I(t), N(t) and S(t) are constants (their derivatives are 0). Since $N'(t) = \beta - \gamma$, in this case $\beta = \gamma$. In this case, we can obtain the following analytical solution (with I(0)=1):

$$I(t) = \frac{be^{bt}}{\lambda(e^{bt} - 1) + b}$$

Where b is:

$$b = \frac{\lambda N_0^2 - \beta}{N_0}$$

MODEL CONSIDERING TRANSMISSION TIME: We started analyzing the least similar situation with a real scenario and we are getting closer, each step at a time, to the most real one. We will consider th transmission time of the messages, $T_c = T_s + T_t$. The left side is the time needed for the message to be forwarded, the first time on the right side is the time needed for two nodes to set up a connection, and the later is the time needed for the message to be completely sent ($T_t = m/Bw$) (m=message size and Bw=bandwidth). We have to take into account that when a node is establishing a connection with another node, it cannot establish a connection with another one, and our model should reflect it. Due to that effect, we include another type of node, the communication nodes (C), and a subdivision of it can also be made, we have the nodes that are forwarding the message (R) and the ones that are receiving it (P). R and P are not allowed to leave the place why they are in this state, they will move to the I class whenever they finish the process. No analytical solution

3.3 Open place with fixed nodes model

Many models describe real-life situations. In a recent paper, the authors describe the following scenario. In the area, nodes can move in the direction that they choose, and they have a contact rate of $\lambda > 0$, and we will also use a similar notation as the one used before. New nodes arrive with a rate of β, those nodes entering the scenario are susceptible nodes (S), meaning that they do not have the message yet. And nodes leave the place at a rate γ, in biological processes those to parameters would be the birth rate and the death rate. Therefore, the number of nodes at a given time, this is $N(t)$, depends on the initial number of nodes $N(0)$ and the two rates.

To make it more realistic, we have to introduce a subtle change that makes it a bit more complex than the basic biological process and this is that susceptible nodes can "die", meaning that they can leave the area, in biological processes, nodes can only disappear if they were previously infected, since here we want to take into account nodes that leave the area, the best way to do so is counting them as dead. We then have the two first equations of our system, which are equal to the ones obtained in the previous section:

$$N(t) = N_0 + (\beta - \gamma)t$$

$$N(t) = I(t) + S(t)$$

And $N(t) > 0$ for every t, since N(t) represents the number of nodes in the area at any given time, it makes no sense to make it negative, so any analytical solutions which gives us a negative value should be modified to represent our real system. As the author mention, the following diagram shows the flows that make the different populations vary.

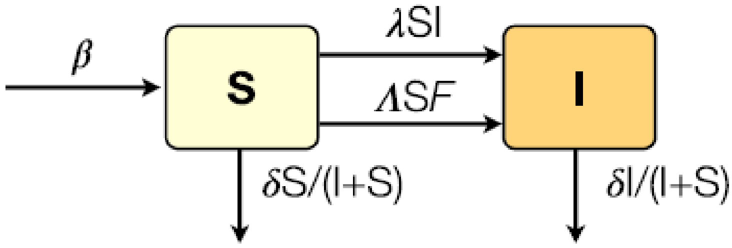


Figure 3.2: Diagram explaining the different flows

So up until now, there is no difference between this scenario and the basic model first explained before. We now introduce the concept of fixed nodes. First of all, those fixed nodes have a contact rate of δ (between fixed and mobile nodes). Fixed nodes act as some sort of infected nodes, this means that when a mobile nodes (a susceptible one) and a fixed node establish contact, the mobile node gets infected. For this reason, we assume that only susceptible nodes can contact the fixed ones.

The following transitions are possible in this simulation:

Regarding to all the different transitions described above, the 3 differential equations that describe the evolution of the system are the following:

$$S'(t) = -\lambda S(t)I(t) + \beta - \delta S(t)/N(t) - \Lambda S(t)F$$

$$I'(t) = \lambda S(t)I(t) - \delta I(t)/N(t) + \Lambda S(t)F$$

$$N'(t) = \beta - \delta$$

This system of equations is in some manner similar to the previous obtained before, with the difference in the F and δ.

3.4 Dynamics of the fixed nodes model

After carefully describing the fixed node model, and writing down the equations that dictate the evolution over time of itself, we will now go further into that evolution. First of all, we have to set some variables to make it possible to get real values from the equations. The next plots all start with the same number of nodes at the beginning of the simulation ($N_0 = 100$) and $F = 2$, we have no nodes with the message at $t = 0$, which means that $I(0) = 0$. We set also the contact probabilities to 0.001. We now distinguish the three most interesting scenarios.

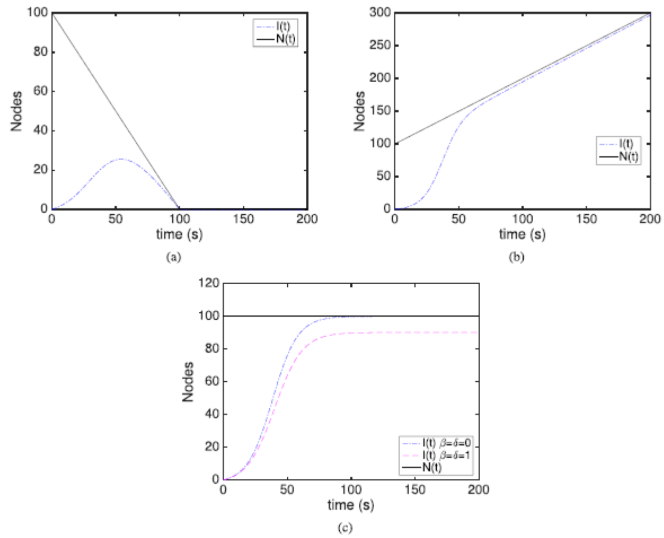


Figure 3.3: Representation of the evolution of the different groups over time for different parameters (a) $\beta < \delta$, (b) $\beta > \delta$, and $\beta = \delta$.

The previous simulations were carried out with Euler's method, since there is no (as far as I know) analytical solutions to it. The first plot (a), is the evolution of the groups when there is no entry rate and exit rate equal to 1. As it is obvious the population keeps going down until it reaches a minimum of $N(100) = 0$, since 1 person leaves the place at each second, and there are 100 people at the beginning, this result makes sense.

The second plot shows the evolution for an entry rate of 1 and an exit rate of 0. The number of total people will increase as long as we keep running the simulation, since the number of people entering exceeds that of people leaving. It is also interesting to see how the infected (I) population behaves, it increases at a varying rate until it gets closer and closer the total node number, this makes sense, since the crowdier it gets (with infected nodes) the easier it is for a new node to get infected, this will happen up until each entering node will be almost instantly infected.

The last two simulations were carried out for the same value of β and γ, and those values were 0 and 1. We reached equilibrium at different values, which makes sense since in the case where the values are different to zero, there will always be some susceptible nodes in the area. For the last case, we have an analytical solution to the system of differential equation, and it is:

$$I_p(t) = \frac{b - \sqrt{b^2 - 4\lambda\Lambda N_0 F}}{2\lambda}$$

3.5 Waiting time

If we had to choose two metrics to evaluate the performance of our method, we would probably choose the diffusion time and the infected nodes. But as we are trying to improve the feeling from the users point of view, we should try to understand which variables he values the most, and I think one of those is the waiting time, which is the time that needs the susceptible node entering the area to get the message. We assume here that the number of nodes with the message is constant and equal to I_e, also, we assume that a mobile device carrier does not leave the message until the information has been forwarded to him. So before leaving, the susceptible node should contact either an infected node (through the probability λ) or a fixed one (through the probability Δ). From a markovian perspective, this can be seen as a situation in which we have to states, the state in which the S node has not been contacted yet either by an infected node or by a fixed one ($k = 0$), or in the state in which he has been and then the state is $k = 1$. We also have that the birth rate is constant:

$$\lambda_b = \lambda I_e + \Delta F$$

And we deduce that the Poisson process describes the metric through the following equation:

$$T_w = E[T_1] = \frac{1}{\lambda_b} = \frac{1}{\lambda I_e + \Delta F}$$

3.6 How different values of β and δ affect our system

Some of the assumptions that we previously made are hardly possible to happen in real life. For example, the fact that we assume that the entry and exit rate were constant, do not make it closer to reality, in fact, if we observe closely to any square, mall or restaurant, we will soon realize that this is far from reality, people usually have dinner at 7 pm, goes to church to a specific moment and so on. For this reason, we will now focus on how the variability of those two rates affects our work at hand. How this will be accomplished is by entering some noise in those two variables. This is one of the most common tools used to evaluate the dynamics of some diffusion processes in which there is some random variability, which is then transmitted to the variables that depend on the first noisy ones.

Since the entry and exit rates vary constantly, no equilibrium can take place. The equations written down before transform in the following ones:

$$S'_{\epsilon}(t) = -\lambda S_{\epsilon}(t)I_{\epsilon}(t) + \epsilon_{\beta}(t) -_{\gamma}(t)S_{\epsilon}(t)/N_{\epsilon}(t) - \Lambda S_{\epsilon}(t)F$$

$$I'_{\epsilon}(t) = \lambda S_{\epsilon}(t)I_{\epsilon}(t) - \epsilon_{\gamma}I_{\epsilon}(t)/N_{\epsilon}(t) + \Lambda S_{\epsilon}(t)F$$

$$N'_{\epsilon}(t) = \epsilon_{\beta}(t) - \epsilon_{\gamma}(t)$$

Even though we can not evaluate it in the equilibrium state (to which never arrives), we can do something similar done in noisy systems, which consists of evaluating the quasi-stationary equilibrium. We set the variables for the noisy distribution and we then solve the equations using the Eulers method.

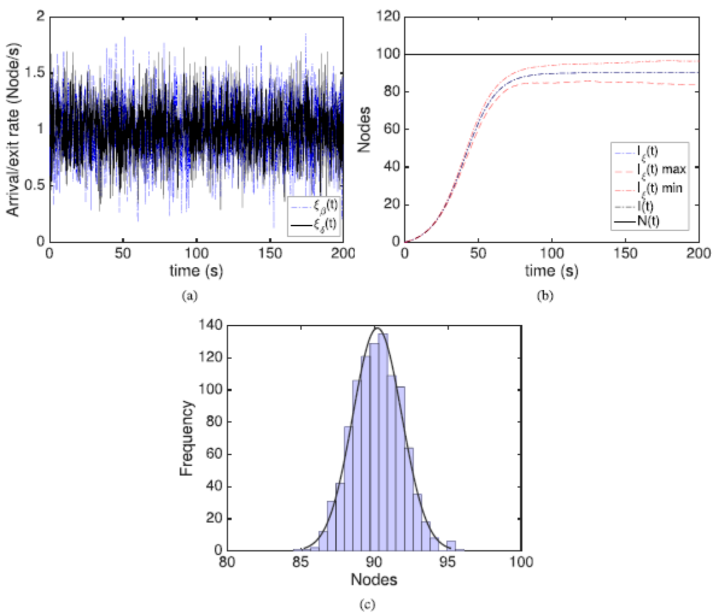


Figure 3.4: How noise affects the inputs and outputs

In the image (a) we can see a sample of the arrival and/or exit rate, it follows a normal distribution with an average of 1 and a standard deviation of 0.25^2. On the plot shown in the (b) part, we can clearly see how the noisy variables affect the solution. And finally, the (c) image shows how the I_ϵ in $t = 200$ follows a normal distribution.

Chapter 4

Scenarios

4.1 Introduction

In this chapter we will explain where the measurements were done will take place which variables we measured to obtain our traces later on. Since this is a crucial point in our project, we will detail precisely the characteristic of the square.

After careful observation and measurements, we were able to get enough data to create the correct digital scenario, and enough information about the pedestrian going through it.

There are many ways in which PedSim can be used. Open and closed areas, crowded or not, etc. In this case, we worked in an open scenario where people could enter and leave freely.

For example, if we were to simulate how human moves in a club, we would have an almost closed space without people entering or going out. In this case, the square "San Vicente Ferrer", we have many people going in and out.

4.2 Location

As we mentioned before, our area is the square "San Vicente Ferrer", it is located in Valencias downtown. Near the town hall square, "La plaza de la Reina" and the River, so it is usually as crowded as it can get.

Also, in the square itself, we have one government building, one big church, and some other smaller businesses. We decided not to measure in mass time since it only happens twice a day and it would have been an unrealistic representation of how the square usually is.

We have five roads going into our studied area of almost equal width. Although pedestrians are not allowed to walk into those roads, but since this is Spain, this will not be a matter of concern.

Figure 4.1: San Vicente Ferrer Square

4.3 Measurements

We will divide this section into two parts, the measurements related to the square itself, and those related to the pedestrians. Since the ones related to the square are way more simple, we will start with those first.

The scenario is quite simple. We have five roads going into the square. The widest ones are "Carrer de les comèdies" and "Carrer del mar", and those will usually be the most crowded ones, since they also connect with two important places in Valencia, those are: "Carrer de la pau" and "La plaza de la Virgen". Moreover, we have a small fountain in the middle of the square, with two benches siding it, and some trees surrounding the fountain. Besides that, even tho we have part of the surface for pedestrians and other part for cars, pedestrians walk as if it was allowed to go in every place of the square since most of the people here are tourist and they are usually mainly focused on the church.

Figure 4.2: San Vicente Ferrer Square 2D

We will now indicate the data collected regarding the pedestrians. I first stayed a couple of days looking at how people behave, common paths, number of people per group, speed of pedestrian, etc. Later on, I decided that the best place in which I could take the measures is in the bottom right corner, since I could see clearly from where to where people displaced.

I started writing down in a notebook, for each group of people walking through the square: number of people in the group, at what time they went in, at what time they left, and from where to where they moved. With this information, and with some help from Google Maps, the speed can easily be calculated.

Figure 4.3: Common path used to calculate the speed

4.4 PedSim

This is finally how our scenario looks like in PedSim. We draw the fountain as a square since we are not allowed to design obstacles shaped as circles in PedSim.

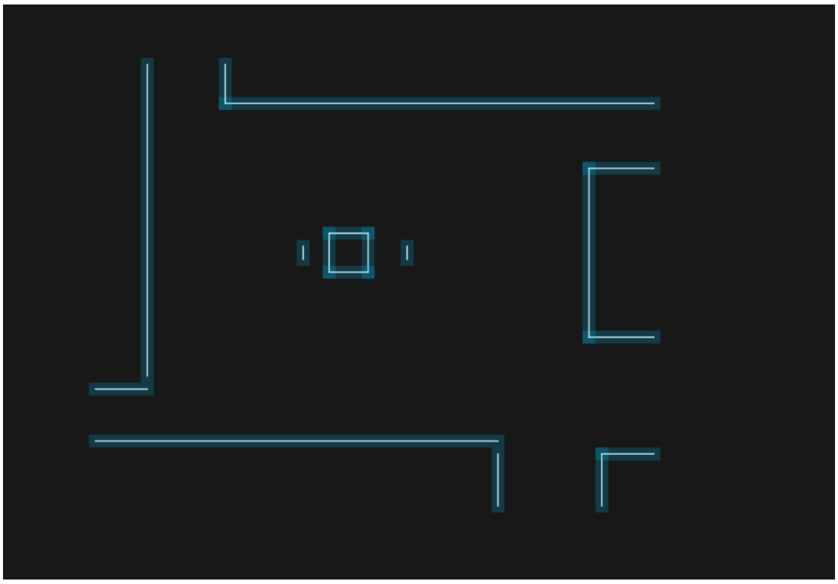


Figure 4.4: Final results

Chapter 5

PedSim

5.1 State of the art

PedSim is a tool to simulate how pedestrians move. It can be used in indoor evacuation simulations, fire evacuation protocols or many other scenarios. Since it is completely based on C++ it can be used in every operating system.

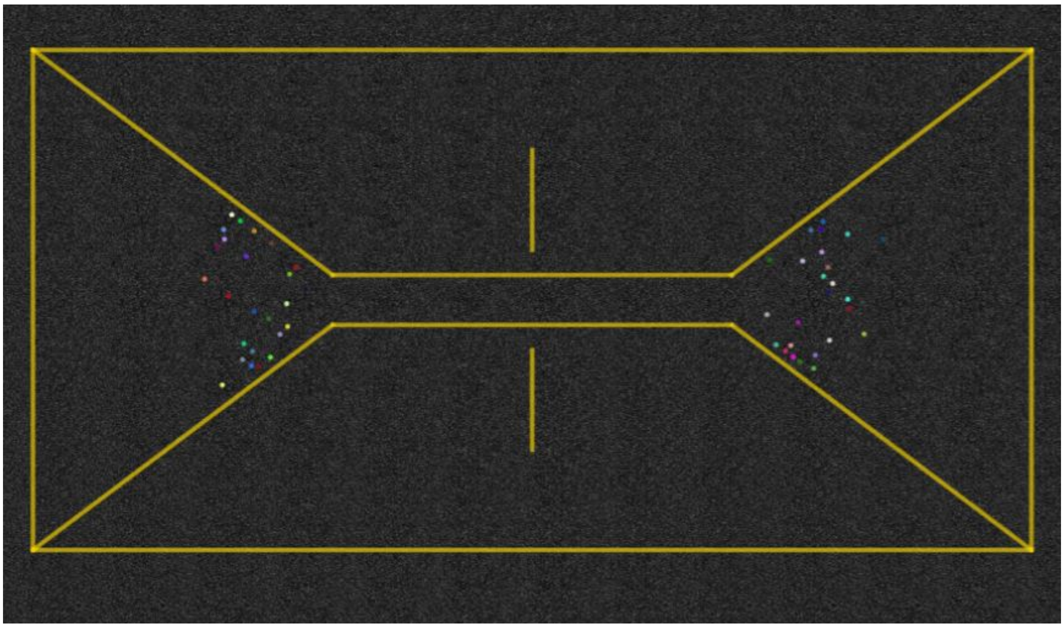


Figure 5.1: Example of a scenario built in PedSim

There are multiple ways to simulate this kind of situations. In PedSim, the particles are individual, making individual decisions. PedSim uses a coupled differential equation to calculate the movement of each of the nodes. More specifically, the one used is the one showed in the paper social force mode by Helbing et al

$$m_i \frac{d\mathbf{v}_i}{dt} = m_i \frac{\mathbf{v}_i^0 - \mathbf{v_i}}{\tau_i} + \sum_{j \neq i} \mathbf{f}_{ij} + \sum_{W} \mathbf{f}_{iW} \tag{5.1}$$

Here m_i is the mass of the $i_t h$ pedestrian, v its velocity and $\mathbf{v_i^0}$ its desired velocity. τ is a time constant. The first term on the right side of the equation models the force with which the pedestrian tries to get closer to the desired velocity, the bigger the difference, the bigger the force that tries to restore it to the desired value. The second term models the interaction of different pedestrians, hence we do not count it for $i = j$, since there is no repulsion of a pedestrian with itself. And lastly, the third term models the interaction of the pedestrian with the environment.

There are many variables which can be changed, they can be used to model how strong pedestrian should repel each other, or how they interact with the obstacles. There are many others that makes this tool highly valuable: pedestrian speed, obstacle force, etc.

From PedSim we can obtain both the traces (a set of numbers) and the video of the simulation, which is of great use to check whether the simulation is right or wrong. The output file given by PedSim gives us for each time step, the position of each of the individual nodes, which is itself highly valuable, and that can later be used to test many different things, in our case, the spread of messages in MANETs.

In our simulation, PedSim uses the above-mentioned differential equation, nevertheless, as it was said by the creator itself, later versions might use another algorithm to make those calculations.Because that the equation is an equation used in physics, here, pedestrians will show behavior close to subatomic particles behavior. This is, short-range interactions make pedestrians that are close enough repel each other, and long-range interactions make pedestrians separated by a distance big enough attract each other. This attraction to pedestrians far away should be replaced by a more complex function which better describes how humans behave.

Thanks to the third term in the equation, we can luckily model the interaction of pedestrians with objects. Whether those objects, obstacles or walls are outdoors or inside of a building, the behaviour displayed depends on how the environment is designed and the value that we give to the variable.

Any movement simulator should take into account not only the physical part of the problem, this is how forces act and how pedestrians move accordingly to them but somehow model the way pedestrians choose a path, this is, the psychological part. If we can model this close enough to reality, we should be able to completely cancel the second term of the equation, this was the social interaction force one. There are many ways this can be implemented, a creative one is the look ahead one, where the pedestrian looks ahead and counts how many other pedestrians are placed on the right of its path and how many on the left, then he chooses the path with less pedestrians. More innovative ways to model the interaction with obstacles are being researched until those are developed, we will keep using our differential equation.

5.2 Inputs

Now that we have enough information to run a simulation and get the traces with PedSim, we need to set the different free variables. Since the measures in "San Vicente Ferrer" square took us 1 hour, then the simulation time will last for 3600 seconds. After multiple simulations, we set the obstacle force, and pedestrian repulsion force in the value that we think looks most realistic, since low values makes the pedestrians get way too close to obstacles to look real, and too high values affect the simulation in the other way around.

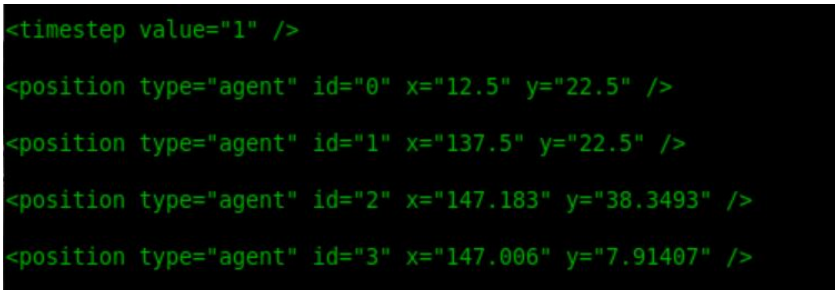
5.3 Code

After deciding which inputs should we feed PedSim with, we are now ready to write down the code. To accomplish this task, we should write a code which reads the data from a .csv file. We think this is highly valuable since other people could use our code easily.

We first read the .csv and store the important variables in a matrix in c. Then, we do a loop that for each time step does the following, it checks the matrix to check whether in this time step a group of pedestrian should be created or not, if so, then it checks how many people should be created together, at which speed they move, and were they appear and where they head to. For the next time step, PedSim moves all the nodes accordingly to the differential equation above mentioned and checks again the matrix. The simulation finishes when the counter gets to 3600.

5.4 Results

We can now visualize the simulation and also get the traces. This is how the file given by PedSim looks like.



```
<timestep value="1" />
<position type="agent" id="0" x="12.5" y="22.5" />
<position type="agent" id="1" x="137.5" y="22.5" />
<position type="agent" id="2" x="147.183" y="38.3493" />
<position type="agent" id="3" x="147.006" y="7.91407" />
```

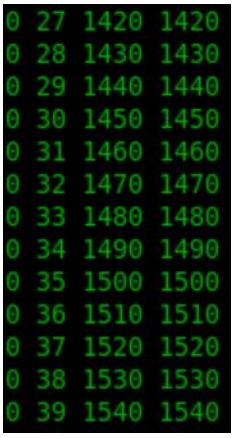
Figure 5.2: Output file given by PedSim

5.5 From PedSim to ONE

After successfully accomplishing one of our main goals, this is, obtaining traces alike to real ones, we have to do some changes to this file before advancing to the next step, which will involve the simulator ONE.

We have used different scripts and altogether with Matlab we modified this file to make it possible for ONE to read it. This is how it finally looks like, now we are in a good position to start using ONE.

We have 4 columns which their meaning are the following: time, node ID, x position and y position.



```
0 27 1420 1420
0 28 1430 1430
0 29 1440 1440
0 30 1450 1450
0 31 1460 1460
0 32 1470 1470
0 33 1480 1480
0 34 1490 1490
0 35 1500 1500
0 36 1510 1510
0 37 1520 1520
0 38 1530 1530
0 39 1540 1540
```

Figure 5.3: File fed to ONE

Chapter 6

ONE

6.1 State of the art

As we mentioned before, there are some situations in which traditional networking does not work. So there is an increasing demand to develop a new tool to deal with those scenarios, one of the proposed solutions is OpNet.

This approach has completely different outcomes depending on the dynamics of the nodes, therefore, the need for a appropriate simulator is obvious. This need is filled by the simulator called ONE (Opportunistic Networking Environment)

This tool allows us to use real-world traces to apply different types of protocols. In concordance with PedSim, we obtain at the end of the simulation process visualization of the mobility and connectivity of the nodes, and a set of matrices containing different types of information.

ONE is a discrete agent simulation engine, due to the complexity of the matter at hands, no analytical solutions can be found, rather, we work with numerical solutions which are "renovated" at each step calculation. Therefore, nodes movements are predetermined, but connectivity is calculated out of other multiple variables, such as location or communication range. Furthermore, We are also free to define how messages will be created.

Nodes in this simulator represent message carriers, those can be seen as cars, bicycles or any other device with the sufficient technology before mentioned. Those have different traits represented but variables, some of the most important are: movement, energy consumption or storage capability.

(photo) There are two main ways to define nodes movement:

- Data load from an external source

- Mobility calculated by algorithms

Lastly, similarly to PedSim, we can visualize the results and we can also obtain a file with a different set of information.

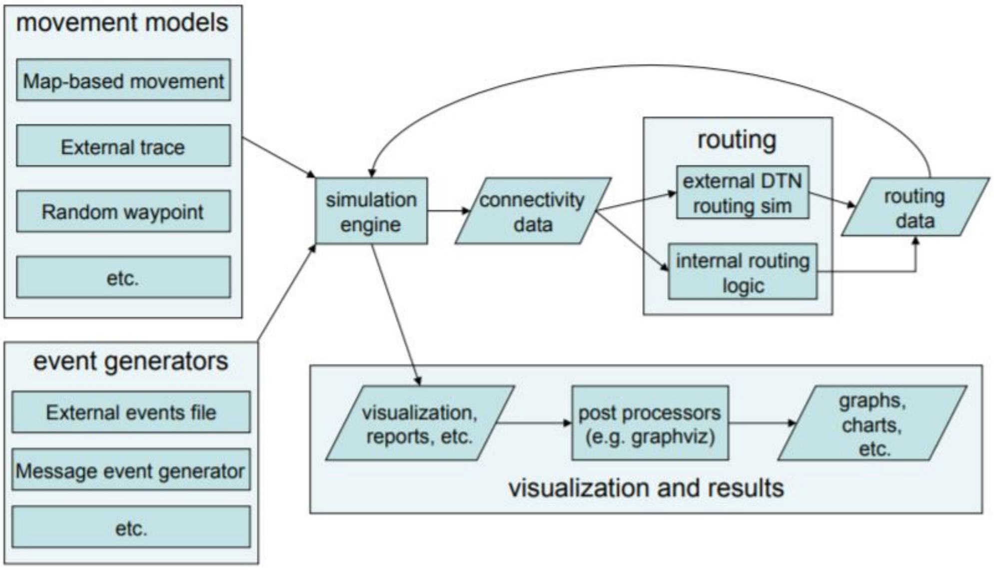


Figure 6.1: Scheme of how ONE works

6.2 Introduction

In this chapter, we will analyze our traces with ONE.

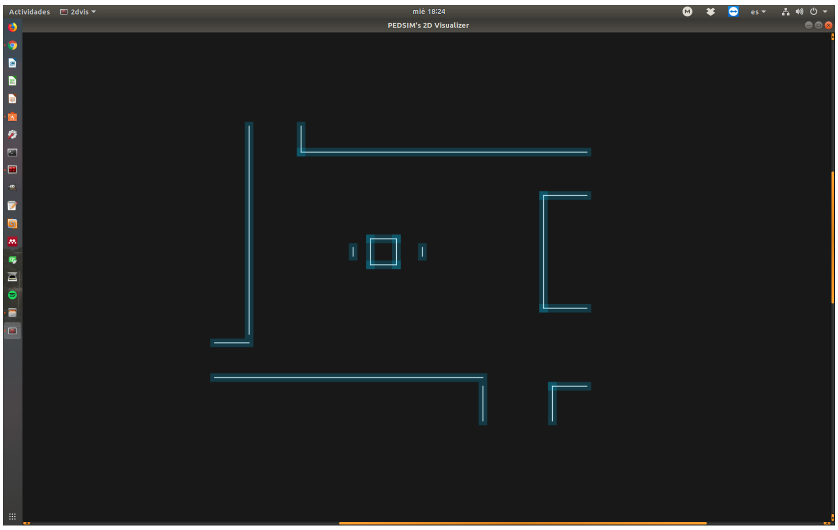


Figure 6.2: Our sceneario

We will simulate different cases. The protocol used here is the epidemic one. It is important to state clearly which routing protocol we will use in this simulation. Those protocols are algorithms that tell us how the information will be shared among the different nodes, there is not a better option, the power of each of the protocols shines in different situations.

There are many protocols with big difference among them, to better understand how they related to each other, it is useful to classify them. There are many ways in which this classification can take place, attending to which of the traits of the protocols are analysed. The most common differentiation consist on grouping them depending on whether they create a replica of the message or not. If the protocol never replicates the message, then it is a forwarding-based protocol. And otherwise, they are called replication-based.

The epidemic routing protocol is a replication-based one. It has its foundation in nature, and it is somehow similar to how some diseases spread. Nodes constantly replicate the message and send it to nodes which do not still have the message.

Lastly, we will simulate 10 times each one of the different scenarios. Each scenario is differentiated from the previous one by the size of the message. We have done the simulations with the following sizes: 1k, 2M, 2M, 4M, and 6M.

6.3 Metrics

Before we go into detail with the different metrics calculated with the output files, we need to set some parameters to feed ONE, those are the following ones, and we also state which protocol we used, in our study we used the epidemic protocol.

Information	
Parameter	Value
Pedestrians speed range	0.24-2.87 m/s
Tx range radio (m)	8
Tx speed (Mbps)	2
Dispositive memory	1GB
Simulation time	3600s

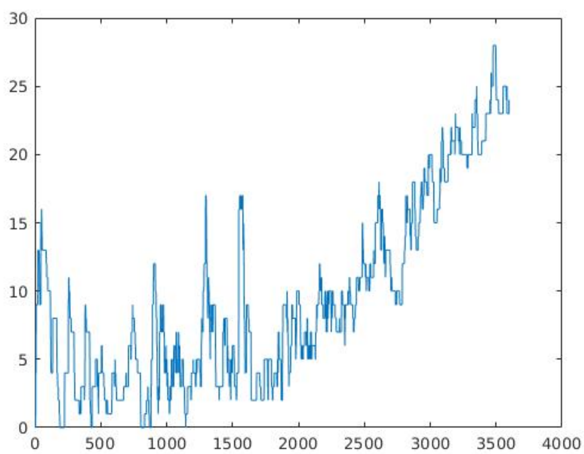


Figure 6.3: S

In this image, we can see the evolution of the number of people inside of the square. I started doing the measurements at 4 pm, so the shape of the graphic is understandable since that Spanish people usually take a nap for a short period after having lunch, and then we start moving to other places after we wake up. This is seen at the end of the graphic, by 5 pm most people have woken up and started doing some other tasks or hanging out with friends.

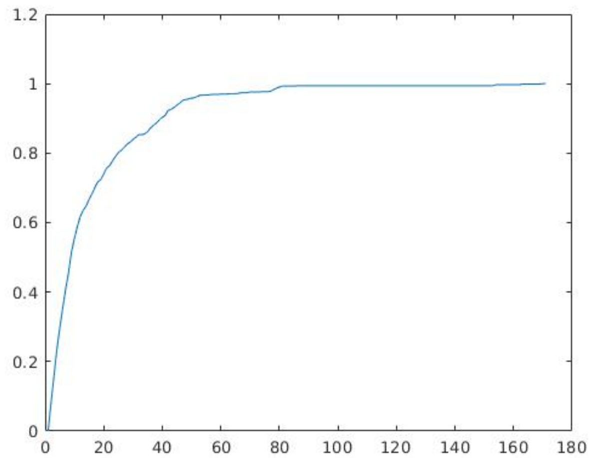


Figure 6.4: S

In this image is plotted the cumulative distribution function for the contact duration. As the visualizations showed me, most of the contacts usually last for a small amount of time and then people stop being near enough to be counted as a contact. Later on, this will prove to be a problem since the bigger the message the more time it needs to be sent.

Another incredibly useful metric to measure how good or bad the performance of our protocol is the delivery ratio. Which is the quotient of the nodes which has received the message over the total number of nodes.

Delivery ratio	
Message size	Ratio
1K	50.8
1M	38.12
2M	18.79
4M	4.14
6M	0.83

From this table, we can observe that the delivery ratio decreases with the message size. This is an expected thing to happen since the bigger the message is, the more time is needed for the message to be sent. This effect will appear again in the overhead ratio, which is the quotient of the messages successfully sent over the number of messages which were started to be sent.

Overhead ratio	
Message size	Ratio
1K	1
1M	77.97
2M	47.89
4M	16.13
6M	3.7

The last important metric that we calculated is the time that passes between the entrance of the node in the square and the time in which the message is received. This calculation has been done for the simulation with a size message of 2M, and the duration is 26.41 seconds.

Chapter 7

Networks science

After using some differential equation models, we will move into further modelings, the next paragraphs will be used as a bridge between the SIR modeling and the networks/graph modeling. We can generalize the previous models to describe more realistic situations. This can be done by adding more nodes groups and allowing more transitions between them. Since by doing that the equations grow on complexity, the analysis will primarily be done numerically. We will review models that directly or indirectly come from graph theory.

7.1 Differential equations models

Non-Markovian epidemics, the methods used previously mainly are based upon some Poisson assumptions. Those models assume that, as we said before, the rates at which the message is forwarded is constant. Due to that, the probability for a node to stay infected or susceptible decays exponentially over time. The fact that those rates are held constant, makes this a Markovian process, this is a process in which each of the nodes have no history (it does not affect its present). This assumption makes the process easier to be dealt with mathematically, but we should only use it when we can somehow know, or at least have a good approximation, of the average of the rates.

The SIRS model, this model has been previously studied by Pastor-Satorras and Bancal, applied to complex networks. It has been studied both analytically and numerically, the later one was studied on a small-world network with discrete-time, in which infected nodes remained like that for a fixed amount of time.

The SEIR model, this model was primarily designed to describe the spread of flu and other respiratory sicknesses, it has been used in network theory by Small and Tse, they studied it with a deterministic approach as well as with a stochastic one. Nodes that transition from susceptible to infected, developed an immunity to the infection that made it impossible for them to be infected again.

7.2 Traits of our network

Realistic networks, the previous examples were all designed to work under some strong assumptions, such as the random undirected and uncorrelated behaviour of nodes, as everyone knows this is far from describing a real scenario. We will need to use a list of metrics to better describe the singularities of each of the networks. Authors in the network theory field of study usually assume that the probability that a vertex of degree k with another vertex follows a given equation (which we will not show here since it does not fall under our scope), but this is indeed more complex and makes the dynamics of the process have a different behaviour. One of the most important results of the SIS model is that the critical number or density of susceptible hosts required for an epidemic to occur is related to the largest eigenvalue of a given matrix.

Effects of clustering, how cluster affects the message/infection spread has been of great interest, this is because that social networks, which are also one of the primary factors deciding how the infection will be transmitted, are highly clustered. This is an empirical experience which I bet everyone knows, you usually meet almost the same people most of the days. Some authors have tried to analyse how a variable that measures the "amount" of clustering affects the transmission of the infection, and they all concluded that for highly clustered networks, even when the transmissibility was low, most of the people will eventually be infected.

7.3 Different tools used in network science to make it more powerful

Weighted networks, in our study to make the modelling more and more real, we now focus on how different contacts happen, everyone can relate that even if you meet your boyfriend each day and you also meet one of you coworkers almost daily, the type of contact that you have with each of them is different, and by different we do not only mean by the amount of time. This can be better expressed in network theory by giving different weights to different nodes connections. SIS models are generalized giving making the infection rate some function of the weight of the edge that connects both nodes. We should also note the results found by Eames, Read, and Edmunds, they took empirical data and analysed the effect of weight distribution and infection risk, they found it to be an important parameter when determining which nodes where riskier to get infected. More results were found by Britton, Deijfen, and Liljeros, they observed that the outbreak probability and the epidemic threshold were highly affected by the values of the weights. **Directed networks**, some infections need this type of approach. Those epidemic transmission are common blood-related or fluids related. Node positioning and node connectivity is highly important when it comes to how many other nodes this one can infect.

Bipartite networks, those networks are extremely usually to understand the spreading of sexually transmitted diseases, since there are mainly two types of nodes (males and females), and this disease is commonly only transmitted between individuals of different groups.

Epidemics in adaptive networks, the previous authors all studied static networks, which means that connectivity among nodes (which nodes are connected with which nodes) did not change over time. This assumption is far from reality, and as we go further into detail we are getting closer and closer to real-life situations, the next paragraphs explain how dynamical networks affect the analytical solutions of epidemic diseases. For human diseases, the structure of the connections varies with the evolution over time of the studied disease. For example, ill people will change their habits dramatically in a short amount of time when they notice that they are sick, and therefore the connectivity with other people will change as fast as their habits. Here we can also differentiate how the behaviour changes regarding how the information about the infection is communicated. If the node is the one that decides locally by itself to change its habits, will have some precise effects, and on the other hand, if it is for example the government the one who communicates that some people is infected and that some measures will be carried out, then the connectivity will change according to it. In our present text, we want to stress the importance of the evolution of the connectivity while the disease spreads, and how the spread is affected by this new connectivity, those two processes feed one each other making the new dynamics usually more complex than before. Some authors have also included strategies implemented by susceptible nodes to avoid being infected, such as defensive strategies (for example staying at home or buying some drugs). There is a very narrow area in which oscillatory dynamics take place. There have been some studies which compared the numerical solutions with the analytical ones and they made sense, so this proves that we are heading in the right direction. Now we will see the power of this modelling, in 2010 Saw and Schwartz proved that the amount of time needed to be effectively vaccinated in adaptive networks was way different than in static ones, so this is a highly valuable application of network theory. We can define in this environment one metric called effective transmissibility T, which decreases with the rewiring probability, under some values of the rewiring probability, the society behaves well enough to choke the spread of the disease. It is an interesting task to develop policies to make it happen.

If we want to go one step further, we should consider how people behave, in the past scenarios when one node disconnected one of its edges, it was immediately connected to other node, which is a behaviour far from real. There have been some strategies to better describe our real world, such as creating and destroying links at different rates. Or establishing again connectivity with past nodes whenever the infected node heals, this strategy successfully represents human behaviour since people usually stop meeting acquaintances when they get sick but they go back to their old habits as fast as they get healthy.

Competing pathogens, this type of process explains how different epidemic processes (infections for example) compete with each other, this is indeed what happens in nature and in our society nowadays (unluckily). The main characteristic in this model is what is called cross-immunity. Imagine that you are infected by flu, after you heal from it, you are for sure immune to flu, but you could potentially be immune to other infections. In 2005 Newman studied the properties of a static network in which total cross-immunity was a characteristic of the nodes. First was released the first disease, and after that one, it was released the second one, that as since nodes had cross-immunity, the second pathogen could only be forwarded to survivors of the first infection.

Epidemic processes in temporal networks, we study the spread of the disease in networks in which if we compare the speed at which the networks changes, is extremely slower than the speed at which the infections spreads, therefore the networks looks almost frozen (with the exemption of some of the adaptive networks with the ones we dealt before). There is also the DBMF theory which models the spread process in a way that nodes rewire its connection much faster than the speed of the spread of the disease, this is also not common in daily life, since the behaviour is extremely personal and varies from person to person, also those networks models have nodes that are usually simultaneously connected to multiple persons, which is also far from reality.

Temporal networks, those networks are defined as a contact sequence, in which for a given time t we have a combination of edges. If we were to take a picture of each of the edges for each time, and then add them up together, we would get a static network projection. Therefore an edge connecting the i and j nodes is represented in the static network if there was a time $t_{ij} < T$ in which there was a contact between the nodes i and j. There are more ways to evaluate the importance of different contact in temporal networks, for example, we could give a weight to an edge depending on how many times those nodes contacted each other during the observation time. Unfortunately, neither of those models do not account for the dynamics of the system and a great amount of different processes with different pattern could end up being represented by the same static network.

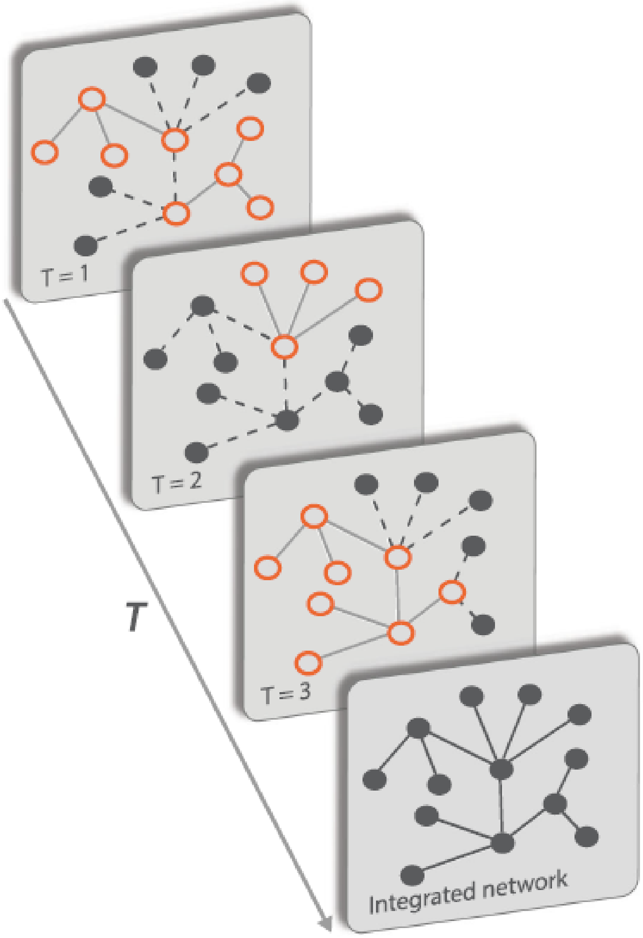


Figure 7.1: Our scenario

Lately, the development of new technologies allowed us to obtain huge amounts of information on social temporal networks such as social network, we are now able to analyze this data and obtain metrics and heterogeneous contact patterns. Recent discoveries have cataloged the real human interactions in heavy-tail statistical distributions, which is completely different by the predictions done by the Poissonian approach, this difference is primarily in the unreal assumptions done by this approach, that humans move randomly. Another significant trait of human behaviour that affects the spread of the infection is how connectivity patterns vary, and the effects that they have. The most interesting effect of the time networks is represented in the projection image shown above. Even when in the static network two nodes appear connected, the dynamics of the spread of the infection are heavily determined by the order in which those connections were "turned on" and "of". From those reasonings, we can derive the fact that in non-Poissonian temporal networks the spread of the disease is way slower. There has been empirical data to support this idea, and we got a confirmation of the analysis that told us that the spread is slower in this case. In another try to better model the epidemical processes, an activity-driven network class of models was developed. This model was based on the idea of activity potential, this metric is defined as the probability that a node engages in a social activity per unit of time. As everyone expects, this value is hardly generalised to every person so it is personal and is highly volatile between different people.

Chapter 8

Contact Network Analysis with Gephi

8.1 Introduction

Gephi is a tool mainly used for scientist that want to analyze data, it also gives an insight into graphs. Similar to other software, it gives to user the ability to interact with and manipulate the structures in order to reveal hidden patterns. It can also be used as a complementary tool to common statistics analysis. This is because it has been proved that visual thinking with interactive interfaces helps reasoning.

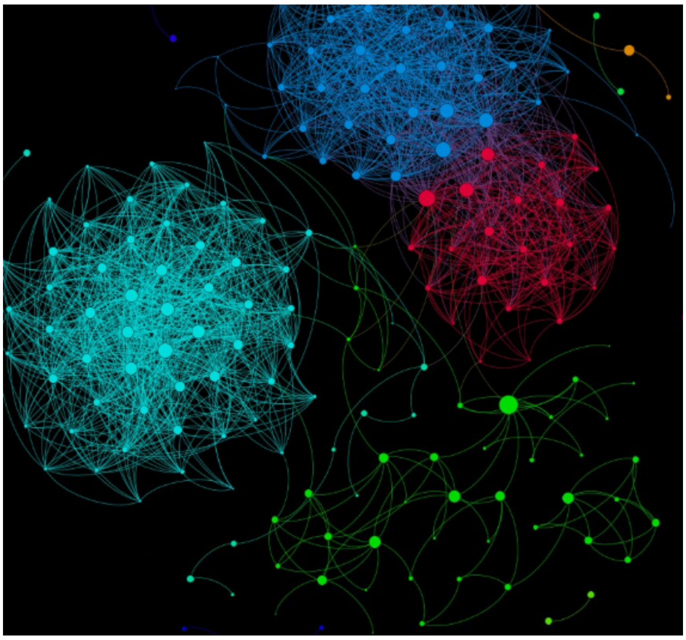


Figure 8.1: Facebook Network visualized with Gephi

Up to 100.000 nodes and 1.000.000 edges can be visulized at the same time. Moreover, state-of-the-art algorithms are used to provide a smoth experience to the user.

8.2 Important metrics

Some important metrics that we can obtain with Gephi are the following ones:

⟶ **Connected component:** this is a metric from graph theory, it is a set of nodes in a graph taht are connected to each other by paths. In nature we usually have one huge connected component and other smaller ones. For example this happens with sexual relationships.

⟶ **PageRank:** with this metric we can reveal which nodes are the most important ones. Important nodes are not only those connected to many nodes, but also nodes connected to important ones. This metric has been widely used in multiple fields, but it was first used to rank webpages.

⟶ **Average page length:** this metric is the average number of steps along the shortest paths for all possible pairs of network nodes. This value is important because it gives us information about how a processes can be spread through the network.

8.3 Results

We modified a bit one of the files given by ONE in order to use it as an input into Gephi. We then obtained the following graph: This figure shows the behavior

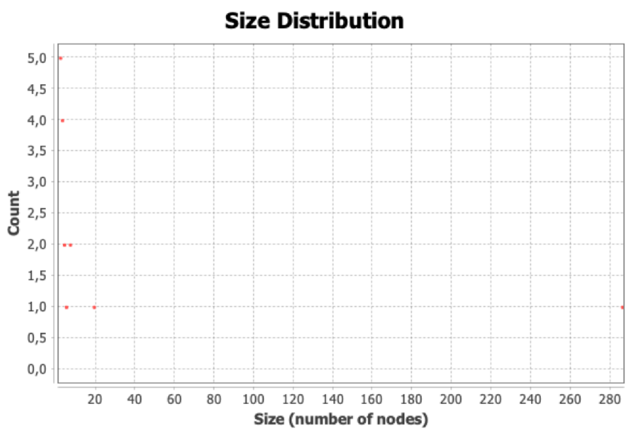


Figure 8.2: Number of groups vs number of members

and the dynamics of our system. As it commonly occurs in realistic cases, the vast majority of our nodes were connected between each other, and there were some other minor groups. We can conclude that if we want the messages to be forwarded to every single node we will need some structure that can do that for us, since mobile devices alone do not achieve this goal (as we can see, it would be perfect if we had all the nodes connected in the same group). For example, we have one group of 280 nodes, and 5 groups of one node.

Moreover, there were 3 nodes that had way more connections than any of the other nodes, this is because some nodes stay longer in the square, take a rest in one of the benches or they just walk slower. This is perfect for us since these kind of people make it possible for our message to spread to more people.

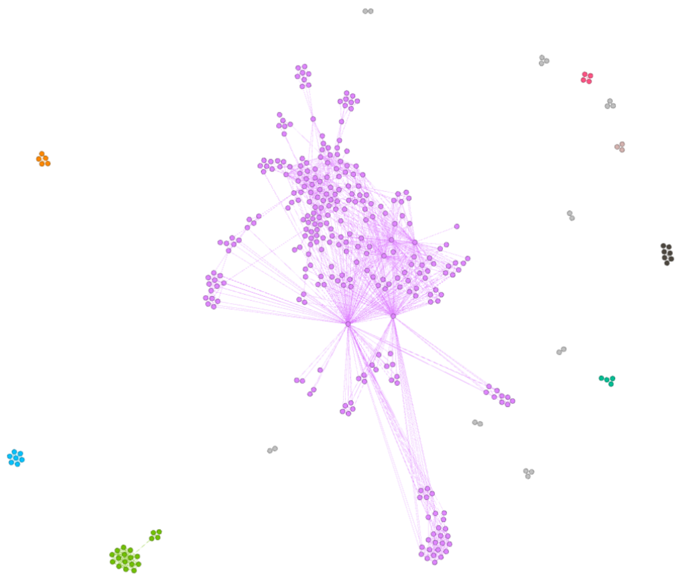


Figure 8.3: Visualization of the groups

This figure shows how our nodes were connected ones to each others. We coloured each of the connected groups with one colour. For example, a message jumping from a node to a node that was connected to the previous one in the pink group will never be able to get to other colour group. Each of the nodes represents a person, and each of the edges represents a contact between the two nodes.

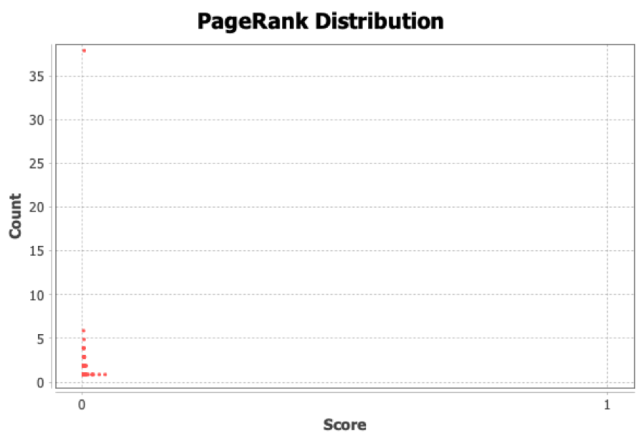


Figure 8.4: Number of nodes for a given value of PageRank

This figure shows how many nodes we have for a given value of a PageRank. Since our network has less than 400 nodes, we should not expect high value of the PageRank variable. We can see in the picture how we have 3 nodes apart from the rest, those were the nodes that we mentioned before that were walking slowly or stayed in the square longer than any other for reasons that we do not know.

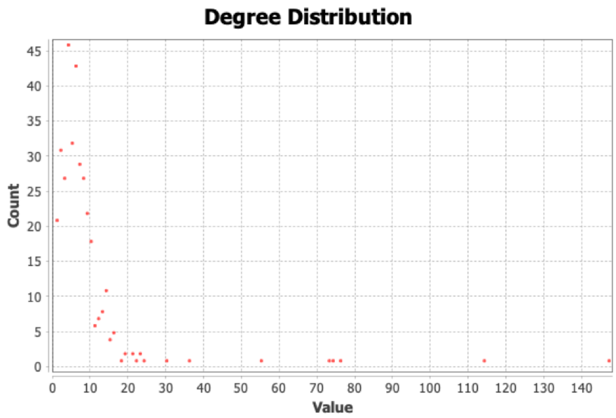


Figure 8.5: Number of nodes for different degrees

Lastly, we have the degree distribution. In this plot we can see that our network is similar to a scale-free network. A scale-free network is a network whose degree distribution follows a power law. Moreover, we have one node with a degree of 144, another with a degree of 113 and 3 nodes with its degree valued between 70 and 80.

Chapter 9

Conclusions

In this work we went a step further than what the author at [40] did since we accounted for the speed and timing of people entering the area, this is since our model read from a .csv file real data we think we were closer to a real simulation. Moreover, this is an interesting field in which there is no doubt that will keep growing in importance as scientist keep thinking about novel methods to implement the technologies and if the other parts that affect our study (such as battery, bandwidth etc) keep evolving as fast as they have been doing we may use this technologies in our daily life soon.

To sum up, the contact time was a decisive variable, especially in the cases where messages where bigger, since the time needed to send the message increases with the size of the message, we found that for this protocol and our scenario and people at study, for a certain message size the successful message delivery rate drop down to neglective levels. One way to bridge this problem is to make the bandwidth bigger or to develop novel protocols. Lastly, we believe that the network theory has much potential in this field and even now that it has not been worked thoroughly it has proved to forecast and explain quite a lot of processes in epidemic spread processes.

Bibliography

[1] S. M. Tornell, C. M. T. Calafate, J.-C. Cano and P. Manzoni. *DTN Protocols for Vehicular Networks: An Application Oriented Overview.*. IEEE Communications Surveys and Tutorials 17.2 (2015), pages. 868-887

[2] L. Pelusi, A. Passarella and M. Conti. *Opportunistic networking: data forwarding in disconnected mobile ad hoc networks.* IEEE communications Magazine 44.11 (2006) (pages. 1, 7, 8).

[3] B Poonguzharselvi and V Vetriselvi. *Survey on routing algorithms in opportunistic networks.* Computer Communication and Informatics (ICCCI), 2013 International Conference on. IEEE. 2013, pages 1-5

[4] N. Chakchouk. *A survey on opportunistic routing in wireless communication networks.* IEEE Communications Surveys Tutorials 17.4 (2015), pages 2214-2241

[5] T. Spyropoulos and A. Picu. *Opportunistic Routing.* Mobile Ad Hoc Networking. John Wiley Sons, Ltd, 2013. Cap. 11, pages. 419-452.

[6] M. Grossglauser and D. Tse. *Mobility increases the capacity of ad-hoc wireless networks.* Mobile Ad Hoc Networking. John Wiley Sons, Ltd, 2013. Cap. 11, pages. 419-452.

[7] T. Spyropoulos and A. Picu. *Opportunistic Routing* INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. Vol. 3. IEEE. 2001, pages. 1360-1369

[8] S. Jain, K. Fall and R. Patra. Routing in a delay tolerant network. Vol. 34. 4. ACM, 2004 (vid. pages. 9, 25).

[9] A. Vahdat and D. Becker *Epidemic routing for partially connected ad hoc networks.* Duke University, 2000 (vid. pages. 10, 25)

[10] J. Niu, J. Guo, Q. Cai, N. Sadeh and S. Guo. and spread: An efficient routing algorithm for opportunistic networking. Wireless Communications and Networking Conference (WCNC), 2011 IEEE. IEEE. 2011, page 498-503

[11] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain and R. Krishnan. *Prioritized epidemic routing for opportunistic networks.* Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking. ACM. 2007, pages 62-66

[12] T. Spyropoulos, K. Psounis and C. S. Raghavendra. *Spray and wait: an efficient routing scheme for intermittently connected mobile networks*. Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking. ACM. 2005, pages 252-259

[13] A. Lindgren, A. Doria and O. Schelén. *Probabilistic routing in intermittently connected networks*. ACM SIGMOBILE mobile computing and communications review 7.3 (2003), pages 19-20

[14] A. Lindgren, E. Davies, S. Grasic and A. Doria. *Probabilistic routing protocol for intermittently connected networks*.

[15] J. Burgess, B. Gallagher, D. D. Jensen and B. N. Levine. *axProp: Routing for Vehicle-Based Disruption-Tolerant Networks.*. Infocom. 2006 (pages 11, 25).

[16] P. Hui, J. Crowcroft and E. Yoneki. *Bubble rap: Social-based forwarding in delay-tolerant networks*. IEEE Transactions on Mobile Computing 10.11 (2011), pages 1576-1589

[17] C. Souza, E. Mota, L. Galvao, P. Manzoni, J. C. Cano and C. T. Calafate. *Fsf: Friendship and selfishness forwarding for delay tolerant networks*. Computers and Communication (ISCC), 2016 IEEE Symposium on. IEEE. 2016, pages 1200-1207

[18] R. Karp, C. Schindelhauer, S. Shenker and B. Vocking. *Randomized rumor spreading*. Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on. IEEE. 2000, pages 565-574

[19] C. T. Calafate, G. Fortino, S. Fritsch, J. Monteiro, J.-C. Cano and P. Manzoni. *An Efficient and Robust Content Delivery Solution for IEEE 802.11P Vehicular Environments*. J. Netw. Comput. Appl. 35.2 (mar. de 2012), pages 753-762

[20] S. Jain, M. Demmer, R. Patra and K. Fall. *Using Redundancy to Cope with Failures in a Delay Tolerant Network*. Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. SIGCOMM '05. New York, NY, USA: ACM, 2005, pages 109-120

[21] J. Widmer and J.-Y. Le Boudec. *Network Coding for Efficient Communication in Extreme Networks*. Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking. WDTN '05. New York, NY, USA: ACM, 2005, pages 284-291

[22] C. S. De Abreu and R. M. Salles. *Modeling message diffusion in epidemical DTN*. Ad Hoc Networks 16 (2014), pages 197-209

[23] E. Hernández-Orallo, M. Murillo-Arcila, C. T. Calafate, J. C. Cano, J. A. Conejero and P. Manzoni. *Analytical evaluation of the performance of contactBased messaging applications*. Computer Networks 111 (2016), pages 45-54

[24] T. Spyropoulos, K. Psounis and C. S. Raghavendra. *Efficient routing in intermittently connected mobile networks: The multiple-copy case*. IEEE/ACM Transactions on Networking (ToN) 16.1 (2008), pages 77-90

[25] M. Karaliopoulos. *Assessing the vulnerability of DTN data relaying schemes to node selfishness.* IEEE Communications Letters 13.12 (2009) (page 14).

[26] S. Gaito, E. Pagani and G. P. Rossi. *Fine-grained tracking of human mobility in dense scenarios.* Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2009. SECON Workshops' 09. 6th Annual IEEE Communications Society Conference on. IEEE. 2009, pages 1-3

[27] A. Passarella and M. Conti. *Characterising aggregate inter-contact times in heterogeneous opportunistic networks.* International Conference on Research in Networking. Springer. 2011, pages 301-313

[28] U. of Dartmouth. CRAWDAD *Mobility increases the capacity of ad-hoc wireless networks.* A Community Resource for Archiving Wireless Data (pages 14, 17).

[29] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim and S. Chong. *On the levywalk nature of human mobility.* IEEE/ACM transactions on networking (TON) 19.3 (2011), pages 630-643

[30] A. Mei and J. Stefa. *SWIM: A simple model to generate small mobile worlds.* INFOCOM 2009, IEEE. IEEE. 2009, pages 2106-2113

[31] J. Ghosh, S. J. Philip and C. Qiao. *ociological orbit aware location approximation and routing (SOLAR) in MANET.* Ad Hoc Networks 5.2 (2007), pages 189-209

[32] C. Bettstetter and C. Wagner. *The Spatial Node Distribution of the Random Waypoint Mobility Model.*. WMAN 11 (2002), pages 41-58

[33] F. Ekman, A. Keränen, J. Karvo and J. Ott. *Working day movement model.* Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models - MobilityModels '08 (2008), pages 33-40

[34] K. Lee, S. Hong, S. J. Kim, I. Rhee and S. Chong. *SLAW: A New Mobility* INFOCOM 2009, IEEE. 2009, pages 855-863

[35] T. C. Tsai and H. H. Chan. *NCCU Trace: Social-network-aware mobility trace.* IEEE Communications Magazine 53.10 (2015), pages 144-149. issn: 01636804. doi: 10.1109/MCOM.2015.7295476

[36] F. A. Silva, C. Celes, A. Boukerche, L. B. Ruiz and A. A. Loureiro. *Filling the gaps of vehicular mobility traces.* Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. ACM. 2015, pages 47-54

[37] Z.J. Haas and T. Small. *A new networking model for biological applications of ad hoc sensor networks.* IEEE/ACM Trans. 14 (1) (2006) 27–40 .

[38] X. Zhang , G. Neglia , J. Kurose and D. Towsley. *Performance modeling of epidemic routing* Comput. Netw. 51 (10) (2007) pages 2867–2891.

[39] C.S. De Abreu and R.M. Salles. *Modeling message diffusion in epidemical DTN* Ad Hoc Netw. 16 (2014) pages 197–209 .

[40] Leonardo Chancay García. *Evaluación y modelado de redes oportunistas*