Connectivity in MANETs

Paulien Out

BMI paper
Supervisor: Sandjai Bhulai

August 2007

Vrije Universiteit Amsterdam
Faculty of Sciences
Business Mathematics and Informatics
De Boelelaan 1081
1081 HV Amsterdam

Abstract

In this paper, I study the connectivity in MANETs. First a one-dimensional model is studied, where two points are placed on the ends of a straight line. Then extra nodes are added randomly. Then I derive an expression for the probability that the two endnodes are connected to each other, depending on the number of added nodes and the range of the signal of all nodes. This range is assumed to be the same for all nodes. The expression found is a simple recursive expression.

The second model studied is a two-dimensional model, where nodes are added randomly on the plane. The technique used in the first model turns out not to work, so I use another technique. For the model on the infinite plane I show that there is a critical radius for the range of the signal of the nodes; only if the radius is larger than this critical value, there is a possibility of an infinitely large group of connected nodes. For the model in a finite square, I give the requirements for a certain fraction of the nodes inside the square to be connected. Also I give the rate of growth of the radius of the signal necessary to have full connectivity of all the nodes inside the square as the size of the square goes ot infinity.

I end the paper with a discussion of the results, and some suggestions for further research.

Preface

The BMI paper is one of the last parts of the BMI course. The idea is to let a student research a problem or question, and present the results. This paper presents the results of my research into the connectivity of MANETs.

Because a project like this is never done alone, there are a few people I would like to thank here:

- Lisette and Elizabeth for reading an early version of this paper and giving helpful comments
- My friens and family for their support during the writing of this paper (and the rest of my studies)
- Sandjai Bhulai for all his help and enthusiasm while supervising this paper

Finally, I hope the reader wil enjoy reading this paper and maybe find something useful in it.

Contents

Chapter 1

Introduction

A communication network consists of nodes, which are connected to each other in some way. Some well-known examples of networks are the Local Area Network and the Wide Area Network, and of course the World Wide Web. All these networks need an infrastructure of some sort, so that communication within the network is possible.

Mobile AD-hoc NETworks (MANETs) are wireless networks that differ from these more traditional kinds of networks in that the nodes can communicate directly with each other without the need for any fixed infrastructure. Possible nodes are, for example, laptops or mobile phones, or a combinations of different kinds of devices that can be used for communication. The nodes connect directly to each other and they can move randomly, which means that the network does not have a fixed topology. The nodes must be close enough to each other, because their signals can only be received in a circle around the node with a certain radius. The topology of the network may change rapidly and unpredictably as nodes move, new nodes appear and other nodes disappear, in other words, if someone takes his laptop to another place, turns it on or shuts it down.

The fact that MANETs do not need any fixed infrastructure makes them useful in situations where no such infrastructure is available. Thus, they can be used to communicate in emergency operations after an earthquake has destroyed the infrastructure, or during military expeditions where it is not safe to use the infrastructure of the enemy. Another possible use of a MANET is a network in a conference or meeting, if setting up an infrastructure for just a short time is too costly. It can also be used by police and ambulance services to communicate in the case of a major accident or a terrorist attack.

And what is more, the same techniques as used for MANETs can also be used for completely different situations. Imagine, for example, a number of farms spread out in a certain area. When on one of the farms there is

an outbreak of a contagious desease, then we would of course be interested in the probability that neighbouring farms will also catch the disease. Here the farms may be viewed as nodes, like the communication devices in a MANET. The range of the signal will then be replace by the area in which the disease could spread if another animal came close enough to a sick one.

However useful it may be, the flexibility of MANETs does also raise problems that need to be solved. For example, it is difficult to keep up the security level, the routing in the network has to be done in a different way from traditional networks, and it is difficult to predict whether or not all nodes can actually be connected to each other.

In this paper I will study the connectivity of MANETs. As a starting point, a one-dimensional model is discussed. This model is interesting in itself, but hopefully it can also serve as a starting point for the next step, a two-dimensional model. I will try to determine the probability that every node can communicate with every other node in the network. This probability will, of course, depend on the number of nodes, the distance between these nodes, and the radius of the circle in which their signal can be received. This probability is interesting; in cases where a fully connected network is needed, there needs to be a sufficient number of nodes and a sufficiently wide-ranged signal to make sure that the network has the required probability of connection. In order to do this, an estimate of this probability for a given situation is needed, and of how any changes made will affect the probability. First I will study the one-dimensional case, and then the two-dimensional case. Relevant theory will be discussed where needed. I will try to derive expressions for the probability of connection in each of these two models, depending on the range of the signal and the number of nodes. These expressions can then be used to make estimates of the probability in a real-life situation. So the central question in this paper will be: what is the probability that two nodes can communicate with each other?

Chapter 2

The one-dimensional model

This model is the simplest one of the two models discussed in this paper. It supposes that all nodes are located on a straight line. Therefore it is not very realistic, but it does provide a good starting point for more complicated models such as the two-dimensional model that will be discussed in the next chapter. The structure of the solution to this problem turns out to be very nice and simple.

First, I will give a description of the problem and state the assumptions I make. Then I will pose the question that needs to be answered, and derive a solution.

2.1 Model description

Two nodes are at a distance d from each other, and they both have a signal that can be received within a circle with radius r around the nodes. I assume that this radius is the same for all the nodes. The model looks as in Figure 2.1.

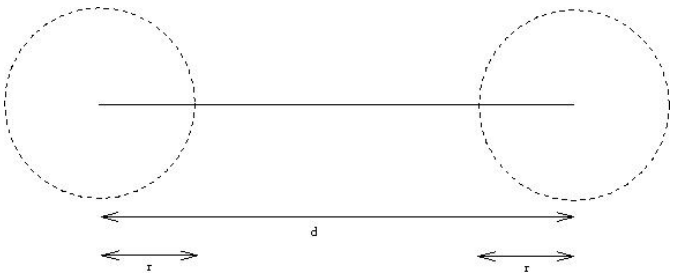

Figure 2.1: 1-dimensional model

In this case, the two nodes cannot make a connection to each other, since their ranges do not have any overlap. Figure 2.2 shows an example where there is a connection.

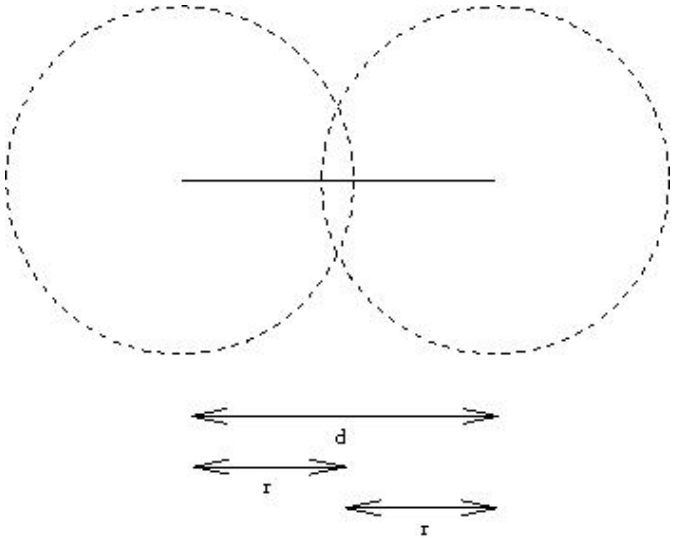

Figure 2.2: 1-dimensional model with connection

This is the starting point for the model; there are two nodes at a certain distance d from each other, and they each have a signal which reaches a circle with radius r around the node. As I said above, I assume that the nodes have the same signal range, i.e., the radius is the same for each node. Also I assume d to be equal to 1, but this does not pose a real restriction: if this is not the case, just divide the r by d and set d to 1. The exact figures do not matter, only the ratio of r to d is important. This is an important observation, and it will be used in the solution later on. The assumption that the radius is the same for all nodes is important here. The results for this model cannot be easily generalised to a case with variable radii.

The question that we address in this paper is: are the two end nodes connected to each other or not? It is easily seen that they are connected if r is larger than $\frac{1}{2}$, so this is not the interesting case. If r is smaller, there is no connection. But what happens if an extra node is added? Or two nodes, or three? These are the important questions.

Extra nodes are added, at a random place on the straight line that runs between the first two nodes. I conceive the line as an interval between 0 and 1, and then pick a random number according to a standard uniform distribution and add a node in that place on the line. The number of these extra nodes is called n, and their places are selected independently according to the same probability distribution. The extra nodes are all assumed to have the same radius r as the original nodes.

2.2 Analysis of the problem

As I said, if $n = 0$ it is known when there is a connection: the nodes are connected if and only if r is larger than $\frac{1}{2}$. But if a node is added, it is not so easy. If r is smaller than or equal to $\frac{1}{4}$ there still can be no connection. The value $\frac{1}{4}$ is the minimal radius needed to cover the distance between the two endnodes with one added node, if this node is placed at its 'ideal' place as seen in Figure 2.3. But if the radius is somewhere between $\frac{1}{4}$ and $\frac{1}{2}$, it will depend on the place of the node. And this place is chosen randomly.

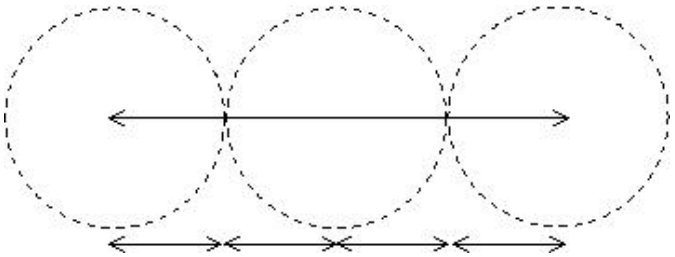

Figure 2.3: With one added node, the mimimum length of r needed for connection is $\frac{1}{4}$.

This leads to a problem. The probability that the two original nodes are connected, will depend on the radius of the signal of the nodes, and on the number of nodes added. In this section, I will derive a formula for this probability. It is denoted by $\mathbb{P}_n(r)$. That probability means the probability that the two end-nodes, at a distance 1 from each other, are connected to each other, assuming that a fixed number of n nodes are added along the line between the end-nodes and seen as a function of the radius r.

The idea used to addres the problem is this: if we take the utmost left of the added nodes, this is the minimum of the n independent stochastic variables with the Uniform$(0, 1)$ distribution. The minimum of n standard uniform stochastic variables has a density function $f(x) = n(1 - x)^{n-1}$. If that node is more than $2r$ away from the left end-node, there is no connection between them, and the two end-nodes are connected with probability 0. If that node is within distance $2r$ of the left end-node, there is a connection between these two nodes.

Now, if we take this minimum as the next left end-node, we can view the remaining problem as the same problem as before, but now with $n-1$ added nodes and distance $1-x$ between the two end-nodes. We can scale this new problem, to have again a case with distance 1 between the end-nodes. To do this we have to divide the distance $1-x$ and the radius r each by $1-x$. The new problem then has again distance 1, but radius $\frac{r}{1-x}$, and $n-1$ added nodes. Thus we get a recursive expression for $\mathbb{P}_n(r)$:

$$\mathbb{P}_n(r) = \int_0^{2r} n(1-x)^{n-1}\mathbb{P}_{n-1}\left(\frac{r}{1-x}\right)dx.$$

Now that the $\mathbb{P}_0(r)$ is known, we can compute the $\mathbb{P}_1(r)$. We know that

$$\mathbb{P}_0(r) = \begin{cases} 1 & \text{if } r \geq \frac{1}{2}, \\ 0 & \text{if } r < \frac{1}{2}, \end{cases}$$

so $\mathbb{P}_1(r)$ will change where $r = \frac{1}{2}$. This means that the integral will have to be split at the point where $\frac{r}{1-x} = \frac{1}{2}$, that is where $x = 1-2r$. Then we get

$$\mathbb{P}_1(r) = \begin{cases} \int_0^{1-2r} 0\,dx, & \text{if } r < \frac{1}{4}, \\ \int_{1-2r}^{2r} \mathbb{P}_0\left(\frac{r}{1-x}\right)\,dx = \int_{1-2r}^{2r} 1\,dx = 4r-1, & \text{if } \frac{1}{4} \leq r < \frac{1}{2}, \\ 1, & \text{if } r \geq \frac{1}{2}. \end{cases}$$

Now that we know $\mathbb{P}_1(r)$, we can proceed to $\mathbb{P}_2(r)$. For $\mathbb{P}_2(r)$ we will have breakpoints at $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{6}$. It is easily seen that $\mathbb{P}_2(r) = 0$ if $r < \frac{1}{6}$, and of course $\mathbb{P}_2(r) = 1$ if $r \geq \frac{1}{2}$. The intervals that remain are $\frac{1}{6} \leq r < \frac{1}{4}$ and $\frac{1}{4} \leq r < \frac{1}{2}$. So we have to compute

$$\mathbb{P}_2(r) = \int_0^{2r} 2(1-x)\mathbb{P}_1\left(\frac{r}{1-x}\right)dx,$$

for $\frac{1}{6} \leq r < \frac{1}{4}$ and $\frac{1}{4} \leq r < \frac{1}{2}$.

Above we have seen that $\mathbb{P}_1(r)$ has breakpoints at $\frac{1}{4}$ and $\frac{1}{2}$. This means that the integral in the expression for $\mathbb{P}_2(r)$ will have to be split where $\frac{r}{1-x} = \frac{1}{4}$ and where $\frac{r}{1-x} = \frac{1}{2}$, that is where $x = 1-4r$ and where $x = 1-2r$.

If $\frac{1}{6} \leq r < \frac{1}{4}$, then $\frac{r}{1-x} \in [\frac{1/6}{1-0}, \frac{1/4}{1/2}) = [\frac{1}{6}, \frac{1}{2})$, so we only need the breakpoint at $\frac{r}{1-x} = \frac{1}{4}$. Then

$$\begin{aligned} \mathbb{P}_2(r) &= \int_0^{1-4r} 2(1-x)\mathbb{P}_1\left(\frac{r}{1-x}\right)dx + \int_{1-4r}^{2r} 2(1-x)\mathbb{P}_1\left(\frac{r}{1-x}\right)dx = \\ &= \int_0^{1-r4} 2(1-x)0dx + \int_{1-4r}^{2r} 2(1-x)(4\frac{r}{10x} - 1)dx = \\ &= 36r^2 - 12r + 1. \end{aligned}$$

Again, if $\frac{1}{4} \leq r < \frac{1}{2}$, then $\frac{r}{1-x} \in [\frac{1/4}{1-0}, \frac{1/2}{0}) = [\frac{1}{4}, \infty)$. That means that we now only need to split at $\frac{r}{1-x} = \frac{1}{2}$. So

$$\begin{aligned} \mathbb{P}_2(r) &= \int_0^{1-2r} 2(1-x)\mathbb{P}_1\left(\frac{r}{1-x}\right)dx + \int_{1-2r}^{2r} 2(1-x)\mathbb{P}_1\left(\frac{r}{1-x}\right)dx = \\ &= \int_0^{1-2r} x(1-x)(4\frac{r}{1-x} - 1)dx + \int_{1-2r}^{2r} 2(1-x)1dx = \\ &= 12r - 12r^2 - 2. \end{aligned}$$

All this taken together means that

$$\mathbb{P}_2(r) = \begin{cases} 1, & \text{if } r \geq \frac{1}{2}, \\ 12r - 12r^2 - 2, & \text{if } \frac{1}{4} \leq r < \frac{1}{2}, \\ 36r^2 - 12r + 1, & \text{if } \frac{1}{6} \leq r < \frac{1}{4}, \\ 0, & \text{if } r < \frac{1}{6}. \end{cases}$$

Now that I have shown the idea, I will go on with the general case $\mathbb{P}_n(r)$ for any $n \in \mathbb{N}$. We have already seen that the number of separate intervals increases with n. For $n = 0$ there were two intervals for r, $[0, \frac{1}{2})$ and $[\frac{1}{2}, \infty)$. For $n = 1$ there are three intervals and for $n = 2$ four. For $n = i$ the number of intervals will be $i + 2$.

The reason for this is that if $n = i$, there will be all the breakpoints there are for $n = i - 1$ because of the recursive expression, plus one extra. This new breakpoint will be at $r = \frac{1}{2+2i}$, because $2 + 2i$ is the minimum length of the radius needed to fill the distance between the two endnodes if all the addes nodes are placed at the 'ideal' place. An example for $n = 3$ is given in Figure 2.4.

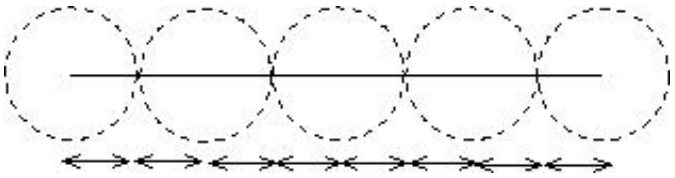

Figure 2.4: With three added nodes, the mimimum length of r needed for connection is $\frac{1}{8}$.

Thus we know that for $n = i$ the breakpoints between intervals will be at $r = \frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \ldots, \frac{1}{2i+2}$.

More intervals mean more calculation, because more integrals have to be computed. The first and last are always simple:
$\mathbb{P}_n(r) = 0$ if $r < \frac{1}{2i+2}$ and $\mathbb{P}_n(r) = 1$ if $r \geq \frac{1}{2}$ for all $n \in \mathbb{N}$.
So only for the second up to the $(n+1)^{th}$ interval an expression needs to be found. For the i^{th} interval we need to compute

$$\mathbb{P}_n(r) = \int_0^{2r} n(1-x)^{n-1} \mathbb{P}_{n-1}\left(\frac{r}{1-x}\right) dx \text{ with } r \in \left[\frac{1}{2i}, \frac{1}{2i-2}\right).$$

If $r \in [\frac{1}{2i}, \frac{1}{2i-2})$, then $\frac{r}{1-x} \in [\frac{1}{2i}, \frac{1}{2i-4})$. This means that the only breakpoint we need from $\mathbb{P}_{n-1}(r)$ is the one at $\frac{1}{2i-2}$. We get

$$\mathbb{P}_n(r) = \int_0^{\frac{1}{2i-2}} n(1-x)^{n-1} \mathbb{P}_{n-1}\left(\frac{r}{1-x}\right) dx + \int_{\frac{1}{2i-2}}^{2r} 2(1-x) \mathbb{P}_{n-1}\left(\frac{r}{1-x}\right) dx$$

$$\text{for } r \in \left[\frac{1}{2i}, \frac{1}{2i-2}\right), \; i = 2, 3, \ldots, n+1.$$

We can conclude that though the number of intervals for which we have to compute something increases with n, the complexity of the computations does not increase.

2.3 Conclusions

For each value of n the $\mathbb{P}_n(r)$ can be computed using a simple recursive expression:

$$\mathbb{P}_0(r) = \begin{cases} 0, & \text{if } r < \frac{1}{2}, \\ 1, & \text{if } r \geq \frac{1}{2}, \end{cases}$$

and for $n = 1, 2, \ldots$

$$\mathbb{P}_n(r) = \begin{cases} 0, & \text{if } r < \frac{1}{2n+2}, \\ \int_0^{\frac{1}{2i+2}} n(1-x)^{n-1}\mathbb{P}_{n-1}(\frac{r}{1-x})dx + & \\ \quad + \int_{\frac{1}{2i-2}}^{2r} n(1-x)^{n-1}\mathbb{P}_{n-1}(\frac{r}{1-x})dx, & \text{if } r \in [\frac{1}{2i}, \frac{1}{2i-2}), \\ & \quad i = 2, 3, \ldots, n+1, \\ 1, & \text{if } r \geq \frac{1}{2}. \end{cases}$$

This is a recursive expression, which means that to compute $\mathbb{P}_n(r)$ one first has to compute $\mathbb{P}_0, \ldots, \mathbb{P}_{n-1}$. And what is more, the number of computations needed per value of n increases with n, so the process will become slower and slower for large values of n.

What we expect to see is that the $\mathbb{P}_n(r)$ will become almost equal to one even for very small values of r when n becomes really large. Already with $n = 1$ and $n = 2$, in Figure 2.5, we see that the probability that the two end nodes are connected gets larger for the same values of r.

If n is infinite, the probability will be one from $r = 0$. This means that the minimum radius needed to reach a required level of probability decreases rapidly when the number of nodes increases.

One last remark I want to make is that this method does not only work if the nodes are added according to a standard uniform distribution, but also for any other continuous distribution on $(0, 1)$. The only thing that will have to change in the recursive expression for $\mathbb{P}_n(r)$ is the part where the distribution of the minimum of n independent standard uniform variables is used. Simply substitute the expression for the new distribution and the expression will be correct again.

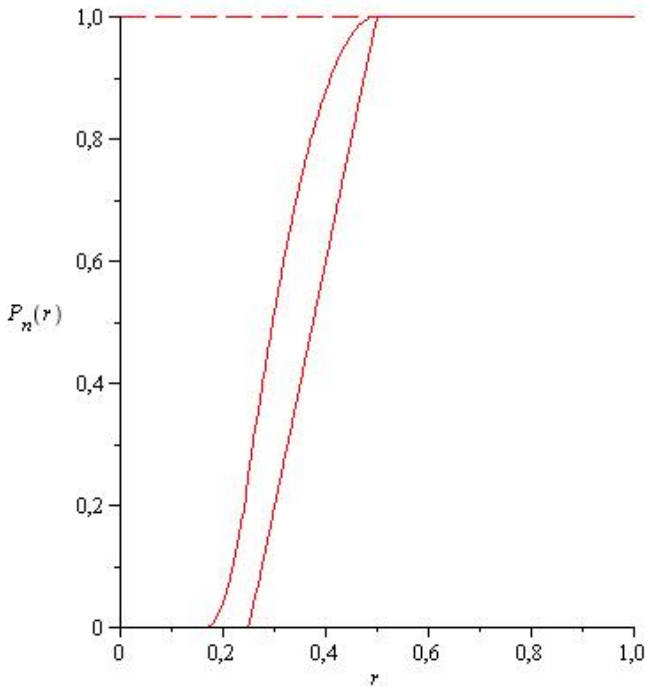

Figure 2.5: $\mathbb{P}_n(r)$ for $n = 0$, $n = 1$ and $n = 2$.

Chapter 3

The two-dimensional model

Now that the one-dimensional model has been solved it is time to move on to a model that is a bit more realistic and useful: the two-dimensional model. This model has some of the same assumptions as the one-dimensional model, namely that there are two nodes that need to be connected, new nodes are added randomly and independently, and that all nodes have the same radius. The difference is that the new nodes will not be added along a straight line, but in a plane.

It would be nice if the same idea as in the one-dimensional case could be used in the two-dimensional case. However, this does not work. In the last chapter the idea was to use a recursive expression by choosing the minimum of all added nodes and then viewing the remaining problem in exactly the same way. This works, because communication can only take place from left to right along the line. But in the two-dimensional case, there will be more than one path to use, and it is not clear which node to view as the "first" node.

For this reason, I used a different technique. This is found in the forthcoming book by Franceschetti and Meester. Most of the ideas and results in this chapter are taken from this work.

In the first section, I will give a description of the model and state the question I will try to answer. Then in the second section I will discuss some theory needed to solve the problem. In section three I will analyse the problem. Section four gives the conclusions.

3.1 Model description

In this model, the nodes will be added on a plane in $\mathbb{R}^2$. Consider a square with sides of length $\sqrt{n}$. This value is chosen because of its convenience, as will become clear later on. In this square nodes are added according to a Poisson process with density λ. This means that nodes are added randomly and independently from each other, and that the probability distribution of

the number of points in a certain domain D, $X(D)$, is given by

$$\mathbb{P}(X(D) = k) = e^{-\lambda|D|}\frac{(\lambda|D|)^k}{k!}.$$

A connection between two nodes x, $y \in \mathbb{R}^2$ will be made if the distance $||x - y||$ between them is smaller than or equal to $2r$. So the connection function, the probability that two points are connected will be $g(x-y)$ with $g(z)$

$$g(z) = \begin{cases} 1, & \text{if } ||z|| \leq 2r, \\ 0, & \text{if } ||z|| > 2r. \end{cases}$$

This model is called the Boolean model because of the zero-one nature of the connection function.

3.2 The infinite plane

In this section the boolean model will be considered on the infinite plane. One reason to do this is that on the infinite plane sharp transitions in the behaviour of the model can be observed. These sharp transitions are known as phase transitions. These occur when small changes in the parameters of the model have huge effects on the overall behaviour.

Another important reason to first look at the infinite plane is that the results derived here can be used in studying the model on a finite part of the plane. This will be done in the next section.

As said in the last section, the starting point is a Poisson point process X on the plane with density λ. Two points are connected according to the connection function, that is, they are connected if and only if the distance between them is smaller than or equal to $2r$. It is always assumed that there is a point at the origin.

A group of nodes that are connected to each other is called a connected component. The number of nodes in the connected component at the origin is denoted by $|C|$. This number can be one if the node at the origin is not connected to any other point, that is, if there is no point at all in a circle with radius $2r$ from the origin. The percolation function is the probability that the connected component at the origin is infinitely large, that is, $\theta(\lambda) = \mathbb{P}_\lambda(|C| = \infty)$. Of course $\theta(\lambda)$ increases when λ increases.

I have already said that the boolean model is a special case of the random connection model. The phase transition theorem for the random connection model is the following theorem.

Theorem 3.2.1 *There exists a $0 < \lambda_c < \infty$ such that $\theta(\lambda) = 0$ for $\lambda < \lambda_c$, and $\theta(\lambda) > 0$ for $\lambda > \lambda_c$.*

This theorem is proven by Franceschetti and Meester. The average number of connections of a node in a boolean model is called the node degree, and it is given by $\xi = 4\pi r^2\lambda$. The phase transition theorem for the boolean model can now be given in three formulations, which are all equivalent.

Theorem 3.2.2 *(i) In a boolean random network of radius r, there exists a critical density $0 < \lambda_c < \infty$ such that $\theta(\lambda) = 0$ for $\lambda < \lambda_c$, and $\theta(\lambda) > 0$ for $\lambda > \lambda_c$.*
(ii) In a boolean random network of density λ, there exists a critical radius $0 < r_c < \infty$ such that $\theta(r) = 0$ for $r < r_c$, and $\theta(r) > 0$ for $r > r_c$.
(iii) In a boolean random network, there exists a critical node degree $0 < \xi_c < \infty$ such that $\theta(\xi) = 0$ for $\xi < \xi_c$, and $\theta(\xi) > 0$ for $\xi > \xi_c$.

Exact values for ξ_c, λ_c and r_c are not known. They can be approximated by computer simulation. This gives $\xi_c \approx 4.512$.

The proof of this theorem is based on Theorem 3.2.1. In the random connection model, the connection function $g(z)$ is a function from $\mathbb{R}^2$ into $[0, 1]$ that only depends on the distance between nodes. Every two nodes x and y are connected with probability $g(x - y)$.

The connection function in the boolean model is indeed a special case of this $g(x - y)$, so that it is already enough to know that Theorem 3.2.1 is valid in our model.

Now the critical value of the density λ_c, in a boolean random network, will of course depend on the value of r. This is said more formally in the following proposition.

Proposition 3.2.1 *In a boolean random network it is the case that*

$$\lambda_c(r) = \frac{\lambda_c(1)}{r^2}.$$

Proof of Proposition 3.2.1. Consider a boolean random network with radius 1 and take a realisation G of this network. If you scale all distances in this realisation with r, you get a scaled realisation G_s. This G_s is a realisation of a boolean random network with radius $\frac{1}{r}$ and density $\frac{\lambda(1)}{r^2}$. The connections in both realisations are the same, which means that if there is an infinite connected component at the origin in G, there also is one in G_s, and if there is not any in G, then also not in G_s. Then we have that the critical value λ_c of the density in G_s is $\lambda_c(G_s) = \frac{\lambda_c(1)}{r^2}$, and the proposition is proven.

By Theorem 3.2.1 we know that for a given connection function, i.e., for a given value of r, there exists a critical value λ_c of the density where a phase transition occurs. Now if we use the same scaling method as in the proof of Proposition 3.2.1, we can turn this around and say that for a given density λ there exists a critical value r_c of the radius.

Say we have a realisation G_1 of a boolean random network with radius r_1 and at the critical value λ_{c1} of the density. Now if this model is scaled with a factor x, we get a realisation G_2 of a boolean random network with radius $r_2 = \frac{r_1}{x}$ and density $\lambda_2 = \frac{\lambda_1}{x^2}$. Now, with Proposition 3.2.1, the critical density of G_2 is $\lambda_{c2} = \frac{\lambda_{c1}}{x^2}$. This is true for every $x > 0$, so every value of r has a correspondig critical density and vice versa. The node degree is another way of saying this.

Theorem 3.2.2 tells us that the larger the radius, the lower a density is needed to have a positive probability for an infinite connected component. And reversely, the larger the density, the smaller the critical value of the radius. This is of course completely as expected. What has not yet been considered is how quickly the critical value of r decreases as λ increases and vice versa.

The first observation to be made is that when the density is high, finite clusters of points tend to consist of single isolated points. If $\lambda \to \infty$ those isolated points are the last finite clusters that remain until there are no finite clusters left. This is not only the case in the boolean model, but in any random connection model, as the following theorem shows.

Theorem 3.2.3 *In a random connection model at high density, points tend to be either isolated, or part of an infinite connected component. More precisely*

$$\lim_{n\to\infty} \frac{\log[1-\theta(\lambda)]}{\lambda \int_{\mathbb{R}^2} g(x)dx} = 1.$$

This theorem means that when λ goes to infinity, $1 - \theta(\lambda)$ behaves as $\exp(-\lambda \int_{\mathbb{R}^2} g(x)dx)$, which is the probability that a point is isolated. So the rate at which $\theta(\lambda)$ tends to one is the same as the rate at which the probability that a point is isolated goes to zero. A complete proof of this theorem is not given by Franceschetti and Meester.

Since the boolean model is a special case of the random connection model, Theorem 3.2.3 is also valid in our case. But it is possible to refine the statement a bit.

In order to have a connected component of size k it is necessary and sufficient that these k points form a connected component and are surrounded by a certain area of empty space containing no points at all. The area of empty space needed will be smaller if the k points are closer together. This means that though $\mathbb{P}_\lambda(|C| = k)$ is small when the density is high, it is larger when the k points are very close together and the approximately circular area around them (with radius r) contains no points.

Now, on the condition of a point being at the origin, let S_α be the event that a disc with radius α contains k additional points (apart from the one at the origin) and an annulus outside of the disc with width $2r$ is empty.

This is a sufficient condition for a finite component of $k+1$ points at the origin. We get

$$\begin{aligned}\mathbb{P}_\lambda(S_\alpha) &= \frac{(\lambda\pi\alpha^2)^k}{k!}\exp\left(-\lambda\pi\alpha^2\right)\exp\left(-\lambda[\pi(\alpha+2r)^2-\pi\alpha^2]\right) = \\ &= \frac{(\lambda\pi\alpha^2)^k}{k!}\exp\left(-\lambda\pi(\alpha+2r)^2\right).\end{aligned}$$

If S_α occurs, there is a finite cluster of $k+1$ points at the origin, so

$$\mathbb{P}_\lambda(|C|=k+1) \geq \mathbb{P}_\lambda(S_\alpha) \qquad \text{for all } \alpha,\ k,\ \lambda. \tag{3.1}$$

This can be refined a bit more to get a better estimate for $\mathbb{P}_\lambda(|C|=k+1)$. First maximize over α to get

$$\mathbb{P}_\lambda(|C|=k+1) \geq \max_\alpha \mathbb{P}_\lambda(S_\alpha) \qquad \text{for all } k,\ \lambda. \tag{3.2}$$

Now let λ go to infinity and rewrite a bit, then

$$\lim_{\lambda\to\infty}\frac{\mathbb{P}_\lambda(|C|=k+1)}{\max_\alpha \mathbb{P}_\lambda(S_\alpha)} \geq 1 \qquad \text{for all } k. \tag{3.3}$$

In this equation, the maximum of $\mathbb{P}_\lambda(S_\alpha)$ is reached at the point where

$$\alpha = \frac{k}{2\pi r\lambda} + O(\frac{1}{\lambda^2}),$$

where $O(\frac{1}{\lambda^2})$ denotes some function of λ of which the growth is smaller than that of $(\frac{1}{\lambda^2})$ as λ increases. If this expression for α is substituted in (3.3) we get with (3.1)

$$\lim_{\lambda\to\infty}\frac{\mathbb{P}_\lambda(|C|=k+1)}{\exp\left[-\lambda\pi(2r^2)-k\log\frac{\lambda}{k}-O(1)\right]} \geq 1 \qquad \text{for all } k. \tag{3.4}$$

This inequality provides a good approximation of $\mathbb{P}_\lambda(|C|=k+1)$ for all values of k. This can be seen as follows. First, the value of α for which $\mathbb{P}_\lambda(S_\alpha)$ is maximal goes to zero as $\lambda\to\infty$. This means that the annulus in the condition for a finite cluster of $k+1$ points becomes a circle with radius $2r$. The disc on the inside of the annulus, with radius α, contains all $k+1$ points, so when α goes to zero the points will be very close to each other.

From the necessary condition for an isolated component of size k, as stated above, it is clear that the empty area that is required becomes a perfect circle with radius $2r$. So the two conditions, the necessary and the sufficient condition, come to mean the same and this means that the inequality provides a good approximation.

It is now possible to make Theorem 3.2.3 of the random connection model a bit more precise for the case of the boolean model. From (3.4) it is clear

that for larger values of k the probability of finite components goes to zero at a higher rate. So the last finite components to disappear as the density increases are those of size one, or the isolated points. This gives the new theorem.

Theorem 3.2.4

$$\lim_{\lambda\to\infty} \frac{1-\theta(\lambda)}{\exp\left(-\lambda\pi(2r)^2\right)} = 1.$$

Now in the case of a boolean model on the infinite plane we know that there is, for any given value of r, a critical density λ_c, which is the minimum density required for at least a positive probability of an infinite connected component. Also we know something about the behaviour of the model at high density.

However, now it is time to study the same model on a finite part of the plane. This is what we are interested in, since real-life situations will always be finite.

3.3 Connectivity in a finite square

In this section the boolean model is considered on a finite square B_n of size $\sqrt{n} \times \sqrt{n}$. The starting point is a boolean model with density $\lambda = 1$ and radius $r > 0$ on the whole plane. The restriction $G_n(r)$ of this network is formed by all the nodes that are inside the $\sqrt{n} \times \sqrt{n}$ square and any connections between these nodes.

Just as in the last section, we always have the condition that there is a point at the origin, and again $\theta(r)$ denotes the probability of an infinite connected component being at the origin. The number of nodes inside B_n is called $N(B_n)$. Because $\lambda = 1$, $\mathbb{E}(N(B_n)) = n$. The number of nodes inside B_n that are part of an infinite connected component on the whole plane is denoted by $N_\infty(B_n)$.

To take the model with $\lambda = 1$ does not make a real restriction, because everything that is said in this section also holds for a model with a square of size 1×1, density $\lambda = n$ and all distances divided by $\sqrt{n}$. This is because the model can be scaled and remains essentially the same, as was already used in the last section.

The percolation function $\theta(r)$ is the probability that a point is part of an infinite connected component in the boolean model on the whole plane. The following proposition tells something more about $\theta(r)$.

Proposition 3.3.1 *We have* $\theta(r) = \mathbb{E}(N_\infty(B_1))$.

The proof of this proposition is given by Franceschetti and Meester. In the rest of this section, first almost connectivity will be discussed. This is about the requirements needed to have a connected component in the square of at least a certain size. The second subsection will discuss the property of full connectivity, where all of the nodes inside the square must be connected.

3.3.1 Almost connectivity

Almost connectivity means that at least a certain fraction of all the nodes inside the square must be connected. Because the percolation function $\theta(r)$ for the boolean model on the whole plane represents the probability that a node is part of an infinite connected component, it is reasonable to expect that the fraction of the nodes inside the square that is connected is about the same as this $\theta(r)$. So, if r is at least above the critical value of the radius, it is possible to make an estimate of the fraction of the nodes that are connected.

For $\alpha \in (0,1)$ the restriction $G_n(r)$ of a boolean random network is α-almost connected if it contains a connected component of at least αn nodes. Because the expected number of nodes inside the square B_n is n, α is the fraction of the nodes that is connected if $n \to \infty$.

The following theorem states that it is not only possible to say for a certain value of r what fraction of the nodes will be connected, but also to turn this around.

Theorem 3.3.1 *Let*

$$r_\alpha = \inf_r(\theta(r) > \alpha).$$

We have that for any $\alpha \in (0,1)$, if $r > r_\alpha$ then $G_n(r)$ is α-almost connected asymptotically almost surely, while for $r < r_\alpha$ it is not.

The term "asymptotically almost surely" means that an event occurs with probability tending to one as $n \to \infty$. This theorem and its full proof are given by Franceschetti and Meester.

This theorem means that for any chosen value α of the fraction of nodes to be connected, there exists some critical value r_α of the radius, and only if the radius is larger than this critical value the desired fraction of the nodes will be connected if $n \to \infty$. Note that though this critical value does not depend on the size n of the square, because $\theta(r)$ does not depend on n, the statement only holds for very large values of n. In the next subsection it will become clear that the requirements for full connectivity of all the nodes inside the square do depend on n.

3.3.2 Full connectivity

In the last section is was shown that for a larger fraction of the nodes to be connected, a larger radius is needed. And it was also shown at what rate

the radius has to grow as the required fraction increases. For all the nodes to be connected the radius must be even larger than in the last section, and it must also increase with the size of the square B_n. But how fast does the radius have to grow?

The following theorem gives a first indication of the required ratio of the radius to the length $\sqrt{n}$ of the side of the square.

Theorem 3.3.2 *Let $\pi r_n^2 = \alpha \log n$. If $\alpha > \frac{5}{4}\pi$ then $G_n(r)$ is connected with high probability, while for $\alpha < \frac{1}{8}$ it is not.*

This theorem and the proof are given by Franceschetti and Meester. It is done in two parts, one to show that the restriction of the boolean random network is not connected if $\alpha < \frac{1}{8}$ and the other to show that it is fully connected if $\alpha > \frac{5}{4}\pi$. The idea for the second part is that, after dividing the square into subsquares, for larger values of α all the subsquares will have at least one node inside it. And by choosing the right size for the subsquares, it is shown that for $\alpha > \frac{5}{4}\pi$ the nodes in adjacent subsquares always make a connection to each other. These two things taken together mean that all the nodes inside the square are connected to each other.

The other part can be shown by filling the square with annuli of inner radius r_n and outer radius $3r_n$, and saying that if there is a node inside the inner radius but none between the radii, the node inside the inner radius is isolated. Now if α is small enough, there fit enough of these annuli inside the square to make the probability that there is an isolated point in at least one of these annuli positive, and then the restriction will not be connected.

From Theorem 3.3.2 it is clear that in order to reach full connectivity inside the square, the radius must grow at least as quickly as the logarithm of the length of the side of the square, $\sqrt{n}$. It also gives some boundaries for values of r_n where there is a full connection at a given n, and for which values there is not. But what if α is in between $\frac{1}{8}$ and $\frac{5}{4}\pi$?

The following theorem makes the necessary rate of growth of r more precise. It also tells us also something about the intermediate values, if only in the asymptotic case.

Theorem 3.3.3 *Let $\pi(2r)^2 = logn + \alpha_n$. We have that $G_n(r)$ is connected with high probability if and only if $\alpha_n \to \infty$.*

The theorem and a proof, which is quite long and difficult, are given by Penrose. Franceschetti and Meester give the theorem without a full proof but with a short outline of the most important steps of the proof. These steps I will also try to explain here.

The first thing needed is to show that with high probability, isolated nodes do not appear if and only if $\alpha_n \to \infty$. The sum of all the events that a single point is isolated, which has a low probability, is approximated by a Poisson distribution.

The second step is to show that no possibility of isolated nodes is equivalent with having full connectivity of all nodes inside the box. Seen from one way, this statement is very logical. Recall from Theorem 3.2.4 that isolated nodes are the last finite components to disappear when approaching when the density, or in this case the radius, tends to infinity. So when there are no isolated nodes left, there will not be any other finite components left and there is full connectivity. Unfortunately, this is not enough to prove the truth of the statement; the theorem does not rule out the possibility that there are very large groups of nodes connected to each other, that are not connected to nodes inside the box, even if there are no single isolated nodes in the box left.

It turns out that the behaviour of the restriction of the model inside the square is exactly the same as that of the longest edge of the nearest neighbour graph that has the nodes in the square as vertices. And finally this longest edge converges asymtotically to the longest edge of the minimum spanning tree of the nodes inside the square. Then it is possible to say that the restriction $G_n(r)$ is connected if and only if the longest edge in the minimum spanning tree is smaller than $\frac{1}{2}r_n$.

All these parts together outline the proof of Theorem 3.3.3.

3.4 Conclusions

The results stated in the last two sections are all that is known about the two-dimensional case. For partial or α-almost connectivity, we can give an estimate of the critical radius r_α. Then for very large values of n it is possible to say that, if the radius is larger than the critical radius, a fraction α of all the nodes inside the square will be part of an infinite connected component. That is, a fraction α of the nodes inside the square will be connected to each other.

About full connectivity, we can for some combinations of r and n say whether there will be full connectivity or not, by Theorem 3.3.2. If the combination gives a value of α for which this theorem does not say anything, we have to resort to Theorem 3.3.3. But then we can only look at very large values of n and scale r accordingly.

These are the only things that can be said about the connectivity in the two-dimensional case. As I said above, it is not possible to say for any value of n and r what the probability will be that there is a connection. This would be what we ideally wanted to know, because this could then be used to make an estimate of the number of nodes needed to have a certain (high)

probability of connection for a given r in a real-life situation. What we can do now is to make sure to arrange our nodes and radii so that the value of α as in Theorem 3.3.2 is larger than $\frac{5}{4}\pi$. Because all the results also hold in the scaled version for any value of the density λ, the value of α can also be influenced by changing the density.

Chapter 4

Discussion and suggestions

In the two last chapters we have seen that for the one-dimensional model we can give the precise probability that two nodes are connected. This can be done, with some computation, for any value of the radius and for any number of addes nodes. In the two-dimensional case, we have seen that making these precise statements was impossible. We could say something about a certain fraction of the nodes inside a square being connected, or about full connection of all the nodes inside the square in the asymptotical case.

The first thing then that would be interesting for further research is whether something can be said in the non-asymptotical case. Is it possible to make an estimate or lower or upper bounds for the probability that two nodes are connected given a certain radius and number of nodes? Or for the expected fraction of the nodes in a certain area that is connected? These are interesting questions, that have not been answered yet.

The other thing that I want to discuss here are the applications and modifications of the theory discussed in this paper. The models as they stand at the moment are certainly interesting, and a good starting point for real-life situations. However, there are certain assumptions in the models that will not be met in real situations. First, there is the assumption that all the nodes have the same radius of their signal. But that will not be the case in any situation. It would be nice to have a model that can do the same things as these models, but with a variable radius. What about a stochastic radius with some chosen probability distribution? Or choosing randomly from three different values, one for laptops, one for palms and one for cell phones? I am sure there are more variations that can be interesting to study.

Also it would make the model a little bit more realistic if it would be three-dimensional. In a real situation, not all nodes, or appliances, will be at the same height. If the differences are small this may not make a huge

difference, but imagine a network in the mountains. This will probably be a hard thing to do, seeing how going from one to two dimensions complicated the model a lot. The last thing that may be important or interesting is disturbances and failure of the signal. If an appliance fails, this could be seen as the same model with one node less. So by estimating the probability that a certain number of nodes fails, this problem could be addressed. But disturbances or static are not covered yet.

Bibliography

[1] Franceschetti, M, and Meester, R., (to appear). *Random networks for communication.*

[2] Penrose, M.D., 1997. The longest edge of the random minimal spanning tree. *Annals of Applied Probability* 7(2), 340-361.