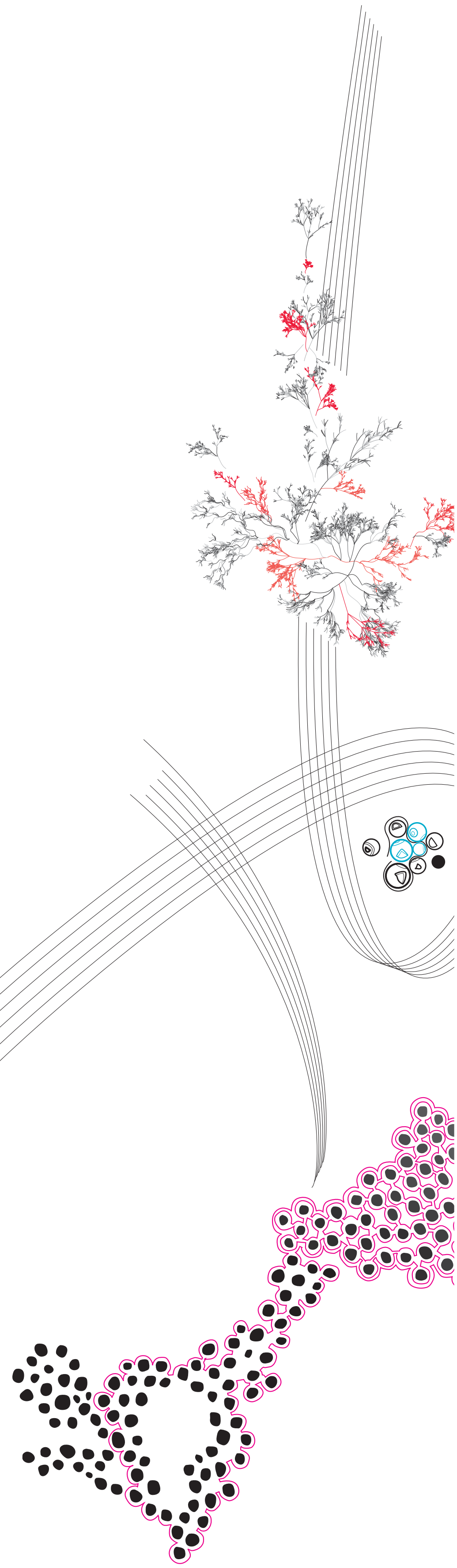


BSc Thesis Applied Mathematics

Deploying base stations in a 5G network

Jorrit Wilmink

Supervisors: L. Weedage & C. Stegehuis

July, 2022

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Preface

I want to thank Lotte Weedage and Clara Stegehuis for supervising me and for all their help and guidance during the last weeks. Without their feedback and assistance I would not be able to achieve this.

Deploying base stations in a 5G wireless network

Jorrit Wilmink*

July, 2022

Abstract

In this paper, we model a fifth generation wireless network as a random geometric graph, using Poisson point processes. Furthermore, we calculate the performance measures average signal-to-noise ratio, number of disconnected users and coverage area. Then we model the effect of damaged base stations, as rain or other random events, such as failure could cause that one is not able to connect to the base station. We investigate where we should deploy new base stations in order to minimize the probability of having no internet connection. The results are accomplished by simulating the model multiple times and verifying which method yields the best solution. In the long run, it is best to deploy base stations while maximizing the coverage area.

Keywords: base station deployment, Poisson point process, random geometric graph, SNR, coverage area, 5G network

*Email: j.wilmink@student.utwente.nl

Contents

I Introduction

5G is the 5^{th} generation mobile network. Compared to older generation networks, 5G enables a network that uses higher and wider bandwidths compared to the previous generations. 5G is meant to deliver higher peak data speeds, which is the number of bytes per second that data travels from the one device to the other, more reliability and a massive network capacity. One should want higher data speeds and a more reliable connection, in order to be able to connect with multiple devices at a good internet speed, such that, for example, one can stream videos in high quality. A base station is a fixed station that uses radio waves to communicate with mobile devices. In order to satisfy these 5G requirements, the deployment density of base stations affects the probability users can connect with the base stations, as a particular connection can fail due to random events as rain, or the line-of-sight probability. These random events can cause a different deployment strategy.

In the literature, [7] a lot of research has already been done on the analysis and design of wireless networks, such as sensor networks and cellular networks. Random graphs were first introduced in 1959 [3], where Paul Erdös assigned each edge a fixed probability that is exists. In 1961, Gilbert was the first to model wireless networks using random graphs [5]. Connecting stochastic geometry and random graphs to base station deployment in wireless networks [2], we will construct a model to determine where to deploy base stations in a 5G wireless network.

In this paper the following research question will be addressed.

"Where should we deploy new base stations to minimize the probability of having no internet connection in a 5G network?"

The base stations in a 5G-network and the links with different users can be modelled as a random geometric graph where the base stations and users are modelled as nodes, and the links as edges. There already quite some techniques which led to results on the connectivity, the capacity, and the outage probability of wireless networks [6], which can be used to calculate for example the outage probability. The outage probability is the probability that the signal-to-noise ratio exceeds a given threshold, such that a connection can be established.

The following sub-questions need to be answered.

- *"How to determine where base stations need to be deployed?"*
- *"How do failures affect the deployment of new base stations?"*

The outline of this paper is as follows. Firstly, necessary theory is explained in Section II. The description of the model is introduced in Section III, such that in Section IV the simulations of different deployment strategies on the model can be described as well as their results. In Section IV.A we perform 100 simulations to compare results of different deployment strategies and in Section IV.B we perform 100 simulations of different deployment strategies when base stations do not operate or if the connection radius of base stations decreases. In the remainder of the paper, we discuss the acquired results, we discuss the model and we draw up conclusions.

II Theory

We want to model the 5G wireless network mathematically. A network consists of users and base stations. If a connection is established between user and base station, it means that the signal-to-noise ratio (SNR) from base station to user is high enough for the user to be connected to the internet at a proper speed. The SNR is a measure that compares the level of the received signal to the level of background noise. For a good reliable connection, the SNR needs to be large enough, as level of the received signal needs to be good enough, considering the background noise. In order to do that, we first need to define when there is a possible connection between a base station i and a user j. To achieve this, we start by looking at the transmission power P_t of a base station. The received power at a user is the strength of the transmitted signal from the base station. We want to obtain an expression for the received power P_r at a user, as a connection can be established when P_r is sufficiently large. The power is in Watt.

A Received power

The received power P_r of a user rapidly decays over distance. Denote the distance in meters between the antennas of base station i and user j by r_{ij}. The gain describes how well the antenna converts radio waves into electrical power. Let the wavelength of the transmitted signal be λ_c. Denote G_t and G_r as the gains of the transmitting antenna and of the receiving antenna respectively. The gains give the signals power and strength. Note that the gains are unitless. Friis equation [4] gives an expression for the received power P_r of a user over a distance r_{ij} in terms of the transmitted power

$$P_r = P_t G_t G_r \left(\frac{\lambda_c}{4\pi r_{ij}} \right)^2 . \tag{1}$$

This is called the Friis transmission formula [12]. Most models are generalized such that the distance dependent term of P_r is of the form $r_{ij}^{-\alpha}$, [13] where the path loss exponent α is a parameter that can be chosen to fit in the specific environment. Thus in (1), α equals 2.

dBm is a unit of level used to indicate that a power level is expressed in decibels (dB) with reference to one milliwatt (mW). Let K_0 be the catch-all constant such that at a distance of $r_{ij} = 1$, we get a path loss in decibel-milliwatts (dBm). For example, In Friis transmission formula (1), $r_{ij} = 1$ yields $K_0 = G_t G_r \left(\frac{\lambda_c}{4\pi}\right)^2$ [13]. In other words, substituting K_0 into (1) yields a path loss with a constant slope.

$$P_r = P_t K_0 r_{ij}^{-2}.$$

This is called a standard path loss model; the slope of the path loss in dBm is constant and is determined only by α. As the dBm is on a logarithmic scale, the received power can be expressed as the transmitted power plus the catch-all constant minus the path loss in dBm.

$$\log_{10}(P_r) = \log_{10}(P_t) + \log_{10}(K_0) - 20 \log_{10}(r_{ij}).$$

B Path loss function

The standard path loss model is the basis for the analysis, design and simulation of cellular networks. Considering $\alpha = 2$, see Equation (1), we define the standard (power-law) path loss function as

$$l_1(r_{ij}) := (r_{ij})^{-2}.$$

The subscript 1 is to indicate its single slope property. This property often leads to some questionable results in some special cases approaching infinity and leading to singularities for the path loss [8]. Therefore, we will use the two-ray model to model a more realistic scenario [11]. The two-ray model has one direct path and one ground-reflected path, which results in a dual slope path loss behaviour, as the waves of the base station are headed downwards. Once the waves reach the ground, the power will decrease more rapidly, implying there exists a critical distance $R_c > 0$. The near-field path loss exponent $\alpha = 2$ is used until a critical distance R_c. This distance R_c is based on the transmission and receiver antenna heights, h_t and h_r respectively. The critical distance is approximately $R_c \approx \frac{4h_t h_r}{\lambda_c}$. Above this distance, the far-field path loss exponent becomes $\alpha = 4$. Note that there is a large difference between $(r_{ij})^{-2}$ and $(r_{ij})^{-4}$ for feasible values of r_{ij}, see Figure (1). In addition, approximating the difference using $\alpha = 3$, yields large errors on the received power for the users too [13]. Therefore, we define the dual slope (power-law) path loss function [11] as

$$l_2(r_{ij}) := \begin{cases} (r_{ij})^{-2}, & r_{ij} \leq R_c \\ R_c^2 (r_{ij})^{-4}, & r_{ij} > R_c. \end{cases}$$

The constant R_c^2 is introduced to maintain continuity. Again, the subscript 2 is to indicate its dual slope property.

As it is more convenient to work in dB, we also convert the dual slope path loss function to dB. This follows immediately after taking the log with base 10,

$$\ell_2(r_{ij}) := \begin{cases} -20\log_{10}(r_{ij}), & r_{ij} \leq R_c \\ 20\log_{10}(R_c) - 40\log_{10}(r_{ij}), & r_{ij} > R_c. \end{cases} \tag{2}$$

A plot of the dual slope path loss $\ell_2(r_{ij})$ is given in Figure (1).

C Signal-to-noise-ratio

We have defined a path loss function and therefore, we are able to describe the signal-to-noise-ratio (SNR). The SNR compares the level of the received power to the power of the background noise. We want to have an expression for the SNR, as the SNR needs to exceed a given threshold in order to establish a connection. The SNR is the signal-to-noise-ratio, implying $SNR := \frac{P_{r_i}}{N}$.
Where P_r is the received power at a user, P_{r_i} is the power of the signal the user receives from base station i. N is the noise in Watt. Assuming $K_0 = G_t G_r \left(\frac{\lambda_c}{4\pi}\right)^2$ is constant, i.e., the signal is only distance-dependent, we obtain

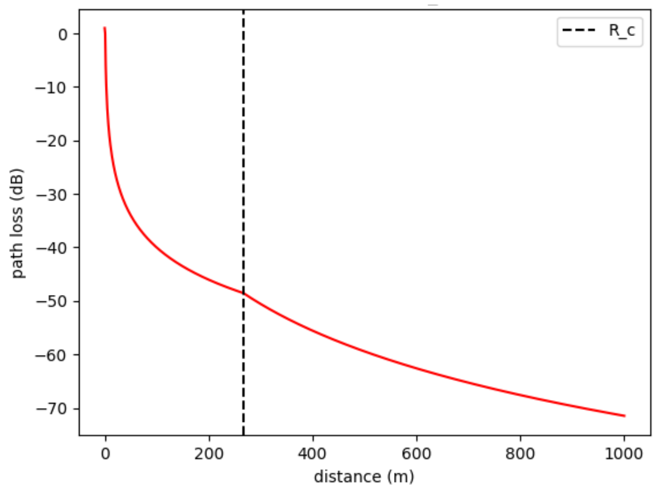


FIGURE 1: Dual slope path loss $\ell_2(r_{ij})$ with a critical distance of $267m$.

$$\text{SNR}_{ij} = \frac{P_{r_i}}{N} = \frac{K_0 P_{ti} l_2(r_{ij})}{N}.$$

with P_{t_i} the transmitted power of base station i and the critical distance is assumed to be $\frac{4h_t h_r}{\lambda_c}$. Taking the logarithm with base 10, we obtain an expression for the SNR in dB terms,

$$\log_{10}(\text{SNR}_{ij}) = \log_{10}(P_{t_i}) + \log_{10}(K_0) + \log_{10}(l_2(r_{ij})) - \log_{10}(N), \quad (3)$$
$$\text{SNR}_{ij} = P_{t_i} + K_0 + \ell_2(r_{ij}) - N. \quad (4)$$

Now, (4) is in dB. As we have an expression for the SNR, we can describe mathematically when there is a connection between user and base station. We set a value for the SNR which the SNR at a particular user needs to exceed in order to have a connection. So if the SNR at user j exceeds the threshold, say t, we model a link with base station i.

D Random geometric graph

As we have described when there is a connection between user and base station, we can start modelling the 5G network as a random geometric graph [10]. A random geometric graph has nodes with a position. In our model, base station i and user j are modelled as nodes, and the connection between a user and a base station denotes an edge, which exists if the SNR exceeds the given threshold t. We denote the edge between base station i and user j as ij. So, if $\text{SNR}_{ij} > t$, a connection exists.

The positions of the nodes are modelled as Poisson point processes (PPP). A Poisson point process is a random collection of points in space. Most important about Poisson point processes is that it independently and uniformly generates points in space according to a Poisson distribution based on some density λ [7].

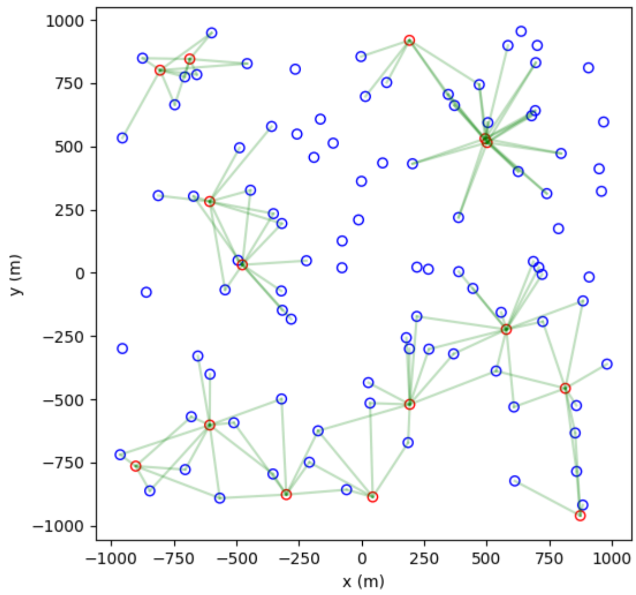


FIGURE 2: Standard graph of users and base stations.
Base stations are red, users are blue and connections are green.

III Model

It is desired to have to have a model be able to simulate base stations, users and the connection between them. In our model, the base stations i, as well as the users j are assumed to be distributed as Poisson point processes $\Phi_b = \{i\} \subset \mathbb{R}^2$ and $\Phi_u = \{j\} \subset \mathbb{R}^2$ with densities λ_b and λ_u respectively. We are interested in a method of deploying base stations where we aim to minimize the outage probability. We are first looking if we can maximize the average SNR and if we can minimize the users without connection by deploying a single new base station, as both methods could possibly minimize the outage probability.

A Set-up of the model

Denote the position of base station i and user j as (x_{b_i}, y_{b_i}) and (x_{u_j}, y_{u_j}), respectively. Take $\lambda_b = 15$ and $\lambda_u = 100$ over the total area. Let $x_{b_i}, y_{b_i}, x_{u_j}, y_{u_j} \in (-1000, 1000)$. Using the parameters $P_{r_i} = -30\ dB \quad \forall i \in \Phi_b, \quad N = -100\ dB$ and $R_c = 267m$, an example of a random geometric graph is given in Figure 2. An edge exists when the SNR at the user from a particular base station exceeds the threshold $t = 15\ dB$.

The performance measures we look at are average SNR, number of disconnected users and coverage area. The performance measures of the standard graph are given in Table 1.

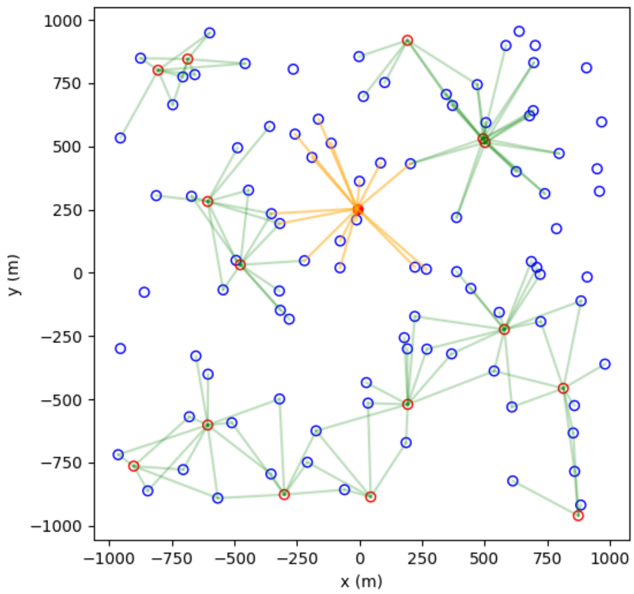


FIGURE 3: Adding a single base station when maximizing the average SNR considering 60×60 possible positions.
One can see that an extra base station is deployed at $(-10, 254)$.

Method 1: optimizing average SNR

We want to compute which coordinate is the best place to deploy a base station such that the average SNR is maximal compared to all other coordinates. For sake of computational time, we do not consider all coordinates, but we do consider $n \times m$ possible locations to deploy a new base station. We start at the bottom-left of the particular area and we go from left to right and from bottom to top, to verify which coordinate is optimal.

In Figure 2 all possible connections are plotted, but we assume that a user connects to the base station that provides the highest SNR for the user. Therefore, we compute the best connection each user has and we calculate the average over all users. If a user has no connections, we set the SNR at zero. For $n \times m$ evenly divided coordinates, we compute the average SNR if we deploy a base station at that specific location. Comparing the values of the average SNR for all $n \times m$ positions, we deploy a new base station at whatever point results in the highest average SNR, see Figure 3. For convenience, the connections of the new base station are coloured orange.

We call the algorithm to maximize the average SNR whilst adding a single new base station Method 1 and we provide a pseudo-code for this method, see Pseudo-code 1. E is a matrix with dimension the number of base stations by the number of users. The performance measures of the graph after applying Method 1 are given in Table 1.

Pseudo-code 1: **Method 1: optimizing average SNR**

```
For n×m coordinates add a base station
    For all i base stations
        For all j users
            If SNR_ij > t
                E_ij = SNR_ij
            Else
                E_ij = 0
    For all j users
        Take the maximum E
    Compute the average of the maximum E
Deploy a base station at the coordinate corresponding
to the highest average E
```

Method 2: optimizing number of connected users

Another performance measure we look at is the number of disconnected users. If for base station i and user j, $\mathrm{SNR}_{ij} < t$, then i and j are not connected. If we count the number of users who do not have a single connection, we know the number of disconnected users.

We want to know at which coordinate we need to deploy a base station in order to have the highest number of connected users. To achieve this, we again look at $n \times m$ possible coordinates for the same reasons as in Method 1. Let E be a matrix with dimension the number of base stations by the number of users. To obtain the number of disconnected users, we assign the value 1 to E_{ij} if $\mathrm{SNR}_{ij} > t$ and we assign the value 0 to E_{ij} if $\mathrm{SNR}_{ij} \leq t$. Then, we count the number of disconnected users. For Method 2, we compute the number of disconnected users for all $n \times m$ possible coordinates where we can deploy a new base station and we deploy a new base station at the coordinate that results in the least disconnected users. see Figure 4. Note that this coordinate is not unique, and therefore the algorithm deploys a new base station at the first coordinate where the number of connected users is the highest.
We call the algorithm to maximize the number of connected users whilst adding a single new base station Method 2 and we provide a pseudo-code for this method, see Pseudo-code 2. The performance measures of the graph after applying Method 2 are given in Table 1.

Pseudo-code 2: **Method 2: optimizing number of connected users**

```
For n×m coordinates add a base station
    For all i base stations
        For all j users
            If SNR_ij > t
                E_ij = 1
            If SNR_ij ≤ t
                E_ij = 0
    For all j users
        Take the maximum E
    Count the number of zeros of E
Deploy a base station at the first coordinate corresponding
to the lowest number of disconnected users
```

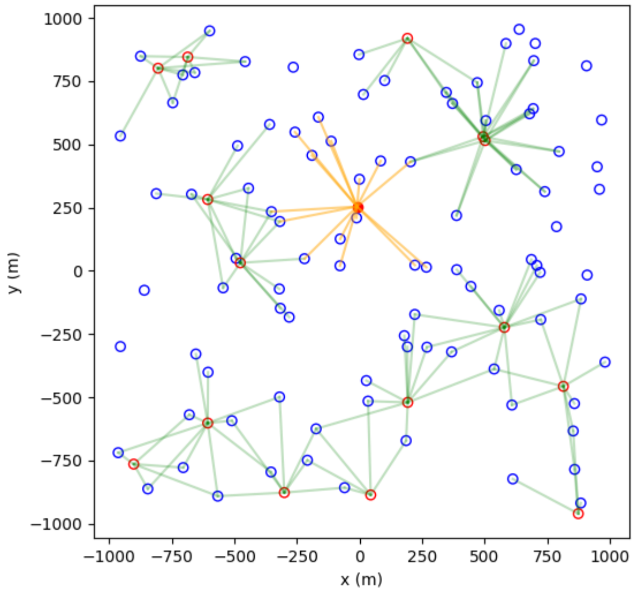


FIGURE 4: Adding a single base station when maximizing the number of connected users considering 60×60 possible positions.
One can see that an extra base station is deployed at $(-10, 254)$.

B Method 3

We described two methods to determine where to deploy a new base station. We want to minimize the outage probability, but we do not know yet which method is better. Before we simulate the first two methods, we first discuss a third method, such that we are able to compare multiple methods.

Visualisation

Instead of drawing each connection as an edge, we can also look at the coverage area of all base stations combined by considering the radius where users can connect with a base station. For this we need to know for what radius the SNR is above the given threshold. The maximum radius for which users can connect can be computed by solving Equation (4) for r_{ij}. Denoting the threshold for the SNR as t, substituting Equation (2) into Equation (4) and solving for r_{ij} yields

$$r_{ij} = 10^{-\frac{t-P_{t_i}+N-20\log_{10}(R_C)}{40}}, \tag{5}$$

given that the connection radius is larger than the critical distance. Denote the maximum radius for base station i in which user j is able to connect as r_{c_i}.
As each base station has a radius in which users are able to connect, there are i distinct circles c_i where users are able to connect, see Figure 5. For each base station, the area users are able to connect is given by

$$c_i : (x - x_{b_i})^2 + (y - y_{b_i})^2 \leq r_{c_i}^2. \tag{6}$$

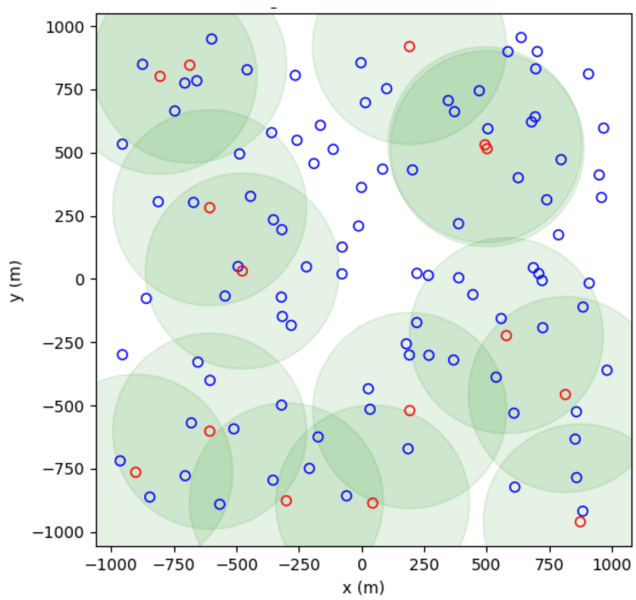


FIGURE 5: Coverage area of all base stations for a critical distance of 267 m.

Intersection points

To calculate the total coverage area, we need to add up the areas of all distinct circles and subtract the overlapping areas of all two circles that intersect with each other. In order to calculate the overlapping area, we first need to calculate the intersection points if there are any.
Denote d_{ij} as the distance between base station i and base station j. We distinguish three cases,

- if $d_{ij} > r_{c_i} + r_{c_j}$, then the circles are too far apart and do not overlap and if $d_{ij} = r_{c_i} + r_{c_j}$, then the circles have one intersection point and do not overlap. In both cases, we are not interested, as there is no overlapping area.
- If $d_{ij} < |r_{c_i} - r_{c_j}|$, then one circle lies in the other circle, and the overlapping area is the area of the smaller circle.
- However, if $d_{ij} < r_{c_i} + r_{c_j}$, then the circles do overlap and there are two intersection points, which we need to know in order to compute the overlapping area.

The equations of the circles where users can connect to base station i and j are obtained by substituting an equal sign for the less than or equal to sign in Equation (6). We need to solve the system of two equations to locate the intersection points. To find a general formula for the intersection points of two overlapping circles, we first look at the special case where the centre of circle c_i lies at the origin, and the center of circle c_j lies at the point $(d_{ij}, 0)$ on the x-axis, see Figure 6.

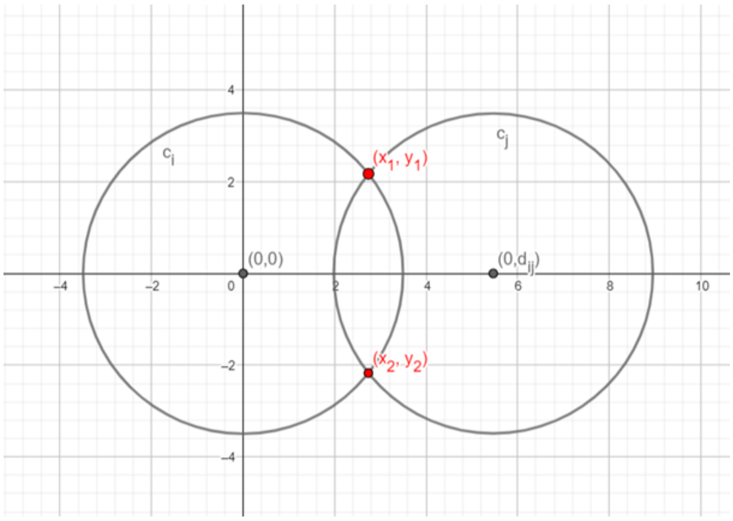


FIGURE 6: c_i centered at the origin and c_j centered at $(d_{ij}, 0)$.

Then the system of equations simplifies to

$$\begin{cases} x^2 + y^2 = r_{c_i}^2 \\ (x - d_{ij})^2 + y^2 = r_{c_j}^2. \end{cases} \tag{7}$$

After subtracting the equations from each other and solving for x, we obtain:

$$x = \frac{r_{c_i}^2 - r_{c_j}^2 + d_{ij}^2}{2d_{ij}}$$

And from Pythagoras we derive:

$$y = \sqrt{r_{c_i}^2 - x^2}$$

Looking back to the general case, we need to solve the system of equations:

$$\begin{cases} (x - x_{b_i})^2 + (y - y_{b_i})^2 = r_{c_i}^2 \\ (x - x_{b_j})^2 + (y - y_{b_j})^2 = r_{c_j}^2 \end{cases} \tag{8}$$

The midpoint between the two centers is $M_{ij} = (\frac{x_{b_i}+x_{b_j}}{2}, \frac{y_{b_i}+y_{b_j}}{2})$. We can use change of coordinates, such that M_{ij} is located at the origin and the centers of c_i and c_j are laying on the x' axis, see Figure 7. This is convenient, as the system of equations in (8) can be transformed into a simpler system of equations similar to system of equations in (7).

$$\begin{cases} (x' + \frac{d_{ij}}{2})^2 + (y')^2 = r_{c_i}^2 \\ (x' - \frac{d_{ij}}{2})^2 + (y')^2 = r_{c_j}^2. \end{cases}$$

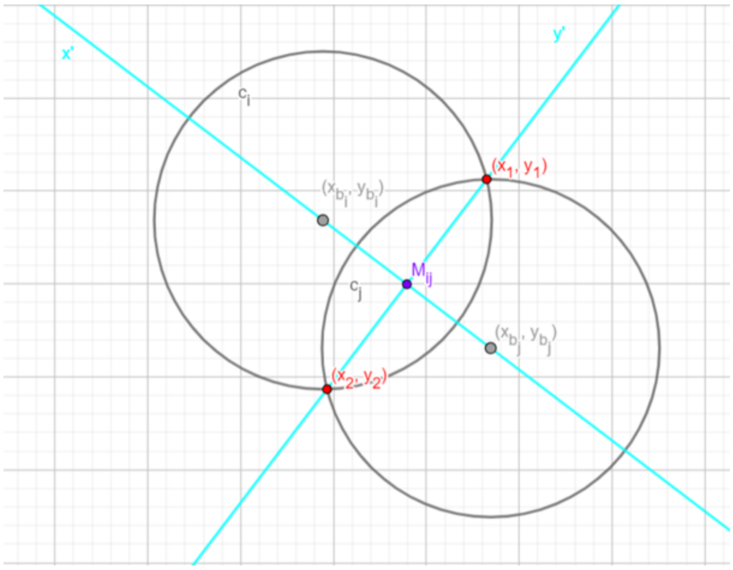


FIGURE 7: Circles c_i and c_j shifted to x' and y' with M_{ij} as origin.

This system of equations can be transformed to, using the change of variable

$$x = \frac{x_{b_j} - x_{b_i}}{d_{ij}} x' + \frac{y_{b_j} - y_{b_i}}{d_{ij}} y' \quad \& \quad y = \frac{y_{b_j} - y_{b_i}}{d_{ij}} x' - \frac{x_{b_j} - x_{b_i}}{d_{ij}} y'.$$

Let

$$l = \frac{r_{c_i}^2 - r_{c_j}^2 + d_{ij}^2}{2d_{ij}} \quad \& \quad h = \sqrt{r_{c_i}^2 - l^2}.$$

In the current coordinate system, $(\frac{x_{b_j}-x_{b_i}}{d_{ij}}, \frac{y_{b_j}-y_{b_i}}{d_{ij}})$ and $(\frac{y_{b_j}-y_{b_i}}{d_{ij}}, -\frac{x_{b_j}-x_{b_i}}{d_{ij}})$ are orthogonal unit vectors, so we can rotate, translate and apply the knowledge from the special case to get the general solution [1]:

$$(x_1, y_1) = \left(\frac{l}{d_{ij}}(x_{b_j} - x_{b_i}) - \frac{h}{d_{ij}}(y_{b_j} - y_{b_i}) + x_{b_i}, \frac{l}{d_{ij}}(y_{b_j} - y_{b_i}) + \frac{h}{d_{ij}}(x_{b_j} - x_{b_i}) + y_{b_i} \right)$$
$$(x_2, y_2) = \left(\frac{l}{d_{ij}}(x_{b_j} - x_{b_i}) + \frac{h}{d_{ij}}(y_{b_j} - y_{b_i}) + x_{b_i}, \frac{l}{d_{ij}}(y_{b_j} - y_{b_i}) - \frac{h}{d_{ij}}(x_{b_j} - x_{b_i}) + y_{b_i} \right)$$

Here, (x_1, y_1) and (x_2, y_2) are the two intersection points of circles c_i and c_j.

Overlapping areas

The overlapping area of two intersecting circles can be computed by dividing the area of overlap into two semi-ovals.

If we look at the overlap of circles c_i and c_j, then we can divide the overlapping area into two semi-ovals by the y'-axis, see Figure 7. We assume that every base station has the same radius. Consequently, by symmetry, both semi-ovals have the same area. Hence, if we compute the area of one of the semi-ovals, we know the total area of overlap.

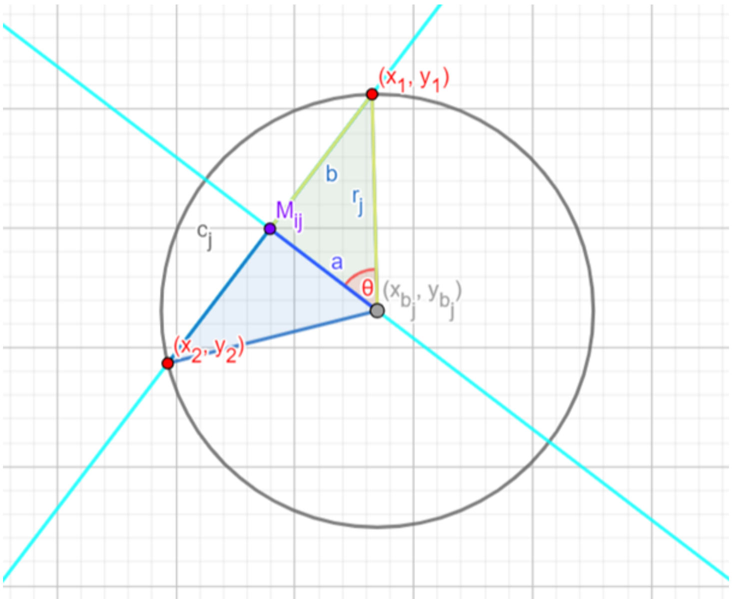


FIGURE 8: Area of the semi-oval lying on the negative x'-plane.

The area of the semi-oval lying on the left side of the y'-axis can be computed by noting that the semi-oval consists of the line in between the two intersection points (x_1, y_1) and (x_2, y_2) and the circular arc of circle c_j in between the same intersection points.

The area of this semi-oval can be calculated by computing the area of the circle sector with angle θ, subtracting the the area of the triangle with vertices M_{ij}, (x_{b_j}, y_{b_j}) and (x_1, y_1), see Figure 8.

Denote the sides of the triangle as a, b and r_j, where a is the side from M_{ij} to (x_{b_j}, y_{b_j}) and b is the side from M_{ij} to (x_1, y_1). Then the area of this triangle can be computed by Heron's formula [9]:

$$\text{Area} = \sqrt{s(s-a)(s-b)(s-r_{ij})}. \tag{9}$$

$$s = \frac{a+b+r_j}{2}. \tag{10}$$

An alternative formula for the area of the triangle is

$$\text{Area} = \frac{1}{2} a r_j \sin\theta. \tag{11}$$

We derive the value of θ by substituting Equation (9) into Equation (11), as we can compute the area of the circle sector by taking $\frac{\theta}{2\pi}^{th}$ part of the area of the circle. Subtracting the area of the triangle from the are of the circle sector leads to a quarter of the area of the overlap.

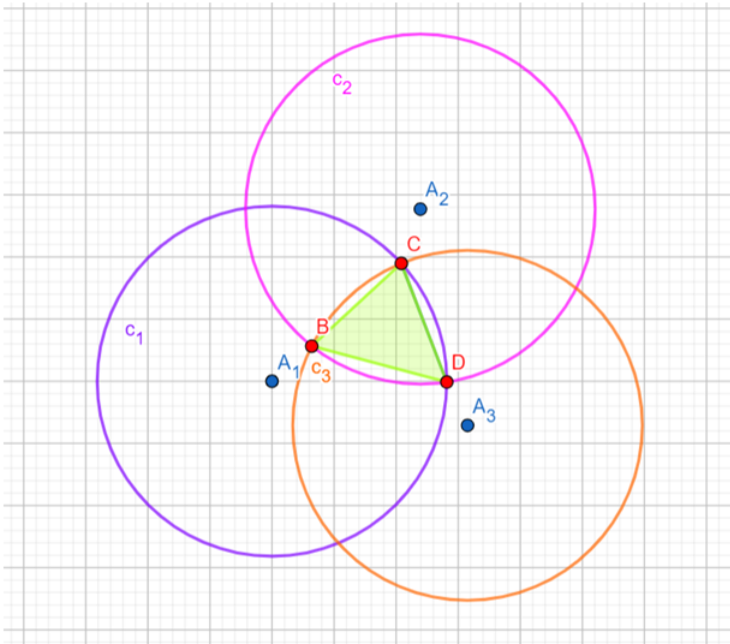


Figure 9: Three overlapping circles

Overlapping areas of multiple circles

For more than two circles, the total coverage area becomes more complicated to compute. If we add an extra circle to two circles that are already overlapping, we obtain something similar to Figure 9.

However, if we subtract the area that circle c_3 overlaps with circle c_1 and the area that circle c_3 overlaps with circle c_2, then the area where all three circles overlap, will be subtracted one time too many. In order to compensate that, the area where all three circles overlap, has to be added to obtain the actual coverage area of the three circles.

The area of overlap of the three circles can be computed similarly to the case where two circles overlap. The three distinct areas outside triangle BCD can be computed with the same method as illustrated in Figure 8. The only thing left to do is to add the area of the triangle BCD.

Note that it is possible that the area of overlap between circles c_1 and c_2 lies entirely inside circle c_3. In that case, we can repeat the same approach as in Figure 9, but the area of the three overlapping circles, is equal to the area of overlap of circles c_1 and c_2.

Also note that for more than three circles, the area of overlap of all circles can be computed by calculating the area of the polygon with vertices the intersection points of the circles. Then the only thing left to do is to add the area outside the polygon and inside the area of overlap.

If three circles overlap, we are adding the area of overlap, after subtracting every area of overlap between two circles to obtain the coverage area. If we do the same calculations for four overlapping circles, the common area of overlap of the four circles will be added two times too many. Therefore, we need to subtract the area where the four circles overlap two times to obtain the coverage area.

This process continues for more overlapping circles, if we consider the same calculations as done to lesser circles. For example, in the case that five circles overlap, we need to add three times the common area of overlap to obtain the coverage area. In the case that six circles overlap, we need to subtract four times the common area of overlap to obtain the coverage area.

For our model, we implemented the calculations up till four overlapping circles. This is due to the fact that in the simulation we use for the standard graph, at most three circles overlap, implying we only need to implement the calculations up till four overlapping circles, as we compute the coverage area when adding a single new base station, resulting in that at most four circles can overlap. Moreover, we assume that it is very unlikely that four base stations need to cover the same area. For these reasons we will not consider five or more overlapping circles.

To account for the area outside of the graph, we compute the area which need to be subtracted from the total area similarly to computing the area of overlap of multiple circles.

Method 3: optimizing coverage area

Considering again $n \times m$ possible coordinates where we can deploy a new base station, we compute for every position what the total coverage area results in and we deploy a base station at the position that results in the highest coverage area. This yields the graph which can be seen in Figure 10. For convenience, the coverage area of the newly deployed base station is coloured orange instead of green.

We call the algorithm to deploy a new base station based on the coverage area Method 3. The performance measures of the standard graph after applying Method 3 are given in Table 1, as well as Method 1, Method 2 and the standard graph. For the pseudo-code of Method 3, see Pseudo-code 3.

PSEUDO-CODE 3: Method 3: optimizing coverage area

```
For n×m coordinates add a base station
    Compute the coverage area
Deploy a base station at the coordinate which
results in the highest coverage area
```

C Failures

We suggested three different methods to deploy a new base station. In Section IV, we will simulate the standard graph for different positions of the users to verify which method is the best. However, failures could have an impact on the deployment strategy. Therefore, we also look at case where the critical distance of a base station decreases and at the case where base stations fail to connect to users with a certain probability.

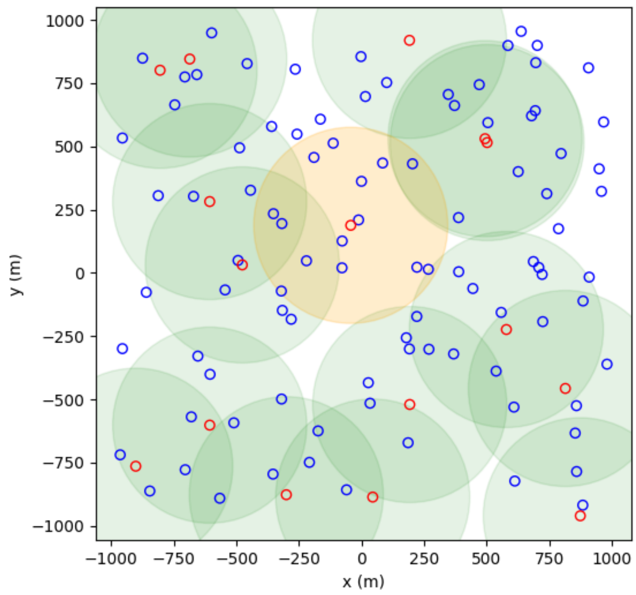


FIGURE 10: Adding a single base station when maximizing the total coverage area considering 60×60 possible positions.
One can see that an extra base station is deployed at $(-43, 188)$.

TABLE 1: Performance measures of the standard graph

	Average SNR (dB)	Disconnected users	Coverage area (m^2)
No methods	18.83	22	3303200
Method 1	21.42	11	3572000
Method 2	21.42	11	3572000
Method 3	21.10	13	3572800

Critical distance

It is possible that the critical distance of a base station decreases due to random events, for example, rainfall. We model this effect by multiplying the critical distance with a given factor smaller than 1.

Connection failure

If a base station is defect or has connection problems, it is possible that a base station cannot connect to users. In our model, we assume that every base station has a probability p that it fails to connect to users and the failures are independent between the base station. Resulting in a decrease of average SNR, connected users and coverage area.

IV Results

First of all, we look at the results of the graph generated in Figure 2. This figure shows the average SNR, the number of disconnected users and the coverage area. These results are shown in Table 1. For these specific positions of users, Method 1 and Method 2 deploy a base station, where the average SNR is the highest and the number of disconnected users is the lowest, while Method 3 deploys a base station where the coverage area becomes the highest. In this case, it is better to improve the average SNR and the number of connected users, as it is a short-term solution to the users that need a connection.
For this one specific simulation, Method 1 and Method 2 deployed the new base station at the same coordinate. In general, this is not the case, as for the deployment of a new base station, the highest average SNR and the highest number of connected users do not need to be at the same coordinate.

The coverage area is estimated by dividing the area in 100×100 squares and looking for each square separately if it is being covered or not. For the results we compute the coverage area numerically, as the analytical approach in Method 3 provides an area which is too low.

We compare the methods with each other. As one single simulation only tells us what to do in the case that the users are located at the specific positions, we use the same positions of the base stations where in every simulation the positions of the users are different.

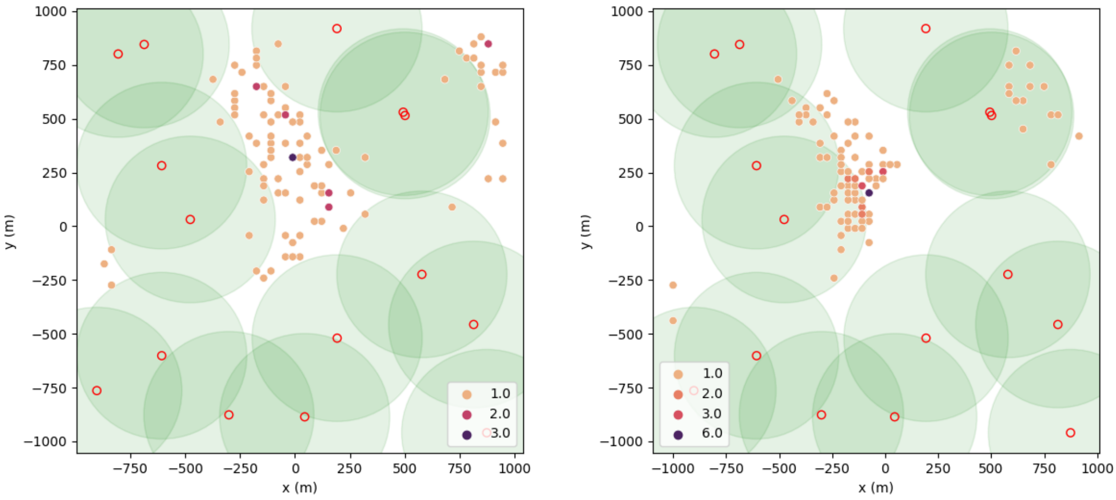
A Deploying a base station for different users

We look at the distribution of base stations as in the standard graph, see Figure 2. However, from now on, we will not use the previously generated PPP points for the users anymore, as we want to know where the new base station will be deployed if the users are located at a different positions. Therefore, we generate the users as a PPP with the same density, but instead of doing only one simulation, we do 100 simulations to obtain coordinates to deploy a base station by Method 1, Method 2 and Method 3.

We can generate the positions of users as PPP multiple times to see what the effect is on the deployment of a base station. This is not interesting for Method 3, as it does not depend on the positions of the users. However, if we run the simulation 100 times with the same positions of the base stations and for different realisations of user positions, we can see where the base station will be deployed if we are either using Method 1 or Method 2, see Figure 11.

For convenience, the coverage areas of the base stations of the standard graph are plotted, to give an intuition of where the base stations are getting deployed.

We see that the deployment of base stations in Method 1 is fairly distributed over the uncovered area, while the deployment of base stations in Method 2 is denser on a more specific area. At the middle-left part of the graph, there are in both methods some outliers. The users are generated as PPP after all, hence, it is possible that in some simulations there is a cluster at a corner in the plot. Most of the time, the base stations are getting deployed in the middle, with little deviation from the positions where a base stations would get the highest coverage area. It is remarkable that in Method 2, it relatively often happens that a base station is being deployed inside the radius of a base station where a



(a) Method 1. (b) Method 2.

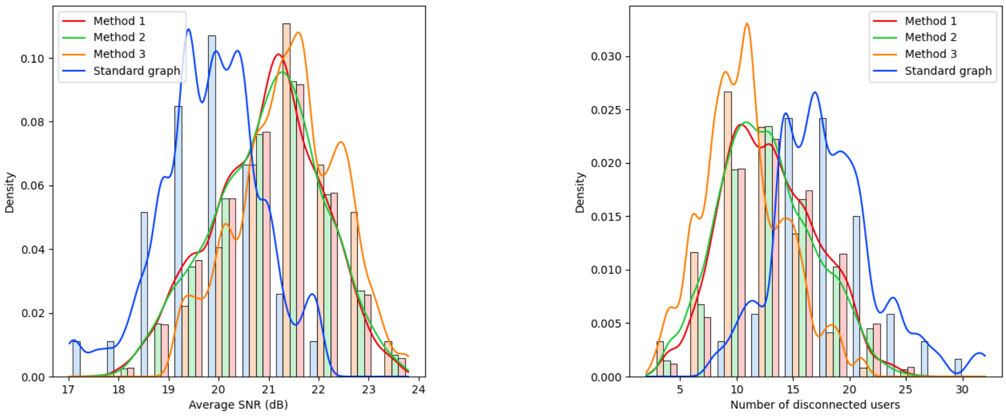
FIGURE 11: The positions of the users in the standard graph are simulated 100 times and the dots are the coordinates of the newly deployed base stations.

TABLE 2: Performance measures of simulating the standard graph with coordinates obtained in Method 1 against the coordinates obtained in Method 2 and Method 3 compared to the standard graph 100 times.

	Average SNR (dB)	Disconnected users	Coverage area (m^2)
No methods	19.85	17.8	3303200
Method 1	20.96	12.9	3485980
Method 2	20.99	12.6	3497840
Method 3	21.41	10.7	3572800

user can connect. This could be due to the fact that the algorithm in Method 2 chooses the first coordinate that has the least amount of disconnected users if there are multiple positions where the number of disconnected users is the lowest. Moreover, in almost every simulation, the coordinate where the number of connected users is the highest is not unique.

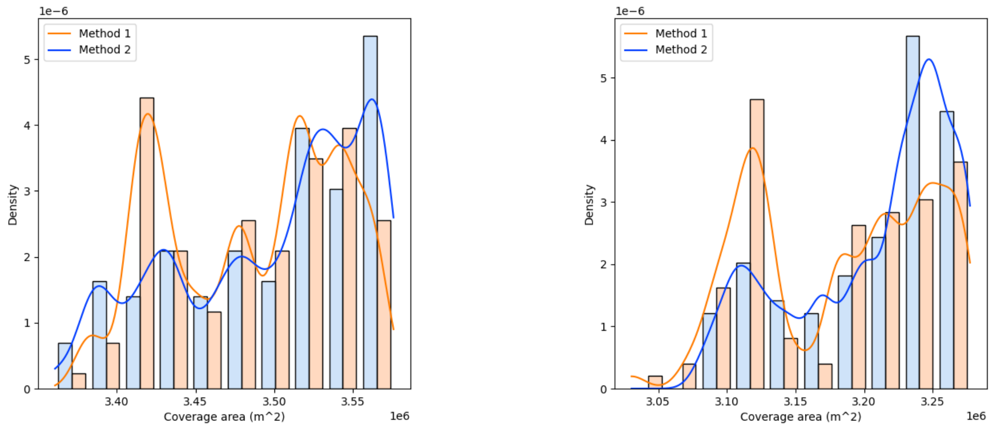
If we compute the performance measures of the 100 points obtained in Method 1, Method 2 and Method 3, as can be expected, Method 1 will always provide the highest average SNR, Method 2 will always provide the lowest number of disconnected users and Method 3 will always provide the highest coverage area. Therefore, with the 100 coordinates obtained in each method, we generate again 100 times the positions of the users, to see which method is best on average. The averages of the 100 simulations of the 100 points are given in Table 2 and a histogram of the density is given for the average SNR and the number of disconnected users in Figure 12 and a histogram for the density is given for the coverage area in Figure 13(a). For the coverage area, only Method 1 and Method 2 are plotted, as the coverage area of the standard graph and the coverage area of Method 3 do not change. The line plotted along the histograms is a kernel density estimate plot (KDE). A KDE uses a continuous probability density curve to represent the data.



(a) Density plot of the average SNR.

(b) Density plot of the number of disconnected users.

FIGURE 12: Performance measures of simulating the standard graph with coordinates obtained in Method 1 against the coordinates obtained in Method 2 and Method 3 compared to the standard graph 100 times.



(a) Density plot of the coverage area of Method 1 against Method 2 for the standard graph.

(b) Density plot of the coverage area of Method 1 against Method 2 for the standard graph with a reduced critical distance of 20%.

FIGURE 13: Coverage area of simulating the coordinates obtained in Method 1 against the coordinates obtained in Method 2 100 times.

Table 3: Performance measures of simulating the standard graph with coordinates obtained in Method 1 against the coordinates obtained in Method 2 and Method 3 compared to the standard graph 100 times. The standard graph has a reduced critical distance of 20%.

	Average SNR (dB)	Disconnected users	Coverage area (m^2)
No methods	18.11	25.4	3184008
Method 1	19.38	20.3	3184008
Method 2	19.48	19.8	3207780
Method 3	19.81	18.1	3278800

We conclude that Method 3 provides the highest coverage area, which is no surprise, otherwise the algorithm does not work. On the contrary, in every performance measure, Method 3 performs clearly better than Method 1 and Method 2, as the average SNR in Method 3 is higher than methods 1 and 2 and the number of disconnected users is lower. Additionally, Method 2 outperforms Method 1 just slightly in every performance measure.

B Deploying base stations considering failure of base stations

Method 3 performs the best under normal circumstances. However, we want to know if our base station deployment strategy changes if the critical distance decreases and how failure of base stations affects the deployment strategy.

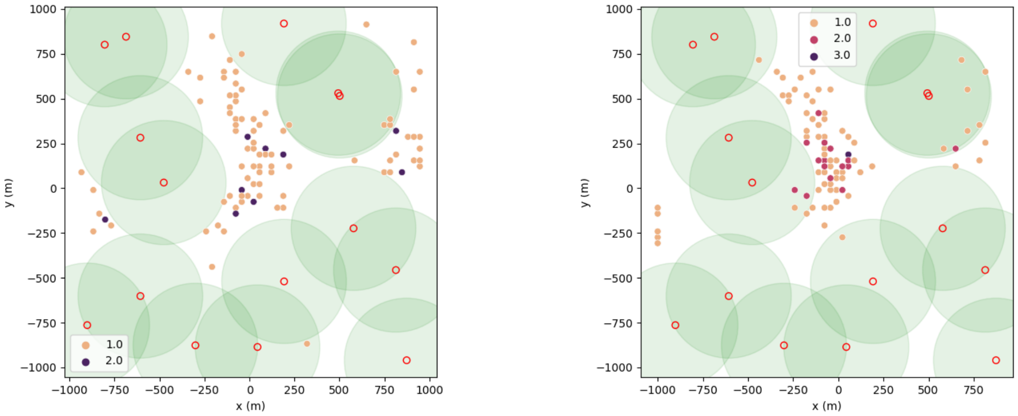
Reduced critical distance

If due to rainfall or another random event the critical distance decays, it could influence our deployment strategy. We assume that the critical distance decreases by 20%, which results in a decay of the radius where users can connect in with approximately 10%, see Equation (5). Using the same approach as for the standard graph, we can simulate the standard graph with a reduced critical distance 100 times to obtain 100 coordinates, see Figure 14. Simulating this 100 points 100 times provides the results in Figures 15 and 13(b) and the averages in Table 3.

As the total coverage area is less than in the standard graph, the coordinates where to deploy base stations tend to move towards the center of uncovered areas, as the base stations are able to cover there more users in the long run. This causes that the difference in performance measures of the three methods becomes smaller compared to the graph under normal circumstances. Nonetheless, Method 3 is still significantly better than Methods 1 and 2, as all performance measures are better for Method 3, as shown in Table 3. Moreover, Method 2 is again slightly performing better than Method 1.

Random failures

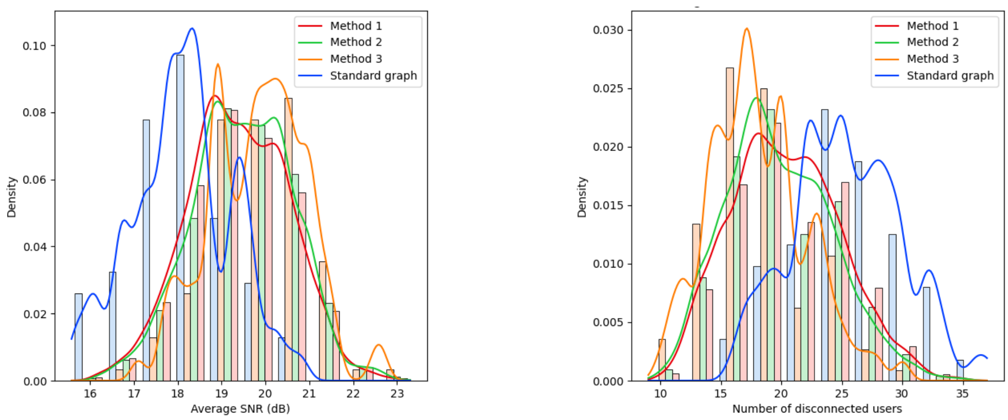
Method 3 is in general under normal circumstances and in case of reduced critical distance outperforming Methods 1 and 2 qua performance measures, but what happens if each base station has a certain probability that it fails to connect with users? We want to simulate the graph under normal circumstances, but with random failures to see which Method performs best in case of failures.



(a) Method 1.

(b) Method 2.

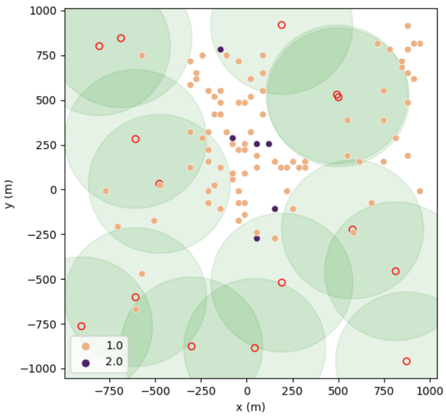
FIGURE 14: The positions of the users in the standard graph are simulated 100 times and the dots are the coordinates of the newly deployed base stations. It is assumed that rain causes that the graph has a reduced critical distance of 20%.



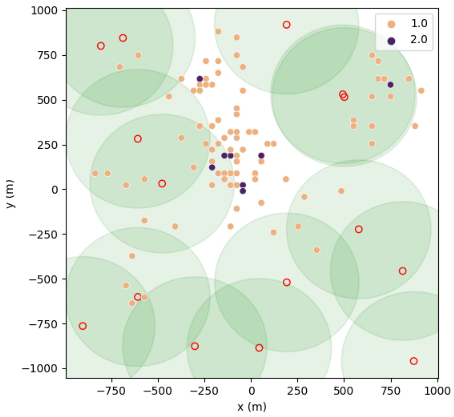
(a) Density plot of the average SNR.
The standard graph has a reduced critical distance of 20%.

(b) Density plot of the number of disconnected users.
The standard graph has a reduced critical distance of 20%.

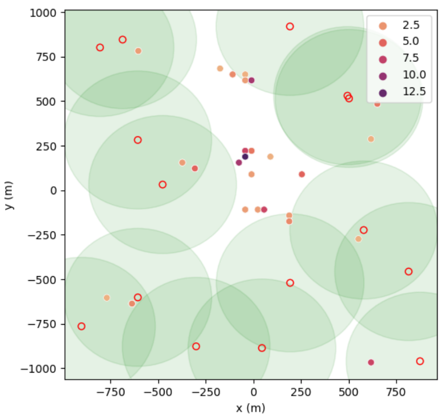
FIGURE 15: Performance measures of simulating the standard graph with coordinates obtained in Method 1 against the coordinates obtained in Method 2 and Method 3 compared to the standard graph 100 times. It is assumed that rain causes that the graph has a reduced critical distance of 20%.



(a) Method 1.

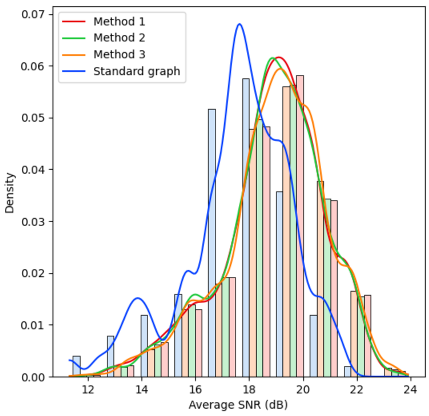


(b) Method 2.

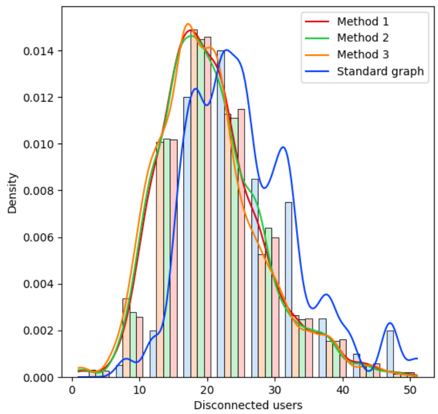


(c) Method 3.

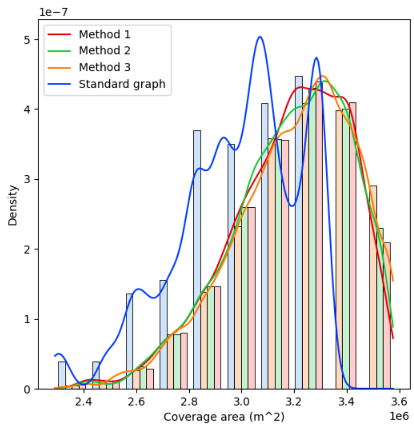
FIGURE 16: The positions of the users in the standard graph are simulated 100 times and the dots are the coordinates of the newly deployed base stations. Each base station in the standard graph has a probability of 20% that it fails.



(a) Density plot of the average SNR.
Each base station in the standard graph has a probability of 20% that it fails.



(b) Density plot of the number of connected users.
Each base station in the standard graph has a probability of 20% that it fails.



(c) Density plot of the coverage area.
Each base station in the standard graph has a probability of 20% that it fails.

FIGURE 17: Performance measures of simulating the standard graph with coordinates obtained in Method 1 against the coordinates obtained in Method 2 and Method 3 compared to the standard graph 100 times. Each base station in the standard graph has a probability of 20% that it fails.

Table 4: Performance measures of simulating the standard graph with coordinates obtained in Method 1 against the coordinates obtained in Method 2 and Method 3 compared to the standard graph 100 times. Each base station in the standard graph has a probability of 20% that it fails.

	Average SNR (dB)	Disconnected users	Coverage area (m^2)
No methods	17.58	25.6	2989220
Method 1	18.88	20.5	3185110
Method 2	18.89	20.4	3189172
Method 3	18.96	20	3203134

We assume that each base station in the standard graph has a probability of 20% that it fails to connect. In that case, it is also interesting to collect coordinates for Method 3, as the coverage area changes per simulation. Doing so yields the coordinates that are plotted in Figure 16. Method 1 and Method 2 produce similar coordinates compared to the graph under normal circumstances, but they fluctuate a lot more due to the failures of base stations. Conversely, Method 3 produces coordinates that fluctuate a little, but every possible coordinate to deploy a new base station is chosen at least twice to deploy a new base station according to Method 3. It is remarkable that in the right bottom corner of Figure 16(c), a base station is deployed ten times. This can be justified, as in the case the base station where the point lies in fails, together with another base station nearby, then deploying a new BS there will increase the coverage area more than a central point.

The results of simulating the 100 obtained points 100 times are plotted in Figure 17 and the averages of the performance measures are given in Table 4. Method 1 and Method 2 are performing even closer, although Method 2 is still slightly better in every performance measure. The performance measures of Method 3 are also closer to the performance measures of Methods 1 and 2. Nevertheless, Method 3 is still performing better in every performance measure than Methods 1 and 2.

V Discussion

In this paper we modelled a 5G wireless network as a random geometric graph where the positions of users and base stations were modelled as Poisson point processes.

We assumed that PPP of the users and of the base stations had the same density over the whole area, while in real life you often see that at specific places, more people are present and more connections are needed, resulting in a higher density on specific areas and a lower density elsewhere.

The density of users could also affect the deployment strategy. A higher density of users will cause the performance measures of the 3 methods to become closer, while a lower density will probably cause higher fluctuations in performance measures, which also would be interesting to investigate.

For the calculation of the coverage area, we used a numerical computation, as the analytical computation in Method 3 failed to provide the correct coverage area. The subtraction of the areas outside the square could have caused this error. The overlap of circles outside the plot has the same issue as the overlap of circles inside the plot. In Method 3, we did not account for the overlap outside the plot. We tried to compute the coverage area while adding the overlap of two overlapping circles, but this still resulted in a coverage area that is too low. Hence, we decided to neglect the overlapping areas of circles outside the plot for Method 3 and we computed the coverage area numerically for the performance measures. We verified that the algorithm in Method 3 deploys a base station at the same coordinate if we used a numerical approach in the algorithm, which is indeed the case.

We could also look at what happens if we try to deploy multiple base stations at the same time or one by one. We did write an algorithm to deploy two base stations at the same time and an algorithm to deploy two base station one by one based on Methods 1, 2 and 3. We wanted to compare the two methods, however, deploying two base stations at the same time took way too long in terms of computational time, as we did not have to compute for $n \times m$ possible coordinates, but for $(n \times m) \times (n \times m)$ possible coordinates.

An additional thing that could affect the deployment strategy is that in Method 2, the coordinates for the number of disconnected users are not unique. In other words, at multiple coordinates the number of disconnected users is the lowest. In our algorithm we did choose the very first coordinate of the coordinates with the lowest number of disconnected users. As we started to compute at the left bottom at the plot, the newly deployed base stations tend towards the left-bottom of the plot, which could influence the performance measures.

We assumed that a user connects to the base station that provides him the highest SNR if that user has a base station nearby that provides a SNR higher than the threshold. We did not consider users that can connect with multiple base stations if needed to deploy a new base station under normal circumstances. We could change the algorithms to deploy new base station based on users who could connect to multiple base stations, that those users do not need an improvement of the SNR.

We also assumed that every base station has the same transmission power and for every user that connects to a particular base station, the same transmission power is used. However, if multiple user connect to the same base station, the base station cannot provide the same power to each user as if only one user connects. So for our model we did not consider that the transmission power of base stations decreases if more users connect, which could definitely influence the performance measures.

Every base station is assumed to be the same, whereas not every base station has the same transmission power and coverage area. It is possible that a base station has a very high transmission power, but for only a short distance or a base station has a relatively low transmission power, but has a very widely range. Not only considering a graph with different base stations could influence the outcome of the deployment strategies, but also considering different base stations to deploy. Especially as base stations are expensive and in some cases it is more advantageous to deploy a cheaper base station, as an expensive base station is unnecessary.

In future research, we can look more to probabilities that base stations fail or that the critical distance decreases. For this research we assumed every base station has a fixed probability that it fails independent of other base stations, but in reality it could be the case that in a particular area something is wrong with the base stations. Moreover, random events such as rain are not uniformly distributed over the whole area. Apart from this, the probabilities that a user cannot connect due to for example a truck blocking the line of sight or a building that stands in front of the base station, could also be simulated in the model self instead of computing probabilities a user cannot connect.

Altogether, there are several factors that could influence the results and the model can be improved on some assumptions. In future research we could refine assumptions and add more complexity to the model. Also probabilities could be used more instead of solely simulating different deployment strategies and obtaining performance measures.

VI Conclusion

In this paper we determined three methods to deploy base stations and we simulated the model to obtain performance measures of each method. These methods are based on maximizing the average SNR, maximizing the number of connected users and maximizing the total coverage area for a given scenario. We found that in the long run under normal circumstances it is best to deploy a new base station at the position that maximizes the coverage area of all base stations. Considering the two other deployment strategies, it is slightly better to try to increase the number of connected users than to maximize the total average SNR. Failures causing a decrease in critical distance or causing base stations to fail to connect to users cause the performances of the three methods to become closer, while it is still best to try to maximize the coverage area when deploying a new base station. To conclude, we have seen that it is best to deploy base stations at the position that results in the highest coverage area in order to minimize the probability of having no internet connection in a 5G network.

References

[1] Salix alba (https://math.stackexchange.com/users/139342/salix alba). How can i find the points at which two circles intersect? Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/1033561 (version: 2014-11-22).

[2] Wenming Cui, Yuan Zhang, Lei Wei, and Min Lu. Optimizing base stations deployment in wireless access networks for smart grids. In *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, pages 1–5. IEEE, 2015.

[3] Paul Erdös. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.

[4] H. T. Friis. A note on a simple transmission formula. *Proceedings of the IRE*, 34(5):254–256, 1946. Cited By :1389.

[5] Edward N Gilbert. Random plane networks. *Journal of the society for industrial and applied mathematics*, 9(4):533–543, 1961.

[6] Martin Haenggi. *Stochastic geometry for wireless networks.* Cambridge University Press, 2012.

[7] Martin Haenggi, Jeffrey G Andrews, François Baccelli, Olivier Dousse, and Massimo Franceschetti. Stochastic geometry and random graphs for the analysis and design of wireless networks. *IEEE journal on selected areas in communications*, 27(7):1029–1046, 2009.

[8] Hazer Inaltekin, Mung Chiang, H Vincent Poor, and Stephen B Wicker. On unbounded path-loss models: effects of singularity on wireless network performance. *IEEE Journal on Selected Areas in Communications*, 27(7):1078–1092, 2009.

[9] Keith Kendig. Is a 2000-year-old formula still keeping some secrets? *The American Mathematical Monthly*, 107(5):402–415, 2000.

[10] Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

[11] Tapan K Sarkar, Zhong Ji, Kyungjung Kim, Abdellatif Medouri, and Magdalena Salazar-Palma. A survey of various propagation models for mobile communication. *IEEE Antennas and propagation Magazine*, 45(3):51–82, 2003.

[12] Martin R Stiglitz and Julie Callahan. Antenna engineering handbook. *Microwave Journal*, 36(10):169–170, 1993.

[13] Xinchen Zhang and Jeffrey G Andrews. Downlink cellular network analysis with multi-slope path loss models. *IEEE Transactions on Communications*, 63(5):1881–1894, 2015.