Characterisation of a Connectivity Measure for Sparse Wireless Multi-hop Networks

Srinath Perur and Sridhar Iyer
KReSIT, IIT Bombay
Mumbai–400076
India
Email: [srinath, sri]@it.iitb.ac.in

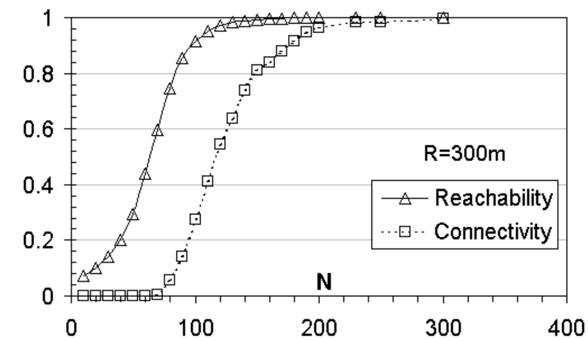
Abstract

The extent to which a wireless multi-hop network is connected is usually measured by the probability that all the nodes form a single connected component. This measure is called connectivity. We find this unsuitable for use with sparse networks since it is not indicative of the actual communication capability of the network, and can be unresponsive to changes in network parameters. We propose an alternative measure called reachability, defined as the fraction of node pairs in the network that are connected. We claim that it is more intuitive and expressive than connectivity when dealing with sparse networks. We obtain analytical expressions for reachability for two and three nodes in the static case. We identify reachability as growing according to the logistic growth model and present a regression model for reachability in terms of number of nodes and normalised transmission range. This model is applicable for static networks with up to 500 nodes. We also extend this model to larger networks using an approximation. These characterisations of reachability can be used by a network designer to estimate the trade off between how connected the network is, the number of nodes, the area of operation, and transmission range of nodes.

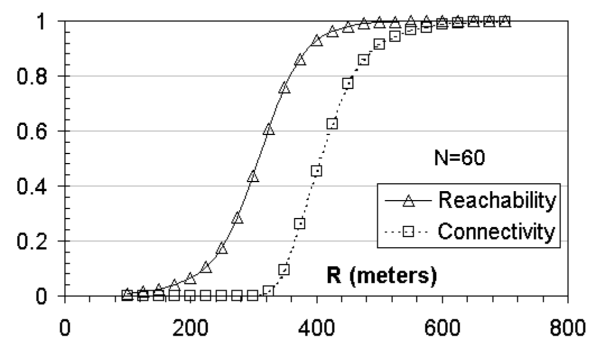
1 Introduction

The extent of communication that a wireless multi-hop network can support is a matter of interest for designers and users of the network. In a wireless multi-hop network, a fundamental limiting factor is the absence or presence of routes between nodes. It is beyond this that factors like channel capacity and interference affect the extent of communication. A popular measure for determining the degree to which a wireless multi-hop network is connected is the probability that the network graph forms a single connected component. This probability is called *connectivity*, and has been extensively studied.

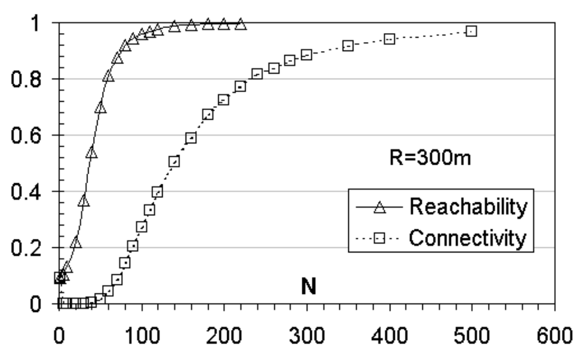
A sparse wireless multi-hop network is one in which connectivity with high probability is not ensured. For our purposes, we call a network sparse if its connectivity is less than 0.95. Such a network can arise in various ways: a vehicular ad hoc network in an area with low traffic density, an initially connected sensor network after some of its nodes have failed, and an ad hoc communications network that is being deployed incrementally can all be sparse networks. Occasionally, in a constrained deployment scenario, we may even wish to deploy a multi-hop network that trades off connectivity for cost. In such sparse networks, we claim that using the probability of connectivity as a design metric can prove inadequate because i) connectivity is not indicative of the actual extent to which the network can support communication; and ii) it is unresponsive to fine changes in network parameters. For example, it is possible that a sparse network that allows a significant number of nodes to communicate has a connectivity close to 0. Further, an increase in some network parameter such as number of nodes, or transmission range, may increase the ability of nodes to communicate, but it may not be reflected by a corresponding increase in connectivity. We believe that a property of the network graph better suited for use with sparse networks is the *fraction of node pairs that are connected*. We call this quantity *reachability*. We consider both connectivity and reachability to be different *connectivity measures* of a network graph.



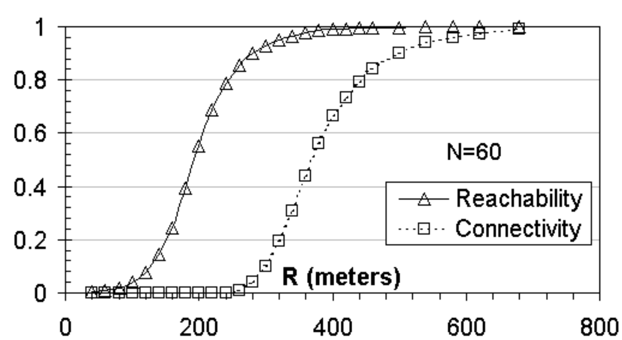
(a) Increasing N, no mobility



(b) Increasing R, no mobility



(c) Increasing N, with mobility



(d) Increasing R with mobility and asynchronous communication

Figure 1. Reachability and Connectivity growth curves

To illustrate the behavior of reachability and connectivity in a sparse multi-hop wireless network, consider this simulation study from [15] for a sparse rural deployment: *N* nodes, each with a uniform radio transmission range of *R* meters, are to be deployed in a 2000m x 2000m area. The nodes are distributed uniformly at random in the area of operation. Figure 1(a) shows growth curves for connectivity and reachability for increasing *N* when $R = 300m$. The connectivity curve remains at 0 for this network till *N* grows to almost 70. But even with fewer than 70 nodes, the network affords a significant degree of communication as can be seen from the corresponding reachability curve. For example, with 60 nodes, 45% of node pairs have a multi-hop path connecting them. Reachability also lets us quantify the difference between having, say, 30 nodes and 50 nodes: 14% of node pairs can communicate when there are 30 nodes, and 30% when there are 50 nodes. Similar observations can be made from Figure 1(b) which plots the growth of reachability and connectivity for 60 nodes as R increases. Low connectivity in sparse networks is often handled by exploiting mobility and asynchronous communication between nodes. Using asynchronous communication, two nodes that may never have had a path between each other at a single instant, can communicate using store-and-forward arrangements with other nodes. Figure 1(c) shows an increased difference between reachability and connectivity when mobility is introduced in the scenario of Figure 1(a). Figure 1(d) shows the comparison when asynchronous communication is enabled in the scenario of Figure 1(b). In this case, almost 80% of node pairs are able to communicate before connectivity increases from 0. Details regarding these simulations and a case-study in the use of reachability for designing sparse multi-hop wireless networks can be found in [15].

Our main contribution in this paper is the characterisation of reachability for wireless multi-hop networks: we present analytical expressions for small cases (Section 5), identify reachability as consistent with the *logistic growth model* (Section 6.2), and obtain a closed form expression through regression analysis of simulated data that is valid for values of N from 2 to 500 when nodes are static and distributed uniformly at random (Section 7). We also extend this model to be applicable to larger networks within a bounded factor of error (Section 8).

2 Related Work

Gupta and Kumar showed in [4] how throughput per source-destination pair in a multi-hop wireless network decreases as node density increases. Grossglauser and Tse in [3] showed that in the presence of mobility, multi-user diversity could be used to achieve a trade off between throughput and delay. This would allow throughput to be maintained almost constant even with increasing node density. A further quantification of this trade off is found in [12]. Similarly, a trade off has also been achieved between *connectivity* and delay. Delay tolerant routing [6] and Message Ferrying [24] are representative examples of work that studies the use of node mobility to achieve asynchronous communication between disconnected nodes in sparse networks. This background forms the context for our work on reachability by motivating the need for connectivity measures that better express the communication capabilities of a network, and allow fine-grained evaluations of trade offs.

It is shown in [20] that forming a connected component with 90% of the nodes in the network requires a much lower minimum transmission range than for complete connectivity. A number of sparse sensor network applications are described in [18]. Regression has been used in [8] to model the threshold transmitting range for k-connectivity in wireless multi-hop networks, and in [22] to model the probability of obtaining a fully connected network. [22] also defines a *connectivity index* in terms of number of nodes in each connected component, that is identical with reachability. However, no model or characterisation is offered. We use the term *reachability* since it is more intuitive in the context of communication in sparse networks. Although this term has been used before in several other areas, to the best of our knowledge it has not been used to denote a connectivity measure.

The characterisation presented in this paper has earlier appeared in a shorter form in [14]. An introduction to reachability and an overview of its role in designing sparse wireless multi-hop networks can be found in [16] and [13] respectively.

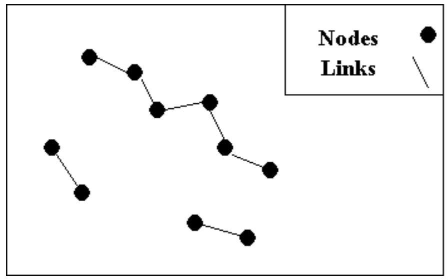


Figure 2. A network instance with Reachability = 0.378

3 Reachability

The reachability of a static network is defined as the *fraction of connected node pairs* in the network. It is a property of the network graph, with no assumptions made regarding the distribution of nodes. Using this definition we can calculate reachability for a network of *N* nodes as:

$$Reachability = \frac{\text{No. of connected node pairs}}{\binom{N}{2}} \tag{1}$$

A pair of nodes is considered connected if there is a path of length one or greater between them. Figure 2 shows one instance of a network with 10 nodes. We count the number of node pairs that can reach each other, that is, nodes that are connected either directly or through other nodes, as 17. Substituting $N = 10$ in the denominator of Equation 1, we obtain the reachability for this network instance as $17/45$ or 0.378.

Note that for the same 10 nodes, it is possible to have a different value of reachability in another instance. We define a *network* by the number of nodes, their bounding area, and the transmission range of the nodes. The same network can have different *instances* depending on how the nodes are arranged. The network's reachability can be measured as the average of reachabilities across several instances of that network. This value is significant since it represents the probability that a pair of nodes chosen randomly from the network are connected. Note that the procedure for measuring reachability is analogous to that for connectivity: a single instance of a network is either fully connected or not, and connectivity is measured as the fraction of a large number of network instances that are connected. When nodes are mobile, the fraction of connected node pairs varies depending on node movement, but a single value can be obtained for any time *instant*. We can measure reachability for a mobile network as the *average of instantaneous reachability values* measured at frequent intervals during the operation of the network. When a theoretical value of reachability is known for a mobile network, the measured reachability is expected to converge to this value with time.

4 Network model and notation

The network model we use for our characterisation of reachability is as follows: *N* nodes are distributed uniformly at random in a square area of side *l*; two nodes can communicate directly with each other if the distance between them is not greater than *R*, the transmission range of every node. Since the network graph remains unchanged when *R* and *l* vary proportionally, we combine the two into a normalised transmission range, $r = R/l$ without loss of generality. While this model takes a simplistic view of radio propagation, it promotes better defined behavior of topological properties, and is useful for an initial study. Note that the assumptions regarding the network model help in *characterising* reachability. The *definition* of reachability is independent of these assumptions. For a network with *N* nodes, normalised transmission range *r* and a set of mobility model parameters *M* in a cube of d dimensions, we denote the corresponding value of reachability as $Rch_{N,r}^{M,d}$. In this work, since we deal only with characterisation of the static case, we use the notation $Rch_{N,r}^{d}$. In the case of most interest, when $d = 2$, we drop the superscript altogether for convenience and write $Rch_{N,r}$.

5 Analysis of reachability for small cases

In this section we derive closed form expressions for reachability of two and three static nodes located uniformly at random along a line of length l: $Rch_{2,r}^{1}$ and $Rch_{3,r}^{1}$.

5.1 $Rch_{2,r}^{1}$

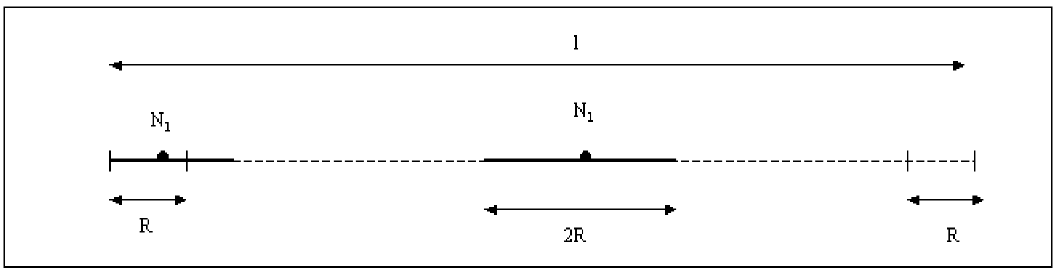


Figure 3. Positions of a node on a line segment

Let N_1 and N_2 be two nodes that can take positions uniformly at random on a line of length l. For instances of this network, $Rch_{2,r}^{1}$ is 1 in cases when they are connected and 0 when they are not. The reachability for this network is therefore equivalent to the probability that two nodes with transmission ranges R are connected when they are distributed randomly on a segment of length l. (As this implies, reachability and connectivity are identical for a network with two nodes.)

We define the *coverage* of a node as the length of the line segment that is covered by the transmission range of the node. The probability that N_1 and N_2 are connected is given by the fraction of the length l covered by N_1:

$$Rch_{2,r}^{1} = \frac{Coverage(N_1)}{l} \tag{2}$$

We first consider the case when $l \geq 2R$. As seen in Figure 3, the coverage of N_1 varies depending on where it is positioned on the line segment. The coverage of N_1 is $2R$ if it is not within a distance R from either end point of the line segment. If it is placed in one of the edge segments of length R, its coverage on one side would remain R, while the coverage on the other side would vary between 0 and R. Considering all positions along the edge segments equally likely, the coverage of N_1 in an edge segment is R for the side away from the edge, and the expected coverage is $\frac{R}{2}$ for the side near the edge[1]. Therefore, the total expected coverage of N_1 on an edge segment of length R is $\frac{3R}{2}$, and the total coverage of N_1 in the middle segment of length $l - 2R$ is $2R$. The expected coverage of N_1 when it takes any position on the line of length l is given by:

$$\begin{aligned} Coverage(N_1) &= \frac{2R}{l}.\frac{3R}{2} + \frac{l-2R}{l}.2R \\ &= \frac{2Rl - R^2}{l}, \quad (l \geq 2R) \end{aligned}$$

For the case when $2R > l > R$, we divide the line of length l into three segments of lengths $l - R$, $2R - l$ and $l - R$. When N_1 is located in the central segment of length $2R - l$, coverage is 1 because N_1's transmission range extends beyond the end-points on either side. When N_1 is located on either of the edge segments of length $2R - l$, it extends to a length R on the side of the farther end-point. On the side of the nearer end-point, N_1's coverage is between $l - R$ and, when it is exactly on the end-point, 0. The expected value for coverage on the side of the nearer endpoint is $(l - R)/2$. Therefore, when $2R > l > R$,

$$\begin{aligned} Coverage(N_1) &= 2.\Big(\frac{l-R}{l}\Big).\Big(R + \frac{l-R}{2}\Big) + \frac{2R-l}{l}.1 \\ &= \frac{2Rl - R^2}{l}, \quad (2R > l \geq R) \end{aligned}$$

Since the coverage is the same for both cases, we can write

$$Coverage(N_1) = \frac{2Rl - R^2}{l}, \quad (l > R)$$

Substituting in Equation 2:

$$\begin{aligned} Rch^1_{2,r} &= \frac{2R}{l} - \frac{R^2}{l^2}, \quad (l \geq R) && (3) \\ &= 2r - r^2, \quad (r \leq 1) && (4) \end{aligned}$$

5.2 $Rch^1_{3,r}$

Finding $Rch^1_{3,r}$ is more involved than finding reachability for two nodes. We start by enumerating the different ways in which three nodes can be positioned in one-dimension:

A. All three nodes are isolated

B. One node is isolated and two are connected

C. All three nodes are connected with one intermediate hop

D. All three nodes are directly connected to each other

[1] If c is a random variable representing coverage on the side near the edge, the expected coverage when the node is located in the edge segment of length R is $\frac{1}{R}\int_0^R c\,\mathrm{d}c$ or $\frac{R}{2}$.

In case A, $Rch^1_{3,r}$ is 0 by definition. In case B, it follows from our definition of reachability (Equation 1) that $Rch^1_{3,r}$ is $\frac{1}{3}$. Similarly, in cases C and D, $Rch^1_{3,r}$ is 1. Since any of these cases is possible, the sum of $Rch^1_{3,r}$ for all possible cases after weighting with the probability of occurrence of each case gives us the expected value of $Rch^1_{3,r}$:

$$Rch^1_{3,r} = \frac{1}{3}P(B) + P(C) + P(D)$$

Since the cases A, B, C, and D are exhaustive, $P(A) + P(B) + P(C) + P(D) = 1$. We can now rewrite the above equation as:

$$Rch^1_{3,r} = \frac{1}{3}[1 - P(A) + 2P(C) + 2P(D)] \tag{5}$$

Derivations for $P(A)$, $P(C)$ and $P(D)$ follow the general approach used for $N = 2$, and are given in Appendix A. For the case $l \geq 4R$ we get:

$$P(D) = (2r - r^2)\left(\frac{3r}{2} - \frac{3r^2}{8}\right) \tag{6}$$

$$P(C) = (2r - r^2)(r - r^2) + (2r - 3r^2)\left(\frac{r}{2}\right) \tag{7}$$

$$P(A) = (1 - 4r + 4r^2)(1 - 4r + 2r^2) + (2r - 3r^2)\left(1 - \frac{7r}{2} + \frac{14r^2}{3}\right) \tag{8}$$

Having obtained expressions for $P(A)$, $P(C)$, and $P(D)$, we can substitute for these in Equation 5 to obtain $Rch^1_{3,r}$.

The expression obtained for $Rch^1_{2,r}$ also gives the probabilistic *connectivity* for two static nodes uniformly distributed on a line segment. This can be understood as follows: with two nodes, every network instance has either one or zero connected node pairs, resulting in a reachability of either one or zero for that instance. Each instance with a reachability of one also is completely connected since there are only two nodes. Therefore, the values of reachability and connectivity are identical for $N = 2$. When $N \geq 3$, there can be connected node pairs in an instance without all nodes being part of a single component. Therefore, reachability is always greater than connectivity for networks with $N \geq 3$.

6 Modelling $Rch_{N,r}$ in the finite domain

The above analysis for $N = 3$ involves the handling of multiple cases, and is significantly more involved than the analysis for $N = 2$. It is evident that this method of analysis is impractical for use with larger values of N, and other methods will have to be explored to characterise reachability. There is work that gives asymptotic probabilistic bounds on *connectivity* in a one-dimensional network by characterising the conditions required for a single node to be left out of the connected component [19, 21]. Such an approach is difficult to use with reachability since the metric by definition tries to capture communication capabilities in a network that can be separated by disconnections. In any case, asymptotic results for one dimensional networks, while of theoretical interest, are unlikely to be of practical use in networks with smaller numbers of nodes.

If the *N* nodes form *k* components with m_i nodes in the i^{th} component, we can rewrite Equation 1 as

$$Rch_{N,r} = \frac{\sum_{i=1}^{k} \binom{m_i}{2}}{\binom{N}{2}} = \frac{\sum_{i=1}^{k} m_i(m_i-1)}{N(N-1)} \tag{9}$$

It may be possible to use results for number of components and distributions of nodes for a Random Geometric Graph [11] to obtain asymptotic bounds (as *N* tends to infinity) for $Rch_{N,r}$.

Because sparse networks often involve small numbers of nodes, we are particularly interested in characterisations of $Rch_{N,r}$ in the finite domain. Since we could generate accurate data for $Rch_{N,r}$ from simulations, we decided to obtain a finite domain characterisation using empirical regression.

6.1 Empirical modelling of $Rch_{N,r}$

We explored data from simulations to see if reachability obeyed any known growth models. We studied the relationship between *r* and $Rch_{N,r}$ for various values of *N*. We chose *r* as our independent variable since it is continuous and allows greater flexibility in choice of data points. $Rch_{N,r}$ grows from zero at $r = 0$ and reaches an asymptote of one for some value of *r*. We visually explored several growth models consistent with this behavior, and found that the logistic growth model consistently fit the simulated data for a wide range of *N* and *r*. Among models we considered and rejected were power law models, sum of exponentials, the Gompertz model, and various logarithmic functions [17].

6.2 The Logistic Growth Curve

The logistic model is often used to fit sigmoidal curves with a lower asymptote of zero and a finite upper asymptote. Its most popular application has been in modeling the growth of populations over time. Intuitively, logistic growth models a system that grows rapidly beyond a threshold, and slows down as it approaches its maximum limit. Figure 4 shows a logistic curve expressed by the equation:

$$y = \frac{k}{1 + e^{\alpha - \beta x}} \tag{10}$$

where *k* is the limiting value that *y* can take, β is the maximum rate of growth, and α is a constant of integration [7]. The curve is skew-symmetric and has a point of inflexion at $x = \alpha/\beta,\ y = k/2$, where the growth rate is maximum [17].

We use the logistic equation to model the growth of $Rch_{N,r}$ as *r* increases for a fixed value of *N*. Since the maximum value of reachability is one, it becomes our upper asymptote. α and β vary with *N*, and we denote them by α_N and β_N. We use Equation 10 in the form:

$$Rch_{N,r} = \frac{1}{1 + e^{\alpha_N - \beta_N r}} \tag{11}$$

Figure 5 shows the close correspondence between simulated data and Equation 11 for the case $N = 100$. The values of α_{100} and β_{100} used were 9.58 and 79.2 respectively. We see how these values were obtained in Section 7.

7 Simulation and Regression Modeling

After having identified reachability as consistent with the logistic model, our approach towards characterising $Rch_{N,r}$ was as follows:

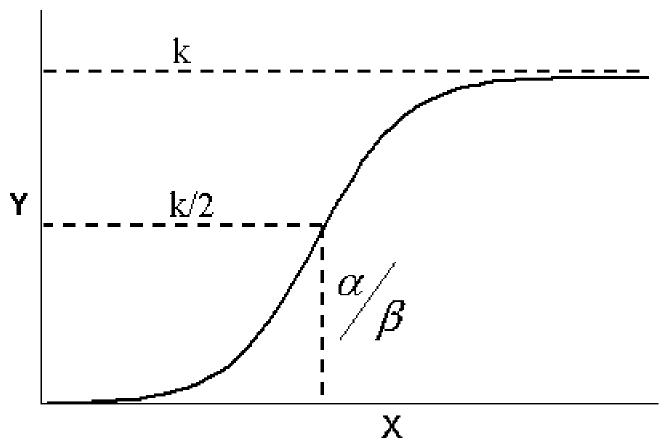


Figure 4. A general logistic curve

- We conducted extensive simulations to obtain data that represented the growth of $Rch_{N,r}$ from 0 to 1 as *r* increased, while keeping *N* fixed.
- We used Equation 11 as a regression function for simulated data, and obtained the coefficients α and β for the corresponding value of *N*. This allowed us to characterise reachability as a function of *r* for one value of *N*.
- We repeated the above two steps for values of *N* ranging from 2 to 500, and performed a second level of regression on the estimated values of α_N and β_N. This gave us a set of equations that allows us to obtain reachability as a function of *N* and *r* for values of *N* ranging from 2 to 500.

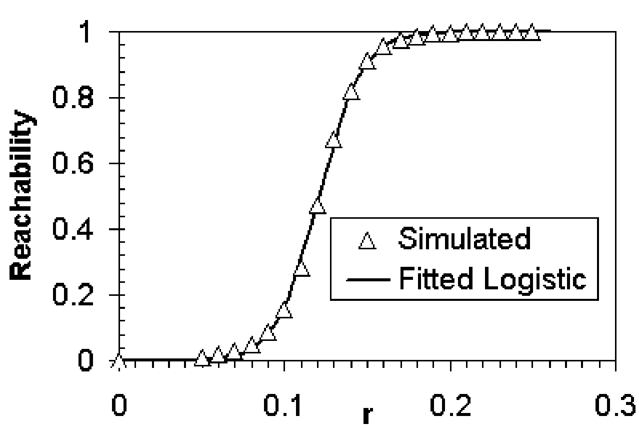


Figure 5. Logistic fit for N=100

7.1 Simulations

We conducted extensive simulations to generate the data required for fitting the regression function. Since we were looking to characterise reachability for small to medium sized networks, we chose, as representative points, 55 values of *N* between 2 and 500. For each of these values of *N*, we varied *r* in increments from zero to a value where reachability was at its maximum value of one. For each such value of *r*, we conducted simulations over 1000 randomly generated network graphs and calculated the mean value of $Rch_{N,r}$ across those instances. We know that the error of the mean is within $1.96s/\sqrt{n}$ with 95% confidence where s is the standard deviation of the samples, and n is the number of samples [5]. A worst case bound for s would be the case when the samples are uniformly distributed in the interval $[0, 1]$. The variance for a uniform continuous distribution in the interval $[a, b]$ is given by $(b - a)/12$ [23]. The worst case standard deviation for the interval $[0, 1]$ is given by $s = \sqrt{1/12} = 0.2887$. Using this value of s, and with $n = 1000$ we find that the error in the mean is within 0.018 with a confidence of 95%. At the end of our simulations, we had 55 tables each containing *r* and reachability values for the corresponding value of *N*. For illustration, one of these tables, for *N*=60, is shown in Table 1.

7.2 Fitting the Logistic Curve

Our next step was to fit each of those 55 tables of values to Equation 11. We transform the non-linear equation to a linear form in order to use the linear least-squares regression. Applying logarithms to both sides of Equation 11 we get:

$$log(\frac{1}{Rch_{N,r}} - 1) = \alpha_N - \beta_N r$$

Substituting $t = log(\frac{1}{Rch_{N,r}} - 1)$,

$$t = \alpha_N - \beta_N r$$

which allows us to estimate α_N and β_N using linear least-squares regression.

We estimated α and β for each of the 55 selected values of *N*. Goodness of fit as measured by the R-squared statistic was almost one, with the lowest value being 0.996. This corroborates the close agreement of simulated values and the fitted equation seen in Figure 5. At this point, we obtained a table with estimated α and β values for the 55 values of *N* we had chosen. Some rows of this table are shown in Table 2.

Table 1. $N = 60$

r	$Rch_{60,r}$
0.11	0.097306765
0.12	0.144781929
0.13	0.214324298
0.14	0.313522569
0.15	0.436204508
0.16	0.572368896
0.17	0.703084160
0.18	0.811325984
0.19	0.880296608
0.20	0.928937296

Table 2.

N	α_N	β_N
2	3.255884789	6.283736818
5	3.977056234	9.870638140
10	4.691024580	14.53923918
.	.	.
.	.	.
55	8.145698174	50.98543867
60	8.263521833	53.85171640
.	.	.
.	.	.
175	11.47178670	124.4936168
200	12.03414482	138.8969787
.	.	.
.	.	.
450	16.21675101	278.7307447
500	16.69687608	302.2307067

7.3 Fitting the Logistic Coefficients

Having estimated the logistic coefficients α_N and β_N for several values of *N*, we performed a second level of regression on the estimated coefficients to express α and β as a function of *N*. Doing this allows us to interpolate α_N and β_N for values of N we have not simulated, and lets us express α_N and β_N concisely in terms of *N*. This can also reduce error by staying faithful to a general trend, mitigating the effect of any anomalous data points.

We fit values of α to a sum of exponentials function, and values of β to a sixth degree polynomial. In the absence of physically significant models, we chose models that gave us maximum accuracy. The

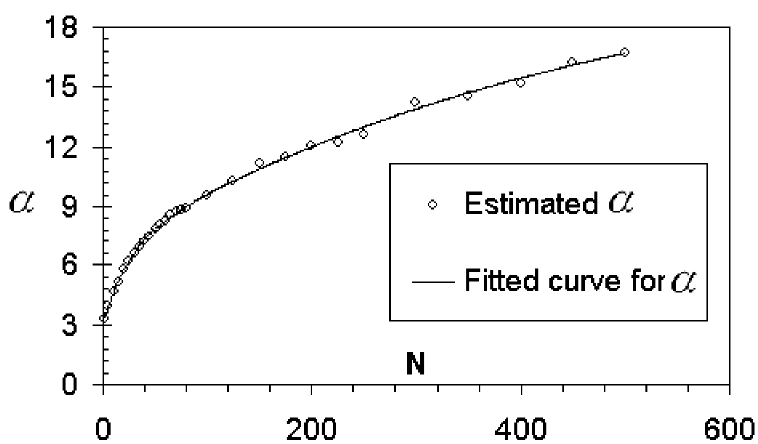


Figure 6. Estimated and fitted α

expressions in terms of *N* for $2 \leq N \leq 500$ are:

$$\alpha_N = 3.004 + 3.815(1 - e^{-4.091\times 10^{-2}N}) + 15.4(1 - e^{-2.055\times 10^{-3}N}) \quad (12)$$

$$\beta_N = 5.141 + 0.9421N - 2.597 \times 10^{-3}N^2 + 8.42 \times 10^{-6}N^3 - 1.37 \times 10^{-8}N^4 + 1.058 \times 10^{-11}N^5 - 3.209 \times 10^{-15}N^6 \quad (13)$$

Figures 6 and 7 plot the estimated values of α and β along with the curves represented by equations 12 and 13.

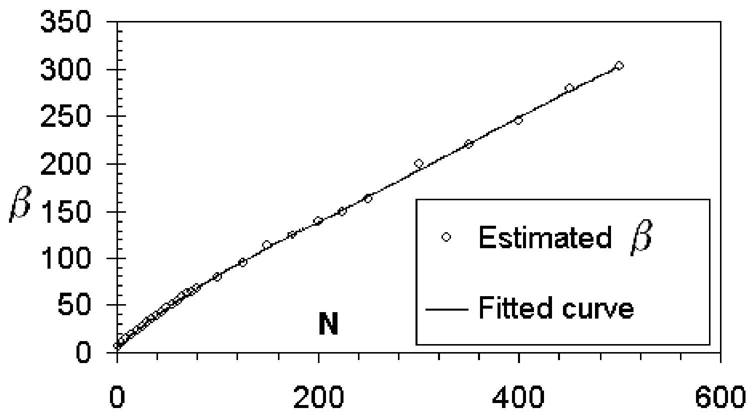


Figure 7. Estimated and fitted β

7.4 Validation

Equations 11, 12 and 13 form a model for reachability. Given a value of *N* and *r*, we obtain the corresponding value of reachability as follows:

- obtain α_N and β_N by substituting *N* in equations 12 and 13
- substitute α_N, β_N and *r* in Equation 11.

We chose 20 values of *N* between 2 and 500 at random, which were not among the 55 values of *N* chosen to build the regression model. For each value of *N*, we chose five values of *r* that would roughly correspond to a reachability value between 0.05 and 0.95. This choice of *r* is necessary because a random selection of *r* is very likely to result in a reachability of either zero or one, since reachability takes on values in between only for a narrow range of values of *r*. We calculated the reachability corresponding to these hundred pairs of *N* and *r* values using equations 11, 12 and 13, and compared them with values obtained from simulation. We calculated absolute and relative errors between the simulated and estimated values of reachability and found an average relative error of 3.5% in the model. We did not observe a single instance where the value of reachability predicted by the model was in error by more than 0.05.

Table 3. Beyond $N = 500$

N	$g_N=\frac{\alpha_N}{\beta_N}$	$Rch(N,g_N-0.01)$	$Rch(N,g_N+0.01)$
500	0.055	0.0515	0.9418
600	0.0495	0.0315	0.9470
700	0.0451	0.0201	0.9518
800	0.0413	0.0129	0.9518
900	0.0381	0.0086	0.9515
1000	0.0354	0.0060	0.9505
1200	0.0308	0.0031	0.9414

8 Extending the model

As N grows, smaller changes in r suffice for $Rch_{N,r}$ to increase from a value near 0 to a value near 1. For example, when $N = 10$, the increase of $Rch_{10,r}$ from 0.1 to 0.9 corresponds to an increase in r of 0.3. But when $N = 500$, it corresponds to an increase in *r* of only 0.015. As N grows larger, $Rch_{N,r}$ begins to resemble a step function by transitioning from a value of almost 0 to a value of almost 1 at a threshold value of *r*. Such phase transition behaviour [9] is a known property of multi-hop networks, and the critical transmitting range is a well-studied problem for connectivity [20].

In our model, the transition of $Rch_{N,r}$ for large values of N takes place at $g_N = \frac{\alpha_N}{\beta_N}$ which is the point of inflexion for the logistic curve. Note that in figures 6 and 7, the shape of the curves seems relatively stable for N greater than 200. We use data for N between 200 and 500 to find a rough estimate for the critical transmitting range for $Rch_{N,r}$ up to $N = 1000$. We approximate α_N using a simple exponential function, and β_N using a linear function as

$$\alpha_N = 16.16(1 - e^{-1.947\times10^{-3}N}) + 6.658 \tag{14}$$

$$\beta_N = 27.8844 + 0.5522N \tag{15}$$

for $500 \leq N \leq 1000$. While these estimates do not exactly predict the point of inflexion, they are close enough that setting $r = g_N - 0.01$ results in a $Rch_{N,r}$ value close to 0, and setting $r = g_N + 0.01$ results in a $Rch_{N,r}$ value close to 1. Table 3 illustrates this: the second column contains g_N values obtained from equations 14 and 15, and the third and fourth columns contain $Rch_{N,r}$ values obtained from simulations by setting *r* to $g_N - 0.01$ and $g_N + 0.01$ respectively.

9 Conclusions

This paper has dealt with characterising reachability for a static network. In a static network, probabilistic reachability is of limited use since the measured reachability of a network instance can vary from the expected reachability. But in the presence of mobility and asynchronous communication in a multi-hop network, the measured value of reachability would tend towards its expected value over time. As evidenced by figures 1(c) and 1(d), it is also in such networks that reachability would be most applicable. There is work that allows us to determine the stationary distributions for the locations of mobile nodes [10, 1, 2]. This distribution is a function of the mobility model used and its parameters. It should be possible to use the models obtained in this paper when the stationary distribution of mobile node locations is close to the uniform distribution. When this is not the case, or when mobile nodes are not necessarily restricted to a square area, we can use simulation. We have developed a simulator for sparse wireless multi-hop networks called *Simran* that includes support for reachability and several other topological properties of multi-hop networks. It is available for download from *http://www.it.iitb.ac.in/∼srinath/simran/*.

The assumptions made regarding the propagation model in this paper are quite idealised, and reachability in a real deployment would almost certainly be worse than that obtained by using such a model. However, it should be possible to use the reachability model presented in this paper for designing networks after choosing suitably conservative parameters. It would be useful in estimating trade-offs between number of nodes, transmission range, and required communication capability at different operating points. To this end, we have built a design tool incorporating the reachability model presented in this paper. The tool can be accessed from *http://www.it.iitb.ac.in/∼srinath/tool/rch.html*.

Acknowledgements

We are grateful to Prof. Anurag Kumar and Dr. Krishna Paul for discussions that helped improve this work. We also thank anonymous reviewers for their helpful comments.

References

[1] C. Bettstetter, G. Resta, and P. Santi. The node distribution of the random waypoint mobility model for wireless ad hoc networks. *IEEE Transactions on Mobile Computing*, 2(3):257–269, 2003.

[2] J.-Y. L. Boudec and M. Vojnović. Perfect simulation and stationarity of a class of mobility models. In *Proc. of IEEE INFOCOM*, pages 2743–2754, 2005.

[3] M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *Proc. of IEEE INFOCOM '01*, volume 3, pages 1360–1369, 2001.

[4] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, 2000.

[5] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. John Wiley & Sons, Inc., New York, 1991.

[6] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proc. of ACM SIGCOMM '04*, pages 145–158, 2004.

[7] S. Kingsland. The refractory model: The logistic curve and the history of population ecology. *The Quarterly Review of Biology*, (57):29–52, 1982.

[8] H. Koskinen. Quantile models for the threshold range for k-connectivity. In *Proc. of ACM MSWiM '04*, pages 1–7, 2004.

[9] B. Krishnamachari, S. B. Wicker, and R. Bejar. Phase transition phenomena in wireless ad hoc networks. In *Proc. of SAWN, IEEE Globecom*, 2001.

[10] W. Navidi and T. Camp. Stationary distributions for the random waypoint mobility model. *IEEE Transactions on Mobile Computing*, 3(1):99–108, 2004.

[11] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[12] E. Perevalov and R. S. Blum. Delay-limited throughput of ad hoc networks. *IEEE Transactions on Communications*, 52(11):1957–1968, 2004.
[13] S. Perur. Designing sparse wireless multi-hop networks. In *Student Workshop of IEEE INFOCOM*, April 2006.
[14] S. Perur and S. Iyer. Characterization of a connectivity measure for sparse wireless multi-hop networks. In *Proceedings of Workshop on Wireless Ad hoc and Sensor Networks, in conjuction with ICDCS*, 2006.
[15] S. Perur and S. Iyer. Sparse multi-hop wireless for voice communication in rural India. In *Proc. of the* 12^{th} *National Conference on Communications*, pages 534–538, 2006.
[16] S. Perur and S. Iyer. Reachability: An alternative to connectivity for sparse wireless multi-hop networks. In *Poster session of IEEE INFOCOM*, April 2006.
[17] D. A. Ratkowsky. *Handbook of Nonlinear Regression Models*. Marcel Dekker, Inc., New York, 1993.
[18] K. Römer and F. Mattern. The design space of wireless sensor networks. *IEEE Wireless Communications*, 11(6):54–61, Dec. 2004.
[19] P. Santi and D. M. Blough. An evaluation of connectivity in mobile wireless ad hoc networks. In *DSN '02: Proceedings of the 2002 International Conference on Dependable Systems and Networks*, pages 89–102, Washington, DC, USA, 2002. IEEE Computer Society.
[20] P. Santi and D. M. Blough. The critical transmitting range for connectivity in sparse wireless ad hoc networks. *IEEE Transactions on Mobile Computing*, 2(1):25–39, 2003.
[21] P. Santi, D. M. Blough, and F. Vainstein. A probabilistic analysis for the range assignment problem in ad hoc networks. In *MobiHoc '01: Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, pages 212–220, New York, NY, USA, 2001. ACM Press.
[22] A. Tang, C. Florens, and S. H. Low. An empirical study of the connectivity of ad hoc networks. In *Proc. of the IEEE Aerospace Conference*, volume 3, pages 1333–1338, 2003.
[23] K. S. Trivedi. *Probability & Statistics with Reliability, Queuing, and Computer Science Applications*. Prentice-Hall of India Pvt. Ltd., New Delhi, 2001.
[24] W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *Proc. of MobiHoc '04*, pages 187–198, 2004.

A Derivations for $P(A)$, $P(C)$ and $P(D)$

Expressions for $P(A)$, $P(C)$ and $P(D)$ in terms of r when $l \geq 4R$ are obtained as follows.

A.1 $P(D)$: All three nodes are directly connected

Observe that when three nodes are directly connected, every two nodes are connected to each other. We derive the combined coverage for two nodes which is the length in which a third node must be located to be connected to both other nodes.

Let x be the average distance between two nodes that are connected. Let N_1 be the node closest to the left end of the line segment of length l, and let N_2 be the other node connected to N_1 on its right. Note that the coverage area for N_3 to be connected to both of them is the length between the two nodes, x, and an overlap of $R-x$ on N_2's right and an overlap of $R-x$ on N_1's left. We also need to accommodate the reduction of this overlap in the end segments of the length of operation, as we did for the case $N=2$. Coverage for the initial $R-x$ segment of l is $(R-x)/2$ to N_1's left, x in between them, and $R-x$ to N_2's right. Coverage for the rightmost segment of length R, after compensating for reduction of overlap is obtained as $5R/4$, and coverage for the central $l-2R+x$ segment is $2R-x$. The total coverage is:

$$\left(\frac{R-x}{2}+x+R-x\right)\frac{R-x}{l}+\frac{l-2R+x}{l}(2R-x)+\frac{R}{l}.\frac{5R}{4}$$

Substituting $x = R/2$ (here we use x without factoring in boundary conditions, as the resulting error is small and allows us to obtain an equation of lower degree):

$$Coverage(D) = \frac{3R}{2} - \frac{3R^2}{8l}$$

Since we have assumed that N_1 and N_2 were connected, we multiply the coverage with the probability of N_1 and N_2 being connected, i.e., Equation 3, and divide by l to obtain:

$$P(D) = (2r - r^2)\left(\frac{3r}{2} - \frac{3r^2}{8}\right)$$

A.2 $P(C)$: Probability that three nodes are one-hop connected

There are two ways in which three nodes can be one-hop connected:

1. By nodes N_1 and N_2 being connected, and N_3 then occurring in a position such that it is connected directly to only one of N_1 or N_2; and

2. By nodes N_1 and N_2 being located such that they are unconnected but can potentially be connected through another node, and then with N_3 positioned such that it connects the two.

Case 1:

Let the average distance between two connected nodes be x. Since N_1 and N_2 are given to be connected, N_3 can be one-hop connected with N_1 only by being located to the right of N_2 in a segment that does not overlap with N_1's coverage. This segment is of length x. We do not consider N_3 being located to the left of N_1 since that case is covered by the symmetrical nature of our analysis. (The analysis proceeds from left to right of the line segment with N_2 always to the right of N_1. We could perform another analysis proceeding from right to left and weight both results by half, but the two analyses would be identical except for the nomenclature of the nodes.) The segment of length l is divided into four segments of length $R - x$, $2x$, $l - 2R - x$, and R, from left to right to account for boundary conditions. After identifying the coverages for each of those segments, taking the product of coverages and segment lengths, summing, and substituting $x = R/2$, we get the coverage within which a node would one-hop connect two already connected nodes as $\frac{R}{l} - \frac{R^2}{l^2}$. Since the first two nodes are already connected, the probability for Case 1 is given by dividing the product of the coverage and Equation 3 by l:

$$P(C_1) = (2r - r^2)(r - r^2)$$

Case 2:

Here N_1 and N_2 are not connected, but are such that they can possibly be connected by a node located between them. First, we find the probability of two nodes being located such that they are not connected, but can possibly be connected. Note that the criterion for this is that the two nodes should be separated by at least a distance of R, and not more than a distance of $2R$. After analysis similar to previous cases, this probability is obtained as: $2r - 3r^2$.

If y is the distance between nodes that are not connected, but can be connected by a third node, the common coverage between the two nodes where the third node should be located can be seen to be $2R - y$. Substituting $y = 3R/2$, the coverage comes to $R/2$.

$$P(C_2) = (2r - 3r^2)\left(\frac{r}{2}\right)$$

Since $P(C) = P(C_1) + P(C_2)$,

$$P(C) = (2r - r^2)(r - r^2) + (2r - 3r^2)\left(\frac{r}{2}\right)$$

A.3 $P(A)$: Probability of all three nodes being isolated

There are two ways in which three nodes can be located such that each node is unconnected to any other node:

1. When the first two nodes are located such that they can never be connected, and the third node is not located in either of their coverage segments.
2. When the first two nodes are located such that they are not connected, but can possibly be connected by a third node, but the third node is not located in the overlapping coverage length of the two nodes, or in their individual coverages.

Case 1:

Here, analysis proceeds similar to previous cases and we first obtain the probability that two nodes are located such that they can never be connected as: $1 - 4r + 4r^2$. Next, we find the combined coverage area of those two nodes and obtain the probability that the third node will not be located in that coverage area. This is found to be $1 - 4r + 2r^2$. The product of these two terms gives us:

$$P(A_1) = (1 - 4r + 4r^2)(1 - 4r + 2r^2)$$

Case 2:

Again we use the probability that two nodes fall such that they are not connected, but can be connected by a third node: $2r - 3r^2$.

We also find the coverage length within which the third node could be connected to one or both the nodes, and obtain from it the probability that the third node will not be located in this coverage length: $1 - \frac{7r}{2} + \frac{14r^2}{3}$. We get:

$$P(A_2) = (2r - 3r^2)\left(1 - \frac{7r}{2} + \frac{14r^2}{3}\right)$$

Since $P(A) = P(A_1) + P(A_2)$,

$$P(A) = (1 - 4r + 4r^2)(1 - 4r + 2r^2) + (2r - 3r^2)\left(1 - \frac{7r}{2} + \frac{14r^2}{3}\right)$$