# Characterization 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

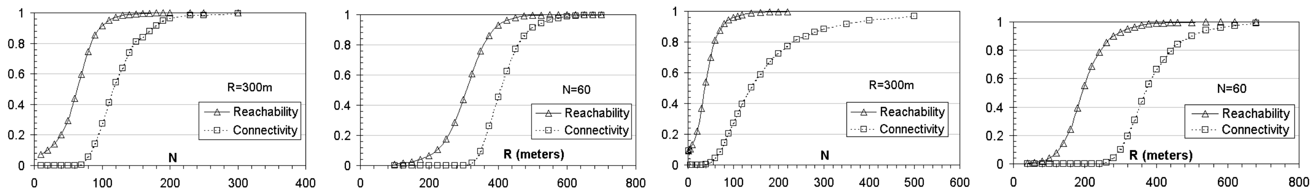
*The extent to which a wireless multi-hop network is connected is usually measured by the probability that all the nodes form a single connected component. We find this measure, called connectivity, unsuitable for use with sparse networks since it is not indicative of the actual communication capability of the network, and can be unresponsive to changes in network parameters. An alternate connectivity measure that we find useful in sparse networks is the fraction of node pairs in the network that are connected. We call this term reachability and claim that it is more intuitive and expressive than connectivity when dealing with sparse networks. 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 normalized transmission range. This can be used by a network designer to estimate the tradeoff 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. 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 parameter 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 probability of 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.

To illustrate the behavior of reachability and connectivity in a sparse multi-hop wireless network, consider this simulation study for a hypothetical deployment in a rural scenario: $N$ 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




(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**

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 Fig. 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. Fig. 1(c) shows an increased difference between reachability and connectivity when mobility is introduced in the scenario of Fig. 1(a). Figure 1(d) shows the comparison when asynchronous communication is enabled in the scenario of Fig. 1(b). In this case, almost 80% of node pairs are able to communicate before connectivity increases from 0. More details regarding these simulations can be found in [9].

In this paper, we identify reachability as a more expressive connectivity measure for sparse wireless multi-hop networks. Our main contribution in this paper is the characterization of reachability for wireless multi-hop networks: we identify reachability as obeying the *logistic growth model* (Sec. 4.2), and obtain a closed form expression through regression analysis that is valid for values of *N* from 2 to 500 when nodes are static and distributed uniformly at random (Sec. 5).

## 2 Related Work

Gupta and Kumar showed in [2] how throughput per source-destination pair in a multi-hop wireless network decreases as node density increases. Grossglauser and Tse in [1] showed that in the presence of mobility, multi-user diversity could be used to achieve a tradeoff between throughput and delay. This would allow throughput to be maintained almost constant even with increasing node density. A further quantification of this tradeoff is found in [8]. Similarly, a tradeoff has also been achieved between *connectivity* and delay. Delay tolerant routing [3] and Message Ferrying [14] 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 tradeoffs.

A study of connectivity in sparse multi-hop networks can be found in [12]. The part of this work most related to ours deals with finding the Minimum Transmission Range (MTR) of nodes that ensures a large enough connected component. For example, the paper shows through simulation that forming a connected component with 90% of the nodes in the network requires a much lower MTR than that required for complete connectivity. While this is a significant observation, [12] does not provide a model for evaluating this trade-off. A number of sparse sensor network applications are described in [11].

Regression has been used in [5] to model the threshold transmitting range for k-connectivity in wireless multi-hop networks, and in [13] to model the probability of obtaining a fully connected network. [13] also defines a *connectivity index* in terms of number of nodes in each connected component, that is identical with reachability. However, no model or characterization 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.

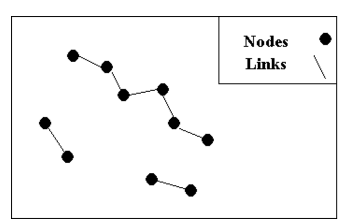



**Figure 2. A network instance with Reachability = 0.378**

## 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:

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

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 Eqn. 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. A network for our purposes can be defined by the number of nodes, their bounding area, and the transmission ranges of the node. The network's reachability would be the expected value of the average reachability across several instances. This value is significant since it represents the probability that a given pair of nodes in the network are connected. Similarly, note that 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. In the presence of mobility, the measured reachability value converges to the expected value with time.

## 4 Characterizing Reachability

Our network model 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 normalized transmission range, $r = R/l$ without loss of generality. While this model takes an overly 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 *characterizing* reachability. The *definition* of reachability is independent of these assumptions. For a network with *N* nodes, normalized transmission range *r* and a set of mobility model parameters *M*, we denote the corresponding value of reachability as $Rch_{N,r}^{M}$. In this paper, we deal only with characterization of the static case, $Rch_{N,r}$.

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

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

It may be possible to use results for number of components and distributions of nodes for a Random Geometric Graph [7] to obtain asymptotic bounds (as *N* tends to infinity) for $Rch_{N,r}$. But since sparse networks often involve small number of nodes, we are particularly interested in characterizations of $Rch_{N,r}$ in the finite domain.

### 4.1 Modeling Reachability

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 [10].

### 4.2 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 3 shows a logistic curve expressed by the equation:

$$y = \frac{k}{1 + e^{\alpha - \beta x}} \quad (3)$$

where *k* is the limiting value that *y* can take, $\beta$ is the maximum rate of growth, and $\alpha$ is a constant of integration [4]. The curve is skew-symmetric and has a point of inflection at $x = \alpha/\beta,\ y = k/2$, where the growth rate is maximum [10].

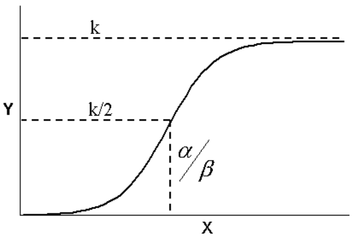



**Figure 3. A general logistic curve**

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. $\alpha$ and $\beta$ vary with *N*, and we denote them by $\alpha_N$ and $\beta_N$. We use Eqn. 3 in the form:

$$Rch_{N,r} = \frac{1}{1 + e^{\alpha_N - \beta_N r}} \quad (4)$$

To illustrate, Fig. 4 shows the result of fitting simulated data with Eqn. 4. The values of $\alpha_{100}$ and $\beta_{100}$ used were 9.58 and 79.2 respectively. We see how these values were obtained in Sec. 5.

# 5 Simulation and Regression Modeling

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

- 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 Eqn. 4 as a regression function for simulated data, and obtained the coefficients $\alpha$ and $\beta$ for the corresponding value of *N*. This allowed us to characterize 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 $\alpha_N$ and $\beta_N$. 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.

## 5.1 Simulations

We conducted extensive simulations to generate the data required for fitting the regression function. Since we were looking to characterize reachability for small to medium sized networks, we chose 55 values of *N* between 2 and 500 as representative points. For each of these values of *N*, we

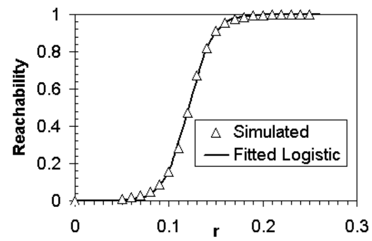



**Figure 4. Logistic fit for N=100**

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. It can be shown that with 1000 runs, the error in the mean is within 0.018 (1.8%) 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.

## 5.2 Fitting the Logistic Curve

Our next step was to fit each of those 55 tables of values to Eqn. 4. It is desirable in general to be able to transform a non-linear equation to a linear form in order to use the statistical properties of linear least-squares regression. Applying logarithms to both sides of Eqn. 4 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 $\alpha_N$ and $\beta_N$ using linear least-squares regression.

We estimated $\alpha$ and $\beta$ for each of the 55 selected values of *N*. Goodness of fit as measured by the R-squared statistic was 0.99 when averaged, with the lowest value being 0.996. This corroborates the close agreement of simulated values and the fitted equation seen in Fig. 4 . At this point, we obtained a table with estimated $\alpha$ and $\beta$ values for the 55 values of *N* we had chosen. Some rows of this table are shown in Table 2.

## 5.3 Fitting the Logistic Coefficients

Having estimated the logistic coefficients $\alpha_N$ and $\beta_N$ for several values of *N*, we performed a second level of regression on the estimated coefficients to express $\alpha$ and $\beta$ as a function of *N*. Doing this allows us to interpolate $\alpha_N$ and $\beta_N$ for values of *N* we have not simulated, and lets us express

**Table 1.**

| 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 | $\alpha_N$ | $\beta_N$ |
|---|---|---|
| 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 |

$\alpha_N$ and $\beta_N$ 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 $\alpha$ to a sum of exponentials function, and values of $\beta$ to a sixth degree polynomial. In the absence of physically significant models, we chose models that gave us maximum accuracy. The expressions in terms of *N* for $2 \leq N \leq 500$ are:

$$\begin{aligned}\alpha_N &= 3.004 + 3.815(1 - e^{-4.091\times10^{-2}N}) \\ &\quad +15.4(1 - e^{-2.055\times10^{-3}N}) \qquad (5)\end{aligned}$$

$$\begin{aligned}\beta_N &= 5.141 + 0.9421N - 2.597\times10^{-3}N^2 \\ &\quad +8.42\times10^{-6}N^3 - 1.37\times10^{-8}N^4 \\ &\quad +1.058\times10^{-11}N^5 - 3.209\times10^{-15}N^6 \qquad (6)\end{aligned}$$

Figures 5 and 6 plot the estimated values of $\alpha$ and $\beta$ along with the curves represented by Eqns. 5 and 6.

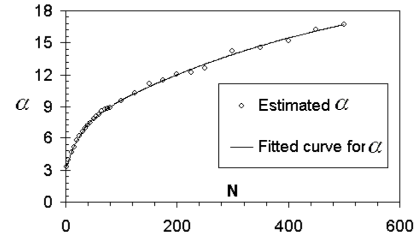



**Figure 5. Estimated and fitted $\alpha$**

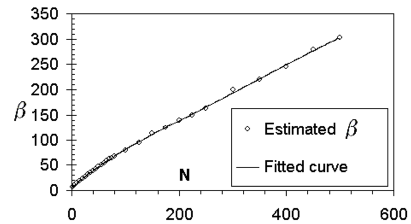



**Figure 6. Estimated and fitted $\beta$**

## 5.4 Validation

Equations 4, 5 and 6 form a model for reachability. Given a value of *N* and *r*, we obtain the corresponding value of reachability as follows:

- obtain $\alpha_N$ and $\beta_N$ by substituting *N* in Eqns. 5 and 6
- substitute $\alpha_N$, $\beta_N$ and *r* in Eqn. 4.

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 Eqns. 4, 5 and 6, 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.

## 5.5 Extending the model

As *N* increases, $Rch_{N,r}$ tends to increase from 0 to 1 more quickly. For example, for $N = 10$, the increase of $Rch_{10,r}$ from 0.1 to 0.9, corresponds to an increase of 0.3 in *r*. But for $N = 500$, it corresponds to an increase in *r* of

**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 |

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 [6] is a known property of multi-hop networks, and the critical transmitting range is a well-studied problem for connectivity [12].

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 inflection for the logistic curve. Note that in figures 5 and 6, 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 $\alpha_N$ using a simple exponential function, and $\beta_N$ using a linear function as

$$\alpha_N = 16.16(1 - e^{-1.947\times10^{-3}N}) + 6.658 \qquad (7)$$

$$\beta_N = 27.8844 + 0.5522N \qquad (8)$$

for $500 \leq N \leq 1000$. While these estimates do not exactly predict the point of inflection, 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 Eqns. 7 and 8, 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.

## 6 Conclusions and Future Work

The assumptions made regarding the network model in this paper are quite idealized, and performance 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 also 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*.

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 Figs. 1(c) and 1(d), it is also in such networks that reachability would be most applicable. The static characterization presented in this paper is a step towards characterizing reachability for more dynamic networks.

## 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] 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.

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

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

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

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

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

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

[8] E. Perevalov and R. S. Blum. Delay-limited throughput of ad hoc networks. *IEEE Transactions on Communications*, 52(11):1957–1968, 2004.

[9] 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.

[10] D. A. Ratkowsky. *Handbook of Nonlinear Regression Models*. Marcel Dekker, Inc., New York, 1993.

[11] K. Römer and F. Mattern. The design space of wireless sensor networks. *IEEE Wireless Communications*, 11(6):54–61, Dec. 2004.

[12] 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.

[13] 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.

[14] 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.