Connectivity of WANETs (1)

There are three fundamental properties of Wireless Ad Hoc Networks (WANETS) which are extensively used in the literature to characterize the performance of the WANET: *Connectivity*, *Capacity* and *Coverage* (C^3 Problems). We will first discuss connectivity problem because it provides the necessary foundation for the discussion on capacity and coverage problems. There are many kinds of capacity or coverage concepts proposed in the literature, all of them are based on some assumptions on connectivity. At the first sight, connectivity is simpler than coverage and capacity problems. However, to fully understand connectivity problems need some highly technical mathematics like percolation theory in statistical physics or Chen-Stein method in advanced probability theory. In fact, the most difficult part of proving theorems in C^3 is more or less related to the connectivity, and the techniques developed here can be reused in our research, especially in the theoretical analysis part. In EPFL, there is even a one semester course for this topic [Grossglauser06], covered the materials from the discrete percolation theory to random geometric graph and their applications to WANETs. In our course, we will use 2-3 classes to cover the following topics:

- Connectivity for arbitrary finite networks;
- Connectivity for homogeneous random geographic graphs;
- Graph evolution for homogeneous random geographic graphs;
- Connectivity for general random geographic graphs;
- k-connectivity problems;
- Connectivity based on more realistic assumptions for radio propagation;
- Connectivity with Bernoulli nodes;

Pan will also help me to present our own work on connectivity of WANETs with direction antenna. The most important part is the second part: *connectivity for homogeneous random geographic graphs.* it is the most extensively studied subtopic, and the results in this subtopic give the theoretical foundation for other subtopics. I will go through this part slowly, and discuss one of Gupta and Kumar's classical paper [Gupta98], the second-most-cited paper of P. R. Kumar, with the greatest patience.

The symbol $\mathbb{R}$ stands for the set of real numbers. $\mathbb{Z}$ is the set of integers, $\mathbb{N}$ the set of integers strictly greater than 0. The symbol $\mathbb{P}$ stands for probability of, and $\mathbb{E}$ is the corresponding expectation. A graph G is determined entirely by its set of *vertices*, V , and *edges*, $E \subseteq V \times V$. We say that an edge $e \equiv (u, v)$ is present in graph G, or $e \in G$ if and only if $e \in E$. Check our handout for related concepts in graph theory.

1 Problem 1: CTR for Arbitrary Network

1.1 Critical Transmission Range (CTR)

The simplest form of connectivity problem in the literature is the characterization of the so called *critical transmission range* (CTR). All the network nodes are assumed to have the same *transmitting range* r_c, and the problem is to identify the minimum value of r_c (the critical transmitting range) such that the resulting communication graph is connected. We recall that an undirected graph G is *connected* if and only if there exists at least one path connecting any two nodes in the graph.

The interest in finding the minimum value of r_c that guarantees certain properties is motivated by energy consumption and network capacity concerns. A low transmission range reduces the energy consumption and increase spatial reuse but if it becomes too low then the network can become disconnected.

The assumption that all the nodes use the same transmitting range reflects all those situations in which transceivers use the same technology and no transmit power control. This is the case, for instance, for most of the 802.11 wireless cards currently on the market. In this scenario, using the same transmitting range for all the nodes is a reasonable choice, and the only way to reduce energy consumption and increase capacity is to reduce r_c as much as possible (we will explain this point in detail when we discuss Gupta and Kumar's capacity paper [Gupta00]).

1.2 CTR for Arbitrary Network

Arbitrary network here means that we do not make any assumption on the distribution of network nodes.

We denote the set of network nodes to be $X = \{x_1, \ldots, x_i, \ldots, x_n\}$ where x_i represents the position of node i and n is the total number of nodes (or *network size*). Obviously, the CTR of X, i.e. r_c^*, is the function of X, and can be denoted as $r_c^*(X)$.

We first give the upper-bound and lower-bound of the CTR for a given arbitrary network.

Theorem 1 (upper-bound and lower-bound of CTR) *(1)* $r_c^*(X) \leq \max_{x_i, x_j \in X} \|x_i - x_j\|$*; (2)* $r_c^*(X) \geq \min_{x_i, x_j \in X \wedge i \neq j} \|x_i - x_j\|$.

From the upper-bound of CTR, we know that $r_c^*(X) \leq \sqrt{2} \cdot l$ when all network nodes are in a square region of side length l and $r_c^*(X) \leq D$ when all network nodes are in a disk region of diameter length D. For arbitrary networks, it is a tight bound since nodes could be concentrated at the opposite corners of the square or the opposite ends of any diameter of the disk. To get this upper-bound is trivial, however, it is the foundation of applying *Percolation Theory* or *Occupancy Theory* to the analysis of connectivity and capacity of WANETs (See Figure 1 for an example).

The lower-bound of CTR reflects a simple and useful relation between the connectivity of a graph G, a global property, and the degree $deg(u)$ of an arbitrary node $u \in V$, a local property. The implication $\{G \text{ is connected}\} \Rightarrow \{\min_{u \in V} deg(u) \geq 1\}$ is always true. The opposite implication is not always true, however, because a network can consists of separate, disconnected clusters containing nodes each with minimum degree larger than 1 (See Figure 2 for an illustration). Connectivity property is a global property which cannot

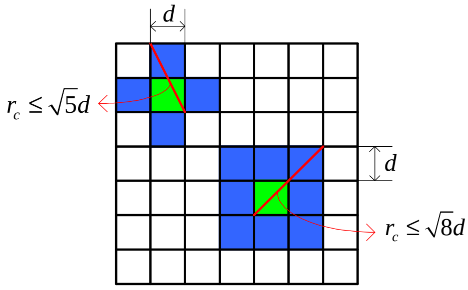

Figure 1: Construction of the percolation model. We divide the plane in squares of size $d \times d$. By setting the relation between r_c and d, we can make sure that any pair of nodes located in two "neighbor" squares are connected.

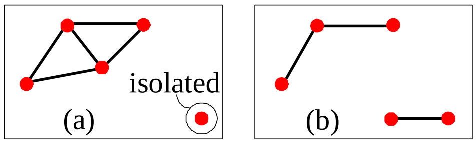

Figure 2: Illustration of the difference between isolated node and disconnectedness. (a){There exists isolated node} $\Rightarrow${The network is disconnected}; (b){the network is disconnected}$\not\Rightarrow${There exists isolated node}.

be guaranteed in a distributed or localized way, however, whether a node is an isolated one only reflects the local property of each node. Many sophisticated techniques are based on this relation, and we will see it many times in the following.

The following theorem shows that the CTR for arbitrary network equals the length of the longest edge of the *Euclidean Minimum Spanning Tree* (EMST) built on the network nodes. We recall that the EMST is a connected graph that contains all the nodes and minimizes the sum of the Euclidean distances of edges (or links).

Theorem 2 (relationship between EMST and CTR) *The CTR for connectivity, i.e. r_c^*, is equal to the length of the longest edge of the EMST built on the nodes.* [1]

PROOF: Consider an arbitrary set of nodes and assume that their EMST is known.

Assume first that r_c^* is shorter than the EMST's longest link. By the definition of the EMST, its longest link (as well as every other link) is the shortest possible way to connect the two subsets of nodes separated by the link. (Otherwise the link sum of the tree could be made even smaller by changing the link to a shorter one.) The assumption made thus implies that r_c^* is too short to connect the two subsets separated by the EMST's longest link, which contradicts the definition of r_c^*.

[1]This elegant result is presented in many papers and books without proof. Try to prove it by yourself, using the "proof by contradiction". You will find a little later that many elegant theorems about the global properties of ad hoc network is proved by contradiction in the literature.

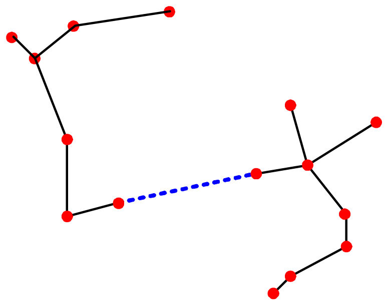

Figure 3: A sample set of 15 nodes and their EMST. The longest link is shown with a dash line.

The assumption that r_c^* is longer than the EMST's longest link is trivially wrong, since in this case all the nodes have been connected using distances shorter than r_c^* by the EMST. This completes the proof. □

According to Theorem 2, computing the CTR is equivalent to computing the EMST on the network nodes, and finding the longest edge in the EMST. Several algorithms exist for finding the EMST; in this study, the Prim algorithm was used: starting with any single node, new nodes are added to the tree one by one, so that at each step the node closest to the nodes included so far is added. One realization with 15 nodes as well as their EMST is depicted in Figure 3.

Algorithm 1 Find $r_c^*(X)$

Require: $X = \{x_1, \ldots, x_n\}$ // n: network size, $x_i \in \mathbb{R}^2$ $\forall i$: position of node i
Ensure: $R = r_c^*(X)$
  Calculate $\mathbf{D} \in \mathbb{R}^{n \times n} : d_{ij} = \|x_j - x_i\|$ // the distance matrix
  $R \leftarrow \max_i\{\min_j\{d_{ij} \mid d_{ij} > 0\}\}$ // the threshold range for minimum degree 1
  $S \leftarrow \emptyset$ // the set of connected subnetworks
  $P \leftarrow \{1, 2, \ldots, n\}$ // the set of nodes not yet included
  while $P \neq \emptyset$ **do** // Note that P cannot contain only 1 element
    $C \leftarrow \{P(1)\}$ // the first element in P
    $P \leftarrow P \setminus \{P(1)\}$ // the first element in P
    $i \leftarrow 0$
    repeat
      $i \leftarrow i + 1$
      $N \leftarrow \{j \mid j \in P, d_{C(i)j} \leq R\}$ // nodes within range R from $C(i)$
      $C \leftarrow C \cup N$ // Append N at end of C
      $P \leftarrow P \setminus N$
    until $P = \emptyset \vee i = \text{card}(C)$
    $S \leftarrow S \cup \{C\}$ // Maintain C as a set
  end while
  $N_S = \text{card}(S)$
  if $N_S > 1$ **then** // run the Prim algorithm for the connected subnetworks
    Calculate $\mathbf{M} \in \mathbb{R}^{N_S \times N_S} : m_{ij} = \min\{d_{kl} \mid k \in S(i), l \in S(j)\}$ // $\mathbf{M}$ is the distance matrix for the connected subnetworks
    $C \leftarrow \{1\}$
    $P \leftarrow \{2, \cdots, N_S\}$ // the subnetworks not yet included
    while $P \neq \emptyset$ **do**
      $s \leftarrow \text{argmin}_{i \in P}\{m_{ij} \mid i \in C\}$ // the closest subnetwork not yet included
      $r \leftarrow \min\{m_{ij} \mid i \in C, j \in P\}$ // and its distance from the included subnetworks
      $C \leftarrow C \cup \{s\}$
      $P \leftarrow P \setminus \{s\}$
      $R \leftarrow \max\{R, r\}$
    end while
  end if

Unfortunately, this way of calculating the CTR is not apt to distributed implementation, since building the EMST requires global knowledge (the exact positions of all the nodes in the network), which can be acquired in a distributed setting only by exchanging a considerable amount of messages. Furthermore, the requirement of knowing exact node positions is very strong: in fact, in many situations, node locations cannot be determined a priori (for instance, when sensors are dispersed on the field using a moving vehicle), and obtaining exact location information when nodes are already deployed is, in general, quite expensive (for instance, because many network nodes should be equipped with GPS receivers).

2 Problem 2: CTR for Homogeneous Random Geographic Graph (RGG)

2.1 Graphic Connectivity and Asymptotic Connectivity

For the reasons described above, considerable attention has been devoted to characterizing the CTR in the presence of some form of uncertainty about node positions. A typical approach is to assume that nodes are distributed in region $\mathcal{R}$ according to some probability density function $\mathcal{F}$, and to study the conditions for asymptotically almost sure connectivity.

The probabilistic characterization of the CTR can be of great help in answering fundamental questions that arise at the network planning stage, such as: given a number n of nodes to be deployed in a certain region $\mathcal{R}$, and given distribution $\mathcal{F}$, which resembles real-world node distribution, which is the minimum value $r_c^*(n, \mathcal{F})$ of the transmitting range that ensures connectivity with high probability? Conversely, given a transmitter technology (i.e. the value of r_c) and distribution $\mathcal{F}$, which is the minimal number $n^*(r_c, \mathcal{F})$ of nodes to be deployed in order to obtain a connected network with high probability? The answer to the questions above depends on the shape of $\mathcal{R}$ and on the distribution $\mathcal{F}$ used to distribute nodes in $\mathcal{R}$. In particular, we consider two probabilistic formulations of the CTR problem:

- *Fixed deployment region*: In this version of the problem, the side l of the deployment region $\mathcal{R}$ is fixed (e.g. $\mathcal{R}$ is the unit square), and the asymptotic value of the CTR as $n \to \infty$ is investigated. In principle, results obtained for this version of the problem can be applied only to dense networks. In fact, the value of the CTR is characterized as the node density $\lambda \equiv \frac{n}{l^2}$ grows to infinity, since l is an arbitrary constant.

- *Deployment region of increasing side*: In this version of the problem, the side l of the deployment region is a further model parameter, and the asymptotic value of the CTR as $l \to \infty$ is investigated. In this model, l can be seen as the independent variable, and both r_c and n are expressed as a function of l (and of the distribution $\mathcal{F}$). Since in this version of the problem the node density $\lambda \equiv \frac{n(l,\mathcal{F})}{l^2}$ can either converge to a constant $c \geq 0$ or diverge as $l \to \infty$, the theoretical results obtained using this model can be applied to networks with arbitrary density.

2.2 Mathematical Preliminaries

2.2.1 Homogeneous Poisson Point Process

As already made apparent by the problem of connectivity, the geographical locations of network nodes are an important factor affecting the performance of wireless multihop networks, which underlines the need to model these locations. To take into account the possibility of practically any con- figuration of nodes, the locations are usually treated as random. Furthermore, unless more specific information is given, it is reasonable to assume that, a priori, the locations are uniformly distributed.

To this end, let us assume that n nodes are randomly and independently located according to the uniform distribution over some bounded domain $\mathcal{A} \subseteq \mathbb{R}^2$ with an area $\|\mathcal{A}\| = A$ (generalization to a higher number of dimensions is straightforward). Then the number of nodes in any subdomain $\mathcal{D} \subseteq \mathcal{A}$ is random, with the distribution $Bin(n, \|\mathcal{D}\|/\|\mathcal{A}\|)$.

However, given the number of nodes n' in any other non-intersecting subdomain $\mathcal{D}'$, the conditional distribution is different; hence, the two numbers are not independent. Keeping our attention on the arbitrarily selected domains $\mathcal{D}$ and $\mathcal{D}'$, let us consider the effect of letting the domain $\mathcal{A}$ become larger and larger, while keeping the average node density $\lambda \equiv n/A$ constant. Since this makes $\|\mathcal{D}\|/\|\mathcal{A}\|$, the probability that an arbitrary node is in $\mathcal{D}$, diminish but keeps the expected number of nodes therein $n\|\mathcal{D}\|/\|\mathcal{A}\|$ constant, in the limit $n, A \to \infty$ the above binomial distribution tends to a Poisson distribution with the parameter $\lambda\|\mathcal{A}\|$. In the same limit, the number of nodes in $\mathcal{D}$ depends less and less on n' (because this leaves $n - n'$ nodes outside $\mathcal{D}$, which also tends to infinity).

The point process that results in the limit is the homogeneous Poisson point process. It can be interpreted as points "uniformly distributed" over the whole plane with average density λ . It is completely characterized by the following two properties:

Definition 1 (homogeneous Poisson point process) *A homogeneous Poisson point process is defined by the following two properties:*

- *The number of nodes N in each finite subarea $\boldsymbol{A}$ of size $\|\boldsymbol{A}\| = A$ follows a Poisson distribution, i.e.,*

$$\mathbb{P}(n \text{ nodes in } \boldsymbol{A}) = \mathbb{P}(N = n) = \frac{\mu^n}{n!}e^{-\mu}; n \in \mathbb{N} \cup \{0\},$$

  with a mean value $\mathbb{E}(N) = \mu = \lambda A$.

- *The number of nodes N_i in disjoint (non-overlapping) areas $\boldsymbol{A}_i$, $i \in \mathbb{N}$, are independent random variables, i.e.,*

$$\mathbb{P}(N_1 = n_1 \wedge N_2 = n_2 \wedge \cdots \wedge N_k = n_k) = \prod_{i=1}^{k} \mathbb{P}(N_i = n_i).$$

We denote this process as being homogeneous, if λ is constant over the entire infinitely large area. In other words, the outcome of the random variable N only depends on the size of the subarea $\boldsymbol{A}$ but not on its particular location or shape.

Random geometric graphs are easily described. A set of nodes is randomly scattered over a region of space according to some probability distribution, and any two nodes separated by a distance less than a certain specified value are connected by an edge.

In the literature, homogeneous random geometric graph means that the nodes are distributed as a homogeneous Poisson point process, or n nodes are placed uniformly and independently in a region $\mathcal{D}$, with homogeneous transmission range $r(n)$. Based on above discussion, we know that when $n = \lambda \to \infty$, they are almost the same.

2.2.2 Asymptotic Notation

In this subsection, we recall the standard notation regarding the asymptotic behavior of functions. Let f and g be functions of a certain parameter x. We are interested in characterizing the asymptotic behavior of f and g as $x \to \infty$.

Definition 2 (big Oh notation and variants) *Suppose f and g are two functions then we shall write $f(n) = O(g(n))$ or equivalently, $g(n) = \Omega(f(n))$, if there exists $K > 0$ such that $\lim_{n\to 0} \frac{f(n)}{g(n)} \leq K$. We write $f(n) = o(g(n))$ if $\lim_{n\to 0} \frac{f(n)}{g(n)} = 0$. The notation $f(n) = \Theta(g(n))$ will mean that both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. Finally, if $f(n) = o(g(n))$, we shall also say $f \ll g$ or equivalently, $g \gg f$. The last might seem an unnecessary extension, but it permits the simpler and obvious idiom: "$\cdots$ for any $r(n) \gg r_c(n) \cdots$" instead of "$\cdots$ for any $r(n)$ such that $r_c(n) = o(r(n)) \cdots$"*

Definition 3 (a.a.s. event) *Let E_x be a random event that depends on a certain parameter x. We say that E_x holds asymptotically almost surely (a.a.s.) or with high probability (w.h.p) if $\lim_{x\to\infty} \mathbb{P}(E_x) = 1$.*

Most work in the literature is concerned with asymptotic properties (a.a.s) for large graphs (network size $n \to \infty$). Here we also focus on asymptotic connectivity properties (In Gupta and Kumar's capacity paper, the "capacity" they discussed is also in the a.a.s. sense). Why? Determine the probability that a random network[2] is connected is equivalent to knowing the distribution of the critical transmission range r^*, which is also a random variable. For a finite number of nodes n, the distribution of r^* is not known even in the simplest cases such as uniform distribution on a domain with a simple shape. If you do not believe me, you can try by yourself, just calculating the distribution of r^* when $n = 3$ or 4, you will find it is very difficult. The distribution of r^* only known asymptotically, as the number of nodes in the network tends to infinity. All the existing precise analytical results are asymptotic in nature. Frankly speaking, we are interested in the case when $n \to \infty$, not because we are considering scalability problem, just because we can get analytical results in this case. Precise computation of probabilities for properties of random network is usually unfeasible except for small values of n, and understanding their average case behavior may be a useful alternative to exact computation, this motivates our interest in asymptotic theory.

Why when we assume $n \to \infty$, everything become simpler? Let me expain it using an example taken from [Spencer01]. We will be looking at labeled graphs G on n vertices. For convenience we'll call the vertices $\{1, \cdots, n\}$. Imagine that every pair i, j of vertices flips a fair coin to decide whether or not to be connected. We call the outcome the random graph $G(n, \frac{1}{2})$. In fact $G(n, \frac{1}{2})$ is a set of graphs, and the number of such graphs is precisely $2^{\binom{n}{2}}$ as each of the $\binom{n}{2}$ pairs i, j can be either connected or not connected. Consider a graph property — for example, the property of containing a triangle, we call it the property A. We set $\mu_n(A)$ equal the proportion of graphs in $G(n, \frac{1}{2})$ that have the property A. A precise evaluation of $\mu_n(A)$ might be very difficult. However, we can easily prove the following theorem about $\mu_n(A)$ when $n \to \infty$ (prove it by yourself before you read the proof):

Theorem 3 $\lim_{n\to\infty} \mu_n(A) = 1$

PROOF: Split the vertices into $s = \lfloor \frac{n}{3} \rfloor$ disjoint triples. A triple i, j, k forms a triangle with probability precisely $\frac{1}{8}$. These are independent events as they involve distinct coin flips.

[2]Here, we say *random network* in the general sense, means that there are some randomness in the positions of nodes.

Thus the probability that none of the s triples form a triangle is $(\frac{7}{8})^s$. This goes to zero as n, and therefore s, goes to infinity. □

Sometimes a "silly" example can be instructive. In the same sprit, you can prove the following old saying:

The infinite monkey theorem *states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type or create a particular chosen text, such as the complete works of William Shakespeare.* (Note that "almost surely" in this context is a.a.s. we just defined, and that the "monkey" is not an actual monkey; rather, it is a vivid metaphor for an abstract device that produces a large, random sequence of letters.)

2.3 Gupta and Kumar's Asymptotic Connectivity Paper

In this section, we discuss Gupta and Kumar's asymptotic connectivity paper [Gupta98] in detail. In fact, except a result borrowed from *Continuum Percolation Theory*, all the details can be understood with the background of undergraduate mathematics. Even you may forget those things, we will provide the related background knowledge when necessary, and try to explain the basic ideas behind the proof procedures as clear as possible. So, please try to read the original paper by yourself with the help of the explanations provided in this section.

2.3.1 Main Result

Theorem 4 (Gupta & Kumar's Asymptotic Connectivity) *Let $\mathcal{D}$ be a disc in the plane $\mathbb{R}^2$ having unit area. Let $G(n, r(n))$ be the network (graph) formed when n nodes are placed uniformly and independently in $\mathcal{D}$, with homogeneous transmission range $r(n) = \sqrt{\frac{\log(n)+c(n)}{\pi n}}$. Then the resulting network $G(n, r(n))$ is connected a.a.s. if and only if $c(n) \to +\infty$.*

Theorem 4 can be expressed as follows which may be more clear: $G(n, r(n))$ is connected a.a.s. if and only if $\lim_{n\to\infty} \big(\pi r^2(n)n - \log n\big) \to +\infty$. Note that the theorem do not give any constriction on the speed when $c(n)$ tends to infinity. Since in practice, we want a small $r(n)$, the speed of $r(n)$ tending to infinity can be set arbitrary small. In most simulation study, the researcher usually sets $c(n) = \log\log n$.

2.3.2 Outline of Proof of Necessity

We first discuss Section 2 of [Gupta98], the proof of necessary condition on $r(n)$ for connectivity. We need to prove:

$$\text{When } r(n) = \sqrt{\frac{\log(n)+c(n)}{\pi n}},\ \Big\{\lim_{n\to\infty} \mathbb{P}\{G(n, r(n)) \text{ is connected}\} = 1\Big\} \Longrightarrow \Big\{\lim_{n\to\infty} c(n) = +\infty\Big\}. \tag{1}$$

Because the event $\{G(n, r(n))$ is connected$\}$ is difficult to handle, we try to consider the event $\{G(n, r(n))$ is disconnected$\}$. Let $P_d(n, r(n)) \equiv \mathbb{P}\{G(n, r(n))$ is disconnected$\}$, then

$$\Big\{\lim_{n\to\infty} \mathbb{P}\{G(n, r(n)) \text{ is connected}\} = 1\Big\} = \Big\{\lim_{n\to\infty} P_d(n, r(n)) = 0\Big\}. \tag{2}$$

If we can get a tight lower bound of $P_d(n, r(n))$, i.e., $f(n, r(n))$, then the necessary condition for connectivity is that this lower bound must tend to zero when n goes to infinity.

$$\text{When } f(n, r(n)) \equiv \liminf_{n\to\infty} P_d(n, r(n)),\ \left\{\lim_{n\to\infty} P_d(n, r(n)) = 0\right\} \Longrightarrow \left\{\lim_{n\to\infty} f(n) = 0\right\}.^3 \tag{3}$$

If we want to prove (1), this lower bound must be expressed as the function of $c(n)$. The problem is how to get $f(n, r(n))$, the lower bound of $P_d(n, r(n))$? Recall that we mentioned before that connectivity is a global property which is difficult to handle while isolation is a local property which is easy to investigate, we can begin from the isolation. Let $P^{(1)}(n, r(n)) \equiv \mathbb{P}\{G(n, r(n)) \text{ has at least one isolated node}\}$. Obviously, $P_d(n, r(n)) \geq P^{(1)}(n, r(n))$. Now, let's concentrate on analyze the lower-bound of $P^{(1)}(n, r(n))$, and try to get its relationship with $c(n)$.

$$\begin{aligned} P^{(1)}(n, r(n)) &\geq \sum_{i=1}^{n} \mathbb{P}\{i \text{ is the only isolated node in } G(n, r(n))\} \\ &\geq \sum_{i=1}^{n} \mathbb{P}\{i \text{ is isolated in } G(n, r(n))\} - \sum_{j\neq i} \mathbb{P}\{i \text{ and } j \text{ are isolated in } G(n, r(n))\} \end{aligned}$$

From the definition of HPPP (Definition 1), we have

$$\mathbb{P}\{i \text{ is not the neighbor of } j\} = \mathbb{P}\{i \text{ is not in the disk centred at } j \text{ with radius } r(n)\} = (1-\pi r^2(n))$$

Therefore, $\mathbb{P}\{i \text{ is isolated in } G(n, r(n))\} = (1 - \pi r^2(n))^{n-1}$ and

$$\begin{aligned} &\mathbb{P}\{i \text{ and } j \text{ are isolated in } G(n, r(n))\} \\ = &\underbrace{\big(4\pi r^2(n) - \pi r^2(n)\big)}_{\text{prob. when } j \text{ is in the shadow area}} \big(1 - \delta\pi r^2(n)\big)^{n-2} + \underbrace{\big(1 - 4\pi r^2(n)\big)}_{\text{prob. when } j \text{ is not in the shadow area}} \big(1 - 2\pi r^2(n)\big)^{n-2} \end{aligned}$$

where $1 \leq \delta < 2$. Figure 4 give an illustration on how to calculate this probability.

[3] Let S be a set of real numbers. A lower bound for S is a number B such that $x \geq B$ for all $x \in S$. The infinum (inf, greatest lower bound) of S is the greatest lower bound for S. An lower bound which actually belongs to the set is called a minimum.

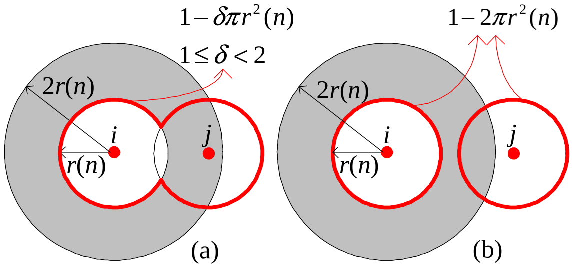

Figure 4: Calculating $\mathbb{P}\{i \text{ and } j \text{ are isolated in } G(n, r(n))\}$. (a) the case when j is at a distance between $r(n)$ and $2r(n)$ from i; (b) the case when j is $2r(n)$ away from i.

We get the lower bound of $P^{(1)}(n, r(n))$:

$$\begin{aligned}
P^{(1)}(n, r(n)) &\geq n(1-\pi r^2(n))^{n-1} - n(n-1)\big(3\pi r^2(n)(1-\delta\pi r^2(n))^{n-2} + (1-4\pi r^2(n))(1-2\pi r^2(n))^{n-2}\big) \\
&\geq n(1-\pi r^2(n))^{n-1} - n(n-1)\big(3\pi r^2(n)(1-\delta\pi r^2(n))^{n-2} + (1-2\pi r^2(n))^{n-2}\big) \\
&\quad \text{(Note: } (1-4\pi r^2(n)) \to 1 \text{ when } n \to \infty) \\
&\geq \theta e^{-c} - n(n-1)(3\pi r^2(n)e^{-\delta(n-2)\pi r^2(n)} + e^{-2(n-2)\pi r^2(n)}) \\
&\quad \text{(Note: using Lemma 2.2 and Lemma 2.1 (i) in [Gupta98])} \\
&\geq \theta e^{-c} - n^2(3\pi r^2(n)+1)e^{-2(n-2)\pi r^2(n)} (\text{ Note: } 1 \leq \delta < 2) \\
&= \theta e^{-c} - (3\pi r^2(n)+1)\Big(n\big(e^{-\pi r^2(n)}\big)^{n-2}\Big)^2 \\
&\geq \theta e^{-c} - (3\pi r^2(n)+1)\Big(n\big(1-\pi r^2(n)\big)^{n-2}\Big)^2 (\text{ Note: using Lemma 2.1 (ii) in [Gupta98]}) \\
&\geq \theta e^{-c} - \kappa e^{-2c} (\text{ Note: using Lemma 2.2 in [Gupta98]})
\end{aligned}$$

Then we get $P_d(n, r(n)) \geq \theta e^{-c} - \kappa e^{-2c}$. Therefore, the necessary condition for connectivity is that $c = +\infty$. In completing the proof for $c = \lim_{n\to\infty} c(n)$, we use the fact that if $r(n) = \sqrt{\frac{\log(n)+c}{\pi n}}$, then the probability of disconnectedness is monotonically decreasing in c.

2.3.3 Outline of Proof of Sufficiency

We then discuss Section 3 of [Gupta98], the proof of sufficient condition on $r(n)$ for connectivity. We first correct an error in this paper. The **equation (1.11)** in [Gupta98] should be:

$$\lim_{\lambda r^2(\lambda)\to+\infty} \frac{1}{q_1(\lambda, (\lambda))} \sum_{k=1}^{\infty} q_k(\lambda, (\lambda)) = 1.$$

We need to prove:

$$\text{When } r(n) = \sqrt{\frac{\log(n)+c(n)}{\pi n}}, \Big\{\lim_{n\to\infty} c(n) = +\infty\Big\} \Longrightarrow \Big\{\lim_{n\to\infty} \mathbb{P}\{G(n, r(n)) \text{ is connected}\} = 1\Big\}. \tag{4}$$

We still use the relation in equation (2). The idea is to get the upper-bound of $P_d(n, r(n))$, i.e. $g(n, r(n)) \equiv \limsup_{n\to\infty} P_d(n, r(n))$. If we can prove that when $c(n) \to \infty$, $g(n, r(n)) \to 0$, we are done.

How to get the upper-bound of $P_d(n, r(n))$? We need to borrow a result from *continuum percolation theory*, and this theory is derived based on a little different model. Suppose nodes are thrown with Poisson intensity λ in the plane $\mathbb{R}^2$. From continuum percolation theory (Theorem 6.3 in [Meester96]), we know that $G^{Poisson}(\lambda, r(\lambda))$ has at most one infinite-order cluster. Furthermore, the node at the origin lies in either an infinite-order cluster or it is isolated.

Consider the restriction $G^{Poisson}_{\mathcal{D}}(n; r(n))$ of the point process to unit disc $\mathcal{D}$. Then, the probability $P^{Poisson}_d(n, r(n)$ that $G^{Poisson}_{\mathcal{D}}(n, r(n))$ is disconnected is asymptotically same as the probability $P^{Poisson;(1)}(n, r(n))$ that it has at least one isolated node. So, again we just need to analyze the probability related to the isolated nodes. The difference here is that in order to get sufficient condition, we need to get upper-bound of the probability.

The last thing we need to do is to relate this result to n nodes uniformly i.i.d. on the unit disc ($G(n, r(n))$ model) by restricting $G^{Poisson}(\lambda, r(\lambda))$ to the disc and showing that difference between the restriction and $G(n, r(n))$ is asymptotically negligible. The difference between $G^{Poisson}_{\mathcal{D}}(n, r(n))$ and $G(n, r(n))$ is that, for $G(n, r(n))$, the number of nodes in the unit disk is a constant number λ while for $G^{Poisson}_{\mathcal{D}}(n, r(n))$ is a random variable following the Poisson distribution with the mean λ. However, when $n = \lambda \to \infty$, intuitively we know the difference is negligible. Therefore, in the following, we just discuss the details of the first step: proving that $\limsup_{n\to\infty} P^{Poisson;(1)}_d(n, r(n)) \leq e^{-c}$.

$$\limsup_{n\to\infty} P^{Poisson;(1)}_d(n, r(n)) = \sum_{j=1}^{\infty} P^{(1)}(j, r(n)) \qquad \underbrace{e^{-n}\frac{n^j}{j!}}_{\text{prob that } G^{Poisson}_{\mathcal{D}}(n,r(n)) \text{ has } j \text{ nodes}} \tag{5}$$

For network $G(j, r(n))$, let A_i be the event that node i is isolated (i.e., it is not in the transmission range of any other $j-1$ node), and let $A \equiv \cup_{1=1}^{j} A_i$ be the event that there is at least one isolated node in the network $G(j, r(n))$. Recall that $\mathbb{P}(A_i) = (1 - \pi r^2(n))^{j-1}$. From the union bound, $\mathbb{P}(\cup_{i=1}^{j}(A_i)) \leq \Sigma_{i=1}^{j}\mathbb{P}(A_i)$, we can write

$$P^{(1)}(j, r(n)) = \mathbb{P}(A) \leq j(1 - \pi r^2(n))^{j-1}. \tag{6}$$

Substituting (6) in (5), we get

$$\begin{aligned}
P^{Poisson;(1)}_d(n, r(n)) &\leq \sum_{j=1}^{\infty} j(1 - \pi r^2(n))^{j-1} e^{-n}\frac{n^j}{j!} \\
&= ne^{-n} \sum_{j=0}^{\infty} \frac{\left(n(1 - \pi r^2(n))\right)^j}{j!} \\
&= ne^{-n} e^{n(1-\pi r^2(n))} (\text{ Note: using formula } \sum_{j=0}^{\infty} \frac{x^j}{j!} = e^x) \\
&= ne^{-n\pi r^2(n)} = e^{\log n - n\pi r^2(n)} = e^{c(n)}
\end{aligned}$$

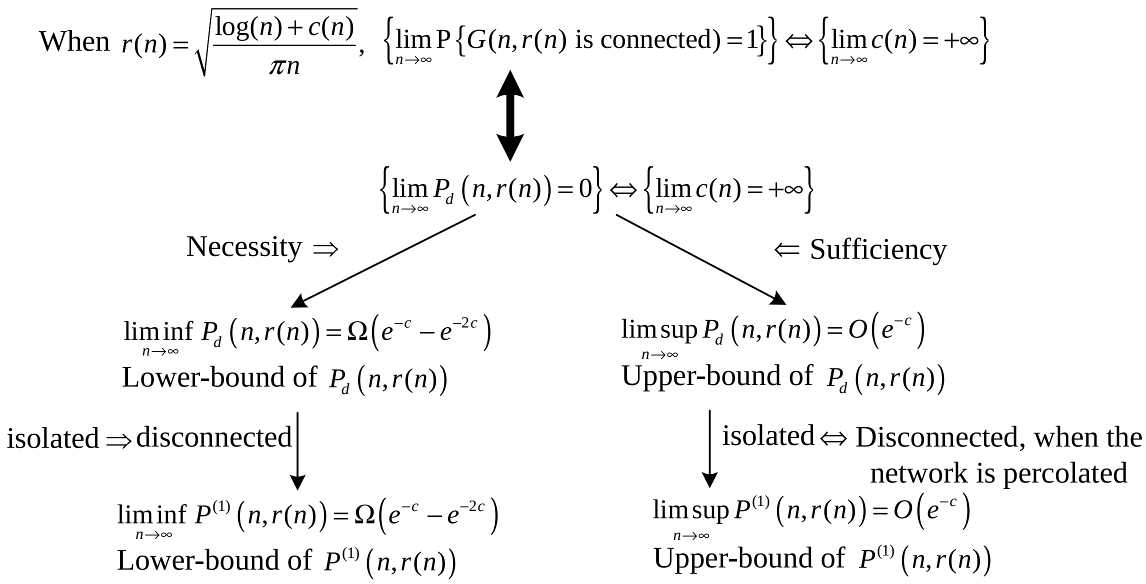

Figure 5: The proof procedure of Gupta & Kumar's asymptotic connectivity theorem.

2.3.4 Summary of the Proof Procedure

The procedure of the proof is summarized in Figure 5. It will be helpful when you feel lost in the forest of inequalities. The basic idea is to convey the connectivity problem to the calculation of network disconnected probability $P_d(n, r(n))$, and the calculation of the lower and upper bound of $P_d(n, r(n))$ can be implemented by calculating the probabilities about isolated nodes, which is easier to handle.

The proof is not given in a constructive way, so even we understand every step, it is still difficult for us to guess how the author get the result of $\pi r^2(n)n - \log n \to \infty$ at the very beginning? What is the key point here to guess this result before you prove it? The answer is the percolation result. In fact we can get this result in a very natural way if we treat the fact that when the network is percolated, the network has no isolated nodes is equivalent to the network is connected, as the starting point of the thinking procedure. We will give our own simple proof following this idea in the class.

2.4 Penrose's Asymptotic Connectivity Papers

In fact, before the publication of [Gupta98], Penrose already published a paper in 1997 with a more general result about asymptotic connectivity properties of RGGs [Penrose97]. For a long time it is not noticed by networking guy, since it is a pure mathematical paper. However, results in this paper and other papers by Penrose are very useful when we discuss more sophisticated cases, such as ununiform distribution of nodes. The proofs in those papers are highly technical and uses the Chen-Stein method, so we cannot treat those papers as Gupta's, and explain every step in detail. Instead, we can only explain the exact meaning of their results.

References

[Penrose03] M. D. Penrose, *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*, Oxford University Press, May 2003.

[Penrose97] M. D. Penrose, "The longest edge of the random minimal spanning tree," The Annals of Applied Probability 7(2), 340C361, 1997.

[Meester96] R. Meester and R. Roy, *Continuum Percolation*, volume 119 of *Cambridge Tracts in Mathematics*, Cambridge University Press, 1996.

[Grossglauser06] M. Grossglauser and P. Thiran, *Networks out of Control: Models and Methods for Large Scale Random Networks*, Class Notes, EPFL, 2006. http://icawww1.epfl.ch/class-nooc/nooc2006.pdf

[Grimmett99] G. Grimmett, *Percolation* (2nd edn.), Springer, Berlin, 1999.

[Hall88] P. Hall, *Introduction to the Theory of Coverage Processes*, Wiley, New York, 1988.

[Spencer01] J. Spencer, *The Strange Logic of Random Graphs*, Springer, Berlin, 2001.

[Stoyan95] D. Stoyan, W.S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, John Wiley & Sons, 1995.

[Gupta98] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming, 1998.

[Gupta00] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," IEEE Transactions on Information Theory, vol. IT-46, no. 2, pp. 388-404, March 2000.

[Xue04] F. Xue and P. R. Kumar, "The number of neighbors needed for connectivity of wireless networks," Wireless Networks, 2004.

[Santi05] P. Santi, *Topology Control in Wireless Ad Hoc and Sensor Networks*, Chap. 4-6, John Wiley and Sons, Chichester, UK, July 2005.

[Hekmat06] R. Hekmat, *Ad-Hoc Networks: Fundamental Properties and Network Topologies*, Springer, 2006.

[Wan04] P. Wan and C. Yi, "Asymptotic Critical Transmission Radius and Critical Neighbor Number for k-Connectivity in Wireless Ad Hoc Networks," In MobiHoc'04.

[Zhang05] H. Zhang and J. Hou, "On the Critical Total Power for Asymptotic k-Connectivity in Wireless Networks," In INFOCOM'05.