Percolation Theory and Network Connectivity

Jie Gao
Computer Science Department
Stony Brook University

STONY BROOK
STATE UNIVERSITY OF NEW YORK

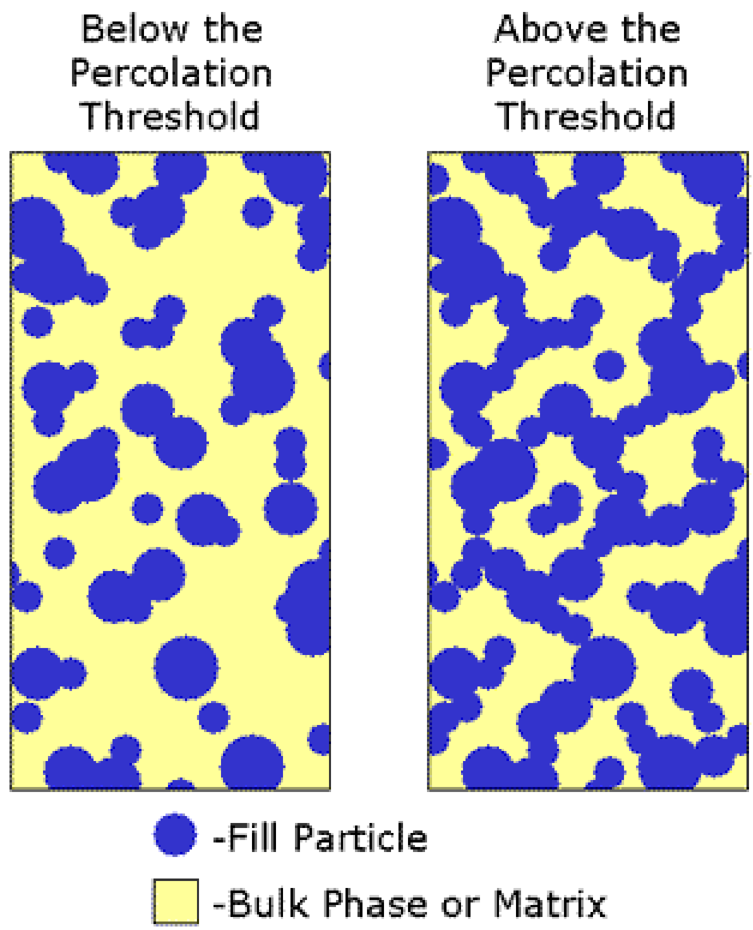
Papers

- Geoffrey Grimmett, **Percolation**, first chapter, Second edition, Springer, 1999.
- Massimo Franceschetti, Lorna Booth, Matthew Cook, Ronald Meester, and Jehoshua Bruck, **Continuum percolation with unreliable and spread out connections**, Journal of Statistical Physics, v. 118, N. 3-4, February 2005, pp. 721-734.

On a rainy day

- Observe the raindrops falling on the pavement. Initially the wet regions are isolated and we can find a dry path. Then after some point, the wet regions are connected and we can find a wet path.

- There is a critical density where sudden change happens.



Phase transition

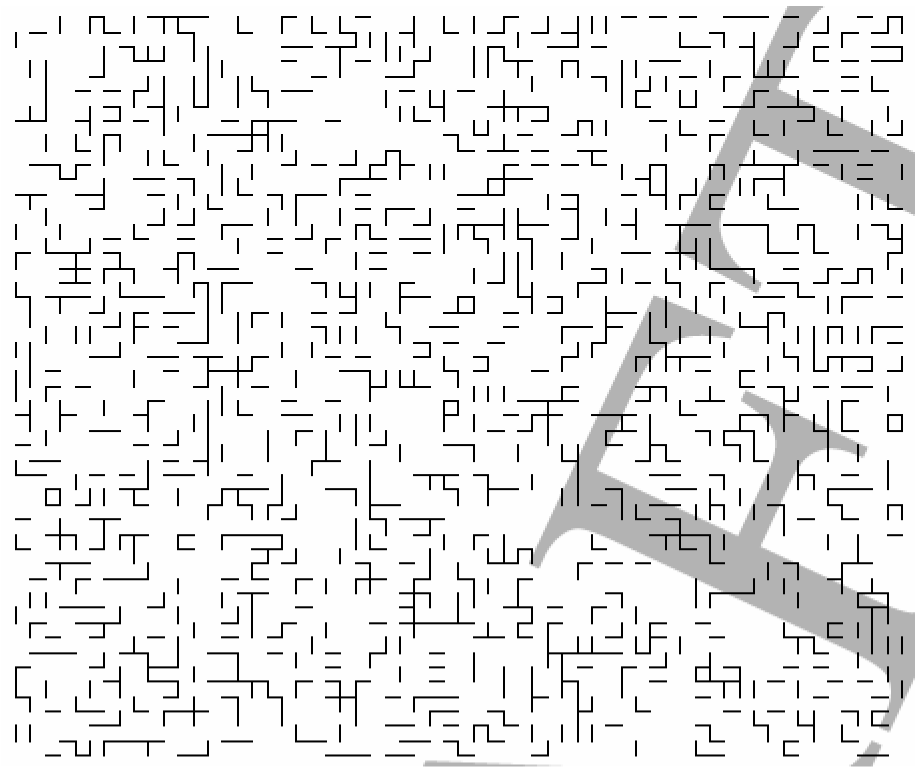
- In physics, a **phase transition** is the transformation of a thermodynamic system from one **phase** to another. The distinguishing characteristic of a **phase transition** is an abrupt sudden change in one or more physical properties, in particular the heat capacity, with a small change in a thermodynamic variable such as the temperature.

- Solid, liquid, and gaseous phases.
- Different magnetic properties.
- Superconductivity of medals.

- This generally stems from the interactions of an extremely large number of particles in a system, and does not appear in systems that are too small.

Bond Percolation

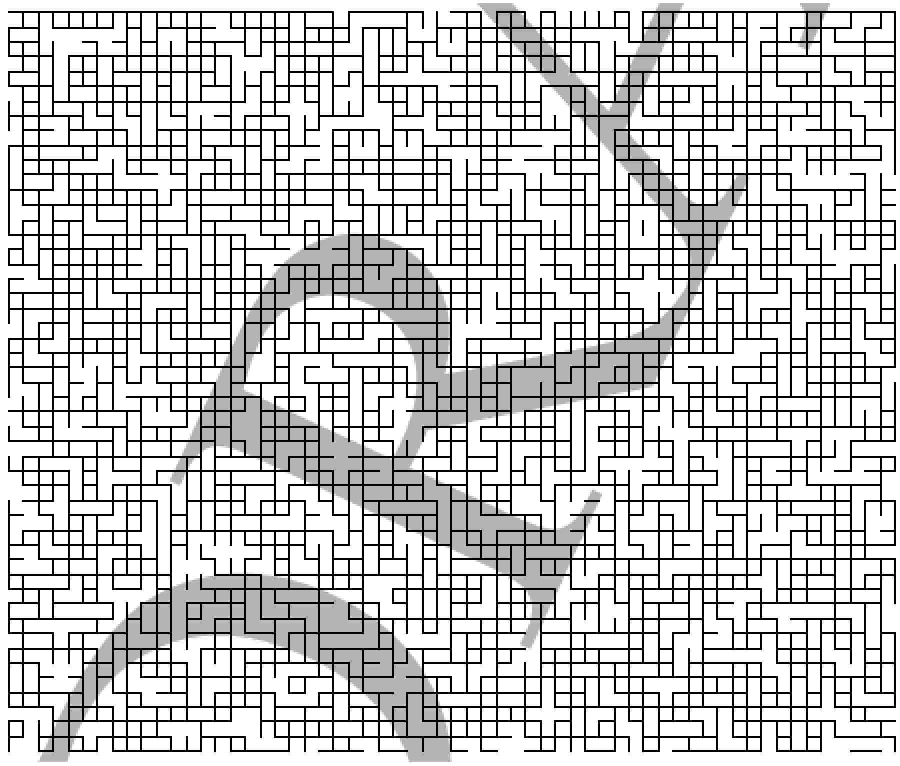
- An infinite grid Z^2, with each link to be "open" (appear) with probability p independently. Now we study the connectivity of this random graph.



p=0.25

Bond Percolation

- An infinite grid Z^2, with each link to be "open" (appear) with probability p independently. Now we study the connectivity of this random graph.

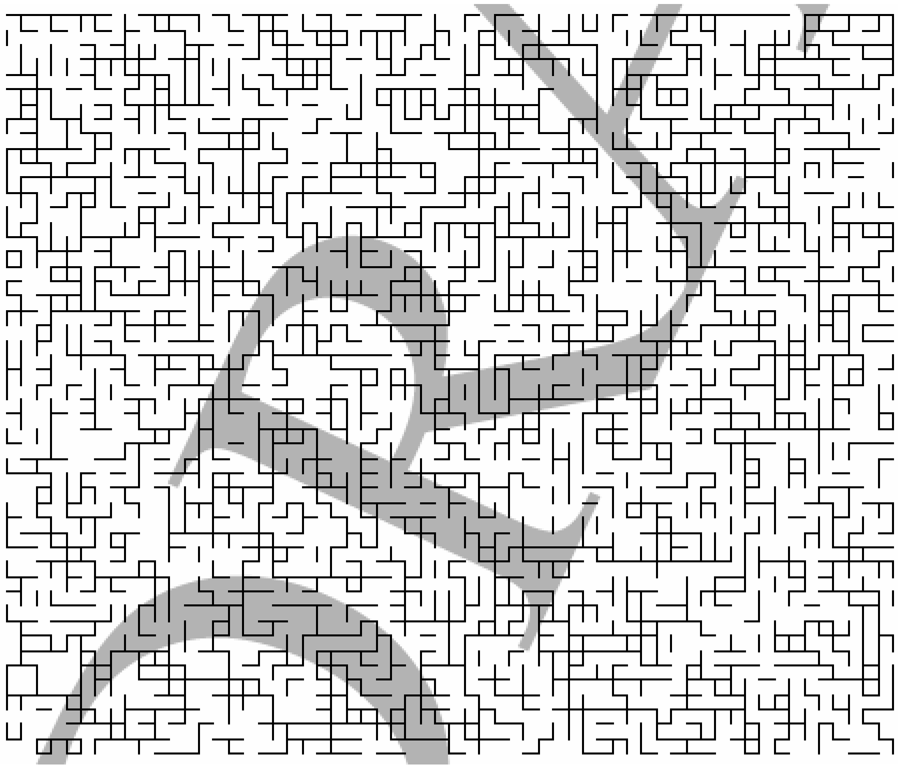


p=0.75

Bond Percolation

- An infinite grid Z^2, with each link to be "open" (appear) with probability p independently. Now we study the connectivity of this random graph.

No path from left to right

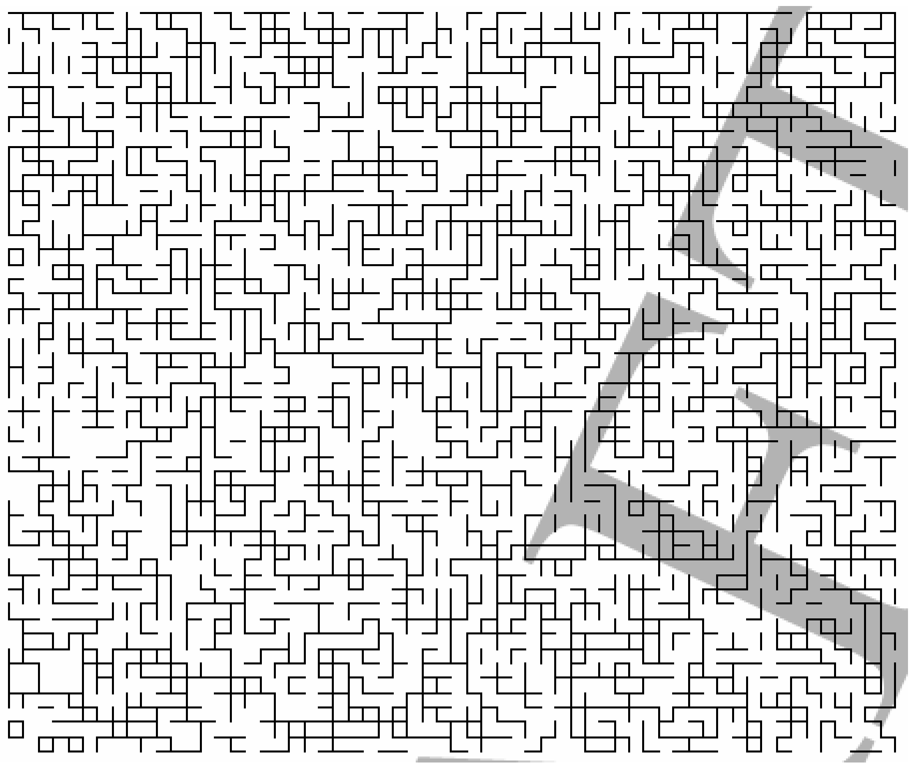


p=0.49

Bond Percolation

- An infinite grid Z^2, with each link to be “open” (appear) with probability p independently. Now we study the connectivity of this random graph.

There is a path from left to right!

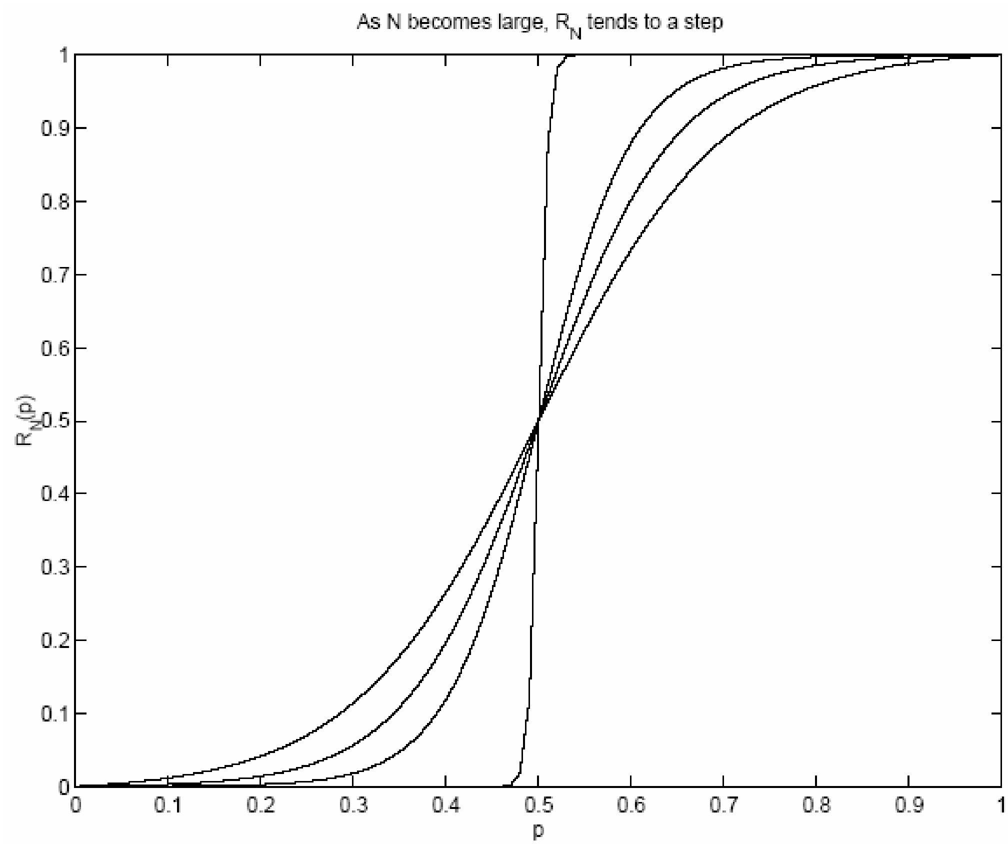


p=0.51

Bond Percolation

- There is a critical threshold p=0.5.

The probability that there is a "bridge" cluster that spans from left to right.

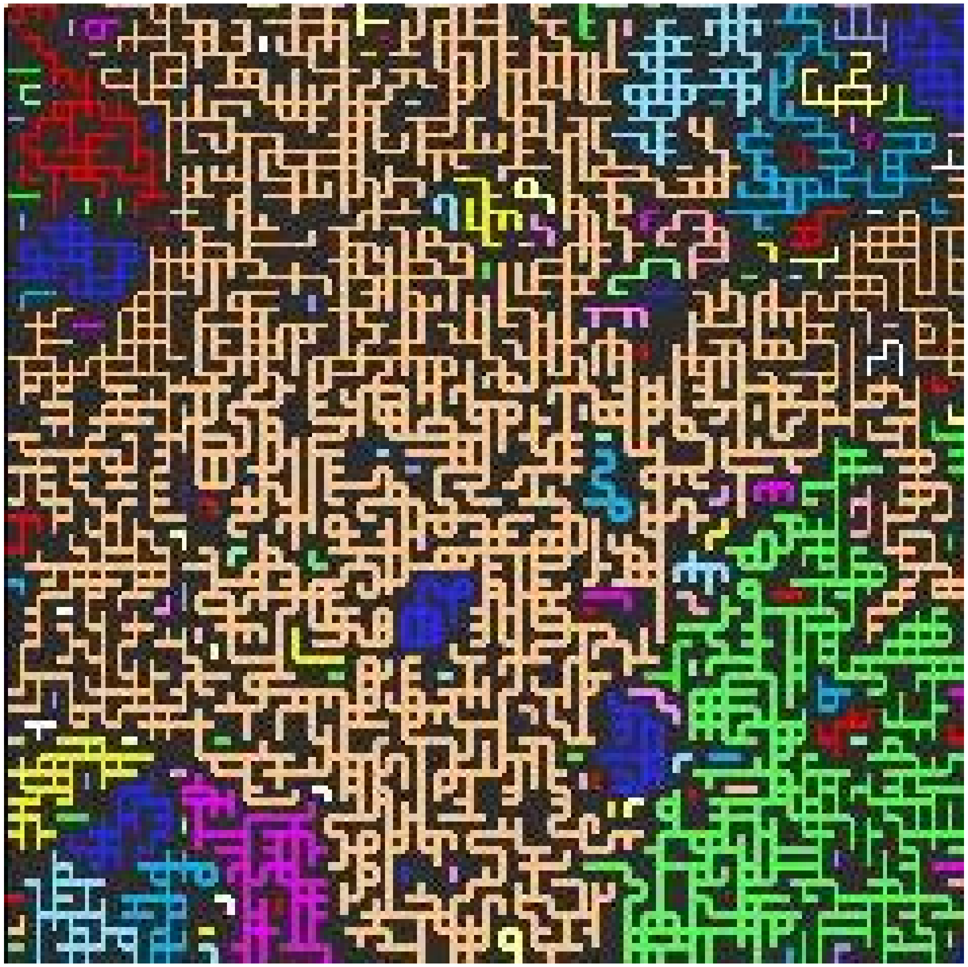


Bond Percolation

- There is a critical threshold p=0.5.
- When p>0.5, there is a unique infinite size cluster almost always.
- When p<0.5, there is no infinitely size cluster.
- When p=0.5, the critical value, there is no infinite cluster.

- Percolation theory studies the phase transition in random structures.

Infinite cluster ≠ connected graph



Main problems in percolation

- The critical threshold for the appearance of some property, e.g., an infinite cluster?
- Behavior below the threshold:
  - We know all clusters are finite. How large are they? Distribution of the cluster size?
- Behavior above the threshold?
  - We know there exists an infinite cluster? Is it unique? What is the asymptotic size with respect to p and n (the network size)?
- Behavior at the threshold?
  - Is there an infinite cluster or not? What is the size of the clusters?

Examples of Percolation

- **Spread of epidemics, virus infection on the Internet.**
  - Each “sick” node has probability p to infect a neighbor node.
  - Denote by p the contagious parameter. If p is above the percolation threshold, then the disease will spread world wide.
  - The real model is more complicated, taking into account the time variation, healing rate, etc.
- **Gossip-based routing, content distribution in P2P network, software upgrade.**
  - The graph is important in deciding the critical value.
  - An interesting result is about the “scale-free” graphs (also called power-law) that model the topology of the Internet or social network: in one of such models (random attachment with preferential rule), the percolation threshold vanishes.

More examples

- **Connectivity of unreliable networks.**
  - Each edge goes down randomly.
  - Is there a path between any two nodes, with high probability?
  - Resilience or fault tolerance of a network to random failures.
- **Random geometric graph, density of wireless nodes (or, critical communication range).**
  - Wireless nodes with Poisson distribution in the plane.
  - Nodes within distance r are connected by an edge.
  - There is a critical threshold on the density (or the communication range) such that the graph has an infinitely large connected component.

Bond percolation

- A grid Z^d, each edge appears with probability p.

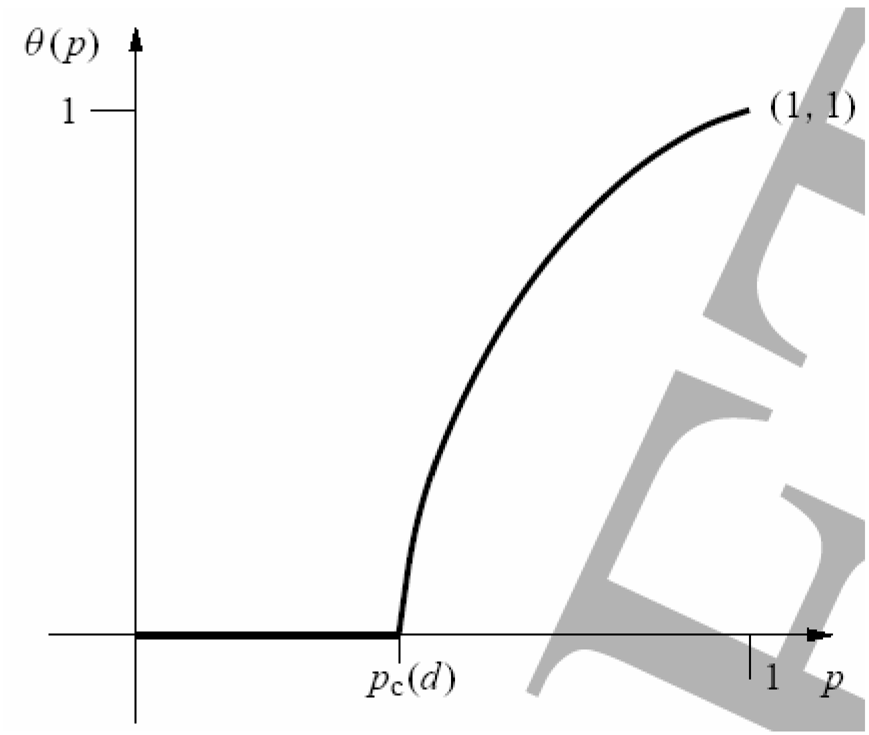
- C(x): the cluster containing the grid node x.
- By symmetry, the shape of C(x) has the same distribution as the shape of C(0), where 0 is the origin.

- $\theta(p)$: the probability that C(0) has infinite size.
- Clearly, when p=0, $\theta(p)=0$, when p=1, $\theta(p)=1$.

- Percolation theory: there exists a threshold $p_c(d)$ such that
  - $\theta(p)>0$, if $p> p_c(d)$;
  - $\theta(p)=0$, if $p< p_c(d)$.

Bond percolation

- This is people's belief on the percolation probability θ(p), It is known that θ(p) is a continuous function of p except possibly at the critical probability. However, the possibility of a jump at the critical probability has not been ruled out when $3 \leq d < 19$.



An easy case:1D

- 1D case: a line. Each edge has probability p to be turned on.
- If p<1, there are infinitely many missing edges to the left and to the right of the origin. Thus $\theta(p)=0$.
- The threshold $p_c(1) = 1$.
- For general d-dimensional grid Z^d, it can be embedded in the (d+1)-dimensional grid Z^{d+1}.
- Thus if the origin belongs to an infinite cluster in Z^d, it also belongs to an infinite cluster in Z^{d+1}.
- This means: $p_c(d+1) \leq p_c(d)$. In fact it can be proved that $p_c(d+1) < p_c(d)$.

2d: interesting things start to happen

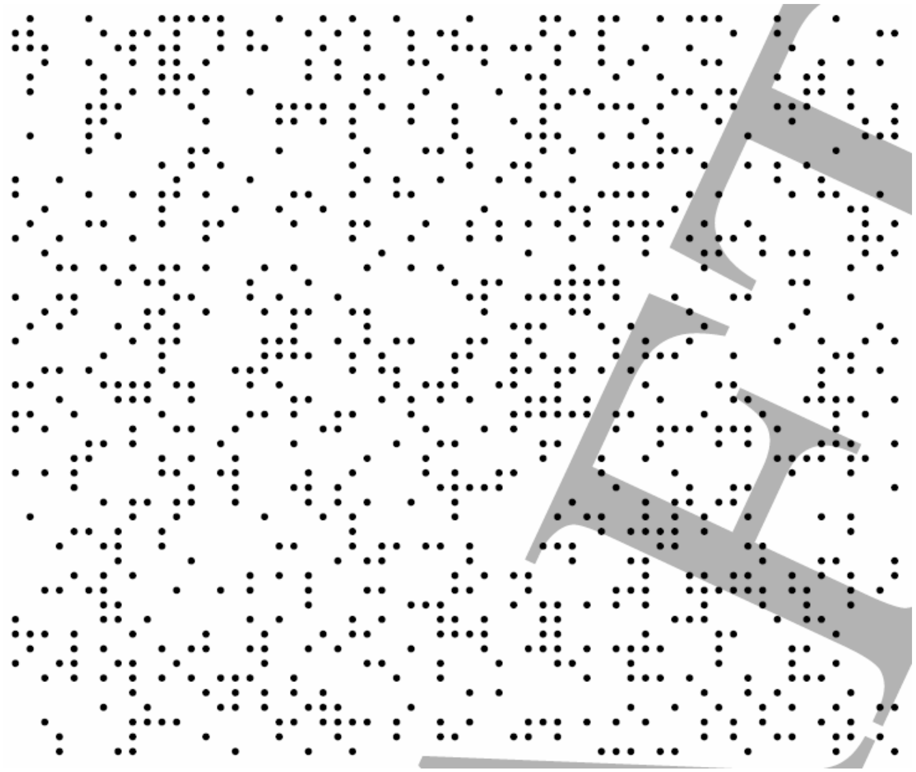
- Theorem: For $d \geq 2$, $p_c(d) = 1/2$.

- There are 2 phases:
- Subcritical phase, $p < p_c(d)$, $\theta(p)=0$, every vertex is almost surely in a finite cluster. Thus all the clusters are finite.
- Supercritical phase, $p > p_c(d)$, $\theta(p)>0$, every vertex has a strictly positive probability of being in an infinite cluster. Thus there is almost surely at least one infinite cluster.

- At the critical point: this is the most interesting part. Lots of unknowns.
- For $d=2$ or $d \geq 19$, there is no infinite cluster. The problem for the other dimensions is still open.

Site Percolation

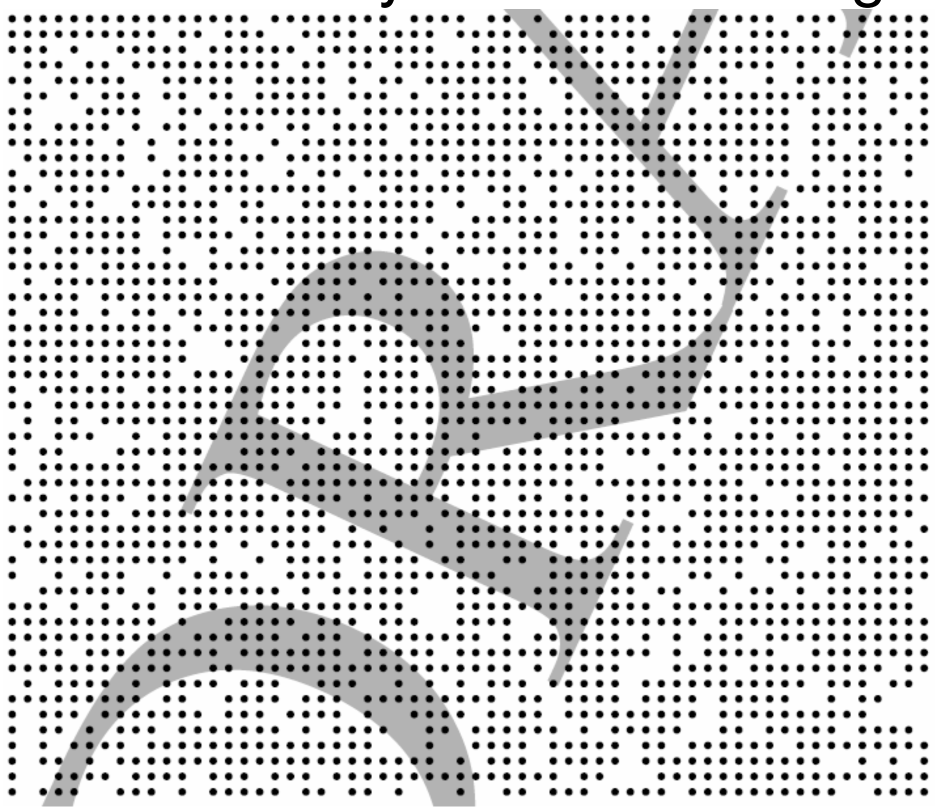
- An infinite grid Z^2, with each vertex to be "open" (appear) with probability p independently. Now we study the connectivity of this random graph.



p=0.3

Site Percolation

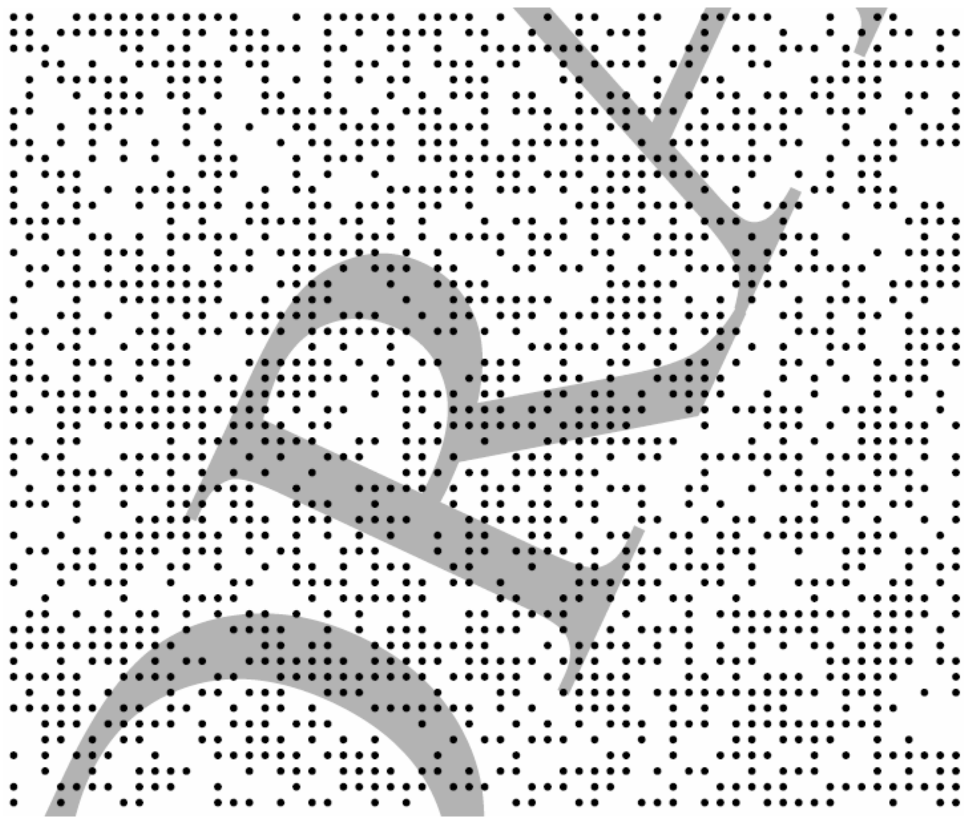
- An infinite grid Z^2, with each vertex to be "open" (appear) with probability p independently. Now we study the connectivity of this random graph.



p=0.80

Site Percolation

- Percolation threshold is still unknown. Simulation shows it’s around 0.59. (note this is larger than bond percolation)



p=0.58

Site Percolation

- Site percolation is a generalization of bond percolation.
- Every bond percolation can be represented by a site percolation, but not the other way around.

- Percolation in an infinite connected graph G(V, E).
- Bond percolation: each edge appears with probability p.
- Site percolation: each vertex appears with probability p.

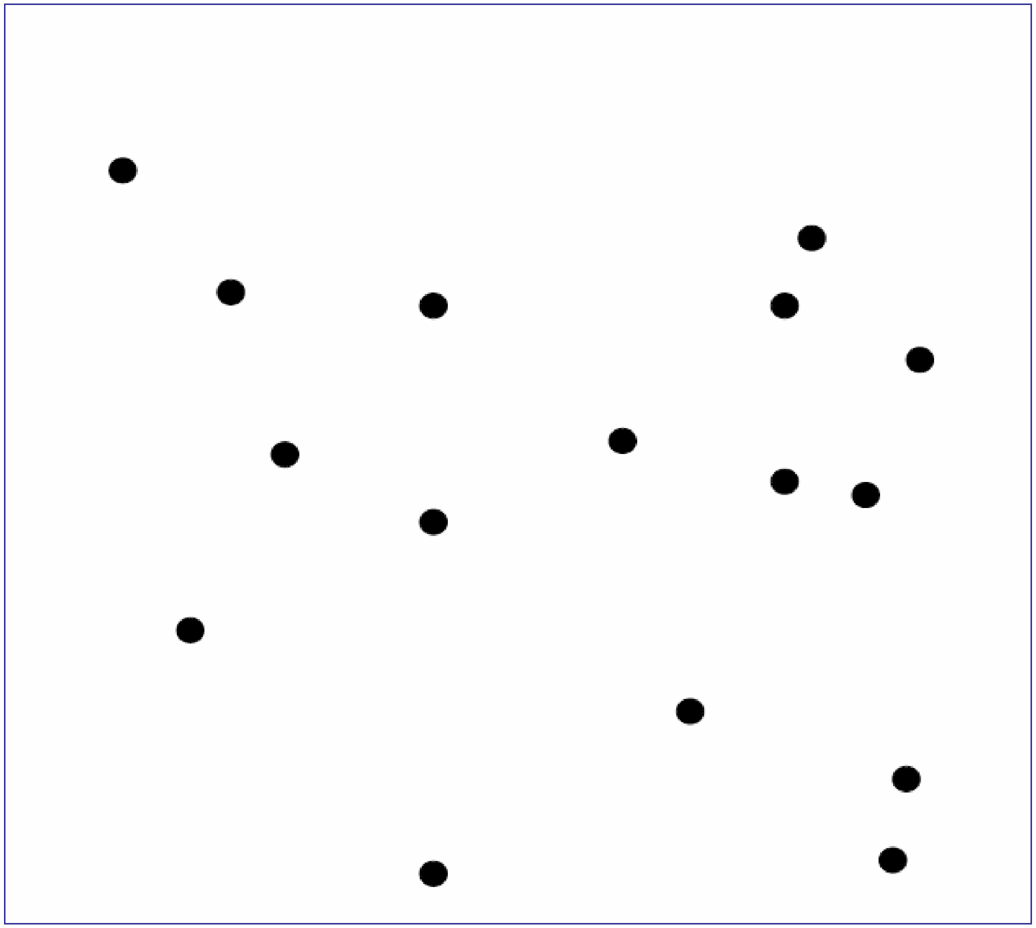
- Denote an arbitrary node as origin, study the cluster containing the origin.

- The percolation threshold of site percolation is always larger than bond percolation.

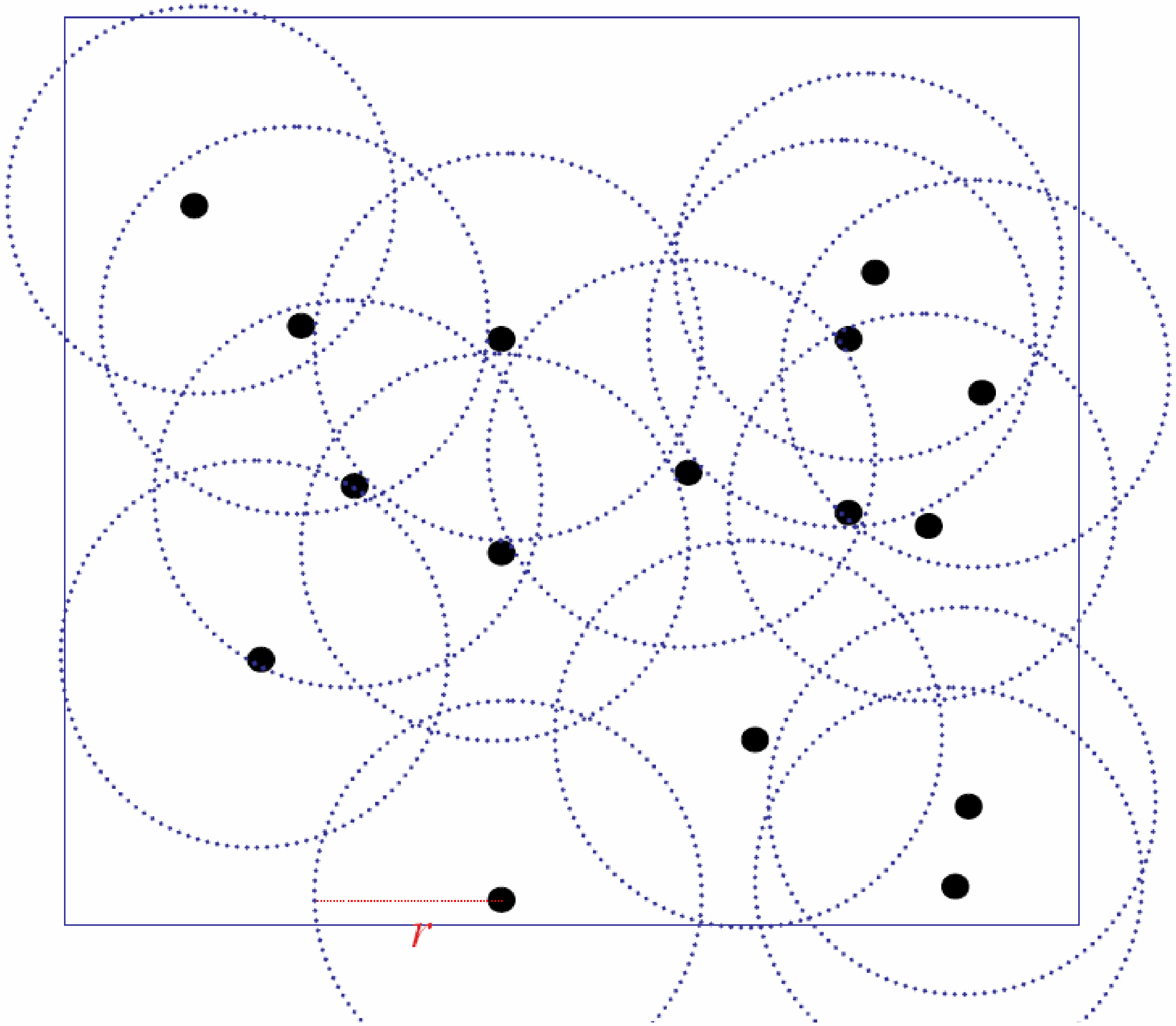
Continuum Percolation

- **Random plane network**, by Gilbert, in J. SIAM 1961.
- Pick points from the plane by a Poisson process with density λ points per unit area.
- Join each pair of points if they are at distance less than r.
- Equivalently,
- In the unit square [0, 1] by [0, 1], throw n points uniformly randomly.
- Connect two nodes with distance less than r.
- This graph is denoted as G(n, r).

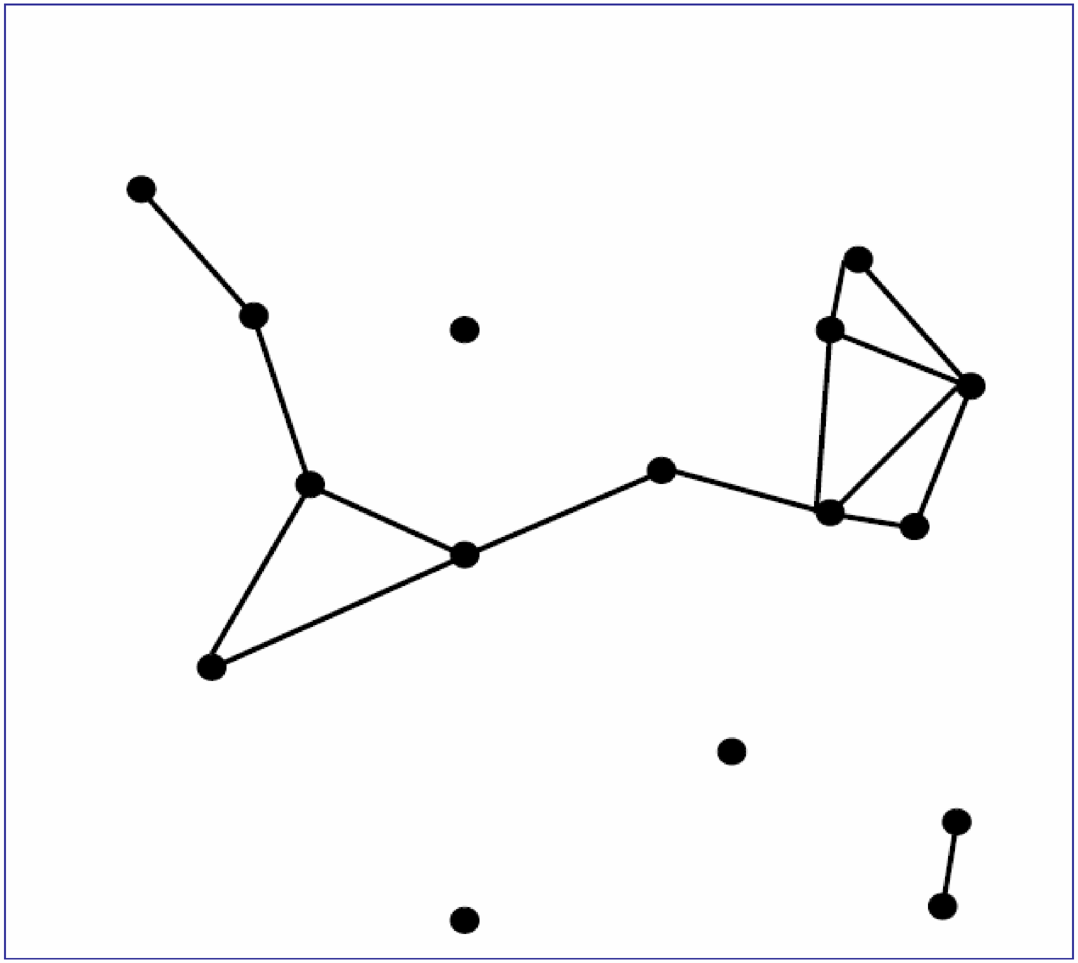
Random geometric graph



Random geometric graph



Random geometric graph



Connectivity in random geometric graphs

- G(n, r): n nodes randomly distributed in a unit square, each node has transmission range r.
- Claim: the network is connected with r=

$$\Theta(\sqrt{\frac{\log n}{n}})$$

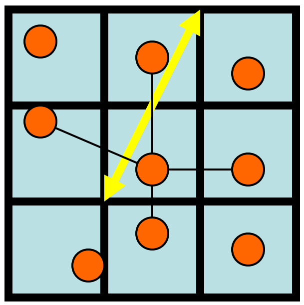
- Proof:
  - Partition the square into small squares of side length

$$\alpha(n) = \sqrt{2\frac{\log n}{n}}$$

  - # squares = n/(2logn).

Connectivity in random geometric graph

- Proof cont.
  - Now m=n/(2logn) squares. Throw n=2mlogn nodes randomly in the squares.
  - With high probability each square gets at least one node --- coupon collector problem.
  - Set $$r(n) = \sqrt{10\frac{\log n}{n}}$$
  - r(n)=5α(n)
  - Each node can connect to its neighboring 4 squares.
  - Thus the graph is connected.



Coupon collector problem

- You want to gather n coupons. Each time you receive a random one. How many coupons do you get until you get coupons of all kinds?
- Answer: nlogn.
  - N_i = time until one gets i different coupons.
  - A random draw gives a new coupon with prob=(n-i)/n
  - $E(N_{i+1}-N_i)$ = Expected time to get a new coupon, with already i different coupons = n/(n-i).
  - Thus $E(N_n)=1+n/(n-1)+n/(n-2)+\ldots+n$ =nlogn.
  - High probability analysis: see notes.

Random geometric graph v.s. random graph

- Erdos-Renyi model of random graphs (Bernoulli random graphs): each pair of vertices is connected by an edge with probability p.
- Random geometric graph: the probability is dependent on the distance.
- One of the main question in random graph theory is to determine when a given property is likely to appear.
  - Connectivity.
  - Chromatic number.
  - Matching.
  - Hamiltonian cycle, etc.

Random geometric graph v.s. random graph

- Erdos-Renyi model of random graphs (Bernoulli random graphs): each pair of vertices is connected by an edge with probability p.

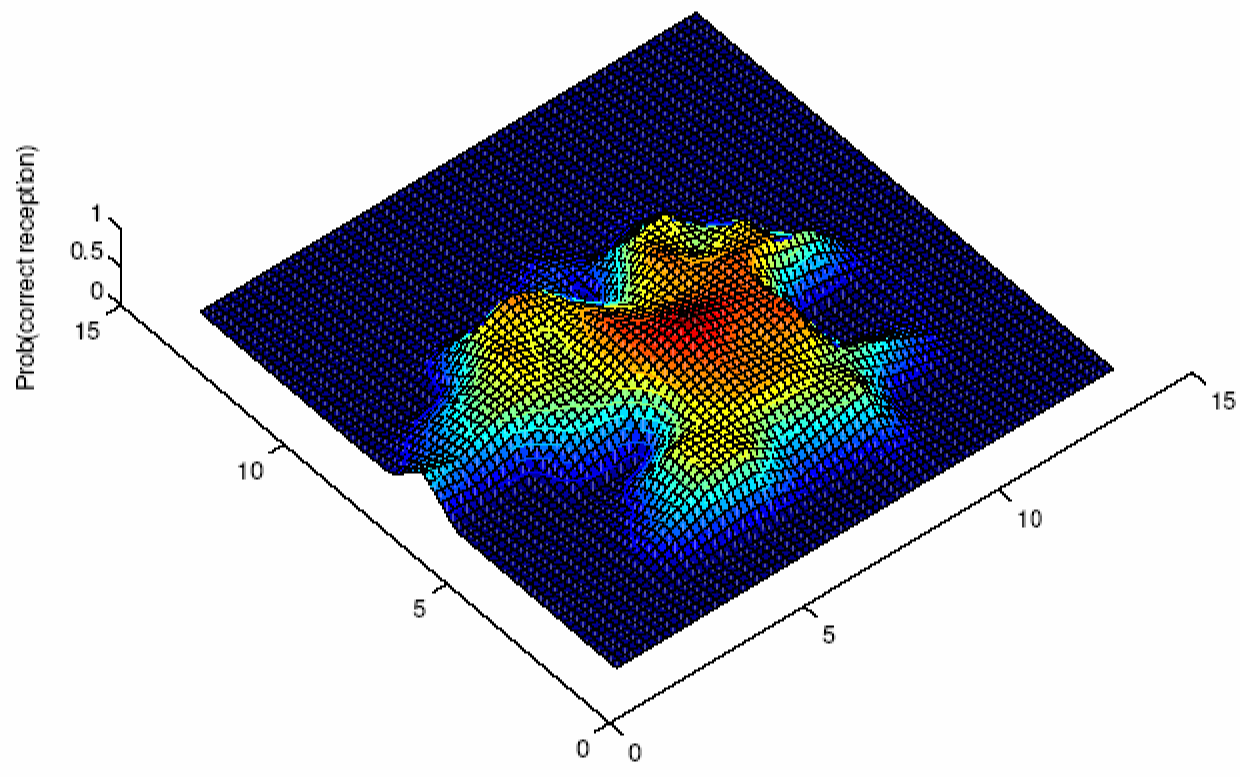
- Friedgut and Kalai in 1996 proved that all monotone graph properties have a sharp threshold in Bernoulli random graphs.

- Monotone graph property P: more edges do not hurt the property.

- This is also true in random geometric graphs. Proved by Ashish Goel, Sanatan Rai and Bhaskar Krishnamachari, in STOC 2004.

Percolation in the real world?

- Communication range is not a perfect disk.



Percolation with noisy links

- Each pair of nodes is connected according to some (probabilistic) function of their (random) positions.
- A pair of points (i, j) is connected with probability $g(x_i - x_j)$, where g is a general function that depends only on the distance.

- In order to keep the average degree the same, fix the effective area $e(g) = \int_{x \in \mathbb{R}^2} g(x) dx$

- The average degree = λ e(g).

Percolation with noisy links

- Percolation threshold

$$0 < \lambda_c(g) = \inf\{\lambda : \exists \text{ infinite connected component a.s.}\} < \infty.$$

- Question: what is the relationship between the percolation threshold and the function g?

Percolation with noisy links

- Question: what is the relationship between the percolation threshold and the function g?

- Each node is connected to the same number of edges on average. So whom should the node be connected to, in order to have a small percolation threshold?

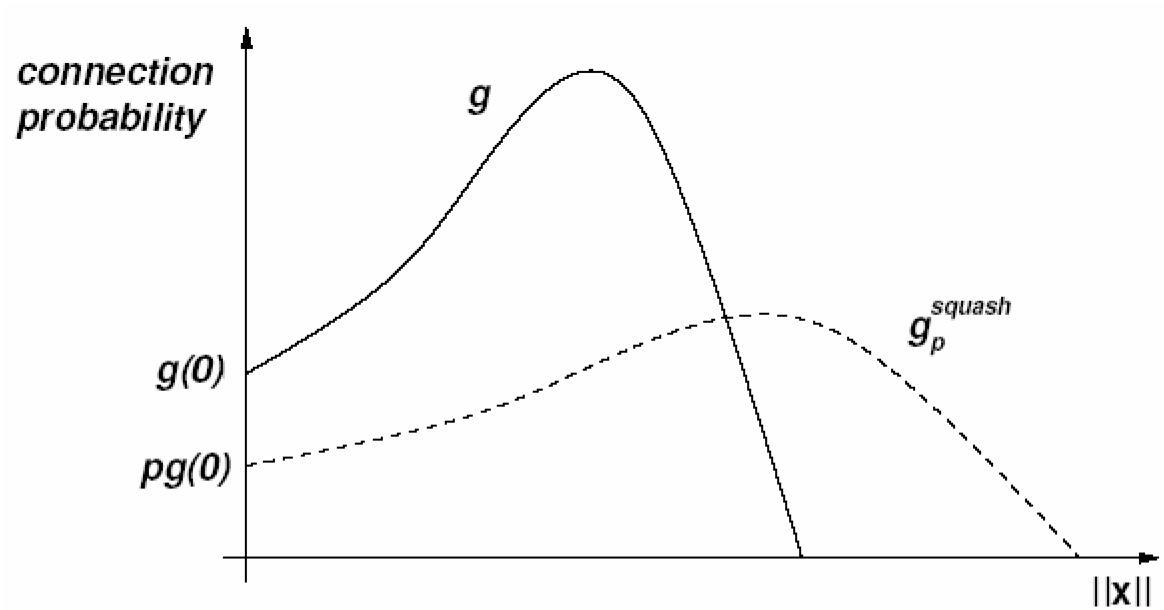
- Which distribution has the best graph connectivity?

- Should I use reliable short links? Or unreliable long links? Or something more complex, say an annulus?

Squashing

- Probabilities are reduced by a factor of p, but the function is spatially stretched to maintain the same effective area (e.g., the same average degree).

$$g_p^{squash}(x) = p \cdot g(\sqrt{p}x).$$



Squashing

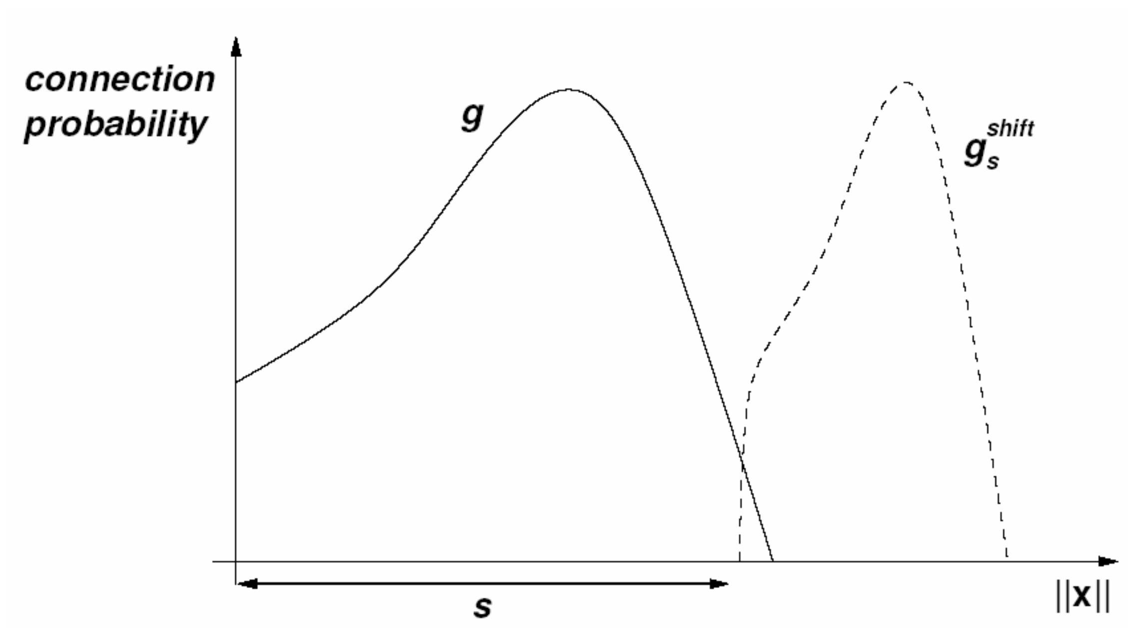
- Probabilities are reduced by a factor of p, but the function is spatially stretched to maintain the same effective area (e.g., the same average degree).

$$g_p^{squash}(x) = p \cdot g(\sqrt{p}x).$$

- Theorem: $$\lambda_c(g) \geq \lambda_c(g_p^{squash}).$$
- It's beneficial for the connectivity to use long unreliable links!
- If the effective area is spread out, then the threshold density goes to 1.
- Question: what makes the difference? The guess is the existence of long links.

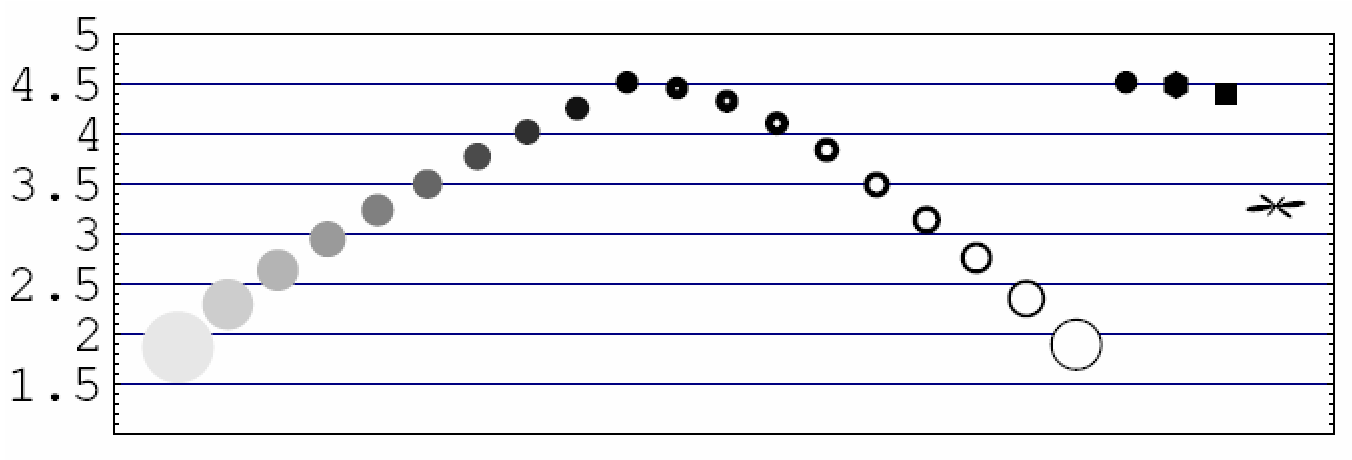
Shifting and squeezing

- Shift the function g outward by a distance s, but squeeze the function after that, so that it has the same effective area.
- Goal: use long links.



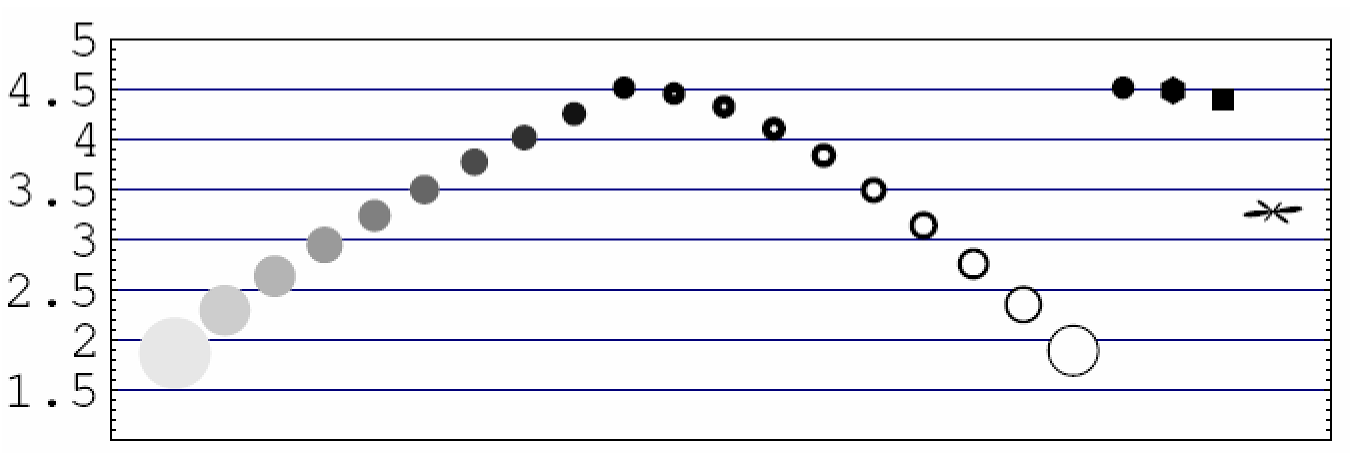
Shifting and squeezing

- Yes it helps percolation! The density threshold goes down.



Connections to points in an annulus

- Points are distributed in the plane by a Poisson process with density λ. Each node is connected to all the nodes inside an annulus A(r) with inner radius r and area 1.
- Theorem: for any critical density λ, one can find a r such that any density above the threshold percolates.



Summary

- Interesting things happen when the # elements is large.
- Useful in simulations – it helps you to understand the results better.
- You can write a short program to visualize and understand percolation.