Phase Transitions in Random Geometric Graphs, with Algorithmic Implications

Ashish Goel

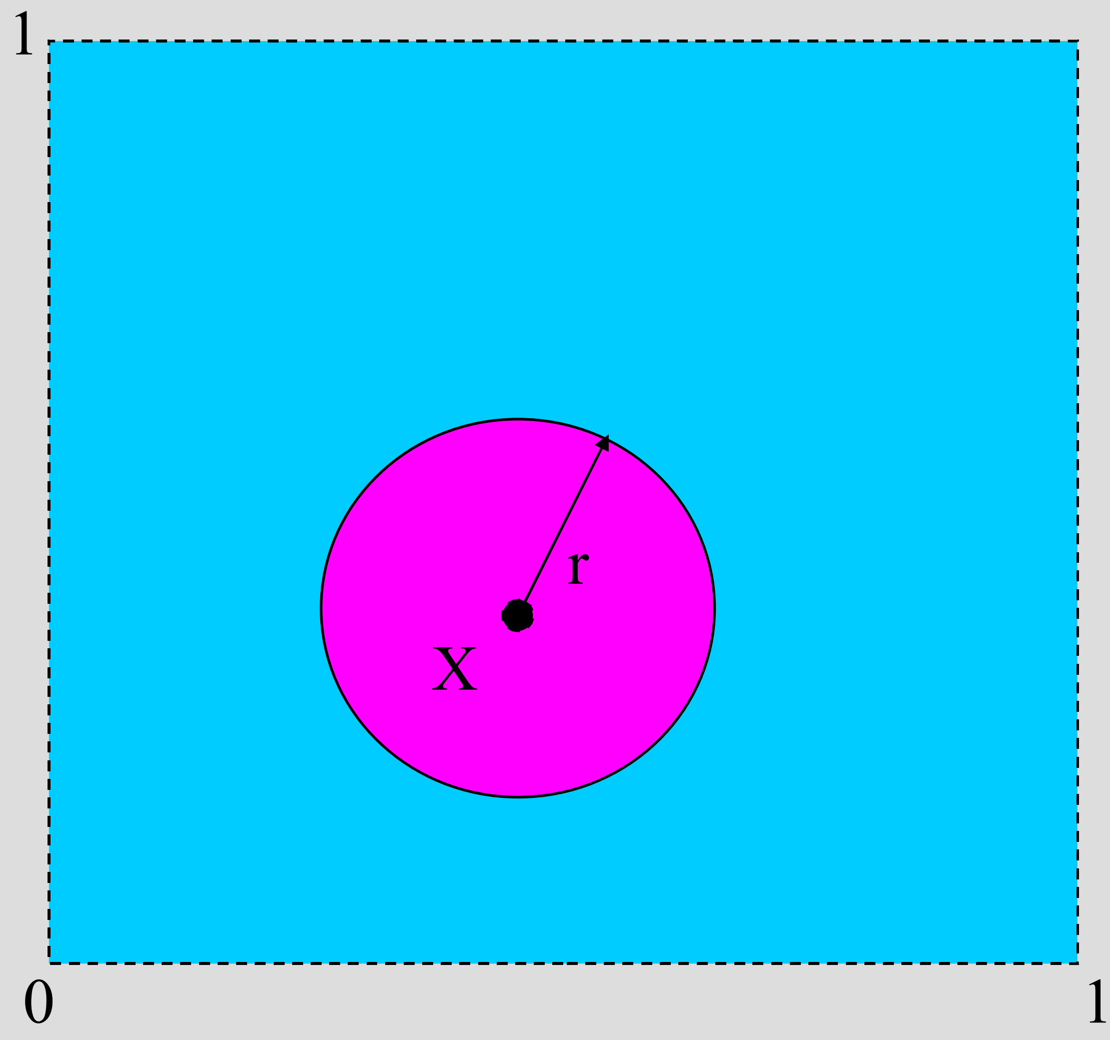
Stanford University

Joint work with Sanatan Rai and Bhaskar Krishnamachari

http://www.stanford.edu/~ashishg

Geometric Random Graphs

- $G(n;r)$ in d-dimensions:
  - n points uniformly distributed in $[0,1]^d$
  - Two points are connected if their Euclidean distance is less than r
- Sensor networks can often be modeled as $G(n;r)$ with $d=2$
  - Eg. sensors “sprinkled” from a helicopter over a corn field
  - The wireless radius corresponds to r
- Question: How should n and r be chosen to ensure that $G(n;r)$ has a desirable property Π (eg. connectivity, 2-connectivity, large cliques) with high probability?



Any other point Y is a neighbor of X with probability πr^2

Expected degree of X is $\pi r^2 (n-1)$

Thresholds for monotone properties?

- A graph property Π is monotone if, for all graphs $G_1=(V,E_1)$ and $G_2=(V,E_2)$ such that E_1 µ E_2,

  G_1 satisfies Π) G_2 satisfies Π

  - Informally, addition of edges preserves Π

- Examples: connectivity, Hamiltonianicity, bounded diameter, expansion, degree ¸ k, existence of minors, k-connectivity ….

- Folklore Conjecture: All monotone properties have “sharp” thresholds for geometric random graphs

  [Krishnamachari, PhD Thesis ’02]

Example: Connectivity

Define c(n) such that $\pi\, c(n)^2 = \log n/n$

- Asymptotically, when d=2
  - G(n;c(n)) is disconnected with high probability
  - For any $\varepsilon > 0$, G(n; $(1+\varepsilon)$c(n)) is connected whp
  - So, $c(n) = \sqrt{\frac{\log n}{\pi n}}$ is a "sharp" threshold for connectivity at d=2 [Gupta and Kumar '98; Penrose '97]

- Similar thresholds exist for all dimensions
  - $c_d(n)$ ¼ $(\log n/(nV_d))^{1/d}$, where V_d is the volume of the unit ball in d dimensions
  - Average degree ¼ log n at the threshold

Width and sharp thresholds

- For property Π, and $0 < \varepsilon < 1$, if there exist two functions $L(n)$ and $U(n)$ such that

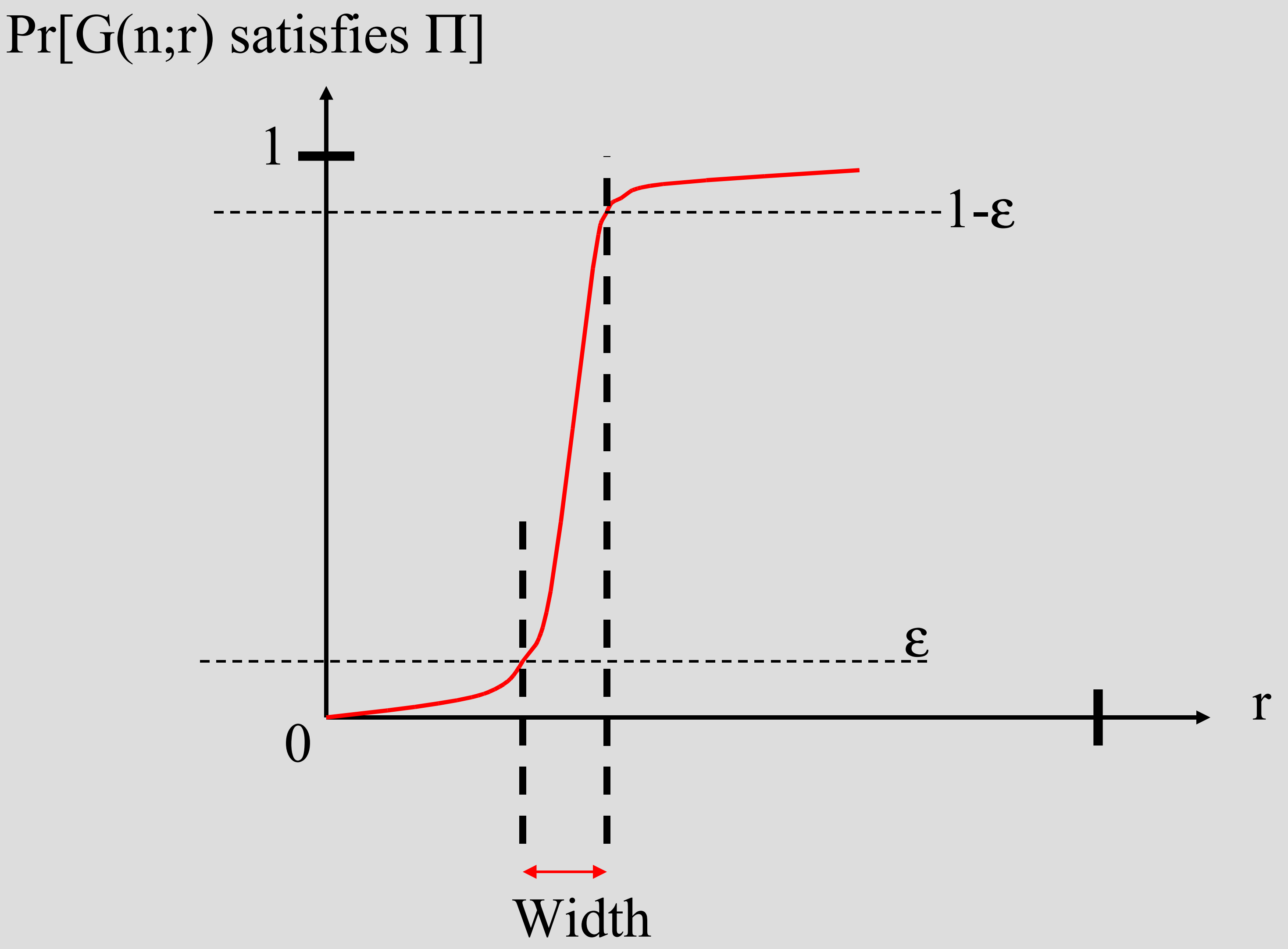
  $Pr[G(n;L(n))$ satisfies $\Pi] = \varepsilon$, and

  $Pr[G(n;U(n))$ satisfies $\Pi] = 1 - \varepsilon$,

then the ε-width $w_\varepsilon(n)$ of Π is defined as $U(n)-L(n)$

- If $w_\varepsilon(n) = o(1)$ for all ε, then Π is said to have a sharp threshold

Example



Connections (?) to Bernoulli Random Graphs

- Famous graph family $G(n;p)$
  - Also known as Erdos-Renyi graphs
  - Edges are iid; each edge present with probability p
  - Connectivity threshold is $p(n) = \log n/n$
    - Average degree exactly the same as that of geometric random graphs at their connectivity threshold!!
  - All monotone properties have ε-width = $O(1/\log n)$ for any fixed ε in the Bernoulli graph model

    [Friedgut and Kalai '96]
  - Can not be improved beyond $O(1/\log^2 n)$
    - Almost matched [Bourgain and Kalai '97]
  - Proof relies heavily on independence of edges
    - There is no edge independence in geometric random graphs => we need new techniques

Our results

$$\mathbf{c}_d(n) = \Theta\left(\frac{\log^{1/d} n}{n^{1/d}}\right)$$

- The ε-width of any monotone property is

$$\begin{array}{rll} O(c_d(n)) & \text{for} & d \geq 3 \\ O(c_2(n)\log^{1/4} n) & \text{for} & d = 2 \\ O(\sqrt{(\log 1/\epsilon)/n}) & \text{for} & d = 1 \end{array}$$

  - Sharp thresholds in the geometric random graph model
  - Sharper transition (inverse polynomial width) than Bernoulli random graphs (inverse logarithmic width)

- There exist monotone properties with width

$$\begin{array}{rll} \Omega(c_d(n)/\log^{1/d} n) & \text{for} & d \geq 3 \\ \Omega(c_2(n)/\log^{1/2} n) & \text{for} & d = 2 \\ \Omega(\sqrt{(\log 1/\epsilon)/n}) & \text{for} & d = 1 \end{array}$$

  - Tight for d=1, sub-logarithmic gap for d>1

Why $c_d(n)$?

- Why express results in terms of $c_d(n)$?
  - Width gives a sharp “additive” threshold
  - We are typically interested in properties that subsume connectivity
  - For such properties, an additive threshold in terms of $c_d(n)$ also corresponds to a “multiplicative” threshold
  - The exact sharpness of the multiplicative threshold depends on $L(n)$ and on the exact additive bounds (details omitted)

Bottleneck Matchings

- Draw n "blue" points and n "red" points uniformly and independently from $[0,1]^d$
  - B_n,R_n denotes the set of blue, red points resp.
- A minimum bottleneck matching between R_n and B_n is a one-one mapping

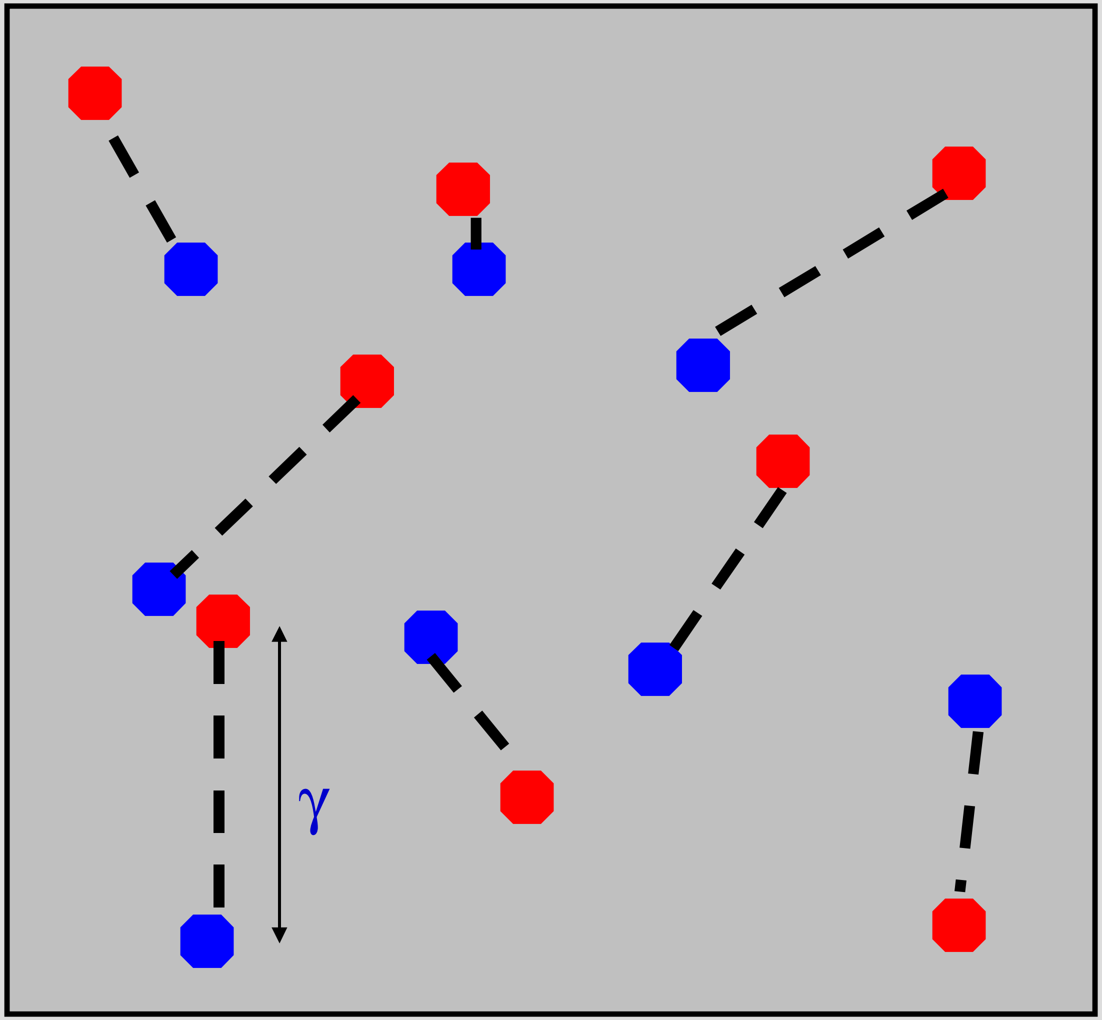
$$f:B_n ! R_n$$

which minimizes

$$\max_{u2\ Bn}||f(u)-u||_2$$

- The corresponding distance ($\max_{u2\ Bn}||f(u)-u||_2$) is called the minimum bottleneck distance
  - Let X_n denote this minimum bottleneck distance

Example



Bottleneck distance = γ

Bottleneck Matchings and Width

Theorem: If $Pr[X_n > \gamma] \cdot p$ then the sqrt(p)-width of any monotone property is at most 2γ

Implication: Can analyze just one quantity, X_n, as opposed to all monotone properties (in particular, can provide simulation based evidence)

Proof: Let Π be any monotone property

- Let ε = sqrt(p)
- Choose $L(n)$ such that $Pr[G(n;L(n)$ satisfies $\Pi] = \varepsilon$
- Define $U(n) = L(n) + 2\gamma$
- Draw two random graphs G_L and G_U (independently) from $G(n;L(n))$ and $G(n;U(n))$, resp.
- Let B_n, R_n denote the set of points in G_L, G_U resp.

Bottleneck Matchings and Width (proof contd.)

Assume $X_n \leq \gamma$.

Let f be the corresponding minimum bottleneck matching between R_n and B_n.

For any $u,v \in B_n$:

$$\|f(u)-f(v)\|_2 \leq \|f(u)-u\|_2 + \|u-v\|_2 + \|f(v)-v\|_2 \leq 2\gamma + \|u-v\|_2$$

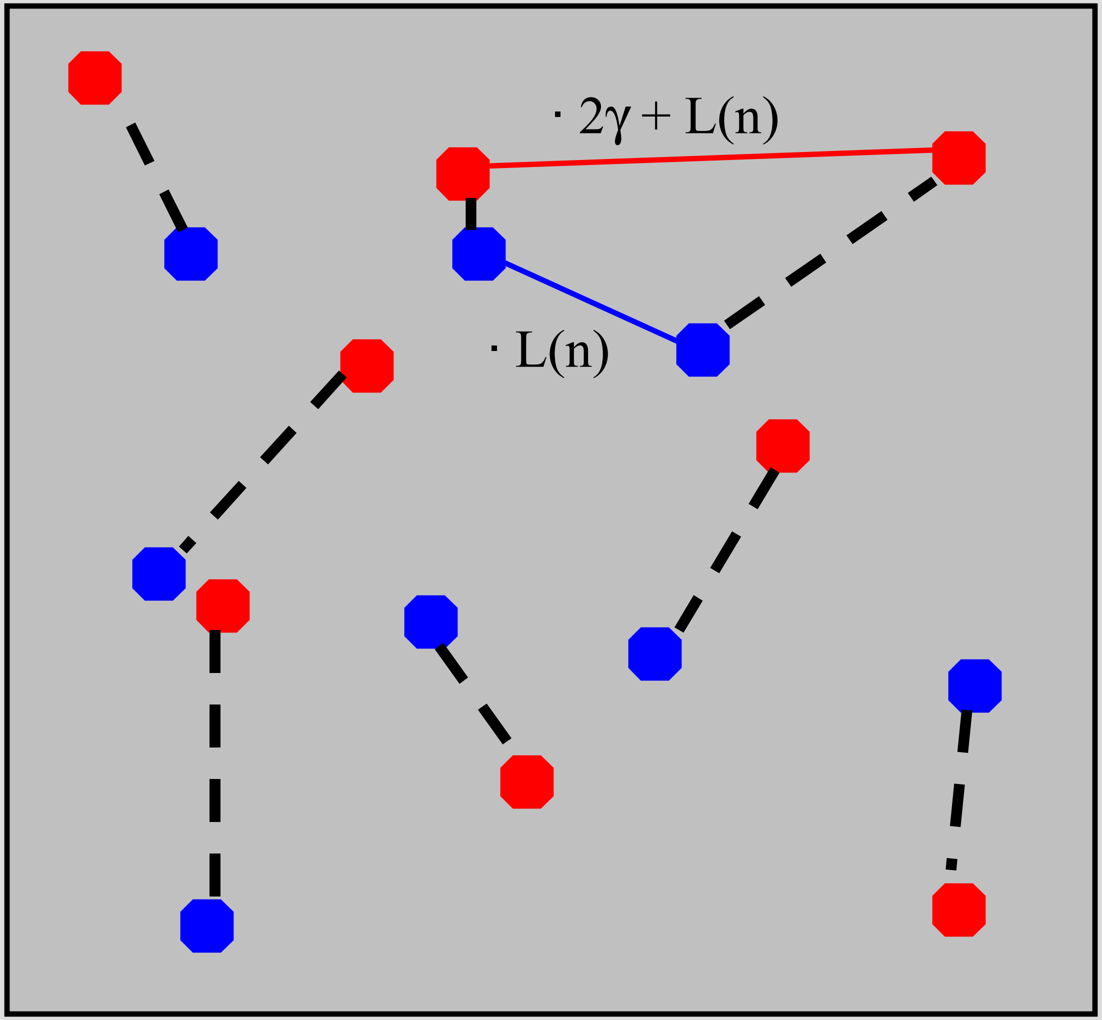
Hence, (u,v) is an edge in $G_L \Rightarrow (f(u),f(v))$ is an edge in G_U

i.e. G_L is a subgraph of G_U

By definition, $\Pr[X_n > \gamma] \leq p$

$\Rightarrow \Pr[G_L$ is not a subgraph of $G_U] \leq p = \varepsilon^2$ (1)

Illustration I: Triangle Inequality



Suppose bottleneck matching · γ

Bottleneck Matchings and Width (proof contd.)

Pr[G_L is not a subgraph of G_U] · p = ε^2 (1)

Let q = Pr[G_U does not satisfy Π]

Π is monotone, Pr[G_L satisfies Π] = ε,

) Pr[G_L is not a subgraph of G_U] ¸ ε¢q (2)

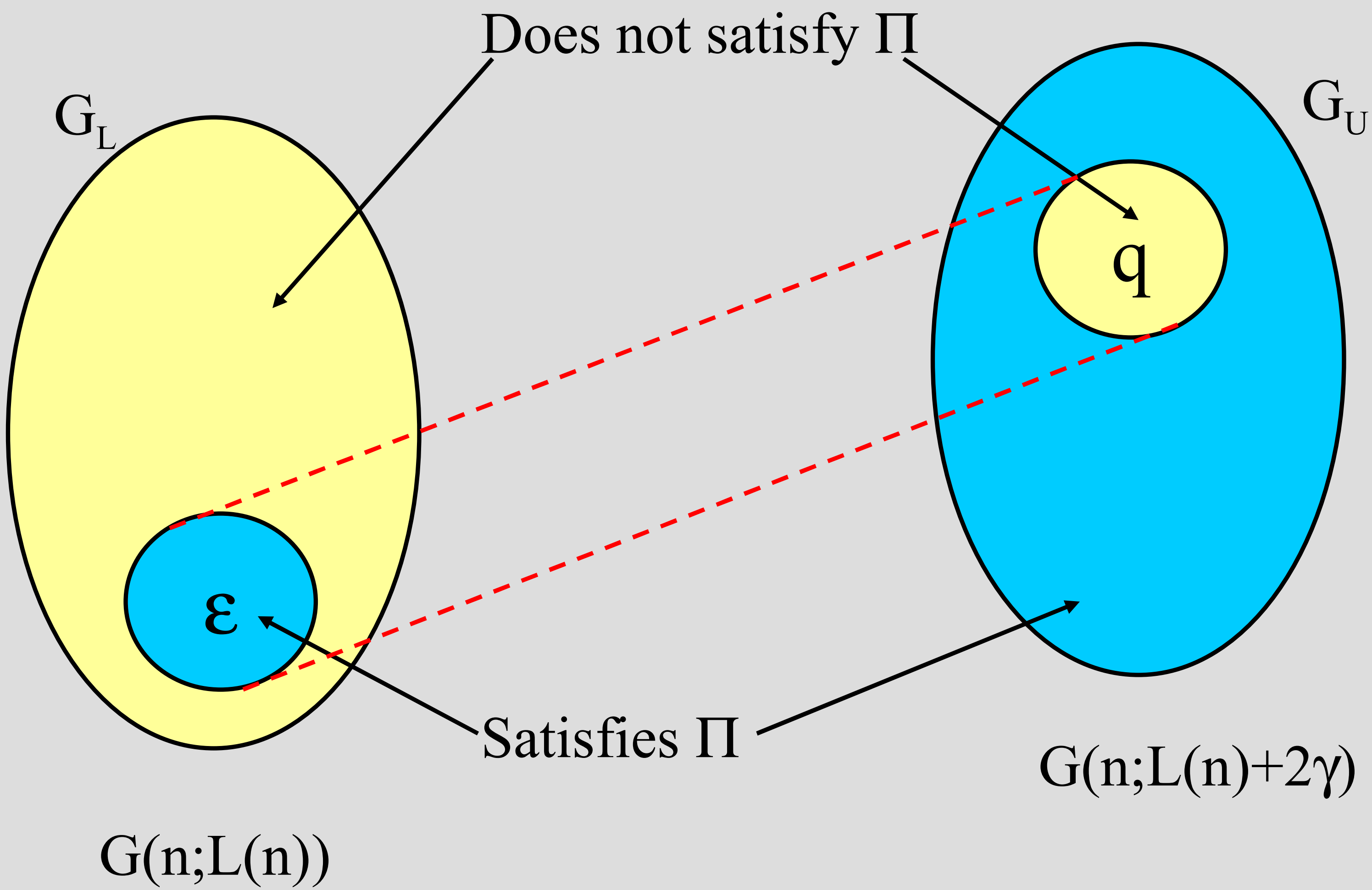
Combining (1) and (2), we get ε¢q · p i.e. q · ε

Therefore, Pr[G_U satisfies Π] ¸ 1-ε

i.e. the ε-width of Π is at most U(n) – L(n) = 2γ

Done!

Illustration II: Probability Amplification



G_L is not a subgraph of G_U with probability at least εq

Our Goal now

Analyze the bottleneck matching distance X_n

Specifically, we are done if $X_n = O(\gamma(n))$

with high probability, for some "small" $\gamma(n)$

Comparison with Bernoulli Random Graphs?

- We are attempting to show something quite strong
  - $G(n;r)$ is a subgraph of $G(n;r+\gamma)$ whp, for small γ
  - "Laminar" structure
- Corresponding result is NOT true for Bernoulli random graphs even for $\gamma = ½$
- If small bottleneck matchings exist whp, we will get stronger thresholds than for Bernoulli random graphs

Existence of Small Bottleneck Matchings

- The bottleneck matching length is

  $O(c_d(n))$ whp for d ¸ 3

  [Shor and Yukich 1991; we present a simpler proof]

  $O(c_2(n) \log^{1/4} n)$ whp for $d = 2$

  [Leighton and Shor 1989]

  O (sqrt(log(1/ε))/sqrt(n)) with probability 1-ε for $d = 1$

  [Our paper (folklore?)]

- This gives us the desired widths
- We will now present the main idea behind our d=1 and d ¸ 3 proofs

Demonstrating Small Bottleneck Matchings: d=1

The Stretch-Shrink-Divide Algorithm

- Let h $x_1, x_2, \ldots, x_n$ i be the coordinates of the red points, in increasing order (assume $n = 2^k$)
  - The coordinates are uniformly distributed in [0,1]
- Multiply the first $n/2$ coordinates by

  $$1/(2x_{n/2+1})$$

  - The first $n/2$ coordinates are now uniformly distributed in [0,1/2]
  - Let δ_1 denote $|1/2 - x_{n/2+1}|$. No point in the left "half" moves by more than δ_1
- Perform a symmetric transformation on the last $n/2$ coordinates (now uniform in [1/2,1])
  - Two regions of equal size and equal density
  - Recurse

For higher d

- Divide using each coordinate in turn
  - After d steps, we have 2^d sets of $n/2^d$ points, each set uniformly distributed in cubes of side ½.
- After $\log n$ steps, there is a red point in each cell of a uniform grid superimposed on the unit cube.
  - Run the same algorithm on the blue points
  - The (unique) red and blue point in each cell are then matched to each other

Analysis: The Basic Idea

- Consider $\delta_1 = |1/2 - x_{n/2+1}|$.
  - Intuition: δ_1 looks like a normal variable
- Lemma: $\text{Prob}[\delta_1 \geq \alpha\beta\, n^{-1/2}] \leq \exp(-\beta^2)$ for an appropriate constant α
- Recursive application of this lemma at different scales gives tight results for $d=1, d \geq 3$
  - Details omitted
  - Shor and Yukich used a similar recursion but did not "re-uniformize", resulting in a more complex proof

Lower bound examples

- For d=1, the property

  Π = {G: 8 v2 V(G), degree(v) ¸ |V(G)|/4}

has width Ω(sqrt(log 1/ε)/sqrt(n))

  - Basic idea: Just the two endpoints on the line are interesting for the purpose of finding the minimum degree

- For d¸ 2, the property Π = {G is a clique} has width Ω $(1/n^{1/d})$
- Open problem: plug the gap in the upper/lower bounds on the width for d ¸ 2
  - Also, all our lower bound examples undergo phase transitions at r = Θ(1). Is there something interesting and different in the region where r is of the order of the connectivity threshold?

Implications – Mixing Time

Recent result: Fastest mixing Markov chain defined on $G(n;r)$ has mixing time $\Theta(r^{-2} \log n)$ for large enough r [Boyd, Ghosh, Prabhakar, Shah ‘05]

Alternate proof:

- $GRID(n;r)$: n points are laid on a grid in $[0,1]^2$ and two points are connected if they are within distance r.
- Fastest mixing time of $GRID(n;r) = \Theta(r^{-2}\log n)$ [Trivial]
- $G(n;r)$ is a super-graph of $GRID(n;r-\delta)$ and a sub-graph of $G(n;r+\delta)$ whp for small enough δ [Our result]

) Fastest mixing time of $G(n;r)$ is $\Theta(r^{-2}\log n)$ whp

Implications – Spectra

Our techniques can be extended to show that the spectrum of random geometric graphs converges to the spectrum of the grid graph. [Rai ’05]

Implications – Coverage

- Coverage: Any point in the unit square must be within a distance r from one of the n sensors
  - Known: there is a sharp threshold in r [Shakkottai, Srikant, Shroff '04]
  - Coverage is NOT a graph property, so it does not fall within our framework
  - But the laminar structure in our proof implies a sharp threshold for coverage as well (weaker than the sharpest known)

Conclusions

- Monotone properties in G(n;r) have sharp thresholds
  - Much sharper than for Bernoulli Random Graphs
- Much stronger too: Random geometric graphs exhibit a laminar structure
  - Useful for recovering several known results/proving new ones
  - Randomness is often a red-herring since the deterministic grid often yields asymptotically tight upper and lower bounds
- Open problem: Does laminarity imply anything about throughput (via separators)?