Complex Contagion and The Weakness of Long Ties in Social Networks: Revisited

Jie Gao

Stony Brook University

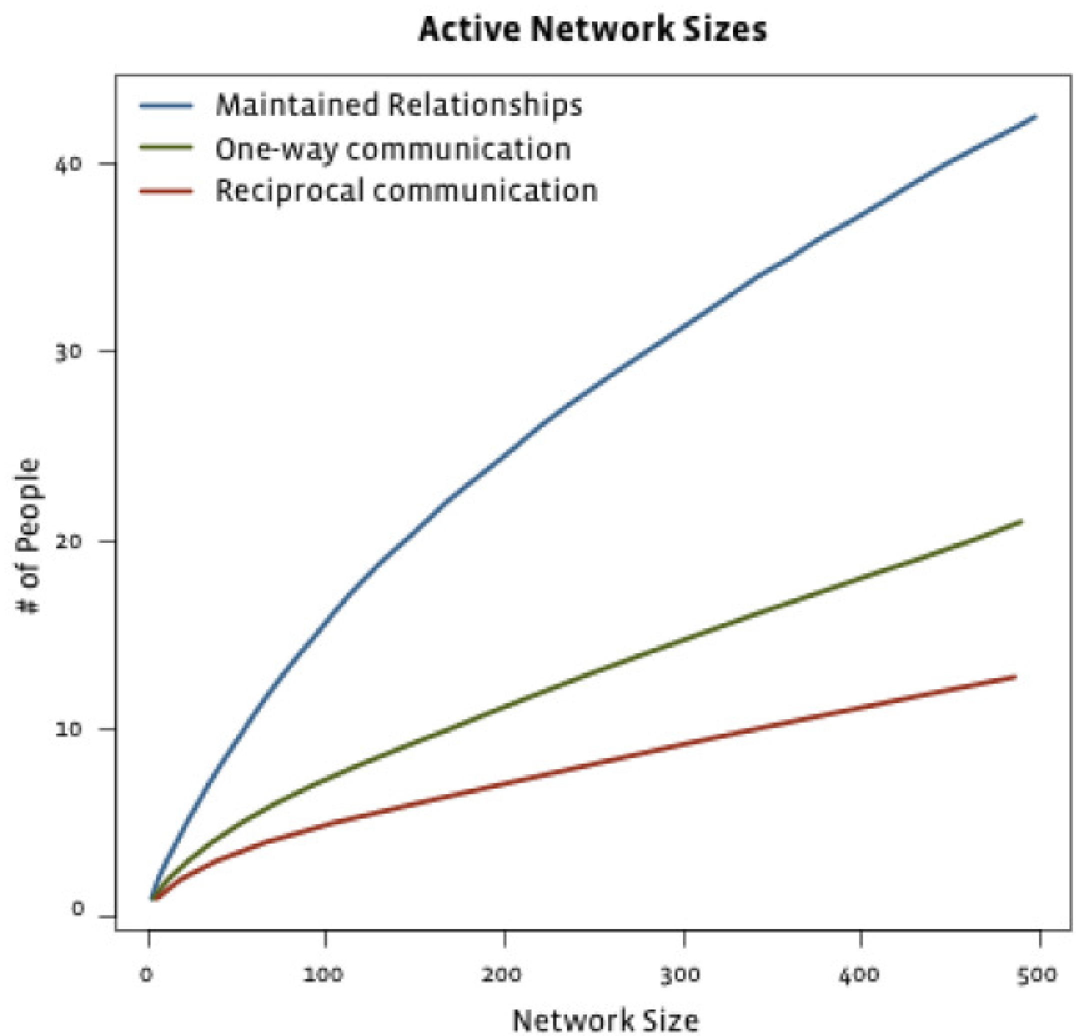
Social Ties and Tie Strength

- Strong ties
  - Family members, close friends, colleagues
  - People who regularly spend time together
  - Typically a small number
- Weak ties
  - People you know, acquaintances
  - Could be a lot

How to Measure Tie Strength?

- Infer from frequency of interactions
- Facebook
  - Reciprocal communication: A, B send msg to each other;
  - One-way communication: A sends msg to B;
  - Maintained relationship: A follows info of B (click content, visit profile page).

Examples: facebook data



Cameron Marlow, Lee Byron, Tom Lento, Itamar Rosenn. Maintained relationships on Facebook, 2009.

Strong Ties: Triadic closure

- Your friends are likely friends of each other.
  - More opportunities to meet
  - Higher level of trust
  - Incentive
- Lots of triangles, small cliques
- High clustering coefficient
  - Prob{two friends of A being friend of each other}
  - # edges between n friends / (n choose 2)

Strong Ties: Homophily

- Friends are alike, they share similar traits
  - Live close; go to same school; have the same hobbies, etc.
- Two forces leading to homophily
  - Selection: people who are alike become friends.
  - Influence: one adopts behaviors from friends.

Strength of Weak Ties

- [Granovetter 1960s] ask "how do you find your new job?"
  - Mostly through personal contacts;
  - Mostly through acquaintances rather than close friends.

Weak Ties: Bridges & Brokers

- Information broker, "structural hole"
- Connects different communities
- Local measure: neighborhood overlap
  - N(A): set of neighbors of A
  - $|N(A) \cap N(B)|/|N(A) \cup N(B)|$
- Global measure: betweeness centrality
  - # shortest paths through a link

Outline

1. Social ties & tie strength
2. Small world phenomenon
3. Network models
4. Complex contagion & weakness of strong ties
5. Our analysis

Small World Phenomenon

- [Milgram 1967] Ask randomly chosen people in Kansas to mail letter to a target person living in MA.
  - Info of target: name, address, occupation.
  - Forward to **ONE** friend known on a first name basis
- 1/3 letters arrived with a median of 6 hops;
- Six degree of separation.

Implication of Small World Experiments

- Network diameter is small!
- Can it be strong ties?
- No – due to triadic closure.
- So it must be the weak ties.

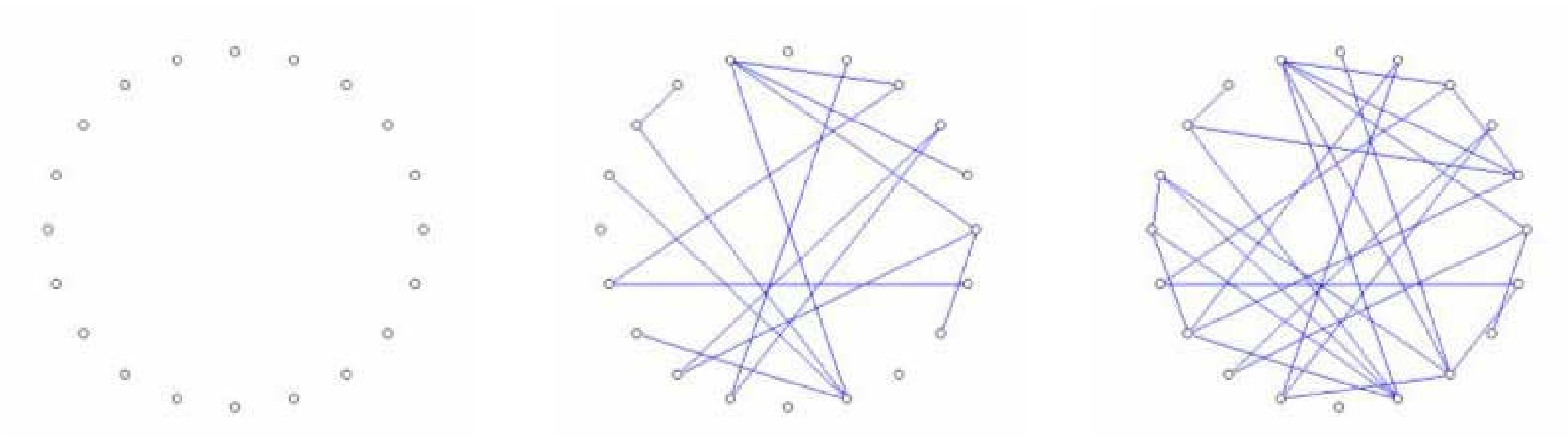
[Granovetter 1973] *"Whatever is to be diffused can reach a larger number of people, and traverse a greater social distance, when passed through weak ties rather than strong."*

Outline

1. Social ties & tie strength
2. Small world phenomenon
3. Network models
   – How to generate graphs with prescribed properties?
4. Complex contagion & weakness of strong ties
5. Our analysis

Random Graphs: Erdös-Renyi Model

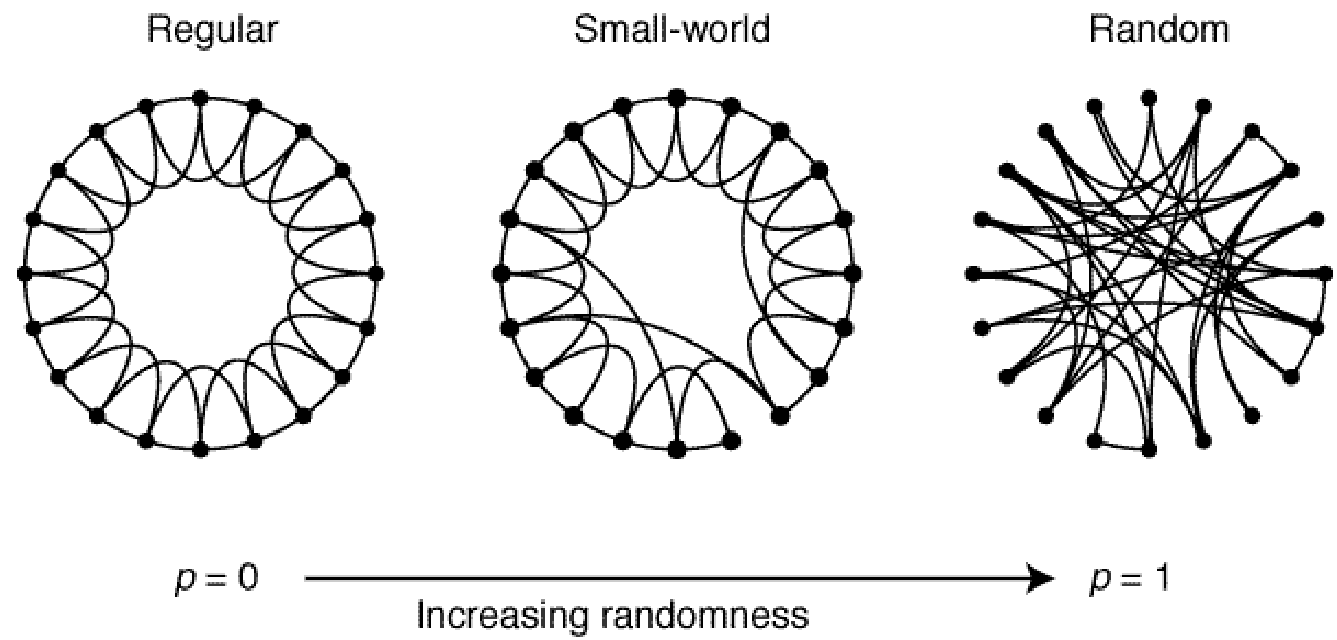
- G(n, p): a random graph on n vertices; each edge exists with probability p.
- Has small diameter.
- **But, clustering coefficient is small.**



$p = 0$ (a) $p = 0.1$ (b) $p = 0.2$ (c)

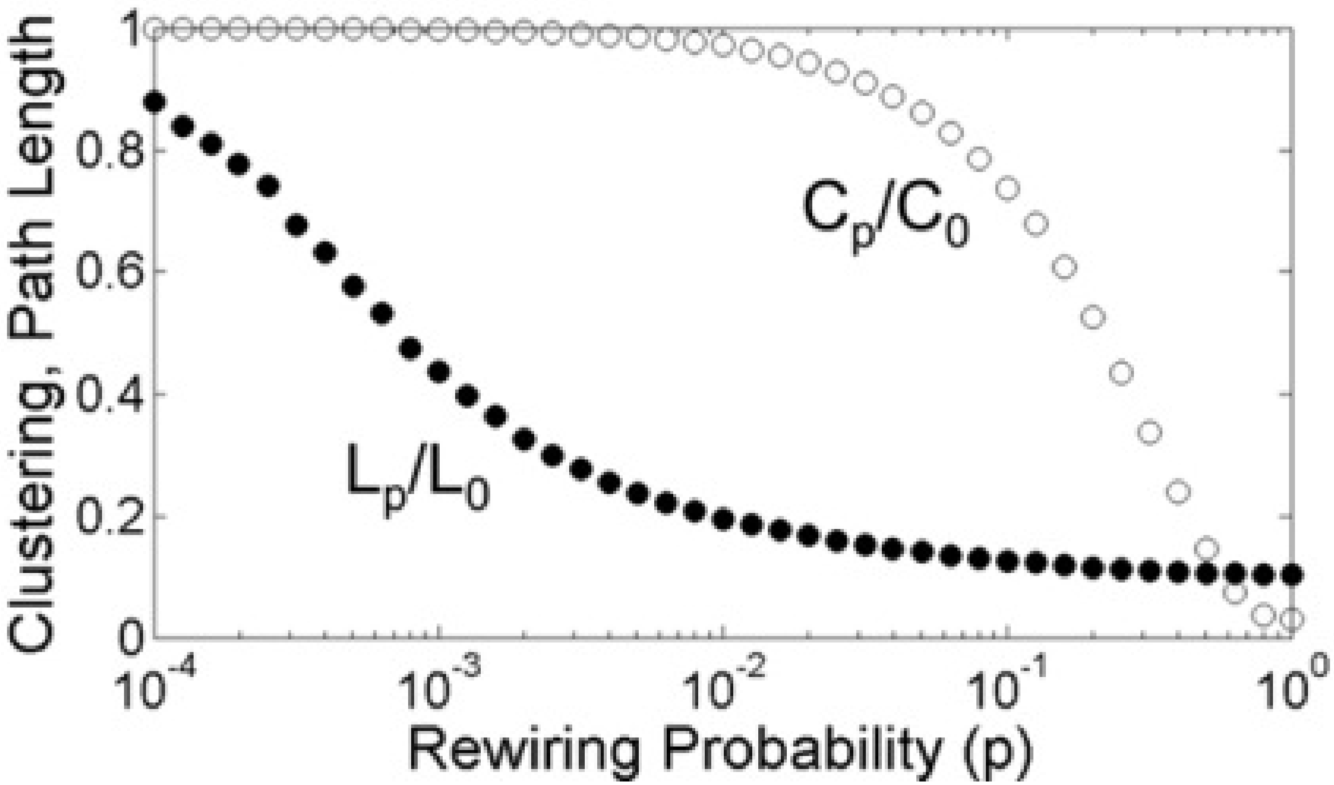
Watts-Strogatz Model

- Start with nodes on a ring
- k-hop neighbors on the ring are connected.
- Randomly "rewire" the endpoint by prob p.



Watts-Strogatz Model

- For a suitable range of p, clustering coefficient is large; graph diameter is small.

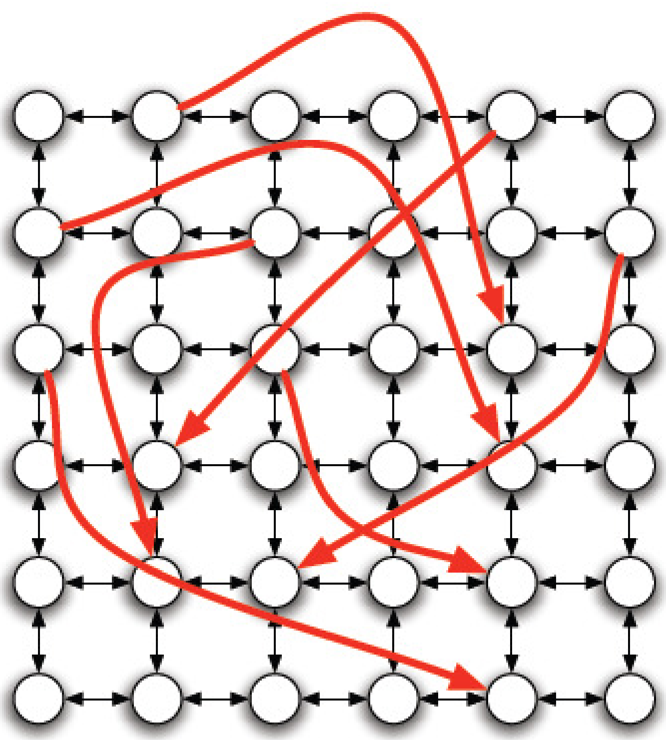


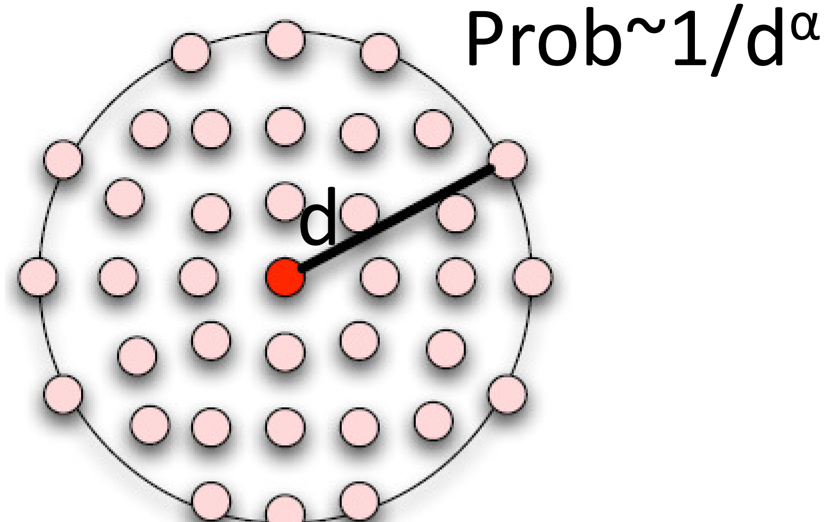
Re-examine Milgram's Experiment

- [Milgram 1967]
  - Forward to ONE acquaintance on a first name basis
- Forwarding decisions are purely local.
- No global knowledge is available.
- Watts-Strogatz Model: there **exists** a short path.
- Question: can we **find** it using local info?

Kleinberg's Small World Model

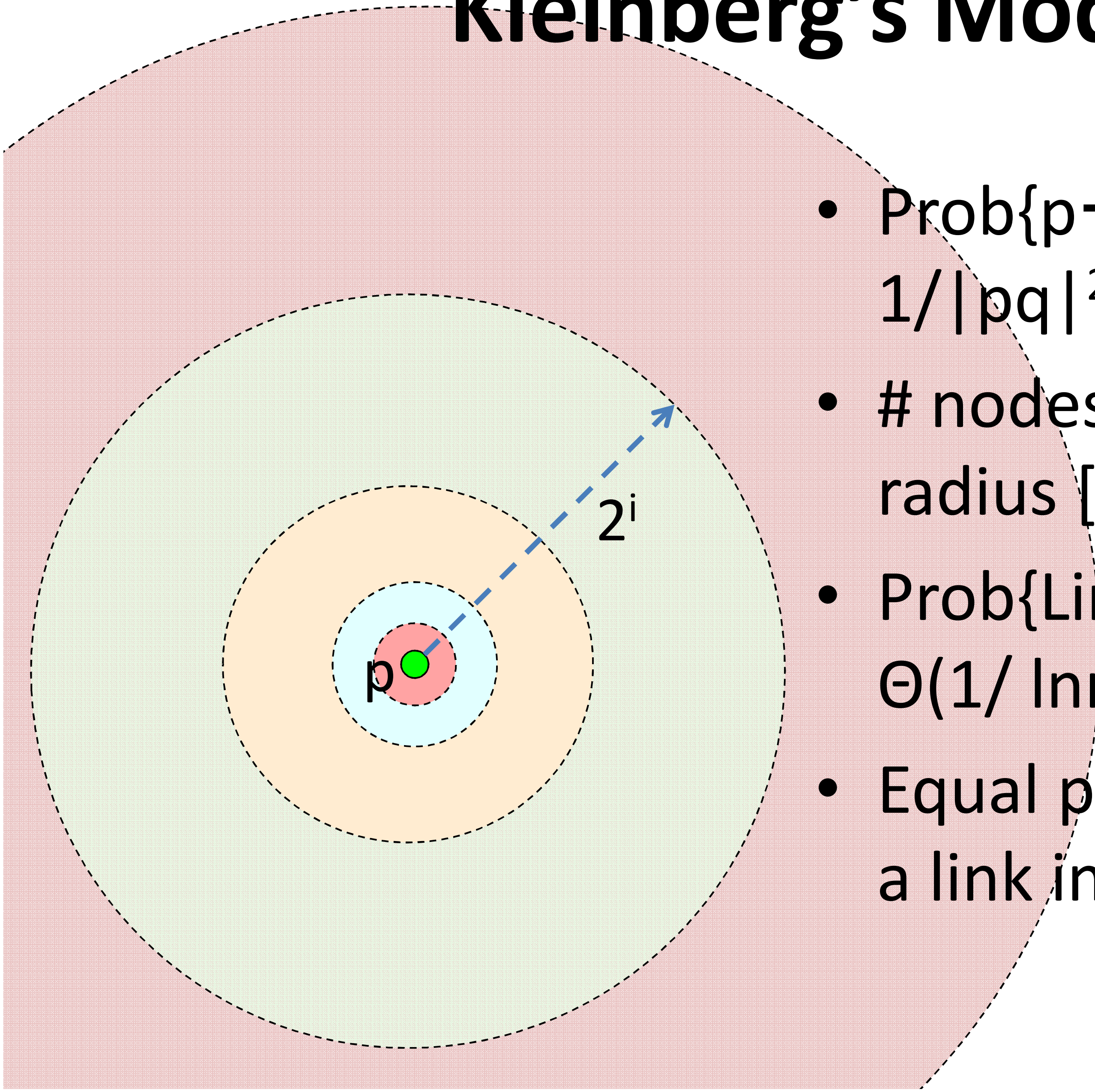
- Add random edges, with a spatial distribution
- When **α=2**, greedy routing ~ $O(\log^2 n)$ hops





The Small-World Phenomenon: An Algorithmic Perspective, STOC'00.

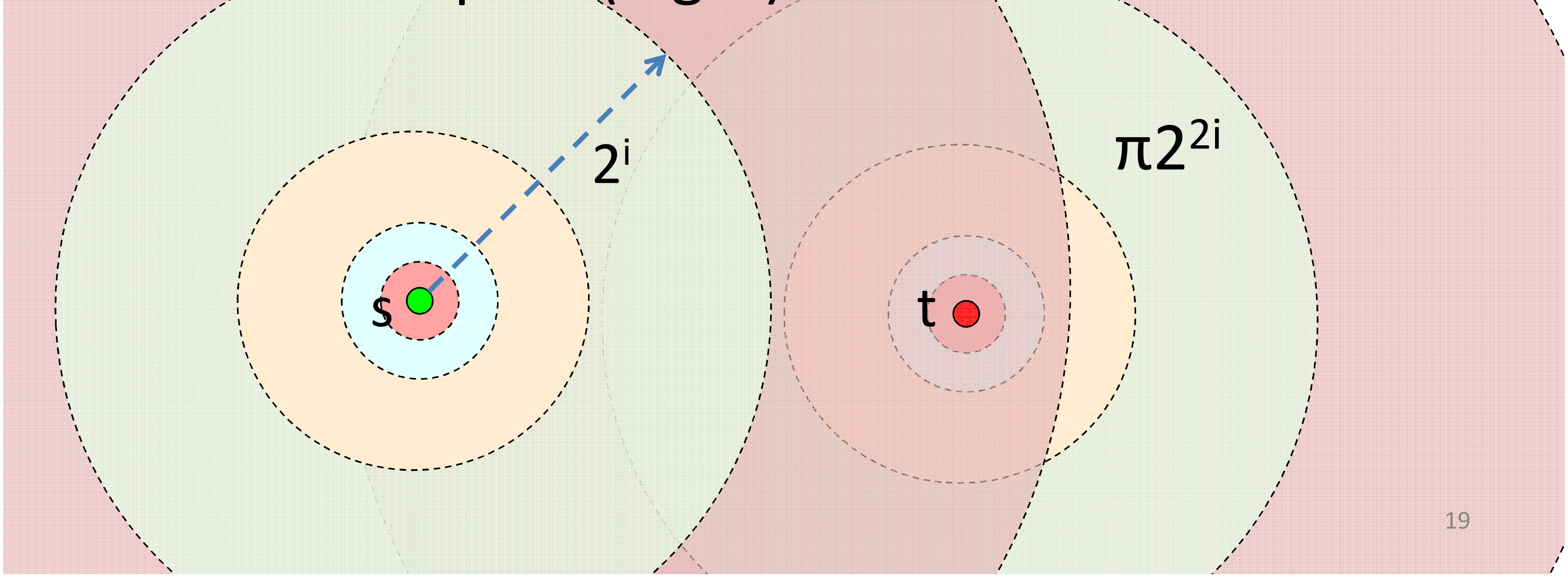
Kleinberg's Model



- Prob{p→q} =1/(πlnn) · 1/|pq|^2
- # nodes inside a ring of radius [2^i, 2^{i+1}] = $3\pi 2^{2i}$
- Prob{Link to ring i} ≈ Θ(1/ lnn)
- Equal prob of choosing a link in each annulus

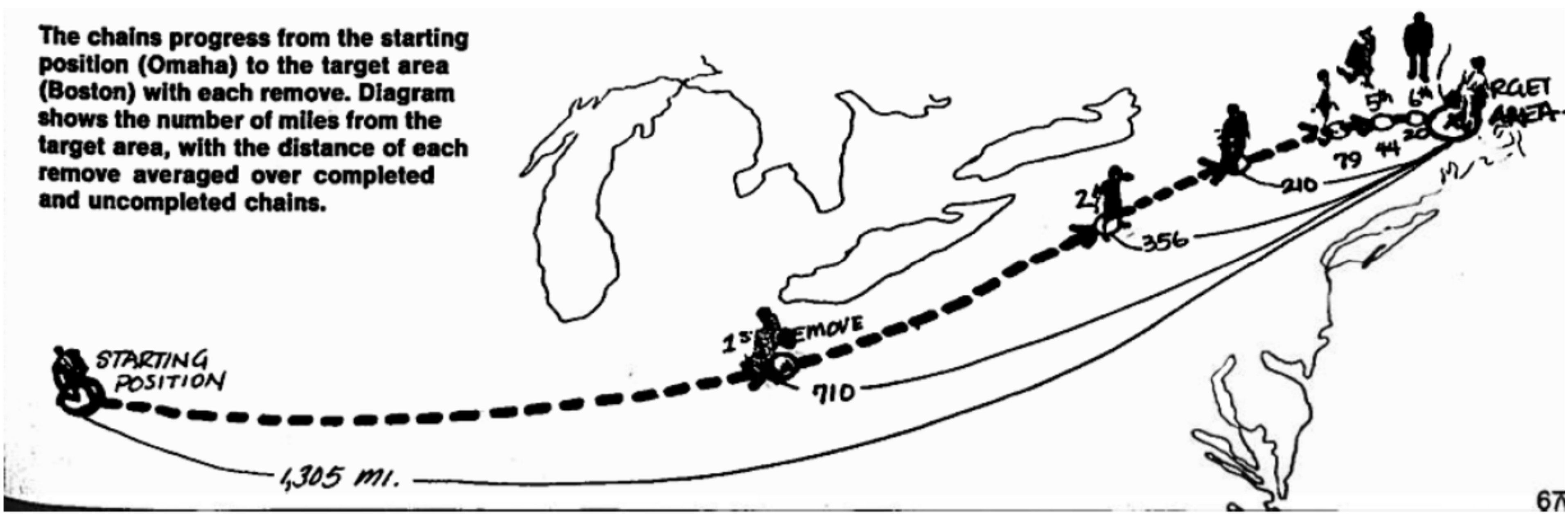
Why Greedy Routing Works?

- Path is distance decreasing & loop-free
- With expected O(logn) steps, the message gets to within 2^i of t.
- Total # steps: $O(\log^2 n)$



An Example

Milgram: “*The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain*”



The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.

Magic Exponent

- Spatial probability ~ $1/d^{\alpha}$
- Greedy routing with short paths: $\alpha=2$.
- For α too big, most random links are too short.
- For α too small, links are too random and lack of direction.

Outline

1. Social ties & tie strength
2. Small world phenomenon
3. Network models
4. Complex contagion & weakness of strong ties
5. Our analysis

Our Problem: Contagion in Social Networks

- Simple contagion
  - Spreads through a **single** contact
  - Virus infection, rumor, information
- Complex contagion
  - Needs **multiple** confirmations/contacts
  - Pricey technology innovations, social behavior changes, immigration [CKM66, MM64].

Speed of Diffusion

- Simple contagion
  - Spreads through a **single** contact
  - **Fast, speed ≈ diameter**
  - Strength of weak ties.
- Complex contagion
  - Needs **multiple** confirmations/contacts
  - Need **wide** bridges.
  - **Slow? How slow?**

"The Weakness of Long Ties"

- Watts-Strogatz Model
- Requiring two active neighbors to be affected
    1. require a substantially large number of random ties to even create one single wide bridge;
    2. Random rewiring erodes the capability of spreading a complex contagion.

"How is it possible that complex contagions are able to spread through real social networks?"

Damon Centola and Michael Macy. Complex Contagions and the Weakness of Long Ties. American Journal of Sociology, 113(3):702–734, November 2007.

Our Results: (I)

- On Kleinberg model, complex contagion can spread in speed **O(polylogn)**.
- The distribution of weak ties are important.

Network Model

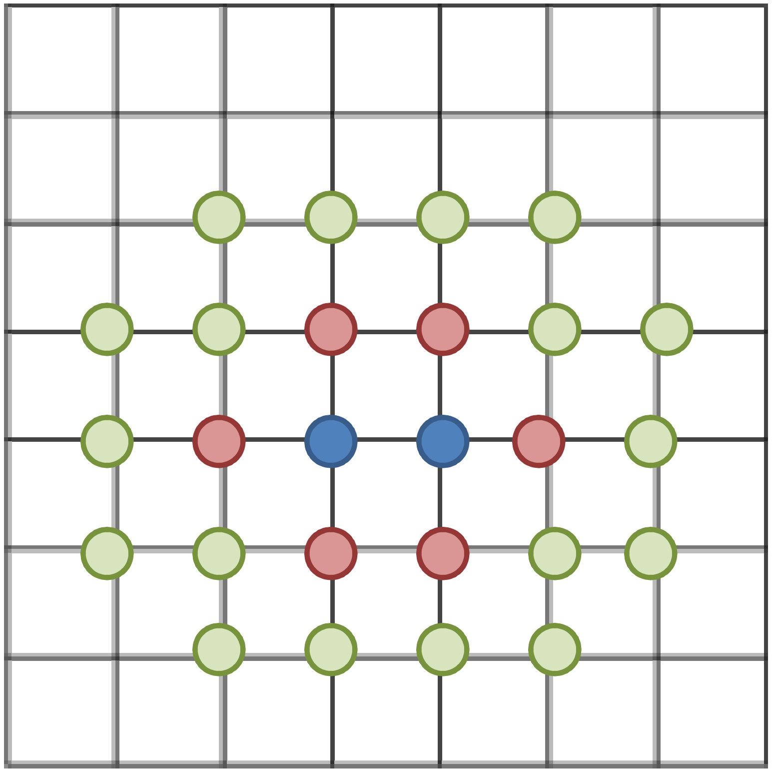
- Network model
  - 2D Grid of n nodes wrapped as a torus;
  - Strong ties: nodes within Manhattan dist of 2.
  - Weak ties: each choosing 2 additional random edges with Prob{p→q} $=\Theta(1/\ln n) \cdot 1/|pq|^2$
- Initial seeds
  - A pair of neighboring active nodes

Model of Diffusion

- Complex contagion requiring two active neighbors to be affected
- Proceed in rounds.
- A node with ≥ two active neighbors in round i become active in round i+1.
- Goal: bound # rounds to cover the whole network.

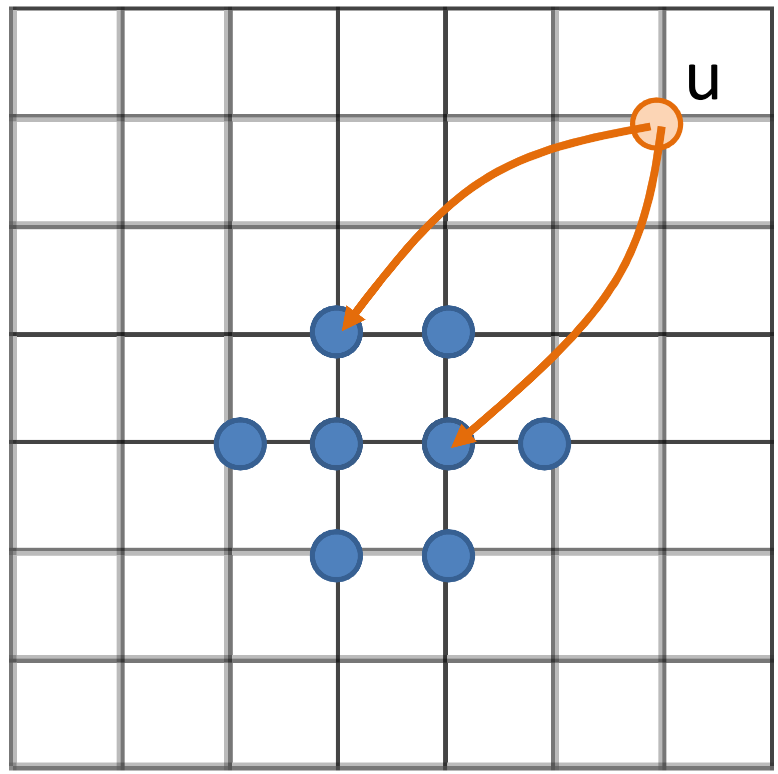
3 Types of Diffusion

- Local diffusion
  - Through strong ties
  - Slow, local
  - Each round: nodes on periphery are activated



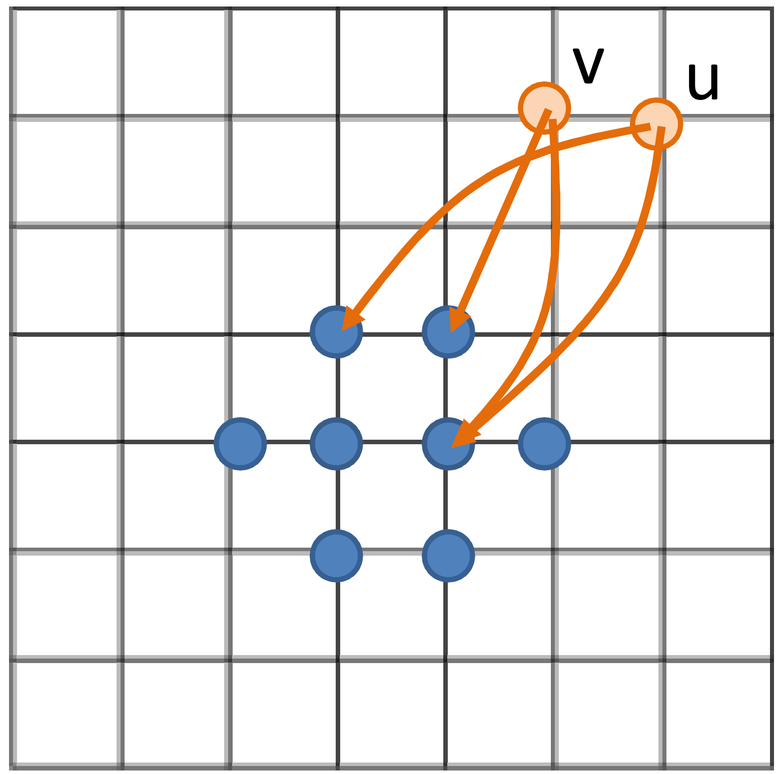
3 Types of Diffusion

- Local diffusion
  - Through strong ties
  - Slow, local
- Random diffusion
  - Through weak ties
  - Isolated active nodes.



3 Types of Diffusion

- Local diffusion
  - Through strong ties
  - Slow, local
- Random diffusion
  - Through weak ties
  - Isolated active nodes.
- Generating new seeds
  - Propagation speed doubles



Observations

- Set of active nodes S_i monotonically increases
- Edges that activate u can be
  - Weak ties built by u.
  - Weak ties built by other nodes to u.
  - Strong ties of u.

Observations

- Set of active nodes S_i monotonically increases
- Edges that activate u can be
  - Weak ties built by u. ← random diffusion
  - ~~Weak ties built by other nodes to u.~~
  - Strong ties of u. ← local diffusion

Ignored – we get an upper bound.

Bound Rate of Diffusion: Two Phases

- Phase 1: using local diffusion, after **$\log^{2.5} n$** rounds, a disk of radius **$R=\log^{2.5} n$** is activated.

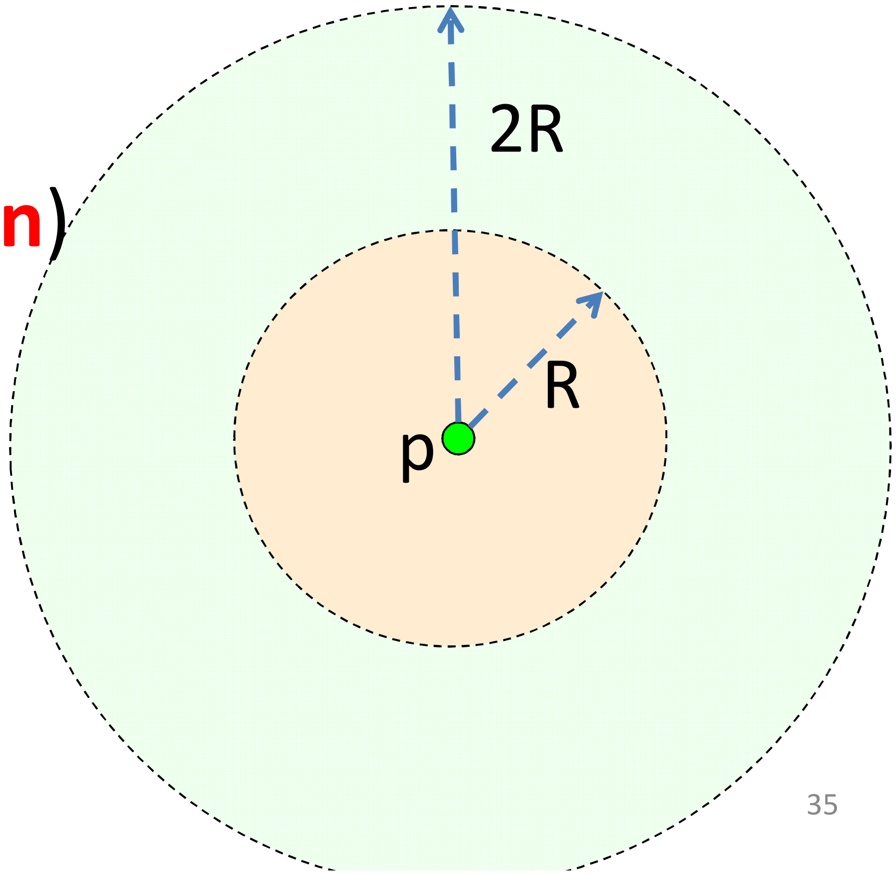
- Phase 2: After a disk of radius **$R \geq \log^{2.5} n$** is activated, # rounds to cover a disk of radius **2R** is **$\log^{2.5} n$**.

- Total # rounds = O(**$\log^{3.5} n$**).

Bound Rate of Diffusion: Phase II

- Suppose that a disk of radius $\mathbf{R \geq \log^{2.5} n}$ is activated.

Claim: # rounds to cover a disk of radius $\mathbf{2R}$ = $O(\mathbf{\log^{2.5} n})$



Proof of the Claim

- Suppose that a disk of radius **R** is activated.

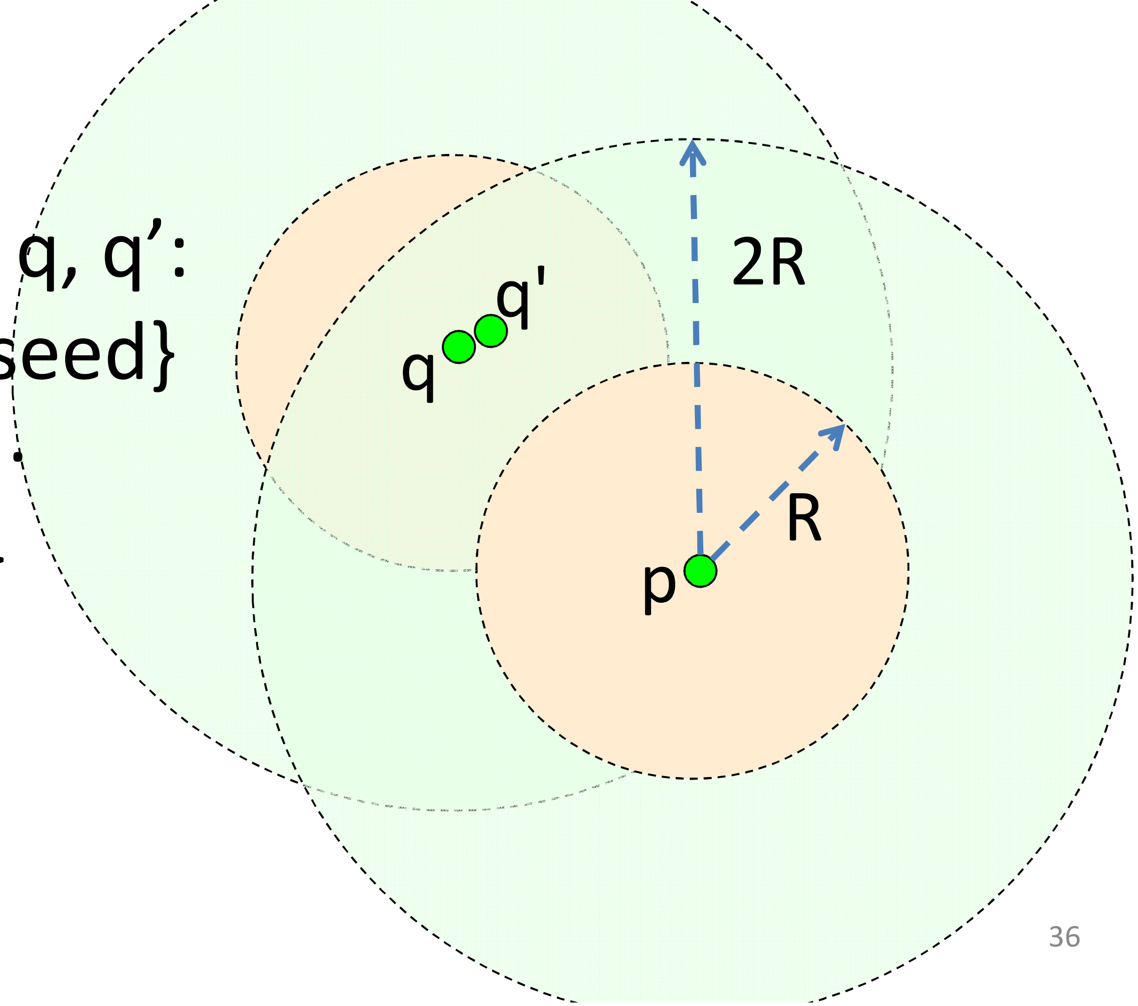
Consider neighbors q, q':

Prob{q, q' is new seed}

= Prob{q activated} ·

Prob{q' activated}

$\geq \Theta(\mathbf{1/\log^4 n})$

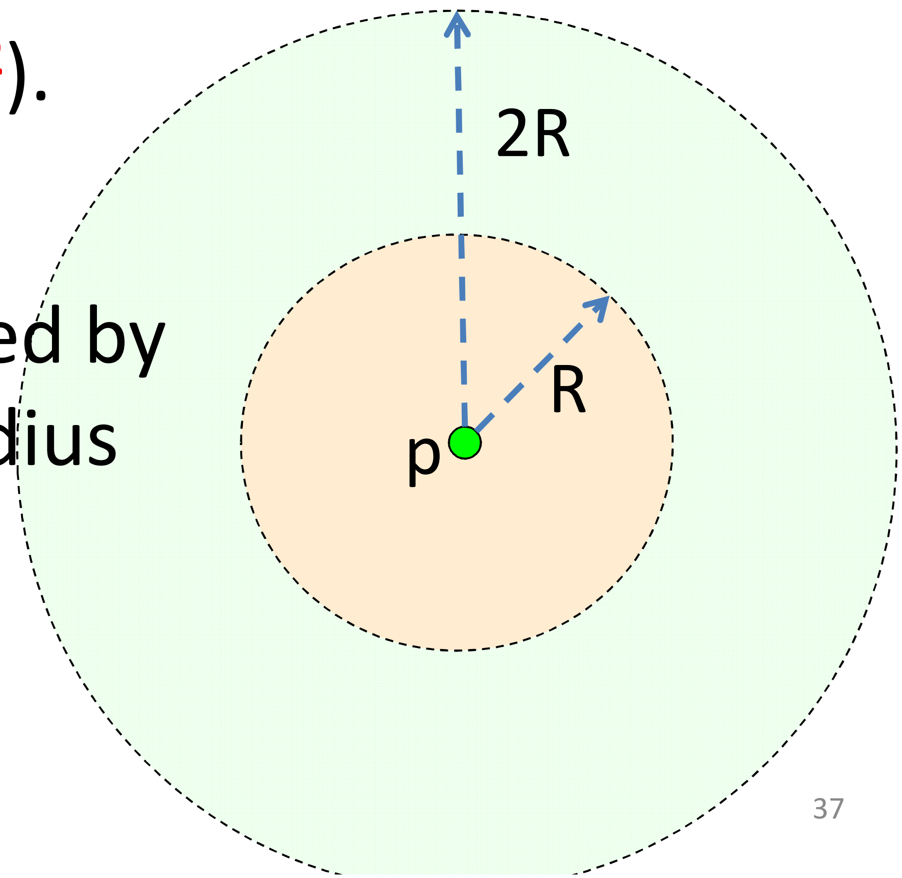


Proof of the Claim

- Suppose that a disk of radius **R** is activated.

The annulus has area $\Theta(\mathbf{R^2})$.

The annulus can be covered by **$R^2/\log^5 n$** disks (bins) of radius **$\log^{2.5} n$** each.



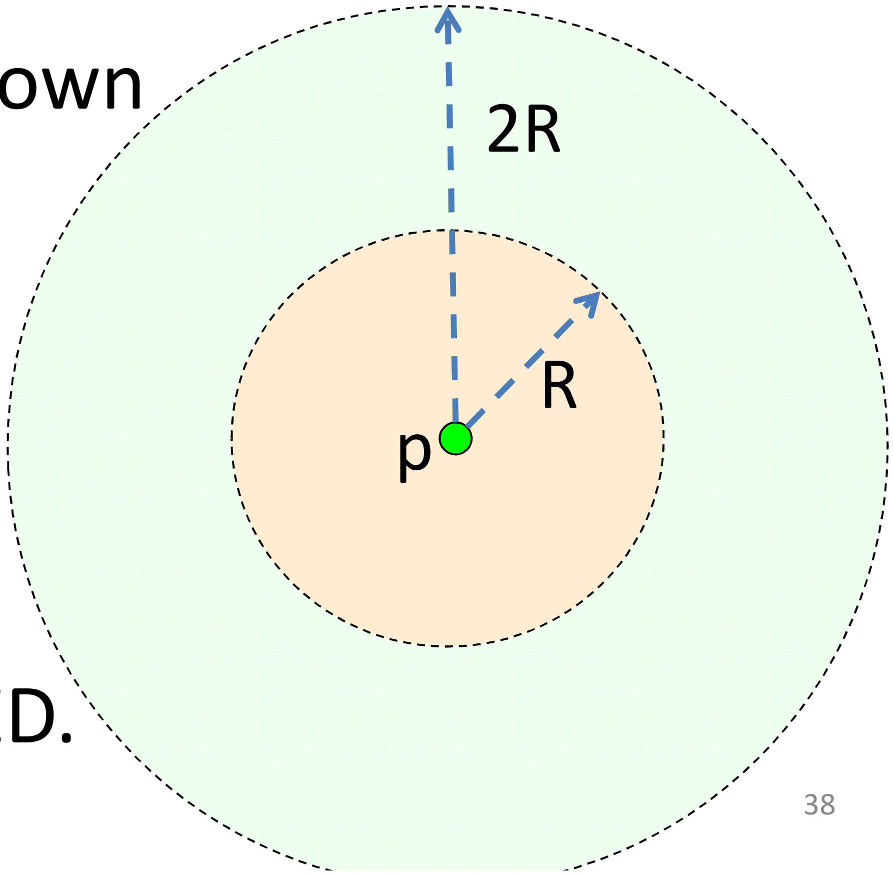
Proof of the Claim

- Suppose that a disk of radius **R** is activated.

seeds generated in the annulus is $\Theta(\mathbf{R^2/log^4n})$, thrown into $\mathbf{R^2/log^5n}$ bins.

W.h.p. each disk of radius $\mathbf{log^{2.5}n}$ has one seed.

After ≤ $\mathbf{log^{2.5}n}$ rounds, the annulus is filled up. QED.

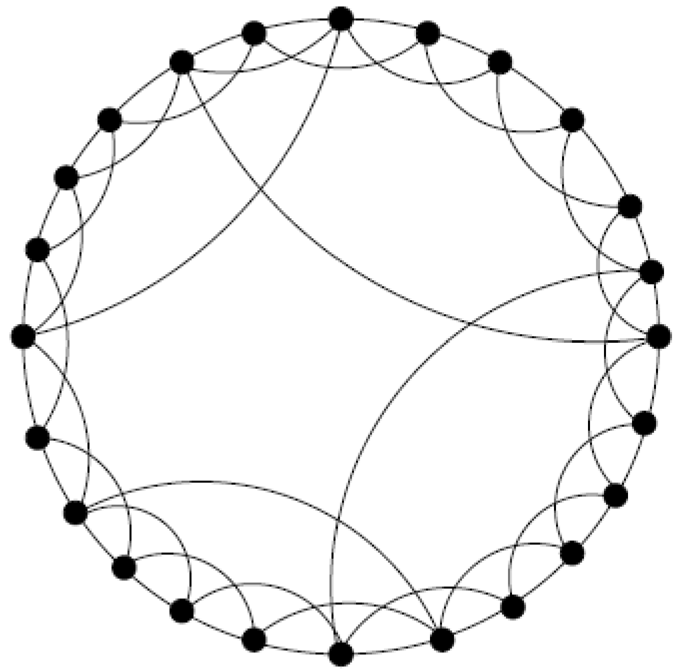


More Results

- Newman-Watts Model
- Kleinberg's Hierarchical Model
- Preferential Attachment Model

Newman-Watts Model

- Similar to Watts-Strogatz model
  - Each node has 2 additional edges to randomly chosen nodes.
- What we show
  - # rounds is **$\Omega(\sqrt{n}/ \log n)$**.
  - Unable to generate new seeds

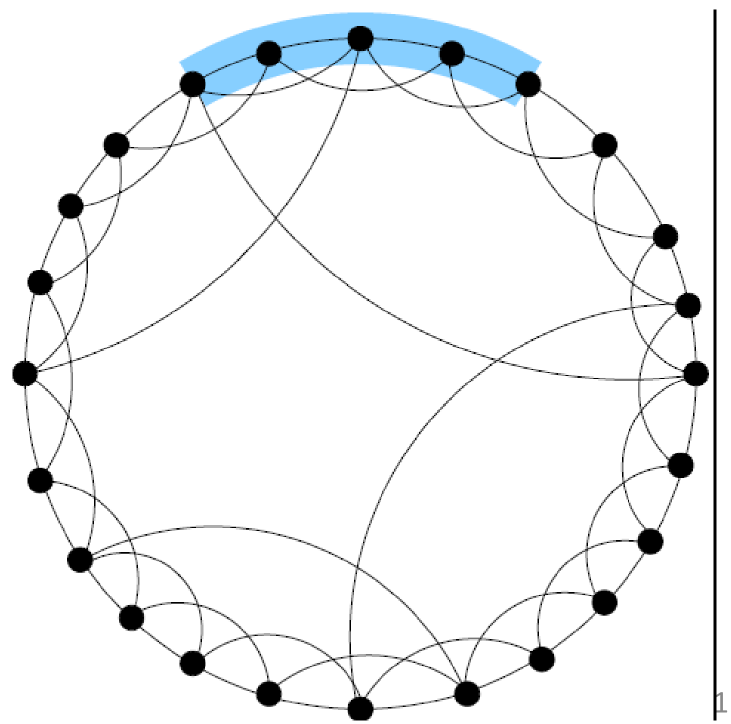


Proof Sketch

- Consider the interval ***F*** of length **√n/ log n** centered at the seeds

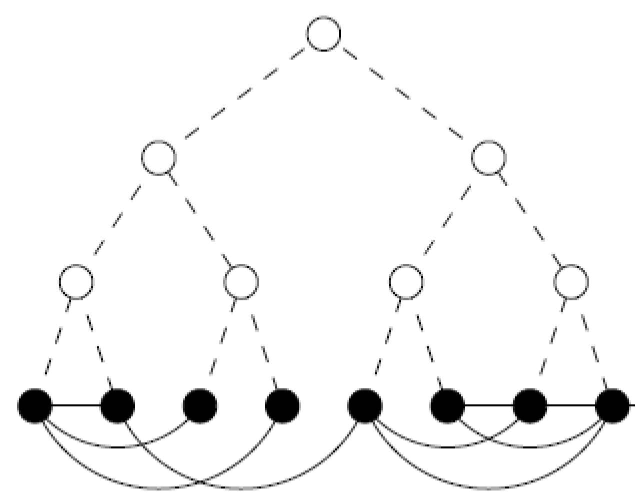
- Prob{any node having 2 weak ties to ***F***} is small

- Diffusion within ***F*** is local & slow.



Kleinberg's Hierarchical Model

- Kleinberg's hierarchical model
  - Hierarchy: b-ary tree;
  - $h(u, v)$: height of LCA of u, v
  - Prob{uv} ≈ $b^{h(u, v)}/\log n$
  - Each node has j random edges
- What we show
  - $j=\Theta(\log^2 n)$: # rounds= $O($**logn**$)$



Generalization

- K-complex contagion
- Different model parameters: # strong/weak ties
- Directed graphs
  - E.g. Twitter network

Recap

- Newman-Watts vs.. Kleinberg's models.
  - Distribution of weak ties: uniform random vs. spatial distribution
  - Speed of diffusion: slow vs. fast.
- Simple contagion vs. complex for Newman-Watts
  - Fast (~diameter, polylog) vs. slow (poly)

Ongoing Work

- Graphs with power law degree distribution
  - Preferential attachment model: O(log n).
- Complex contagion in real data sets
- Different threshold for different users
- How to choose initial seeds
  - NP-hard [KKT'03].

Questions & Comments

- Joint work with my students Golnaz Ghasemiesfeh, Roozbeh Ebrahimi @ Stony Brook