Small-world Networks

An algorithmic perspective

It's a small world!

- Technological advances of the early 20th century led to the theory of “six degrees of separation”.
- The theory states: for any two individuals there is a chain of acquaintances linking them of length at most 6.
- Stanley Milgram (famous for the “Milgram experiment”) decided to conduct an experiment in 1967 to confirm this hypothesis.

Milgram's small-world experiment

- Methodology: two random individuals are chosen, one in Nebraska and another in Massachusetts.

- A letter from the first reaches the second through a chain of acquaintances. The chain's length turns out to be an average of 5 or 6.

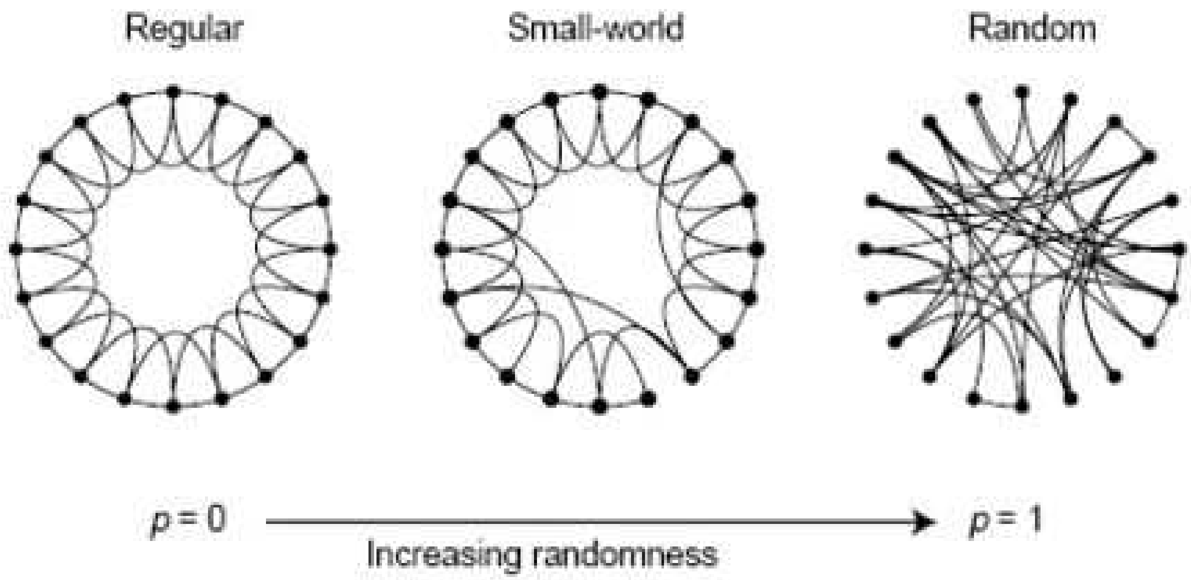
Modeling small worlds with graphs

- There has been a lot of research on modelling and explaining the small-world phenomenon.
- We could model society by a random graph. Then, a “simple” explanation for this phenomenon is: random networks have low diameter.
- But social networks are not really random networks: the relation of A and B is influenced by the relations of A,C and B,C.
- Therefore, social networks must be clustered.

Watts and Strogatz

- Watts and Strogatz proposed a model with low mean shortest path and high clustering coefficient.
- Basic idea: begin with n nodes placed on a ring. Each node is connected to k of its closest neighbors. (Local structure)
- Add to that a number of random edges. (Long-range contacts)
- Result: a “re-wired ring lattice” with low diameter.

Watts and Strogatz



Extending Watts and Strogatz

- The previous model explains one aspect of the small-world phenomenon: the existence of short paths.
- But there is also another important aspect: The ability of individuals to find those paths without full knowledge of the network.
- This *algorithmic* aspect of the phenomenon must also be studied.

Outline of results

We will show the following:

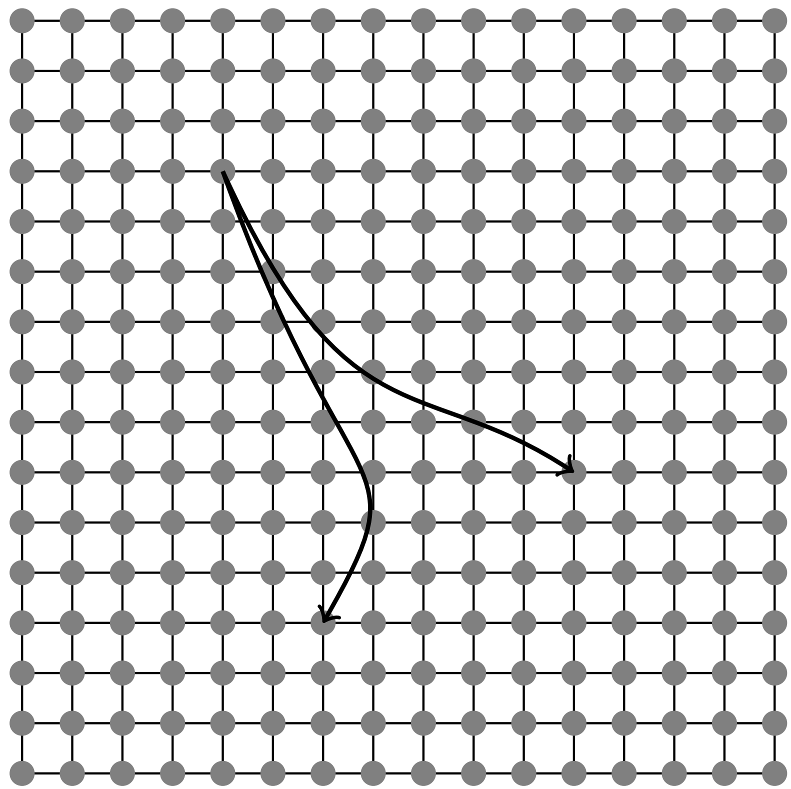
- Networks based on the Watts-Strogatz model (basic graph + uniformly random edges) fail to capture the algorithmic aspect of the phenomenon.
- A variation of this model where random edges are added with probability depending on the distance of the two nodes does capture this aspect.
- In fact, for grid networks, a specific family of networks is shown to be the only correct model for small-world networks.

Basic model

- Variation of Watts-Strogatz model. Start with $n \times n$ vertices on a grid. Every vertex is connected to all other vertices with grid distance at most p, for some constant p. (Local structure).
- Every vertex u also has q long-range contacts: q other vertices are chosen independently at random with probabilities proportional to $(\frac{1}{d(u,v)})^r$. (q, r constants).
- For fixed p, q we have a single-parameter model (on r).

Basic model - illustration

Example of the network model



$(p = 1, q = 2)$.

The problem

- Pick randomly two vertices s, t. A message starting from s must reach t using only local information.
- All vertices know their coordinates on the grid, and the location of t.
- All vertices know the locations of their neighbors but not the locations of their neighbors' neighbors.

Restatement of results

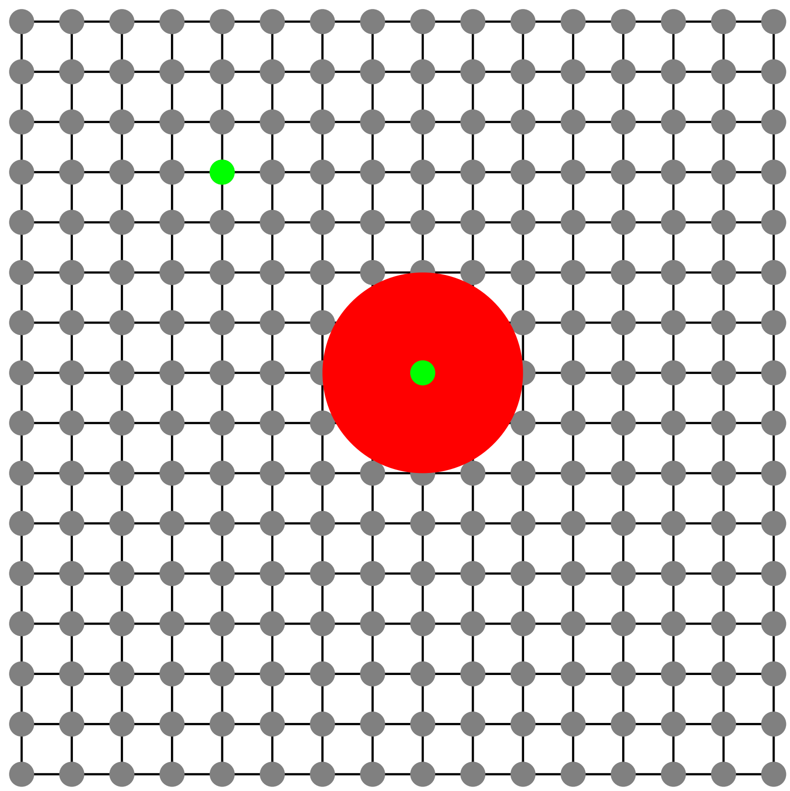
- For $r = 0$ (uniformly pick long-range contacts) there is no decentralized algorithm to find short chains (of length $polylog(n)$).
- For $r = 2$ and $p = q = 1$ there is one, with expected path length $O(\log^2 n)$.
- $r = 2$ is the only value of r where this happens!

Intuition

Why $r = 0$ doesn't work.

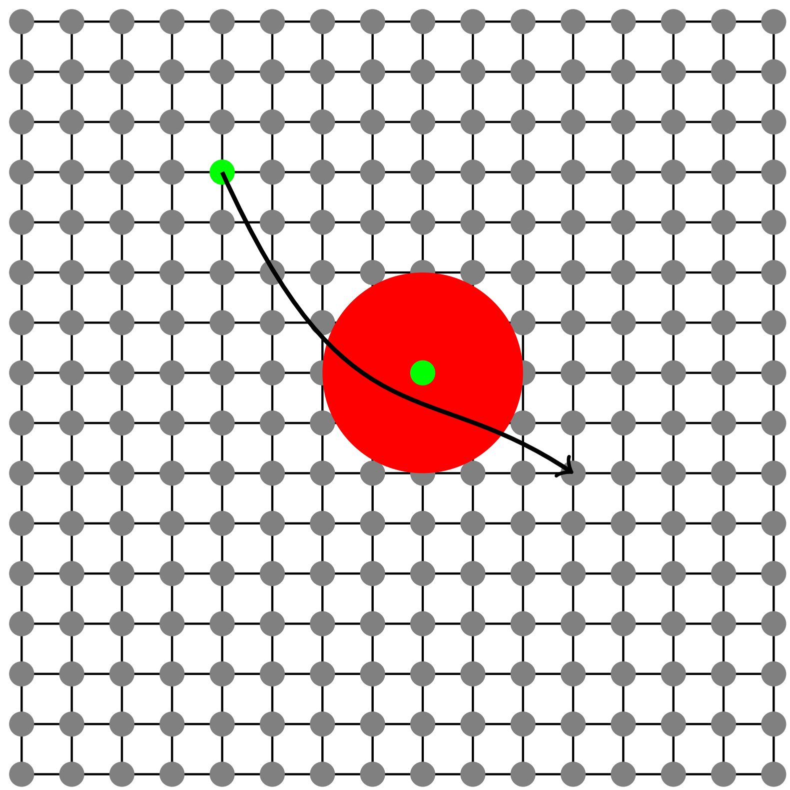
- In order to reach our target fast we *need to use long-range edges.*
- *But these edges are completely random, so there is no smart way to decide where to go!*

Intuition



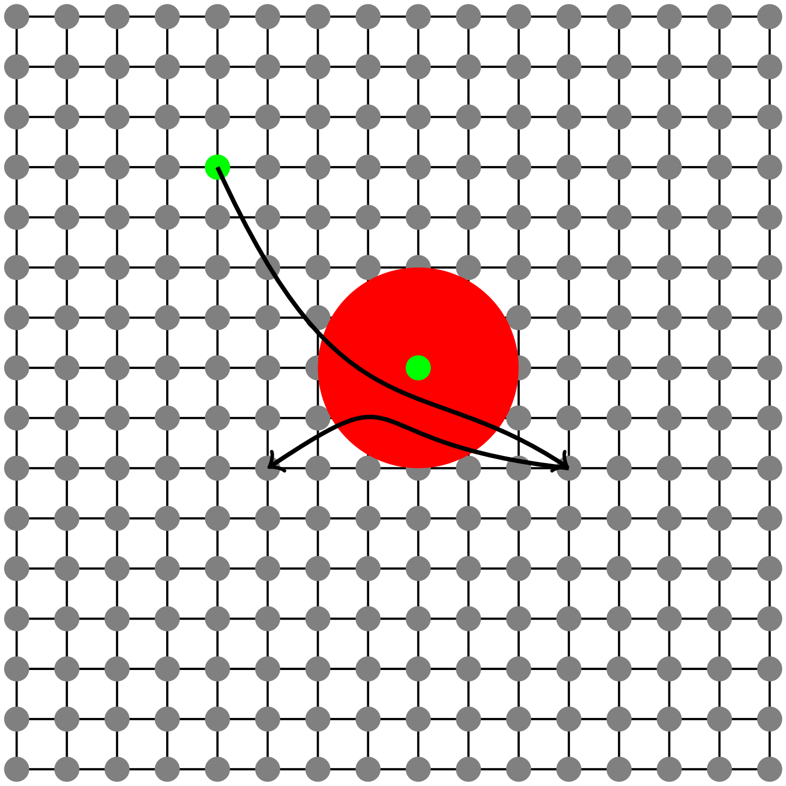
Uniformly random long-range contacts take us to uniformly random places!

Intuition



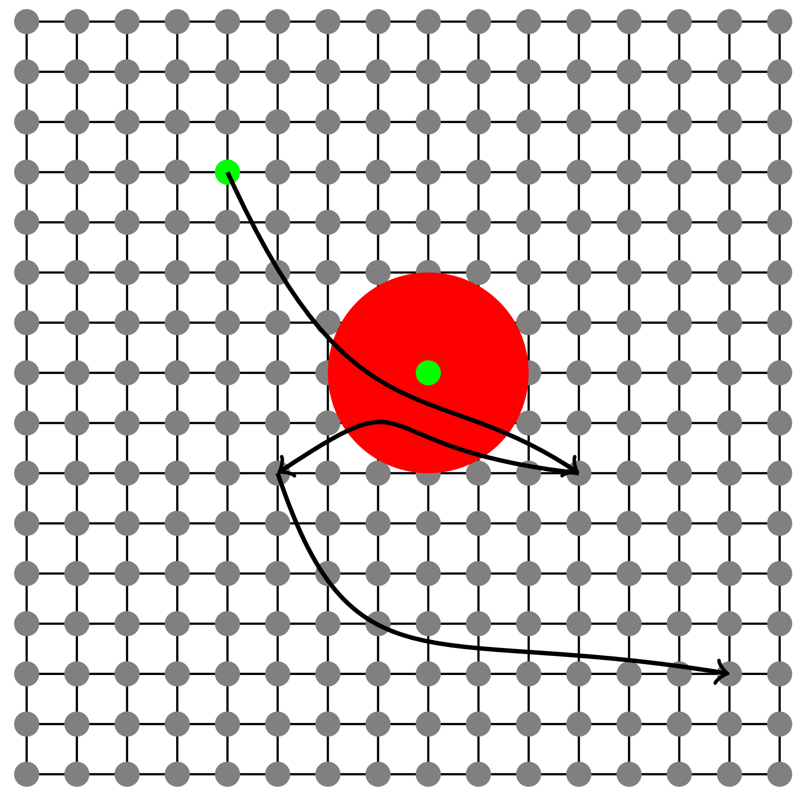
Uniformly random long-range contacts take us to uniformly random places!

Intuition



Uniformly random long-range contacts take us to uniformly random places!

Intuition



Uniformly random long-range contacts take us to uniformly random places!

More Intuition

Why $r = 2$ works.

- Now long-range edges are not randomly placed, but instead closer vertices are favored. Therefore, forwarding the message geographically makes sense, because vertices closer to the target have a better chance of having it as a long-range contact.

Yet More Intuition

But why only $r = 2$?

- For $r > 2$ the network becomes too clustered because nearby vertices are favored too much and long-range contacts don't take you far enough to make a short chain.
- For $r < 2$ long-range vertices appear too random-like and therefore it's hard to find the right path.
- Geometric intuition: The number of neighbors of the target at distance at most k is $\approx k^2$.

The algorithm

A very simple idea.

- At each step, forward the message to the neighbor that is closest to t, in terms of lattice distance.

Proof

- When the message is at vertex u, we say that the algorithm is in phase j if $2^j < d(u,t) \leq 2^{j+1}$. ($j \leq \log n$).
- Proof strategy: we will show that the probability that a phase ends at a given step is $\Theta(\frac{1}{\log n})$. Thus, each phase is expected to last $\Theta(\log n)$ steps, and since we have $O(\log n)$ phases the algorithm has expected running time $O(\log^2 n)$.

Proof

- $p = q = 1$ and $r = 2$. Thus, every vertex has 4 short-range neighbors in the grid, plus one long-range contact.
- The probability that u has v as its long-range contact is proportional to $\frac{1}{d(u,v)^2}$. Normalize:

$$P[u \to v] = \frac{1}{d(u,v)^2} / \sum_{u \neq v} \frac{1}{d(u,v)^2}$$

Proof

- But

$$\sum_{u \neq v} \frac{1}{d(u,v)^2} \leq \sum_{j=1}^{2n-2} \frac{4j}{j^2} = 4 \sum_{j=1}^{2n-2} \frac{1}{j} \leq \ldots \leq 4 \ln(6n)$$

- Thus, $P[u \to v] \geq \frac{1}{4 \ln(6n)} \frac{1}{d(u,v)^2}$

Proof

- Suppose we are in phase j, $\log\log n \leq j < \log n$. In order for this phase to end at this step the message must enter the set of vertices B_j with distance at most 2^j from t. These are at least $1 + \sum_{i=1}^{2^j} i > \ldots > 2^{2j-1}$.
- For $v \in B_j$, $d(v,t) \leq 2^j$. For the current holder u, $d(u,t) \leq 2^{j+1}$. Thus, (triangle inequality) $d(u,v) \leq 2^{j+2}$.
- $P[u \to v] \geq \frac{1}{4\ln(6n)} \frac{1}{2^{2j+4}}$. Therefore, $P[u \to B_j] \geq \frac{1}{4\ln(6n)} \frac{2^{2j-1}}{2^{2j+4}} = \frac{1}{128\ln(6n)}$

Proof

- For $j < \log\log n$ every phase lasts at most $\log n$ steps even if only short-range contacts are used.
- Therefore, each phase is expected to last at most $O(\log n)$ steps and the result follows from linearity of expectations.

Lower bound, $r < 2$

- For the lower bounds we allow algorithms more power: each node knows all the contacts of all previous nodes in the chains, and can forward the message to any of them in one step.
- Proof strategy: identify an area U with small lattice distance around t. Any algorithm (w.h.p.) must use a long-range contact to enter U. We show that is is hard to find such a long-range edge.

Proof

- $$P[u \to v] = \frac{1}{d(u,v)^r} \Big/ \sum_{u \neq v} \frac{1}{d(u,v)^r}$$

- $$\sum_{u \neq v} \frac{1}{d(u,v)^r} \geq \sum_{j=1}^{n/2} \frac{j}{j^r}$$

- For simplicity, set $p = q = r = 1$

$$\sum_{u \neq v} \frac{1}{d(u,v)^r} \geq \sum_{j=1}^{n/2} 1 = n/2$$

Proof

- Let U be the set of nodes at distance at most $n^{1/3}$ from t. Then $|U| \leq 1 + \sum_{j=1}^{n^{1/3}} 4j \leq 4n^{2/3}$.
- At step i we are at node u. The probability of getting from there to U is

$$P[u \rightarrow U] \leq \sum_{v \in U} \frac{1}{d(u,v)} / (n/2) \leq \frac{2|U|}{n} \leq 8n^{-1/3}$$

- Therefore, it will take an expected number of $\Omega(n^{1/3})$ steps to use a long-range contact in U. The result follows.

Lower bound, $r > 2$

- Proof strategy: show that it will take many steps to find a long-range contact that takes the message very far.
-

$$
\begin{aligned}
P[(u \to v) \wedge (d(u,v) > m)] &\leq \sum_{v:d(u,v)>m} \frac{1}{d(u,v)^r} \\
&\leq \sum_{j=m+1}^{2n-2} \frac{4j}{j^r} \\
&\leq \ldots \leq \frac{m^{2-r}}{r-2}
\end{aligned}
$$

Proof

- For simplicity set $p = q = 1$, $r = 3$.
- $P[(u \to v) \wedge (d(u,v) > \sqrt{n})] \leq \frac{1}{\sqrt{n}}$. The probability of finding a long-range link that takes us $\sqrt{n}$ steps away in the first $\frac{1}{4}\sqrt{n}$ steps is at most

$$\frac{1}{4}\sqrt{n} \cdot \frac{1}{\sqrt{n}} = \frac{1}{4}$$

- Thus, with probability $\geq \frac{3}{4}$ we move less than $\frac{n}{4}$ in the first $\frac{1}{4}\sqrt{n}$ steps.

Proof

- But this is enough to reach only $\leq \frac{n^2}{2}$ vertices.
- Therefore with probability $\geq \frac{3}{8}$ we need more than $\frac{1}{4}\sqrt{n}$ steps. Making the expected number of steps greater than $\frac{3}{8} \cdot \frac{1}{4}\sqrt{n}$.

THE END