Small World
An algorithmic perspective

Niko Vuokko

October 24th 2007
Information Networks Seminar
Introduction

Small-World phenomenon

Problematic networks

Too fast, too imprecise: $r < 2$
Too introvert: $r > 2$

The navigable network: $r = 2$

Balance in all things
Introduction

Small-World phenomenon

Problematic networks

Too fast, too imprecise: $r < 2$
Too introvert: $r > 2$

The navigable network: $r = 2$
Balance in all things
From legends to theory

- The small-world phenomenon: in a social network each pair of nodes is connected with a fairly short path.
- First significant scientific attention in the 1960’s.
- Milgram *et al.*: people are connected to each other with paths of length six on average.
- Path lengths give the average diameter of the network.
- The claim is a strong requirement for the denseness and the homogeneity of the network.
First attempts on an explanation

- Pool and Kochen gave ground to the claims already before Milgram’s tests.
- They showed that random graphs have very often short diameters, of size $O(\log n)$.
- They didn’t use transitivity: if Anna and Bob both know Cecil, then Anna and Bob probably know each other too.
- But this may easily lead to a strongly-clustered network where the claim can’t hold.
A new model emerges
Searching a good balance

- In 1998 Watts and Strogatz published a network model that tried to balance between these two problems.
- They created networks with both local and long-range links.
- Local links used the $K$-closest-neighbours rule and the long ones were chosen uniformly at random.
- This seems to match the ideas of transitivity and homogeneity quite well.
- This model actually fits to many real-world networks.
## Small-World phenomenon

### Twisting the question
Not just why, but how?

- The random graph theory successfully explains the existence of short diameters.
- But in Milgram’s tests the letters actually found the recipients in those six steps.
- How are strangers able to find these short paths with their very limited information?
- The graph is huge and quite dense. There’s a whole lot of paths and most of them cannot be short.
- Thus the latent information of the network must be more important than it seems at first.
Defining the model

Idea of Kleinberg

- Let the edges be directed.
- Model the network as a two-dimensional $n \times n$ grid and use the Manhattan distance.
- Each element has an outgoing edge to each node within distance $p \geq 1$.
- Each element also has $q$ randomly selected long-range outgoing edges.
- The length of these long-range edges will be decisive.
Pin-pointing the problems

- If we just select the long-range edges uniformly at random, there will be no small-world.
- Look at the nodes at most $\sqrt{n}$ away from target $t$.
- Probability of hitting one of them is $1/\sqrt{n}$.
- It would take $O(\sqrt{n})$ steps to get there in average.
- The problem here is that the closer we are to $t$, the more probably the long-range edges will take us to totally elsewhere.
Defining the model continued

Selecting the long jumps

- Say we are selecting the long-range edges of $u$. A node $v$ will be selected with probability proportional to $d(u, v)^{-r}$.
- This $r$ will be the *concentration exponent*.
- The model now has parameters $p, q$ and $r$, but only $r$ has any real effect on the model’s behaviour.
The goal is to examine decentralized algorithms.

- An entity knows only what it has been told.
- It knows the location of the target, its own links and the grid structure of the lattice.
Networks without a small-world

- When would there be some problems?
- For large $r$ this might be quite obvious.
- In that case the close neighbours of $t$ will be proportionally quite far away from everything else.
- Therefore getting to the neighbourhood will easily take too long, because the long links are not long enough.
What about small $r$’s?

- In the case of a small $r$ there should be no problems with converging on the target.
- So why shouldn’t it work?
  - Problem is that we need precision to hit the proportionally small neighbourhood.
  - Small $r$ makes the algorithms to easily overshoot.
  - This means that the long links don’t give enough advantage.
What about small $r$’s?

- In the case of a small $r$ there should be no problems with converging on the target.
- So why shouldn’t it work?
- Problem is that we need precision to hit the proportionally small neighbourhood.
- Small $r$ makes the algorithms to easily overshoot.
- This means that the long links don’t give enough advantage.
Introduction

Small-World phenomenon

Problematic networks

Too fast, too imprecise: $r < 2$

Too introvert: $r > 2$

The navigable network: $r = 2$

Balance in all things
Too fast, too imprecise: $r < 2$

Seeking the chokepoint

Link lengths

- Remember the uniform case $r = 0$.
- The closer we are, the farther the graph will take us.
- Probabilities of long links should be too large and short links too small:

$$\sum_{v \neq u} d(u, v)^{-r} \geq \frac{n^{2-r}}{(2-r)2^{3-r}}.$$
Seeking the chokepoint

Neighbourhood

- Select a neighbourhood $U$ for $t$ with radius $pn^\delta$.
- We get easily $|U| \leq 4p^2n^{2\delta}$.
- Next let’s calculate how easy it is to find a long-range link to $U$ in $\lambda n^\delta$ steps.
- Define this event to be $E$. 
In a certain step we’ll find a long link to $U$ with probability at most
\[
\frac{q|U|}{(2-r)2^{3-r}n^{2-r}} \leq \frac{(2 - r)2^{5-r}qp^2n^{2\delta}}{n^{2-r}}.
\]

Doing this in $\lambda n^\delta$ steps thus has probability
\[
P(\mathcal{E}) \leq \lambda n^\delta \frac{(2 - r)2^{5-r}qp^2n^{2\delta}}{n^{2-r}} \leq \frac{1}{4}
\]
when selecting $\lambda$ suitably and $\delta = (2 - r)/3$. 
Scrapping parts

- Next we’ll forget the not-so-obviously problematic parts.
- Let $\mathcal{F}$ be the event for $d(s, t) \geq n/4$.
- Easily one sees that $P(\mathcal{F}) \geq 1/2$.
- Now we can conclude that

$$P(\overline{\mathcal{F}} \cup \mathcal{E}) \leq \frac{1}{2} + \frac{1}{4} \implies P(\mathcal{F} \land \overline{\mathcal{E}}) \geq \frac{1}{4}.$$
Scrapping parts continued

- Suppose $F \land \overline{E}$.
- Then $d(s, t) \geq n/4 > p\lambda n^\delta$.
- Getting to $t$ in $\lambda n^\delta$ steps requires now at least one long jump to $U$.
- This is a contradiction. In this case thus all paths to $t$ have length more than $\lambda n^\delta$. 
Cleaning house

- Now we can concentrate on the substantial part of situations where we have the most problems.
- If $X$ denotes the number of steps needed to reach $t$, then

$$E(X) \geq E(X|\mathcal{F} \land \overline{\mathcal{E}}) \cdot P(\mathcal{F} \land \overline{\mathcal{E}}) \geq \frac{1}{4} \lambda n^\delta.$$
Outline

Introduction
Small-World phenomenon

Problematic networks
Too fast, too imprecise: $r < 2$
Too introvert: $r > 2$

The navigable network: $r = 2$
Balance in all things
Brainstorming the solution

- The links should now be more tightly concentrated.
- This means that getting far will be hard.
- Our aim is to prove that most paths are much too short.
Gathering pieces

- Let $\varepsilon = r - 2$ be the number of problems we have.
- If $v$ is a long-range contact of $u$ then we can easily say that $P(d(u, v) > m) \leq m^{-\varepsilon}/\varepsilon$.
- Define $F$ and $X$ similarly as before.
- $E$ will be the event that we find a link longer than $n^\gamma$ in $\lambda n^\beta$ steps.
- We’ll progress just as we did in the $r < 2$ case.
Too introvert: \( r > 2 \)

**Probability of \( \mathcal{E} \)**

- The union bound will give us

\[
P(\mathcal{E}) \leq \lambda n^\beta q n^{-\varepsilon \gamma} / \varepsilon \leq \frac{1}{4},
\]

when choosing \( \lambda \) suitably and \( \beta = \varepsilon \gamma \).

- Now we once again see that \( P(\mathcal{F} \land \overline{\mathcal{E}}) \geq 1/4 \).

- In that case the first \( \lambda n^\beta \) steps will take us only
\( \lambda n^{\beta + \gamma} = \lambda n < n/4 < d(s, t) \) steps closer. (Choose \( \beta + \gamma = 1 \))
Endgame

Requirements $\beta = \varepsilon \gamma$ and $\beta + \gamma = 1$ imply

$$\beta = \frac{\varepsilon}{\varepsilon + 1} \quad \text{and} \quad \gamma = \frac{1}{\varepsilon + 1}.$$ 

We achieve the desired bound using the same tricks as before:

$$E(X) \geq E(X|\mathcal{F} \land \overline{\mathcal{E}}) \cdot P(\mathcal{F} \land \overline{\mathcal{E}}) \geq \frac{1}{4} \lambda n^\beta = \frac{1}{4} \lambda n^{\frac{r-2}{r-1}}.$$
<table>
<thead>
<tr>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Small-World phenomenon</td>
</tr>
<tr>
<td><strong>Problematic networks</strong></td>
</tr>
<tr>
<td>Too fast, too imprecise: ( r &lt; 2 )</td>
</tr>
<tr>
<td>Too introvert: ( r &gt; 2 )</td>
</tr>
<tr>
<td><strong>The navigable network:</strong> ( r = 2 )</td>
</tr>
<tr>
<td>Balance in all things</td>
</tr>
</tbody>
</table>
Introduction

Problematic networks

The navigable network: $r = 2$

Summary

Balance in all things

Plot of the lower bounds

What’s happening in $r = 2$?
Going through a phase

- The probability that $u$ has $v$ as its long-range link is at least $d(u, v)^{-2}/(4 \log(6n))$.
- We say that the algorithm is in phase $j$ if for the current node $u : 2^j < d(u, t) \leq 2^{j+1}$.
- Suppose $B_j$ is the set of nodes $v : d(v, t) \leq 2^j$.
- We easily get $|B_j| > 2^{2j-1}$ and $\forall v \in B_j : d(u, v) < 2^{j+2}$.
- What is the probability of changing phase?

$$P(\text{we move to } B_j) \geq \frac{2^{2j-1}}{4 \log(6n)2^{2j+4}} = \frac{1}{128\log(6n)}.$$
Problematic networks

The navigable network: \( r = 2 \)

Summary

Balance in all things

Phase-shift

- \( X_j \) is now the time spent in phase \( j \):

\[
E(X_j) = \sum_{i=1}^{\infty} P(X_j \geq i) \leq \sum_{i=1}^{\infty} \left( 1 - \frac{1}{128 \log(6n)} \right)^{i-1} = 128 \log(6n).
\]

- There are \( \log n \) phases in total, therefore the expectation of the path lengths is \( E(X) = \mathcal{O}(\log^2 n) \). \( \square \)
The problem in the first cases was that either the closer nodes were too close or the farther nodes were too far.

In the $r = 2$ case all the phases were homogeneous.

The magic behind this is that 2 is the only exponent for which the long-range links are uniformly distributed over distance scales.

Links of length $2^j$ to $2^{j+1}$ have the same probabilities for all $j$. Thus we have enough precision in every case.
Summary

- In a large network one has to manage local and global relations simultaneously.
- Heisenberg uncertainty principle for networks: you can’t have both at the same time, but you can trade them.
- The paper states the balance enabling a subject to grasp the whole and still observe the vicinity.