Adventures in random graphs:
Models, structures and algorithms

Armand M. Makowski

ECE & ISR/HyNet
University of Maryland at College Park
armand@isr.umd.edu

LECTURE 5
Small Worlds

Letter-relaying experiment by S. Milgram

Objective: A source individual must forward a letter to a destination individual

---

Rules:

- **Limited** information about destination individual: Address, name and profession
- Letter **cannot** be addressed directly to the destination individual
- Can only forward letter to someone known on a first name basis

Outcome (1967): A significant fraction of the letters reached their destination in at most **six** hops!

"Six degrees of separation"

Relevance:

- Structure of social networks and importance of social ties
- Routing with limited information in communication networks
- Browsing behavior in WWW

Modeling, modeling, modeling!

- Model social networks as (random) graphs
- Erdős-Renyi graphs $\mathbb{G}(n;p)$?

NO!

Because

In Erdős-Renyi graphs $\mathbb{G}(n;p)$:

- Small diameter:

$$\text{Diam}(\mathbb{G}(n;p)) = O(\log n) \quad \text{when } p \sim \frac{\lambda}{n}$$

- No structure such as geography, professional occupation, etc
- Clustering:

$$\text{Clustering coefficient } (\mathbb{G}(n;p)) = p$$

  and this will be small when $p \sim \frac{\lambda}{n}$

Small World: Random graph with small diameter but high clustering!

Random graph models of small worlds

Q: How do we construct a random graph model of small worlds?

Regular networks + Random short cuts

Typically,

- Regular networks: High average path length but high clustering
- Random shortcuts: Shortens average path length but may decrease clustering

According to Strogratz and Watts

- One-dimensional lattice organized on a ring – L nodes
- Each bound is “rewired” with probability ϕ – Shortcuts

Difficulties: Lack of uniformity, Rewiring can make the network disconnected

According to Newman and Watts

- One-dimensional lattice organized on a ring – L nodes
- Add shortcuts at random

The random graph SW$(n; p)$

- Two-dimensional lattice or grid with $n = L^2$ nodes

$$G_L = \{1, \ldots, L\} \times \{1, \ldots, L\}$$

- **Local** edges to grid neighbors – Deterministic

$$\boldsymbol{x} \sim \boldsymbol{y} \quad \text{if and only if} \quad d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_1 = 1$$

  so between 2 and 4 neighbors

- **Long haul** edges – With probability $p \in (0, 1)$, each node $\boldsymbol{x} \in G_L$ creates a random connection with **one** node selected **uniformly** from $G_L - \{\boldsymbol{x}\}$.

Theorem 1 *For each $p \in (0,1)$, there exists $A = A(p) > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}\left[\mathrm{Diam}\left(\mathrm{SW}(n;p)\right) \le A \log n\right]$$

In the case $p = 0$:

$$\mathrm{Diam}\left(\mathrm{SW}(n;0)\right) = 2L = 2\sqrt{n}$$

and

$$\text{Clustering coefficient } \left(\mathrm{SW}(n;0)\right) = 0$$

According to Kleinberg

- Two-dimensional lattice or grid

$$G_L = \{1, \ldots, L\} \times \{1, \ldots, L\}$$

- **Local** edges to grid neighbors – Deterministic

$$\boldsymbol{x} \sim \boldsymbol{y} \quad \text{if and only if} \quad d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_1 = 1$$

- **Long haul** edges – Each node $\boldsymbol{x}$ has q random connections with destinations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q$ (in $G_L - \{\boldsymbol{x}\}$) selected with probability

$$\frac{d(\boldsymbol{x}, \boldsymbol{y})^{-\alpha}}{\sum_{\boldsymbol{z} \in G,\ \boldsymbol{z} \neq \boldsymbol{x}} d(\boldsymbol{x}, \boldsymbol{z})^{-\alpha}}, \quad \boldsymbol{y} = \boldsymbol{y}_1, \ldots, \boldsymbol{y}_q$$

for some $\alpha > 0$.

The impact of $\alpha > 0$

With $\alpha \simeq 0$, increasingly uniform selection

With $\alpha = 0$ and $q = 1$, we recover SW$(L; p)$ (with $p = 1$)

With $\alpha \uparrow \infty$, only very close neighbors are selected with very high probability

Navigation on small worlds

Distributed routing

A routing algorithm is **distributed** if the decisions made t step t depends only on the knowledge of

- the nodes $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_t$ visited so far
- the coordinates of the destinations of shortcuts generated at these nodes

Figure of merit:

$$\sup_{\boldsymbol{x} \neq \boldsymbol{y}} \left(\mathbb{E}\left[T_{\text{Algo}}(\boldsymbol{x}, \boldsymbol{y}) \right] \right)$$

where

$$T_{\text{Algo}}(\boldsymbol{x}, \boldsymbol{y}) = \begin{array}{l} \text{Number of steps used by Algo} \\ \text{ to reach } \boldsymbol{y} \text{ from } \boldsymbol{x} \end{array}$$

Three cases concerning **navigability**

- $\alpha < 2$
- $\alpha = 2$
- $\alpha > 2$

Efficient routing when $\alpha = 2$

Greedy (decentralized) algorithm: Always forward the message to a grid node as close to the target node as possible

---

This greedy decentralized algorithm makes the small world navigable!

Theorem 2 *We have*

$$\sup_{\boldsymbol{x}\neq\boldsymbol{y}} \left(\mathbb{E}\left[T_{\text{Greedy}}(\boldsymbol{x},\boldsymbol{y})\right]\right) \leq O((\log n)^2)$$

Algo generates iterates

$$\{\boldsymbol{x}_t,\ t = 0, 1, \ldots\} \quad \text{with } \boldsymbol{x}_0 = \boldsymbol{x}$$

In phase j at iterate t if

$$2^j < d\,(\boldsymbol{x}_t, \boldsymbol{y}) \leq 2^{j+1}$$

How many phases?

$$\text{No more than } \log_2 2L \text{ phases}$$

How long in a phase? On the **average**, no more than

$$144\,(1 + \log 2L)\,\text{steps}$$

Thus

$$\begin{aligned}\mathbb{E}\left[T_{\text{Greedy}}(\boldsymbol{x},\boldsymbol{y})\right] &\leq 144\,(1+\log 2L)\cdot\log_2 2L \\ &\leq C\,(\log n)^2 \qquad (1)\end{aligned}$$

for some $C > 0$

If $\boldsymbol{x}_t$ is in phase j, then the probability that a shortcut to a node $\boldsymbol{w}$ leads to a phase $k < j$ is bounded below by

$$\min_{\boldsymbol{x}_t:\ 2^j < d(\boldsymbol{x}_t,\boldsymbol{y}) \leq 2^{j+1}} \frac{\sum_{\boldsymbol{w}:\ d(\boldsymbol{y},\boldsymbol{w})\leq 2^j} d(\boldsymbol{x}_t,\boldsymbol{w})^{-2}}{\sum_{\boldsymbol{z}\neq\boldsymbol{x}_t} d(\boldsymbol{x}_t,\boldsymbol{z})^{-2}}$$

Impossibility of efficient routing when $\alpha \neq 2$

- When $0 < \alpha < 2$, short paths exists but individuals **cannot** determine them in a decentralized manner, i.e.,

$$\mathbb{E}\left[T_{\text{Algo}}(\boldsymbol{x}, \boldsymbol{y})\right] \geq \delta L^{\frac{2-\alpha}{3}}$$

  for some $\delta > 0$

- When $2 < \alpha$, short path no longer exist!