Distributed Routing and in Small Worlds - Double Clustering

Oskar Sandberg - Göteborg University

General Navigability

- The basic structure need not be a circle, but can be a higher dimensional grid, a more general graph, or even a pointset in a metric space.

General Navigability

- The basic structure need not be a circle, but can be a higher dimensional grid, a more general graph, or even a pointset in a metric space.
- The general results extends to:

Efficient routing is possible when:

$$\mathbf{P}(x \rightsquigarrow w) \propto \frac{1}{\#\text{ nodes closer to } x \text{ than } w}$$

Where “closer” is with respect to the same distance function as the greedy routing.

Dynamics

- The question I have been trying to answer: how do navigable networks form?

Dynamics

- The question I have been trying to answer: how do navigable networks form?
- My first attempt: networks are navigable when

$$\mathbf{P}(x \rightsquigarrow z) \propto \mathbf{P}(\text{route to } z \text{ hits } x).$$

Dynamics

- The question I have been trying to answer: how do navigable networks form?
- My first attempt: networks are navigable when

$$\mathbf{P}(x \rightsquigarrow z) \propto \mathbf{P}(\text{route to } z \text{ hits } x).$$

- Curious observation, and leads to a nice way of creating navigable networks.

Dynamics

- The question I have been trying to answer: how do navigable networks form?
- My first attempt: networks are navigable when

$$\mathbf{P}(x \leadsto z) \propto \mathbf{P}(\text{route to } z \text{ hits } x).$$

- Curious observation, and leads to a nice way of creating navigable networks.
- Perhaps difficult to apply to natural settings.

Another Idea

Where else does a sequence of events $A_1, A_2, \ldots$ such that

$$\mathbf{P}(A_k) = \frac{1}{k}$$

occur in probability?

Another Idea

Where else does a sequence of events $A_1, A_2, \ldots$ such that

$$\mathbf{P}(A_k) = \frac{1}{k}$$

occur in probability?

Let $X_1, X_2, \ldots, X_n$ be random variables with an exchangeable joint distribution such that $\mathbf{P}(X_i = X_j) = 0$ when $i \neq j$.

$$\text{Let } A_k = \{X_j < X_k \; \forall j < k\}$$

Interest Model

This observation leads directly to a method for generating navigable graphs.

- Let x associate with each other node w a random variable X_w from such a distribution.

Interest Model

This observation leads directly to a method for generating navigable graphs.

- Let x associate with each other node w a random variable X_w from such a distribution.
- X_w is x's “interest” in w.

Interest Model

This observation leads directly to a method for generating navigable graphs.

- Let x associate with each other node w a random variable X_w from such a distribution.
- X_w is x's "interest" in w.
- Let $x \rightsquigarrow w$ if w is more interesting to x than any node which is closer.

Interest Model

- It follows directly that

$$\mathbf{P}(x \rightsquigarrow w) = \frac{1}{1 + \#\ \text{nodes closer to } x \text{ than } w}$$

This is now independent for each w.

Interest Model

- It follows directly that

$$\mathbf{P}(x \rightsquigarrow w) = \frac{1}{1 + \#\text{ nodes closer to } x \text{ than } w}$$

  This is now independent for each w.
- Expected number of shortcuts from each node is $\log n$.

Interest Model

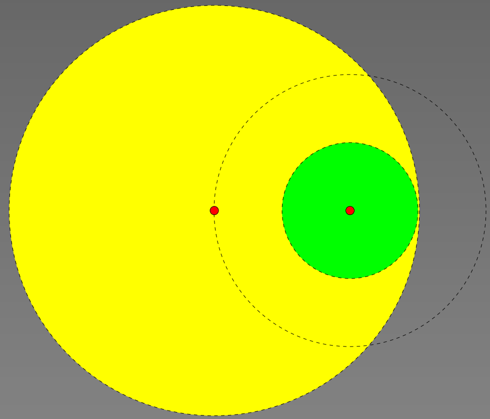
- It follows directly that

$$\mathbf{P}(x \rightsquigarrow w) = \frac{1}{1 + \#\ \text{nodes closer to } x \text{ than } w}$$

This is now independent for each w.

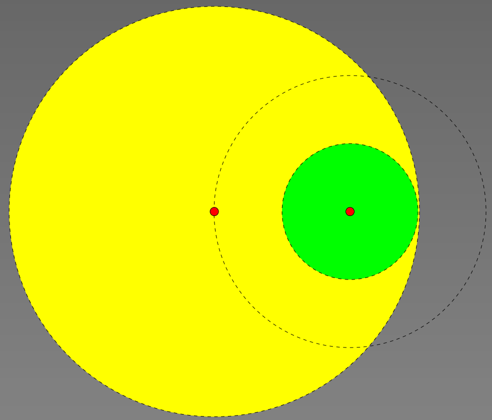
- Expected number of shortcuts from each node is $\log n$.
- From the same argument as above, or directly, one can see that greedy routing takes $O(\log n)$ steps on a graph generated like this.

Interest Model



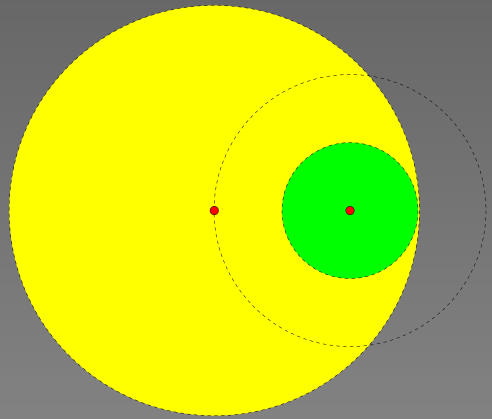
- If d is the distance between x and z, the yellow disk is the vertices within $(3/2)d$ of x and the green within $d/2$ of z.

Interest Model



- If d is the distance between x and z, the yellow disk is the vertices within $(3/2)d$ of x and the green within $d/2$ of z.
- x must have a shortcut to the "most interesting" vertex in the yellow disk.

Interest Model



- If d is the distance between x and z, the yellow disk is the vertices within $(3/2)d$ of x and the green within $d/2$ of z.
- x must have a shortcut to the "most interesting" vertex in the yellow disk.
- This lies in the green disk with constant probability.

Double Clustering

- Let X_w be a node x's interest in w and Y_w be another node y's.

Double Clustering

- Let X_w be a node x's interest in w and Y_w be another node y's.
- X_w and Y_w are **not** likely to be independent.

Double Clustering

- Let X_w be a node x's interest in w and Y_w be another node y's.
- X_w and Y_w are **not** likely to be independent.

A better model:

- With each node x we associate a position $p(x)$ in some metric "space of interests".

Double Clustering

- Let X_w be a node x's interest in w and Y_w be another node y's.
- X_w and Y_w are **not** likely to be independent.

A better model:

- With each node x we associate a position $p(x)$ in some metric "space of interests".
- Let X_w be the inverse of $|p(x) - p(w)|$.

Double Clustering

- Let X_w be a node x's interest in w and Y_w be another node y's.
- X_w and Y_w are **not** likely to be independent.

A better model:

- With each node x we associate a position $p(x)$ in some metric "space of interests".
- Let X_w be the inverse of $|p(x) - p(w)|$.
- That is: $x \rightsquigarrow w$ if $p(w)$ is closer to $p(x)$ than p of any node closer to x.

The Double Clustering Graph

Definition 1 *Let* $(x_i)_{i=1}^n$ *and* $(y_i)_{i=1}^n$ *be two sequences of points without repetition in possibly different spaces* M_1 *and* M_2 *with distance functions* d_1 *and* d_2 *respectively. The digraph* $G = (V, E)$ *is constructed as follows:*

- $V = \{1, 2, \ldots, n\}$.
- $(i, j) \in E$ *if for all* $k \in V$, $k \neq i, j$:

$$d_1(x_i, x_k) < d_1(x_i, x_j) \Rightarrow d_2(y_i, y_k) \geq d_2(y_i, y_j)$$

(Make undirected by removing directionality of the edges.)

The Double Clustering Graph

- The two sets of points $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ are symmetric in the definition.

The Double Clustering Graph

- The two sets of points $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ are symmetric in the definition.
- It becomes a random (spatial) graph when the points are chosen randomly.

The Double Clustering Graph

- The two sets of points $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ are symmetric in the definition.
- It becomes a random (spatial) graph when the points are chosen randomly.
- Simple case: let the two spaces be the same, x_i be some sequence and y_i be a random permutation it. For instance, $M_1 = M_2 = [n]$, and $x_i = i$ and $y_i = \pi(i)$.

The Double Cycle

A simplest case:

Let M_1 be a *directed* cycle ($k \rightarrow k+1 \mod n$), and M_2 be a second cycle with positions permuted by π ($k \rightarrow \ell$ if $\pi(\ell) = \pi(k) + 1 \mod n$).

The Double Cycle

A simplest case:

Let M_1 be a *directed* cycle ($k \to k+1 \mod n$), and M_2 be a second cycle with positions permuted by π ($k \to \ell$ if $\pi(\ell) = \pi(k) + 1 \mod n$).

Let G be the double clustering graph constructed from these.
Lemma: Greedy routing in G for z with respect to M_1 produces a path that monotonically approaches z in M_2. (And vice versa.)

The Double Cycle

Routing for z:

- y had a link to $x \Rightarrow x$ was M_2-closest to it "so far".

The Double Cycle

Routing for z:

- y had a link to $x \Rightarrow x$ was M_2-closest to it “so far”.
- Thus the first node after x which is closer to y than x in M_2 will have an edge from y to it.

The Double Cycle

Routing for z:

- y had a link to $x \Rightarrow x$ was M_2-closest to it "so far".
- Thus the first node after x which is closer to y than x in M_2 will have an edge from y to it.
- If such a node existed before z, y would have routed to it.

The Double Cycle

Routing for z:

- y had a link to $x \Rightarrow x$ was M_2-closest to it "so far".
- Thus the first node after x which is closer to y than x in M_2 will have an edge from y to it.
- If such a node existed before z, y would have routed to it.
- Thus all nodes between x and z in M_1 are further from y than x in M_2.

The Double Cycle

In the double cycle, whether a node x has a link that halves the distance to the destination z is independent of any previous steps.

- Let y be a previous node in the path, R the nodes between x and z in M_1.

The Double Cycle

In the double cycle, whether a node x has a link that halves the distance to the destination z is independent of any previous steps.

- Let y be a previous node in the path, R the nodes between x and z in M_1.
- By the previous lemma, any node in R is further from y than x in M_2 (*less interesting*).

The Double Cycle

In the double cycle, whether a node x has a link that halves the distance to the destination z is independent of any previous steps.

- Let y be a previous node in the path, R the nodes between x and z in M_1.
- By the previous lemma, any node in R is further from y than x in M_2 (*less interesting*).
- Thus the order of these nodes in M_2 has no effect on where y routed.

The Double Cycle

- This occurs with probability $1/2$ and by the previous argument is independent of the earlier path.

The Double Cycle

- This occurs with probability $1/2$ and by the previous argument is independent of the earlier path.
- Halving takes two steps $\Rightarrow$ routing takes $2 \log n$.

The Double Cycle

- This occurs with probability $1/2$ and by the previous argument is independent of the earlier path.
- Halving takes two steps $\Rightarrow$ routing takes $2 \log n$.

- This independence does **not** hold in other double clustering graphs.

The Double Cycle

- This occurs with probability $1/2$ and by the previous argument is independent of the earlier path.
- Halving takes two steps $\Rightarrow$ routing takes $2 \log n$.

- This independence does **not** hold in other double clustering graphs.
- We conjecture that double clustering graphs are navigable in all graph families with limited growth.