Network Mathematics - Why is it a Small World?

Oskar Sandberg

Networks

- Formally, a network is a collection of points and connections between them.

Networks

- Formally, a network is a collection of points and connections between them.
- This is an abstraction which can be used to describe a lot of different systems (technical, physical, biological, sociological, etc. etc.).

Networks

Math	**Graph**	**Vertex**	**Edge**
CS	Network	Node	Link
Physics	System	Site	Bond
Sociology	Social Network	Actor	Tie
		Individual	Friendship
	WWW	Webpage	Link (d)
	Internet	Site	Connection
		Network	Bridge
	Road System	Crossing	Road

Definition and Types

$$G = (V, E)$$

Definition and Types

$$G = (V, E)$$

- V is a set (collection) of vertices.

Definition and Types

$$G = (V, E)$$

- V is a set (collection) of vertices.
- E is a set of edges (pairs (u, v) where $u, v \in V$).

Definition and Types

$$G = (V, E)$$

- V is a set (collection) of vertices.
- E is a set of edges (pairs (u, v) where $u, v \in V$).
- Structured and designed: Corporate LANs, telephone networks.

Definition and Types

$$G = (V, E)$$

- V is a set (collection) of vertices.
- E is a set of edges (pairs (u, v) where $u, v \in V$).
- Structured and designed: Corporate LANs, telephone networks.
- Randomly evolved: The Web, social networks.

Definition and Types

$$G = (V, E)$$

- V is a set (collection) of vertices.
- E is a set of edges (pairs (u, v) where $u, v \in V$).
- Structured and designed: Corporate LANs, telephone networks.
- Randomly evolved: The Web, social networks.
- Somewhere in between: The Internet, P2P Networks.

Network Mathematics

The questions depend on the type of network.

Network Mathematics

The questions depend on the type of network.

- When designing structured networks, questions are usually algorithmic. (How do I create a network with this property?)

Network Mathematics

The questions depend on the type of network.

- When designing structured networks, questions are usually algorithmic. (How do I create a network with this property?)
- When studying randomly generated networks questions tend to analytic. (Does the network have this property?)

Random Graph Theory

The simplest model for a random graph $G(n, p) = (V, E)$:

Random Graph Theory

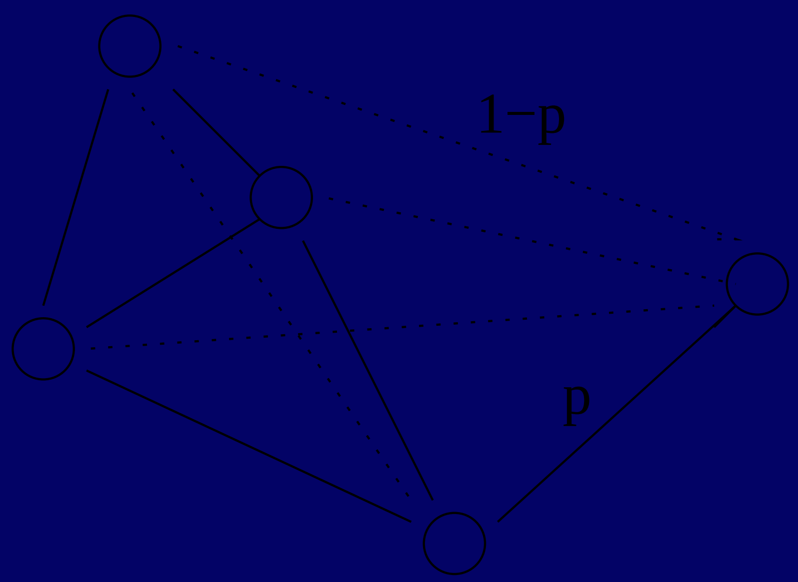
The simplest model for a random graph $G(n, p) = (V, E)$:

- $V = \{0, 1, 2, \ldots, n\}$

Random Graph Theory

The simplest model for a random graph $G(n,p) = (V,E)$:

- $V = \{0,1,2,\ldots,n\}$
- $u \leftrightarrow v$ (that is $(u,v) \in E$) independently and with probability p for every pair of vertices u and v.



Random Graph Theory, cont.

There are a lot of interesting results regarding this type of graph. Relevant properties include:

Random Graph Theory, cont.

There are a lot of interesting results regarding this type of graph. Relevant properties include:

- If $p > 1/n$ “most” of the vertices form one connected cluster.

Random Graph Theory, cont.

There are a lot of interesting results regarding this type of graph. Relevant properties include:

- If $p > 1/n$ “most” of the vertices form one connected cluster.
- If $p > \log n/n$ all of the vertices are connected.

Random Graph Theory, cont.

There are a lot of interesting results regarding this type of graph. Relevant properties include:

- If $p > 1/n$ “most” of the vertices form one connected cluster.
- If $p > \log n/n$ all of the vertices are connected.
- The “diameter” of the connected cluster is $\log n$.

Random Graph Theory, cont.

- However, this isn't a great model for studying real world networks.

Random Graph Theory, cont.

- However, this isn't a great model for studying real world networks.
- The vertex *degree* is highly concentrated (varies little).

Random Graph Theory, cont.

- However, this isn't a great model for studying real world networks.
- The vertex *degree* is highly concentrated (varies little).
- Triangles are relatively rare.

Random Graph Theory, cont.

- However, this isn't a great model for studying real world networks.
- The vertex *degree* is highly concentrated (varies little).
- Triangles are relatively rare.

In recent years, new models have been introduced for networks with various properties.

Small World Phenomenon

- The “Small World Phenomenon” is that many naturally occurring networks have a small graph diameter.

Small World Phenomenon

- The "Small World Phenomenon" is that many naturally occurring networks have a small graph diameter.
- It was famously illustrated for social networks by Stanley Milgram in 1967.

Stanley Milgram

Small World Phenomenon

- The “Small World Phenomenon” is that many naturally occurring networks have a small graph diameter.
- It was famously illustrated for social networks by Stanley Milgram in 1967.
- He experimented by having volunteers in Omaha, Nebraska forward letters to a stockbroker in Boston through friends.

Stanley Milgram

Small World Phenomenon

- The "Small World Phenomenon" is that many naturally occurring networks have a small graph diameter.
- It was famously illustrated for social networks by Stanley Milgram in 1967.
- He experimented by having volunteers in Omaha, Nebraska forward letters to a stockbroker in Boston through friends.
- Milgram reported that on average the packages reached their destination in only six steps.

Stanley Milgram

An Example.

Using social networking websites (in this case orkut) we can create similar paths. For example:

An Example.

Using social networking websites (in this case orkut) we can create similar paths. For example:

From:

Miss Fernanda Trincado, Santa Cruz del Sol, Brazil.

An Example.

Using social networking websites (in this case orkut) we can create similar paths. For example:

From:

Miss Fernanda Trincado, Santa Cruz del Sol, Brazil.

To:

Oskar Sandberg, Göteborg, Sweden.

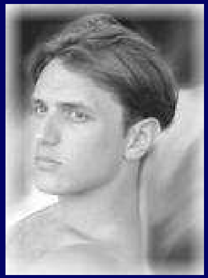
An Example. cont.

Fernanda
Trincado

An Example. cont.

Fernanda Trincado

Knows:

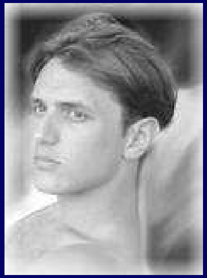


Leandro Toledo

An Example. cont.

Fernanda Trincado

Knows:



Leandro Toledo

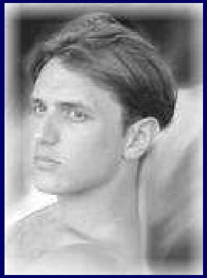
Is related to:

Marcelo Toledo

An Example. cont.

Fernanda Trincado

Knows:



Leandro Toledo

Is related to:

Marcelo Toledo

Who knows:

Nat Friedman

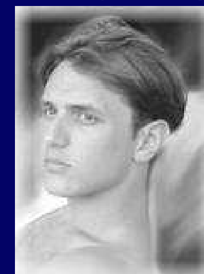
An Example. cont.

Fernanda Trincado

Who knows:

Nat Friedman

Knows:



Leandro Toledo

Who knows:

Roger Dingledine

Is related to:

Marcelo Toledo

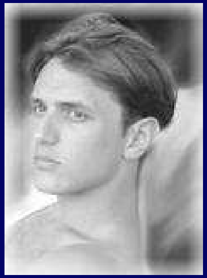
An Example. cont.

Fernanda Trincado

Who knows:

Nat Friedman

Knows:



Leandro Toledo

Who knows:

Roger Dingledine

Is related to:

Marcelo Toledo

Who Knows:

Oskar Sandberg

Mathematical Models.

- The simple type of random graphs discussed before have low diameter.

Mathematical Models.

- The simple type of random graphs discussed before have low diameter.
- As noted, however, they are not a good model for social networks.

Mathematical Models.

- The simple type of random graphs discussed before have low diameter.
- As noted, however, they are not a good model for social networks.
- It isn't possible to *search* in them.

Kleinberg's Model

- For searching to be possible, vertices need to have locations, and whether $u \leftrightarrow v$ should depend on the distance between them ($d(u,v)$).

Kleinberg's Model

- For searching to be possible, vertices need to have locations, and whether $u \leftrightarrow v$ should depend on the distance between them ($d(u,v)$).
- Let $\mathbf{P}(x \leftrightarrow w) \propto 1/d(x,w)^{\alpha}$, where $d(x,w)$ is the distance between them.

Kleinberg's Model

- For searching to be possible, vertices need to have locations, and whether $u \leftrightarrow v$ should depend on the distance between them ($d(u,v)$).
- Let $\mathbf{P}(x \leftrightarrow w) \propto 1/d(x,w)^{\alpha}$, where $d(x,w)$ is the distance between them.
- α tunes the degree of "locality" the shortcuts.

Kleinberg's Model

- For searching to be possible, vertices need to have locations, and whether $u \leftrightarrow v$ should depend on the distance between them ($d(u,v)$).
- Let $\mathbf{P}(x \leftrightarrow w) \propto 1/d(x,w)^{\alpha}$, where $d(x,w)$ is the distance between them.
- α tunes the degree of "locality" the shortcuts.
- Route using *greedy routing*: step to the neighbor which is closest to destination.

Kleinberg's Model, cont.

Efficient routing is possible when α is such that:

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

This can be seen to be $\alpha = d$, where d is the dimension of the space (2 in the simulations).

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?
- Kleinberg’s result is mostly negative: for the vast majority of networks, searching is not possible.

Dynamics

- The question I have been trying to answer: how do navigable networks form?
- Kleinberg’s result is mostly negative: for the vast majority of networks, searching is not possible.
- Why should one expect real-world networks to have the necessary edge distribution?

Some Math

Take the numbers $1, 2, \ldots, n$ and draw them in a random order. What is the probability that the k-th number drawn is the biggest yet?

Some Math

Take the numbers $1, 2, \ldots, n$ and draw them in a random order. What is the probability that the k-th number drawn is the biggest yet?

- Consider only the relative size of the first k numbers drawn.

Some Math

Take the numbers $1, 2, \ldots, n$ and draw them in a random order. What is the probability that the k-th number drawn is the biggest yet?

- Consider only the relative size of the first k numbers drawn.
- These have a random order: each is equally likely to be the biggest of them.

Some Math

Take the numbers $1, 2, \ldots, n$ and draw them in a random order. What is the probability that the k-th number drawn is the biggest yet?

- Consider only the relative size of the first k numbers drawn.
- These have a random order: each is equally likely to be the biggest of them.
- Thus the k-th number has probability $1/k$ of being the biggest one yet.

Interest Model

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

- Let u associate with each other node v a random quantity representing u's interest in v.

Interest Model

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

- Let u associate with each other node v a random quantity representing u's interest in v.
- Let $u \leftrightarrow v$ if u is more interesting to v than any node which is closer.

Interest Model

- It follows that

$$\mathbf{P}(u \leftrightarrow v) = \frac{1}{1 + \#\text{ nodes closer to } u \text{ than } v}$$

This is now independent for each v.

Interest Model

- It follows that

$$\mathbf{P}(u \leftrightarrow v) = \frac{1}{1 + \#\text{ nodes closer to } u \text{ than } v}$$

This is now independent for each v.

- Expected number of shortcuts from each node is $\log n$.

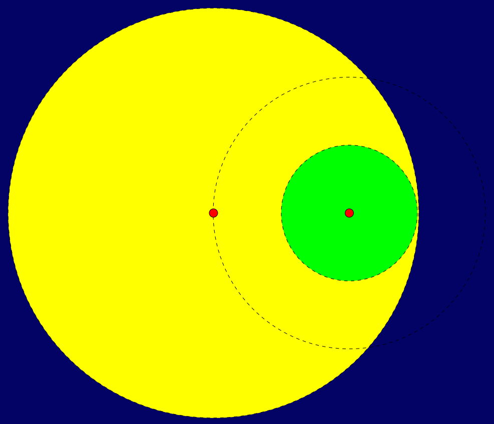
Interest Model

- It follows that

$$\mathbf{P}(u \leftrightarrow v) = \frac{1}{1 + \#\text{ nodes closer to } u \text{ than } v}$$

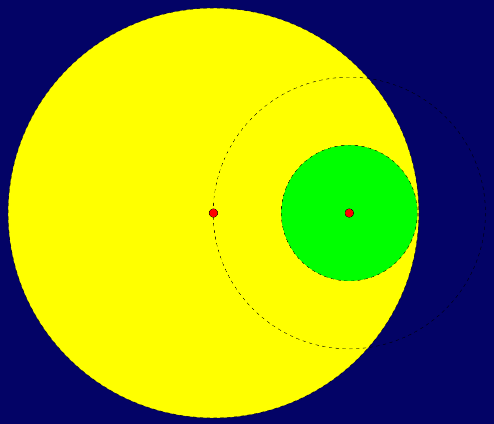
  This is now independent for each v.
- Expected number of shortcuts from each node is $\log n$.
- One can see that greedy routing takes $O(\log n)$ steps on a graph generated like this.

A Proof



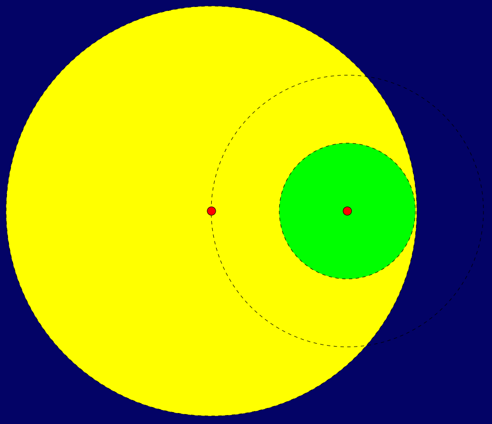
- If d is the distance between u and v, the yellow disk is the vertices within $(3/2)d$ of u and the green within $d/2$ of v.

A Proof



- If d is the distance between u and v, the yellow disk is the vertices within $(3/2)d$ of u and the green within $d/2$ of v.
- u must have a shortcut to the very “most interesting” vertex in the yellow disk.

A Proof



- If d is the distance between u and v, the yellow disk is the vertices within $(3/2)d$ of u and the green within $d/2$ of v.
- u must have a shortcut to the very “most interesting” vertex in the yellow disk.
- The probability that that vertex is in the green part is $1/9$.

Double Clustering

- The above model is still not very realistic. For example, u's interest in v and v's interest in u are **not** likely to be independent.

Double Clustering

- The above model is still not very realistic. For example, u's interest in v and v's interest in u are **not** likely to be independent.

A better model:

Double Clustering

- The above model is still not very realistic. For example, u's interest in v and v's interest in u are **not** likely to be independent.

A better model:

- With each vertex u we associate a position $p(u)$ in some "space of interests".

Double Clustering

- The above model is still not very realistic. For example, u's interest in v and v's interest in u are **not** likely to be independent.

A better model:

- With each vertex u we associate a position $p(u)$ in some "space of interests".
- Let u's interest in v be the inverse of $|p(u) - p(v)|$.

Double Clustering

- The above model is still not very realistic. For example, u's interest in v and v's interest in u are **not** likely to be independent.

A better model:

- With each vertex u we associate a position $p(u)$ in some "space of interests".
- Let u's interest in v be the inverse of $|p(u) - p(v)|$.
- That is: $u \leftrightarrow v$ if $p(u)$ is closer to $p(v)$ than p of any node closer to u to than v.

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.)