Lecture 4: Connect (4/4)

How the friendship we form connect us?
Why we are within a few clicks on Facebook?

COMS 4995-1: Introduction to Social Networks
Tuesday, September 18th
This course is now officially “sexy [kinda]” congratulations!

1st assign. due Thursday 4:10pm
- Part A+C on papers!
- Part B(+raw results of C) on dropbox
- Sign the cover sheet
- 1 late days: 5% (you have 3 free during semester)
* Milgram’s “small world” experiment

* It’s a “combinatorial small world”
* It’s a “complex small world”
* It’s an “algorithmic small world”
Main idea: social networks follows a structure with a random perturbation

Formal construction:
1. Connect all nodes at distance in a regular lattice
2. Rewire each edge uniformly with probability $p$ (variant: connect each node to $q$ neighbors, chosen uniformly)

Collective dynamics of ‘small-world’ networks.
Main idea: social networks follows a structure with a random perturbation

Collective dynamics of ‘small-world’ networks.
* Milgram’s “small world” experiment

* It’s a “combinatorial small world”
  * It’s a “complex small world”
* It’s an “algorithmic small world”
Where are we so far?

Analogy with a cosmological principle

– Are you ready to accept a cosmological theory that does not predict life?

In other words, let’s perform a simple sanity check
1. Consider a randomly augmented lattice (N nodes)
1. Consider a randomly augmented lattice (N nodes)
2. Perform “small world” Milgram experiment

Can you tell what will happen?

(a) The folder arrives in 6 hops
(b) The folder arrives in $O(\ln(N))$ hops
(c) The folder never arrives
(d) I need more information

\[ \text{the folder arrives in } \leq \sqrt{N} \]
(a) The folder arrives in 6 hops

NOT TRUE

* It actually does look like a naive answer
* More precisely:
  – By previous result we know that shortest paths is of the order of \( \ln(N) \), which contradicts this statement.
(b) The folder arrives in $O(\ln(N))$

ACCORDING TO OUR PRINCIPLE, OUGHT TO BE TRUE BECAUSE IT WAS OBSERVED BY MILGRAM

* A sufficient condition for this to be true is:
  – Milgram’s procedure extract shortest path

* Answering this critical question boils down to an algorithmic problem
A thought experiment

(c) The folder never arrives

SEEMS UNLIKELY

unless the procedure is badly designed (cycle)
or we model people dropping
or if the grid contains hole
(d) I need more information

* In particular, how to model Milgram’s procedure
  * “If you do not know a target, forward the folder to your friend or acquaintance that is most likely to know her.”
What is Greedy Routing?

* A mathematical model of what Milgram measured
  - Participants know where the target is located
  - They use grid information + shortcuts “incidentally”
  
  N.B.: Grid “dimensions” can describe geography or other sociological property (occupation, language)

* Example:
How does greedy routing perform?

* Does it extract the shortest path?
  - Not necessarily, this is why we need to analyze it!
* Case study: dimension \( k=1 \), target \( t \), starting from \( u_0 \)
  - We introduce interval: \( I_l = \{ u \in V \text{ such that } |u - u_0| \leq l \} \)
  - The greedy routing constructs a path \( s = U_0, U_1, U_2 \)
  we denote the end-point of the \( i \)-th shortcuts as \( X_i \)

\[
\forall n \forall E \Pr \left[ \bigcup_{i=1,\ldots,n} \{ X_i \in I_l \} \right] \leq \sum_{i=1,\ldots,n} \Pr [X_i \in I_l] \leq \frac{n \cdot \binom{2r+1}{2}}{N}
\]
\[ \forall n \forall e \]

* CLAIM: If none of \{ X_1, X_2, \ldots, X_n \} are in \( I_l \) and we start from \( u_0 \) outside \( I_l \)

  – Then greedy routing needs at least \( \min(n, l) \) steps
How does greedy routing perform?

\[
\Pr\left[ \bigcup_{i=1,\ldots,n} \{ X_i \in I_l \} \right] \leq \sum_{i=1,\ldots,n} \Pr [X_i \in I_l] \leq \frac{n \times (2014)}{2 \sqrt{N}}
\]

* Fixing \( n = l = \frac{1}{2} \sqrt{N} \), this event has proba \( \leq 1/2 \)
  
  – So with proba \( \geq 1/2 \), \{ \( X_1, X_2, \ldots, X_n \) \} are not in \( I_l \)

* On this event, assuming s not in \( I_l \)
  
  – Greedy routing needs more than \( n \) steps
  
  – Or it has to reach \( t \) from boundary of \( I_l \), using \( l \) steps

Greedy routing needs \( \frac{1}{2} \sqrt{N} \) steps
A thought experiment

* In a line Milgram’s uses $\frac{1}{2} \sqrt{N}$ steps
  - square root is not satisfying for small world
  - Not much better when $k>1$! $\frac{1}{4} N^{\frac{1}{k+1}}$
  - even worse, the proof applies to any distributed alg.

* Our sanity check test has grandly failed!
  - “Small world” results explain that short paths exist
  … finding them remains a daunting algorithmic task
Milgram’s “small world” experiment

It’s a “combinatorial small world”

It’s a “complex small world”

It’s an “algorithmic small world”

– Beyond uniform random augmentation
**Autopsy of “Small-world” failure**

* In a uniformly augmented lattice shortcuts do exist
  – About $\sqrt{N}$ shortcuts leads to $I_l$ when $l = \sqrt{N}$.

* But they are dispersed among $N$ nodes

* Moreover, previous steps do not lead to progress
  – So need about $N/\sqrt{N} = \sqrt{N}$ trials

* Is there another augmentation?
The 10 papers that will make you a social expert
10 sociological must-reads

Homophily

- People “love those who are like themselves”, “Similarity begets friendship”
  - Nichomachean Ethics, Aristotle & Phaedrus, Plato

- Do you think homophily produces or hinder small world?

Homophily in Online Dating: When Do You Like Someone Like Yourself?
Andrew T. Fiore and Judith S. Donath
MIT Media Laboratory
20 Ames St., Cambridge, Mass., USA
{fiore, judith}@media.mit.edu

ABSTRACT
Psychologists have found that actual and perceived similarity between potential romantic partners in demographics, attitudes, values, and attractiveness correlate positively with attraction and, later, relationship satisfaction. Online dating systems provide a new way for users to identify and communicate with potential partners, but the information they provide differs dramatically from what a person might glean from face-to-face interaction. An analysis of dyadic interactions of approximately 65,000 heterosexual users of an online dating system in the U.S. showed that, despite these differences, users of the system sought people like them much more often than chance would predict, just as in the offline world. The users’ preferences were most strongly same-sexing for attributes related to the life course, like marital history and whether one wants children, but they also demonstrated significant homophily in self-reported physical build, physical attractiveness, and smoking habits.

Author Keywords
Online personals, attraction, computer-mediated communication, online dating, relationships

ACM Classification Keywords
H.5.3. Group and Organization Interfaces; Asynchronous communication

NATURE OF ONLINE PERSONALS DATA
We analyzed data from one online dating system in particular. Through an agreement brokered by the Media Laboratory with an online dating Web site (the “Site”), we obtained access to a snapshot of activity on the site over an eight-month period, from June 2002 through February 2003. The data included users’ personal profile information, their self-reported preferences for a mate, and their communications via the site’s private message system with other users. Anonymous ID numbers distinguished unique users.

Table 1 indicates which profile characteristics users could specify about themselves and about the partners they would like to meet.

Data about private messages exchanged by the users included the sender, recipient, subject, text, date and time of delivery, and whether the recipient had read the message.
Augmenting lattice with a bias

* What if the augmentation exhibits a bias
  – Most of the people you know are near,
  – Occasionally, you know someone outside

* Does this break the lower bound proof?

Does finding a neighborhood of $t$ becomes easier?
How to model augmentation bias

* Formal construction:
  1. Connect nodes at distance p in a regular lattice
  2. Connect each node to q other nodes, chosen with a biased probability
  3. p=q=1 to simplify

The small-world phenomenon: An algorithmic perspective.
How to model augmentation bias

* Formal construction:
  1. Connect nodes at distance $p$ in a regular lattice
  2. Connect each node to $q$ other nodes, chosen with a biased probability

\[
\Pr [u \leftrightarrow v] = \frac{1}{\sum_{v \neq u} \frac{1}{\|u-v\|^r}}
\]

* $r$ may be called the clustering coefficient
  * If a node is twice further, probability is $2^r$ times less

The small-world phenomenon: An algorithmic perspective.
Impact of clustering coefficient

Small values of $r$
Approaches uniform augmentation

Large values of $r$
Approaches original lattice
(a) Yes, finding a neighborhood of $t$ becomes easier

A PRIORI NOT TRUE

- It is easier **only if** you are already near the target
- In general, it can take a larger number of steps
(b) Yes, for another reason
  – All positions are not equal, hence progress is possible
  – As shortcut are used recursively, probability increases
  – So we need to study the sequence of progress