Lecture 3: Connect (3/4)

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 13th
Please complete the survey

How do you feel about the pace of the class?
- SAMPLE
- OK!

Which programs are you attending?
- SAMPLE

What is your primary interest for taking this course?
- SAMPLE
Or the “shadow” survey

Overall what were the reason(s) you choose not to take the class?

- The instructor the lectures did not seem to make sense
- The angle: I think the idea is important but I want to do coding
- The load; although I think it's useful, it's just too demanding. I love...
- External reason: I wish I could take it, but I had to focus. I may be...
- Other

SAMPLE OUF!
* Milgram’s “small world” experiment

* It’s a “combinatorial small world”
  o End of the proof of “3 connected nodes”,
  o More properties
  o A words on regular graph

* It’s a “complex small world”

* Thursday: It’s an “algorithmic small world”
How to look at large graphs?

* Consider increasing sequence of graphs:
  - $G_n=(V_n,E_n)$, where $V_n=\{1,2,\ldots,n\}$
  - $E_n$ is random (e.g., edge $\{i,j\}$ occurs w.p. $p(n)$)

* When is an assertion $A$ satisfied?
  - Find a threshold function: a function $t(n)$ such that
    
    $$(i) \quad \text{When } \lim_{n \to \infty} \frac{p(n)}{t(n)} = 0, \text{ then } P[G_n \text{ satisfies } A] \to_{n \to \infty} 0.$$  

    $$(ii) \quad \text{When } \lim_{n \to \infty} \frac{p(n)}{t(n)} = \infty, \text{ then } P[G_n \text{ satisfies } A] \to_{n \to \infty} 1.$$
Example 2:

* Let A be \( \{ G_n \text{ has at least 3 connected nodes} \} \)
  
  o i.e., \( G_n \) has a pair of adjacent edges

* Thm: Let \( N_n \) be \( \# \{ \text{adjacent edges} \} \)

\[
N_n = \sum_{\text{e, e' adj.}} X_{e, e'}
\]

\[
\frac{p(n)}{1/n^{3/2}} \rightarrow_{n \to \infty} 0 \quad \Rightarrow \quad P[N_n > 0] \rightarrow_{n \to \infty} 0
\]

\[
\frac{p(n)}{1/n^{3/2}} \rightarrow_{n \to \infty} \infty \quad \Rightarrow \quad P[N_n > 0] \rightarrow_{n \to \infty} 1 \quad (i.e., P[N_n = 0] \rightarrow 0)
\]

o What we have shown so far

\[
\frac{p(n)}{1/n^{3/2}} \rightarrow_{n \to \infty} 0 \quad \Rightarrow \quad E[N_n] \rightarrow_{n \to \infty} 0
\]

\[
\frac{p(n)}{1/n^{3/2}} \rightarrow_{n \to \infty} \infty \quad \Rightarrow \quad E[N_n] \rightarrow_{n \to \infty} \infty
\]

\[
\Rightarrow P[N_n > 0] \rightarrow 0 \quad \text{with Markov ineq.}
\]
THM: \[ P[X = 0] \leq \frac{E[(X - E[X])^2]}{E[X]^2} = \frac{Var(X)}{E[X]^2} \]

So we want \( \frac{Var(N_n)}{1E[N_n]^2} \to 0 \) to conclude.

Lemma: \[ Var[N_n] \leq E[N_n] + \sum_{e,e' \text{adj.} \neq f,f' \text{adj.}} Cov(X_e,e', X_{f,f'}) \]

- Where \( Cov(X,Y) = E[ (X - E[X])(Y - E[Y]) ] \)
- In particular, \( X \) and \( Y \) independent implies \( Cov(X,Y) = 0 \)
Proof of the lemma

\[ \text{Var}(N_n) = \mathbb{E} \left[ (N_n - \mathbb{E}[N_n])^2 \right] = \mathbb{E} \left[ \left( \sum_{e,e'} (X_{e,e'} - \mathbb{E}[X_{e,e'}]) \right)^2 \right] \]

we can develop the product into

\[ = \mathbb{E} \left[ \sum_{e,e'} X_{e,e'}^2 - 2 \mathbb{E}[X_{e,e'}] \sum_{e,e'} X_{e,e'} + \mathbb{E}[X_{e,e'}]^2 \right] + \mathbb{E} \left[ \sum_{e} (X_{e,e} - \mathbb{E}[X_{e,e}]) X_{e,e} \right] + \mathbb{E} \left[ \sum_{(f,e')} (X_{e,e'} - \mathbb{E}[X_{e,e'}]) X_{e,e'} \right] \]

A₁

A₂
Proof of the lemma (2)

A. \[ \left| E \left[ \sum_{e,e'} (X_{e,e'} - \mu[X_{e,e'}])^2 \right] \right| \leq \sum_{e,e'} \mu[X_{e,e'}^2] = \sum_{e,e'} \mu[X_{e,e'}] = \mu[N_n] \text{ as } X_{e,e'} = X_{e,e'} \]

\[ A_2 = \sum_{e,e \neq f,f} \mu \left[ (X_{e,e'} - \mu[X_{e,e'}]) \times (X_{f,f'} - \mu[X_{f,f'}]) \right] \]

\[ = \sum_{e,e' \neq f,f} \text{Cov}(X_{e,e'}, X_{f,f'}) \] which proves the result.
What is $\text{Cov}(X_e,e', X_f,f')$?

- If $\{e,e'\} \cap \{f,f'\} = \emptyset \Rightarrow$ disjoint edges $\Rightarrow X_{e,e'}$ independent of $X_{f,f'}$
  
  hence $\text{Cov}(X_{e,e'}, X_{f,f'}) = 0$

- If $|\{e,e'\} \cap \{f,f'\}| = 1 \Rightarrow e,e'$ and $f,f'$ have exactly one edge in common

$$\text{Cov}(X_{e,e'}, X_{f,f'}) = |E[(X_{e,e'} - \mathbb{E}[X_{e,e'}])(X_{f,f'} - \mathbb{E}[X_{f,f'}])]$$

$$\leq |E[X_{e,e'} - X_{f,f'}] + \mathbb{E}[X_{e,e'}] | \mathbb{E}[X_{f,f'}]| \leq 2p(n)^3$$

$$= p(n)^3$$
So \( \sum \text{Cov}(X_{e,e'}, X_{f,f'}) \leq \#\{ (e,e') \neq (f,f) \text{ such that } |\{ e, e' \} \cap \{ f, f' \} | = 1 \} \times 2p(n)^3 \)

\( \leq \frac{n(n-1)(n-2)}{2} \times 2 \times 2 \times x(n-3) \) for choice of \( e,e' \) and choice of which common edge adjacent.

Hence

\[ \frac{\text{Var}(N_n)}{\text{E}[N_n]^2} \leq \frac{|\text{E}[N_n]|}{\text{E}[N_n]^2} + \frac{c n^4 p(n)^3}{n^2} \]

\[ = \frac{1}{n^2 p(n)^4} \to 0 \] as \( n \to \infty \).
Milgram’s “small world” experiment

It’s a “combinatorial small world”
  - End of the proof of “3 connected nodes”,
  - More properties
  - A words on regular graph

It’s a “complex small world”

Thursday: It’s an “algorithmic small world”
**Properties of Unif. Rand. Graph**

<table>
<thead>
<tr>
<th>Flavor 1 and 2: series of phase transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conn. Comp. &lt;ln(n)</td>
</tr>
<tr>
<td>isolated nodes</td>
</tr>
<tr>
<td>Diameter O(ln(n))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p</th>
<th>0</th>
<th>1/n</th>
<th>ln(n)(1+ω(1))/n</th>
<th>ω(ln(n))/n</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/n</td>
<td>ln(n)(1+ω(1))/n</td>
<td>ω(ln(n))/n</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- All monotone properties have sharp threshold

E. Friedgut, G. Kalai, Proc. AMS (1996) following Erdös-Rényi’s results
* Milgram’s “small world” experiment

* It’s a “combinatorial small world”
  - End of the proof of “3 connected nodes”,
  - More properties
  - A words on regular graph

* It’s a “complex small world”

* Thursday: It’s an “algorithmic small world”
Flavor 3 (each node with d random neighbors)

How to construct?
1. Assign d “semi-edge” to each node
2. Pair “semi-edge” into edge randomly
3. Remove self-loops and multi-edges (rejection)
(NB: This construction can be used for any degree dist.)

Degree constant: should we expect connectivity?
**Properties of Unif. Rand. Graph**

* Transitions is even faster:
  - d=1: collection of disjoint pairs (not connected)
  - d=2: collection of disjoint cycles (a.s. disconnected)
  - d=3: connected, small diameter, in fact much more ...
* Thm: for d≥3 there exists γ>0 such that for any n
  \( G_n \) is a \( γ \)-expander with high probability:

For any subset \( A \subseteq V \),

\[
\frac{|\partial (A, A^c)|}{\min (|A|, |A^c|)} \geq γ.
\]

\( \partial(A, B) = \{ (u, v) \in E \mid u \in A \text{ and } v \in B \} \)
Assuming random uniform connections
  o Phase transitions (i.e. tipping points) are the rule.
  o Connectivity, small diameter, even expander property can be guaranteed without strict coordination
    ... this may outclass deterministic structure!
  o Elegant and precise mathematical formulation

Is small world only a reflection of the combinatorial power of randomness?
Outline

* Milgram’s “small world” experiment
* It’s a “combinatorial small world”
* It’s a “complex small world”
* It’s an “algorithmic small world”
Social networks are not flat

* Do not underestimate the effect of the inbreeding!
  o How many of your friends are friends (strong ties)?
  o Which are friends with different people (weak ties)?
* Several evidences that these play different roles
  o Strong ties: resistance to innovation and danger (e.g., reduce risk of teen suicide)
  o Weak ties: propagate information, connect groups (e.g., who to ask when looking for a job)


Suicide and Friendships Among American Adolescents.
Can we quantify the difference?

- A metric of strong ties
  - Clustering coefficient
  - Also writes \( P \left( (v, w) \in E \mid (u, v) \in E \text{ and } (u, w) \in E \right) \)
- Structured networks: rings, grids, torus
  - Clustering remains constant in large graph
  - As indeed, in most empirical data set
- Random graphs
  - Clustering coefficient goes to zero as \( n \to \infty \)
    ... most results are biased towards weak ties
Main idea: social networks follows a structure with a random perturbation

Formal construction:
1. Connect nearby nodes in a regular lattice
2. Rewire each edge uniformly with probability $p$
   (variant: add a new uniform edge with probability $p$)

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

Collective dynamics of ‘small-world’ networks.
Between order and randomness

As a function of $p$:
- $C(p)$: clustering coefficient
- $L(p)$: median length of the shortest path
- Both normalized by $p=0$.

For small $p$ ($\sim 0.01$)
- large clustering
- small diameter
A few rewire suffice!
Thm: ∀p>0, augmented lattice has small diameter, there is A such that \( P[D≤A \ln(N)] \rightarrow 1 \)

- Elementary proof if assuming regular augmentation
- Otherwise proof similar to Uniform Random Graph

More on “randomly perturbed structures”
- Degree distribution,
- assortative mixing,
* Milgram’s “small world” experiment

* It’s a “combinatorial small world”
* It’s a “complex small world”
* It’s an “algorithmic small world”
  o UPCOMING NEXT WEEK