Lecture 3.-1

Jan 30, 2024

1. How's homework.
2. starting from features of a "social network."
   (emphirical)

Goal:
- Set assumptions
- see if the assump. generates interesting patterns.

"random"

network property	real-world	ER	configuration
deg. distrib.	heavy tailed	Poisson($\langle k \rangle$)	
diameter	"small" ($\propto \log n$)	$\Theta(\log(n))$	
clustering coeff.	social : moderate non-social : low	$\Theta(1/n)$	...
reciprocity	high	$\Theta(1/n)$	
giant component	very common	$\langle k \rangle > 1$	

How should we interpret the value of $\langle k \rangle$ $\langle l \rangle$ l_{max} C $r_{(i)}$ Σ etc ?

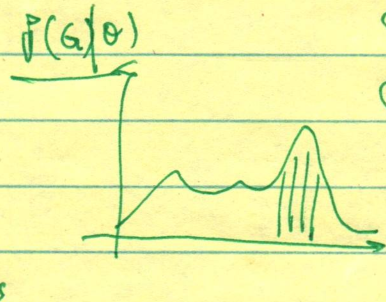
$f : \theta \rightarrow G$

$\Downarrow$

$\mathbb{P}(G \mid \theta)$

$\uparrow$

new random graph models



① mathem. analysis

② null modeling (statistical modeling)

3-1 2

$$\langle x \rangle = \sum_{G,\theta} x(G)\, P(G|\theta)$$ — if you will.

Four main types of random graph models.

1. Erdös–Rényi random graph (simplest of all R.G models) $P(i \to j \mid p)$

2. configuration model — random graph conditioned on a fixed degree $\{k_i\}$ sequence.

3. modular random graphs — models of "communities" "modules" "clusters"

$$P(i \to j \mid z_i, z_j, \theta).$$

4. latent space models $P(i \to j \mid \underline{d(i,j)})$ — distance in the latent space

not in this course.

All ~~assump~~ assume $P(i \to j \mid \theta)$ are conditionally independent.

③

The Erdös-Rényi Random Graph.

- Definition
- Properties.

Mean degree

Degree distribution

motifs / reciprocity / clus. coeff.

- Surprise #1 : phase transition in connectedness

- Surprise #2 : "small world" networks / shortest paths.

" Question : Are ER realistic? "

④

Def:

$G(n, p)$

iid prob. that $i \leftrightarrow j$

(simple graphs only)

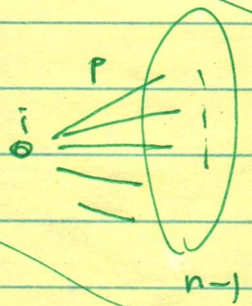
nodes

- defines an ensemble $P(G|p)$
- "how many? $2^{\binom{n}{2}} \approx 2^{n^2}$ (BIG !!)

$$\forall\, i, j \quad A_{ij} = A_{ji} = \begin{cases} 1 & \text{if prob. } p \\ 0 & \text{otherwise} \end{cases}$$

$c := \langle k \rangle$ often $c = 2 \sim 4$.

$$p = \frac{c}{(n-1)} = O\left(\frac{1}{n}\right)$$



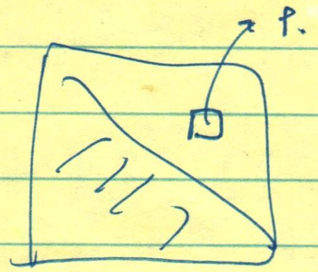
what do we get when $p=1$?

$p=0$?

How to "generate $G(n,p)$" with a computer?

$\Theta(n^2)$

```
for i = 1 .. n
    for j = i+1 .. n
        if rand() > p
            A_ij = A_ji = 1
```



⑤

Prop. of $G(n,p)$ "the random graph model"

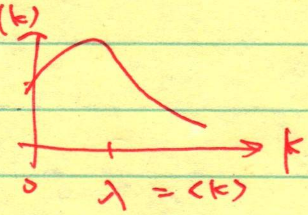
- trad. "simple graph" but generalizable to directed graphs, bip., multilayer.

  w/ w/o self-loops.

  even there hypergraph version, etc.

- connectivity is homogeneous

- $P(k)$ is Poisson



- l_{max} & $\langle l \rangle \sim O(\log n)$ "small-world like"

- $C = O(1/n)$ few triangles (locally "<u>tree-like</u>")

- LCC $\propto n$ if $\langle k \rangle > 1$

⑥ mean degree $c = \langle k \rangle = \mathbb{E}[k]$.

for general p, $\mathbb{P}(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}$

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m \, \mathbb{P}(m)$$

$$\vdots$$

$$= \binom{n}{2} \cdot p$$

$$\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} \mathbb{P}(m) \quad \cdots \quad = \frac{2}{n} \binom{n}{2} \cdot p = \boxed{(n-1)\, p}$$

recall $\langle k \rangle = 2m/n$

Degree distribution

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$\log\left((1-p)^{n-1-k}\right)$$

$$= (n-1-k)\log\left(1-\frac{c}{n-1}\right)$$

$$\cong (n-1-k)\frac{-c}{n-1}$$

$$\cong -c$$

$\log(1+x) \cong x$ if $x \ll 1$

$\binom{n-1}{k} \approx \frac{(n-1)^k}{k!}$

$$P(k) \cong \binom{n-1}{k} p^k e^{-c}$$

$\cancel{\cong \frac{(n-1)^k}{k!}}$

— What values of p are realistic?

— fact: most real-world are **sparse**

— implies that we need to assume a small p!

$$\frac{(n-1)!}{k!\,(n-k-1)!} \quad \text{if } n \gg k$$

$$\cong \frac{(n-1)^k}{k!}$$

$$p = \frac{\langle k \rangle}{n-1}$$

$n! \approx n^n$

$$\cong \frac{(n-1)^k}{k!} p^k e^{-\langle k \rangle}$$

$$\cong \frac{(n-1)^k}{k!}\left(\frac{c}{n-1}\right)^k e^{-c} = \frac{c^k}{k!} e^{-c}$$

Poisson

Lecture 3-2 Feb 1, 2024.

Prop. of $G(n,p)$

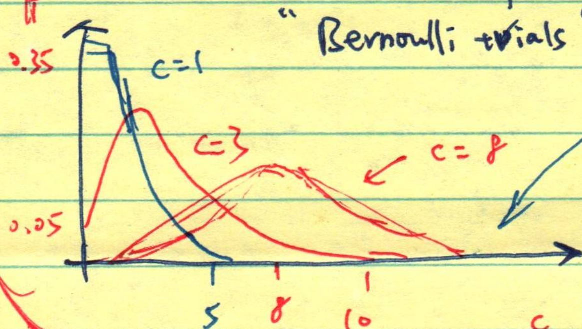
- trad. simple graph
- connectivity is homogeneous
- $\mathbb{P}(k)$ is Poisson
- l_{max} & $\langle l \rangle \sim O(\log n)$ "small-world like"
- $c = O(\frac{1}{n})$ few triangles
- LCC $\propto n$ if $\underline{c>1}$

"mean-field"

$$= \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$\mathbb{P}(k) = f(\langle k \rangle) \qquad p = \frac{\langle k \rangle}{n-1}$$

"Bernoulli trials"



is it reasonable?

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

§ Reciprocity

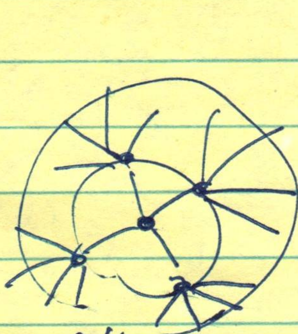
$$r = \frac{\#\ \text{mutual}}{\#\ \text{one-way}} = \frac{(n^2-n)\,p^2}{(n^2-n)\,p} = p = \frac{c}{n-1}$$

say this denominator first. (introduce)

§ clustering coefficient

$$C = \frac{\#\ \triangle}{\#\ \angle} = \frac{\binom{n}{3} p^3}{\binom{n}{3} p^2} = p = \frac{c}{n-1}$$

how about higher motifs?



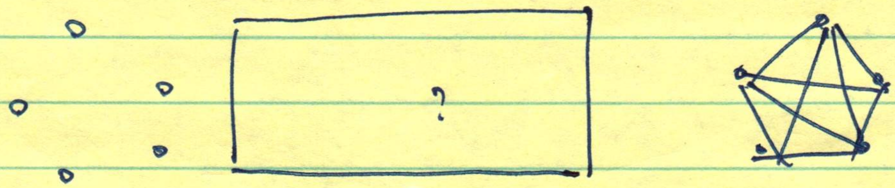
GIST:

① $r = O(1/n)$; $c = O(1/n)$ "density of motifs" $\to 0$ if $n \to \infty$

② "locally tree-like"

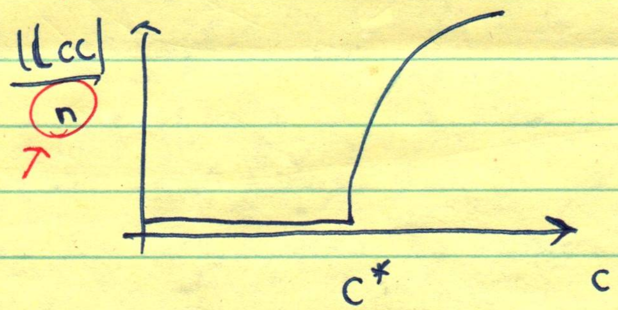
§ Can you repeat these calculations for r_i on C_i ?

§ Surprise #1 : a phase transition in connectedness.



$c = 0 \longrightarrow c = 1$

the giant comp. (LCC = GC)



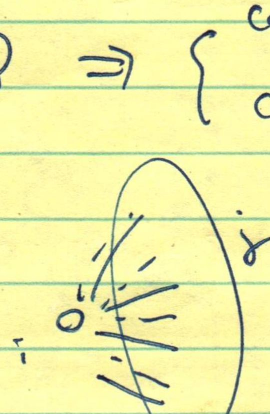
§§ Let u be the average fraction of nodes $G(n, p)$ not in the GC.

When is a node i not in the GC ? via j $(1-p)$

~~if j ∈ GC~~

Given j (not necessarily in GC)

(i) — (j) $\Rightarrow$ { case 1 : i is not connected to j
case 2 : i is connected to j but $j \notin GC$. (pu)



③

$$u = \left(1 - p + pu\right)^{n-1}$$

$$\lim_{n\to\infty}\left(1 - \frac{x}{n}\right)^n = \exp(-x)$$

$$u \cong \exp\left(-c(1-u)\right) \quad \text{when } n \to \infty$$

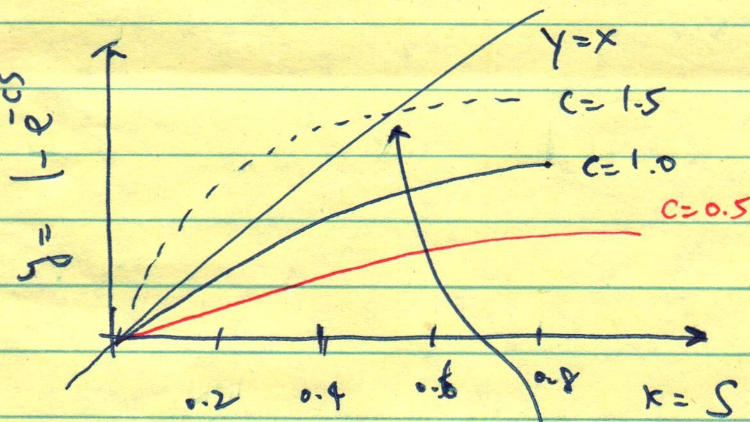
Let $S = 1 - u$ (avg. frac. in GC).

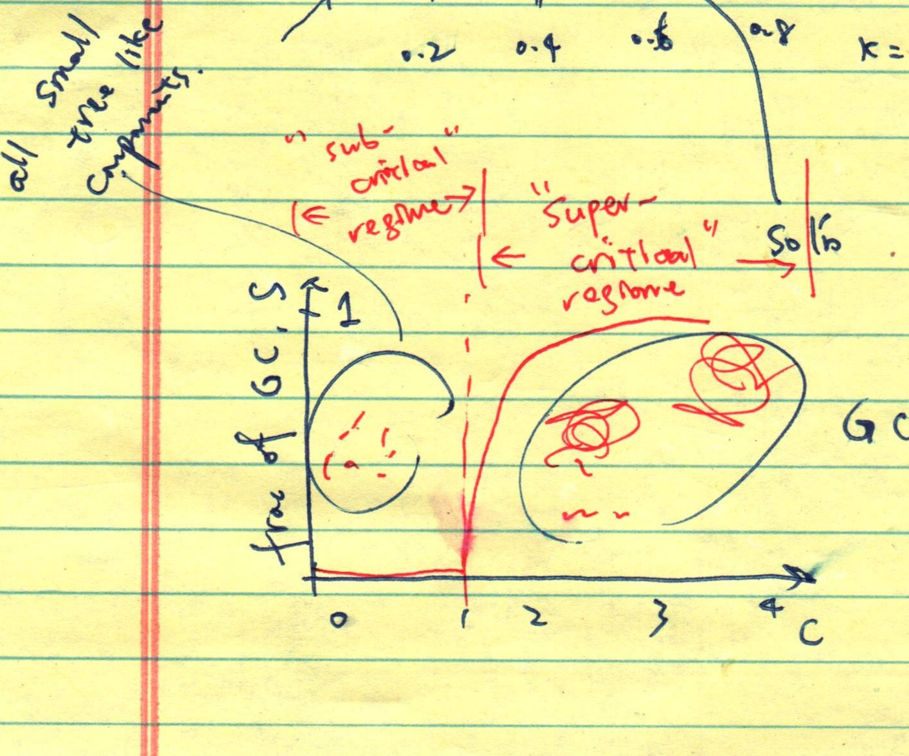
prob. $i \in GC$

$$\boxed{S = 1 - \exp(-cS)}$$

transcendental equation

(i.e., no simple closed-form exists)





GC + small-tree like components.

regime → phase

phase transition

critical point $c = 1$

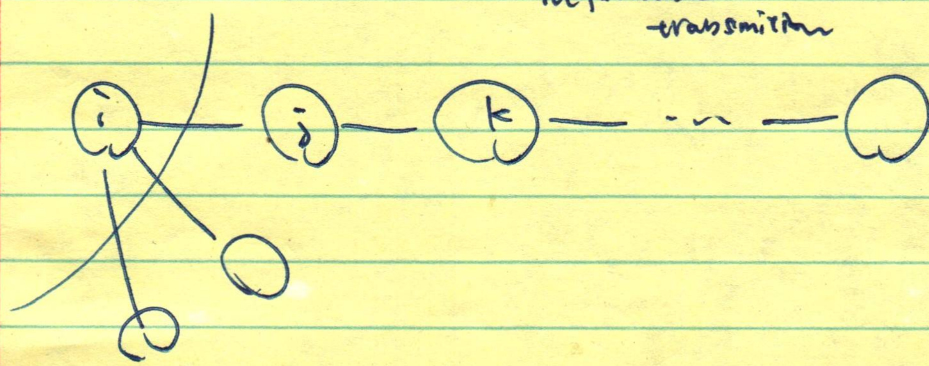
④

longest geodesic paths are short. ⇒ this "explains" a pattern in real-world networks.

∴ random model

} Surprise 2. : "Small world # and shortest paths"

branching process infections/ information transmition key "iid"



c

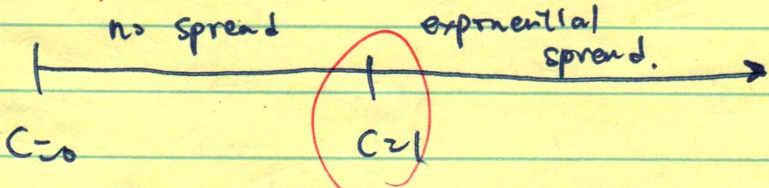
$c \quad c^2 \quad c^3 \quad \cdots \quad c^{l_{max}}$

nodes in a tree grows like $O(c^l)$ → $n = c^{l_{max}}$

$\Rightarrow l_{max} = O(\log n)$

} i.e., short paths.

percolation / spreading processes



(all fractured trees) (GC size ~ $\Theta(n)$, length of chain ~ $O(\log(n))$)

fluctuation dominates "~~scale free~~" power law size-distribution

(5)

Key Q's now:

1. How much of observed pattern is generated by degrees alone?

simpler version:

2. How much observed pattern generated by edge density alone?
(mean degree)

Let's now see the "notes"