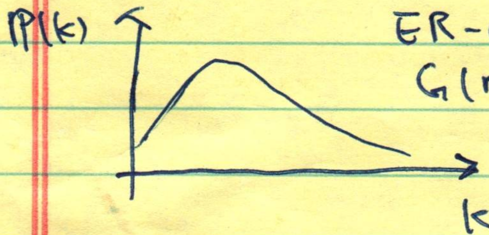
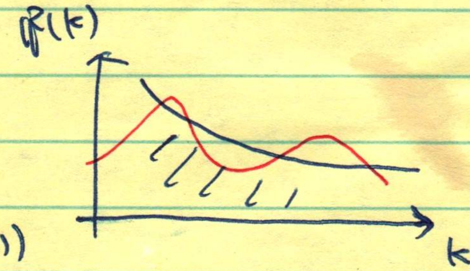
we still follow normal order, i.e., ① → ② → ③ ···
<u>not</u> ① → (*1) → (*2) → ② → ③

Lecture 4-1 Feb 6, 2024.

§ Random Graphs — w/ heterogenous degrees.



ER-Graph
$G(n, p) \Longrightarrow$



$G(n, P(k))$

the configuration model $G(n, \{k\})$

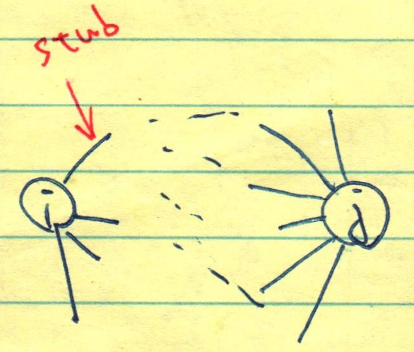
} specified degrees
$(k_1, k_2, \cdots, k_n)$

See (*)

$$\forall i<j \quad A_{ij} = A_{ji} = \begin{cases} 1 & \text{w/ prob} \propto k_i k_j \\ 0 & \text{otherwise} \end{cases}$$

Ask: Can the empirical value of some [illegible] measure $x = f(G)$ be explained by the observed deg structure ?
(and randomness)

move down



k_i k_j

$$\forall i<j \quad A_{ij} = A_{ji} = \begin{cases} 1 & \text{w/ prob } k_i k_j / 2m \\ 0 & \end{cases}$$

Two ways to choose $\{k_i\}$

① from prob distri, ~~ki~~ like $\mathrm{Exp}(\lambda)$, lognormal(μ,σ), Poisson(λ), Power-law(α)

- math fun

$\langle k\rangle, \langle k^2\rangle, \langle k^3\rangle$
generating functions.

② From an empirical network.

- inference prob!

Big Picture

4 graph spaces

↓ draw

(8 graph spaces)

(Configuration* / Molloy-Reed* → computation.

Chung-Lu) → graph simplification math.

③

Generating a config. model graph:

Two constraints ① which graph space

② Is it graphical $\ulcorner$ HH (Havel-Hakimi)

G^*

$G(n, \{k\})$

via

G^*

$\downarrow$

G_1

"proposal" picked from $\downarrow$ "markov chain"

G_2

proposal distribution

$\downarrow$

G_3

$\vdots$

$\downarrow$

G_n

$$\sum_G f(G)\,\mathbb{P}(G \mid \{k\})$$

$$\parallel$$

$$\sum_G f(G)\cdot\mathbb{P}(G)$$

$$\parallel$$

so, with $f(G^*)$ we can now compute $\langle f(G) \rangle_{G \in \{G_i\}}$.

Use: the Fosdick et al meme.

Procedure:

1) obtain G_0

2) $G_t \to G_{t+1}$

use double-edge swap.

$\Rightarrow$ "the double edge swap"

④

Prop. of configuration model random networks $G(n, \{k\})$

simple ft. self-loop.

- Four flavors of this model :
- connectivity is specified by $\{k\}$
- diameter : $O(\log n)$ (could be $O(\log \log n)$ depending on $\langle k^2 \rangle$)
- motif freq. $= \Theta(1/n)$, if $\{k\}$ is sparse

△ □ locally-tree-like structure

if $\langle k^2 \rangle \gg 1 \Rightarrow$ then bigger motif freq.

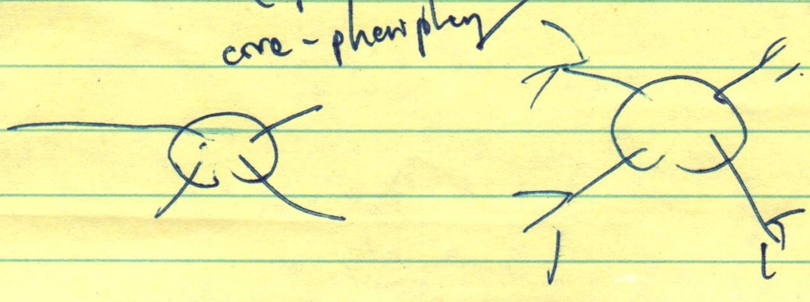
- size of LCC usually $\Theta(n)$

Question: Does $\langle k^2 \rangle$ large

$\Downarrow$

high degree nodes are "central"

CP
core-periphery



Yes!

LD-core
low-deg

($*_1$) Let's Try Something Different!

» Look at "examples" first! «

AAron's notes:

But first, introduce the notion of MCMC

see $*_2$

karate
Zachary Club:

① ~1000 samples

$P(l)$, $P(l_{max})$

geodesic distance, diameter.

② Harmonic centrality $h_i = \frac{1}{n-1} \sum_{j \neq i} l_{ij}^{-1}$.

↳ maybe not interesting ?!

Dutch high school

① $r^* = 0.74 > \langle r \rangle_{\{G\}}$

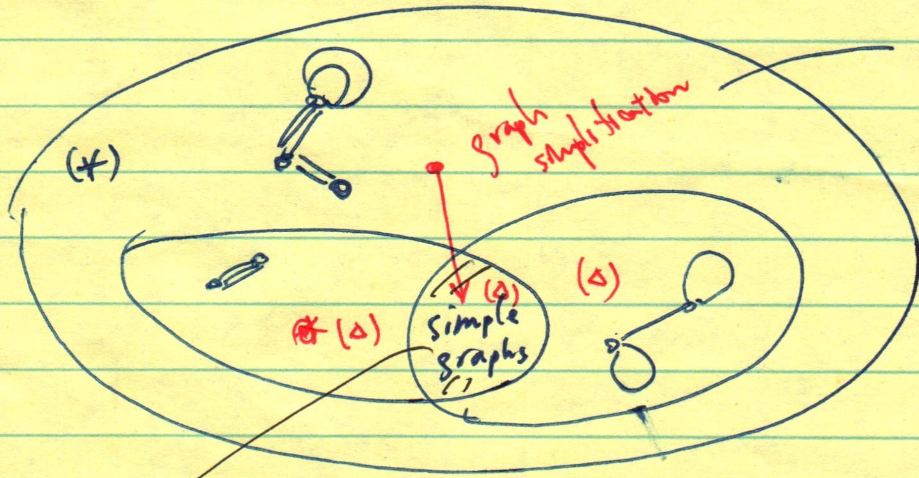
② $p(r^* = 0.74) < 10^{-4}$. hypothesis test

Next Class: Chung-Lu, Molloy-Reed, Math expressions of graph descriptors.

Lecture 4 - 2

Feb 8, 2024

Review



erase self-loops & collapse multiedges

⇓

preferentially impacts high-deg nodes.

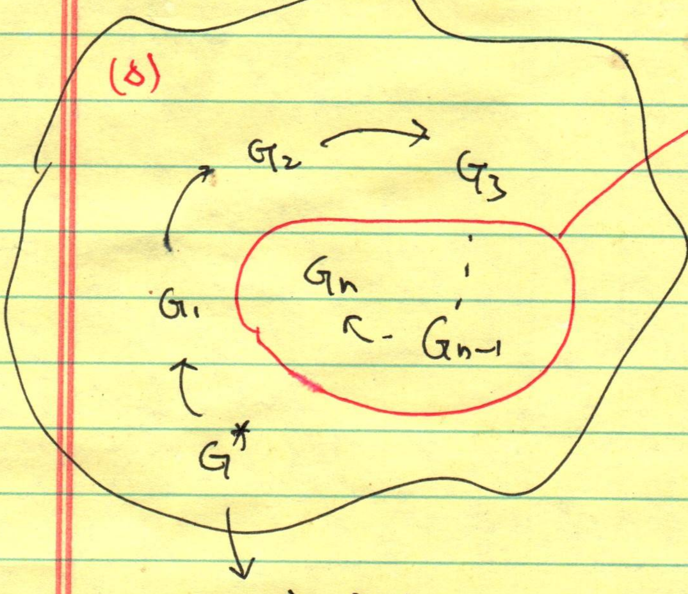
Question:

~~How to choose a~~ (Which) graph space to choose?

≫ one that matches your data.

- ① check data for self-loops / multiedges
- ② think about underlying system.

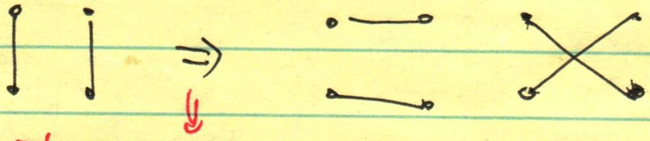
2018



config model samples

" possible "rejection" "

" sampling gap "



from
1) Data
2) degree-seq. (HH algm)

Three steps:

select → swap → accept. (or reject)

(*) Actully, depending on what we aim to randomize, we should consider 8 graph spaces.

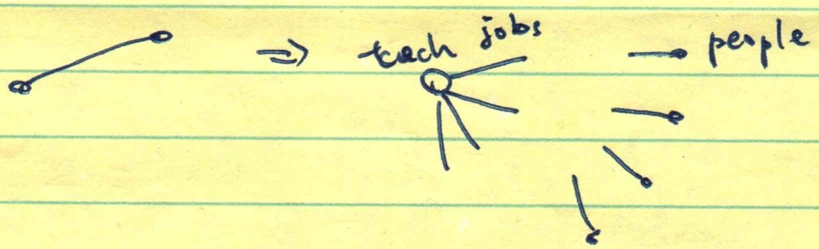
Vertex labeled

ex: me + my friends

1 arrangement

Stub labeled

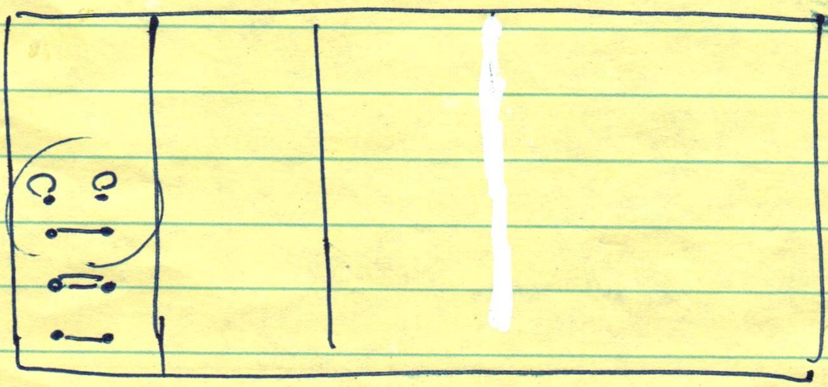
ex: people + jobs @ tech.com



4! arrangements.

* choose the graph space that matches your data !

* Next five min : ~~see slides~~ Actully, draw !



(3)

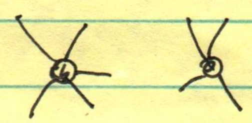
(1995)

(*) Molloy-Reed model (stub-matching algm.)

» loopy, multi- graphs / directed or undirected.

" uniform random matching on stubs "

$k_1 = 5 \quad k_2 = 4$



⇒ (Draw on-site ...)

node ID = 1 2

stubs = 1 1 1 1 1 2 2 2 2 ...

make array S

S' ← random_permutation(S)

for i = 1 to length(S)/2

E ← E ∪ { S(i, i+2) }

(+) " Counting "

$$f_{\text{loopy multi}}(G) = f_{\text{simple}}(G) \frac{1}{\prod_{i=1}^{n} w_{ii}!\, 2^{w_{ii}}} \times \frac{1}{\prod_{i<j} w_{ij}!}$$

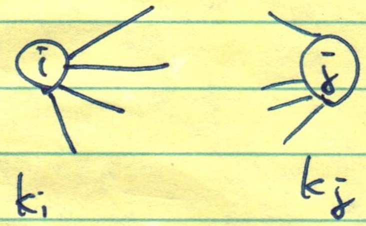
$$f_{\text{loopy}}(G) = f_{\text{simple}}(G) \frac{1}{\prod_{i=1}^{n} w_{ii}!\,(2^{w_{ii}})}$$

$$f_{\text{multi}}(G) = f_{\text{simple}}(G) \frac{1}{\prod_{i<j} w_{ij}!}$$

2002

§ Chung-Lu model (simple graphs or directed graphs)

⇒ like conf. model, but produces $\vec{k}$ only in expectation.



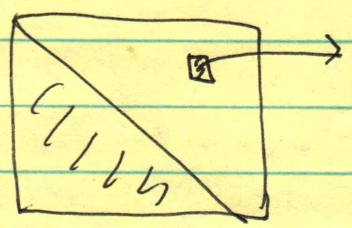
$$p_{ij} = k_i \left(\frac{k_j}{2m-k_i}\right) \approx \frac{k_i k_j}{2m}$$

$$\forall i>j \quad A_{ij} = A_{ji} = \begin{cases} 1 & \text{w/ prob } p_{ij} \\ 0 & \end{cases}$$

make $p_{ij} \le 1$

we req. $\max_i k_i^2 < 2m$

Generating Chung-Lu:



p_{ij} $\Theta(n^2)$ running time

Nuance:

iid assumption

simple graph

multigraph

actual outcome v.s. rigorous outcome.

negligible.

(5)

§ Clustering coeff.

$$C = \sum_{k_i=0}^{\infty} \sum_{k_j=0}^{\infty} q_{k_i} q_{k_j} \left(\frac{k_i k_j}{2m} \right)$$

$$= \frac{1}{n} \frac{\left(\langle k^2 \rangle - \langle k \rangle \right)^2}{\langle k \rangle^3} = O\left(\frac{1}{n} \right)$$

locally tree-like

excess degree

$$q_k = \frac{(k+1)\, p_{k+1}}{\langle k \rangle}$$

$\mathbb{P}$(neighbors deg k)

$$= \left(\frac{k}{2m} \right) \cdot n p_k$$

frac of nodes w/ deg k.

Other stuff,

see handout

(giant component, diameter, etc)

$$\langle k_{neighbor} \rangle = \sum_k k \, \mathbb{P}(\text{nei has deg } k)$$

$$= \frac{\langle k^2 \rangle}{\langle k \rangle}$$ friendship paradox!

⑥

Conclusion

Choose your $\mathbb{P}(G|\theta)$ $\nearrow$ edge density $\to G(n,p)$

$\searrow$ (degrees $\to G(n,\{k\})$)

more practical !