(0)

Lecture 2-1

(0) Highlight "Big Picture"

wk1

wk2 — network (summary) statistics

w3 — random graph models

w4

w5 — prediction

w6

w7 — ~~compu~~ community structures **

w8

parameters Θ

↓

"ensemble"

$G(\Theta)$

↑

typical #'s described by the same Θ

Different parameters

↓

diff. model

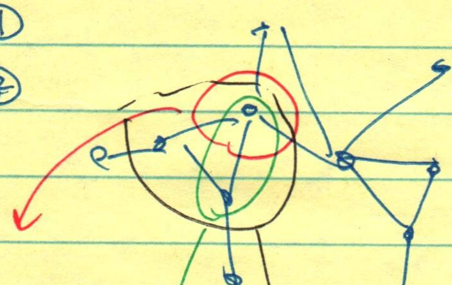
↓

diff. ensemble.

TODAY / THUR.

①

②



(connectivity) degree

(motif) reciprocity (2-clique) 3-clique

$D(i,j)$

paths!

⇐ * positional (centrality measures / k-cores

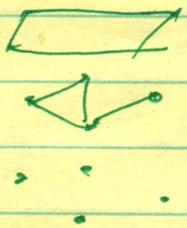
("where is a node located in a #, if it has some geometry" onion-decomposition)

①

Lecture 2-1

Jan 23, 2024

Announcements :



(1)) → (1) "We know the edges" / interaction (5352)
(2) "we want to infer the edges !
(week 13)

(1) A_{ij} computation & storage

(2) Ω / ω // O / o // Θ / θ →

(3) "clique prob"
"that sprouts computation example" > ① make sure just defined right
② take home message :
> quantum
> approx.
> heuristics

"Hard" prob's is usually either "solvable" in a practical way

↳ better pitch ?! So, don't lose hope ! $P \stackrel{?}{=} NP$.

(4) ~~Opener : Topology class in math / examples /~~ let's see ~~some examples!~~ (of importance measures)

Today [Note : Measures can be "global" / "local" | "undirected" or "directed"].

— Degree / Degree Dist

scope / graph

deg.	L	U	$k_i = \sum_{j=1} A_{ij}$
in-deg	L	D	$k_i^{in} = \sum_{j=1}^{n} A_{ji}$
out-deg	L	D	$k_i^{out} = \sum_{j=1} A_{ij}$

$$k_i^{tot} = \sum_{j=1}^{n} A_{ji} \vee A_{ij}$$

"counts"

— Intro to. edge count / arc count / mean deg / mean in- or out- degree

① - 1

weighted graph

$A_{ij} = \begin{cases} w_{ij} \in \mathbb{R} \\ 0 \end{cases}$

degree $k_i = \sum_j A_{ij}$

$s_i = \sum_{j=1}^{n} A_{ij} = \sum_{j=1}^{n} w_{ij}$ "~~streng~~ strength"

edges

$m = \frac{1}{2} \sum_i k_i = \frac{1}{2} \sum_{ij} A_{ij} = \sum_{i<j} A_{ij}$

$\sum k_i = 2m$

$\langle k \rangle = \frac{1}{n} \sum_i k_i = \frac{2m}{n}$

in ecology

simple #'s

$\rho = \frac{M}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{\langle k \rangle}{n-1}$

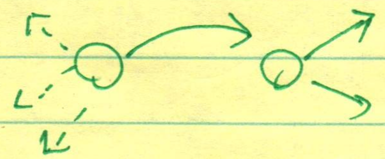
conductance / density

then go to "directed networks"

reciprocity

How often ~~does~~ is the arc "reciprocated"?

$(i \to j)$



un-

$A_{ij} = 1$

$A_{ji} = 0$

"frac. of all arcs that are reciprocated"

"Global reciprocity" $r = \frac{1}{m} \sum_{ij} A_{ij} A_{ji}$

$tr(A) = \sum_{i=1}^{n} a_{ii}$ $\quad = \frac{1}{m} \mathrm{Tr}\, A^2$ show them!

total # of arcs.

$0 \le r \le 1$

WWW $r \approx 0.57$* ~~sup~~ surprisingly high!

email/addr. book $r \approx 0.23$

High school friend $r \approx 0.3 - 0.5$

~~Samp~~ Same website page!

e.g., within CU links!

③

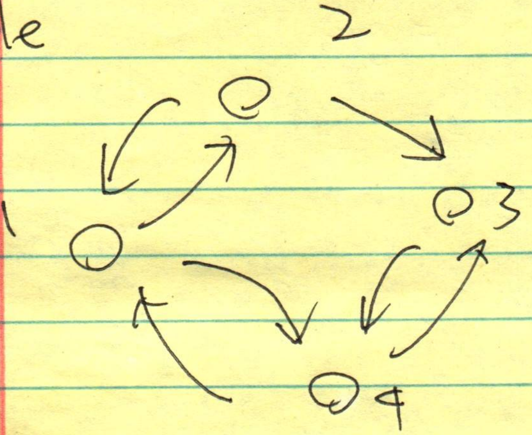
reciprocity (local version)

$$r_i = \frac{1}{k_i^{out}} \sum_j A_{ij} A_{ji} \qquad \left(\text{if } k_i = 0 \Rightarrow r_i = 0 \right)$$

$$0 \leq r_i \leq 1$$

" useful to describe a empirical # "

Example



A	1	2	3	4
1	0	1	0	1
2	1	0	1	0
3	0	0	0	1
4	1	0	1	0

$$\text{global} : r = \frac{1}{7} \left(2 + 1 + 1 + 2 \right) = \frac{6}{7} = 0.857$$

$$\text{mean local} : \langle r \rangle = \frac{1}{4} \left(\frac{2}{2} + \frac{1}{2} + \frac{1}{1} + \frac{2}{2} \right) = \frac{7}{8} = 0.875$$

4. See Clauset 2022 notes. (2.2.1)

Highlights

political blogs (2004)

$n = 1490$

$m = 19090$

*1 most real-world networks have a heavy-tailed degree distribution (why?)

$$\sigma^2 \gg \langle k \rangle$$

*2 linear-linear not good! eg here $\langle k \rangle = 25.6$, but $k_{max} = 351$

↓ *2

log-log

*3 CCDF

complementary ~~cdf~~ ~~cumu~~ cumulative dist. function

CDF $F_X(x) = P(X \leq x)$

cCDF $\bar{F}_X(x) = P(X \geq x)$

(1) papers → see Aaron's external website

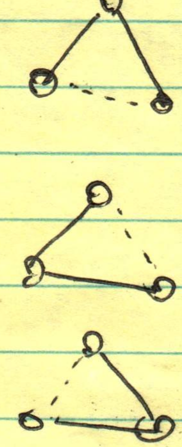
Lecture 2 / 2^{nd}-half

Jan 25, 2024

§ Clustering coeffi.

connected triple = open ~~triads~~ triads.

How often "triads" closed & from triangles ?



global $C = \#(\triangle) \times 3 \Big/ \#(\wedge)$

$0 \leq C \leq 1$, but often $C \approx 0.2$

$$= \sum_{ijk}^{n} A_{ij} A_{jk} A_{ki} \Big/ \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \neq i}^{n} A_{ij} A_{jk}$$

$$\left(\sum_{ij,k\neq i}^{n} \right) \qquad \underline{\Theta(n^3)}$$

i j k   1 2 3 4

123

124

234

134

"density of triangles" in a #

$$\rightarrow \mathrm{Tr}(A^3) \Big/ \sum_{i \neq j} A^2$$

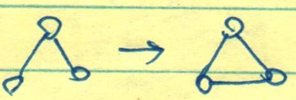
$\mathrm{diag}(A^{\ell})$ counts ℓ-cycles, $\ell \in \mathbb{Z}^+$

Highlight 1   originate from each node.

Highlight 2 : What # have C=1 (complete graph)

" have C=0 (tree, bipartite)

$*_1$ Social networks triadic closure



$\widehat{\approx} \; 10^{\times} \; \# \triangle$ for a given size n

vs. non-social #'s

2

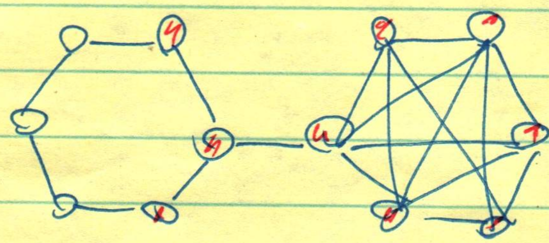
local $C_i = \dfrac{(\#\text{ pairs of nb's of } i \text{ that are connected})}{(\#\text{ pairs of nb's of } i)}$

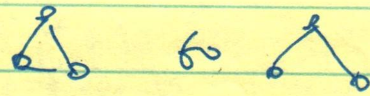
$$= \sum_{jk} A_{ij} A_{jk} A_{ki} \Big/ \binom{k_i}{2}$$

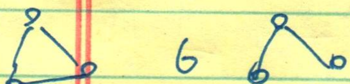
$0 \le C_i \le 1$

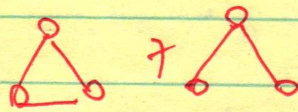
---

example



20  60

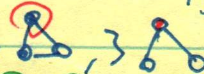
0  6

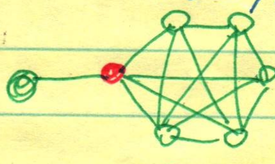
0  7

$C = \dfrac{60}{73} = 0.82$

local ex.

$C_i = \frac{0}{3} = 0.$

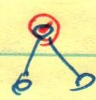
0,  3



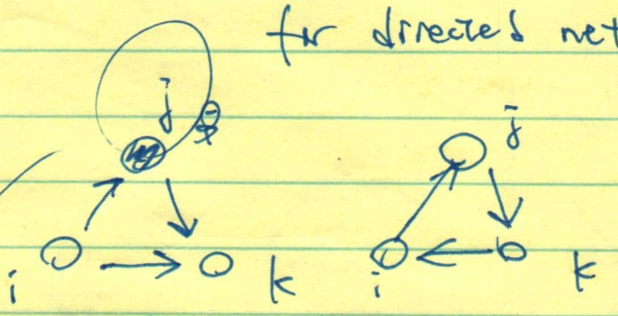
$C_j = \frac{10}{15}$

$= 0.67$

10 

15 

See notes for "feed-forward" / feed-back loops.
for directed networks.



FFL. FBL

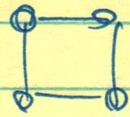
feed forward node

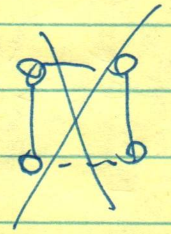
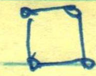
$$\# FFL = \sum_{i=1} \sum_{j} \sum_{k \neq i} A_{ij} A_{jk} A_{ik}$$

$$\# FBL = \sum_{i<j<k} A_{ij} A_{jk} A_{ki}$$

④

larger motifs.

$\boxed{n=4}$ # 

 $E\left(\#\right.$  $\left|\ \text{random graph}\right)$

$\boxed{n=5}$

positional measures "paths" n # $\underline{\Theta(n^3)}$

Floyd-Warshall algm. → All pair shortest paths (APSP) prob.

Dijkstra's algm.
"single-source shortest path,"
apply to all.

input A
output "pairwise distance matrix" L.

l_{ij} gives the length of geodesic distance. between i, j

$l_{ij} = \begin{cases} \text{the length of geodesic distance between } i, j \\ \infty \quad \text{if no path exists.} \end{cases}$

G: diameter. $= \max_{ij} l_{ij}$

L: eccentricity $\max_{j} l_{ij} = \varepsilon_i$

5

highlights.

mean geodesic path length

$$\langle \ell \rangle = \frac{1}{Z} \sum_{ij} \ell_{ij}$$

important !

"small world" property

if diameter (L) grows as $\Theta(\log n)$.

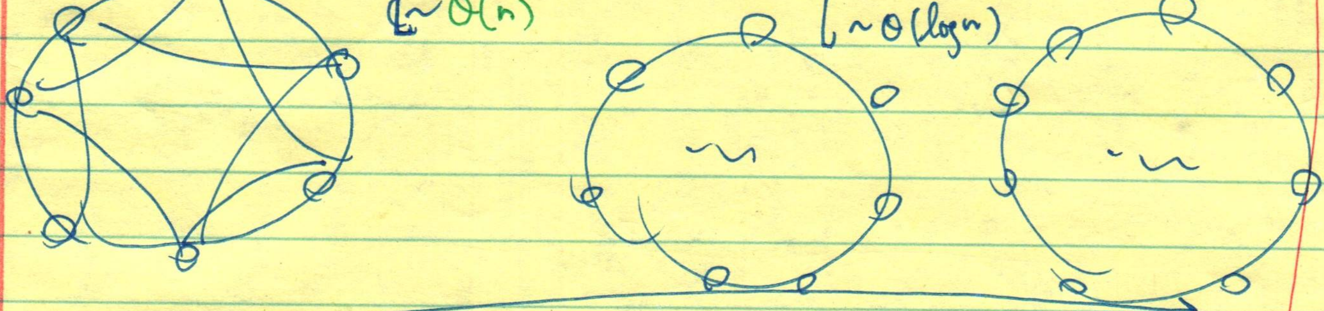
— Stanley Milgram (1933-1984)

Duncan Watts / Steve Strogatz 1998 Science

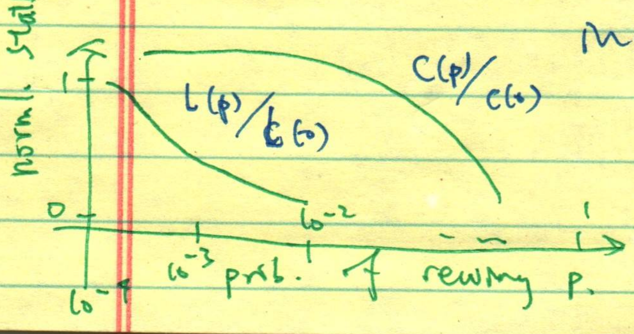
regular — High L, High C — $L \sim O(n)$

Small world — Low L, High C — $L \sim O(\log n)$

random — Low L. Low C.



increasing rand. p connectivity p



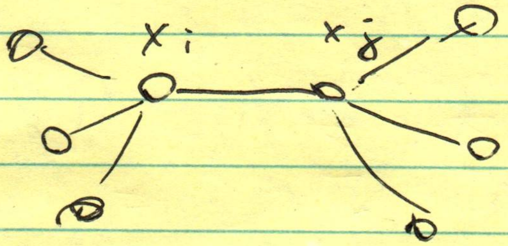
* highlights: "locally clumpy"

— "early starts" "small world"

— sampling error v.s. sensitivity of parameters

6

assortative mixing



highly assortative

Given $(i, j) \in E$, how similar are x_i, x_j ?

— assortative mixing (like links like) “social [illegible]”

homophily / age, status, beliefs

— disassortative mixing (~~[illegible]~~ like links dislikes)

examples ?

economic / ecological

producer / consumer.

How to know if the mixing pattern is typical ?

⇒ use random graph model.

(reg. though in terms of how you specify it)

— local assortative mixing. (Show 3 paper figures)