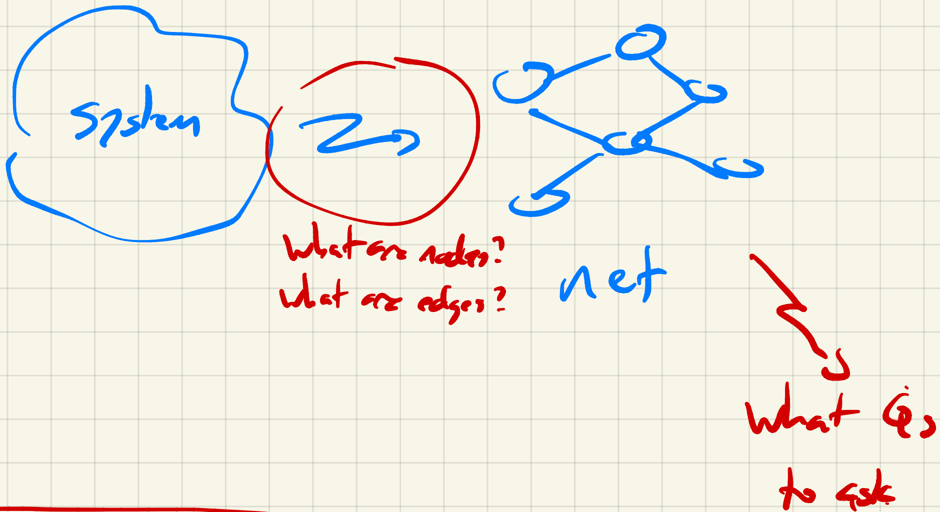
Welcome to 3352 Bio Nets

Jan 14, 2021

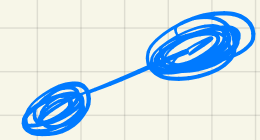
networks are a model



the two fundamental questions

1) what is a node?

2) what is an edge?



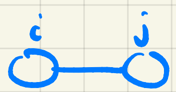
defining a network

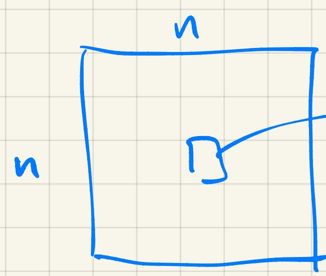
$G = (V, E)$

vertices, edges

$E \subseteq V \times V$

$(i,j) \in E$





$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

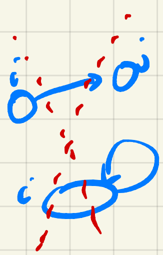
binary adjacency matrix

$\Theta(n^2)$ memory

the most basic kind of network?

"simple graph"

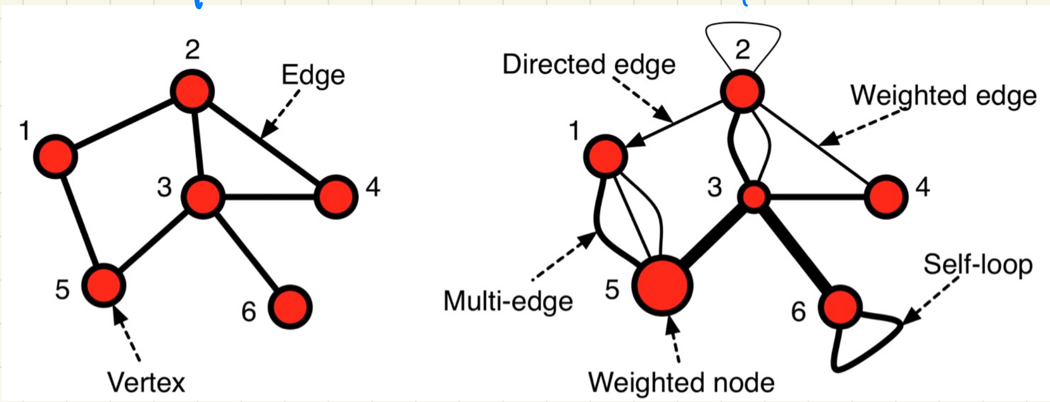
1) edges are undirected
2) " unweighted
3) no self-loops



19 Jan 2021

watch the supplemental videos

simple | non-simple network



network attributes

on the edges

on the nodes

characteristics of a network

edge	*node*	*network*
○ **unweighted**	metadata or attributes	⋆ sparse
○ weighted	locations or coordinates	⋆ dense
○ signed	state variables	◇ bipartite
• **undirected**		◇ projection
• directed		† connected
multigraph		† disconnected
timestamps		acyclic
		temporal
		multiplex
		hypergraph

see lecture notes

edges $m = \theta\binom{n}{2}$ dense

edges $m = O(n)$ sparse

most networks are here

adjacency matrix

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

$$A_{\text{simple}} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}}_{n} \; n$$

$$A_{\text{exotic}} = \begin{array}{c|c} & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 & \{1,1,2\} & 0 \\ 1 & \frac{1}{2} & \{2,1\} & 1 & 0 & 0 \\ 0 & \{1,2\} & 0 & 2 & 3 & 3 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ \{1,2,1\} & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 2 \end{pmatrix} \end{array} \; n \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

node metadata

n

$\Theta(n^2)$ space

$A_{ij} = A_{ji}$) undirected

adjacency list

$\Theta(n+m)$ space

$A_{ij} \neq A_{ji}$ directed

$$A_{\text{simple}} = \begin{array}{lcl} [1] & \rightarrow & (2,\ 5) \\ [2] & \rightarrow & (3,\ 1,\ 4) \\ [3] & \rightarrow & (2,\ 5,\ 4,\ 6) \\ [4] & \rightarrow & (2,\ 3) \\ [5] & \rightarrow & (1,\ 3) \\ [6] & \rightarrow & (3) \end{array}$$

$$A_{\text{exotic}} = \begin{array}{lcl} [1] & \rightarrow & ((5,1),(5,1),(5,2)) \\ [2] & \rightarrow & ((1,1),(2,\tfrac{1}{2}),(3,2),(3,1),(4,1)) \\ [3] & \rightarrow & ((2,1),(2,2),(4,2),(5,3),(5,3)) \\ [4] & \rightarrow & ((2,1),(3,2)) \\ [5] & \rightarrow & ((1,1),(1,2),(1,1),(3,3)) \\ [6] & \rightarrow & ((3,3),(6,2)) \end{array}$$

edge list

each element (i,j) s.t. $i \in V, j \in V$

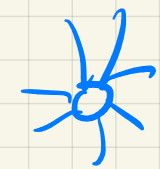
$\Theta(m)$ space

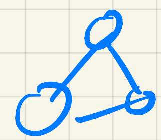
$$A_{\text{simple}} = \{(1,2),(1,5),(2,3),(2,4),(3,5),(3,6)\}$$

$$\begin{aligned} A_{\text{exotic}} = \{&(1,5,1),(1,5,1),(1,5,2),(2,1,1),(2,3,2),(2,2,1/2),(2,3,1), \\ &(2,4,1),(3,2,1),(3,2,2),(3,4,2),(3,5,3),(3,5,3),(4,2,1), \\ &(4,3,2),(5,1,1),(5,1,2),(5,1,1),(5,3,3),(6,3,3),(6,6,2)\}\ . \end{aligned}$$

Summarizing a network (statistically) → how are networks shaped?

3 types:

1) connectivity – # of connections (directly or indirectly) 

* 2) motif – specific subgraphs of network 

3) positional – where in the network does a node sit 

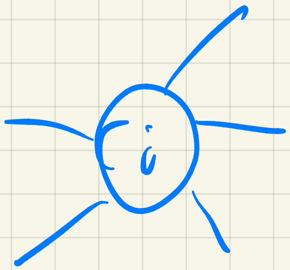
---

2 flavors

* 1) node level – calculated for some node $i \in V$

* 2) network level – calculated across entire network

degree k_i node lvl

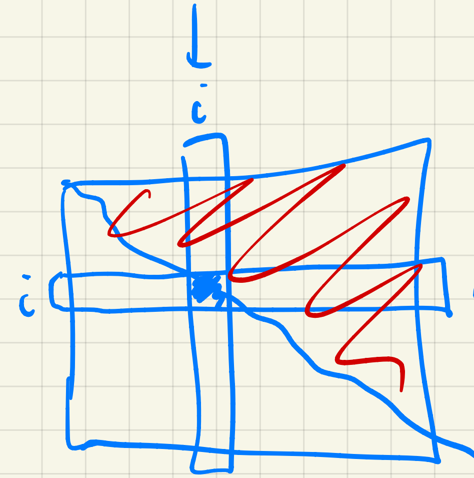


$$k_i = \sum_{j=1}^{n} A_{ij}$$

"

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

$\rightarrow$



$\sum A_{ij} = k_i$

$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ij} \right) = 2m$$

"strength" $\equiv$ weighted degree

$$m = \frac{1}{2} \sum_{i=1}^{n} k_i = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} = \sum_{i=1}^{n} \sum_{j>i} A_{ij}$$

$$\boxed{\sum_{i=1}^{n} k_i = 2m}$$

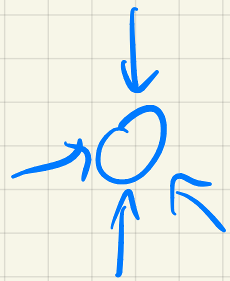
$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n}$$

mean degree

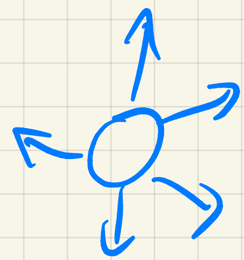
network lvl

net lvl connectance (density) $\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{\langle k \rangle}{n-1}$ (simple graph)

directed degree



in-degree k^{in}



out-degree k^{out}

nice (v)

$$\boxed{k_i^{out} = \sum_{j=1}^{n} A_{ij}}$$

$$k_i^{in} = \sum_{j=1}^{n} A_{ji}$$

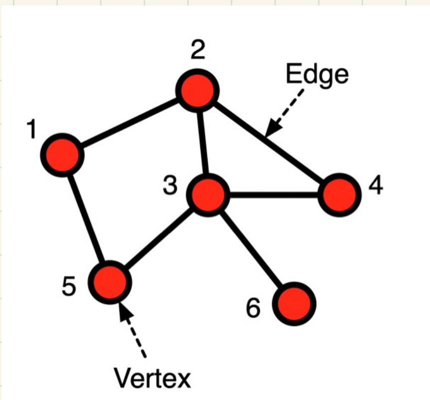
$$k_i^{tot} = \sum_{j=1}^{n} A_{ji} \vee A_{ij} \qquad [A_{ij} \text{ OR } A_{ji}]$$

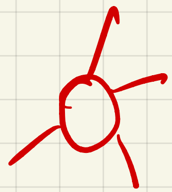
not (v)

$$\langle k^{out} \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^{out} = \frac{1}{n} \underline{\sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ij} \right)} = \frac{m}{n} = \langle k^{in} \rangle$$

21 January 2021

degree distributions





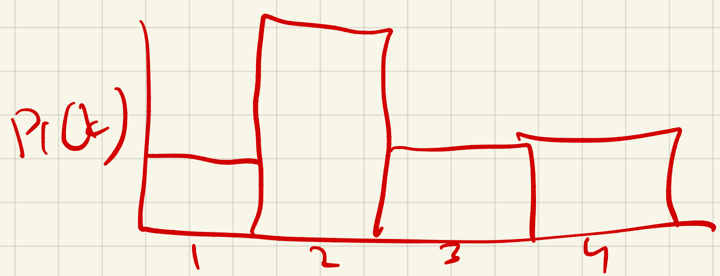
degree $\equiv$ # of connections

deg. sequence: $(2, 3, 4, 2, 2, 1)$

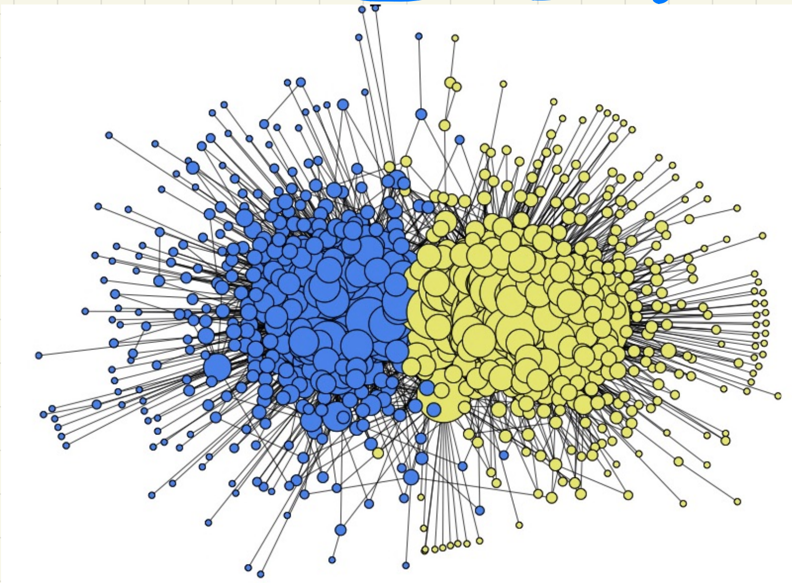
$(1, 2, 2, 2, 3, 4)$

$p(k) = P(k) = Pr(k) = Pr(K=k)$

k	1	2	3	4
$Pr(k)$	$\frac{1}{6}$	$\frac{3}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



exploring the degree distribution



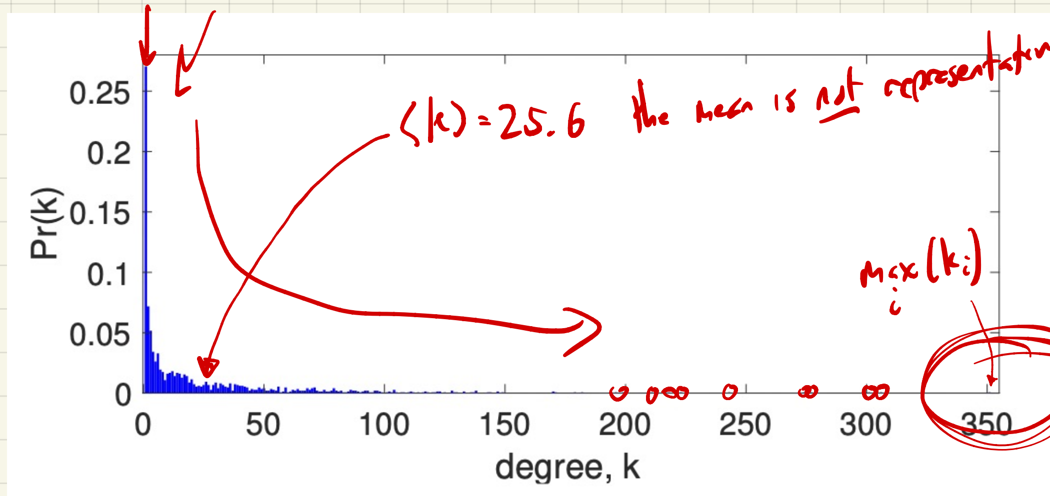
political blogs (2004)

$n = 1490$

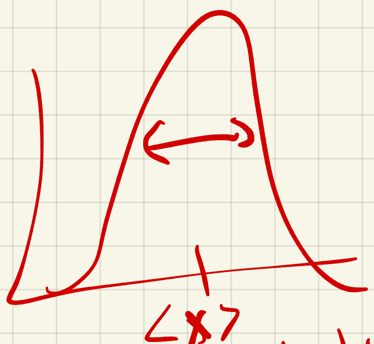
$m = 19090$

→ most real-world networks have a heavy-tailed degree distribution

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



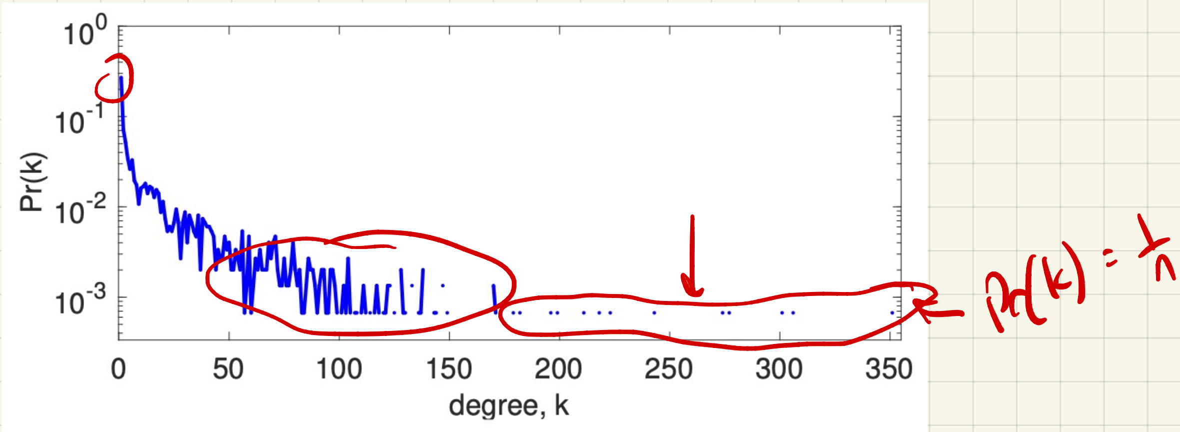
"linear-linear" plot



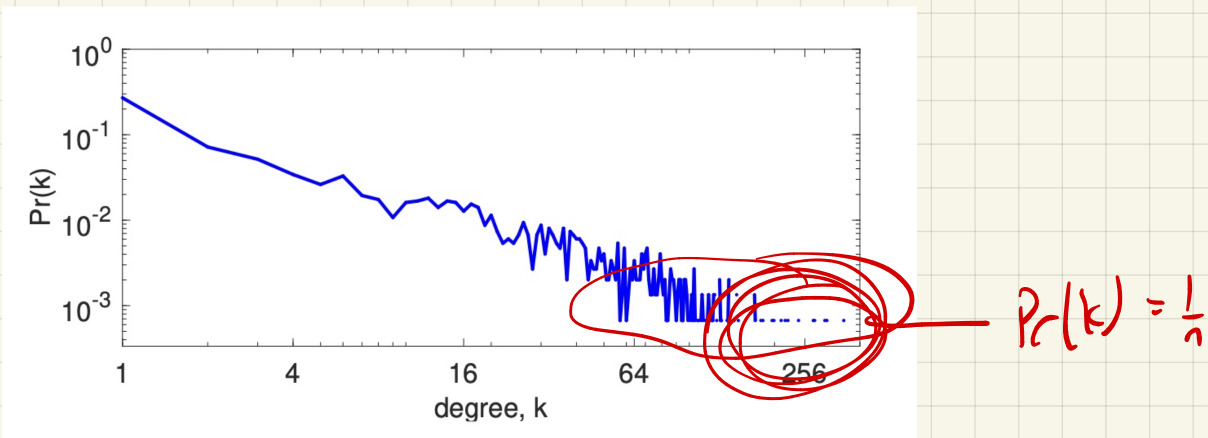
distribution of heights are like a normal
the mean ⟨x⟩ is representative

1) log-log plot

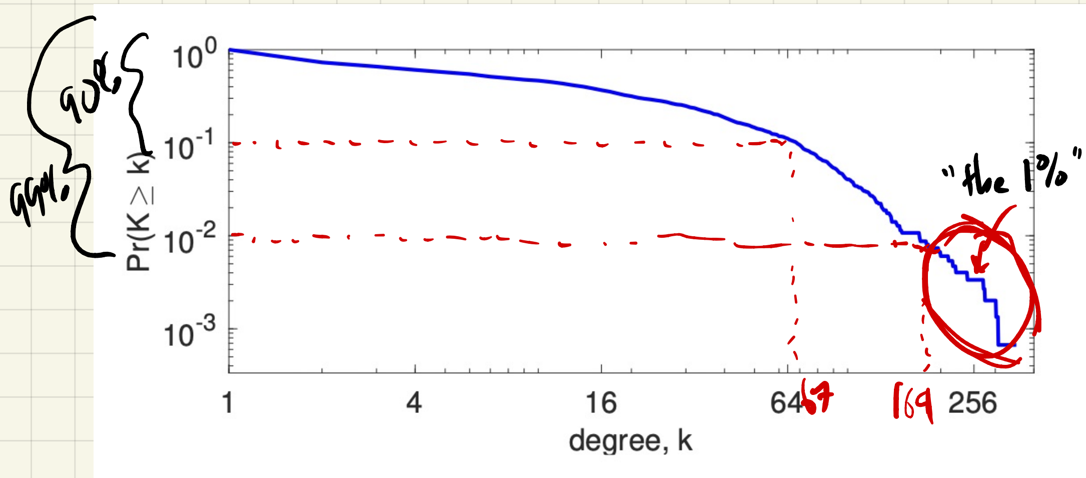
2) CCDF $\Pr(K \geq k) = 1 - \underbrace{\Pr(K < k)}_{\text{CDF}}$



"log-linear" plot



"log-log" plot of $Pr(k)$



"log-log" plot of $\Pr(K \geq k)$

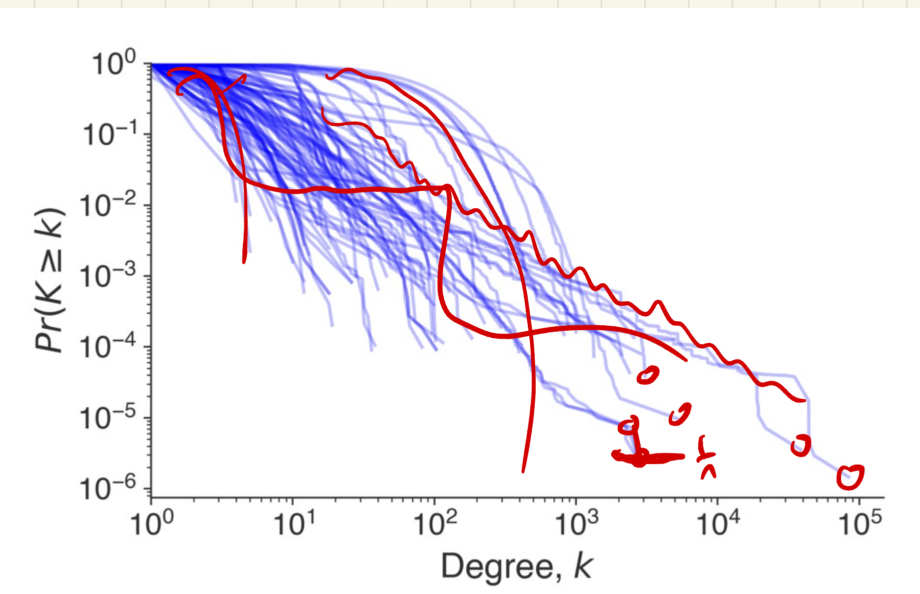
CCDF

power-law dist. $\Pr(k) \propto k^{-\alpha}$

log-normal

exponential $\Pr(k) \propto e^{-k}$

ways to model the $\Pr(k)$

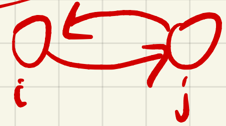


100 degree distributions

enormous diversity in degree distribution structure!

Small motifs

reciprocity

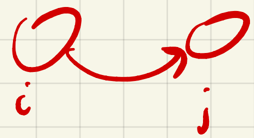


reciprocated link

$A_{ij} = 1$

$A_{ji} = 1$

$A_{ij} = 1$
$A_{ji} = 0$



unreciprocated

net. lvl.

$n=2$

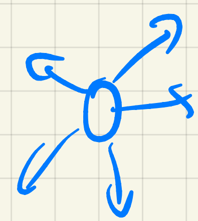
$\Theta(n^2)$ time to compute using adj. matrices

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

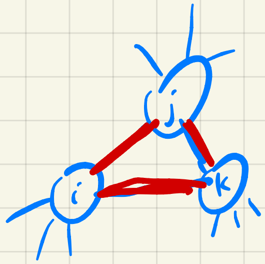
$m = \sum_{ij} A_{ij}$

node lvl

$r_i = \frac{1}{k_i^{out}} \sum_{j} A_{ij} A_{ji}$

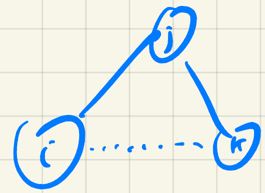


$n=3$



triangle

clustering coeff.



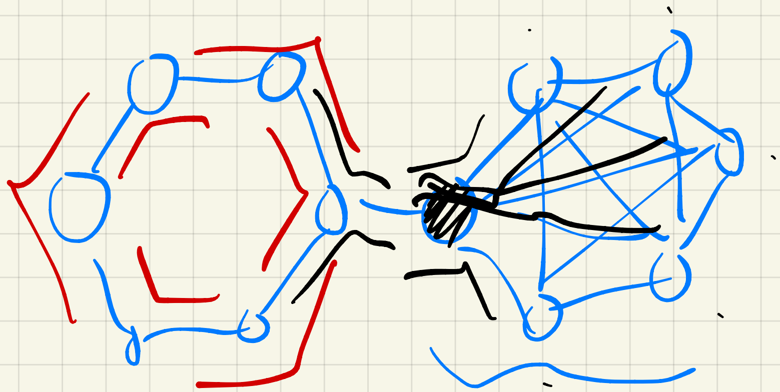
Connected triple
("wedge" or "open tr.")

$\wedge$

net lvl

$$C = \frac{\#\Delta \times 3}{\#\wedge} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} A_{ij}A_{jk}A_{ki}}{\sum_{i}\sum_{j}\sum_{k\neq i} A_{ij}A_{jk}}$$

$\Theta(n^3)$ time?



$\binom{6}{3} = 20\ \Delta$

$20\ \wedge \times 3 = 60\ \wedge$ | clique contribution

$0\ \Delta$

$6\ \wedge$ | ring contribution

$7\ \wedge$ | bridge contribution

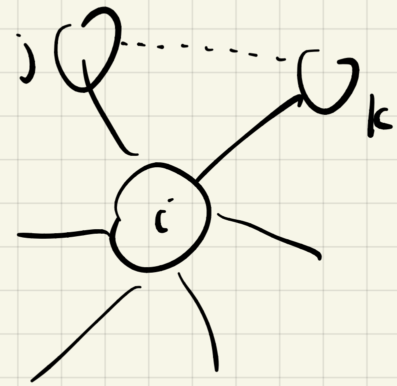
plugging in...

$$C = \frac{3 \times \#\Delta}{\#\wedge} = \frac{20 \times 3}{20 \times 3 + 6 + 7} = \frac{60}{73}$$

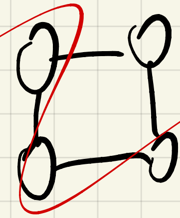
$$= \underline{0.82}$$

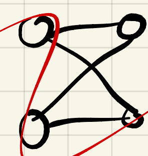
local clust. coeff. (node lvl)

$$C_i = \frac{\text{\# of pairs of neighbors of } i \text{ that are connected}}{\binom{k_i}{2}}$$



~~larger motifs~~





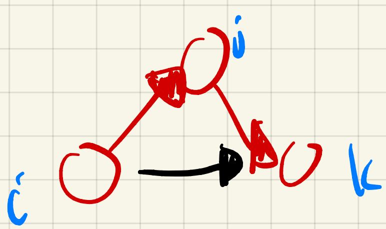
motif of interest

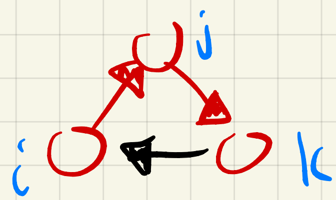
max #

feed-forward or feed-back loop

$n = 3$ directed

$\Theta(n^3)$ time





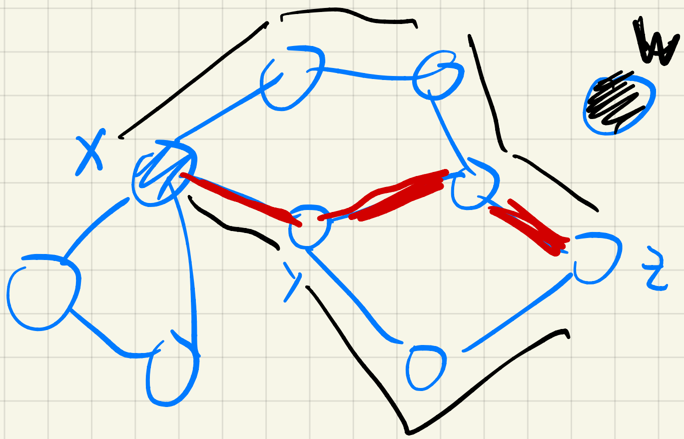
$$FFL = \sum_{i=1} \sum_{\substack{j=1 \\ j \neq i}} \sum_{k \neq i} A_{ij} A_{jk} \underline{A_{ik}}$$

$$FBL = \sum \sum \sum A_{ij} A_{jk} \underline{A_{ki}}$$

path enumeration algorithm $\longrightarrow O(n + n\langle k \rangle^2)$

geodesic paths / notions of distance

path : sequence of distinct vertices $x \to y \to \dots \to z$
s.t. for every $i \to j$ $(i,j) \in E$



geodesic ≡ shortest path

calculated using a
APSP algorithm
input : G, a graph
out : ℓ, $n \times n$ matrix of distances
ℓ_{ij} = length of geodesic path from $i \to j$

BUT $\ell_{ij} = \infty$ if j is unreachable from i

diameter : $\max_{ij} \ell_{ij}$ (ignoring $\ell_{ij} = \infty$) net lvl

eccentricity : $\varepsilon_i = \max_j \ell_{ij}$ (") node lvl

MGD $\langle \ell \rangle = \frac{1}{Z} \sum_{ij} \ell_{ij}$ (")

Z ≡ # of non-inf. paths
non-zero
net lvl

pol blogs

1) 268 components

largest comp 1222 (82%)

smallest " 1 node

2) $\langle \ell \rangle = 2.74$

$\ell_{max} = 8$

$$\frac{\text{one path}}{\underbrace{1492064 \text{ paths}}_{Z}}$$