Q: What is a random graph?
- How many $|V|, |E|$
- How are $\underbrace{\text{edges configured}}$

↳ assumed random

Big idea:

- Edges are configured randomly
- Random graphs as a whole are considered $\underbrace{\text{implicit}}$ / $\underbrace{\text{explicitly}}$,

modularity calculation & theoretical study

actually generating graphs to study

Q2: why do we care?
- Mirror properties of real networks
- Use random graphs as "null models" for hypothesis testing or otherwise

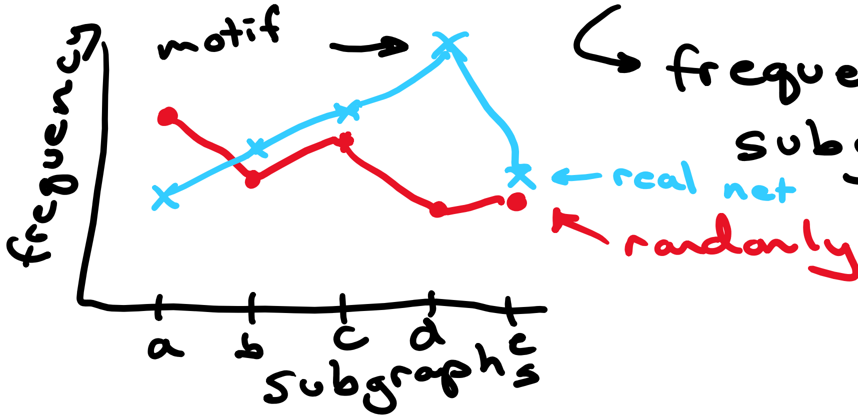
→ E.g., modularity, we compare

→ E.g., modularity, we compare our observation relative to a randomly configurated graph

$\left(\frac{d_i d_j}{2m}\right)$ ← attachment probabilities

→ E.g. 2: <u>motif</u> finding

↳ frequently appearing subgraph in same G



One last thing

↳ we can study random graphs to get a better understanding of real graphs

---

Random Graph Models

classic model: Erdös-Rényi

O.G.: $G(n,m)$ $\langle k \rangle = \frac{2m}{n}$

O. G.: $G(n, m)$ $\quad \langle k \rangle \quad \frac{\quad}{n}$

$|V|$ $\quad |E| \quad$ avg. degree

We have n vertices and m edges

→ we randomly select 2 endpoints for each edge

Issue: generates loopy multigraphs

"Newer" model: $G(n, p)$ $\quad \langle k \rangle = p(n-1)$

attachment probability

↳ we define edges buy flipping a weighted coin for all u, v vertex pairs

Note: we can generate simple graphs

Note 2: This is a Bernoulli process

↳ Hence, we end up with a binomial distribution for degrees

Our <u>degree distribution</u>:

For our E-R model

For our E-R Model

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

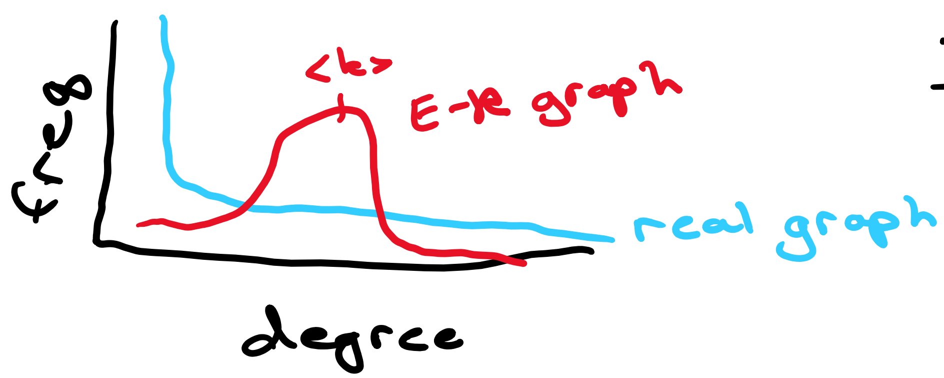
prob. of degree

as $n \to \infty$ and k is fixed

$$p(k) = e^{-\langle k \rangle} \frac{\langle k \rangle}{k!}$$

mean value

Binomial → poisson



Issue: not represenative of real ODs

Let's Analyze E-R graphs in terms of connectivity and a GIANT component

Q: What $\langle k \rangle$ for a giant component?

$\langle k \rangle = 0$ → fully disconnected

$\langle k \rangle = n-1$ → fully connected

Q2: Where do we switch from mostly disconnected → connected?

"critical point"

"critical point"

we can observe around $\langle k \rangle = 1$, a giant component quickly appear

=> This mirrors real networks ✓

AKA why we study RGs

---

what about other real-world properties:

* small-world
* low diameter
* existence of hubs
* skewed degree distributions
* Large component

Can we theoretically quantify/understand these?

- consider vertex v
- v has degree $\langle k \rangle$

- Each $u \in N(v)$ has $d(u) \approx \langle k \rangle$
- Each $w \in N(N(v))$ has $d(w) \approx \langle k \rangle$

→ 1-hop neighborhood: $\langle k \rangle$
2-hop neighborhood: $\langle k \rangle^2$
3-hop: $\langle k \rangle^3$

$$|N_d(v)| \approx \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^d$$

$$\approx \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$$

N_d: d-hop neighborhood

To get our diameter:

set $|N_d(v)| = n$, solve for d

$$\frac{\langle k \rangle^{d+1} - \cancel{1}}{\langle k \rangle \cancel{- 1}} = n$$

$$\langle k \rangle^d \approx n$$

$$d \approx \frac{\ln(n)}{\ln(\langle k \rangle)}$$

assume $n \ggg \langle k \rangle$

$$d \approx \ln(n)$$

diameter log of # vertices

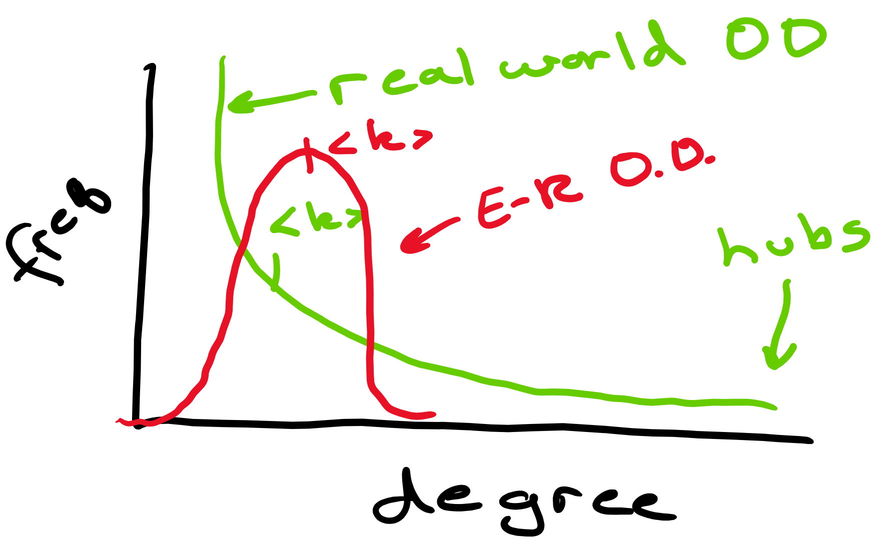
SO: our diameter grows logarithmically with n

↳ AKA small-world □

One big issue with E-R graphs

↳ degree distribution is <u>flat</u> → close to ⟨k⟩-regular

∴ there are no hubs



In reality: $\langle k \rangle \lll k_{max}$

↳ E-R doesn't capture this

↳ E-R doesn't capture this

Introducing:

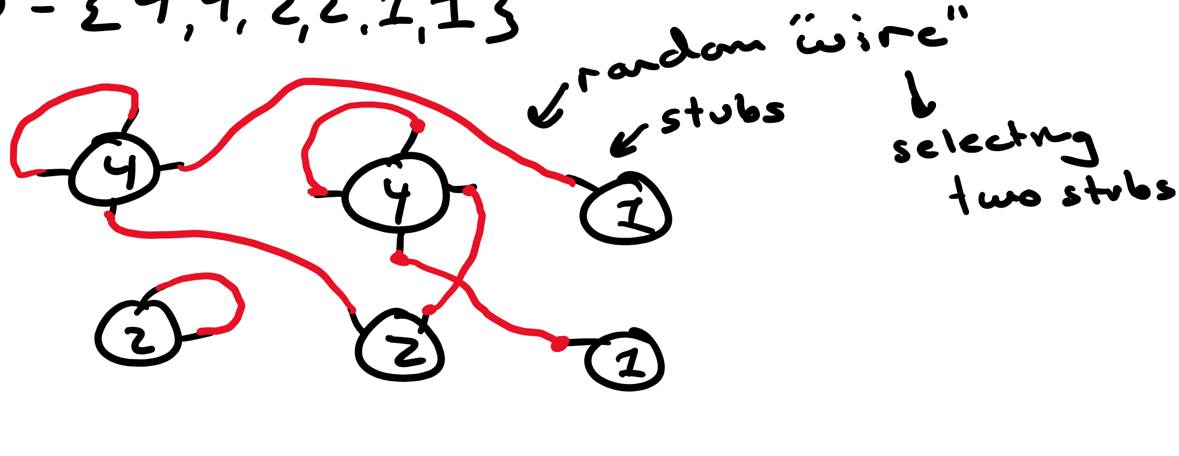
The configuration model
(/ Chung-Lu model)

↳ A random graph with an assumed degree distribution

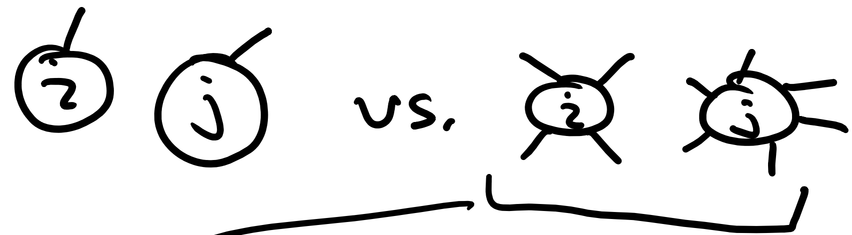
Basic Idea:

- We have n vertices each with d_i "stubs", where d_i is a value in the D.D.
- We randomly select stubs and attach them

$DD = \{4, 4, 2, 2, 1, 1\}$



What is the p_{ij} attachment probabilities

Consider 

Note: more likely to select a stub from a higher degree vertex

AKA: attachment probabilities are a function of d_i, d_j and the overall degree sum

Probability of edge (i, j)

= (prob. of selecting i's stub)

* (prob. of selecting j's stub)

* 2 [$(i,j), (j,i)$]

* m [total attempts]

$$= \frac{d_i}{2m} \frac{d_j}{\cancel{2m}} \cancel{2m}$$

$2m \quad \cancel{2m}$

$$= \boxed{\frac{d_i d_j}{2m}}$$ we've seen this with modularity

Note: A lot of the same properties of E-R graphs still hold

→ small-world, low diameter, giant component

(a function of <k> insted of D.D.)

Final Model: Chung-Lu

$G(n,m) \rightarrow$ Configuration Model

$G(n,p) \rightarrow$ Chung-Lu Model

Chung-Lu model

- Input degree distribution
- we have pairwise probabilities for attachment for all i,j pairs
- We flip a biased coin with

- We flip a biased coin with $\left(\frac{d_i d_j}{2m}\right)$ probability to create edge (i,j)

→ Constructs a simple graph

Another issue still remains:

→ No clustering (as with E-R)

Prob. of a triangle (and clustering)

$$P(\triangle) = (P_{ij})(P_{jk})(P_{ik})$$

For E-R $\Rightarrow P(\triangle) = p^3$

For C-L $= p(\triangle) = \frac{(d_i d_j d_k)^2}{(2m)^3}$

$\Rightarrow$ Triangle probabilities are very very low relative to reality

Recall: clustering coefficient is often $\frac{1}{3}$ or greater

Next class: we'll be going more detail of null models