Lecture 16 - Random Graphs

Monday, March 18, 2024 11:41 AM

Q: What is a random graph?

How many $|V|, |E|$

How are edge configured?

(attachment probability)

Big idea:

- Edges are configured randomly
- Random graphs as a whole are considered implicit/explicitly

Theoretical study of properties

actually generating graphs to study

Q2: Why do we care:

- Mirror properties of real networks
- Use RGs as "null models", for hypothesis testing or otherwise

→ E.g.: modularity, we compare our observation relative to randomly configured network our observation relative to a randomly configured network with same degree sequence

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

→ E.g. 2: motif finding

↳ a "frequently" appearing subgraph in some G

One last thing

↳ we can study random graphs to better understand real network behavior

---

Random Graph Models

Classic Model: Erdös-Rényi

Classic Model: Erdös-Rényi

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

n = $|V|$, m = $|E|$ $\qquad \langle k \rangle$ = avg. degree

we have $|V|$ vertices, and randomly connect them via $|E|$ edges

↳ selecting endpoints u, v

Issue: generate loopy multigraphs

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

p = attachment probability

↳ we flip a coin for each unique u, v pair of vertices

Note: we have a simple graph

Note 2: This is a Bernoulli process

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

<u>Degree distribution</u> = how many vertices of a given degree

Degree distribution = how many vertices of a given degree

For our E-R model:

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

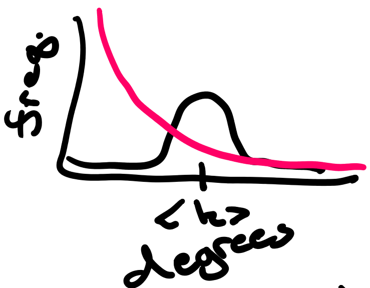
prob. of degree k

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

Binomial → Poisson

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

mean value



Why? → $n \ggg \langle k \rangle$ in real networks, and 1 fewer parameter

---

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

Where do we switch from

Where do we switch from mostly disconnected to mostly connected?

"critical point"

(see B. section 3.C)

we can theoretical observe that around $\langle k \rangle = 1$, a giant component is expected to appear

=> this mirror real networks

AKA why we study RGs

What about other real-world properties:

* Small-world
* Low diameter
* Existence of hubs
* Skewed degree distributions

Can we quantify these?

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

- consider vertex v
- v has degree $\langle k \rangle$
- each of $N(v)$ has degree $\langle k \rangle$
- each of $N(N(v))$ has deg. $\langle k \rangle$

...

$\rightarrow$ 1-hop neighborhood: $\langle k \rangle$

2-hop neighborhood: $\langle k \rangle^2$

3-hop: $\langle k \rangle^3$

$$|N_d(v)| = \langle k \rangle + \langle k \rangle^2 + \dots + \langle k \rangle^d$$

d-hop neighborhood

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

To estimate our diameter:

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

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

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

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

since $n >>> \langle k \rangle$

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

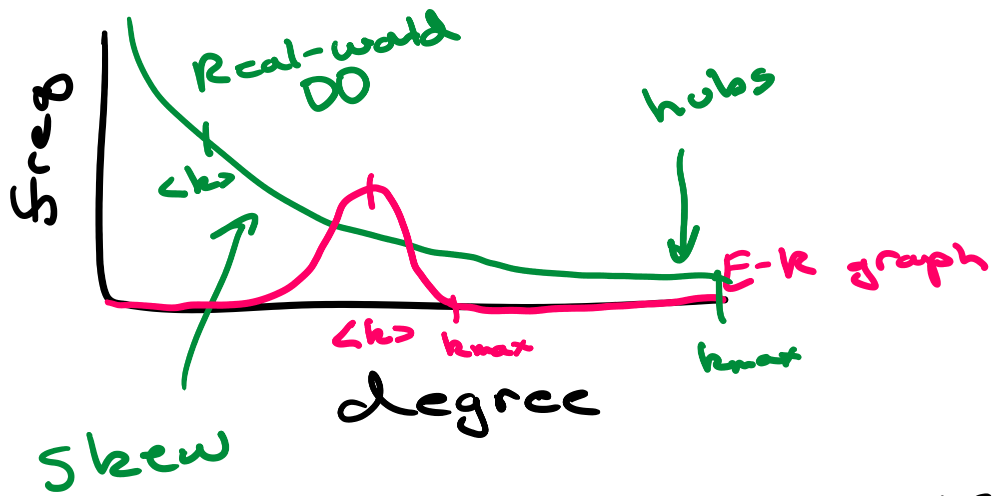
diameter — order of G

SO: our diameter grows logarith. as a fuction of n

↳ AKA a small world graph □

One Big issue with E-R graphs

↳ degree distribution and lack of hubs



In reality $\langle k \rangle <<< k_{max}$

↳ E-R doesn't capture this

Introducing:

Introducing:

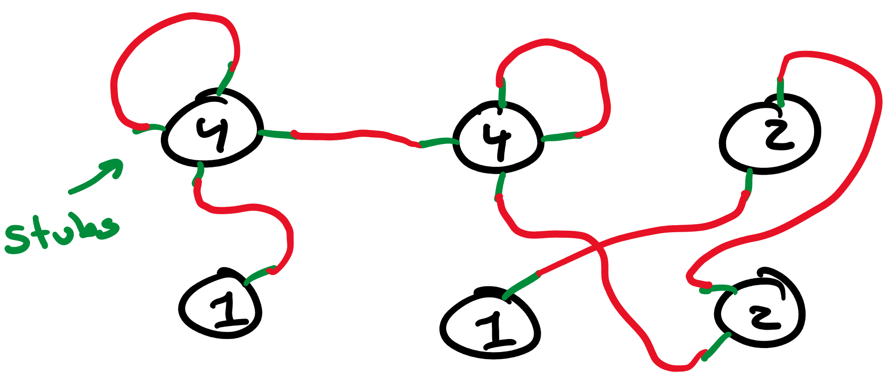
The Configuration Model

↳ A random graph with a given degree distribution

Basic Idea:

- We have n vertices each with d_i "stubs", where d_i is a value in some degree distribution
- We then randomly select stubs and attach them

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



What are the attachment probs?

Consider

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

AKA: Attachment probabilities are a function of $d(i), d(j)$ and the sum of the degrees

Prob of edge (i, j)

= (prob of selecting i's stubs)

* (prob of selecting j's stubs)

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

* m (total attempts)

$$= \left(\frac{d(i)}{2m}\right)\left(\frac{d(j)}{\cancel{2m}}\right)(\cancel{2m})$$

$$= \frac{d(i)\,d(j)}{2m}$$

as we've seen with modularity

Note: A lot of the same properties of E-R graphs hold, such as small-world, giant comp, etc.

small-world, giant comp, etc.

AND

we can also capture the skewed DDs and hubs

Issue 1: not simple graphs

Issue 2: no clustering

↳ E-R had same issue