Lecture 17 - More random graphs and null models

Thursday, March 21, 2024 11:50 AM

Review

Erdös-Renyi

$G(n,m) \to$ n verts
m edges by selecting with replacement 2 vertices

$G(n,p) \to$ n verts
$\forall x,y \in V(G) \to (x,y)$ exist with prob p

Configuration Model

$\hookrightarrow p_{u,v} = \dfrac{d(u)\,d(v)}{2m}$
(attachment probabilities)

Issues:

E-R don't have a realistic degree distribution

E-R and CM don't have realistic clustering

$G(n,m)$ and CM do not define simple graphs

Chung-Lu Model

consider $p_{u,v} = \dfrac{d(u)\,d(v)}{2m}$

consider $p_{u,v} = \frac{d(u)d(v)}{2m}$

and a Bernoulli / $G(n,p)$

process forall $x,y \in V(G)$

For $v \in V(G)$

create edges for all $u \in V(G)$

with probability $p_{u,v}$

$\to$ Chung-Lu model

Note: we won't match the exact input degree distribution

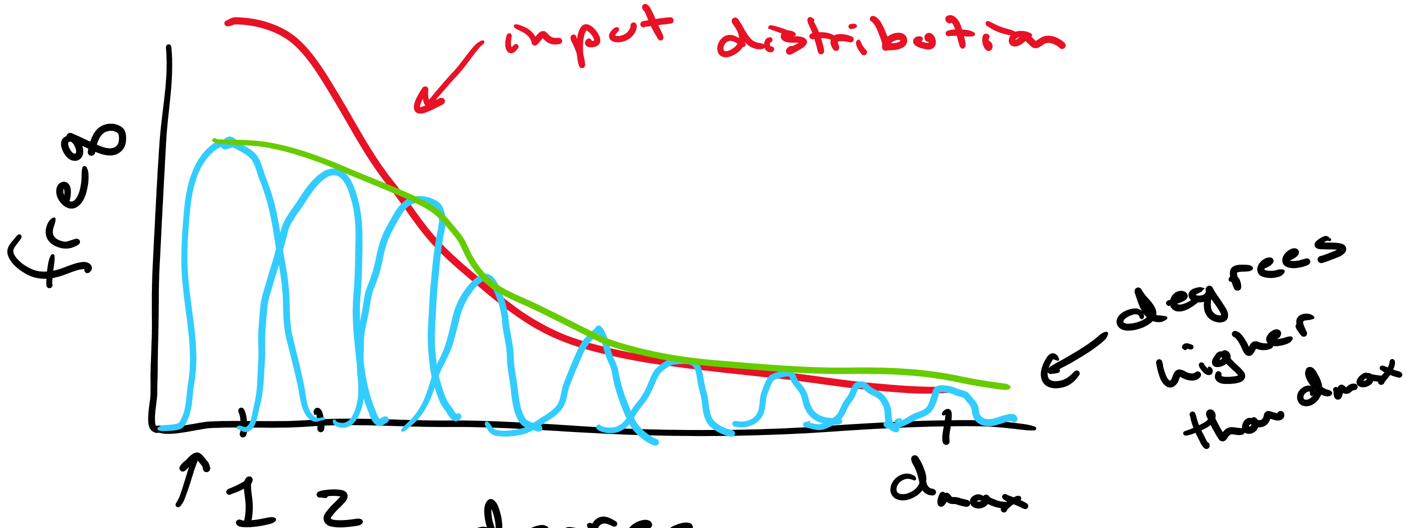
$\Rightarrow$ Theorists: but in expectation we will

(also as $n \to \infty$, the expected # of self loops and multi edges $\to 0$)

Unfortunately: the above isn't true in practice at all

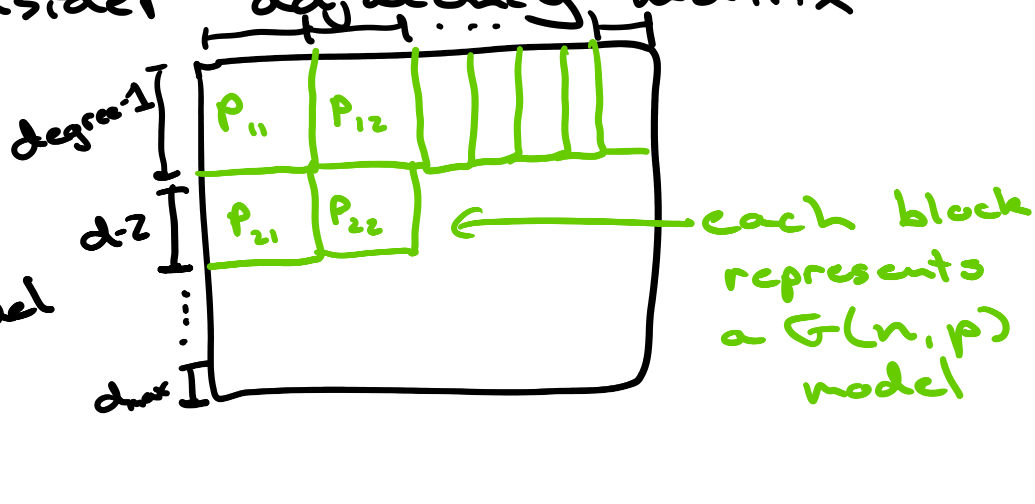
In reality: our output distribution

In reality: our output distribution is a sum of Poisson dists for all $d_i d_j$ degree pairs and vertices



consider adjacency matrix

In general: this is a block model



Takeaway from above: almost no degree distribution is actually realizable via a Chung-Lu process realizable via a Chung-Lu process

Side effect: we can't modify our probabilities or input degree distribution to exactly get to the desired output distribution

(Brissette & Slota)

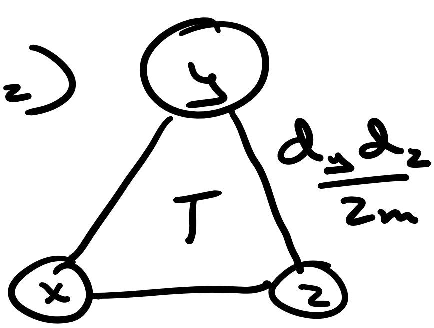
AKA Theory $\rightarrow$ practice sometimes doesn't work out

STILL: no clustering

Why: consider a triangle

$P_T = f(p_{xy}, p_{yz}, p_{xz})$

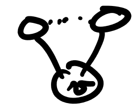
$$P_T = \frac{d_x^2 d_y^2 d_z^2}{(2m)^3}$$



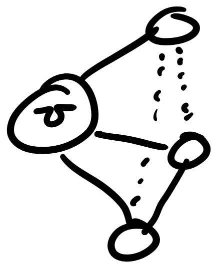
$\rightarrow 0$ as $n \rightarrow \infty$

Also consider clustering coefficient

Also consider clustering coefficient more directly



$$C_v = \frac{\text{triangle containing } v}{\text{total \# triangle that } v \text{ could be in}}$$



$C_v = \frac{1}{3}$

Consider Erdös-Rényi:

$$C_v = \frac{p\,\frac{d(v)(d(v)-1)}{2}}{\frac{d(v)(d(v)-1)}{2}}$$

take away

$$\langle k \rangle = p(n-1)$$

$$C_v = p = \frac{\langle k \rangle}{n-1}$$

generally $\langle k \rangle \lll (n-1)$

so $C_v \to$ small)

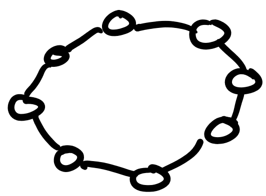
as $n \to \infty$ $C_v \to 0$ $\square$

Usually $C_v = 0.25 \leftrightarrow 0.33$ ish (observe)

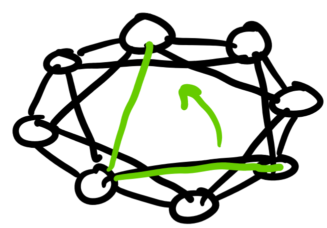
Introducing:

Watts-Strogatz model
aka "small-world" model

Basic idea



ring graph
$k=2$



$k=4$

k is the degree of vertices, where we connect to $\frac{k}{2}$ neighbors in a ring topology

WSM: also rewires each edge (select a new endpoint) with probability β to a randomly selected vertex

$\beta \to 0$ we have the original model (clustered, high diameter)

$\beta \to 1$ we result in a E-R graph

$\beta = 1$ we result in a E-R graph (no clusters, low diameter)

$0 < \beta < 1$ we have a clustered network w/a low dia.

Unfortunately:

We still have no hubs

We still have a whack degree distribution

Modeling network growth

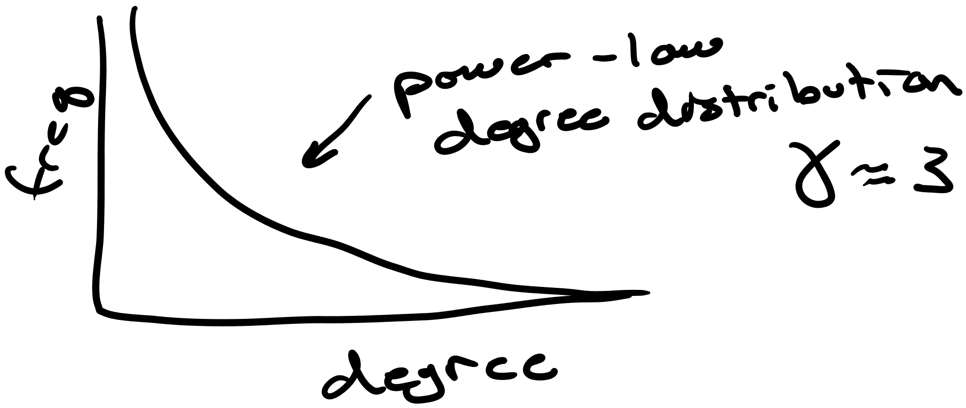
Barabasi-Albert Model:

→ we add a new vertex and attach it to existing vertices with a probability as a function of the degrees of existing verts

(aka preferential attachments)

→ we grow a graph with preferential attachment

$p_{u,v} = \dfrac{d(v)}{\sum_{i \in V(G)} d(i)}$ we add u, and attach it via the given $p_{u,v}$



But no clustering ☹

Q: Is there a model that captures all of our real-world properties

A: Yes

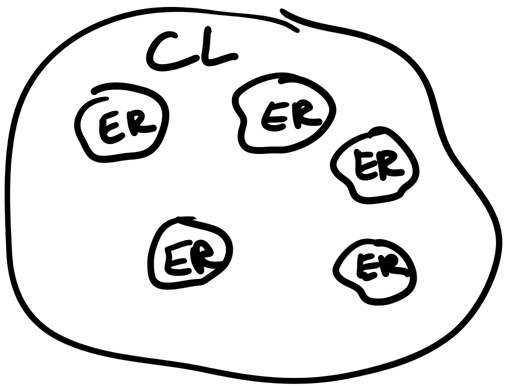
Ap2: BTER

(block two-level Erdös-Rényi)

We construct E-R blocks with high p
(communities, clustering)

We layer a C-L graph on top
(small-world, degree distribution)

(small-world, degree distribution)



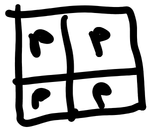
Problem: tough to analytically study

Takeaway:

complexity of model <=> complexity of analysis

Other graph models

- Defined by matrix products
  - → RMAT, Krönecker
  - → fast to generate
- Other block models



- Other benchmark graphs
  - → LFR graphs

Null Models

aka null graph models

↳ Graph that is <u>randomly configurated</u> having same set of properties

(n, m, D)

Why: hypothesis testing

- → We measure something on G
- → How does this measurement compare to random G', which has same set of G's properties?

→ null hypothesis

Is what we're observing the result of randomness or not?

What graphs can we use for this?

For something related to a degree distribution:

$\dfrac{d_u d_v}{2m}$ => Chung-Lu probabilities or configuration models

However: C-L graphs are somewhat biased and don't fit a degree distribution

C-M fit the distribution but are not simple

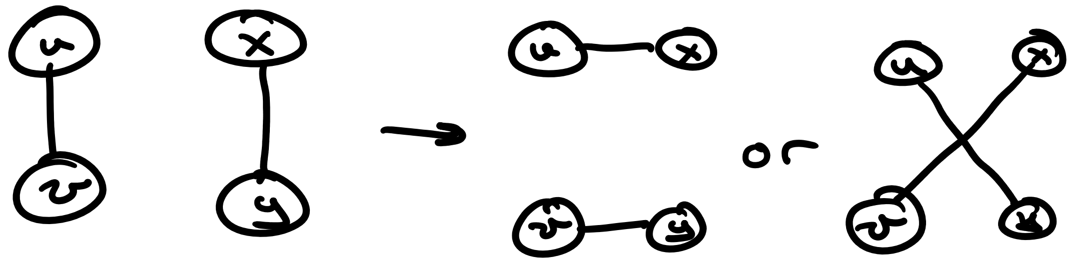
Theorists: $n \to \infty$, p_{loops} and $p_{multi} \to 0$

Reality: not quite accurate

So: How can we actually

So: How can we actually generate simple null model with a given deg. dist?

A: we can take any simple graph with deg. dist. (via H-H) and do <u>double-edge-swaps</u>



Randomly select edge pairs and double-edge-swap and repeat:

→ We get a Markov process that can <u>eventually</u> realize any simple graph w/ deg. dist.

Mixing time: unknown (maybe a lot)

One final thing:

a null model graph needs to be <u>uniformly randomly</u> sampled from the entire topolgical space

(pretty hard to do)