Lecture 15 - Random graphs and null models

Sunday, March 20, 2022 2:02 PM

Review:

$G(n,m)$
$G(n,p)$ model

Configuration model

↳ attachment probability

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

Chung-Lu model

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

and a Bernoulli process for all vertex pairs

For $v \in V(G)$:
  create edges for all $u \in V(G)$
  with prob. $p_{u,v}$

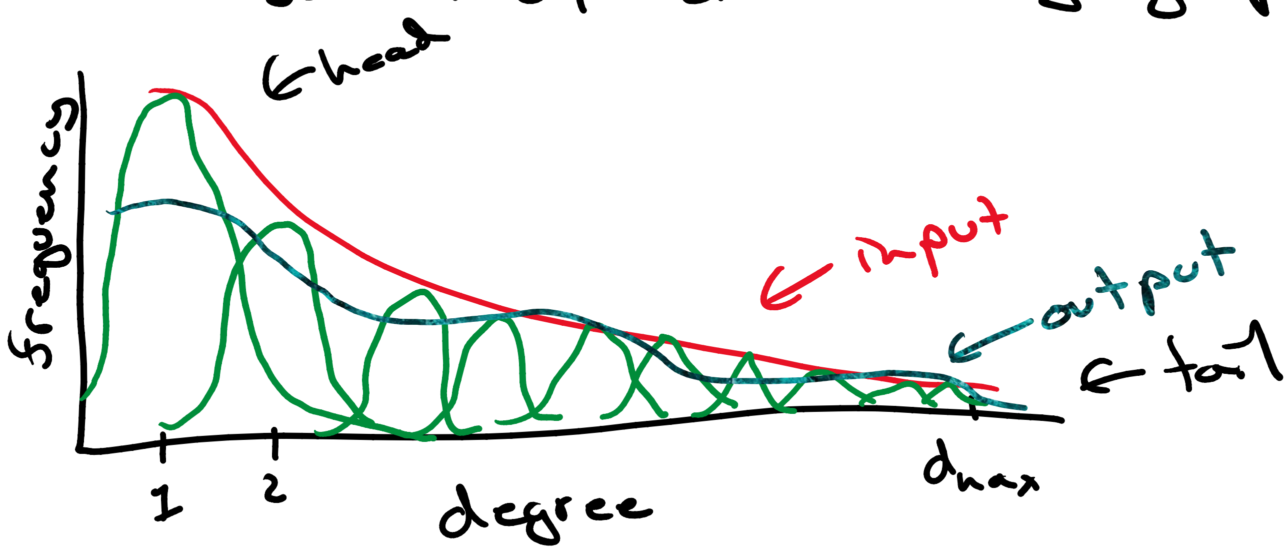
→ Chung-Lu model

Note: we won't match the exact input distribution

=> but in expectation we will (not really)

In reality: our output distribution is a sum of Poisson distributions

↳ we're really just layering a bunch of Erdös-Renyi graph



In effect: almost no input degree distribution can be generated

Effect p2: we also can't modify our

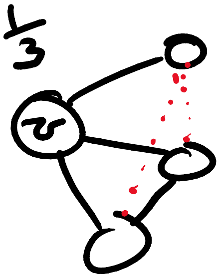
Effect p2: we also can't modify our input distribution to get same desired output distribution

However $\rightarrow$ no clustering

Clustering

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

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



Consider Erdös-Rényi

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

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

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

as $\langle k \rangle \ll (n-1)$

$C_v \to$ small

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

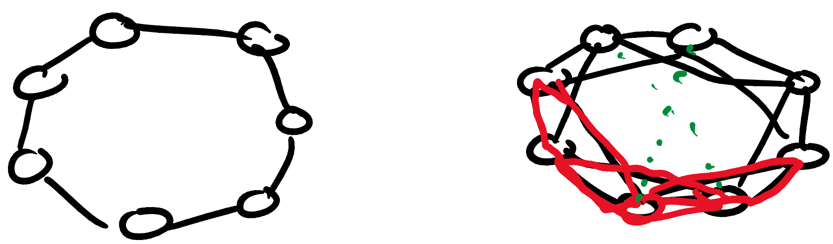
Usually $C_{ER} \approx 0.25 \leftrightarrow 0.33$

Introducing:

Watts - strogatz model

→ small - world model

Basic idea:



ring graph
$k=2$

$k=4$

→ k refers to degree, where we connect to $\frac{k}{2}$ neighbors on each side

WSM: also rewires each edge with some probability β

$\beta \to 0$, we have a tightly clustered graph w/ high diameter

$\beta \to 1$, we approach an Erdös-Renyi

Unfortunately:

Unfortunately:

- No hubs
- No degree distribution in general

Note: we haven't modeled any growth process

Barabasi-Albert model:

→ we add a new vertex, and attach it to existing vertices

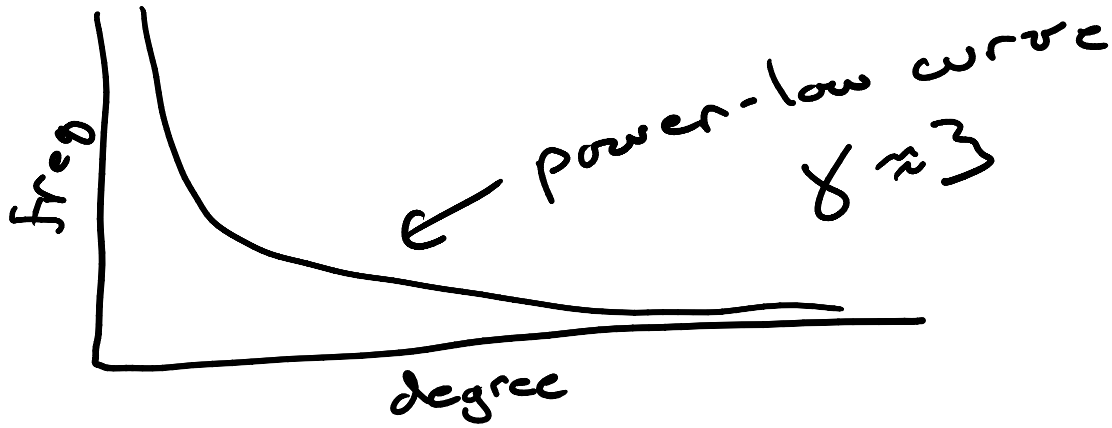
→ attachment probability is proportional to current degrees

→ constructs a graph that grows via preferential attachment

aka "rich get richer"

attachment for new vertex v to existing vertex u

$$P_{v,u} = \frac{d(u)}{\sum_{i \in V(G)} d(i)}$$



↳ aka power-law or scale-free graphs

But: no clustering :c

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

A: YES

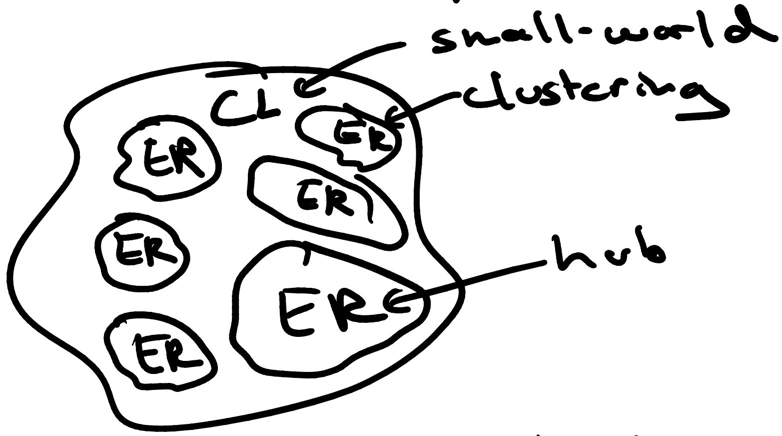
Ap2: BTER

Block two-level Erdös-Renyi

Basic idea

→ we construct dense Erdös-Renyi "blocks" aka communities

→ we layer a Chung-Lu graph over top of all blocks over top of all blocks

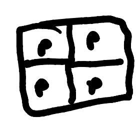


Problem: tough to analytically study

Takeaway: trading complexity/accuracy of each model for complexity of analysis

---

Other models:

- Defined by matrix products
  - → RMAT, Kröneckeer
  - → fast to generate
- Other block models 
- Other benchmark graphs
  - LFR graphs

Null models

Null models

aka null graph models

Graphs that are randomly configured and have some fixed property (n, m, D)

Why: hypothesis testing

We measure something on G

How's it compare to a random graph with G's degree distribution?

→ null hypothesis

What graph models can we use for this?

If we want to theoretically study something with regards to degree distribution

→ we can use Chung-Lu probabilities

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

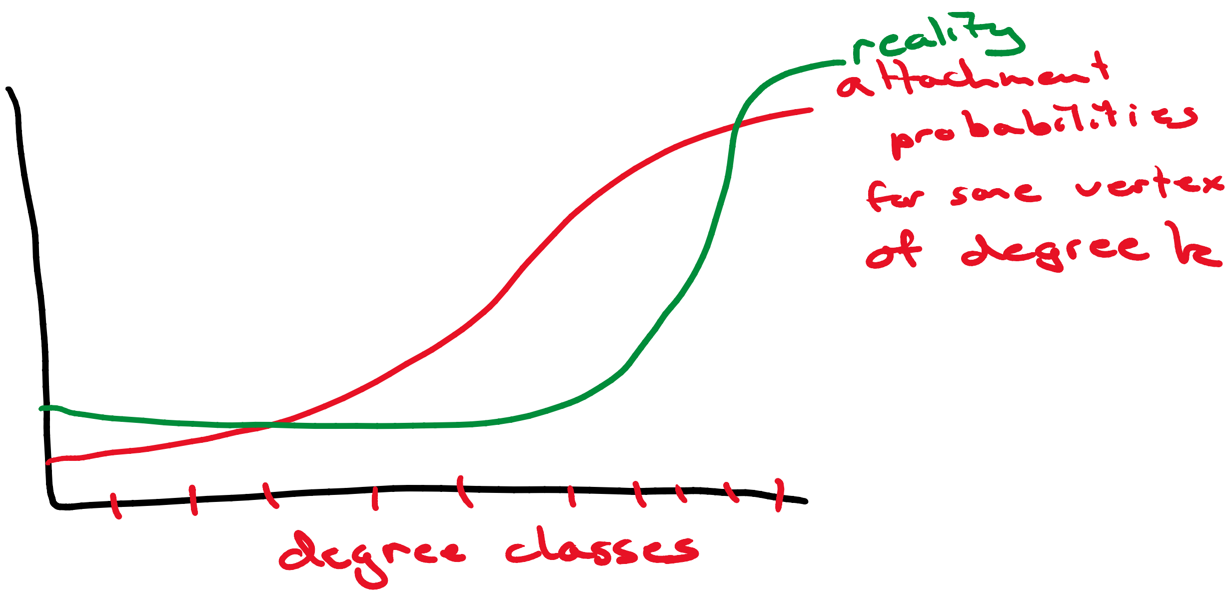
However: C-L graphs contain multi-edges and self loops, and this is contained in the attachment probabilities

Issue is if we want to consider

Issue is if we want to consider simple graphs specifically

Theorists: as $n \to \infty$ probability of self-loops and multiedges $\to 0$

Reality: only the case when n is much larger than most graphs we study

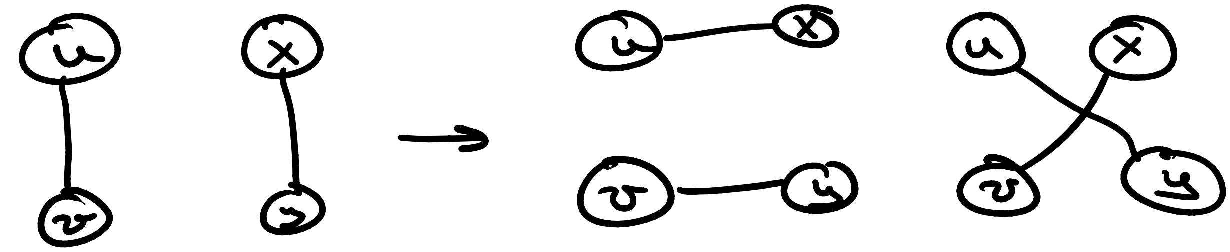


attachment probabilities for simple graphs are <u>not</u> C-L probabilities

So how can we actually generate a null model for a simple graph?

A: we take any graph with the given degree distribution and do double edge swaps

given degree distribution and do
<u>double edge swaps</u>



If we randomly select edges and swap them, ignoring multi/loop swaps, and <u>repeat this</u>:

markov process that will sample from the simple graph space in a uniformly random way

Mixing time: unknown