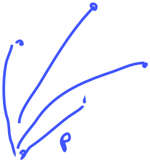
MITx: Statistics, Computation & Applications

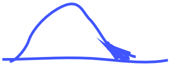
Criminal Networks Module

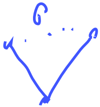
Lecture 3: Network Models

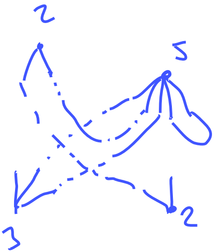


$$(n-1)p = \mathbb{E}[\text{degree}]$$

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






2
5
3
2

MITx:
Statistics, Computation & Applications

Criminal Networks Module

Lecture 3: Network Models

Network models

Develop models of networks in order to:

- determine interesting structural properties in a network
- study how networks come to exist and how this changes the structural properties
- study processes on networks and be able to make statements for many networks at once

Erdös-Renyi model

- Random graph models: network models that have some specified parameters, but otherwise the edges in the network appear at random
- Simple network model: Fix number of nodes and number of (undirected) edges and place edges uniformly at random

Erdös-Renyi model

- Random graph models: network models that have some specified parameters, but otherwise the edges in the network appear at random
- Simple network model: Fix number of nodes and number of (undirected) edges and place edges uniformly at random
- Erdös-Renyi model: $G(n, p)$, where each edge between the n nodes is formed with probability $p \in [0, 1]$ independently of every other edge

Erdös-Renyi model

- Random graph models: network models that have some specified parameters, but otherwise the edges in the network appear at random
- Simple network model: Fix number of nodes and number of (undirected) edges and place edges uniformly at random
- Erdös-Renyi model: $G(n, p)$, where each edge between the n nodes is formed with probability $p \in [0, 1]$ independently of every other edge
  - similar but easier to handle mathematically
  - expected number of edges is $\mathbb{E}[\text{number of edges}] = \binom{n}{2} p$
  - expected degree of a node i is $\mathbb{E}[k_i] = (n-1)p$
  - degree distribution is $\mathrm{Binomial}(n-1, p)$,
    i.e. $\mathbb{P}(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$
  - degree distribution does not follow a power law:

    approximation of binomial distribution by Poisson distribution:
    $\mathbb{P}(k) = \frac{e^{-\lambda}\lambda^k}{k!}$, where $\lambda = (n-1)p$

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$\log p_k = -\lambda + k\log(\lambda) - \underline{\log(k!)}$$

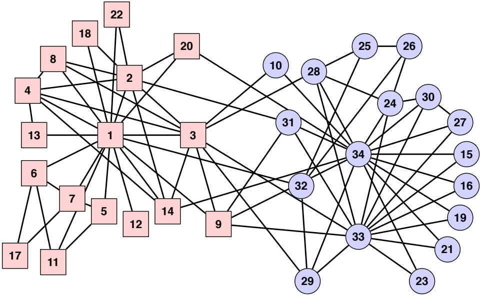
$$\approx -k\log k - k \quad + k\log(\lambda)$$

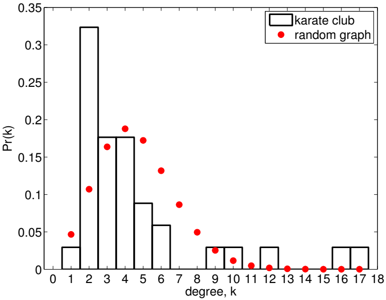
$$\approx \underline{-k\log k}$$

power law degree distr

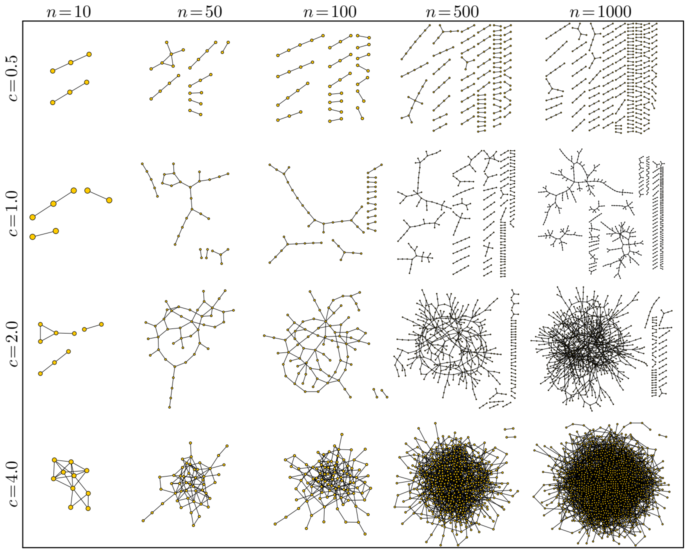
$$\log p_k \approx \underline{-\alpha\log(k)}$$

Binomial degree distribution





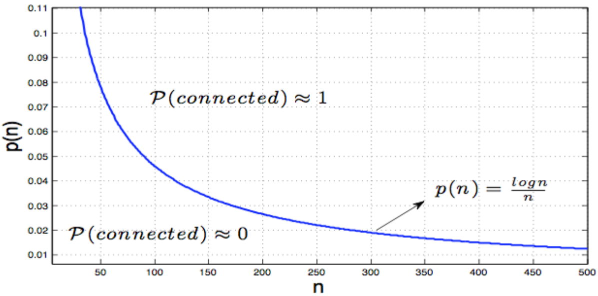
Erdös-Renyi graphs



$c = (n-1)p$ (expected degree)

Erdös-Renyi graphs: connected components

- study structural graph properties as $n \to \infty$
- Erdös-Renyi graphs show phase transition:
  - $p < \frac{1}{n}$: no connected component of size $\geq c\log(n)$
  - $p > \frac{1}{n}$: giant component emerges, i.e. component of size cn
  - $p > \frac{\log(n)}{n}$: graph is connected
- proof of phase transitions in seminal work of Erdös & Renyi (1959)



Erdös-Renyi graphs: clustering coefficient

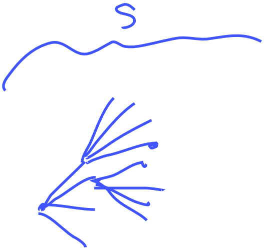
Expected clustering coefficient in Erdös-Renyi graphs is $\frac{\binom{n}{3}p^3}{\binom{n}{3}p^2} = p \approx \frac{c}{n}$

	Network	Type	n	m	c	S	ℓ	α	C
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980	3.48	2.3	0.20
	Company directors	Undirected	7 673	55 392	14.44	0.876	4.60	–	0.59
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822	7.57	–	0.15
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838	6.19	–	0.45
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918	4.92	–	0.088
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16			2.1	
	Email messages	Directed	59 812	86 300	1.44	0.952	4.95	1.5/2.0	
	Email address books	Directed	16 881	57 029	3.38	0.590	5.22	–	0.17
	Student dating	Undirected	573	477	1.66	0.503	16.01	–	0.005
	Sexual contacts	Undirected	2 810					3.2	
Information	WWW `nd.edu`	Directed	269 504	1 497 135	5.55	1.000	11.27	2.1/2.4	0.11
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914	16.18	2.1/2.7	
	Citation network	Directed	783 339	6 716 198	8.57			3.0/–	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977	4.87	–	0.13
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000		2.7	
Technological	Internet	Undirected	10 697	31 992	5.98	1.000	3.31	2.5	0.035
	Power grid	Undirected	4 941	6 594	2.67	1.000	18.99	–	0.10
	Train routes	Undirected	587	19 603	66.79	1.000	2.16	–	
	Software packages	Directed	1 439	1 723	1.20	0.998	2.42	1.6/1.4	0.070
	Software classes	Directed	1 376	2 213	1.61	1.000	5.40	–	0.033
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000	11.05	3.0	0.010
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805	4.28	2.1	0.012
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996	2.56	2.2	0.090
	Protein interactions	Undirected	2 115	2 240	2.12	0.689	6.80	2.4	0.072
	Marine food web	Directed	134	598	4.46	1.000	2.05	–	0.16
	Freshwater food web	Directed	92	997	10.84	1.000	1.90	–	0.20
	Neural network	Directed	307	2 359	7.68	0.967	3.97	–	0.18

$n = |\text{nodes}|$, $m = |\text{edges}|$, c: mean degree, S: prop. largest component, ℓ: mean geodesic, α: exp. power-law degree distribution, C: clustering coeff.

Erdös-Renyi graphs: diameter

- for constant average degree c, one can show that diameter is $\log(n)$



$$c^s = n$$

$$s\log(c) = \log(n)$$

$$s = \frac{\log(n)}{\log(c)}$$

Erdös-Renyi graphs: diameter

- for constant average degree c, one can show that diameter is $\log(n)$
- heuristic argument:
  - average number of nodes s steps away from randomly chosen node: c^s
  - all nodes can be reached when $c^s \approx n$, or equivalently when $s \approx \frac{\log(n)}{\log(c)}$
  - hence, the diameter is $s \approx \frac{\log(n)}{\log(c)}$

Erdös-Renyi graphs: diameter

- for constant average degree c, one can show that diameter is $\log(n)$
- heuristic argument:
  - average number of nodes s steps away from randomly chosen node: c^s
  - all nodes can be reached when $c^s \approx n$, or equivalently when $s \approx \frac{\log(n)}{\log(c)}$
  - hence, the diameter is $s \approx \frac{\log(n)}{\log(c)}$

Summary of structural properties of Erdös-Renyi graphs:

- small-world property, i.e. diameter of the order of $\log(n)$
- locally tree-like, i.e. few triangles
- and binomial degree distribution, i.e. no power law

Configuration model

How to address unrealistic degree distribution of Erdös-Renyi graphs?

Configuration model

How to address unrealistic degree distribution of Erdös-Renyi graphs?

- generate random networks with a given degree distribution (e.g. power law)

Configuration model

How to address unrealistic degree distribution of Erdös-Renyi graphs?

- generate random networks with a given degree distribution (e.g. power law)

Configuration model (Bender and Canfield, 1978)

- One of the most widely used models for this purpose
- Specify degree sequence $(k_1, \dots, k_n)$ and attach k_i stubs to node i
- Choose 2 stubs at random and create edge connecting them and continue process with remaining stubs

1 1 1 2 2 3 3 3 3 4 4 4 5 5

1 2 | 3 5 | 1 2 | 3 4 | 3 5 | 4 3 | 1 4

Configuration model

How to address unrealistic degree distribution of Erdös-Renyi graphs?

- generate random networks with a given degree distribution (e.g. power law)

Configuration model (Bender and Canfield, 1978)

- One of the most widely used models for this purpose
- Specify degree sequence $(k_1, \ldots, k_n)$ and attach k_i stubs to node i
- Choose 2 stubs at random and create edge connecting them and continue process with remaining stubs
- **Drawback:** It is possible to have self-loops and multi-edges
  - but average number of self-loops and multi-edges is a constant as $n \to \infty$
- **Drawback:** One can show that clustering coefficient is of order of $\frac{1}{n}$

Preferential attachment model

- other type of network models: generative network models
- explore hypothesized generative mechanisms to see what structures they produce

Preferential attachment model

- other type of network models: generative network models
- explore hypothesized generative mechanisms to see what structures they produce

Preferential attachment model (Price, 1976)

- In context of bibliographic network:
  - newly appearing paper cites previous ones with probability proportional to their number of citations
  - give every paper β citations for free
- one can show that this model has power law degree distribution: $p_k \sim k^{-\alpha}$, where $\alpha = 2 + \frac{\beta}{c}$

$$p(\text{edge to node } i) = \frac{\deg(\text{node } i) + \beta}{\sum_i (\deg(\text{node } i) + \beta)}$$

Preferential attachment model

- other type of network models: generative network models
- explore hypothesized generative mechanisms to see what structures they produce

Preferential attachment model (Price, 1976)

- In context of bibliographic network:
  - newly appearing paper cites previous ones with probability proportional to their number of citations
  - give every paper β citations for free
- one can show that this model has power law degree distribution: $p_k \sim k^{-\alpha}$, where $\alpha = 2 + \frac{\beta}{c}$
- **drawback:** leads to acyclic network
- **drawback:** small clustering coefficient

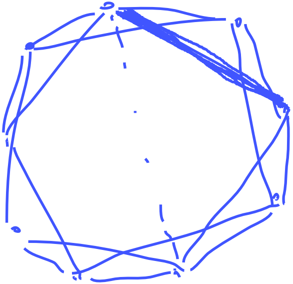
Small-world model

How to address unrealistic clustering coefficients?

Small-world model

How to address unrealistic clustering coefficients?

- generate circulant graph, where every node around a cycle is connected to its c closest neighbors
- **drawbacks:** diameter is of order n, no power law degree distribution



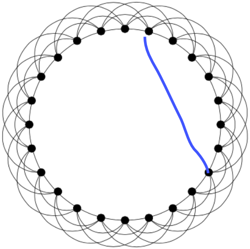
\# triangles $O(n)$
diameter $O(n)$

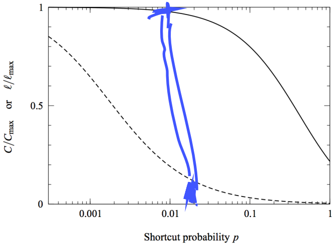
\# triangles $O(n)$
diameter $O(\log(n))$

Small-world model

How to address unrealistic clustering coefficients?

- generate circulant graph, where every node around a cycle is connected to its c closest neighbors
- **drawbacks:** diameter is of order n, no power law degree distribution
- **Small-world model** (Watts and Strogatz, 1998)
  - rewire edges at random with probability p
  - leads to networks with high clustering coefficient and small diameter
  - **drawback:** degree distribution does not follow power law





Summary: Network models

- Erdös-Renyi model
  - realistic diameter, but small clustering coefficient and binomial degree distribution
- configuration model
  - power law degree distribution, realistic diameter, but small clustering coefficient
- preferential attachment model
  - generative model with power law degree distribution, realistic diameter, but small clustering coefficient
- small-world model
  - generative model with high clustering coefficient, realistic diameter, but no power law degree distribution

How to construct a network model that leads to power law degree distribution, realistic diameter, and high clustering coefficient?

References

- Chapters 12-15 in

  M. E. J. Newman. *Networks: An Introduction*. 2010.