References

These notes are drawn from the following sources:

- Newman, The structure and function of complex networks. SIAM Review (2003). 45(2). pg. 167-256.
- Boccaletti, Latora, Moreno, Chavez, Hwang, Complex networks: structure and dynamics. Physics Reports (2006). 424. pg. 175-308
- Barrat & Weight, On properties of small-world network models. Eur Phys J B (2000). 13. pg. 547-560.
- Albert & Barabasi, Statistical mechanics of complex networks. Rev Mod Phys (2002). 74(1). pg. 74-97.
- Lada Adamic's course notes for SI 508. Available online as part of University of Michigan's Open Educational Resources Initiative.

Notation:

$N =$ # of nodes

$K =$ # of edges/links

$\mathcal{N} = \{n_1, n_2, \ldots, n_N\} =$ set of nodes

$\mathcal{L} = \{l_1, l_2, \ldots, l_K\} =$ set of links

$G_{N,K} = (\mathcal{N}, \mathcal{L}) =$ graph (directed or undirected)

$l_{ij} = (i,j) =$ link between nodes i & j

$k_i =$ degree of node i = # of neighbors

$p_k =$ degree distribution = probability that a node chosen at random has degree k

$L =$ average shortest path $= \frac{1}{N(N-1)} \sum_{\substack{i,j \in \mathcal{N} \\ i \neq j}} d_{ij}$,

where d_{ij} = shortest path from i to j

$c_i =$ local clustering coefficient $= \dfrac{\text{\# of edges between } k_i \text{ neighbors of } i}{\text{max possible \# of edges between them}}$

$= \dfrac{e_i}{k_i(k_i-1)/2}$

$C =$ Clustering coefficient $= \frac{1}{N} \sum c_i$

Today I'll talk about two alternatives to the Erdos-Renyi random graphs Veronica talked about last time. We will begin by reviewing ER graphs so that we can compare them to the real world and today's networks.

First recall that the # of possible links in a graph is bounded in between 0 and $\frac{N(N-1)}{2}$, where N is the number of nodes.

Recall degree:

k_i = degree (# of neighbors) of node i

p_k = degree distribution (encodes a lot of info about the topology of the graph)

For ER,

$$p_k = e^{-Np} \frac{(Np)^k}{k!}, \qquad \text{(Poisson)}$$

where p = the probability of connecting any 2 nodes.

Recall characteristic path:

This is the average shortest distance between nodes.

For ER,

$$L = \text{characteristic path length} \sim \frac{\log N}{\log(\text{mean degree})} = \frac{\log N}{\log Np}.$$

Lastly, recall the clustering coefficient:

This quantity tells you how clique-ish your graph is (how likely are your friends to also be friends).

For ER,

$$C = \text{clustering coefficient} = O\left(\frac{1}{N}\right),$$

which notably $\rightarrow 0$ as $N \to \infty$. Further note that C is always a fraction between 0 and 1.

In comparison to random graphs, we can look at quantities of real world networks:

ER	Real-World
• p_k is Poisson	• Scale-free power law distribution, $p_k \sim k^{-\alpha}$
• $L \sim \frac{\log N}{\log Np}$	• small L ($L \sim \log N$ at most)
• $C \to 0$ as $N \to \infty$	• Highly clustered, $C = O(1)$

In particular, real-world networks often display degree distributions with $p_k \sim k^{-\alpha}$ for $\alpha \in [2,3]$.

We conclude that ER does well with L, but has poor p_k and C.

Now we will move on to talking about 3 different models in turn that focus on reproducing real-world values of p_k, C, and L.

$\boxed{\text{Small-World Network (Watts \& Strogatz)}}$

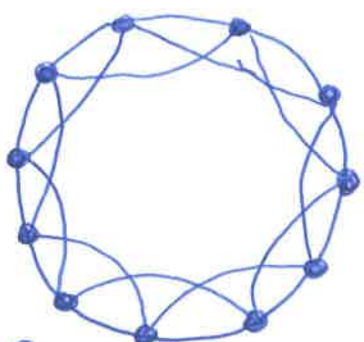
- 1998
- one parameter model
- represents a middle ground between a random graph and a regular lattice
- first successful attempt to create networks with real-world C & L.

WS Small World Algorithm:

1. Start with order: consider a rectangular lattice with N nodes.
   → We consider 1D with periodic BC's, so a ring.

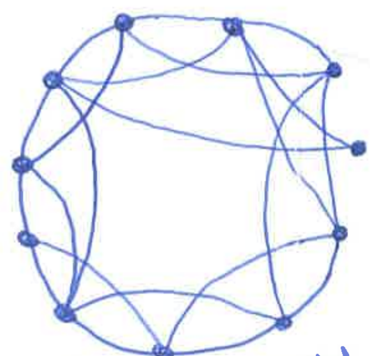
2. Connect every node to its first $2m$ neighbors (m on either side).
   → Note this leads to $K = mN$ total (undirected) edges.

3. Randomize: for every node and every link at that node, with probability p break the link and rewire it to a randomly chosen node (assuming no double or self edges).
   → with probability $1-p$ leave the link in place
   → This introduces pNm long-range edges.



Regular lattice

Example with $m=2$



Small world

Changing p lets us regulate between structure & randomness.

We will first look at $p_k(p)$, $L(p)$, and $C(p)$ for the limiting cases $p=0$ & $p=1$.

<u>$p=0$: Ring Lattice</u>

- All nodes have the same degree $(2m) \longrightarrow p_k(0) = \delta_{2m}(k)$

- Average node is a quarter of the circle away, so

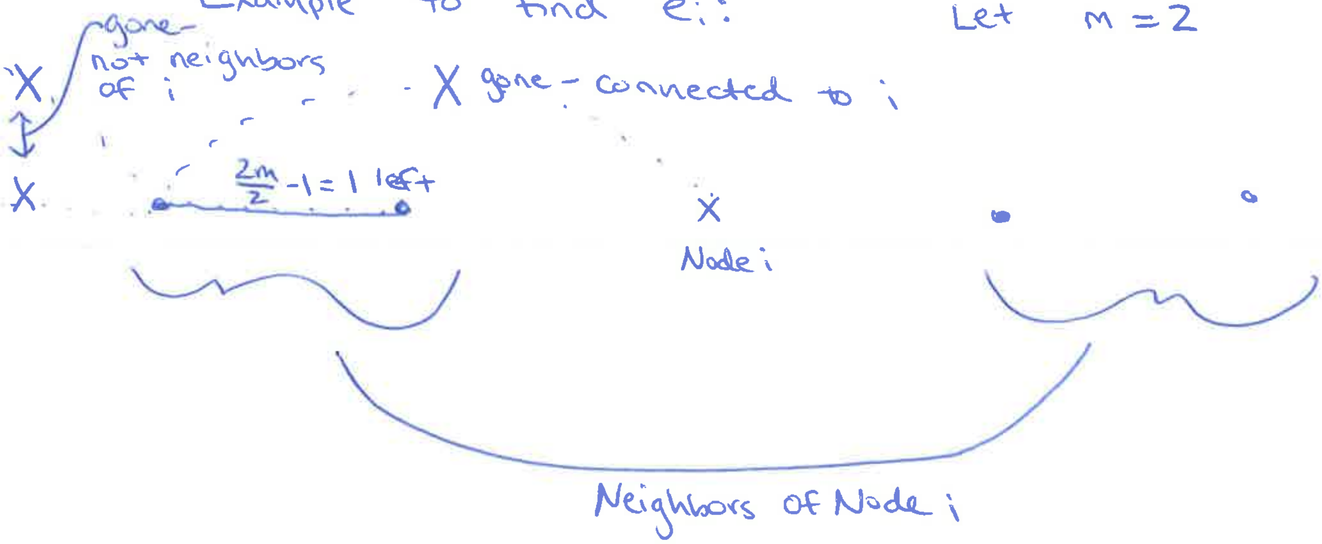
$\frac{N}{4}$ hops away. Plus you can hop m nodes at a time, so $L(0) \approx \frac{N}{4m} \gg 1$.

- Clustering coefficient: since the local clustering C_i is the same for all nodes, the average $C = C_i$ and we only need to find it for one node, as follows.

$$C = C_i = \frac{e_i}{k_i(k_i-1)/2} = \frac{\text{ratio of \# of edges that actually exist between } k_i \text{ neighbors}}{\text{possible number}},$$

where we know $k_i = 2m$.

Example to find e_i: Let $m = 2$



- At m jumps away from i, you have at most $m-1$ edges $\left(\frac{2m}{2} - 1\right)$. This is because half are gone and the one connected to i is subtracted.
- At $m-1$ jumps away, you have $m-1+1$ connections.
- At $m-j$ jumps away, you have $m-1+j$ connections.

$$\Longrightarrow e_i = \frac{1}{2} \cdot 2 \sum_{j=0}^{m-1} m-1+j = \frac{3}{2} m(m-1)$$

multiply by 2 for both sides

divide by 2 for undirected

It follows that

$$C = \frac{2e_i}{k_i(k_i-1)} = \frac{3(m-1)m}{2m(2m-1)} = \frac{3(m-1)}{4m-2} \to \frac{3}{4}.$$

P=1 Random Graph:

At $p=1$, we have a random graph with the extra constraint that every node has minimum connectivity $k_{min} = m$.

- P_K is Poisson
- $L(1) \approx \frac{\log N}{\log 2m}$ (since $2m$ is the average degree)
- $C(1) \approx \frac{2m}{N}$ (because the probability that 2 nodes are connected is $\frac{2m}{N}$ and the probability that their neighbors are connected is $\frac{2m-1}{N} \approx \frac{2m}{N}$)

similar to ER graphs

Now we move on to studying intermediate values of p.

$P_K(p)$ degree distribution:

We will show that $p>0$ introduces disorder into the network and spreads out $P_K(0) = \delta_{2m}(k)$.

Note: Since a single edge of each link is rewired, we maintain $k_{min} = m$ edges and introduce pmN long-range edges.

The average degree $= 2m$.

Note: For $m \geq 1$, we always get a connected network.

Now, given that each vertex has min degree $k_{min} = m$, we can write the degree of vertex i as

$$K_i = \underset{\substack{\uparrow \\ \text{guaranteed} \\ \text{min \# of} \\ \text{nodes}}}{m} + \underset{\substack{\uparrow \\ \text{additional nodes due to random swaps}}}{c_i},$$

where $c_i \geq 0$. We further split c_i into 2 parts, with $c_i = c_i^1 + c_i^2$:

- $c_i^1 \leq m$ is the # of edges that have been left in place (happens with probability $1-p$)

and

- c_i^2 is the # of rewired edges that have been rewired toward i from other nodes (happens with probability $\frac{1}{N}$).

Then the probability distribution of c_i^1 is $\mathrm{Bin}(m, 1-p)$:

$$\mathbb{P}(c_i^1) = \binom{m}{c_i^1} (1-p)^{c_i^1} p^{m-c_i^1}$$

(we choose m nodes and keep them in place with probability $1-p$).

Similarly, c_i^2 is also Binomial with $c_i^2 \sim \mathrm{Bin}(pNm, \frac{1}{N})$:

$$\mathbb{P}(c_i^2) = \binom{pNm}{c_i^2} \left(\frac{1}{N}\right)^{c_i^2} \left(1 - \frac{1}{N}\right)^{pNm - c_i^2}.$$

In the limit of large N,

$$\mathbb{P}(c_i^2) \to \mathrm{Poisson}\left(pNm \cdot \frac{1}{N}\right) = \frac{(pm)^{c_i^2} e^{-pm}}{c_i^2!}.$$

Using that the distribution of a sum is the convolution of the distributions, we obtain

$$\mathbb{P}(c) = \sum_j^{\min(c,m)} \mathbb{P}(c^1 = j)\, \mathbb{P}(c^2 = c-j) = \sum_j \binom{m}{j} (1-p)^j p^{m-j} \frac{(pm)^{c-j} e^{-pm}}{(c-j)!}$$

And then, since $k = m + c$, we conclude that

$$P_k(p) = \sum_{j=0}^{\min(k-m,m)} \binom{m}{j} (1-p)^j p^{m-j} \frac{(pm)^{k-m-j}}{(k-m-j)!} e^{-pm}.$$

$C(p)$ clustering:

Recall that $C(0) = \frac{3(m-1)}{2(2m-1)}$. Rather than calculate $C(p)$ directly, we will calculate C', a different quantity with the same general meeting:

$$C'(p) = \frac{\text{mean \# of links between neighbors of vertex}}{\text{mean \# of possible links between neighbors}}$$

$$\approx C(0)\,(1-p)^3$$

initial clustering coefficient — $C(0)$

probability that 2 neighbors of i at $p=0$ are neighbors of i later and also neighbors of each other — $(1-p)^3$

Barrat & Weight (2000) showed that $C'(p) \to C(p)$ as $N \to \infty$.

$L(p)$ characteristic path length:

Note: L does not begin to decrease until $p \geq \frac{1}{Nm}$, which guarantees the existence of at least one shortcut, since # of shortcuts $= Nmp$. Conversely,

$$N \geq \frac{1}{mp} \quad \text{for } L \text{ to decrease}$$

$\Rightarrow$ suggests $\exists N^*(p)$ such that

$$N < N^* \Rightarrow L \sim N$$

$$N > N^* \Rightarrow L \sim \log N$$

Now it is widely accepted that $L(N,p) \sim \frac{N}{2m} f(p2mN)$

with $f(u) = \begin{cases} \text{constant} & \text{if } u \ll 1 \\ \ln u / u & \text{if } u \gg 1 \end{cases}$ (has been confirmed using numerics, series expansions, etc.)

This says that

$$L(N,p) \sim \frac{N}{2m} \quad \text{for } p2mN \ll 1$$

$$\sim \frac{\ln(2mpN)}{4m^2} \quad p2mN \gg 1.$$

and it has been shown using a mean field method that is exact for u small or large but not $u \simeq 1$ that

$$f(u) = \frac{4}{\sqrt{u^2+4u}} \tanh^{-1}\left(\frac{u}{\sqrt{u^2+4u}}\right)$$

for 1D small world. (See Newman, Moore, & Watts (2000) or Barbour & Reinert (2001))

Notice that $u = 2pmN = 2 \times \#$ of shortcuts and

$f(u)$ = average fraction by which the distance between 2 nodes is reduced by a given u.

Recall that:

$\underline{p=0}$

✓ high $C \to \frac{3}{4}$

✗ high $L \approx O(N)$

$\underline{p=1}$

✗ low $C \approx \frac{\log N}{\log 2m}$

✓ low $L \approx \frac{2m}{N}$

A true small world should have high C and low L, but these limiting cases suggest that high C is associated with high L and low with low. In fact, there is a region with high C and low L for small

p. Because p slightly greater than 0 immediately reduces L while leaving clustering the same locally, we have a region of high C and low L.

→ The take-away for WS small world is then

X p_k
✓ C
✓ L.

Now we will move on to talking about 2 networks that are able to produce scale free power law distributions.

1. Price Model (1965)
2. Barabasi-Albert Model (1999)

Price Model Algorithm:

Consider a directed graph with N vertices (or papers, because Price was interested in understanding citation networks). Let

p_k = in-degree
in-degree = # of citations of your paper
out-degree = # of references in your paper's bibliography

There are two parts to creating the network:

1. Growth: new vertices are added continuously (not necessarily at a constant rate). Each new vertex is given a fixed out-degree at appearance/publication. (Note - we assume out-degree can vary, but that mean out-degree m is constant. This also means mean in-degree = m).

2. Cumulative Advantage: This is the rule to determine which papers to cite in the new node's bibliography (according to "rich get richer" idea). In its simplest form, cumulative advantage says that

$$\mathbb{P}(\text{new paper cites } i) \propto k_i,$$

so that

$$\mathbb{P}(\text{new paper cites } i) = \Pi_i(k_i) = \frac{k_i}{\sum_j k_j}.$$

→ but this would mean all new vertices have $\Pi_i = 0$ forever.

Price's remedy: $\mathbb{P}(\text{new paper cites } i) \propto k+k_0$, with $k_0 = 1$ (think: each new paper's existence is its first citation). Then

$$(*)\quad \mathbb{P}\left(\begin{array}{l}\text{new node attaches to}\\ \text{any vertex of degree } k\end{array}\right) = \frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1}.$$

Degree distribution for Price:

We will consider $p_{k,N}$ = degree distribution as a function of N = # of vertices. We will use the master equation approach to find $p_{k,N}$:

Note: mean # of new citations to vertices of current degree $k = \frac{(k+1)p_k m}{m+1}$.

Note: $N p_{k,N}$ = # of vertices of in-degree k at size N.

Goal: Understand how # of vertices of in-degree k changes as N goes to $N+1$.

$$\underbrace{(N+1)p_{k,N+1}}_{\substack{\#\text{ of vertices}\\ \text{of degree } k \text{ at}\\ \text{size } N+1}} - N p_{k,N} = \begin{array}{c}\text{mean \# of old}\\ \text{degree } k-1 \text{ papers}\\ \text{that were}\\ \text{cited}\end{array} - \begin{array}{c}\text{mean \# of old}\\ k \text{ papers that}\\ \text{were cited}\end{array}$$

Using (✱) on the last page, we obtain

$$(N+1)\, p_{k,N+1} - N p_{k,N} = \left(k p_{k-1,N} - (k+1) p_{k,N}\right) \frac{m}{m+1}$$

for $k \geq 1$

and

$$(N+1)\, p_{k,N+1} - N p_{k,N} = 1 - \text{any old papers of degree 1 that were cited}$$

$\uparrow$ one new paper cites itself

$$= 1 - p_{0,N} \frac{m}{m+1} \quad \text{for } k = 1.$$

Looking for stationary solutions, we set $p_{k,N+1} = p_{k,N} = p_k$ and rearrange to obtain

$$p_k = \begin{cases} \left[k p_{k-1} + (k+1) p_k\right] \frac{m}{m+1} & \text{for } k \geq 1 \\ 1 - p_0\, m(m+1) & \text{for } k = 0 \end{cases}$$

$$\Longrightarrow \quad p_0 = \frac{m+1}{2m+1} \quad \text{and}$$

$$p_k = \left(1 + \frac{1}{m}\right) B\left(k+1,\, 2 + \frac{1}{m}\right),$$

where $B(a,b)$ is Legendre's beta function. Note $B(a,b) \to a^{-b}$ for large a and fixed b. Thus, we obtain

$$p_k \sim k^{-(2+\frac{1}{m})},$$

which is a power law with $\alpha = 2 + \frac{1}{m}$! This will typically give values in agreement with real-world networks ($\in [2,3]$).

Now we turn to a more simplified model that lends itself to more analysis.

Barabasi-Albert Model:

The BA model is a simplification of sorts of the Price model that makes it less challenging to analyze.

BA Algorithm:

1. Growth: Start with m_0 nodes. At each time step, add a new node with $m \le m_0$ edges.
   → Unlike Price, the graph is undirected and m is set constant $\forall$ nodes.

2. Preferential attachment: Assume that
   $\mathbb{P}(\text{new node is connected to } i) \propto k_i$ with

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

   → Unlike with Price, we don't have to set this $\propto k + k_0$, because each vertex that appears in the graph starts with nonzero m degree.

Notice that $N = t + m_0$ at t time steps later.

$\underset{\#\text{ of nodes}}{N} = \underset{t \text{ added}}{t} + \underset{\text{initial nodes}}{m_0}$

Furthermore, $K = \#$ of edges $= mt$.

One can apply the same master equation equation approach as with Price to find p_k, but we will use a slightly different method for variety.

Continuum approach for p_k

We are interested in $k_i(t)$, the time-dependence of k_i for a given node i.

Note that $k_i(t)$ increases every time a new node appears and links to i (this happens with probability $\Pi(k_i)$. Suppose k_i is a continuous random variable.

Then
$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) \quad \text{and}$$
$$\frac{\partial k_i}{\partial t} = m\,\Pi(k_i) = m \frac{k_i}{\sum_j k_j} = \frac{m k_i}{\text{sum over all degrees}} = \frac{m k_i}{2 \times \#\text{ of edges}} = \frac{m k_i}{2mt}.$$

Thus, $\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$. With IC $k_i(t_i) = m$, we have
$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}.$$

(notice that all nodes evolve the same way)

Now we calculate the probability that node i has degree $k_i(t) < k$:
$$\mathbb{P}[k_i(t) < k] = \mathbb{P}\left[t_i > \frac{m^{1/2}}{k^{1/2}}t\right],$$

because $k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2} \Rightarrow t_i = \frac{m^{1/2}}{k^{1/2}}t.$

We assumed we add the nodes at equal time intervals, so that
$$t_i \sim \text{uniform}$$
$$P(t_i) = \frac{1}{m_0 + t}$$

$$\Rightarrow \mathbb{P}[k_i(t) < k] = \mathbb{P}\left[t_i > \frac{m^{1/2}}{k^{1/2}}t\right]$$
$$= 1 - \left(\frac{m^{1/2}}{k^{1/2}}t\right)\left(\frac{1}{m_0 + t}\right).$$

Then $p_k = \frac{\partial}{\partial k}\mathbb{P}(k_i(t) < k) = \frac{m^{1/2}t}{k^{3/2}(m_0 + t)}.$

As $t \to \infty$, $p_k \sim 2m^{1/2}k^{-3}.$

So we obtain a scale free degree distribution, but we lose by simplifying the Price model is the ability to adjust k. Instead, $k=3$.

As for characteristic path length, $L \sim \frac{\ln N}{\ln \ln N}$, which satisfies small world.

There is no analytical result for clustering, but it seems that $C \sim N^{-0.75}$. This is better that $C \sim \frac{1}{N}$ for ER, but is still poor.

In conclusion, BA model fairs well with p_k and L, but clustering is too low.