IV: Section (ix): Degree Distributions, Random Networks, Power Laws & Scale-Free Networks.

Midway through the 20th century, the first formal attempts to study complex networks were undertaken. Early pioneering work was done by Erdös & Rényi (1959), which ushered in a first paradigm for the formal study of particular complex networks known as ER graphs (for Erdös-Rényi). The governing principle behind ER graphs is that their edge connections are determined randomly.

Consider, to this end, a simple, connected graph G in which pairs of vertices, u & v are connected with probability p_{uv}. We recall that for a graph with n vertices, there are a total of $\binom{n}{2}$ possible edges, and so the parameterization of such an ER graph could require at most the specification of $\binom{n}{2}$ probability parameters. For simplicity of exposition, we fix $p_{uv} = p$ (a constant) for all $u, v \in V(G)$.

Define $\boxed{ER(n,p)}$ as an ER random graph on n vertices, in which each two distinct vertices are connected by an edge with probability p.

We now consider a brief statistical analysis of ER graphs, beginning with a short review of some concepts from elementary statistics. In particular, we wish to investigate the nature of degree distributions, average distance/path length & connectivity in very large complex networks.

Concepts from Elementary Statistics

A $\boxed{\text{random variable}}$, X, is a variable that is assigned a numerical value based on the outcome of some random "event". If the set of all such possible numerical assignments for X is finite, we call this random variable a discrete random variable. Here we only consider the discrete case.

For instance, for the "event" of flipping a coin we might say: assign $X=1$ for Heads, $X=0$ for Tails.

Similarly ~~The~~ for the "event" of tomorrow's weather, we might say: assign $X=2$ if it is sunny, $X=1$ if it is cloudy & $X=0$ if it rains.

Once we properly define a random variable for some probabilistic event, we can then render a $\boxed{\text{probability distribution}}$ for X. In general, a probability distribution satisfies (2) intuitive conditions:

1. $0 \leq P(X) \leq 1$, all probabilities are bounded by zero & one, &
2. $\sum_n P(X) = 1$

where property (2) says that the sum of the probabilities over all possible values of X equals 1 (This basically says some outcome had to occur!).

$\boxed{\text{Ex.}}$ Let $X=1$ for Heads, $X=0$ for Tails for the event of flipping a fair coin. Then:

X	$P(X)$
0	$\frac{1}{2}$
1	$\frac{1}{2}$

Gives a probability distribution. Note that:

$$\underbrace{P(X=1)}_{\text{Read: prob. } x=1} + \underbrace{P(X=0)}_{\text{Read: Prob. } x=0} = 1, \text{ as required.}$$

Note that we will henceforth write $P(k)$ for $P(X=k)$, as is customary.

$\boxed{\text{Ex.}}$ Find the prob. distribution modeling the event of rolling a 6-sided die.

X	$P(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Note that:

$$\sum_{k=1}^{6} P(k) = \frac{1}{6} + \cdots \frac{1}{6} = 1. \checkmark$$

One common calculation is statistics in regard to random variables involves the computation of what is called $\boxed{\text{expectation}}$, denoted: $\boxed{E[X]}$ or sometimes simply abbreviated: μ, for Mean.

Simply put, $E[X]$ is equal to the value we "expect" X to acquire following a single $\boxed{\text{Trial}}$ of the referenced probabilistic event. Note that $E[X]$ need not equal, in a strict sense, one of the values assigned to X when defining this particular random variable.

$\boxed{\text{Def}}$ $\boxed{E[X] = \sum_{i=1}^{n} x_i \cdot P[x_i]}$, for X, a discrete R.V.

$\boxed{\text{Ex.}}$ Consider the coin-flipping experiment: $X = 1$ for Heads, $X = 0$ for Tails.

$$E[X] = \sum_{i=1}^{2} x_i \cdot P[x_i] = 0 \cdot \underbrace{P(0)}_{=\frac{1}{2}} + 1 \cdot \underbrace{P(1)}_{=\frac{1}{2}} = 0 + \frac{1}{2} = \boxed{\frac{1}{2}}.$$

Roughly speaking, this results says that for a single flip of a coin, we expect $X = \frac{1}{2}$, since the outcomes of Heads / Tails are equally likely.

$\boxed{\text{Ex.}}$ Consider the die-rolling experiment for a fair six-sided die.

$$E[X] = 1 \cdot P(1) + 2 \cdot P(2) + 3 \cdot P(3) + 4 \cdot P(4) + 5 \cdot P(5) + 6 \cdot P(6)$$
$$= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}$$
$$= \frac{21}{6} = \boxed{3.5}$$

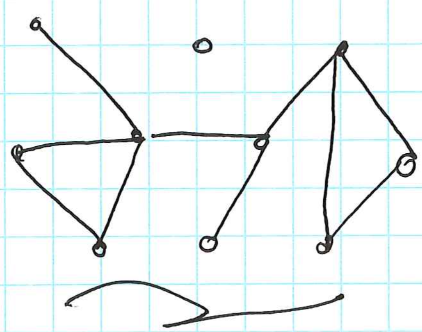
← This says that we "expect" to see, roughly, 3.5 (the mean) appear on the up-turned die face.

→ Note: From the previous discussions, we can see that $E[X]$ is essentially a weighted (prob.) average.

Let's now apply some of these core statistical ideas to the analysis of networks & graphs.

Next we construct the $\boxed{\text{Degree Distribution}}$ of a graph/network.

$\boxed{\text{Ex.}}$
Consider the graph below. Let $\underline{X}$ be defined as $\underline{\deg(v)}$ for a randomly (i.e. uniform random) chosen vertex from G.



G

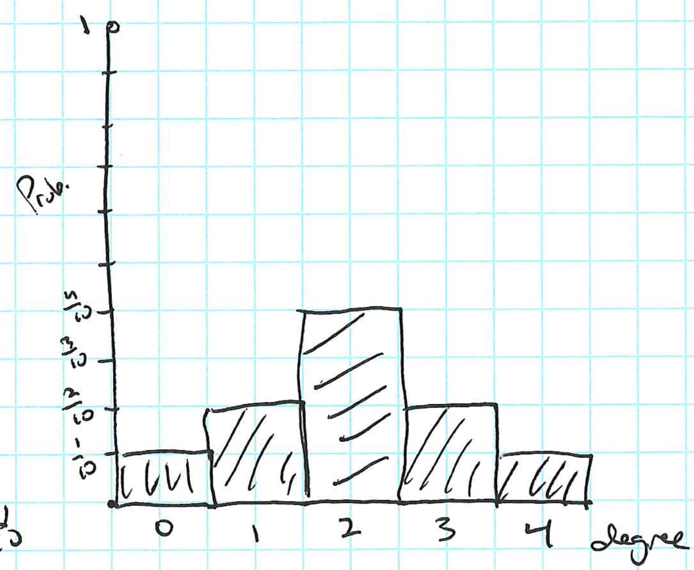
$|V(G)| = 10$

Note that (1) deg. vertex has degree $\boxed{0}$, (2) have degree $\boxed{1}$, (4) have degree $\boxed{3}$, and (1) has degree $\boxed{4}$.

Thus the <u>probability distribution</u> of X, i.e. the $\boxed{\text{degree distribution}}$ of the graph is:

X	$P(X)$
0	1/10
1	2/10
2	4/10
3	2/10
4	1/10

Histogram of degree distribution →



$$E[X] = 0\cdot\frac{1}{10} + 1\cdot\frac{2}{10} + 2\cdot\frac{4}{10} + 3\cdot\frac{2}{10} + 4\cdot\frac{1}{10}$$

$$= \frac{2}{10} + \frac{8}{10} + \frac{6}{10} + \frac{4}{10} = \boxed{2}.$$

So the "expected", or average degree of a vertex in G is $\underline{2}$.

Next we determine the <u>degree distribution</u> for an <u>ER(n,p)</u> random graph.

Let $P(\delta(u)=k)$ denote the probability that degree $u \in V(G)$ equals k.

Then, $\boxed{P(\delta(u)=k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}}$

<u>Why?</u> $\binom{n-1}{k}$ counts the number of ways to choose k vertices from a set of $n-1$ possible neighbors; the p^k factor is the multiplier obtained from the k neighboring vertices, each with probability p; the $(1-p)^{n-1-k}$ factor is due to the non-neighbor set.

Classically, this formula is known as a ~~probab~~ $\boxed{\text{binomial probability distribution}}$.

Note that this formula allows us to compute the <u>degree distribution</u> for $ER(n,p)$ in its entireity.

$\boxed{\text{Ex.}}$ Find the probability a vertex in $ER(20,.3)$ has degree measure 6.

$$P(\delta(u)=6) = \binom{20-1}{6} (.3)^6 \cdot (.7)^{13} \approx \boxed{.19}$$

So there is a roughly <u>19%</u> chance a randomly chosen vertex in $ER(20,.3)$ has degree $=6$.

A deeper question related to degree distributions for ER random graphs pertains to the computation of expectation. On average, what can we expect the degree of a vertex in ER(n,p) to be?

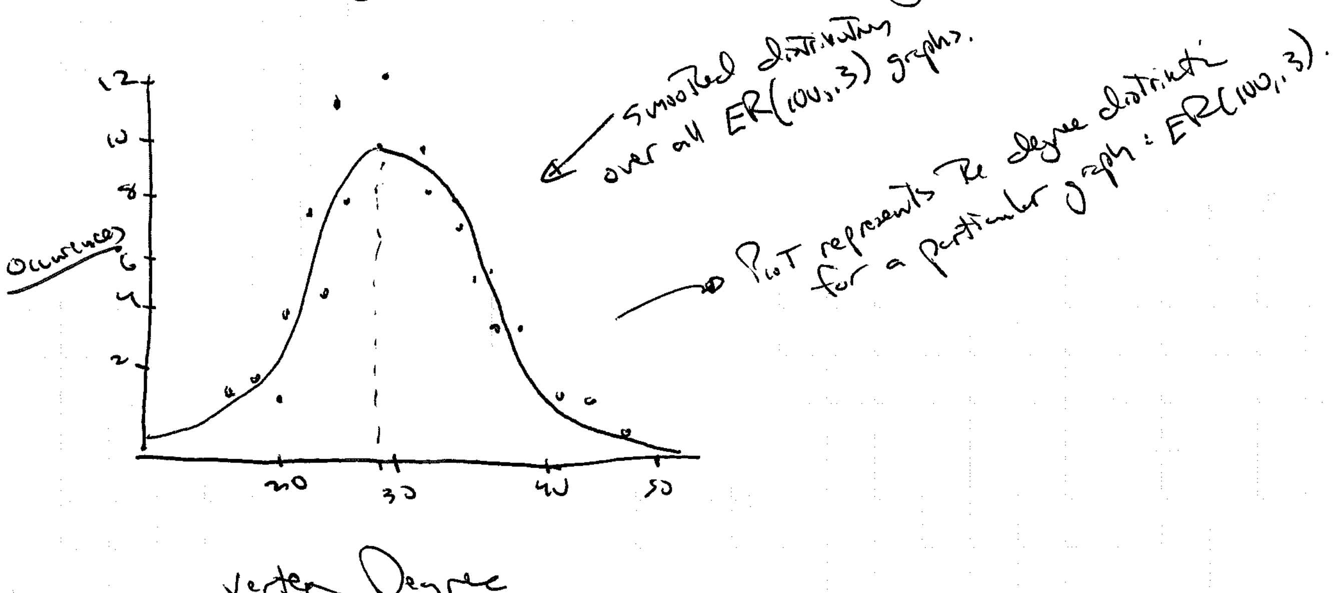
$$\boxed{E[\delta] = \sum_{k=1}^{n-1} k \cdot P(\delta = k) = p(n-1)}$$

(The proof uses several identities related to the "choose" function)

Thus for, say, ER (100,.3), we have:

$E[\delta] = .3(100-1) \approx \underline{30}$, so the mean vertex degree for ER(100,.3) is roughly 30.

In fact, it may surprise the reader to discover that if we plotted the vertex distribution, say, over all possible choices of ER(100,.3) networks, this graph would appear approximately <u>bell-shaped</u>, i.e. <u>Normal</u>.



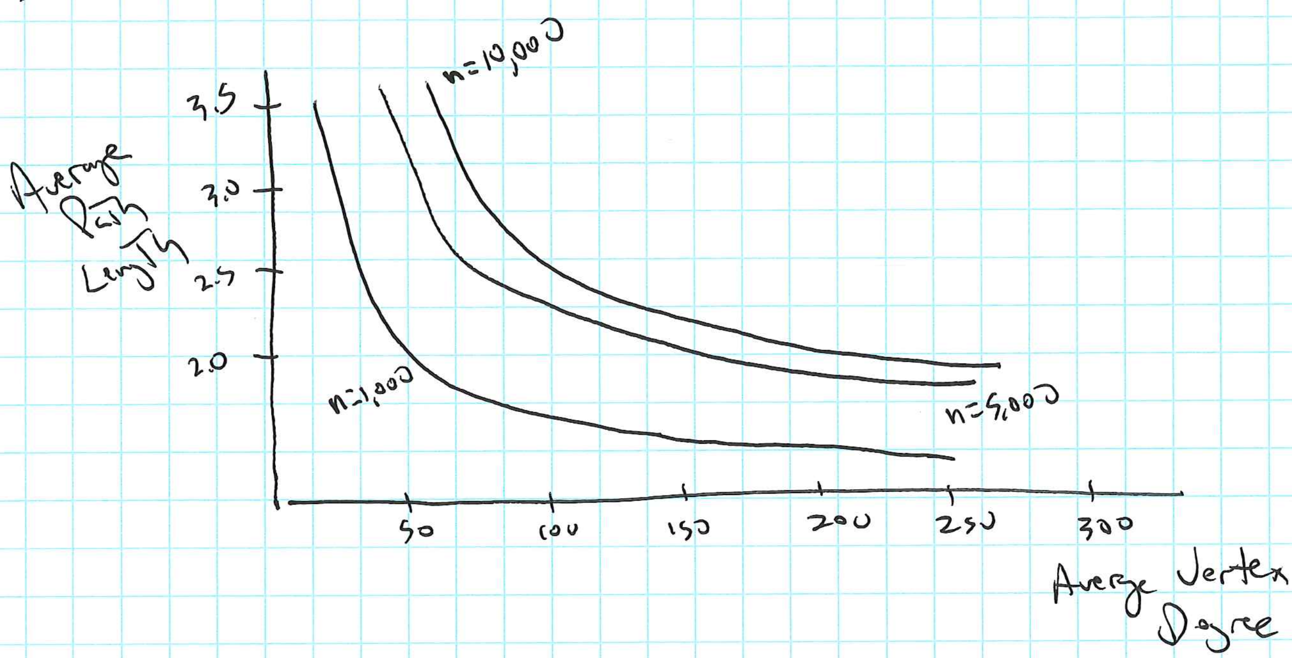
Of course, and indicated by the plot shown, a single, particular graph will not <u>exactly</u> conform to this distribution. However, for larger n (with fixed $E[\delta]$), the <u>Central Limit Theorem</u> guarantees convergence to this smoothed curve.

In addition, in 2004, Fronczak showed that for large random graphs: $G \in ER(n,p)$, the average path length can be estimated as:

$$\bar{d}(G) = \frac{\ln(n) - \gamma}{\ln(pn)} + .5 \approx \frac{\ln(n) - \gamma}{\ln(E[\delta])} + .5$$

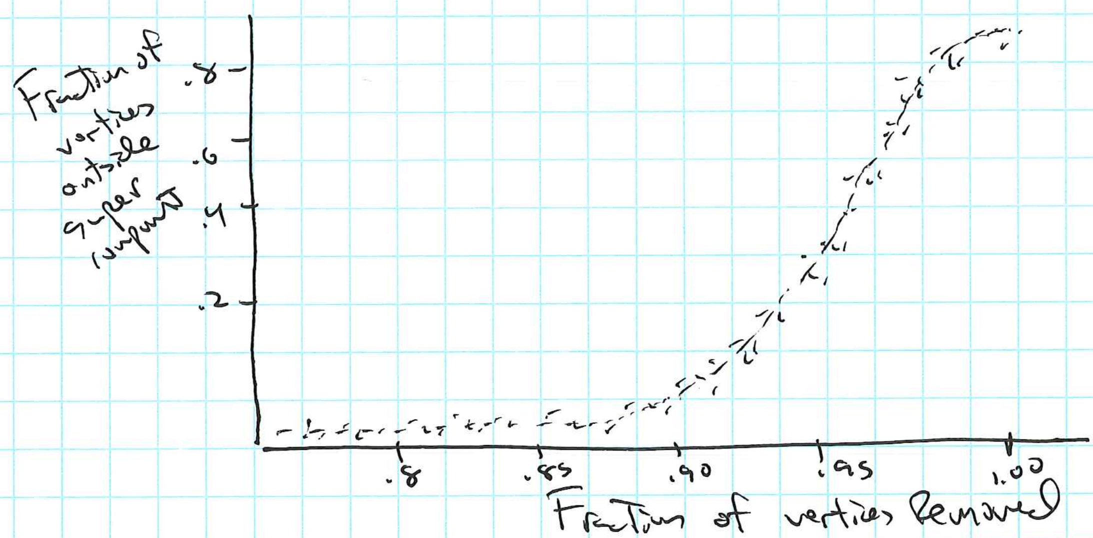
where $\boxed{\gamma}$ is the so-called $\boxed{\text{Euler constant}}$, $\gamma \approx .5572$.

This result can be used to demonstrate an important property of random complex networks - their average path length drops logarithmically (which is to say they exhibit a high level of connectivity).



As further evidence of the high level of connectivity exhibited by random networks, empirical evidence suggests that increasing

The probability parameter $\boxed{p}$ not only increases the $\boxed{\text{density}}$ of the network but there seem to emerge graphs containing an enormous, $\boxed{\text{giant/ super component}}$ (to which the majority of vertices belong), in addition to a few very small components. It is noteworthy that these super components often appear (precipitously) even for small p values. Below we show that the systematic removal of large quantities of vertices from a random graph will still leave it largely connected - in this way we can see that random graphs are remarkably resilient to large node failures.



for ER(2000, .15) graph

Additional attempts to codify a field of "complex networks" followed Erdös & Rényi's early peregrinations in the second half of the 20th century. Noteworthy was 1967 study performed by $\boxed{\text{Stanley Milgram}}$, a social psychology professor. Milgram was studying at that time the nature of social networks. Through a now-famous "experiment," Milgram asked individuals to send letters to some of their acquaintances (these were pre-email days) to determine the "degree of separation" of two randomly chosen people (Note that this notion relates to average path length in a network, as previously seen - also, such a measure is sometimes known as the diameter of a graph). Remarkably, he found that most people were linked through acquaintances & mutual acquaintances by relatively short paths - "5.5 hops", sufficed, as it turns out. This result lead of course to the colloquial "six degrees of separation" notion. In complex network theory this phenomenon is known as the small world phenomenon.

More interestingly still, the late 1990s saw a revival of research interest in complex networks. In part this renaissance in complex network theory was inspired by a great proliferation of newly "discovered" real-world networks, including the World Wide Web, the human genome, neural networks, ecological & "food webs", cellular metabolic processes, large-scale social networks, global financial networks, and air traffic networks, to name but a few examples.

At this time a watershed moment in complex network theory occurred which would spawn a new paradigm for large-scale networks. Two researchers, Albert-László Barabási & (then student) Réka Albert were attempting to build a topological model of a small subset of the WWW. Through this analysis they expected to find a random graph topology (in the vein of an ER graph). Surprisingly, though, they did not. Instead of the relatively "democratic" degree distributions common to ER graphs, this complex network ~~exhibited~~ contained a subset of vertices with extraordinarily large degrees - such vertices are called **hubs** in complex networks. In addition, while the tendency for "communities" to form in ER random graphs (sometimes called the **clustering coefficient** of the graph) is independent of the size of the vertex set, the same cannot be said for the graphs studied by Barabási-Albert. For B-A graphs, clustering becomes less prominent as the ~~te~~ vertex size increases. In one sense, then, the existence of vital hubs in the WWW network revealed that a few ~~prominent~~ highly-linked webpages were in essence holding the entire internet together (rejoice, Googles & Twitters of the world).

But perhaps this discovery was merely an anomaly – a mere artifact of the synthetically-formed network of the internet? As it turns out, this is probably <u>not</u> the case. Concurrently, the presence of hubs, small clustering & the so-called super small world effect were detected in a myriad of other "real-world" networks, including many of those previously mentioned. This consonance of complex networks that cut across disciplinary lines inspired an attempt (<u>still on-going</u>) to uncover a <u>well-defined architecture</u>, or a <u>set of fixed "laws"</u> that would seemingly apply across the board to networks of <u>cells</u>, <u>computers</u> & we dare say, the <u>Kevin Bacons</u> of the world. Perhaps, the discovery of such organizing principles around complex networks would lend itself to, among other applications: <u>securing power networks against catastrophic failures</u>, <u>halting the spread of epidemics</u>, <u>combatting cancer</u> & better <u>understanding the brain</u>.

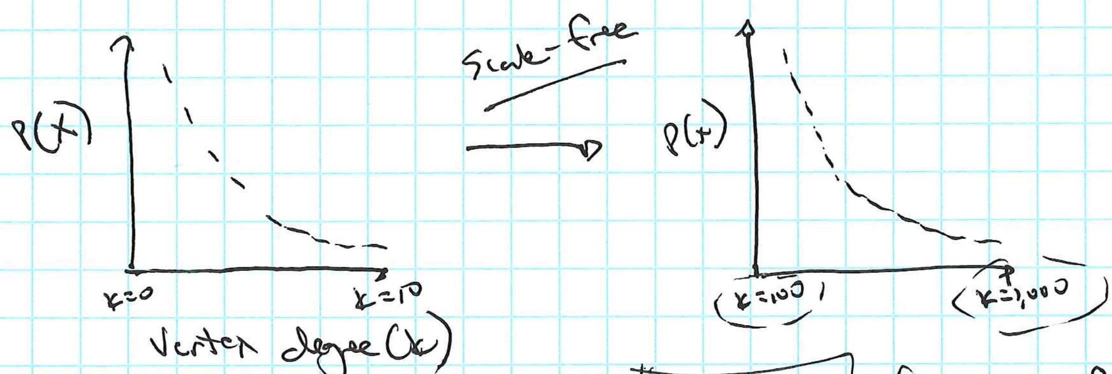
While the study of complex networks is still, we think, in its infancy, many of the key characteristics described above are captured in what have come to be known as scale-free networks. We describe these networks & their mathematical properties below.

A network is called $\boxed{\text{scale-free}}$ if the vertex degree distribution follows a $\boxed{\text{power law}}$. This means that the probability that an arbitrary node has degree k is proportional to $(1/k)^{\alpha}$ for some $\alpha > 1$. Put mathematically,

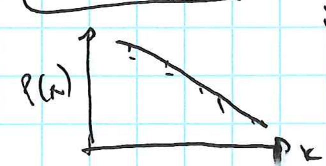
$$\boxed{P(k) \underbrace{\propto}_{\text{"proportional to"}} \left(\frac{1}{k}\right)^{\alpha}} \qquad \alpha > 1$$

(say if $\alpha = 2.5$, then
$P(2) \approx .176$
$P(10) \approx .003$)

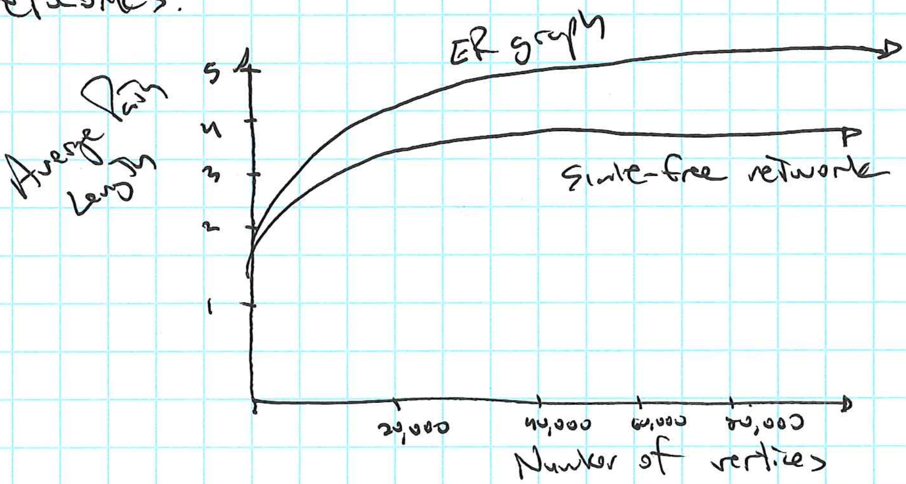
In practice, most real-world complex networks obey this power law for $\boxed{2 < \alpha < 3}$. The power law distribution guarantees a asymptotically long-tailed distribution with a high left-end peak, which accounts for the presence of the aforementioned $\boxed{\text{hub}}$ vertices. These networks are called scale-free because the form of their degree distribution basically looks the same, regardless of scale.



Note the $\boxed{\text{log-log plot}}$ for scale-free networks is roughly linear.



Importantly, as was mentioned, scale-free networks tend to exhibit a super small world property surpassing even that of random graphs. This disparity can be seen directly when we compare <u>average path lengths</u> in random graphs & scale-free networks.



(Note: The internet consists of over <u>3 billion</u> webpages, most of which are on average no more than <u>19</u> "clicks" away from one another.)

Many researchers have recently begun to investigate & question the reason behind the seeming proliferation of scale-free networks in real-world applications. A few cogent explanations have surfaced. If, for instance, we consider the "evolutionary" development of the internet – new webpage links are not all created equally. Old nodes, particularly, <u>hubs</u> with dense connectivity tend to acquire new links at a higher frequency than do newer nodes. The penchant for hubs to attract a greater volume of links in this fashion is termed preferential attachment in complex network theory.

Another key insight related to the "why" of scale-free networks ~~Lies~~ relates to the resilience of scale-free networks

To catastrophic (random) node failure. In the case, for instance, of living organisms, we would hope that "failures"/mutations of some cells does not lead to a total system failure (e.g. the development of cancer). Indeed, for many scale-free networks, the random failure of up to 80% of all the nodes in the system can still leave the network fully connected (i.e. functioning). However, if, say a malicious (& intelligent) agent specifically attacks a small number of vertices of high importance (i.e. hubs), a scale-free network is in general very susceptible to these sorts of attacks. Current research in, for example, cancer metastasis, economic market vulnerabilities, internet security attacks, and the spread of infectious diseases, has all profited from the analysis of scale-free & complex networks.