CITS4403
Computational Modelling

Lecture 3: Graph II

Dr. Siwen Luo

Semester 2, 2024
School of Computer Science,
University of Western Australia

0 LECTURE PLAN

THE UNIVERSITY OF WESTERN AUSTRALIA

Lecture 3: Graph II

1. **Recap**
2. Small world phenomenon
   - Regular Graph
   - Watts-Strogatz Graph and Experiment
   - Implementation
3. WS Graph on Real Social Network Data
   - Degree Distribution
   - Heavy-tailed Distribution
4. Scale-free Network
   - BA model

THE UNIVERSITY OF WESTERN AUSTRALIA

Random Graph

A graph with nodes and edges generated from random.

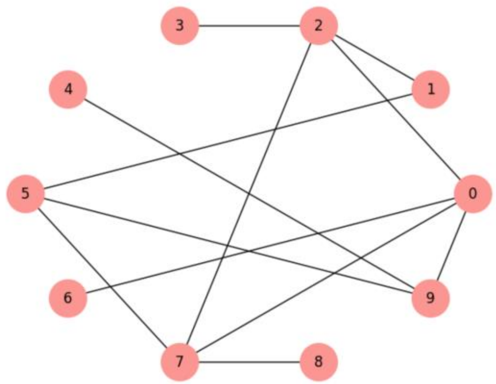
Erdos Renyi (ER) Graph:

$$G(n, p)$$

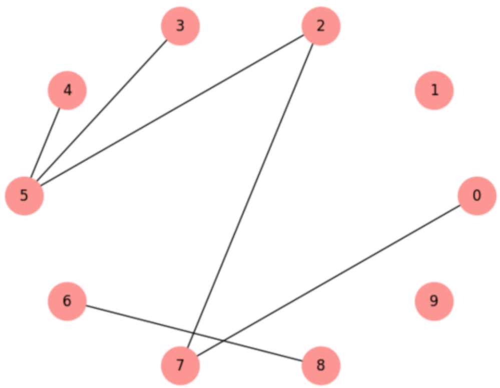
n is the number of nodes and p is the probability that there is an edge between any two nodes.

Note that this is not the same format as we just saw with $G = (V, E)$. This is saying that the Graph, G, is a function of n and p. i.e. it depends on these two parameters that we need to choose.

1 Recap



An ER random graph of $n = 10, \boldsymbol{p} = \mathbf{0.3}$



An ER random graph of $n = 10, \boldsymbol{p} = \mathbf{0.1}$

THE UNIVERSITY OF WESTERN AUSTRALIA

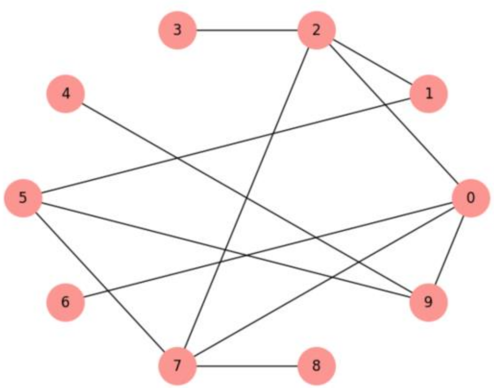
Graph Property

Connectivity

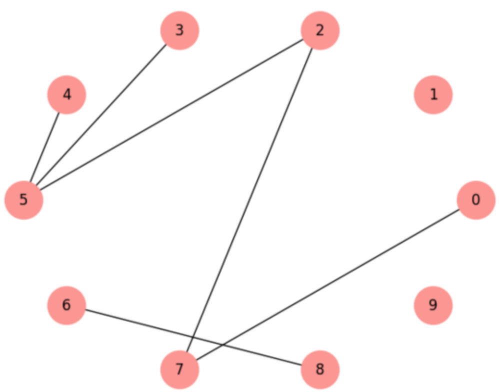
An undirected graph is ***connected*** *if there is a path from every node to every other node.*

Path: a sequence of nodes with an edge between each consecutive pair.

Start at any node and whether you can reach all other nodes.



Connected



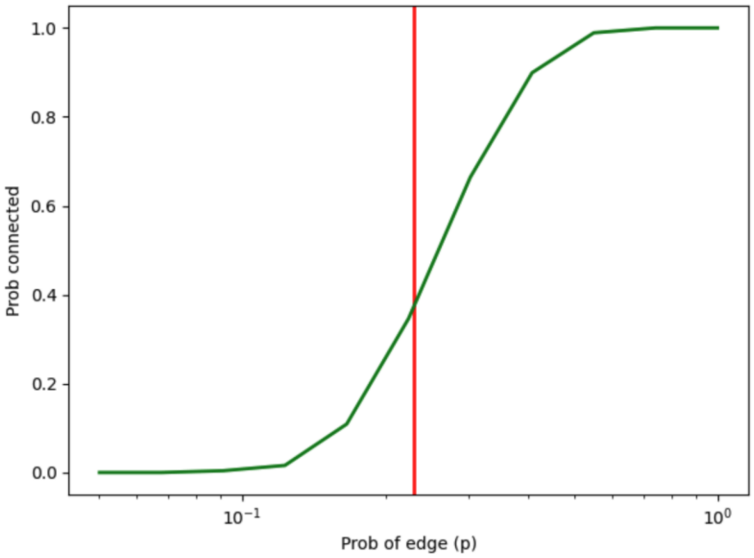
Not connected

THE UNIVERSITY OF WESTERN AUSTRALIA

Connectivity

$\boldsymbol{G(n,p)}$ is unlikely to be connected if $\boldsymbol{p < p^*}$ and very likely to be connected if $\boldsymbol{p > p^*}$

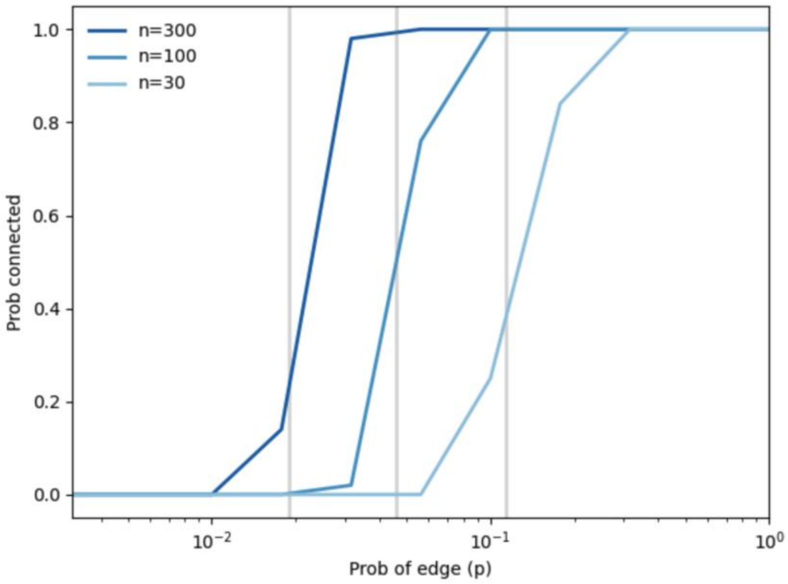
Run experiments to estimate by generating a large number of random graphs and counting how many are connected.



Connectivity

$G(n, p)$ is unlikely to be connected if $p < p^*$ and very likely to be connected if $p > p^*$

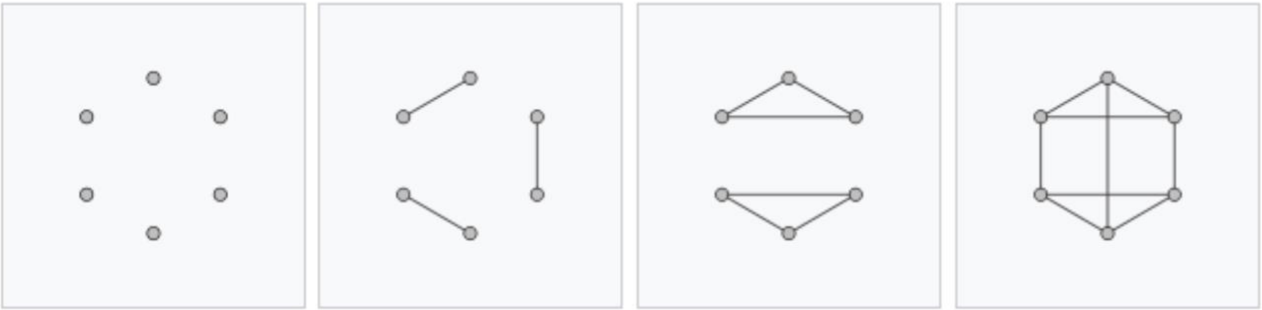
As n increases, the critical value p^* gets smaller, and the transition gets more abrupt.



1 Recap

Regular graph is a graph that each node has the **same** number of neighbours, or every node has the same degree.

k*-regular graph or regular graph of degree *k: a regular graph with vertices of degree k.



0-regular graph 1-regular graph 2-regular graph 3-regular graph

THE UNIVERSITY OF WESTERN AUSTRALIA

Random Graph vs. Regular Graph

Random Graph	Regular Graph
Low clustering Low path lengths	High clustering High path length

Neither is good for modelling graphs of high clustering and short path length

0 LECTURE PLAN

THE UNIVERSITY OF WESTERN AUSTRALIA

Lecture 3: Graph II

1. Recap
2. **Small world phenomenon**
   - Regular Graph
   - Watts-Strogatz Graph and Experiment
   - Implementation
3. WS Graph on Real Social Network Data
   - Degree Distribution
   - Heavy-tailed Distribution
4. Scale-free Network
   - BA model

Small World Phenomenon

Small social world

We expect social networks to be localised — people tend to live near their friends — and in a graph with local connections, path lengths tend to increase in proportion to geographical distance.

But we live in a much smaller world than expected.

2 Small World Phenomenon

American social psychologist, best known for his controversial experiments on obedience conducted in the 1960s during his professorship at Yale.

His other small-world experiment, while at Harvard, led researchers to analyse the degree of connectedness, including the ***six degrees of separation*** concept.

Stanley Milgram

(August 15, 1933 – December 20 1984)

https://www.youtube.com/watch?v=a99ry70CnRs

Watts & Strogatz Graph

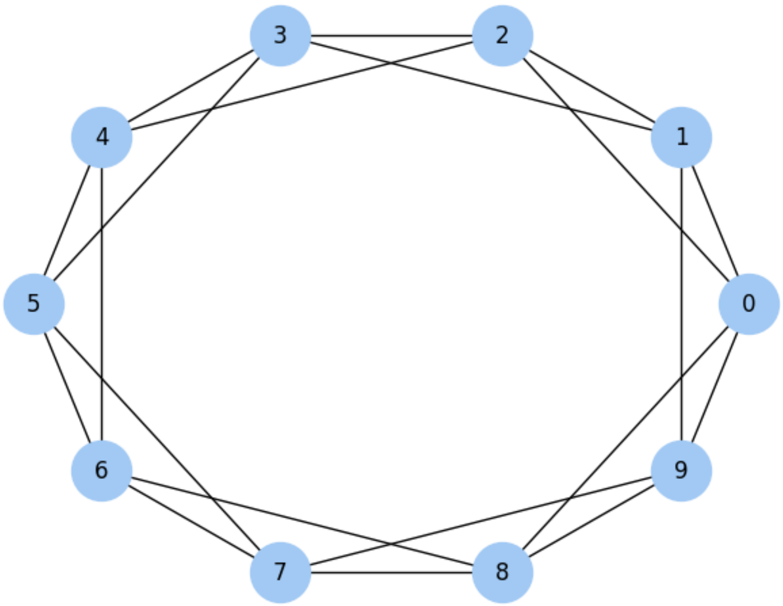
Watts and Strogatz tried to explain the **small world phenomenon** by modelling the process that leads to the phenomenon and building a small-world graph, **Watts-Strogatz (WS) Graph.**

- Start with a **regular graph** with n nodes and each node connected to k neighbours.
- Choose a subset of the edges and “**rewire**” them by replacing them with **random** edges.
- The probability that an edge is rewired in this process is controlled by a **parameter** $\boldsymbol{p}$

Purpose: The WS experiment is to find a region where $0 < p < 1$ to let the graph have **high clustering and low path length.**

1. Start with a Regular Graph

Ring lattice: a regular graph of n nodes that are arranged in a circle with each node connected to the k nearest neighbours.



Watts & Strogatz Experiment

2. Rewire Edges

Go through each edge of the ring lattice graph, keep source node fixed and rewire destination node. No self-loops or multiple edges between nodes are allowed.

3. Compute Clustering Coefficient & Path Length

There are two characteristics to quantify the small world property, **high clustering and low path lengths.**

Review of Clustering Coefficient

Clustering Coefficient

To quantify the tendency for the nodes to form cliques, which is a set of nodes and there are edges between all pairs of nodes.

For a node u, has k neighbours. If all the neighbours are connected to each other, there would be $k(k-1)/2$ edges among them. The local clustering coefficient for u, denoted as C_u is the **fraction of those edges that actually exist.**

The graph average clustering coefficient, $\bar{C}$, is the average of C_u over all nodes.

Path Length

Average length of the shortest path between each pair of nodes.

THE UNIVERSITY OF WESTERN AUSTRALIA

WS Graph:

$$G(n, k, p)$$

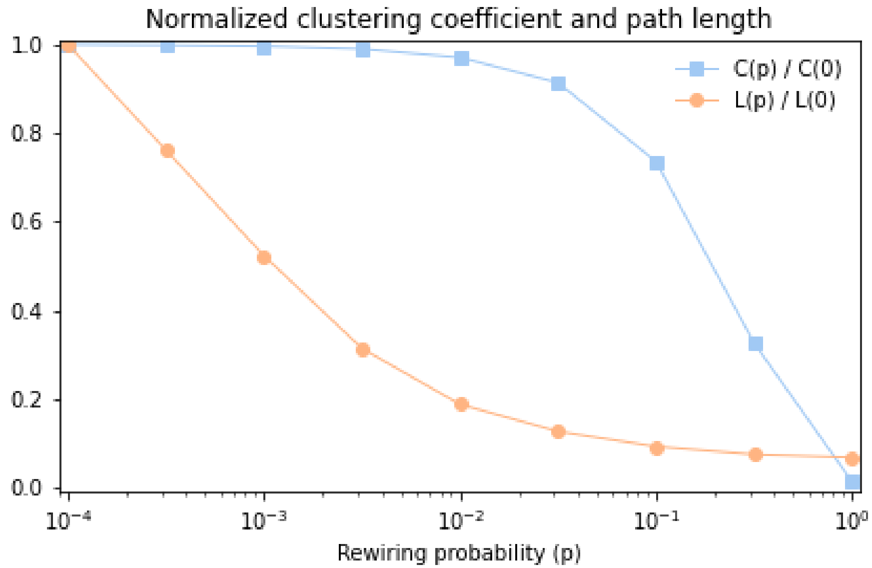
n: the number of nodes
k: degree of each node
p: the probability to rewire the edge between a pair of nodes

Note that this is saying that the Graph, G, is a function of n, k and p. i.e. it depends on these three parameters that we need to choose.

Watts & Strogatz Experiment

Run WS Experiment:

with a range of p

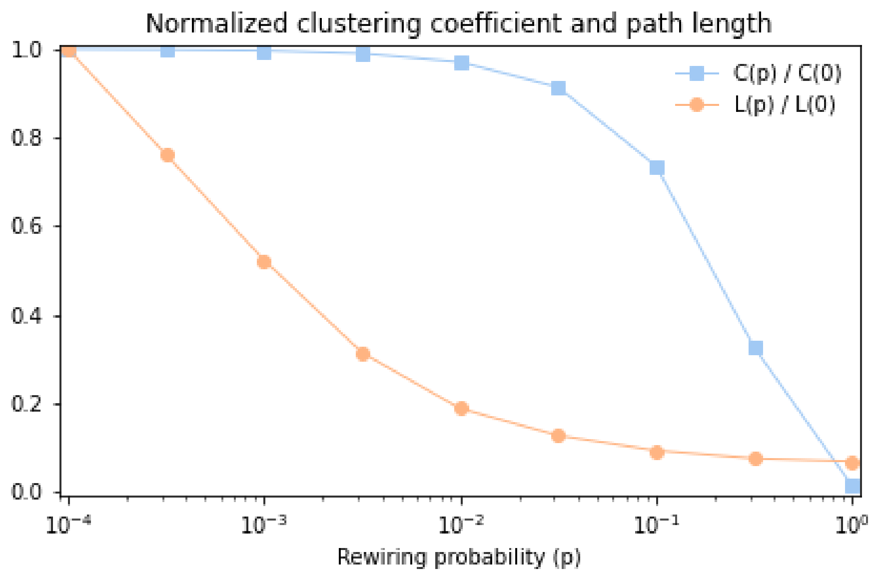


There is a wide range of p where a WS graph has the properties of a small world graph, high clustering and low path lengths.

Watts & Strogatz Experiment

Run WS Experiment:

with a range of p



That's why the proposed WS graphs can be used as a model for real-world networks that exhibit the small world phenomenon.

THE UNIVERSITY OF WESTERN AUSTRALIA

Lecture 3: Graph II

1. Recap
2. Small world phenomenon
   - Regular Graph
   - Watts-Strogatz Graph and Experiment
   - Implementation
3. **WS Graph on Real Social Network Data**
   - Degree Distribution
   - Heavy-tailed Distribution
4. Scale-free Network
   - BA model

Facebook Social Network Data

How WS Graph performs on Real-world data?

Let’s try with data of modern social network from the real-world of Facebook, check whether this dataset has the characteristics of a small world graph: high clustering and low path lengths.

SNAP Facebook Data:

- Consists of circles/friend lists from Facebook collected from survey participants.
- **4039** nodes
- **88234** edges

https://snap.stanford.edu/data/egonets-Facebook.html

3 Facebook Social Network Data

WS Graph:

$$G(n, k, p)$$

n: the number of nodes
k: degree of each node
p: the probability to rewire the edge between a pair of nodes

In SNAP Facebook Data:

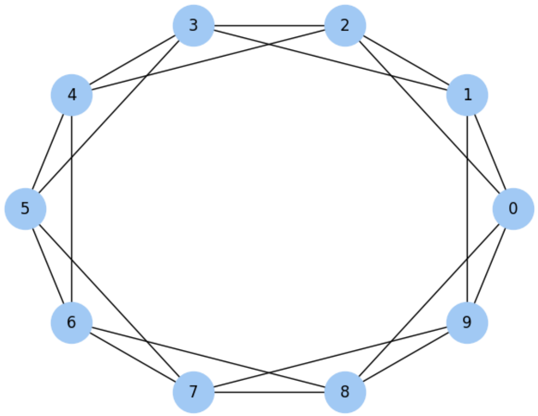
$n = 4039, m = 88234$
$\boldsymbol{k}$**: ?**
p: can be a range of probability to test

Estimate k

To estimate a constant k (degree of node), we can take the value as the average degree of node of the graph.

Average degree is the **sum of degree** divided by the **number of nodes**.

When count the sum of degree (of all nodes), **an edge will be counted twice.**



E.g. ring lattice of 10 nodes, the sum of degree is 40.
(the degree of each node is 4).

Estimate k

To estimate a constant k (degree of node), we can take the value as the average degree of node of the graph.

Average degree is the **sum of degree** divided by the **number of nodes**.

When count the sum of degree (of all nodes), **an edge will be counted twice.**

In SNAP Facebook Data:

Sum of degree (of all nodes) = $2m$

Average degree = $2m/n$

$$\boldsymbol{k} = 2 \times 4039/88234 = 44$$

WS Graph on SNAP Facebook Data:

$$G(n, k, p)$$

n: 4039
k: 44

We can again run the experiment over different p values to check if this graph models the small world properties well by showing high clustering values and small path lengths.

THE UNIVERSITY OF WESTERN AUSTRALIA

WS Graph on SNAP Facebook Data:

$$G(n, k, p)$$

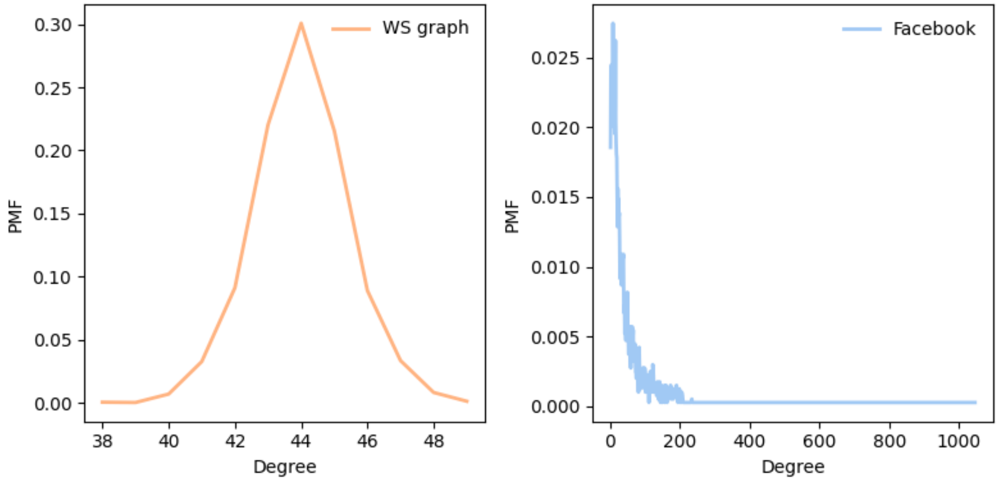
n: 4039
k: 44

We can again run the experiment over different p values to check if this graph models the small world properties well by showing high clustering values and small path lengths.

In experiment, when p is 0.05, the build WS graph has average clustering coefficient around 0.6 and average path length around 3.6, **which models the small world properties well.**

But…

If we plot the degree distribution of the actual dataset and the constructed WS Graph, $G(4039, 44, 0.05)$



If the WS graph is a good model for the Facebook network, it should have the same average degree across nodes, and ideally the same variance in degree.

THE UNIVERSITY OF WESTERN AUSTRALIA

Probability Mass Function (PMF)

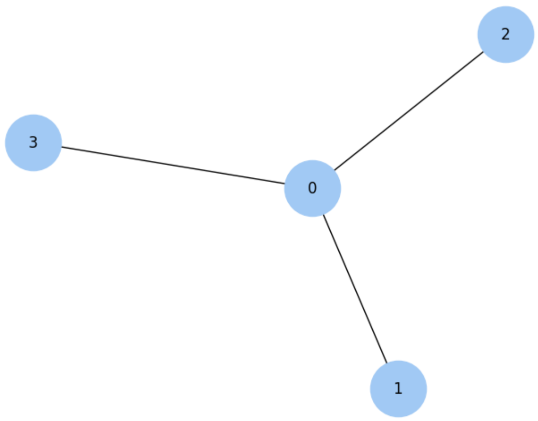
PMF maps from values to their probabilities

A PMF of degree maps from each degree to the fraction of nodes with that degree.

E.g. if we have a graph with nodes *1, 2, 3* connected to a central node, *0*, then:

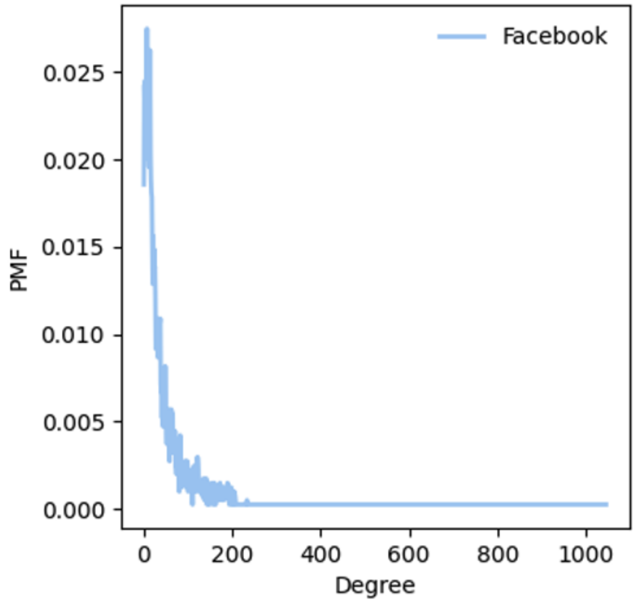
degree of node 0 is **3**, degree of node *1, 2, 3* is **1**

In this example, 75% of the nodes have degree 1 and 25% have degree 3.



3 Heavy-tailed Distribution

Distributions like this, with many small values and a few very large values, are called **heavy-tailed**.



Heavy-tailed Distribution

Among all heavy-tailed distributions, there is one type: **power law distribution**.

Mathematically, a distribution obeys a power law if:

$$PMF(k) \sim k^{-\alpha}$$

Where α is a parameter, and the symbol $\sim$ indicates that the *PMF* value of degree k is asymptotic to $k^{-\alpha}$ as k increases.

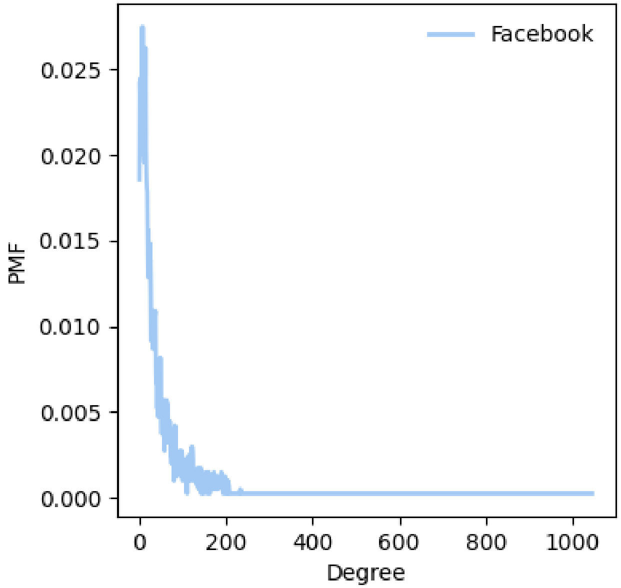
If we take the log of both sides, we get:

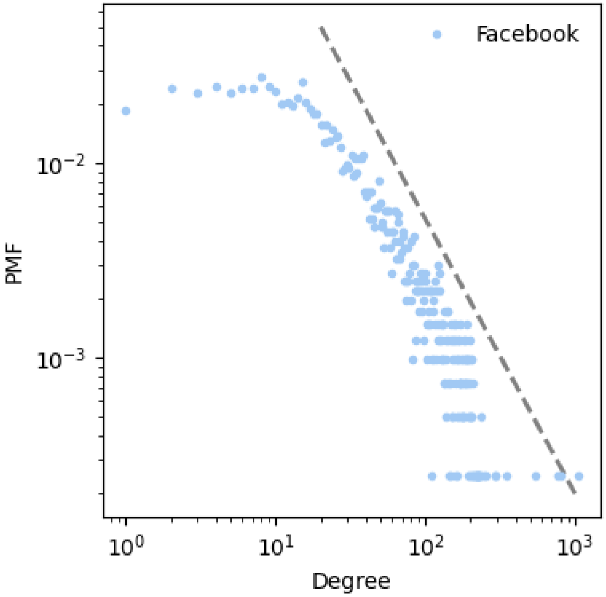
$$\log PMF(k) \sim -\alpha \log k$$

If a distribution follows a power law and we plot $PMF(k)$ versus k on a log-log scale, we expect a straight line with slope $-\alpha$**, at least for large values of k.**

Heavy-tailed Distribution

If we plot $PMF(k)$ on **log-log axis**, we have:

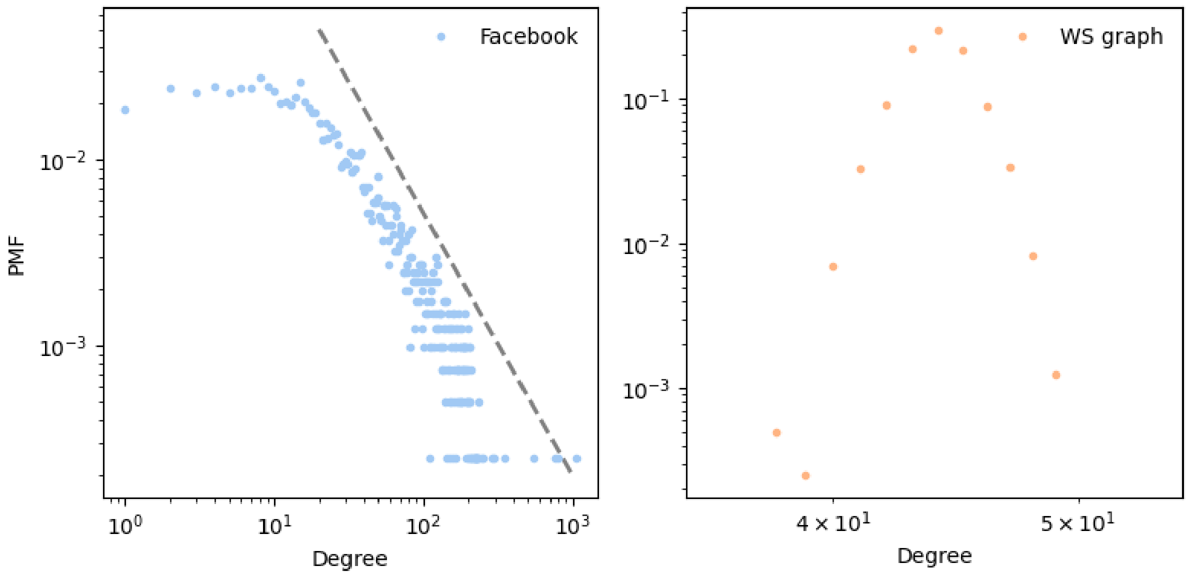




Under this transformation, the Facebook data fall approximately on a **straight** line, which suggests that there is a **power law** relationship between the largest values in the distribution and their probabilities.

Heavy-tailed Distribution

If we plot $PMF(k)$ on **log-log axis**, we have:



However, the built WS Graph cannot reveal the power law distribution as in the actual Facebook dataset.

THE UNIVERSITY OF WESTERN AUSTRALIA

Lecture 3: Graph II

1. Recap
2. Small world phenomenon
   - Regular Graph
   - Watts-Strogatz Graph and Experiment
   - Implementation
3. WS Graph on Real Social Network Data
   - Degree Distribution
   - Heavy-tailed Distribution
4. **Scale-free Network**
   - BA model

Scale-free Network

A **scale-free network** is a network whose degree distribution follows a **power law**, at least asymptotically.

Barabasi & Albert (BA) model: generate random graphs with the scale-free property.

Different from WS Graph:

- **Growth** -- Instead of starting with a fixed number of vertices, the BA model starts with a small graph and adds nodes one at a time.
- **Preferential attachment** -- When a new edge is created, it is more likely to connect to a node that already has large number of edges.

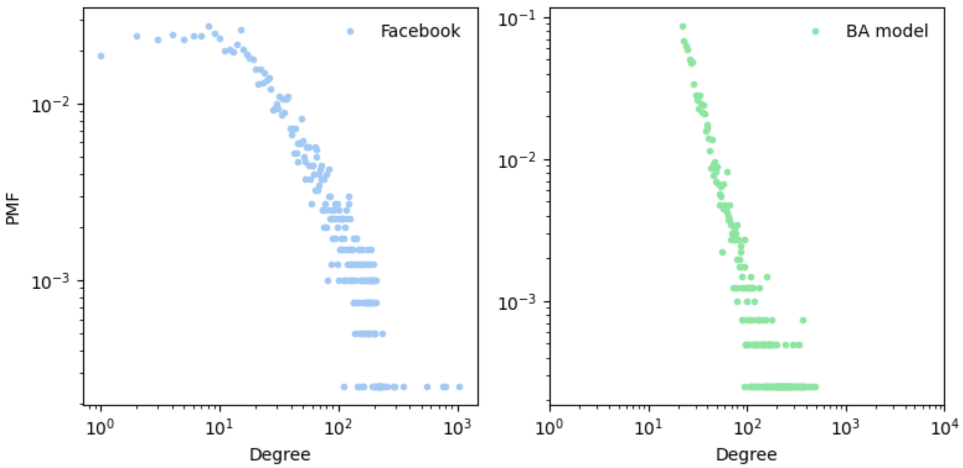
Scale-free Network

Build BA model and check scale-free property:

1. Build BA graph $G(n, m)$
   n: number of nodes to generate;
   m: the number of edges each node starts with when it is added to the graph.

2. Measure the degree of each node and compute $PMF(k)$

3. Plot $PMF(k)$ of k on a log-log axis



WS model vs. BA model

WS model	BA model
• High clustering • Short path length • Power law distribution ✗	• Low clustering • Short path length • Power law distribution ✓

THE UNIVERSITY OF WESTERN AUSTRALIA

WS model vs. BA model

BA model and WS model both explain aspects of small-world behavior, but they offer different explanations.

WS model	BA model
Social networks are 'small' because they include clusters and have short path length.	Social networks are small because they include nodes with high degree that act as hubs, and hubs grow over time due to preferential attachment.

NEXT WEEK PREVIEW…

Lecture 4: Cellular Automata I

- Cellular Automata
  - 1D CA
- Rules of 1D CA
  - Classes of different rules