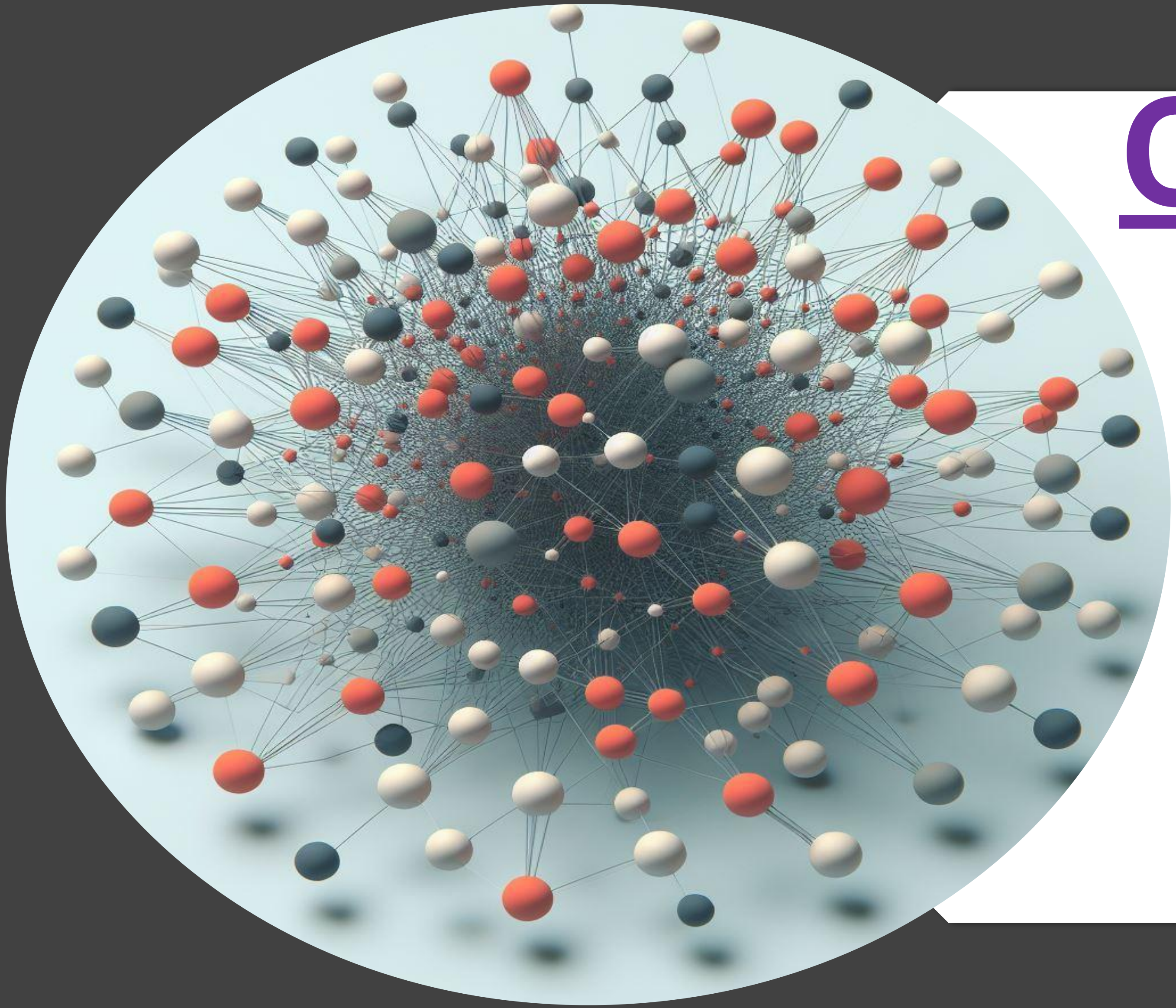


CASE STUDY ON

RANDOM GRAPH MODELS

PRESENTATION OUTLINE

- What properties does real networks have ?
- Revisiting ER-model (Erdos & Renyi, 1959)
- What ER-model support and does not support?
- Preferential attachment (Barabasi & Albert (BA) model, 1999)
- What BA model supports and does not support ?
- Summary

Clustering coefficient CC(v):

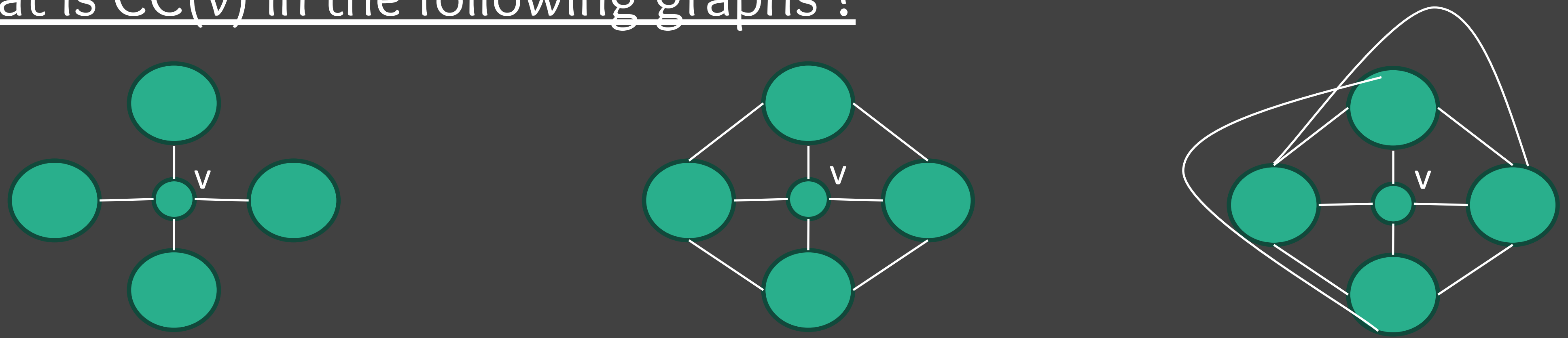
Clustering coefficient of a vertex v indicates how close it neighbours are to being a complete clique.

$$CC(v) = \frac{2 * e(v)}{\deg(v) * (\deg(v) - 1)}$$

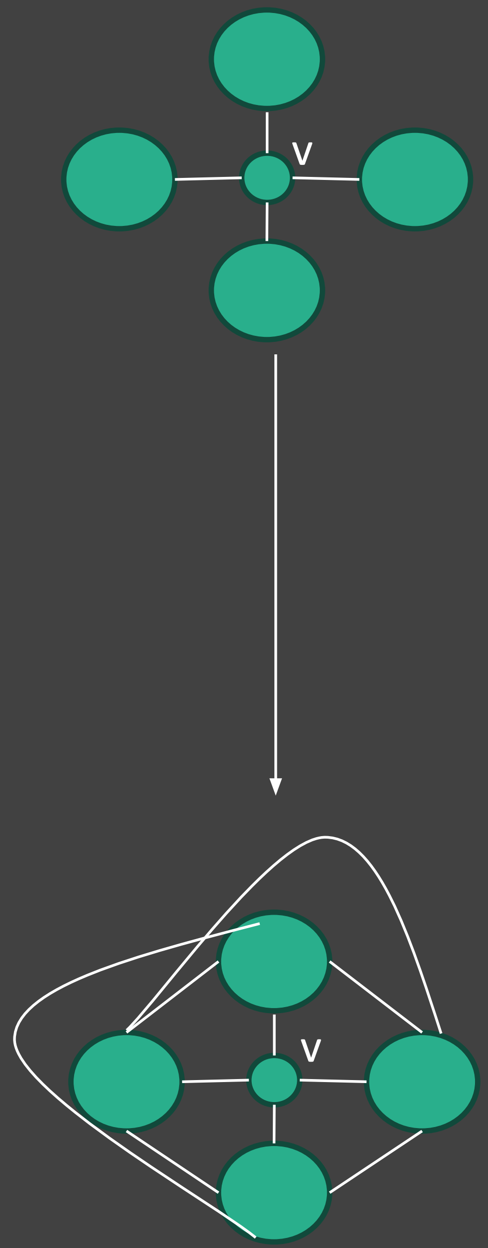
e(v) denotes the number of edges between all neighbours of v.

Note that isolated vertices are ignored and CC of degree 1 vertices is defined to be 0.

What is CC(v) in the following graphs ?



CC(ν) = 0



CC(ν) = 1

Note that:
0 <= CC(ν) <= 1

Clustering coefficient of a graph G is the average of all the clustering coefficients of all the vertices in a graph.

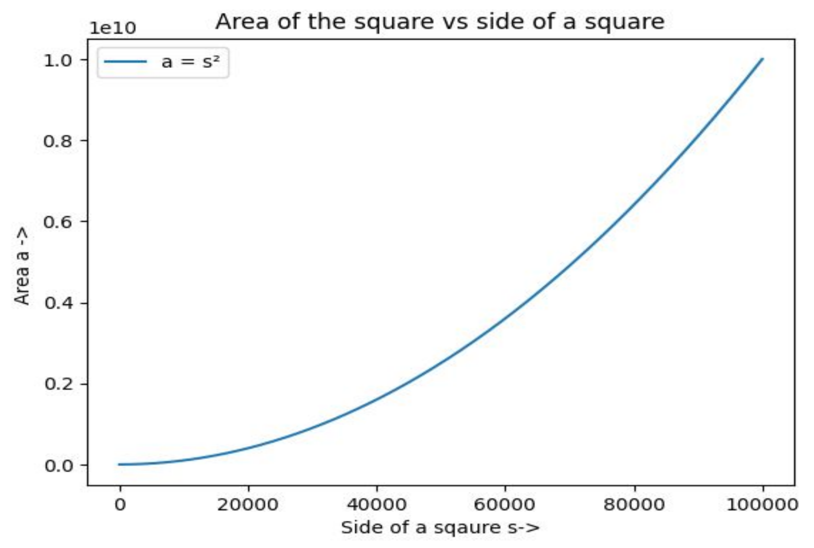
Power law:

Power law is a functional relationship between two quantities, where one quantity varies as a power of another quantity. Of the form:

$$y = k\,x^{a}$$

- ➔ where y and x are quantities which are said to be related by power law.
- ➔ k is a constant
- ➔ a is a power law exponent.

Simple example:



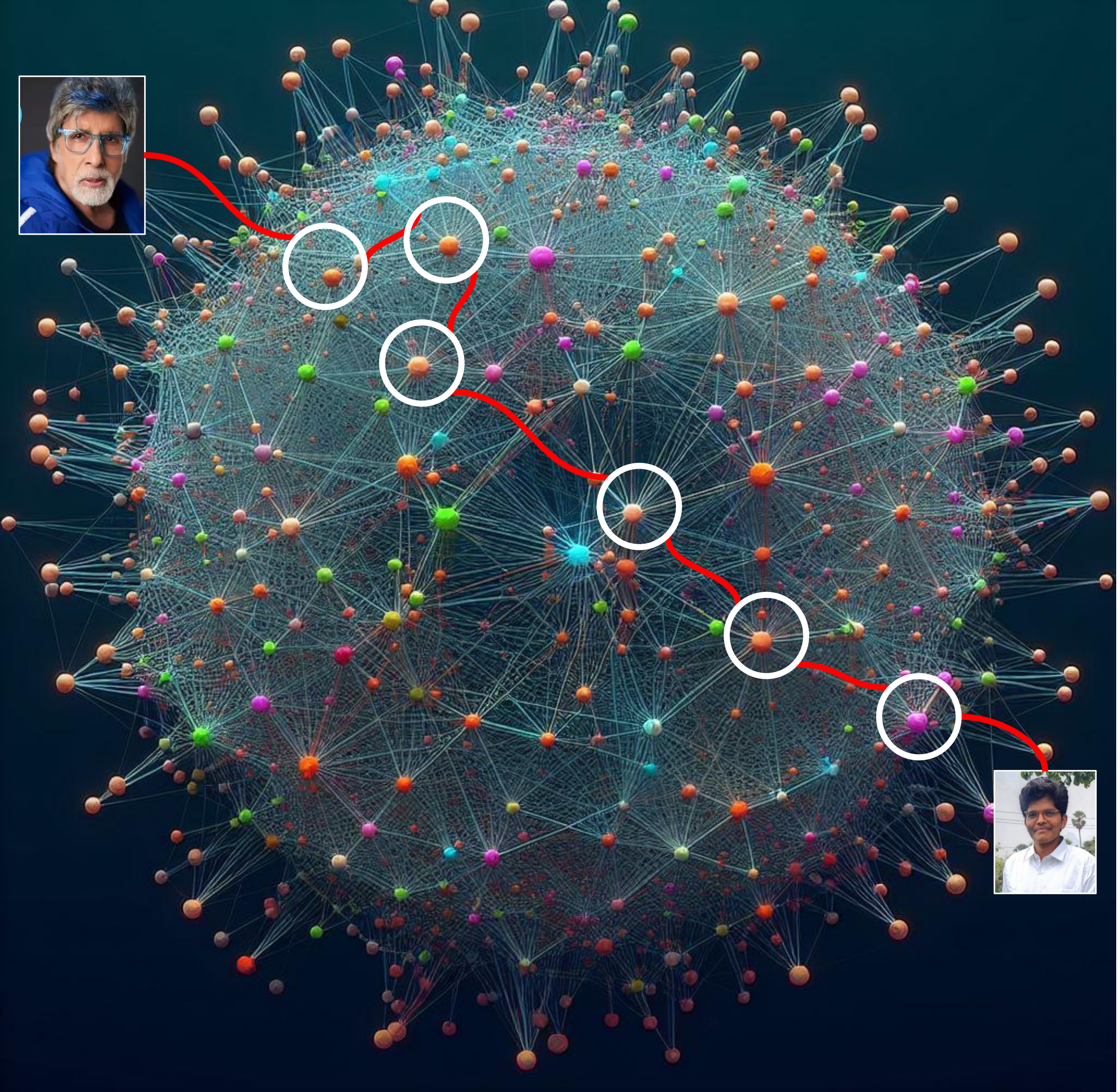
Degree distribution of graph G(V, E):

Degree distribution refers to the probability distribution of the degrees of vertices in the graph. P(k) = probability that a vertex has degree k

Power law graphs:

In a power law graph, the degree distribution follows a power law.

$$P(k) \propto k^{-\gamma}$$



Small world property:

Despite having large network size, most nodes can be reached from any other node through a relatively small number of connections !! (many times can be expressed in terms of logn)

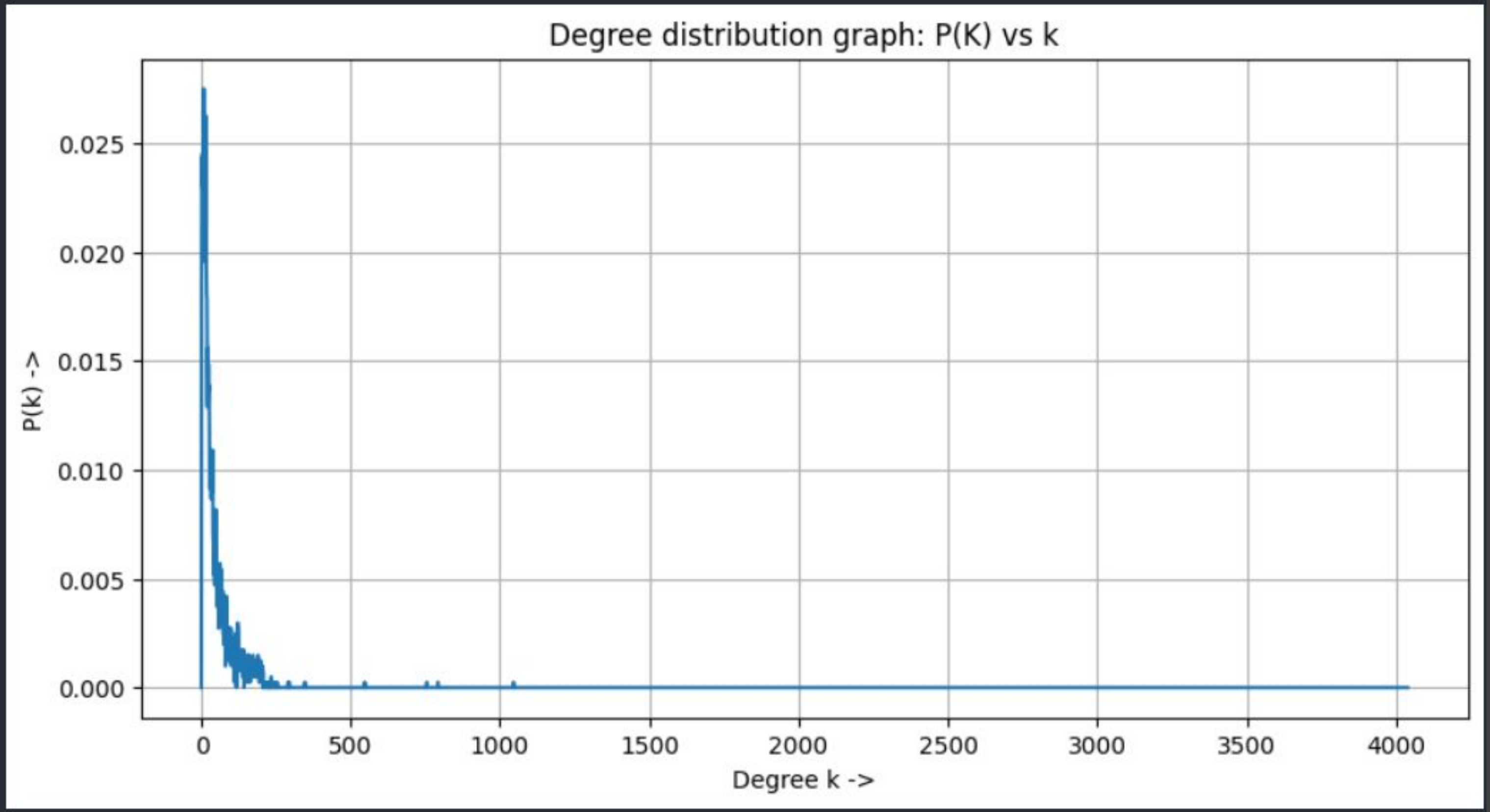
Six degrees of separation:

Any two people in the world can be connected by a chain of no more than **six** intermediaries on average.

Real networks:

1) Real networks have high clustering coefficient (i.e many clusters, triangles)
2) Real networks hold power law.
3) Real networks hold small world property.

```
Number of vertices in the graph is n =  4039
Number of edges in the graph is n =  176468
Average distance between any two nodes is =  3.6925068496963913
Clustering coefficient of graph G =  0.6055467186200876
```



[Link->]

Erdos-Renyi random graphs:

G(n, p) model

n - number of nodes

p - probability that a pair of nodes has edge.

G(n, m) model:

1. G(n, m) is a randomly selected graph from the set of $\binom{\frac{n * (n-1)}{2}}{m}$ graphs.
2. Erdos showed that when n -> ∞, (i.e for large graphs) G(n, m) is same as G(n, p) model !!
3. **Important thing to note is that G(n, m) is computationally intensive than G(n, p) model.**

ER model

1) **Does not hold power law on degree distribution.**

   Binomial / Poisson distribution

2) Holds small world property

   Average path length L ~ $\ln n / \ln(\langle k \rangle)$

3) **Clustering coefficient**

   C = p = <k>/n

Link->

Real world networks like citation networks, social networks, web networks evolve over time, that is dynamic in nature. (this cannot be captured by ER model) This is the motivation to discover preferential attachment model. (Hoping to hold real world properties than ER model)

Preferential attachment (Barabasi & Albert (BA) model, 1999)

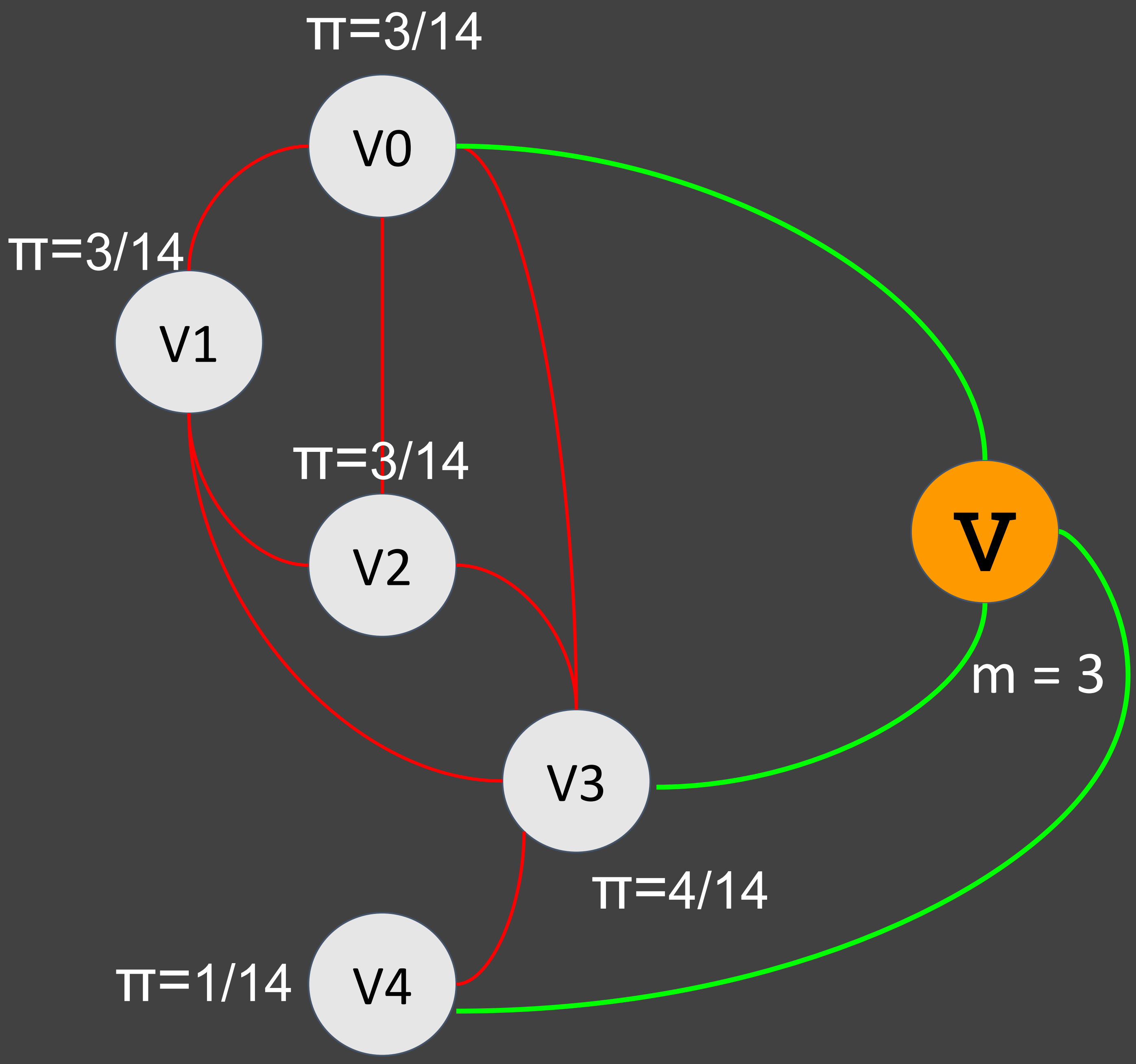
Let at t = 0, the graph G has n_0 nodes and $m_0 (>= n_0)$ edges

Growth:

At the beginning of each second add a new node with **m** edges ($m <= n_0$) connecting to m already existing nodes

Preferential attachment:

The probability of linking to existing node i is proportional to the node degree k_i (k_i denotes the degree of node i)



$$\Pi(k_i) = \frac{k_i}{\sum k_i}$$

Yay!!
probability

After t seconds (or timesteps):

Number of nodes in the graph = $t + n_0$

Number of edges in the graph = $mt + m_0$

Note that this is a growing model. But can be modified to shrink as well (i.e deleting nodes (or) some edges at some random time)

Degree of a node at any time is dependent on the time and configuration of graph at time 0.

Let $K_i(t)$ be the expected value of degree of node i at time t. For ease of calculation let $m_0 = 0$, $n_0 = 0$ (which is okay when a network starts, standard assumption)

$$K_i(t + \delta t) = K_i(t) + m\pi(K_i)\delta t$$

$$K_i(t + \delta t) - K_i(t) = m\pi(K_i)\delta t$$

$$\frac{\mathrm{d}k_i(t)}{\mathrm{d}t} = m\frac{k_i(t)}{\sum k_i} = m\frac{k_i(t)}{2mt} = \frac{k_i(t)}{2t}$$

Let node i is added at time ti : Ki(ti) = m

$$\int_{m}^{k_i(t)} \frac{\mathrm{d}k_i}{k_i} = \int_{t_i}^{t} \frac{\mathrm{d}t}{2t}$$

$$\Rightarrow \ln\left(\frac{k_i}{m}\right) = \frac{1}{2}\ln\left(\frac{t}{t_i}\right)$$

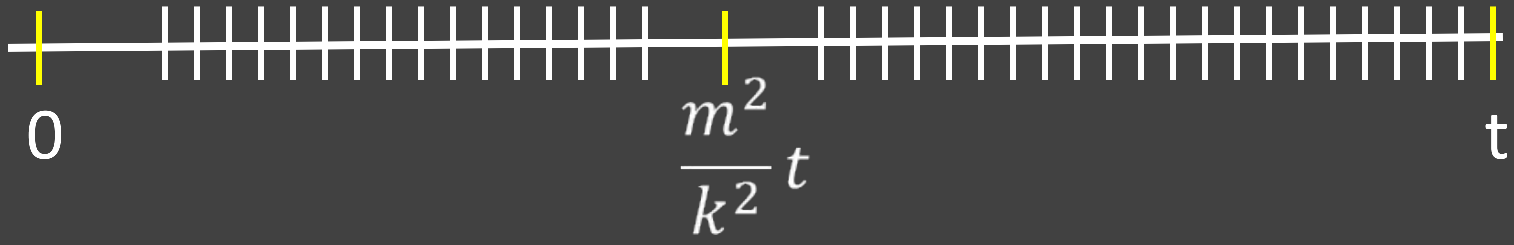
$$\Rightarrow k_i\ (t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}}$$

Let P(Ki(t) <= K) be the probability that a randomly selected node to have Ki(t) <= K at time . (Which is in fact degree distribution)

$$\Rightarrow k_i\ (t) \leq \mathrm{K}$$

$$\Rightarrow\ m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} \leq K$$

$$\Rightarrow\ t_i \geq \frac{m^2}{k^2}t$$



What if K < m ??

Let us find cumulative distributive function F(k) = P(Ki(t) <= K)

$$F(k) = p(k_i(t) \leq k)$$

$$F(k) = \frac{t - \frac{m^2 t}{k^2}}{t} = 1 - \frac{m^2}{k^2}$$

$$p(k) = \frac{\mathrm{d}}{\mathrm{d}k} F(k) = 2\frac{m^2}{k^3}$$

What we proved??

BA model

1) Holds power law on degree distribution.

$$P(K) = 2\frac{m^2}{k^3}$$

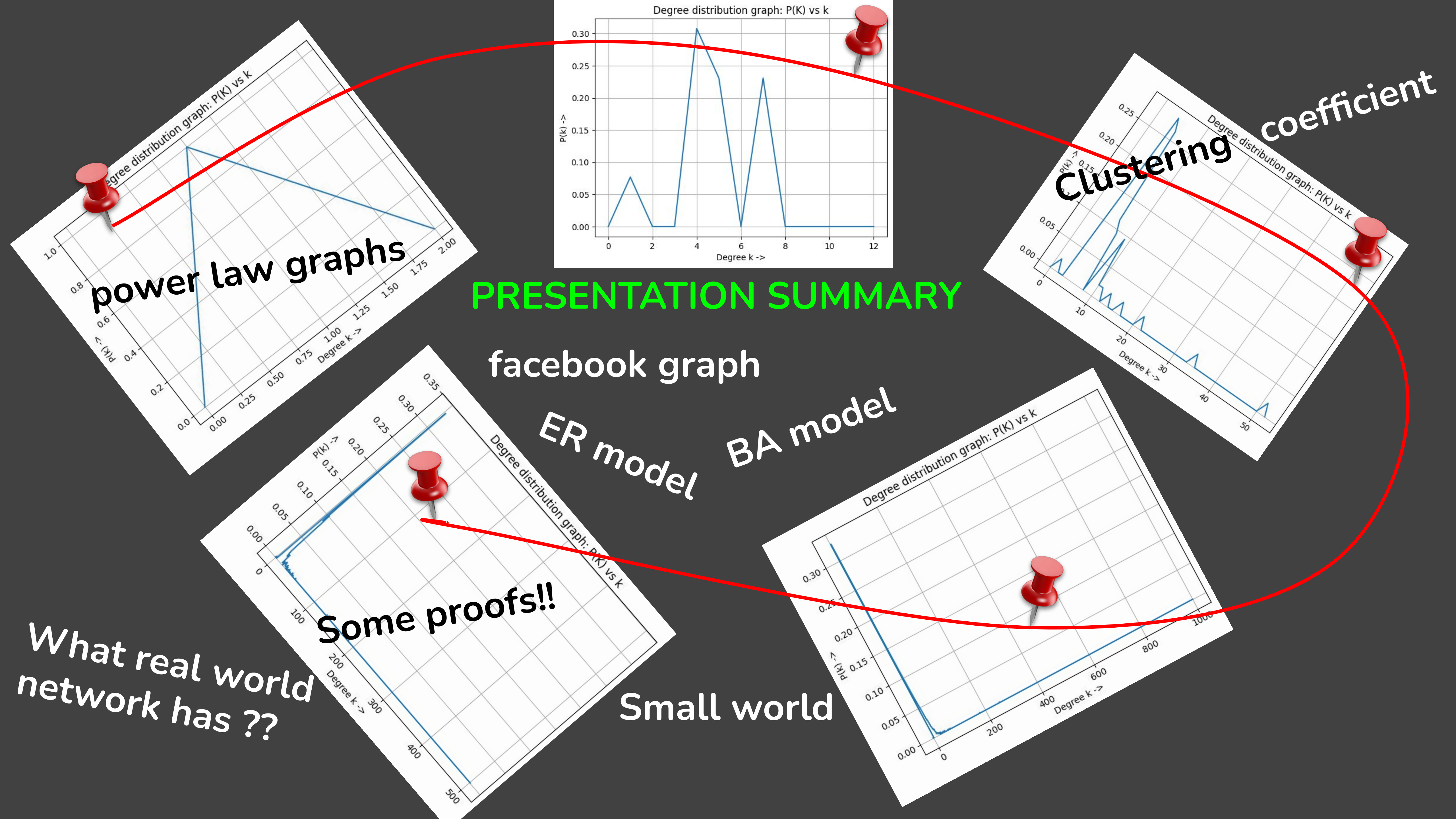
2) Holds small world property

   Average path length $L \sim \dfrac{\log N}{\log(\log N)}$

3) **Clustering coefficient**

   $C \sim N^{0.75}$

Link->


PRESENTATION SUMMARY
power law graphs
Clustering coefficient
facebook graph
ER model
BA model
Some proofs!!
What real world network has ??
Small world
Degree distribution graph: P(K) vs k
Degree k ->
P(k) ->