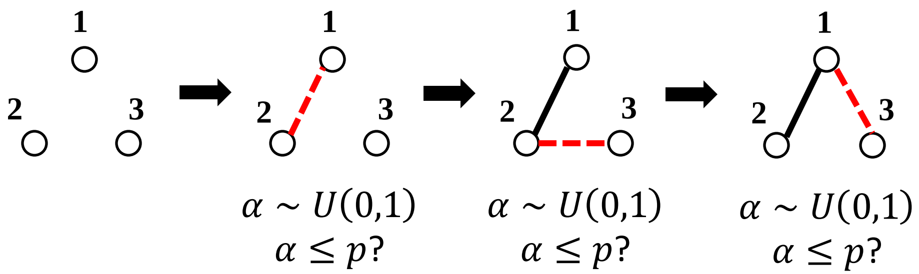
Statistical Graph Model

Barabasi-Albert
Configuration Models

Yu Wang, Ph.D.
Assistant Professor
Computer and Information Science
University of Oregon
CS 410/510 - Fall 2024

Statistical Graph Model – Random Graph Models – $G(n, p)$

In $G(n, p)$, a graph is constructed by connecting labeled nodes randomly.
Each edge is included in the graph with probability p.



How many steps should we consider?

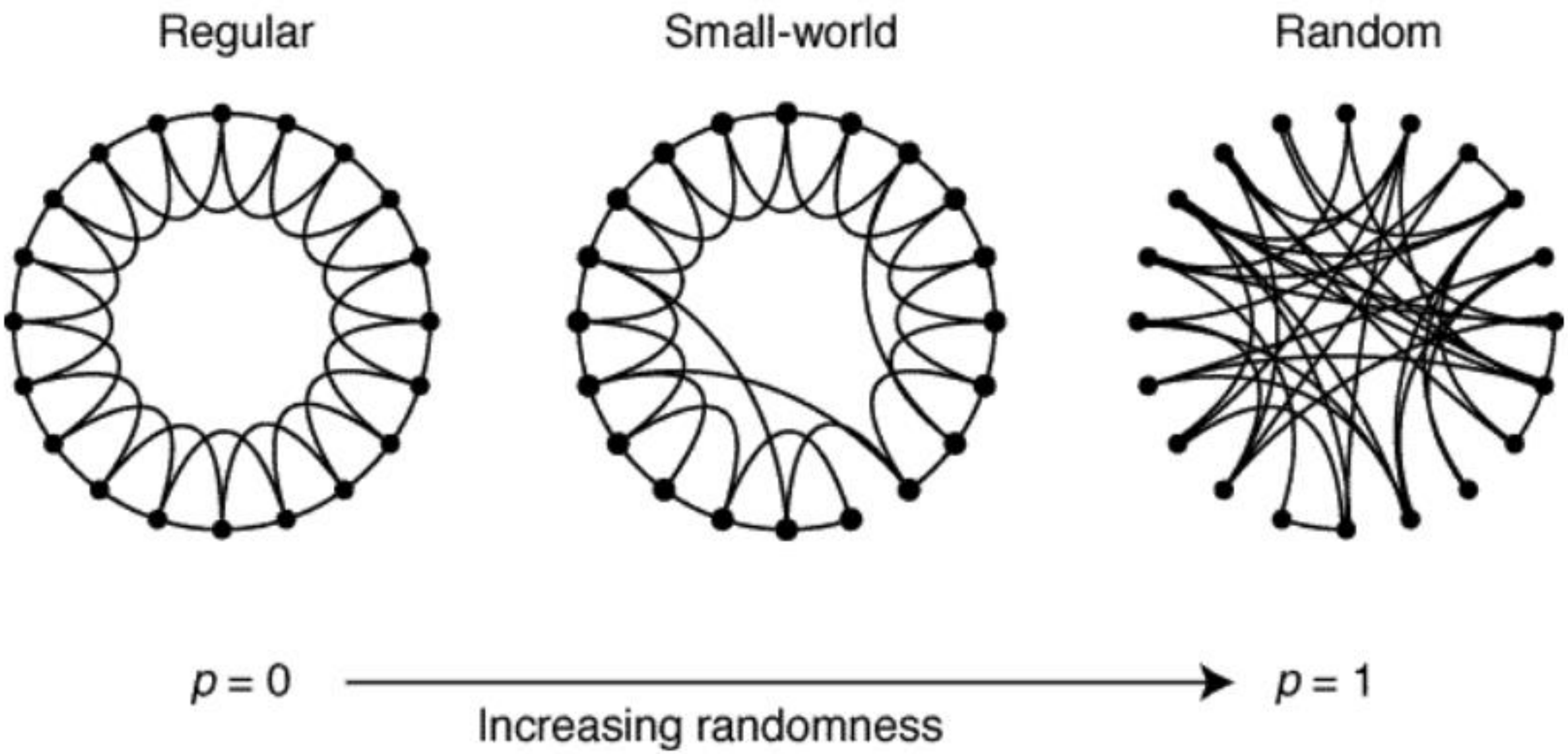
$$C_n^2 = \frac{n(n-1)}{2}$$

$G(n, p)$: a random graph with totally n nodes and among each pair of nodes, the edge is added with the probability of p

What is the expectation of the number of edges?

$$\bar{m} = p\frac{n(n-1)}{2}$$

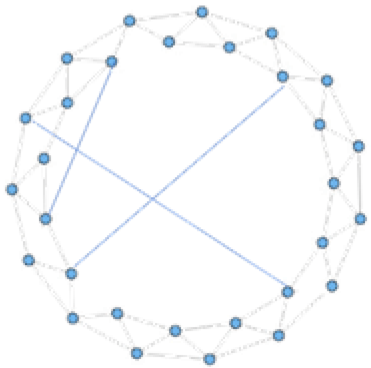
Statistical Graph Model – Small World



Statistical Graph Model – Small World

Idea:

- Randomly reconnect some links



Watts and Strogatz 1998.

Single parameter model

- Go between regular lattice and random graph
- Start with regular lattice of n nodes, k edges per vertex where k<<n
- Randomly reconnect with other nodes with probability p
  - Creates pnk/2 “long distance” connections
- p=0 regular lattice and p=1 random graph

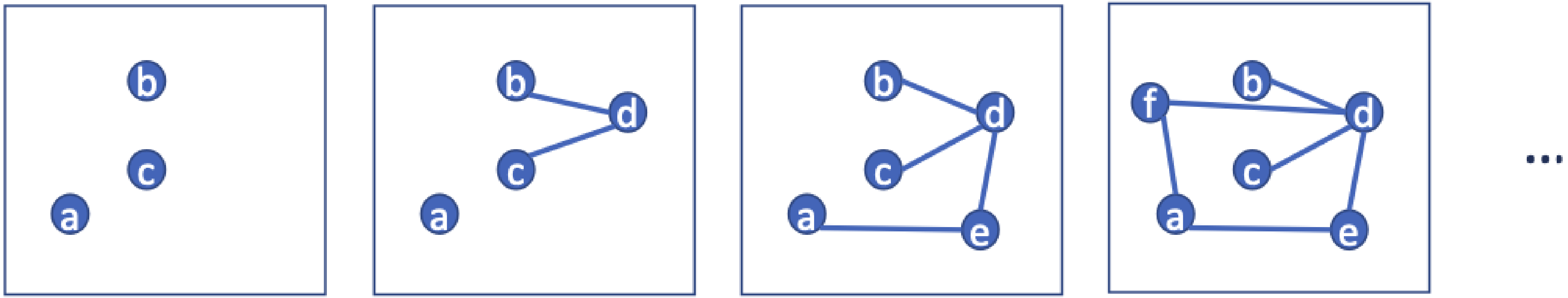
Statistical Graph Model – Dynamic Graph

- What does a real graph look like?
- What is normal/abnormal?
- Are real graphs random?

- Real Networks are growing!
  - Evolving with time
    - New nodes and edges
  - Citation/collaboration networks
  - Web
  - Social networks

Growing random graph

- Simple model with no nodes/edges being removed, only adding nodes/edges



$t = 0$ $t = 1$ $t = 2$ $t = 3$

Statistical Graph Model – Dynamic Graph

Stochastic growth model:

- Starting point
  - $t = 0$, n_o unconnected nodes
- Growth
  - On every time step $t = \{1,2,3, \ldots\}$ we add a new node with $\mathrm{m} \leq \mathrm{n}_0$ edges
    - i.e., at time $t = i$ the new node will have degree $k_i = m$
- Attachment
  - Form m edges with nodes existing in the graph uniformly at random
  - Probability to attach to any node already in the network

$$\Pi(k_i) = \frac{1}{n_0 + t - 1}$$

  - Example:

  $t = 11$ with $n_0 = 3$

  There are an initial $\mathrm{n}_0 = 3$ nodes and already $t - 1 = 10$ other new nodes added, so $\frac{1}{13}$

Statistical Graph Model – Dynamic Graph

Stochastic growth model:

- Starting point
  - $t = 0$, n_o unconnected nodes
- Growth
  - On every time step $t = \{1,2,3, \ldots\}$ we add a new node with $\mathrm{m} \leq \mathrm{n}_0$ edges
    - i.e., at time $t = i$ the new node will have degree $k_i = m$
- Attachment
  - Form m edges with nodes existing in the graph uniformly at random
  - Probability to attach to any node already in the network

$$\Pi(k_i) = \frac{1}{n_0 + t - 1}$$

Expected i-node degree at time t is $< \mathrm{k_i(t)} >$

$$k_i(t) = m + \frac{m}{n_0 + i} + \frac{m}{n_0 + i + 1} + \ldots + \frac{m}{n_0 + t - 1}$$

Initial m edges

As time goes by new nodes are added and pick m nodes to connect to…

Statistical Graph Model – Dynamic Graph

$$k_i(t) = m + \frac{m}{n_0 + i} + \frac{m}{n_0 + i + 1} + \ldots + \frac{m}{n_0 + t - 1}$$

$$= m + m\sum_{k=i}^{t-1} \frac{1}{n_0 + k} = m + m\left(\sum_{k=1}^{t-1} \frac{1}{n_0 + k} - \sum_{k=1}^{i-1} \frac{1}{n_0 + k}\right)$$

$$= m + m\left(\sum_{k=n_0+1}^{n_0+t-1} \frac{1}{k} - \sum_{k=n_0+1}^{n_0+i-1} \frac{1}{k}\right)$$

$$= m + m(\log(n_0 + t - 1) - \log(n_0 + 1) - (\log(n_0 + i - 1) - \log(n_0 + 1)))$$

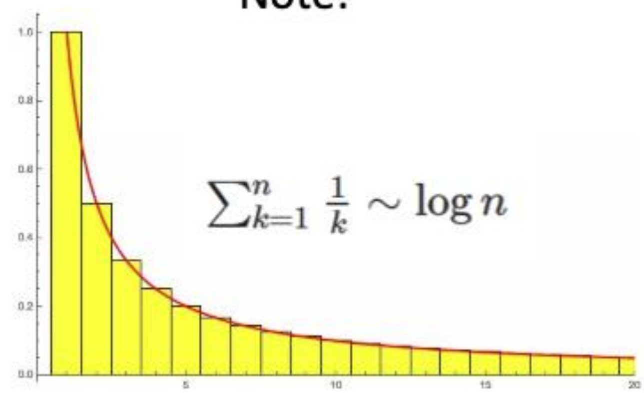
$$= m + m(\log(n_0 + t - 1) - \log(n_0 + i - 1))$$

$$= m\left(1 + log\frac{n_0+t-1}{n_0+i-1}\right)$$

$$\approx m\left(1 + log\frac{t-1}{i-1}\right), \text{when t is very large}$$

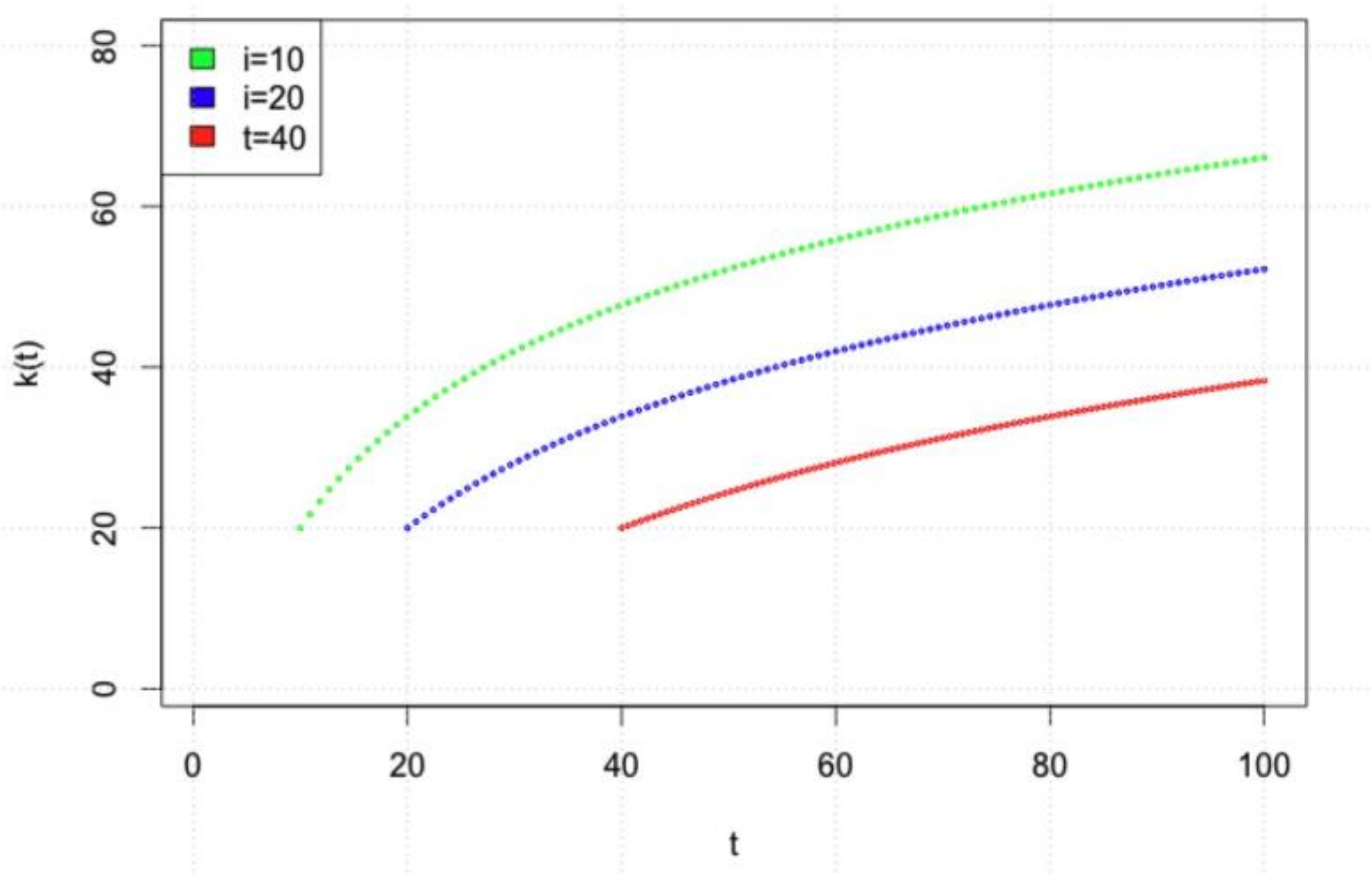
$$\approx m\left(1 + log\frac{t}{i}\right)$$

Note:



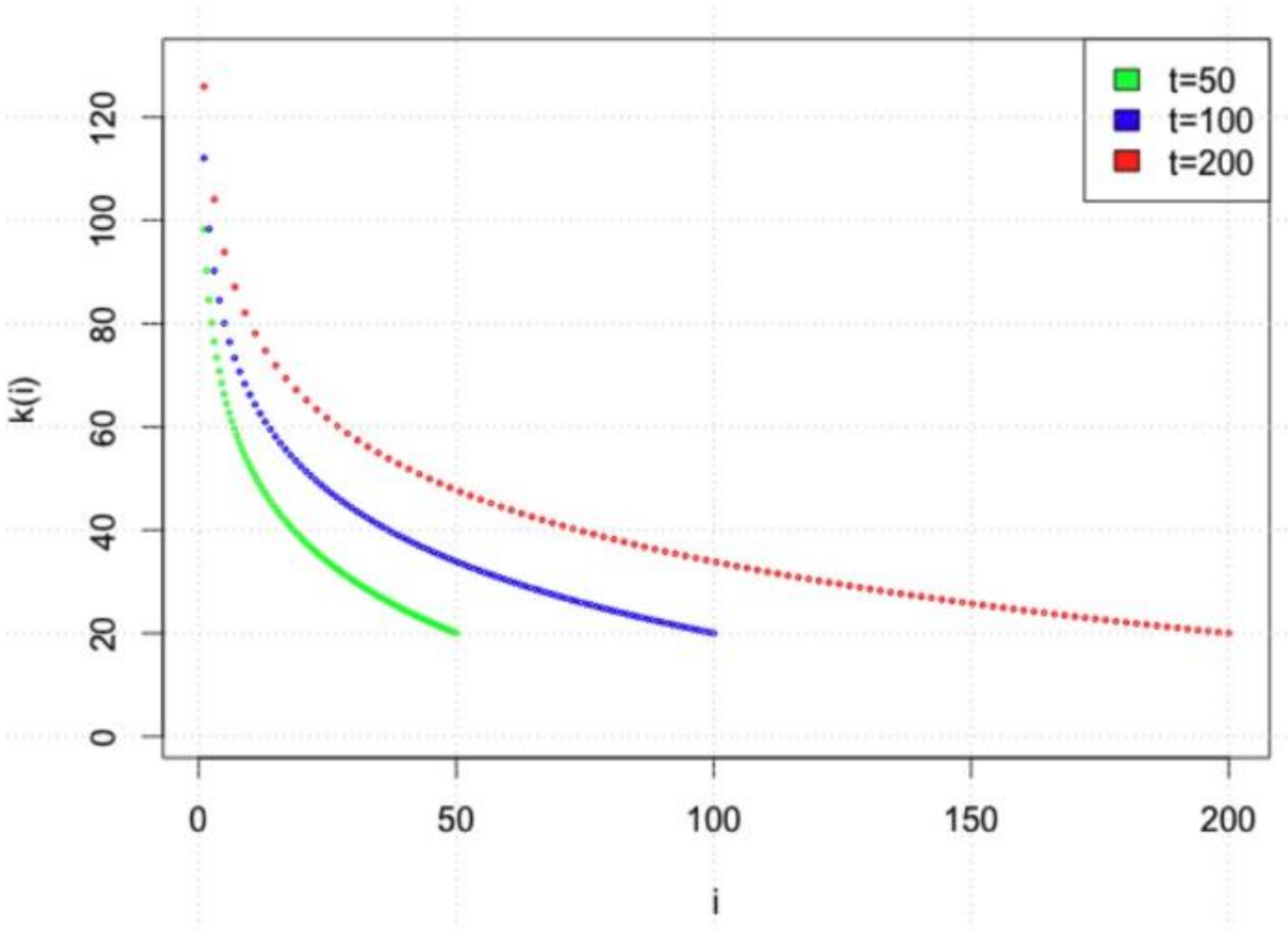
$\sum_{k=1}^{n} \frac{1}{k} \sim \log n$

Statistical Graph Model – Dynamic Graph



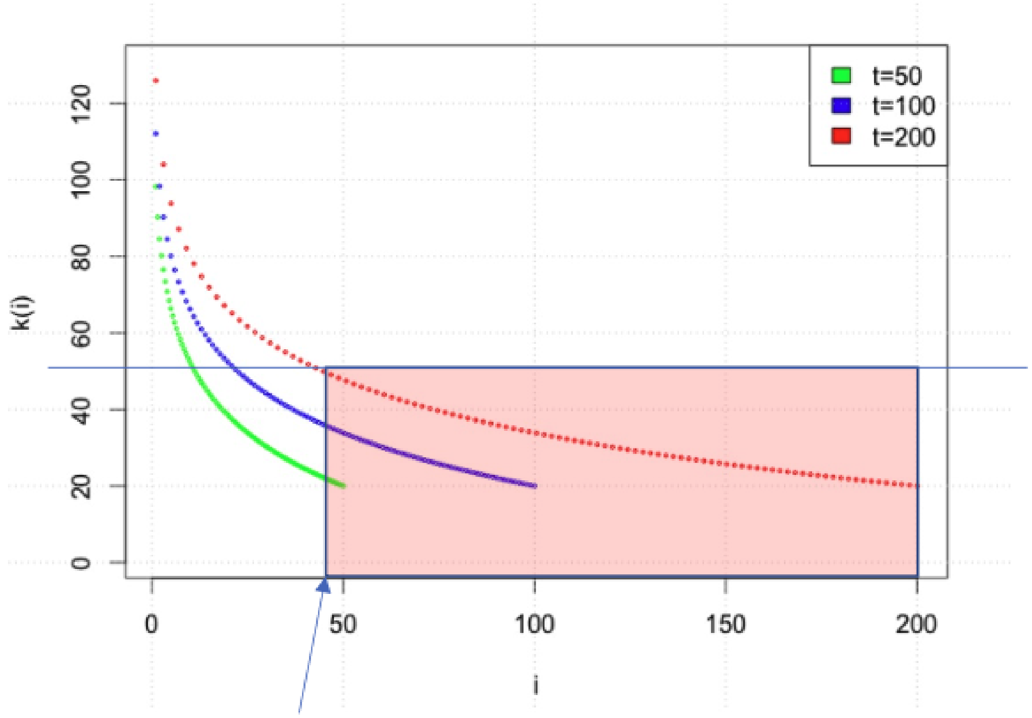
$$k_i(t) = m\left(1 + \log\left(\frac{t}{i}\right)\right), \qquad m = 20, \qquad i = 10, 20, 40, \qquad t \geq i$$

Statistical Graph Model – Dynamic Graph



Find all nodes that at time t has degree less than k

i.e., $k_i(t) \leq k$? e.g., $k_i(t) \leq 50$ t=200



Searching for this point i

Find all nodes that at time t has degree less than k

i.e., $k_i(t) \leq k$?

Note: expected i-node degree at time t is $\mathrm{k_i(t)} = m(1 + \log\left(\frac{t}{i}\right))$

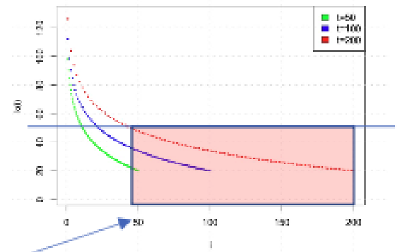
$$k_i(t) = m\left(1 + \log\left(\frac{t}{i}\right)\right) \leq k$$

$$\left(\log\left(\frac{t}{i}\right)\right) \leq \frac{k}{m} - 1$$

$$\frac{t}{i} \leq e^{\frac{k-m}{m}}$$

$$\frac{t}{i} \leq e^{\frac{k-m}{m}}$$

$$i > te^{\frac{m-k}{m}}$$



Growing random graphs

Find all nodes that at time t has degree less than k

i.e., $k_i(t) \le k$?

$$i > te^{\frac{m-k}{m}}$$

Fraction of nodes with degrees $k_i(t) < k$ (CDF):

$$F(k) = P(k_i(t) < k) = \frac{n_0+t-i}{n_0+t} = \frac{n_0+t-te^{\frac{m-k}{m}}}{n_0+t} =$$

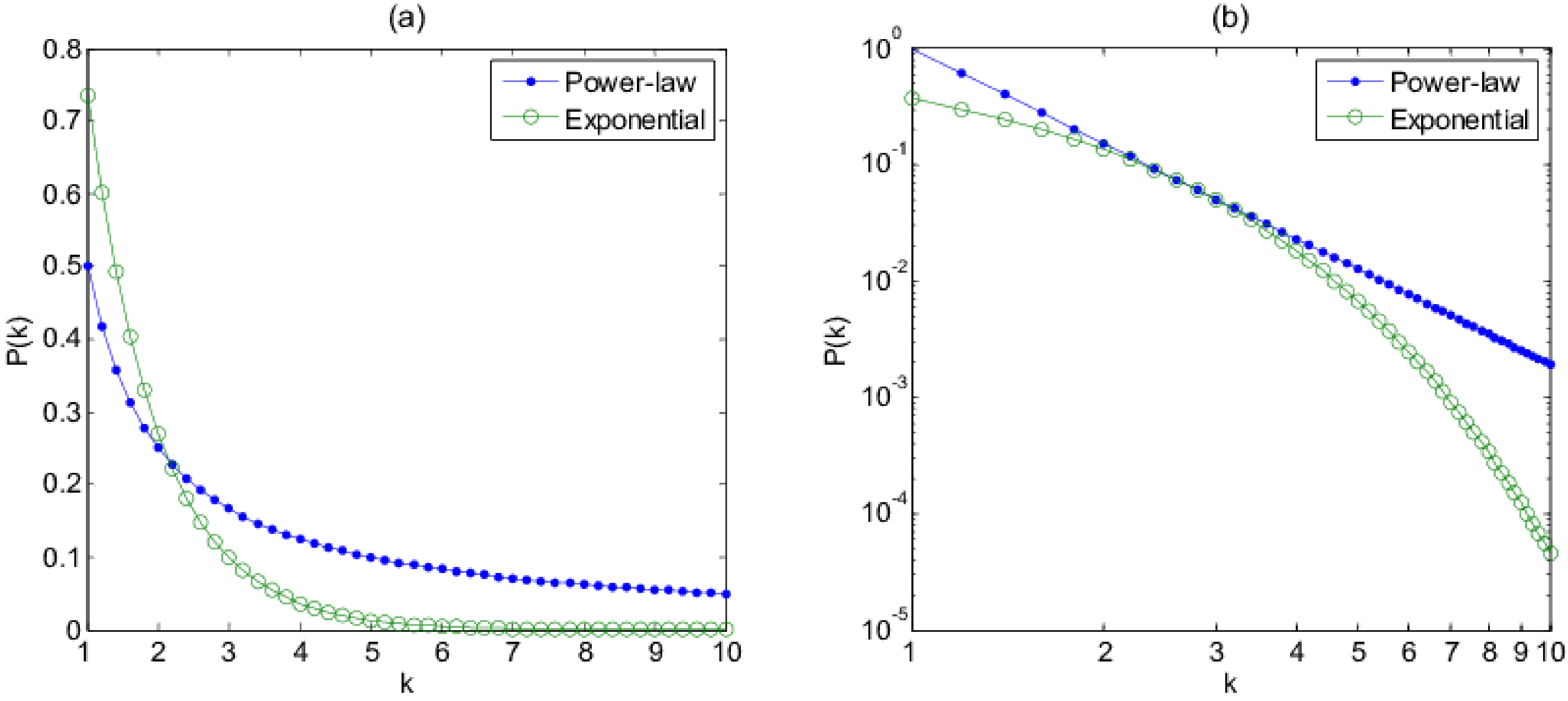
$$\frac{n_0+t(1-e^{\frac{m-k}{m}})}{n_0+t} \approx 1 - e^{\frac{m-k}{m}}$$ then convert CDF to PDF

$$P(k) = \frac{d}{dk}F(k) = \frac{1}{m}e^{-\frac{k-m}{m}} = \frac{e}{m}e^{-\frac{k}{m}}, \quad k \ge m$$

Because initially they all have at least m edges

This is exponential, not power law…

Statistical Graph Model – Dynamic Graph



Preferential attachment model

Barabasi and Albert, 1999.

- Dynamic growth model
  - Starting point
    - t = 0, n_o unconnected nodes
  - Growth
    - On every time step $t = \{1,2,3, \ldots\}$ we add a new node with $\mathrm{m} \leq \mathrm{n}_0$ edges
      - i.e., at time $t = i$ the new node will have degree $k_i = m$
  - **Preferential** Attachment
    - Form m edges with nodes existing in the graph proportional to the node existing degrees k_i
    - Probability to attach to any node already in the network

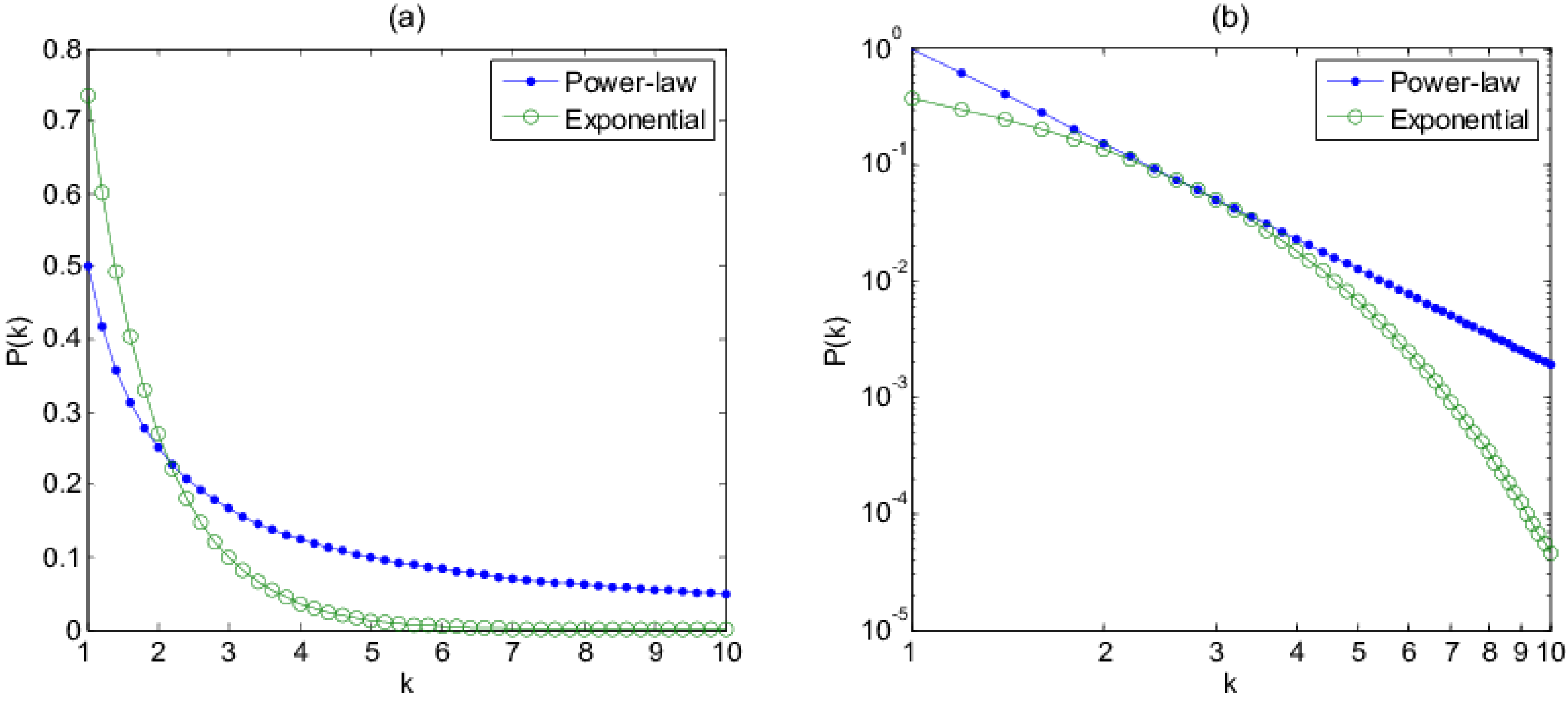
$$\boldsymbol{\Pi}(\boldsymbol{k_i}) = \frac{\boldsymbol{k_i}}{\sum_j \boldsymbol{k_j}}$$

Stochastic Growth Model

$$\Pi(k_i) = \frac{1}{n_0 + t - 1}$$

After t steps: $n_0 + t$ nodes and mt edges

Statistical Graph Model – Barabasi-Albert (PA)



- Idea: We want to generate a random graph $G = (V, E)$ with a fixed degree distribution
- Let D be a sequence of node degrees $\{k_1, k_2, \dots, k_n\}$ where $|V| = n$
  - Then we know the number of edges

$$|E| = m = \frac{1}{2}\sum_{i=1}^{n} k_i$$

Statistical Graph Model – Configuration Model

- Idea: We want to generate a random graph $G = (V, E)$ with a fixed degree distribution
- Let D be a sequence of node degrees $\{k_1, k_2, \ldots, k_n\}$ where $|V| = n$
  - Then we know the number of edges

$$|E| = m = \frac{1}{2}\sum_{i=1}^{n} k_i$$

Generation process: Connect the stubs from each node (seen below)

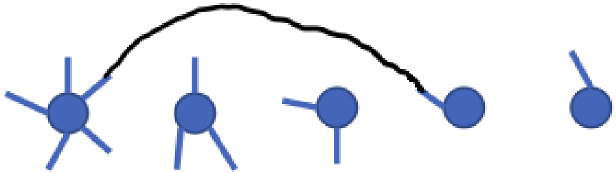
Example: D = {5,3,2,1,1}

- Idea: We want to generate a random graph $G = (V, E)$ with a fixed degree distribution
- Let D be a sequence of node degrees $\{k_1, k_2, \dots, k_n\}$ where $|V| = n$
  - Then we know the number of edges

$$|E| = m = \frac{1}{2}\sum_{i=1}^{n} k_i$$

Generation process: Connect the stubs from each node (seen below)

Example: D = {5,3,2,1,1}

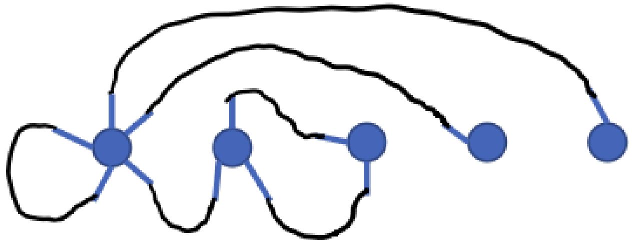


- Idea: We want to generate a random graph $G = (V, E)$ with a fixed degree distribution
- Let D be a sequence of node degrees $\{k_1, k_2, \dots, k_n\}$ where $|V| = n$
  - Then we know the number of edges

$$|E| = m = \frac{1}{2}\sum_{i=1}^{n} k_i$$

Generation process: Connect the stubs from each node (seen below)

Example: D = {5,3,2,1,1}



Note: Allows for self-loops and multiple edges between a pair of nodes.

Statistical Graph Model – Configuration Model

- Idea: We want to generate a random graph $G = (V, E)$ with a fixed degree distribution
- Let D be a sequence of node degrees $\{k_1, k_2, \dots, k_n\}$ where $|V| = n$
  - Then we know the number of edges

$$|E| = m = \frac{1}{2}\sum_{i=1}^{n} k_i$$

What is the probability that two nodes i and j are connected?

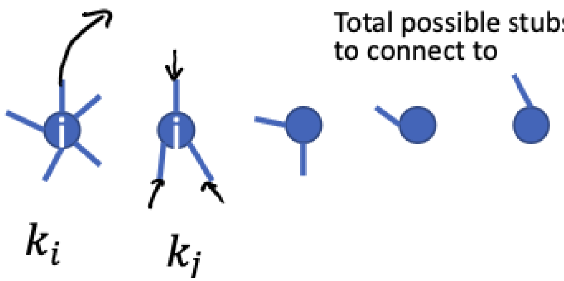
Total possible stubs to start from i

Total possible stubs to connect to for j

$$k_i\left(\frac{k_j}{2m - 1}\right)$$

Total possible stubs to connect to

$$p_{ij} = \frac{k_i k_j}{2m - 1}$$



Any Question?