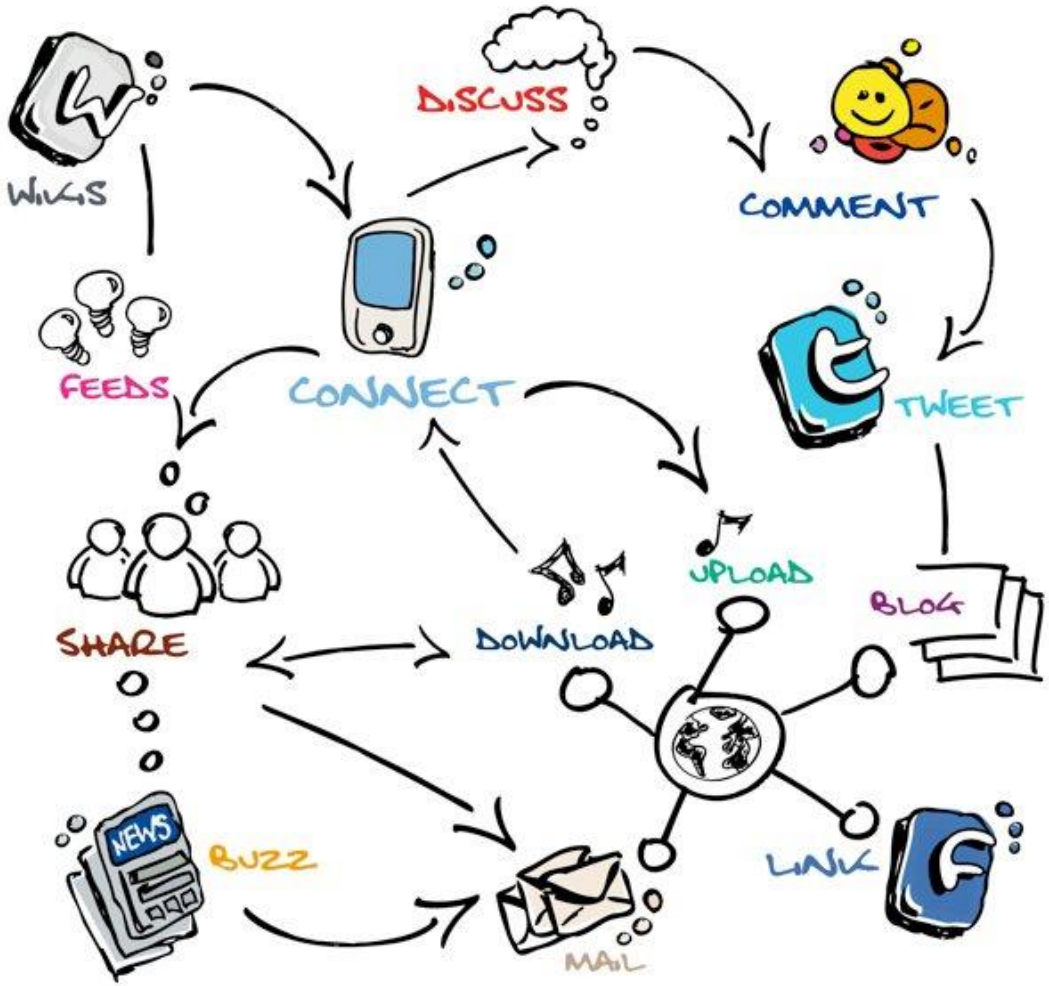
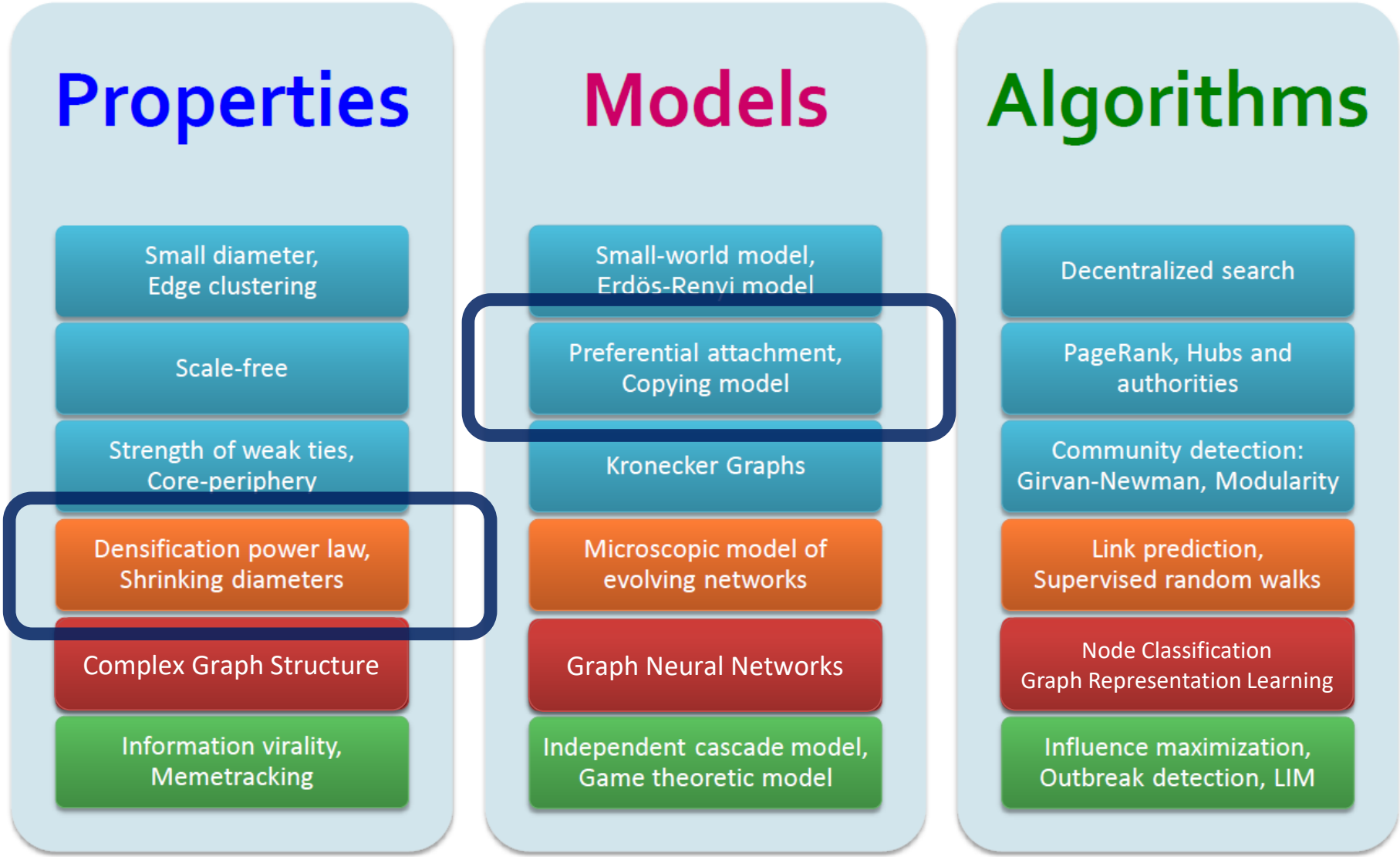


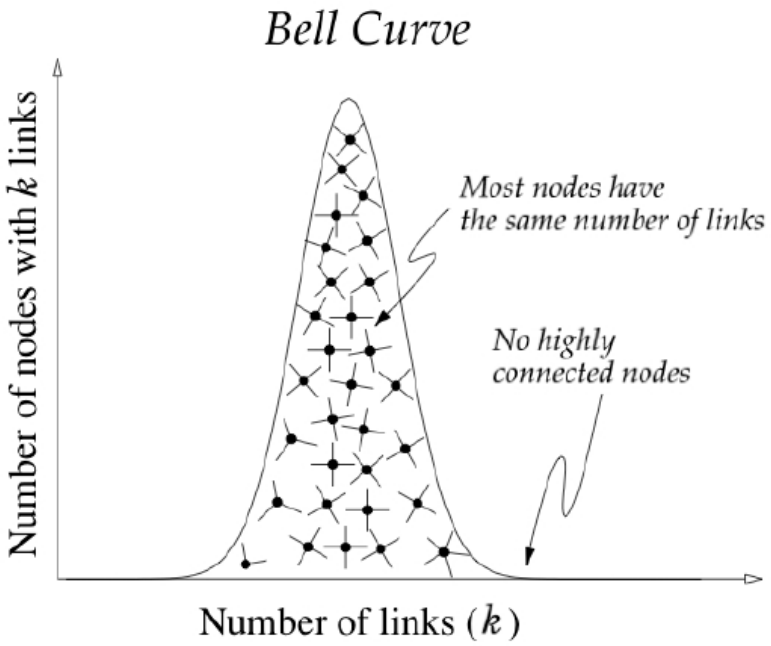
Preferential Attachment and Network Evolution

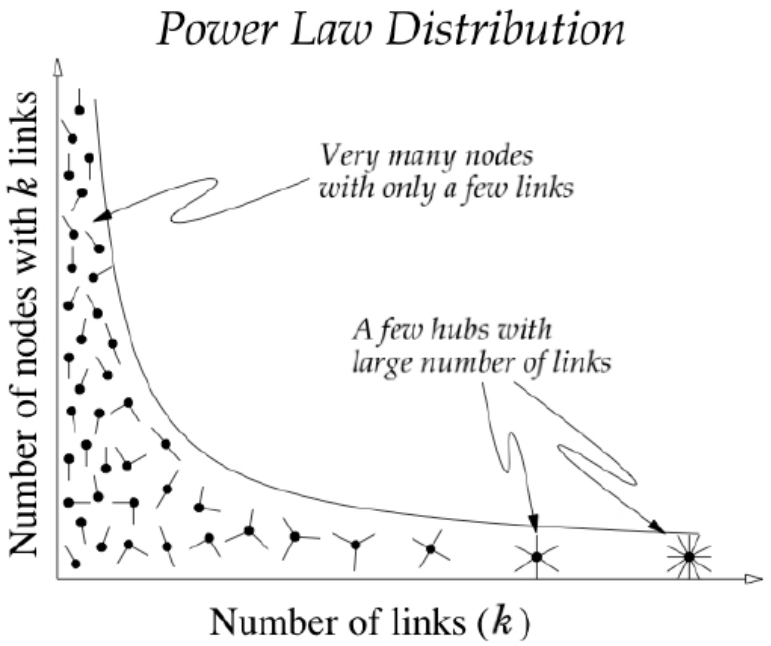
Ahmet Onur Durahim

How the Class Fits Together

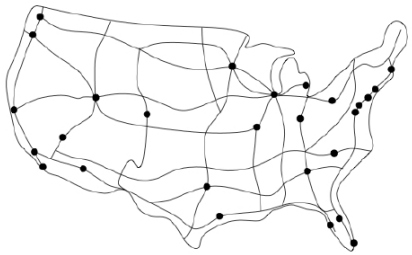


Exponential vs. Power-Law





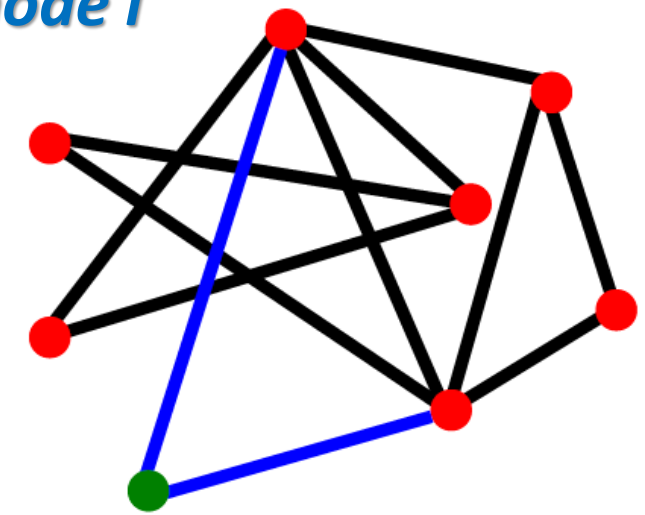
Model: G_{np} **?**



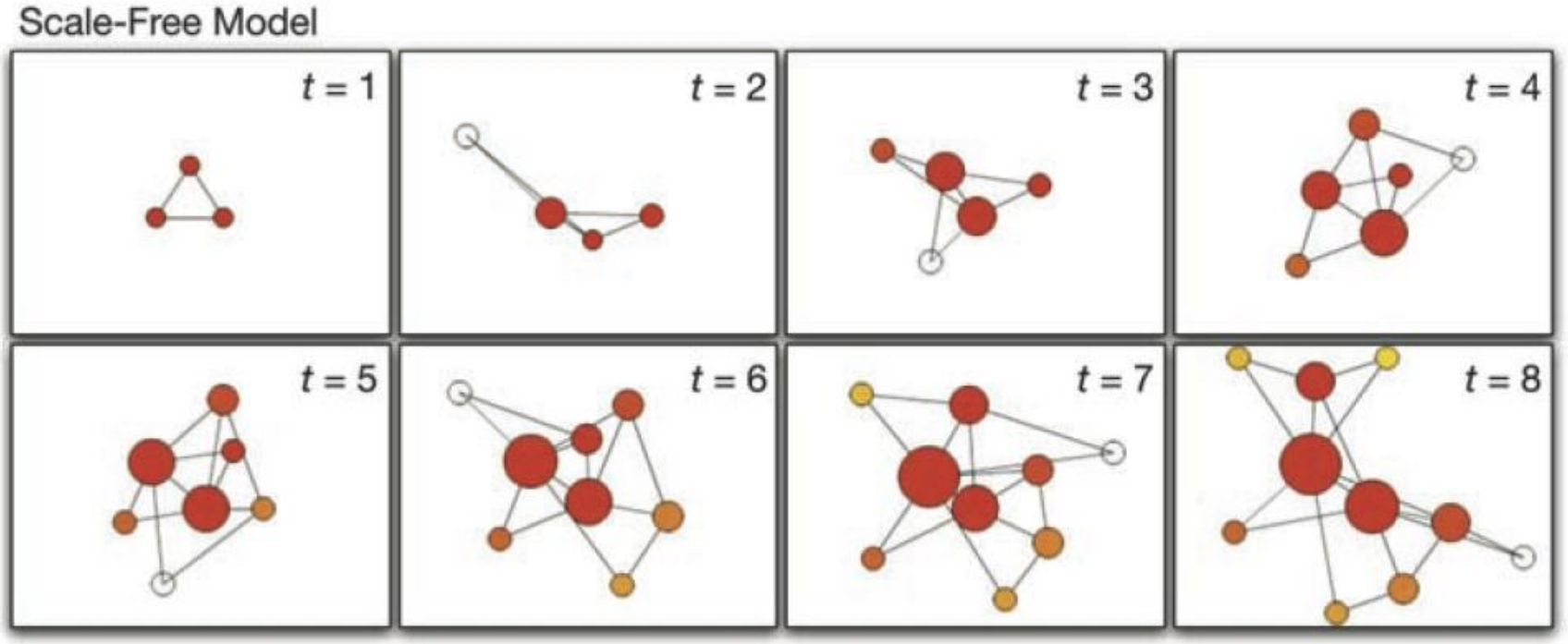
Model: Preferential Attachment

- **Preferential attachment**
  [Price '65, Albert-Barabasi '99, Mitzenmacher '03]
  - Nodes arrive in order ***1,2,…,n***
  - At step **j**, let $\mathbf{d_i}$ be the degree of node **i < j**
  - ***A new node j*** arrives and creates ***m out-links***
  - Prob. of j linking to a previous node ***i*** is ***proportional to degree d_i of node i***

$$P(j \to i) = \frac{d_i}{\sum_k d_k}$$



Preferential Attachment Model



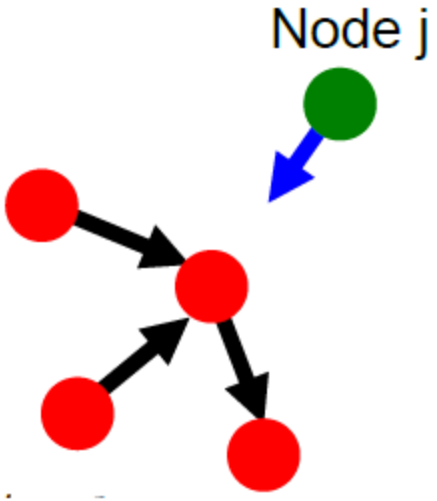
Barabasi, 1999

Rich Get Richer

- **New nodes are more likely to link to nodes that already have high degree**
- **Herbert Simon's result:**
  - Power-laws arise from *"Rich get richer"* (cumulative advantage)
- **Examples** [Price '65]
  - **Citations:** New citations to a paper are proportional to the number it already has
    - **Herding:** If a lot of people cite a paper, then it must be good, and therefore I should cite it too
  - **Sociology:** ***Matthew effect***
    - Eminent scientists often get more credit than a comparatively unknown researcher, even if their work is similar
    - http://en.wikipedia.org/wiki/Matthew_effect

[Mitzenmacher, '03]

The Exact Model



We will analyze the following model:

- Nodes arrive in order 1,2,3,…,n
- When **node *j* is created** it makes a ***single out-link*** to an earlier node i chosen:
  - 1) with prob. p, node j links to i chosen **uniformly at random** (from among all earlier nodes)
  - 2) with prob. $1-p$, node j chooses i uniformly at random and links to node l that i points to
    - **This is same as saying:** with prob. $1-p$, node j links to node l with prob. proportional to d_l (the in-degree of l)
  - **Our graph is directed:** Every node has out-degree **1**

Degree Distribution

- **What is $F(k)$, the fraction of nodes that has degree less than k at time t?**
  - How many nodes have degree < k?

$$d_i(t) = \frac{p}{q}\left(\left(\frac{t}{i}\right)^q - 1\right) < k$$

  - Solve for i and obtain: $i > t\left(\frac{q}{p}k + 1\right)^{-\frac{1}{q}}$

- **There are t nodes total at time t so the fraction $F(k)$ is:**

$$F(k) = 1 - \left[\frac{q}{p}k + 1\right]^{-\frac{1}{q}}$$

Degree Distribution

- **What is the fraction of nodes with degree exactly k?**
  - **Take derivative of $F(k)$:**
    - $F(k)$ is CDF, so $F'(k)$ is the PDF!

$$F(k) = 1 - \left[\frac{q}{p}k + 1\right]^{-\frac{1}{q}}$$

$$F'(k) = \frac{1}{p}\left[\frac{q}{p}k + 1\right]^{-1-\frac{1}{q}} \Rightarrow \alpha = 1 + \frac{1}{1-p}$$

q.e.d.

Pref. Attachment: Good news

- **Preferential attachment gives power-law degrees!**
- Intuitively reasonable process
- Can tune ***p*** to get the observed exponent
  - On the web, **P[*node has degree d*] ~ $d^{-2.1}$**
  - ***2.1 = 1 + 1/(1-p)*** → ***p ~ 0.1***

Pref. Attachment: Bad news

- **Preferential attachment is not so good at predicting network structure**
  - **Age-degree correlation**
    - ***Solution:*** Node fitness (virtual degree)
  - **Links among high degree nodes:**
    - On the web nodes sometime avoid linking to each other
- **Further questions:**
  - **What is a reasonable model for how people sample through network node and link to them?**
    - ***Short random walks***
    - ***Effect of search engines – reaching pages based on the number of links to them***

Many Models lead to Power-Laws

- **Copying mechanism (directed network) (Kleinberg)**
  - Select a node and an edge of this node
  - Attach to the endpoint of this edge
- **Walking on a network (directed network)**
  - The new node connects to a node, then to every first, second, … neighbor of this node
- **Attaching to edges**
  - Select an edge and attach to both endpoints of this edge
- **Node duplication**
  - Duplicate a node with all its edges
  - Randomly prune edges of new node

Distances in Preferential Attachment

Ultra small world

$$\bar{h} = \begin{cases} const & \alpha = 2 \\ \dfrac{\log\log n}{\log(\alpha - 1)} & 2 < \alpha < 3 \\ \dfrac{\log n}{\log\log n} & \alpha = 3 \\ \log n & \alpha > 3 \end{cases}$$

Small world

Avg. path length

Degree exponent

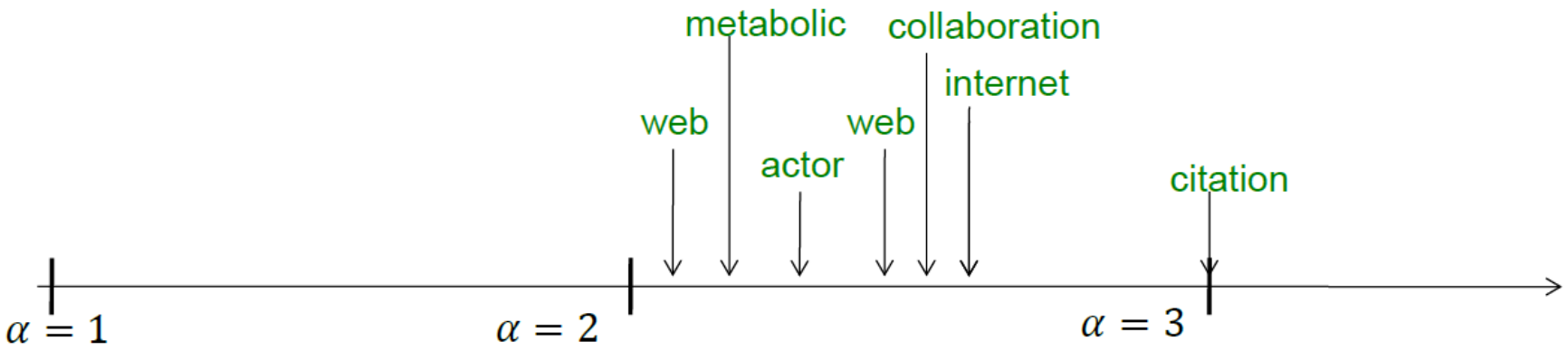
Size of the biggest hub is of order *O(N)*. Most nodes can be connected within two steps, thus the average path length will be independent of the network size.

The average path length increases slower than logarithmically. In G_{np} all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

Some models produce $\alpha = 3$. This was first derived by Bollobas et al. for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

Summary: Scale-Free Networks



Second moment $\langle k^2 \rangle$ diverges		$\langle k^2 \rangle$ finite
Average $\langle k \rangle$ diverges	Average $\langle k \rangle$ finite	
Ultra small world behavior		Small world
Regime full of anomalies…	The scale-free behavior is relevant	Behaves like a random network

Model Comparison

	Random	BA model	WS model	Empirical networks
P(k)	$\frac{\lambda^k e^{-\lambda}}{k!}$	k^{-3}	poisson like	power law
C	$\langle k \rangle / N$	$N^{-0.75}$	const	large
$\langle L \rangle$	$\frac{\log(N)}{\log(\langle k \rangle)}$	$\frac{\log(N)}{\log\log(N)}$	$\log(N)$	small