Overlay (and P2P) Networks

Part II

Samu Varjonen

Ashwin Rao

Birds Eye View of Complex Networks

- Milgram's Experiment
  - Small diameter
- Duncan Watts Model
  - Random Rewiring of Regular Graph
  - Graph with small diameter, high clustering coefficient
- Scale Free Model
  - Preferential attachment

Outline for this lecture

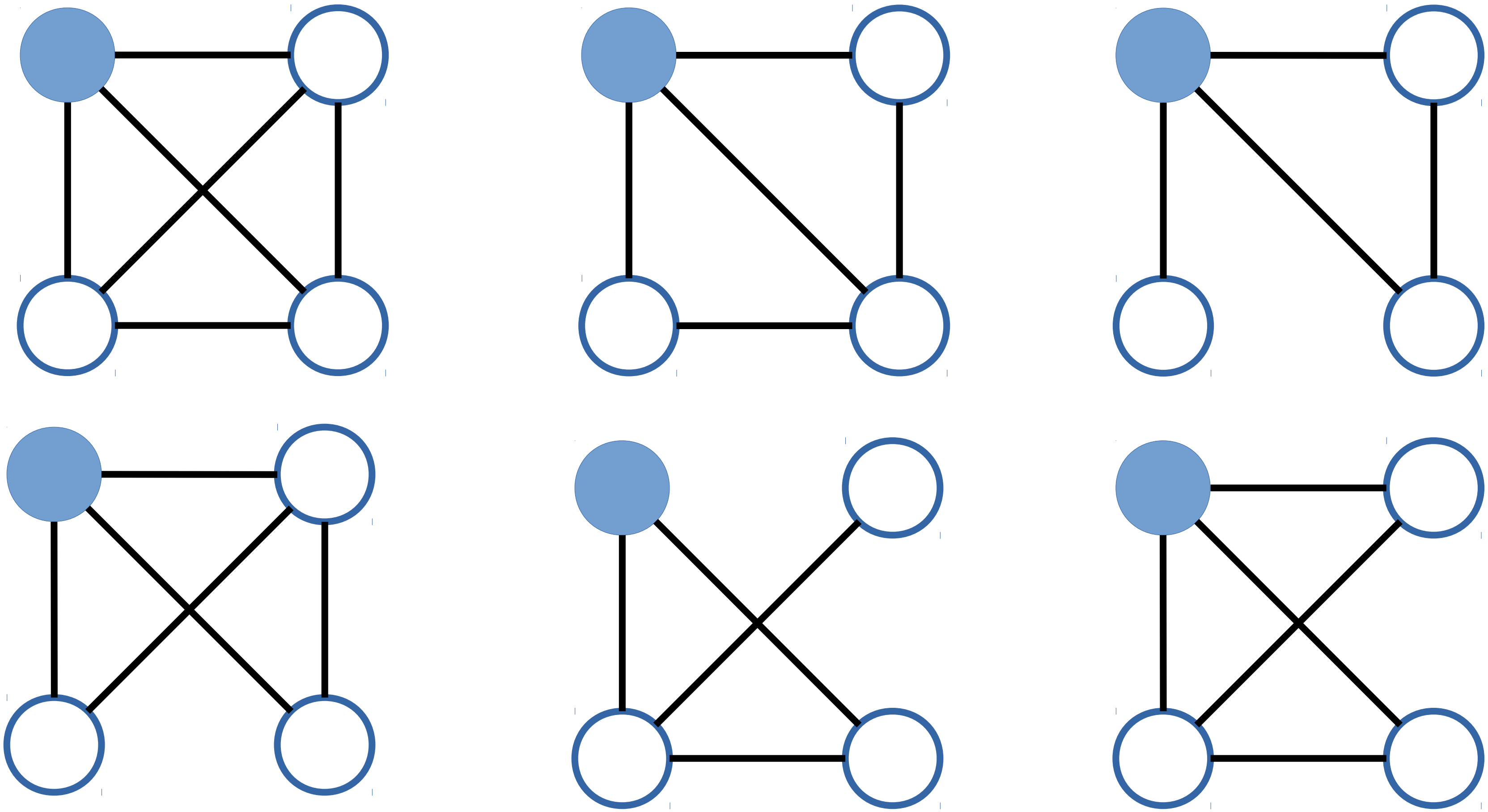
- Graph Properties
- Scale-Free Networks
- Navigation in Small World Networks

Local Clustering Coefficient

- Degree to which vertices of a graph tend to cluster together
- Quantifies how close neighbors of a vertex are to being a clique (complete graph)
- Computing clustering coefficient of vertex i
  - k_i : the number of neighbors of vertex i
  - N_i : set of vertices that are neighbors of vertex i
  - Number of edges in neighborhood of i : $\binom{k_i}{2} = \frac{k_i(k_i-1)}{2}$
  - Local Clustering Coefficient = $\frac{|e_{pq} : p \in N_i, q \in N_i, e_{pq} \in E|}{\frac{k_i(k_i-1)}{2}}$

Local Clustering Coefficient

Example



Average Clustering Coefficient

$$\hat{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

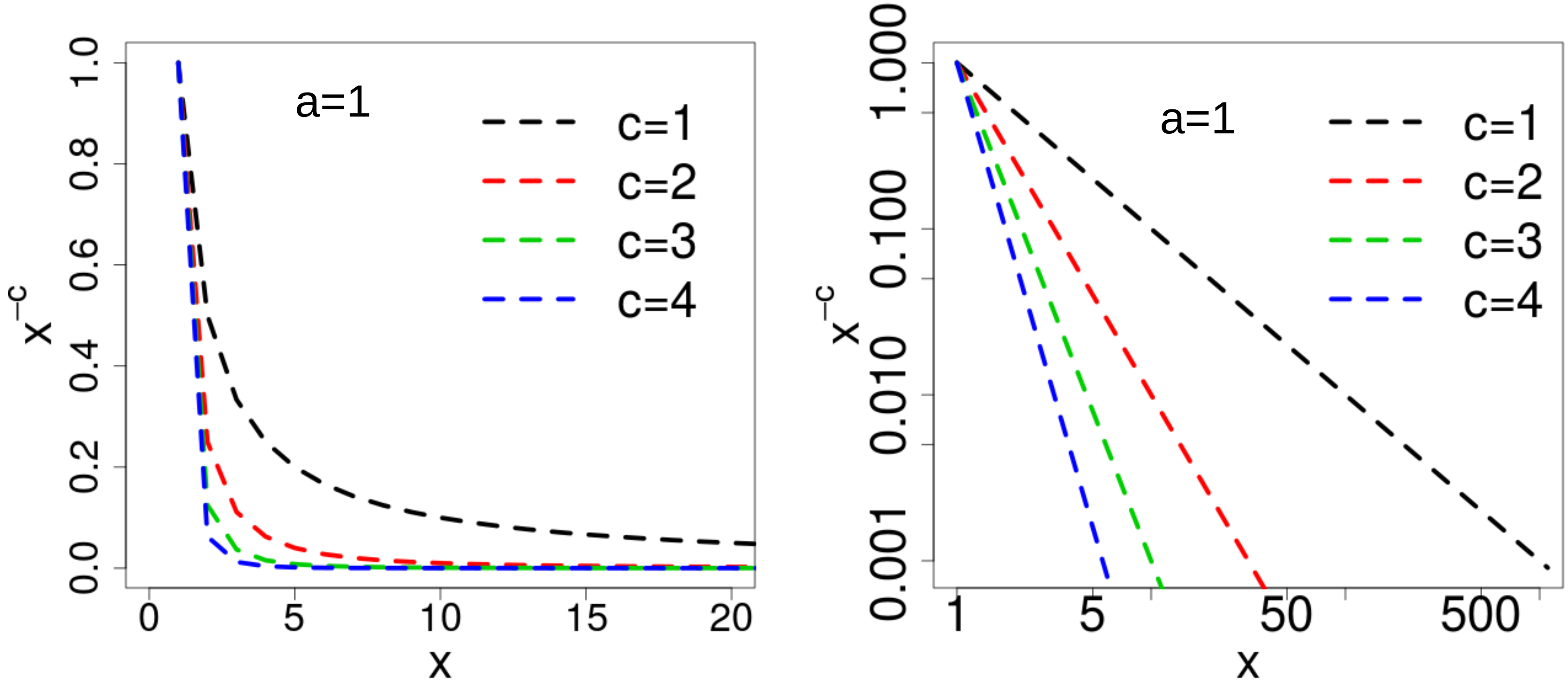
Graph Diameter

- Distance between two vertices i, j
  - Number of vertices in the shortest path between i, j
- Diameter: Greatest distance between any pair of vertices
- Radius: Minimum distance between any pair of vertices

Power-law

$$y = ax^{-c}$$
$$log(y) = -c\,log(x) + log(a)$$

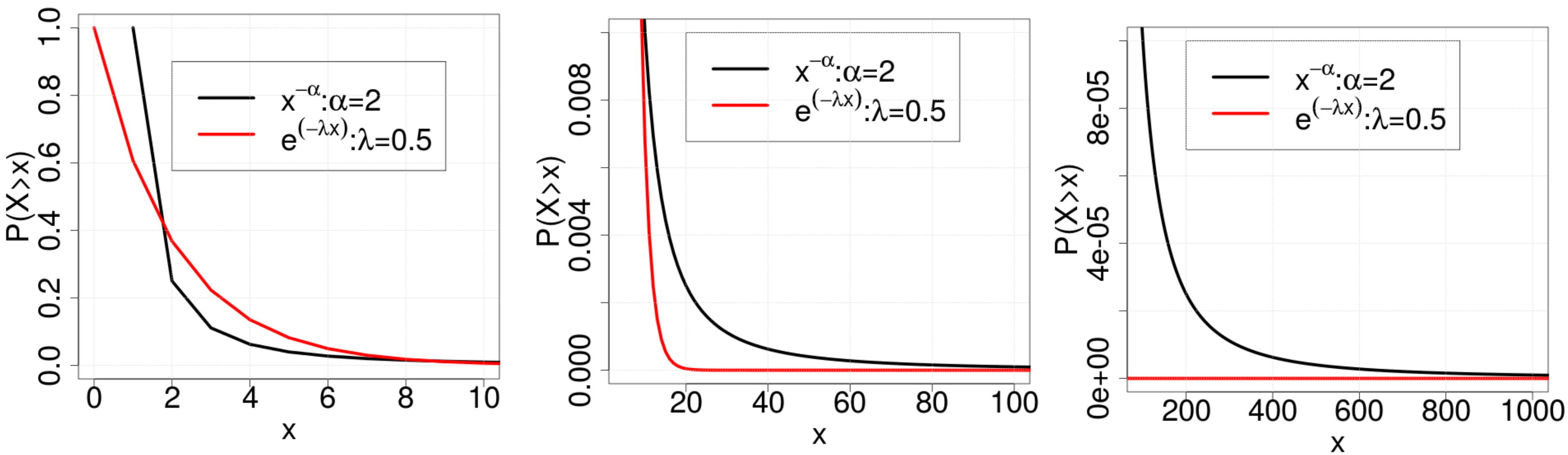
c : slope of log-log plot
a : constant (uninteresting)



Power-law distribution

$$Pr[X > x] \sim c\,x^{-\alpha}$$

$$f(x) \sim g(x) \implies \lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$$

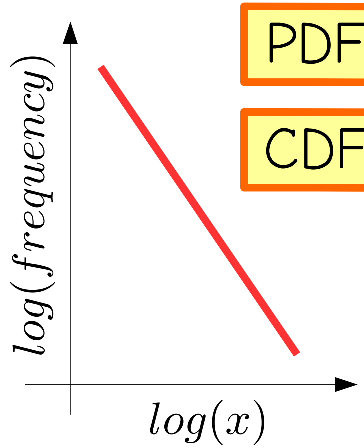


Self-similar & Scale-free
Heavy tail
Small occurrences (x) ***extremely*** common
Large occurrences (x) ***extremely*** rare

What is the CDF of power-law?

Micheal Mitzenmacher. "**A brief history of generative models for power law and lognormal distributions**." Internet mathematics 1.2 (2004): 226-251.

Exponent of PDF vs Exponent of CDF



PDF $p(x) = cx^{-\alpha}$

CDF $P(x) = \int_x p(x)dx = \int_x cx^{-\alpha}dx$

$$P(x) = \frac{c}{-\alpha+1}x^{-\alpha+1} = c'x^{-(\alpha-1)} = c'x^{-\alpha'}$$

$$P(x) \sim x^{-\alpha'}$$

$$\alpha' = \alpha - 1$$

What is the exponent of the PDF if you have read

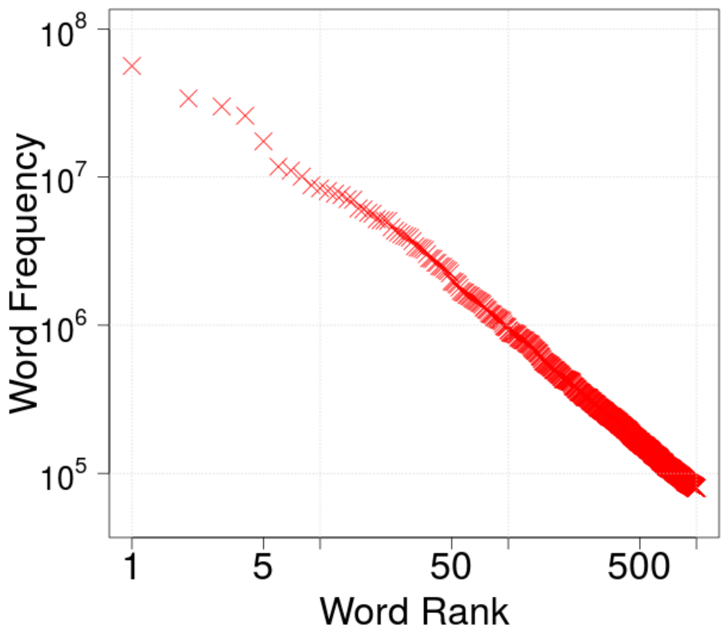
$$Pr[X > x] \sim c\,x^{-\alpha}$$

Zipf's Law

Example of discrete power-law

- Sort words in a corpus in the decreasing order of their frequency
- The frequency of any word is inversly proportional to its rank in the frequency table

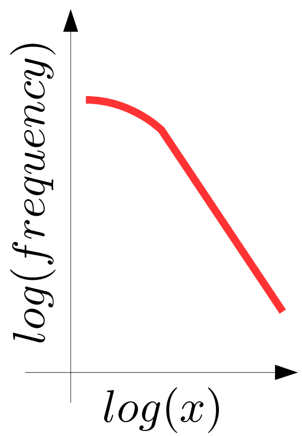
https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/PG/2006/04/1-10000



Ranking	Word	Frequency
1	the	56271872
5	in	17420636
10	he	8397205
50	when	1980046
100	men	923053
500	paid	162605
1000	names	79366

Quantities following Power-law

xmin: the smallest **x** in the measured samples for which the power-law behaviour holds



	Quantity	Minimum	Exponent
		$\boxed{x_{\min}}$	α
(*a*)	frequency of use of words	1	2.20(1)
(*b*)	number of citations to papers	100	3.04(2)
(*c*)	number of hits on web sites	1	2.40(1)
(*d*)	copies of books sold in the US	2000000	3.51(16)
(*e*)	telephone calls received	10	2.22(1)
(*f*)	magnitude of earthquakes	3.8	3.04(4)
(*g*)	diameter of moon craters	0.01	3.14(5)
(*h*)	intensity of solar flares	200	1.83(2)
(*i*)	intensity of wars	3	1.80(9)
(*j*)	net worth of Americans	$600m	2.09(4)
(*k*)	frequency of family names	10000	1.94(1)
(*l*)	population of US cities	40000	2.30(5)

Mark Newman. **"Power laws, Pareto distributions and Zipf's law."** Contemporary physics 46, no. 5 (2005): 323-351

Preferential Attachment (Growth in Scale-Free Network)

- Initial number of nodes $\mathbf{m_0}$
- At time **'t'** add a new vertex **j** with degree **m (< m_0)**
- The probability that vertex **j** creates a link with vertex **i** is the probability **π** which depends on $\mathbf{k_i}$ (the degree of vertex **i**)

$$\pi(k_i) = \frac{k_i}{\sum_l k_l}$$

- Probability that a node has **k** links is $P(k) \propto k^{-\gamma}$
- γ exponent for the degree distribution

Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." Science 286, no. 5439 (1999): 509-512.

Evolving Copying Model

Observations & Assumptions

- Observations of the Web-graph
  - The graph is a directed graph
  - Average degree of vertex is largely constant over-time
  - Disjoint instances of bipartite cliques
  - On creation, vertex adds edge(s) to existing vertices
- Assumptions made by the model
  - The directed graph evolves over discrete timesteps
  - Some vertices choose their outgoing edges independently at random, while others replicate existing linkage patterns by copying edges from a random vertex

Ravi Kumar et al. "**Stochastic models for the web graph.**" In Annual Symposium on Foundations of Computer Science, 2000.

Generic Evolving Graph Model

$$t = 1,2,3 \ldots \text{ (discrete timesteps)}$$

$$G_t = < V_t, E_t >$$

$$|V_{t+1}| = |V_t| + f_v(V_t, t)$$

$$E_{t+1} \cup E_t + f_e(f_v, G_t, t)$$

An evolving graph model can be completely characterised by

$$< f_v, f_e >$$

Ravi Kumar et al. "**Stochastic models for the web graph.**" In Annual Symposium on Foundations of Computer Science, 2000.

Evolving Copying Model

Linear Growth Copying

- Parameters
  - Constant out-degree $d \geq 1$
  - Copy factor α
- At time instant t
  - Add a single new vertex u $[f_v(V_t, t) = 1]$
  - u is a copy of a prototype vertex $p \in_R V_t$
  - The i^{th} outlink from u is randomly chosen from V_t with a probability α and with probability $1 - \alpha$ it is the i^{th} outlink from p

Ravi Kumar et al. "**Stochastic models for the web graph.**" In Annual Symposium on Foundations of Computer Science, 2000.

More Reading

- Variants of Evolving Copying Model
  - Multiple Variants of Linear Growth
  - Exponential Growth
  - Include death process to cause edges and vertices to disappear

Ravi Kumar et al. "**Stochastic models for the web graph.**" In Annual Symposium on Foundations of Computer Science, 2000.

Scale Free Networks

- Node degree distribution follows Power-law
- Can be created using
  - Preferential attachment (undirected graphs)
  - Copying Generative Models (directed graphs)
  - ...
- How realistic are these models?
  - Do graphs undergo densification (degree increases with time)?
  - Does the diameter decrease with time?

Jure Leskovec et al. "**Graphs over time: densification laws, shrinking diameters and possible explanations.**" In ACM SIGKDD, pp. 177-187. 2005.

When to use a particular model?

- What is the objective of your study?
- Why do you need the model?
- Which model will you use for

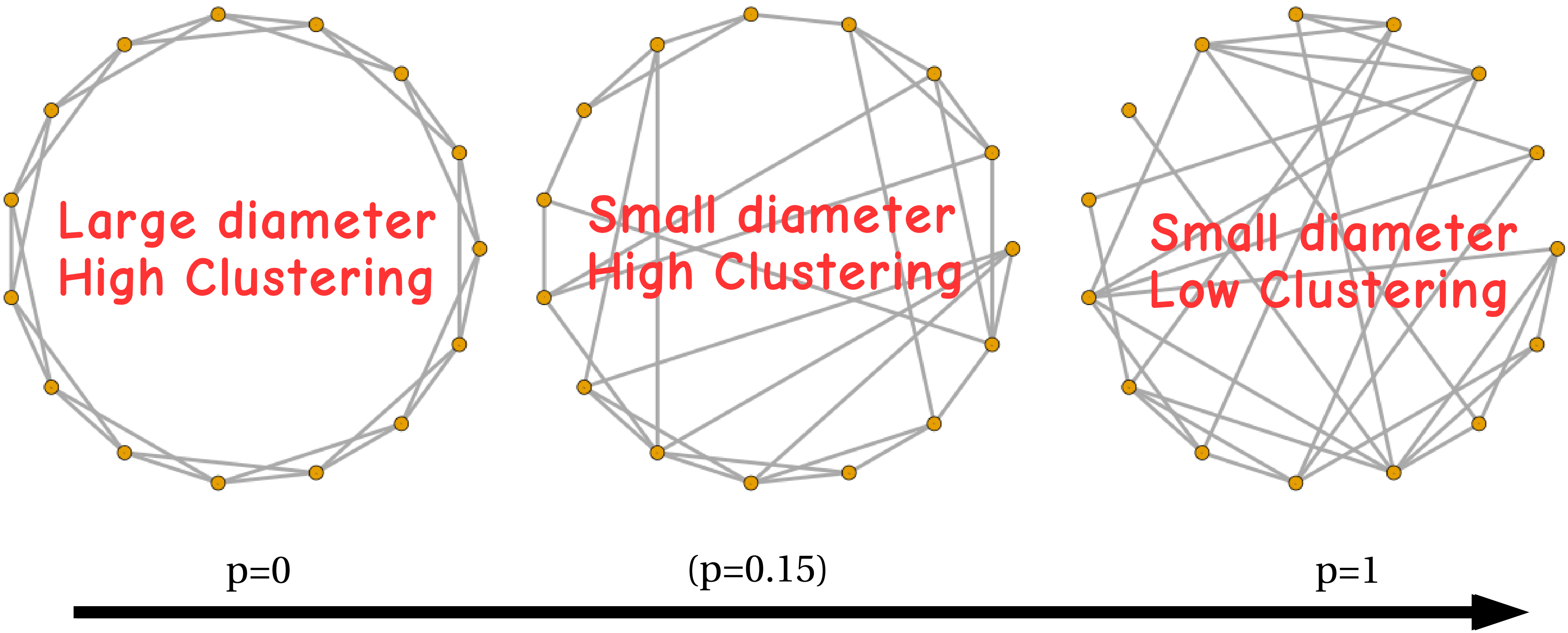
  (a) Properties of a snapshot of a graph

  (b) Evolution of the graph

  (c) ...

Random Rewiring Procedure

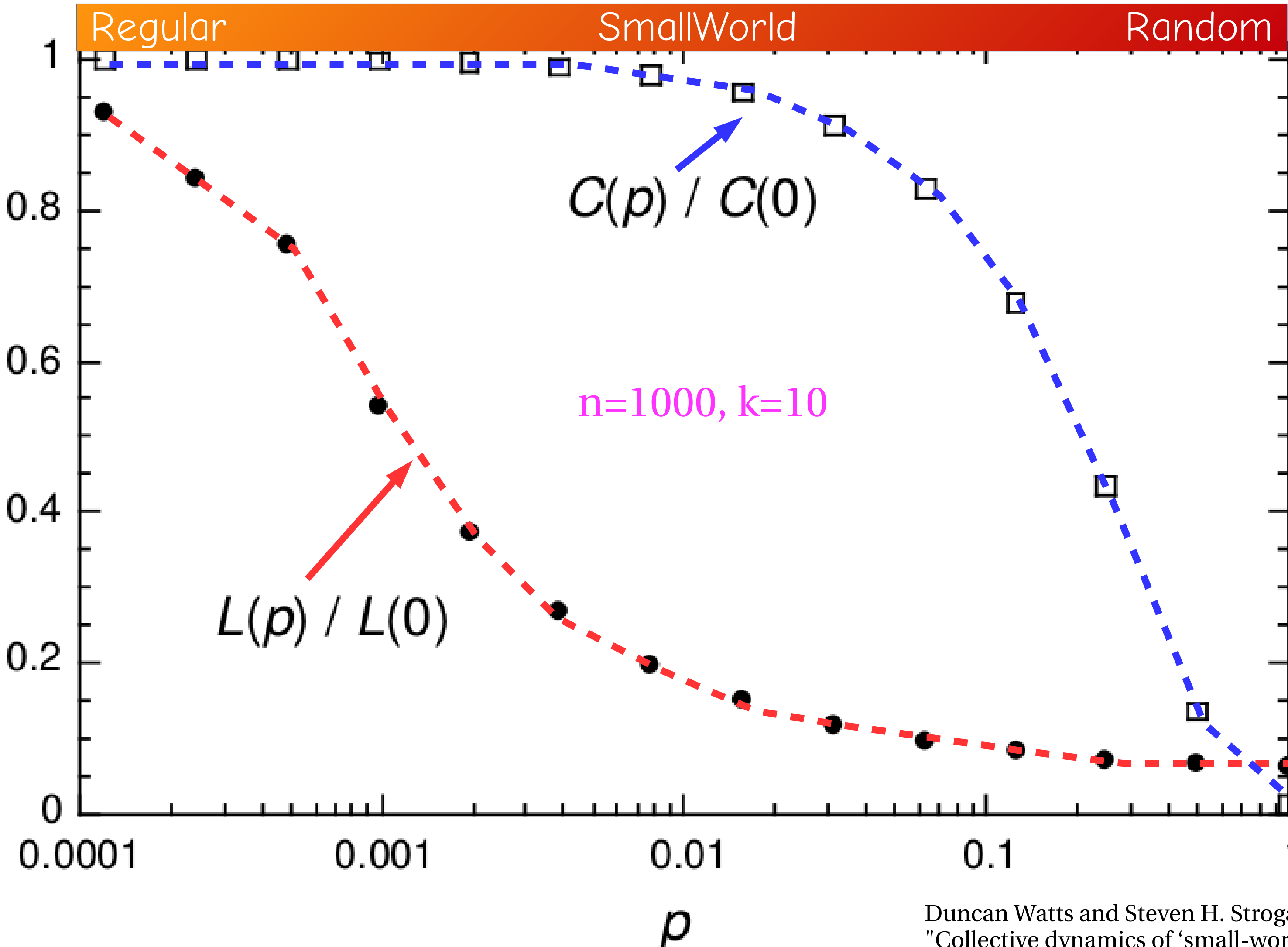
n=10, k=4



n: number of vertices
k: degree of each node, i.e., neighbours of a vertex (n >> k >> ln(n) >> 1)
p: probability of rewiring

Duncan Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." nature 393, no. 6684 (1998): 440-442.

Path Length & Clustering Coefficient in Small-World Networks



Duncan Watts and Steven H. Strogatz.
"Collective dynamics of 'small-world' networks."
nature 393, no. 6684 (1998): 440-442.

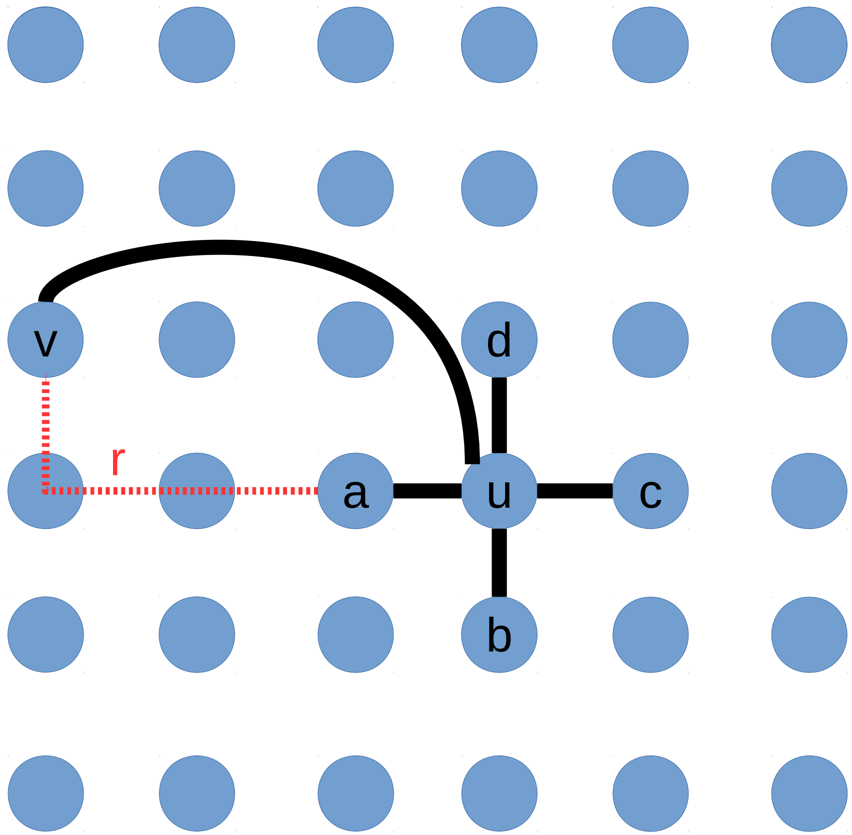
Decentralized Search in Small World Networks

(Motivation)

- How would you find the shortest path if you have the complete picture of the graph?
  - What will be the information required to be present on each node?
- Can you find the shortest path of small world graph if nodes only have **local information**?
  - My coordinates on the graph
  - Coordinates of my neighbors
  - Which algorithm would you use?

Klienbergs Small World

- $n \times n$ Lattice
- Each node has short-range links to its neighbors
- Long range connection to a node v selected with probability proportional to $r^{-\alpha}$. Where r is the ***Manhattan distance*** between u and v
- α : the clustering exponent, probability of connection as a function of lattice distance



Jon Kleinberg. "**Navigation in a small world.**" Nature 406, no. 6798 (2000): 845-845.

What does Klienbergs Model abstract?

- α abstracts the randomness of long links
  - $\alpha = 0$: uniform distribution over long-range contacts (similar to rewiring graph)
  - As α increases, long range contacts become more clustered
- Model can be extended to **d**-dimensional lattices

Kleinberg's Result

- Assumption:
  - Each node only has local information (neighbors)
  - **d** : lattice dimension **n × n × .... (d times)**
  - **p** : short-range links to vertices in **p** lattice steps
  - **q** : long-range links computed with exponent **α**
- **Greedy Algorithm:** Each message holder forwards message across a connection that brings it as close as possible to the target in lattice distance
- Message delivery time **T** of **any** decentralized algorithm
  - $\mathbf{T \geq cn^{\beta}}$ **for β = (2 – α)/3** for **0 ≤ α < 2** and **β = (α-2)/(α-1)** for **α** > 2
  - **T** bounded by polynomial of **logN** when **α = 2**
  - **c** depends on **α**, **p**, and **q** but not **n**

Implications of the result

- Decentralized Search requires a large number of steps (much larger than the shortest path)
  - Small world captures clustering and small paths but not the ability actually find the *shortest paths*
  - *""When this **correlation** (between local structure and long-range connections) is **near a critical threshold**, the structure of the **long-range connections forms a type of gradient that allows individuals to guide a message efficiently towards a target**. As the correlation drops below this critical value and the social network becomes more homogeneous, these cues begin to disappear; in the limit, when long-range connections are generated uniformly at random, **the result is a world in which short chains exist but individuals, faced with a disorienting array of social contacts, are unable to find them**."*

Important Papers

- Ravi Kumar et al. "**Stochastic models for the web graph.**" In Annual Syposium on Foundations of Computer Science, 2000.
- Jon Kleinberg. "**Navigation in a small world.**" Nature 406, no. 6798 (2000): 845-845.
- Jure Leskovec et al. "**Graphs over time: densification laws, shrinking diameters and possible explanations.**" In ACM SIGKDD, pp. 177-187. 2005.

Sources for these slides

- Sasu Tarkoma ”**Overlay and P2P Networks**”, 2015
- Daron Acemoglu et al. ”**6.207/14.15: Networks. Lecture 7: Search on Networks: Navigation and Web Search.**” 2009.