CLOUD COMPUTING CONCEPTS

with Indranil Gupta (Indy)

STRUCTURE OF NETWORKS

Lecture A

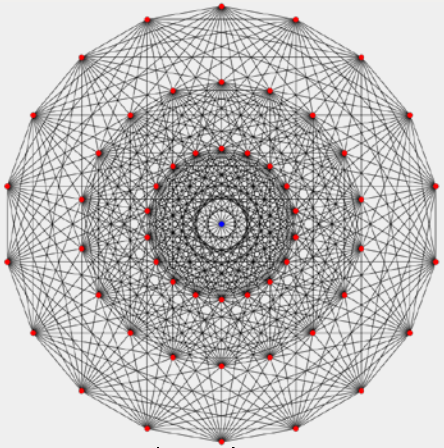
STRUCTURE OF NETWORK

What's a Network/Graph?

- Has vertices (i.e., nodes)
  - E.g., in the Facebook graph, each user = a vertex (or a node)
- Has edges that connect pairs of vertices
  - E.g., in the Facebook graph, a friend relationship = an edge

Lots of Graphs/networks

- Large graphs/network are all around us
  - Internet : vertices are routers/switches and edges are links
  - World Wide Web: vertices are webpages, and edges are URL links on a webpage pointing to another webpage
    - Called “Directed” graph as edges are uni-directional
  - Social networks: Facebook, Twitter, LinkedIn
  - Biological networks: DNA interaction graphs, ecosystem graphs, etc.



Source: Wikimedia Commons

Complexity of Networks

- Structural: *human population has ~7 B nodes, there are millions of computers on the Internet…*
- Evolution: *people make new friends all the time, ISP's change hands all the time…*
- Diversity: *some people are more popular, some friendships are more important…*
- Node Complexity: *Endpoints have different CPUs, Windows is a complicated OS, Mobile devices …*
- Emergent phenomena: simple end behavior ➔ leads to ➔ complex system-wide behavior.
  - *If we understand the basics of climate change, why is the weather so unpredictable?*

Network Structure

- “Six degrees of Kevin Bacon”
- Milgram’s experiment in 1970

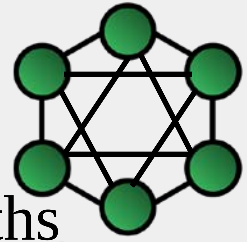
- Recent work on shows similarities between the structures of: Internet, WWW, human social networks, p2p overlays, Electric power grid, protein networks
- These networks have “evolved naturally”
- Many of these are “small world networks”

Two Important Network Properties

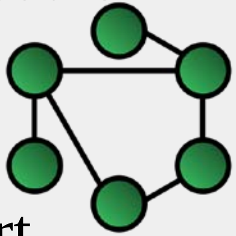
1. Clustering Coefficient: CC

   Pr(A-B edge, given an A-C edge and a C-B edge)

2. Path Length of shortest path

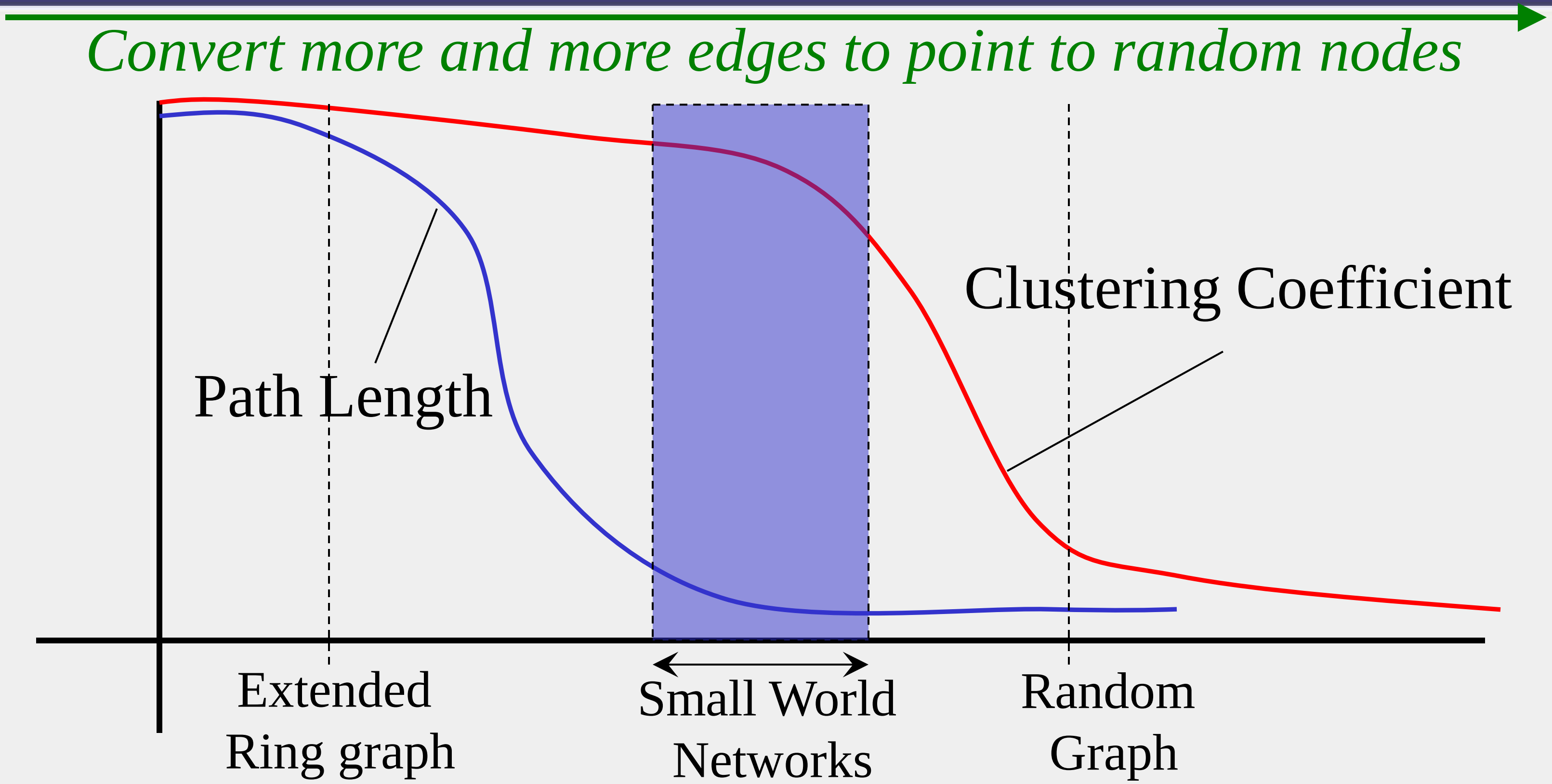


- Extended Ring graph: high CC, long paths



- Random graph: low CC, short paths
- Small World Networks: high CC, short paths

Deriving Small-world Graphs



Small-world Networks All Around

Most “natural evolved” networks are small world

- Network of actors → six degrees of Kevin Bacon
- Network of humans → Milgram’s experiment
- Co-authorship network → “Erdos Number”
- World Wide Web, the Internet, …

Many of these networks also “grow incrementally”

“Preferential” model of growth

- When adding a vertex to graph, connect it to existing vertex v with probability proportional to num_neighbors(v)

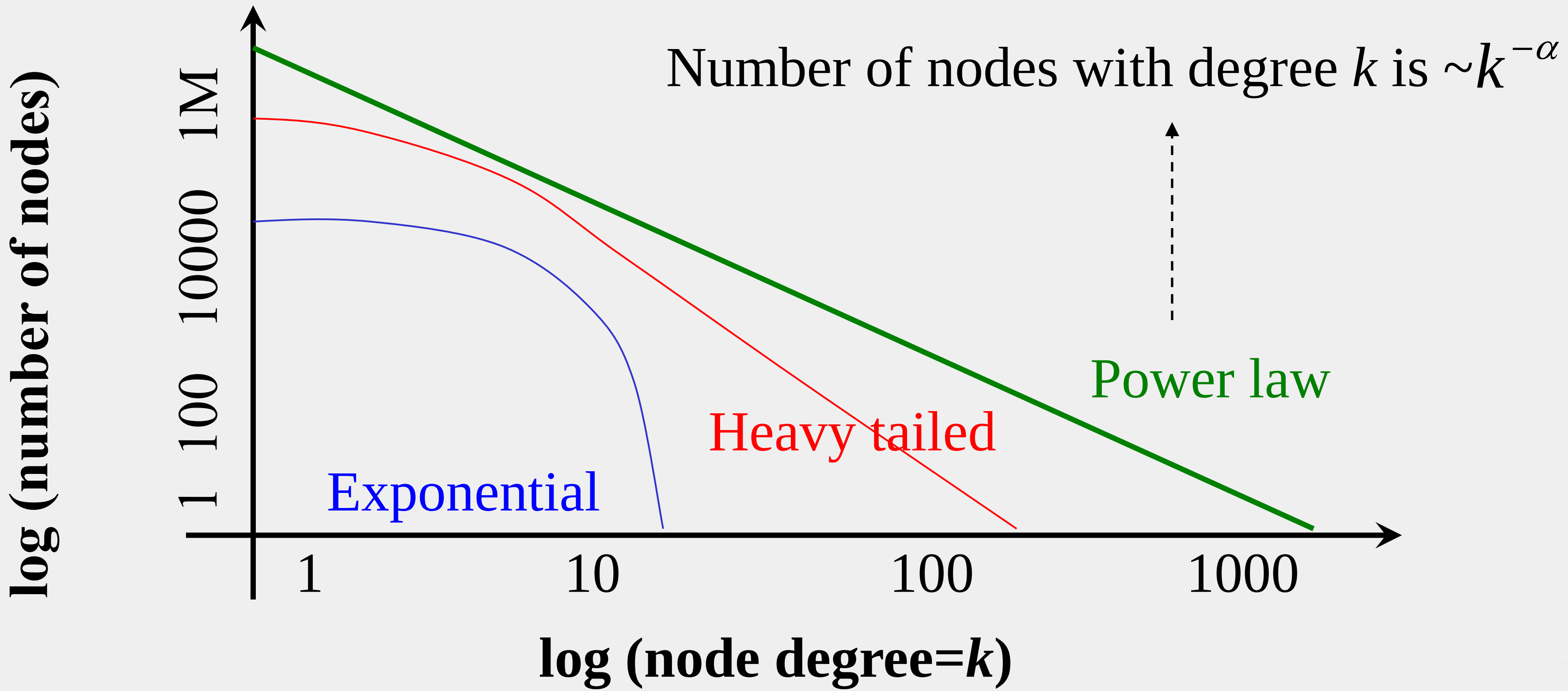
Degrees

Degree of a vertex = number of its immediate neighbor vertices

Degree distribution – what is the probability of a given node having k edges (neighbors, friends, …)

- Regular graph: all nodes same degree
- **Gaussian**
- Random graph: **Exponential** $e^{-k.c}$
- **Power law:** $k^{-\alpha}$

Power Law Graphs



Small-world and Power-Law

- A lot of small world networks are power law graphs
  - Internet backbone, telephone call graph, protein networks
  - WWW is a small-world graph and also a power-law graph with α=2.1-2.4
  - Gnutella p2p system network has heavy-tailed degree distribution
- Power law networks also called *scale-free*
  - Gnutella has 3.4 edges per vertex, *independent of scale (i.e., number of vertices)*

Small-world $\neq$ Power-Law

- Not all small world networks are power law
  - E.g., co-author networks
- Not all power-law networks are small world
  - E.g., Disconnected power-law networks

Resilience of Small-world+Power-Law

Most nodes have small degree, but a few nodes have high degree

Attacks on small world networks

- Killing a large number of randomly chosen nodes does not disconnect graph
- Killing a few high-degree nodes will disconnect graph

"A few (of the many thousand) nutrients are very important to your body"

"The Electric Grid is very vulnerable to attacks"

Routing in Small-world/Power-law Networks

- Build shortest-path routes between every pair of vertices
- => Most of these routes will pass via the few high-degree vertices in the graphs
  - => High-degree vertices are heavily overloaded
  - High-degree vertices more likely to suffer congestions or crash
- Same phenomenon in Electric power grid
- Solution may be to introduce some randomness in path selection; don't always use shortest path

Summary

- Networks (graphs) are all around us
  - Man-made networks like Internet, WWW, p2p
  - Natural networks like protein networks, human social network
- Yet, many of these have common characteristics
  - Small-world
  - Power-law
- Useful to know this: when designing distributed systems that run on such networks
  - Can better predict how these networks might behave