Countries = nodes
Treaties = edges

The Internet (Internet Mapping Project, color coded by ISPs)

Food Web of Little Rock Lake, WI

Electric Power Grid
Metabolic reaction network
This Lecture: Common Thread

Networks
   – Structure of,
   – Dynamics within,

• We’ll study networks at three different “levels”

Complexity of Networks

• Structural: human population has 6 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: PCs have different CPUs, Windows is a complicated OS...
• Emergent phenomena: simple end behavior ➔ complex system-wide behavior. If we understand the basics of climate change, why is the weather so unpredictable?

1. Network Structure

• “Six degrees of Kevin Bacon”
• Milgram’s experiment in 1970
• Watts and Strogatz Model
• Kleinberg’s algorithmic results

• Recent work on mapping Internet, WWW, p2p overlays, electric power grid, protein networks, co-authorship among scientists
• These networks have “evolved naturally”

A Scientist’s Perspective

• Two important metrics
  – Clustering Coefficient: CC
    • Pr(A-B edge, given an A-C edge and a C-B edge)
  – Path Length of shortest path
• Ring graph: high CC, long paths
• Random graph: low CC, short paths
• Small World Networks: high CC, short paths
Most “natural evolved” networks are (probably) small world
- Network of actors → six degrees of Kevin Bacon
- Network of humans → Milgram’s experiment
- Co-authorship network → “Erdos Number”
Many of these networks also “grow incrementally” [Faloutsos and Faloutsos]

Another Scientific Viewpoint
That was about “nature of neighbors”; what about number of neighbors?
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** \( e^{-k \gamma} \)
- **Power law** \( k^{-\gamma} \)

Power law vs. Small World
- A lot of small world networks are power law graphs (Internet backbone, telephone call graph, protein networks)
- Not all small world networks are power law (e.g., co-author networks)
- Not all power law networks are small world
- **Preferential Model** for network growth generates power law distributions – special way of incremental growth
  - e.g., Web pages linking to each other
- Power law networks also called scale-free
Power law + Small world

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 terrorists"

2. Network Dynamics

 • Strogatz goes on to discuss dynamics of many “natural networks”
 • We’ll focus on dynamics w.r.t. the Internet and P2P networks in the papers [Akella et al] and [Ripeanu et al]
 • But let’s just touch a bit on oscillation dynamics in networks…

Networks of coupled dynamical systems

 • If each node is a dynamical system, and is affected by its neighbors, what behaviors emerge from the entire network?
 • E.g., Social networks, network of neurons in the brain, protein networks, …
 • An example of emergent behavior: self-synchronization

Self-Synchronizing Fireflies

 • Synchronizing Fireflies of Malaysia
 • Each firefly: \( \dot{\theta} = \omega \)
 • Is driven by an external stimulus \( \dot{\Phi} = \Psi \)
 so \( \dot{\theta} = \omega + A \sin(\Phi - \theta) \)
 • Can show self-sync occurs when \( \omega - A < \Psi < \omega + A \)
 • For more details see [Strogatz’s non-linear dynamics textbook]

Why the heart beats by itself

 • Consists of a few thousand sinoatrial cells
 • Each oscillating at its own frequency
 • Peskin’s model: when a cell fires, all other cells have a small jump in voltage
 • [Why does this self-synchronize?] – Think of two sinoatrial cells first
 • For more details, see [Strogatz’s book “Sync”]
Discussion

• What is one problem where a self-synchronizing system could be used to design a distributed protocol?
• Why is the co-authorship network different from the Internet if both follow an incremental / preferential construction?

A Level Up: The Internet

• [Faloutsos et al] showed that the Internet backbone follows a power law distribution
• [One kind of Dynamism over such a network?]

Main Result

• Take a power law network (node has degree $k$ with probability $k^{-\alpha}$)
• Shortest path routing, with ties broken by higher degree
• With uniform traffic model for all pairs of end nodes, maximum edge congestion grows as $O(n^{\frac{1}{\alpha}})$ [Theorem 1]
Policy Routing

• Due to ISP to ISP financial contracts, AS to AS edges are either
  – Customer-provider edges, or
  – Peering edges
• Policy routing prefers customer → provider traffic
• Gives “valley free” paths: most edges are customer → provider

Clout Traffic Model: well-connected nodes generate more end to end traffic

+ Shortest-path routing has worse max edge congestion than with uniform traffic.

Clout Traffic Model: well-connected nodes generate more traffic

+ Policy routing also gives superlinear growth, but is better than shortest path routing

A Solution: Add redundant edges between selective node pairs
Discussion

• Metrics: max edge congestion (why not average?)
  – Instability from single source could spread
    • Think “Electric Power Grid failures”
    • Think “self-synchronizing routers”
• Why is Shortest Path Routing always worse than Policy Routing?
  – Shortest Path Routing is supposed to be “efficient”
  – Outrageous Opinion: Are policies the reason why the Internet stays up and robust? Should the design of Internet be left to non-technicians?

Another Level Up: Applications

Study of Gnutella

• [Ripeanu et al]
• Gnutella
  – Peer to peer Overlay
  – Users download songs from other users, upload their own songs
  – Each computer host = “peer”
  – Completely decentralized

Gnutella Structure

- Servents (“Peers”)
- Connected in an overlay graph
- Store their own files
- Also store “peer pointers”
- Queries flooded out, ttl-restricted

Study of Gnutella

• 6 month period 10/00-5/01
  – Before revision of Gnutella protocol (late 2001)
• Characteristics
  – System Size
  – Network Traffic
  – Node Connectivity

Gnutella Network Growth

95% of nodes in largest connected component
Quick Growth over time (exponential?)
Spikes
Churn Characteristics

- 40% of nodes logged in for less than 4 hours
- 25% nodes alive for more than 24 hours

Traffic Volume

- 170 K connexns for 50 K node Gnutella
- 6 KBps per connexn → 1 GBps total → 330 TB/month
- 1.7% of total traffic in Internet Backbone
- Recall: [3/00] 25% UWisc traffic from Napster

Average degree of node is scale-independent
On average 3.4 edges / node

Traffic Volume

- 170 K connexns for 50 K node Gnutella
- 6 KBps per connexn → 1 GBps total → 330 TB/month
- 1.7% of total traffic in Internet Backbone
- Recall: [3/00] 25% UWisc traffic from Napster

Average degree of node is scale-independent
On average 3.4 edges / node
Overlay-Network Match

- Does the Gnutella Overlay reflect the underlying Internet structure?
- Entropy technique in paper
  - Nodes identified with their domain names
  - Gnutella graph structure \( \rightarrow \) clustering of nodes
  - Calculate entropy of above clustering and compare with entropy of a random selection of nodes from across domains
  - If same, Gnutella graph is random, otherwise it is more ordered
- Authors find Gnutella clustering entropy to be 8% lower than random clustering entropy
- *Gnutella structure is independent of underlying Internet* (hence the term "application overlays")

Discussion

- Do overlays really reflect the application?
  - Concept of malleable overlays
- Are application-dependent overlays "unfriendly" to the network and other applications?
- What if the overlays are very large? (think PlanetLab)
- What if they are small and there are millions of them? (think overlay hosting services)
- What if they are large *and* many? (think overlay hosting services on top of PlanetLab-style clusters)

Another level Up: The Users, Humans, …

Summary

- Humans, and the networks connecting them… societal networks, actor networks, co-authorship graphs…
- And we’re back full circle!
- We’ve discussed
  - Network structure
  - Network dynamics
- Many commonalities
  - power law, small world … among “naturally evolved” networks
  - social nets, metabolic nets, electric power grid, Internet, WWW,…
- Can look at in awe, but systems design also has to deal with it
- Large opportunity to design distributed protocols inspired by nature (rigorously!) …

Discussion

- Questions?

Final Report Grading

- Final report will be graded just like a conference reviewer would:
  - Importance of problem
  - Novelty of solution
  - Evaluation of solution
  - Clarity of Presentation
  - Nits (grammar, references, etc.)
Semester’s Final Lecture

• 3 papers from website + 2 handouts
  – Read as many as you can – you’ll enjoy them
  – None of them is technical!
• Next Tuesday’s lecture (last lecture) – we’ll close the discussion we started in the semester’s first lecture