Peer-to-peer networks

Pioneers, self-organisation, small-world-phenomenons

Patrick Baier

Ferienakademie im Sarntal 2008
FAU Erlangen-Nürnberg, TU München, Uni Stuttgart

September 2008

Overview

1. Introduction
2. The first peer-to-peer networks
   - Napster
   - Gnutella
3. Self-organisation
   - Definition
   - Pareto distributions
   - Small world phenomens
4. Outlook

What is a peer-to-peer network

Types of networks I

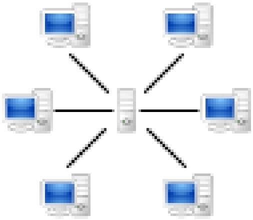

Figure: Client server based network, *Wikipedia.de*

What is a peer-to-peer network

Types of networks II

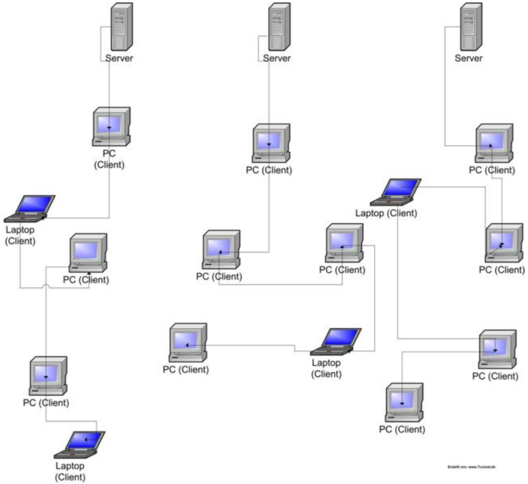

Figure: Peer-to-peer network, *Wikipedia.de*

What is a peer-to-peer network
Definition

- A connection between several network participants, for exchanging data.
- Two particpants directly communicate with each other.
- Nodes are called *peers*.
- All peers have equal rights.

Overview

- Released by *Shawn Fanning* in 1999.
- Actually a very simple protocol.
- First massively popular peer-to-peer network (download of the year 2000).
- It allowed to share audio data in the *.mp3* format with thousands of other users.
- Because of offering copyright protected files, Napster was sued severel times.
- Therefore, Fanning started a cooperation with Bertelsmann Ecommerce in the year 2000.
- Today Napster is based on a client-server architecture.

Napster
Design I

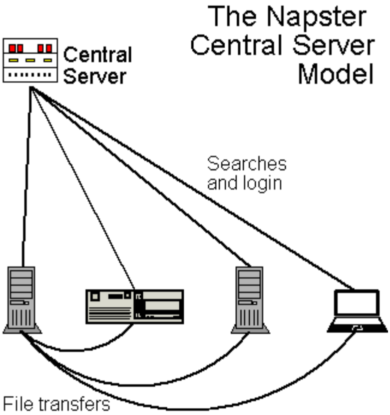

Figure: Napster infrastructure

Napster
Design II

- Napster is actually no peer-to-peer network (a server is necessary!).
- Peer tells server which media it offers and its address.
- Server keeps track of all peers and their media files.
- Procedure:
  - Peer contacts server with a request for a file.
  - Server responds with a list of addresses containing peers sharing the requested media.
  - The requesting peer contacts a peer from the list it receives and downloads the file directly.

Napster

Assets and drawbacks

- Assets:
  - Simple design concept.
  - Good solution for distributing files over several nodes.
- Drawbacks:
  - Server is a single point of failure.
  - Procotol is not scalable.

Overview

- Developed by *Justin Frankel* and *Tom Pepper* in early 2000.
- Gnutella protocol is open-source.
- Therefore, many different clients arose within the next years.
- Most famous clients: *LimeWire* and *Morpheus*.

- Gnutella is a real peer-to-peer network (there exists no server).
- The protocol has an important parameter (*TTL*) to control the network structure.
- First problem is that the connection setup for a peer is complicated.
- Which peer should be called on connection setup, avoiding that the connected peer becomes a server?

Gnutella
Design II

- Solution: *Bootstrapping*
  1. Initially a peer has a preexisting list with peers.
  2. On startup, peer contacts all other peers from list until an active peer answers.
  3. The contacted peer sends the message further and so on, until peer *TTL* away is reached.
  4. Every peer on this path sends a response back to the starting peer.
  5. Starting peer updates his neighborhood list for the next start.

Gnutella
Design III

- Use of five basic messages while operating (used on top of *TCP*):

Ping	Used for finding other nodes.
Pong	Response to the ping-message.
Query	For querying the network for files.
QueryHit	Answer to a query, when the node shares the queried file. Contains the IP and port address of the sender.
Push	Technical message for sharing behind firewalls.

Design IV

- After sending the Ping-message on startup, the peer receives Pong-messages.
- The peer chooses randomly from the answering peers *k* as its neighbors.
- A typical value for *k* is 5.
- The Query-message is send to all neighbors and so do they, until peer *TTL* away is reached.
- If one of this peers answers with a QuerHit-message a direct connection is established and this is added as neighbor.

- Assets:
  - No single point of failure (no server).
  - Scalable and stable structure.
- Drawbacks:
  - Peer can only reach the peers *TTL* away.
  - Therefore, rare files might not be found in large networks.

Overview

1 Introduction

2 The first peer-to-peer networks
- Napster
- Gnutella

3 Self-organisation
- Definition
- Pareto distributions
- Small world phenomens

4 Outlook

"...the ability of systems comprising many units and subject to constraints, to ***organize themselves*** *in various spatial, temporal or spatiotemporal activities. These emerging properties are pertinent to the system as a whole and cannot be seen in units which comprise the system..."*

- A. Babloyantz

Overview

Pareto distributions

Short introduction

- Several measurements have analyzed the average number of neighbors of a peer in a Gnutella network.
- As a result, the average number of peers with d neighbors can be approximated by $\frac{C}{d^k}$, with C and k as network specific constants.
- Such a relation is called a *power law*...
- ... and the according probability distribution is called a *pareto distribution*.
- Pareto distributions are an often observed phenomen in social processes.
- **Remember:** This is not the result of an algorithm, it actually is the result of a social phenomen.

Pareto distributions
The diameter of Gnutella

- Only asking the next *TTL* neighbors, we must consider the diameter of the network.
- Actually five measurements in the year 2000 have shown:

$$\text{Diameter of Gnutella} \approx 8...12 \text{ peers}$$

- This also can be explained by a social phenomen.

Overview

Small world phenomens

Description I

- Idea: Every member in a social networks is connected with every other member only over a small chain of members.
- Phenomen was first described by *Stanley Milgram* who made the following experiment:
  - Some letters were given to 60 people in Omaha, Nebraska addressed to a special destination in Sharon, Massachusetts.
  - The letter could only be forwarded to known persons.
- Result:
  - Most of the letters actually arrived.
  - The average number of station passed was only 5.5 .

Small world phenomens
Description II

- Networks with a small diameter are therefore called small-world networks.
- To understand how they emerge, we look at three modeling approaches:
  - Watts und Strogatz's approach
  - Kleinberg's approach
  - Barabasi und Albert's approach

Small world phenomens

Watts und Strogatz's approach I

- Starting with a ring network with *n* nodes.
- Every node is connected to the next *k/2* neighbors to the left and the right.
- Network consists of cliques and has a relatively large diameter.
- Now we replace every edge with the probability $p \in [0, 1]$ by a random edge, leading to a random node.
- Result (for small *p*):
  - Most of the cliques persist.
  - The diameter of the network decreases significantly.

Small world phenomens

Watts und Strogatz's approach II

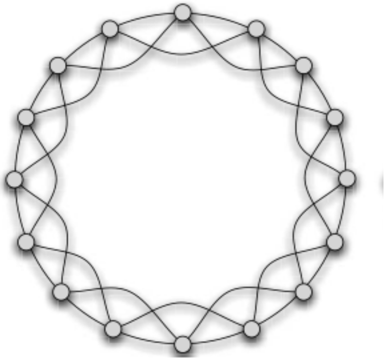

Figure: Starting with $k = 4$

Small world phenomens

Watts und Strogatz's approach III

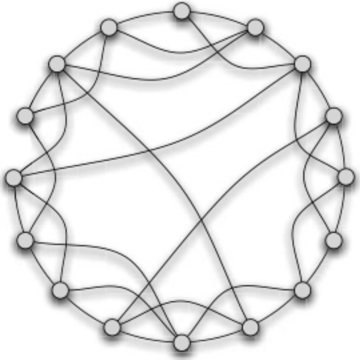

Figure: After alternating some edges

Small world phenomens

Kleinberg's approach I

- Network is a grid network.
- Every node is connected to its direct neighbor.
- Special distant edges are added according to a special probability distribution.
- Kleinberg could prove that the diameter of the network is within $\mathcal{O}(\log^2 n)$.

Small world phenomens

Kleinberg's approach II

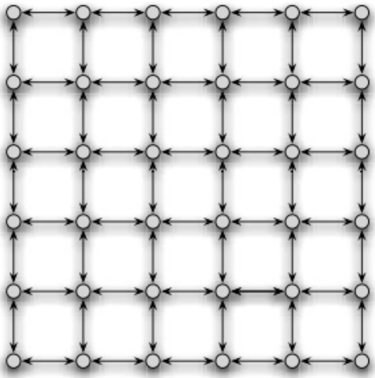

Figure: Starting position

Small world phenomens

Kleinberg's approach III

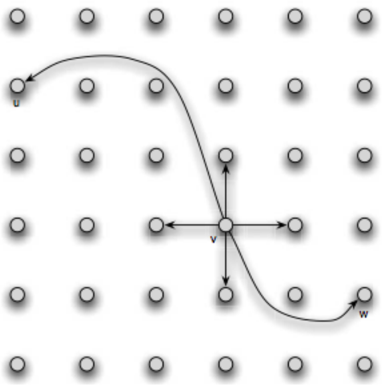

Figure: Adding distant edges

- Network is a small arbitrary graph.
- New nodes are added with m (as a constant) edges.
- Edges of the new nodes lead to old nodes according to a probability distribution, in which nodes, which already have many edges, are more likely to get a new one (rich gets richer).
- They could prove that the diameter of such a network is within $\mathcal{O}(\log n)$.

Small world phenomens

Barabasi und Albert's approach II

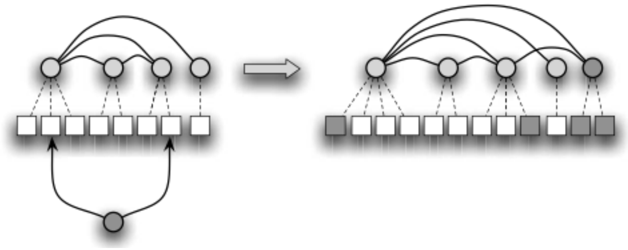

Figure: Adding a node

In the year 2000 several measurements took place, which compared these three approaches with the actual Gnutella network, regarding the average distance between two peers. As a result we can say that:

Barabasi und Albert Biggest correspondence with Gnutella. This is justified by the similar methods used in this approach and the real Gnutella protocol.

Watts Strogatz Only moderate correspondence with Gnutella.

Kleinberg Only few correspondence with Gnutella.

Outlook

- Napster and Gnutella aren't in use any more.
- But the basic concepts of Napster and Gnutella are still in use today.
- Gnutella was detached by *Kademila*.
- Kademila introduced the concept of distributed hash tables.
- Kademila is still popular today. Famous clients: *eDonkey*, *BitTorrent* and *Azureus.*