

# Peer-to-peer networks: pioneers, self-organisation, small-world-phenomenons

Patrick Baier

October 10, 2008

## Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introduction</b>                            | <b>1</b> |
| 1.1      | Preamble . . . . .                             | 1        |
| 1.2      | Definition . . . . .                           | 1        |
| <b>2</b> | <b>The first peer-to-peer networks</b>         | <b>2</b> |
| 2.1      | Napster . . . . .                              | 2        |
| 2.1.1    | History . . . . .                              | 2        |
| 2.1.2    | Design . . . . .                               | 2        |
| 2.1.3    | Assets and drawbacks . . . . .                 | 2        |
| 2.2      | Gnutella . . . . .                             | 2        |
| 2.2.1    | History . . . . .                              | 3        |
| 2.2.2    | Design . . . . .                               | 3        |
| 2.2.3    | Assets and drawbacks . . . . .                 | 4        |
| 2.3      | Conclusion . . . . .                           | 4        |
| <b>3</b> | <b>Self-organisation</b>                       | <b>4</b> |
| 3.1      | Definition . . . . .                           | 4        |
| 3.2      | Pareto distribution . . . . .                  | 4        |
| 3.3      | The diameter of the Gnutella network . . . . . | 5        |
| 3.4      | Small world phenomen . . . . .                 | 5        |
| 3.4.1    | Watts und Strogatz's approach . . . . .        | 5        |
| 3.4.2    | Kleinberg's approach . . . . .                 | 6        |
| 3.4.3    | Barabasi und Albert's approach . . . . .       | 6        |
| 3.4.4    | Gnutella and small-world networks . . . . .    | 6        |
| 3.4.5    | Conclusion . . . . .                           | 6        |
| <b>4</b> | <b>Outlook</b>                                 | <b>6</b> |

# 1 Introduction

## 1.1 Preamble

Within the last decade a phenomenon appeared, which changed the way of exchanging data over the internet. Peer-to-peer networks made it possible to easily find and exchange any kind of data over a network, without even knowing the person who is offering the data. Together with the rise of the audio format MP3, peer-to-peer networks encouraged the exchange of audio files which are protected by copyright laws. Faced by this new way to get music for free, a wave of protests and suits were organised by several music companies. But instead of closing down the existing peer-to-peer networks, more and more new networks arose and more and more people became aware of this new technology.

To give a fundamental introduction into this technique this seminar paper gives, on the one hand, a short introduction to the historical important peer-to-peer networks (*Napster* and *Gnutella*) and explains, on the other hand, the important concept of self-organisation.

## 1.2 Definition

A peer-to-peer network is, as every usual network, a connection between several network participants (*nodes*), for exchanging data. The key feature of a peer-to-peer network is that two participants directly communicate with each other, instead of using a central instance for the communication (like in a server based network). Therefore, a peer-to-peer network only contains *peer nodes* and all nodes have equal rights. The nodes are offering different kind of services which can be used by other nodes.

# 2 The first peer-to-peer networks

## 2.1 Napster

In 1999 *Shawn Fanning* released his first version of the famous peer-to-peer network *Napster*. Despite Napster is not a real peer-to-peer network, it introduced the basic concepts of peer-to-peer networks and laid the foundation for the following peer-to-peer networks.

### 2.1.1 History

With its publication in the year 1999, Napster was the first massively popular peer-to-peer network, which made it possible to easily share audio data in the .mp3 format with thousands of other users. According to this fact, Napster become as soon as possible a very popular application.

At the end of the year 2000 Fanning started a cooperation with the Bertelsmann Ecom-

merce and started to rearrange Napster to a chargeable system which is now mainly based on a client-server architecture.[Wik08b]

### 2.1.2 Design

Although Napster is considered to be the first peer-to-peer network, it is actually based on a client-server architecture. Its concept is pretty simple and uses the peer-to-peer approach only very rudimentary.

Each peer offers his media data to the whole network. A central server keeps track of all peers and the media offered by them. If a peer requests a special media file, it first contacts the server, which replies a list with peers offering the requested file. The peer connects know to one of the peers in the list and directly downloads the file from the peer. [Wik08b]

### 2.1.3 Assets and drawbacks

The drawback of this solution is the still existing client-server architecture, because the server is a single point of failure, it can be aim of a denial-of-service attack, for example. With regard to throughput, a server is something like a bottleneck, it doesn't scale when a large number of users are trying to contact the server.

The asset from Napster is it simple concept as shown in section 2.1.2. [MS07]

## 2.2 Gnutella

Impressed by Napsters technology and its popularity, the Gnutella peer-to-peer network became, one year after the release of Napster, the first important network, which exclusively used the peer-to-peer technique without a central server. Although it couldn't reach the popularity of Napster, Gnutella showed that a peer-to-peer network could operate without a single server instance.

### 2.2.1 History

The Gnutella network was developed by Justin Frankel and Tom Pepper in early 2000. Since the Gnutella source code was open-source, many different clients were developed within the next few years, using the Gnutella protocol. The most popular ones were *LimeWire* and *Morpheus*, which were mainly used to share media data. But instead of sharing only audio files, these clients also allowed to share other data, like movies and applications. [Wik08a]

### 2.2.2 Design

An operating Gnutella networks uses five basic messages for the communication between its nodes:

**Ping** Is used like in the most common operating systems. It helps to find other nodes in the peer-to-peer network. Once a node receives a *ping-message* it responds with a *pong-message*.

**Pong** Is send after a node has received a *ping-message*. It includes the nodes address and a list of its files which are shared in the network.

**Query** For searching the network for special files, a node can use the *query-message*. If a node has a file which fits the queried data, it responds with a *queryhit-message*.

**QueryHit** Is the response for the *query-message*. It includes the nodes address and a details about the queried files.

**Push** Is a technical way of enabling the sharing of data from nodes behind a firewall.

The first problem when operating in a network without a central node is the *bootstrapping problem*. It deals with the challenge to find a first node, which can be contacted, for joining the network. The most common idea of the Gnutella clients is a pre-existing address list of nodes operating in the network.

When the client wants to join the network, it contacts successively all nodes on his pre-existing address list, until the first node responses by sending a *ping-message* to them. After that, this node sends a message to all of his neighbours, which send again a message to their neighbours and so on. The protocol defines a constant *TTL* (*Time to Live*) which influences how many neighbours should be asked, starting from the first node. Every node decreases the *TTL*-value before sending the message to its neighbours, until the *TTL*-fields reaches zero. Every found neighbour sends a response to the starting node using the *pong-message*. The starting node then generates a new list of active peers, which can be used for the next start mechanism instead of the old one.

Therefore the structure of the graph follows a random schema, which is determined by the active peers and their neighbours. Later we will see, that with a *TTL* of five, a user can almost reach the entire network by the so called *small world phenomenon* . [MS07]

### 2.2.3 Assets and drawbacks

The assets of Gnutella, compared to Napster, is its decentralised structure with no single point of failure. Therefore a crashing node doesn't cause a crash of the entire network and the network is more scalable.

The drawback of the introduced technique is, that a peer can only reach the next *TTL* neighbours. Assuming that the *TTL* is five and a very rare file is shared by a node *b* which is the sixth node on the shortest path to node *a*. If node *a* is looking for that rare file, it can't find it although the file is inside of the network, because it's too far away from node *a*. [MS07]

## 2.3 Conclusion

After introducing the two pioneers of peer-to-peer networks and showing their capabilities and problems, we can conclude that the peer-to-peer concept is a very efficient and robust way to work without a central instance such a server. This technique enables sharing of a very large amount of data, distributed over many peers, without a significant loss of performance. In the next section a very important concept of peer-to-peer networks is introduced, the self-organisation.

# 3 Self-organisation

## 3.1 Definition

Self-organisation in the fields of peer-to-peer networks can be seen as mechanisms to organise the structure of the network by the network itself, which means that it is organised by its peers. There are a great range of different definitions, one that fits the term of self-organisation with regard to peer-to-peer networks very good can be found in [Bab91]:

”...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...”

To show a way of self-organisation the concept of the pareto distribution is shown in the following.

## 3.2 Pareto distribution

If we consider the number of neighbours of a peer in a Gnutella network, several analysis have shown that the number of peers with  $d$  neighbours is linearly dependent on the number of the neighbours  $d$ . With the two constants  $k$ ,  $C$  and with  $y$  as the number of Peers with  $d$  neighbours we get:

$$y = \frac{C}{d^k}$$

The probability distribution of such a relation is called a pareto distribution. The characteristic of such a distribution is the *heavy tail* property, which means that a relatively small part of the value set contributes more to the overall value than the larger number of small values.

If the number of neighbours in a graph is pareto distributed, we call it a *pareto graph*. Later we will see, that the structure of a Gnutella network really follows a pareto graph.

### 3.3 The diameter of the Gnutella network

As already mentioned in 2.2.3, a Gnutella client asks only its next *TTL* for a certain file. Therefore if the diameter of a average Gnutella network is very large, a client might not be linked to a queried file, although this file is within the network. Actually an empiric measuring in the year 2000 showed that the average diameter of the Gnutella network is between 8 and 12. To understand how such a small diameter can emerge, although the average number of participants in such a network is very high, we now look a theory called the *small world phenomen*.

### 3.4 Small world phenomen

The small world phenomen describes a phenomen in social networks, which claims that every member of the network knows every other member of the network only over a small chain of other members.

The idea was formulated by *Stanley Milgram* who observed this phenomen in a simple experiment. Milgram gave 60 letters with a specific destination to different persons who should forward their letter to the destination, only by forwarding the letter to a person which is known to them. Most of the letters actually arrived and the average number of stations passed was only 5.5. According to this result, Milgram formulated the theory of the *six degrees of separation* which claims that every person is known to an other person only over six other persons worldwide. Therefore networks with such a small diameter are called small-world networks.

To understand how such networks emerge, we look at three approaches to model this phenomen.

#### 3.4.1 Watts und Strogatz's approach

Watts und Strogatz tried to find a description for a random network with a relatively small diameter. They started by creating a ring network with  $n$  nodes and every node is connected to the next  $k/2$  neighbours on the left and on the right side. As a consequence, the network consists of a set of clique and still has a relatively large diameter. The idea was now to replace every edge of the network with probability  $p \in [0, 1]$  by a random edge, which leads to a random node. As a result, the diameter of the network decreases significantly while most of the cliques persist, if a small value for  $p$  is chosen.

#### 3.4.2 Kleinberg's approach

Kleinberg's network modell consist of a grid network in which every node is connected to its direct neighbour. By adding special edges, called *distant edges*, which connect nodes over a longer distance he could show, that diameter of the network is bounded by  $\mathcal{O}(\log^2 n)$ . The choice, which distant edges are actually added to the network, is made by a special probability distribution, which considers the distance between two nodes.

### 3.4.3 Barabasi und Albert's approach

Starting with a small and arbitrary graph, new nodes are added to the networks with  $m$  edges. These new nodes choose their edges with a probability to old nodes, which considers the numbers of neighbours of these old nodes. This means that nodes, which already have a large number of neighbours, are more likely to get new neighbours. It can be shown that the diameter of the resulting network is bounded by  $\mathcal{O}(\log n)$ .

### 3.4.4 Gnutella and small-world networks

Now we can compare these three approaches with the empirical found values for a Gnutella network. Matching the individual parameters of the three models to the real Gnutella parameters we can compare the different approaches by comparing the *characteristical path length*, which is the average distance of two nodes in the network.

The following results show how good each of the three approaches fits the real Gnutella network:

**Barabasi und Albert** This approach has the biggest correspondence with the real Gnutella network. This is mainly justified by the reason how new nodes are added, which fits the method used by the Gnutella network.

**Watts Strogatz** This approach has only a moderate degree on correspondence with the Gnutella network.

**Kleinberg** This approach has the fewest correspondence with the real Gnutella network.

### 3.4.5 Conclusion

The results in 3.4.4 showed that the real Gnutella network is similar to the three shown approaches, which are all pareto distributed small-world networks and therefore has a relatively small diameter. Hence the drawback, mentioned in 2.2.3, doesn't matter if the diameter of the networks is small.

## 4 Outlook

Although the Napster and Gnutella protocol are out of date and not used any more, the main ideas and concepts of these protocols are still implemented in the latest peer-to-peer protocols.

Gnutella was gradually detached by the *Kademila* protocol, which resolved some weaknesses of the Gnutella protocol and introduced the concept of distributed hash tables. It most known clients are *eDonkey*, *BitTorrent* and *Azureus*, which are still very popular today.

## References

- [Bab91] A. Babloyantz. *Introduction in self-organization, emerging properties and learning*. Plenum, 1 edition, 1991.
- [MS07] Peter Mahlmann and Christian Schindelbauer. *Peer-to-Peer-Netzwerke*. Springer, 1 edition, 2007.
- [Wik08a] Wikipedia. Gnutella. <http://de.wikipedia.org/wiki/Gnutella>, July 2008.
- [Wik08b] Wikipedia. Napster. <http://de.wikipedia.org/wiki/Napster>, July 2008.