

# Applying Social Network Theory to a small Weblog Community

## Hubs, Power Laws and the Community Core

Michael Schuster

Knallgrau New Media Solutions GmbH

August 6<sup>th</sup>, 2004

---

### Abstract

Weblogs and Weblog Communities are interesting examples of Social Networks. Sites are connected through links that form a dense Network of people. Social Network Theory has developed rapidly throughout the last decade. Networks seem to be everywhere. The emergence of Weblog communities offers the possibility to look at such a network with reasonable resources. This paper analyses a small Weblog community and visualises the Network of people.

**Keywords:** Social Networks, Weblog Communities, twoday.net, Scale-Free Networks

---

## 1. Introduction

Weblogs and Weblog Communities are getting more and more popular, the phenomenon of blogging is gaining more ground every day. Really interesting about Weblogs is, that their very nature provides a great field to prove some basic concepts of social network theory, as Weblogs are nothing but a virtual social network, making implicit connections explicit through links, comments, stories and the blogroll. The goal of this paper is, to apply some of those concepts to a small Weblog community, named twoday.net (<http://www.twoday.net>), run by Knallgrau New Media Solutions GmbH, based in Vienna, Austria. twoday.net is providing Weblogs as a hosting service, currently hosting over 3.000 Weblogs with nearly 20.000 registered users. Compared to other hosting services it can be considered small, making it possible to analyse the community with reasonable amount of resources. The first part of the paper gives a short introduction into the Basics of Weblogs, the second sums up the governing principles of Social Networks, leading to the results of the analysis and the final conclusion. This paper has been prepared for the second International Conference on Weblogs, BlogTalk 2.0, which took place in Vienna.

## 2. Weblogs and what it is all about

For a detailed and nicely written introduction into the history of Weblogs it seems advisable to read Rebecca Bloods "Weblogs: a history and perspective".<sup>1</sup> The emergence of Weblogs is in some way a step back to the roots of the Internet. The way Tim Berners-Lee envisioned the Internet, it should be a space where we, as individuals, write, work, communicate and socialise:

*"The dream behind the Web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. There was a second part of the dream, too, dependent on the Web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize."*<sup>2</sup>

That is what people do in Weblogs. They form a virtual self, publish bits and pieces about themselves, their lives and their surroundings. They tell stories, rant, chat, curse and listen to others. As David Weinberger puts it:

---

<sup>1</sup> Blood, R. (2000)

<sup>2</sup> Berners-Lee, T. <http://www.w3.org/People/Berners-Lee/ShortHistory>

*"...the Web is binding not just pages but us human beings in new ways. We are the true 'small pieces' of the Web, and we are loosely joining ourselves in ways that we're still inventing."*<sup>3</sup>

It is this interconnected, heavily linked, intermingled nature that is a new way of looking at social networks.

Weblogs consist of postings or entries, made by the blogger (the owner of the Weblog), where she or he writes about something. Those postings are often consisting links to interesting things on the web, as the term Weblog (a "log" of the "web") indicates. It provides a path through the Internet:

*"Weblogs are often-updated sites that point to articles elsewhere on the web, often with comments, and to on-site articles. A Weblog is kind of a continual tour, with a human guide who you get to know."*<sup>4</sup>

On most Weblogs it is possible for others to comment or connect in some other way (e.g. Trackback) and most of the sites maintain something called a "blogroll", which is a list of frequently visited Weblogs giving the reader a view of the "neighbourhood" of the Blogger. Through this links the Bloggers form a network.

### **3. Social Networks and their governing principles**

For quite a few years academic research worked with a network theory developed already in 1959 when Erdős and Rényi described random networks.<sup>5</sup> They proposed that *"that such systems could be effectively modelled by connecting their nodes with randomly placed links."*<sup>6</sup> A large field of mathematics concentrating on this kind of networks emerged, until in 1998 physicists at the University of Notre Dame began to map the World Wide Web. They found out, that different to their expectations, the World Wide Web isn't a random network.<sup>7</sup> By measuring outgoing and incoming links of web pages, they discovered that the nature of that network is very different from being random. In a random network every node has the same probability of having  $n$  links, leading to a distribution where a large fraction of nodes has almost the same number of links. This leads to a bell curve type of distribution, well known as Gaussian distribution (See Figure 1).

---

<sup>3</sup> Weinberger D. (2003), p. X

<sup>4</sup> Winer D. (2002), <http://newhome.Weblogs.com/historyOfWeblogs>

<sup>5</sup> Cohen R., Shlomo H., ben-Avraham D. (2002), p. 3

<sup>6</sup> Barabasi A., Bonabeu E. (2003), p. 52

<sup>7</sup> Albert R., Jeong H., Barabasi A. (1999), p. 130

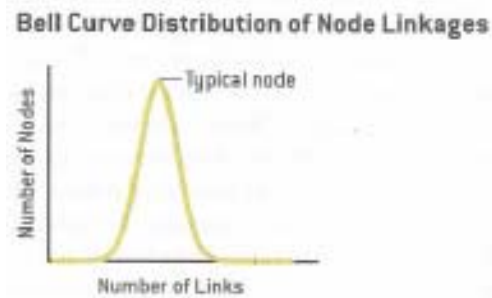


Figure 1: Bell Curve Distribution of Node Linkages in a random network

(Source: Barabasi A., Bonabeu E. (2003), p. 53)

The World Wide Web is following a different law, now known as "Scale-Free Network". The reason is simple: people in the WWW are free to link wherever they want. Therefore some sites get drastically more links than others. Quickly this principle had been labelled "preferential attachment",<sup>8</sup> meaning that the actors (nodes) of those networks choose carefully where to link, and some nodes are apparently more popular than others. Those nodes are described as Hubs.<sup>9</sup> The distribution within these networks follows a Power Law, where some nodes have almost all the links and most other nodes have few to no links. Printed on an Log-Log-Scale the distribution shows a straight line (See Figure 2).

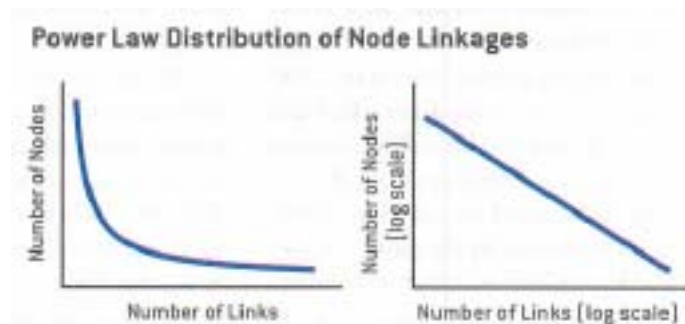


Figure 2: Power Law Distribution of Node Linkages in a Scale-Free Network

(Source: Barabasi A., Bonabeu E. (2003), p. 53)

---

<sup>8</sup> Barabasi A., Albert R. (1999); Jeong H., Neda Z., Barabasi A. (2002), p. 567

<sup>9</sup> Barabasi A., Bonabeu E. (2003), p. 54

The Power Law Distribution can be written as<sup>10</sup>

$$P(k) = c k^{-\lambda}, \quad k = m, \dots, K$$

where  $c$  is a normalization factor,  $k$  is the connectivity of a node and  $m$  and  $K$  are lower and higher cutoffs for this connectivity. The cutoff is necessary, because Power Law distributions are infinite by definition. Some governing principles can be derived from the structure of those networks, which shall be described briefly:

- **Rich get richer<sup>11</sup>**

In order to explain Hubs, the theory of preferential attachment proposes that those nodes that have more links have a higher chance to gain more links. This benefits nodes that are in the network longer but also those that seem to be “fitter”.

- **Small World Properties<sup>12</sup>**

Although some networks consist of million nodes, the diameter is rather small. This can be explained through the nature of scale-free networks themselves, because any node just needs connection to the hub, in order to reach many other nodes. Stanley Milgram carried out a famous experiment in 1967, discovering that everybody in the world could be linked to anybody else by just 6 links. Although this might not be true to really everybody, recent studies using emails instead of written letters have proven the theory.<sup>13</sup>

- **Clustering<sup>14</sup>**

A further development in the theory of scale-free networks has been the work of Watts and Strogatz on Clustering. They discovered that in a scale-free environment heavily connected clusters can exist, partly explaining the small-world phenomenon.

- **Achilles Heels<sup>15</sup>**

As a direct consequence of the structure mentioned above, Hubs play a crucial role in the stability and security of the network. The weakness of scale-free networks is not due to random nodes but the position and high connectivity of Hubs.

---

<sup>10</sup> Cohen R., Havlin S., ben-Avraham D. (2002), p. 4

<sup>11</sup> Barabasi A., Bonabeu E. (2003), p. 54

<sup>12</sup> Newman M. (2002), p. 41

<sup>13</sup> Later on the term „Six degrees of separation“ had been coined. More in Watts D. (2003)

<sup>14</sup> Newman M. (2002), p. 41

<sup>15</sup> Barabasi A., Bonabeu E. (2003), p. 55

- **Hubs and Diffusion**<sup>16</sup>

Diffusion of any kind of information through a network is very dependent on the Hubs in the network. Pastor-Satorras and Vespignani showed that Computer Viruses can remain active for a long time, just by residing at certain hubs and reinfecting nodes. This is interesting for any kind of epidemic spreading, from Computer Viruses to Diseases to Innovations.

- **Growing Difference**

The larger the network gets, the larger the difference between those nodes with high connectivity and those with low connectivity.

After the discovery of Scale-Free Networks many scientist embarked on analysing the nature of different networks. It has been proven that many networks have a Scale-Free Nature and follow a power law. This is true for the World Wide Web,<sup>17</sup> scientific collaborations,<sup>18</sup> Hollywood actors,<sup>19</sup> directors of companies,<sup>20</sup> sexual partnerships,<sup>21</sup> email communication,<sup>22</sup> most languages<sup>23</sup> and even cells and proteins.<sup>24</sup> It is therefore also quite sure that many other real world networks have scale-free nature. Although not proved economy itself could be viewed as a network of interdependent actors (nodes) with some very important hubs.<sup>25</sup>

#### **4. Methodology**

We analysed the content of all Weblogs hosted on twoday.net in June 2004. Stories (meaning the postings or entries), Comments and the Blogroll where scanned for Hyperlinks. Only links within the Hosting Service where counted (meaning all links going to a site that had "twoday.net" in the URL). Because it is not that trivial to determine where the Blogroll of a Website resides, we used the specific structure of twoday.net, where the pages is built out of little pieces of HTML Code, called Skins. They are much more structured than normal Templates available in Weblog Software, therefore we could

---

<sup>16</sup> Pastor-Satorras R., Vespignani A. (2001), p. 3201

<sup>17</sup> Barabasi A., Albert R., Jeong H. (2000)

<sup>18</sup> Barabasi A. et al. (2002); Newman M., Watts D., Strogatz S. (2002)

<sup>19</sup> Newman M., Watts D., Strogatz S. (2002)

<sup>20</sup> Newman M., Watts D., Strogatz S. (2002)

<sup>21</sup> Schneeberger A. et al. (2004)

<sup>22</sup> McArthur R., Bruza P. (2003)

<sup>23</sup> Steyvers M., Tenenbaum J. (2001)

<sup>24</sup> Barabasi A., Bonabeu E. (2003)

<sup>25</sup> Cohen R., Havlin S., ben-Avraham D. (2002), p. 20

localise the Blogroll quite well. To cater to the specifics of Weblogs, we used weights for the different pieces of content:

<b>Source</b>	<b>Information</b>	<b>Weight</b>	<b>Why?</b>
Comments	Links to other Weblogs, probably User Names?	40 %	Because commenting on a Weblog is the strongest link between two persons, meaning they engage in a conversation. You don't talk with a person, that you haven't got some kind of link to.
Skins	Links to other Weblogs	40 %	We assume that people that list other Weblogs in their skins have produced some kind of Blogroll (or equivalent). Listing somebody permanently on your Weblog is a obvious sign that there exists some kind of link between two persons.
Stories	Links to other Weblogs	20 %	Placing a link in a story/posting on a Weblog is some kind of link you establish, but it can as well be just a sign of dissent or a temporary link you want to establish.

Table 1: Explanation of Wheights

Due to limitations of the available Visualisation tool (we wanted to browse the data set), we had to set some thresholds. We computed two Data Sets with the following characteristics, to filter out those blogs that seem to be "dead" and those not linking to anyone else.

<b>Type</b>	<b>Number of Stories (Minimum)</b>	<b>Active longer than</b>	<b>Weight Threshold</b>	<b>Number of Links</b>	<b>Number of Weblogs</b>
Data Set 1	10	14 Days	Yes (3 %)	1	88
Data Set 2	10	14 Days	No	0	585

Table 2: Thresholds for the Data Sets

Data Set 1 had also a Weight Threshold, where lowest 3 % with the smallest weight where excluded from the Data Set. It is quite obvious that both Data Sets exclude dead Weblogs and fairly new Weblogs.

## 5. Analysis

For most of the Analysis we used the Data Set 2 to have more general results. For the visualisation we used mostly the Data Set 1, to get a result set showing the most active community members and their linkage, while still being able to browse the result set. We analysed incoming and outgoing links and found the following values

Type	Average	Median
Incoming	6 Links	2 Links
Outgoing	6,12 Links	3 Links

Table 3: Values for Incoming and Outgoing Links

Both distributions follow a Power Law, which can easily be seen in Figure 3 and Figure 4. Additionally we tried to estimate the distance between an average node in the network to any other average node using a Dijkstra algorithm. The shortest path for Data Set 1 is 5 steps, for Data Set 2 it is 6 Steps. This is also a hint to the Scale-Free Nature of the Network, as the distance grows slower than the number of nodes within the network.

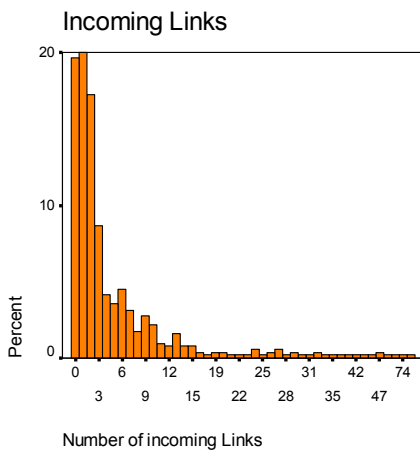


Figure 3: Distribution of Incoming Links

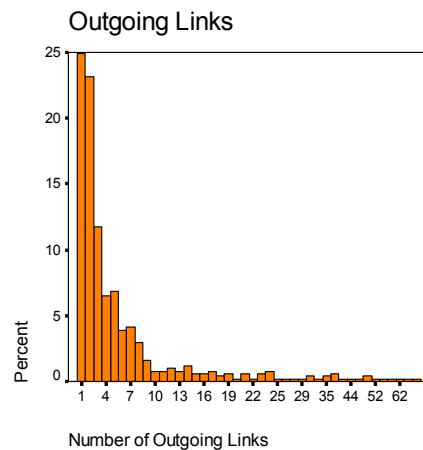


Figure 4: Distribution of Outgoing Links

For the visualisation we used Touchgraph, a freely available tool.<sup>26</sup> The Data Set 1 shown in Figure 5 can be downloaded at <http://www.knallgrau.at/blogtalk>.

---

<sup>26</sup> <http://www.touchgraph.com>

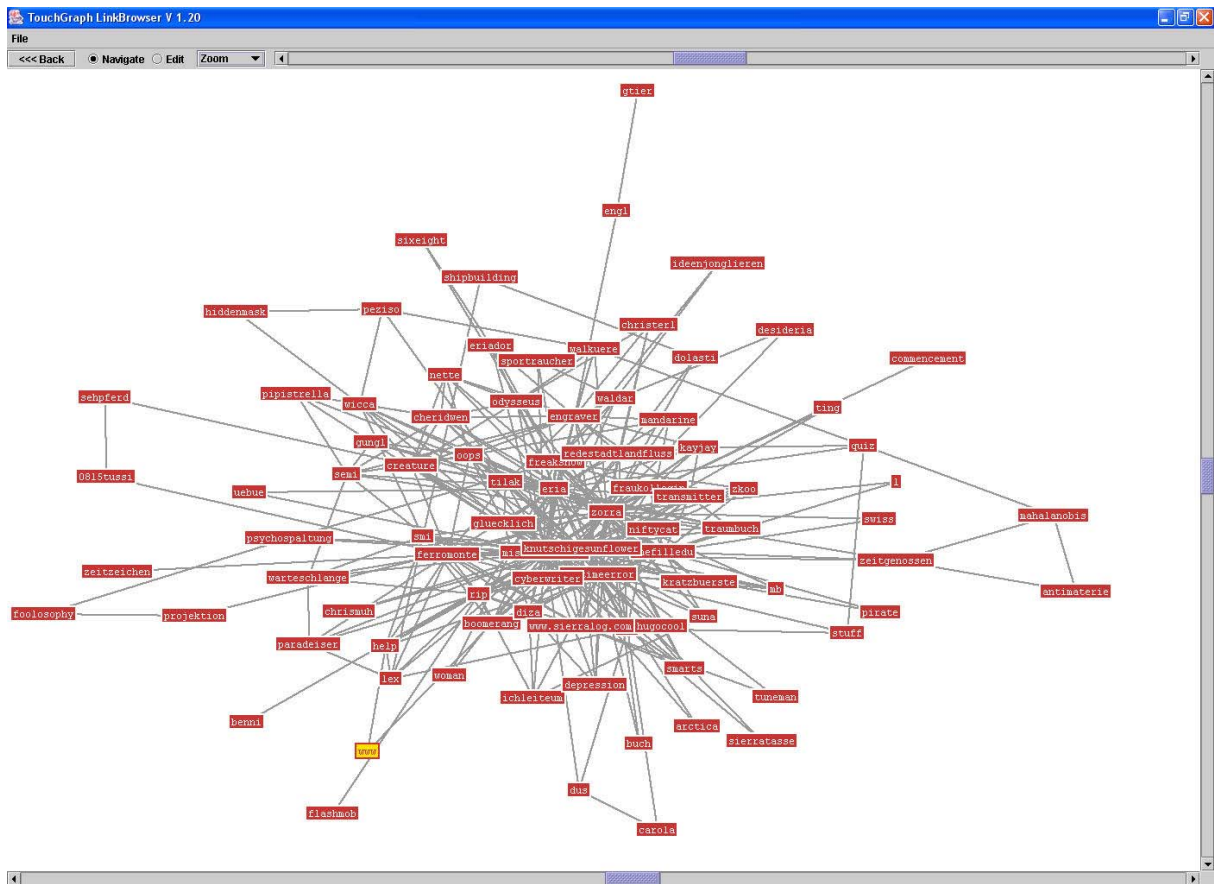


Figure 5: Visualisation of Data Set 1

## 6. Conclusion

It is easily arguable that Weblog communities form a Scale-Free Network, due to the very nature of Weblogs. They serve as a virtual representation of our social self and replicate our communication patterns as well. Probably it is possible to study the Blogosphere concerning the propagation of memes, the emergence of such networks and other factors relevant to Networks.

## 7. References

- Barabasi A., Albert R. (1999)**, Emergence of scaling in random networks. *Science*, Volume 286, p. 509-512
- Barabasi A., Albert R., Jeong H. (2000)**, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A*, Volume 281, p. 69-77
- Barabasi A. et al. (2002)**, Evolution of the social network of scientific collaborations, *Physica A*, Volume 311, p. 590-614
- Barabasi A., Bonabeu E. (2003)**, Scale Free Networks, *Scientific American*, May 2003
- Blood, R. (2000)**, Weblogs: a history and perspective, September 2000,  
[http://www.rebeccablood.net/essays/Weblog\\_history.html](http://www.rebeccablood.net/essays/Weblog_history.html)
- Cohen R., Havlin S., ben-Avraham D. (2002)**, Structural Properties of Scale-Free Networks, *Handbook of graphs and networks*, Wiley
- McArthur R., Bruza P. (2003)**, Discovery of implicit and explicit connections between people using email utterance, *Proceedings of the Eighth European Conference of Computer-supported Cooperative Work*, Helsinki, p. 21-40
- Newman M., Watts D., Strogatz S. (2002)**, Random graph models of social networks, *Proceedings of the National Academy of Science*, Volume 99, p. 2566-2572
- Newman M. (2002)**, The structure and function of networks, *Computer Physics Communications*, Volume 147, p. 40-45
- Schneeberger A. et al. (2004)**, Scale-Free Networks and Sexually Transmitted Diseases, *Sexually Transmitted Diseases*, June 2004, Volume 31, Number 6, p. 380-387
- Steyvers M., Tenenbaum J. (2003)**, Small Worlds in Semantic Networks, Working Paper,  
<http://psiexp.ss.uci.edu/research/papers/small9formatted.pdf>
- Pastor-Satorras R., Vespignani A. (2001)**, Epidemic spreading in Scale-Free Networks, *Physical Review Letters*, Volume 86, Number 14, p. 3200-3203
- Watts D. (2003)**, *Six Degrees. The science of the connected Age*, Norton & Company, New York
- Weinberger D. (2003)**, *Small Pieces loosely joined*, Perseus Publishing, Cambridge