MSc Modern Applications of Mathematics

Growth of Value in Social Networks

Author:
Ruth Jenkins

Supervisors:
Professor Chris Budd
Shail Patel

September 2008

Declaration: I certify that all work contained within this document is my own, unless otherwise referenced.

Ruth Jenkins

Abstract

Metcalfe's Law states that the value of a communications network grows with the square of the size of the network. Metcalfe, however, could well have been misquoted. The validity, or not, of Metcalfe's Law is critical in a time when companies are valuing themselves, using this rough rule of thumb as a justification. Does Metcalfe's Law actually hold for online social networks?

This project seeks to review current social network models and look into methods of valuing them. We will explain the basis of Metcalfe's Law, as well as other theoretical value models and examine empirical data to test the validity of both the network models and the value metrics.

Finally we present a dynamic model of an online social network and measure its value growth. In fact, it turns out that Metcalfe is a long way off the mark.

Contents

Chapter 1

Introduction

The dot-com bubble of the late nineteen-nineties was due, in large part, to the market's belief in Metcalfe's Law. Metcalfe's Law states that the value of a communications network grows proportionally to the square of its size. The dot-com entrepreneurs chased market share, not profits, since it was believed that once the size of the business was big enough Metcalfe's Law would kick in and provide profits on a quadratic scale. Unfortunately for these small businesses they had misinterpreted Metcalfe - the bubble burst. Between March 2002 and October 2002 five trillion dollars in market value was lost from technology companies, triggering a mini recession in the Western economies.

We are now entering a new era of online provision. The web is moving away from information dissemination and toward networking; with interactions between groups. These social networking applications, for example Facebook, Secondlife, Myspace and Flickr, are currently seeing the same unrealistically high valuations being placed on them as was placed upon their nineteen-nineties counterparts [19]. What is the relationship between growth and value? Does Metcalfe's Law now apply?

This report aims to investigate various existing models of social networks and explore several value forming processes. This report will then look at empirical data to determine which models best reflect reality. Finally we will develop a model of online social network value growth.

Chapter 2

Social Network Structure

Mathematicians seriously began to involve themselves in social network analysis about half a century ago. Network analysis had been around as a branch of pure mathematics for some time before this, but Pool and Kochen [31] were the first to apply a mathematical model to the social sciences. Pool and Kochen investigated chain lengths connecting people and the probability of two people knowing each other. This early work used the random network as a model. We will see below that a random network models a few, but not many, of the features of a social network. More accurate and refined models have now been produced.

The first question is naturally then, how do we know what an accurate model looks like? This chapter begins with a section addressing that question. We will look at empirical data from real world networks and define measures that will be important to retain in the models.

This chapter will then examine several different models and consider their strengths and weaknesses as candidates for a social network. The models to be reviewed are:

- The random network
- The Watts-Strogatz 'small world' network
- The Barábasi-Albert model of preferential attachment
- The Davidsen model of introduction

2.1 Observed Characteristics of a Social Network

Real networks are complicated objects. We need to extract their key features, whilst retaining a manageable model. We consider the three features considered most important [26]:

- average path length, *how many intermediary acquaintances exist between two randomly chosen people?*
- clustering coefficient, *how much more likely are you to know someone if you have a mutual friend?*
- degree distribution, *are we all equally popular or do some people have more friends than others?*

Some real world analysis [23, 26, 28] informs us that social networks are characterised by:

- high clustering, *people tend to form friendship groups*
- short average path length, *the famous, and perhaps surprising fact, realised in Milgram's 'six degrees of separation' experiment [23]*
- a power law degree distribution, *we like to make friends with popular people*

We will firstly define these measures more precisely and then go on to look at some actual empirical data.

2.1.1 Definitions

Average Path Length

The path length, $l(i, j)$, is the fewest number of edges between nodes i and j, with the average path length, L, being the average taken over all pairs of nodes with finite path length.

Clustering

Clustering feels intuitively straightforward; groups of friends will form within a larger network. It is however, fairly difficult to measure and has been defined in different ways by different people. Watts and Strogatz [35] define it as follows:

If node i has k_i neighbours, then the maximum number of connections between these neighbours is $\frac{1}{2}k_i(k_i - 1)$. Let C_i denote the fraction of these connections that actually exist. Then the clustering coefficient across the network is the sum of these, i.e.

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i$$

The problem with this definition is that a person with very few friends is given the same weighting as a person with many friends (and hence a few small, tightly knit cliques will outweigh larger, more loosely connected sub-networks). This can be seen to erode the value of C for networks with a large variance across the degree distribution [28]. Newman *et al* [27] redefined the clustering coefficient to take account of this. They proposed the following:

$$C_{new} = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}$$

where a connected triple is node connected to two other nodes, and a triangle demands that these two other nodes are also connected to each other.

The intuition of this definition is similar, what percentage of a person's friends are themselves friends? Whilst it is better suited to networks with large variance at degree, it is more difficult to calculate. We will use the Watts-Strogatz definition unless otherwise stated as it is the more commonly understood definition in literature [8].

Degree Distribution

Thinking qualitatively about social networks is does not seem reasonable that all nodes should have equal degree, it is just a fact of life that some people are more popular than others. The degree distribution tells us how many nodes have a given degree.

Analysis of real networks tells us that popular people haven't just got a lot of friends, they are also good at making new friends, whilst those with fewer friends are not: the situation becomes self-perpetuating. This phenomena is refered to as 'preferential attachment' by Barábasi and Albert [5], and more casually as 'the rich get richer'. Since we are considering online social networks, this effect will also be impacted upon by the type of user. A heavy user of an social networking site is likely to have more online friendships, and since they have more connections to maintain they will spend more time online, this is again, self-perpetuating. It tends to be the case that early adopters of new technology are the heavy users, and there is a large difference between the amount of time spent online between the heavy and light users [14].

Degree distribution is a complicated measure, real networks actually tend to show different distributions at different points [8]. Consequently there is no 'correct' distribution for us to use, but a power law is the best estimation available to us [1, 5, 10].

Power law degree distributions have been well studied, see for example [1, 5, 27]. This type of network is such that

$$P(k) = k^{-\gamma}$$

where $P(k)$ is the probability that a node has degree k. The other parameter, γ depends on the network, and we will need to find the correct value for online social networks.

Networks that follow the power law model tend to be 'scale free' and many use the terms interchangeably [1].

2.1.2 Characteristic Values

There has been a fair bit of analysis of systems that can, loosely, be described as social networks. Data on friendship connections is difficult to come by, therefore the network of film actors (who has acted with whom) and collaboration networks (which academics have written a paper with which others) are used as pseudo social networks in most research. A summary of the findings from these is found in table 2.1.2.

[1] It is the degree distribution that is scale free, since $k^{-a\gamma} = bk^{-\gamma}$

Table 2.1: Characteristic values from literature

Network type	Nodes	Average path length (L)	Clustering coefficient (C)	Average Degree	γ	Citation
Film actors	212 250	3.65	0.79	28.78	2.3	[5, 35]
Physicists	55 627	4	0.726	173	1.2	[27]

Table 2.2: Characteristics values found for online social networks

Network type	Nodes	Average path length (L)	Clustering coefficient (C)	γ
Flickr	1 846 198	5.67	0.313	1.74
LiveJournal	5 284 457	5.88	0.330	1.59
Orkut	3 072 441	4.25	0.171	1.50
Youtube	1 157 827	5.10	0.136	1.63

We are, however, specifically interested in online social networks, which may well be different from their offline counterparts. There has been a recent surge of research in this area, since it is both of contemporary interest and data is more readily available.

Mislove *et al* have very recently performed a large scale measurement study of several online networks [24]. They collected data from over 11.3 million users across four well known websites (Flickr, LiveJournal, Orkut and YouTube). The results confirm that these networks display the power-law and small-world properties we might expect, along with high clustering. Actual figures from the paper are given in Table 2.1.2.

The main problem with Mislove's data is that only Orkut is a purely social networking site. Flickr's main function is an online photo album and LiveJournal is primarily a blog site (although both of these offer social networking services). Youtube is a video sharing website, and with its many unregistered users and little social networking, serves as a poor proxy. Unfortunately, the Orkut data is the least complete due to the large amount of bandwidth required for this search. Three million users were surveyed, which is around 11% of the total number of registered users.

If we concentrate on the Orkut data we see an average friends per user of 106.1 (within Dunbar's bounds), an average path length of 4.25 (of the same order as Milgram [23]) and clustering that, if not high, is way above what one would expect for a random graph. It also has a power law degree distribution, with $\gamma \approx 1.5$, possibly indicating an underlying system of preferential attachment (see section 2.2.3).

2.2 Historical and Current Social Network Models

2.2.1 The Erdős and Rényi Random Network

Random graphs were initially developed within the bounds of pure mathematics, with seminal work in the field coming from Erdős and Rényi [25]. Despite this, it was the random network that first acted as a model of a social network for Pool and Kochen. We do not expect the random network to be a very realistic model, but it is a starting point.

A random network has two parameters: number of nodes, n, and probability of two nodes being connected, r. The type of social networks we are considering will have large n and small r. If we know n and our required average degree, k, then r is simply

$$r = \frac{k}{n}$$

The average path length in a random network is small. It is by introducing some randomness into a regular network that Watts and Strogatz were able to induce the small world effect. The average path length in a random network can be calculated using

$$L = \frac{\log n}{\log nr}$$

[8]. Using characteristic values of $n = 1000000$, and $r = 0.0001$, we calculate $L = 4$, which is very low, and a bit lower than the eponymous six degrees of separation.

There is no clustering as such, since the probability that B and C are friends with each other is, by definition of a random graph, independent from the fact that they both know A. In fact if we use the clustering coefficient as defined in the previous section, it is simply equal to r for a truly random network. We are working with $n = O(1000000)$ and $k = O(100)$, which leads to $C = r \approx 0.0001$, this is of orders of magnitude less than the values in tables 2.1.2 and 2.1.2.

All nodes in a random network have the same expected degree, therefore the degree distribution certainly does not follow a power law. In fact

$$P(k) = \binom{n-1}{k} r^k (1-r)^{n-1-k}$$

which can be derived by considering the number of other nodes that are,(k), and are not, ($n-1-k$), connected to our initial node. This is the binomial distribution.

2.2.2 The Watts and Strogatz Small World Model

Watts and Strogatz set up their model by starting with a regular 1D ring network with a given number of nodes, n, and degree of each node, k. They then introduced a third parameter, $0 \leq q \leq 1$, where q is the probability that an edge is 'rewired' from its neighbour to another random node. Therefore $q = 0$ is the original, regular network and $q = 1$ gives the fully rewired network [2].

Watts and Strogatz observed that since small q affects only a very low percentage of the nodes, the clustering coefficient remains largely unaffected. In contrast, average path length is significantly affected by the introduction of just a few 'shortcuts', since these shortcuts have an impact, not only on the path length between the newly connected nodes, but also between their neighbours, and their neighours' neighbours. It seems then, that a small amount of rewiring has created the network we were looking for: low average path length and high clustering, in fact as $n \to \infty$, then infinitesimally small q is sufficient to induce small world average path lengths [6].

Random rewiring does nothing to affect degree distribution, and consequently the Watts-Strogatz model still displays a binomial degree distribution.

[2] It should be noted that, because only one end of an edge is rewired with probability q, even $q = 1$ does not give a fully random network [6]. This has made analysis of the system difficult, and is one of the reasons that others have suggested modifications.

2.2.3 The Barábasi-Albert Model of Preferential Attachment

In [5] Albert and Barábasi set out to examine the degree distribution of a network modelled on preferential attachment. There network evolved as follows.

The model takes n_0 initial nodes and at each time step an action is performed:

- *With probability* $0 \leq p_{add} < 1$, $m \leq m_0$ new edges are added. For each new link one node is chosen at random and connected to a second node selected with a probability

$$P(k_i) = \frac{k_i + 1}{\sum_j k_j + 1}$$

- *With probability* $0 \leq p_{rewire} < 1 - p_{add}$, m edges are 'rewired'. An edge is selected at random and one of its ends moved to a new node, selected preferentially with probability $P(k_i)$ as above. This is done for each of the m edges.

- *With probability* $1 - p_{add} - p_{rewire}$ a new node is added, this node is given m edges connected to existing nodes with the above defined probability, $P(k_i)$.

Since the aim of the model was to produce a power law degree distribution we check this has been achieved.

Let k_i be the degree of node i, then the probability that a new node will connect to k_i is

$$P(k_i) = \frac{k_i}{\sum_j k_j}$$

and therefore k_i increases at a rate proportional to k_i. In fact the rate of growth is given by

$$\frac{dk_i}{dt} = mP(k_i) = \frac{mk_i}{\sum_j k_j} = \frac{k_i}{2t} \tag{2.1}$$

where m is the number of edges of the newly added node, and t is the time step. If a node is added at each time step, the the total number of edges is mt, and hence

$$\sum_j k_j = 2mt$$

since summing over all nodes counts each end of an edge. Therefore (2.1) becomes

$$\frac{dk_i}{dt} = \frac{k_i}{2t}$$

which can be solved to give

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} \tag{2.2}$$

where t_i is the time at which node i was added. Therefore, if we wish to calculate the probability that a node i has fewer connections that a set value k, we may use (2.2), and rearrange

$$P(k_i < k) = P\left(m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} < k\right) = P\left(t_i > \frac{m^2 t}{k^2}\right)$$

which for equal time steps becomes

$$P\left(t_i > \frac{m^2 t}{k^2}\right) = 1 - P\left(t_i \leq \frac{m^2 t}{k^2}\right) = 1 - \frac{m^2 t}{k^2(t + m_0)}$$

The probability density function can be found via

$$p_k = \frac{d(P(k_i < k)}{dk}$$

which via (2.2.3) is

$$p_k = \frac{2m^2 t}{k^3(t + m_0)} \rightarrow \frac{2m^2}{k^3} \text{for large t}$$

We therefore see that modelling a network based on preferential attachment produces a power law degree distribution. This reflects real life as seen in tables 2.1.2 and 2.1.2.

This model was setup with the intention of modelling the degree distribution and did not have the aim of modelling high clustering or short path-length. The authors do not mention these characteristics within their paper. During our matlab simulations the clustering coefficient can be calculated and we can see how well this model represents this phenomenon.

2.2.4 The Davidsen Model of Introduction

This model is based around a simple premise: we often meet new friends through our old friends. This basic assumption, along with the fact that people join and leave a network, is enough to formulate the model.

All of the other models reviewed here are static models with fixed nodes and edges, whilst this makes them easier to study it is clearly an oversimplification. Dynamical models are at the forefront of current network research [28] but this model will be the only one consider within this report. In the next section we will look at value within a social network. It seems likely that the network might grow in such a way as to optimise it's value and therefore a network that grows organically as its value grows, seems to provide a better basis.

Taking a fixed number of nodes, n, with an initial number of edges between them, we impose the following dynamics onto the system:

- A person is selected at random and two of their randomly selected friends are introduced, creating a new edge in the network if there was not already one in place. If the original person does not have enough friends to create an introduction, then they introduce themselves to a stranger chosen at random.
- A person is chosen at random and, with probability q, they are removed from the network and all their edges are removed. A new person is introduced to the network and given one, random, acquaintance.

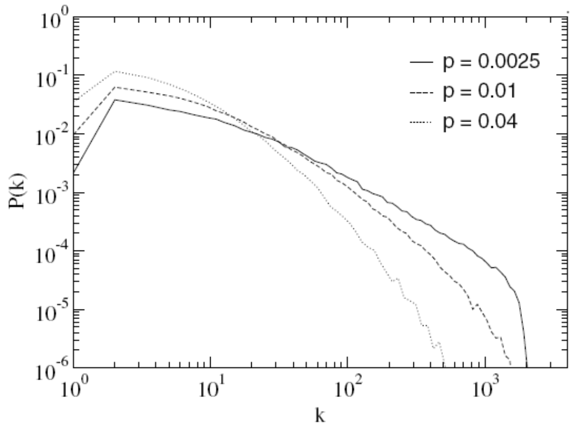
The probability, q, gives us a time scale. In general, for a real social network, people are far less likely to die, or move away, than they are to meet new people, therefore a value of $q \ll 1$ will be considered. As this is our main parameter (initial number of nodes and edges playing a less significant role in a dynamic network that has reached steady state) it has an impact on the degree distribution.

The preferential attachment built into this system comes from the fact that a person with many acquaintances has a increased probability of being selected to be introduced to a new person. The shape of the degree distribution changes as we vary q, with a power law distribution for small p; the exponent being inversely related to q, (where the preferential linking dominates) through to an exponential degree distribution for large p. We would like a model with a degree distribution that follows a power law, and we have discussed above that small p is the more intuitive for social networks. For $q = 0.0025$ we have

$$p_k = k^{-1.35}$$

which fits in well with the values we see in table 2.1.2. There is a cutoff for large k due to the fact that the nodes have a limited lifetime and consequently their number of acquaintances cannot grow without bound. This means our distribution is only a power-law up to a point; it does not have the long-tail that we would otherwise expect. Figure 2.1 is the original graph from [9] showing both the power law regime and cutoff for large k.

Figure 2.1: The degree distribution p_k of the model in statistically steady state. For this graph $n = 7000$, although this does not have much affect. Original graph from [9].



When we consider the clustering coefficient of this model there are no rigorous results, but very strong numerical evidence to suggest that the network demonstrates realistic values. Table 2.3 shows the clustering coefficients of this model, alongside that for a comparable random model. These values can also be compared to tables 2.1.2 and 2.1.2.

Table 2.3: Clustering coefficient for different values of q, with $n = 7000$. C_{rand} is the coefficient of a random network with the same n and k.

p	k	C	C_{rand}
0.04	12.9	0.45	0.0021
0.01	49.1	0.52	0.0070
0.13	149.2	0.63	0.021

Average path length, L, can also be calculated for the network and is found to be consistent with logarithmic behaviour, as required. For $q = 0.0025$ and $n = 7000$ we find that $L = 2.38$. This is higher than we would expect for an entirely random graph where $L_{rand} = 1.77$, and in fact a little lower than would be ideal.

Chapter 3

Calculating Value Growth

3.1 Theoretical Approaches to Value Growth

Not many people have looked at network value from an empirical research perspective, but there are plenty of theoretical models in current circulation. Here we will review them.

3.1.1 Linear growth

Sarnoff's Law

The value of a broadcast network grows linearly with the number of users. [32, 29]. This is a taken as fact for pure broadcast networks (e.g. television and radio) and we assume it is also true for broadcast websites (for example news based).

Linear growth in a communication network

There is some empirical evidence that linear growth might hold for communication networks [30], despite the fact that it is counter-intuitive.

The two main theoretical arguments in support of this are based on convergent value distributions and consumption limits [37]. We have so far assumed that each node derives equal value from each other node (many do [32, 33]), but this, of course, could be far from true. If the distribution value were to be convergent, for example, if you valued connection with each node half as much as you did the last node, then the total value to any node would be

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = \sum_{x=1}^{\infty} \frac{1}{2^n} \rightarrow 2$$

Therefore when we sum over all nodes we see a total value of $2n$, i.e. $O(n)$. We may now accept this or disregard it, but let us now cap the number of connections that are possible for each user (there are only so many hours in the day), lets call this number C. Whilst the size of the network is still growing below C then we will see quadratic growth (or some other type dependent on our value distribution), but once the cap is reached then each new user will instantly connect to the maximum number of people they wish to and network value will be, at most, nC, which is again $O(n)$ since C is fixed.

3.1.2 Metcalfe's Law

As mentioned briefly in the introduction, Metcalfe's Law states that the value of a communication network grows in proportion to the square of the number of users [15]. The theory behind this is sound; the value of a communication network depends on communications, that is, interactions between users. The number of possible pairwise interactions in a network of n users is the n^{th} triangle number, i.e. $\frac{n(n-1)}{2}$, which is clearly quadratic.

The devil is in the detail. Metcalfe's original presentation was to try and encourage companies to install LANs with at least three machines [29, 33], he did not say 'users' he said 'compatibly communicating devices'. The problem came when, during the dot-com boom, companies where keen to value their networking businesses as highly as possible. They applied Metcalfe's Law where it did not belong. In a large network, not all the users are going to communicate with all the others - it is simply impossible in a network with millions of users (which is the order we are talking about with online social networks).

Another arguement against Metcalfe's Law is put forward by Briscoe, Odlyzko and Tilly [7, 29]. Their case is a logical one. If there is exist two networks, the first of size m and the second, n, then the values will be m^2 and n^2 respectively. If the two networks were to combine then the total value would now be $m^2 + n^2 + 2mn$, and value has been gained 'for free': it would be illogical for them not to merge. An important factor in this arguement is the value gained by each network is equal (mn); if one network were to profit more than other from the deal, then this may cause a stale-mate, since this is not the case then a merger must surely occur.

3.1.3 Reed's Law

Any user of a social networking site will realise that they are not just about communicating with other people, but about communicating with groups of other people. Despite Metcalfe's Law seeming a bit ambitious, Reed argues than growth in some networks is even faster, in fact, it is exponential [32].

If a network supports the forming of groups then we consider how many groups exist. The number of non-trivial subsets that can be formed in a network of size n is $2^n - n - 1$, which is clearly $O(2^n)$.

A key argument in Reed's paper is that value can be derived from potential connectivity, rather than actual connectivity. This is a critical assumption that validates not only his findings, but Metcalfe's Law too. Reed assumes that each potential connection is worth as much as any other, which is simply not borne out in real life since we have not seen exponential value growth in any network.

Reed's Law is intuitively crazy. It does not seem at all feasible that a network with 101 people is worth double a network of 100 people, or the even more extreme case that the birth of one child doubles the value of the world. We use the merger arguement again here [7]; if network value really did grow exponentially then two separate networks would almost be forced to join up.

3.1.4 Zipf functions and Value as a Harmonic Series

A key assumption for Sarnoff, Metcalfe and Reed is that all connections are equally valuable. This simply isn't true. In a large network most potential connections are never utilised [7], and this becomes obvious when one considers the network of the World Wide Web, where not everyone speaks the same language.

Studies have shown that the number of interactions between two populations is inversely proportional to the distance [29]. This is often called a 'gravity law' since it is the same pattern as that of the gravitational attraction between two masses. Specifically, if there are two populations, A and B at a distance d apart, then traffic is proportional to $\frac{AB}{d^\alpha}$, where α is a situation dependent constant, usually between 1 and 2. If $\alpha = 2$ (and that is a common value for it to take in many real-life situations [37]) then the total value is proportional to $n \log(n)$, for the details of this derivation please see Appendix A.

Since we do not know what the distances are in an online network, indeed 'distance' may be defined in many ways, we try to derive the $n \log n$ value via alternative means.

We argued above that if the value distribution were convergent then overall growth would be linear. If the value is constant across all nodes we see Metcalfe's quadratic growth. We require something in between. Zipf's Law says that if we order a large collection by popularity, the second on the list will be worth about half of the first (in terms of whatever we are measuring) and the third on the list is worth one third the first, i.e. that the harmonic series is generated [37]. In this case we do not have a convergent value distribution, rather a slowly diverging one,

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots = \sum_{x=1}^{\infty} \frac{1}{x}$$

It is clear that this series diverges when we consider the integral test for convergence. Since $f(x) = \frac{1}{x}$ is a non-negative monotone decreasing function then the sum $\sum_{x=1}^{\infty} \frac{1}{x}$ converges if and only if the integral, $\int_{x=1}^{\infty} \frac{1}{x}\, dx$ is finite. Since

$$\int_{x=1}^{n} \frac{1}{x}\, dx = \ln x|_1^n = \ln n - \ln 1 \to \infty \text{as n} \to \infty$$

then the integral is not finite and hence the sum diverges. It is a standard result that the sum is bounded:

$$\sum_{1}^{n} \frac{1}{k} \leq \ln n + 1$$

please see Appendix B for a proof of this. This is the value for each of the n nodes, and hence the total value will be of order $n \log(n)$.

3.2 Empirical Research of Value Growth in a Social Network

3.2.1 Dunbar's Chistmas Card Questionnaire

Data for actual social networks (as opposed to the proxies of citation and actor networks) is difficult to come by, especially when we become interested in valuing these networks. The main reason being that it is difficult to collect the data. Even if people are willing to sit down and tell you who all of their friends are, it is likely that they will forget some. It is even more likely that not all of those friends will be prepared to participate and hence the network will be largely incomplete. Recent approaches for estimating these network sizes have used smaller sub-populations to generate reliable figures, but are unable to provide information on value [22]. Psychologists Hill and Dunbar conducted a recent experiment to investigate this issue [11]. This paper uses the exchange of Christmas cards and subjective questionnaire responses to value relationships. A graph from this paper, showing the distribution of network size across respondents, is reproduced in Figure 3.1. The mean network size is 153.5 ($\pm$84.5). This resonates well with previous work by Dunbar [12], which uses human neocortex size to estimate a social group size of 150.

This paper also found further evidence to back Dunbar's previous observations that social networks are hierarchically different [11]. That is that one has a core of very close friends and larger numbers of increasingly less intense relationships. The previously published figures show clusterings of five (support cliques), 12-15 (sympathy groups), 35 (bands) and 500 (mega bands). These figures have been verified by many others (for a list of references see the bibliography for [11]).

3.2.2 Value Growth of an Online Social Network

The impetus for this project was the research paper by Shail Patel of Unilever [30]. Patel was interested to discover whether Metcalfe's Law did indeed hold for a social network, and if it did not then what did hold. In his paper his summarises the theoretical arguements and looks to empirical data for support. The use of empirical data to aid the understanding of network value growth does not appear to have been tackled by others as of yet.

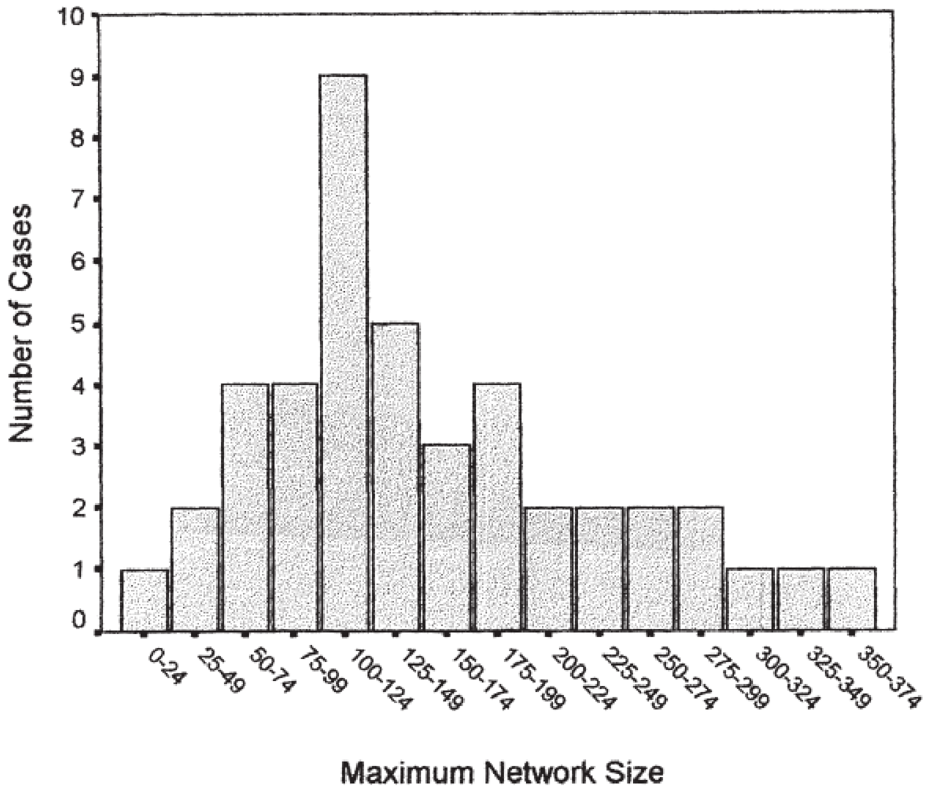


Figure 3.1: Network sizes as found by Dunbar and Hill, [11]

One problem is defining what we mean by value; value to whom? Value itself is theoretical concept and therefore we are going to require a proxy to measure it. Patel takes data from Ebay and Secondlife, and uses total global transaction (in US dollars) as a proxy for business value in each case.

The benefits of using Ebay and Secondlife is that this data is readily available and accurate. The restriction of using Ebay is that it is not a social networking site, as it is the union of the (non-disjoint) sets of sellers and buyers, with communication happening between each pair perhaps only once. Secondlife is a social networking site, but since it is a virtual world, we cannot be sure it will act in the same way as a real world network.

Patel expects that Ebay should show quadratic growth, whereas Secondlife can also benefit, not just from pairwise interactions, but also the group interactions suggested by Reed [32], and therefore could have greater than $O(n^2)$ growth. The numerical and graphical analysis of the Ebay results is inconclusive. There is similarity across all the possible models (linear, quadratic, $n\log(n)$ and n^2) in terms of accuracy of fit, although the linear and quadratic models far slightly better. Patel concludes that, by Occam's Razor, the value growth is probably linear. The story is similar with Secondlife; no conclusive results but a more likely fit with the linear and $n\log(n)$ models than the other two.

The most interesting aspect of Patel's report is that he distinguishes between active and registered users (for Ebay) and standard and premium residents (in Secondlife). The phenomenum of early adopters/heavy users having a significantly different profile to the masses is a well known one [38]. In Secondlife particularly, there is an indication that for premium residents the value growth may be better than linear, but that the slow down to linear growth comes from the large bulk of more casual users.

3.2.3 Primary Data From 'Facebook': Investigating Value per User

Despite an extensive search no papers could be found that looked at value within an online network, at a user level. Patel's paper looks at the global perspective, and whilst it hints that different users may have different value profiles, this needs further investigation.

We chose to perform a small scale questionnaire of Facebook users to see if we could learn anything from data on a user level.

This time the proxy for value to a user is the amount of time they spend online. It is value to a user in the sense that that is the value they place on the network. It can be translated into value to the network (or at least value to the business) as a user with more hours online will provide more eyeballs in terms of advertising - there are standard techniques used in market research to convert social value into a financial amount [37].

The main critic

There are clearly huge drawbacks with the data. The three main factors being sample size, sample selection and self-reporting errors. The sample size is very, very small, only 45 data points are used, whereas Facebook has a network size of over 90 million active users. Whilst many would argue that this is a pointlessly small sample, these are only initial findings and we would encourage others to follow this up. The sample is not random, in that all the invited people are this author's Facebook friends [1]. This clearly narrows the field to a small demographic, although in terms of age and sex the sample is broadly representative of Facebook as a whole [13]. Since the questionnaire is a voluntary one, perhaps the sample is biased because of the people who chose to partake. Finally we have the problem of self-reporting. The number of friends a user has is a factual statistic, whereas number of hours spent online is volunteered information. We expected under-reporting of this value, as people tend to be embarrassed at high usage levels. We tried to minimize this by making the questionnaire as anonymous as possible. In the end most people chose to share their data publically and under-reporting does not appear to be a problem as the figures are, if anything, higher than expected.

The initial graph is shown in Figure 3.2. There is clearly a positive correlation between number of friends and time spent online. We tried linear, quadratic and exponential regression and all had similar R^2 values, although visually the quadratic model fits best.

It is clear, that there are two components to the graph, one looking quadratic and one linear. Our initial thought was that this third variable may be either sex, or age, of respondent. Dunbar discovered that there is a near quadratic relationship between age and number of friends [11], whilst anecdotal evidence would suggest that sex is a factor in a communication network. A

[1] although friends of these friends began to join as well: this is the power of social networking!

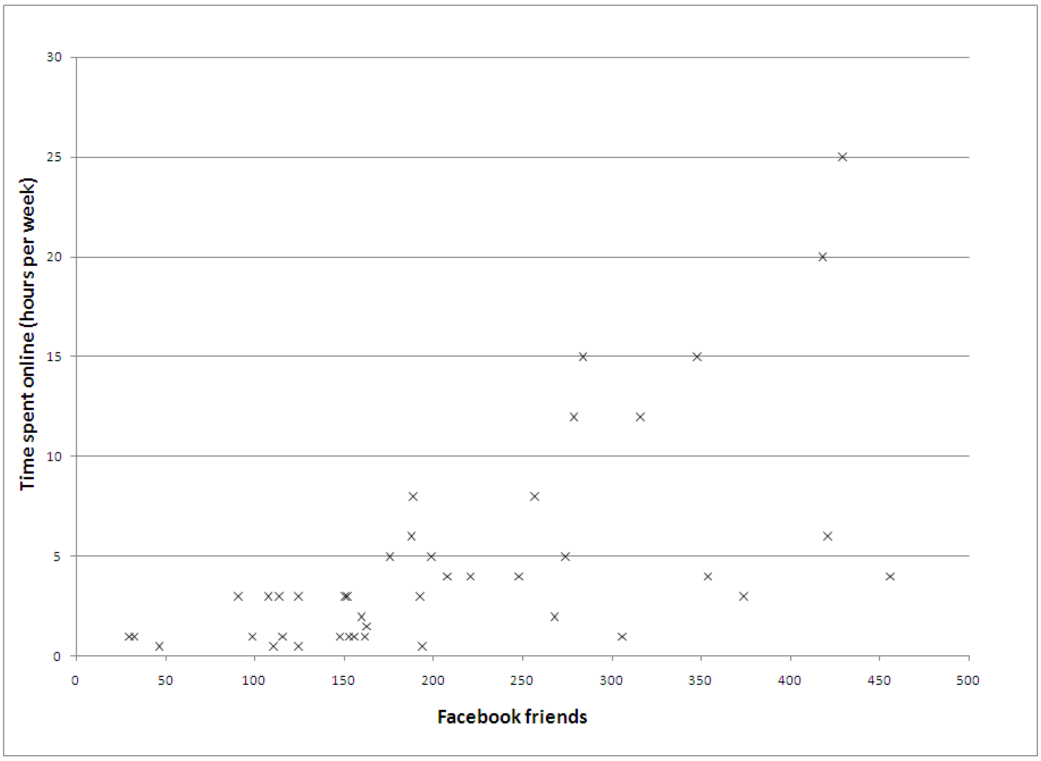


Figure 3.2: Initial graph of data collected from Facebook.

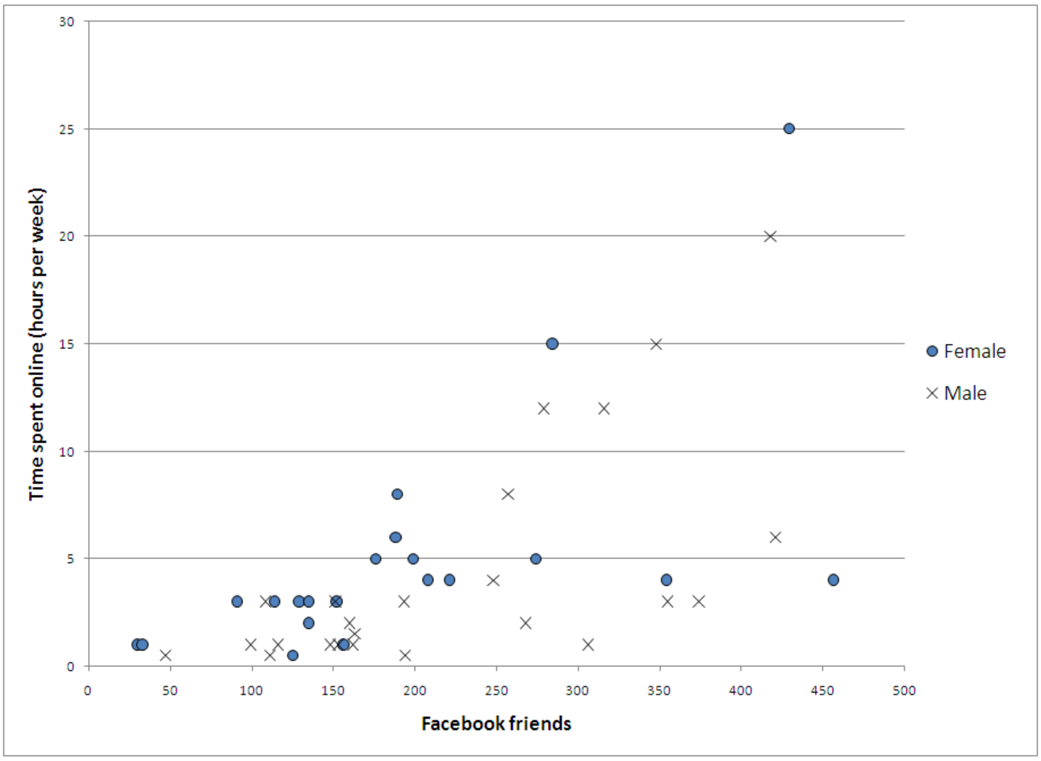


Figure 3.3: Facebook data with sex overlaid.

second graph was plotted to show sex of the respondent, this can be seen in Figure 3.3. This showed no difference between the sexes, and we conclude that sex may not be a factor. More data is required here to verify results.

Comparing number and friends and hours online with age shows similarly uninteresting results. See Figures 3.4 and 3.5. Dunbar's results tell us to expect more friends with age, but since Facebook has mainly young users this may be the overriding factor.

Having discounted the obvious factors of age and sex, we consider Patel's distinction between premium and regular users of Secondlife. Facebook has three components. Firstly it allows you to send public or private messages to other users or to 'poke' them or otherwise interact on a one-to-one basis. If this function is the main driver we should see a linear relationship between number of friends and value; each friend adds a fixed amount of value for the user. Secondly it allows you to view the public messages between your friends, view photos of friends taken by other friends etc. This function should will show quadratic value growth, since if a person has n friends, then the number

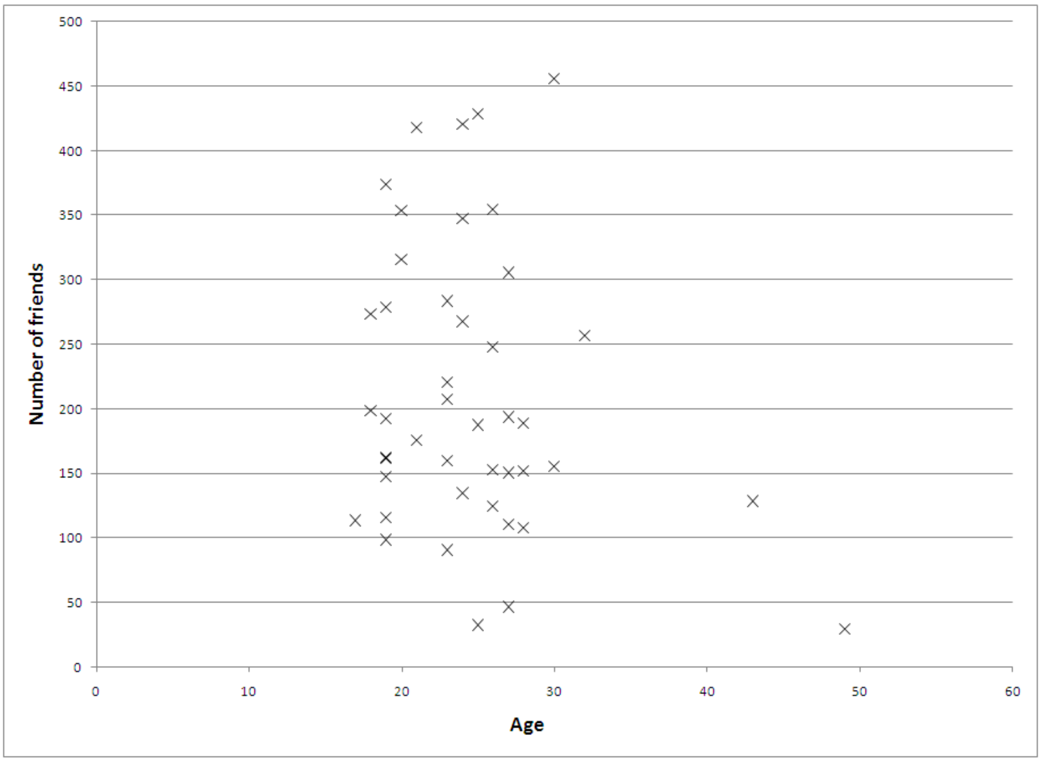


Figure 3.4: There is no link between age of user and number of Facebook friends.

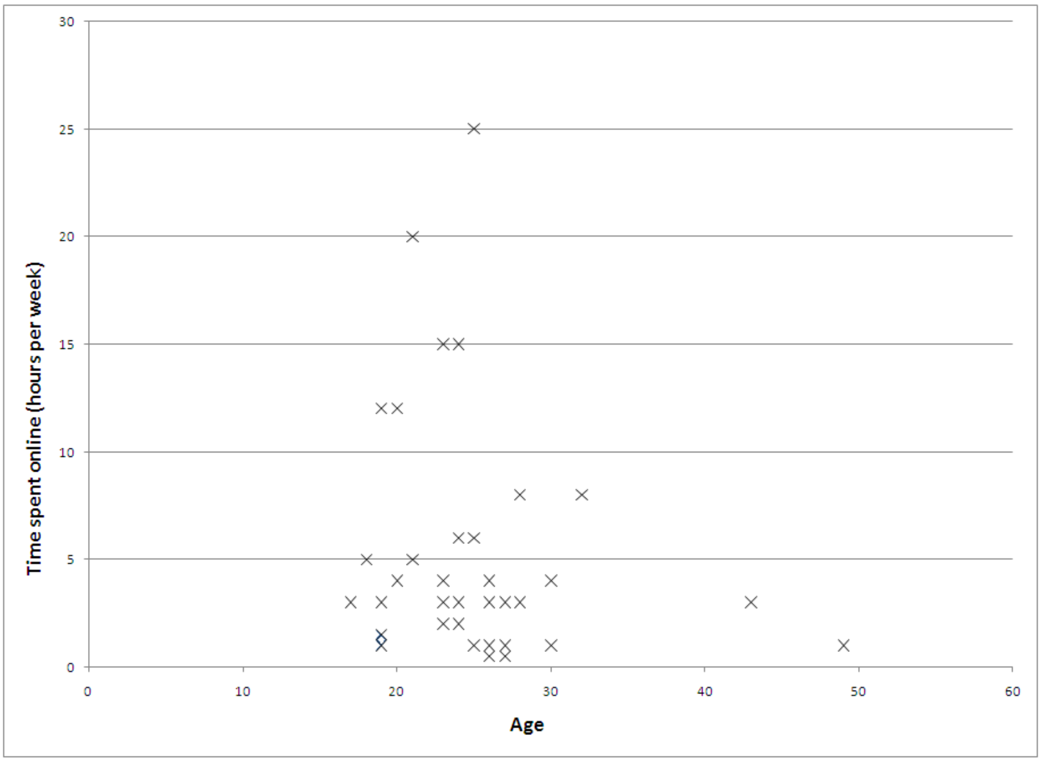


Figure 3.5: There is no link between age of user and amount of time spent online.

of 'interconnections' they may view will be $O(\frac{1}{2}n(n-1))$ and hence as they add friends to their list, value of the service will grow quadratically. Thirdly it allows you to create and join groups, and communicate with the members of that group, it is possible that where we will see exponential growth as suggested by Reed [32].

It is important to note here that this linear/quadratic/exponential growth is not in relation to the size of the entire network, only in relation to the size of an individual's personal network. The implications of this will be discussed further in Section 3.3.

The data was split into two groups, those who are considered 'linear' users (i.e. use Facebook primarily as a second form of email) and 'group' users (those who utilise more of Facebook's applications). These split was subjectively done, either by asking the user how they used the site, or from observing their interactions. The results here were quite astounding. The 'group' users appear to have clear quadratic value growth, the R^2 value quoted being 0.912, this is shown in Figure 3.7. We also performed linear and exponential regression for the 'group' users, shown in Figure 3.8 and found fairly high R^2 values, although not such a good visual fit. For the 'linear' users, the linear regression model shows the best fit, although it does show an unconvincing R^2 value, see Figure 3.6.

Mislove *et al* found that 13% of Orkut users used the group forming features of the site [24]. Our results have the figure at almost 50%. The fact that our data collection was via a group may, of course, have biased the data to this extent. Another arguement may be that Facebook is more keyed in to group formation than Orkut, and therefore may have greater value potential.

We conclude, taking into account the work of Patel, that low-level users add linear value to the network, whilst heavy users add better than linear, likely quadratic value.

It cannot be stressed enough that the sample size used here is way below anything required for significant results.

Finally we examine the degree distribution of our Facebook data. The mean average number of friends per person is 211 with a standard deviation of 110. This compares well with Dunbar's figures of 150, and 84.5, with our larger values possibly due to the ease of communicating with more people online, or due to a bias in our sample towards 'friendly' people. A frequency diagram showing how numbers of friends are distributed (Figure 3.9) also has

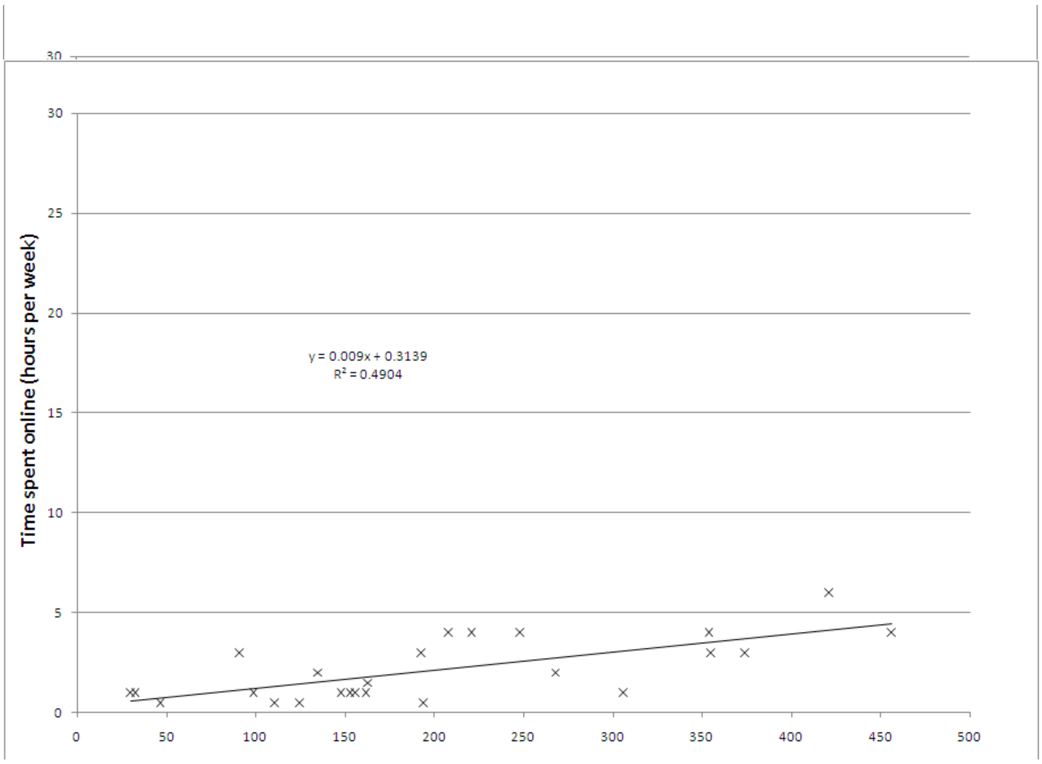


Figure 3.6: 'Linear users' show linear friends/value relationship

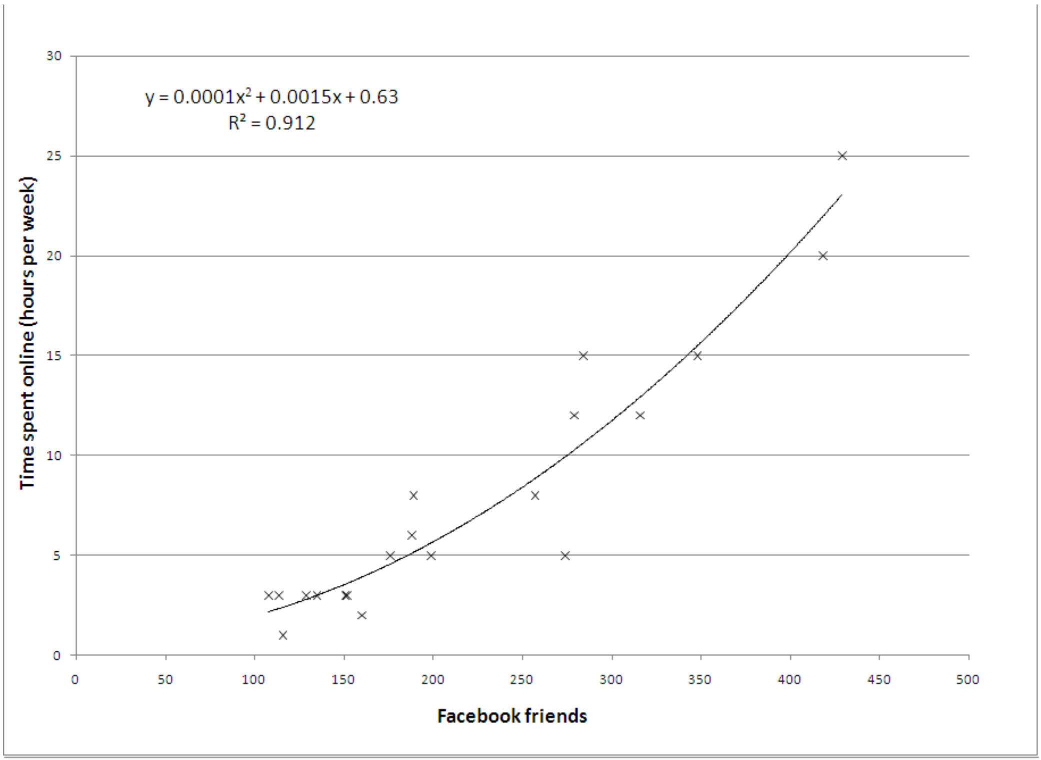


Figure 3.7: 'Group' users show a quadratic friends/value relationship

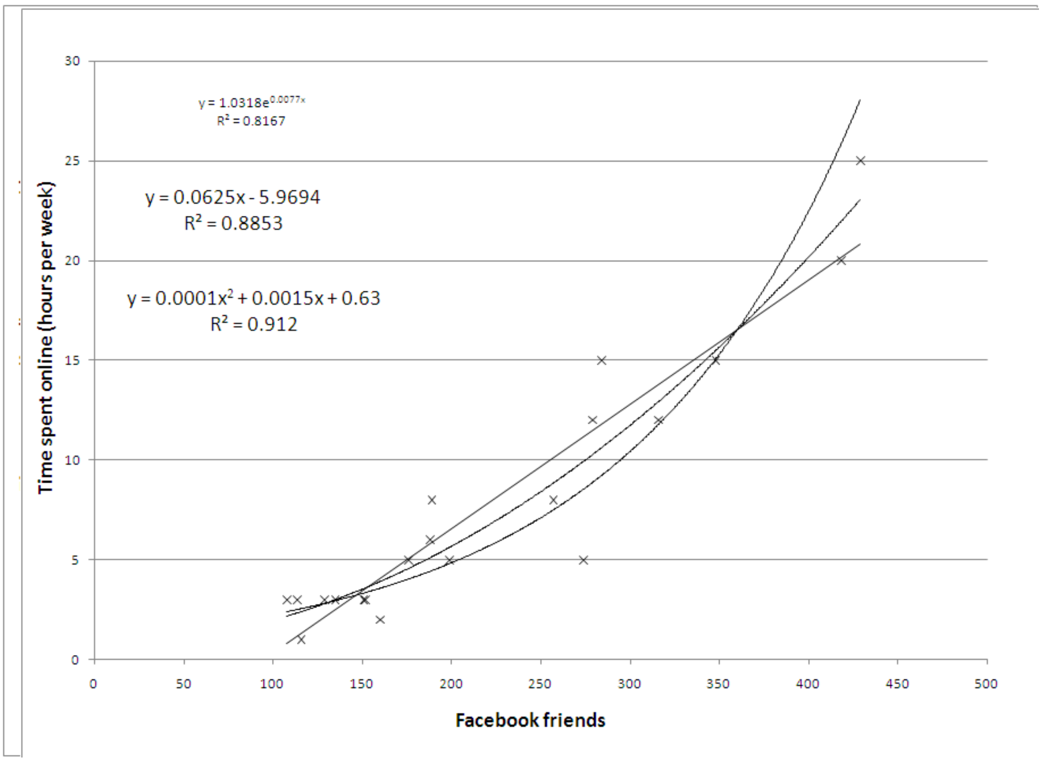


Figure 3.8: Evidence for linear, quadratic and exponential value growth for 'group' users

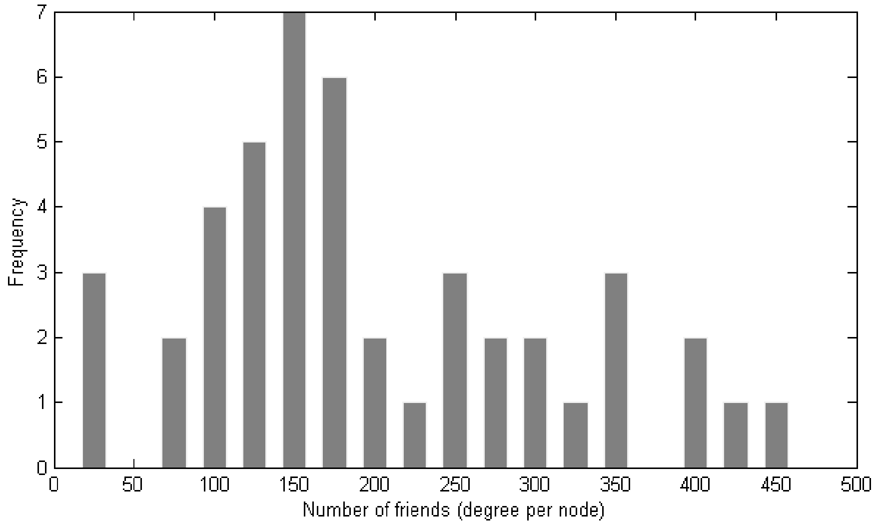


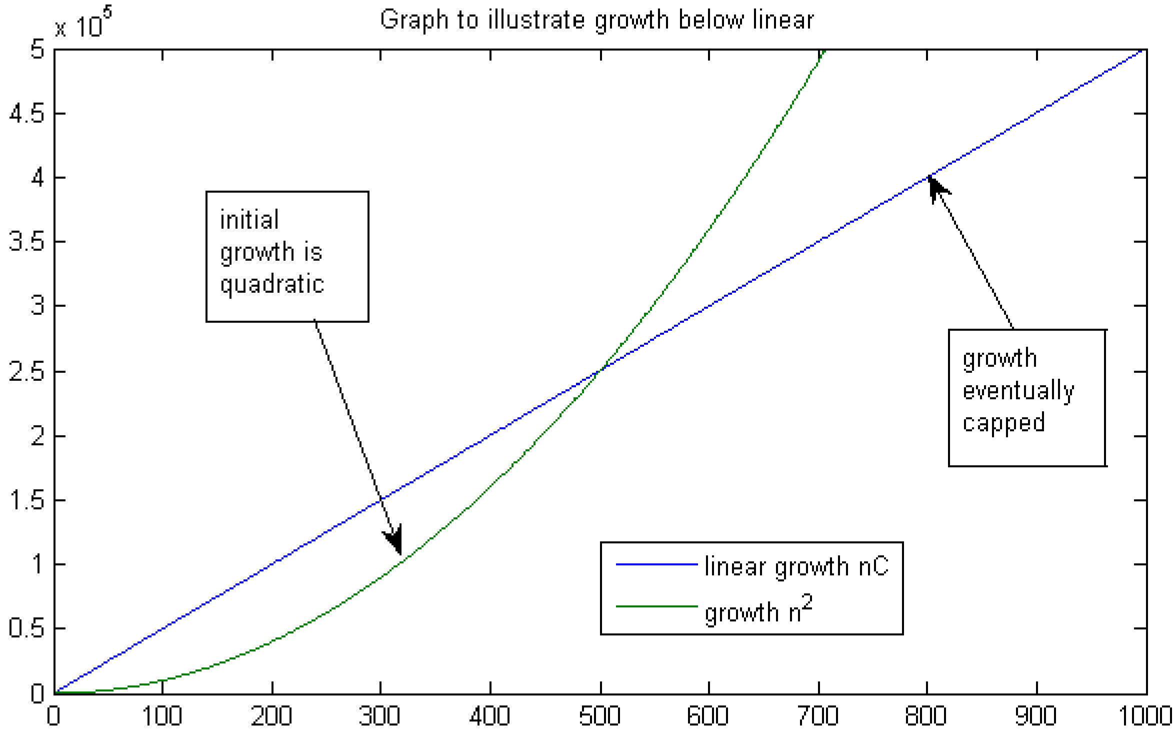
Figure 3.9: Degree distribution for nodes in our Facebook sample

a striking resemblance to Dunbar's graph (see Figure 3.1), and is, roughly, a power-law shape, as expected.

3.3 From Value to the Individual to Value of the Network

We briefly mentioned above that value to an individual and value to the network are clearly related, but not identical concepts. We have seen that, for our Facebook data, there are people for whom the value of *their* network grows quadratically with size, or linearly with size, but how do we scale this up to the entire network.

We must, surely, assume that there is an upper bound on the value a personal network can have. There are two arguements for this. Firstly we consider Dunbar's figures; people cannot (on average) maintain more than around 500 friendships, perhaps the internet makes this easier, and the number could rise, but it is surely not infinite. Secondly we use an exhaustive arguement. Each connection (if it is to have any value) must require a finite amount of time to maintain. There are only twenty-four hours in each day, so even if we elimate all other activities (eating, sleeping etc.) there is a finite limit on the number of connections that can be managed.



We must therefore conclude, that if C is the maximum amount of value any person can derive from the network, then nC is the upper bound for the value of the network. C is fixed (in our example, since we consider hours per week as a proxy for value, we take $C = 7 \times 24$), therefore value grows linearly with the size of the network (at best).

The final comment we need to make here is that the linear growth will be seen only when the network becomes so large that n dominates C. Remember that C is an upper limit, and most people are a long way off reaching it, therefore in the early stages of network growth, the network will be growing at a non-linear rate, but below the level of nC. See figure 3.3 for a depiction of this.

Chapter 4

Developing A Model

In order to measure value growth within a social network it was important to both build a model of the network and create a means of valuing. We began by developing Matlab models of each of the four models discussed in Section 2.2. The coding for each of these can be found in Appendix C.

Each model consists of A, an $n \times n$ matrix, with friendship between nodes i and j being created by a non-zero value in $A(i, j)$. The precise value of $A(i, j)$ is indicative of the value that i places upon j, and as such is not necessarily reciprocal (in fact it rarely is and in some cases $A(i, j) \neq 0, A(j, i) = 0$). The maximum value any node can place on any other is 1, with others being a fraction of this. To calculate the total network value we sum across all individual node values. The impact of this is that a model with equal value across all nodes will always have greater value than one with, for example, a Zipf value profile. Since we are only interested in the shape of the growth however, this is unimportant to us.

To value these networks we developed four value profiles which could be used with each of the four networks. The four value profiles are

- Equal value weighting *each connection is valued at* 1
- Truncated *assuming there is a limit to the number of friendships that can be maintained, a certain number of connections are valued at* 1 *and after this no more connections can be made*
- Zipf value *the first connection a node makes is valued at* 1*, the second at* $\frac{1}{2}$*, the* n^{th} *at* $\frac{1}{n}$

- Truncated Zipf value *as for Zipf, but assuming there is a limit to the number of friendships*
- Dunbar value *following on from Dunbar's work, different groups are given different values, with value decreasing as group size increases*

4.1 Verifying Existing Results

We realised early on that the Small World model has identical value properties to the random model since average path length does not affect the value within the network. Whilst it was felt that, in fact, connecting to a node that created short path lengths would be a potentially beneficial move for another node, and hence one that could be highly valued, there are two issues to consider. Firstly there is no data to support this hypothesis. Secondly, the calculation of average path length within Matlab was far too calculation heavy and could not even be implemented once, never mind at each iteration to facilitate connection choices. For this reason the Small Word network does not feature in any analysis.

If an equal value weighting is used across all nodes then, irrelevant of the network structure, total value growth follows Metcalfe's Law and is quadratic. Figures 4.1 and 4.1 clearly show this.

When a truncated, equal, value weighting is used the results are, again, as expected. A short, initial, quadratic growth phase is seen and then as all nodes reach their maximum saturation growth becomes linear, with each new node adding a set amount of value (proportional to the maximum number of friends). This can be seen in Figure 4.1.

When we consider Zipf function value growth, the results are as we expect from theory (proved in Appendix A). That is, growth is proportional to $n \log n$. These growth graphs, along with $n \log n$ for comparison can be seen in Figures 4.1 and 4.1.

4.2 Model Requirements

Following our review of empirical evidence in Sections 2.1.2 and 3.2. We decide upon the following features as desirable in the creation of a valued

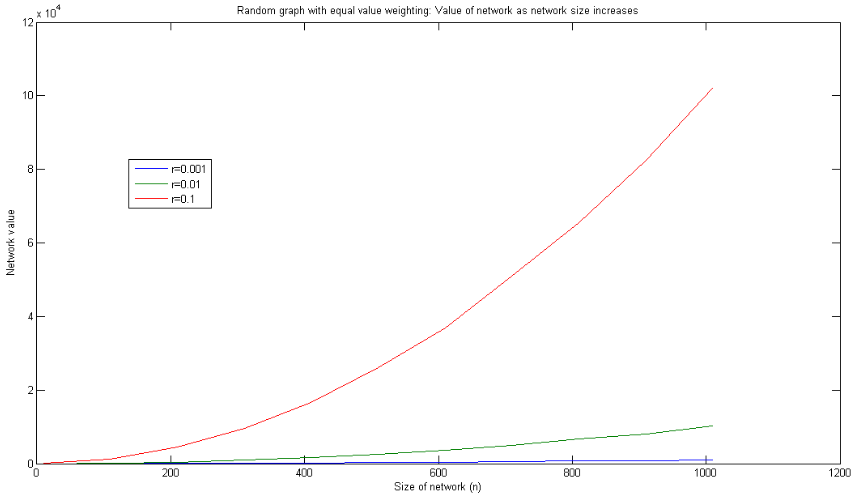


Figure 4.1: Growth is quadratic when equal value weighting is used

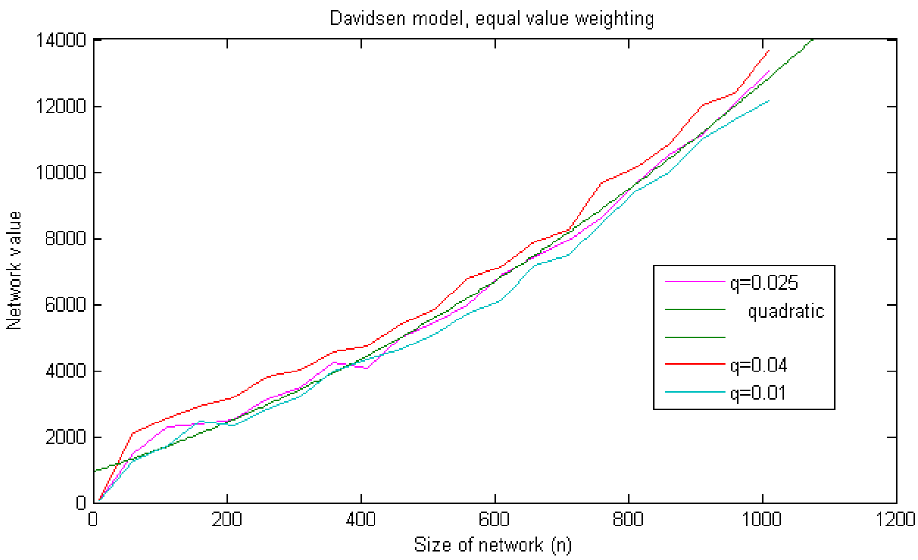


Figure 4.2: Growth is quadratic when equal value weighting is used

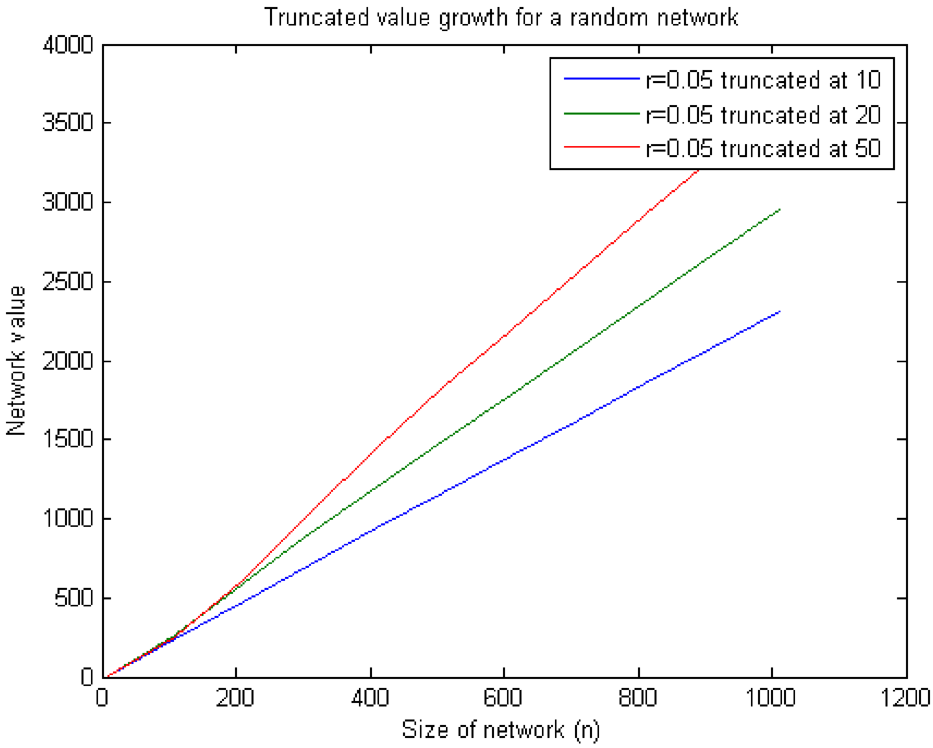


Figure 4.3: Truncated value results in an initial quadratic growth phase becoming linear

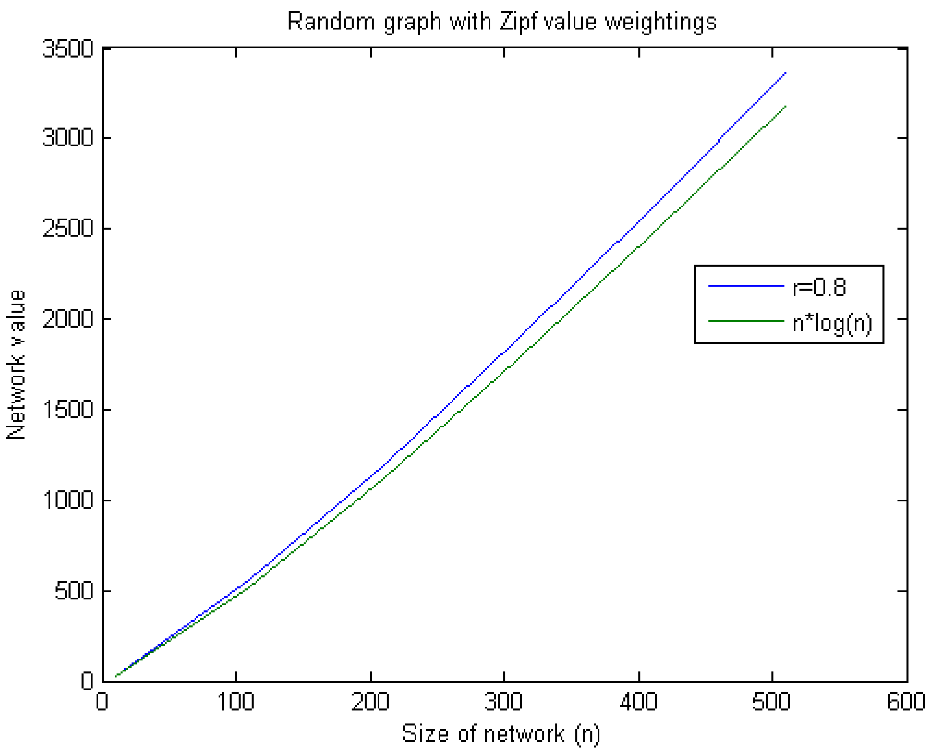


Figure 4.4: Zipf grows value O(log(n))

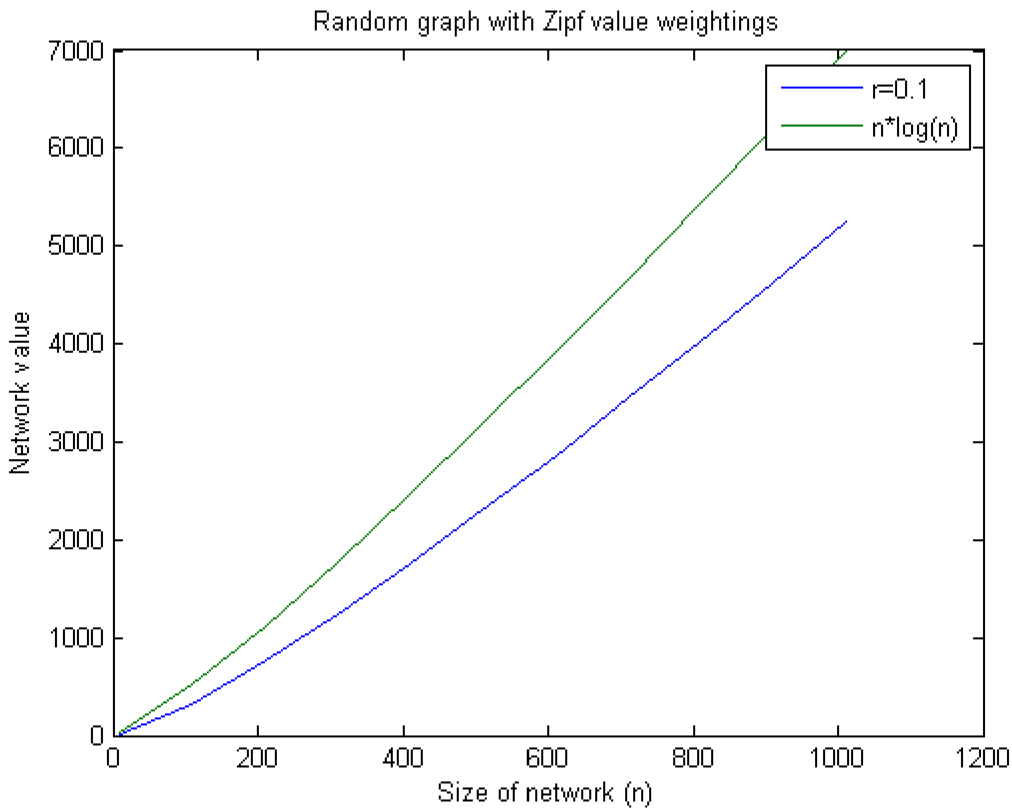


Figure 4.5: Zipf grows value O(log(n))

social network model.

- A large number of nodes (ideally $O(10^6)$, although this will be impossible on Matlab)
- An average number of friends of between,approximately, 100 and 150
- A power law degree distribution with an exponent of around $\gamma = 1.5$.
- A clustering coefficient of approximately 0.3, and certainly higher than that expected at random
- Short average path-lengths of around 5
- A value profile that distinguishes between different levels of friendship (from Dunbar's groupings), truncated at 500 connections per user
- A value profile that distinguishes between two types of users, with around one half deriving value from direct connections only and the other half also finding value in connections between friends

4.3 Final Model

For the final model we elected to use Davidsen's model. This was chosen for several reasons.

Firstly it is a dynamic model. Since we are measuring value growth, this is considered an important factor. To measure the growth of all the static models, many different sized models were created and valued. Whilst this gives a good indication of how value varies with size it isn't quite the same as an actual 'growing' network.

Secondly it offers us a power law degree distribution. The final model uses $q = 0.0025$, which results in $\gamma \approx 1.35$, which is close to the required 1.5.

It also, for $q = 0.0025$, offers us an average degree of 149.2 [9], which fits perfectly with our requirements.

Thirdly it provides high clustering. The clustering coefficient for $q = 0.0025$ and $N = 7000$ is quoted in [9] as 0.63.

Fourthly, average path lengths are too expensive to calculate on Matlab, but Davidsen quotes a value of $l = 2.38$, which, whilst shorter than observed, is still of a similar order, and does, at least, demonstrate the 'small world' nature we would like to see.

It is the high clustering and low path lengths that made the Davidsen our model of choice above the Barábasi-Albert model (which is dynamic and has a power law degree distribution).

Our Davidsen model requires four inputs, n, q, r and $maxtime$. The matrix dimension, n determines the maximum number of people within the network, although since the network is growing dynamically not all of these are 'live'. We calculate the actual size of the network at any given time by finding the number of nodes of non-zero degree. The probability q gives our model a time scale, it is the probability that, at each timestep, a node will be deleted. This was the main parameter that needed adjusting as it affected the degree distribution of the model. In the final model we found that the input value r (used as the probability of connection in the initial setup of a random network) had little effect, except when $r >> 0$ in which case the random graph dominated the preferential attachment model. For this reason we set $r = 0$. Therefore at timestep one the matrix is empty and there are no connections, and the network grows from here. Finally we consider $maxtime$. This is the number of iterations through which the model will run.

Initially we attempted to use the value distribution based directly on Dunbar's observed group sizes, but there were two main drawbacks. Firstly it proved difficult to decide on precise values for each group level (Dunbar gives us none). Secondly because we were unable to attain large enough network sizes (due to restrictions in computing power), average numbers of friends are stuck at around 30. Dunbar's figures require many more friends than this and it was unclear how to scale down the group sized accordingly. It was concluded that a Zipf function showed many similarities to Dunbar's figures and that, in fact, Dunbar's empirical figures are quite likely an approximation of a underlying Zipf distribution. We therefore grow value in accordance with Zipf.

To distinguish between the two types of user we split the network into two. This was implemented in two separate ways. Initially we set all even numbered nodes to be of the linear/email user type and odd nodes to be group users. After some thought we realised that, as suggested by Patel [30], it is most likely the early adopters who will have higher usage levels. We therefore adapted the model so that the first fifty percent of users are of the group use type, and the second half are the more casual email users.

When calculating the value of the network to a 'linear' user, i, we simply summed along the i^{th} row, where the values $A(i,j)$ indicate the value that i places upon j. Since the valuations were made using a Zipf function this approximates to $\log n$. For the 'quadratic users' we initially sum the individual valuations as we did for linear users. To calculate the value to i of two friends communicating with each other we firstly need to know if these two friends of i are indeed friends with each other. This proved too calculation heavy, and we decided to approx the situation by assuming that all friends were mutual. Whilst this is far from being true, with such a high clustering coefficient, it is not too far fetched either. To calculate the third party connection we calculated $A(i,j) \times A(i,k)$ to find the value to i of j and k communicating with each other. This values connections between two valued friends more highly than connections between those less important, which fits with qualitative opinions of Facebook's users on this matter.

Despite the fact that a large network was at the top of our list of criteria, $N = O(10^6)$ was never going to be achieved with our matrix setup on Matlab. We found that $N = O(10^3)$ was achievable, which in fact, is all that Davidsen *et al* worked with in [9]. It appears that modelling large enough networks is a universal problem.

Chapter 5

Conclusions

5.1 Results of the Model

The amount of calculation and memory required to handle the matrices within Matlab, even utilising the sparse matrix setup, has restricted us to networks with $N = 600$ and 10000 iterations of the growth cycle. This had huge implications for us regarding the upper limit on the number of friends, which was set, from Dunbar's figures, at 500. The average number of friends per user remained at around 3, and hence the cap was useless. Figure 5.1 shows the degree distribution for this case. This meant that with the cap in place we were seeing greater than quadratic growth. See Figure 5.1 to see this.

Enlarging the network was not possible, and therefore, in an attempt to see more realistic results, we reduced the cap. Unfortuanately, due to the size of the required networks, we could not get any results worth printing. With a cap of one in place, growth is linear, but this result is trivial, since if each person in a network can have only one friend, of course growth will grow linearly (no interactions are possible between mutuall friends either). Even with a cap of just two, we could not grow a large enough network.

Whilst it is very disappointing that no graphical or numeric results could be produced, we still believe that this research suggests that networks go through an initial phase of fast value growth, better than quadratic, before all users become saturated and then growth continues linearly. This is in fact the *opposite* of Metcalfe's Law (which implied that growth may start slowly

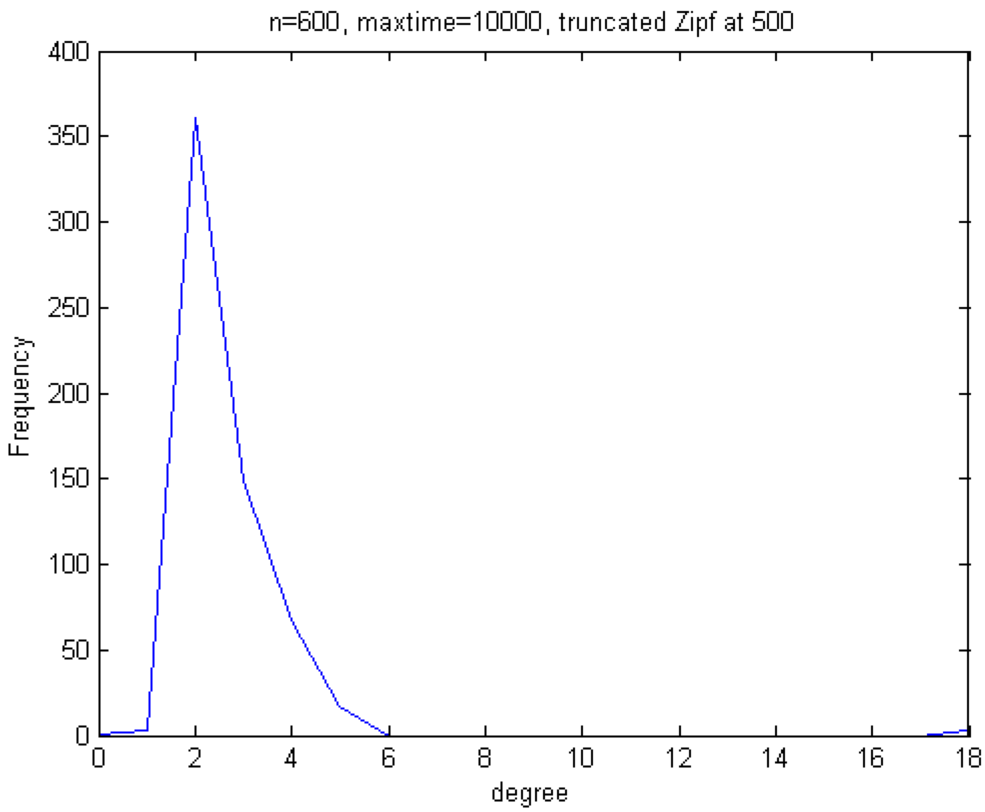


Figure 5.1: Computing restrictions resulted in low average degree per node

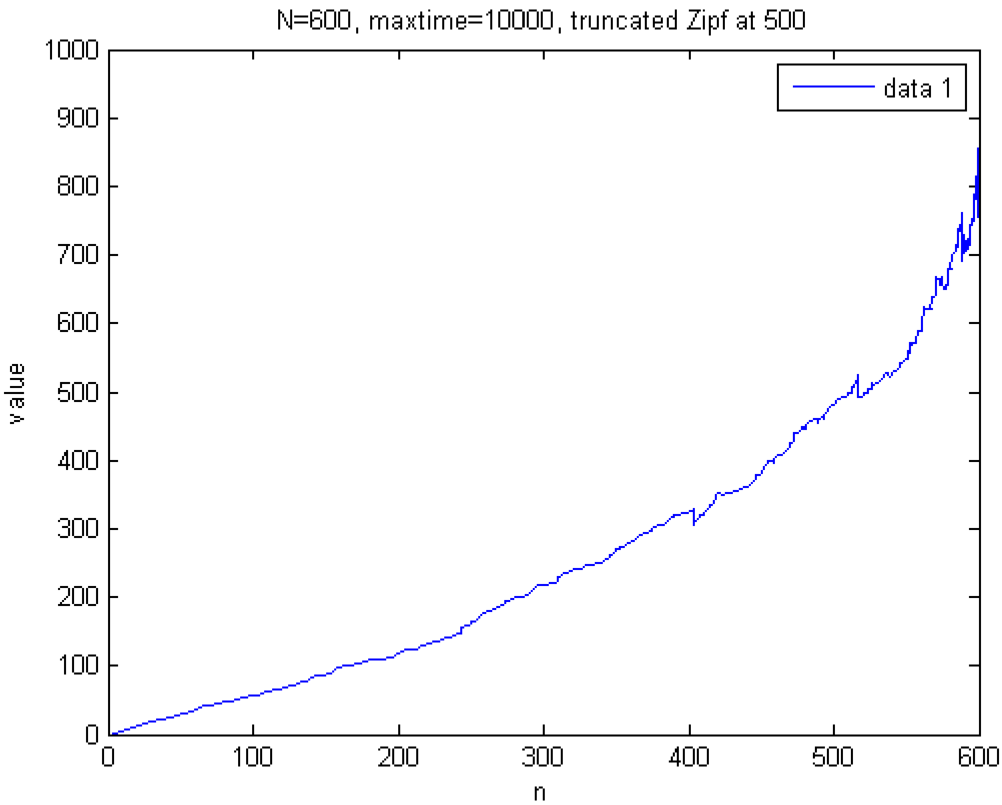


Figure 5.2: Growth is better than quadratic when the truncation is set at 500

and then would become quadratic once a certain critical mass was achieved).

The question is, when is this point reached in real online social networks? We were not able to achieve realistic values due to computing constraints, it would be interesting to try this model on a more powerful machine.

Whilst our results are far from conclusive, we do not believe that Metcalfe's Law is valid for online social networks.

5.2 Review of the Model and Suggestions for Further Work

This project has thrown up far more questions that it has managed to answer.

We feel that the most exciting idea, worthy of further consideration, is that of the two (or quite possibly more) user types within a network behaving differently in their approaches to valuing the network. We split our Facebook users into 'email' and 'group' types, although it is immediately clear that three types of communicators (one-to-one, those interested in the communication of others, and group users) may have been more appropriate. It is, of course, possible that people move between different types of usage, and this could be factored in as well.

The very small sample size for the Facebook analysis, along with the many possible biases in the sample, is an obvious issue and could certainly be addressed in a larger survey. Intrinsically, any questionnaire of this nature will attract more heavy users which will bias the data. Mislove *et al* [24] got around this problem very well in their survey with screen scraping the data. It would be simple to find a users number of friends, age and sex using this method, in addition, Facebook pages do contain the information about a users current online/offline status, and whilst massively time-consuming it would be possible to collect the hours online data in this way too. The user type was, as discussed, a subjective decision on either their part, or ours. It would be possible to put stricter definitions on this, to avoid any bias here.

This report uses data from Dunbar's investigation of social group sizes [11]. It is not known how relevant this data is to online social groups. We would expect a certain, probably large, amount of similarity, but it would be interesting to see similar surveys being conducted of online friendships. The data would be easier to find online, and it seems likely, with current interest in online networking, that this research will appear soon.

Bibliography

[1] Albert and Barábasi, *Topology of Evolving Networks: Local Events and Universality*, Physical Review Letters, **85**, 2000, 5234–5237

[2] Amaral, Scala, Barthélémy and Stanley, *Classes of small-world networks* PNAS, **97**, 2000, 11149–11152

[3] Arthur, *Increasing Returns and the New World of Business*, Harvard Business Review, July 1996

[4] Backstrom, Huttenlocher, Kleinberg and Lan, *Group Formation in Large Social Networks: Membership, Growth and Evolution*, KDD, **'06**, 2006, August 20–23

[5] Barábasi and Albert, *Emergence of Scaling in Random Networks*, Science, **286**, 1999, 509–512

[6] Barrat and Weigt, *On the properties of small-world network models*, European Physical Journal B, **13**, 2000, 547–560

[7] Briscoe, Odlyzko and Tilly *Metcalfe's Law is Wrong*, IEEE Spectrum, 2006, 26–31

[8] Chung and Lu, *Complex Graphs and Networks*, CBMS, 2004

[9] Davidson, Ebel and Bornholdt, *Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks*, Physical Review Letters, **88**, 2002, 128701-1–4

[10] Dorogovtsev, Mendes and Samukhin, *Structure of Growing Networks with Preferential Linking* Physical Review Letters, **85**, 2000, 4633–4636

[11] Dunbar and Hill, *Social network size in humans* Human Nature, **14**, 2002, 53–72

[12] Dunbar, *Coevolution of Neocortical Size, Group Size and Language in Humans*, Behavioural and Brain Sciences, **16**, 1993, 681–735

[13] Facebook, *http://www.new.facebook.com/press/info.php?statistics*

[14] Fichman, *Information Technology Diffusion: A Review of Empirical Research*, Proceedings of the International Conference of Information Systems, **13**, 1992, 195

[15] Gilder, *Metcalfe's Law and Legacy*, Forbes, September 1993

[16] Grossman, *The Business Value of Social Networking*, Forbes, July 2008

[17] Goyal, *Connections*, Princeton UP

[18] Holme and Newman, *Nonequilibrium phase transition in the coevolution of networks and opinions* Phys. Rev. E 74, 2006

[19] Lauria, *MySpace Loves Facebook Value*, New York Post, 26 October 2007

[20] Liebowitz and Margolis, *Network Externalities (Effects)*, The New Palgraves Dictionary of Economics and the Law, MacMillan, 1998

[21] Leskovec, Singh and Kleinberg, *Patterns of Influence in a Recommendation Network*,

[22] McCarty, Killworth, Bernard, Johnson and Shelley, *Comparing two methods for estimating network size*, Human Organization **60**, 2001, 28–39

[23] Milgram, *The Small World Problem*, Psychology Today, **2**, 1967, 60–67

[24] Mislove, Marcon, Gummadi, Druschel and Bhattacharjee, *Measurement and Analysis of Online Social Networks*, IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, 2007, 29–42

[25] Newman, Watts and Strogatz, *Random graph models of social networks*, PNAS, **99**, 2002, 2566–2572

[26] Newman, *The Structure and Function of Complex Networks*, SIAM Review, **45**, 2003, 167–256

[27] Newman, Strogatz and Watts, *Random graphs with arbitrary degree distrubutions and their applications*, Physical Review E **64**, 2001,

[28] Newman, Barábasi and Watts, *The Structure and Dynamics of Networks*, Princeton UP, 2006

[29] Odlyzko and Tilly, *A refutation of Metcalfe's Law and a netter estimate for the value of networks and network connections*, www.dtc.umn.edu/ odlyzko/doc/metcalfe.pdf

[30] Patel, *Report on Value Within A Communication Network*, Confidential internal Unilever document, 2007

[31] Pool and Kochen, *Contacts and Influence*, Social Networks **1**, 1978, 1–48

[32] Reed, *That Sneaky Exponential - Beyond Metcalfe's Law to the Power of Community Building*, Context, 1999

[33] Simeonov, *Metcalfe's Law: more misunderstood than wrong?*, 2006, http://simeons.wordpress.com/2006/07/26/metcalfes-law-more-misunderstood-than-wrong/

[34] Vega-Redondo, *Complex Social Networks*, Cambridge UP

[35] Watts and Strogatz, *Collective Dynamics of 'Small-world' Networks*, Natture, **393**m 1998, 440–442

[36] Watts, *A simple model of global cascades on random networks* PNAS, **99**, 2002, 5766–5771

[37] Weinman, *Is Metcalfe's Law Way Too Optimistic?*, Business Communications Review, 2007, 18–27

[38] Wikipedia, www.wikipedia.org/wiki/Diffusion-of-innovations

Appendices

Appendix A

Derivation of $n \ln(n)$ Value Growth

We assume that, on average, the value of being able to communicate with someone is in inverse square proportion to their distance, i.e.

$$v_i = \frac{1}{r_i^2}$$

where v_i is the value of communicating with an individual at distance r_i. Then, as long as the distribution of the population is uniformly distributed in a disk of radius R, the total value to an individual, v, can be found be integrating over the disk,

$$v = \int_0^{2\pi} \int_0^R \frac{1}{r^2} r \, dr \, d\theta = \int_0^{2\pi} \ln r|_0^R d\theta = \int_0^{2\pi} \ln R d\theta = 2\pi \ln R$$

The total value, V, of the network will be n times the value to each individual,

$$V = nv = 2\pi n \ln R$$

and since the total population is proportional to the size of the disk, i.e. $an = R^2$,

$$V = 2\pi n \ln \sqrt{an} = \pi n \ln a + \ln n \propto n \ln n$$

Appendix B

Finite Sum of the Harmonic Series

For any monotonically increasing function, $f(k)$, we know that

$$\int_{m-1}^{n} f(x)dx \leq \sum_{k=m}^{n} f(k) \leq \int_{m}^{n+1} f(x)dx$$

Therefore, for our sequence,

$$\sum_{k=2}^{n} \frac{1}{k} \leq \int_{1}^{n} \frac{dx}{x} = \ln(n)$$

$$\sum_{k=1}^{n} \frac{1}{k} = 1 + \sum_{k=2}^{n} \frac{1}{k} \leq \ln n + 1$$

Appendix C

Matlab Code