Structure of an On-Line Community: Bands in MySpace

Dietmar Offenhuber[1] & Whitman Richards[2]
{dietmar, whit}@media.mit.edu
(1) Media Lab; (2) CSAIL
Mass. Inst. of Technology
E15-390, Cambridge, MA 02139

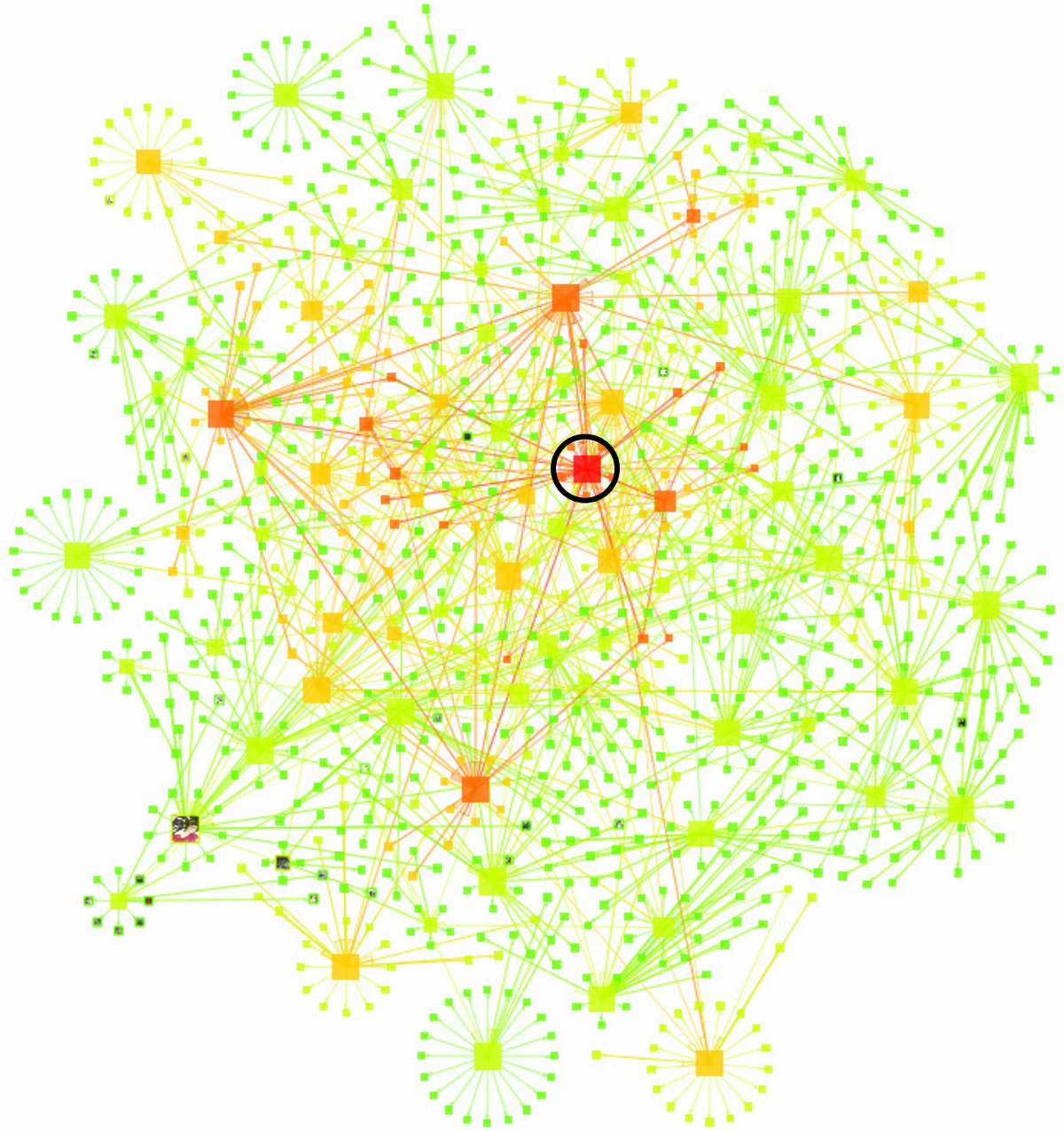

Figure 1 Visualization of a 4-hop network sampled from the MySpace. The marked red node represents the start profile, the color gradient shows the topological distance – dark orange is distance one, light orange distance two, light green distance three and dark green distance four. The size of the vertices indicates the vertex degree.

Abstract

With over 170 million users, MySpace is the currently the most popular social networking site, reaching the status of a mass medium on its own. The range of users spans high school students to fan clubs, gourmet, games, and investor groups. We thus have many communities, each with different interests and objectives. One such community are bands and composers, who use MySpace as a promotional platform. To our surprise, their network structure and activity appears different from networks previously studied [1]. Although in the public domain, MySpace communities may provide insights into how member objectives influence network structure and behavior.

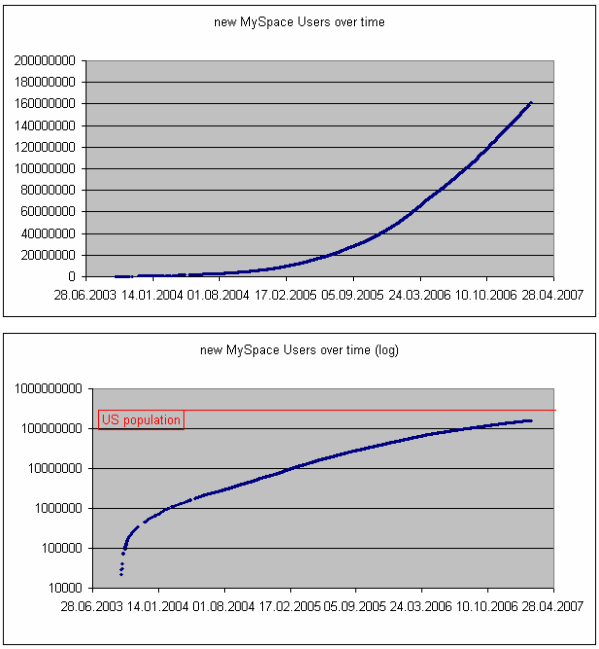

Figure 2 Number of members in the MySpace network (based on sequential id numbers) Jun. 2003 – mar. 2007. Note the exponential growth of the service (left) which is replotted (right) using a logarithmic scale that makes clearer an approaching asymptote.

Introduction

Over the last several years, online social networking platforms have become widely popular. Burgeoning communities have been emerged with similar interests. Participants create self-descriptive profile pages that often include their tastes and preferences, with links to other members of the same community, usually called ”friends”. MySpace [3] is the most popular service of its kind with over 170 million users (see Fig. 2.)[7]. With its status as an emerging mass medium, MySpace is especially attractive to artists and musicians for the promotion of their work, not only offering samples, but more importantly allowing them to target new audiences with compatible tastes. D. Boyd describes it as a “symbiotic relationship between bands and Fans”, since “bands wanted to gather fans and fans wanted to be connected to their favorite bands.” [8] One might expect the structure of this social network to be similar to many others examined, especially peer-to-peer networks, or perhaps citation networks [1]. However, MySpace users have different objectives. Consider the aspiring musician, who sets up a profile in order to maximize her audience and to stay in contact with her fans. The following design of a profile raises several issues:

1. How can people find my profile and which paths in the network would they use?
2. Which friends should I connect to in order to maximize visitors on my profile page?
3. How many links to friends should I show on my profile (more is not necessarily better)?
4. Is there a way to predict the number of page visitors?

This trade off between promotion, search and navigation brings together in one social community several aspects that are often considered independently. How will these different objectives affect the structure of the network? Will the network be a typical multi-scale structure with parameters similar to peer-to-peer networks? If not (which will be seen the case), then how should they be characterized? Answers to these questions help illuminate structural variations for on-line communities, and how these variations relate to the objectives of the community members.

Description of MySpace Network

Member profiles

Every Member of the MySpace network maintains a self-descriptive profile page, containing personal information, media files such as music, photo collections or videos, contributions and guest

book entries from other friends, and, most important, links to the profiles of friends.

While a profile can have an unlimited number of friends, only a small number of them can be featured on the front page of a profile, the so called top-n area. An important factor in our analysis is the number of places in the top-n area an owner chooses in the design of their profile page. Looking at the histogram in Figure 2, we find a peculiar oscillating pattern. The explanation is simple and encouraged by the standard layout of MySpace profiles: most users choose a number that can be neatly arranged in a rectangular table of four columns: 8, 12, 16, 20 and 24.

If your objective is to attract new visitors, obviously not all friend links are equally valuable. Therefore many band profiles establish a friend link with an influential artist, in order to attract people sharing a compatible taste to her own profile. Clearly, establishing the right social neighborhood is important - typically the top-n areas of artists and bands on MySpace include a few famous colleagues serving as references for the musical tastes, a number of links to other "friendly" bands they are connected to and often share audiences with in real life, as well as a small number of personal friends or fans from outside of the musical world. Additional information about friends can also appear in a comment area located immediately below the top-n profiles, where any friend can post messages and images.

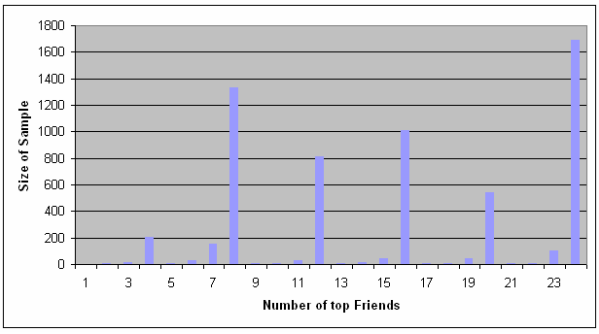

Figure 3 Frequencies of profiles with a specific size of the top-n Box (number of top friends).

Links to Friends

In addition to friends that appear as a profile, or in the comment area, there is a special link to an expanded set of friends. As will be seen, this list can number in the hundreds. To the casual visitor, who does not search through the depths of a site, they are invisible. The consequence is that many communities, especially the bands to be examined, typically have a very large number of friends, usually more than the number of social connections an individual can be expected to maintain. One thousand friend links are not uncommon; but also profiles with over a million friends are found. For example, every new profile starts with "Tom", the profile of MySpace founder and CEO Tom Anderson as a friend by default, a profile with over 160 million friends, so the shortest distance between any two users is usually two (some profile owners choose to remove this link from their friend list).

The surprisingly high number of friend connections becomes more comprehensible when looking at the nature of MySpace profiles: A profile may stand for a person, but could also represent companies and interest groups, historical personalities, cartoon characters, general topics or a commercial. Profiles often serve the purpose of identification – a profile owner might have the profile of i.e. Karl Marx among his friends, in order to signal a personal preference to visitors. This allows her to connect to other friends of Marx, who supposedly share similar interests. Distinguishing between them is often hard, since the service offers no way to differentiate between kinds of relationships beyond the generic "friend" status, whereas in real life we have a multitude of different terms to describe relationships between people.

Data Acquisition & Methods

For our study of the network structure, we collected around 30502 profile pages, starting with profile pages of around 10 different bands and subsequently traversing their social network. Band profiles, usually representing the professional identity of a band or a musician, differ from basic generic MySpace profiles in a number of ways: They are listed in a special directory, the music section of MySpace, include information and promotional material and, finally, publish some figures associated with the popularity of the band, such as the number of visitors on the profile. A generic profile can be converted into a band profile through a special function offered by the MySpace site. From our sample of 30502 profiles, 12708 are band profiles, 14371 generic profiles. The remaining 3425 profiles were set to "private", only accessible for their friends

with only their name and profile picture visible to the public.

For reconstructing the social neighborhood of a profile, we start from a single profile page and extract all the links from the top-n and the comment areas, adding them as linked nodes to the graph (see Fig. 1). Following these links we check whether the original profile is listed under the top-n area, or in other words, whether the top-friendship is bi-directional or not. Including bi- and unidirectional links, the graph becomes prohibitively large beyond distance 3, while still being too sparse, since most incoming links are missing. During each iteration we check for bi-directionality and subsequently remove all unidirectional friend links. This allows us to conduct more iterations and results in a denser network.

Results

To provide some preliminary statistics based on our sampling of generic MySpace profiles, we found a median friend count of 205 with S^2= 1,88722 10^{12}**.** Band profiles, on the other hand, had a median value of 1080 friends, with S^2= 3139912465. The median number of profile views for musicians within this sample is 6579. This figure is unavailable for generic MySpace profiles.

Friend (or degree) Distribution.

Common measures for describing networks are the distribution of vertex degrees, the clustering coefficient, and the mean vertex-to-vertex distance [1]. The vertex degree for each person is simply the number of friends; the clustering coefficient is a likelihood measure of whether the friend of your friend is also your friend; and the mean vertex to vertex distance is the mean of the minimum number of hops to reach another person. For scale-free networks, the degree distribution (i.e. the number of vertices vs. the vertex degree) falls off as a power law, with an exponent typically in the range of 2. For these graphs, the clustering coefficient can range from 0.1 or less (e.g. peer to peer networks) to 0.4 or so (e.g. www). Mean distance between pairs of vertices also varies widely in scale-free graphs from about 3 to 10. Small world graphs have slightly different properties, and are noted for their high clustering coefficients, in the 0.7 range. Our estimate for the cluster coefficient for the MySpace network is about 0.4, which is above expected peer to peer values. Finally, random graphs constitute another class, distinguished by an asymptotic mean vertex-to-vertex distance of 2 for very large graphs. For MySpace, because the founder and CEO Tom Anderson is typically a friend by default, the vertex to vertex distance is about 2. These parameter values for MySpace do not appear to fall into an obvious network category [1].

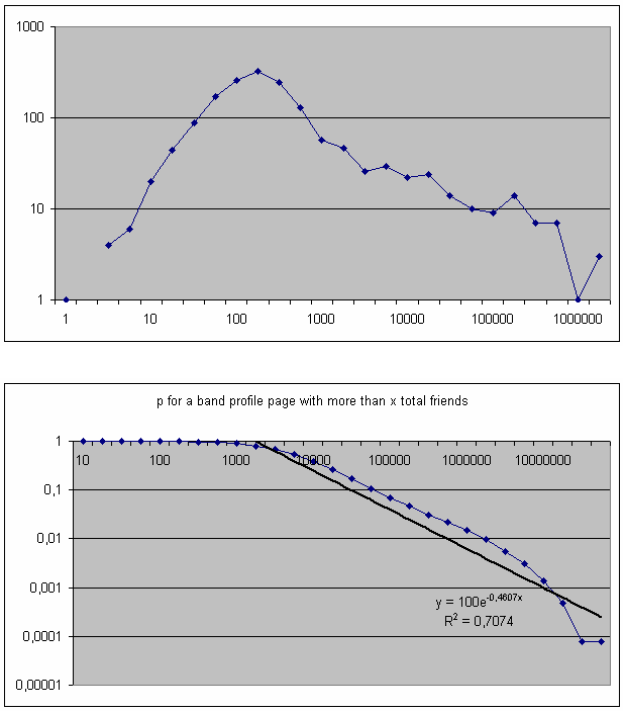

Figure 4 Histograms of Friend Counts (Top). The probability of a band profile with more friends than shown on the abscissa (bottom).

Figure 4 shows the degree distribution for the MySpace network. The top distribution is just a log-log plot of the number of members (ordinate) having the number of friends shown on the abscissa. The average number of friends is about 200, as mentioned earlier. In Fig 5, we show similar data for the band profiles, with a mean of about 1000. In the bottom of both figures, the same results are replotted as a cumulative distribution. This clearly shows a power-law tail. The exponent for the cumulative plot is –0.4, which is equivalent to a power law exponent of 1.4, which is in the range of email or scientific citation networks. But preceding this tail is a broad region, mostly below the mean, which is quite uncharacteristic of all scale-free graphs. Indeed, the distribution is roughly Poisson on the log-log plot. MySpace itself, as well as the musicians' community,

thus appears atypical of the four classes of 26 networks cited in Newman's review (Social, information, technological, biological.)

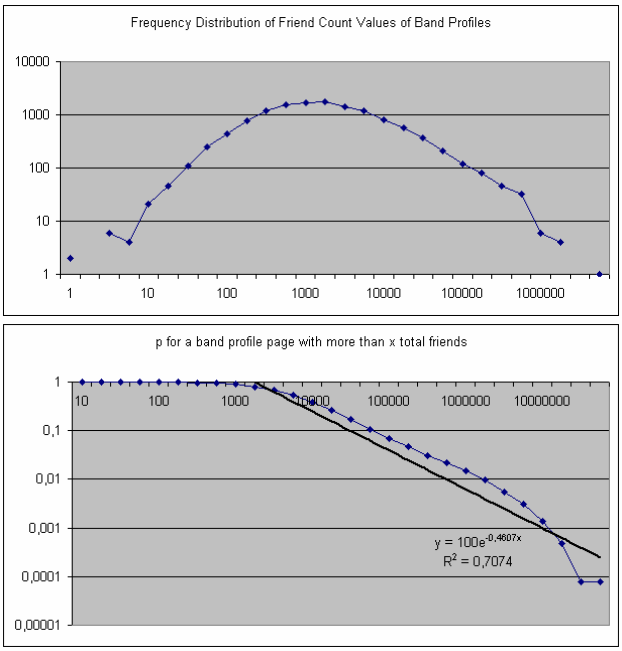

Figure 5 Histograms of Friend Counts (Top) and probability for a band profile with more than x total friends (bottom).

Site Activity and Number of Friends

One would expect that the more friends one has, the more activity or hits (i.e. page views) will occur on the site. In turn, this implies that older members should have more total hits (the number of page hits from day 1 to now). Indeed although there is a high variation, page views and friend counts correlate with the age of the membership, especially over the past year (Figure 6.)

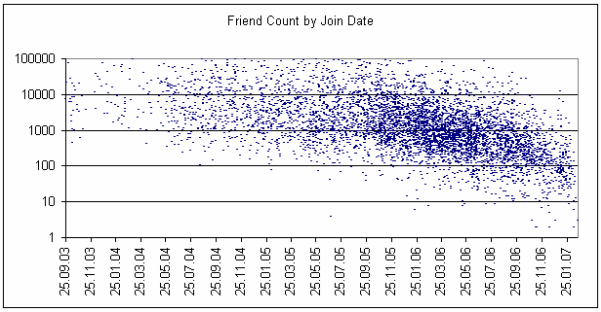

Figure 6 (top) Number of friends vs. age of membership.

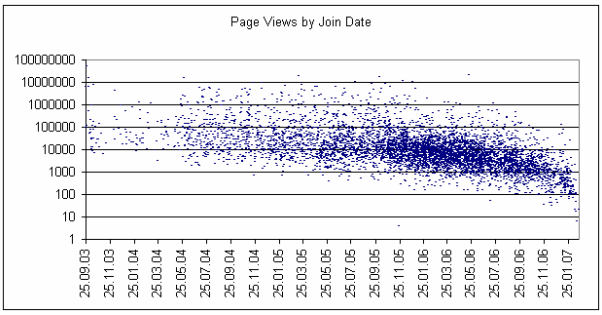

Figure 6, bottom: number of page views vs. age of membership

If we now plot the total number of page views vs. the total friend count, we find the correlation shown below in Figure 7 with a linear relation on log-log coordinates. The surprise is that, although no information about a profile's neighborhood or age has been taken into account, the horizontal scatter is much smaller than expected.

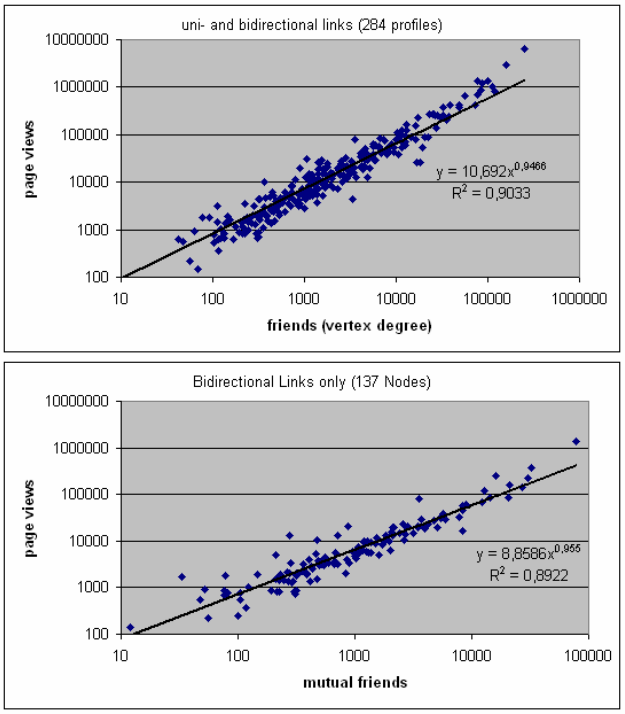

Figure 7 Band Profiles: Number of Page Views vs. Number of Friends: Uni- and Bidirectional Links (top) Bidirectional Links (bottom) - both in log. Space

However note that now we have used profiles from the band community restricting friends to a three hop network. The resulting R^2 correlation values are 0.903 and 0.892 respectively for unidirectional and bidirectional links.

The directionality distinguishes whether a friendship is reciprocal or not. For example, some of the top-n friends may not be reciprocal friends (e.g. two friends may not each be displayed on the other's profiles). In order to account for this ambivalence, we present two versions of the distance 3 network. There is little difference between unidirectional and bidirectional links (Fig. 7).

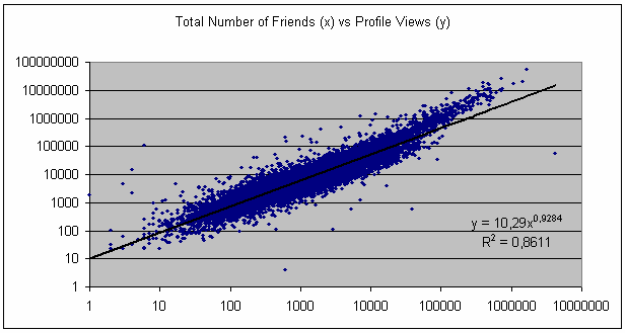

Figure 8 Distribution of Friends vs. Page Views across 12708 Band Profiles (excluding Generic Profiles).

The observed correlation is still present if the distance 3 hop constraint is removed. In this case, for bands alone, there will be 12708 profiles. Figure 8 shows the same linear correlation with R^2 of 0.861, with the appearance of a few outliers.

We can now examine the spread of the data points along the regression lines to determine the mean and distribution of friends as well as page views. These distributions of course will be highly correlated, and have a similar Gaussian form (Figure 9.) The median value is 1080 for the vertex degree (friend count) and 6579 for the page view count. These distributions characterize nominative expectations for site membership and activity.

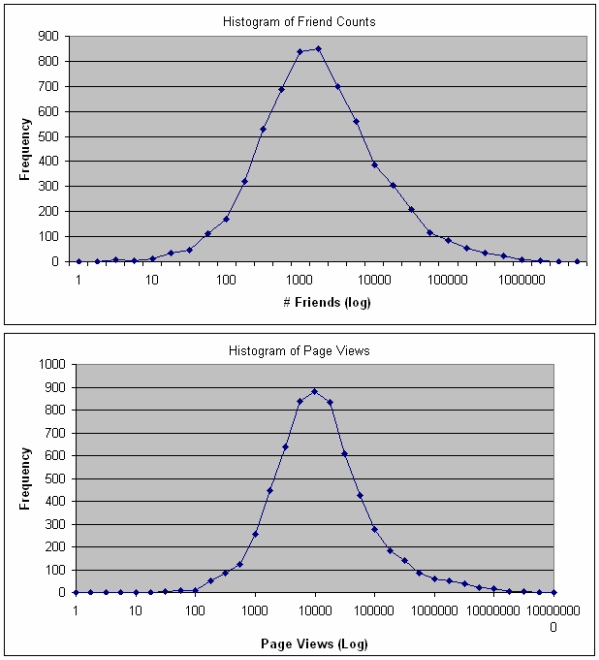

Figure 9 Histograms of Friend Counts (top) and Page Views (bottom)

Design Choices for Profiles

Finally, we take the collected music profiles plotted in Figure 9 and separate them by the size of their top-n box. Hence we have 24 bins of data, and for each bin we can compute the regression line for the relationship between friend count and profile views (similar to Fig. 7.) The result is that for profiles with a smaller numbers of featured friends, the correlation value is higher than if a site features many friends (Fig 10, top.) This result is initially a bit puzzling. The implication is that when there is a large number of featured friends, there is no correlation between the number of friends and the page views. In other words, if we have many featured friends on the profile page, we are unable to predict the number of hits (page views).

We can also look at the slope of the regression line of page views vs. friends. This will show the efficacy of a site in eliciting hits (Fig 11, bottom.) These slopes show a clear peak at 12-16 top friends. This indicates that the gain in hit rate per friend will increase up to 12 - 16 top friends, but decrease thereafter.

The above findings reflect more on the activity of reciprocal relationships between friends and little about how the network is traversed. If we

had information about the total number of incoming links, then we could make a more definitive statement about the relation between navigation and friendship activity.

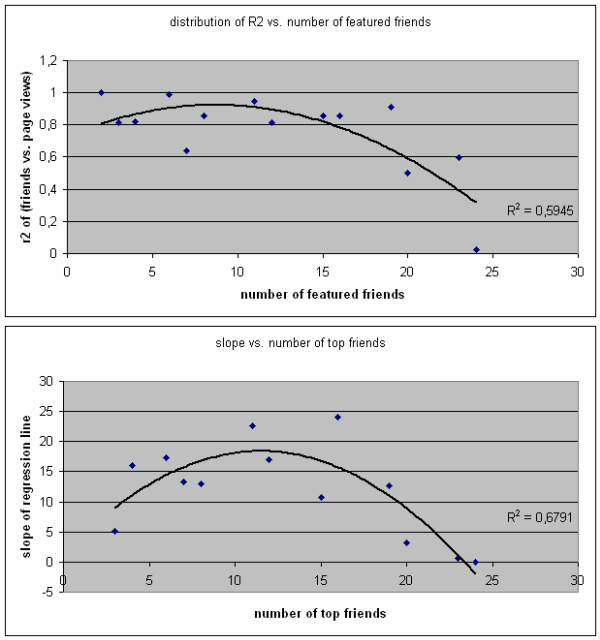

Figure 10 Top: comparing the correlation values of the linear friend / page view regressions to the size of the top friends area. Bottom: comparing the slope of the same regression lines to the size of the top friends area.

Discussion of the Results

An obvious open question is how people navigate the MySpace network in order to meet new people. Unfortunately, this question is difficult to answer by counting hits (page views). First, the reciprocal use of present links to friends may dominate any navigational hits. In this regard, the MySpace search engine is very advanced and offers many options for filtering. For example, with a few clicks it is possible to list all profile pages of single females with an age between 17 and 25, who live in Cambridge, MA.

Hence navigation on MySpace probably takes place primarily through browsing via the internal search engine instead of traversing the links of the network. The MySpace network is for interaction not for exploration.

The interactive dynamic of MySpace is clearly influenced by many variables: the interface design, the profile content, the group identity selection and the network structure. Earlier we noted that the classical network parameters seem distinctive and are not typical peer to peer values. Although technically the diameter of MySpace network is very low (~2) this is a result of its construction where almost everyone is linked to the founder or other very distinguished members of that group. However, even with this imposition of a hierarchical structure we do not find the expected scale-free network (Fig. 4). Rather, the current evidence suggests that for the band interest group, the clustering about all nodes is very high with short links between all clusters. This means that for any profile the neighbors of that profile have many shared links, which fosters the interaction between the group members.

Although MySpace is mainly a public site, the high activity within its interest groups may lead to insights into other virtual communities, some which may be less public. An important aspect of the MySpace community is that many profiles list friends that may not be people or people that may be deceased. These include for example influential composers from previous epochs or fake profiles of celebrities. These profiles reflect the character of a member ("one is also judged by the company one keeps"[4]). They demonstrate the identity of the group to its members and could easily serve to rally present members and to recruit new members to some cause. But, as we mentioned at the outset, one of the objectives of MySpace is promotion.

References

[1] M. E. J. Newman, The structure and function of complex networks, SIAM Review,
45(2):167-- 256, 2003.

[2] A. Barabasi and R. Crandall. Linked: The New Science of Networks,
volume 71. AAPT, 2003.

[3] see http://www.myspace.com

[4] J. Donath and D. Boyd. Public Displays of Connection. BT Technology
Journal, 22(4):71–82, 2004.

[5] S. Feld. The Focused Organization of Social Ties. American Journal
of Sociology, 86(5):1015, 1981.

[6] L. Freeman. Visualizing Social Networks. Journal of Social Structure, 1(1):4, 2000.

Newman, M.E. J. L. The structure and function of complex networks, SIAM 45, 167-256 , 2003

[7] Pew Research Center, Social Networking Websites and Teens: An Overview, 2005, url: http://www.pewinternet.org/pdfs/PIP_SNS_Data_Memo_Jan_2007.pdf

[8] D Boyd, (in press) “Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life.” MacArthur Foundation Series on Digital Learning, Identity Volume (ed. David Buckingham).