Email Worm Propagation on Random Graphs

Stephan Bohacek
University of Delaware
Department of Electrical and Computer Engineering
Newark, DE 19716
bohacek@udel.edu

Abstract

The theory of random graphs is applied to study the impact of the degree distribution on the spread of email worms. First, the idea of the email address book graph is introduced. It is shown that the structure of this graph plays a critical role in the propagation of worms. Second, convincing evidence is provided that the email address book graph has a heavy tailed degree distribution. The implications of this degree distribution are then investigated. It is shown that this distribution implies that there are some nodes with very high degree. Indeed, the degree is so high, that a small fraction of nodes have enough degree that these nodes could be connected to even node in the graph. It is further shown that these high degree nodes are connected. The result is that these high degree nodes form a highly connect core. This core appears if the degree distribution is heavy tailed and the nodes are connected without any special preference. However, it is shown that graphs such as the IP router graph have connection special preference. Finally, it is shown that this preference does not matter as far as worm propagation goes. As a result, this work shows that the degree distribution plays a critical role in worm propagation while other details of the graph play a limited role.

1 Introduction

In recent years email worms have causes considerable damage to the Internet. For example, the damage caused by the "I love you" worm was estimated to be over 3 billion dollars [Jones, 2001]. While email systems are being improved, email worms provide a useful means to spread worms and will likely continue to do so in the future.

This paper examines the spread of email worms by employing random graph theory as developed in [Erdos and Renyi, 1959]. The principle result is that the degree distribution plays a key role in the propagation of the worm. Specifically, it will be shown that if the degree distribution has a type of heavy tail[1], then there is a connected subgraph we call the *core*. The nodes within the core are relatively small in number, but they have a very high degree and are capable of connected to the entire graph. Thus, the worm can spread very efficiently by infecting the core. This paper also explores the correlation within observed graphs. It is found that there is strong correlation and dependence. As a result, the core does not seem to appear exactly as might be expected. However, further investigation reveals that while the dependence of the how nodes are actually link together effects the core, it does not effect the spread of the worm. It is conjectures that the reason for this behavior is related to the hop count metric used. Specifically, we introduce another metric for the distance between nodes, the worm distance, that allows nodes that are not directly connected, but have many alternative paths between them, to have a small distance between them. We begin with a short investigation of how the graph effects the spread of email worms.

Consider the “I Love You” worm as a classical example of an email worm. This worm was transmitted via email with an attached visual basic script. When the attachment was opened, the script was executed and a copy of the worm was sent to all entries in the MS Outlook address

[1] Since the considered graphs are finite, the moments of the degree distribution are finite. This work requires that some moments grow without bound as the size of the graph increases.

book. This worm also spread other ways [CERT Advisory, 2000], however, the principle method was from email and this is the method of interest here.

The spread of email worms depends on the graph that is defined by address books. By this we mean that each end-user is a node and the adjacencies between nodes is defined by address books. Note that the address book does not necessarily include the addresses in the MS Outlook address book, but could also include the sender's addresses of all received emails. We assume that the address of the sender of receiver emails is placed into the address book, and that the sender also records the destination address of each sent email. Thus, we can assume that the email graph is undirected.

In order to understand the spread of email we develop a simplified model. Specifically, we assume that when a host is infected, it sends an email to all neighbors (as defined by the address book). these neighbors then open the infectious emails after waiting an time period that is exponentially distributed with mean λ. Let M_t be the number of infectious emails at time t. That is, M_t is the number of emails that will infect a host. Let I_t be the number of host that is infected at time t. Then, the dynamics of the worm are

$$\begin{aligned} dM_t &= -dI_t + R\left(M_t + I_t\right) dI_t \\ I_t &\sim \text{Cox Process with rate } \lambda M_t \end{aligned}$$

where $R\left(M_t + I_t\right)$ is the rate of growth. To understand this model, examine the first expression. This defines how the number of infectious emails is incremented. First, when a host is infected, it means that one less infectious email is waiting to be opened, hence the $-dI_t$ term. However, when a new host is infected, it sends emails to its neighbors as defined by the address book. Some of these neighbors will already be infected. Furthermore, some will not be infected yet, but will have infectious emails waiting to be opened. In either case, the email sent by the newly infected host does not play a role in spreading the worm. Therefore, we set the number of hosts that is infected by the newly infected host as a random number $R\left(M_t + I_t\right)$. Note that the dependence on the number of infected host and infectious emails is made explicit. The second part of the model is I_t. Since the emails are opened after an exponential waiting period λ, the rate that new infections occur is λM_t.

It is clear from the above discussion that the rate of growth of the worm depends on the email address book graph over which the worm spread. This dependence is captured in $R\left(M_t + I_t\right)$. Thus, we investigate $R\left(M_t + I_t\right)$ for several types of graphs. Figure 1 shows the value of R as a function of the fraction of infected hosts. This was found through simulation. The graph is a tree with each node having degree 4.

Figure 2 shows the rate of growth over a random graph. Random graphs have been extensively investigated [ref]. A random graph denoted by $G^{n,p}$, is a graph with n nodes and each node is connected with probability p. In the case shown here, the average degree was 4. Note that this is the similar to case of the tree above. However, R is quite different. To some extent, this figure makes sense. For example, when half of the nodes are already infected, a newly infected node picks, on average, at random 3 nodes to send emails to. Note that the node that sent it an email is already infected. However, of the three other neighbors, on average half of them are infected. More generally, if the fraction of infected nodes is x, then the average number of nodes infected by a node if $x \times 3$.

Figure 3 shows the growth rate when the graph is the same as the graph of IP routers. Specifically, each router is a node and the adjacencies are defined by the links between routers. This data is from the data collected by the skitter project. This plot has a peculiar shape in that at first, it is very large, but rapidly decreases and then slowly tapers.

It this type of plot that we are most concerned with. The reason, as we shall see, for this shape is directly related to the graph. In particular, it is well known that the IP graph has heavy-tailed degree distribution. That is, if we model the degree as a random variable, then the probability of high degree nodes decays like

$$P\left(\text{degree} = d\right) \sim \frac{1}{d^{\alpha}}.$$

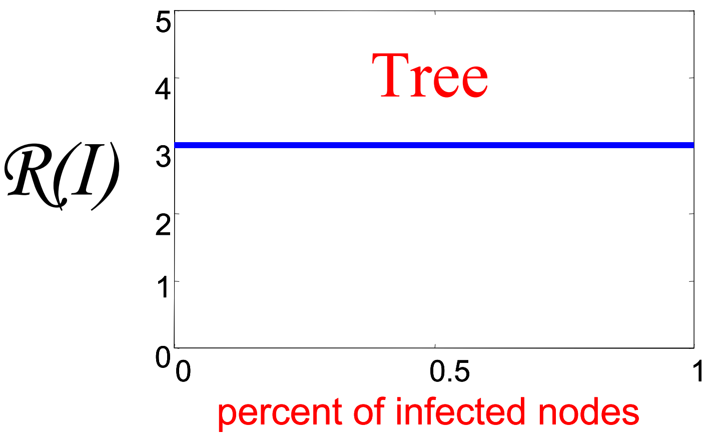

Figure 1: The Rate of Growth on a Tree

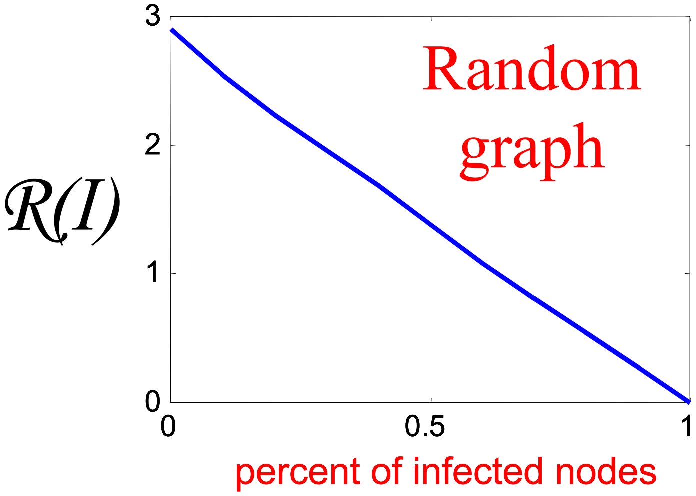

Figure 2: The Rate of Growth on a Random Graph

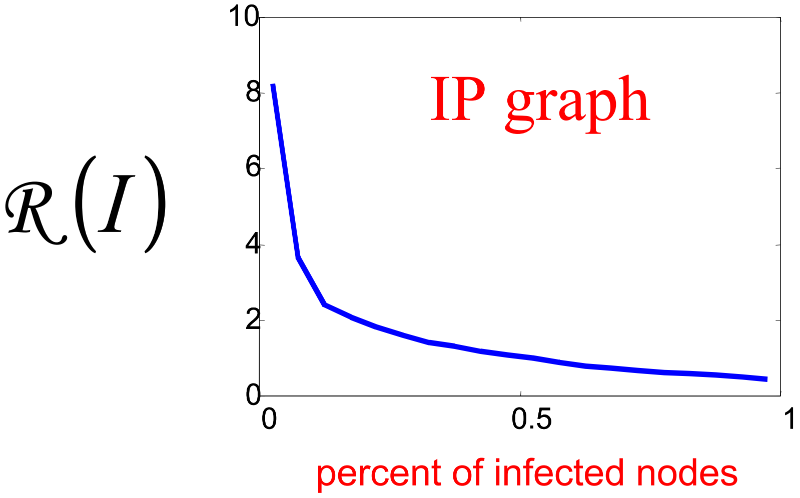

Figure 3: The Rate of Growth on a Random Graph

We claim that this behave leads to the shape shown in Figure 3. One way to understand this behavior is by examining Figure 4. This figure shows a highly connected core which makes up about 10% of the graph. The rest of the graph is a weakly connected. From this schematic, Figure 3 is more understandable. Suppose that a first a node on one of the blue tendrils is infected. This node is in an area of the graph that is weakly connected, and hence, the worm spreads slowly here. eventually, the worm spreads to the highly connected core. at this moment, the worm had infected very little of the graph. Hence, there is nearly 0% of the graph infected when the first node of the highly connected core is infected. However, once the first node in the core is infected, the worm growth becomes explosive because the core is so highly connected. This growth accounts for the high value of R when only a small fraction of the graph is infected. Once the highly connected core if infected, the worm spreads more slowly along the tendrils. This accounts for the slow tapering part of the plot of R.

While the schematic in Figure 4 helps understand the plot in figure 3, it does not provide much information about the spread of an email worm over the email graph. We claim that the above if typical for graphs with heavy tail degree distribution. Furthermore, we claim that it is likely that the email graph does have heavy tailed degree distribution. Thus there are two issues; the degree distribution of the email graph, and the effect of the degree distribution.

2 Graphs with heavy tailed degree distribution

The occurrence of heavy tailed degree distributions has been extensively documented since Hurst first observed that water levels of the Nile river had long range dependencies. We say that a distribution has heavy tail if $P(X = k) \sim \frac{1}{k^{\alpha}}$ for large k. In this case, it is easy to shown that the moments above α do not exits. A case of interest is when $1 < \alpha < 2$. In this case, the mean is finite, but the variance and all high moments are infinite. This infinity variance leads to a high variable. In particular, such distributions allow occurrence of very large events. In terms of degree, this would mean that there are some node that have very large degree. In terms of email address books, this means that some people have very large email address books.

Unfortunately, it is not possible to measure the email address book graph because of privacy reasons. However, many other graphs have been observed and these give some indication of the email address book graph. For example, Figure 5 shows the decay of the degree distribution given by the World-Wide-Web graph from [Broder *et al.*, 2000]. The World-Wide-Web graph is

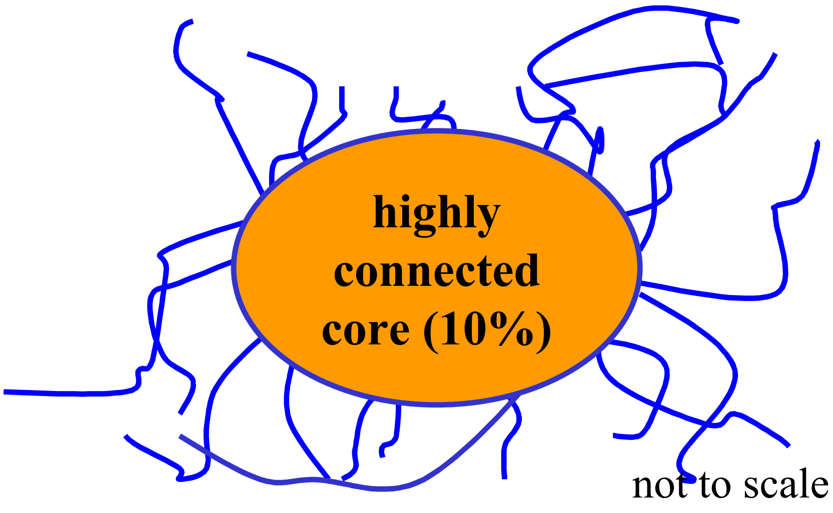

Figure 4: Schematic of Graphs with Heavy Tailed Degree Distribution.

the graph defined by Web pages; each web page is a node and the adjacencies are defined by hot links. In this case, it was found that the α is roughly 1.09, the highly variable case.

Figure 6 shows the degree distribution of the graph defined by actor collaboration from [Amaral *et al.*, 2000]. In this graph, each actor is a node and the adjacencies are defined by whether actors have worked together. In this case, $\alpha \approx 1.3$. Note that as the number of collaborations grows, the distribution decays more rapidly than it would if it was truly a heavy tailed distribution. In the work that follows, this rapid decay does not matter. Figure 7 shows the decay of the distribution of graph defined by sexual partnerships from [Liljeros *et al.*, 2001]. Here a node is a person and adjacencies exist if two people were sexual partners. In this case, when the number of sexual partners over the persons life is considered (the figure on the right), $\alpha = 1.6$. On the other hand, if only the number of sexual partners over the last 6 months is considered, $\alpha = 2.31$.

Figure 8 shows the degree distribution of the IP router graph from the SCAN dataset [SCAN,]. Here we find that $\alpha = 1.36$. Note, again, that for large degrees, the distribution decays more rapidly then a strict heavy tailed distribution would. One reason for this is that the graph is finite. Hence, it is not possible for a node to have too high of a degree. However, even taking this into consideration, the degree decays rather rapidly for large degrees. But, as noted before, the fact that the degree decays in a way that is similar to the heavy tailed degree distribution for smaller degrees is sufficient for the work here.

While we do not have access to the email graph, the above provides indication that this graph may be heavy tailed. In particular, the actor collaboration graph and the sexual partners graphs are, in some ways similar to the email graph as the adjacencies in the sexual partner graph is defined by social relationships of choice while the actor graph is from relationships from work. The email graph would includes these two type of relationships.

There has been some attempts to observe the email address book graph [Newman *et al.*, 2002]. For example, observations of email destinations have been observed. This data indicates that the degree distribution is not heavy tailed. However, it is too early to draw strong conclusions from such data. The problem is that these observations are only of recent email exchanges and, hence, only observe often occurring email exchanges. In figure 7, the left plot show the effect of only consider recent interactions. For example, the web graph was considered, but if only links with a particular popularity were included, then it is not clear if the graph would continue to be heavy tailed. However, there is little doubt that the web graph does indeed have heavy tailed degree distribution. While this early measurement work cannot be discounted, the more complete observations above provide strong evidence in favor of heavy tailed degree distribution for the

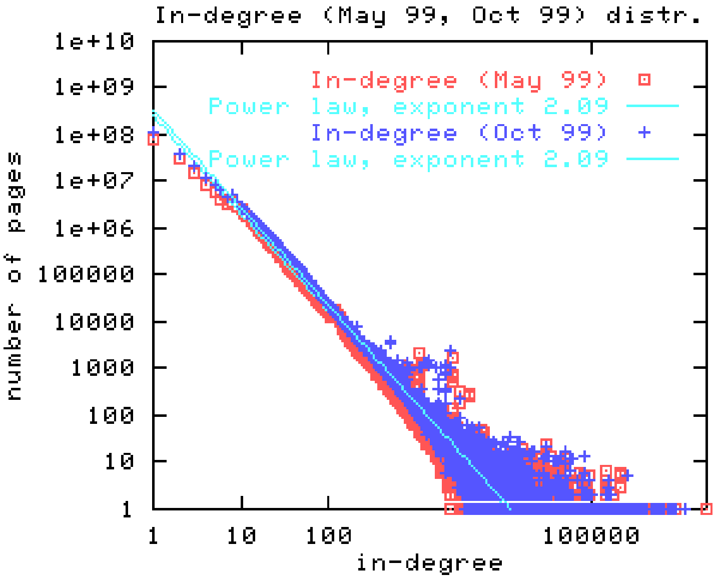

Figure 5: Degree Distribution of World-Wide-Web Page Graph. [Broder *et al.*, 2000]

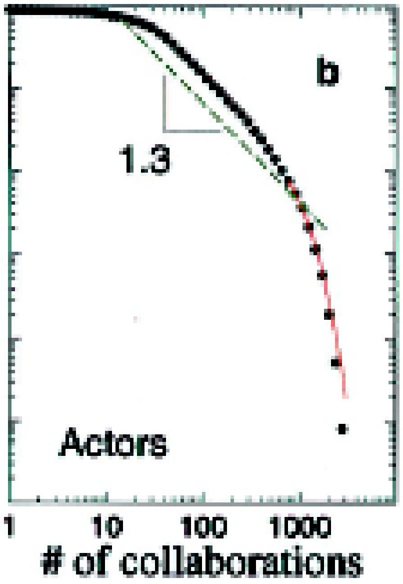

Figure 6: Degree Distribution of Actor Collaborations. [Amaral *et al.*, 2000]

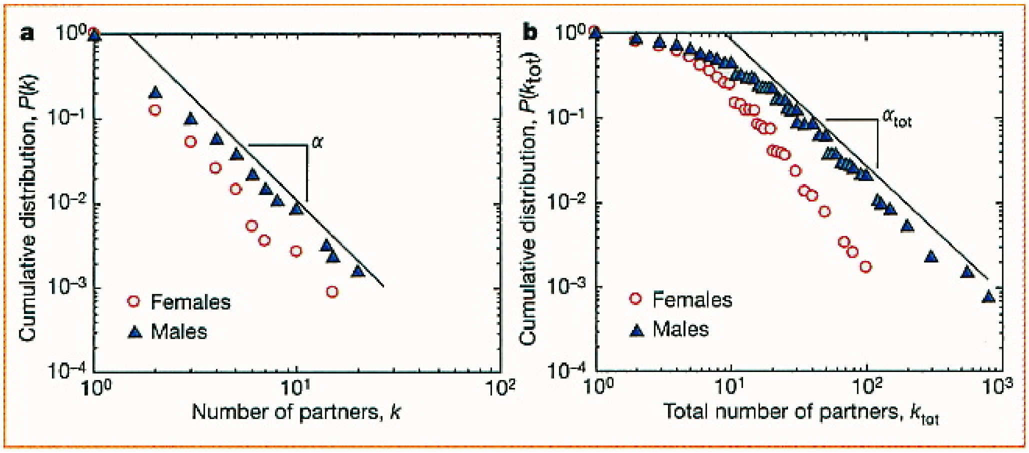

Figure 7: Distribution of the Number of Sexual Partners [Liljeros *et al.*, 2001]. The figure on the left is of the numbe rof partners in the last 6 months, whereas the one on the right is the total number of sexual partners.

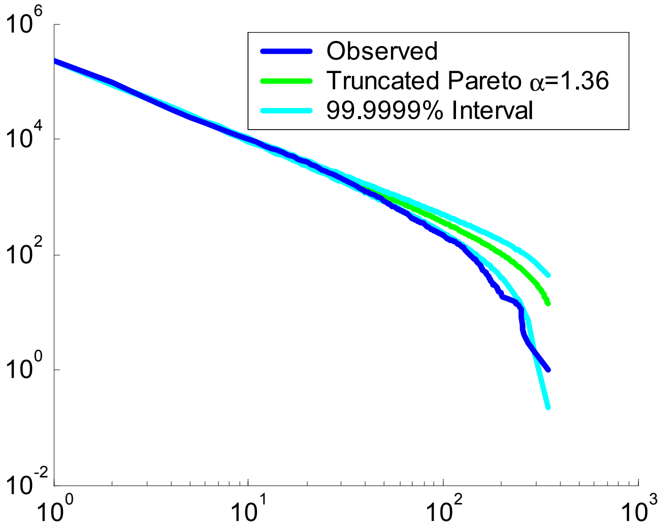

Figure 8: Degree Distribution of the IP Router Graph.

email graph.

While the above indicates that the email graph may be heavy tailed, we should be more precise. Specifically, since the email graph is finite, it is not possible for any of the moments to be infinite. Thus, strictly speaking, the degree distribution cannot be heavy tailed. However, in this paper, the focus is on the behavior of graphs as the number of nodes increases (as is done in the theory of random graphs [Erdos and Renyi, 1959], [Palmer, 1985]). Thus, a heavy tailed degree distribution is one such that some of the moments grow without bound as the number of nodes increases. The truncated Pareto is an example of such a distribution, as is discussed in the next section. The work below utilizes the truncate Pareto distribution. However, the exact behavior of the tail does not appear to be critical. Rather, what seems to be required is that the part of the complimentary distribution that follows the power-law decay grows as the number of nodes increases. That is, all that appears necessarily is that as the number of nodes increases, the linear part of log-log plots above grows.

3 The impact of heavy tailed degree distribution

While large scale measurements of the email graph have not been made, the above provides reason that the email graph does have heavy tailed degree distribution. Thus, we will make the plausible assumption that the degree is heavy tailed distributed. With this assumption, we will investigate the existence of a highly connected core as indicated by the schematic 4. The sexually transmitted disease epidiology research community has also recognized the existence of such a core. In that field it is known as the "core group" or "high frequency transmitters." The epidiology research community has learned that these individuals are responsible for the spread of the disease and hence these are the ones that are must be attended to in order to reduce the propagation of the disease [Yorke *et al.*, 1978]. Thus, in terms of epidiology, the highly connected core is the same as the core group. We will see that the heavy tailed degree distribution results in such a core group.

For concreteness, we will assume that the degree distribution is distributed according to a truncated Pareto distribution, i.e., the probability that the degree is less than d is

$$P\left(D<d\right)=\frac{1-\left(\frac{1}{d}\right)^{\alpha}}{1-\left(\frac{1}{N+1}\right)^{\alpha}},$$

where N is the total number of nodes in the graph.

Let d_1 denote the degree of the node with highest degree.

Theorem 1 *Let the degree be distributed according to the truncated Pareto distribution. Then*

$$\lim_{N\to\infty}\frac{E\left(d_1\right)}{N^{1/\alpha}}\geq 1$$

Theorem 2 *With the same assumptions as above*

$$\lim_{N\to\infty}P\left(d_1\geq N^{1/\alpha}\right)=1$$

To understand these results, suppose that $\alpha=1$. In this case, the degree of the largest node is N, the number of node in the graph. Hence, this single node connects to all other nodes. It is clear that this single node makes up the core group as once it is infected, it spreads the worm to all other nodes. Indeed, because of this very high degree node, the diameter of graph is two. However, it is unlikely that $\alpha=1$. Thus we examine the case $\alpha>1$.

Let d_k be the degree of the node with the k^{th} highest degree.

Theorem 3 *For a fixed k, and with the same assumptions as above*

$$\lim_{N\to\infty}\frac{E\left(d_k\right)}{N^{1/\alpha}}\geq 1$$

and

$$\lim_{N\to\infty} P\left(d_k \geq N^{1/\alpha}\right) = 1.$$

This implies that in a large graph, there will be many nodes with very high degree. It also provides an estimate of the number of nodes required to cover the graph. We define a cover as a set of nodes such that each node in the graph is either a neighbor of a node in the set, or is in the set. From the above we can get a lower bound on the number of nodes required to cover the graph. In particular, let r be the number of nodes in the cover, then $rN^{1/\alpha} = N$ in order for each node in the graph to be covered. Thus, the number of nodes in the cover must be at least $N^{1-1/\alpha}$. If we look at the fraction of nodes in the cover, we see that this fraction is at least

$$\frac{N^{1-1/\alpha}}{N} = \frac{1}{N^{1/\alpha}}.$$

Thus, we see that the fraction of nodes in the cover goes to zero as the size of the graph grows. This is positive result because it means that a very small portion of the node in the graph must be made secure in order to affect the spread of the worm.

Thus, we see that there are nodes with very high degree. But to form a connected core, these nodes must be connected. We will now show that these high degree nodes are indeed connected. Let $n_1$denote the node with highest degree and n_k denote the node with k^{th} highest degree. Furthermore, let $n_1 \sim n_2$ denote the event that n_1 is connected to n_2. To compute the probability of these two nodes being connected note that the n_1 can either be connected to n_2 or not. The number of ways that n_1 can be connected to the graph, but not to n_2 is $\binom{N-d_1-d_2}{d_1}$ and the number of ways that n_1 can be connected with exactly one link between n_1 and n_2 is $\binom{N-d_1-d_2}{d_1-1}d_2$. Thus, the probability that the two nodes with highest degree are connected is

$$P\left(n_1 \sim n_2\right) = \frac{\binom{N-d_1-d_2}{d_1-1}d_2}{\binom{N-d_1-d_2}{d_1} + \binom{N-d_1-d_2}{d_1-1}d_2} = \frac{d_1 d_2}{N - 2d_1 - d_2 + 1 + d_1 d_2}.$$

Since $d_k \approx N^{1/\alpha}$, we find that, if $\alpha < 2$, then

$$\lim_{N\to\infty} P\left(n_1 \sim n_2\right) = 1.$$

Similarly, for k and j fixed

$$\lim_{N\to\infty} P\left(n_k \sim n_j\right) = 1.$$

We now have a clearer idea of this core group or highly connected core. It is made up of high degree nodes that are connected. This core must exist from the Pareto distribution of the degrees. In fact, it the degree are indeed Pareto distributed, then, as the graph grows in size, a cover appears which is connected and connects to each node in the graph. While this sheds light on the graph, this result requires that the nodes be connected in a uniform manner. That is, there should not be any preference given to how nodes are connected. Next we will see that graphs may exhibit strong connection preferences.

4 Node connection preference

Here we examine the IP router graph and show that there are strong preferences for how nodes are connected. We will consider many tests. First we consider the degree that nodes of degree on are connected. Figure 9 shows the frequency that nodes of degree one are connected to nodes of other degree. In total there are 132001 nodes of degree one in the IP router graph examined. In order to determine if these frequencies are to be expected, we compute the probability that the nodes of degree one are connected to m nodes of degree k. Let $R(k)$ be the number of nodes of

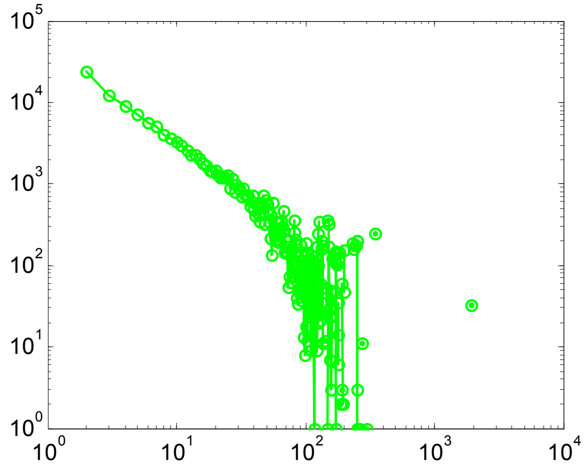

Figure 9: The frequency that nodes of degree one are connected to nodes of other degree.

degree k. Then the probability of the nodes of degree one being connected to m nodes of degree k is

$$P\left(m\right)=\frac{\binom{R(k)}{m}\binom{512953-R(k)}{132001-m}}{\binom{512953}{132001}}.$$

And the expected number of nodes of degree k that the nodes of degree one are connected to is

$$F_k=\sum_{m=1}^{\min(R(k),132001)} mP\left(m\right).$$

Figure 10 is similar to Figure 9 but also shown is F_k. In order to compare the frequency to the expected value see Figure 11. This figure is normalized so the deviation between the expected value and the observed frequency is shown. This figure also shows the plausibility interval defined as the interval such that $P\left(\left|m-F_k\right|<I\right)=0.999999$. We see that the observed frequency is far larger than what one could reasonable expect if the nodes were connected at random. Thus, we conclude that the nodes are not connected at random.

In the previous section we shown that the nodes of high degree should be connected. However, in the IP router graph, the node with highest degree was not connected to the other nodes with high degree. Figure 12 show the probability that the node with highest degree is connected to nodes of high degree. We see that this probability is very near to one, and yet the nodes were not connected. This also indicates that the nodes are connected in a way that is slightly different from what is expected.

It has been found that graphs with structure can be made into graphs without structure by applying a series of transformations [Watts, 1999]. Specifically, define a single application of the transformation as the result of taking two links and switching their destination. Figure 13 shows how this transformation works.

This type of transformation has been shown to get rid of structure in graphs. Indeed, one can think of the configuration of the graph as the state of a Markov chain and this transformation as a means to move from one state to the next. Then, it has been shown that the state of this Markov chain is more likely to take values that occur without connection preferences.

In order to gauge the effect of these transformations, we consider the growth function. The growth function $GR\left(k,A\right)$ is the total number of nodes that can be visited starting at node A and

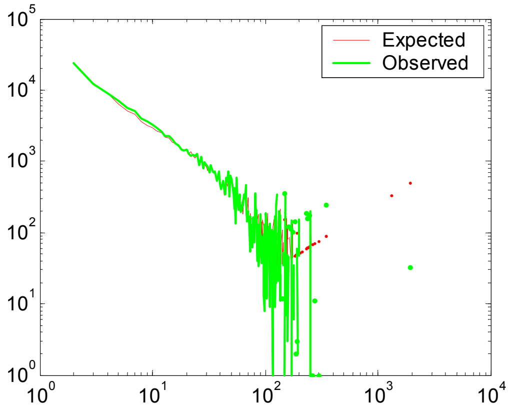

Figure 10: The frequency that nodes of degree one are connected to nodes of degree k as a fucntion of k. Also, in red, is the expected number of nodes of degree k that nodes of degree one are connected to.

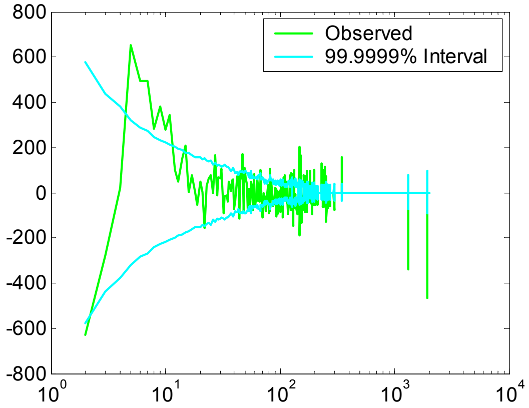

Figure 11: The deviation of the frequency of the nodes with degree k connected to nodes of degree one. Also shown is the plausibility interval.

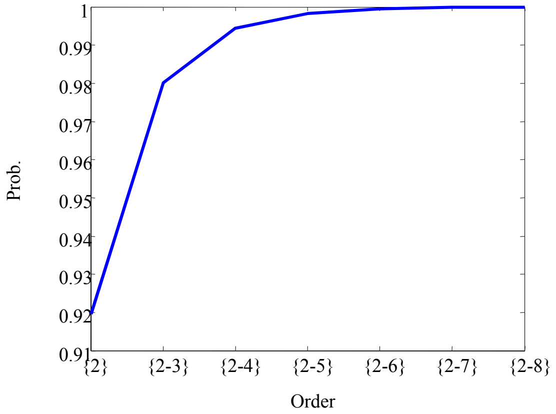

Figure 12: Probability that the node with highest degree is connected to nodes next highest degree. Specifically, the probability denoted by $\{2\}$ is the probability that the node with highest degree is connected to the node with second highest degree. Similarly, the probability denoted by $\{2, 3\}$ is the probability that the node with highest degree is connected to a node of second or third highest degree. In the IP router graph, none of these events occured.

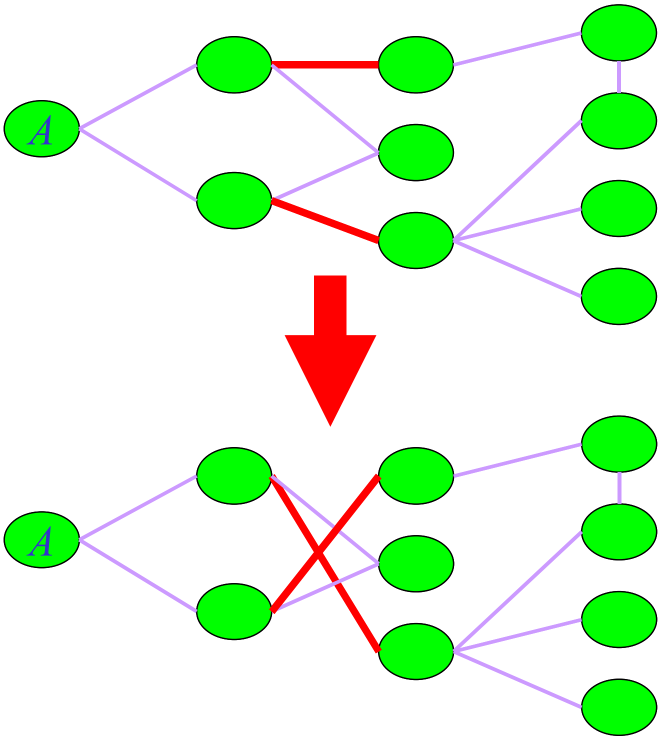

Figure 13: A transformation of the graph is obtained by selecting to links are random and switching their destination.

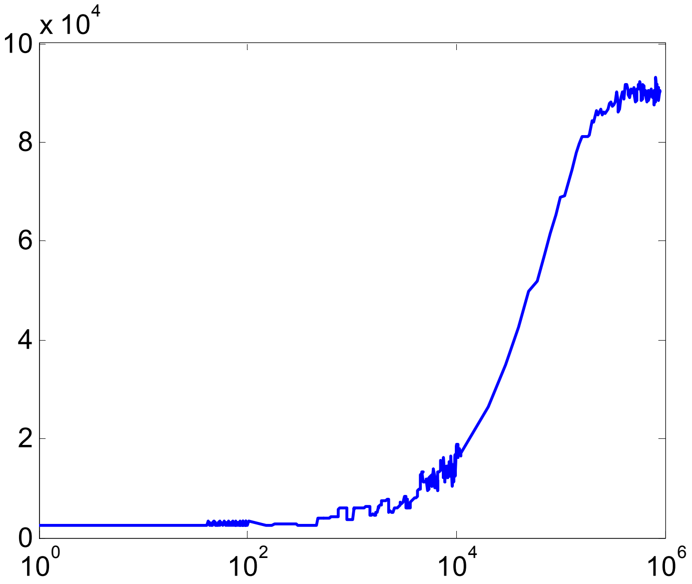

Figure 14: Maximum growth rate as a function of the number of transformations.

only jumping k hops. The maximum growth function is $MAXGR(k) = \max_A GR(k, A)$. Figure 14 shows how the maximum growth rate varies as more transformations are applied. Since it is known that the transformations get rid of structure, we see that there was extensive structure as the maximum growth rate increased by nearly two orders of magnitude.

In order to show that these transformation did indeed get rid of the structure of the graph that occurred because of the node connection preferences, we consider Figure 15. This figure is the same as 11 but when applied to the graph after 10^6 transformations. Also, a much smaller plausibility interval is shown. It is clear that now the nodes of degree one are connected in a fashion that is more what is expected when nodes are connected without preference. Furthermore, in this case, the node with highest degree is connected to the other nodes with high degree. Indeed Figure 16 shows the subgraph of the 8 highest degree nodes. This subgraph is well connected.

5 Implications of node connection preferences in worm propagation

The previous section showed that graph with heavy tailed degree distribution might not behave exactly as expected. This is indeed the case when the IP router graph is considered. It is not clear if it is the case when email graphs are considered. One difference is that network engineering select the degree and the connectivity of many nodes. Since the design is done with a large part of the network in mind, the way in which the nodes are connected depends on the way that other nodes are connected. In the email graph, there are no designers that try to design the graph to meet some performance objectives. However, other processes may be at work.

While the work in Section 3 shows that the heavy tailed degree distribution has a big impact on the spread of worms, the previous section calls into doubt these results since graphs may have node preference. To investigate this effect we ran many simulations of the worm spreading over a graph and computed the average value of R as a function of both the number of infected nodes and the number of transformations as described in the previous section. Figure 17 shows the result of these simulations. We see that the growth rate of the worm does not depend on the

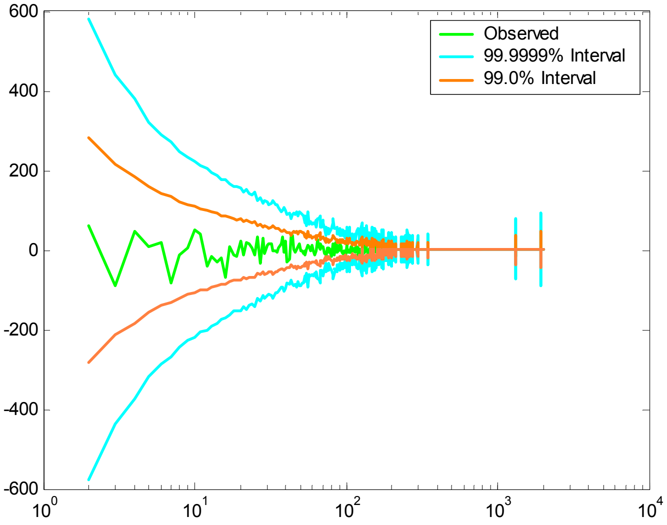

Figure 15: The frequency of nodes of degree k connected to nodes of degree one, after 10^6 transformations.

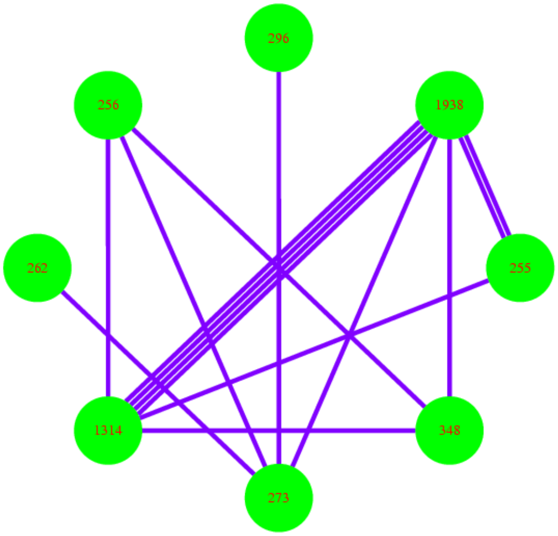

Figure 16: The subgraph of highest degree nodes.

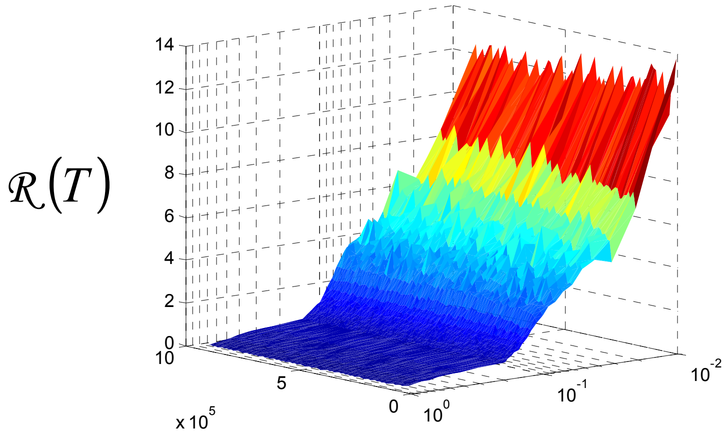

Figure 17: The worm growth rate as a function of the number of transformations, whcih goes from 0 to 10^6 and a fucntion of the fraction of the graph that is infected, which goes from 0 to 1.

transformations. This is quite surprising since the maximum growth rate of the graph, as defined in the pervious section changes when transformations are performed. This result is quite good because it shows that while the graph might be slightly abnormal in the sense that the nodes are connected with preference, the results form Section 3 still seem to hold.

To understand how the spread of the worm could be not effected by changes in the maximum growth rate, we think about how the worm spreads. Consider Figure 18. Here node A and B are not directly connected. However, they are of high degree and there are many alternative paths between them. as a result, the worm will spread form A to B as if they were directly connected. to see this, consider a worm infecting node A. This worm is then sent to all of A's neighbors. Since there are so many neighbors of A, one of these neighbors opens the email very shortly after node A sends it. This node becomes infected and then sends an infectious email to B. Thus, an infectious email is sent to B very shortly after node A is infected. Thus, because of the abundance of alternative paths, the fact that node A and B are not directly connected does not effect how a worm spreads. Metrics such as max growth rate cannot account for the abundance of alternative paths.

In order to quantify the effect of alternative paths, we define the worm distance between nodes A and B as the average time it takes a worm to spread from node A to node B. It turns out that while this distance makes good sense, it is very difficult to compute. Nonetheless, after significant computations, the mean distance between the nodes with highest degree was computed as a function of the number of graph transformations. Figure 19 shows how the worm distance varies with the number of transformations. Also shown is how the average distance between host of high degree changes when the more traditional hop count metric is used. We see that the worm distance is much less effected by the transformations as compared to the hop count distance. However, the worm distance is still effected by the transformations. Further computations are required to determine if some other related metric yields no changes was the number of transformations varies.

6 Conclusions

Section 1 showed how the email address book graph plays a critical role in the propagation of worms. The next four sections try to understand what this role is. In Section 2 provided plausible evidence that the email graph has a degree distribution with a heavy tail. Section 3 showed that the heavy tailed degree distributions can lead to a particular behavior of how the worm spreads.

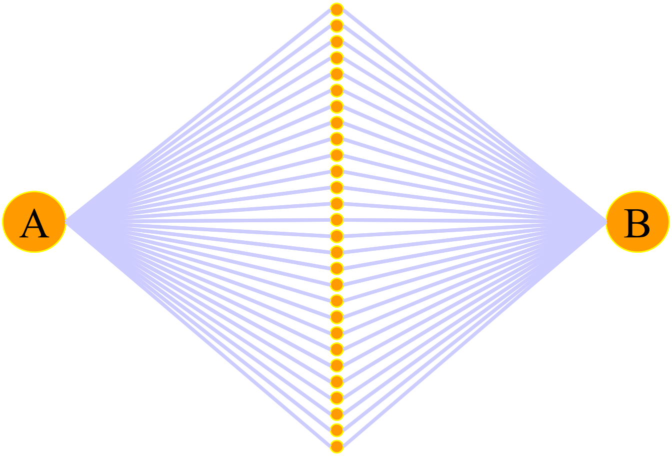

Figure 18: If a worm starts at node A, then it travels to node B in nearly the time it takes to travel one link.

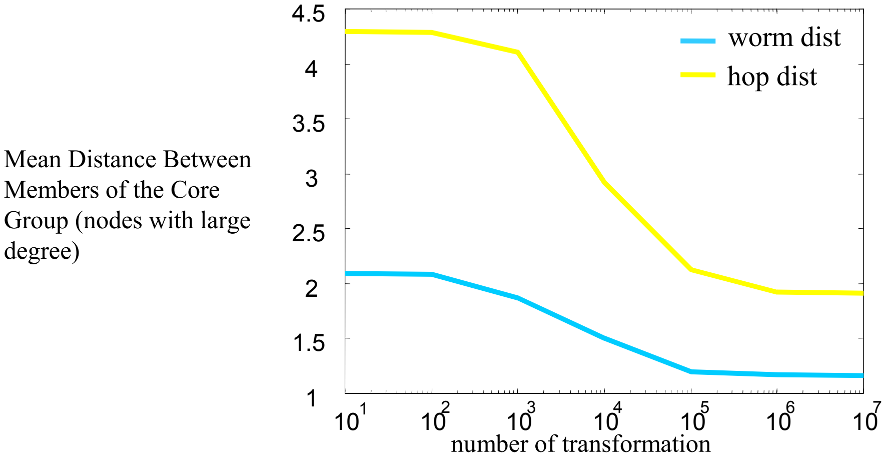

Figure 19: The average distance between node with highest degree. Both hop count and worm distance metrics are shown.

As is discussed, the heavy tailed degree distribution implies that there are some nodes with very high degree. These nodes are of such high degree that only a small fraction of these nodes is required to provide enough links to connect to the entire graph. It was then shown that these nodes should be connected. Thus, the graph has a set of nodes with very high degree that are connected. As a result, these node with high degree form a connected subgraph, this subgraph is responsible for spreading the worm. Worm defense systems should focus on this subgraph. Section 4 showed that not all graphs exactly obey the guidelines set in Section 3. Indeed, the results in Section 3 imply the presents of nodes with large degree, but these nodes are only connected if node are connected without special preference. Section 4 showed that in the IP graph nodes are connected with preference. On the other hand, Section 5 showed that this preference does not play much of a role. Therefore, we make the following conclusion: Assuming that the email address graph is heavy tailed, a reasonable assumption, then a worm on this graph will spread much like a worm spreads on any graph with heavy tailed degree distribution. In particular, a simulation of a worm spreading over a graph of IP routers, World-Wide-Web graph, or a artificially constructed graph are useful ways to study the propagation of email worms.

References

[Amaral *et al.*, 2000] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97, 2000.

[Broder *et al.*, 2000] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. WWW9. *Computer Networks*, 33:309–320, 2000.

[CERT Advisory, 2000] CERT Advisory. Love letter worm (CA-2000-04), 2000.

[Erdos and Renyi, 1959] P. Erdos and A. Renyi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[Jones, 2001] Paul Jones. Copycats of i love you worm spread. *WinPlanet*, 2001.

[Liljeros *et al.*, 2001] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y Berg. The web of human sexual contacts. *Nature*, 411:907–908, 2001.

[Newman *et al.*, 2002] M. E. J. Newman, Stephanie Forrest, and Justin Balthrop. Email networks and the spread of computer viruses. *Phys. Rev. E*, 66, 2002.

[Palmer, 1985] E. M. Palmer. *Graphical Evolution.* 1985.

[SCAN,] SCAN. Available at http://www.isi.edu/scan/mercator/scan.gz.

[Watts, 1999] D. J. Watts. *Small-Worlds.* Princeton University Press, 1999.

[Yorke *et al.*, 1978] J. A. Yorke, H. W. Hethcote, and Anold. Dynamics and control of the transmission of gororrhea. *sexually transmitted diseases*, 5:51–56, 1978.