Network analysis for identifying anomalous patterns of message distribution on Usenet News

Dr. Harry Rolf
harry.rolf@anu.edu.au

This article is adapted from my 2008 honours thesis (unpublished) published online May 2022

This article presents the results of a study that investigated the use of visual analytics for aiding in the detection of a phenomenon called path preload, an anomalous pattern that can appear in message distribution on the Usenet News network. These patterns are thought to be associated with user behaviour such as file sharing or the posting of spam to news server but are notoriously hard to detect using traditional methods of content analysis and manual assessment often relied upon by administrators to enforce news server policy and moderate content.

To overcome this challenge a visual analytic method was designed and used to examine a sample of 82,986 unique messages collected from ninety-six Usenet news groups. Message path information was used to create network visualisations that could be inspected for evidence of path preload and its potential effect on message distribution. Results show that use of the visual analytic method can enable the successful identification of path preload and provide insights into its effect on the network that could aid in the development of techniques for automated detection and prediction of path preload.

Introduction

The Usenet News Network or Usenet (for short) is a publicly accessible distributed (but decentralized) bulletin board system, which predates the internet and email communication. Today despite decreasing user numbers and support from internet service providers Usenet still supports a large body of users scattered around the world who post messages to thousands of bulletin boards. Usenet faces the same challenges experienced by Email, Facebook, Twitter, and other popular platforms for online communication. It is subject to spam and misinformation and has also become a haven for file sharing (offering a viable alternative to P2P networks). This is due to the distributed nature of communication on the network and the high level of anonymity afforded by its message format.

Insights gained from the study of Usenet continue to be relevant to many of the challenges faced by socio-technical communication networks today. Studies of Usenet's dynamics can contribute to our understand of how a network's structure and function affects the flow of information and communication among its users, and how those users also shape the social-technical network.

Understanding user message posting behaviour has the potential to reveal patterns in communication which could identify the sources of spam and other content being distributed on the network. A phenomenon known as path-preload, the focus of this paper, provides an example of the effect user posting behaviour can have in message distribution and network dynamics. The phenomenon is commonly associated with the distribution of spam and other content at risk of being filtered out by protected servers on the network. Path-preload involves faking the path that a message has travelled to reach its current news sever by altering a message's header information. This helps prevent it from being sent to 'protected' servers on the network, by indicating it has already visited them.

There are many reasons behind the occurrence of path-preload on Usenet. It may be caused by users posting anti-social content, or users who load specific news servers into their messages paths to give them high ratings with anti-spam software, or so that their messages will avoid those news servers on the network which may have aggressive spam detection systems in place. Users may also cause path preload unintentionally through the software used to post messages, or deliberately for testing purposes. Path-preload may also be caused by news servers which are incorrectly configured or setup to function differently to accepted standards.

Despite the many reasons for path preload to exist, documented occurrences are rare, possibly because it is notoriously hard to detect, unless the message is picked up by the news server when it is posted. Unfortunately to make matters worse any knowledgeable user may setup their own news server which could easily distribute such messages unquestioned with a few subscribers signed up.

To identify a fake path requires the comparison of that path with the paths from a significant number of other messages. This would force the anomalous message path to standout against the regular paths of other messages arriving at a news server. However, if there is no way of knowing which paths are fake to begin with how can we go about such an analysis?

Study aims

This study aimed to investigate anomalous patterns of message distribution on Usenet, specifically to look for instances of path-preload as it may be associated with the distribution of content such as spam. The study also aimed to evaluate the usefulness of visualisation techniques for the analysis of Usenet data and the detection of patterns in message distribution. The study was guided by the following research questions:

1. Are anomalous patters of message distribution detectable in network visualisations of message distribution?
2. Can instances of an anomalous behaviour known as path-preload be detected in Usenet message distribution data?

The Usenet News network

The Usenet News network is a globally distributed news network built on the Network News Transfer Protocol (NNTP) which provides public access to a bulletin board system where anyone accessing the network can post or reply to messages. Usenet news essentially functions as a publicly accessible social network that allowed people to easily view and reply to messages from discussions on their local news server. Usenet was one of the earliest online systems, created in the mid-1980s, since then Usenet's infrastructure and use has changed significantly, but its core concept of messages being organized into groups and distributed between servers has remained unchanged. Generally, Usenet can be thought of as a broadly decentralized bulletin board system where users are able to read and reply to groups of existing messages that have arrived at their local server (Fisher, 2003).

However, the networks popularity as a social space has declined significantly due a range of factors including the prevalence of spam and the popularity of other new social media services such as Twitter and Facebook. This has led to the closure of many significant news servers. The network has now become a haven for file sharing with binary file distribution. This transformation maybe attributed to the networks unique structure which is well suited to the sharing of information – the NNTP protocol creates a decentralized network, with a high level of anonymity that is only regulated by server administrators (Fellows, 2006; Kim, Schneider, Ager, & Feldmann, 2010; Turner, Smith, Fisher, & Welser, 2005).

Applying a network approach to Usenet News

In 2004 Borgs et al. (2004) explored the community structure of Usenet newsgroups using cross post graphs containing information on instances when messages were posted to two or more newsgroups simultaneously. This work provided evidence that a close relationship exists between the content of a news group and the messages being posted. Borges et al. also observed the appearance of small-world properties in these cross-post graphs. Inspired by this observation Nakano, Y. et al. (2007) collected the path header information of messages gathered from an extensive number of publicly accessible Usenet news servers to model the topological structure of Usenet. Their analysis explored the topological properties of the network, such as degree distribution, correlations, average shortest path distance, clustering coefficient and the community structure of the network. These properties were then compared with other theoretical network properties found in similar networks such as the World Wide Web and E-mail.

From these results Nakano Y. et al. (2007) conclude that the structure of Usenet exhibits strong scale free and small-world properties. They also found that in the network the largest community is formed by only a small number of highly contributive news servers, and that other servers tended to be characterized by only their geographic location. The small-world properties possessed by Usenet suggest that analysis of its structure could provide valuable insights into complex network structure and function, as the vast amount of message data and other information are publicly accessible and relatively easy to acquire.

The network

Usenet is a logical network, originally resting on top of many different physical networks today the internet is responsible for supporting Usenet. Two aspects that are important for understanding the network's function are its topological structure and the distribution algorithm guiding message propagation through the network.

Topological structure

RFC 1036 for the Standard Interchange of USENET Messages (1983) describes the Usenet network topology as resembling a directed graph, where each node in the graph represents a news server and edges are transmission paths between two news servers. These edges or links between news servers are bidirectional with messages from news groups able to be exchanged backwards and forwards between news groups. Usenet is also made up of sub-networks or hierarchies such as comp, sci, aus, and alt. The topological structure of these sub-networks function as connected graphs, where each node is connected to every other node in the sub-network, theoretically making the entire network connected (Horton, 1983).

Message distribution

Messages posted to news servers on the network propagate indirectly from one news server to the next, which is unlike other internet services such as e-mail networks where articles take the shortest path to their destination. On Usenet news servers must subscribe to other news servers (known as feeds) to receive and send messages. When a news server receives and accepts a message from another server that messages will then be forwarded to a list of subscribing news servers. This process will continue until the article reaches all news servers on the network or has propagated as far as possible (Lueg, 2003).

A critical aspect of this propagation algorithm is the prevention of loops within the network. In the above example it is possible for a host to send an article to its neighbour and in turn have the neighbour send it back to the host server (since the network is bidirectional) and for this process to continue infinitely. The problem is solved by news servers keeping records of all messages they have seen and discarding those that they have already seen. This process is further optimized by including path information in each

message. These path strings provide a record of Usenet's topological map, news servers rely on these message paths to discover new servers on the network (Horton, 1983).

Message path strings

A message posted to a news server on Usenet is required to contain a path field which lists the servers that a message has traversed to reach its current destination. Entries are added by a news server as it is forwarded on to another subscribing news server. A message's path is used to prevent news servers from sending a message to neighbours who have already seen it, lessening unnecessary resource usage and traffic between news servers and to establish a path for reaching new news servers (Horton, 1983).

Example of message path information: *news.edisontel.com! nntp.eutelia.it! news.newsland.it! news.ecp.fr!club-internet.fr! feedmesmall.clubint.net! zen.net.uk! demorgan.zen.co.uk! nx01.iad01.newshosting.com! newshosting.com! 216.196.98.140.MISMATCH! border1.nntp.dca.giganews.com! nntp.giganews.com! local02.nntp.dca.giganews.com! nntp.comcast.com! news.comcast.com.POSTED! not-for-mail*

Normally the right most entry in the path will be an entry from the news server that the message originates from, it is also permitted that an entry be added before this, which specifies the name of the sender (used for compatibility reasons).

Several different entries may appear in a message's path:

- Aliases: These entries usually appear as the abbreviated names of news server addresses. Aliases can also be added to represent the sender of a message at the beginning of a path.
- Real addresses: The actual physical address of a news server, which can be reached on the network.
- Feeders: These entries are news servers which form part of a larger internal organization or group of news servers for a single domain. Feeding servers tend to be dedicated to handling either outbound or inbound messages, their purpose is usual indicated in the server's path entry.

Visual analytic process

To address the research questions the study used a visual analytic method that supported the exploration of a dataset containing messages gathered from Usenet. The process was based on an iterative visual analysis process described by Keim, A.D. et al. (2006), that involves the steps of data collection, pre-processing, representation, interaction and decision making.

Data collection

Usenet messages were collected in 2008 from a list of seven publicly accessible news servers (Table 1). News servers were selected based on three criteria: The number of supported news groups, length of message retention, and the amount of bandwidth they provided. News servers were selected using lists of active servers obtained from Newz Bot (1995) and Free and Public Usenet News Servers (1999).

A list of active news groups was compiled using lists provided by the Big 8 Management Board (2007) and Newz Bot (1995) which both kept regularly updated lists of active news groups. A random sample of groups were then selected from the alt and big8 news hierarchies. The alt and big8 hierarchies were chosen to firstly limit the total size of the dataset, and at the same time to provide a wide sample of the Usenet News network. According to Nakano, Y. et al. (2007) these two hierarchies had much wider international distribution compared with the many hundreds of other news hierarchies, of which an extensive list of hierarchies is kept by Lewis, S.E. (2004).

A sample of ninety-six news groups was created for message collection (See Appendix A). Messages were then collected from the selected news groups over a period of approximately four months from January to April 2008. This produced a data set containing a total of 82,986 unique messages, posted by a total of 11,864 users over the four-month collection period, containing 51,933 unique message paths.

News Server Name	News Groups	Speed	Retention
aioe.org	53401	20.30 k/sec	Low
aioe.cjb.net	51007	69.37 k/sec	Low
textnews.news.cambrium.nl	174080	105.17 k/sec	Medium
news.edisontel.com	36849	94.53 k/sec	High
free-text.usenetserver.com	68588	208.66k/sec	High
freenews.netfront.net	53667	63.43 k/sec	High
allnews.readfreenews.net	45358	78.57 k/sec	Medium

Table 1: The seven news servers used for data collection and related information.

Pre-processing

According to analysis conducted by Nakano, Y. et al. (2007) there are several features inherent in a message path which should be cleaned-up before paths are used to form a network graph. Content analysis was run to clean message path data and to generate a list of message common paths for analysis.

Common paths

According to the Usenet News distribution algorithm, a message is initially posted to a single news server, from which it should propagate to every other server on the network. As a message propagates the path field should contain a record of all the servers that it has visited. If the path information from every copy of a single message on the Usenet News network (across multiple servers) is compared, only the first entry should be identical across all paths, since the only news server that every copy of a message should have in common is the news server it was originally posted to, this can be referred to as a message's common path.

The presence of long or unusual common paths is a potential indication of path-preload. When a message's path is preloaded all copies of that message will include the preloaded path which will be different to the paths of other messages arriving at the same server. In some cases, a common path may contain more than one entry because a message may travel through a series of consecutive news servers to begin with, or if not enough copies of a single message are collected and compared. The common path of a message can be identified by collecting and comparing path information from multiple copies of that message, each copy coming from a different news server.

Representation

Once a list of common paths was generated, news groups were selected for visual analysis. The following criteria were used to choose news groups for analysis with common paths that may be related to instances of path preload:

- News groups with comparatively high numbers of common paths found to contain more than one real or alias news server entry in succession starting from the first entry.

- News groups where individual users are found to be responsible for large numbers of identical common paths compared to other users responsible for common paths.

Once news groups were selected, separate network graphs were generated to view a news group's topological structure, along with additional information to aid analysis including a list of the news group's common paths, the news group's origin servers, the number of messages, users, unique paths, and the number of messages posted by each user.

Message path information was used to build an adjacency matrix of nodes which represented news servers and edges that represented which news servers a message had been forwarded through as it propagated the network. The adjacency matrix was produced as a graph file formatted using standard Geographical Mark-up Language (GML) (Himsolt, 1996). The GUESS (Adar, 2006) and TULIP (Auber, 2004) visual exploration packages were used to visualise the adjacency matrix. In the adjacency matrixes generated nodes were weighted using a value calculated from the total number of messages which passed through that server, and edges were weighted using a value calculated from the total number of messages flowing along that link. These graphs were bi-directional as messages flow between servers in both directions, a quirk of the NNTP protocol. Adjacency matrixes were visualised for analysis using a force directed algorithm to create an abstract topological layout of nodes and edges (Withall, Phillips, & Parish, 2007).

Interaction

The software packages used to visualise adjacency matrixes also supported interaction and analysis of the graphs generated. The first step in the visual analysis of a graph involved manually mapping common paths to the network topology. This enabled the topological features of the graph to be taken into consideration when evaluating the potential of common paths to be instances of path preload, and for answering other questions related to the distribution of messages on the network.

Analysis then focused on comparing the distribution of common paths on the network. A set of criteria was used for determining whether a path (containing nodes and links) was anomalous or not:

- The number of news groups a message has been posted to. If a message is posted to numerous news groups, it can be regarded as suspicious.
- The types of news servers included in the common path. For instance, does the server originate from a server with an address or alias, and is this a common server of origin?
- Does the common path stand out against other paths? For instance, is the path it travelled regular or irregular when compared to others?

Any instances of common path identified as suspicious by these criteria were further investigated through querying associated message content stored within the dataset.

Decision making

While an instance of path preload cannot be confirmed with complete certainty, suspicious message paths identified through visual analysis can have their associated message content interrogated by an expert person to understand why an anomalous path may have occurred. If upon viewing a message's content it is determined that it exhibits features associated with the motivations behind path preload such as spam, then it can be considered likely for that message to contain preloaded path entries.

Results of analysis

A total of 15061 instances of common path were identified from message path information, which contained a total of 2228 unique common paths. Table 2 presents the top 8 most frequent common paths in the dataset. The total number of common paths detected (where at least two instances of a message must be compared) is significantly less than the total number of 51933 messages found in the data set. Which means that approximately 71% of all messages collected were only seen once on one of the seven news servers, this suggests that the dataset may not cover a large enough network sample to produce large numbers of significant common paths.

Only 50 of the 2228 unique common paths or approximately 2% contained a single news server entry, leaving the remaining 98% needing further analysis to determine their relevance. 1491 instances of common path were detected with frequencies of 1, making up approximately 67% of the total 2228 unique common paths. This suggests that again not enough message copies were collected from servers to form an accurate list of all common paths. However, this significantly reduces the size of the set of common paths needing to be considered when groups are chosen for further analysis.

Common Paths	Frequency	Proportion
postnews.google.com, news.glorb.com	2928	19%
uni-berlin.de, fu-berlin.de	859	5.7%
local02.nntp.dca.giganews.com,nntp.giganews.com, border1.nntp.dca.giganews.com	230	1.5%
news.motzarella.org, news.glorb.com	219	1.45%
local02.nntp.dca.giganews.com,nntp.giganews.com, border1.nntp.dca.giganews.com, news.glorb.com	176	1.2%
nntp.comcast.com, local02.nntp.dca.giganews.com, nntp.giganews.com, border1.nntp.dca.giganews.com	104	0.7%
uni-berlin.de,fu-berlin.de, news.glorb.com	90	0.6%
nntp.earthlink.com, local02.nntp.dca.giganews.com, nntp.giganews.com, border1.nntp.dca.giganews.com	85	0.56%

Table 2: Top 8 common paths based on their frequency and the proportion of the 15061 common paths they represent.

Selection of news groups

The 15061 instances of common path originated from 51 of the 96 news groups in the dataset. News groups were selected from this list to provide smaller samples for specific analysis to detect path preload. Selection was based on the criteria specified for news group selection which involved, choosing news groups with low numbers of common paths containing single servers, and with users associated with a significant majority of the common paths associated with that news groups.

Four news groups were selected for visual analysis based on these criteria. News groups which contain the majority of common paths tended to have a small number of users associated with a significant number of common paths. However, most of the associated common paths do not contain the right range

of news servers in correct order to make them suspicious. News groups with smaller numbers of common paths tended to have little constancy between their path information as they lack the quantity of common paths needed to make them suspicious. The four news groups selected (Table 3) have an average or less than average number of common paths, which gives them a much better user to message ratio. The results from the analysis of the rec.photo.marketplace news group are presented in this paper.

News groups	Common paths
rec.photo.marketplace	82
rec.heraldry	286
talk.politics.drugs	314
rec.arts.comics.marketplace	101

Table 3: The four selected newsgroups and number of common paths identified.

Rec.photo.marketplace

The rec.photo.marketplace contained a wide variety of common paths which exhibited features such as containing two or more alias or real news server addresses in concession at the beginning of the path, and a significant number of common paths associated with only a single user. Interestingly the common paths analyses from this group posed an exception to the assertion that single instances of common path are irrelevant. When the news group's common paths were grouped by their associated users, many single instances of common path began to form coherent groups of similar path elements. While the news group contains a total of 82 common paths, these paths form a total of approximately 10 groups, with some groups containing overlapping common paths.

The distribution of these common paths was mapped to a network visualisation of the server groups from the rec.photo.marketplace news group (Figure 1), where the groups of common paths (highlighted in the colours red, blue, pink and purple) were compared with the degrees of nodes on the network. From this visualization, the characteristics of common paths can be analysed in consideration with the graph's topological features:

- High traffic links: Identifiable from the weight of links between nodes which is represented by the thickness of the links shown on the graph.
- Hub news servers: These news servers have high degrees of connectivity and are indicated on the graph by red coloured nodes, less connected nodes are shaded a colour closer to yellow.
- Outlying news servers: These news servers are likely to be the news servers from which messages originated and are shaded a light-yellow colour.

The most frequently occurring common paths appeared to be created by messages which travel between two nodes with high degrees of connectivity. Less frequently occurring common paths tend to travel inwards from outlying news servers, eventually ending up at network hubs near the centre of the graph. However, more common paths appear to originate from outlying news servers, which suggests messages with common paths may have travelled long distances and would suggest that many common paths occur from not enough messages being collected as a result.

Through common path analysis it is possible to examine the news servers from which messages have originated. Figure 2 shows the news servers that messages in the rec.photo.marketplace news group

originated from highlighted as red circles, comparing them with the news group's common paths. Results of this comparison show a strong similarity between the origin of messages with common paths and those without, with one exception.

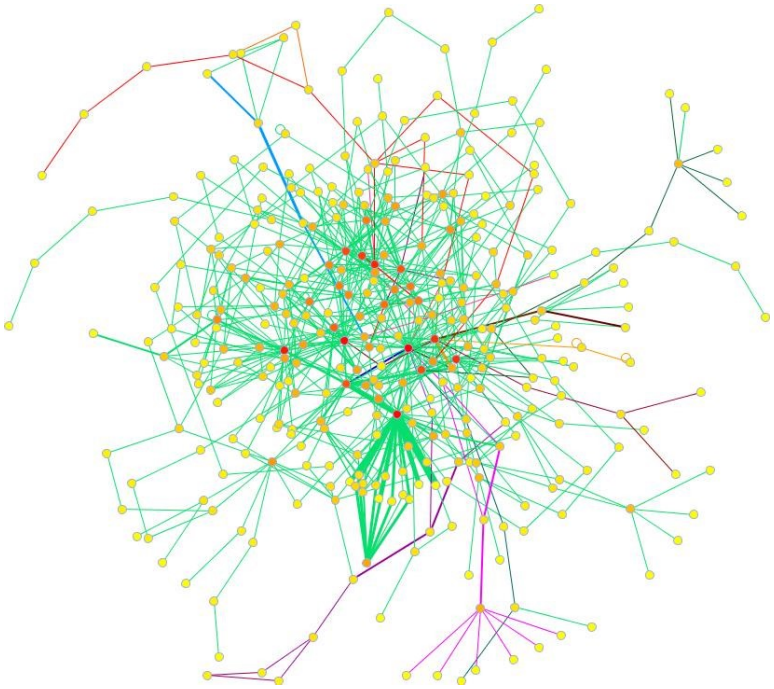

Figure 1: A network visualisation of rec.photo.marketplace news group showing node degree (nodes in red have the highest degree, yellow lowest) and common paths (edges in colours red, blue, pink and purple).

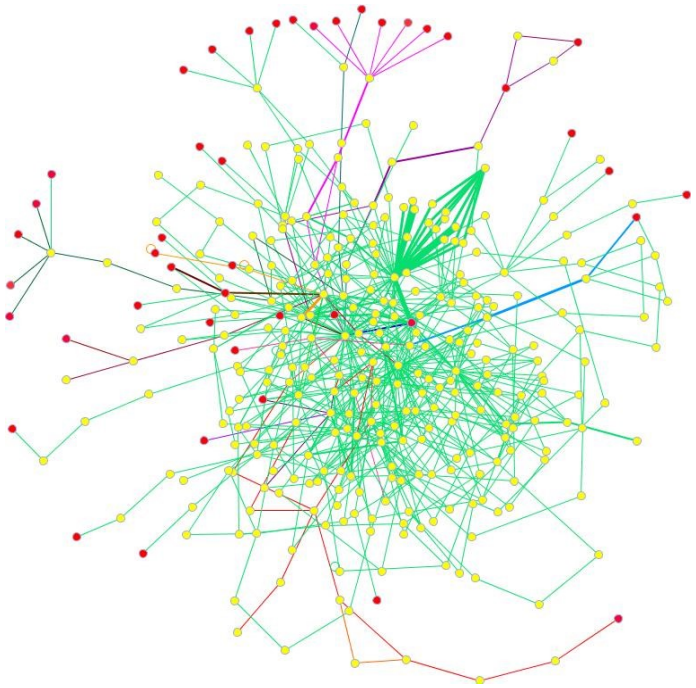

Figure 2: A network visualisation of message paths from the rec.photo.marketplace newsgroup showing a comparison between common paths (edges in colours red, blue, pink and purple) and news servers of origin as red nodes.

An anomalous distribution pattern

One particular common path (Figure 3) was identified as a potential instance of path-preload in the rec.photo.marketplace news group. The path was detected through comparing the distribution of messages containing common paths with the news servers from which other messages originated. The common path (Table 3) contains duplicate path information, which when visualized creates a visible fork on the visualized networks topology.

The full message path is: *fe3.news.blueyonder.co.uk.POSTED! blueyonder! pe1.news.blueyonder.co.uk! blueyonder! pe2.news.blueyonder.co.uk! nntp.giganews.com! border1.nntp.ams.giganews.com! feeder.news*

The alias news server entry ‘blueyonder’ appears twice in alternate positions within the common path. This could be caused by a news server receiving the message twice or adding two entries for itself in the path. Another reason could be that part of, or the entire common path was preload which would also explain the duplicate news server entries. On further examination the message did not contain any exceptional content, suggesting that the exception may have been caused by a news server rather than the user posting the message.

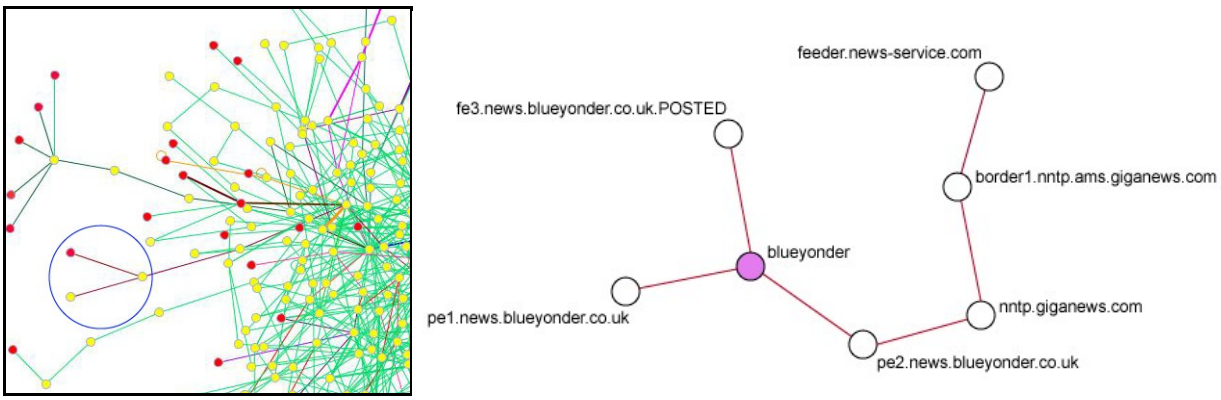

Figure 3: An anomalous path entry in rec.photo.marketplace. **Left:** highlights where on the network graph (Figure 1 and 2) the path can be found. **Right:** presents an illustration of the anomalous path information.

Conclusions

The results from this study demonstrate that when anomalous path information is present in message headers it can be identified through the method of common path comparison and visual analysis of network graphs. Anomalous paths were clearly visually identifiable on network graphs as illustrated by Figures 1, 2 and 3. In Figures 1 and 2 at least one other anomalous configuration is also visible (the purple path at the top right). While the method is unable to confirm whether the anomalous path information may have been deliberately loaded into the message header by a user or placed there by an incorrectly configured news reader or server it provides a foundation for the development of more precise techniques. Closer inspection of how anomalous paths relate to structural features on the network's topology may provide clues to patterns which can be used in the development of methods that can detect certain kinds of behaviour and even link it to certain kinds of content.

Usenet's structure and function has been shaped by the people and technology which underpin it. This analysis illustrates how these social and technical aspects (the dynamics) come together to shape the logical network. Administrators shape the network through the configuration of their news servers, including which news groups they host, the news severs they subscribe to and the policies they choose to enforce. Users may influence the network through their news reader, their choice of news server and the groups they subscribe/contribute to which also shape the network. As the infrastructure has changed so too has how the network is used, conversations have given way to file sharing and the distribution of many kinds of content, milliseconds determine which news server receives a message first and google groups (opened in 2001) while preserving much of the network's recorded history as major news servers cease operation has exposed the network to prolific amounts of spam.

Studies of Usenet often focus on the conversations and social data contained on the network. As this function diminishes new opportunities are emerging to study different kinds of content, social and network behaviour. The establishment of www.usenetarchives.com in 2021 also continues to keep the network's history alive and accessible, it provides access to data from 300 million posts in 9 thousand news groups. Usenet offers an accessible source of data that may provide insights into questions relevant to complex systems and network science, as well as the social and technical challenges faced by social networks such as the spread of misinformation and ideas online.

Appendix: List of News Groups

#	News Group Name	#	News Group Name
1	comp.constraints	51	alt.beach.shag.mp3s
2	comp.dcom.lans.token-ring	52	alt.brhuk
3	comp.games.development.programming.misc	53	alt.chris-caputo-election
4	comp.infosystems.intranet	54	alt.comp.periphs.soundcard.avm
5	comp.lang.c++.leda	55	alt.crimehip.lang.simula
6	comp.lang.perl.modules	56	alt.dcom.telecom.ip
7	comp.music.midi	57	alt.drugs.heroin.addict.jay-denebeim
8	comp.os.linux.misc	58	alt.emircpih.religion.agnostic
9	comp.os.ms-windows.nt.software.services	59	alt.fan.battlefield-earth
10	comp.os.os2.misc	60	alt.fan.dolly-buster
11	comp.society	61	alt.fan.letterman.guests
12	comp.sources.bugs	62	alt.fan.silentbob
13	comp.sys.amiga.programmer	63	alt.flame.bajarat
14	comp.sys.ibm.pc.games.war-historical	64	alt.flame.montgomery-wood.utterly-insane-arse-licking-clotpole-obsessive-compulsive-rmgrouper
15	comp.sys.mac.printing	65	alt.fr.defense.la.gendarmerie.nationale
16	comp.sys.psion.apps	66	alt.fr.luttons.contre.la.tuberculose
17	comp.text.sgml	67	alt.fred-hall
18	comp.unix.xenix.misc	68	alt.games.half-life.editi
19	misc.education.medical	69	alt.give-me-a-wake.before-i-die.on-usenet
20	misc.health.infertility	70	alt.h.i.p.c.r.i.m.e.lang.perl
22	news.misc	71	alt.h0pcr0me.industry.utilities.gas
23	rec.arts.comics.marketplace	72	alt.h1pcr1me.hotrod
24	rec.arts.movies.tech	73	alt.h2pcr2me.future
25	rec.autos.sport.indy	74	alt.hh.ii.pp.cc.rr.ii.mm.ee.admin.policy
26	rec.collecting.cards.non-sports	75	alt.hh.ii.pp.cc.rr.ii.mm.ee.taxes
27	rec.food.baking	76	alt.hipclone.software
28	rec.games.computer.puzzle	77	alt.hipcrime.music.research
29	rec.games.roguelike.adom	78	alt.i-think.therefore.i-dont.listen-to.top-40
30	rec.heraldry	79	alt.israel.la.terre.promise
31	rec.music.artists.extreme	80	alt.ks.wikipedia
32	rec.music.indian.misc	81	alt.mad-newgrouper.who.newgroups.at.midnight
33	rec.music.tori-amos	82	alt.movies.kubrik
34	rec.photo.marketplace	83	alt.music.mr-bungle
35	rec.sport.archery	84	alt.news-admins.japan.fascist.fascist.fascist
36	rec.sport.soccer	85	alt.online-service.internetmci

37	rec.video.releases	86	alt.personals.big-fol
38	sci.bio.immunocytochem	87	alt.pl.paintball
39	sci.engr.joining.welding	88	alt.president.clinton
40	sci.math.symbolic	89	alt.religion.christian-teen
41	sci.physics.particle	90	alt.sci.astro.hale.bopp
42	soc.adoption.parenting	91	alt.sex.girl.watcher
43	soc.culture.burma	92	alt.silly
44	soc.culture.iraq	93	alt.sports.football.pro.baltimore
45	soc.culture.quebec	94	alt.support.marijuana-munchies
46	soc.genealogy.slavic	95	alt.things-to-do-today.scan-apartment
47	soc.religion.quaker	96	alt.tv.strange-luck
48	talk.politics.drugs	97	alt.video.ptv.replaytv
49	alt.afilcm		
50	alt.astrology.pro		

References

Adar, E. (2006). GUESS: a language and interface for graph exploration. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 791–800).

Auber, D. (2004). Tulip—a huge graph visualization framework. In *Graph drawing software* (pp. 105–126). Springer.

B8MB. (2007). The Big 8 Management Board. Retrieved February 4, 2008, from http://www.big-8.org

Borgs, C., Chayes, J., Mahdian, M., & Saberi, A. (2004). Exploring the community structure of newsgroups. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 783–787).

Fellows, G. (2006). Newsgroups reborn--The binary posting renaissance. *Digital Investigation*, *3*(2), 73–78.

Fisher, D. (2003). Studying social information spaces. In *From Usenet to CoWebs* (pp. 3–19). Springer.

Himsolt, M. (1996). GML: A portable graph file format. Retrieved April 10, 2008, from www.infosun.fmi.unipassau.de/Graphlet/GML/gml-tr.html

Horton, M. R. (1983). *Standard for interchange of USENET messages*.

Keim, D. A., Mansmann, F., Schneidewind, J., & Ziegler, H. (2006). Challenges in visual data analysis. In *Tenth International Conference on Information Visualisation (IV'06)* (pp. 9–16).

Kim, J., Schneider, F., Ager, B., & Feldmann, A. (2010). Today's usenet usage: NNTP traffic characterization. In *2010 INFOCOM IEEE Conference on Computer Communications Workshops* (pp. 1–6).

Lewis, S. E. (2004). Master List of Newsgroup Hierachies. Retrieved February 15, 2008, from

Lueg, C. (2003). *From Usenet to CoWebs: interacting with social information spaces*. Springer Science \& Business Media.

McCane, B. (1999). Free and Public Nsenet News Servers. Retrieved February 3, 2008, from

http://freeusenetnews.com

Nakano, Y., Nakamura, M., & Okabe, Y. (2007). Analysis for topological properties of the network feeding usenet news. In *2007 International Symposium on Applications and the Internet* (p. 14).

Newzbot. (1995). Public USENET Resource. Retrieved February 3, 2008, from http://www.newzbot.com

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, *10*(4), JCMC1048.

Withall, M., Phillips, I., & Parish, D. (2007). Network visualisation: a review. *IET Communications*, *1*(3), 365–372.