Forensic Triage of Email Network Narratives through Visualisation

John Haggerty

School of Computing, Science and Engineering,
University of Salford, Salford, Greater Manchester, UK

Sheryllynne Haggerty

Department: School of Humanities,
University of Nottingham,
Nottingham, UK

Mark Taylor

School of Computing and Mathematical Sciences
Liverpool John Moores University,
Liverpool, UK

1. Introduction

Due to the amount of information that email may provide to a forensics examiner, it remains a key source of evidence during a digital investigation. However, the analysis of email adds complexity to a digital investigation due to the amount of data that must be searched for relevant evidence. For example, in one investigation, police officers analysed 100,000 indecent images of children and 10,000 emails for the prosecution of a paedophile ring involving four individuals (BBC, 2012). This is extremely time-consuming because current practice utilises tools and techniques that require manual analysis of email files.

Email is particularly useful during a digital investigation in that it may elucidate quantitative information, i.e. network events and actor relationships, and qualitative information, i.e. email content, to the forensics examiner. Recent research in this area has focused on the quantitative analysis of emails, for example, by analysing actor relationships identified through this medium. However, these approaches are unable to analyse the qualitative content, or narrative, of the emails themselves to provide a much richer picture of the evidence. This paper therefore posits *TagSNet* (**Tag** cloud and **S**ocial **Net**works)*,* a novel approach which combines both quantitative and qualitative analysis of emails using data visualization. As will be demonstrated by the case study, the examiner is able to triage large volumes of emails to identify actor relationships as well as their network narrative. In this way, they will be able to prioritize their search for potential evidence relevant to the investigation.

This paper is organised as follows. Section 2 discusses related work in data visualisation, forensics and analysis of emails. Section 3 posits our approach, *TagSNet*, for the triage of quantitative and qualitative email data in digital forensics investigations. Section 4 presents the results of applying this methodology to a case study using the Enron email corpus. Finally, we make our conclusions in section 5.

2. Data visualisation and digital forensics

Due to the complexity and volume of data available today, there is much interest in data visualization of narratives outside the digital forensics domain. For example, Segel and Heer (2010) and Hullman and Diakopoulos (2011) propose visualization approaches using data produced by media organisations for conveying rhetoric, for example, political discussions in news stories. These approaches posit design strategies for visualization and interpretation of

narratives. Fisher *et al*. (2008) posit *Narratives*, an approach to visualize key words over time. This approach visualizes a sequence(s) of key words as a series of related line graphs. They suggest that this approach could be used for tracking items or actors of interest in news items. Dou *et al.* (2012) posit *LeadLine*, a tool to automatically detect events in news items and social media as well as support their exploration through visualization.

Other visualization approaches extend their analysis beyond Web data. For example, Nair *et al.* (2011) posit how a patient's data may be better represented to clinicians by using documents to produce patient 'stories'. Wang *et al.* (2011) posit a methodology for the analysis of large textual documents. This approach focuses on a central event and then analyses the relationship between this and other events. Ungar *et al.* (2011) propose *IntentFinder*, a tool for the analysis and representation of data which attempts to link document and narrative information with a subject's social networks. What these approaches have in common is that they are not designed for forensics investigations, for example, by only allowing data to be mounted in read-only mode.

The advantages of using data visualization for large data sets have led to such approaches in digital forensics being posited. For example, Schrenk and Poisel (2011) discuss the requirements for visualization in digital investigations due to the volume of data that must be searched. Whilst they do not posit a single approach, they discuss methodologies for a range of visual exploration, such as time-related and email data. Osborne *et al.* (2012) focus on visualizations to support the investigatory process rather than data to identify evidence *per se*. Jankun-Kelly *et al.* (2009) posit an approach to investigate a range of documents, including Webcache files and email. This approach focuses on visualization of textual data rather than relationships between actors. Hai-Cheng Chu *et al.* (2011) suggest an approach for the identification of social networks through Social Network Services, such as Facebook. However, this approach focuses on the extraction of evidence rather than the visualisation of the social networks discovered. Palomo *et al.* (2011) focus on the visualization of network traffic through self-organising maps to identify anomalous behaviour or system intrusions. However, this approach focuses on the identification and visualization of network artefacts, such as source ports, destination addresses, protocols, etc. rather than social interactions between actors or network narratives.

Other approaches to data visualization in digital forensics have focused on email as a potential source of evidence. For example, the Forensic Toolkit (FTK) version 5 (Access Data, 2013) has now integrated social network visualisations of emails into their software. Haggerty *et al.* (2011) use the Enron email corpus as a case study to propose a method for the triage and analysis of actors within an email network. Henseler (2010), who also uses the Enron data set, suggests an approach for filtering large email collections during an investigation based on statistical and visualisation techniques. Wiil *et al.* (2010) provide an analysis of the 9/11 hijackers' network and focus on the relationships between these actors. This study uses a number of measures associated with social network analysis to identify key nodes. However, these approaches only focus on the quantitative analysis of actor relationships rather than the qualitative information within the emails themselves.

The importance of qualitative information in email content is recognised in other research. For example, DeBarr *et al.* (2013) suggest an approach for the analysis of Uniform Resource Locators (URLs) substrings for the detection of "phishing" attacks. Hamid and Abawajy (2011) propose an approach to detect "phishing" emails by combining content with behaviour analysis. Zilberman *et al.* (2011) use content analysis to ensure that topics are common to sender and recipient for the detection and prevention of data leakage via email. Yoshinaga *et al.* (2010) characterise email content by keywords to determine how email activity depends on content. Esichaikul *et al.* (2011) propose an approach to mine the content of emails to determine important emails within a data set. However, all these approaches have in common that they do not address the requirements for triage during a digital investigation. Moreover, they do not provide visualisation of the email content in order to analyse potential evidence.

There is therefore a requirement for combining both quantitative and qualitative data during an investigation to not only visualize the actors involved, but also to analyse what is being discussed, i.e. the network narrative. The next section posits the methodology to meet this requirement.

3. Methodology Overview

As suggested in (BBC, 2012), the volume of email data that a forensics examiner may encounter during a digital investigation may be considerable. Therefore, the key challenges to digital investigations involving emails include: the volume of data that may contain evidence, evidence identification and analysis, identification of potential sources of evidence such as actors or data sources of interest, and representation of the evidence. This section posits the *TagSNet* approach for the automated visualisation of quantitative and qualitative email data to meet these challenges and to triage evidence.

Currently, there is no accepted definition of the term 'network narrative'. In related literature, a *network* comprises a set of actors and the relations between them and the network itself. A *narrative* is the discourse in relation to network events or effects. We therefore define 'network narrative' as; *the discourse with regard to a set of actors, their relationships and events pertaining to them*. Identifying the network narrative allows us to assess the impact of endogenous and exogenous events of interest on the network(s) and content discovered during a digital forensics investigation.

Figure 1 illustrates the framework for the forensic investigation of email data. These processes do not differ much from investigations into other file types. However, the Triage and Analysis stage reflects the need to identify and assess both quantitative and qualitative data. This is achieved through the use of the *TagSNet* software specifically aimed at this type of analysis.

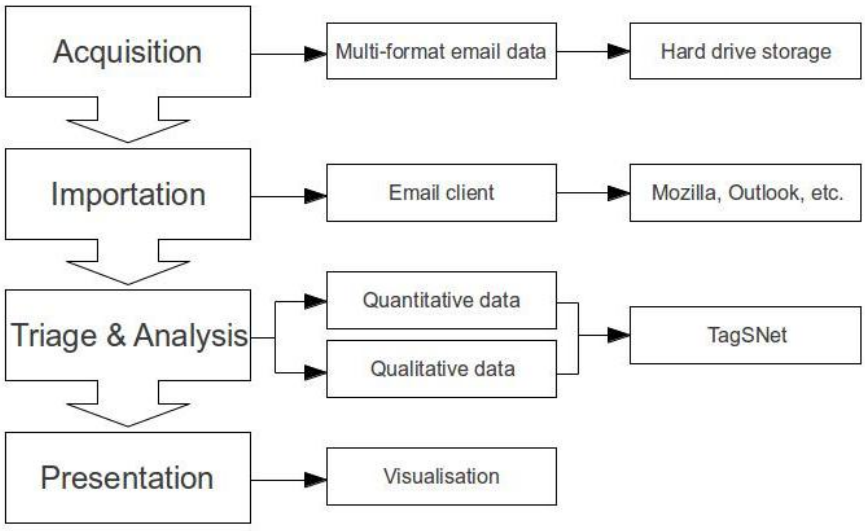

Figure 1. The process for the forensic investigation of emails.

As with any investigation, the data must be acquired in a robust manner, ensuring that the evidence maintains its integrity. Therefore, emails are imported into *TagSNet* in read-only mode to avoid data modification. These email files are located in client-specific directories. For example, Mozilla Thunderbird stores email data in text format in mbox files under the following directories dependent on the operating system: C:\Documents and Settings\[UserName]\ApplicationData\Thunderbird\ Profiles\ (Windows XP), C:\Documents and Settings\[Your user name]\Application Data\Thunderbird\Profiles\[ID].default\Mail\Local Folders (Windows 7) and ~/.thunderbird/xxxxxxxx.default/ (Linux) and ~/Library/Thunderbird/Profiles/ xxxxxxxx.default/ (Mac OS X). Once the data has been imported, the triage and analysis stage

views the files without any direct modification of the original data allowed. The visualisations of the data form the basis for the presentation of evidence by summarising large data sets in an easy-to-view format.

As discussed above, the information that may be retrieved from emails falls into two categories; quantitative and qualitative. Quantitative information refers to the network events and actor relationships that may be inferred from this data. For example, Alice sends Bob an email which is an event in the network. From this event, we may infer that Alice and Bob have some form of relationship in that there is communications between the two actors. With the multiple events that occur in the network over time, a forensics examiner may infer relational information between actors, for example using statistical techniques discussed in (Haggerty *et al.*, 2011) or (Henseler, 2010). This information may be retrieved from the email header. Qualitative information refers to the content of the emails themselves. For example, when Alice sends Bob an email, a message is sent. This content may be in textual form or may be some other data form, such as an attached file. This information may be retrieved from the message body section of an email.

TagSNet has been developed by the authors to meet the requirements of email investigations by visualising the networks and content of emails. This software extends the *Matrixify* (Haggerty and Haggerty, 2011) temporal social network analysis tool. These visualisations are not aimed at answering questions *per se*, but to enable a forensics examiner to triage email data more quickly than a manual trawl or just relying on social network analysis. In this way, the forensics examiner may be able to see not only who is communicating in the network but also what they are talking about and identification of key issues to the actors. Moreover, the triage may present further potential sources of evidence that would be of interest during the investigation.

The software provides two views for the analyst; the social network to which they belong and a tag cloud of the email contents. A social network is an interconnected group or system and the relations, both logical and physical, between the actors. It should be noted that the relationships are derived from the email flow and may be a simplistic representation of the actual relationship that actors within the network have with one another. Moreover, there is a tendency to assume that just because actors are linked they must form a cohesive and positive social network. However, this is not necessarily the case and therefore the relationships between network members must be explored further to fully understand how these networks function (Haggerty *et al.*, 2011). The network views in *TagSNet* are ego-centric in nature due to the source material, i.e. we do not know the relationships between actors beyond those identified in the suspect's emails. Rather than reading individual emails to build up a picture of the discussions and themes in the network narrative for content analysis, *TagSNet* identifies and quantifies the data in an email client's data folder, i.e. it visualises the folder rather than an individual message. Through this data mining, key words are identified as they re-occur, thereby identifying the network narrative concerns.

These two elements combined provide a rich picture of the network events and relationships over time, including reactions to endogenous and exogenous events. Of interest to the forensics examiner are the following:

- Key actors
- Actor relationships in the network at specific times
- Key narratives in the network
- Change over time (e.g. pre- and post-criminal activity)
- The identification of further evidence sources or lines of enquiry in either quantitative or qualitative data

As illustrated in figure 2, the software has, at its most basic level, three main areas of functionality; file reading and processing (data mining), visualization, and graphical output. These functional points are covered in more detail below.

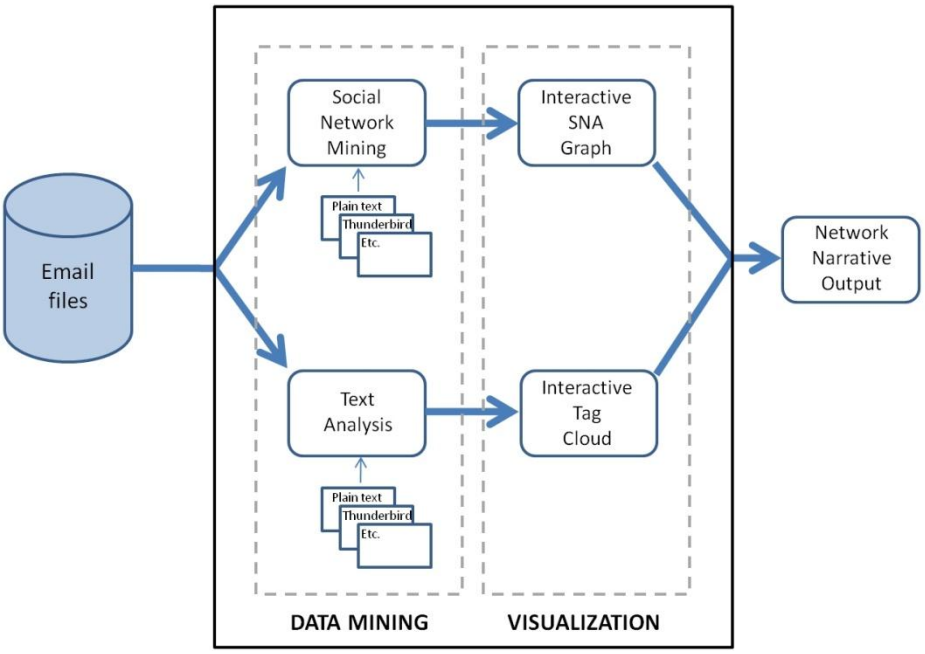

Figure 2. Overview of *TagSNet*.

The email files are processed in two ways; for social network information and narrative analysis. The social networks elucidated by the email files are derived from the FROM, TO, and CC data in both messages sent to and received by the suspect. This data includes the search of forwarded messages located under the main message. As noted above, this view of the network is a suspect-centric snapshot, i.e. as we are analyzing the suspect's computer, the social networks will be from the suspect's point of view. Narrative analysis is achieved by creating a dictionary of all words in the email file and then counting their occurrence. These results are made available to the visualization function as the basis for text sizing. It is posited that the occurrence of words (or lack of) suggests their concern to the network. As such, commonly occurring words, such as 'the', 'a', 'to', etc. are ignored during this process. These words provide a useful function in language but their commonality adds noise to the visualizations without adding to the network narrative analysis. However, this function could be extended to include a user-defined dictionary of words to include or exclude in a search.

The results of this data mining are passed to the two visualization functions. A social network graph is constructed from the data passed from the social network mining function. This graph-building element visualises actors as network nodes, identifies the actors and produces lines to represent relationships between them. A tag cloud is created from the textual analysis results to produce the narrative view of qualitative data. This view sizes words in the email text by frequency of occurrence and these are placed using a random layout. Various sensitivity levels, or thresholds, can be applied to the data, based on popularity of words, to reduce noise, and highlight key concerns within the text. These visualizations together form output in the form of a network narrative. Both these visualizations are interactive in that the forensics examiner may move both actors and text around. This enhances the visualization by ensuring that the results can be explored and that the best layout can be chosen.

This section has provided an overview to the *TagSNet* approach for the analysis of network narratives in email data. In the next section, we demonstrate the applicability of the proposed approach for triaging evidence by applying it to email data from the Enron corpus.

4. Case Study and Results

Enron was a large energy company that employed thousands of workers across 40 countries. The Enron fraud resulted in the bankruptcy of the company and dissolution of a large

accountancy and audit company. The main executives, such as the CEO Jeffrey Skilling, whose emails form the basis of this case study, used a series of techniques to perpetrate the fraud, such as accountancy loopholes, employing special purpose entities and poor accountancy practices, in order to hide billions of dollars of debt that the company had accrued. The email corpus is available online at (EnronData.org, n.d.) and provides a useful test set for methodologies related to email data due to its size and complexity.

Table 1 provides an overview of the Jeffrey Skilling email data set. Folder organisation is important for the forensics examiner as this will provide a rudimentary level of triage. For example, Skilling kept many of his work-related emails in folders such as Inbox and Sent Items. He also kept emails related to work in less obvious folders, such as the Genie folder which is related to attendance at a specific conference. The email data set also provides information about his personal life as well. For example, the Mark folder contains emails related to a family member.

Folder	**No. of emails**	**Mbox file size (KB)**	**No. of actors**	**Visualisation time (secs)**
_Sent Mail	275	748	515	11.292
All Documents	834	2,846	1427	105.425
Deleted Items	483	1,991	727	59.856
Discussion Threads	652	2,353	1330	76.416
Inbox	1253	5,421	2041	333.128
Genie	10	19	10	0.106
Mark	55	291	84	1.929
Notes Inbox	244	761	568	14.643
Sent	276	756	556	11.112
Sent Items	54	180	121	0.586

Table 1. Analysis of the Skilling email account by folder.

As can be seen in table 1, the email folders range in the amount of information that may be returned about Skilling's network narratives. For example, the amount of email content in plaintext varies from 19 KB to 5.4 MB and the time that it takes to visualise the text corresponds to the size of the file. Moreover, the number of actors in the network also varies depending on the amount of emails that are stored, in this case from 10 in the Genie folder to 2041 in the Inbox. In table 1, the average times to process and visualise the network narrative of the mbox files on a Windows 7 computer with a 2 GHz Intel Pentium Processor and 4 GB RAM range from 0.106 seconds to 333.128 seconds. This is significantly faster than reading the emails manually and therefore aids the triage process.

Three folders from the Skilling email account are used to illustrate the ability of *TagSNet* to triage data and prioritise searches. It should be noted that figures 3 to 5 demonstrate this triage process for the identification of potential evidence rather than to provide evidence of the fraud discussed above. Moreover, they do not provide measurements or layouts based on statistical measures of the network, such as centralities suggested in *Haggerty et al.* (2011), as this is outside the scope of this paper. These email folders, Genie, Mark and _Sent Mail, are used for two reasons. First, they represent different aspects of Skilling's email use; a specific set of correspondence related to a business event, personal correspondence with a family member and general business email traffic. This allows us to compare narratives in different contexts. Second, ranging from a small (10 actors, 930 words and 19 KB mbox file) to large (515 actors, 50,198 words and 748 KB mbox file) data set evaluates the impact of data scaling on the approach.

Emails from the Enron corpus are converted to Thunderbird mbox format to aid data mining as they are stored in plaintext. As discussed in section 3, email header data is used to generate

network diagrams whilst the text of the emails, i.e. content, is used to generate the tag clouds, combining to form the network narrative. The two views in *TagSNet* are shown in different windows. However, for aesthetic purposes and comparison, the frames have been removed to focus on the network narrative in this paper. Due to the size of the mbox files, different levels of sensitivity to content data mining have been used. For example, in small files, such as Genie, it is possible to show all keywords. However, in larger files, this creates background noise. Therefore, thresholds of word re-occurrence are used to reduce the amount of information that is returned in the visualization. *TagSNet* allows the user to set the threshold level to provide the best aesthetic view without distorting the evidence.

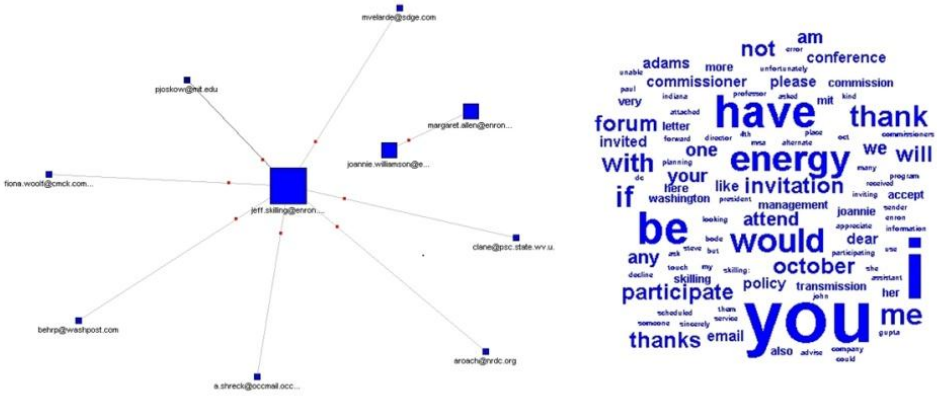

Figure 3. Genie folder network narrative.

Figure 3 illustrates the Genie folder network narrative. This network comprises 10 emails, 10 actors and 930 words. Network nodes are sized by occurrence. All words are included in the visualization and the size of the font indicates their reoccurrence in the mbox file. The words highlighted in this view include; ‘you’, ‘energy’, ‘invitation’, ‘participate’, ‘forum’, ‘October’, ‘attend’, ‘invited’, ‘policy’, �’management’, ‘conference’ and ‘Washington’. A qualitative analysis of the original emails suggests that this folder contains information that relates to the attendance at an energy forum in Washington organised by Skilling, and this is evident in the network narrative visualization.

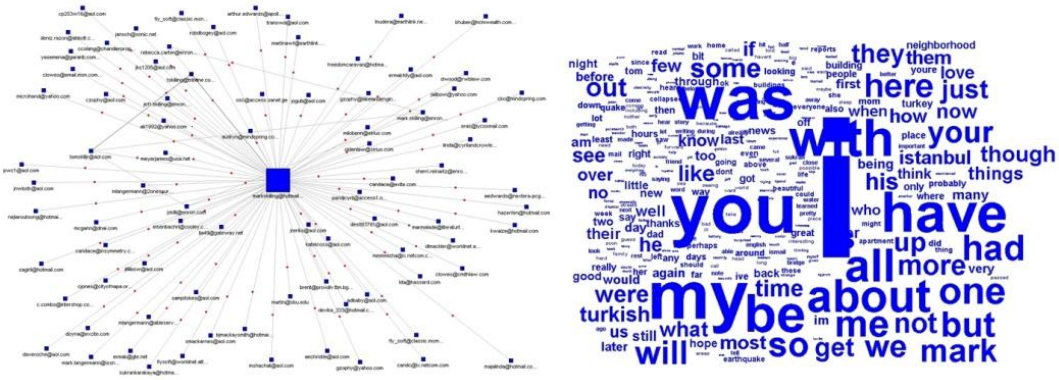

Figure 4. Mark folder network narrative.

Figure 4 illustrates the Mark folder network narrative. This network comprises 55 emails, 84 actors and 28,064 words of correspondence with a relative of Skilling. To reduce noise in the tag cloud, only words that occur more than ten times are included. The personal nature of the content is illustrated by the dominance of ‘I’ and ‘my’ in the network narrative view. In addition, other general but personal words, such as ‘me’, ‘we’, ‘have’, etc. dominate the view. However, within the network view, ‘Istanbul’ and ‘Turkish’ also appear. A qualitative analysis of the original emails

suggests that this a folder associated with personal messages from a family member, and there is some association with Turkey.

Figure 5. Sent_mail folder network narrative.

The _Sent Mail folder, as illustrated in figure 5, comprises 275 emails, 515 actors and 50,198 words. Again, to reduce noise, only words that occur more than 20 times are included in the tag cloud. This folder differs from the other two in that the content is far more general as they are emails related to Skilling's day-to-day business dealings. Two names are immediately apparent in the tag cloud, Jeff Skilling and Sherri Sera (Skilling's personal assistant) and this is supported by the network diagram. Indeed, many of the emails were sent by Sera on behalf of Skilling and were saved to this folder. This is illustrated by the prominence of her email address, 'sherrisera@enron.com' in the content as it appeared in the signature block. The use of words differs to those in the Mark folder in that they are obviously more related to business, for example, 'enron', 'fax', 'business', 'information', 'executive', 'company', 'assistant' and 'message'. Also highlighted is a location, 'Houston', where the business had its headquarters. Moreover, two numbers are also identified; '7136468381' and '7138535984'. These are the phone and fax numbers for Sera.

The three network narratives above quickly identify where to prioritise a manual trawl of emails for evidence using traditional forensic tools and provide a substantially quicker analysis than manually reading the files to triage data. Given the personal nature of the Mark folder's emails, we may place this as a low priority unless the family member was somehow implicated in the case. We could also discount the Genie folder's emails, unless the case was related to the forum that took place in Washington. The highest priority would be the _Sent Mail folder for a number of reasons. First, it highlights the importance of Skilling's personal assistant in his business activities and would indicate that her email account may provide relevant evidence to the investigation. Second, as the emails are associated with business dealings, it may identify other actors of interest in the network views. Third, it highlights further potential sources of evidence, such as the phone numbers that are used, and therefore call logs, which could be beneficial to the investigator. It should be noted that in investigations involving emails, key words highlighted by the network narrative may be misleading as the actors involved may use codes. However, any unusual words would be highlighted in the visualizations and could be followed up in the manual analysis.

5. Conclusions

Due to the amount of information email may provide to a forensics examiner, it remains a key source of evidence during a digital investigation. With our reliance on this medium, a forensics examiner may be required to triage and analyse large email data sets. Current practice utilises tools and techniques that require a manual trawl through such data, which is a time-consuming process. Recent research has focused on data visualization to mitigate the effect of large data

sets on an investigation. The approaches concerned with emails focus on the analysis of emails to identify social networks. However, these approaches are unable to analyse the qualitative, i.e. content (or narrative), of the emails themselves to provide a much richer picture of the evidence.

This paper therefore posits a novel approach, *TagSNet*, to visualise the network narratives present in email data. This approach combines both network events and relational information with content analysis. In this way, it provides a rich picture of a suspect's activities. As demonstrated by the case study, this approach can be used to triage data that may be of interest to the examiner to be followed up with manual searches for evidence specific to the case or to identify further sources of evidence. Further work aims to extend this approach to other media, such as online documents and social media.

References (to be sorted to ensure they fit the style)

Access Data (2013). http://www.accessdata.com. (Accessed 30 October 2013).

BBC (2012), http://www.bbc.co.uk/news/uk-england-19947914. (Accessed 30 October 2013).

DeBarr, D., Ramanathan, V. & Harry Wechsler, H., "Phishing Detection Using Traffic Behavior, Spectral Clustering, and Random Forests", *Proceedings of Intelligence and Security Informatics (ISI 2013)*, June 4-7, 2013, Seattle, Washington, USA, pp. 67 - 72.

Dou, W., Wang, X., Skau, D., Ribarsky, W. & Zhou, M.X (2012), "LeadLine: Interactive Visual Analysis of Text Data through Event Identification and Exploration", *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, Seattle, USA, 2012, pp. 93-102.

EnronData.org (n.d.), http://enrondata.org/content/data/. (Accessed 30 October 2013).

Esichaikul, V., Guha, S. & Juntapoln, C., "Monitoring Email Transaction Logs by Text-Mining Email Contents", *Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA)*, 2011, pp. 255 - 258.

Fisher, D., Hoff, A., Robertson, G. & Hurst, M. (2008), "Narratives: A Visualization to Track Narrative Events as they Develop", *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, Columbus, USA, 2008, pp. 115-122.

Haggerty, J. and Haggerty, S. (2011), "Temporal Social Network Analysis for Historians: A Case Study", *Proceedings of the International Conference on Visualization Theory and Applications (IVAPP 2011)*, Algarve, Portugal, 2011, pp. 207 - 217.

Haggerty, J., Karran, A.J., Lamb, D.J. and Taylor, M.J. (2011), "A Framework for the Forensic Investigation of Unstructured Email Relationship Data", *International Journal of Digital Crime and Forensics*, Volume 3 Number 3, September 2011, pp. 1-18.

Hai-Cheng Chu, Der-Jiunn Deng, and Jong Hyuk Park, "Live Data Mining Concerning Social Networking Forensics Based on a Facebook Session Through Aggregation of Social Data", *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 7, August 2011, pp. 1368 - 1376.

Hamid, I.R.A. & Abawajy, J., "Phishing Email Feature Selection Approach", *Proceedings of the International Joint Conference of IEEE TrustCom-11*, Changsha, China, November 16-18, 2011, pp. 916 - 921.

Henseler, H. (2010), "Network-based filtering for large email collections in E-Discovery", *Artificial Intelligence and Law*, Volume 18 Number 4, pp. 413-430.

Hullman, J. & Diakopoulos, N. (2011), “Visualization Rhetoric: Framing Effects in Narrative Visualization”, *IEEE Transactions on Visualization and Computer Graphics*, Volume 17 Number 12, pp. 2231-2240.

Jankun-Kelly, T.J., Wilson, D., Stamps, A.S., Franck, J., Carver, J. & Swan II, J.E. (2009), “A Visual Analytic Framework for Exploring Relationships in Textual Contents of Digital Forensics Evidence”, *Proceedings of the 6th International Workshop on Visualization for Cyber Security*, Atlantic City, USA, 2009, pp. 39-44.

Nair, V., Kaduskar, M., Bhaskaran, P., Bhaumik, S. & Lee, H. (2011), “Preserving Narratives in Electronic Health Records”, *Proceedings of the International Conference on Bioinformatics and Biomedicine*, Atlanta, USA, 2011, pp. 418-421.

Palomo, E.J., North, J., Elizondo, D., Luque, R.M. & Watson, T. (2011), “Visualization of Network Forensics Traffic Data with Self-Organizing Map for Qualitative Features”, *Proceedings of the International Joint Conference on Neural Networks*, San Jose, USA, 2011, pp. 1740-1747.

Schrenk, G. & Poisel, R. (2011), “A Discussion of Visualization Techniques for the Analysis of Digital Evidence”, *Proceedings of the 6th International Conference on Availability, Reliability and Security*, Vienna, Austria, 2011, pp. 758-763.

Osborne, G., Turnbull, B. & Slay, J. (2012), “Development of InfoVis Software for Digital Forensics”, *Proceedings of the 36th International Conference on Software and Applications Workshop*, Izmir, Turkey, 2012, pp. 213-217.

Segel, E. & Heer, J. (2010), “Narrative Visualization: Telling Stories with Data”, *IEEE Transactions on Visualization and Computer Graphics*, Volume 16 Number 6, pp. 1139-1148.

Ungar, L., Leibholz, S. & Chaski, C. (2011), “IntentFinder: A System for Discovering Significant Information Implicit in Large, Heterogeneous Document Collections”, *Proceedings of the International Conference on Technologies for Homeland Security*, Waltham, USA, 2011, pp. 219-223.

Wang, D., Liu, W., Xu, W. & Zhang, X. (2011), “Topic Tracking Based on Event Network”, *Proceedings of the International Conferences on Internet of Things, and Cyber, Physical and Social Computing*, Dalian, China, 2011, pp. 488-493.

Wiil, U.K., Gniadek, J. & Memon, N. (2010), “Measuring Link Importance in Terrorist Networks”, *Proceedings of the International Conference on Social Networks Analysis and Mining*, Odense, Denmark, 2010, pp. 225-232.

Yoshinaga, N., Itaya, S., Rie Tanaka, R., Konishi, T., Doi, S., Yamada, K. & Davis, P., "Content Propagation Analysis of Email Communications", *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 79 - 82.

Zilberman, P., Dolev, S., Katz, G., Elovici, Y. & Shabtai, A., "Analyzing Group Communication for Preventing Data Leakage via Email", *Proceedings of Intelligence and Security Informatics 2013*, June 4-7, 2013, Seattle, Washington, USA, pp. 37 - 41.