Development of Interactive Visualisations of Network Structure and Data Flow: A Literature Review

ROBERT PASSMORE, UNIVERSITY OF CAPE TOWN

ABSTRACT

National Research and Educational Networks (NREN) are fundamental in providing Internet connectivity between research and education institutions and facilitating collaboration between them. Managers and policy makers who develop and maintain these networks require tools to identify network problems and to plans changes to network structure to improve bandwidth, latencies and reliability. Visualisations of network structure and data flow information would provide overview and detailed insight into the structure and behaviour of the network, and are therefore key to NREN managers and policy makers. This literature review explores the available literature on collection of network data flow information, the use of tools to discover the structure of networks and the design and evaluation of information visualisations to effectively communicate network information to the necessary parties.

Categories and Subject Descriptors

C.2.1 [Computer-Communication Networks]: Network Topology

H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces

General Terms

Design, Measurement, Visualisation

Keywords

National Research and Education Networks, Autonomous Systems, Distributed Network Probing, Visualisations, Visualisation Evaluation

1. INTRODUCTION

The network traffic between African member institutes of the UbuntuNet Alliance, the regional Research and Education Network that connects the NRENs of South-Eastern Africa, suffers from performance setbacks in the form of low latencies due to the intercontinental circuitous routing followed by 75% of the traffic [Chavula et al, 2014]. This problem is difficult to address since the managers and policy makers do not have the tools to effectively analyse the network structure and data flows of UbuntuNet to identify problems areas. With an up to date collection of network information and a suite of visualisations representing this information, network managers and policy makers could easily identify unexpected, circuitous routing of network traffic, as well as problems with latencies, traffic congestion and bandwidth utilisation on routes between member institutes of UbuntuNet. Importantly, the network information data must be kept up to date and accurate for the visualisations to accurately represent the network. Visualisations – as cognitive tools – must be designed to enable informed decisions through advanced insight as to how to improve the performance of the network. Of particular importance is the design and evaluation of information visualisations so as to produce useful and effective tools to the user.

2. TOPOLOGY DISCOVERY

Internet topologies can be represented at different levels of resolution. Ranging from the IP level to the Point-of-Presence (PoP) and Autonomous System (AS) level [Donnet, 2007]. The Ubuntunet Alliance connects NRENs which are AS-level networks [Gilmore, 2014]. However, an AS level representation is not sufficient for understanding the relationship between ASes. This is because ASes interconnect at PoPs and Internet Exchanges. To analyse the network, the network must also be represented at the PoP level to understand where exactly these ASes connect and how they connect.

Topology discovery uses tools that were not built for the purpose topology discovery [Luckie, 2008]. Namely traceroute and Border Gateway Protocol (BGP) logs. Traceroutes are network diagnostic tools that identify networked devices along a route and is used to gather a perspective of the routes which network traffic takes and the time it takes to return from the destination to the source – Round Trip Time (RTT). However, due to the heterogeneous nature of the internet, traceroute encounters network nodes and protocols that affect its accuracy [Donnet, 2007]. Firewalls – which prevent the traceroute ICMP messages from returning to the source – NAT routers and load-balancing routers are examples of network nodes that might impact the accuracy of a traceroute.

BGP tables can be analysed to understand the AS-level routes which traffic might take between a source and destination. This is gives a high-level view of the Internet - the chain of ASes between source and destination.

3. DISTRIBUTED PROBING

To effectively map an Internet-scale network, traceroutes and BGP logs need to be deployed and gathered from sources distributed through the network. This increases the portion of the network explored as it decreases the chances of inadvertently avoiding ASes that are not present between your selection of source/destination pairs [Montamedi, 2013].

Many tools facilitate this probing (ATLAS, RIPE, etc.). Some consideration need to be taken when choosing your source locations and destination addresses to get an accurate and complete view of the network [Montamedi, 2013]. Consideration is also required to avoid the pitfalls of traceroute probing from single locations. Namely the congestion of the network or the redundancy of paths [Montamedi, 2013]. This occurs when a source probes many destinations and the first few hops of each traceroute travel along the same path. This might also occur at hops close to the destination node when many traceroutes redundantly cover the links to the destination node.

4. INFORMATION VISUALISATIONS

An effective visualisation needs to accurately and clearly convey the information needed by the user. Information visualisations can take a variety of forms and presents many challenges.

4.1 Challenges

Network topology and traffic visualisations suffer from crowding due to the large amount data to be visualised. This presents a few design challenges.

Firstly, the visualisations will have to overcome this crowding problem to be effective tools for exploration and analysis. Effective interaction and filtering features can help drill-down and explore detail without crowding the interface, however, the ‘big-picture’ view of all the data that should allow the user to grasp the context of all the data will still be vulnerable to crowding [Becker et al., 1995].

The use of graph visualisations for network topologies (an inherently spatial structure) presents some challenges and opportunities.

The final visualisation will take advantage of the fact that it is not bound to the geo-spatial structure of the network topology and as a result can utilise space more effectively to avoid the occlusion/hiding of data in what would otherwise be a crowded visualisation in the areas with a dense population of network infrastructure (PoPs, IXs).

It is a challenge is to effectively communicate the structure of the network without the use of a geographical map as a background for the visualisation. The spatial reference of a map normally gives the user an immediate comprehension of the location and context of the nodes in the visualised graph. Possible solutions to overcome the challenge include effective colour coding and explanatory legends, helping the user identify different network regions; pop-up information describing the node and its neighbouring nodes upon hovering over the node; as well as fine-grained filters to highlight the structures that the user is looking for.

A design decision must be made with the cooperation of the users to decide what information should be displayed using the different 'channels' in the visualisation. These channels may include colours (of nodes or links), distance (between nodes) and width of links (normally indicating some kind of weight attached to the link). As we are not tied to geographical distances between nodes, the length of links could be used to represent different information other than geographical distance (such as latencies). This would provide the users with an easy to comprehend map showing which nodes exchange information the fastest.

4.2 Visualisation Types

Information visualisations aim to allow a user to digest and comprehend large sets of data using their vision and interaction with the visualisation as cognitive tools [Hegarty, 2011]. Information visualisations should be designed with an understanding of the type of information they plan to visualise. Different visualisation types are more effective or better suited at displaying different types of information or data and for different information retrieval and exploration tasks [Hegarty, 2011]. Hegarty [Hegarty, 2011] highlights this point by discussing how animations - although popular for displaying complex systems - are ineffective. Hegarty [Hegarty, 2011] further discusses how pie graphs - although popularly sidelined as an effective visualisation - have demonstrated that they are better suited at complex comparison tasks to the bar graph which excels at simple judgements [Hegarty, 2011].

4.3 Graphs

Graph visualisations are of particular interest for displaying the network structure and traffic flow information that relates to Internet-scale networks. Graphs are well suited to display the entity and relationship information of networks which are collections of connected nodes [Eick, 1996].

Brath and Jonker explain how to utilise different *layout types*, *visual attributes* and *interaction* to display information to overcome the three most common problems with graph visualisations: *Display Clutter* - visualisations that are dense with nodes and edges become confusing*; Node Positioning* - different node positioning algorithms can dramatically affect the interpretations and insights of the data; and *perceptual tension* - which is the paradox of how the proximity of nodes strongly communicates connection between them but distantly positioned and connected nodes dominate the visualisation since the edge takes up more of the visualisation space [Eick, 1996].

4.4 Graph Layouts

Graph layouts lay the spatial foundation of the visualisation, setting the template for how graph nodes are positioned relative to each other. Different layout techniques are better at overcoming different visual challenges (such as highly connected graphs or dense clusters of nodes) or highlighting particular characteristics of the graph (edge weights, hierarchies, etc.) [Brath et al, 2015].

Some commonly used layouts are force-directed layouts, node-only layouts, time oriented layouts, top-down layouts, radial, maps, chord diagrams, adjacency matrices, treemaps, hierarchy pie diagrams and parallel coordinate graphs [Brath et al, 2015].

4.5 Force-directed

Force-directed layouts utilise "forces" which are factors or parameters such as attraction, repulsion, gravity and edge weight to calculate the relative position of nodes in the graph. Experimenting with these parameters allows the implementer to produce clearer visualisations with the given data [Brath et al, 2015]. Allowing the end user to manipulate the position of nodes or clusters of nodes in the graph gives them the power to make the visualisation clearer for their individual dataset. Although force-directed graphs provide clearer visualisations of large graphs, they are not naturally suited for displaying time data [Brath et al, 2015].

Force-directed layouts use iterative algorithms to position the nodes of the graph. Fruchterman Reingold, Force Atlas and Yifan Hu and examples of common algorithms. These algorithms offer a speed vs quality compromise where the more iterations allowed to the algorithm increases the space-usage of the visualisation but also increases the time taken to generate the visualization [Brath et al, 2015]. This is of key concern in any user interface where the user's experience is negatively impacted by slow feedback [Shneiderman et al, 1996].

4.6 Node-only

Node-only layouts attempt to overcome saturation of the visualisation by edges where the graph is highly connected (each node is connected to a large subset or all of the other nodes). It does so by not displaying edges and successfully avoids clutter and occlusion of nodes by edges. This type of layout substitutes the extra channel of data visualisation for simplicity and clarity [Brath et al, 2015].

Visualisations better suited for displaying hierarchies include **tree-diagrams**, **treemaps**, **hierarchy pie diagrams** (**sunburst diagrams**), **top-down layouts** and **radial graph diagrams**. Key concerns when producing tree layouts is the the breadth and depth of the tree. Wide trees are better visualised in a left-to-right fashion while narrow and deep trees are better visualised with top-down layouts [Brath et al, 2015]. This is due to the space taken up by the node labels. If the child-parent-sibling relationships are what you want to communicate, the tree-diagrams are best suited for that. If instead you want to display the relative size of sub-trees or siblings, treemaps and hierarchy pie diagrams are better suited [Johnson et al, 1991].

4.7 Chord Diagrams and Adjacency Matrices

Both chord diagrams and adjacency matrices are well suited for displaying flow information between nodes. The chord diagram displays flow through the width of the edge between two nodes. The edge can be effectively used to display bi-directional information by changing the width of the edge along its length. The width of the edge at the node indicates the outgoing volume of flow (or edge weight).

The adjacency matrix is effective in visualising highly connected graphs as it displays the accurate quantities with the label of each cell in the matrix. However, the size of the matrix grows exponentially with each additional node, indicating that it is best suited for small highly connected graphs [Brath et al, 2015]. A heat map colouring of the cells can provide an effective and quick understanding of the relationships in the graph.

Maps are effective at displaying spatial graphs since they allow the user to identify nodes if they are familiar with the location of the entity that the node represents [Brath et al, 2015]. Map graphs present some challenges including node occlusion and dense clusters within the visualisation. One could choose to visualise a network topology encoded with geographical information. This places the nodes on a geographical map and the length of the links represents the distance between the nodes' geographical locations. This representation allows the user to identify the location of nodes and the presence or absence of links between these locations [Brath et al, 2015].

4.8 Design considerations
A variety of **visual channels and visual attributes** can be used to communicate different dimensions of quantitative and categorical information. These channels or attributes include size, colour, shape, texture and position. Each channel is usable in different situations and are better suited to displaying certain types of information.

Size is an effective channel to display quantitative information [Brath et al, 2015]. Colours, brightness and hues can also be used to show quantitative information but users are less able to retrieve accurate quantitative information.

Colour and shape can be used to show categorical information very effectively since humans are good at identifying changes in colour and shape [Brath et al, 2015].

5. INTERACTIVITY TECHNIQUES

Shneiderman reminds us that interacting with an information visualisation is key to enabling the user to perform their information retrieval tasks. Shneiderman's Information-Seeking Mantra "Overview first, zoom and filter, then details-on-demand" reflects this importance of interaction for visualisations [Shneiderman et al, 1996].

An overview of the visualisation provides the user with the "big-picture" understanding of the information. This overview needs to clearly display the information with focus on the information needs of the user. The overview also acts as the starting point from which the user begins further information seeking activity. The overview therefore needs to give enough specific detail through labelling or legends to allow the user to direct their exploration of the data.

Zoom and filter are very important in allowing the user to focus on the details and data that interest them and ignoring or hiding the details that do not. Zooming focuses on a smaller window of the visualisation and brings more fine-grained information into focus which may have been hidden in the overview. It is important that zooming is an intuitive interaction and that the zoom is not too fast which will disorientate the user [Shneiderman et al, 1996].

Filtering hides the data that is not of interest to the user. This is a form of information query as the act of filtering out information is similar to retrieving data that matches a search criteria. Filtering is commonly implemented through value sliders for different dimensions of the data or checkboxes that – when unchecked – hide some subset of the data. Advanced filtering techniques allow the user to select some element of the data in the visualisation and the visualisation filters out all other data elements except those that are similar (in some attribute or another) to the selected element [Brath et al, 2015].

Details-on-demand allow the user to answer the questions like "which element is this?" and "what are the values of this element's attributes?". Usually implemented with a pop-up box that presents itself after an element is clicked on, details-on-demand give the user the specifics to satisfy their information need for fine-grained detail.

6. EVALUATION TECHNIQUES

Information visualisations are not easily evaluated. The reductionist thinking of controlled and isolated experiments used to identify the impact of variable independent variables on the measured dependant variables is not as useful for evaluating information visualisations [Shneiderman et al, 2006]. Visualisations are tools used in the creative activity of discovery and to observe users using the visualisation in environments that are foreign to their normal work environments would not produce realistic observations. Users regularly collaborate with other users, look at the same data on many different occasions and formulate and answer questions they didn't originally have. Furthermore, engagement between the observer/developer and the user are key to improving the tool and in helping the domain-expert become an expert user of the tool. This approach is somewhat controversial since some believe that those studying the culture around or human interaction with the system should refrain from interfering with the users in case they influence the situation to the extent that the observations are not reflective of normal usage of the tool.

Shneiderman and Plaisant advise that evaluations of information visualisations take the form of hands-on and intensely interactive engagements and observations of the users in their normal work environment. Their recommended approach is the 'Multi-dimensional In-depth Long-term Case Study' (MILCs) approach to evaluating information visualisations. In their paper [Shneiderman et al, 2006], they also recommend following a guideline for ethnographic observations to minimize the likelihood of the observations going wrong and causing the experiment to fail. The guideline involves detailed preparation work, intensive observation through collection and documentation, accurate analysis of the data and appropriate reporting. The guideline is detailed within the paper.

An alternative approach used in the Visual Analytics VAST contest creates a synthetic dataset where the 'ground truth' is known in advance and the insights developed with the aid of the visualisation can then be objectively evaluated [Shneiderman et al, 2006]. Clearly this approach would involve an elaborate development of a large enough dataset that sufficiently hides the important insights well enough to test the visualisation. It is also vulnerable to a bias in that when the data is created (not measured from observed phenomena), it may not faithfully reflect real-world situations.

7. CONCLUSIONS

There is plenty of literature available discussing the cognitive tools of information visualisations, their attributes, characteristics, strengths and flaws. Furthermore, the growing popularity of information visualisations for scientific and business purposes has created a variety of novel visualisation types. These can be used to guide the design of multiple unique prototypes for a user evaluation. The methods of evaluation of information visualisations are not yet well established but Shneiderman has described the strengths of the intensely interactive and more realistic evaluation approach of Multi-Dimensional In-depth Long-term Case studies which shows promise. From the literature explored, the best type of visualisation for communicating the linked nature of networks, their structure and traffic flow information is a graph (link and node) visualisation. Graph visualisations present many novel approaches to displaying network information – tree maps, node-only, force-directed and hierarchical tree-diagrams each can present information in a manner useful to network managers and organisations policy makers – providing a tool for the insight required to identify problems and plan changes. Exploration of these graph layout styles will help overcome the design challenges of crowded and noisey visualisations often faced in attempts to visualise of large networks with hundreds of nodes and thousands of links between them.

REFERENCES

1. Brath, R., Jonker, D. Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data. John Wiley & Sons, Inc., Indianapolis, 2015.
2. Chavula, J., Feamster, N., Bagula, A., & Suleman, H. (2014). Quantifying the Effects of Circuitous Routes on the Latency of Intra-Africa Internet Traffic: A Study of Research and Education Networks.
3. Donnet, B., & Friedman, T., Internet topology discovery: a survey. Communications Surveys & Tutorials, IEEE Communications Surveys & Tutorials, 9(4), 56-69.
4. Eick, S.G. Aspects of Network Visualisation. Computer Graphics and Applications, IEEE, 16(2), 69-72.
5. Gilmore, J. S., Huysamen, N. F., & Krzesinski, A. E., Mapping the african internet. in *Proceedings Southern African Telecommunication Networks and Applications Conference (SATNAC),* (Mauritius, 2007)*,* South African National Research Foundation.
6. Goodall, J. R., Lutters, W. G., Rheingans, P., & Komlodi, A. (2005, October). Preserving the big picture: Visual network traffic analysis with tnv. InVisualization for Computer Security, 2005.(VizSEC 05). IEEE Workshop on (pp. 47-54). IEEE.
7. Gupta, A., Calder, M., Feamster, N., Chetty, M., Calandro, E., & Katz-Bassett, E., Peering at the Internet's frontier: A first look at ISP interconnectivity in Africa in *Passive and Active Measurement,* (Los Angeles, 2014), Springer International Publishing, 204-213
8. Heer, J., Bostock, M., Ogievetsky, V. A Tour Through the Visualisation Zoo. Communications of the ACM, 53(6), 59-67.
9. Hegarty, M. The Cognitive Science of Visual-Spatial Displays: Implications for Design. Topics in Cognitive Science, 3(3), 446-474.
10. Johnson, B., Shneiderman, B. Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures in Visualisation '91, Proceedings, IEEE Conference on, (San Diego, 1991), IEEE, 284-291.
11. Koutsofios, E. E., North, S. C., Truscott, R., & Keim, D. A., Visualizing large-scale telecommunication networks and services. in *Visualization'99. Proceedings* (San Francisco, 1999)., IEEE, 457-461.
12. Fryer, T., Research and Education Networks around the World and their Use. in *Segunda Conferencia de Directores de Tecnología de Información,* (Lima, 2012), Red Clara, 1-13.
13. Luckie, M., Hyun, Y., & Huffaker, B., Traceroute probe method and forward IP path inference. in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, (Vouliagmeni, 2008), ACM, 311-324.
14. Mao, Z. M., Rexford, J., Wang, J., & Katz, R. H., Towards an accurate AS-level traceroute tool. in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications,* (Karlsruhe, 2003), ACM, 365-378.
15. Montamedi, R., Rejaie, R., Willinger, W., A Survey of Techniques for Internet Topology Discovery, *IEEE Communications Surveys & Tutorials, PP* (99). 1-25.
16. Shneiderman, B., The eyes have it: A task by data type taxonomy for information visualizations. in Visual Languages, 1996. Proceedings., IEEE Symposium on (Boulder, 1996), IEEE,336-343.
17. Shneiderman, B., Plaisant, C. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies in *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, (BELIV, 2006), ACM, 1-7.