UNIVERSITY OF CAPE TOWN

DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours Project
Final Paper 2023

Title: Mapping African Internet Topology

Author: Houston Yu

Project Abbreviation: MAIT

Supervisor(s): Josiah Chavula

Category	*Min*	*Max*	Chosen
Requirement Analysis and Design	*0*	*20*	20
Theoretical Analysis	*0*	*25*	0
Experiment Design and Execution	*0*	*20*	0
System Development and Implementation	*0*	*20*	20
Results, Findings and Conclusions	*10*	*20*	10
Aim Formulation and Background Work	*10*	*15*	10
Quality of Paper Writing and Presentation	*10*		10
Quality of Deliverables	*10*		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	*0*	*10*	
Total marks		**80**	

Mapping African Internet Topology

Houston Yu
Department of Computer Science
University of Cape Town
Cape Town, South Africa
yxxcha002@myuct.ac.za

ABSTRACT

Previous research on visualization tools do not effectively communicate the interconnections in the African internet topology. This is mainly due to the complicated nature of the existing visualization tools and users find it difficult to use these tools, which has resulted in users becoming less interested in the existing visualization tools.

In this paper, we presented the design of a network topology visualization, which visualizes the African internet at AS-level. Such a visualization allows users to identify autonomous systems, internet exchange points and the relationships between them.

We assessed the usability and accuracy of the visualization at communicating the AS-level internet topology of Africa. Usability tests were conducted to evaluate the design of the visualization in terms of usability and accuracy. The results of usability tests are acceptable which indicates that the visualization tool has good usability.

Keywords

Autonomous System, Internet Exchange Point, Internet Service Providers, Topology Visualization, Border Gateway Protocol

1. Introduction

Over the past few years, the internet has continued to evolve and expand, which is why it is extremely important to understand the underlying infrastructures that support it. In order to understand the complexity of the internet, mapping the logical and physical topology of internet infrastructures and visualizing them can provide valuable information into the network's performance and potential bottlenecks [1].

1.1 Problem Statement

Although various visualization platforms are constantly being developed and refined for internet topology visualization, there is still a need for further research into the internet topology of Africa. This is mainly due to little research being done in visualizations tools specifically for the African Internet topology. The tools that have been described in previous research do not effectively communicate the interconnections in the African internet topology. Mapping the internet topology in Africa presents unique challenges due to Africa's limited internet infrastructures and diverse geography. [2]. Therefore, it is necessary to collect BGP datasets and traceroute measurements for logical topology discovery and use it to infer and visualize the relationships between networks and countries across Africa [2][3].

By visualizing the relationships between networks in Africa; network experts or users who are interested in internet networks can learn more about their ISPs. Such a visualization would be useful in identifying what kind of relationships do their ISPs have with other ISPs and at which internet exchange point (IXP) do they exchange data with each other.

1.2 Research Aim

Many researchers have tried to use various techniques and datasets such as Border Gateway Protocol (BGP) datasets and traceroute measurements to map internet topology and visualize the relationships between internet service providers (ISPs) across the world [1]. Despite various approaches to mapping internet topology, the visualization platforms that have been developed in previous research are difficult to use and users often struggle to understand the data that are being visualized due to its complicated nature [3]. One of the possible reasons for this is most of the visualization tools focus more on how appealing the visualizations should be instead of presenting useful information to the viewers. Another possible reason is that some visualization tools lack essential features such as allowing the viewers to filter and find what they want to see [3]. These flaws have either made the visualization tools to become less useful or difficult to comprehend.

The purpose of visualization tools is to present information in a visual and interactive manner [19]. The visualizations should allow users to extract useful information easily such as identifying the ISPs and the relationships between ISPs. Such visualizations should also be simple and easy to use, especially for users who have never used any visualization tools before [6][19].

The following research question has been formulated for this project: How can the visualization of interconnections between ISPs effectively communicate the AS level internet topology of Africa, allowing users to identify ISPs, exchange points and peering links, and how the visualization is able to improve its usability and accuracy to extract useful information?

In this paper, we present the design of a network topology visualization, which visualizes the African internet at AS-level using Border gateway protocol (BGP) and IXP datasets collected from the Center for Applied Internet Data Analysis (CAIDA) and Packet Clearing House (PCH) public repositories. At AS-level, autonomous systems (AS), also called ISPs, and IXPs are being shown in the visualization.

We have evaluated the usability and accuracy of the visualization by conducting usability tests to determine its ability to convey information about ISPs and the type of relationships they have with other ISPs.

2. Literature Review

2.1 Background and Related Work

In the paper by Gilmore et al. [2007], data collected from traceroute measurements and BGP datasets were used to generate router-level and AS-level maps of the African internet. At the AS-level, Walrus tool created by CAIDA was used to display AS level structure of the African internet topology in a 3D space. At the router-level, a java-based tool, called Terrapix, was created specifically to produce router level maps, where 2D and 3D visualizations were used to map nodes and links to geographic locations [3].

Yang et al. [2016] introduced a topology visualization tool called NetVizura to visualize the Network topology of National research and Education Networks (NREN) in Africa using traceroute data [4]. This work adopted a user centered design approach where usability tests were conducted by recruiting human participants to assess the accuracy and the effectiveness of the visualization [4]. The visualization shows traceroute information of all IP addresses to which traceroutes were sent on a map. The design of visualization also allows multiple traceroutes to be sent from different vantage points which can be viewed on a map. It also displays the locations of the internet exchange points and the placement of fibre around Africa continent [4].

In a paper by Periakaruppan and Nemeth [1999], they have created a graphical traceroute tool. called GTrace, which is used to discover routing loops and using traceroute information to help decide how the route should be implemented. The traceroute tool uses several heuristics to determine the location of a node and it also executes traceroute before representing traceroute information on a world map as a series of nodes and links. The traceroute tool also displays more detailed information about the traceroute such as the hop IP address, hop host name and the hop number [5].

2.2 Design guidelines for visualizations

The Visual Information-seeking Mantra by Schneiderman [1996] is often seen as a starting point for information visualization applications. Schneiderman explains 4 major principles in his paper: The first principle is overview ("gain an overview of the entire collection"). The second principle is zoom ("zoom in on items of interests"). The third principle is filter ("filter out uninteresting items"). The last one is details-on-demand ("select an item or group and get details when needed") [7].

2.3 Representation of visualizations

Topology visualization is one of the most common ways of presenting computer networks and this visualization can be presented using a graph representation, which consists of nodes and links (edges) [6][18]. In order to achieve a representative of visualizations, Hatch et al. [2001] discuss what properties are desirable in terms of the representative design of visualizations. For example, one of the desirable properties is low visual complexity with high information content [17]. However, there are challenges when visualizing such data because there are so many nodes and links that are connected to each other. This can cause the display of the visualization to be cluttered and it can also lead to occlusion of nodes or links [6]. One of the solutions for this issue is to aggregate the nodes and links together. Although this may solve the issue, another problem arises from this solution. Beker et al. [1995] believes by doing so will leave out important information. To solve this problem, the visualizations should have an interactive feature to filter certain data [6].

Interactivity is particularly important when representing visualizations of network topology. This is because interactivity helps to address issues where large amounts of clustered points may occur when visualizing large datasets. In the paper by Ellis & Dix's [2007] offers various clutter reduction techniques that can be used to minimize such issues [16].

3. Design and Implementation

3.1 Approach

A User-centered Design approach is being adapted to ensure that the users are involved throughout the design process. This approach ensures that the final visualization product meets the users' needs and requirements [8]. Users of the visualization can be network operators, network administrators and researchers from various ISPs.

The design process is split into three phases: The early envisioning phase, the global specification phase and the detailed specification phase [9].

During the early envisioning phase, requirements gathering is done to identify the needs of users and an analysis of the problem that the visualization tool is trying to solve [8][9]. The main purpose of this phase is to identify the initial visual queries (an information need addressed by a visualization) that the potential users may have.

During the global and detailed specification phase, iterative tasks of analysis, design and evaluation take place. Once the analysis has been done, various solutions can be proposed and presented to potential users based on the results gathered from analysis [9]. The solutions that are being proposed can come in the form of a prototype. There are different types of prototypes such as low-fidelity prototype (paper prototype) and high-fidelity prototype (evolutionary prototype) [9]. After a prototype has been created, it is also necessary to evaluate it. The evaluation can be achieved through various methods such as questionnaires and usability tests [9].

3.2 Early envisioning phase

This phase is the first iteration of the User-centered design approach. The main goal of this phase is to gain an understanding of potential users, what tasks they will perform and the environment in which they will perform these tasks [9]. Using such information can help guide the final design of the visualization tool.

3.2.1 Analysis and Requirements Gathering

In order to identify user tasks and goals, we had weekly meetings with our supervisor, Dr. Josiah Chavula, to discuss what initial features should be expected in our visualization tools to meet the potential user needs. This can help to ensure that the design of our visualization tools is in the right direction even if there aren't any users who we can ask what their real needs are during the early iteration of the design cycle.

The weekly meetings took place in our supervisor's office at the University of Cape Town. During the weekly meetings, our supervisor provided us with some useful datasets or recommended certain public repositories where data is stored. These datasets can be used in our visualization tool. Most of our questions regarding the features that would be included in our visualization tool were answered by our supervisor. Based on the requirements we have gathered; we need a visualization that provides users an overview of ISPs (AS) and their relationships between each other in Africa.

3.2.2 Implementation

The initial prototype was developed using a static map instead of a dynamic map where users can interact with it.

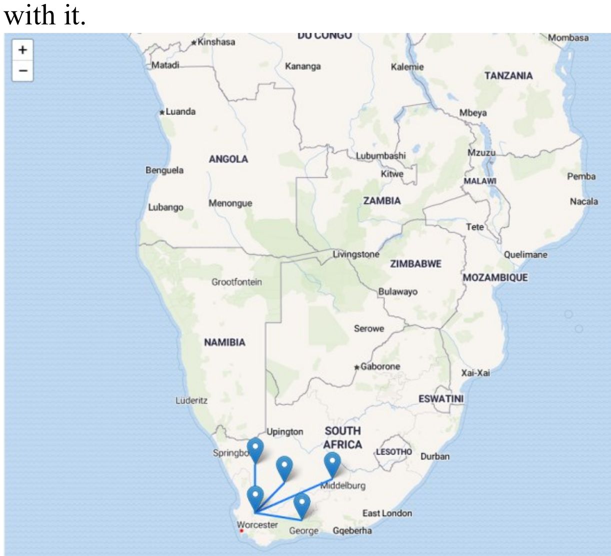

Figure 1. Screenshot of the initial prototype

The initial prototype (Figure 1) was created with limited functionality using Leaflet Javascript library and a small sample of dummy data. The markers represent ASes and the lines connecting the markers are the relationships that each AS may have with each other. Users can click on the markers to see the Autonomous System number (ASN).

3.2.3 Evaluation

In order to get feedback on the initial visualization of our prototype, an informal user feedback evaluation was conducted with our supervisor.

During the feedback session, our supervisor suggested a more interactive map where users can click on a specific country and see all the AS relationships for that country. It is also noted that the locations of ASes are not important since an AS has a large coverage area (IP addresses appear in more than one location), which is nearly impossible to pin the exact locations on a map.

3.3 Global and Detailed specification phase

In these phases, solutions in the form of an improved version of the initial prototype are proposed to our supervisor [9].

3.2.1 Analysis and Requirements Gathering

We had another session in one of our weekly meetings with our supervisor where we try to identify the features that are still missing due to the lack of requirements gathered in the previous iteration. Rather than asking the same questions in our previous meetings, we decided to let our supervisor point out what features we are still missing or provide us with additional ideas and comments.

During the session, it was discovered that potential users may be interested in what type of relationships each AS may have with each other. Another requirement is displaying the number of customers for each AS, also known as customer cone (a set of ASes that an AS can reach using peer-to-customer links) [10]. For these requirements, more details about each AS should be shown in some ways to support ASes that have already been represented by the visualization.

3.2.2 Implementation

In order to improve the previous iteration's prototype, it was necessary to obtain map data for Africa Continent which comes in the form of geojson file. The source of the map data is from a public repository. An interactive map is generated using the freely available geojson file with the addition of d3 javascript library, which is a powerful tool for creating visualizations. A webpage for the visualization tool was created using HTML and javascripts.

MySQL database was used to store data (ASN, country code) for each individual AS and their neighbors (type of relationships). The data would be converted into json files which were used to visualize the network topology for each country.

Users would be able to see a topology visualization for the country that they have clicked on as shown in figure 2. The topology visualization is also created using d3 javascript library. The nodes represent ISPs or ASes and the links connecting them represent the type of relationship. In this case, the blue links or links are the peer-to-customer (p2c) relationships, and the red lines are peer-to-peer relationships (p2p). There is also another hidden colour (grey) which represents unknown relationships.

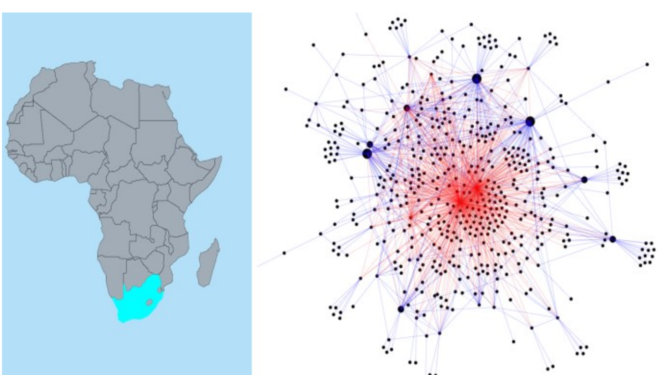

Figure 2. Screenshot of the improved prototype

3.2.3 Evaluation

Another feedback evaluation was conducted with our supervisor during this iteration. Major changes had been made to the previous iteration's prototype since the static map has been completely removed and replaced with an interactive map.

After our supervisor performed simple tasks using the visualization, we received insightful feedback from our supervisor on what he likes the most and least about the visualization. At the end of the evaluation session, our supervisor wants more filtering features where users are to search for what they are looking for.

3.4 Final Visualization Design

Figure 3a and 3b show the final visualization design. Based on the feedback received in the second iteration, users feel confused about the graph (what do the nodes represent and what are the meanings of the colours) and they did not know about the interactive features. This was addressed by adding popover buttons which will provide them with a brief explanation of the graph and what interactive features are available. Users also complained about feeling a bit lost when looking at the graph (Which AS or IXP they are looking at). This was resolved by adding a navigation scheme, known as "breadcrumb", that reveals the type of relationship and the name of AS or IXP the users are looking at. By adding a navigation scheme, users are able to navigate through the visualization more easily.

In addition to the AS relationship, another type of relationship has been added as shown in figure 3b. This relationship contains Internet Exchange Points (IXPs), where ISPs exchange data with each other at the same IXPs. The data was obtained from the IXP directory managed by Packet Clearing House (PCH) [11]. The data only includes active IXPs in Africa.

Schneiderman's four major principles for data visualization (Overview, zoom, filter and details on demand) were applied when designing the user interface for the visualization tool. Users are first presented with an interactive map of Africa. Users can hover onto one of the countries to see its name if they don't where a country is located. After clicking on a country, they would be able to see a topology visualization for that specific country as shown in figure 3a. The visualization can be switched between AS and IXP relationships by clicking on one the labeled buttons at the top. The nodes are draggable which allows users to see certain nodes clearer when nodes are overlapping. They can also hover onto the nodes to see the names of nodes. Users can also hide and show the links connecting the nodes by toggling the switch buttons at the top left if they just want to see a certain type of relationship. There are two radio buttons at the top right which display the size of nodes based on the option that the users select. Next to the topology visualization is a list of AS names and the numbers next to the names are the AS numbers (ASN). There are two modes: Zoom mode and Relationship mode. In Zoom mode, Users can click on one of names to zoom into the node that corresponds to the AS name. In Relationship mode, users must select two AS from the list to check the relationship between the two AS which will be displayed in a text box. There are also zoom buttons below which allow users to zoom in and out.

In terms of filtering, users can double-click on a node to filter out the ASes that are not connecting to the selected node (AS) and only display the ones connecting to it as shown in figure. Another filtering feature is the search box, which searches for an AS by typing in the name of AS. These filtering features also apply to the IXP visualization.

Different colours were used to represent different types of data in the visualization. Blue nodes represent IXPs, and black nodes represent ASes. Similarly with the link colours (section 3.2.2). Colours can also be used to highlight certain nodes to make them stand out which allows users to identify the nodes easily. Nodes become lime to indicate that the node is being dragged or zoomed into.

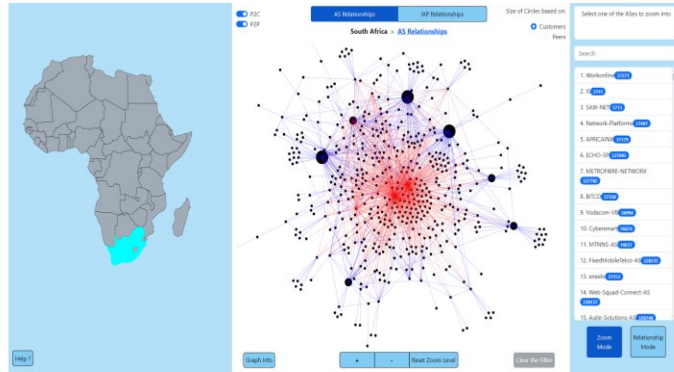

Figure 3a. AS Relationship Screen

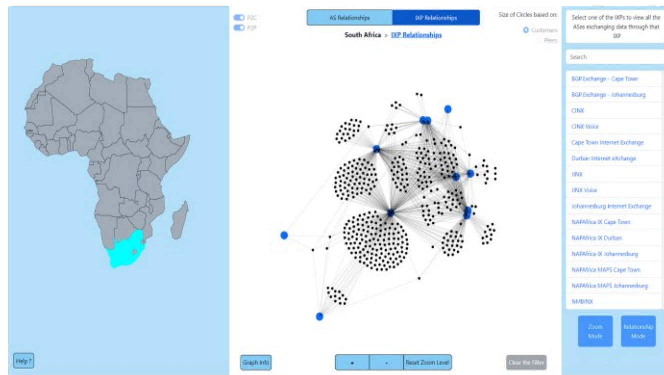

Figure 3b. IXP Relationship Screen

4. Final Evaluation and Results

4.1 Evaluation Metrics

The evaluation methods consist of usability and accuracy. Users' feedback of the visualization tool was taken into account and used to determine the usability and accuracy of the visualization tool.

4.1.1 Accuracy

Accuracy refers to the ability of the visualization tool to convey correct information to the users. This means whether users can obtain the correct answers to their questions that are related to visual queries.

4.1.2 Usability

Usability refers to the ease with which a user can accomplish the given tasks when interacting with the visualization tool [12]. During usability tests, users were given a set of predefined tasks and they would be observed when accomplishing these tasks. Performance measurements such as task completion and task success rate were collected when they are interacting with the visualization tool. Usability tests can help to discover design flaws in prototypes. And these flaws can be addressed in the final design of the prototype [12].

Users' satisfaction and personal opinions were also collected during usability tests using a questionnaire. The questionnaire consists of various questions that aim to collect some demographic information on the users, their opinions on the visualization tool, and there's also a scale where users can rate the visualization tool. The scale can be used as quantitative measurements while users' opinions are used as qualitative measurements. These measurements can therefore be used to evaluate the usability of the visualization tool.

4.2 Usability Tests

In order to assess the usability and accuracy of the visualization tool, the tests were conducted with a total of 14 users. Measurements that were taken during the usability tests were the successful task completion (accuracy) and the ease of use of the visualization tool (usability).

Ethical clearance was obtained from the Science Faculty Research Ethics Committee. Before participating in the usability tests, users were asked to sign a consent form to inform them that their results would be kept anonymous. Users were given snacks as a reward for participating in the tests.

Users were split into two groups. The first user group consists of novice users who have little or no experience in computer networking. The novice users are a mixture of undergraduate and fourth year Honours Computer Science students. All the novice users were identified and recruited from the Computer Science Seniors Laboratory. The second user group consists of expert users who specialise in the field of networking and most likely resemble the actual users (network operators and network administrators). There is only one expert user who is a postgraduate Computer Science student and specializes in computer networking.

The usability tests were conducted in the Computer Science Seniors Laboratory. Participants accessed the visualization tool through a webpage, using either their personal laptop or a laboratory computer. As the visualization requires a basic knowledge of computer network, a brief explanation of the visualization would be provided to those who don't know much about internet.

After the explanation. users would then perform a set of instructed tasks to test whether they can accomplish the given tasks successfully. Rather than assisting users throughout the whole test process, users were allowed to answer the questions after completing the given tasks. This is to avoid spending too much time with each user.

4.2.1 Task Design

There are two objectives when setting the tasks for the usability test: The first one is to test whether users could answer visual queries correctly (addresses the accuracy), and the second one is to determine if users could use the interactive functionality of the visualization easily (addresses the usability).

For the first objective, the visual queries were questions related to identifying ISPs (ASes) and IXPs, the type of relationships and the number of customers. Here is a list of visual queries that users can answer: "Which node is ISP and IXP", Which AS has the most or least customers?", and "What type of relationship does ISP A has with ISP B?".

The second objective focuses on several tasks that aim to test whether users can successfully complete a combination of high-level subtasks of overview, zoom, filter and details on demand as described by Schneiderman [1996]. These tasks can be classified as *locate* tasks which are related to searching and finding information that have already been visualized [12][13].

4.2.2 Questionnaire

After completing the instructed tasks, users would be asked to fill in a questionnaire to determine the usability of the visualization tool. The questionnaire consists of open-ended questions (users' opinions and impressions) to evaluate user experience.

The questionnaire also has rating scales where users can give a rating, which ranges from 1(bad) to 5 (good). The rating scales provide a measure of usability, but it cannot tell us about the reasons behind the rating. This is why open-ended questions were added to the questionnaire.

4.3 Analysis of Results

4.3.1 Visual Query and Task completion

The accuracy of visual queries is determined by whether participants can answer questions related to the visual queries correctly. If participants are able to answer the questions correctly, they have successfully performed the given tasks.

In figure 4, results showed that participants were able to determine which node is AS or IXP in the topology visualization with 100% accuracy. Another visual query achieved 100% accuracy is identifying the type of relationships each AS may have. However, the last visual query was answered with less accuracy which only had 88.8% accuracy.

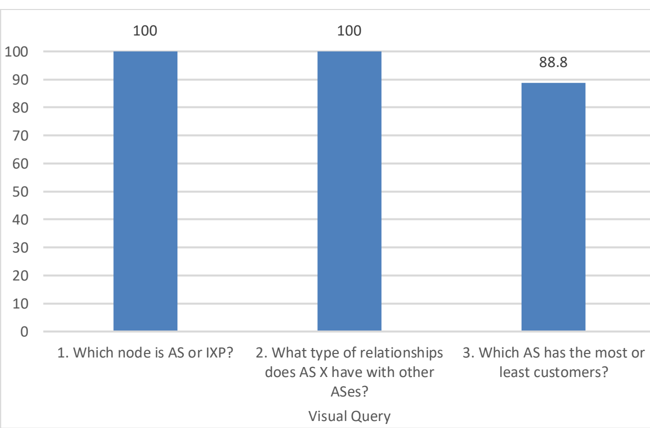

Figure 4. Accuracy of visual queries answered

All the participants were able to answer the first two visual queries (Visual query 1 and 2) without any issues. This is because the answers to visual query 1 and 2 focus on the use of colours in the visualization. For visual query 1, the nodes in the IXP visualization (figure 3) have two colours: Black represents AS and blue represents IXP. Similarly, for visual query 2, the lines or links connecting the nodes have 3 different colours: Blue for c2p relationship, red for p2p relationship, and grey for unknown relationships. Once the meaning of the colours were explained to the participants, they were therefore able to answer the first two visual queries easily due to the apparent contrast between the colours.

A possible reason for the diminished accuracy for the last visual query (Visual query 3) is the difficulty of determining the number of customers for a particular AS. In order to display the number of customers for each AS, the sizes of nodes are used as an indication to how many customers each AS has. The bigger the nodes the more customers it has and the smaller the nodes the less customers it has. Although the sizes of the nodes can be distinguished, participants still struggle to tell which AS has the most or least customers since the sizes of nodes all look quite similar. This problem can be resolved if participants zoom into the nodes to view their exact sizes, but this is unintuitive. This indicates a flaw in the visualization where the sizes of nodes are not apparent and should be addressed to make it more noticeable.

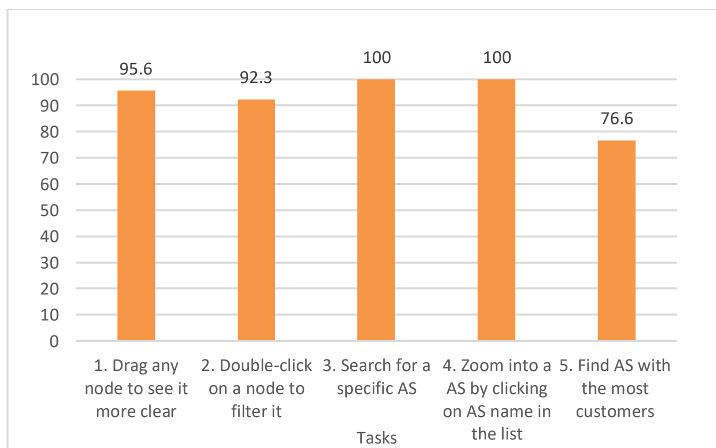

Figure 5. Successful tasks completed

Figure 5 shows that task 3 (Search for a specific AS) and task 4 (Zoom into AS by clicking on AS name in the list) had 100% successful task completion while task 1 and 2 nearly achieved 100% successful task completion. In comparison to these tasks, task 5 has the lowest completion percentage. This is understandable since task 5 needs to answer visual query 3 in figure 4.

Tasks 3 and 4 were easy to perform due to the simple tasks that each participant needed to perform. To search for a specific AS, participants must search it by typing the name of AS in the search box. Once they found what the AS they had been looking for, they can click on the name of AS in the list which would automatically zoom into the node that corresponds to the name of AS. Participants can therefore perform tasks 3 and 4 successfully just by searching and clicking.

The reasons for the lower completion percentage for tasks 1 and 2 are interactivity issues. These tasks require participants to interact with the nodes in the visualization. The links or lines connecting the nodes are preventing participants from dragging or clicking on the nodes. A possible explanation for this is that the links that are connected to the center of nodes instead of the edges. In order to interact with the nodes, participants must click directly on the nodes and not the links. Therefore, Participants sometimes couldn't perform these tasks successfully because the links were overlaying with the nodes.

In task 5, participants were tasked with finding the AS with most customers (largest node). This resulted in a few responses where participants struggled to find AS with most customers. After analyzing the results, it was discovered that participants who were able to perform task 5 successfully relied on the list next to the topology visualization. The list displays the number of customers along with AS names. If participants click on AS with the most customers in the list, they could automatically find it in visualization. However, a few participants failed to realize this and only looked at the size of the nodes. This is understandable since the number next to AS names in the list is not clear to what it means. This demonstrates an error in the design of visualization, and it can be resolved by providing users with an explanation to what the number means.

4.3.2 Rating Scale

There are two rating scales: The first one asks the participants how frequently they would use the visualization tool. The second one asks for participants' impressions of the visualization tool.

Figure 6 displays the scores for the first rating scale (how frequently the users would use the visualization tool). Figure 7 presents the scores for the second rating scale (overall impression of the visualization tool). In both figures, the horizontal axis represents the scores ranging from 1 (lowest) to 5 (highest). The vertical axis represents the number of users who had given the scores. The scores were collected from 13 users who participated in the usability test and there is another separate usability test conducted by an expert user. The score was not given by the expert user because the expert user decided to give insightful feedback instead of providing a score. Therefore, the expert user is excluded and only 13 participants' scores were used to evaluate user experience.

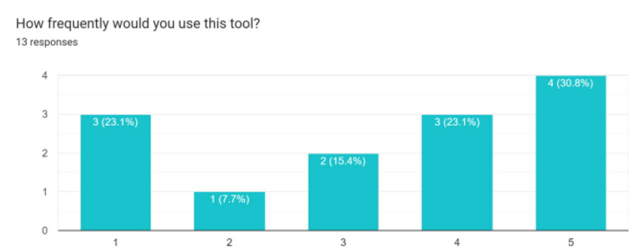

Figure 6. Score for the first rating scale

In figure 6, the scores were almost evenly distributed. Participants who gave the highest score (5) are the ones who are likely to use the visualization tool for exploring internet in their home countries or they are using it for enjoyment. On the other hand, participants who gave the lowest score (1) are likely the ones who are not interested in the visualization tool or they think the visualization will not be useful to them in any ways.

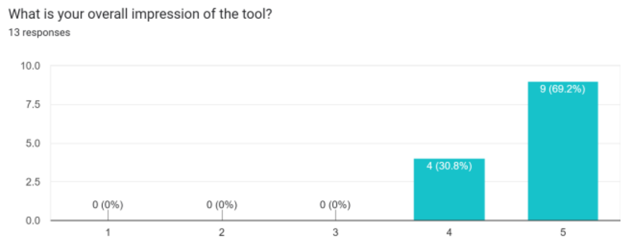

Figure 7. Score for the second rating scale

In figure 7, the scores were more skewed toward to the right side which means most users gave a score of 4 or 5. The possible reasons behind the high scores are: Users find the visualization really fascinating and most users said that the visualization is fun to interact with during the usability tests. Users also find the visualization simple and easy to use. Therefore, these factors contribute to the positive impression of the visualization tool.

4.3.3 User Experience Feedback

Additional open-ended questions were added to the questionnaire to determine user experience and usability problems that participants may experience during the usability tests.

The questions were: “What features did you like the most about the visualization?”, “What issues do you have with the visualization tool?” and “What features would you add to the visualization?”.

In order to analyse the results for each question, thematic analysis was used to review the text and then code the text [14]. Coding the text means to subdivide and label the text to create encodings [14]. This coding process involves extracting themes directly from the text [14]. Only explicit themes were extracted from the text because some users' feedback was short and difficult to interpret. In addition, only themes with more than one occurrence were extracted due to the limited sample size. The themes extracted from users' responses were grouped into 3 different themes for each question.

In figure 8, most users were able to use the interaction features with ease. Ease of use has proved to be a theme that users identified the most. The responses to the question mentioned that the visualization was easy to use after listening to our explanations. Some users also found the filtering features intuitive to use after using it a few times.

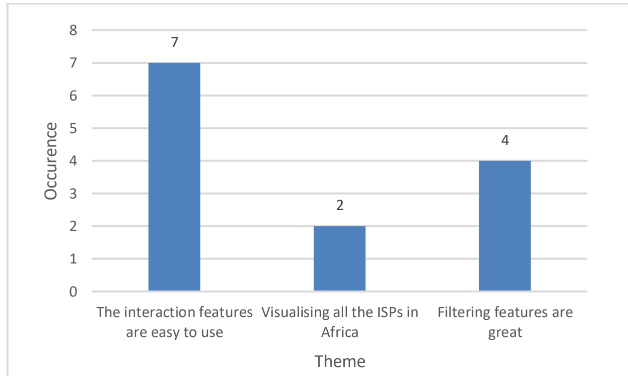

Figure 8. Responses to “What features did you like the most about the visualization?”

In figure 9, it can be seen that the lack of instructions in visualization had caused usability problems which would ruin user's experience with the visualization tool. Users did not know that they could interact with the visualization unless someone explains the interaction features to them.

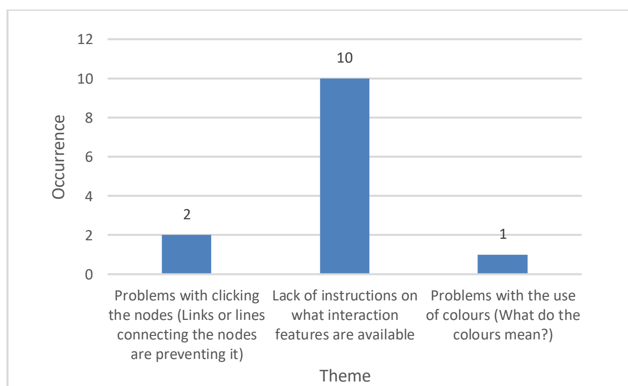

Figure 9. Responses to “What issues do you have with the visualization tool?”

In figure 10, most users wanted to add a feature that would provide them with a clear explanation so that they would know what to do with the visualization. This is understandable since the lack of instructions can cause confusion during the usability tests. As a result, users may miss out on some of the features.

Adding navigation tools could also help to improve user experience because it helps them to keep track of what they were doing while performing certain tasks.

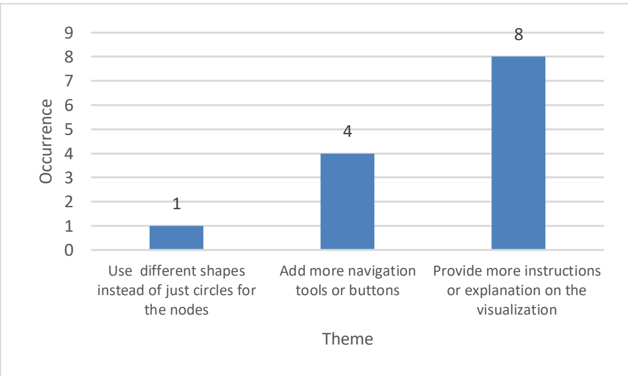

Figure 10. Responses to "What features would you add to the visualization?"

After conducting theme analysis, a few usability issues were raised from these themes. These usability issues can be addressed in the future implementation of the visualization tool. One of the solutions is to apply Human Computer Interaction principles which can assist users in understanding an interactive system.

5. Limitations

5.1 Limited dataset

The datasets that are being used in the visualization have limited data or there are no data at all for certain countries in Africa. This is due to the limitations of the datasets in terms of the amount of usable data. The datasets were fetched from various public repositories such as the Center for Applied Internet Data Analysis (CAIDA) data server and IXP directory owned by Packet Clearing House (PCH) [11][15]. Most data contain useful information (ASN, country code), but some data in the datasets are unusable due to the irrelevant information it provides. In addition, there are instances where data are inconsistent or incomplete and therefore it cannot be used in the visualization.

The visualization tool requires a large amount of data to create an appealing visualization. When there is not enough data, users sometimes are only able to see isolated nodes with no lines (links) connecting it or only a few lines connecting the nodes. Therefore, this can reduce the effectiveness of visualization.

5.2 Usability Testing

When adapting a User-Centered Design approach, it is crucial that users have the appropriate knowledge of computer networks or a high-level understanding of internet. They must also be available during the testing phase of the design cycle to ensure valuable feedback.

Participants consist mostly of novice users who can only perform simple interactions and answer visual queries during the usability test. A limitation of recruiting novice users is that they may not have a clear understanding of what the visualization is trying to achieve (the problem that it is trying to solve). Furthermore, participants may not represent the real users of the visualization (network operators and administrators). Therefore, recruiting an expert user can help to address the needs of real users.

6. Conclusions and future work

In this paper, we used User-Centered Design approach for the design of topology visualization tool in Africa. The final design of visualization was based on feedback from our supervisor, novice users and an expert user. The visualization shows various ISPs (AS) as nodes and the relationships between ISPs as links. The visualization can also show IXP relationships which will display the IXPs and all the ASes exchanging data through the IXPs.

In the final evaluation, we evaluated whether the newly designed visualization tool can accurately communicate the network topology for each country in Africa and allows users to identify the AS that they are looking for. This was evaluated by checking how many users were able to answer the visual queries successfully. These visual queries are related to identifying AS and IXP, the type of relationships and the number of customers each AS may have. It was shown that the visualization was able to communicate the AS with the most customers but not with 100% accuracy due to a design flaw in the visualization. The accuracy of which AS has the most customers can be improved by increasing the size variants of node which will make the nodes (ASes) with different number of customers more visible to users.

After conducting usability tests, the visualization tool was considered partially useful due to the issues raised by novice users. Once the usability issues have been addressed, another separate usability test would need to be conducted with the expert user to determine whether the visualization is presenting the data accurately. After conducting usability test with the expert user, more feature requirements, such as creating another type of visualization, were discovered. This would be useful because users may have different preferences on what type of visualization fits them the best. Therefore, creating a different type of visualization can cater for different user preferences.

Future work may involve creating new types of visualizations or expanding the visualizations to other continents besides Africa.

7. Acknowledgement

Thanks to Josiah Chavula, for his guidance and inputs throughout the project as our project supervisor.

8. References

[1] Kimberly Claffy, Young Hyun, Ken Keys, Marina Fomenkov, and Dmitri Krioukov. 2009. Internet Mapping: From Art to Science. *Cybersecurity Applications & Technology Conference for Homeland Security*. https://doi.org/10.1109/catch.2009.38

[2] Arpit Gupta, Matt Calder, Nick Feamster, Marshini Chetty, Enrico Calandro, and Ethan Katz-Bassett. 2014. Peering at the Internet's Frontier: A First Look at ISP Interconnectivity in Africa. *Passive and Active Measurement*: 204–213. https://doi.org/10.1007/978-3-319- 04918-2_20.

[3] Gilmore, J. S., Huysamen, N. F., & Krzesinski, A. E. 2007. Mapping the african internet. *in Proceedings* Southern African Telecommunication Networks and Applications Conference (SATNAC), (Sept 2007), *Mauritius*, South African National Research Foundation.

[4] Chantal Yang, Hussein Suleman, and Josiah Chavula. 2016. A topology visualisation tool for National Research and Education Networks in Africa. 2016 IST-Africa Week Conference. https://doi.org/10.1109/istafrica.2016.7530605

[5] Periakaruppan, R. and Nemeth, E. 1999. ``{GT}race {A} {G}raphical {T}raceroute {T}ool''. LISA '99: *Proceedings of the 13th USENIX conference on System administration*: 69–78.

[6] R.A. Becker, S.G. Eick, and A.R. Wilks. 1995. Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics 1*, 16-28. DOI:https://doi.org/10.1109/2945.468391

[7] Shneiderman, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations. in *Visual Languages, 1996. Proceedings., IEEE* Symposium on, IEEE, 336-343.

[8] Abras, C., Krichmar D. M., and Preece. J 2004. User-centered design. Bainbridge, W. *Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37, 4: 445–56. http://doi.org/10.1.1.94.381

[9] Valiati, E.R, Pimenta, M.S, and C M D S Freitas. 2006. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. *Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization*, October:1–6. http://doi.org/http://doi.acm.org/10.1145/1168149.1168169

[10] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and Kc Claffy. AS Relationships, Customer Cones, and Validation. CAIDA. Retrieved August 27, 2023 from https://conferences.sigcomm.org/imc/2013/papers/imc039-luckieAemb.pdf

[11] Prefix.pch.net,. Packet Clearing House (PCH) - Internet Exchange Directory. https://prefix.pch.net/applications/ixpdir/.

[12] Valiati, E.R, Pimenta, M.S, and C M D S Freitas. 2006. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. *Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization*, October:1–6. http://doi.org/http://doi.acm.org/10.1145/1168149.1168169

[13] Usability.gov,. Planning a Usability Test. http://www.usability.gov/how-to-and-tools/methods/planning-usability-testing.html.

[14] Carpendale, S. 2008. Evaluating information visualizations. Information Visualization. Springer Berlin Heidelberg, 19–45. http://doi.org/10.1007/978-3- 540-70956-5_2

[15] CAIDA Data server. https://publicdata.caida.org/datasets/as-relationships/

[16] Ellis, G. & Dix, A. 2007. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics* 13, 6: 1216–1223. http://doi.org/10.1109/TVCG.2007.70535

[17] Hatch, A. S., Smith M. P, Taylor, C. M. B., Munro, M., 2001. No silver bullet for software visualisation evaluation. … Issues of Visualization. …, (Cisst), pp.1 – 8. Available at: http://www.dur.ac.uk/malcolm.munro/papers/docs/WFI ofV-2001.pdf.

[18] Winckler, M., Palanque, P and Freitas, C. M.D.S. 2004. Tasks and scenario-based evaluation of information visualization techniques. Proceedings of the 3rd annual conference on Task models and diagrams, October 2015: 165–172. http://doi.org/10.1145/1045446.1045475

[19] Carr, D. A. 1999. Guidelines for designing information visualization applications. Proceedings of ECUE, 99, 1- 3.