CSC/SDS 235: Visual Analytics

Fall 2024

HW 04: Networks

This is a group assignment (3-4 students) – I recommend choosing collaborators with complementary skillsets to yours!

Goals:

- Explore network data via visual analytics
- Work with network (graph) data

Instructions

Overview

Scenario

The year is 2012, and we find ourselves on BankWorld, a planet much like Earth but with a very different geography: one large land mass containing several different nation-states.

The most important organization on BankWorld is the Bank of Money (BOM). BOM has many offices of various sizes across BankWorld. Each of these offices has many computers active throughout the day, with an overall total surpassing 1,000,000 active machines across the entire organization.

The Bank of Money (BOM) Corporate Information Officer (CIO) has asked you to create a situation awareness visualization of the entire enterprise. This is a considerable challenge, given that BOM operates from BankWorld's coast to coast! In addition to observing the global situation, she would also like to be able to detect operational changes outside of the norm, with particular attention to a recent period of network vulnerability.

The Data

To support your analysis, you have been granted access to a dataset documenting activity on one of BOM's subnetworks of approximately 5000 machines:

- Firewall logs [available in both CSV and raw format]
- Intrusion detection system (IDS) logs [available in both CSV and raw format]
- A description of the network

The data is available via Google Drive (with your Smith account). You may use as much or as little of the data in your analysis as you like. For this challenge, it is strongly recommended that you make use of all the available data... it isn't very big compared with previous challenges, but wrangling it may take some effort!

Your Objective

We've recently started to explore techniques for dealing with text-based data, as well as working with connected entities. This challenge will require you to think carefully about how to model (and visualize!) these relationships, which may evolve over time, as well as to identify nefarious patterns in data that seems far removed from human behavior. To help guide your analysis, you may want to consider the following questions:

- What security trends do you notice in the firewall and IDS logs over the course of the two days?

- What noteworthy events took place for the time period covered in the firewall and IDS logs?
- What do you suspect are the root causes of the events you identified?
- Understanding that you cannot shut down the corporate network or disconnect it from the internet, what actions should the network administrators take to mitigate these root issues?

Again, as always: don't worry too much about getting the "right answer" - instead, focus on making sure that the evidence your present should support your hypotheses of what the roles and relationships are, and the motivations of the person(s) involved.

Good luck!

Acknowledgements

This dataset was prepared by Pacific Northwest National Laboratory as part of the VAST Challenge 2012.

Deliverables

You will submit ***four*** deliverables for this assignment:

1. Sketches of the visualization(s) you intended to create.
2. The write up you would present to your supervisor based on your analysis.
3. Code (and a README.txt with instructions for running the code) that generates the visualization(s) in your write up.
4. A reflection (the entire group can write a reflection together, or group members may write individual reflections) that includes:
   - How each group member contributed to the final submission
   - One obstacle you encountered and how you overcame it
   - If you were to do this assignment again, what you would do differently.

Submission

Submit your deliverable(s) in on Gradescope. If you worked on the reflection as a group, submit as a group (https://guides.gradescope.com/hc/en-us/articles/21863861823373-Adding-Group-Members-to-a-Submission), otherwise submit (all pieces) individually.

Rubric

The following matches the rubric you will see on Gradescope. ***Note your sketches and reflection weight most heavily into your grade.***

	Missing / Not Complete (0)	Approaching (3)	Meets (5)	Exceeds (6)
Sketches	Not submitted or not readable.	Sketches are difficult to read and/or need more detail. They do not demonstrate appropriate visual mappings (as discussed in lecture), or clearly support the analysis objectives.	Sketches are difficult to read and/or need more detail. They include some appropriate visual mappings (as discussed in lecture), but not all visual mappings are appropriate. Some visualizations do not support the analysis objectives.	Sketches are detailed, clear, and easy to read. They demonstrate appropriate visual mappings (as discussed in lecture), and clearly support the analysis objectives.
Reflection	Not submitted or not readable.	Reflection does not fully address all three points listed above. And/or needs improvement in one or more of the following areas: formatting, grammar and spelling, clear, concise writing.	Reflection addresses all three points listed above, but answers are not thoughtful. It is well formatted, contains good grammar and spelling, and clear, concise writing.	Reflection thoughtfully addresses all three points listed above. It is well formatted, contains good grammar and spelling, and clear, concise writing.

Continued on next page

	Missing / Not Complete (0)	Approaching (1)	Meets (2)
Code	Not submitted.	Code does not run.	Code runs.
Write-up	Not submitted or not readable.	Write up addresses some but not all the objective(s) of the assignment. It could use improvement in one or more of the following areas: formatting, grammar and spelling, clear, concise writing. Hypotheses are unclear and/or not supported by visualizations shown.	Write up clearly addresses the objective(s) of the assignment. It is well formatted, contains good grammar and spelling, and clear, concise writing. Hypotheses are present and well supported by visualizations shown.