COMS30038 — SECURITY BEHAVIOURS

LAB 3: CRYPTOMARKET DISCUSSIONS

In this session, you will be reading and analysing real cybercrime discussion data from the forums for the famous Silk Road cryptomarket. The aim is that you will gain some understanding of both the opportunities and the challenges of working with online cybercrime data, and explore different security topics and themes, using the sort of qualitative methods often employed in cybercrime research.

1 SILK ROAD TOPICS

In this exercise you will be engaging in a loose form of *thematic analysis*, working with cybercriminal discussion data. This is one approach that can be used to develop an understanding of rich qualitative data such as you might often find in a cybercriminological context. *Please note: this exercise involves you reading discussions between people involved in crime, primarily the online drugs trade. Things you may see casually discussed and recommended in this data may be both dangerous and illegal.*

To begin with, you will need to form small groups of three or four people, and download the `silkroad.tar.bz2` file, which is a compressed version of a single CSV file. Once extracted (`tar -xvjf silkroad.tar.bz2`), each row in the CSV file relates to a single post from the original Silk Road's forums. The 'body' column contains the text of the post, and the other columns contain metadata about the time, author, and so forth. Posts should be grouped by the thread to which they belong, making it easier to follow a conversation. The 'quotes' column extracts text from the original post body that was quoted from another user – in some cases this can help you follow conversations by indicating what a reply was responding to.

FAMILIARISATION A key initial step is always to understand your data. Spend five minutes reading a few conversations selected from the dataset at random. Check your understanding of the data either within your group or by asking a TA for help. Try to help each other understand the apparent communication norms of the discussion boards you are reading.

Topic assignment There are always large numbers of potential avenues for analysis in data such as this. Fully open-ended exploration can be overwhelming, so it is often good to go in with a question or topic to guide your analysis. Assign each member of your group a topic that they will be focused on for the analysis ahead. If you have spotted topics that you find interesting during familiarisation, you can use those. However, a good prepackaged set of topics for this exercise is:

1. Early finalisation / finalising early / FE
2. PGP
3. social engineering
4. 'stealth'

Filtering There are over 650,000 posts in the dataset – manually examining everything in the dataset to check if it relates to your topic is not possible within this lab. Using whichever tool you find easiest to work with, filter the dataset to produce a subset that relates to your topic. The most straightforward approach would be to use a set of keyword filters to select posts – but it's worth checking the terminology used in the discussions and sense-checking your chosen filtering approach.

Sampling Depending on your chosen topic and filtering approach, you may still have tens of thousands of posts potentially related to your topic of interest. Reviewing all of these may still not be possible, so you need to come up with a method for *sampling* posts from your subset which you will then manually review. Absent other considerations the best approach is uniform random sampling, which almost any data tool should be able to achieve. For a first analysis, use this sampling approach to select a reasonable number of posts that you would be able to review without getting too tired (perhaps 30-50).

Themes Individually read your sample and consider how your topic is being discussed in each post. In some cases you may need to refer back to the context of the thread this post was made in to properly understand the individual post. There are a variety of different approaches to generating themes, but for this exercise we can take a loose, atheoretical approach. At least to start with, make general notes about any aspect of how your topic appears in each post, and particularly any elements that strike you as interesting. The purpose of this exercise is not to develop a definition of the terms in your topic, but to identify themes in discussion about that topic. Consider things like the stated or unstated purpose of the post, what the post implies about cryptomarket users' attitudes or behaviour, or what is different between mentions of the topic in different posts.

Coding Once you have read through your sample once, and have your general thematic impressions, you should try to formalise the different themes

you have seen into different 'codes' or labels that you can apply to each post in the sample. Try to apply at least one label to each post, and come up with short one-line definitions that explain each label. Depending on how you are handling the data, you can apply codes by adding a column to a data structure, highlighting rows in different colours, etc.

Presentation Return to the rest of your group. Each member should present back about their topic and the different themes they found, illustrating with what they think are good examples from their sample. Discuss the topics and what themes you did or didn't expect to find.

Second Coding An important part of making qualitative analysis meaningful is checking that others can reliably apply the categorisations you have invented. Generate a new similarly-sized sample from your filtered subset of the data, and swap samples with another member of your group (also sharing with them the definitions of the different labels you developed). Each group member should then use this fresh sample and try to apply the coding scheme developed by the person who worked on the first sample. As you work, look out for cases where you think there is a new theme not captured by the previous coding scheme, and make a note of these, as well as any situations where you are unsure of how to apply the coding scheme.

Review Return to your group and discuss your experiences of applying the approach used by another group member. It is normal for this not to be straightforward! Identify if there have been misinterpretations of themes, and discuss potential new themes. Try to conclude by arriving at a summary of the themes that would cover both samples, for all the topics your group investigated.

2 WRITING PRACTICE

This section isn't part of the lab itself, but an opportunity for you to practice writing essay-style answers to (week-topical) questions. You can show your answer to your TA next week, or share in your group's Teams channel to get feedback on how you're answering questions.

1. Give an example of a situation where you have deployed one of the techniques of neutralisation discussed in the lectures, or have been tempted to do so. Explain the appeal of this form of self-reflection regarding actions, and why it might appear with regards to crime. [≈250 words]

2. Select one the various potential interventions suggested by Hutchings & Pastrana in this week's readings on eWhoring. Explain some possible unintended consequences that might apply to this intervention, and discuss whether this is a serious enough consideration to avoid going ahead with such an intervention. [≈400 words]