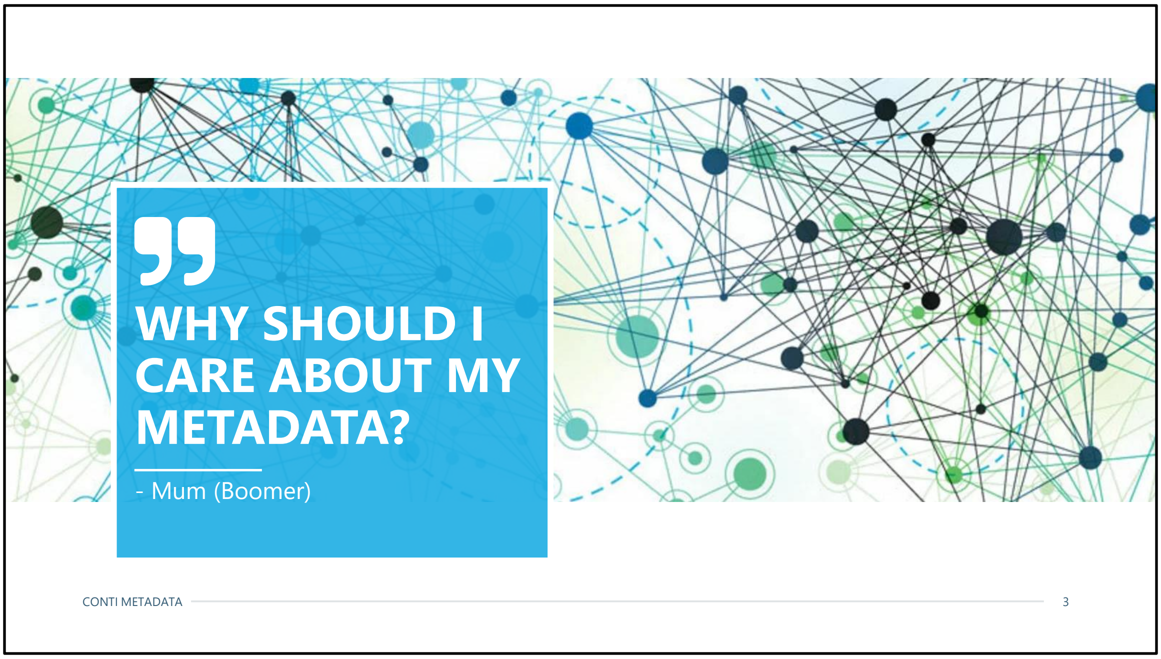
WHY SHOULD I CARE ABOUT MY METADATA?

A DATA VIEW INTO THE CONTI CHAT LEAKS



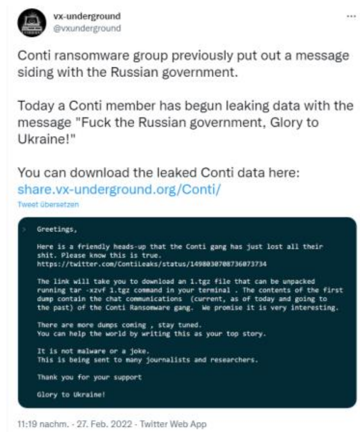
Australia Day 2022
Target tracking purchases predicting pregnancy
Government tracking and tracing
Predictive Algorithms
But really what can you do with it

CONTI LEAKS

- Cybercrime-as-a-service group active since 2019.
- Would utilise “double-extortion”
- Estimated $100 million+ payouts [1]

- 25 Feb 2022: announced support for Russia during invasion of Ukraine
- 27 Feb 2022: thousands of files leaked to vx-underground

- Files included chat logs and source code



Not a Threat Intel Profile

Presented with this range of data I wanted to see what I could do with it.

Ultimately the Conti chat files gave us over 168’000 interactions between 465 unique actors which would have to be reviewed.
Huge task for an analyst
If we could model the metadata, it gives us a place to start and target our analysis efforts.

Diving into the logs, it looks like it provides us enough metadata to get a head start
Time and who is talking to who is all we need. Ultimately what is being said is irrelevant to us.

What we can do is use the metadata to look at the *relationships* between actors in order to assist with further analysis.

WHAT IS A NODE GRAPH?

USE OF NODE GRAPHS

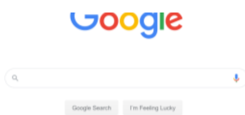
Graph theory and graph analysis should be used to study the **relationship** between objects.

Maps and Pathfinding

Google Maps utilises Graph Theory to find the shortest route to a destination [2]

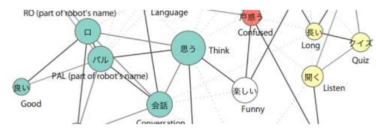
Social Networks

Tracking the connection between people and places.



Google PageRank

Google Search looks at the number of relationships between pages to assign ranking.



Linguistics

Researchers can use graphs to track the relationships between words. [3]

WE USE THESE GRAPHS TO LOOK AT RELATIONSHIPS

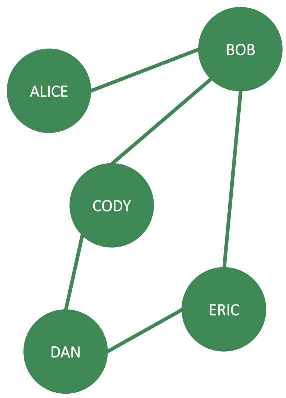
One of the first uses for Graph Theory was mapping how to get from island to island in Konigsburg without crossing a bridge twice.

Linguistics: Tracking words used by patients by non-talkative schizophrenic patients

Bloodhound uses Node Graphs looking at the relationships between different domain objects.
If you collect network data in a SIEM you can build a node graph to characterise that traffic.

WHAT IS A NODE GRAPH?

NODE GRAPH



Alice, Bob, Cody, Dan, and Eric are at a convention

Alice shook Bob’s hand

Cody shook hands with Bob and Dan

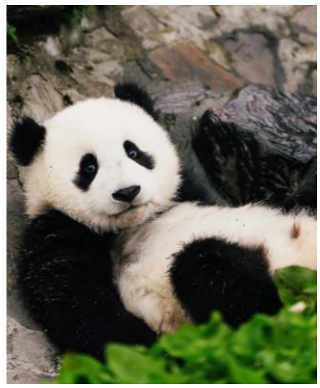
Eric shook hands with Dan and Bob

Read the slide

There are different types of node graphs. You can have directed graphs that show dependencies or interactions. Or weighted graphs.

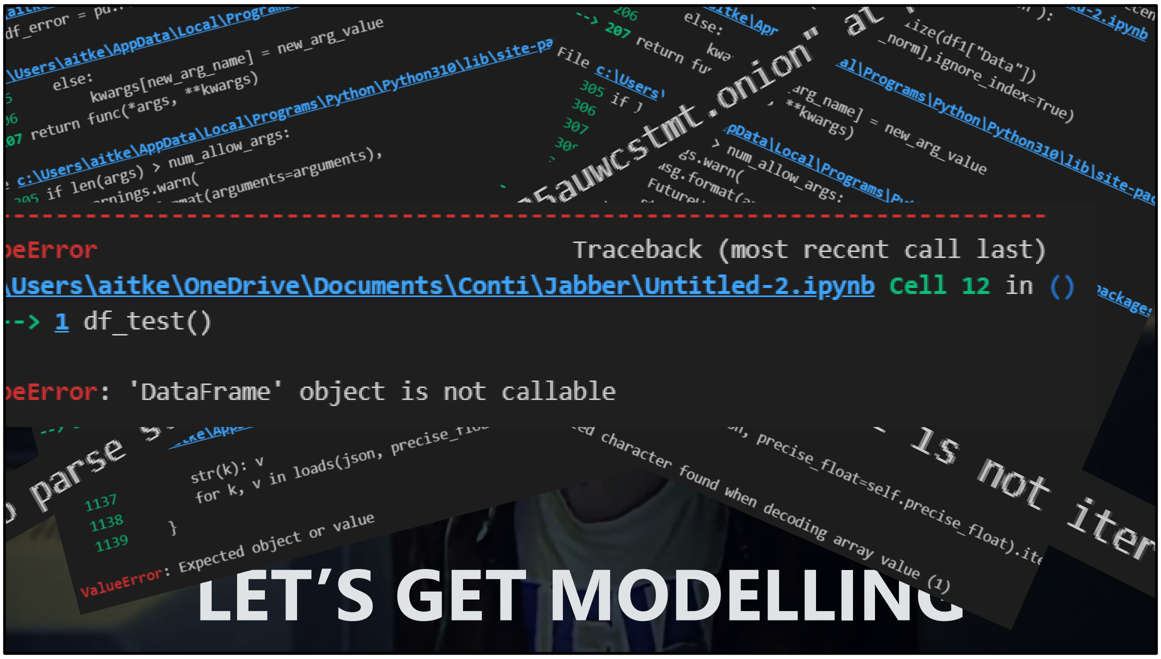
GRAPH ANALYSIS TECHNIQUES

ANALYSIS TOOLS



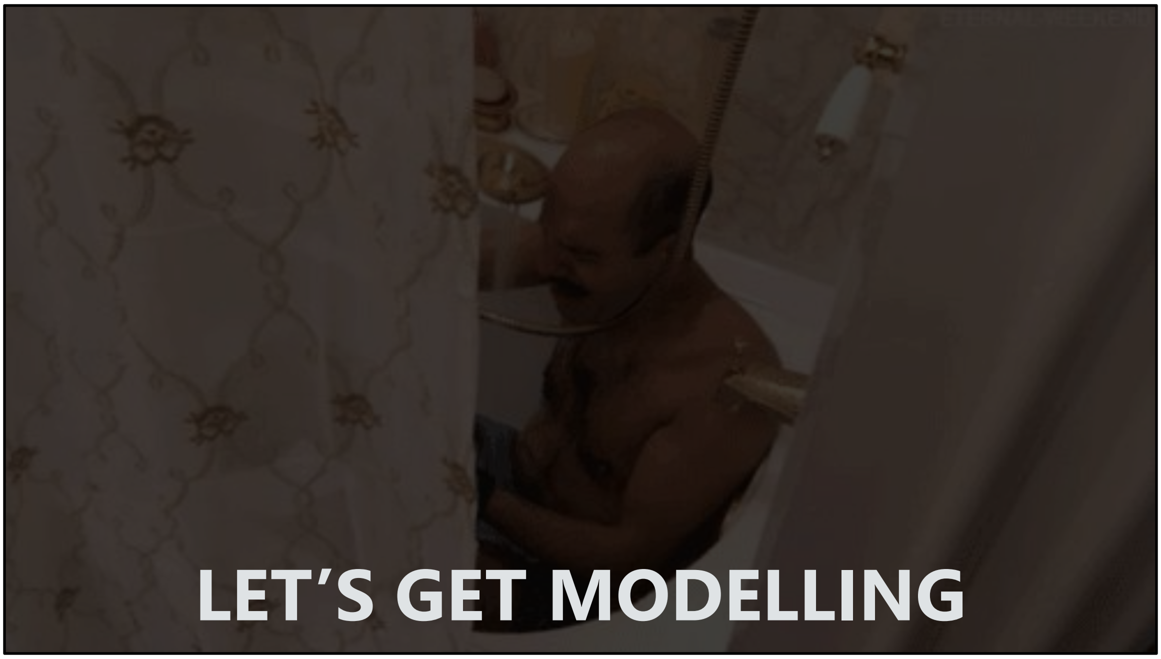
Three primary tools will be used for Analysis.

Python, because it’s better than trying to do this all in PowerShell (you may laugh, but that’s what I used to do. Because we weren’t allowed Python on our work computers)
(this one is a ball python)
Pandas, to help us manipulate the JSON data and it has some basic statistical tools built in.
NetworkX, which doesn’t have a cute animal logo and should be ashamed of themselves. Package made for drafting complex networks.



So, I allocated myself 24 hours to get cracking on this project

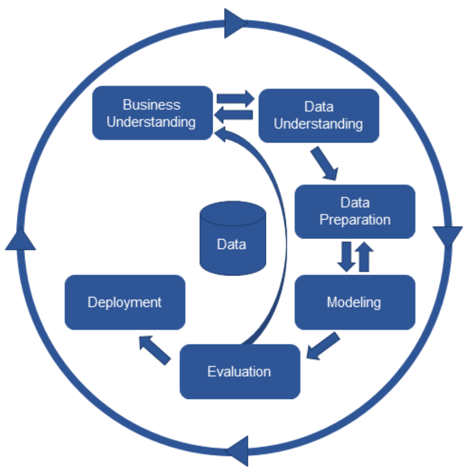
Totally not because I had a university assignment due the next day



So 18 hours into my 24 hours I've finally completed the work I needed to do *before* modelling and my housemate found me crying in the shower.

GRAPH ANALYSIS TECHNIQUES

BEFORE THE MODELLING



CHOOSE THE RIGHT DATA

Make sure you're collecting the right data from the right places.

UNDERSTAND YOUR DATA

Read documentation, understand the nuance of the data. Explore and make sure you know exactly what you're collecting.

CLEAN YOUR DATA

Get rid of data that doesn't matter, make sure it's all the same format. Garbage in, Garbage out

This is CRISP-DM – Reference 4. If you ever work with data in your job for anything, check this model out and use it as a structure for your workflow.
Collect your data, have a play, explore it.
Describe it, understand the qualities and features.
Cleaning the data is the lengthiest task. (side note: there is a standard for JSON – ISO/IEC 21778 – developers who do not follow it should be shot. I will take no questions on this matter)

GRAPH ANALYSIS TECHNIQUES

CREATING THE GRAPH

```
1. node_series = conti_df['to_short']
2. node_series = node_series.append(conti_df['from_short'])
3. node_series = node_series.drop_duplicates()

4. G = nx.Graph()
5. G.add_nodes_from(node_series)
6. G = nx.from_pandas_edgelist(conti_df, source='to_short’, target='from_short’)

7. pos = nx.spring_layout(G,scale=500,k=50/np.sqrt(G.order()))
8. d = dict(G.degree)
9. nx.draw(G, pos, node_size=[d[k]*20 for k in d], with_labels=True)
```

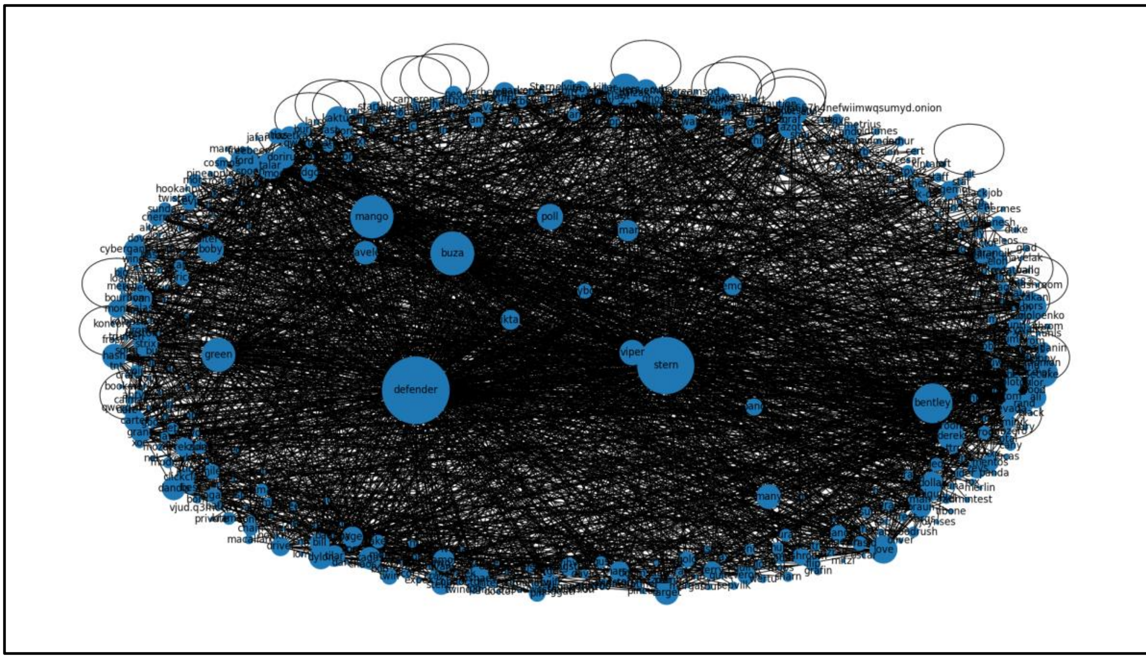

There will be very little code in this presso, I promise. But this is for the people who want to know what happens behind the scenes.

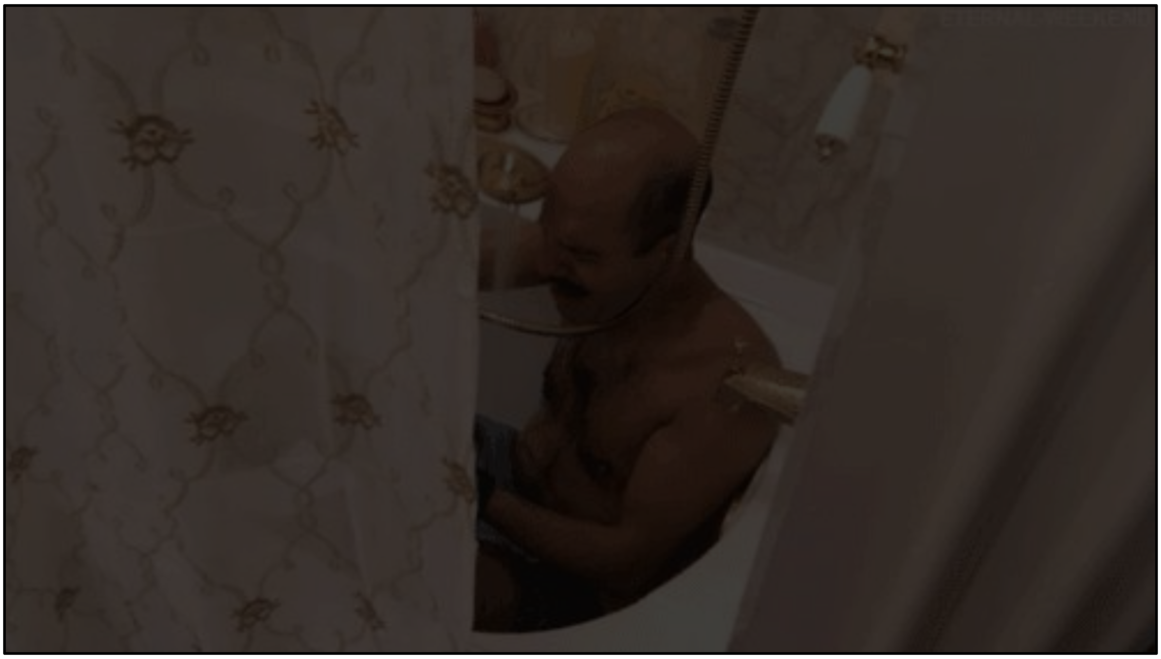
Essentially there are three parts to creating the graph. Identifying the nodes (1-3)

Creating the graph structure (4-6)

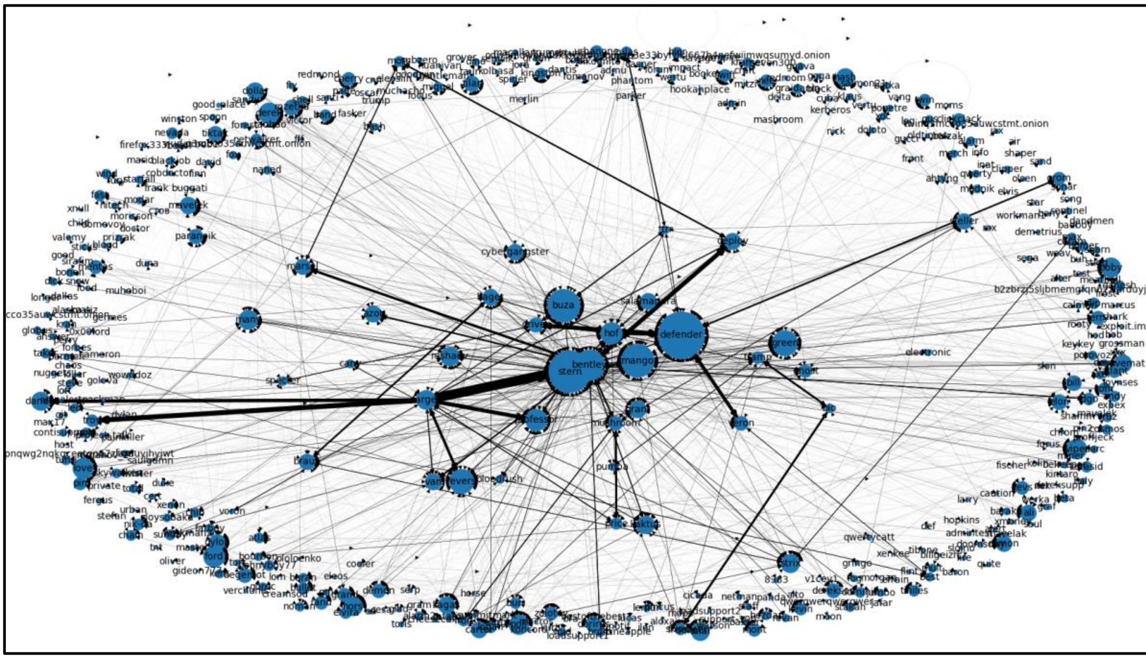
Finally drawing the graph (7-9). What we’re saying in 7-9 is that we want the more important nodes to be bigger.

Great! So now we’re going to have a pretty graph that will give us all the answers!

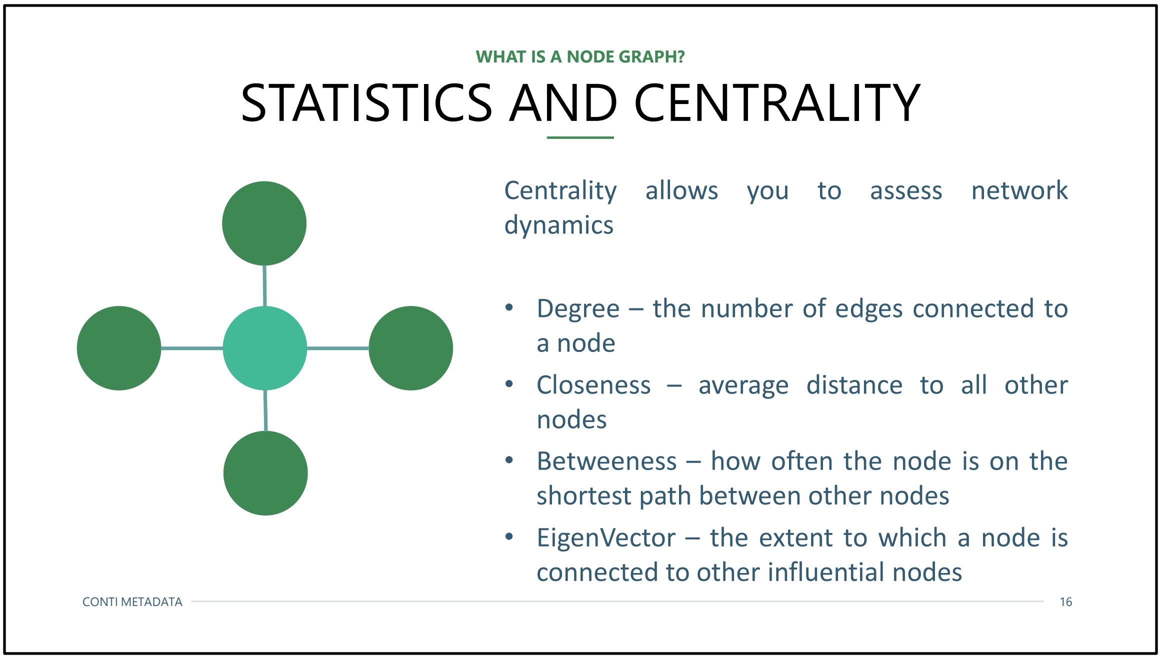

mango
poll
buza
green
defender
viper
stern
bentley
love
cameron
hermes
merlin
duke
admintest



Graphs with this many nodes become tricky to purely use as analysis tools. Those magical answers you thought you would find in nine lines of code probably wont be there. But there are some small tricks…



stern
bentley
defender
mangor
buza
hof
green
deploy
electronic
cyberdungster
merlin
phantom
parker
admin
masbroom
redmond
winston
nevada
dylan
host
fischer
larry
hopkins
moon
oliver
fergus
stefan

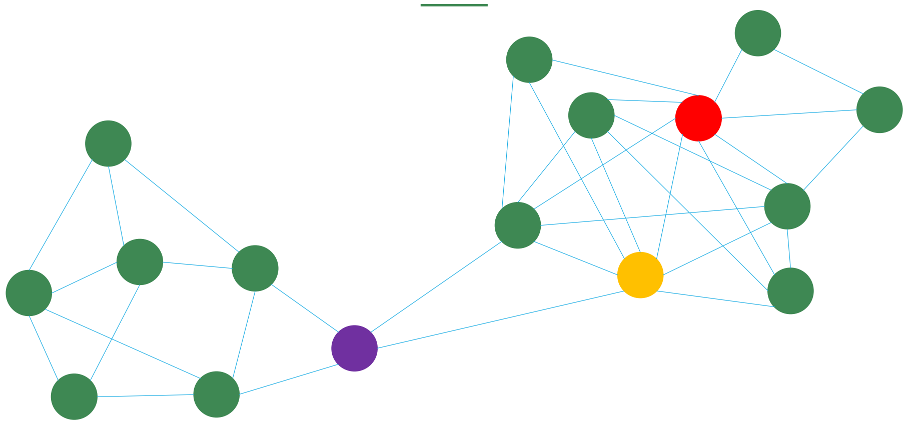


Really the graph is just a means to an end. What matters is the stats you can pull from the graph. Graphs allow you to identify *centrality*. Centrality tells you who is important in a network

Degree: exposure to the network. Opportunity to directly influence.
Closeness: Who diffuses information to the network.
Betweeness: Informal Power. Broker Resources. Controls flow of information.
Eigenvector: Also used to identify informal power. Particularly useful in dense networks. “Not what you know, but who you know”

WHAT IS A NODE GRAPH?

STATISTICS AND CENTRALITY



CONTI METADATA 17

Red: Degree

Yellow: Closeness

Purple: Betweenness

GATHERING STATISTICS

```
df_metrics = pd.DataFrame(dict(
    in_degree = nx.degree_centrality(G),
    eigenvector = nx.eigenvector_centrality(G),
    closeness = nx.closeness_centrality(G),
    betweenness = nx.betweenness_centrality(G)
))
```

GRAPH ANALYSIS TECHNIQUES

CENTRALITY

NAME	CENTRALITY
Defender	1.058
Stern	0.859
Mango	0.522
Buza	0.511
Bentley	0.450
Green	0.303
Revers	0.281
Hof	0.225

DEGREE

Exposure to the network.
Opportunity to directly influence.

NAME	CENTRALITY
Stern	0.463
Defender	0.475
Bentley	0.414
Mango	0.413
Buza	0.407
Green	0.402
Professor	0.372
Revers	0.372

CLOSENESS

Who diffuses information to the network.

NAME	CENTRALITY
Defender	0.293
Stern	0.167
Mango	0.065
Buza	0.058
Bentley	0.043
Ford	0.031
Revers	0.025
Green	0.022

BETWEENNESS

Informal Power. Broker Resources.
Controls flow of information.

Purely by looking at the metadata statistics, we can start to find some information about Conti.

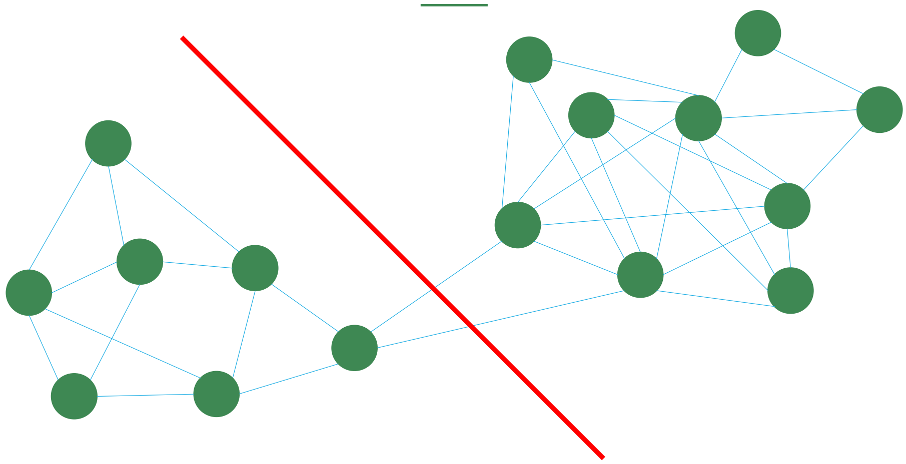
There are a few nodes with high authority. These are obviously highly influential nodes within the network. These nodes are the centre of information. It also means the network isn’t highly resilient. Loss of these nodes could lead to a communication breakdown.

The closeness start middling, but like degree trend downwards rapidly. Our players with the highest degrees are highest on the closeness. Are these individuals regularly in contact with everyone in the group? Interestingly Bentley moves up and Professor appears on this list. Do these players serve roles as middle-managers? Maybe they have larger teams?

Defender and Stern present with the highest betweenness scores. This really represents that there are no other information channels within the network. To get from one side of the network to the other you probably have to go through them.

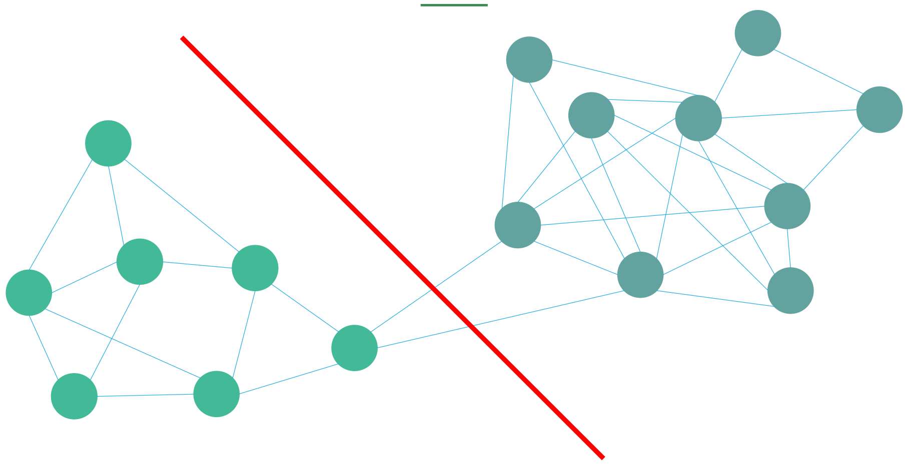
WHAT IS A NODE GRAPH?

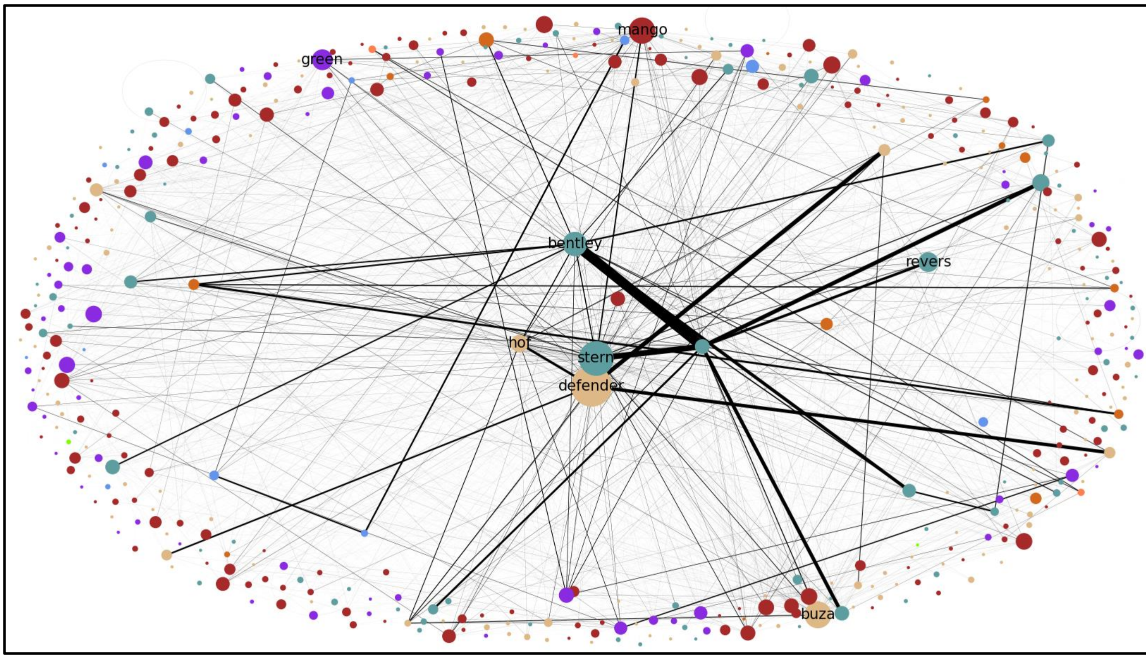
FINDING COMMUNITIES



WHAT IS A NODE GRAPH?

FINDING COMMUNITIES





Here's our network with some community detection applied. For ease of reading I've only allowed the names of the people that we discussed earlier.

Community detection is difficult to apply to this graph. Stern and Defender appear to regularly interact with all elements of the network. The algorithms don't handle that well. Maybe we should drop Stern and Defender and see what shakes out?

Stern and Defender appear to have their own core groups that they manage. Additionally there is an unnamed blue dot that Bentley and Stern both interact with regularly? Why?

Mango and Green appear to be the only nodes of significance in their community, cementing the idea that they might be team leaders.

Who is this orange dot? They appear to be somewhat central, but not within the upper leadership echelons?

OUTSIDE RESEARCH

STERN
Identified as one of the leaders of the group.

Often described as the "CEO"

"Catankerous taskmaster" - KrebsOnSecurity

BENTLEY
Mid-level manager or senior developer.

Instructed junior members on duties.

Supposedly in charge of anti-antivirus works.

MANGO
Mid-level manager and coder.

Handled QA works.

Typically tasked by Stern on random side projects.

BUZA
Conflicting information.

Representative from Emotet (KrebsOnSecurity)

Technical manager in charge of coders. (Check Point)

DEFENDER
Another senior member.

Described as "COO" compared to Stern's "CEO".

Handled finances and internal logistics.

KrebsOnSecurity did a big series of four pieces
Forescout
Check Point
Rapid7

Mango: "Blockchain", "hacker social network"

TAKEAWAYS

EXPLORE CENTRALITY

- Remove key nodes
- Discover pathways
- Identify key individuals

UNDERSTAND COMMUNITIES

- Who is the core of a group?
- Who is missing?

PLAY WITH THE DATA

- Explore what you’re looking at
- Add, remove, chop, change
- Try new things

REFERENCES

1. https://ransomwhe.re/#browse
2. https://www.ijcaonline.org/archives/volume178/number8/mehta-2019-ijca-918791.pdf
3. https://robomechjournal.springeropen.com/articles/10.1186/s40648-020-00161-6
4. https://www.datascience-pm.com/crisp-dm-2/
5. https://krebsonsecurity.com/2022/03/conti-ransomware-group-diaries-part-ii-the-office/
6. forescout.com/resources/analysis-of-conti-leaks/
7. https://research.checkpoint.com/2022/leaks-of-conti-ransomware-group-paint-picture-of-a-surprisingly-normal-tech-start-up-sort-of/
8. https://www.rapid7.com/blog/post/2022/03/04/graph-analysis-of-the-conti-ransomware-group-internal-chats/

WHY SHOULD I CARE ABOUT MY METADATA?

A DATA VIEW INTO THE CONTI CHAT LEAKS