NINJA CORRELATION OF APT BINARIES

EVALUATING THE EFFECTIVENESS OF FUZZY HASHING TECHNIQUES IN IDENTIFYING PROVENANCE OF APT BINARIES

Bhavna Soman

Cyber Analyst/Developer, Intel Corp.

@bsoman3

DISCLAIMERS

Opinions expressed are those of the author and do not reflect the opinions of his/her employer.

-GEORGE P. BURDELL

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

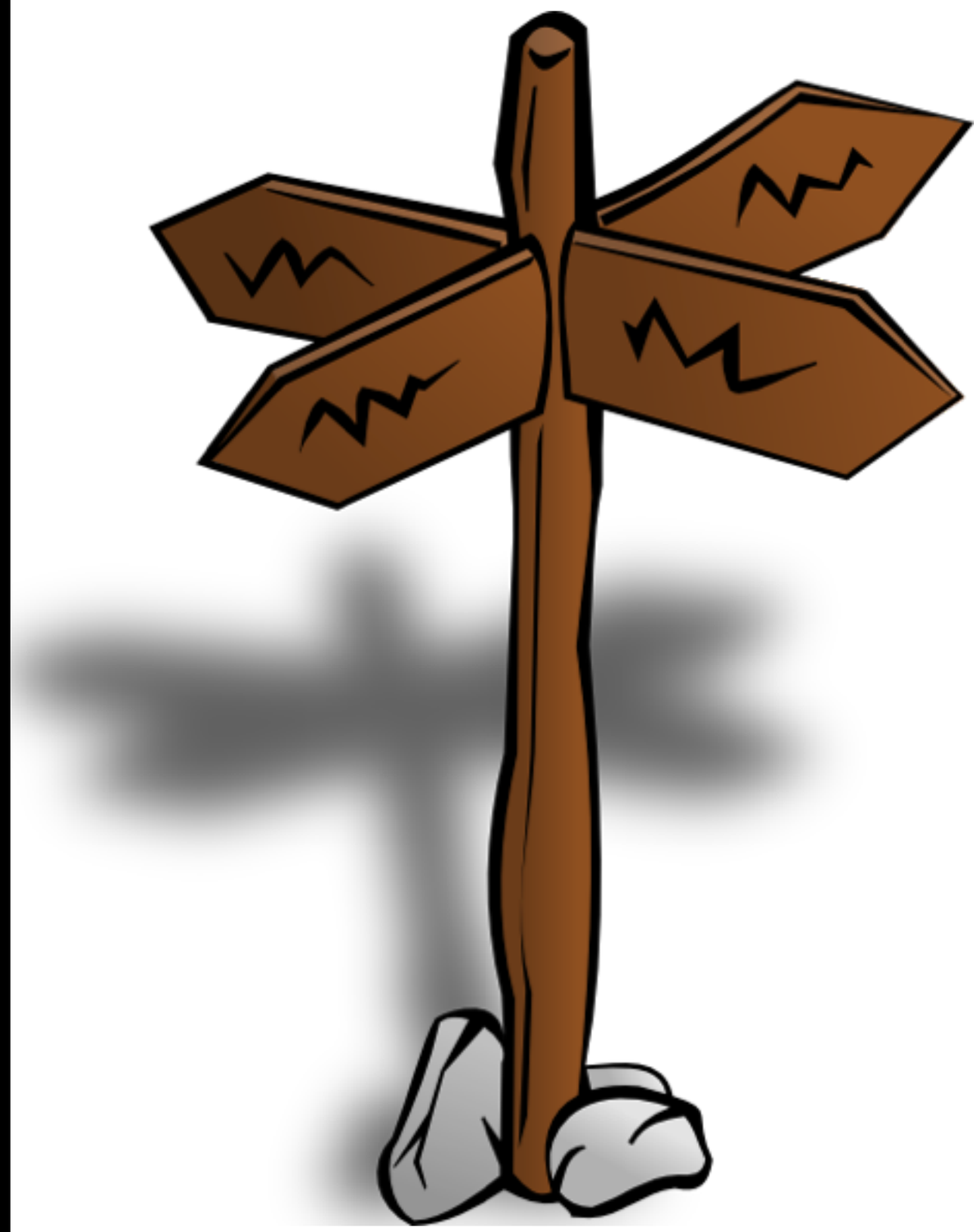
-LEGAL

WHAT ADVANTAGE CAN KNOWING THE ORIGINS OF A MALICIOUS BINARY GIVE YOU??

- We can apply past analyses of motivations and capabilities of adversary
- Connect disparate events into one whole picture
- So what’s the best way to connect the dots?

AGENDA

- Methods to connect binaries
- Getting a test dataset and ground truth
- Results
- Sample clusters found
- Takeaways and Future direction



WHAT IS THE BEST WAY TO CONNECT SIMILAR BINARIES??

- Imphash— md5 hash of the import table
- ssdeep— Context triggered piecewise hashing
- SDhash— Bloom filters

How to :

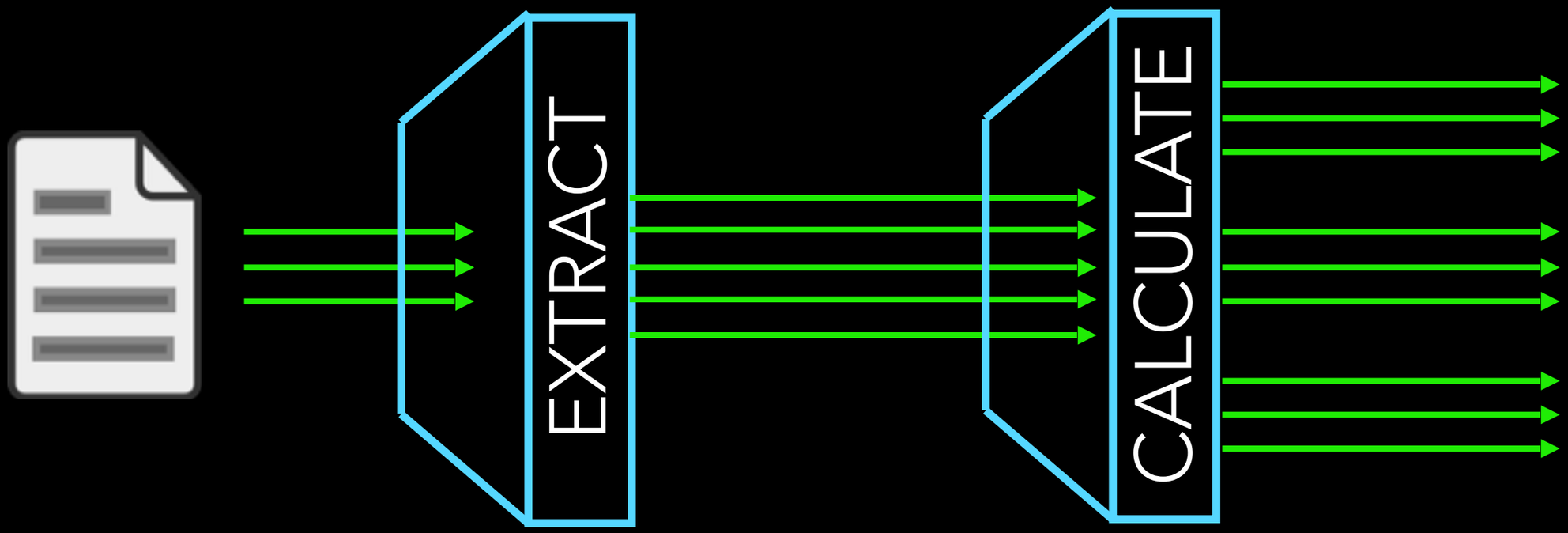
1. *Get non-trivial dataset of binaries related to targeted campaigns*
2. *Establish ground truth without static/dynamic analyses of hundreds of binaries?*

GATHERING DATA

APT Whitepapers

MD5s

Similarity Metrics



- Published Jan-March 2015
- e.g. “Project Cobra Analysis”, “The Desert Falcon Targeted Attacks”

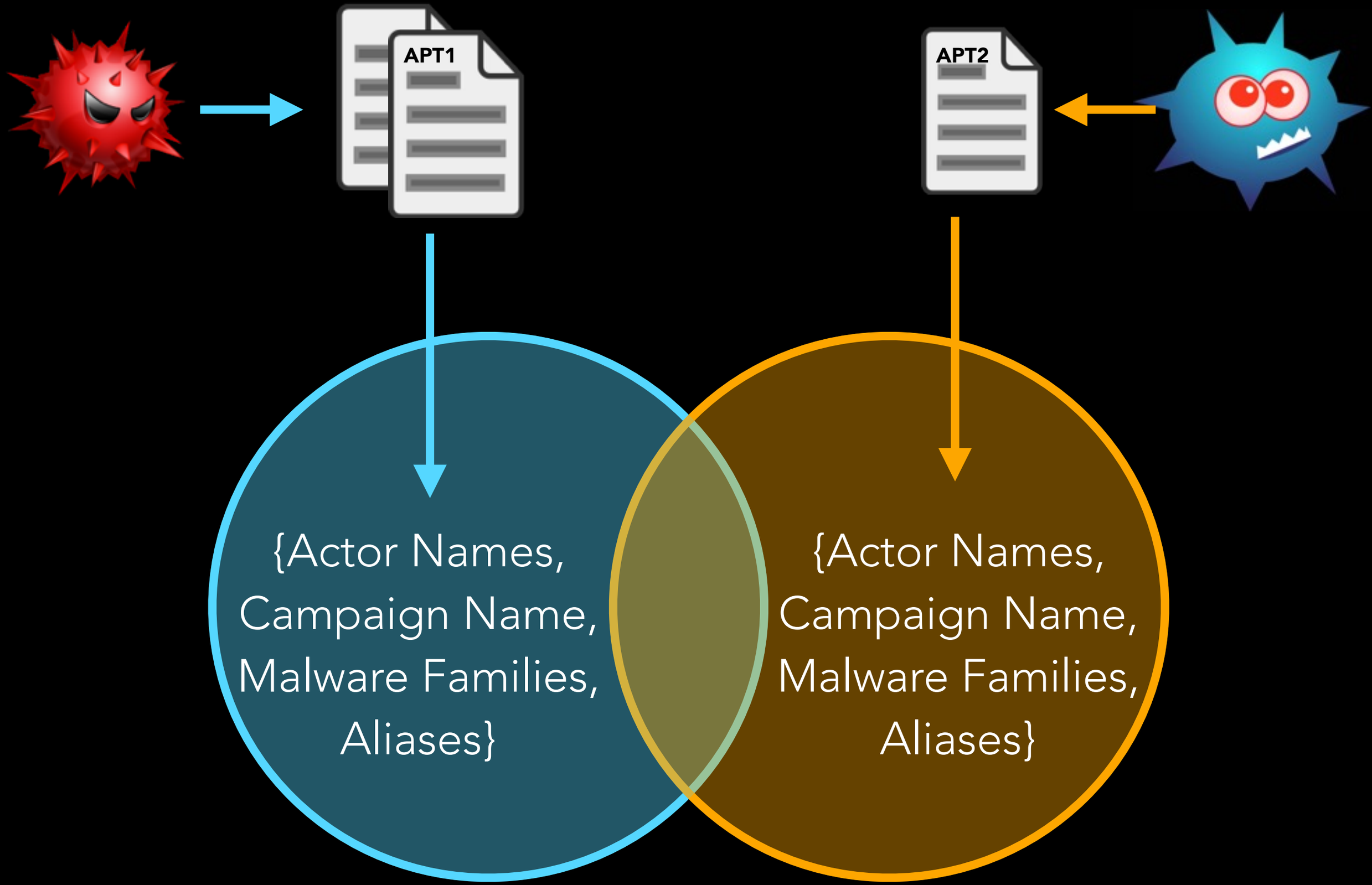
- Extract MD5s
- >10% Malicious on Virus Total

- Calculate for each binary
  - Import hash
  - ssdeep
  - SDhash

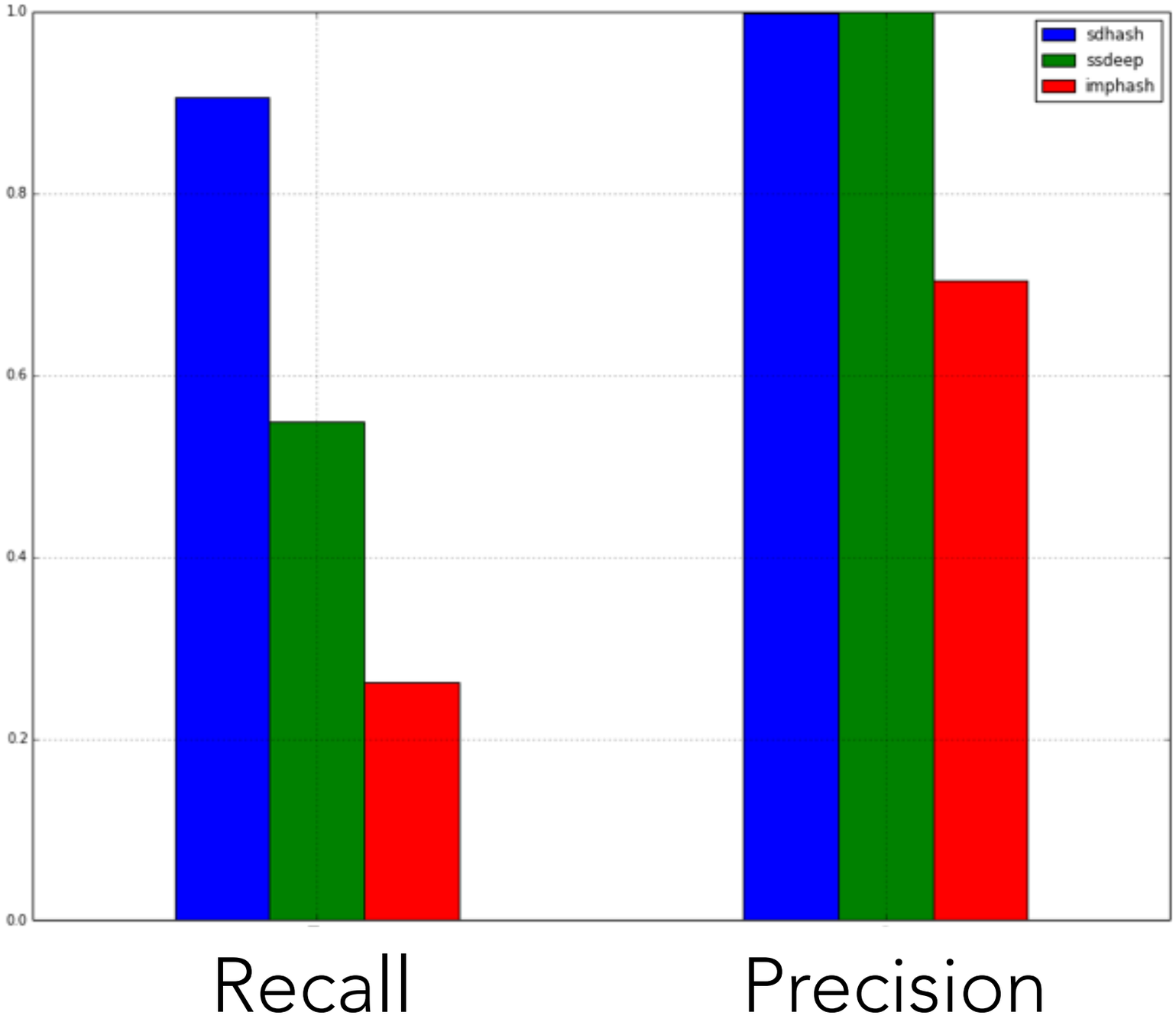
ASSESSING CORRELATIONS

Are these malware related?

ASSESSING CORRELATIONS



SUMMARY RESULTS

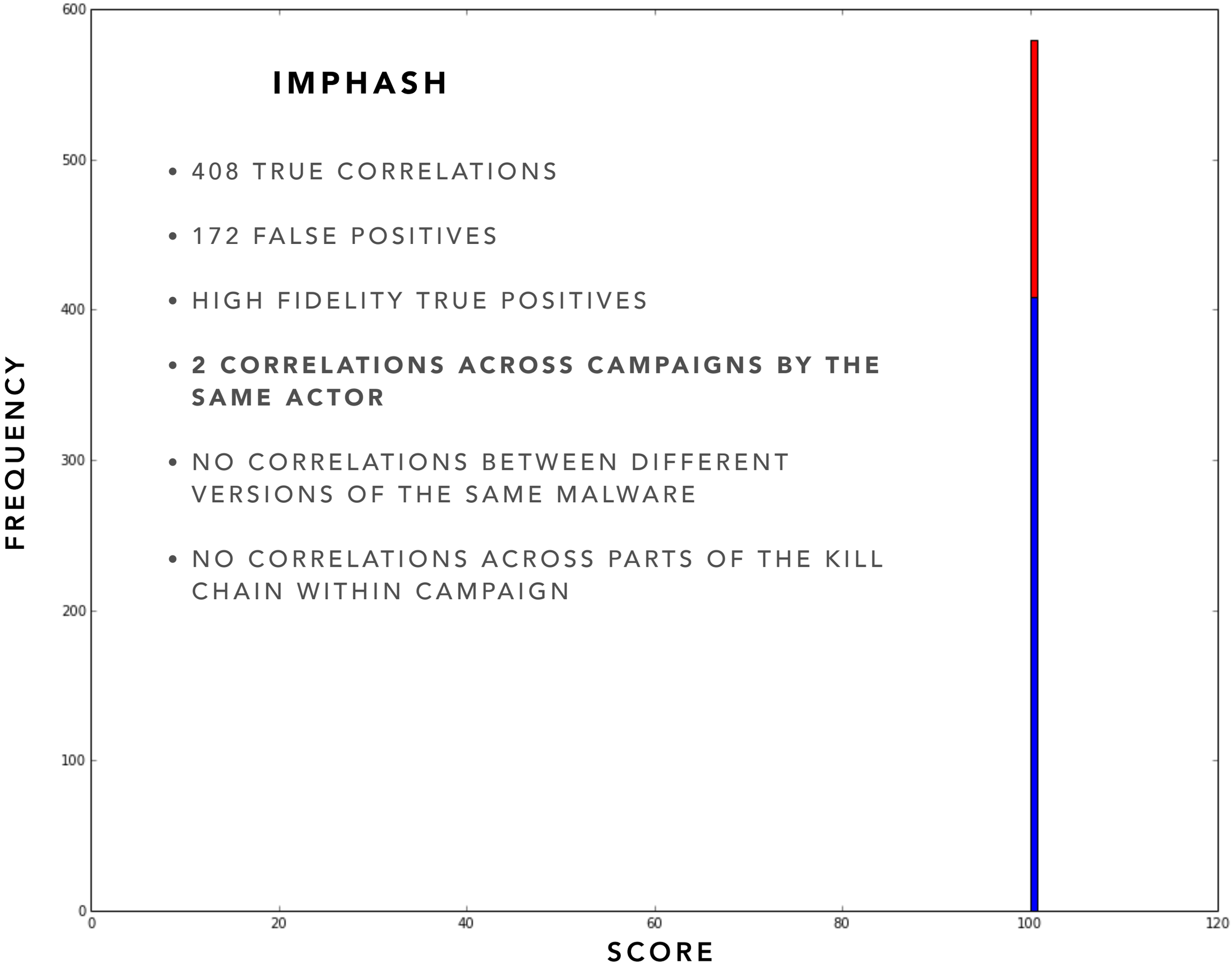


- No one method found all the correlations
- Imphash had the most false positives
- Sdhash had maximum recall
- Both ssdeep and SDhash had near perfect precision

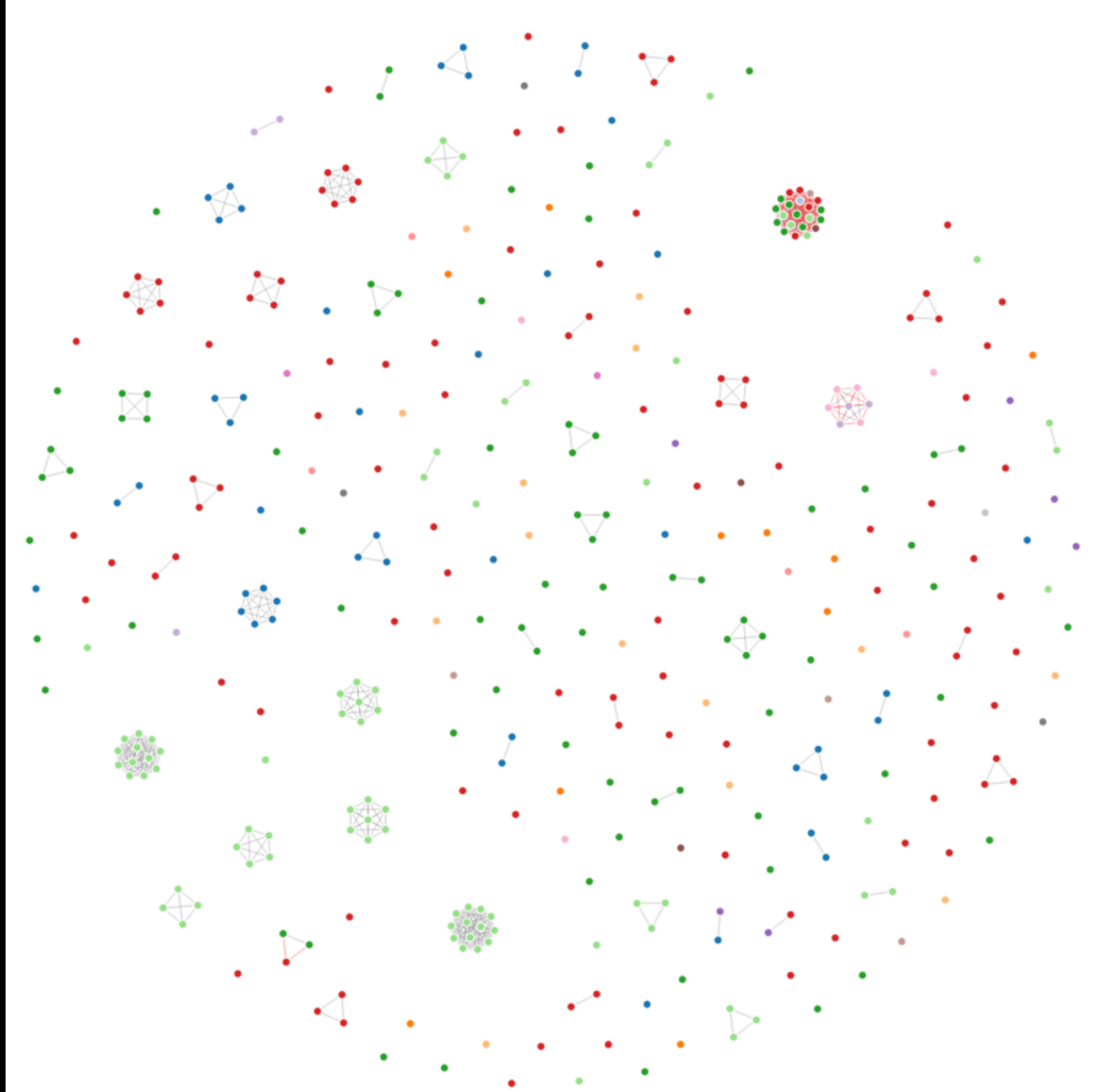
IMPHASHES

IMPHASH

- 408 TRUE CORRELATIONS
- 172 FALSE POSITIVES
- HIGH FIDELITY TRUE POSITIVES
- **2 CORRELATIONS ACROSS CAMPAIGNS BY THE SAME ACTOR**
- NO CORRELATIONS BETWEEN DIFFERENT VERSIONS OF THE SAME MALWARE
- NO CORRELATIONS ACROSS PARTS OF THE KILL CHAIN WITHIN CAMPAIGN



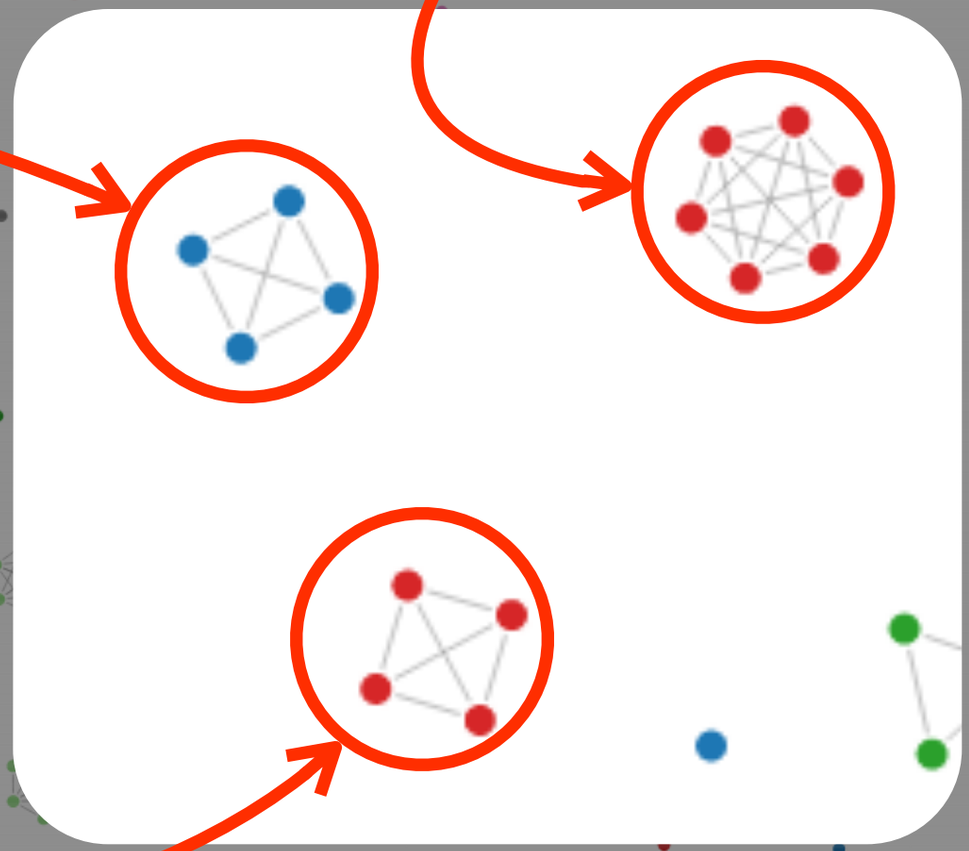
IMPHASH



IMPHASH

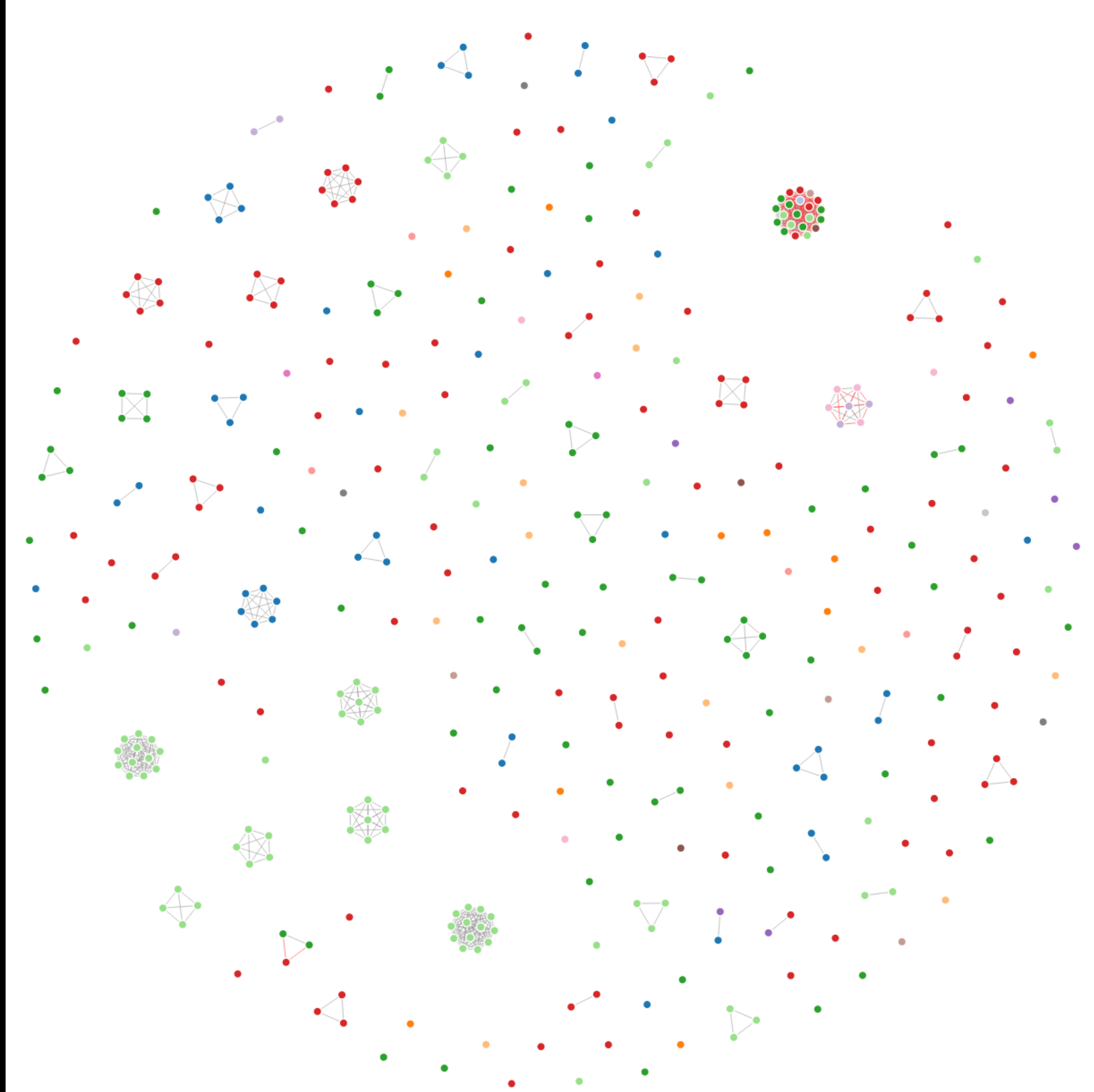
- Version 1.5 of ComRAT (Turla Attackers)
- Compiled on March 25, 2008
- Other versions of the RAT in the dataset were not connected

- Wipbot 2013 Samples
- Used by the Waterbug attack Group
- Compiled on 15-10-2013
- Also referred to as Tavdig/ WorldCupSec/Tadj Makhal



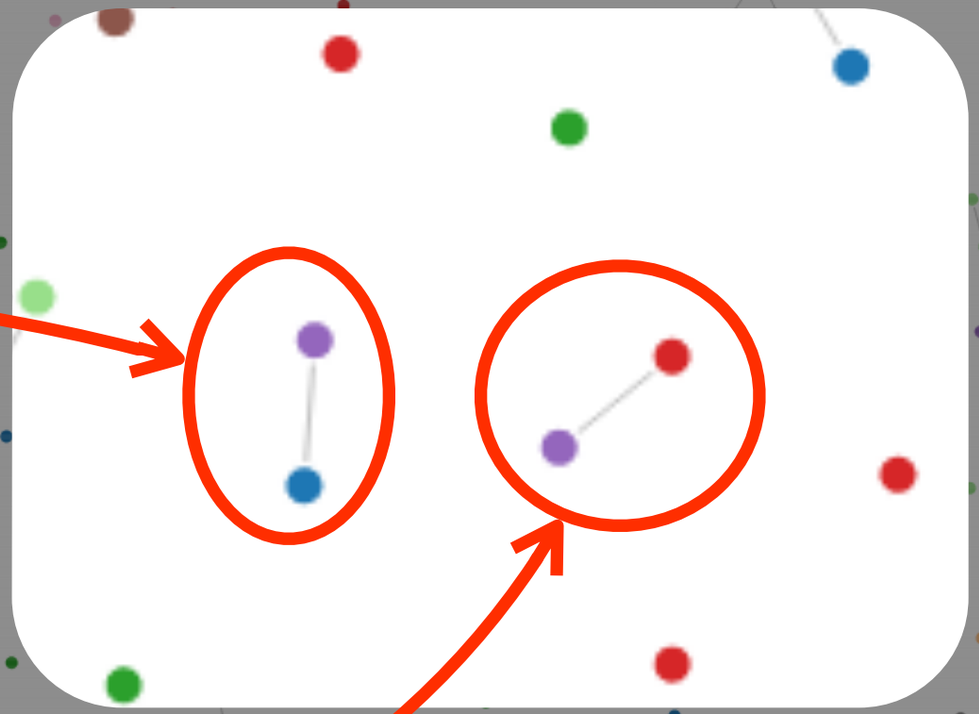
- SAV samples circa 2011
- Used by the Waterbug Attack group
- AKA Turla/Uruboros

IMPHASH



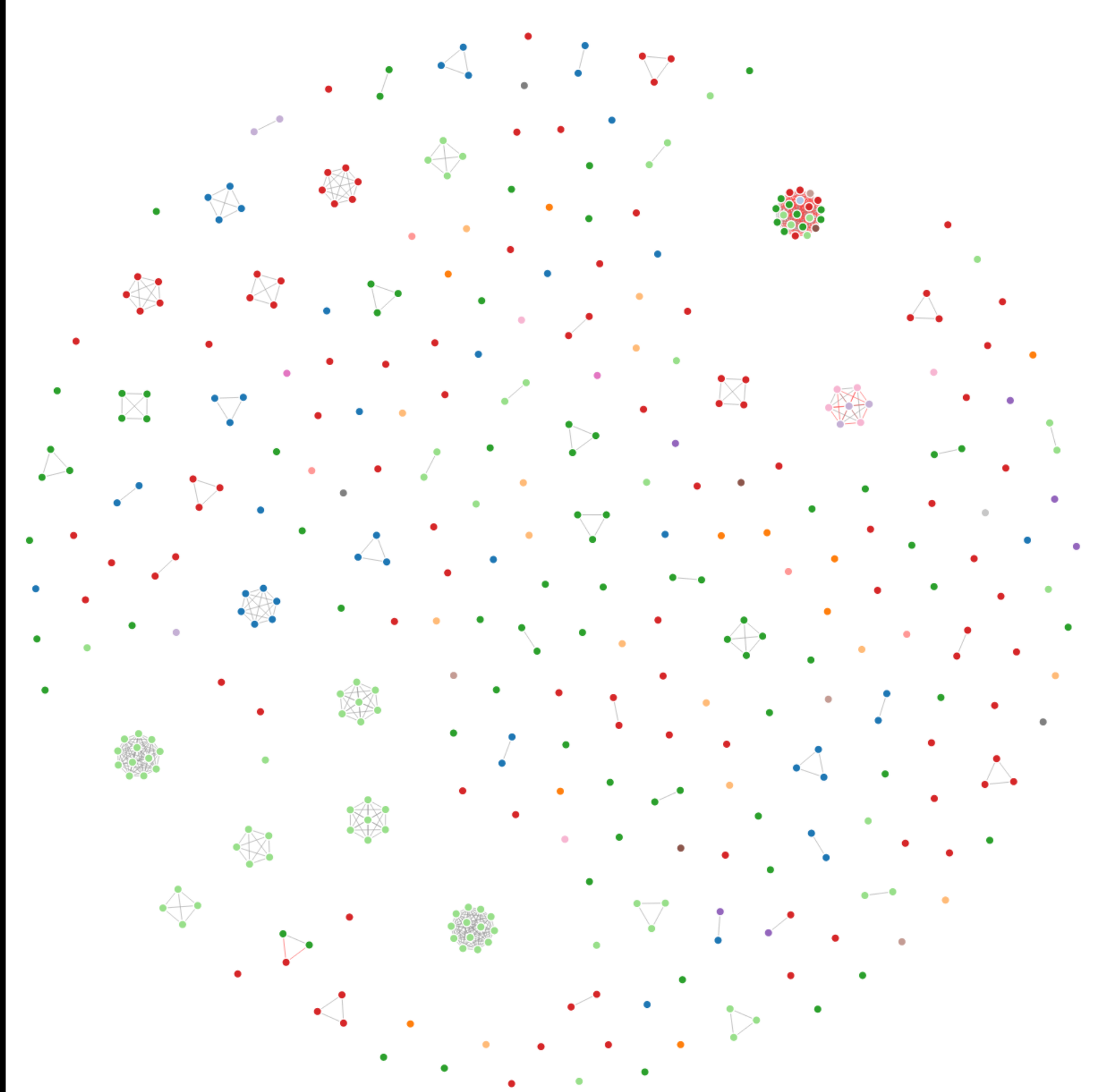
IMPHASH

- Both samples of ComRAT
- Associated with Waterbug Group and Turla Attackers respectively



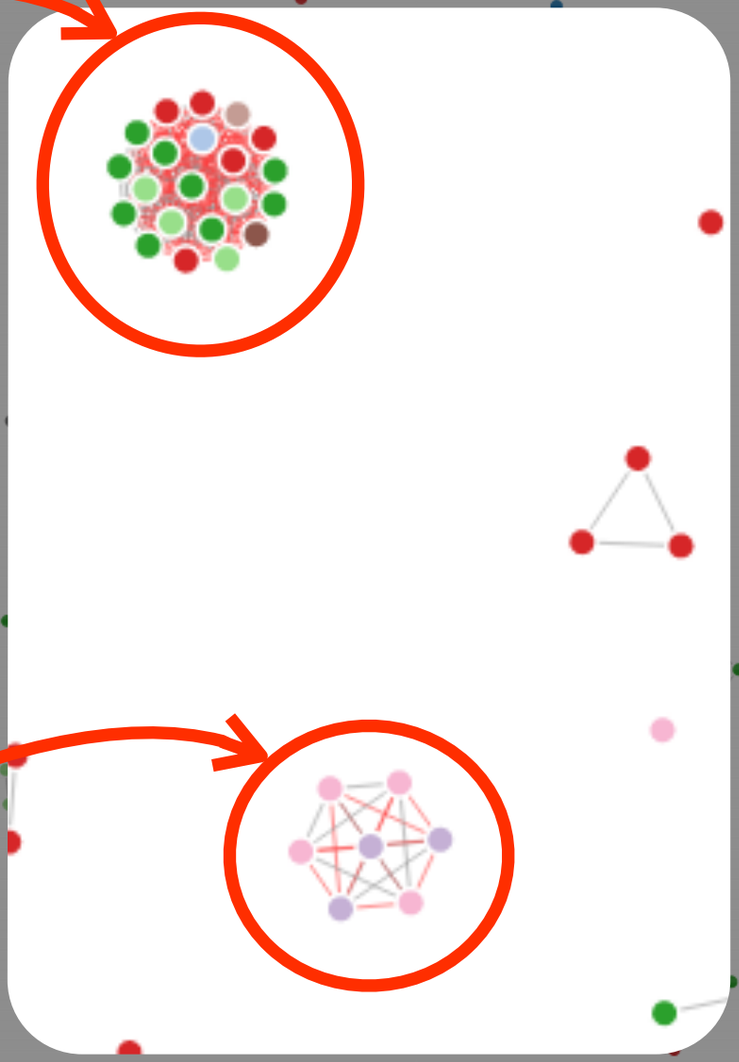
- Samples of the Carbon Malware
- Related to Project Cobra and The Waterbug Attack Group

IMPHASH



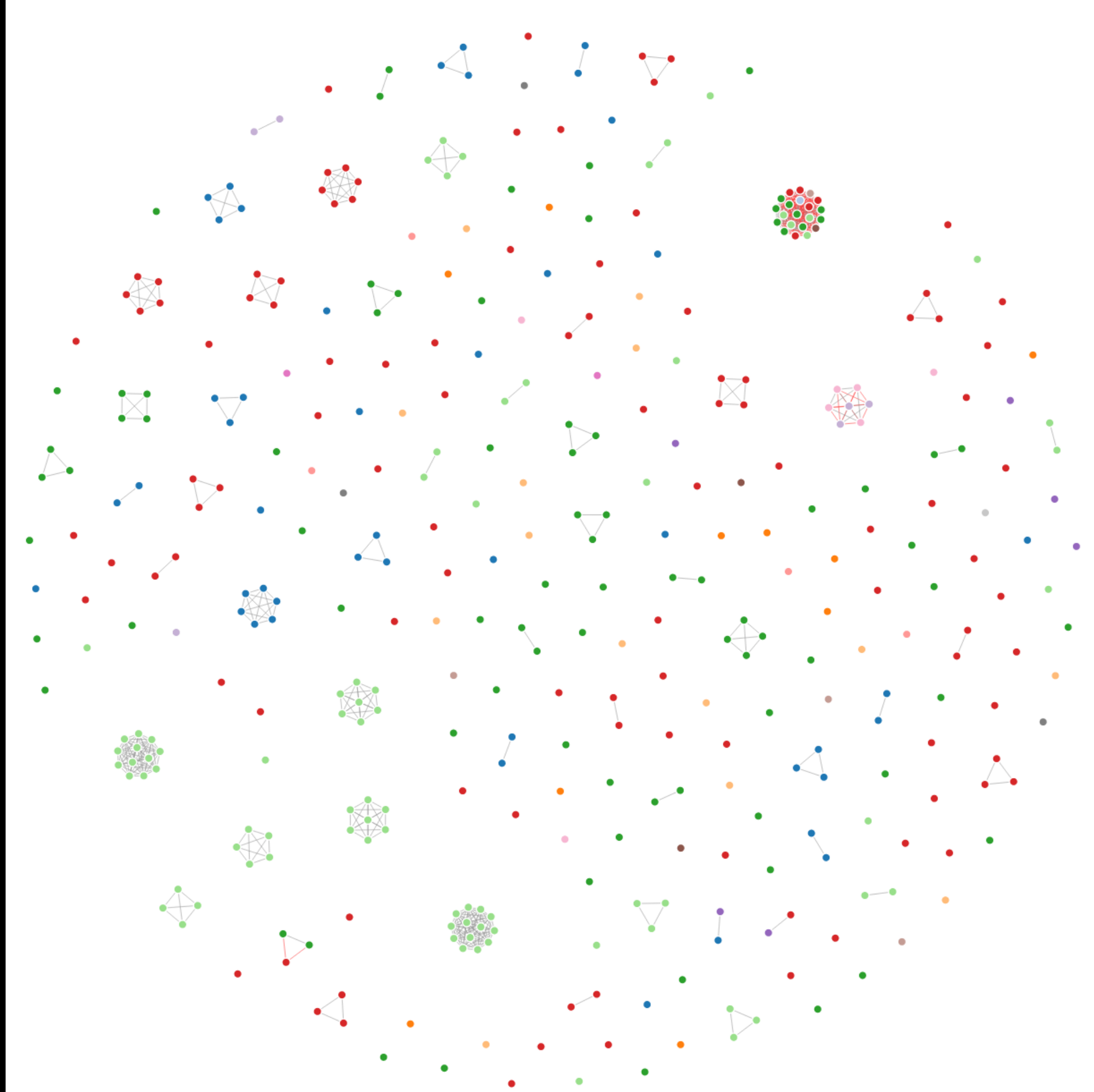
IMPHASH

- Binaries from SIX different campaigns
- No common Actor or Malware Family
- Different parts of the Kill chain



- Credential stealer and dropper from OP Arid Viper
- Vs. Droppers used by Attacks on the Syrian Opposition Forces
- No common attribution or KNOWN link

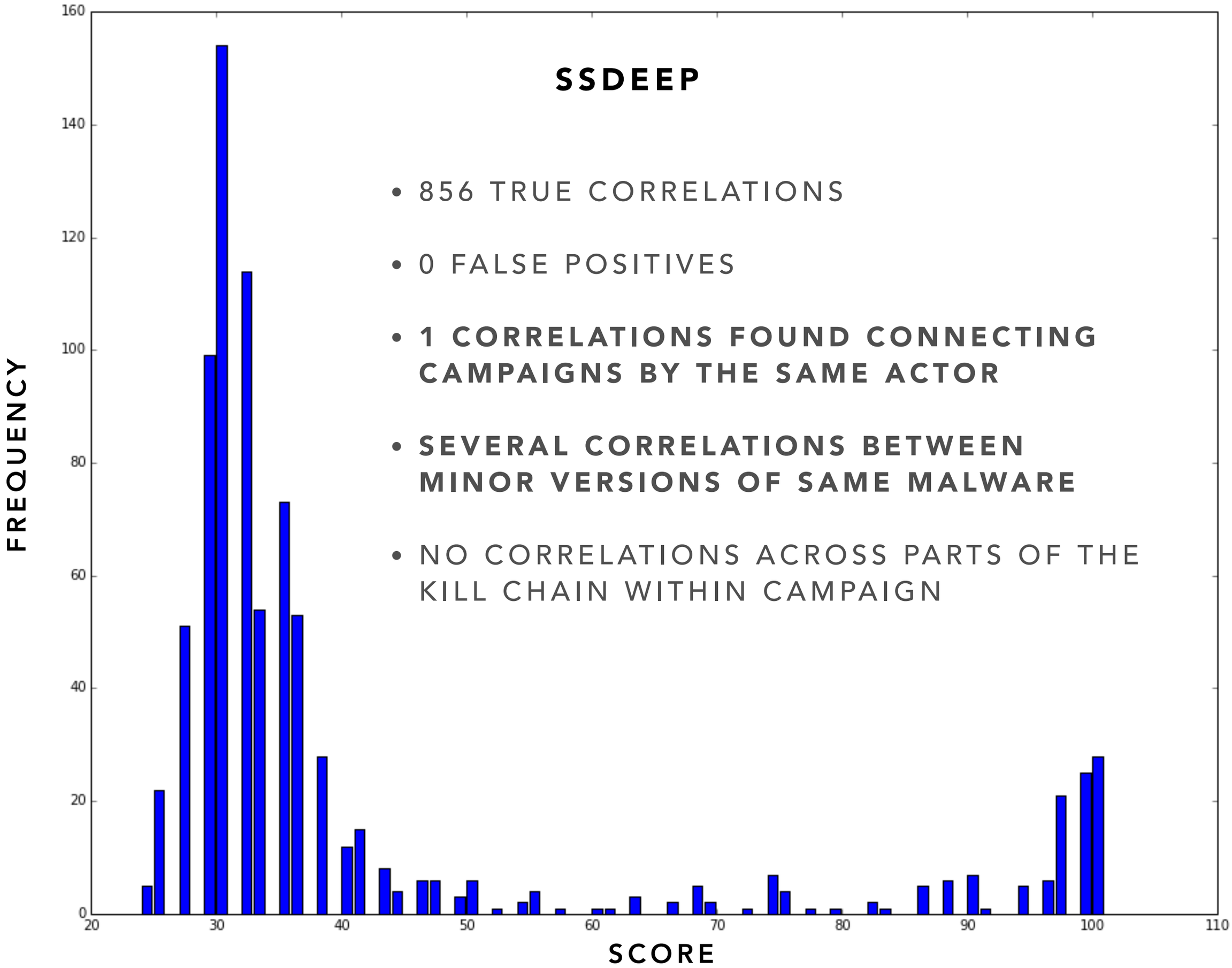
IMPHASH



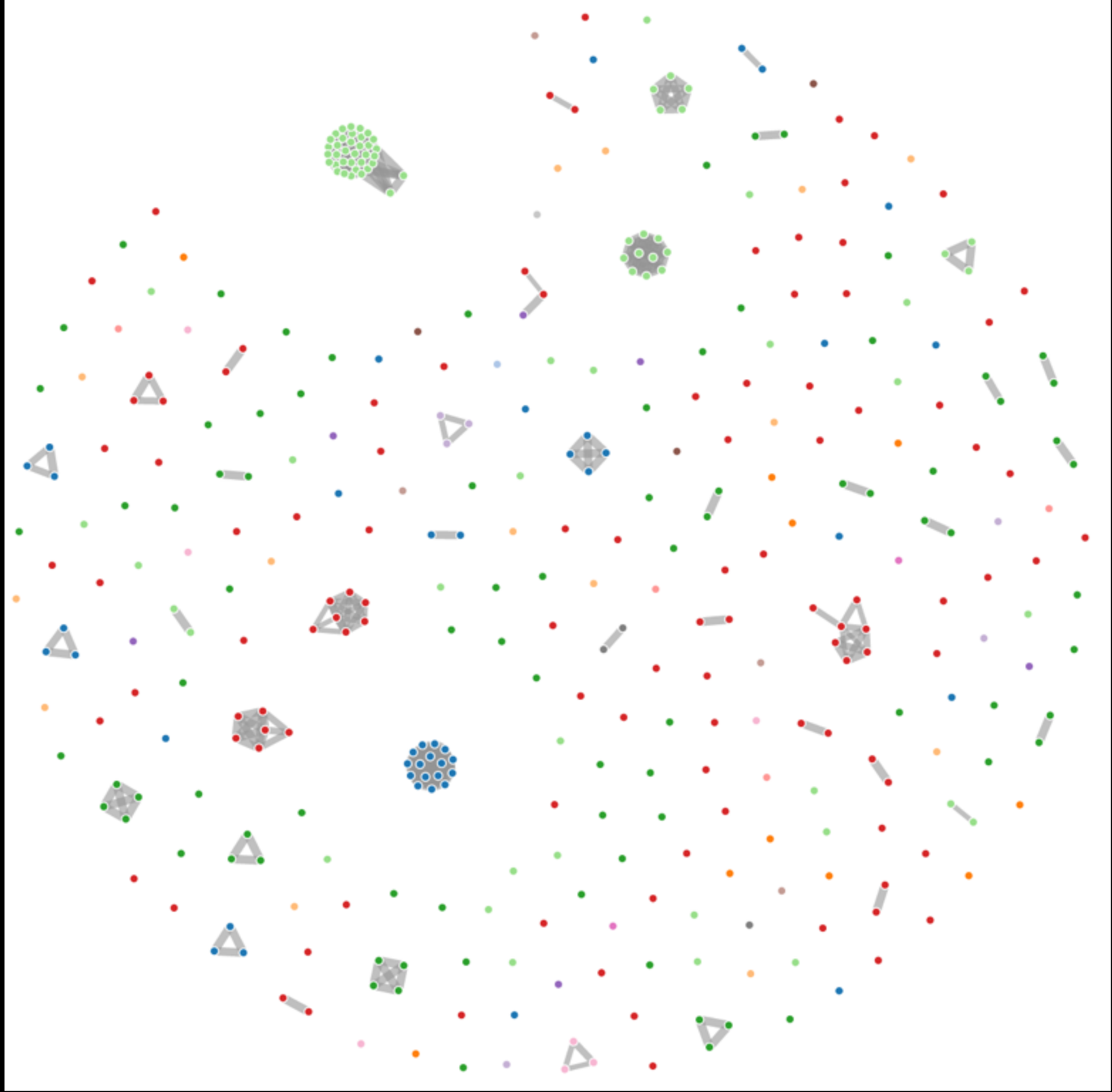
SSDEEP

SSDEEP

- 856 TRUE CORRELATIONS
- 0 FALSE POSITIVES
- **1 CORRELATIONS FOUND CONNECTING CAMPAIGNS BY THE SAME ACTOR**
- **SEVERAL CORRELATIONS BETWEEN MINOR VERSIONS OF SAME MALWARE**
- NO CORRELATIONS ACROSS PARTS OF THE KILL CHAIN WITHIN CAMPAIGN



SSDEEP

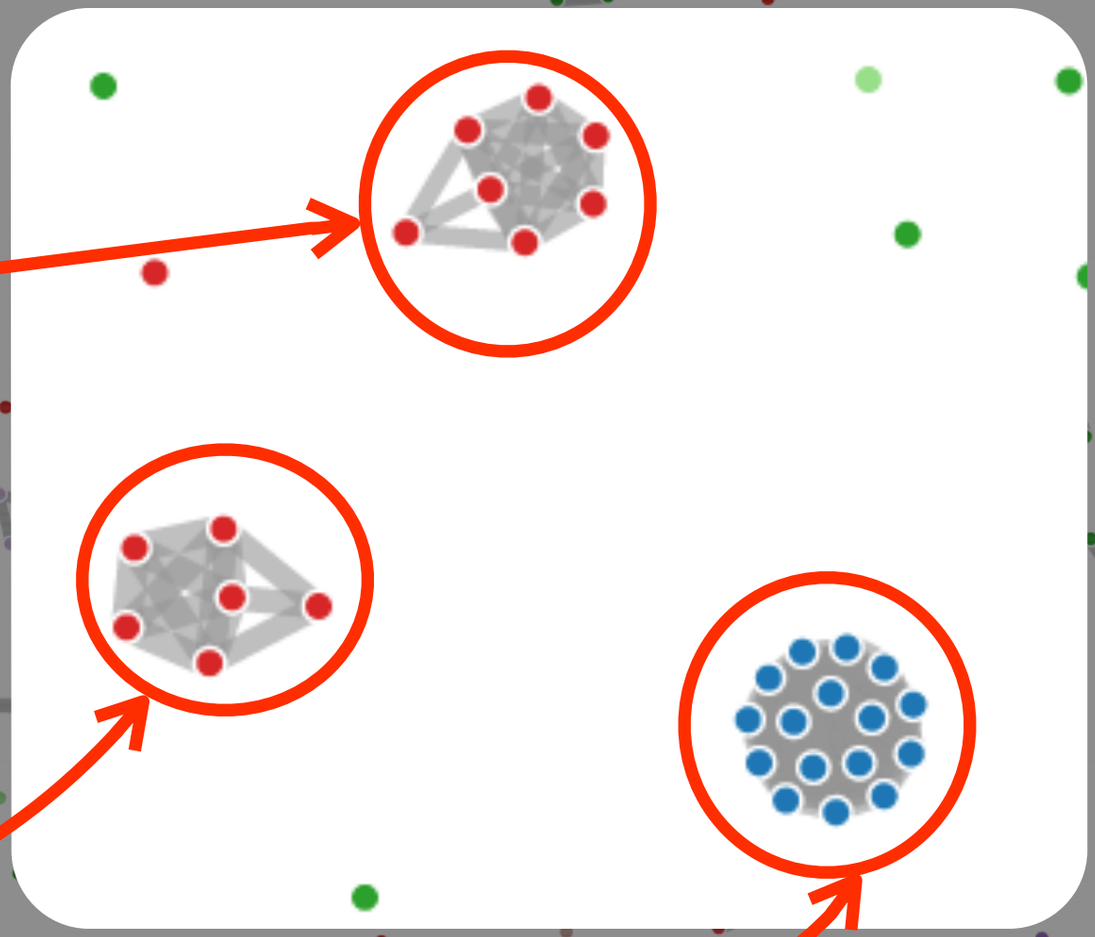


SSDEEP

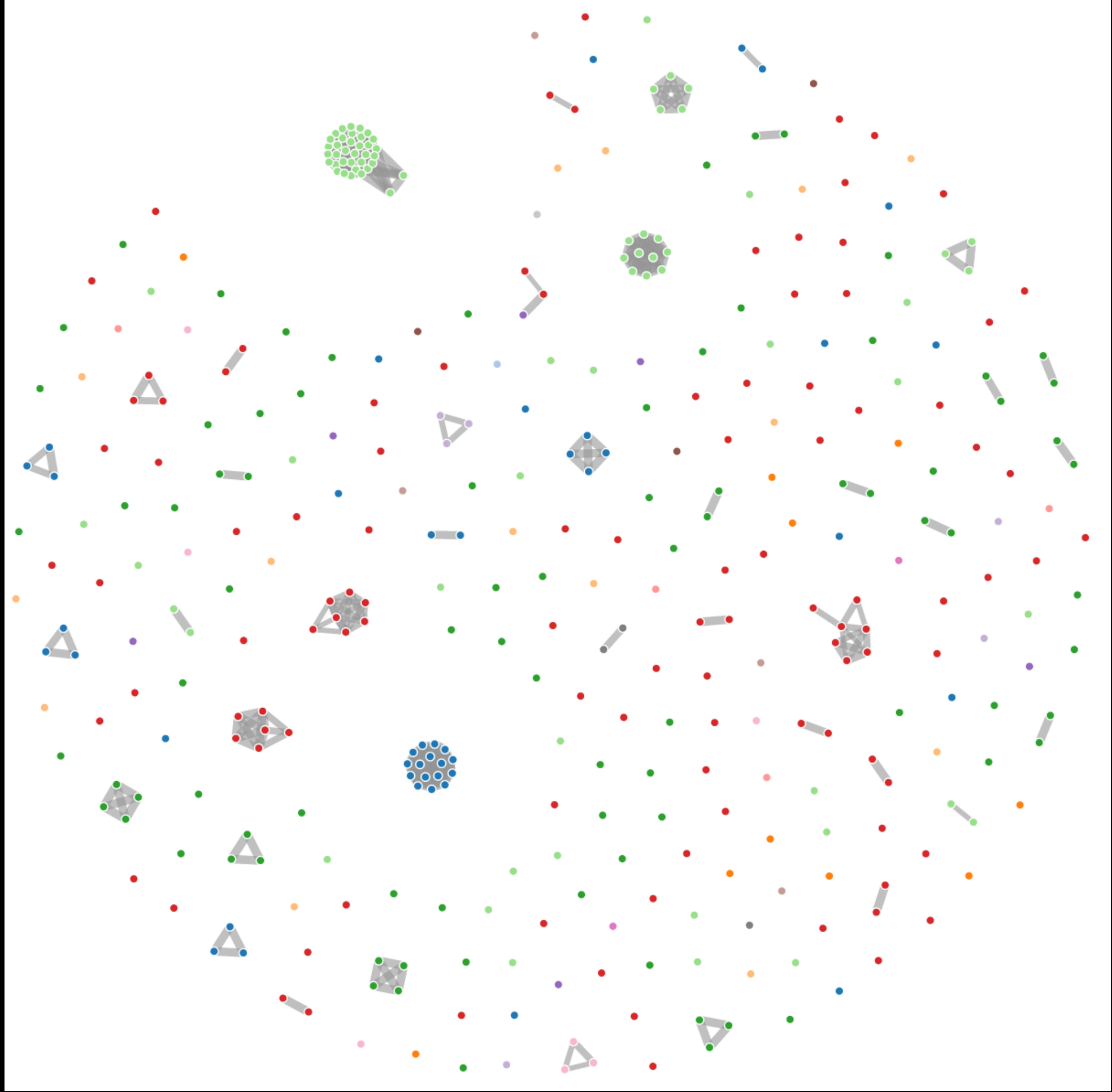
- SAV/Uruboros samples
- Used by the Waterbug Attack group
- Timestamped 2013

- Wipbot 2013
- Used by the Waterbug attack group

- Correlation across minor versions of ComRAT
- Compile dates span over 3 years

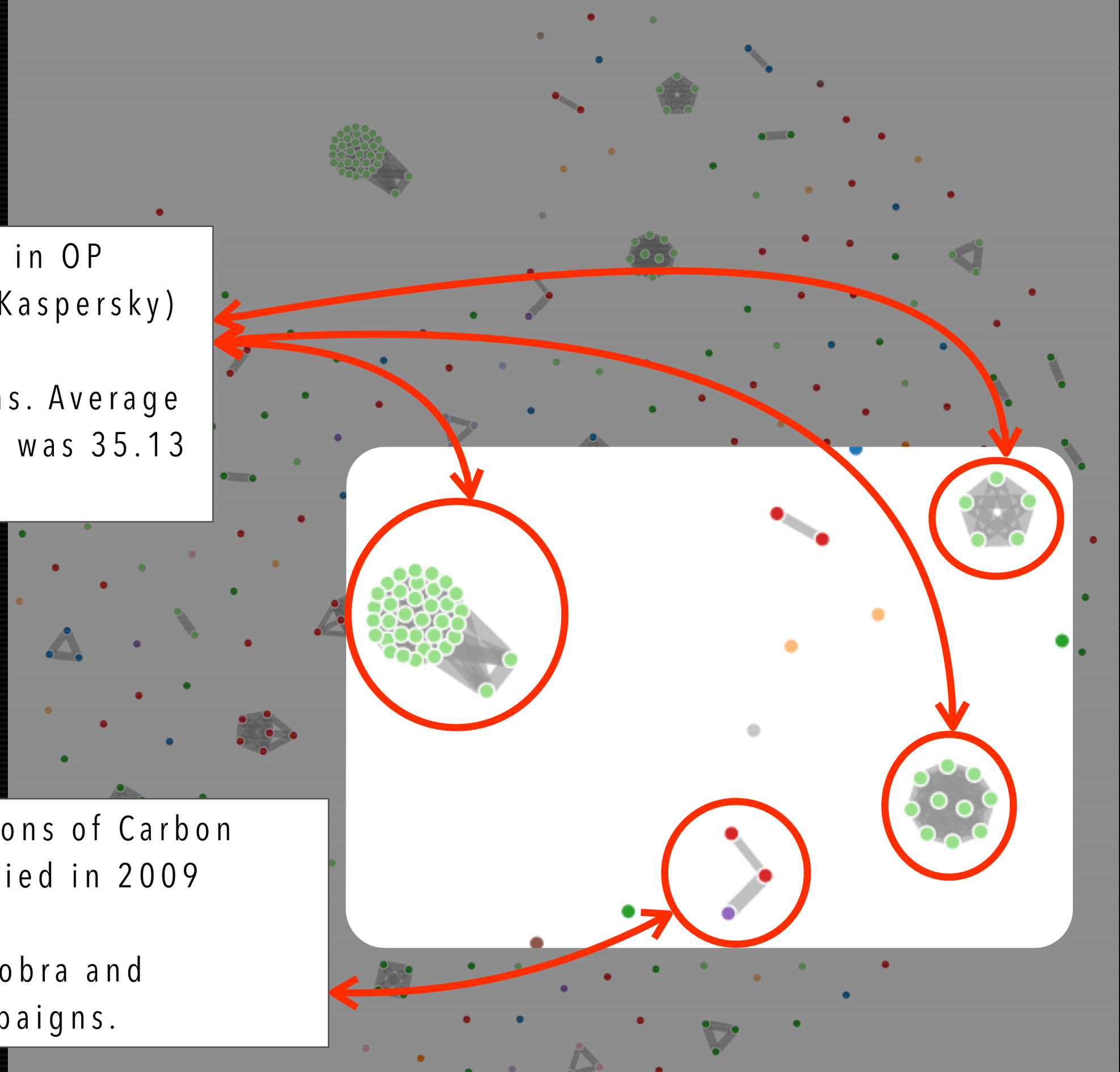


SSDEEP



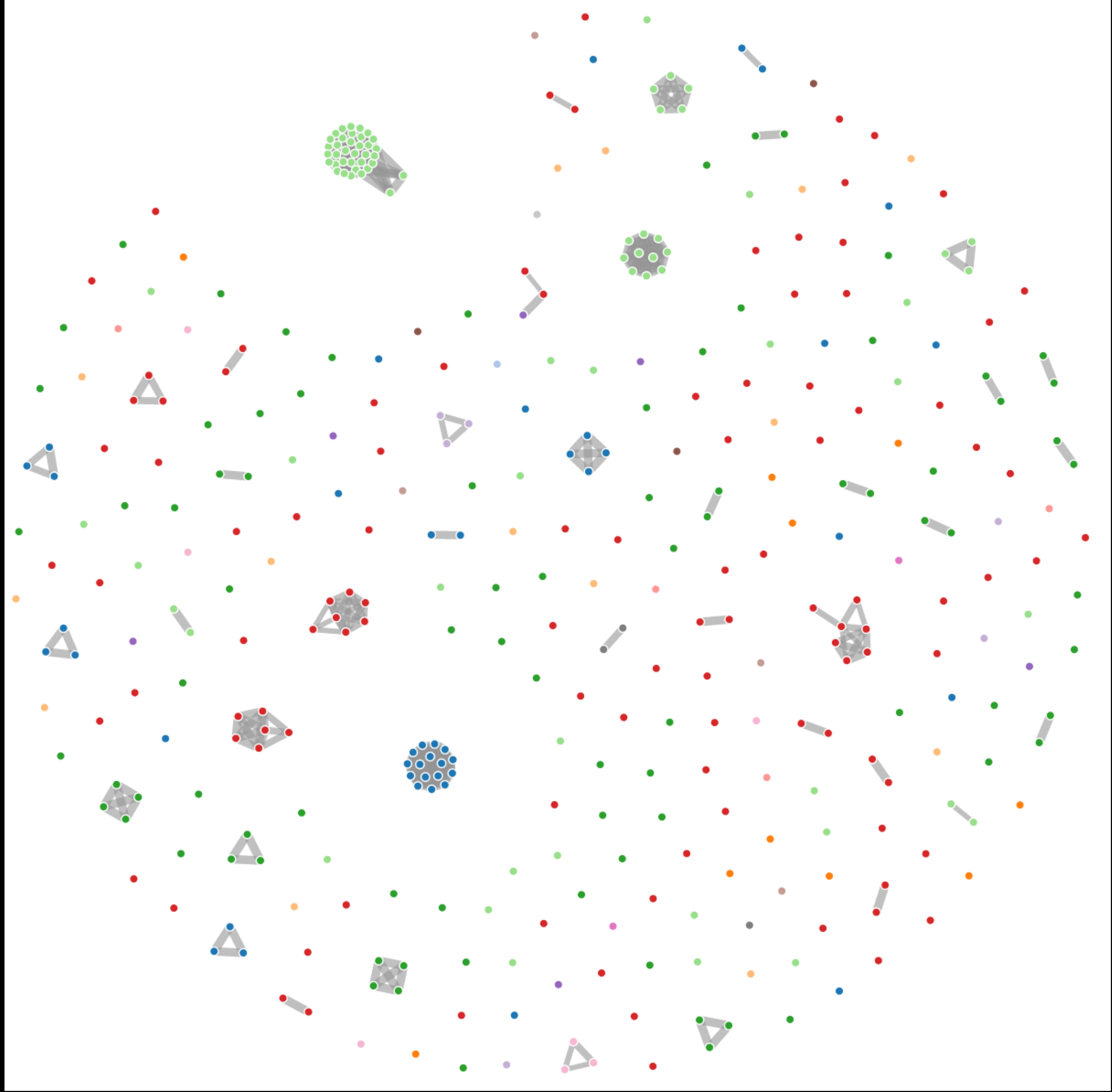
SSDEEP

- Backdoors used in OP Desert Falcon (Kaspersky)
- 630 Correlations. Average similarity score was 35.13



- Different Versions of Carbon Malware complied in 2009
- From Project Cobra and Waterbug Campaigns.

SSDEEP



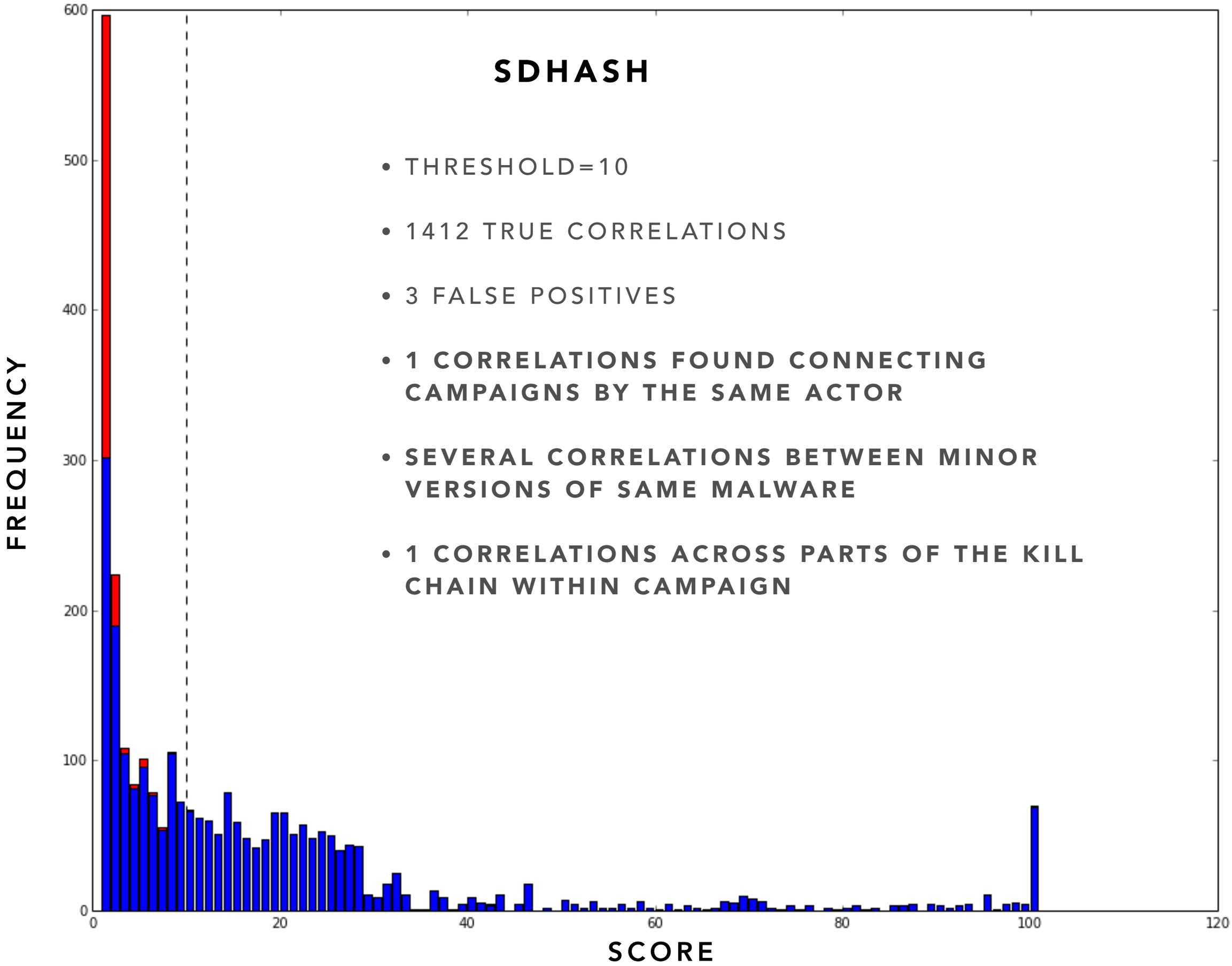
SSDEEP

NO FALSE POSITIVES

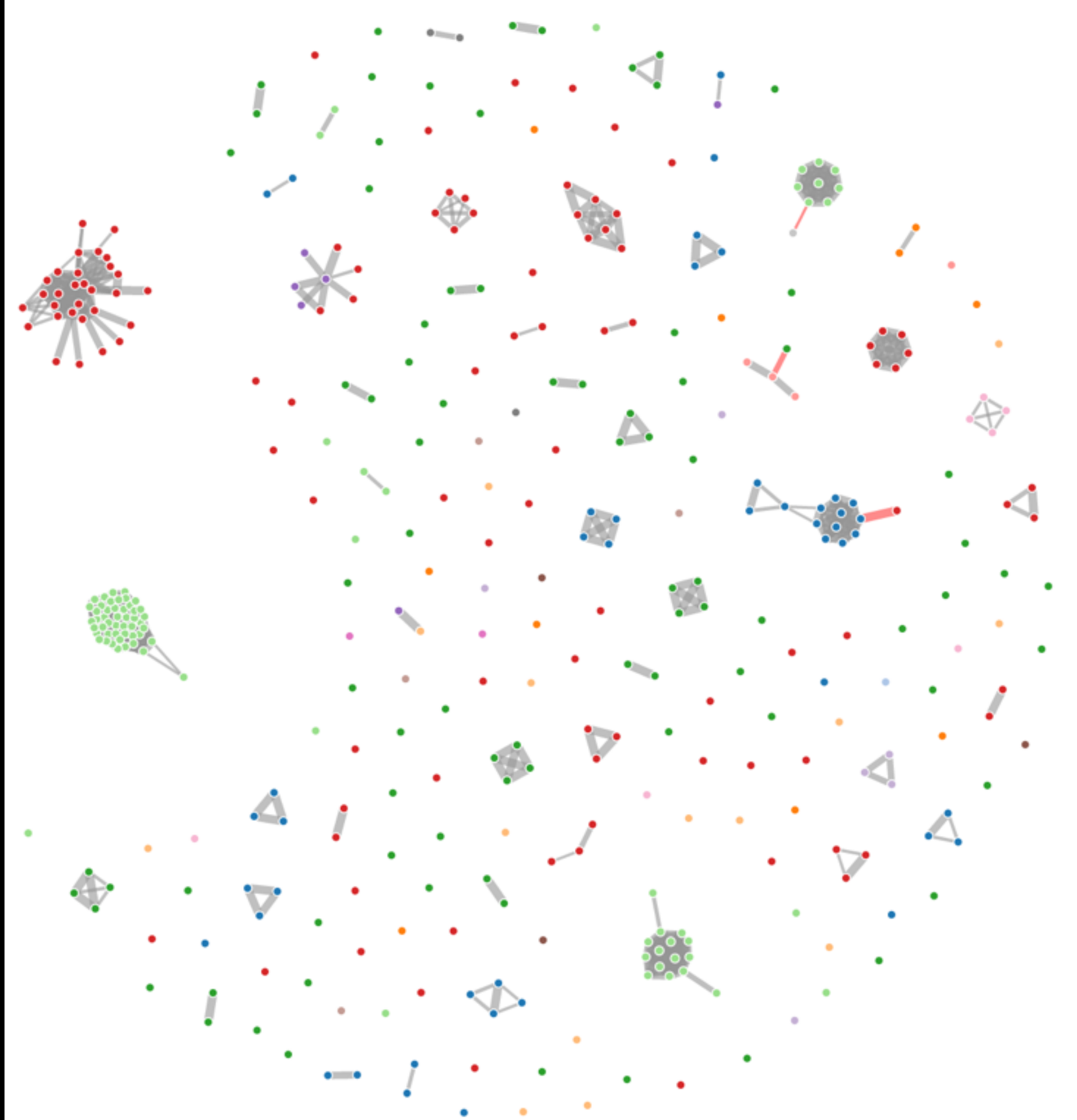
SDHASH

SDHASH

- THRESHOLD=10
- 1412 TRUE CORRELATIONS
- 3 FALSE POSITIVES
- **1 CORRELATIONS FOUND CONNECTING CAMPAIGNS BY THE SAME ACTOR**
- **SEVERAL CORRELATIONS BETWEEN MINOR VERSIONS OF SAME MALWARE**
- **1 CORRELATIONS ACROSS PARTS OF THE KILL CHAIN WITHIN CAMPAIGN**

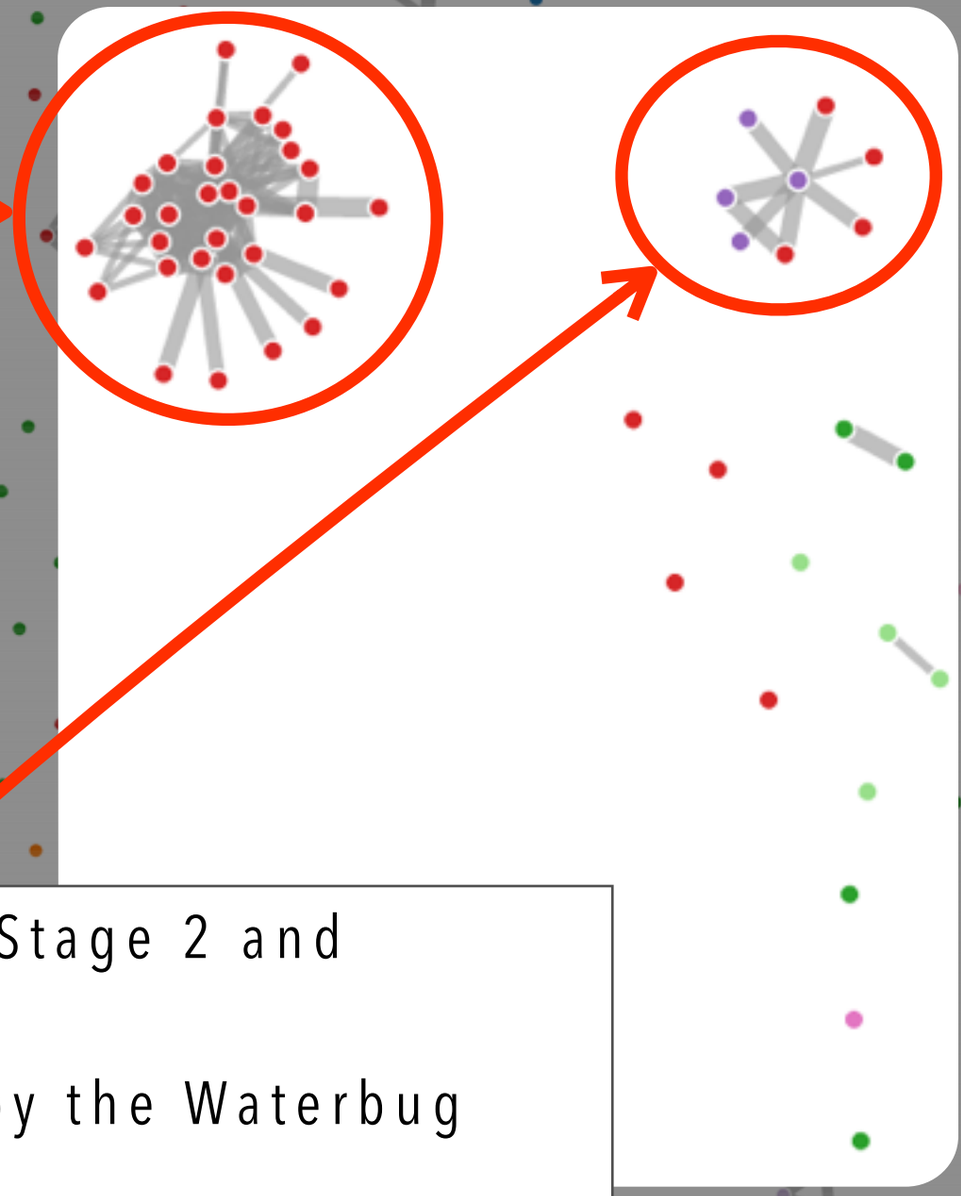


SDHASH



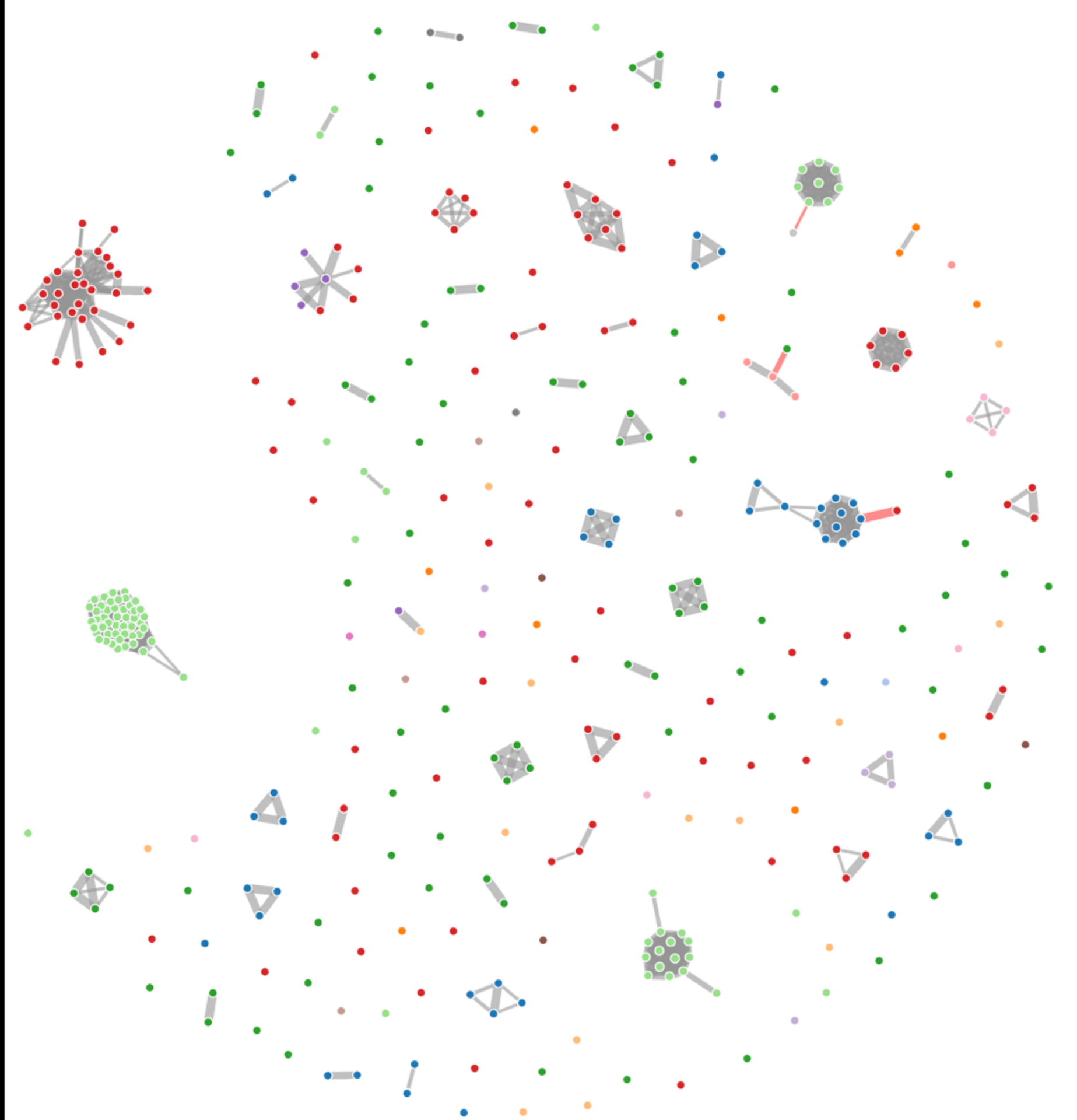
SDHASH

- SAV/Uruboros samples
- 30 different Binaries compiled over 3 months in 2013



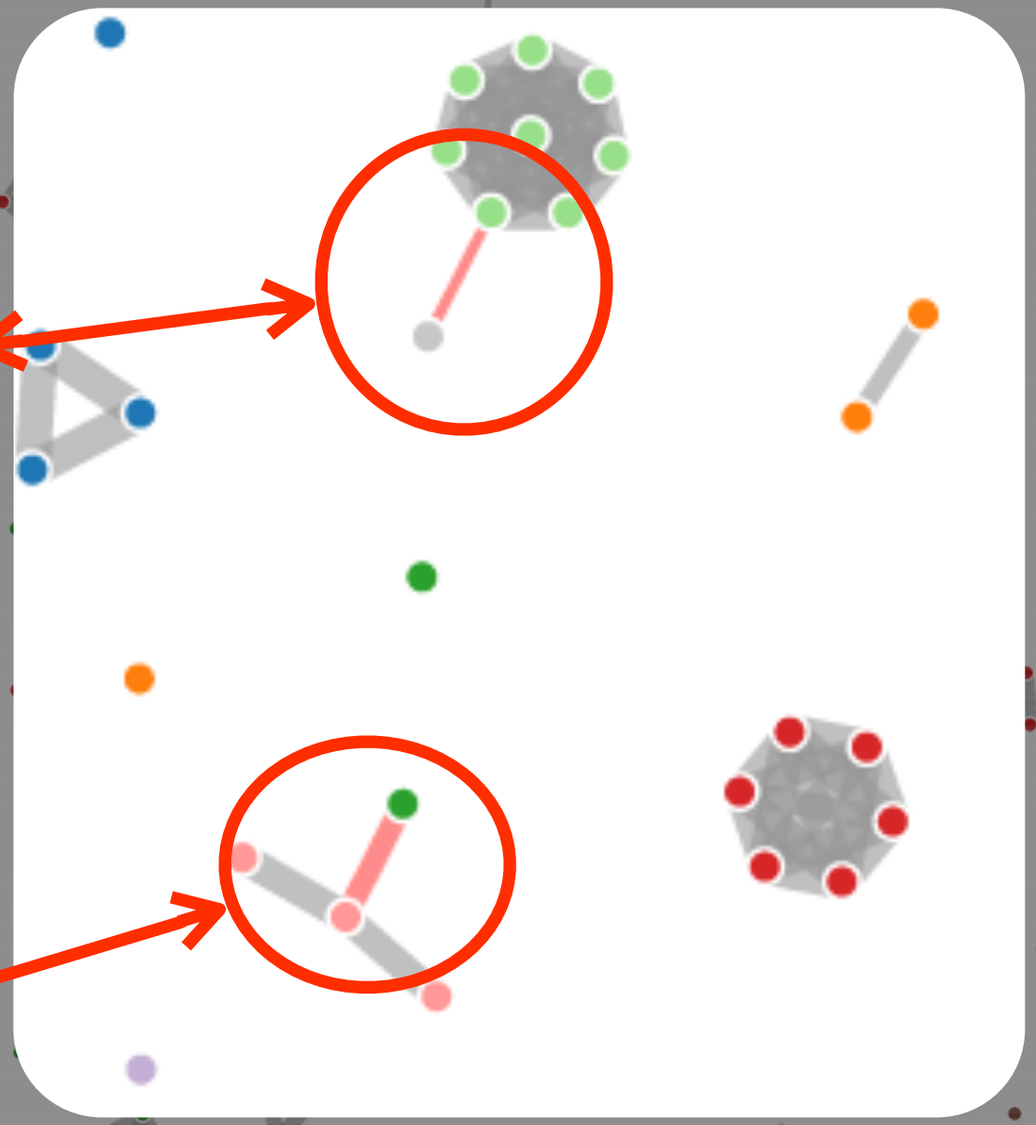
- Correlation between Dropper, Stage 1, Stage 2 and Injected Library of Cobra Campaign
- High similarity with Carbon Tool used by the Waterbug group
- Widely varying AV labels even controlling for vendor
- **Correlations made by sdhash only**

SDHASH



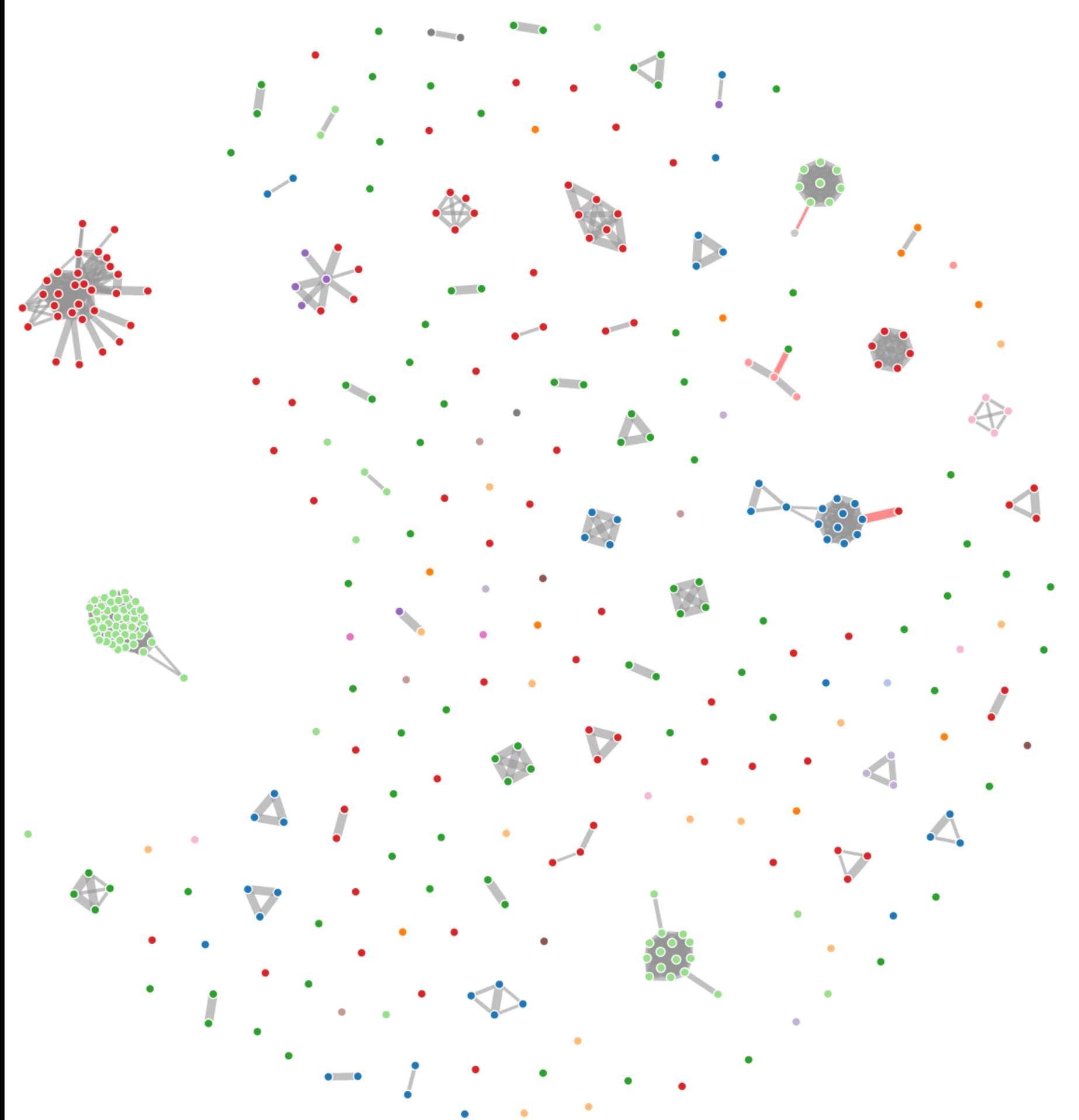
SDHASH

- Backdoor used by OP Desert Falcon
- Vs. Scanbox sample (known to be related to Anthem attacks and Deep Panda)
- No known relationship between those actors/campaigns/malware families



- "HttpBrowser"malware used in Anthem attack
- "AmmyAdmin" tool used by the Carbanak group

SDHASH



WHAT NEXT?

- Imphash, ssdeep, SDhash— which is best?
- It pays to know your adversary
- APT binaries may share code within campaign and actor— Code similarity can be used to connect binaries from the same source.
- Connections can help make strategic decision to respond to an adversary, NOT infection.

ACKS/Q&A/THANKS!

- Chris Kitto and Jeff Boerio for helping me make better slides.
- Wonderful folks that write Security white papers
- @kbandla for creating and maintaining APTNotes
- Virus Total for the great data they provide

@bsoman3, bhavna.soman@ {intel.com, gmail.com}