

USING MALTEGO TUNGSTEN™ TO EXPLORE THE CYBER-PHYSICAL CONFLUENCE IN GEOLOCATION

Shalin Hai-Jew
Kansas State University
SIDLIT 2014 (of C2C)
July 31 – Aug. 1, 2014

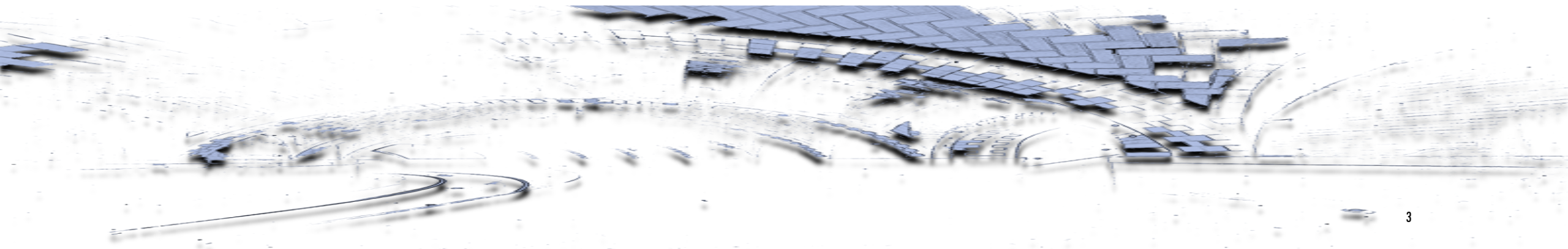
PRESENTATION OVERVIEW



Ever wonder how to map parts of cyberspace (the Surface Web) with the physical world? Maltego Tungsten™ (v. 3.4.0) is a high-end (penetration testing) tool that enables the mapping out of physical locations to cyber ones, and vice versa. Maltego Tungsten (formerly Maltego Radium) enables the identification of various types of location data: GPS (Global Positioning System) coordinates, locations to cities / states / countries, and even more precise geo-locations (like lat. / long.). Physical location data may be identified for accounts on social media platforms (like Twitter, Facebook, and others), websites, disambiguated names (even aliases), and other types of online data. The findings are presented in dynamic and static visual graphs and data tables. The information collected is all publicly available data known as “open-source intelligence” (OSINT). Understanding the ranges of such knowability (through both intended and unintended data leakage) is critical as part of digital literacy. This presentation will show how to move from the physical to the electronic and the electronic to the physical.

CYBERSPACE AS A LEVERAGEABLE “CHOKEPOINT”

- ❖ A tool to reach many for numerous human endeavors: self-expression, socializing (“the place to see and be seen”), inter-communications, commerce, banking, entertainment, learning, research, and political mass mobilization (among others)
- ❖ Hard (impossible?) to remain anonymous or invisible (over time)
- ❖ A (cyber)space where one is vulnerable to identification and interception (to use a geographical metaphor)



@ ... CYBER-PHYSICAL CONFLUENCE

Cyber

Active engagement: email, microblogging (Tweeting), image-sharing, tagging, messaging, purchasing, posting to a social network site, downloading or accessing information, hosting a website

Passive engagement: mobile devices connecting with communications towers, EXIF data collected through digital cameras

Physical

Space-time (geolocation and local time and standard time)

Embodied interactions in space



CYBER-PHYSICAL SPILLOVERS

Cyber

- ❖ Animating messages (fictional and nonfictional; multimedia; by humans and 'bots and cyborgs)
- ❖ Social contagions and cascading events
- ❖ Online socializing and virtual community building



Physical

- ❖ Mental and emotional: Individual and group radicalization (human suggestibility, particularly among vulnerable individuals); sense of identity
- ❖ Physical: Flash mobs; social activism; fund-raising; inter-relationships; meet-ups, and others

CYBER-PHYSICAL SPILLOVERS (CONT.)

Cyber



- ❖ Shared photos, slideshows, audio, and video from conferences on content-based social media platforms
- ❖ Dedicated websites and microsites
- ❖ Real-time microblogging messaging
- ❖ Online sharing based on real-world relationships
- ❖ Data capture and leakage

Physical

- ❖ Real-world events that are expressed and shared (and reported on) electronically
- ❖ Real-world communities
- ❖ Human intercommunications and documentation (like mobile phone communications and digital photography)
- ❖ Human transportation
- ❖ Human shopping

MAIN IDEAS



- ❖ There is no real cyber that exists separate from physical space. All cyber activity may be linked back to physical space (with publicly available and easily usable software tools). It is totally possible to link many aspects of cyber to physical spaces.
- ❖ Using network analysis (and linkage), most cyber activity may be linked to an actual person often in a general (sometimes specific) location.
- ❖ In the cyber-physical confluence, each aspect (both the cyber and the physical) sheds light on the other; aspects of each domain are used to leverage and augment the other.

WHY PHYSICAL SPATIALITY MATTERS

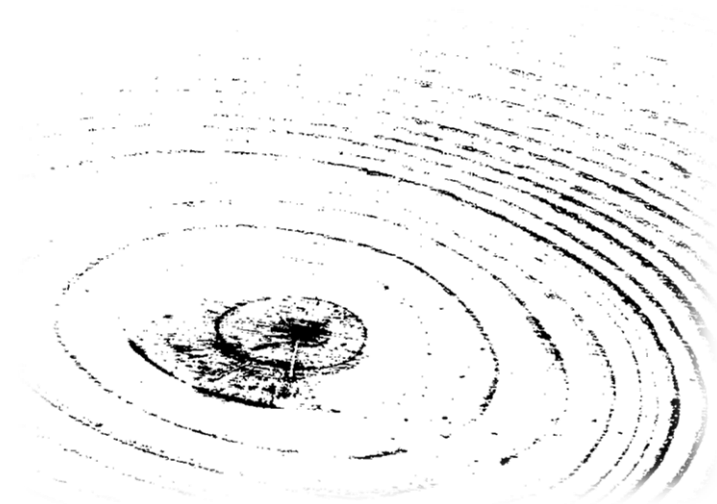
- ❖ Once a physical location or sequence (or map) of locations is identified...
 - ❖ **Additional Data:** A lot of publicly available information about that locale is available, such as demographics, lifestyles, earnings, dominant languages, government, law enforcement, resources, and others.
 - ❖ **Layered Mapped Data:** A lot of map layers (overlays) may be applied to the location, indicating infrastructure like roads and natural spaces, built and natural spaces, and other static data.
 - ❖ **Locations and Human Activities:** Certain human activities are tied to certain physical spaces, so assumptions and inferences may be made about what a person or persons are doing in a particular location at a particular time.
 - ❖ **Relationship Modeling:** A person's co-location in space-time with another or others may suggest affinity, inter-communications, or collusion.

WHY PHYSICAL SPATIALITY MATTERS (CONT.)

- ❖ **Lifestyle Patterning:** Over time, a person's regular appearance at certain locations may indicate a lifestyle. An anomalous and transitory appearance in a certain non-habitual location may also be indicative of yet something else.
- ❖ **Inferred Actions and Intentions (reverse-engineered from sequences of locations and times):** A particular sequence of space-time presences may suggest a particular sequence of actions...and may be suggestive of intentions from the inferred actions. (In the parlance, this is a kind of "inference attack.")
- ❖ **Direct Contacts:** If (malicious, benevolent, or neutral) actors have an interest in a physical encounter with a target, knowing space-time patterns may be a critical part of the surveillance and real-space interactions. Applied in an adversarial way, geolocation awareness of a quarry ("rabbit") may be high-risk.

WHY PHYSICAL SPATIALITY MATTERS (CONT.)

- ❖ **Real-time Data:** Further, dynamic data may be applied showing traffic patterns at different times of day, for example. A virtualized simulation of the real-space may be created with data. With satellite imagery accessible, real-time real-space aerial imagery may be accessible as well. If a person is being sought in real-time, there are certain rules about how far people can move on foot or by vehicle, but the last known location may be richly informative for those striving to locate a person.



ON HUMAN PATTERNING IN PHYSICAL SPACE

❖ **Identification of Data to a Person:** Massachusetts Institute of Technology (MIT) researchers who studied the cell phone records for 1.5 million people found that “for 95 percent of the subjects, just four location data points were enough to link the mobile data to a unique person” (Tucker, 2014, pp. 18 – 19). More than 9 percent of Americans are part of a “geo-social network,” with plenty of data leakage from malware in apps and mass data collection by service providers in order to provide the service—such as pings off cell towers (Tucker, 2014, pp. 21 – 22).

ON HUMAN PATTERNING IN PHYSICAL SPACE (CONT.)

❖ **The Company You Keep:** Researchers have found that even when people choose to go geospatially incognito, by turning off all signaling, they may still be tracked based on the behaviors of their friends and acquaintances (Tucker, 2014, pp. 25 – 26), through a kind of geospatial social network analysis. The author suggests that while people can lower the detectability of their prediction level to others, their efforts do not make them less predictable—because the predictability comes from themselves: “Your life pattern is you. It’s what you do, with whom, and where. It’s the content that fills the vessel of your existence” (Tucker, 2014, p. 29).

❖ **Predictivity:** Further, through a method of “eigendecomposition” from sensor data (collected from subjects over six years), researchers were able to create a model “that could predict a subject’s location with higher than 80 percent accuracy up to eighty weeks in advance” (Tucker, 2014, p. 27) or a year and a half into the future.

“CONVENIENCE SAMPLING” OF DIGITAL DATA

- ❖ “Big data” collections of public (and accessible) information
- ❖ Digitized data points (without additional digitization necessary), which lowers the cost of knowing (even when there are other ways of knowing some of the same information)
- ❖ Ways to leverage (visualize and analyze) metadata and data
- ❖ Simple-use software tools for data extractions and queries
- ❖ Extracting data to serve as part of a “human sensor network” through social media platforms and other mass data collection points
- ❖ Use of physical location (unique identifier) as a way to align mixed data sets



You are Here!

WELCOME! SELF-INTROS!

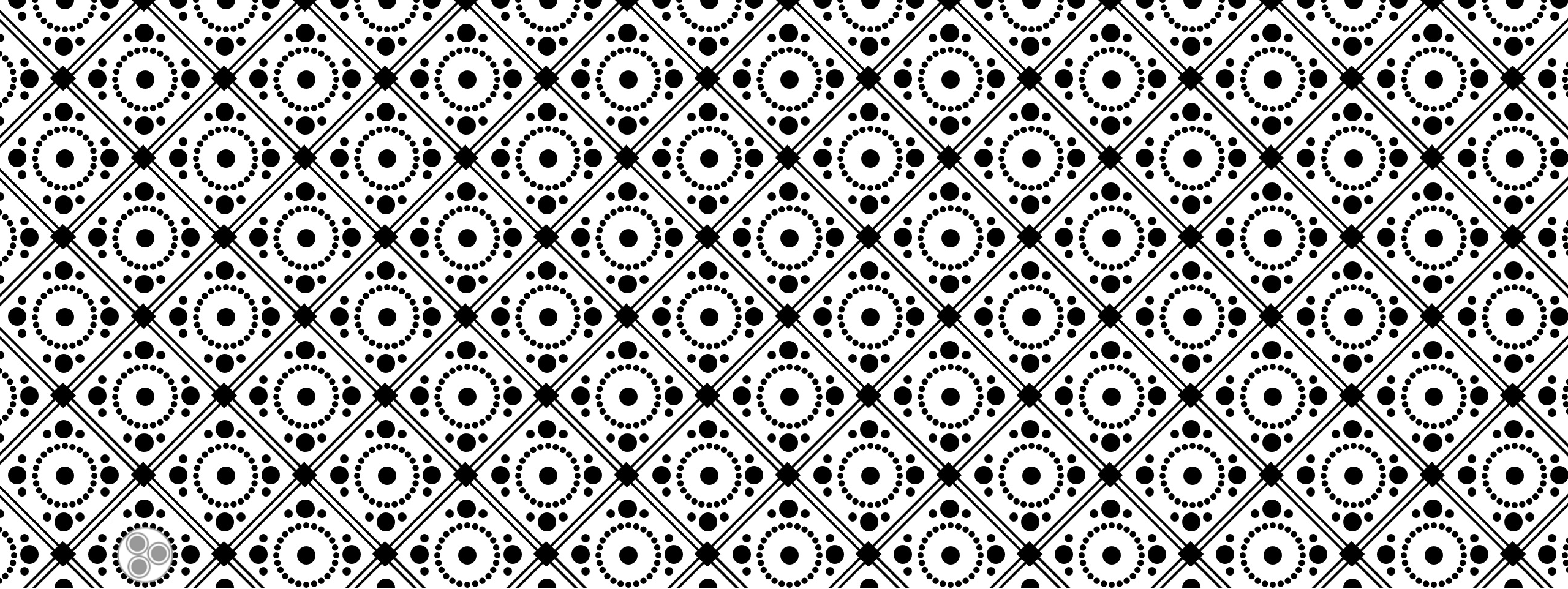
Who are you?

What are your interests in this presentation? What would you like to learn?

What are your experiences with Maltego Tungsten™ (if any)?

What are your experiences with geographical information systems (GIS)? Global Positioning Systems (GPSes)? Mapping?

Clarification: This presentation is designed to introduce geolocational capability, not convey how to directly conduct data crawls and analyses using Maltego Tungsten™. Further, what will be shown is a limited part of the software capabilities only. The tool has greater range than what will be shown.



HOW THIS WORKS... BROADLY SPEAKING

To move from cyberspace to
physical space

DEFINITIONS OF TERMS

Cyber-physical confluence: The overlap between electronic and real-world physical lives

Geolocation: The method for identifying the geographical location of a person or device from Internet-based information

Maltego Tungsten™: A penetration testing software tool (with a limited public version, which limits # of nodes and which goes to a public server for the extraction and which only is activated for a few days) made by Paterva

OSINT (open-source intelligence): Publicly available data that may have informational value

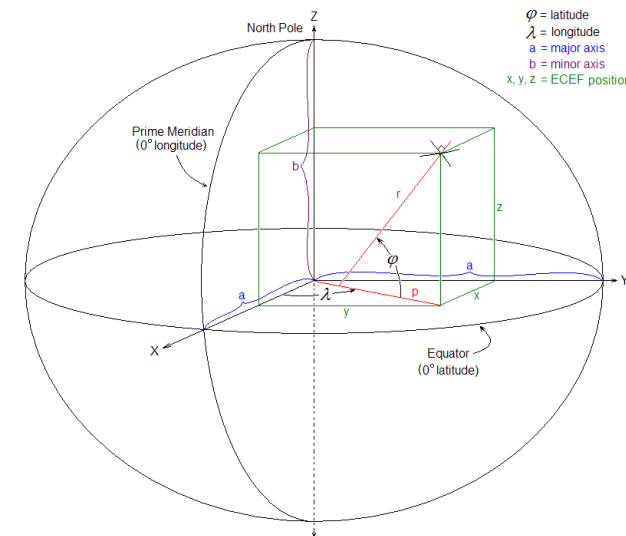
GIS (geographical information system): A computer system that is used to capture, store, manage, visualize, and analyze geographical information

“Machines”: A sequence of code which enables designed data extractions from the Web (in Maltego)

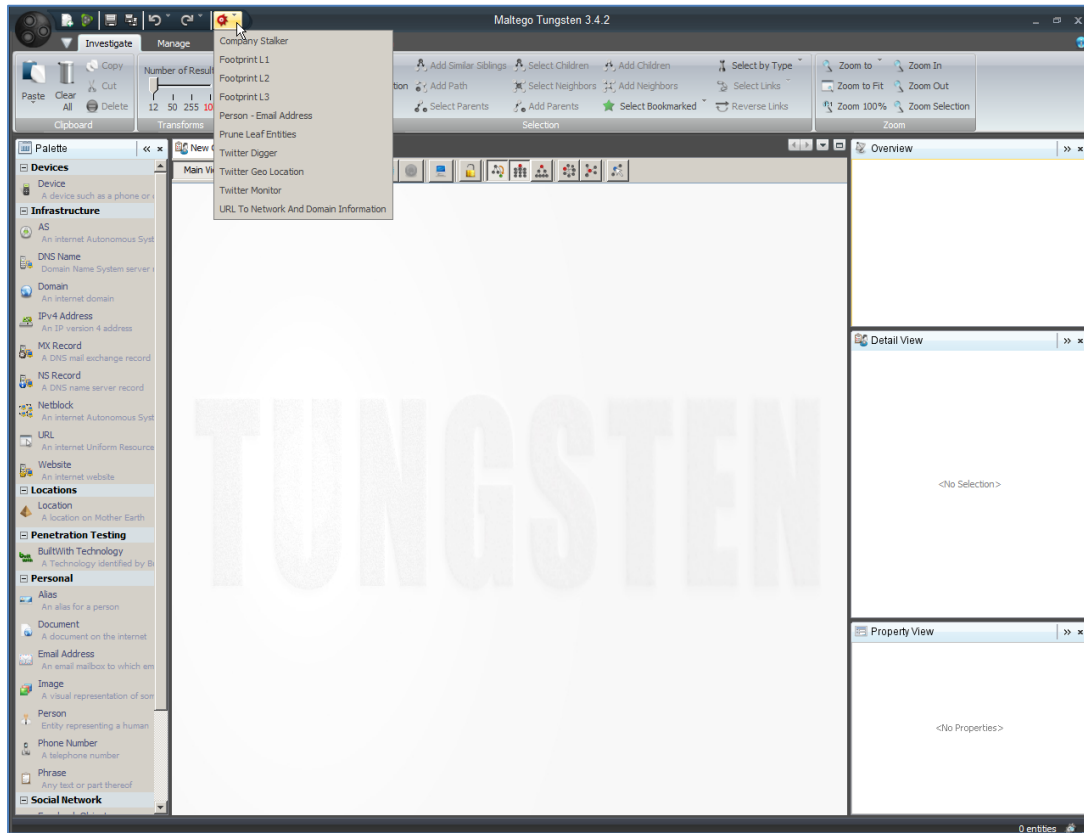
“Transforms”: A sequence of code which “transforms” one type of electronic data to other types (in Maltego)

GENERAL TYPES OF GEOGRAPHICAL DATA (ON SOCIAL MEDIA PLATFORMS)

- ❖ zip codes
- ❖ area codes / dialing codes
- ❖ lat(itude) / long(itude) / elevation or altitude or “sea level”
- ❖ x, y, and z coordinates (based on the ECEF or “Earth-Centered, Earth-Fixed” or ECR or “Earth Centered Rotational” system)
- ❖ city, state; province, state
- ❖ textual or alphanumeric descriptors
- ❖ and others



MALTEGO TUNGSTEN™



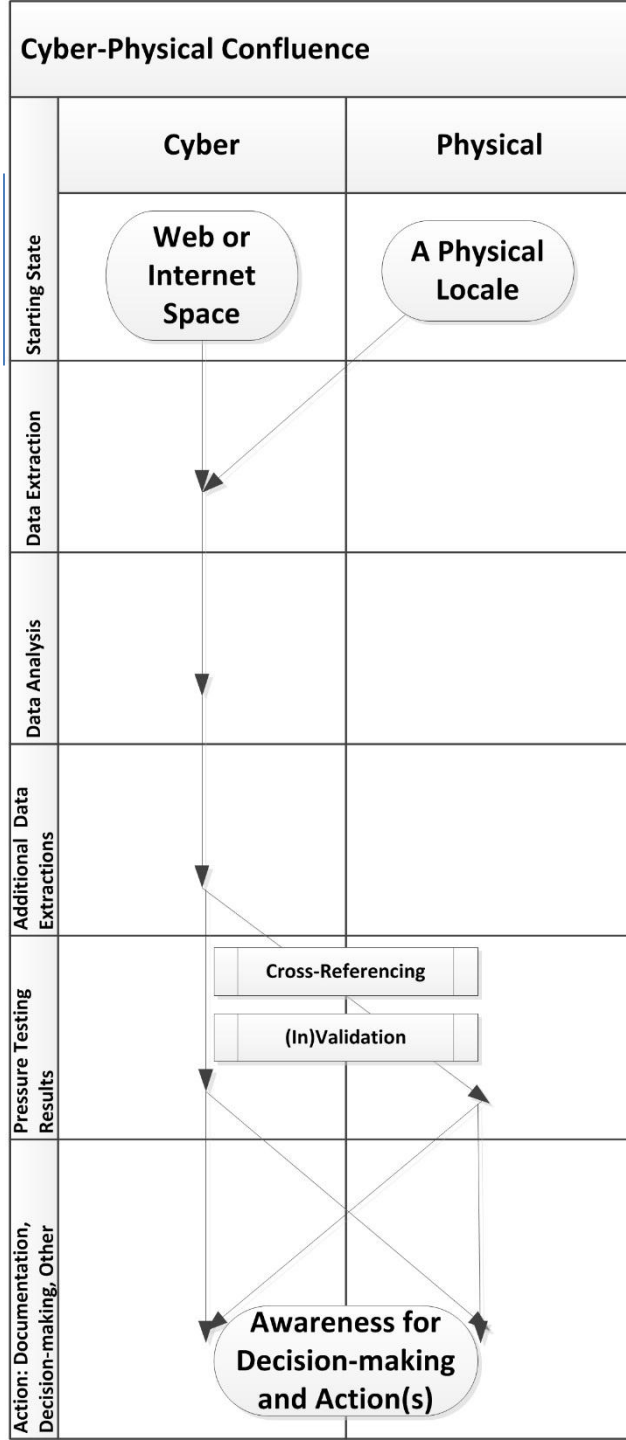
- ❖ A penetration testing tool with...
 - ❖ access to publicly available information on the Web and Internet (including long-ago past data if it is hosted on a server)
 - ❖ “machines”—sequenced codes that enable targeted data extraction of public information from the public Web and Internet
 - ❖ “transforms”—code snippets that enable the “transforming” of one data type to another to chain information
- ❖ may be automated with macros and code
- ❖ may be re-run over time (with the same initial parameters or updated ones)

MALTEGO TUNGSTEN (CONT.)

- ❖ A penetration testing tool with... (cont.)
 - ❖ multi-lingual functionality
 - ❖ data visualization capabilities (2D and 3D graphs)
 - ❖ exports as data tables and reports
 - ❖ For more, you can go to [“Maltego Radium™: Mapping Network Ties and Identities across the Internet”](#)
- ❖ For this presentation, all data extraction runs will be achieved with a limit of 12 results...just to ensure that the dataset and graphs are not totally unwieldy. These are all fairly cursory crawls. More in-depth work would likely result in more effective findings (but would also require much more in the way of human analysis).

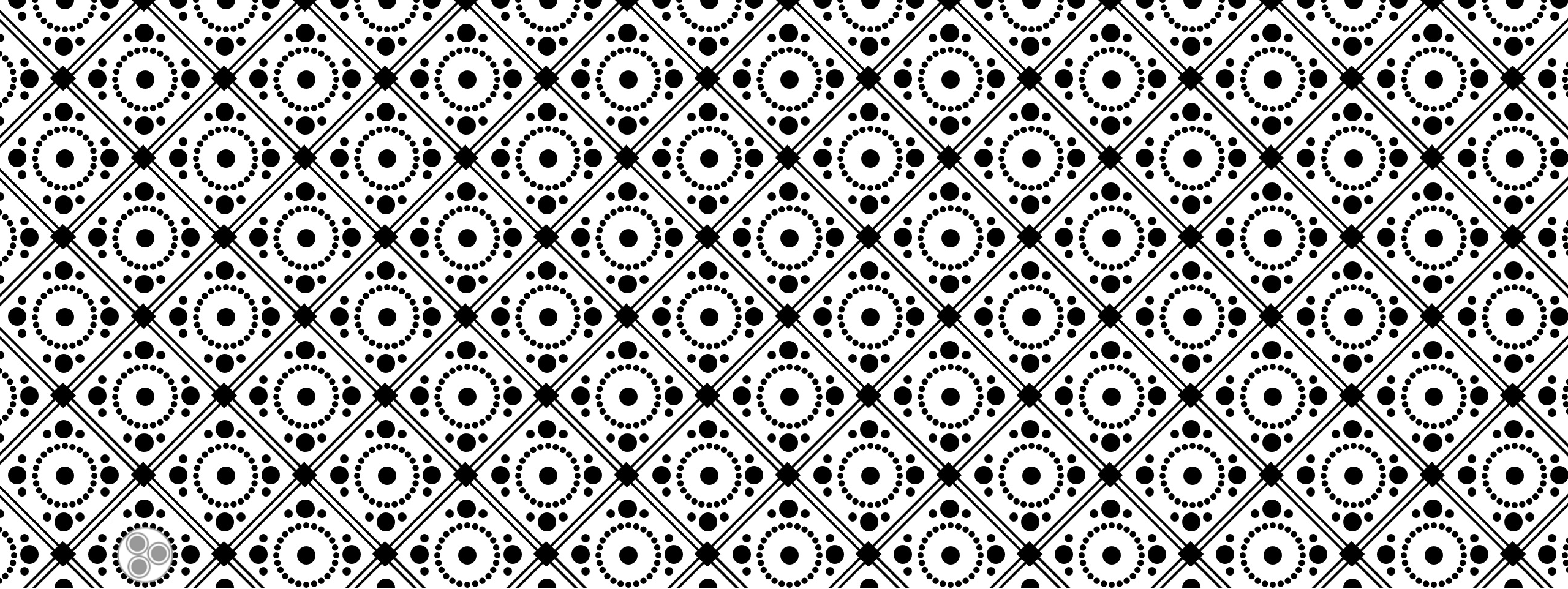
SOFTWARE TOOL ENABLEMENTS BASED ON CONNECTIVITY AND PATTERNS

- ❖ Traffic analysis (interactivity by whom, but not when and not mass and not message contents) > social structure > function of the social networks
- ❖ Data extraction for Internet network visualizations
- ❖ Transcoding between online information types (alias, document, email address, image, person, phone number, phrase, social network accounts, locations, Internet domains, devices)
- ❖ Network visualizations of extracted information



A SWIMLANE (CROSS-FUNCTIONAL) DIAGRAM OF PROCESS

- ❖ **The (porous) swimlanes:** The cyber lane; the physical lane
- ❖ **Phases:** starting state, data extraction, data analysis, additional data extractions, pressure testing results, action (documentation, decision-making, other)
 - ❖ May require a lot of additional work in disambiguation and clarifying
 - ❖ May result in leads for further research
- ❖ Suggests that there is a gap between the cyber and physical and limits of what each can show about the other



EXAMPLES OF MALTEGO TUNGSTEN™ CRAWLS FROM THE REAL

From popular “targets” in these
use cases...

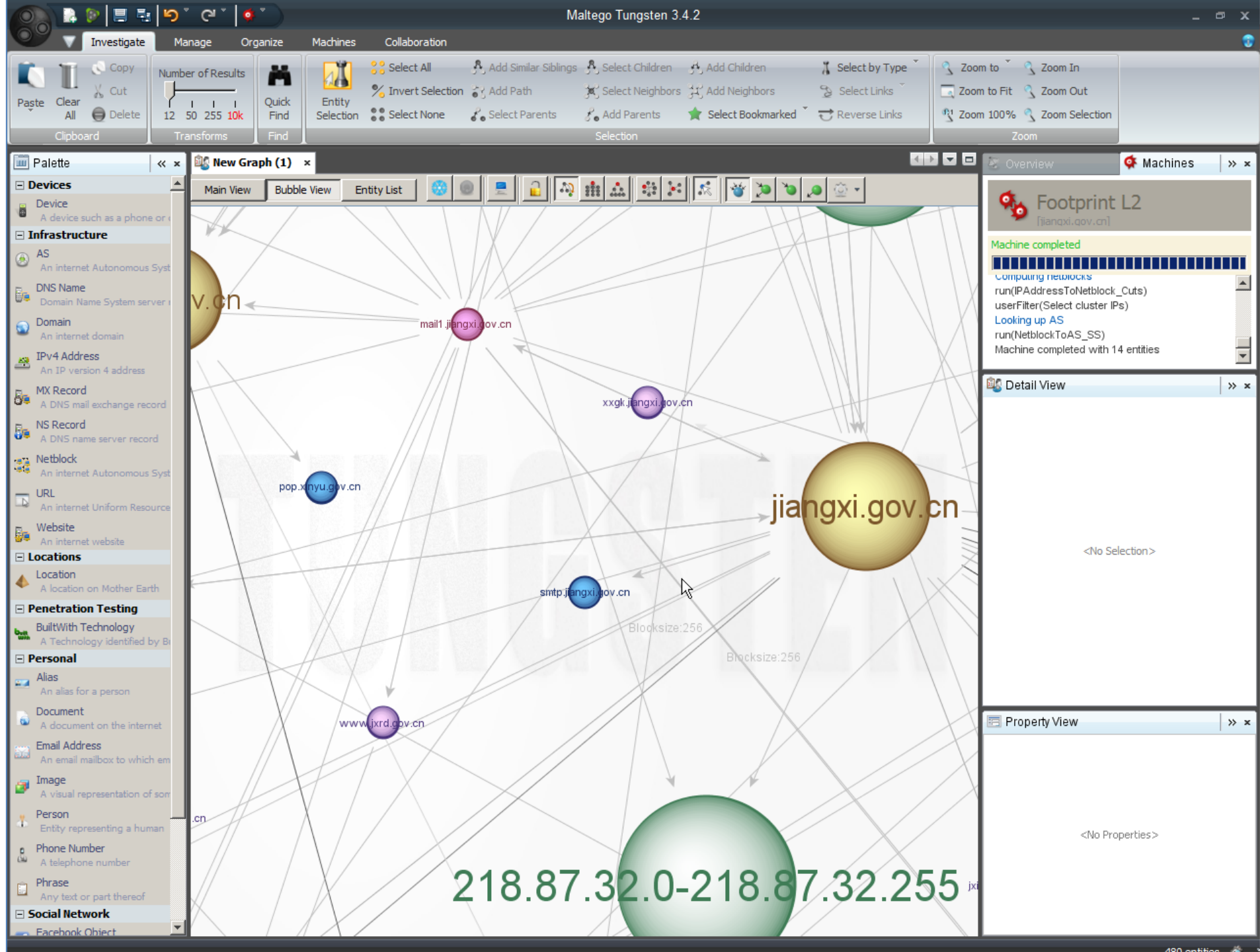


MAPPING PHYSICAL SPACE TO CYBERSPACE

To get a sense of the main
cyber players in a physical
locale

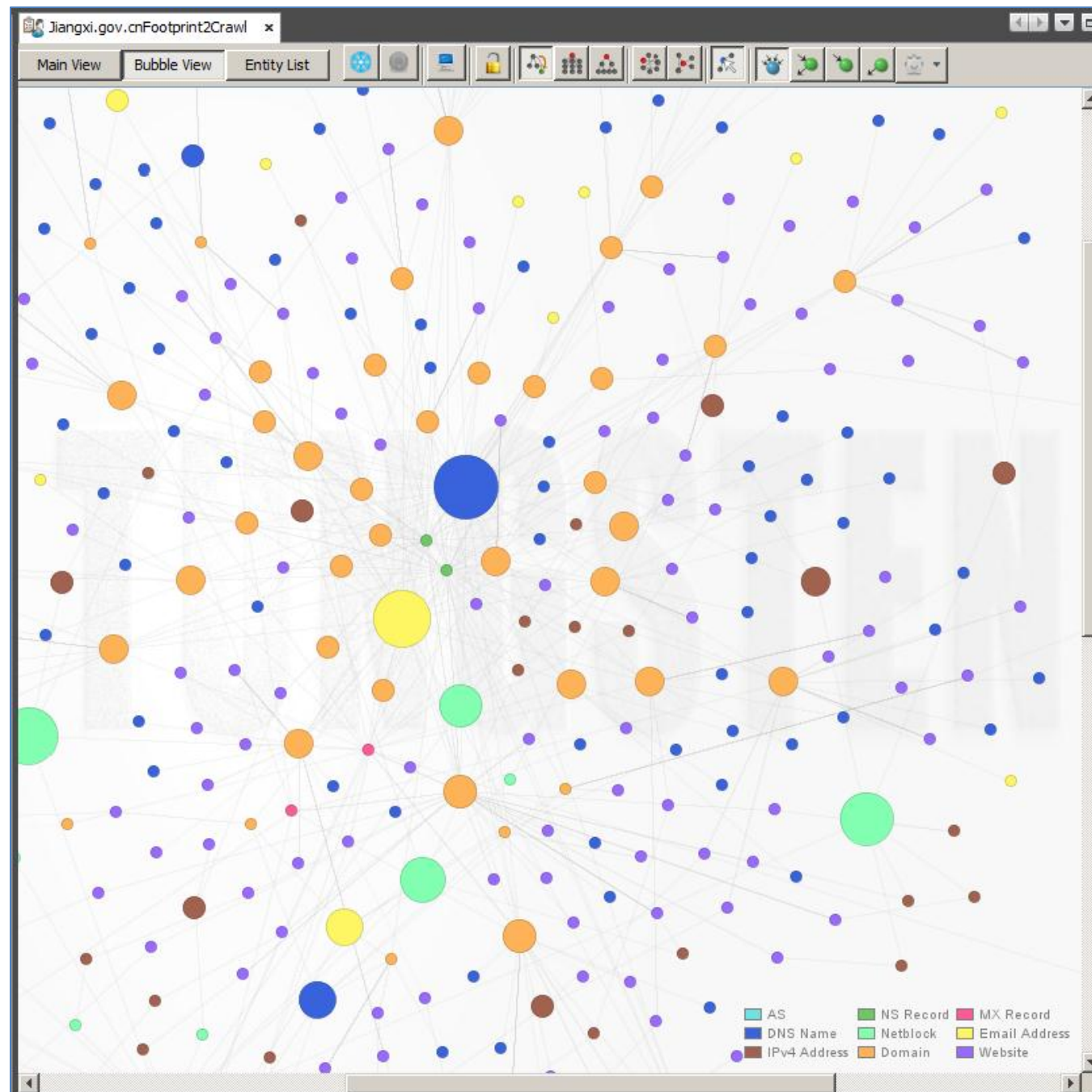
EXTRACTION SEQUENCE

1. Choose a physical locale. Choose what part of the domain you're interested in (such as government or commercial or other).
2. Conduct a Google search of that locale. Identify the relevant domain.
3. In Maltego Tungsten, select one of the footprinter levels (1 is light and 3 is intensive).
4. Conduct the Web domain footprinting. Respond to the “user filter” queries. Prune the graph as necessary.
5. Visualize and re-visualize the data in various ways. Analyze results.



JIANGXI.GOV.CN

- ❖ AS numbers
- ❖ DNS name
- ❖ IPv4 address
- ❖ NS Record
- ❖ Netblock
- ❖ Domain
- ❖ MX record
- ❖ Email address
- ❖ Website



ENTITY LIST

- ❖ Nodes (types) / entities
- ❖ Links (with relational weights)
- ❖ Exportable file for uses in other software programs

Maltego Tungsten 3.4.2

Investigate Manage Organize Machines Collaboration

Clipboard: Paste, Clear All, Cut, Delete
Transforms: Number of Results (12, 50, 255, 10k), Quick Find
Selection: Select All, Invert Selection, Select None, Add Similar Siblings, Add Path, Select Parents, Select Children, Add Neighbors, Add Parents, Select by Type, Select Links, Reverse Links, Zoom to, Zoom In, Zoom to Fit, Zoom Out, Zoom 100%, Zoom Selection

Palette: Devices, Infrastructure, Penetration Testing, Personal, Social Network

Entity List View

Nodes	Type	Value	Weight	Incoming	Outgoing	Bookmark
jiangxi.gov.cn	Domain	jiangxi.gov.cn	93	4	28	★
dns.jiangxi.gov.cn	NS Record	dns.jiangxi.gov.cn	100	1	40	★
jxdns.jiangxi.gov.cn	NS Record	jxdns.jiangxi.gov.cn	100	1	35	★
mail1.jiangxi.gov.cn	MX Record	mail1.jiangxi.gov.cn	100	1	19	★
mail.jiangxi.gov.cn	MX Record	mail.jiangxi.gov.cn	100	1	7	★
jxinet.gov.cn	Domain	jxinet.gov.cn	100	3	6	★
jxjx.gov.cn	Domain	jxjx.gov.cn	100	2	0	★
jxzs.gov.cn	Domain	jxzs.gov.cn	100	2	4	★
jxdaj.gov.cn	Domain	jxdaj.gov.cn	100	3	8	★
jxfy.gov.cn	Domain	jxfy.gov.cn	100	3	8	★
yingtan.gov.cn	Domain	yingtan.gov.cn	100	3	10	★
jxlight.gov.cn	Domain	jxlight.gov.cn	100	2	7	★
ganzhou.gov.cn	Domain	ganzhou.gov.cn	100	4	14	★
jxgs.gov.cn	Domain	jxgs.gov.cn	100	2	9	★
jxrd.gov.cn	Domain	jxrd.gov.cn	100	2	4	★
jxwst.gov.cn	Domain	jxwst.gov.cn	100	3	8	★
jxgrain.gov.cn	Domain	jxgrain.gov.cn	100	2	8	★
xinyu.gov.cn	Domain	xinyu.gov.cn	100	3	15	★
jxsport.gov.cn	Domain	jxsport.gov.cn	100	2	8	★
jxinvest.gov.cn	Domain	jxinvest.gov.cn	100	2	3	★
ic.jiangxi.cn	Domain	ic.jiangxi.cn	100	2	6	★
jxgfgb.gov.cn	Domain	jxgfgb.gov.cn	100	3	7	★
jxwh.gov.cn	Domain	jxwh.gov.cn	100	2	10	★
jxszb.gov.cn	Domain	jxszb.gov.cn	100	2	7	★
jxysdzkcj.gov.cn	Domain	jxysdzkcj.gov.cn	100	3	8	★
jxmzj.gov.cn	Domain	jxmzj.gov.cn	100	3	7	★
jian.gov.cn	Domain	jian.gov.cn	100	3	13	★
jxaudit.gov.cn	Domain	jxaudit.gov.cn	100	2	7	★
jxdpc.gov.cn	Domain	jxdpc.gov.cn	100	3	14	★
jiangxi.cn	Domain	jiangxi.cn	100	2	9	★
jxmkaqjc.gov.cn	Domain	jxmkaqjc.gov.cn	100	3	8	★
jxinvest.cn	Domain	jxinvest.cn	100	2	3	★
jgsc.gov.cn	Domain	jgsc.gov.cn	100	2	4	★

Overview: Footprint L2 (jiangxi.gov.cn)
Machine completed
Computing networks
run(IPAddressToNetblock_Cuts)
userFilter(Select cluster IPs)
Looking up AS
run(NetblockToAS_SS)
Machine completed with 14 entities

Detail View: <No Selection>

Property View: <No Properties>

480 entities

The image shows a screenshot of the Twitter profile for Wolf Blitzer (@wolfblitzer). The profile picture is a headshot of a man in a suit and tie. The header shows statistics: 7,115 tweets, 309 photos/videos, 971 following, 713K followers, and 17 favorites. A 'Follow' button is in the top right. The bio states: 'I'm the anchor of CNN's The Situation Room (@CNNSitRoom), CNN's lead political anchor, and, yes, I do my own tweeting.' Location is 'Washington D.C.', website is 'cnn.com/situationroom', and joined date is 'September 2009'. There are 309 photos and videos. The main feed shows three tweets from Wolf Blitzer, all posted 22h ago, mentioning live coverage of the civil war in Syria, tonight's primaries, and the latest on Bergdahl. The right sidebar includes a 'Follow Wolf Blitzer' section with input fields for full name, email, and password, a 'Sign up for Twitter' button, and a 'Worldwide Trends' section with various trending hashtags.

Wolf Blitzer @wolfblitzer

I'm the anchor of CNN's The Situation Room (@CNNSitRoom), CNN's lead political anchor, and, yes, I do my own tweeting.

Washington D.C.

cnn.com/situationroom

Joined September 2009

309 Photos and videos

Tweets Tweets and replies

Wolf Blitzer @wolfblitzer · 22h
Our @NPWCNN is back covering the civil war in Syria. He joins me live today during 1PM ET hour. #Wolf

Wolf Blitzer @wolfblitzer · 22h
We're looking ahead to tonight's primaries. @GloriaBorger joins me live during 1PM ET hour. #Wolf

Wolf Blitzer @wolfblitzer · 22h
The latest on #Bergdahl w/ @MKosinskiCNN in Poland w/ President

Follow Wolf Blitzer

Full name

Email

Password

Sign up for Twitter

Worldwide Trends · change

#thescoutdown

#NotersAwesome

#youknowtheyreacnotswhen

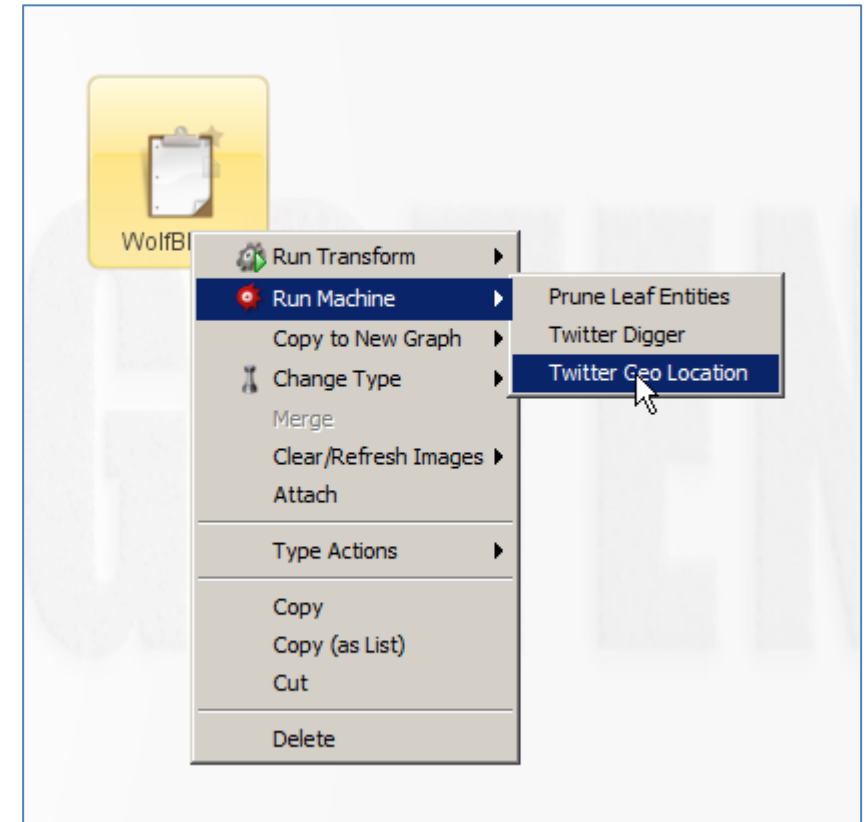
#BuCasaGelisenPisanCimarin

MAPPING A TWITTER ACCOUNT TO PHYSICAL LOCATIONS

To tie the messaging of a Twitter account holder (individual or group) to physical locations

EXTRACTION SEQUENCE


1. Identify target account on Twitter.
2. Under Personal, identify a Phrase. Once an account is identified, run all transforms on that account. What should be extracted should be the following: images taken with particular mobile devices (with EXIF or “exchangeable image file format” data); types of mobile devices used; IP address of accounts
3. Place any GPS coordinates into Google Maps for the mapping of the location.





WHO

International Confederation of Midwives (ICM), WHO and partners reveals major deficits in the midwifery workforce in 73 countries where these services are desperately needed. The report recommends new strategies to reduce these deficits and save millions of lives of women and newborns.

[Read the news release on investing in midwifery](#) 

[Read the feature on midwifery training in Bangladesh](#)

Investing in midwifery
can save millions of
lives of women and
newborns

Progress towards
universal health
coverage in BRICS

Raising taxes on
tobacco: what you
need to know

WHO calls for higher
tobacco taxes to save
more lives



Humanitarian health action



Disease outbreak news

Information about disease outbreaks



Director-General

Director-General and senior management



Governance

Constitution, Executive Board and World Health Assembly



WHO guidelines

A selection of evidence-based guidelines



WHO reform

Addressing public health challenges in the 21st century

Universal health coverage

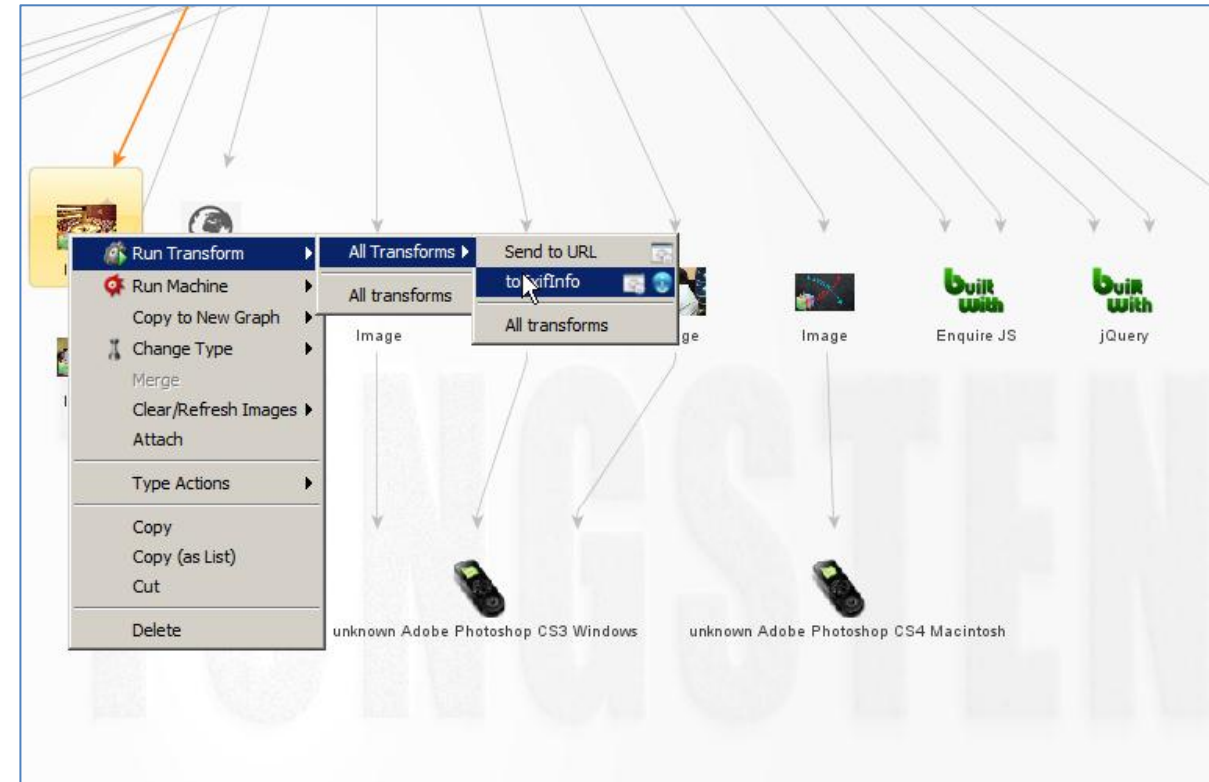


MAPPING A WEBSITE TO PHYSICAL SPACES

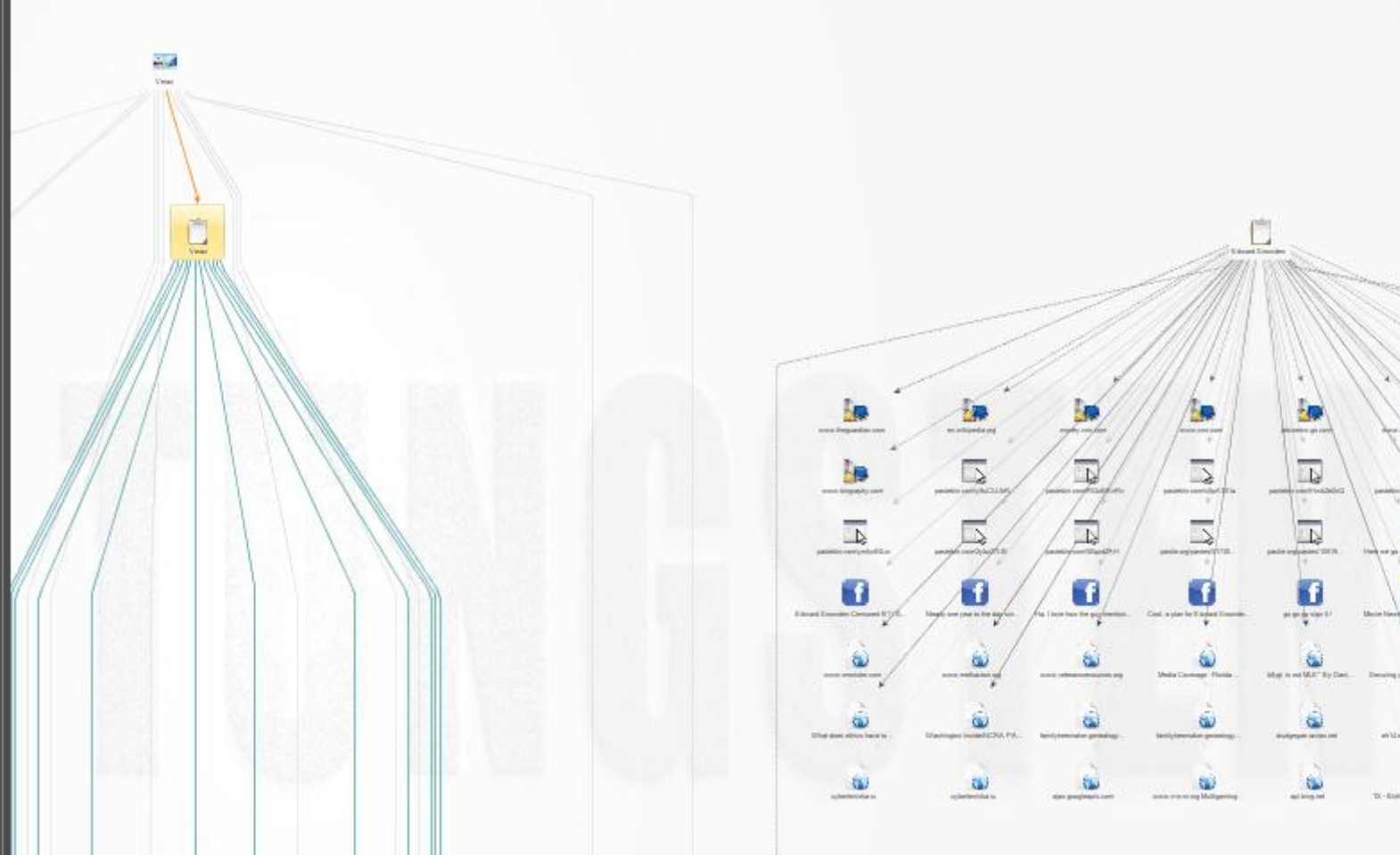
To know where a company or organization or institution is physically based (central and peripheral locations)

EXTRACTION SEQUENCE

1. Choose a website URL (http://www.who.int/en/). Footprint the URL. Or run the URL to Network and Domain information machine.) Apply transforms to the URL.
2. Capture IP addresses, and run those in Google or an IP WHOIS look-up. (There are distributed servers on multiple continents. No surprise there.)
3. Check for EXIF data on images for location tagging. (This information may be blocked by the webmasters.)



- Domain Name System server
- Domain
An internet domain
- IPv4 Address
An IP version 4 address
- MX Record
A DNS mail exchange record
- NS Record
A DNS name server record
- Netblock
An internet Autonomous Syst
- URL
An internet Uniform Resource
- Website
An internet website
- Locations**
- Location
A location on Mother Earth
- Penetration Testing**
- BuiltWith Technology
A Technology identified by B
- Personal**
- Alias
An alias for a person
- Document
A document on the internet
- Email Address
An email mailbox to which em



Detail View

Phrase
maltego.Phrase
Verax

+ Relationships

- Generator detail

Source	Verax	(Alias)
Transform	convertToPhrase	
Gen. date	2014-06-04 13:00:57.713 -0500	

TURNING AN ONLINE ALIAS TO PHYSICAL SPACE AND A REAL IDENTITY

To locate an online alias to physical haunting grounds and unique identifiers (linked to a person)

EXTRACTION SEQUENCE

1. Select an alias (or a handle without a human identifier) (Verax). See if the alias is connected to any other known accounts.
2. In a known case, add “Edward Snowden” as a phrase, and see if there is any bridging node or connectors. Unique identifiers are occasionally available. (Sometimes, the application does not actually help find connections and may seem obtuse.)

Maltego Tungsten 3.4.2

Investigate Manage Organize Machines Collaboration

Clipboard: Paste, Clear All, Cut, Delete
Transforms: Number of Results (12, 50, 255, 10k), Quick Find
Find: Entity Selection, Select All, Add Similar Siblings, Select Children, Add Children, Select by Type, Select Links, Select None, Add Path, Select Neighbors, Add Neighbors, Select Parents, Add Parents, Select Bookmarked, Reverse Links
Selection: Zoom to, Zoom In, Zoom to Fit, Zoom Out, Zoom 100%, Zoom Selection

Overview:

Detail View:

Property View:

Infrastructure:

- AS: An internet Autonomous System
- DNS Name: Domain Name System server record
- Domain: An internet domain
- IPv4 Address: An IP version 4 address
- MX Record: A DNS mail exchange record
- NS Record: A DNS name server record
- Netblock: An internet Autonomous System
- URL: An internet Uniform Resource Locator
- Website: An internet website

Locations:

- Location: A location on Mother Earth

Penetration Testing:

- BuiltWith Technology: A Technology identified by BuiltWith

Personal:

- Alias: An alias for a person
- Document: A document on the internet
- Email Address: An email mailbox to which email is sent
- Image: A visual representation of something
- Person: Entity representing a human
- Phone Number: A telephone number
- Phrase: Any text or part thereof

Social Network:

- Facebook Object: Facebook Object
- Twit: Twit entity

Main View:

Entity List:

1 of 147 entities



MAPPING AN EGO NEIGHBORHOOD OF A FOCAL NODE TO PHYSICAL LOCATIONS

To gain a sense of the main physical locations of a focal node

EXTRACTION SEQUENCE

1. Select an individual with a name that is somewhat disambiguated. From the Maltego Palette, place the “Person” icon on the workspace. Type in the selected name (in this case, “Elon Musk.”) Right click the icon. Run the “transforms.”
2. Add a URL link with known ties to the individual to identify additional sources and overlapping nodes and ties.
3. Resolve URLs to IP addresses (which link directly to geographical space).
4. Resolve images to EXIF information.
5. Resolve Twitter account to Twitpic images.
6. Resolve Facebook accounts to all knowable transforms. Resolve likely email accounts to all knowable transforms.
7. Check area codes / dialing codes of related linked phone numbers.

Maltego Tungsten 3.4.2

Investigate Manage Organize Machines Collaboration

Clipboard: Paste, Clear All, Cut, Delete

Transforms: Number of Results (12, 50, 255, 10k)

Find: Quick Find

Selection: Select All, Add Similar Siblings, Select Children, Add Children, Select by Type, Invert Selection, Add Path, Select Neighbors, Add Neighbors, Select Links, Select None, Select Parents, Add Parents, Select Bookmarked, Reverse Links

Zoom: Zoom to, Zoom In, Zoom to Fit, Zoom Out, Zoom 100%, Zoom Selection

Palette:

- Domain: An internet domain
- IPv4 Address: An IP version 4 address
- MX Record: A DNS mail exchange record
- NS Record: A DNS name server record
- Netblock: An internet Autonomous System
- URL: An internet Uniform Resource Locator
- Website: An internet website
- Locations**
 - Location: A location on Mother Earth
- Penetration Testing**
 - BuiltWith Technology: A Technology identified by BuiltWith
- Personal**
 - Alias: An alias for a person
 - Document: A document on the internet
 - Email Address: An email mailbox to which email is sent
 - Image: A visual representation of something
 - Person: Entity representing a human
 - Phone Number: A telephone number
- Social Network**
 - Facebook Object: Facebook Object
 - Twit: Twit entity
 - Affiliation - Facebook: Membership of the Facebook
 - Affiliation - Twitter: Membership of Twitter

Entity List: Main View, Bubble View, Entity List

Overview: Person - Email Address (Elon Musk)

Machine completed

run(PersonToEmailAddress_SE)
userFilter(Select email addresses)
run(EmailAddressToDNSName_SE)
[Looking up email address on search engines](#)
Machine completed with 24 entities

Detail View: Email Address (maltego.EmailAddress) elon.musk@gmail.com

+ Relationships

- Generator detail

Source	Elon Musk	(Person)
Transform	To Email Address [Verify common]	
Gen. date	2014-06-04 15:12:17.934 -0500	

Property View: Properties

Type	Email Address
Email Address	elon.musk@gmail.com

Graph info

Weight	100
Incoming	1
Outgoing	0
Bookmark	★

Entity List: pastebin.com/hXhBB4Zn, pastebin.com/gpKKfwLq, pastebin.com/QVb6aEdU, pastebin.com/yYUYT9b9, pastie.org/pastes/y/201..., pastie.org/pastes/83368..., http://bit.ly/1o9U0RI Tesla an..., God damn it Ford !!!, Pretty cool. "Intellectual" pr..., elonm@hotmail.com

Send to URL

- To Domain [DNS]
- To Email Addresses [PGP (signed)]
- To Email Addresses [PGP]
- To Email Addresses [using Search Engine]
- To Person [PGP]
- To Phone number [using Search Engine]
- To URLs [Show search engine results]
- To Website [using Search Engine]
- Verify email address exists [SMTP]
- emailToFlickrAccount
- emailToMyspaceAccount
- searchPastebinsForEmail

Run Transform: All Transforms, Related Email addresses, Other transforms, All transforms

Run Machine: All transforms

Copy to New Graph: All transforms

Change Type: All transforms

Merge

Clear/Refresh Images

Attach

Type Actions

Copy

Copy (as List)

Cut

Find

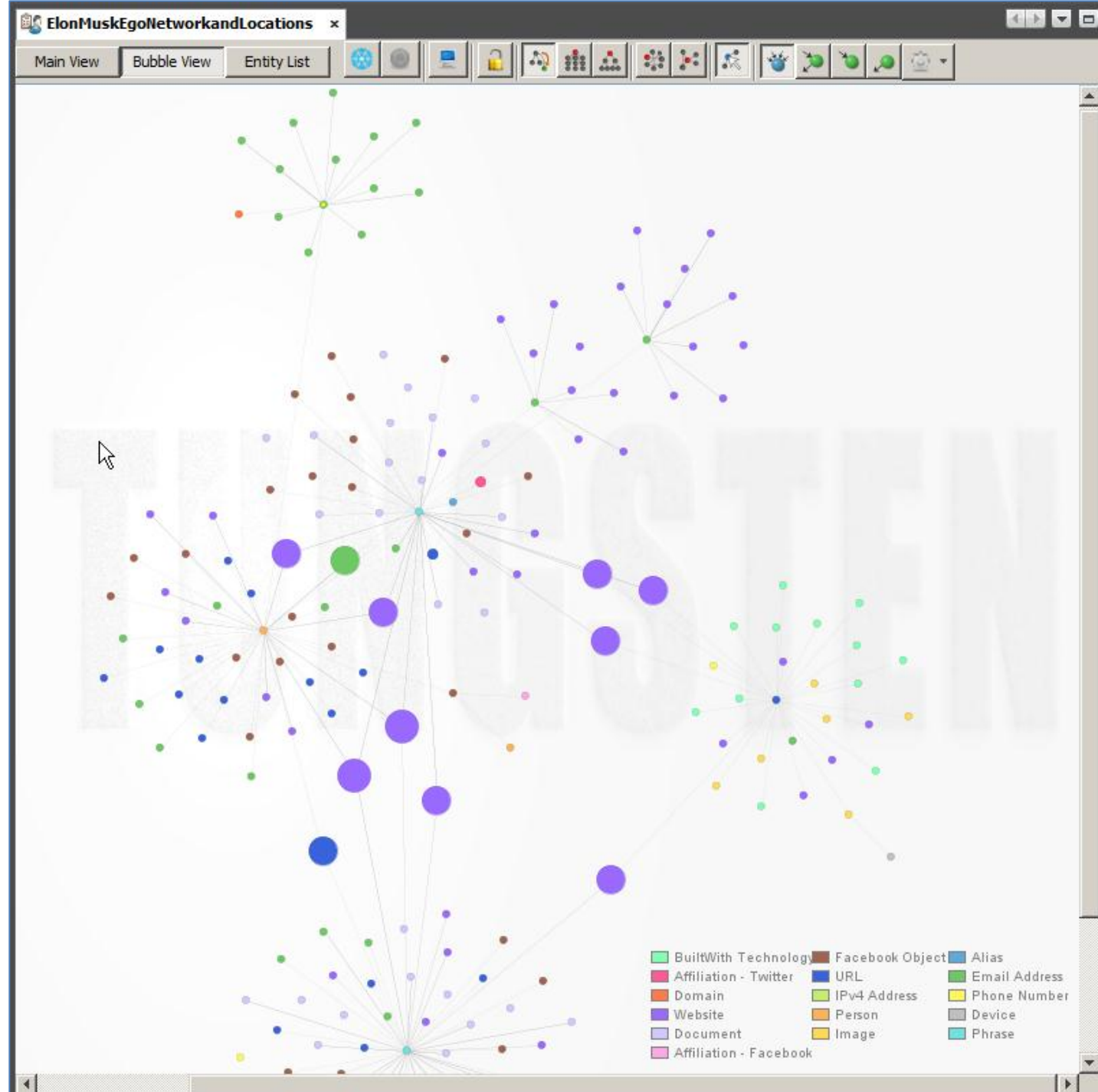
Properties

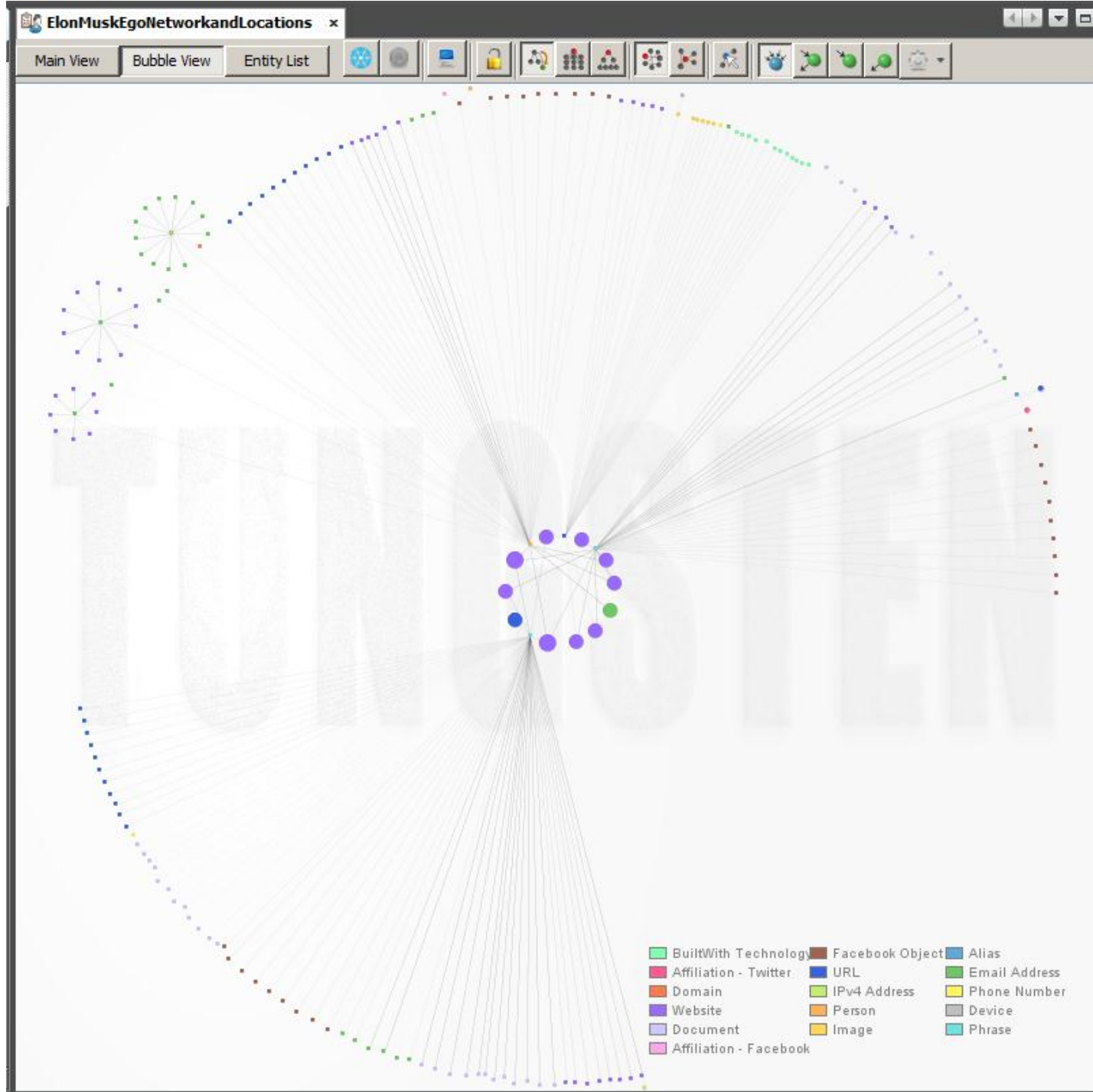
Notes

Display info

Zoom

1 of 189 entities







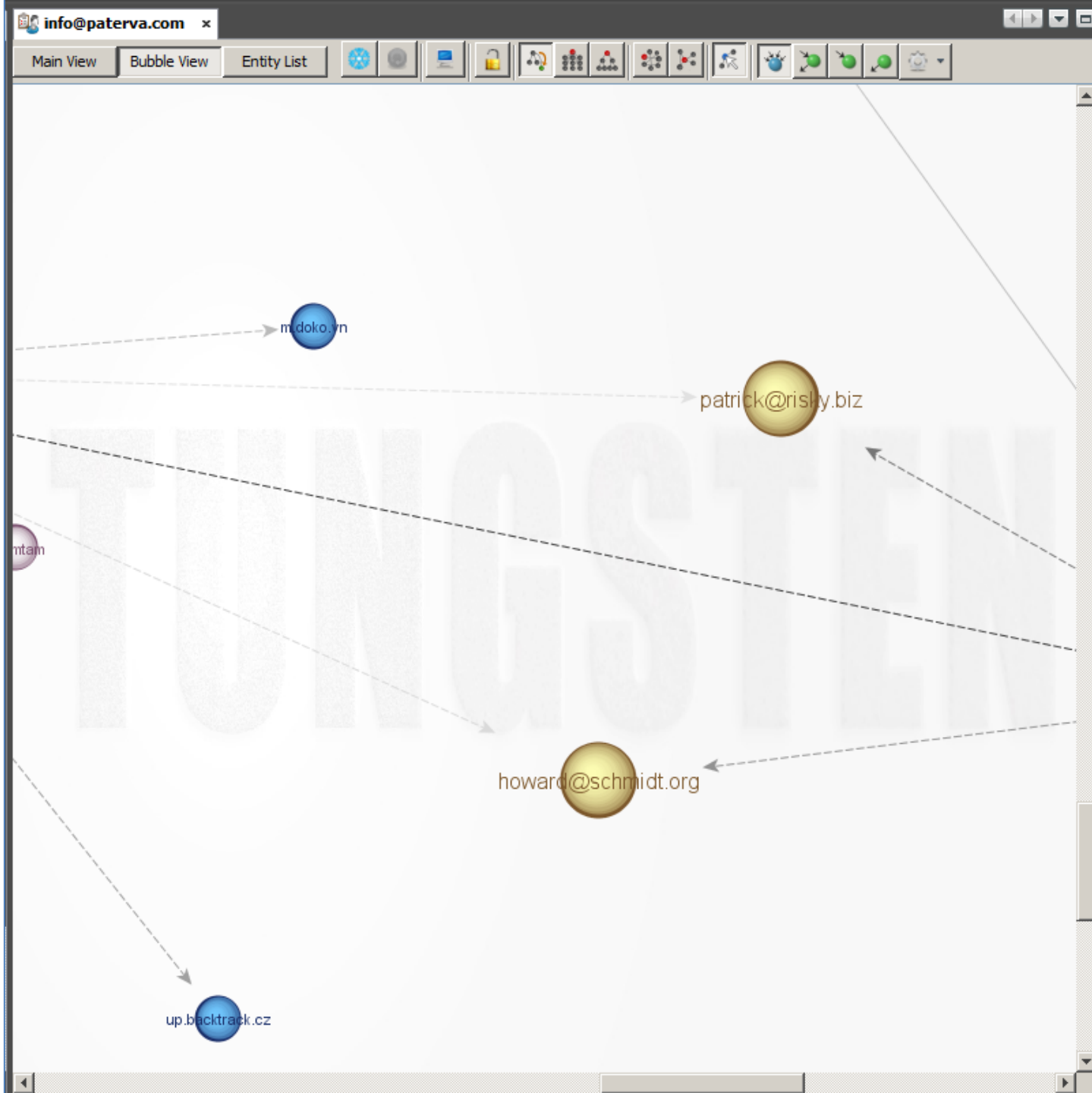
MAPPING AN EMAIL ACCOUNT TO A PHYSICAL LOCATION

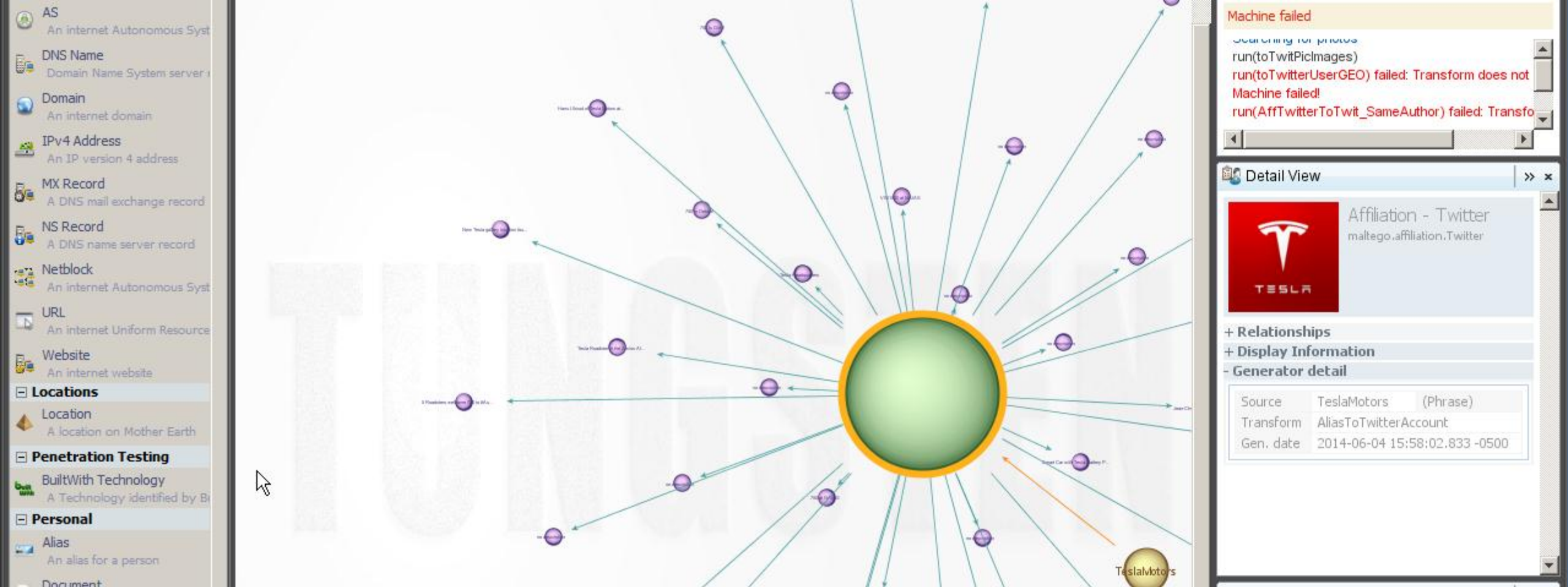
To re-identify an email account by linking it to a location (and a unique identifier)

EXTRACTION SEQUENCE

1. Select an email address (info@paterva.com). Run transforms on the email account.
2. Select out phone numbers and run the dialing codes and area codes on Google for some location data.
3. Transform URLs to IP addresses, which may be translated into geographical space using a simple search engine. Run a WHOIS search for an IP address to look at ownership of the IP address.





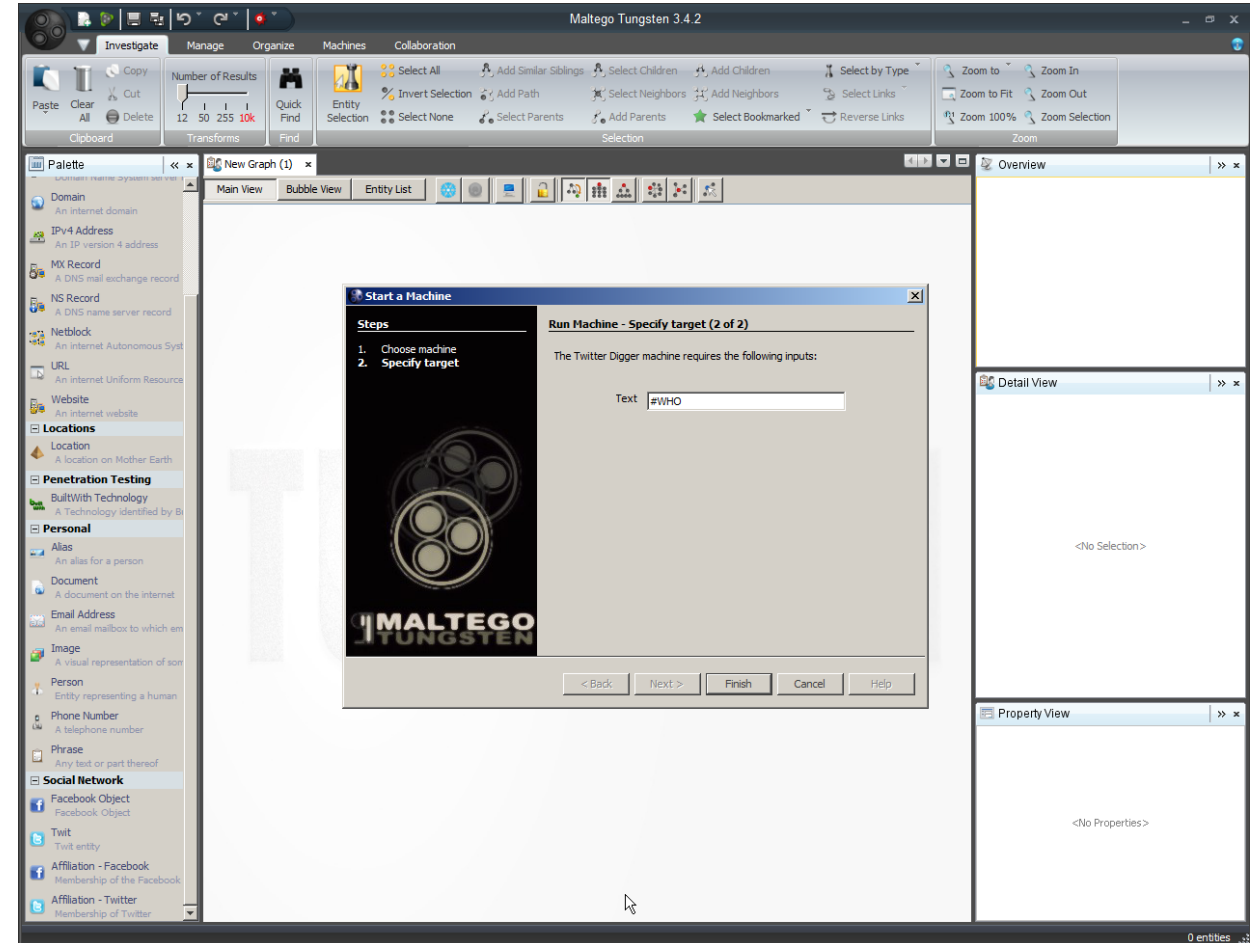


MAPPING A TWITTER (HASHTAG OR KEYWORD) CONVERSATION TO LOCATIONS

To better understand the physical locations of the participants of a Twitter hashtag (#) conversation or event (through a tag-based eventgraph)

EXTRACTION SEQUENCE

1. Start the Twitter Digger. Type the targeted “keyword” text into the text window. (In this case, it was “Tesla Motors”).
2. Run the Twitter Geolocation as well.
3. Link any of the entities (objects) to physical spaces (such as domains to locations; phone numbers to locations; IPs to locations; image EXIF data to locations, and others).



42°19'40.2"N 83°02'55.2"W - Go...

https://www.google.com/maps/place/42°19'40.2"N+83°02'55.2"W/@42.32783,-83.04867,15z/data=!4m2!3m1!1s0x0:0x0

42.32783,-83.04867

42°19'40.2"N 83°02'55.2"W
42.327830,-83.048670

Directions

Street View

Explore this area
Search nearby · restaurants · cafes · bars

42°19'40.2"N 83°02'55.2"W

42°19'40.2"N 83°02'55.2"W

Earth

WINDSOR

Map data ©2014 Google Terms Privacy Report a problem 1000 ft

42°19'40.2"N 83°02'55.2"W - Go...

https://www.google.com/maps/place/42°19'40.2"N+83°02'55.2"W/@42.3278307,-83.04867,3581m/data=!3m1!1e3!4m2!3m1!1s0x0:0x0

Google

42°19'40.2"N 83°02'55.2"W

42°19'40.2"N 83°02'55.2"W

Street View

Map

RiverWalk

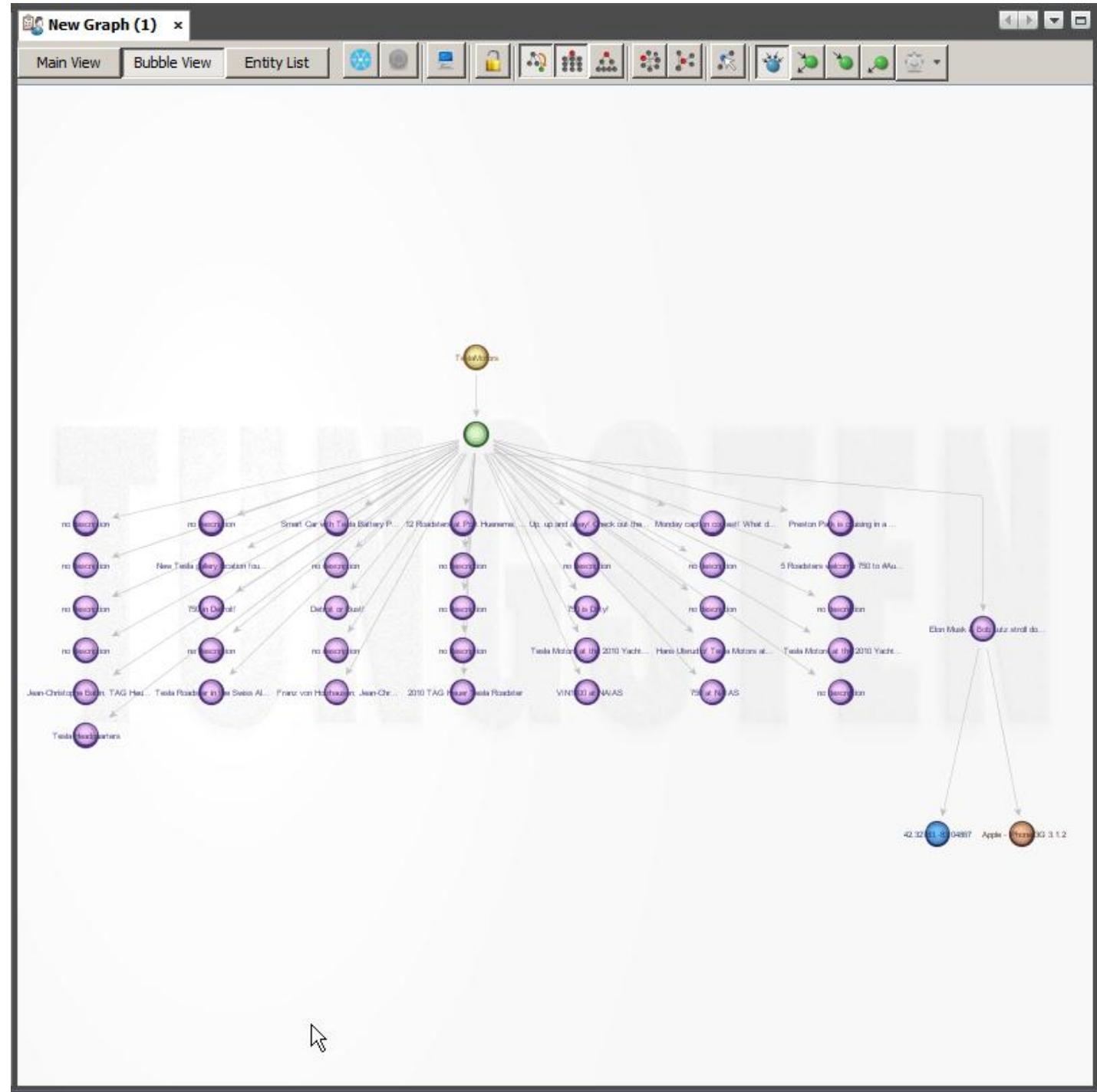
Windsor

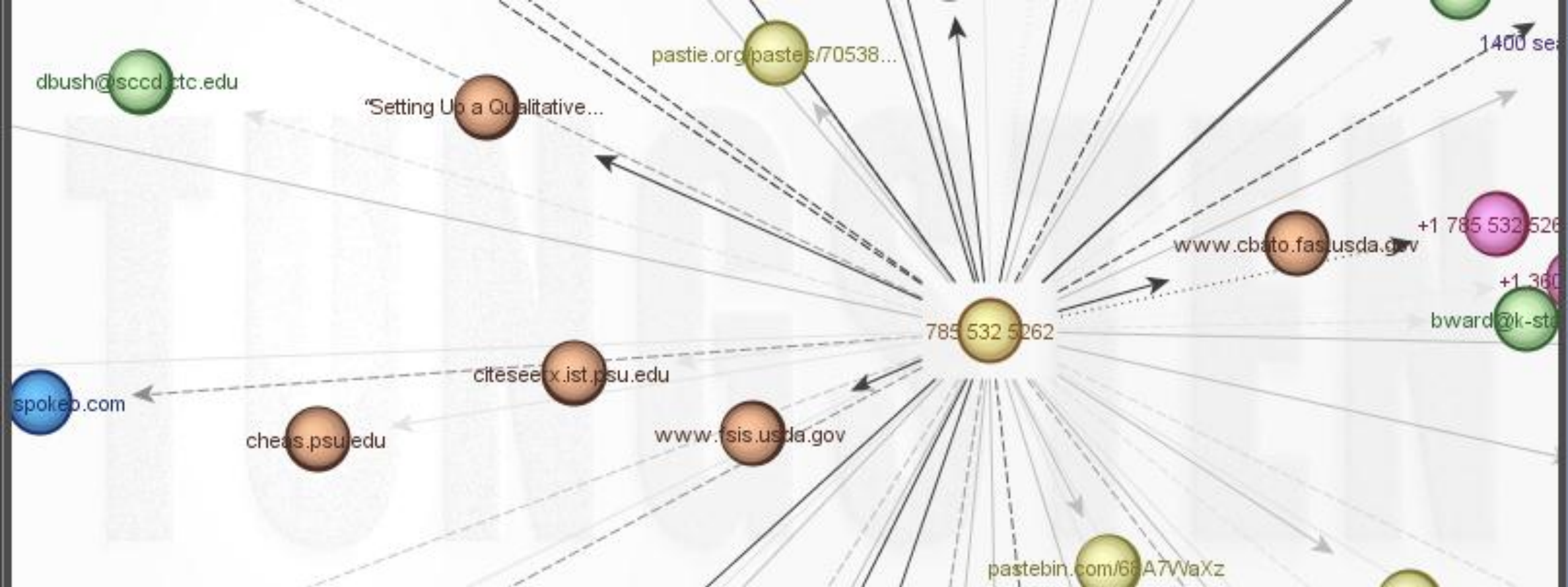
General Motors Co

GM Renaissance Center

Map data ©2014 Google Terms Privacy Report a problem 1000 ft

ONE OF SEVERAL HIERARCHICAL / STRUCTURED VIEWS



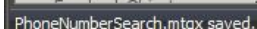


PHONE NUMBER TO “TRANSFORMS” TO LEAD TO PHYSICAL LOCATIONS

To link a telephone number to physical locations (beyond what is knowable by area code)

EXTRACTION SEQUENCE

1. Select a telephone number. (Do not use the Phone Number drop-in from the Palette.) Use the Phrase.
2. Right click the phrase. Run all “transforms” on that Phrase (phone number in this case).
3. There are a range of leads which lead to physical location.





MAPPING A TWITTER EVENTGRAPH (BASED ON #HASHTAGS) TO LOCATIONS

To relate participants in an event that is microblogged to physical locations of the various communicators / participants

EXTRACTION SEQUENCE

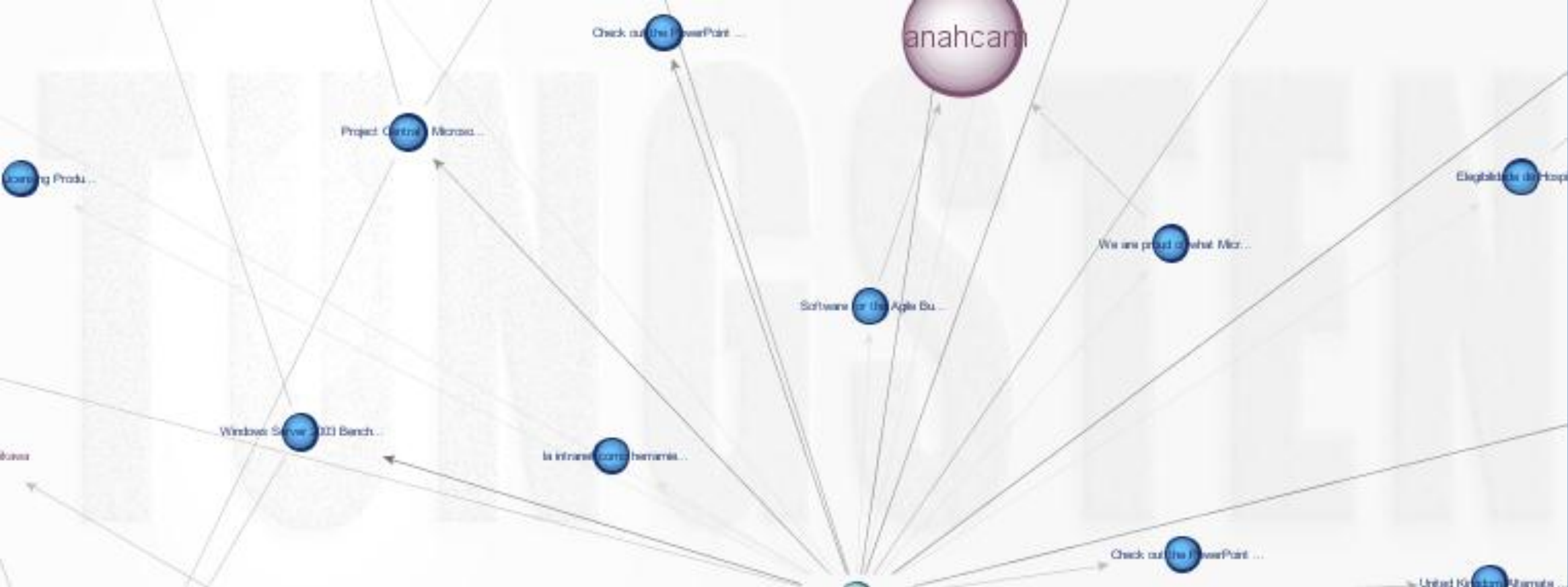
1. Identity a hashtag string that indicates a particular event (or formal conference). In this case, it will be “#NETC2014”. In Maltego Tungsten, drag “Phrase” from the Palette onto the workspace. Type in the hashtag string.
2. Change type to Social Network -> Twit. Run Transforms, including pulling all hashtags.
3. Or run Twitter Digger.
4. TwitterMonitor should offer some continuous tracking, but these were all somewhat problematic when I tried different runs.
5. See the next slide for responses from Maltego Support of Paterva to an email query.
6. (**Note:** The illustration on the prior slide was an eventgraph from Twitter of #NETC2014 using NodeXL, which is a freeware tool that also enables such extractions.)

DRAT!

Reply from Maltego Support (May 2014)

“The problem (that our client devs are currently working on) is that we need to now authenticate for every API call we make to twitter and we are trying to find a way to implement the OAuth for every Maltego user. As soon as this is done you should be able to use these transforms without a problem! The new release should be out in about a month that will fix this!

“However the Alias'/Phrases to twitter accounts should still be functional.”

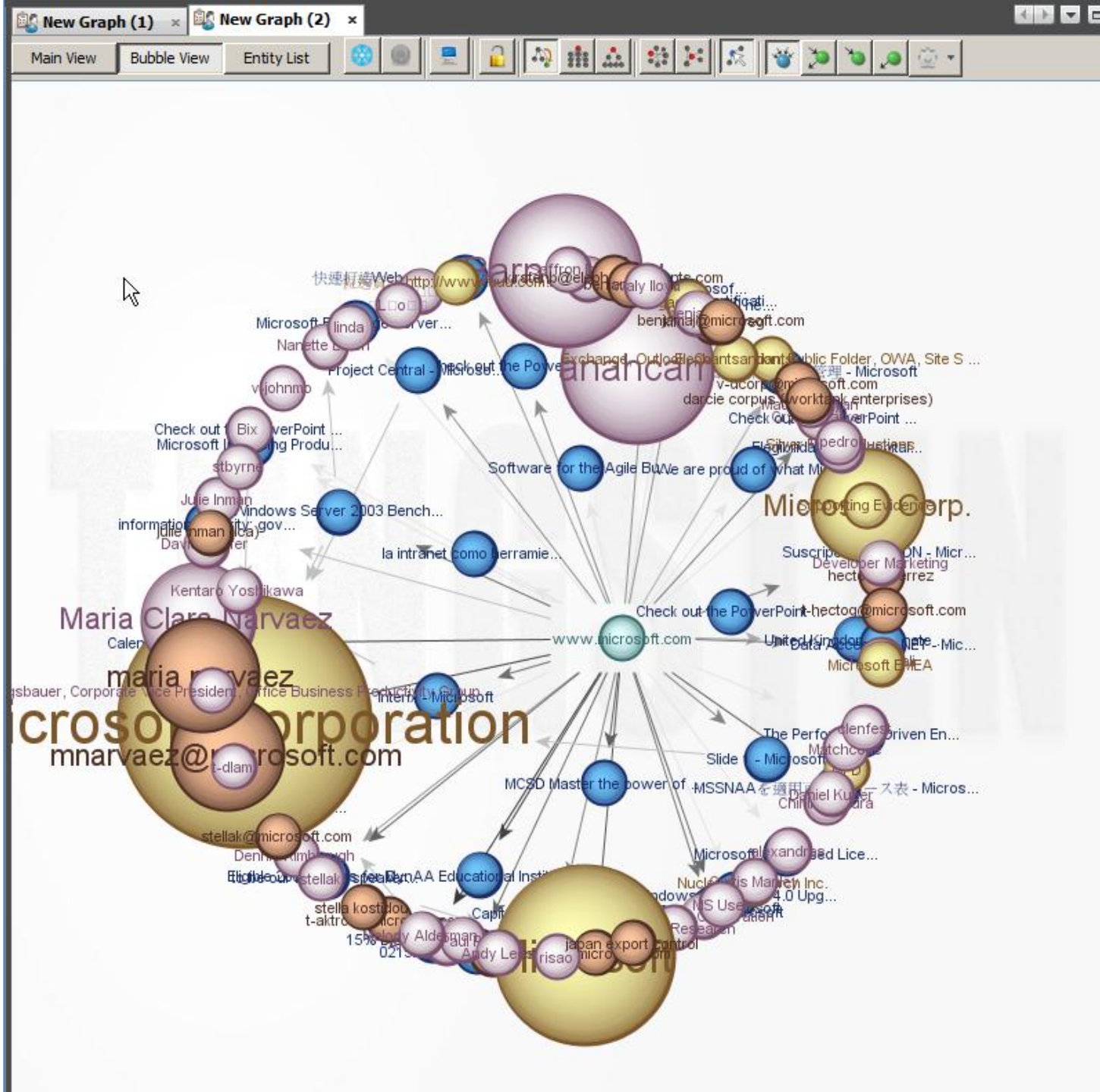


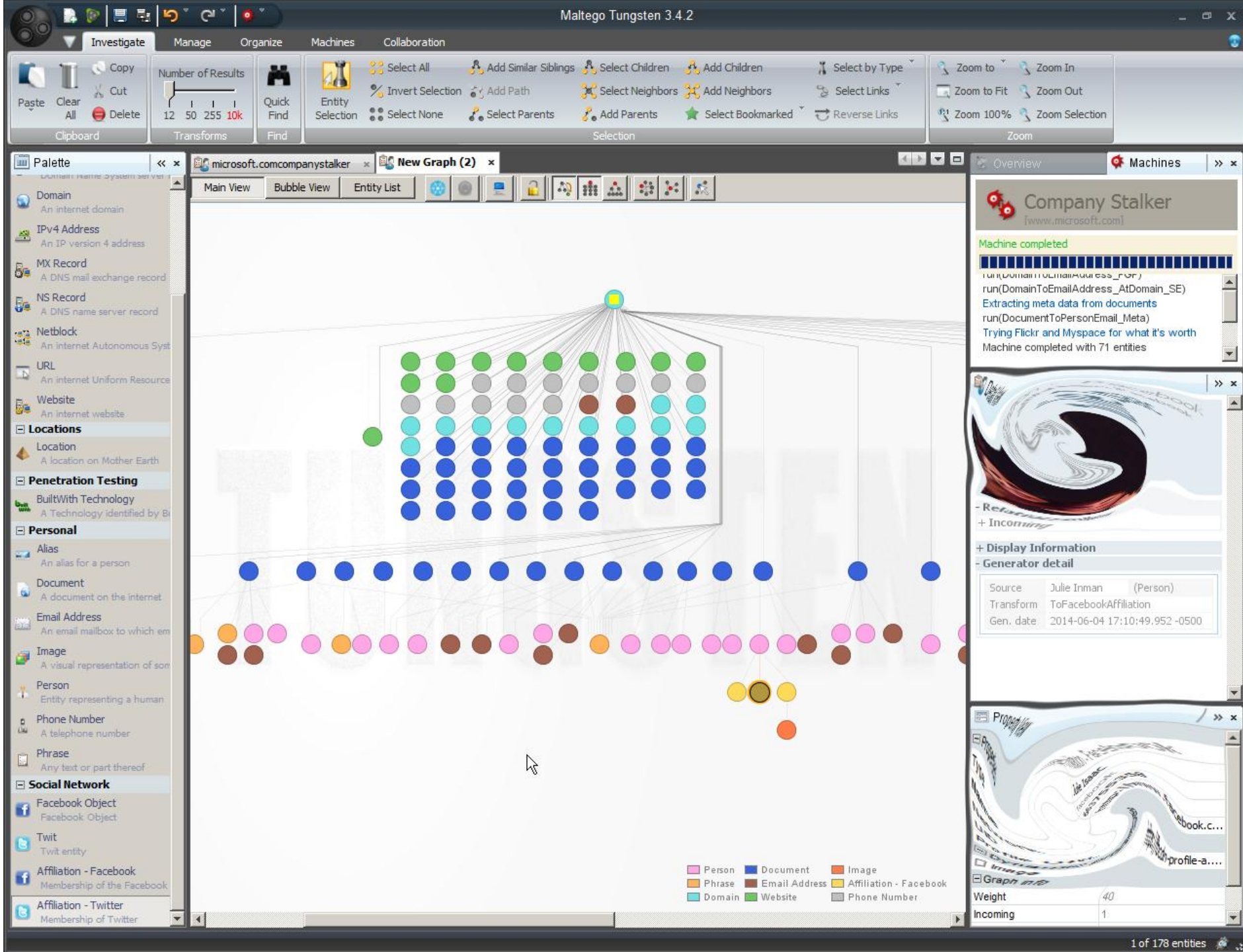
MAPPING A WEB NETWORK TO PHYSICAL LOCATIONS

To map a web network (of
interlinked websites) to
physical locations to
understand proximity and
culture effects

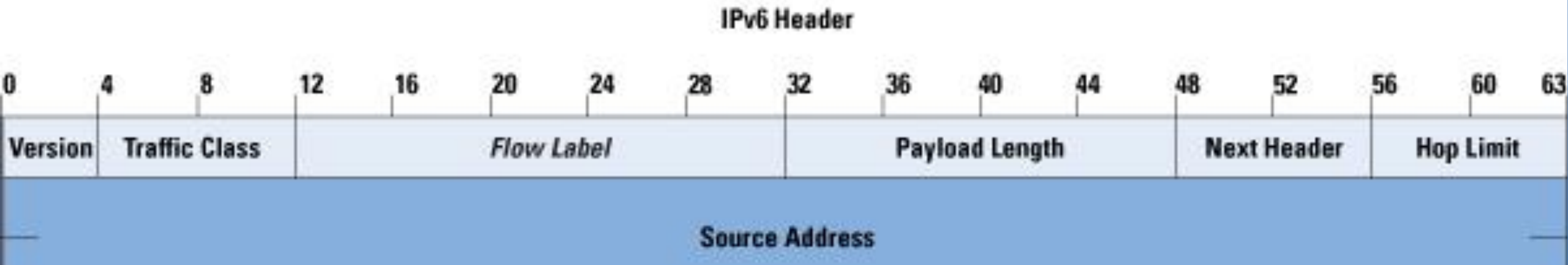
EXTRACTION SEQUENCE

1. Select a uniform resource locator or “URL” (which will serve as the focal node of an ego neighborhood of URLs on the Web; this network is created by URL linkages). In this case, “www.Microsoft.com” will be used. In Maltego Tungsten, run a “Company Stalker.”
2. Run all transforms. Select entities about which to transform into IP addresses and direct physical spaces. Telephone numbers can be directly turned into geolocation data. Images may be probed for EXIF data. Domain information may be tapped for DNS data / records. Identified people may be probed for Facebook and Twitter affiliations. People’s presences may be searched for in pastebins (web application spaces where people may store large amounts of text for a time online).





Identification		Flags	Fragment Offset
Time to Live	Protocol	Header Checksum	
Source Address			
Destination Address			

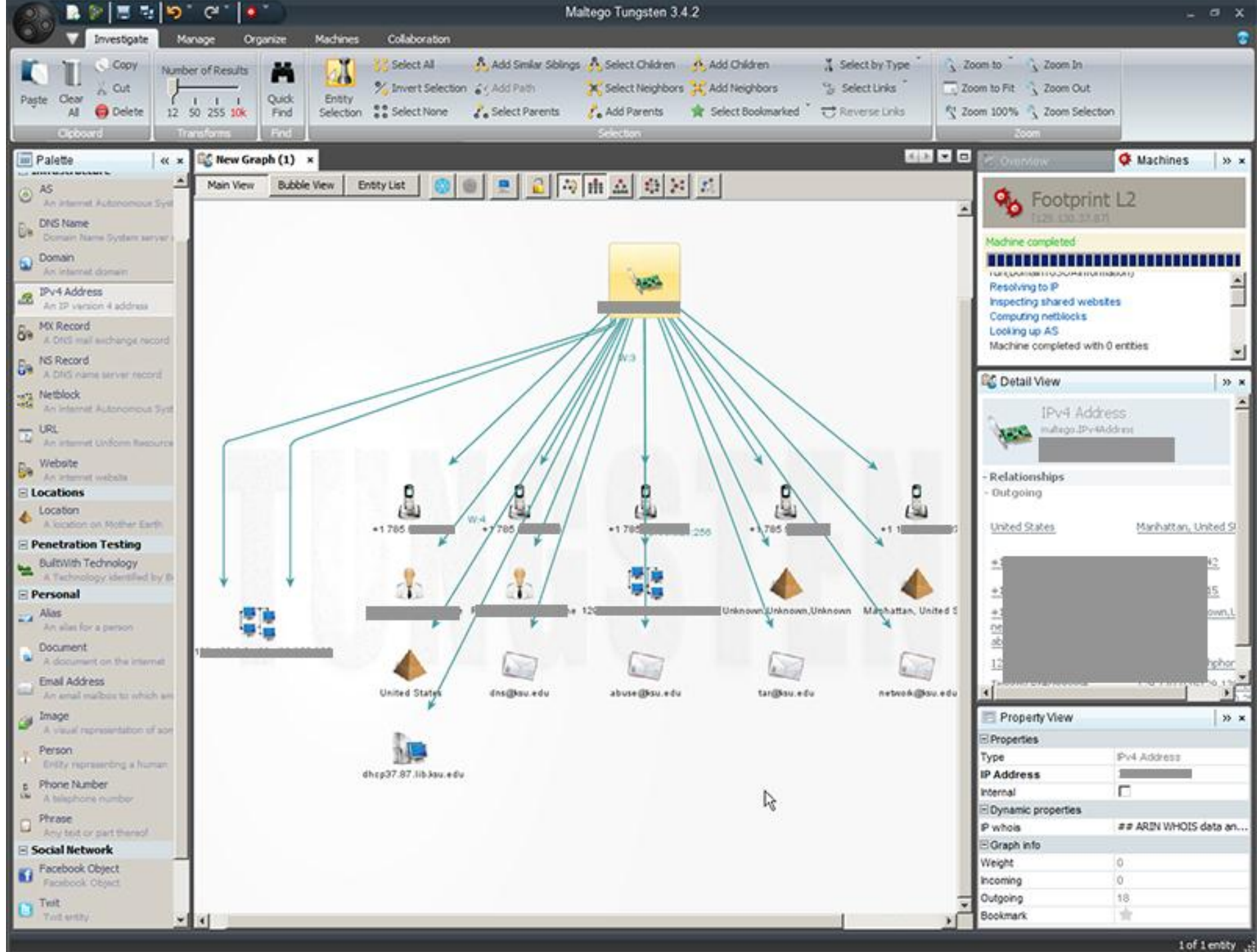


MAPPING AN INTERNET PROTOCOL (IP) ADDRESS TO VARIOUS CONTENTS AND PHYSICAL LOCALE

To link a particular computer on a network to a physical locale (and also related locales based on linked websites and linked email accounts and multimedia contents)

EXTRACTION SEQUENCE

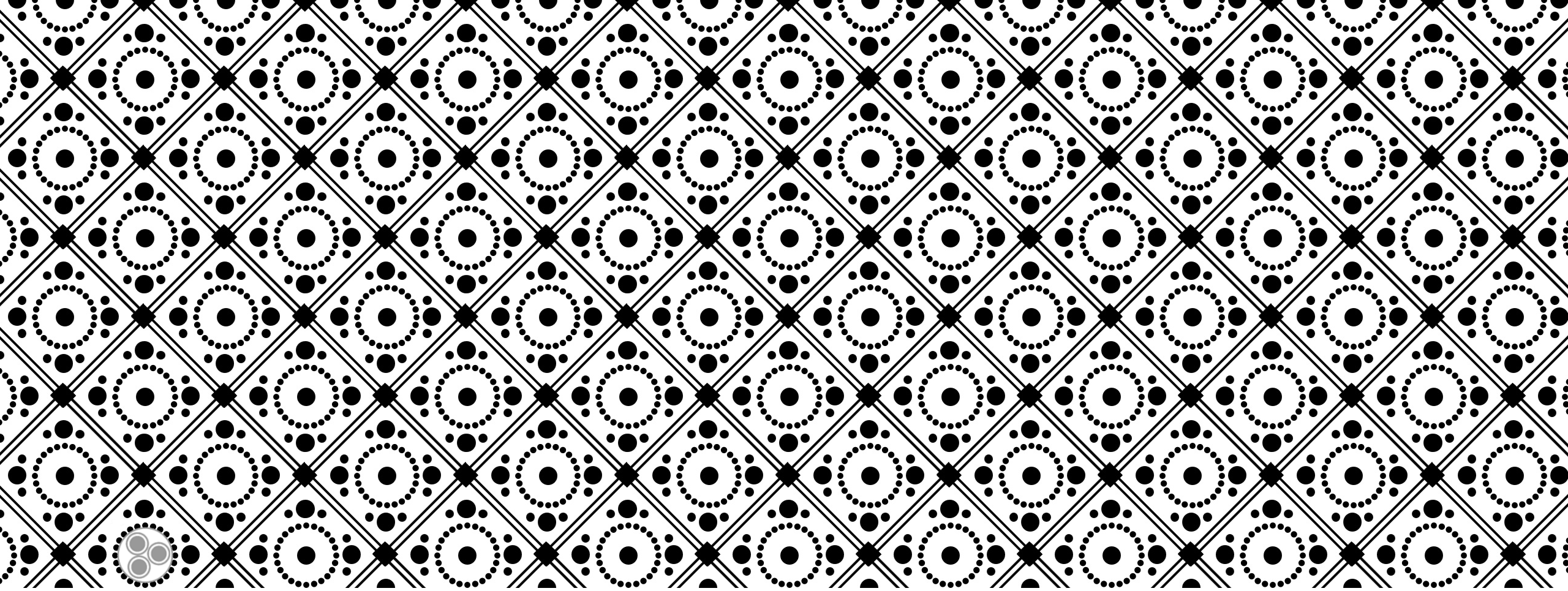
1. Select an IP address. (In a search engine, you can always go to “What’s my IP?” to find out your own.) Paste the IP address into the workspace as a an IPv4 Address (from the Palette).
2. The results will link to locations and telephones, which resolve to locations. There will be people identified linked to the IP address. There will also be emails...





MAPPING TIME STAMPS TO PHYSICAL LOCATIONS

To link posted online contents and messages on social media platforms to physical locations based on time zones and usual day-night activity levels



THE INFERENCE ATTACK ANGLE

To understand what may be
understood from information or
data

LATENCY OR HIDDENNESS IN ELECTRONIC (SOCIAL MEDIA PLATFORM-BASED) GEOLOCATION

- ❖ The handle-identified “mayor” of a hashtag network or microblogging conversation (per Dr. Marc A. Smith’s term), a “mayor” who is invisible without electronic social network analysis...and the tie of the “mayor” to a physical location which is an indicator of identity
- ❖ The space-time co-location pairing of people who do not have an apparent relationship but just a passing one (that may indicate a hidden relationship) or the converse where people may not have physical co-location but have shared virtual spaces
- ❖ The physical convergence of people in a shared physical space for political / social / demonstration / performance or other purposes

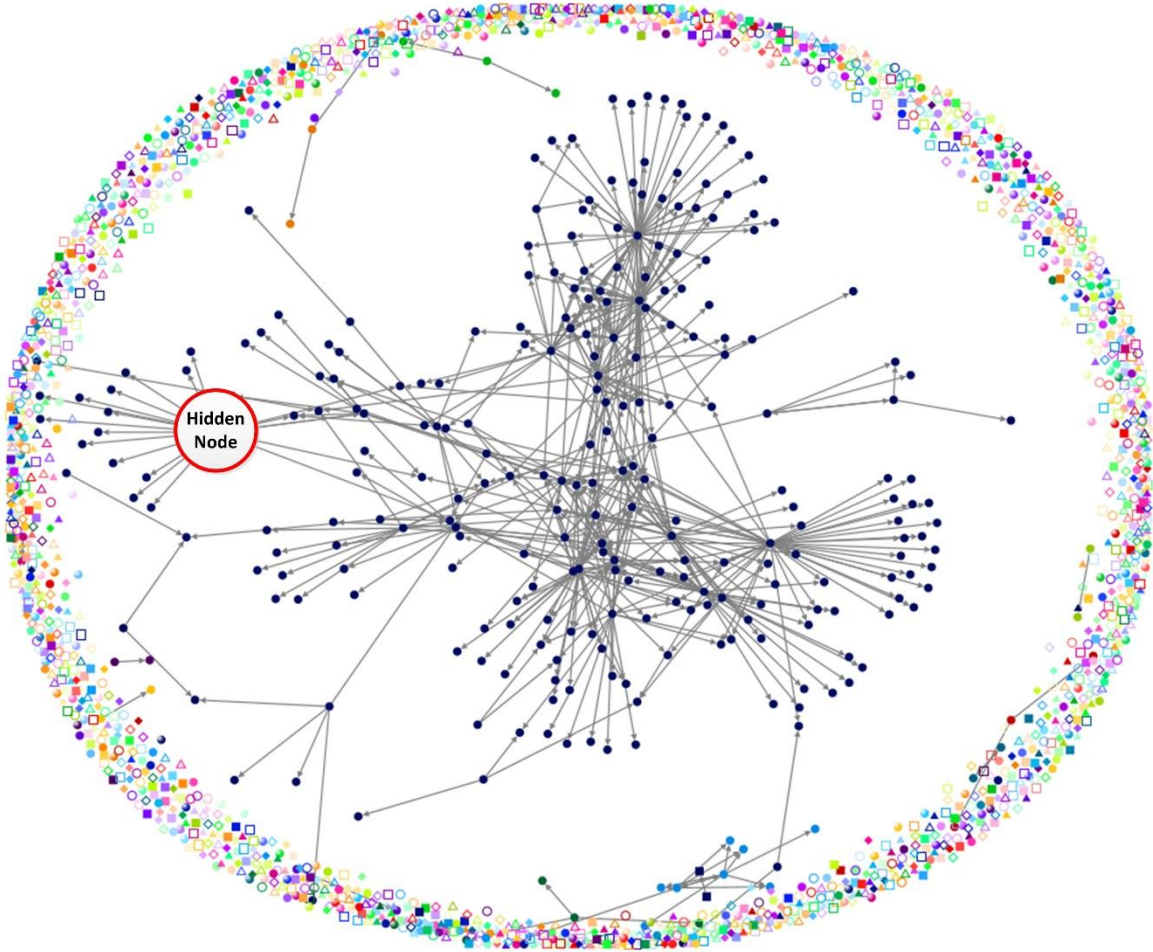
DISCUSSION



You are Here?

- ❖ How would you stay secure and safe in an environment with so much knowability and real-time and physical trackability? Any sophisticated ways to keep secrets?
- ❖ What parts of your locational privacy are you willing to give up and why?
- ❖ What are some potential “risks” of location data sharing or leakage through social media (microblogging, image-sharing, video-sharing, mobile applications, and others)?
- ❖ What are some potential “benefits” of location data sharing or leakage?
- ❖ When does location matter? When doesn't it matter?
- ❖ What would your long-term use of space-time show about who you are and what you do?
- ❖ If you wanted to track a person, is it fair to tap their geolocational data? And if so, how would you do it?

GOING TO GROUND?



- ❖ Be aware of your own patterning and habits...and break them.
- ❖ Be aware of (electronic and other) data leakage. Don't turn anything electronic and put it online if you don't want it possibly leaked and shared.
- ❖ Work out what you will / will not allow your friends and family to share. (People are people, though, and they will inadvertently or advertently leak.) Train your friends and family to take precautions as well. Latent or hidden nodes in a network may be identified by surrounding nodes.

GOING TO GROUND? (CONT.)

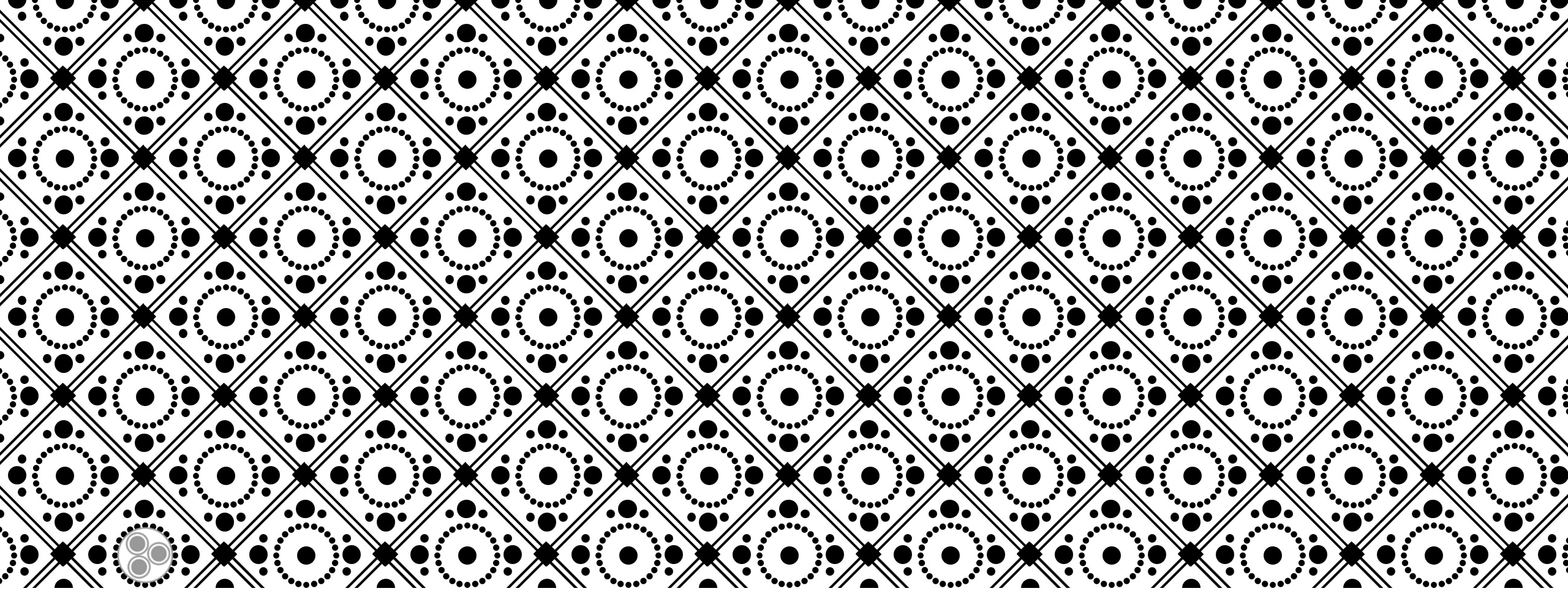
- ❖ Avoid over-sharing. TMI exists.
- ❖ Keep a very small “attack surface” profile—because each vector is a way that people may expose their data and themselves.
- ❖ Turn off the geolocation feature on all apps and digital cameras.
- ❖ Avoid using mobile devices that require regular check-ins to a locational tower.
- ❖ Once data is leaked “into the wild,” it is very hard (impossible?) to bring back.
- ❖ There is no online “invisibility” except for some people who really have some sort of cloaking and do not appear online...except for a few controlled links. They have “scrubbed” identities that have resulted from something more than just being quiet. Apparently, “off grid” is artificially possible.

EVEN SO...

- ❖ There are still other sensors in the environment, such as cameras...and others' cell phones and image / audio captures...
- ❖ There is constant data collection through all sorts of basic behaviors in-world.
- ❖ Anonymity is limited to certain people in certain contexts about certain information.
- ❖ There is no escaping unintended revelations if law enforcement is interested and brings its resources to bear. Likewise, it's hard to fully backstop a faked online identity because of lack of access to official documentation of personhood.

ATTENTION DEFICIT...AS PRIVACY?

- ❖ Laws and their enforcement provide some protection for human privacy. For corporations, the fear of legal liabilities may inhibit privacy infringements.
- ❖ The limits of human attention, particularly law enforcement attention, gives truth to the semblance of privacy. So while it's all collected (pre- and post- Snowden), the question really is about when it's accessed and what it's used for.



AMBIGUITY VS. DEFINITION, EXACTITUDE, AND PRECISION

To clarify what is not knowable
vs. knowable

CAVEATS TO GEOLOCATING WITH CYBER DATA

- ❖ **Faked Data:** Anything electronic can be faked. Anything physical can be faked (albeit with somewhat more effort).
- ❖ **Social Performances and Posing:** Narrow-cast and broadcast communications are social performances, and as such, there is always a degree of posing. (A lot of online accounts are ambiguous, and many are faux. Some online accounts are ‘bots—such as Tweeting and conversational ‘bots.) Avoid over-assertion beyond what is actually verifiable.
- ❖ **Noisy Data:** On social media platforms, geolocational information is often noisy data, particularly if it is human-reported (or “volunteered geographic information”) with mistaken or intentional mis-reporting vs. device-reported (but some devices can be hacked to report inaccurate data).
- ❖ **Ambiguity:** The noise-to-signal ratio can be quite high if using just limited software tools and limited data sources. There is a problem of disambiguation (differentiating valuable signals from general noise).

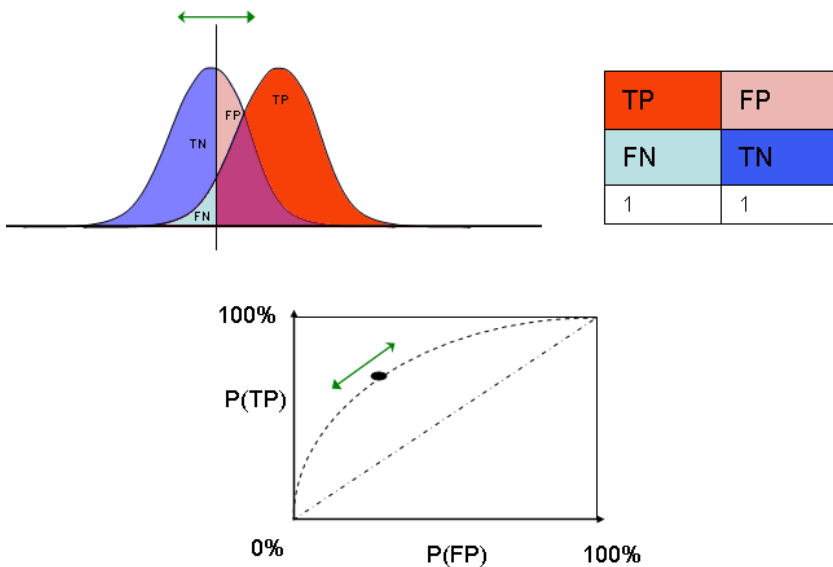
CAVEATS TO GEOLOCATING WITH CYBER DATA (CONT.)

- ❖ **Limits to Information Gathering:** Data extractions are done with a lot of dependencies. There are limits to the computer, the software, the social media platforms (and their APIs / Application Programming Interfaces, with time, amount, and rate limiting oftentimes)... There is forbidden (blocked) data which is not extractable from various social media platforms (whether visibly or invisibly to users).
- ❖ **Challenges of Time:** The data extraction may collect data for defunct accounts that are no longer being used...because there are residual data elements findable online.
- ❖ **Security Countermeasures:** Much geolocational data is now hidden, scrubbed, or encrypted. Many also choose not to share. (One research statistic is that only 1% of Tweets include geo data.)

WAYS TO PRESSURE-TEST FINDINGS

- ❖ Attempt to disconfirm the findings. What evidence is there that your findings are untrue? Is this evidence thoroughly analyzed? Has all relevant evidence been brought to bear?
- ❖ Test data extractions against known information to see how accurate the tool is (in a variety of settings).
- ❖ Test data extractions against parameter settings of the tool. (For example, acquire data samples at the various settings: 12, 50, 255, 10K...) Different settings will always result in different data outcomes.
- ❖ Run multiple data extractions using the same parameters but different software tools to see if the same results are attained. If not, what are the differences? Why might these differences exist?

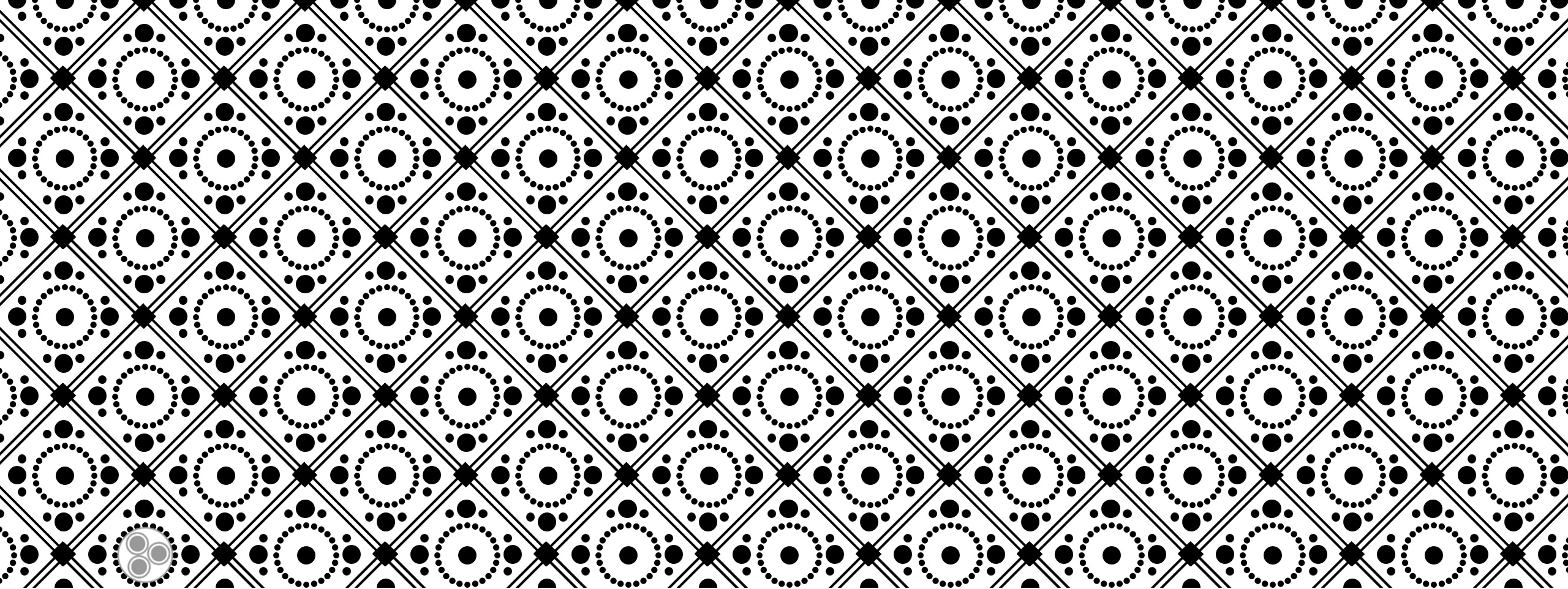
WAYS TO PRESSURE-TEST FINDINGS (CONT.)



- ❖ Use a broad range of data sources. Prune (or cut / delete) relational ties carefully.
- ❖ Analyze the findings both in breadth and depth (the network, the clusters / branches, and the nodes).
- ❖ Check the cyber-physical results (and early intuitions) against the real-world.

AN EMPIRICAL RESEARCH ANGLE

- ❖ Social media relations considered to be based on real-world evidence of interactivity (co-linking between websites, replies and re-Tweeting of online microblogged conversations, following / follower relationships, and other types of interactivity).
- ❖ Such extractions may be considered sampling from “big data”.
- ❖ People’s self-reportage in surveys considered often to be fallible (in terms of memory) and limited (in terms of subjectivities).
- ❖ This type of query is both structured (with a prior questions) and unstructured (inclusive of discovery learning). The latter means that researchers are open to what they will find in the data extraction.
- ❖ Research assertions are virtually never made without qualifiers.



DEMOS

Your call...

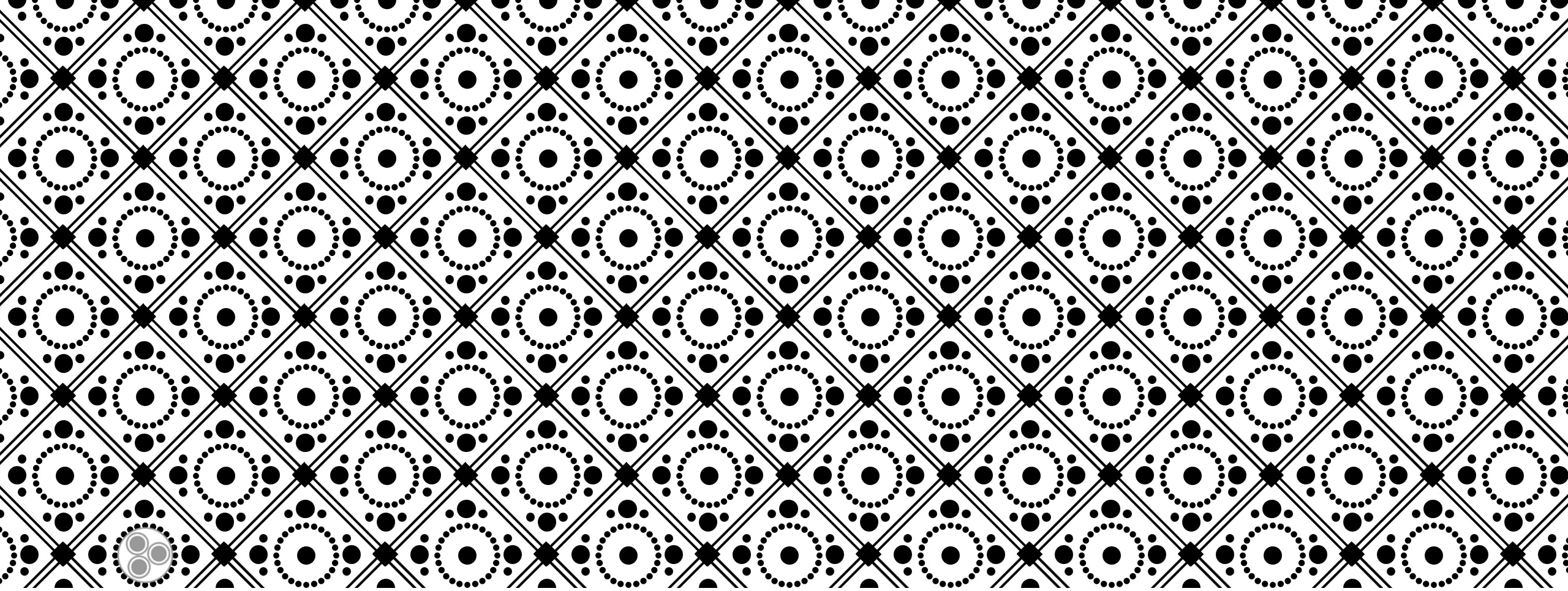
What should we do? What sort of data extraction from the Web should we conduct?

INSIGHTS? OBSERVATIONS? QUESTIONS?



You are Here!

- ❖ What are ways to tailor the initial data query to elicit the right information that you want?
- ❖ What are some strategies to prune unnecessary branches in the Machines?
- ❖ What are some ways to disambiguate the findings?
- ❖ What might be some effective strategies in terms of running the same data extraction over time at regular intervals? When might that be helpful? When not?



SOME RESOURCES AND REFERENCES

To probe further...if you're
interested

A SAMPLER OF ONLINE GEO TOOLS

Government Source

National Map Viewer (USGS): <http://nationalmap.gov/viewer.html> (a viewer for USGS map data)

Others

Veloroutes: <http://veloroutes.org/elevation/> (addresses to elevations)

Map Coordinates: <http://www.mapcoordinates.net/en> (simple interface to find latitude, longitude and sea level of locations on Google Maps)

Address Converter to Latitude/Longitude/Altitude:
<http://stevemorse.org/jcal/latlon.php>

RELATED TOOLS

Basic Complementary Tools to Run
with Maltego Tungsten

Data Hunting: Google

Mapping: Google Maps

Additional Capabilities

Data Extraction and Analysis: NodeXL,
NCapture of NVivo

(Palantir seems like a super dynamic version
of a data extraction and analysis tool.)

Text Data Analysis: AutoMap and ORA-
NetSense (of textual extractions for sentiment
analysis and gist), UCINET matrices

Graphing and Data Visualization:
GraphML, UCINET

REFERENCES

Andrejevic, M. (2007). *iSpy: Surveillance and Power in the Interactive Era*. Lawrence: University Press of Kansas.

Blum, A. (2012). *Tubes: A Journey to the Center of the Internet*. New York: Harper Collins.

Tucker, P. (2014). *The Naked Future: What Happens in a World that Anticipates your Every Move?* New York: Current, A Penguin Random House Company.



CONCLUSION AND CONTACT

Dr. Shalin Hai-Jew

- Instructional Designer

Information Technology Assistance Center (iTAC)

Kansas State University

212 Hale Library

785-532-5262

shalin@k-state.edu

Notes: The presenter has no tie to Paterva.

All images are either released through Creative Commons licensure...or are unique image captures from the presenter...or are clipart from the Microsoft library...or is a small screen capture from Paterva.

