Identity Analytics for Fraud Detection

Stephen Coggeshall
Chief Analytics and Science Officer
ID Analytics/LifeLock

July 30, 2013

Overview

- Who is ID Analytics
  - What data do we see
- What is the Field of Identity Analytics
  - Analytics around big data where the concept of identity is important
- Identity Manipulation
  - What is it, who does it
- Identity Fraud Rings
  - How to find identity fraud rings
  - Some examples of identity fraud rings
  - Examples of the overlap between commercial identity fraud rings and tax fraud rings

Who Is ID Analytics

- Founded in 2002
- Mission: build solution for growing identity fraud problem
- Methodology: cross industry data consortium, advanced analytics
- Identity risk management products to businesses, identity alerts to consumers
- Sold to Lifelock 3/2012. Now a wholly owned subsidiary
- Experts in identity fraud and identity risk management

Data Flow Visibility

- Core data is applications for account openings: credit cards, cell phones, retail credit (e.g., Nordstrom), payday loans, auto loans. ~1.8 billion applications over ~10 years

- Fraud applications: ~3 million fraud attempts

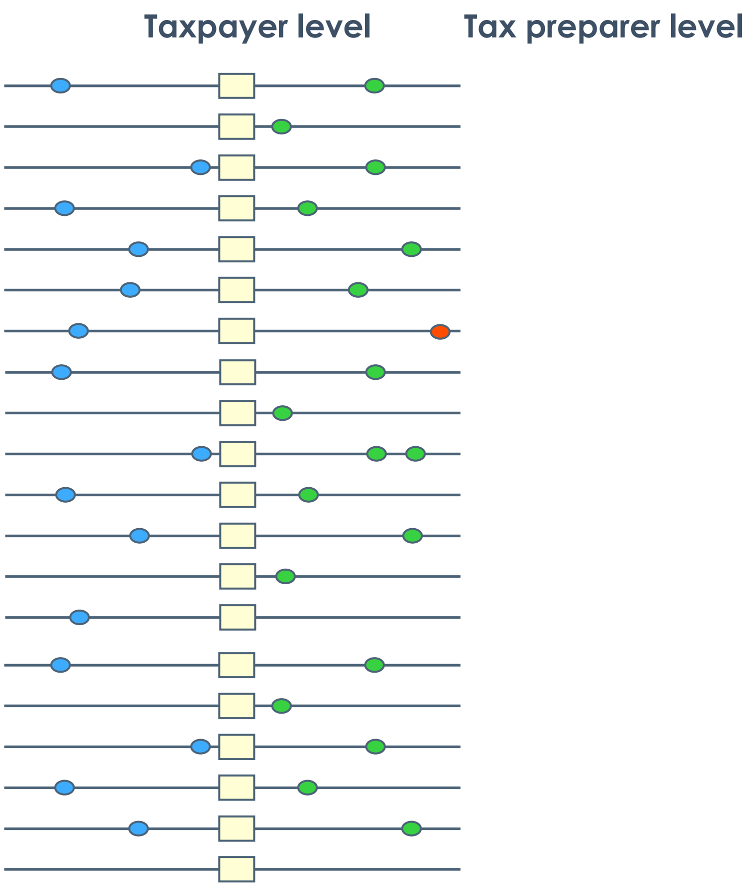
- Other large data sets:
  - U.S. white pages phone book data monthly for about 8 years
  - Credit bureau header files (PII only – SSN, Name, Address, Phone, DOB)(SNAPD)

- Other small data sets:
  - Change of addresses
  - Authentication quizzes
  - Consumer enrollments
  - Warm/hot addresses
  - OFAC lists
  - SSA DMF
  - SSA area group tables

What is the Field of Identity Analytics?

- Algorithms around large data sets to understand the identification details of entities (people, addresses, SSNs, phones, emails…)
- Example algorithms
  - ID Score – is this event a fraud attempt?
  - Identity Resolution – who is this asserted identity?
  - Identity Manipulation – who deliberately and improperly manipulates their identity
  - Identity Fraud Rings – who collaborates to commit fraud?
- These algorithms are combined to solve specific problems

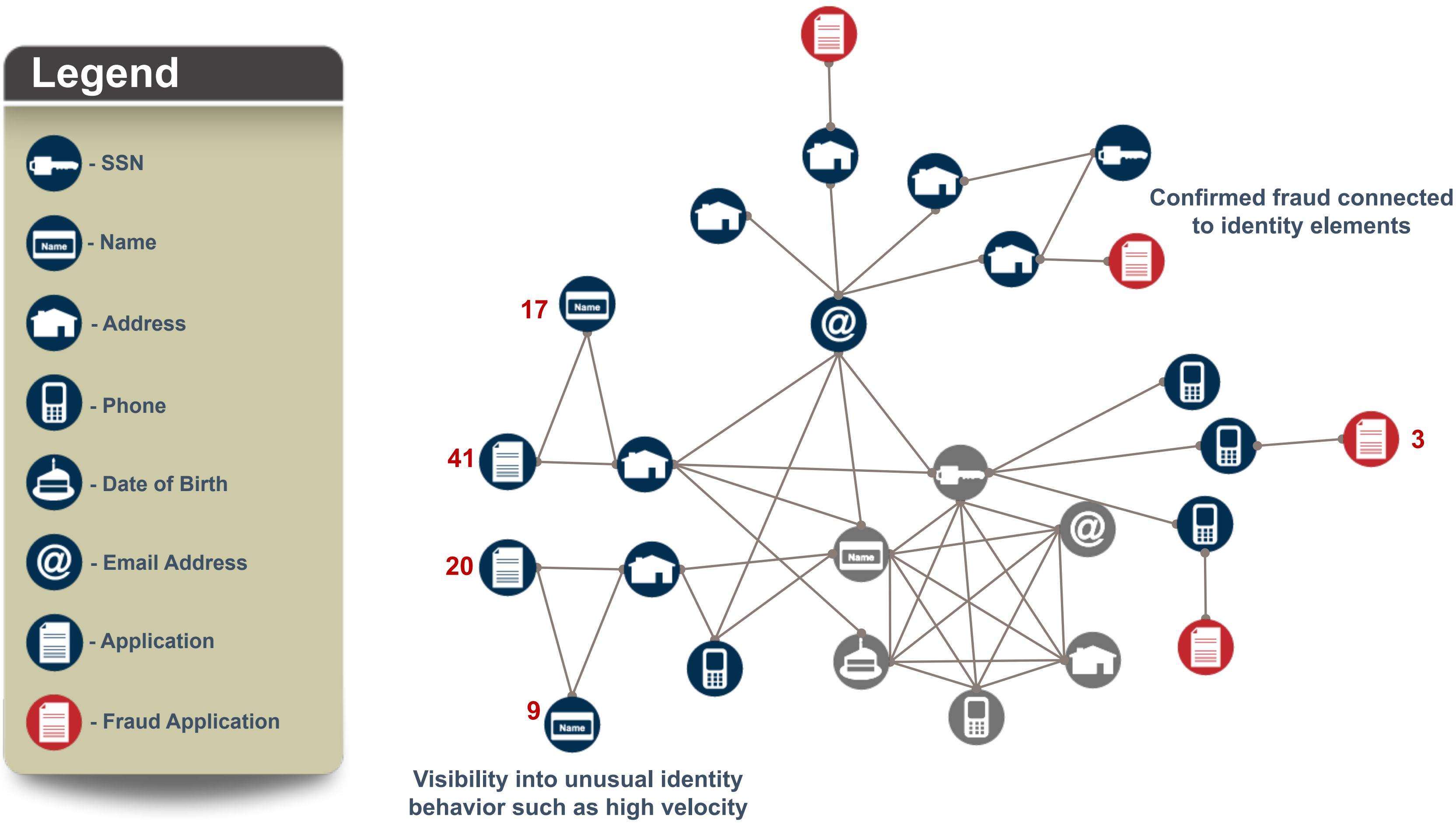
Entity Levels are Important for Solution Design



Identity Fraud Score – Is this event a fraud attempt?

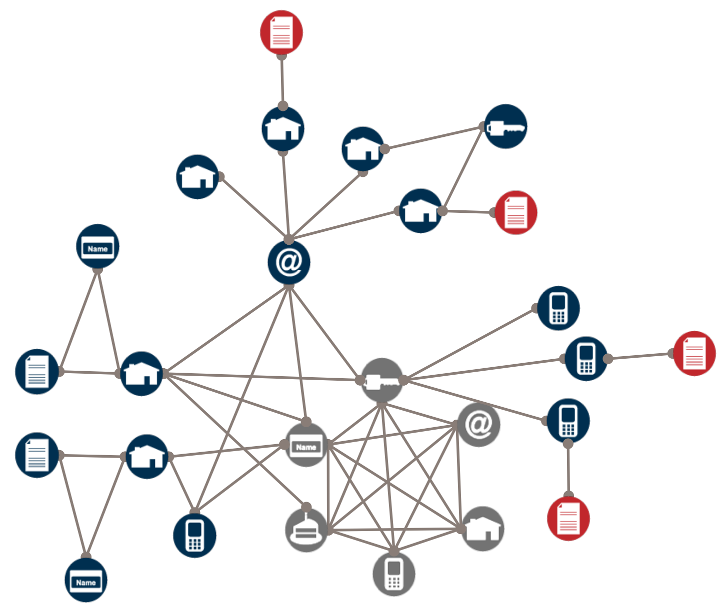
- In real time, businesses send us an account application containing PII (SNAPD)
- We examine the PII, look for unusual associations in the current and past events, build a score
- Return the score and reason codes all within a second
- If the PII on the app is related to a consumer enrolled in a monitoring service (Lifelock) we also send a real time alert to the consumer
- We score a few hundred million applications each year

Algorithms Use a Graph-Based Approach



What's Behind The ID Score?

- Receive the application (SNAPD: SSN, Name, Address, Phone, DOB)
- Build the PII-linked graph
- Translate this graph into numbers
- These features are the inputs to machine learning algorithms
- Calculate the score
- Return the score and reason codes



All this is done within one second

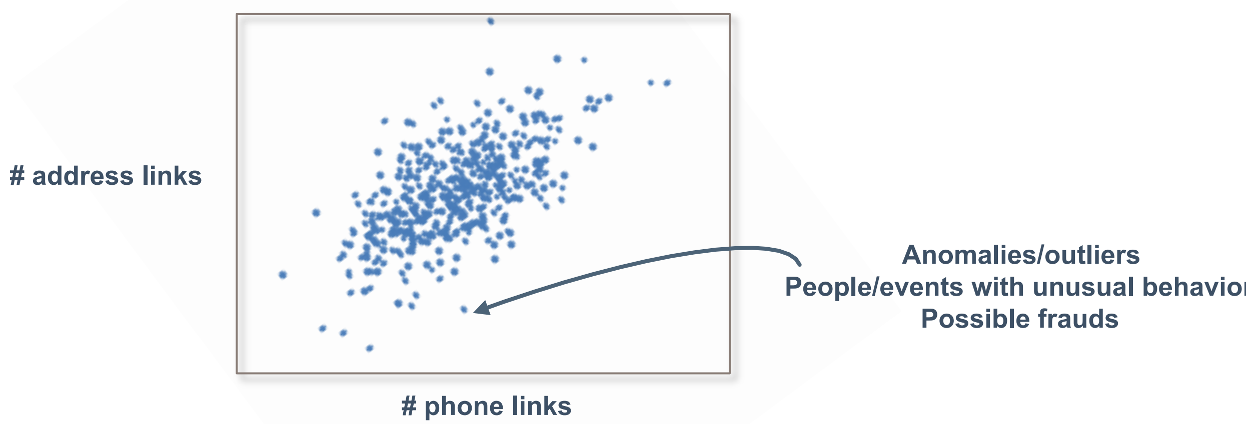
Requires very thoughtful data arrangement

How to Build a Fraud Score

- Clearly define the business objectives, uses, implementation, definitions
- Assemble data: many examples of past events with (hopefully) labels
- Separate data into training, testing and validation
- Encode data for modeling
  - Encode categoricals
  - Clean outliers
  - Encode and normalize continuous/ordinal variables
  - Build special expert variables
- Build preliminary models. Try many technologies. Examine goodness.
- Perform (out of time) validation
- Implement, monitor and improve

What To Do When You Don't Have Tags?

- How can you build a model to predict something when you don't have examples of it? In this situation you can build an **unsupervised model**
- Unsupervised modeling examines how the data points are distributed in the input space only (there is no output/label)
- You simply look at how the data points scatter. Can you find **clusters**? Can you see **outliers**?
  - **Clusters**: groups of people or events that are similar in nature
  - **Outliers**: unusual, anomalous events – potential frauds
- Can work well for fraud models where you're looking for anomalies
- Hardest part is figuring out what variables (dimensions) to examine



Identity Resolution – A Core Requirement

- We frequently see only partial PII (e.g., name, city...)
- We need to be able to resolve to the appropriate person
- Our mainline product ID Score answers “are you who you say you are?”
- To do this we first need to answer “who do you say you are?,” particularly when presented with limited PII.
- We need a capability to do this for hundreds of millions of records, so must be algorithmic, scalable and robust.

Identity Resolution – What

Input is some partial PII:

- John Smith, 123 main street 92130, 8583758374
- Peggy Hodges, 8/14/1978
- Freddy Thompson, 355-85-4857
- Armand Milejas, 4/??/1952, 849-28-3652

Maybe some input characters are wrong (typos, deliberate)

Identity Resolution Machine

Outputs

Unique person ID label, SSN, most common name representation, best address, phone, date of birth

- 6TY7F4DX8, 535-64-2871, John Theodore Smith, 123 Main Street 92130, 8586478493, 5/25/1976
- 5IHG8FD4J, 483-83-3827, Margaret Joyce Hodges, 75 Highway 72 87485, 20395684738, 8/14/1978
- J8GK5FDH9, 355-85-4857, Fred Kirk Thompson, 8374 Chestnut Ave 74982, 31295488763, 6/04/1983
- 5MU43JH8L, 849-28-3642, Armand Milejas, 327 Landcaster Ave 27387, 4838279854, 4/26/1952

- Figures out who this is
- Fills in missing information
- Corrects typos or deliberate variations

Identity Manipulation – Who Does It?

- Identity Resolution determines the person associated with an event
- Sort all applications by person and examine variations of PII that people use

First Name	Last Name	SSNs	DOBs	Address	Zip	Phone
Anitra Latasha Leland	Johnson McWillan McMillan McNulty	857379684 857369608 857379648	5/11/1983 6/11/1983	307 Granada Dr 14 Walker Ave # 31 1974 Spring Dr 939 East St 1938 5th St 312 Winona St	72994 72091 71946	7345639409 7345638758 3879847364 3878748763
3 different First Names	4 different Last Names	3 different SSNs	2 different DOBs			

- We built an "Identity Manipulation Score" to quantify unusual variations
- Scored 300 million people. Here are some bad ones:

Name	City	IM Score	# SSNs	# DOBs	# FNs	# LNs
John	NY	999	56	12	3	2
Wendy	Detroit	998	23	7	5	10
Dawn	Atlanta	997	15	12	8	7

Identity Fraud Rings

- We want a process to systematically find many fraud rings rather than find a few by inspection

- Solution:
  1. Gather a small group of highly-likely "bad" people (high IMs, known ID thieves…)
  2. Look for groups of bad people that are interconnected ("bad" people sharing addresses, phone numbers or email). Remove "singletons" and only keep groups.
  3. Remove likely false positives:
     - Kiosks
     - Business addresses
     - Other common or frivolous addresses, phone numbers or emails

- The remaining groups of interconnected bad people are very good candidates for identity fraud rings. Because they share an address, phone or email they likely know each other and are communicating/ sharing information.

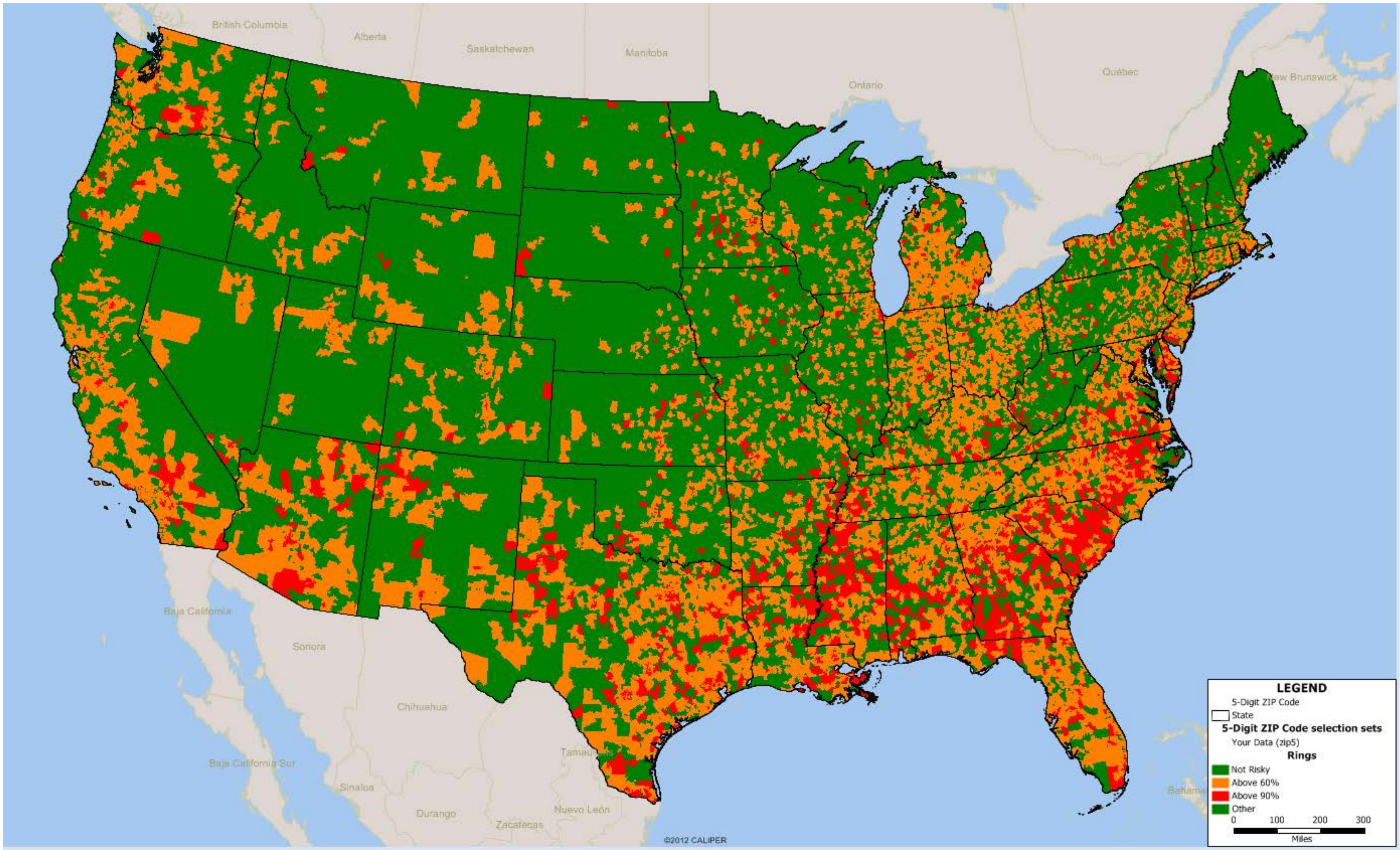
Example Identity Fraud Rings

Ring ID	City	# Applications	# People	# SSNs	# Credit Cards	# Cell Phones	# Retail Credit	# First names	# Last Names	# Dates of Birth	# Addresses
61261	Dearborn, MI	265	15	15	7	253	5	16	16	23	18
24584	Orangeburg, SC	60	4	8	15	34	3	9	6	6	2
48093	Queens, NY	352	6	14	85	7	260	13	19	16	18
36748	Bessemer, AL	283	13	54	42	233	2	30	17	32	8
91340	Las Cruces, NM	285	14	14	55	39	186	14	10	16	8
14673	Miami, FL	133	4	31	7	118	4	12	9	20	7
27323	Tampa, FL	55	4	8	13	10	27	4	2	4	6
12218	Macon, MS	66	4	15	9	48	8	13	8	8	5
100898	Los Angeles, CA	189	4	25	5	174	10	11	15	26	5
50937	Phoenix, AZ	310	5	6	44	1	265	8	3	6	1
80667	Amarillo, TX	302	6	32	41	177	81	11	9	11	4
78629	Philadelphia, PA	178	4	59	4	161	13	6	5	10	3

More than 10,000 identity fraud rings

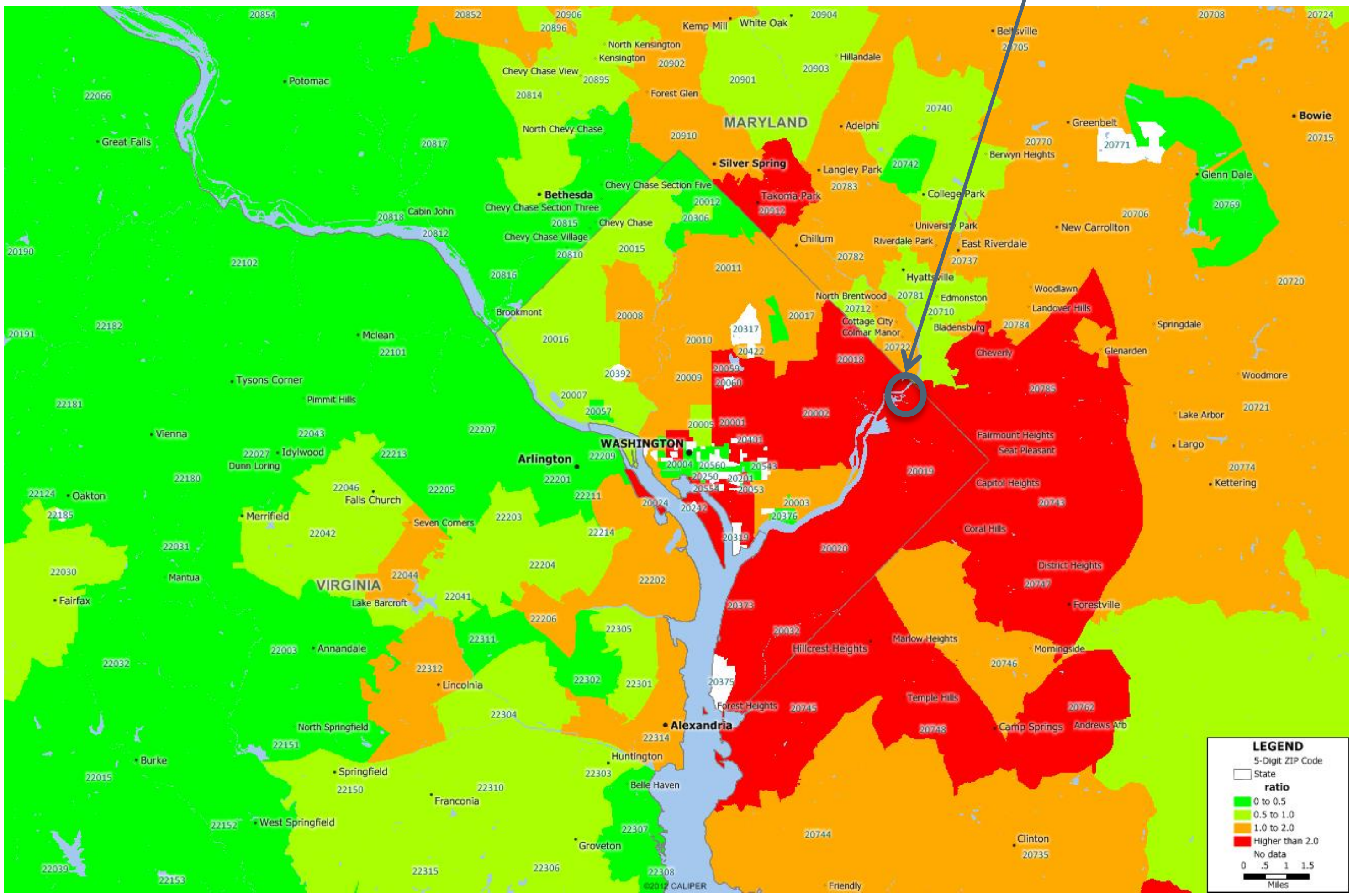
I know who they are, where they live, what they're doing around commercial products

Identity Fraud Ring Locations



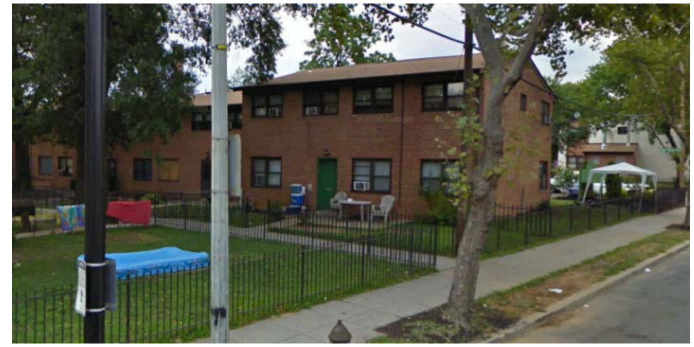
Example Fraud Ring #2101

Fraud ring #2101



Washington, DC area

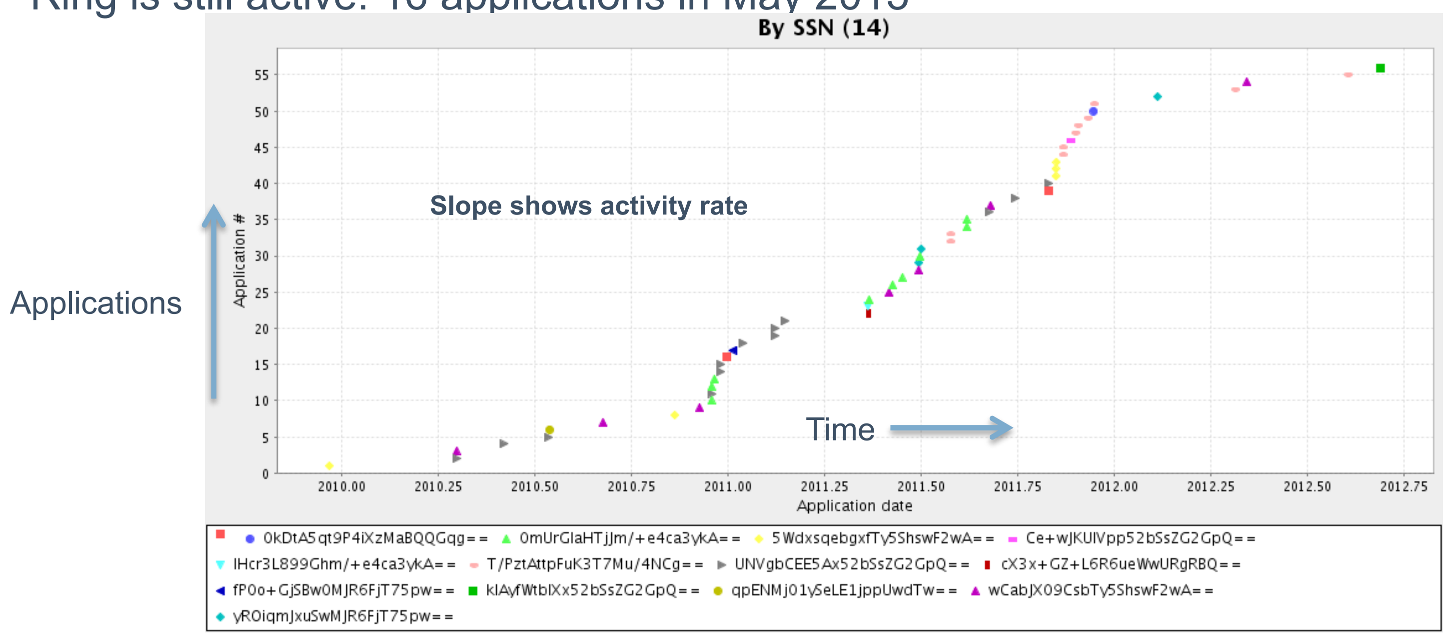
Example Fraud Ring #2101



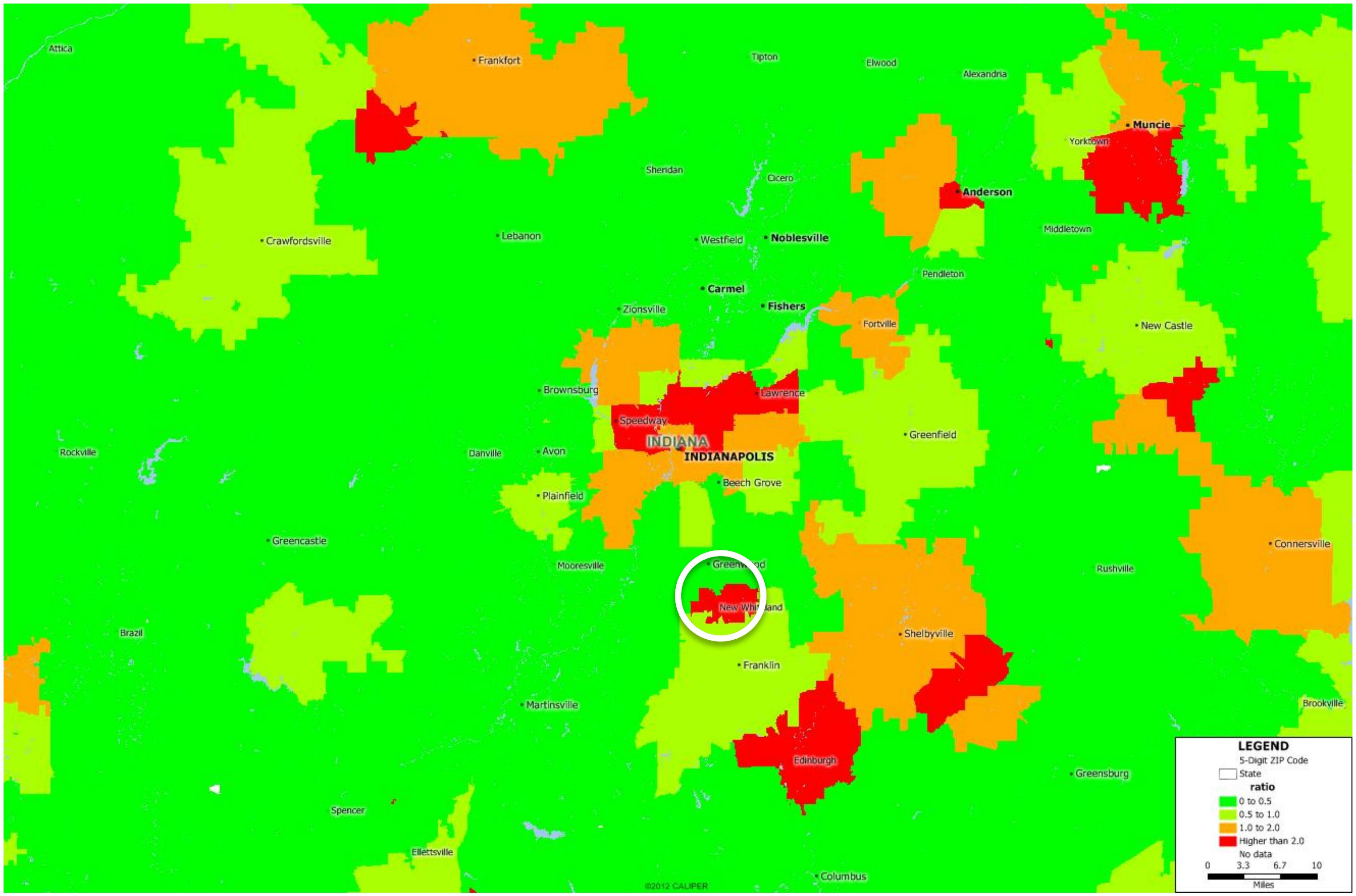
Six people, Family/Friends Identity Manipulation and Identity Theft

1) Gerald Smith, 24. Uses 2 FNs. 10 apps
2) Corona Jones, 24. Uses 2 SSNs, 2 LNs. 12 apps
3) Corona Jones, 52. Uses 3 SSNs, 2 DOBs, 2 FNs 5 apps
4) Monique Jones, 43. Uses 2 SSNs, 3 LNs.
5) Latasha Jones, 21. Uses 3 SSNs, 2DOBs. 26 apps
6) Angel Jones, 22. No identity manipulation 12 apps

- 85 phone & credit card applications from these 6 people from one address (close to Anacostia Park) in Washington, DC over the past 3 years. Sharing of SSNs, DOBs, names.
- An additional 5 ID theft victims; 2 victims are deceased (ID theft of the dead).
- Ring is still active: 16 applications in May 2013



Example Identity Fraud Ring #3062



Indianapolis, IN area

Example Identity Fraud Ring

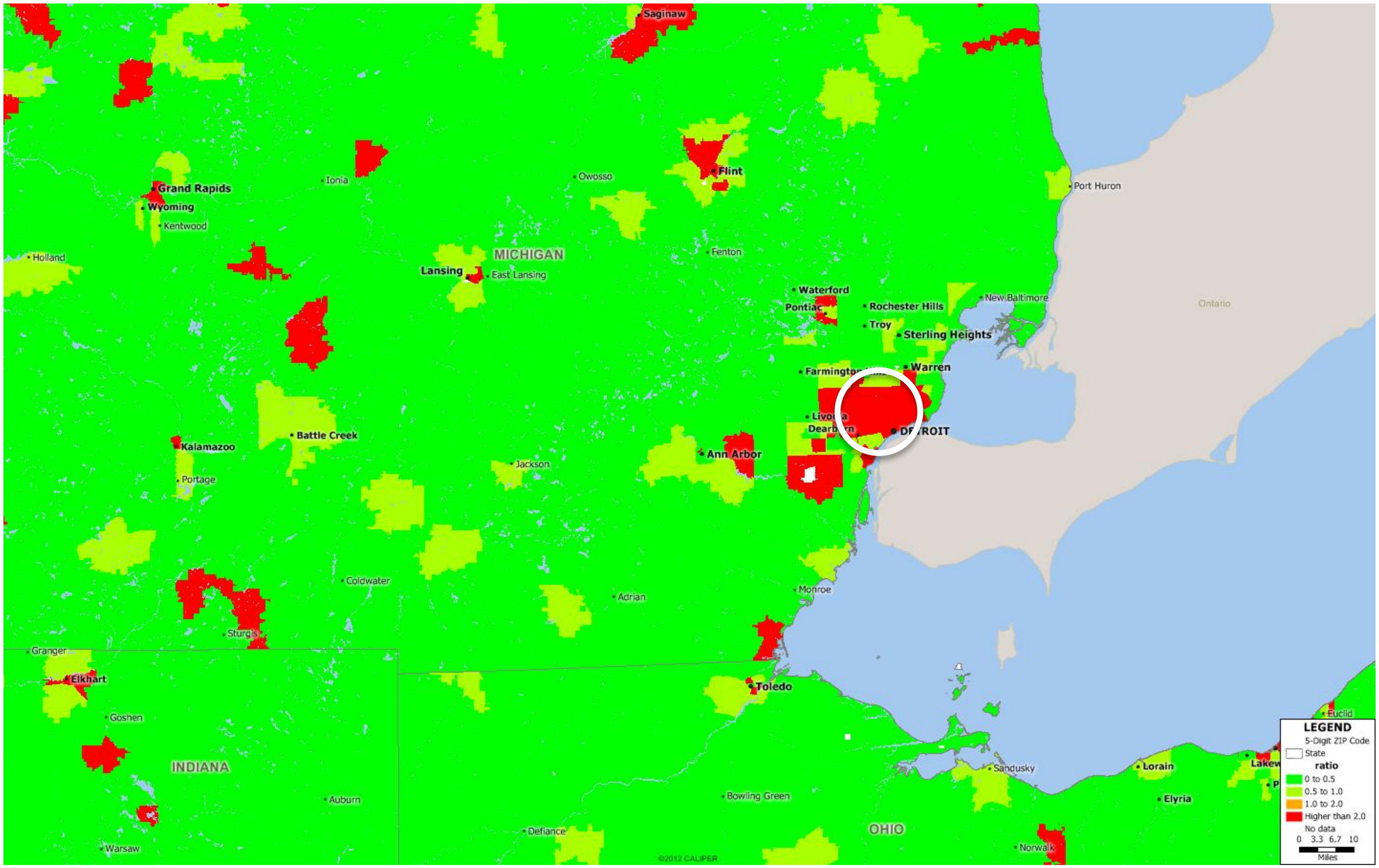
FRAUD RING NO. 3062 | FOUR PEOPLE, FAMILY/FRIENDS IDENTITY MANIPULATION FRAUD

First Name	Last Name	Age	Details	Apps
Hatti	Smith	48	12 SSNs, 3 DOBs, 2 FNs, 3 LNs	194
Frank	Smith	75	7 SSNs, 5 DOBs, 2 FNs, 2 LNs	117
Dottie	Smith	71	2 SSNs, 3 DOBs, 4 FNs, 2 LNs	10
Freida	Jones	48	2 SSNs, 2 LNs	24

- 345 credit card apps, 1 payday loan from these 4 fraudsters from 2 addresses. Frank uses his college email (retired professor).
- Started with Frank and Hatti, then Dottie. Later Freida joined.
- ID Score caught 97%

Indianapolis, IN area

Example Identity Fraud Ring #29235

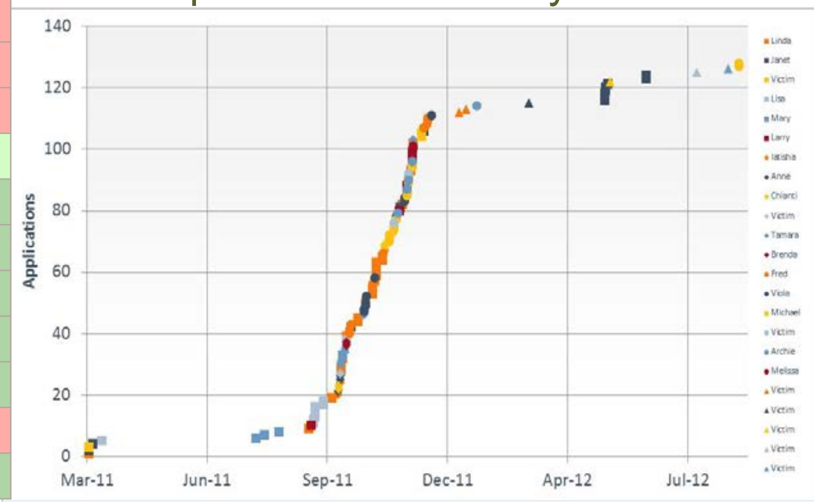


Detroit, MI

Example Identity Fraud Ring #29235

Who	What	Age	# Apps	# SSNs	# DOBs	# FNs	# LNs
Linda	Fraudster	37	28	2	2	1	2
Michael	Fraudster	57	13	2	1	1	2
Janet	Fraudster	18	13	1	2	1	3
Fred	Fraudster	56	12	1	3	1	2
Lisa	Fraudster	22	10	2	1	2	5
Viola	Fraudster	54	10	3	3	3	4
Tamara	Fraudster	21	6	1	2	1	2
Larry	Fraudster	32	6	1	1	1	1
Archie	Fraudster	40	5	1	2	1	1
Chianti	Fraudster	22	4	2	1	4	1
Mary	Fraudster	56	4	1	2	1	2
Latisha	Fraudster	26	3	1	3	4	1
Anne	Fraudster	31	3	1	2	4	3
Melissa	Fraudster	43	2	1	2	3	2
Brenda	Fraudster	36	2	4	2	2	2
Johnny	Victim?	24	3	1	1	1	1
Christopher	Victim	31	2	1	1	1	1
Aaron	Victim	43	2	1	1	1	1
Chuangaree	Victim	40	1	1	1	1	1
Karen	Victim	21	1	1	1	1	1
Michael	Victim	31	1	1	1	1	1
Hazel	Victim	42	1	1	1	1	2
Leon	Victim	61	1	1	1	1	1

- 133 cell phone applications from 2 addresses, ID Score caught 95%
- Mother and 2 daughters. Father(?) and 2 children
- Family and friends doing identity manipulation and identity theft



Can We Find Tax Fraud Rings In Our Commercial Data?

- We have minimal information (name, age, city) on three tax identity fraud rings from newspaper articles
  - Statesboro, GA
  - Long Island, NY
  - Los Angeles, CA

- Can we find the “published” people in our commercial activity data?

- If yes, do we see any connections with the fraud rings we find in commercial activity?

Statesboro Tax Fraud Ring

- 21 individuals in Georgia, 1 in Ohio, and 1 in Florida were charged with illicitly obtaining and utilizing individuals PII from medical records to obtain tax refunds
- Weak overlap with our commercial fraud ring data (Ring #110293), only linked by a previous address.
- But three of these tax fraudsters also have substantial history manipulating their identities on commercial products

We find evidence of identity misuse by some of the arrested people in our commercial data

Long Island Tax Fraud Ring
“Operation Refund Racket”

- 7 people arrested filing $60,000 in fraudulent tax refunds
  - 5 more people sought after in the investigation
- Returns were loaded to debit cards and then used to purchase money orders
- One person has substantial application velocity **after** being arrested, attempting to get a cell phone(s)
- One identity identified as being part of commercial fraud ring (Ring #24790)
- Majority of activity linking to a home in Hempstead, NY
  - Activity significant at end of 2011
  - Cell phones targeted in this activity
  - 18 identities identified in this ring
  - 6 identities have significant history manipulating PII

We see overlap of this tax fraud ring with an identified fraud rings in our commercial data

Los Angeles Tax Fraud Ring

- On February 27, 2013 eight people were charged in connection to using stolen identities to commit $19 MM in tax refund fraud:

Artak (Max) Berberyan, 33	Van Nuys
Armen (Roman) Berberyan, 33	Van Nuys
Suren (Sunny) Gambaryan, 33	North Hollywood
Ashot Karapetian, 47	North Hollywood

Vigen (Vic) Tsaturyan, 47	Sun Valley
Arman Zargaryan, 30	Granada Hills
Akop (Jack) Kantrdzyan, 33	Sylmar
David (Little Guy) Samsonyan, 31	Winnetka

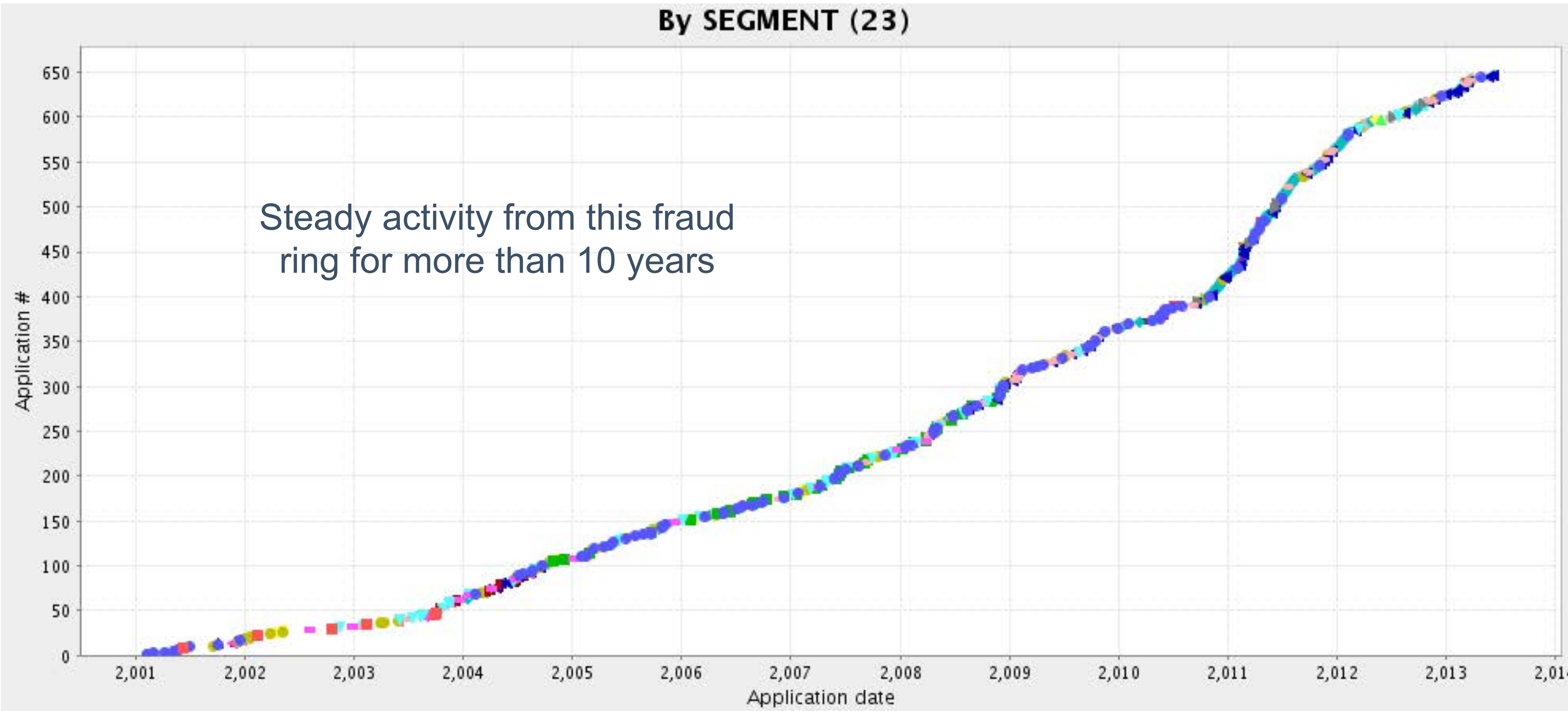
- Accused of fraudulently filing more than 2,500 income tax returns, using more than 1,800 identities
  - Many victims are retirees and residents of homeless shelters

- 6 individuals were arrested and 2 of them were still sought by authorities

- ID Analytics used the names and ages from the article and found that one individual was involved in a fraud ring that we identified (Ring #8736)
  - Akop Kantrdzyan, has a high Identity Manipulation Score

Ring #8736, LA Armenian Ring

Observed Commercial Activity

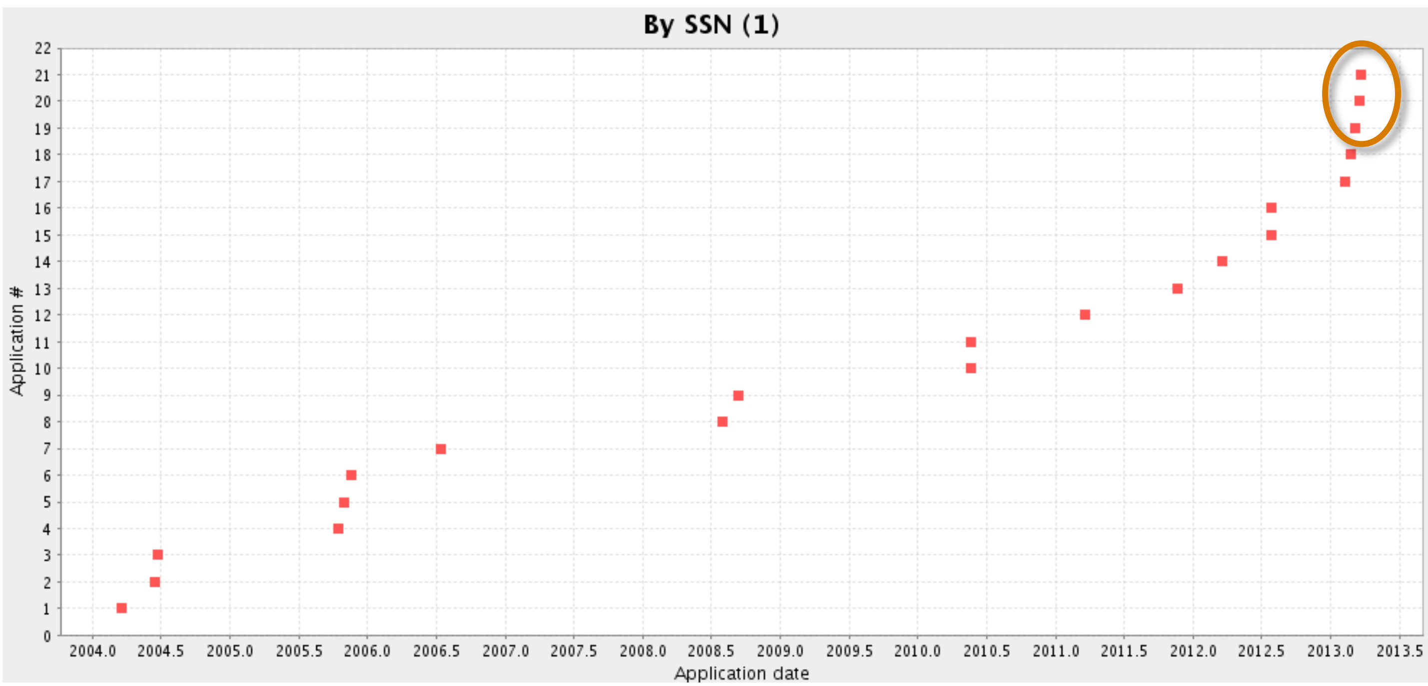


Dark blue – credit card
Light blue – retail credit
Green – cell phone
Red – utility

Ring #8736, LA Armenian Ring

Sought-After Fugitive Is Active After Others Detained

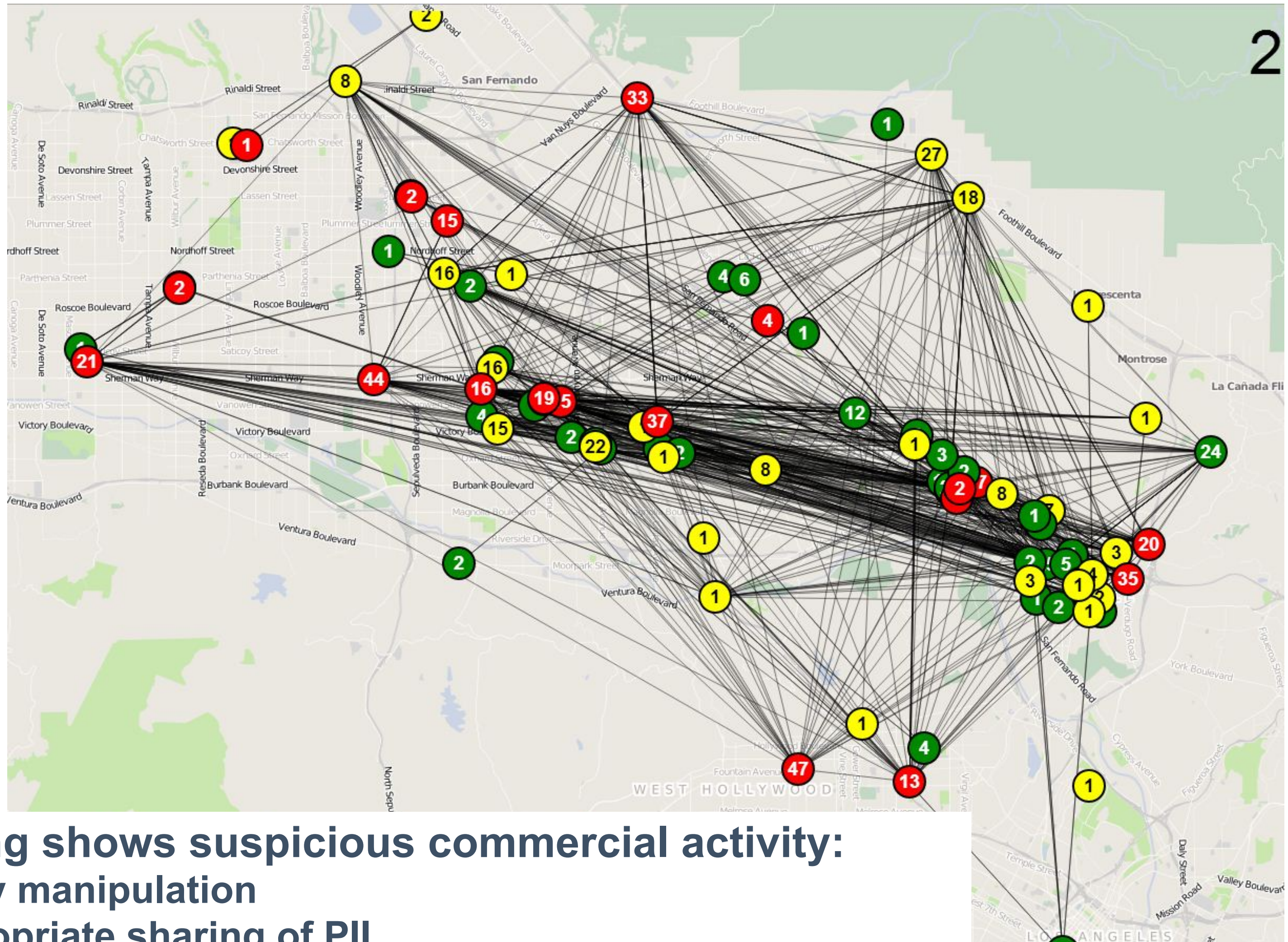
- One individual submitted 3 applications immediately after 2/28 article date calling him a sought-after individual
- He applies for two bank cards and a cell phone between 3/2 and 3/17
- Applications submitted using two addresses (one address twice – once listing unit – one not listing). We know the addresses.



id:analytics

Ring #8736, LA Armenian Ring

Evolution Shows Anomalous Activity and Connections



Fraud ring shows suspicious commercial activity:

- Identity manipulation
- Inappropriate sharing of PII
- Anomalous activity volume

Summary

- There exist well-developed tools in the private sector to find and prevent identity fraud
  - Identity fraud scores
  - Identity manipulation scores
  - Identity fraud ring membership scores

- We see overlap with tax identity fraud rings and commercial identity fraud rings

- This can provide substantial help for
  - Investigating known tax fraud rings
  - Finding additional people associated with known tax fraud rings
  - Finding previously-unknown tax fraud rings

id:analytics®

id:analytics®