Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem

Kyle Soska
Carnegie Mellon University
ECE / Cylab
ksoska@cmu.edu

Nicolas Christin
Carnegie Mellon University
ECE / Cylab
nicolasc@cmu.edu

Carnegie Mellon

Conventional Commerce

Internet Commerce

Conventional Illicit Commerce

Illicit Internet Commerce

Anonymous Marketplaces

- Amazon.com of illegal goods
  - Drugs, CC's & Fake IDs, Weapons, etc.
  - No Child Porn
- Safety
- Convenience
- Variety
- Accountability
- Competition

Anonymous Marketplace Technology

- Hidden Website (Tor Hidden Service, I2P)
  - Customers
    - No cost of creation
    - No information needed
  - Vendors
    - Vendor bonds required
    - Often invite only
    - Public feedback history

- Payments (Bitcoin)
  - Marketplaces often act as escrow agent
  - Escrow sometimes acts as a mixing service

- Hidden Messages(PGP)

Market Transactions

"I'll take the red pill"

Market Transactions

"1 BTC please"

Market Transactions

Deposit 1 BTC
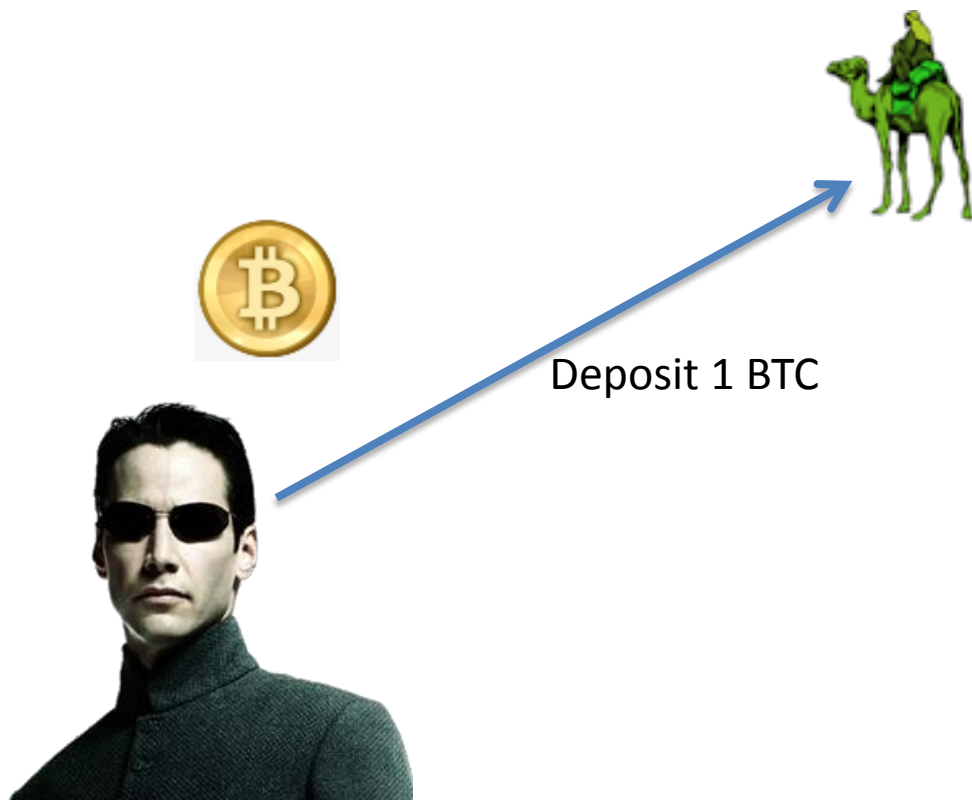
Market Transactions

Funds ok

Carnegie Mellon
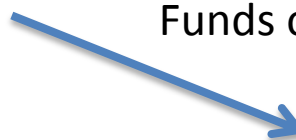
Market Transactions

Market Transactions

Received
"Excellent seller, would do business
with again. A++++"

Market Transactions

Deposit 0.9 BTC
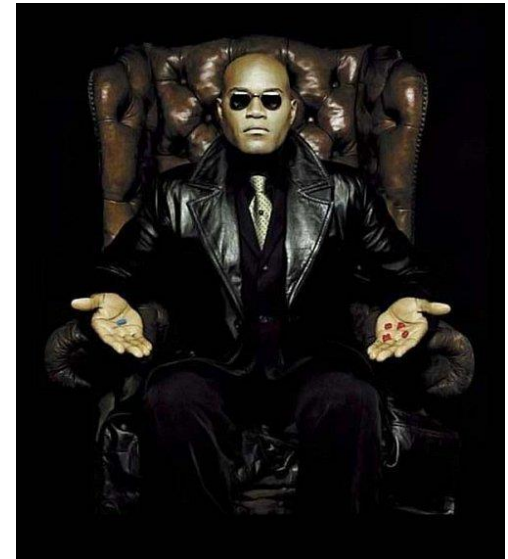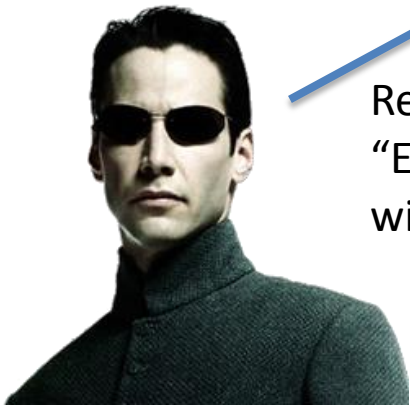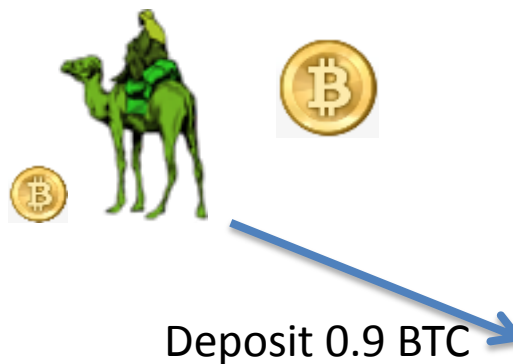
Questions

- How much is being sold?

- What is being sold?

- How many vendors are relevant?

- What do vendors sell?

Measurement Platform Overview

Manual Login / Solve CAPTCHA

Browser

Cookie / Session

Scraper

Tor 1

Tor 2

⋮

Tor 20

Marketplace.onion

HTML Only

Raw DB

Parser

Parsed DB

Analysis

3.2 TB

30 GB
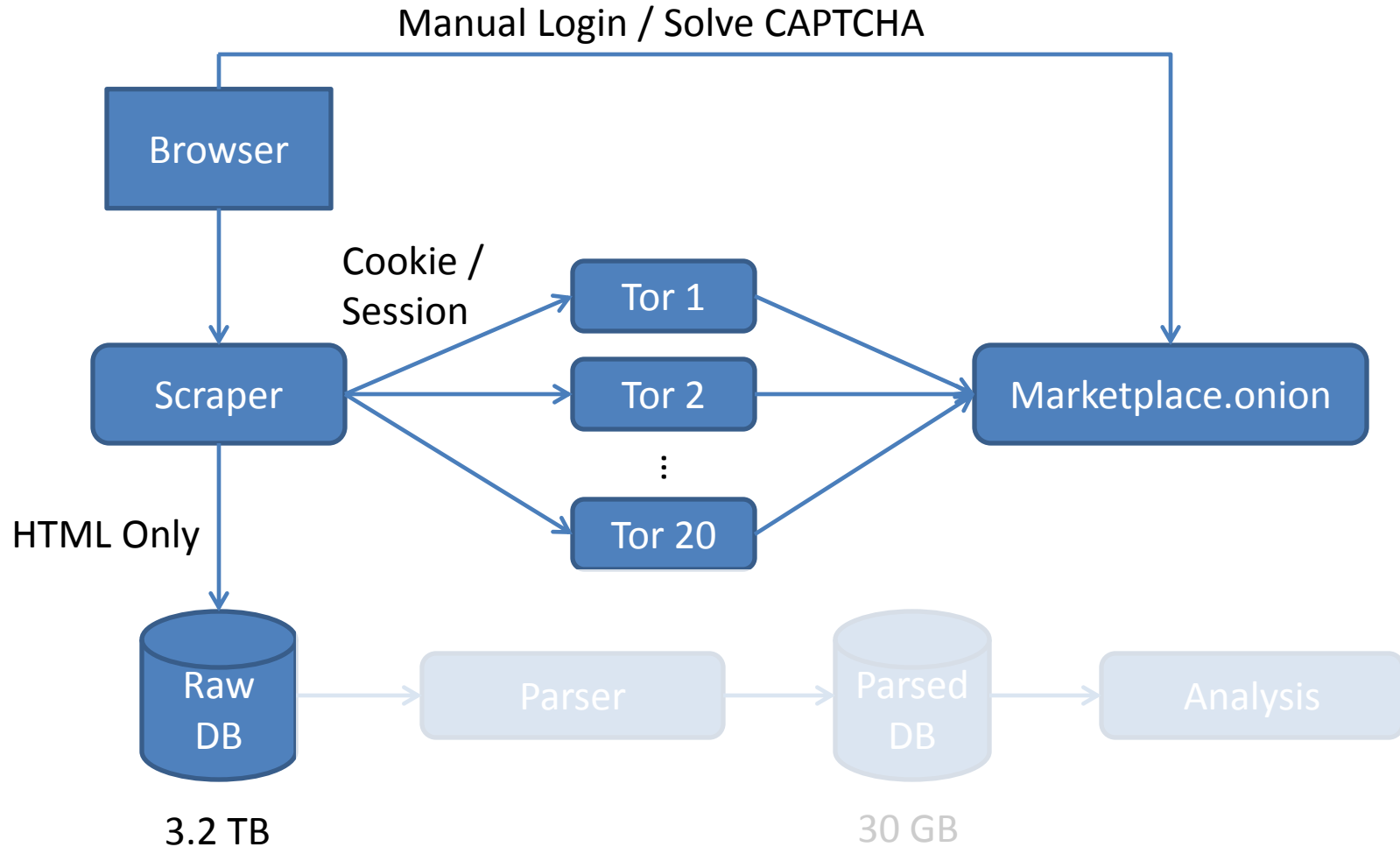
Measurements

- **Stealth**
  - Indistinguishable from real user
  - Random delays, scrape slowly
  - Popular User Agent
  - Browse website "normally"
- **Complete and instantaneous**
  - Dynamic marketplace, moving target
  - Scrape quickly
  - Site availability as low as 70%

Measurements

- **Anti-Scraping Encountered**
  - Rate Limits
  - Cookie Timeout
  - User Account Suspension

- **Totals**
  - 35 Marketplaces 1,908 scrapes total – 3.2 TB
  - 27 – 331,691 pages per scrape
  - 11/22/11 – present

Parsing

Manual Login / Solve CAPTCHA

Config

Browser

Site Layout

Cookie / Session

Scraper

Tor 1

Tor 2

⋮

Tor 20

Marketplace.onion

HTML Only

Raw DB

Parser

Parsed DB

Analysis

3.2 TB

30 GB
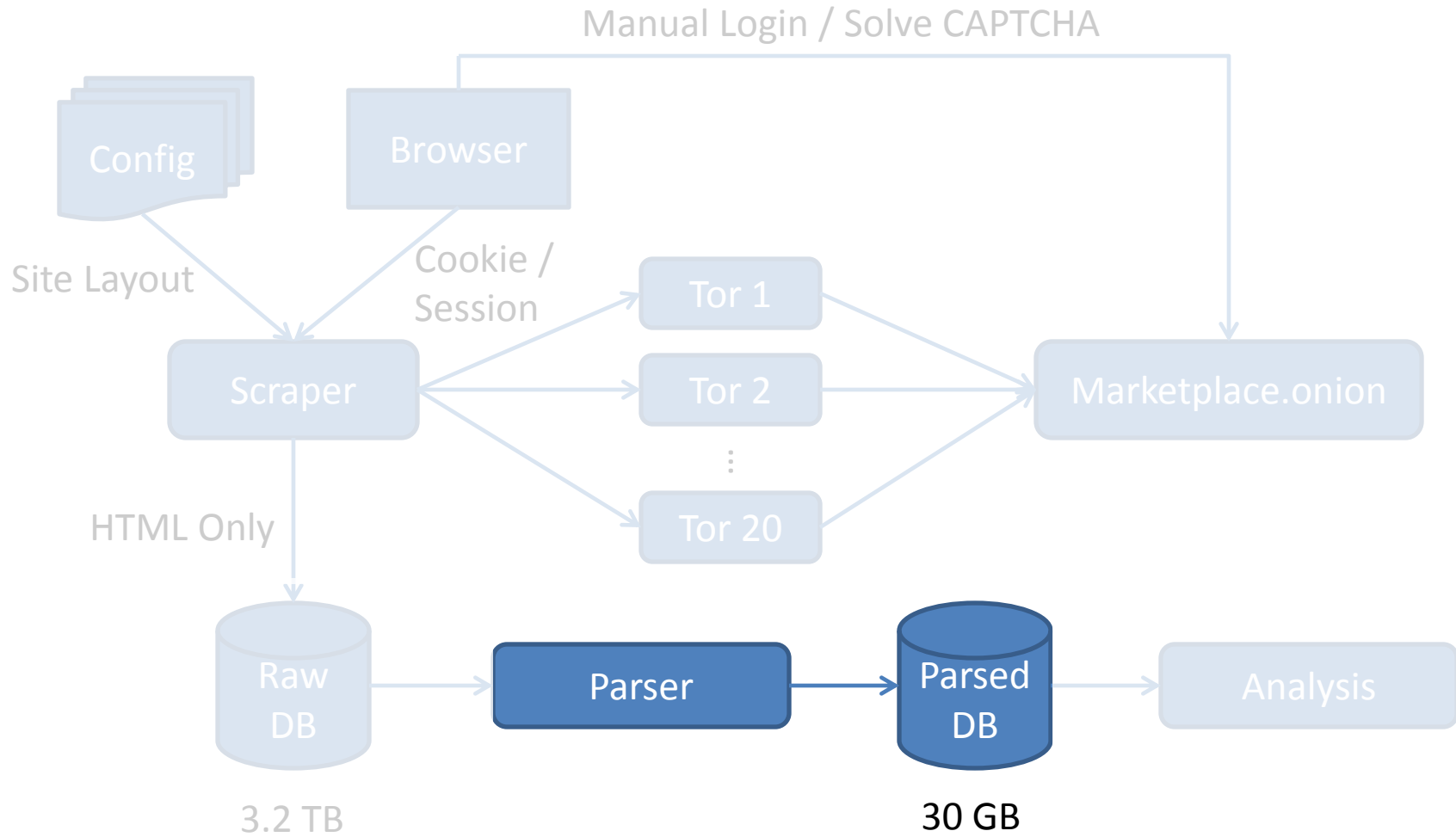
Silk Road Available Data

Books

Hacking for beginners

Seller:
(98)

Price:
฿0.12

Ships from: undeclared
Ships to: Worldwide

Description:
Hacking For Beginners is a reference book for beginners to learn ethical hacking for free and from basic level to clear all the fundamental concepts of ethical hacking.the book has been prepared by Hacking Tech (www.hackingtech.co.tv) website for the users benefit.so enjoy the book and site...

add to cart

Recent feedback

rating	feedback	freshness
5 of 5	Fast delivery	3 days
5 of 5	Thanks!	4 days
5 of 5	Leave feedback here	9 days
5 of 5	Leave feedback here	9 days
5 of 5	5 of 5	10 days
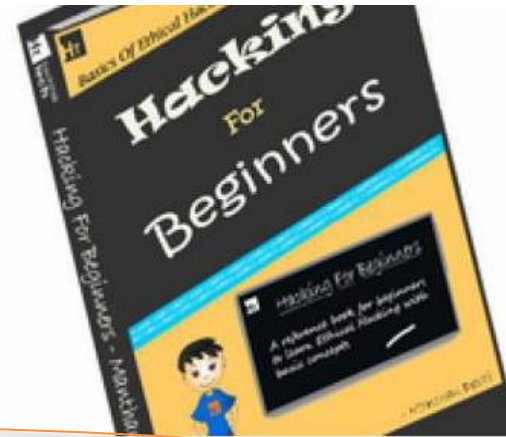
Feedback is often **mandatory!**
➔Acceptable proxy for sales volume

Analysis
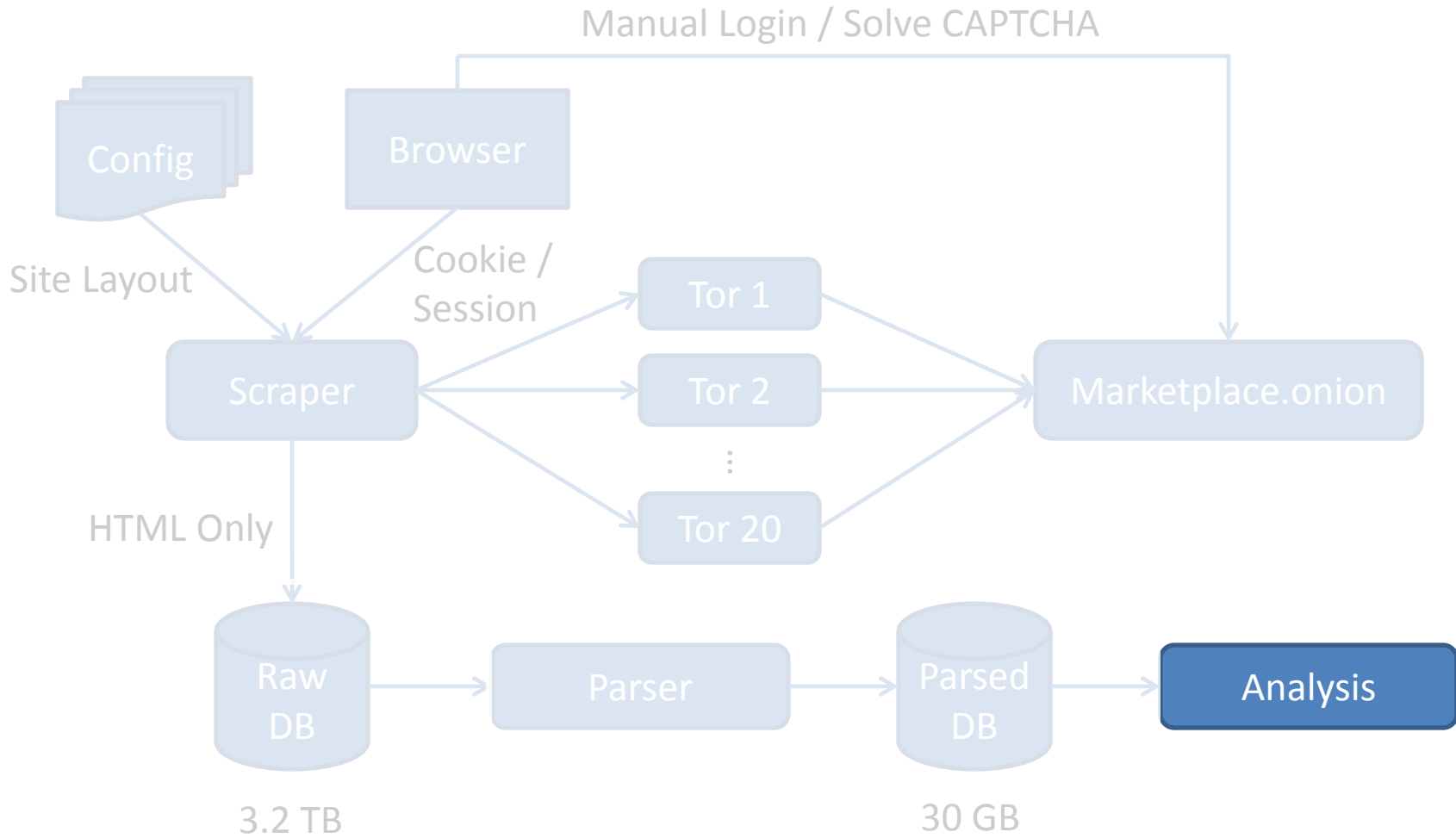
Data Completeness

- **How complete is the data?**
  - Unreliable dynamic marketplaces that take days to scrape
  - Empirical observations - lower bound

- **Idea:** Estimate population via mark and recapture
  - Schnabel Estimator allows multiple recapture

Mark and Recapture

Population Size = 24
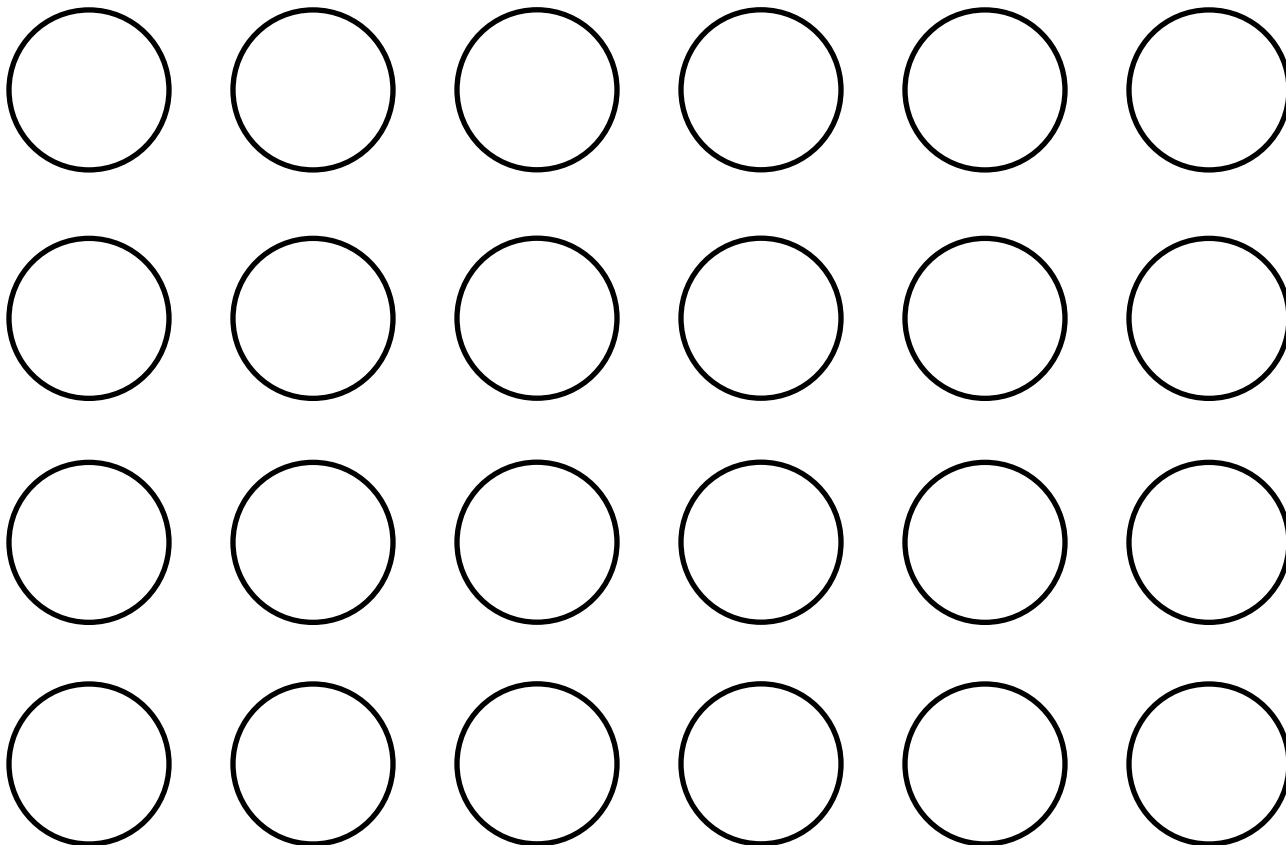
Mark and Recapture

Sample Size = 10
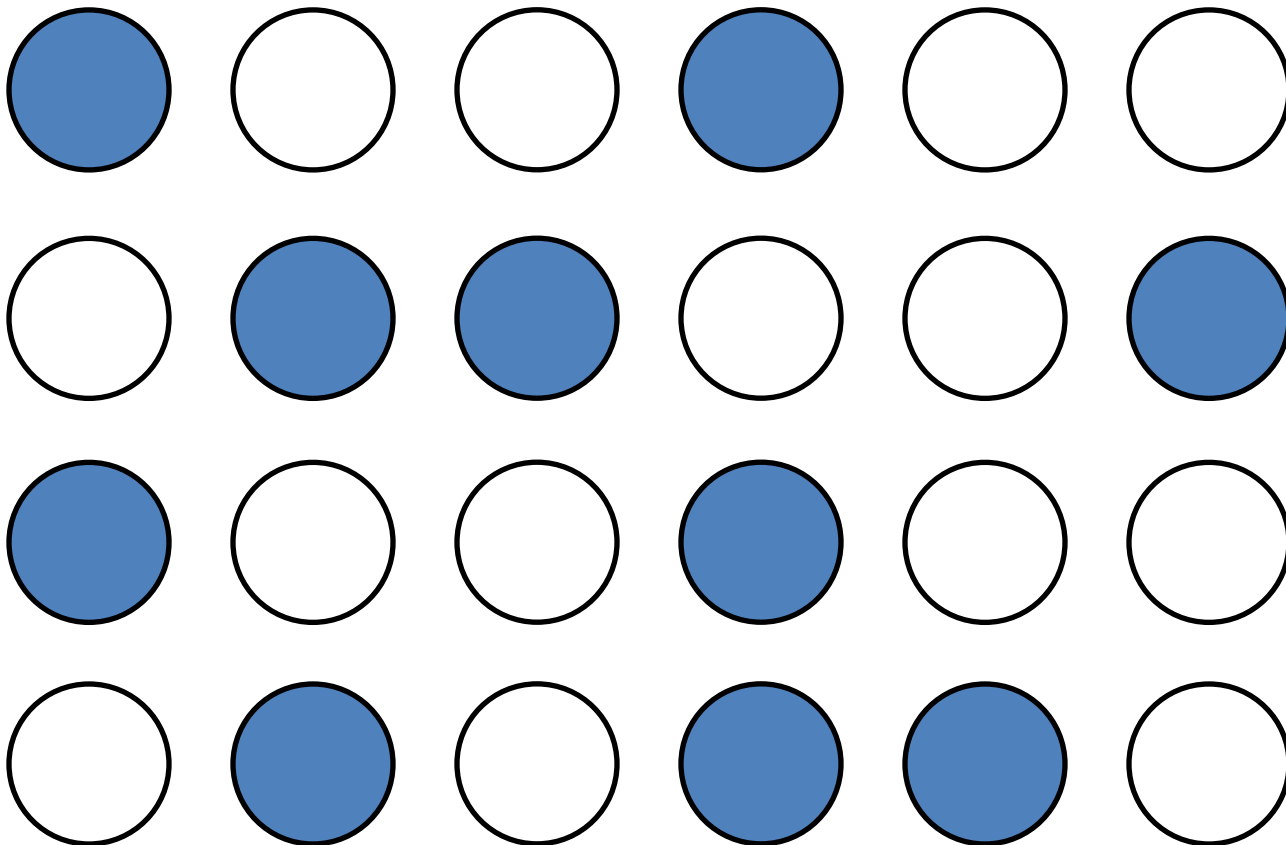
Mark and Recapture

Sample Size = 13
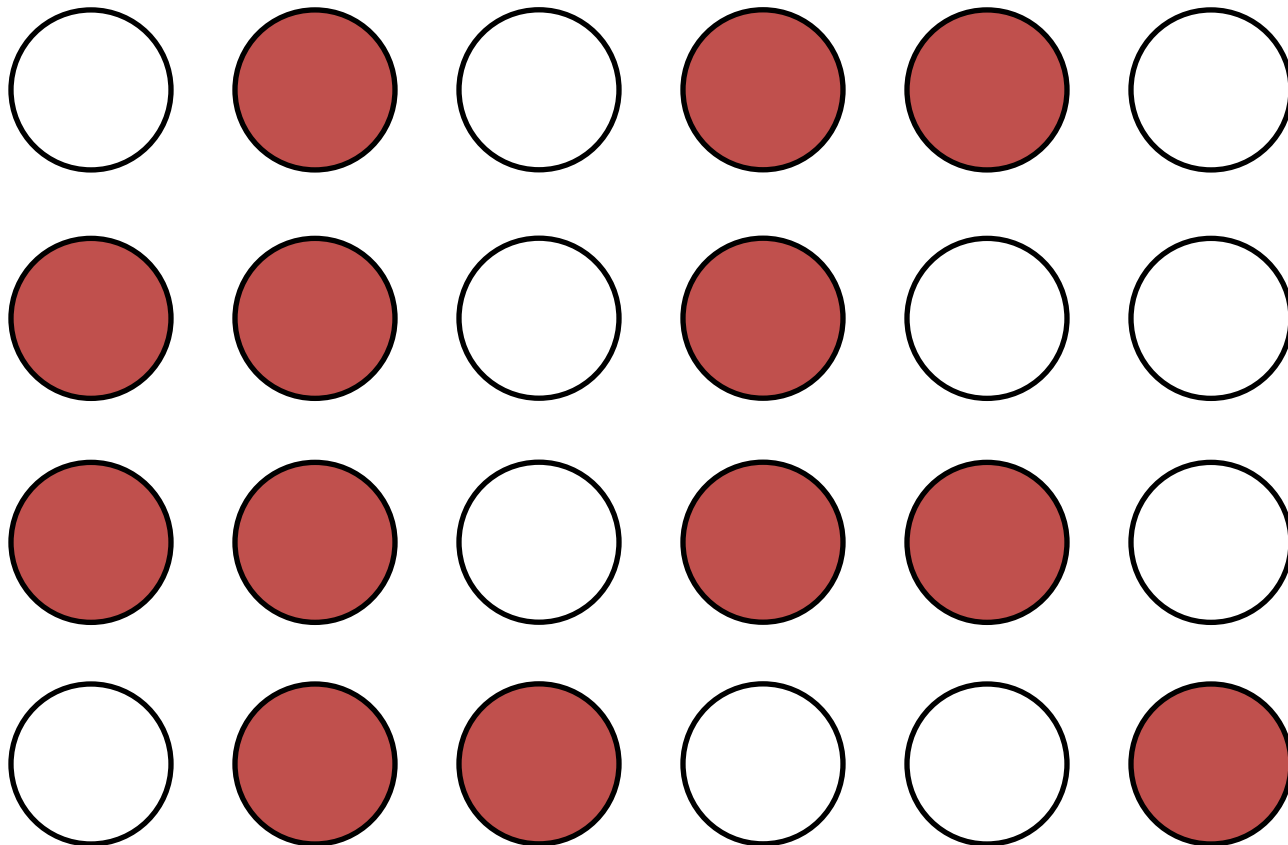
Mark and Recapture

Overlap = 5, Population Estimate = 26
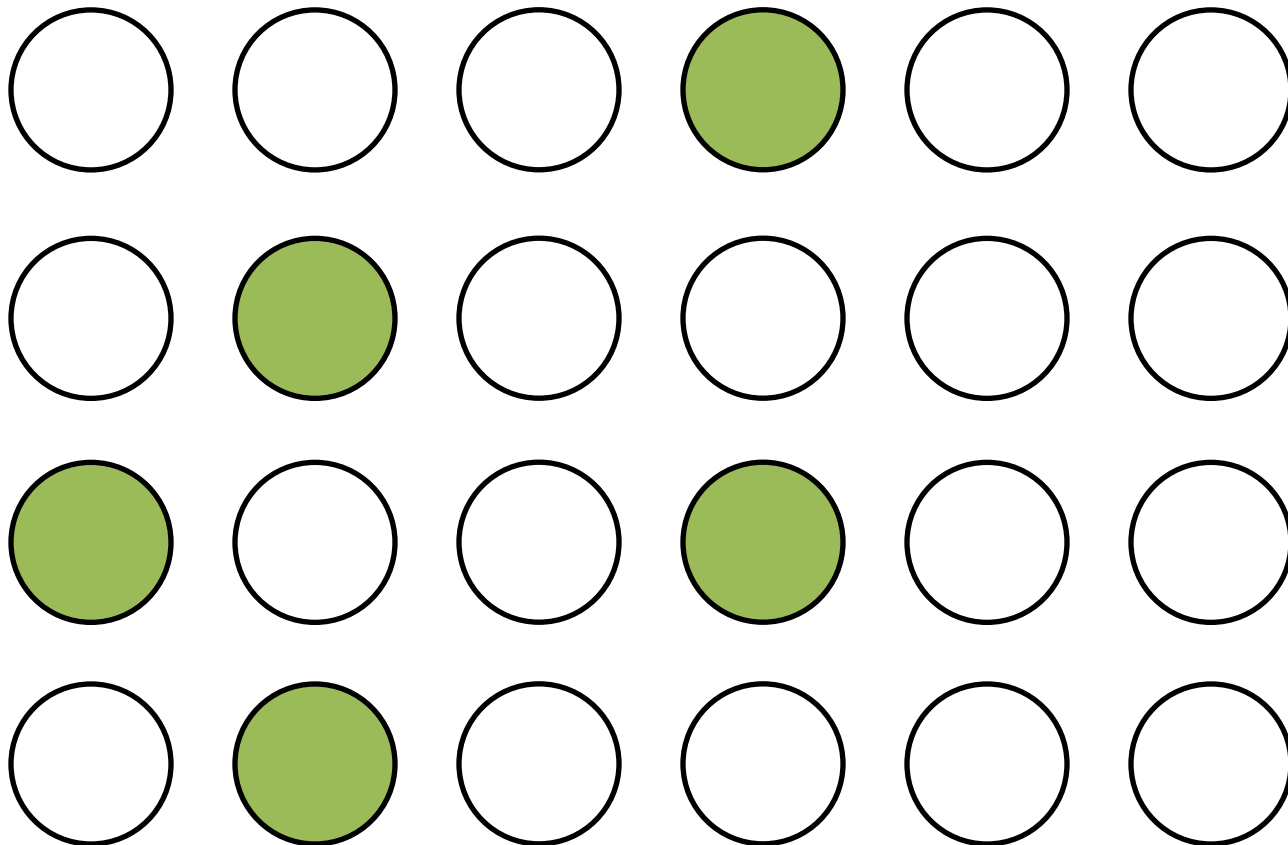
Data Completeness
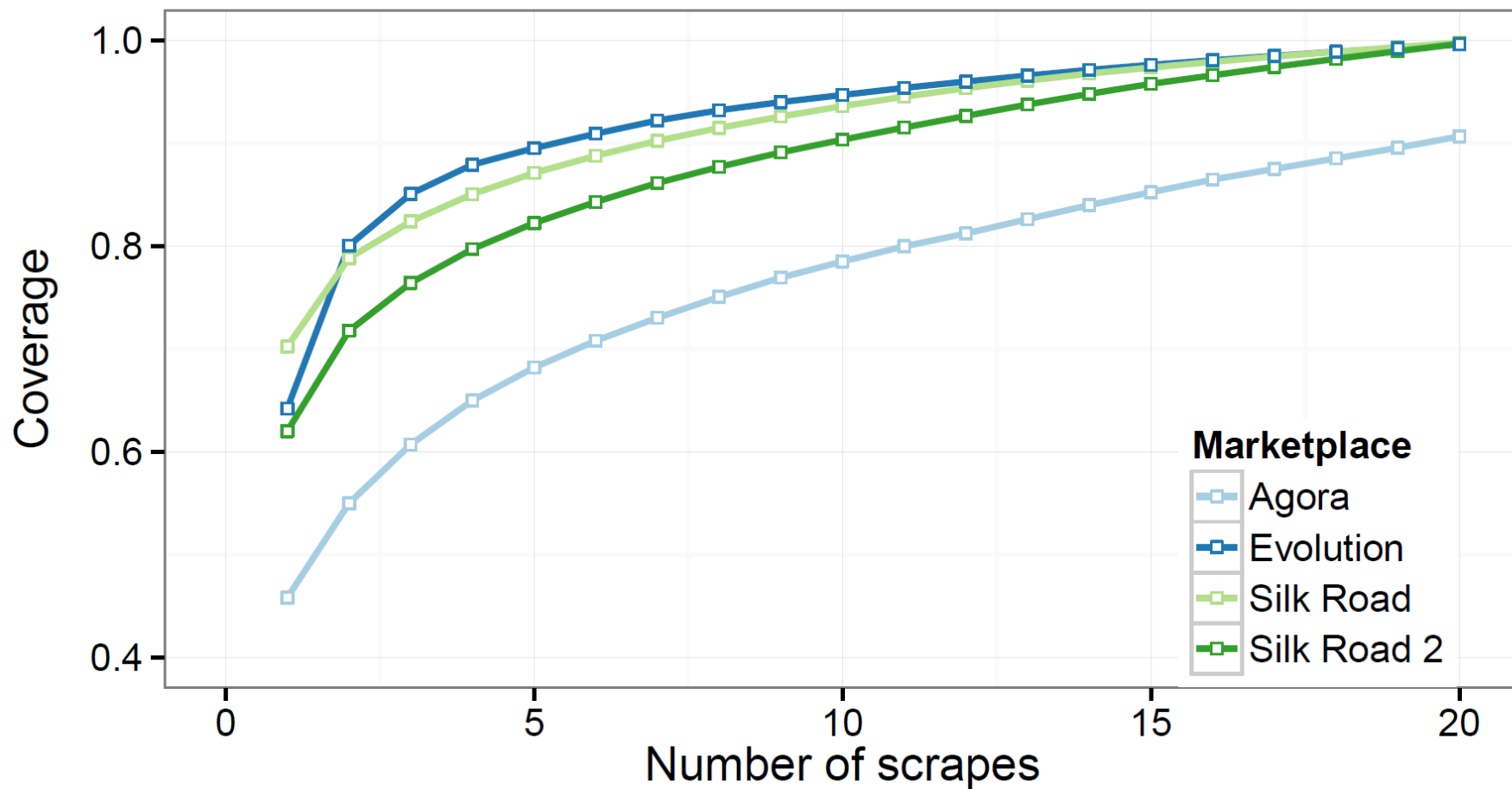
Analysis

- **Assumption:** Each feedback corresponds to precisely one transaction

  - Anonymity requires strictly enforced feedback system to establish reputation

  - Possible on many marketplaces to purchase several quantities of item and leave 1 feedback, conservative estimate

Alternative Transaction Proxies

- **Counting # Item Listings**
  - Very efficient and convenient
  - Assumes that there exists some stable ratio between transaction volume and # listings

  - Daily $\frac{volume}{\#\ Listings}$ for The Evolution Marketplace in July 2014 and September 2014 differ by factor of 4

Uniqueness

- **Problem:**
  - 100s of observations of same feedback
  - Double counting leads to over-estimations
  - Feedback may be updated, deleted

- **Solution:**
  - Automatically detect updated feedbacks
    - Only keep most recent version
  - Hash {timestamp, title, vendor, message, rating}

Holding Prices

- Feedbacks are useful to vendors but are destroyed when the listing is removed

- Vendors raise listing prices prohibitively high

$0.02 -> $1,000.00

$1,100.00 -> $1,000,000.00

- Need to look at historical price for item

31

Holding Prices

- Heuristic A:
  - Remove all free things
  - Remove all things > $100,000
  - Calculate median of remaining prices
  - Remove everything greater than 5x median
  - Remove things less than 25% of median
- Heuristic B:
  - Remove all things > $100,000
  - Remove upper quartile
  - Remove everything greater than 100x cheapest non-zero price

- Evaluation
  - Coefficient of Variation $c_v = \dfrac{\sigma}{\mu}$

Holding Prices CDF
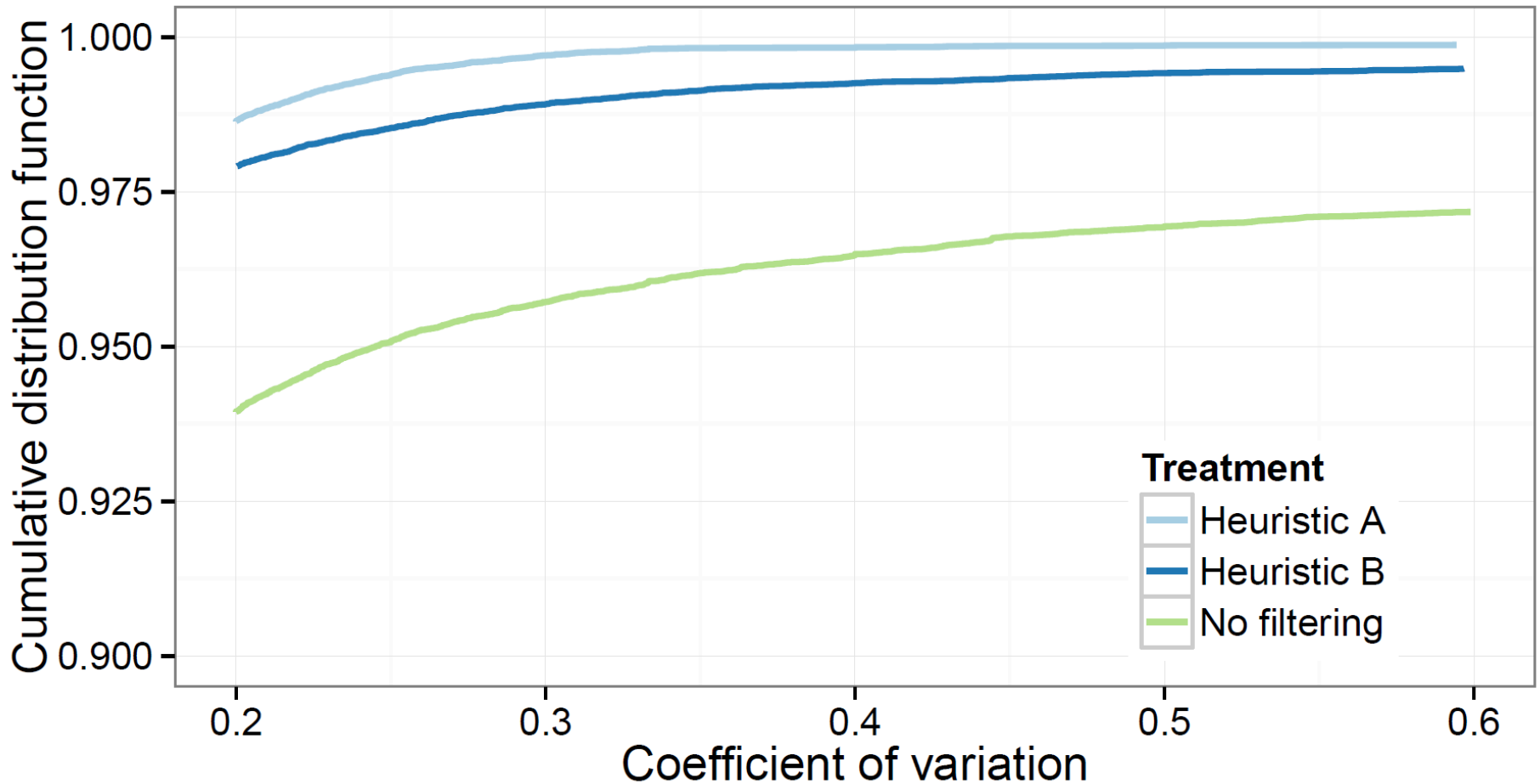
Sales Volume
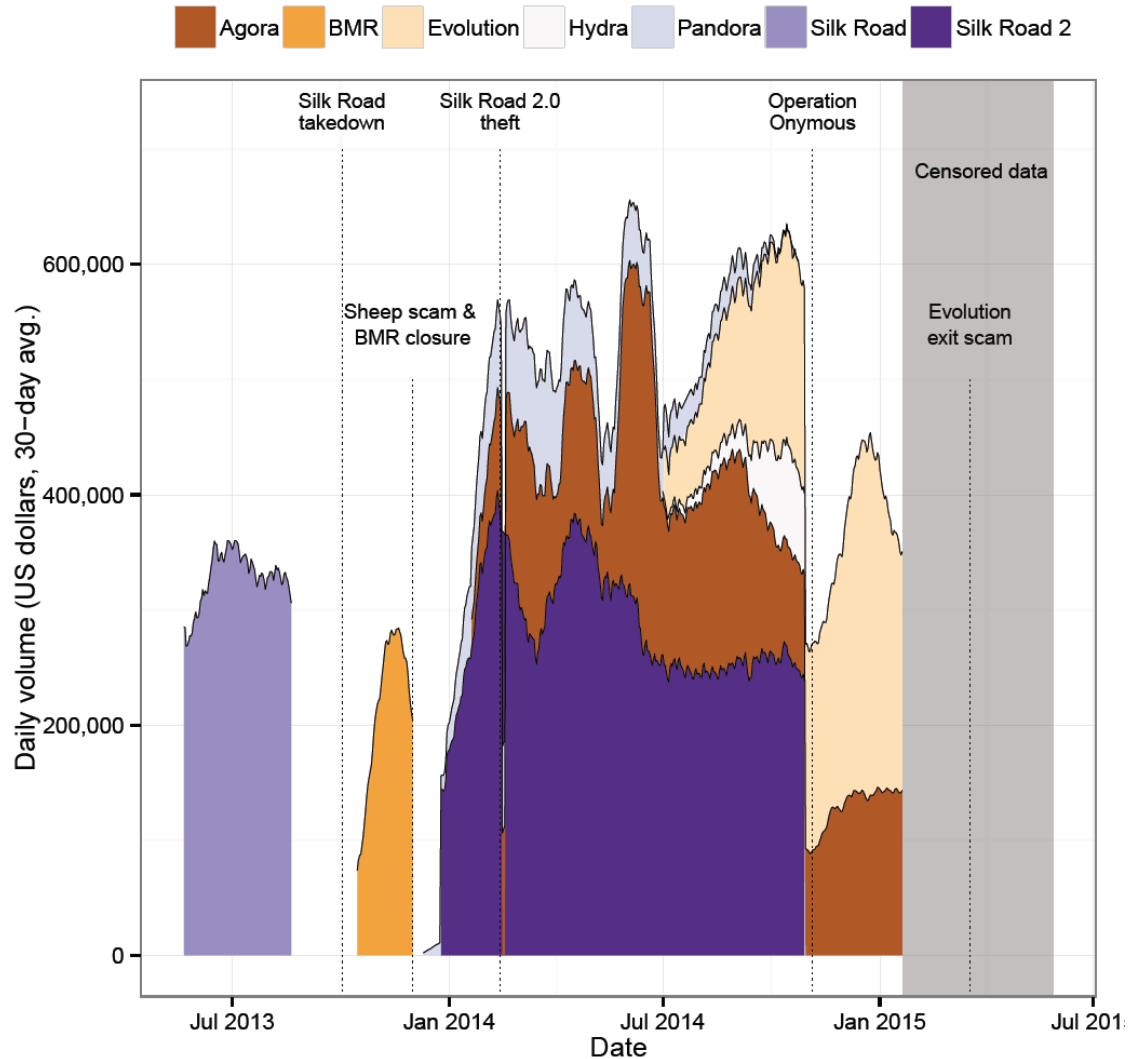
Product Categories

- **What is being sold?**
  - Product labels are often unavailable or inaccurate
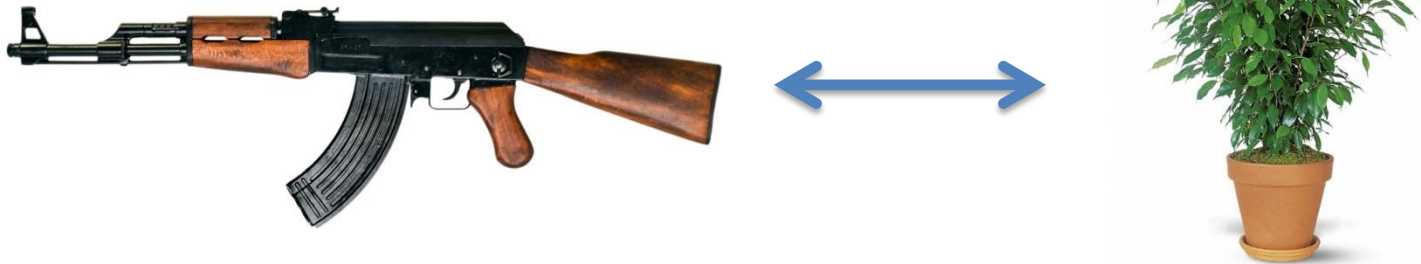
- Classifier trained from Agora and The Evolution Marketplace
  - Listing title and description concatenated and tfidf
  - 1,941,538 unique samples, 162,198 words tokenized
  - Predicts 16 class labels

Confusion Matrix

	BNZ	DG	DIS	ELEC	MISC	OP	PAR	PSY	RX	SL	STI	STR	THC	TOB	WPN	X
BNZ	0.98	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0
DG	0	0.96	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
DIS	0	0	0.99	0	0	0	0	0	0	0	0	0	0	0	0	0
ELEC	0	0.03	0	0.92	0.02	0	0.02	0	0	0	0	0	0	0	0.01	0
MISC	0	0.2	0	0	0.8	0	0.01	0	0	0	0	0	0	0	0	0
OP	0	0	0	0	0	0.98	0	0	0.01	0	0	0	0	0	0	0
PAR	0	0	0	0	0	0	0.99	0	0	0	0	0	0	0	0	0
PSY	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
RX	0.03	0	0	0	0	0.02	0	0	0.93	0	0.02	0	0	0	0	0
SL	0	0	0	0	0	0	0	0	0.02	0.98	0	0	0	0	0	0
STI	0	0	0	0	0	0	0	0	0	0	0.99	0	0	0	0	0
STR	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
THC	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
TOB	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
WPN	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0.98	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99

Item Sales Per Category

Vendor Volumes CDF
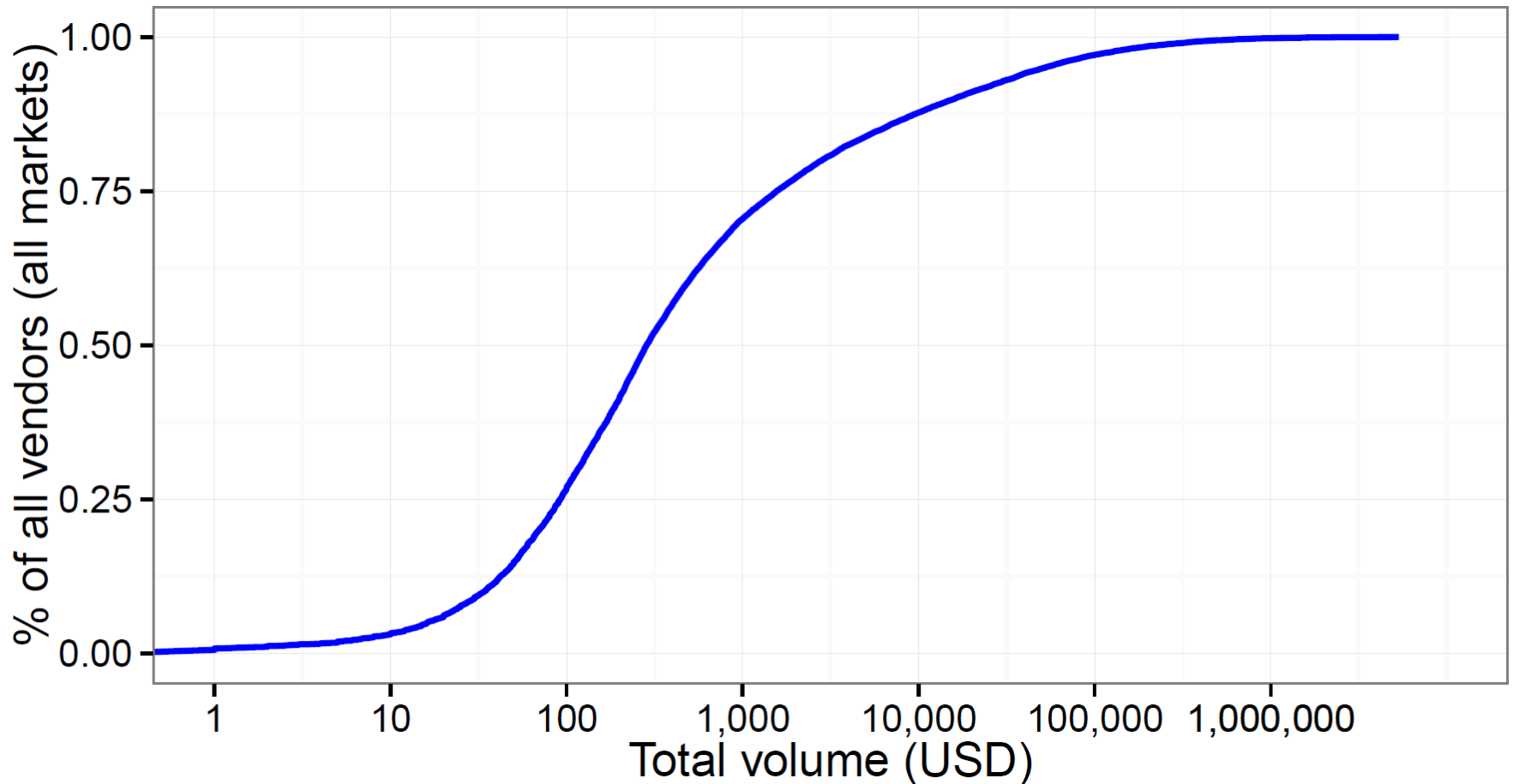
Vendor Diversity

- **Do vendors specialize in what they are selling?**
  - Do vendors sell what they make?
  - Does a single online presence sell goods for several diversified suppliers?

- **Coefficient of Diversity $\in [0, 1]$**
  - 0 – all sales from same category
  - 1 – equal sales from each category
  - Only vendors > $10,000 total sales considered

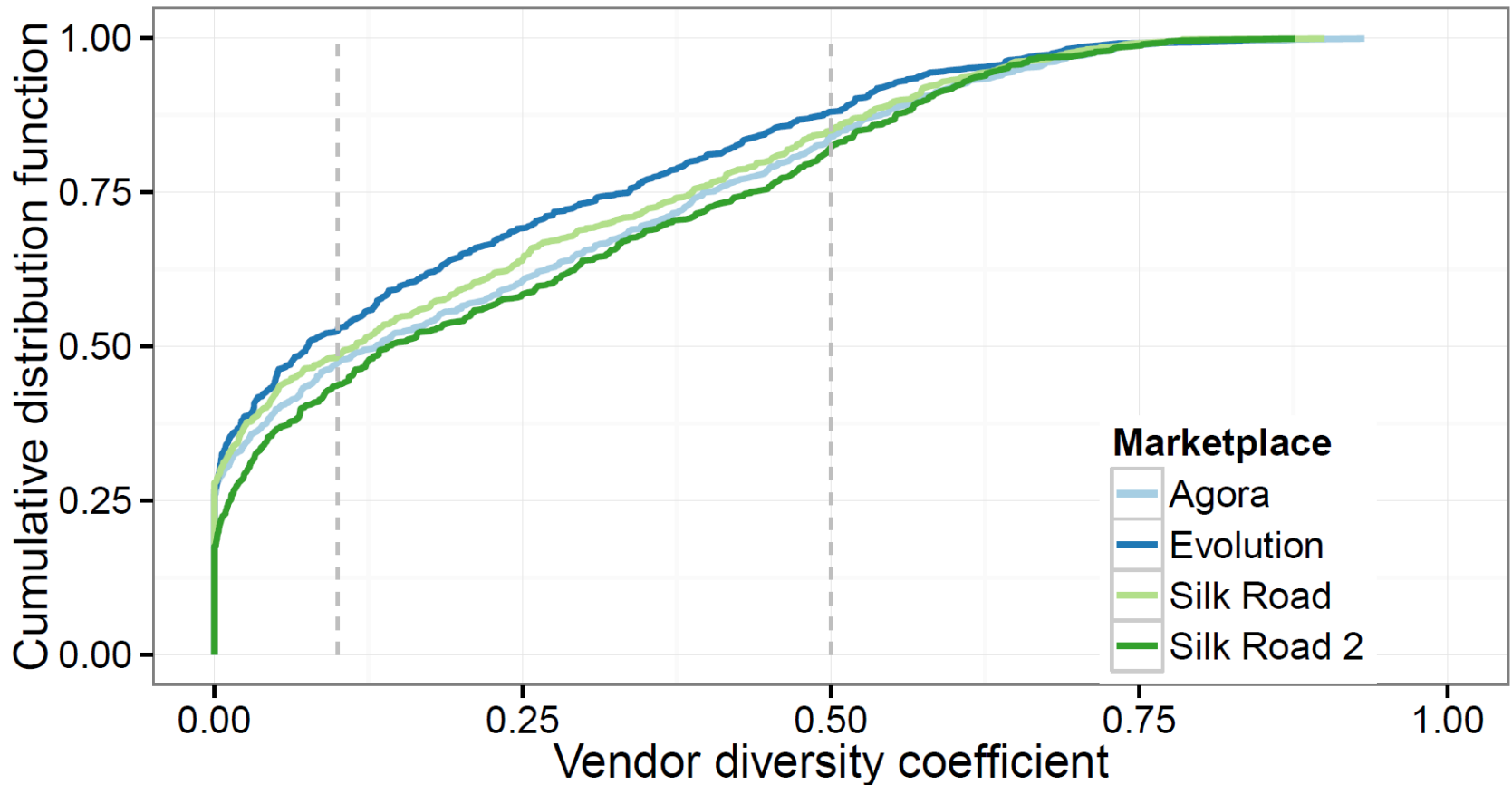# Vendor Diversity CDF

Validation

- Trial evidence GX226A, GX227C places Silk Road 1 weekly volumes at $475,000/week in late March 2012, consistent with our estimates

- Administrator reports Silk Road 2 daily volumes of around $250,000 in September 2014, similar to our estimated $270,000

- Leaked Agora vendor page shows sales total on June 5, 2014 to be $3,460, our observations yielded $3,408

Takeaways

- Anonymous Marketplaces are very easy to setup and use and have wide customer appeal

- Anonymous Marketplace ecosystem transacts in excess of $500,000 / day

- Anonymous Marketplaces are primarily used (~75%) for recreational drugs

- Anonymous Marketplace ecosystem has historically recovered from takedown efforts and scams

- Anonymous Marketplaces are controlled by small set of highly influential vendors

Kyle Soska – ksoska@cmu.edu

Data Completeness - Schnabel Estimator

- F true feedbacks at time t

- n observations

- C_i feedbacks in observation i

- M_i feedbacks in observation i previously seen

- R_i total previously observed feedbacks

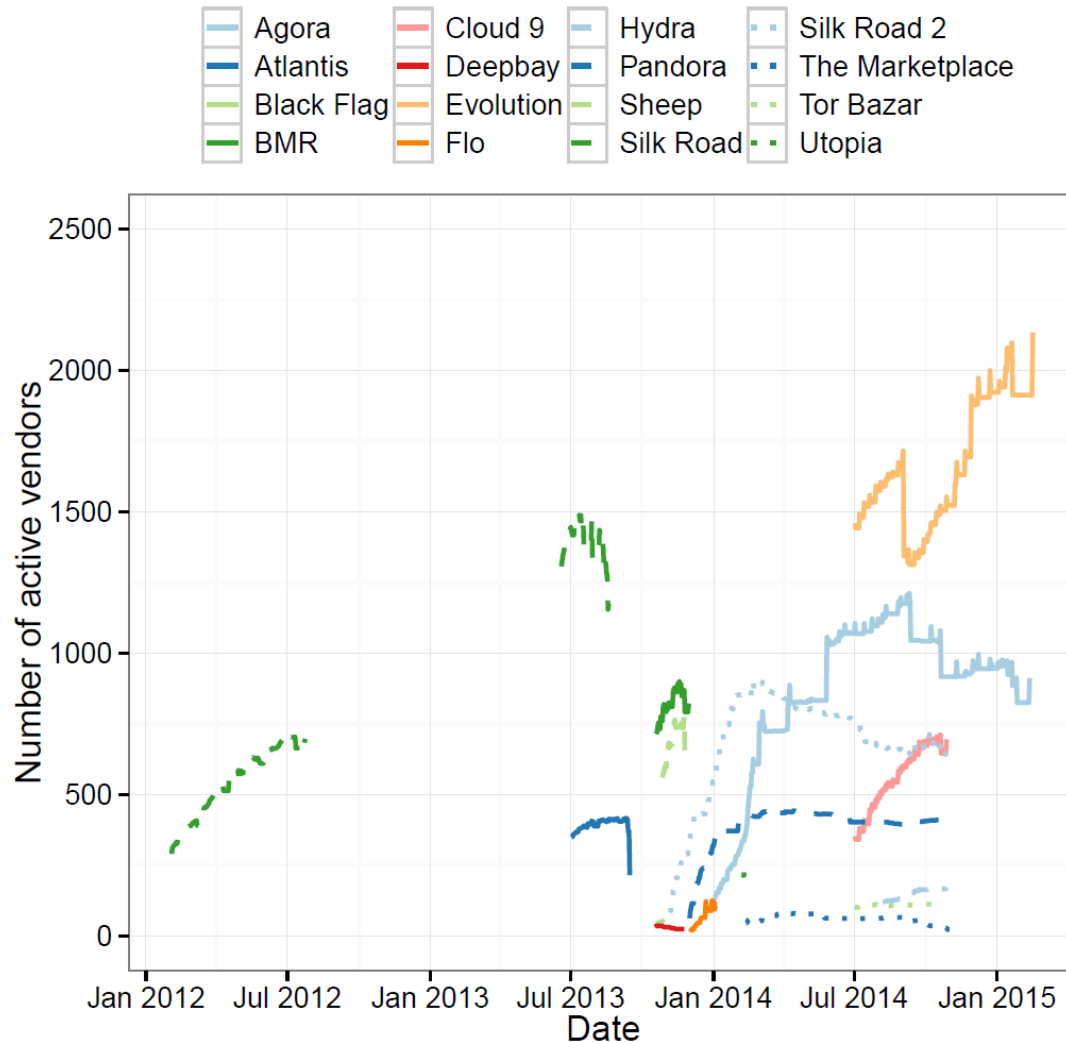- Estimate $\hat{F} = \dfrac{\sum_{t=1}^{n} C_t M_t}{\sum_{t=1}^{n} R_t}$

Vendor Diversity

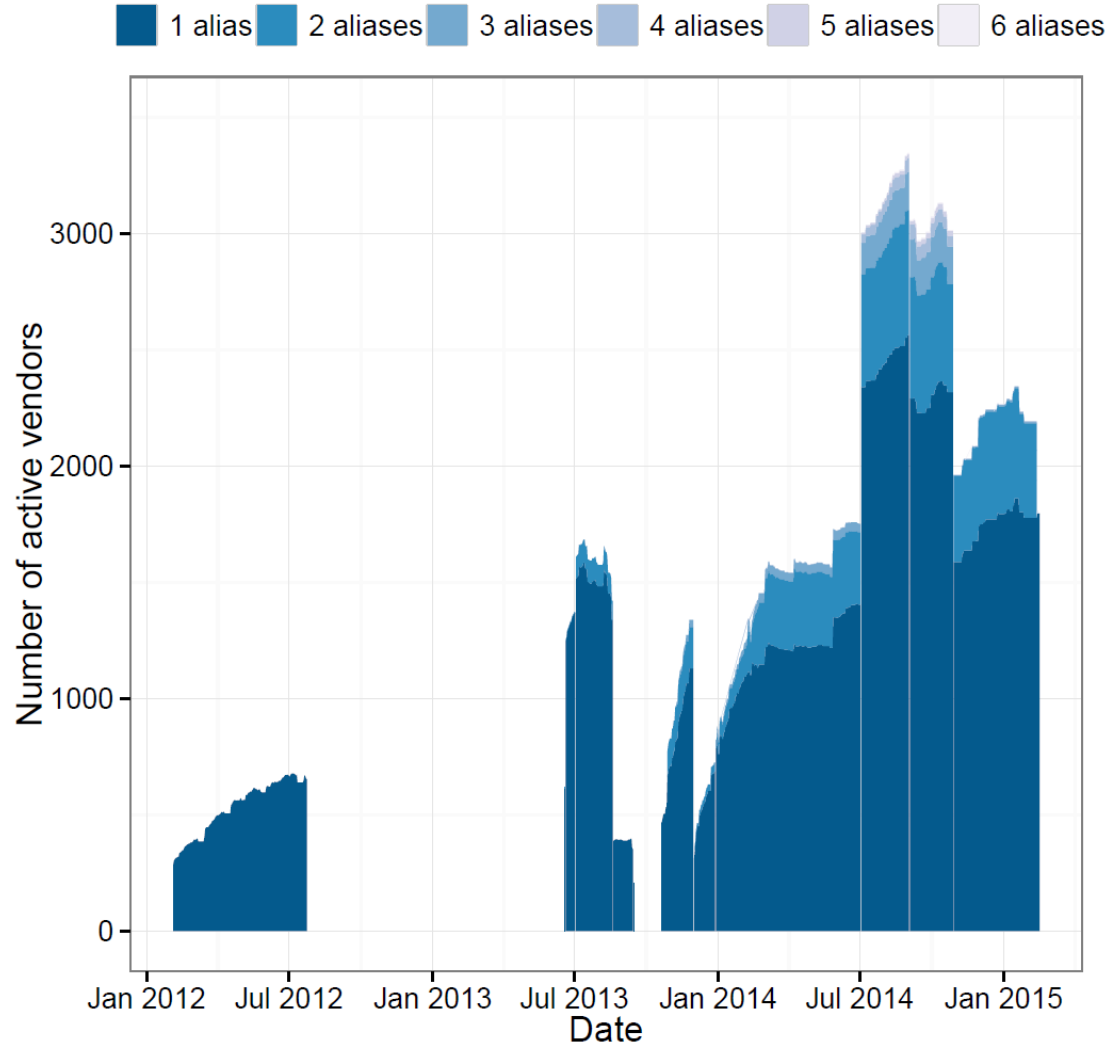- $C_{s_{j,i}}$ = % of vendor j's total sales that came from category i

- Coefficient of Diversity =
$$c_d = \left(1 - \max\left(C_{s_j}\right)\right) \frac{\left|C_{s_j}\right|}{\left|C_{s_j}\right| - 1}$$

Active Sellers Over Time

Aliases Per Sender

PGP Deployment