MEASURING AND FINGERPRINTING CLICK-SPAM IN AD NETWORKS

Vacha Dave *, Saikat Guha★ and Yin Zhang *

* The University of Texas at Austin

★ Microsoft Research India

Internet Advertising Today

- Online advertising is a 31 billion dollar industry *
- Publishers can monetize traffic
  - Blogs, News sites, Syndicated search engines
  - Revenue for content development
- Pay-per-click advertising
  - Advertisers pay per-click to ad networks
  - Publishers make a 70% cut on each click on their site

*Based on Interactive Advertising Bureau Report, a consortium of Online Ad Networks

Click-spam in Ad Networks

- **Click-spam**
  - Fraudulent or invalid clicks
  - Users delivered to the advertiser site are uninterested
  - Advertisers lose money
- Possible Motives
  - Malicious advertisers (or other parties)
    - Deplete competitor's ad budgets
    - Isolated cases
  - Publishers/Syndicated search engines
    - Make money on every click that happens on their site

Click-spam in Ad Networks

Microsoft, which offers pay-per-click ads through its adCenter service, says click laundering -- an offshoot of click fraud, which has plagued the industry for years -- is growing in scale and sophistication. "This is the newest form of criminal activity on the Internet," says Brad Smith, Microsoft's general counsel.

How Click Laundering Works

See a graphic provided by Microsoft on how click laundering works.

In the RedOrbit case, Microsoft says it discovered the alleged scheme after detecting a growing number of suspicious clicks from RedOrbit's site over a two-week period starting in January 2009. The site had previously averaged 75 clicks a day, but the number spiked above 10,000 clicks per day, according to the complaint.

According to Microsoft, RedOrbit was able to manufacture large numbers of bogus clicks on shady sites known as parked domains—Web sites that are legal but exist only to display ads. In this case, many of the ads were invisible to the naked eye. Then, using a technical slight of hand, it submitted the clicks to Microsoft in a way that made them appear to have occurred on RedOrbit's own site, a requirement for getting paid.

Mobile Devices and Ads

Ant Smasher

- Mobile game
- Squish the ant to win the game
- Ads placed close to where user is expected to click

Click-spam Detection

- No ground truth
  - Almost impossible to know if particular click is genuine
  - Need to guess the intent of user
- Different levels of click-spam in different segments
  - Aggregate numbers are meaningless
- Ad networks aren't transparent
  - Security by obscurity
- Real problem – lot of work needed
  - Researchers lack real attack data

Contributions

- First method to independently estimate click-spam
  - As an advertiser
  - For specific keywords
- Test across ten ad networks
  - Search, contextual, social and mobile ad networks
  - Show that click-spam is a problem
    - For Mobile and Social ad networks
- Discover five classes of sophisticated attacks
  - Why simple heuristics don't work
- Release data for researchers

Estimating click-spam – Approach

- Hard to classify any single click
  - Estimate fraction of click-spam
- Designed Bayesian estimation framework
  - Uses only advertiser-measurable quantities
- Cancel out unmeasurable quantities
  - By relating different mixes of good and bad traffic

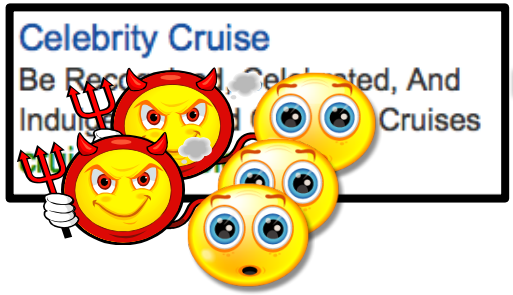
Estimating Click-spam – Main Idea

Both non-spammers and spammers click ads

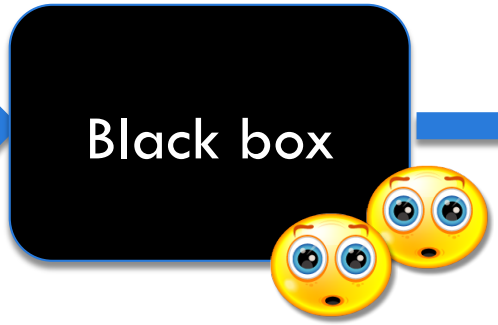
A fraction of non-spammers buy

How many  ?

Equate ratios of buyers to non-spammers



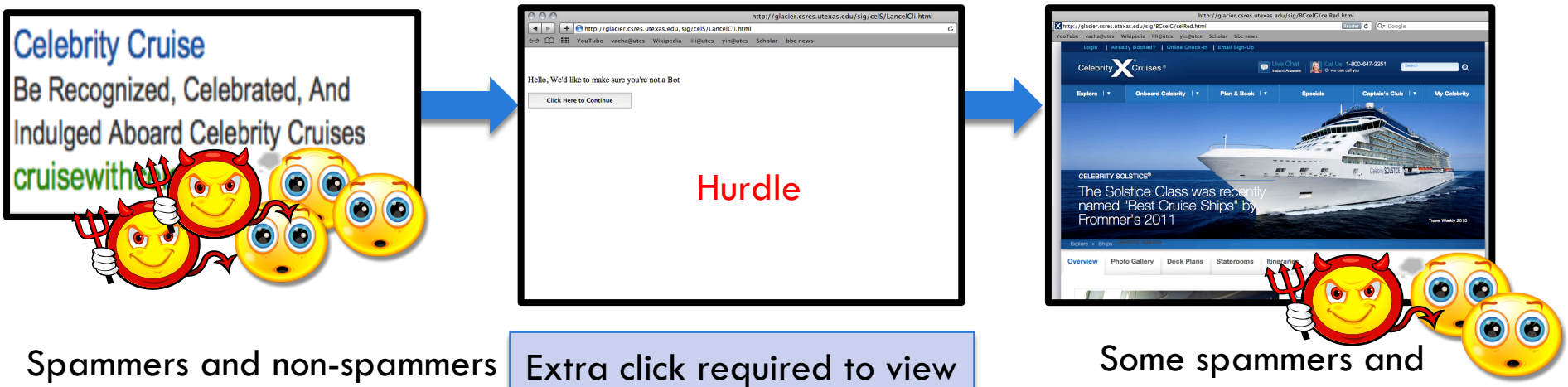
Both non-spammers and spammers click ads



Lose spammers and some non-spammers

Some non-spammers buy

Dissecting Black box – Hurdles



Spammers and non-spammers click on an ad

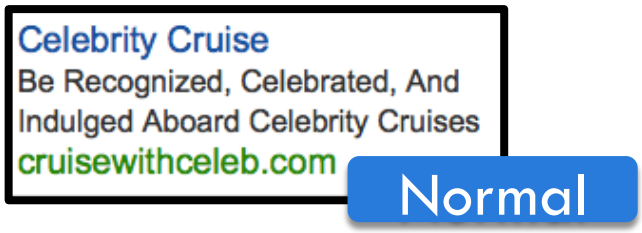
Extra click required to view site

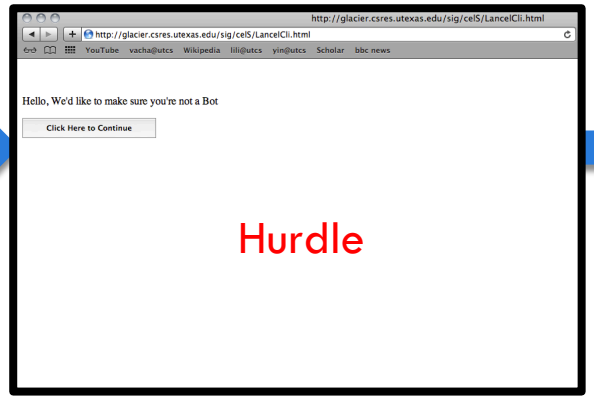
Some spammers and Non-spammers see the content

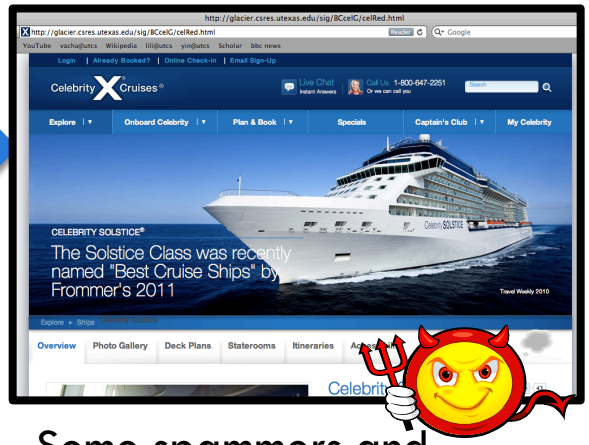
- Different hurdles have different hardness
  - 5 sec wait, Click to continue
- Send only a fraction of traffic through hurdles
  - To minimize impact on user experience
- Perfect hurdle would block all spam
  - In reality, some spammers get through (False Negatives)

Dissecting Black box - Bluff Ads[1]

- Bluff Ads
  - Junk ad text with normal keywords, same targeting
  - Normal users unlikely to click







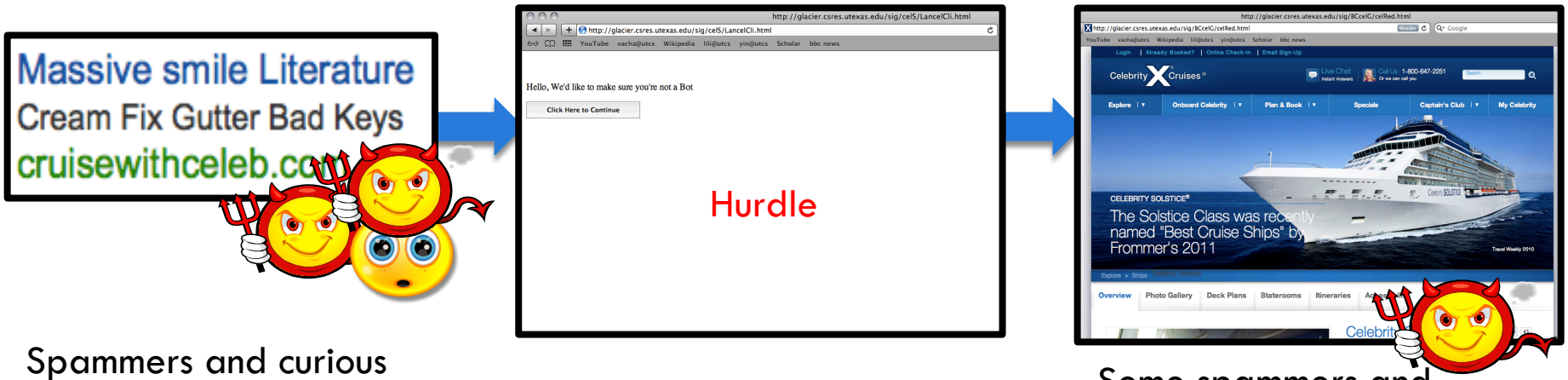
Spammers and curious users click on an ad

Some spammers and users may see the content

[1] Fighting online click fraud using bluff ads [CCR 2010]

Dissecting Black box - Bluff Ads[1]

- Maximum False Negative rate known for each hurdle
- Can be subtracted out



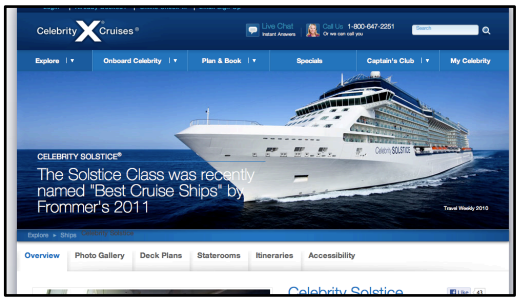
Spammers and curious users click on an ad

Some spammers and users may see the content

[1] Fighting online click fraud using bluff ads [CCR 2010]

Testing Ad Networks

- Sign up as advertisers for ten ad networks
  - Search, Contextual, Mobile and Social
  - Google, Bing, AdMob, InMobi, Facebook and others
- 240 Ads
  - Keywords: Celebrity, Yoga, Lawnmower
  - Hurdles: Click to continue, 5 sec wait
- 50,000 Clicks
  - 30,000 bluff ad clicks
- Cost: $1500

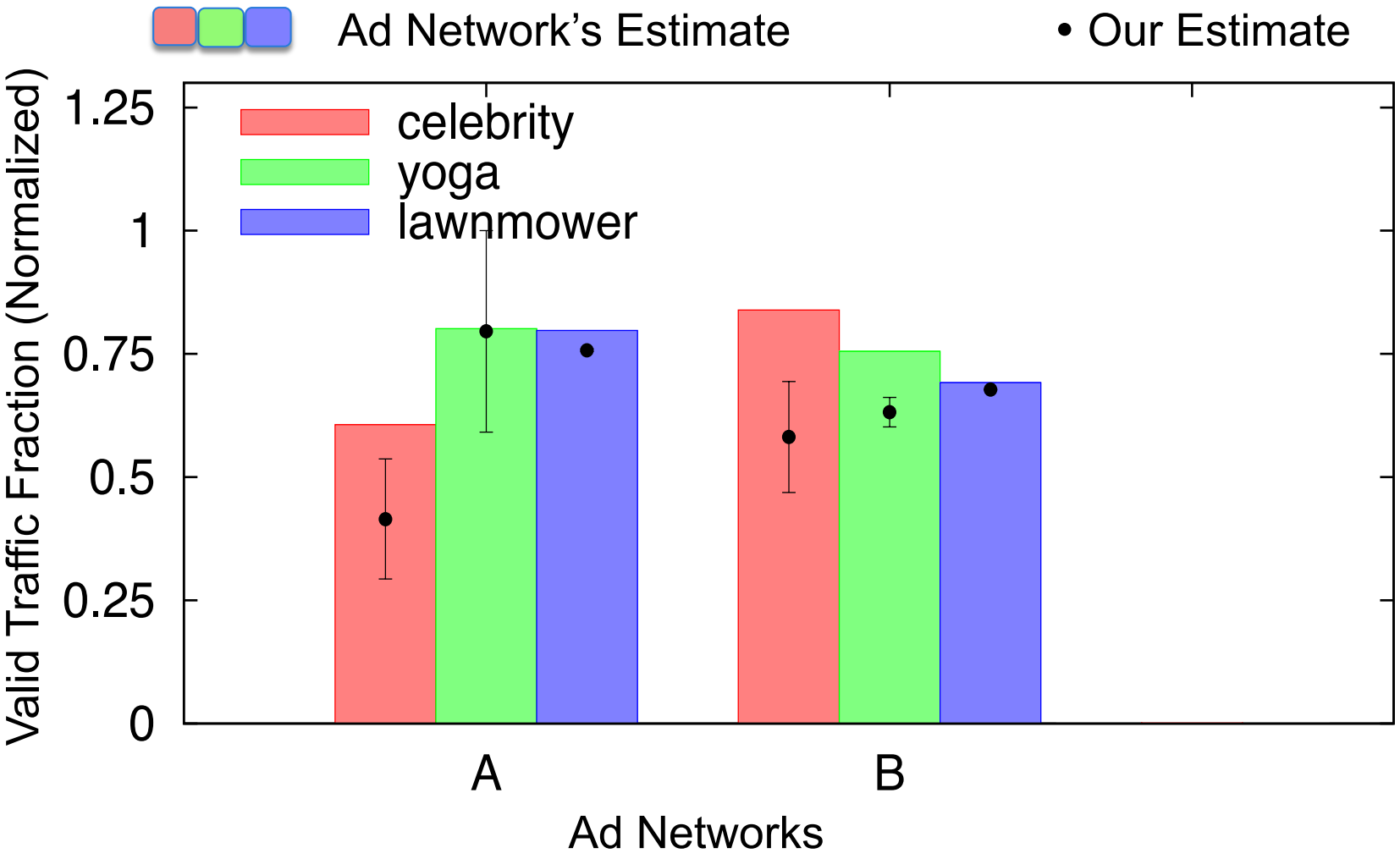


Uh-oh. How do we validate?

No ground truth!

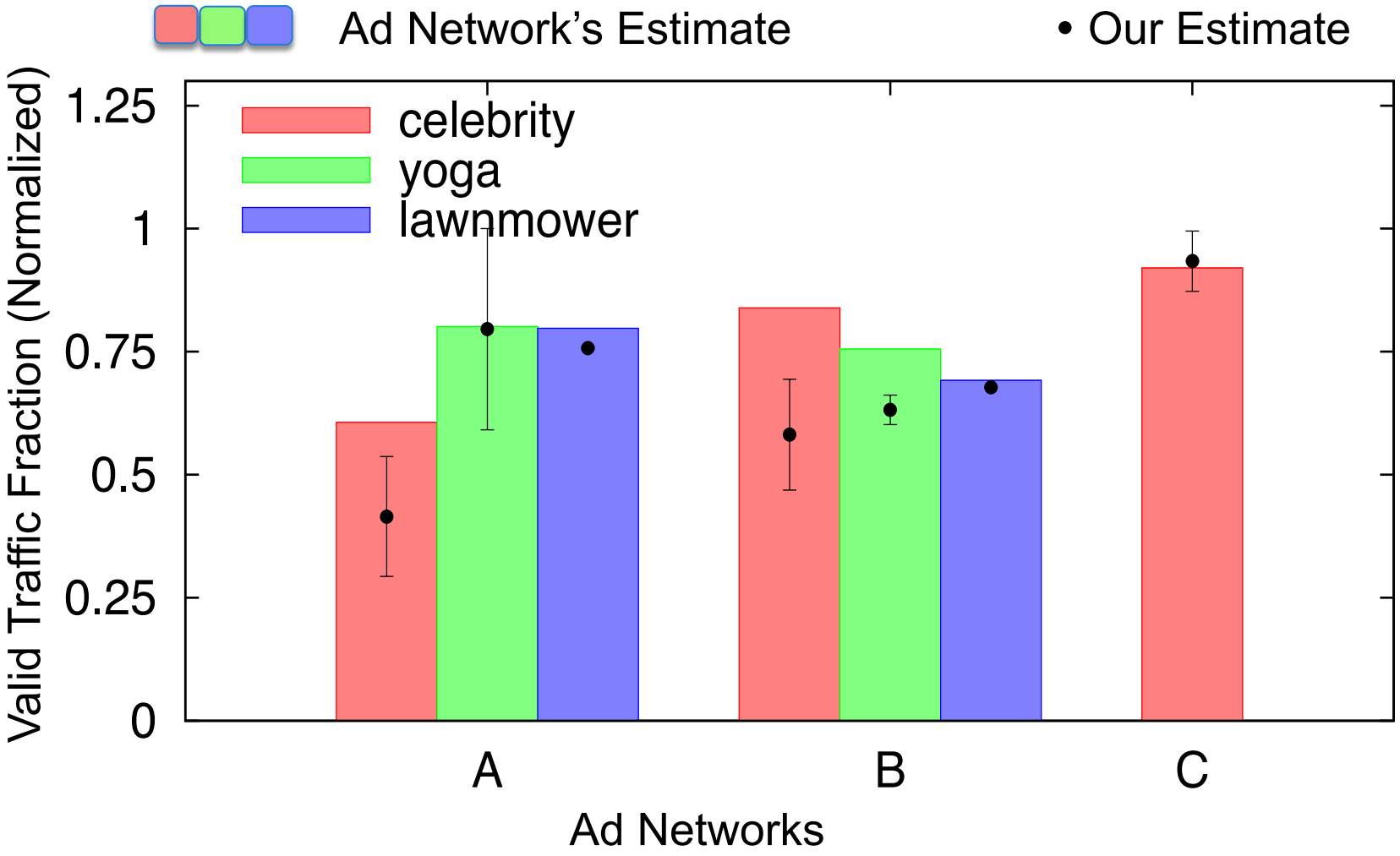
Compare against search ads on Google and Bing

Results – Validation using search ads



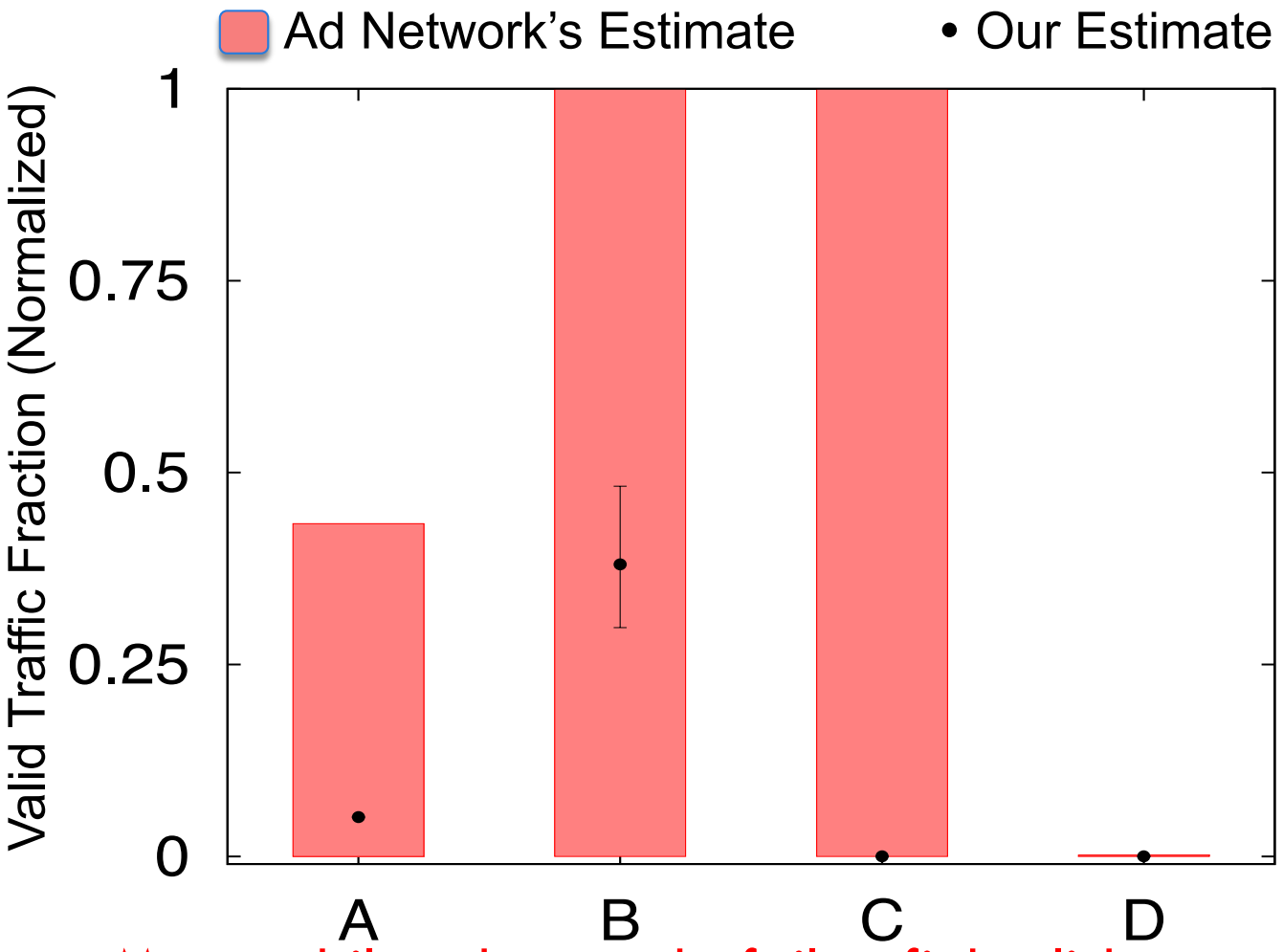
Clicks charged are close to the estimated valid clicks

Results – Validation using search ads



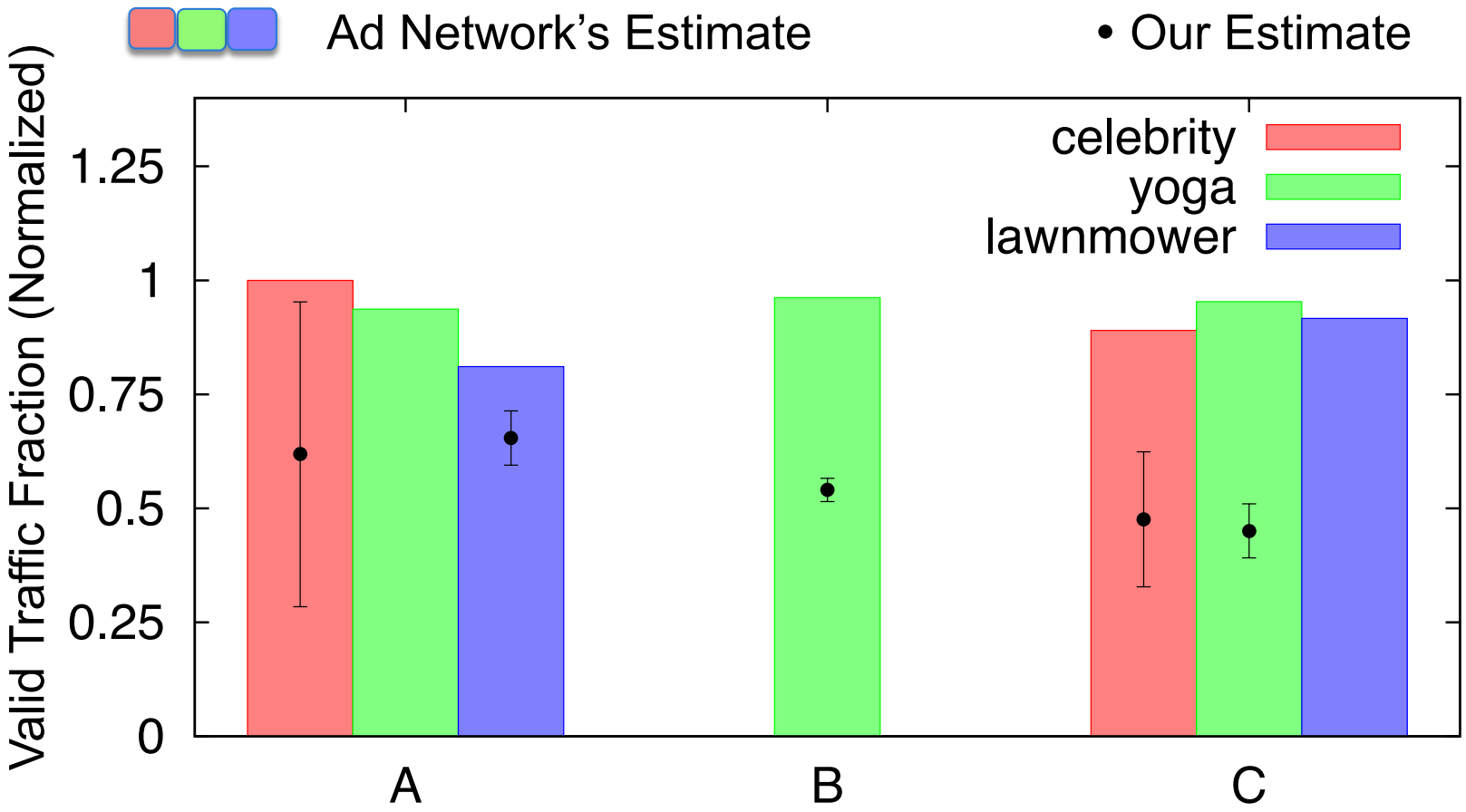
Clicks charged are close to the estimated valid clicks

Results – Estimating Mobile Spam



Most mobile ad networks fail to fight click-spam

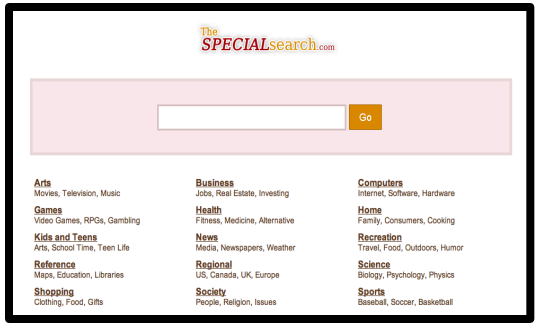
Results – Estimating Contextual Spam

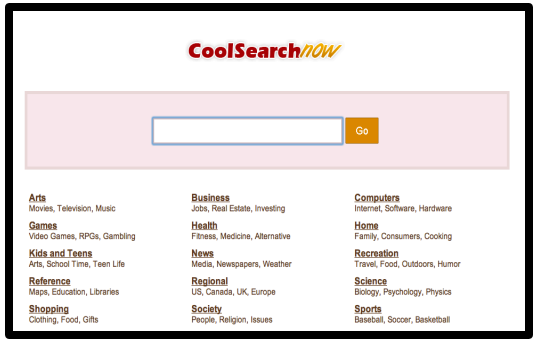


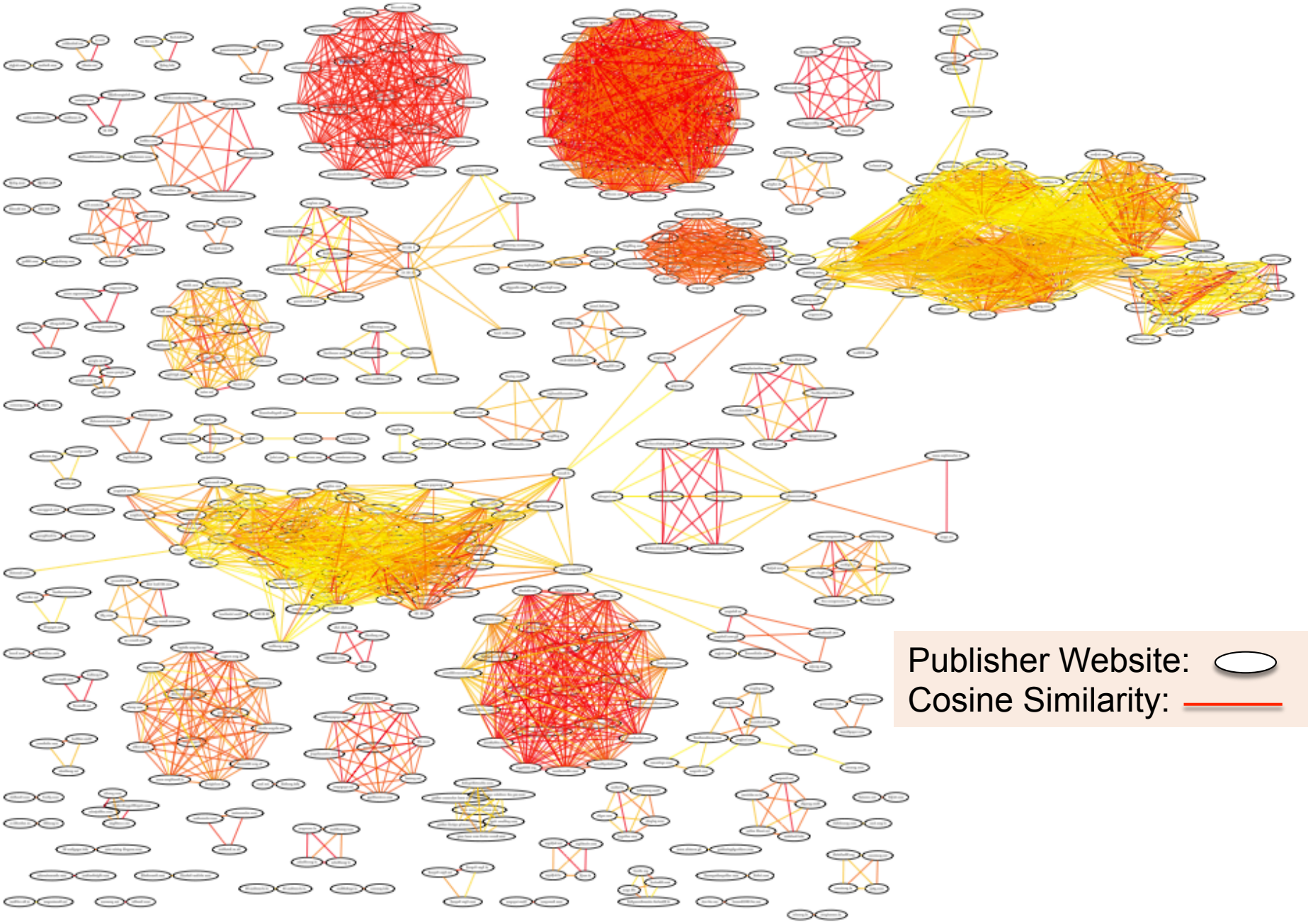
All networks seem to be underestimating the amount of spam

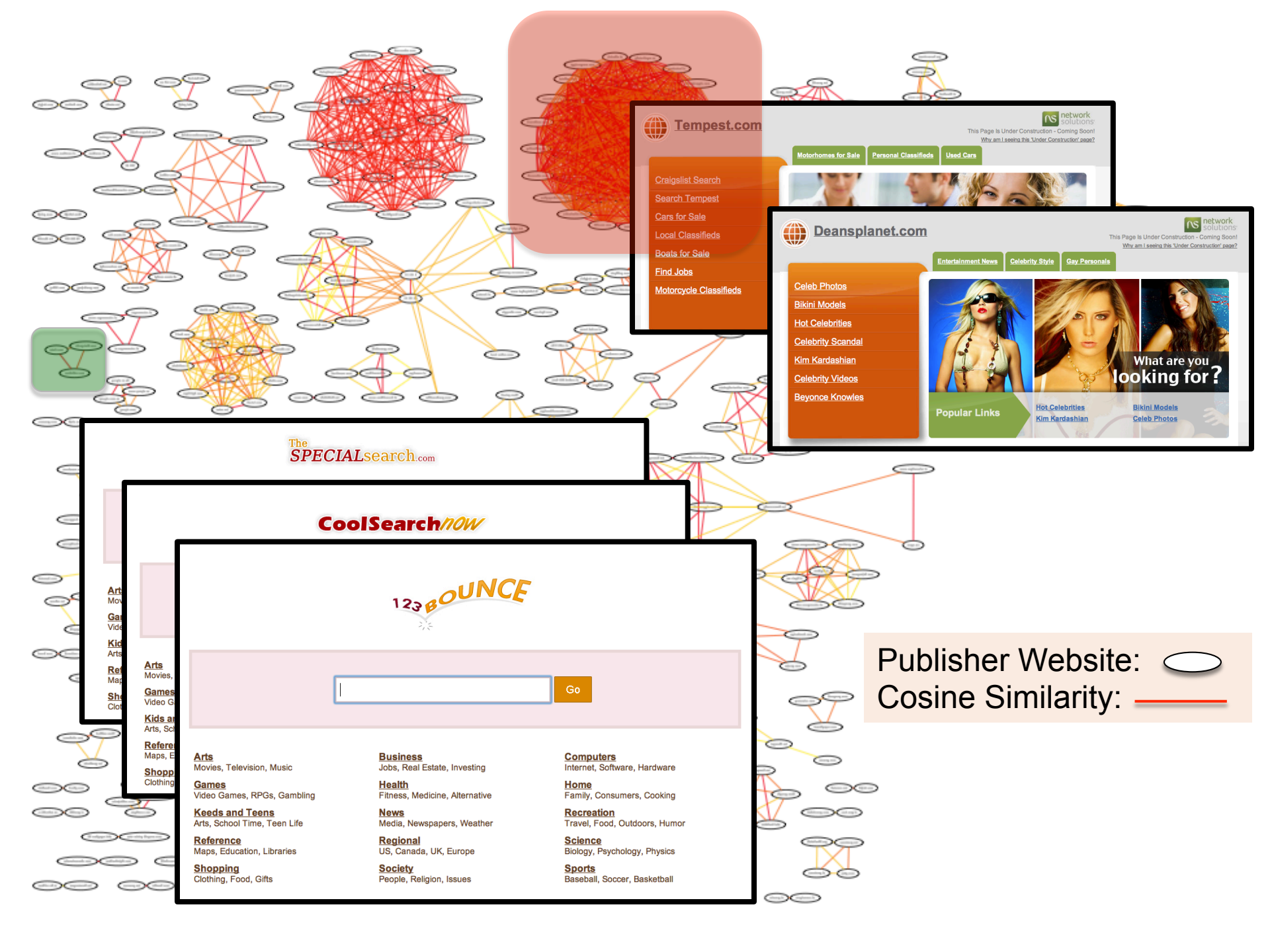
Where is click-spam coming from?

- Analyze bluff ad clicks
  - Publishers: Strong motive
    - Instead of clicks/users
  - Manual Investigation
- Challenge: Scale
  - 3000+ publishers, 30,000 Clicks
- Identical sites!
- Cluster on cosine similarity
  - Feature vector
    - WHOIS , IP Address/Subnet, HTTP parameters






Publisher Website:
Cosine Similarity:


Tempest.com
Craigslist Search
Search Tempest
Cars for Sale
Local Classifieds
Boats for Sale
Find Jobs
Motorcycle Classifieds
Deansplanet.com
Celeb Photos
Bikini Models
Hot Celebrities
Celebrity Scandal
Kim Kardashian
Celebrity Videos
Beyonce Knowles
What are you looking for?
Popular Links
The SPECIALsearch.com
CoolSearchNOW
123BOUNCE
Go
Arts
Movies, Television, Music
Business
Jobs, Real Estate, Investing
Computers
Internet, Software, Hardware
Games
Video Games, RPGs, Gambling
Health
Fitness, Medicine, Alternative
Home
Family, Consumers, Cooking
Keeds and Teens
Arts, School Time, Teen Life
News
Media, Newspapers, Weather
Recreation
Travel, Food, Outdoors, Humor
Reference
Maps, Education, Libraries
Regional
US, Canada, UK, Europe
Science
Biology, Psychology, Physics
Shopping
Clothing, Food, Gifts
Society
People, Religion, Issues
Sports
Baseball, Soccer, Basketball
Publisher Website:
Cosine Similarity:

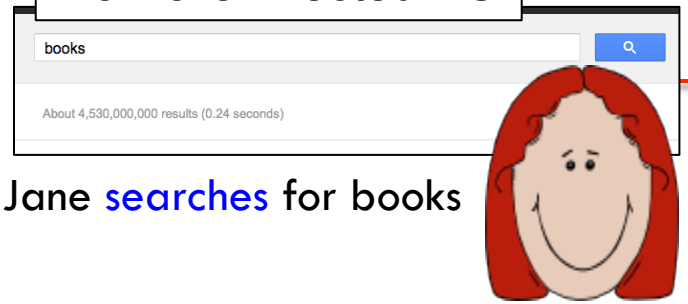
Publishers creating sybil nodes to beat ad network thresholds

Publisher Website:

Cosine Similarity:

Case Study 1 - Malware driven click fraud

Malware infected PC



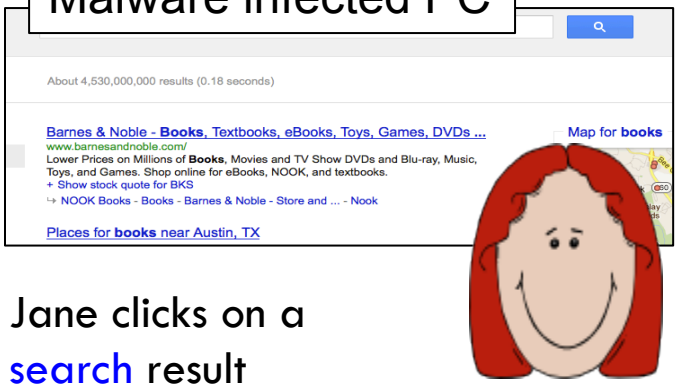
Jane searches for books

(BOTID=50018&SEARCH-ENGINE-NAME&q=books) Base64

Botmaster generates list of publishers

Publisher List

Malware infected PC



Jane clicks on a search result

Publisher URL

www.moo.com

Auto-Redirect (Fraud)

AD URL

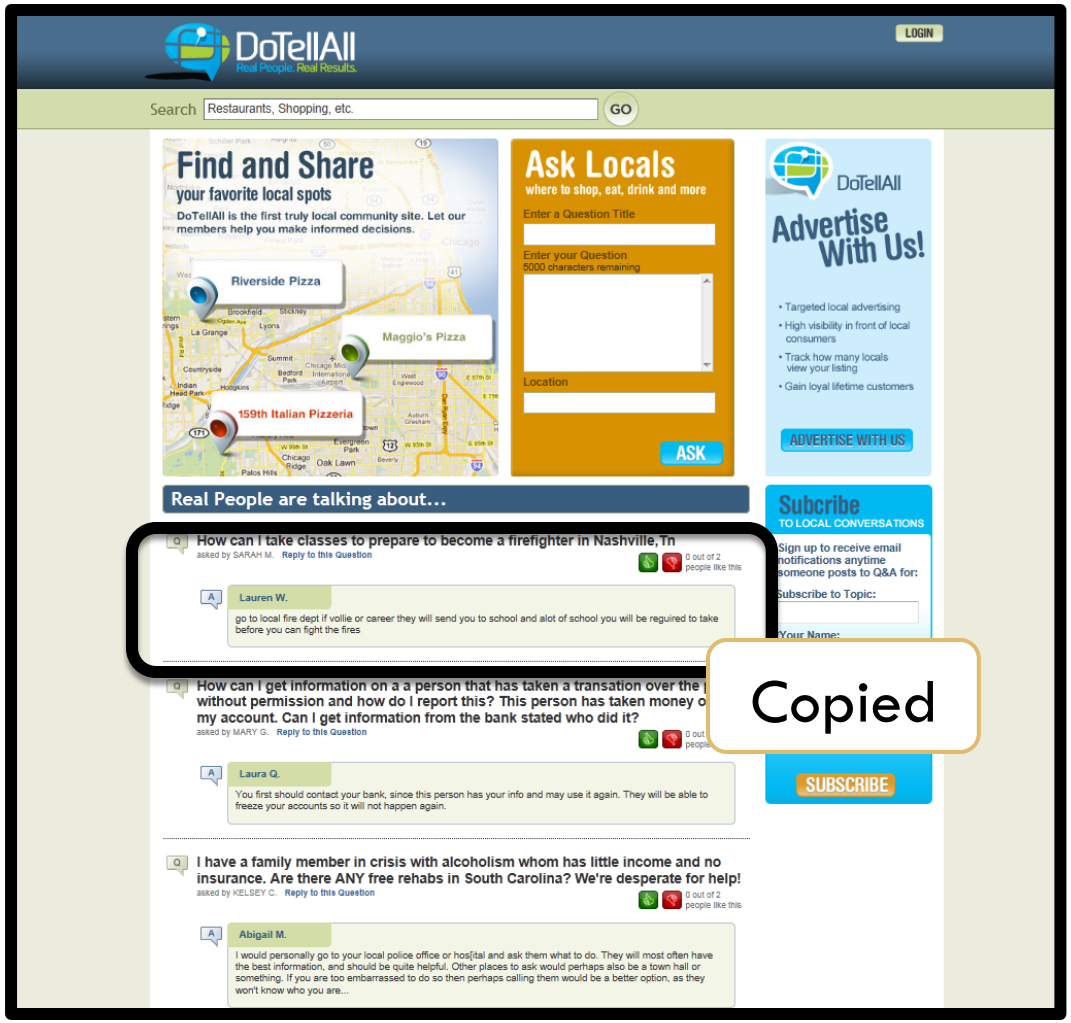
All background traffic – Jane sees nothing

Case Study 1 - Malware driven Click fraud

- Responsible Malware: TDL4
  - Validation: Run malware in VM
- Can intercept and redirect all browser requests
  - Browser specific filtering doesn't work
- Only 1 click per IP address per day
  - Threshold based filtering doesn't work
- Mimics real user behavior
  - Timing analysis doesn't work

ClickSpam and Arbitrage

- Polished forum sites
- Bluff ad clicks on ad network X
- No malware reports
- Not popular
  - Where do they get traffic?
- No ads on the site !!



Click-spam and Arbitrage

Work From Home-Now Hiring
3 Positions Available - $17-21/Hour Based On Experience. Apply Now!
DoTellAll.com/Apply

Site pays $ to Y

- Advertiser on network Y
  - Creates 4500+ ads
- Publisher on network X
- Page now has only ads
  - No questions or answers
- Confusing users into clicks

Click-spam and Arbitrage

Work From Home-Now Hiring
3 Positions Available - $17-21/Hour Based On Experience. Apply Now!
DoTellAll.com/Apply

Site pays $ to Y

- Tricking real users into clicking
  - Bot detection techniques don't apply

Case Study3 - Click Fraud using Parked Domains

Go to icicibank.com

Jane mistypes icicbank.com in her browser and presses enter

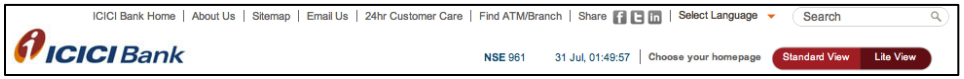
Parked Domain

Auto-Redirect

Pull ads for "icicbank" from a Syndicated Search Engine

Auto-Redirect (Fraud)

AD URL



Jane ends up on icicibank.com

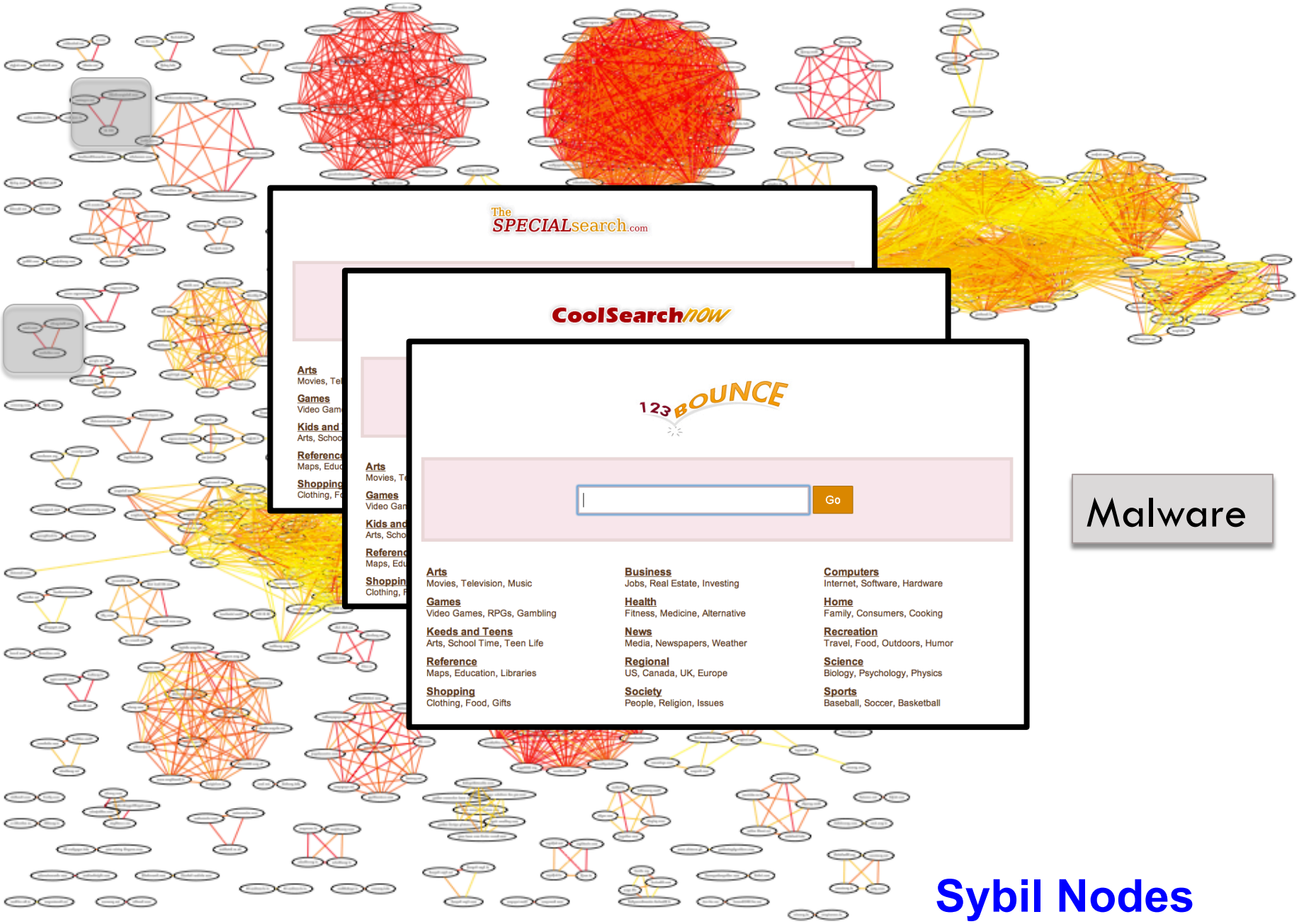
icicibank.com pays for a click

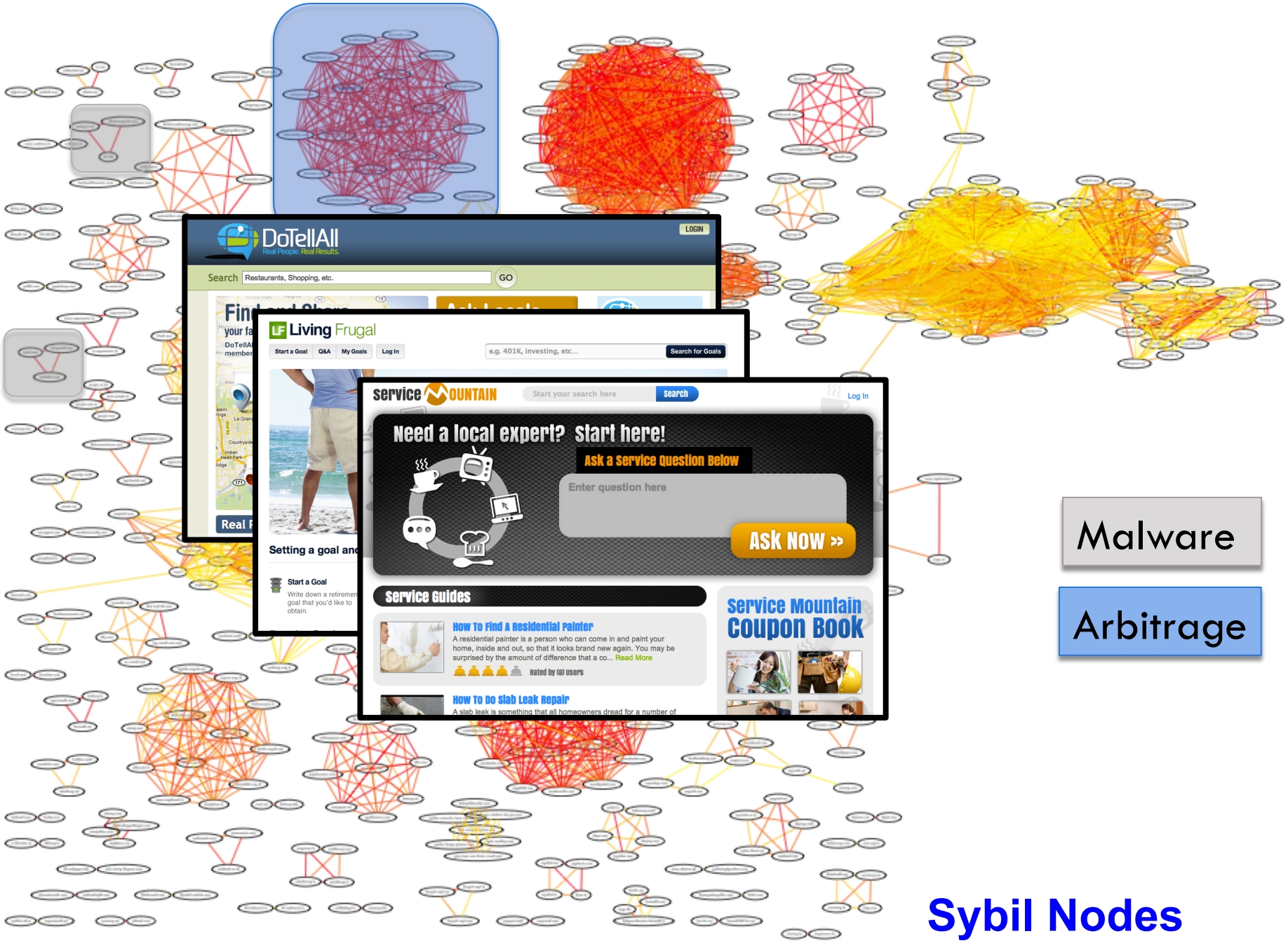
Case Study3 - Click Fraud using Parked Domains

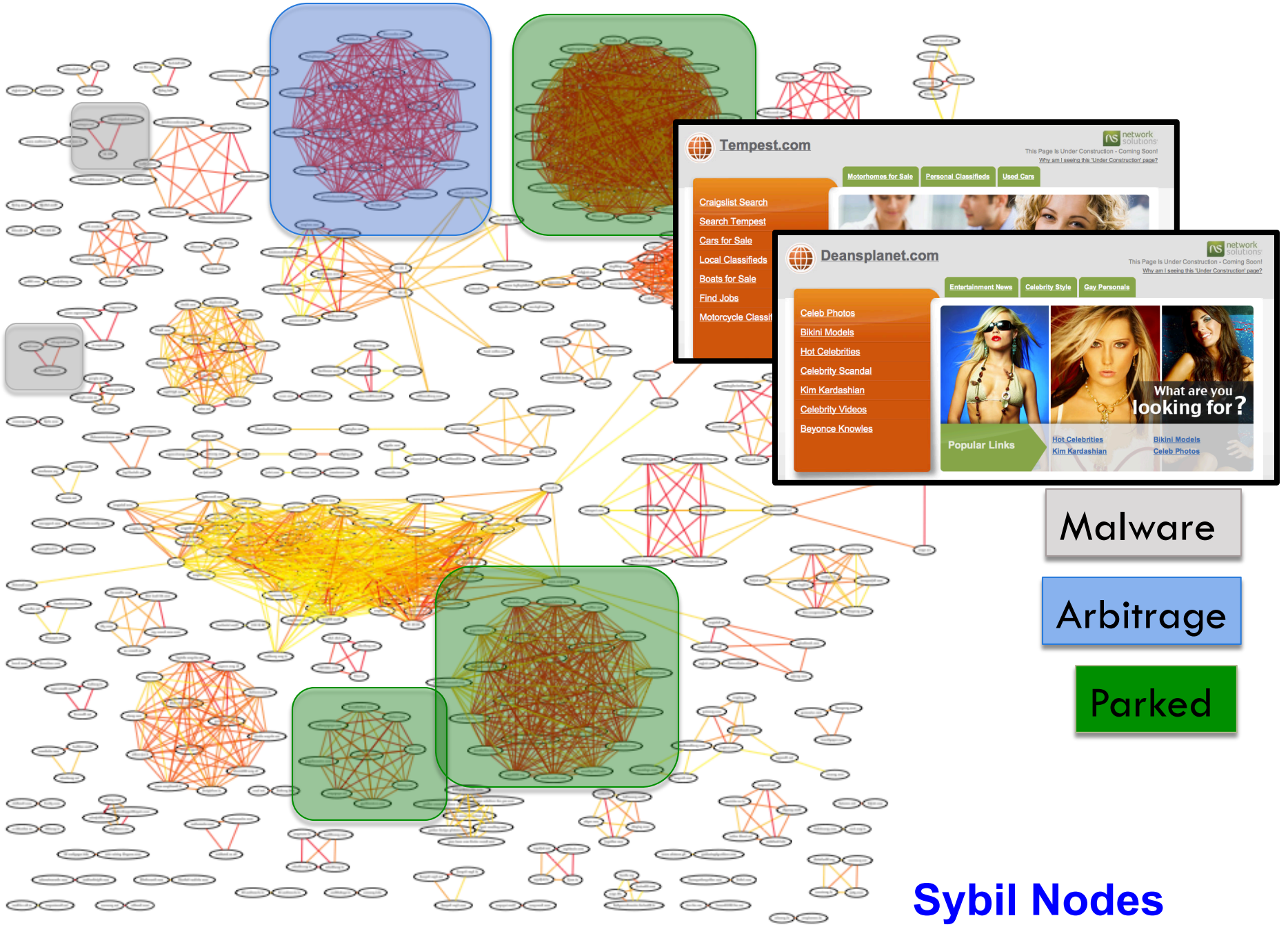
- 41of 400 parked domains hosted on a single IP
  - Misspellings of common websites:
    - icicbank.com, nsdi.com ☺
  - Auto-redirect depends on Jane's geo-location
  - IP hosts 500,000 such domains
- User mistypes a URL
  - Advertiser must pay!
- User behavior indistinguishable from normal traffic
  - Naively using conversions don't work

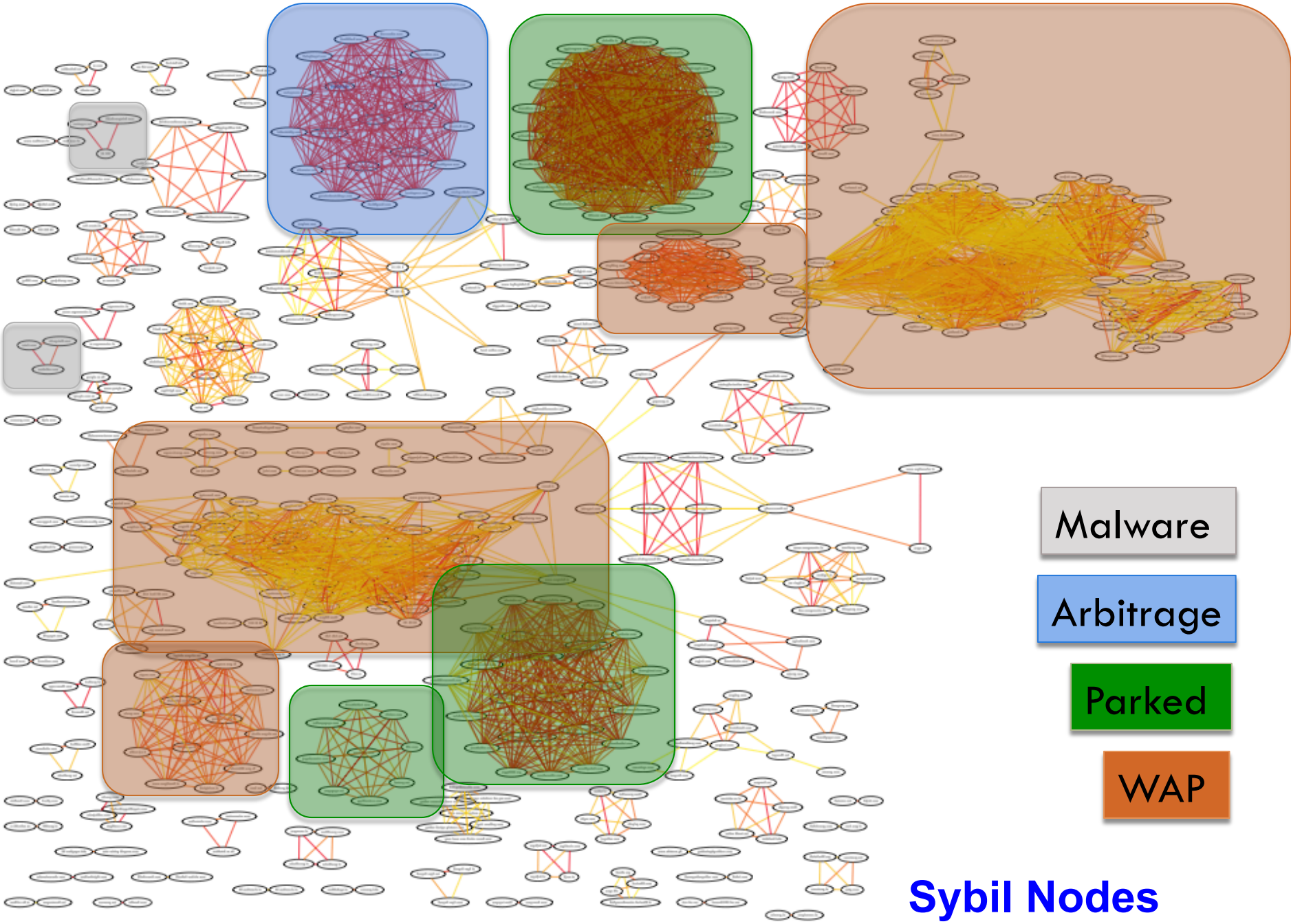
Case Study 4 – Mobile click-spam

- Indian Mobile ad network
  - Supplies WAP Ads to a group of WAP porn sites
  - Ad links indistinguishable from porn video links
- Gaming apps
  - Place ads close to where users are expected to click
  - Ant-Smasher, Milk-the-Cow, and 50 others


The SPECIALsearch.com
CoolSearchnow
123BOUNCE
Go
Arts
Movies, Television, Music
Games
Video Games, RPGs, Gambling
Keeds and Teens
Arts, School Time, Teen Life
Reference
Maps, Education, Libraries
Shopping
Clothing, Food, Gifts
Business
Jobs, Real Estate, Investing
Health
Fitness, Medicine, Alternative
News
Media, Newspapers, Weather
Regional
US, Canada, UK, Europe
Society
People, Religion, Issues
Computers
Internet, Software, Hardware
Home
Family, Consumers, Cooking
Recreation
Travel, Food, Outdoors, Humor
Science
Biology, Psychology, Physics
Sports
Baseball, Soccer, Basketball
Malware
Sybil Nodes


DoTellAll
Real People. Real Results.
LOGIN
Search
Restaurants, Shopping, etc.
GO
Living Frugal
Start a Goal
Q&A
My Goals
Log In
e.g. 401K, investing, etc...
Search for Goals
Setting a goal and
Start a Goal
Service MOUNTAIN
Start your search here
Search
Log In
Need a local expert? Start here!
Ask a Service Question Below
Enter question here
Ask Now »
Service Guides
How To Find A Residential Painter
A residential painter is a person who can come in and paint your home, inside and out, so that it looks brand new again. You may be surprised by the amount of difference that a co... Read More
Rated by (0) Users
How To Do Slab Leak Repair
Service Mountain
Coupon Book
Malware
Arbitrage
Sybil Nodes


Tempest.com
This Page Is Under Construction - Coming Soon!
Why am I seeing this 'Under Construction' page?
Motorhomes for Sale
Personal Classifieds
Used Cars
Craigslist Search
Search Tempest
Cars for Sale
Local Classifieds
Boats for Sale
Find Jobs
Deansplanet.com
This Page Is Under Construction - Coming Soon!
Why am I seeing this 'Under Construction' page?
Entertainment News
Celebrity Style
Gay Personals
Celeb Photos
Bikini Models
Hot Celebrities
Celebrity Scandal
Kim Kardashian
Celebrity Videos
Beyonce Knowles
What are you looking for?
Popular Links
Hot Celebrities
Kim Kardashian
Bikini Models
Celeb Photos
Malware
Arbitrage
Parked
Sybil Nodes


Malware
Arbitrage
Parked
WAP
Sybil Nodes

Summary

- Click-spam remains a problem
- First way of estimating click-spam Independently
  - As an advertiser, for a set of keywords
  - Extensive validation
- Sophisticated click-spam attacks today
  - Sybil sites
  - Malware mimics user behavior
  - Social engineering attacks and others
- Dataset is available for download
  - All clicks (minimally sanitized)
  - http://www.cs.utexas.edu/~vacha/sigcomm12-clickspam.tar.gz

Thanks!

Data at:

http://www.cs.utexas.edu/~vacha/sigcomm12-clickspam.tar.gz