

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

2

- ◆ 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

3

◆ 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.

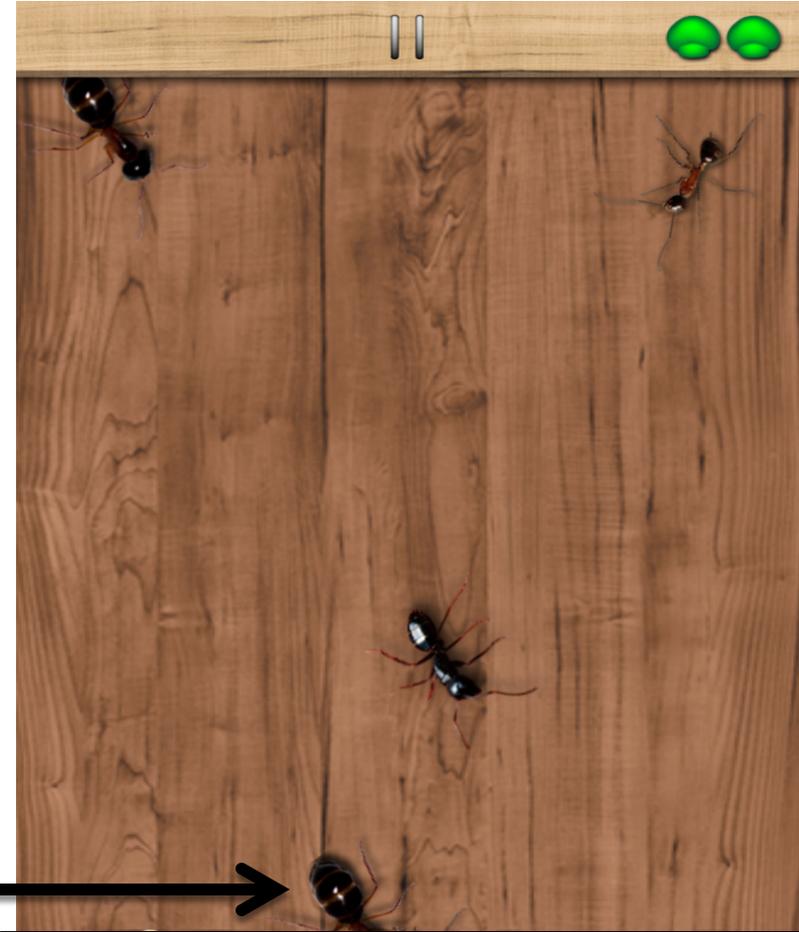
According to Microsoft, RedOrbit was able to **manufacture large numbers of bogus clicks on shady sites known as parked domains**—Web sites that are legal but exist only to display ads. In this case, many of the ads were invisible to the naked eye. Then, using a technical slight of hand, it submitted 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

5

Ant Smasher

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



Click-spam Detection

6

- ◆ 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

8

- ◆ 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



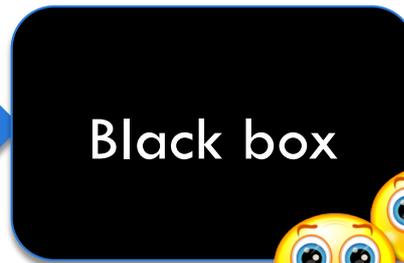
A fraction of non-spammers buy

How many 😬 ?

Equate ratios of buyers to non-spammers



Both non-spammers and spammers click ads

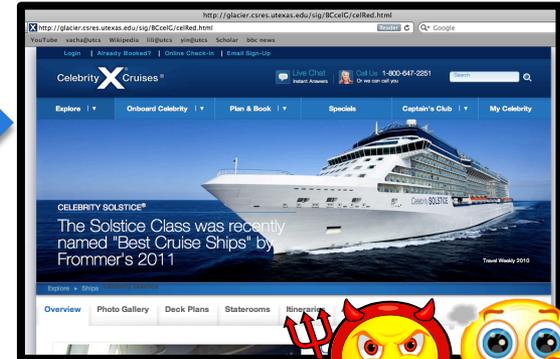
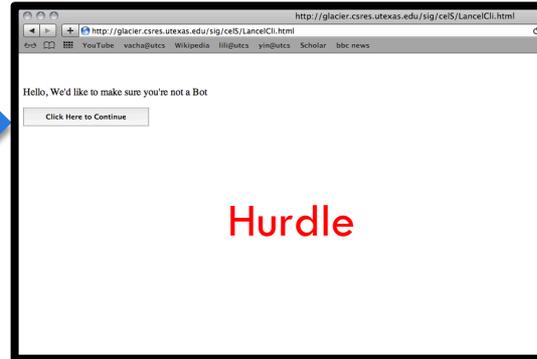


Lose spammers and some non-spammers



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

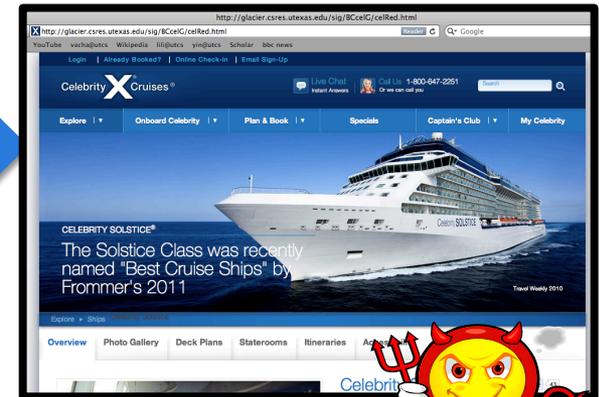
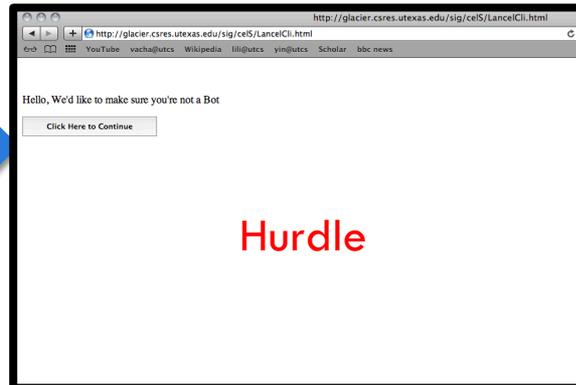
Celebrity Cruise
Be Recognized, Celebrated, And
Indulged Aboard Celebrity Cruises
cruisewithceleb.com

Normal

Massive smile Literature
Cream Fix Gutter Bad Keys
cruisewithceleb.com

Bluff

Massive smile Literature
Cream Fix Gutter Bad Keys
cruisewithceleb.com



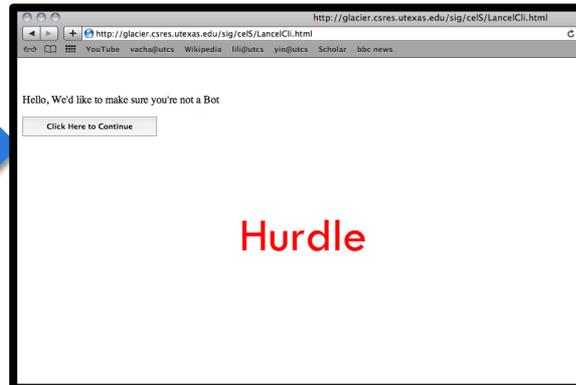
Spammers and curious
users click on an ad

Some spammers and
users may see the
content

Dissecting Black box - Bluff Ads[1]

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

Massive smile Literature
Cream Fix Gutter Bad Keys
cruisewithceleb.com



Spammers and curious users click on an ad

Some spammers and users may see the content

Testing Ad Networks

13

- ◆ 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

Celebrity Cruise
Be Recognized, Celebrated, And Indulged Aboard Celebrity Cruises
cruisewithceleb.com

Gentle Yoga for Beginners
No pretzel poses...just easy yoga for beginners. Award-winning DVDs.
gentleyogaforbeginners.com

Buy Any Zero Turn Mower
Get Free S&H +Pay No Tax \$2,079.99
CALL or Shop Online 4 Lowest Prices
zeroturnlawnmowers.com



Uh-oh. How do we validate?

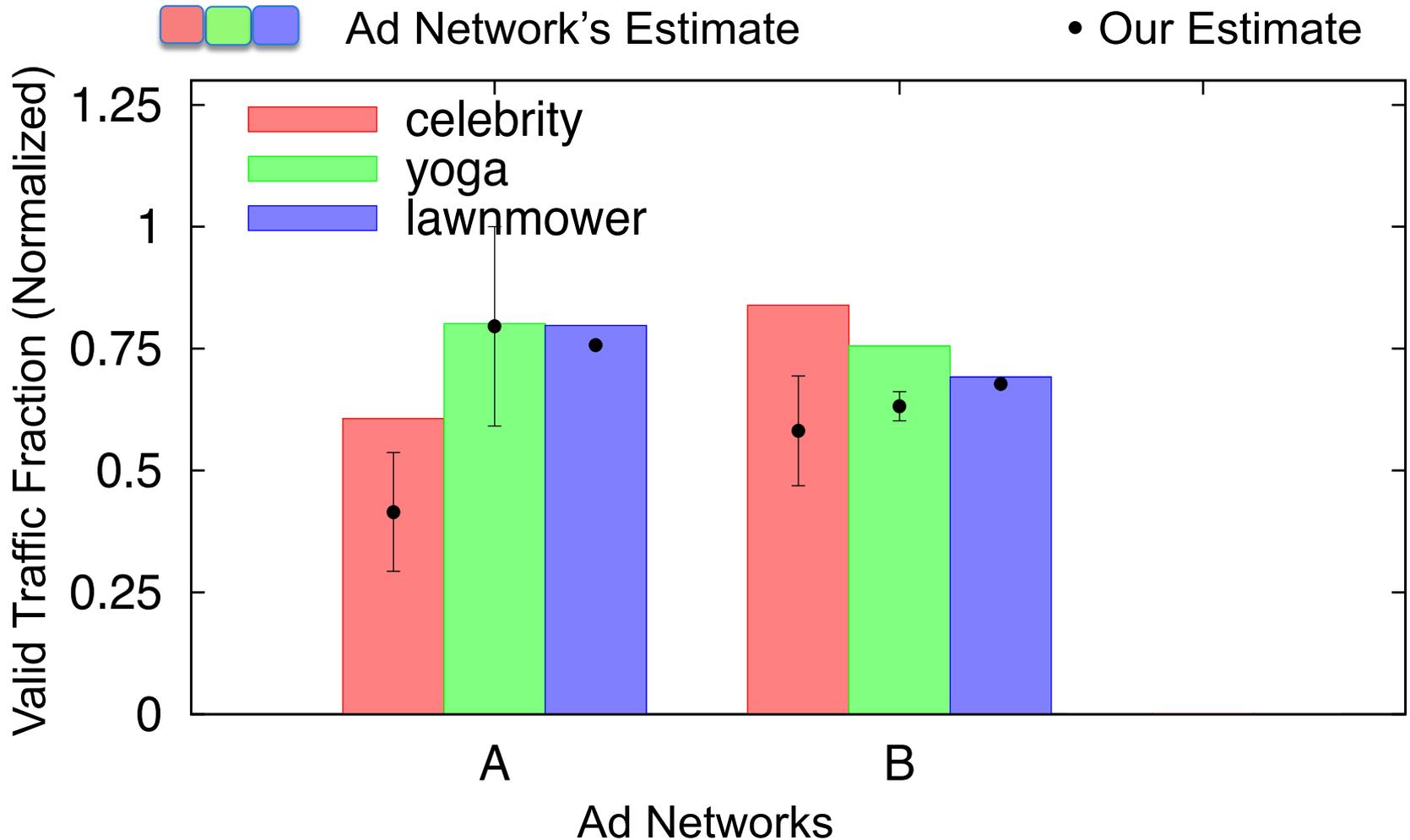
14

No ground truth!

Compare against **search ads** on Google and Bing

Results – Validation using search ads

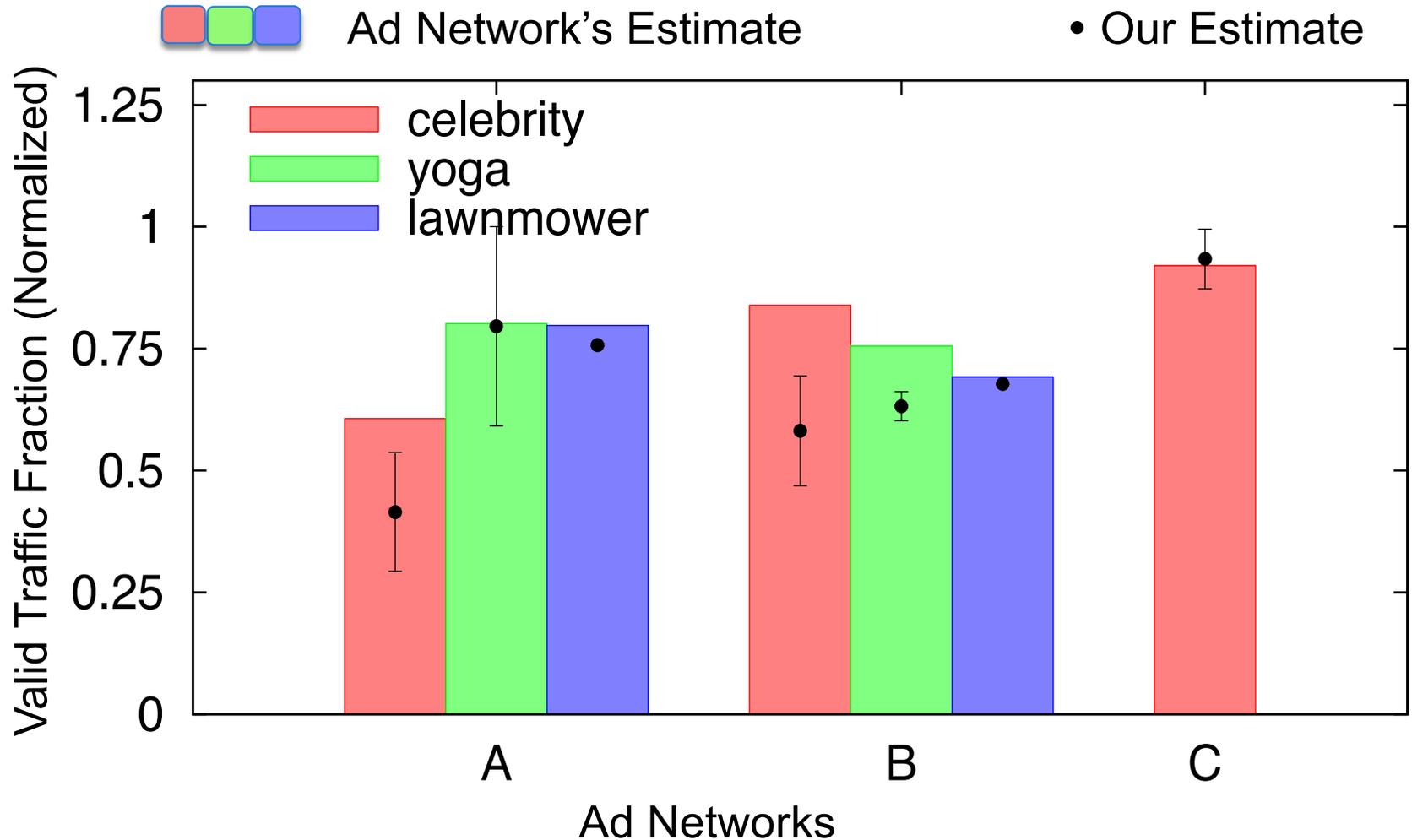
15



Clicks charged are close to the estimated valid clicks

Results – Validation using search ads

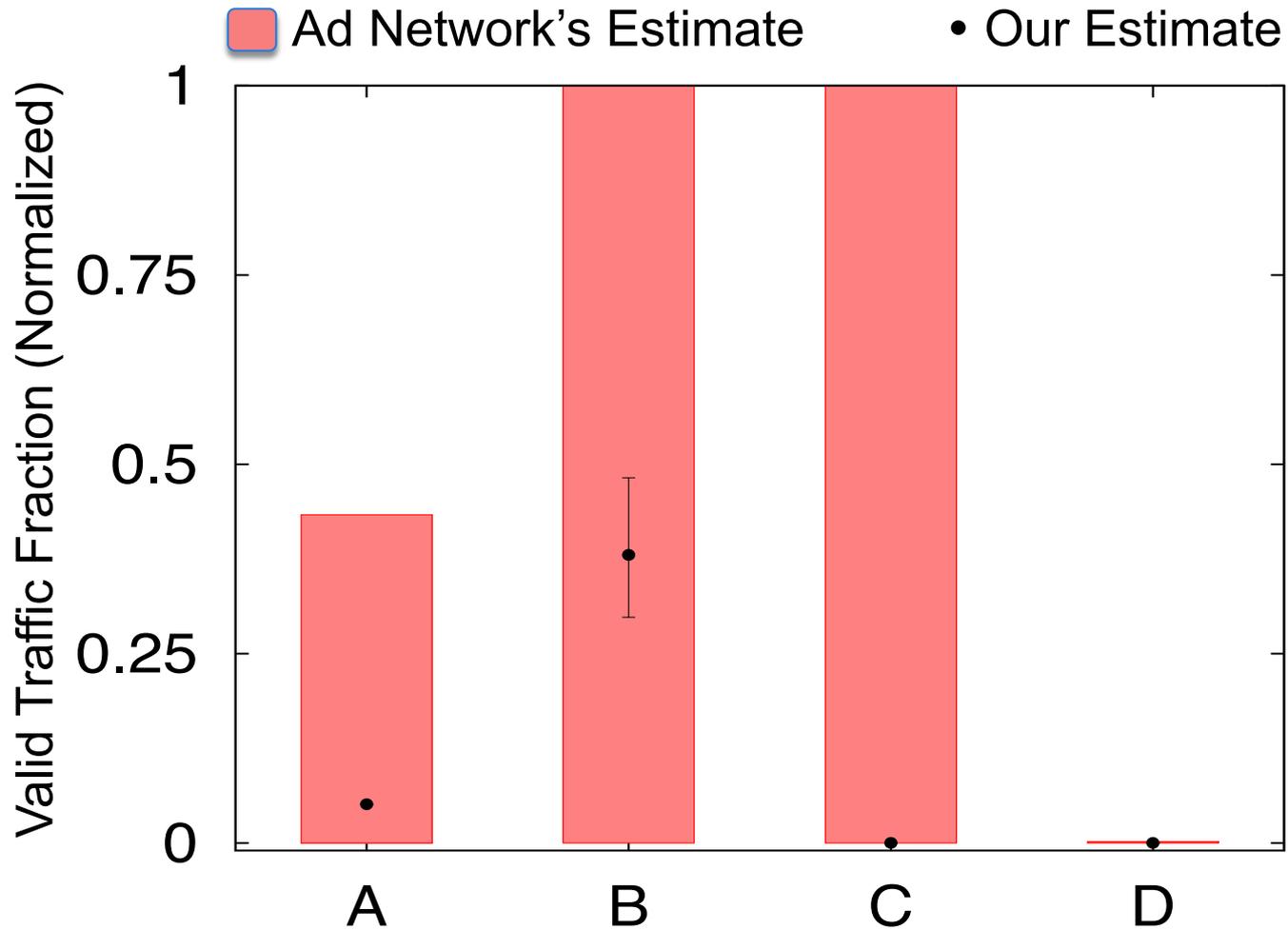
16



Clicks charged are close to the estimated valid clicks

Results – Estimating Mobile Spam

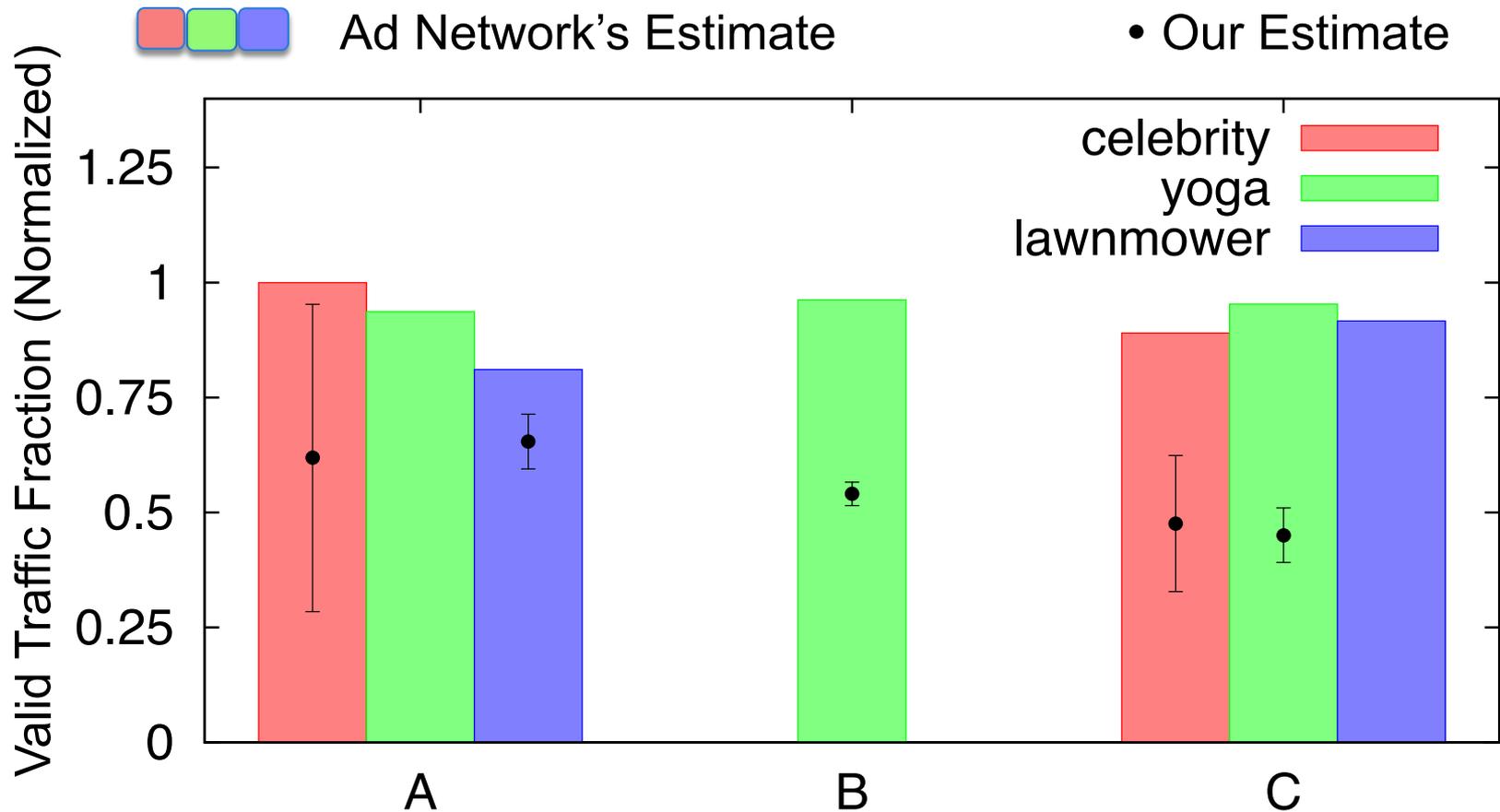
17



Most mobile ad networks fail to fight click-spam

Results – Estimating Contextual Spam

18

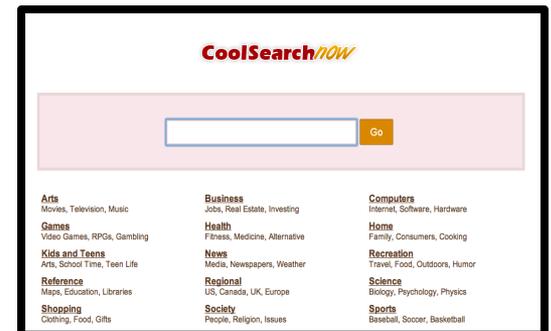
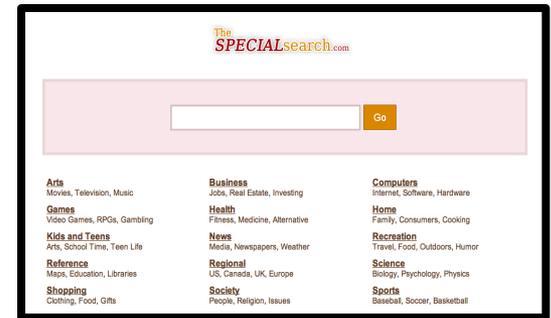


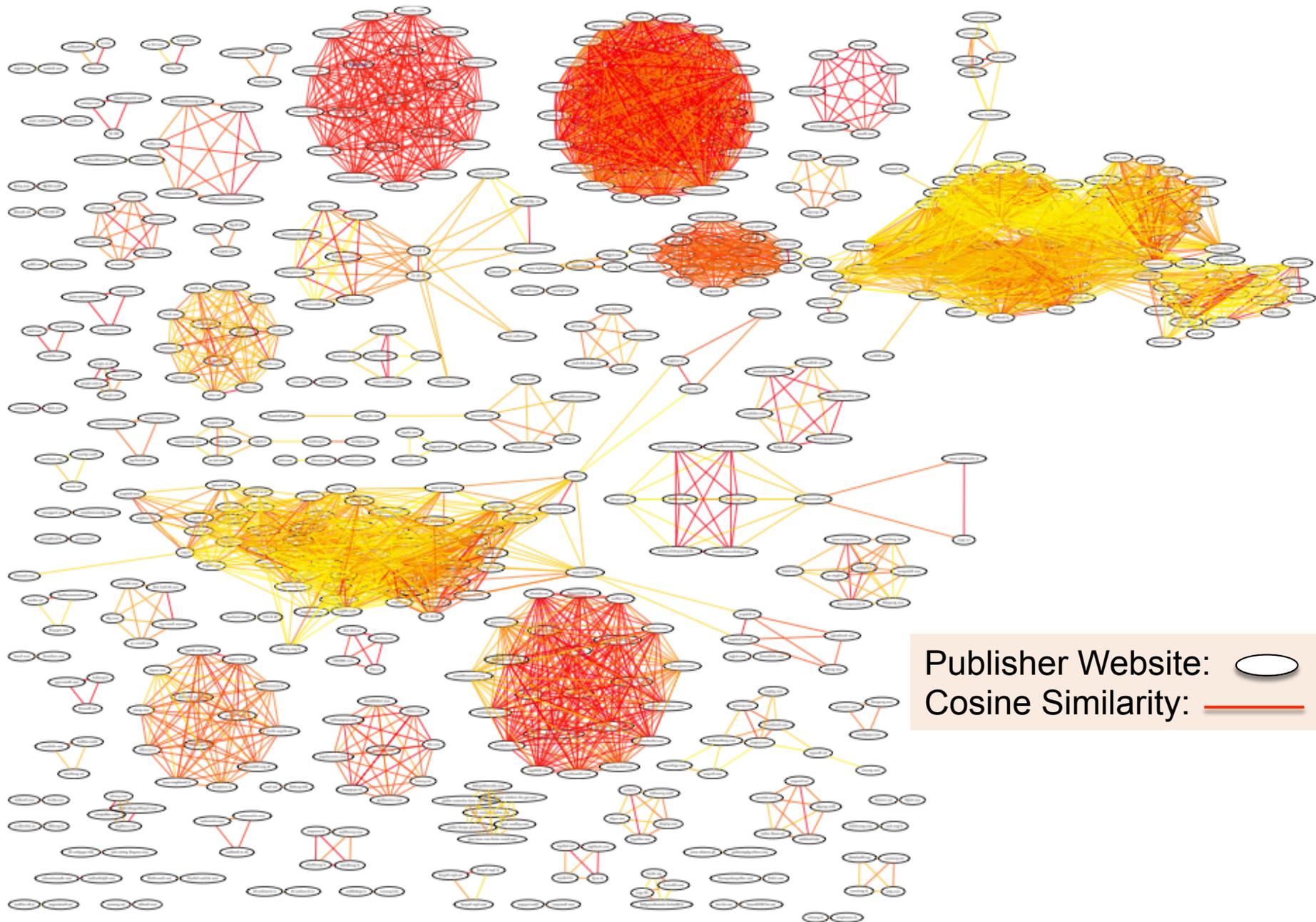
All networks seem to be underestimating the amount of spam

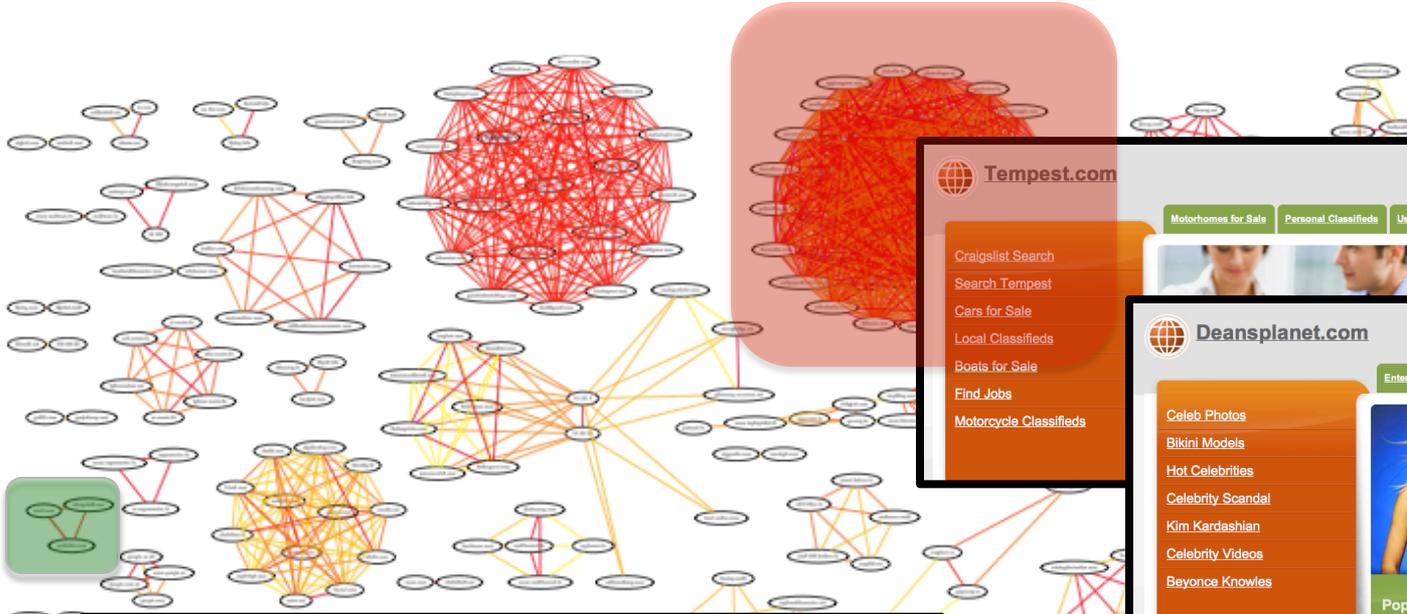
Where is click-spam coming from?

19

- ◆ 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







Tempest.com network solutions

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
Motorcycle Classifieds

Deansplanet.com network solutions

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

The **SPECIALsearch.com**

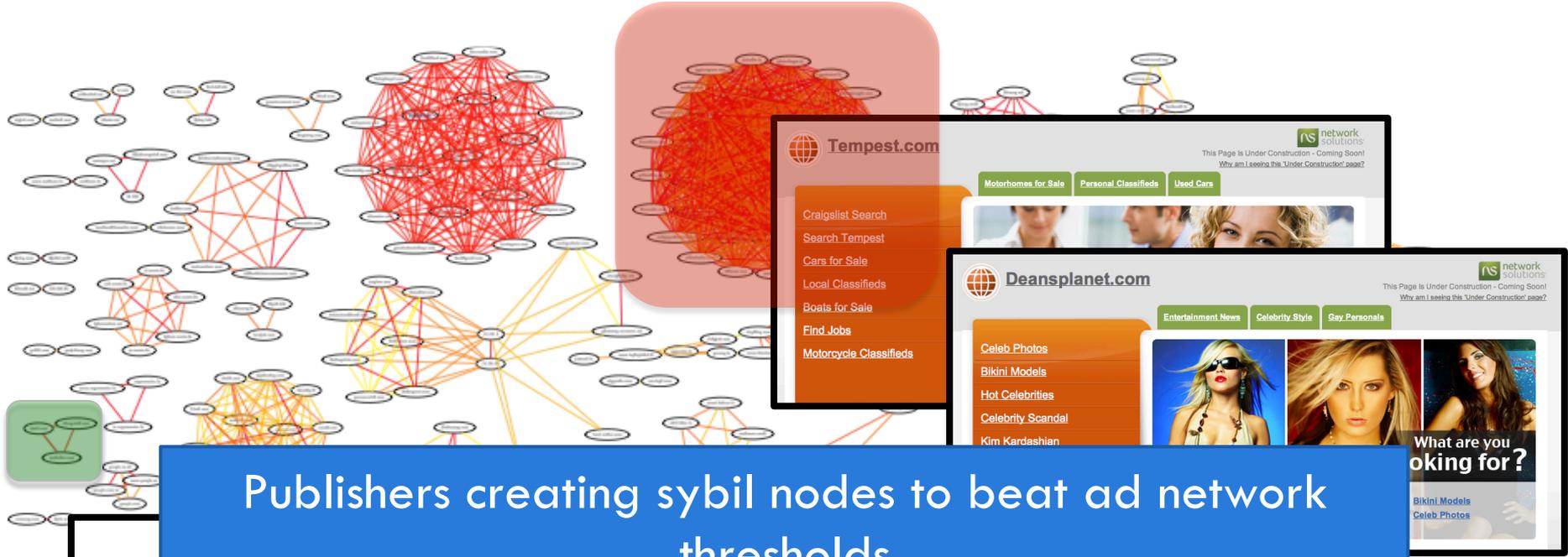
CoolSearchNOW

123BOUNCE

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

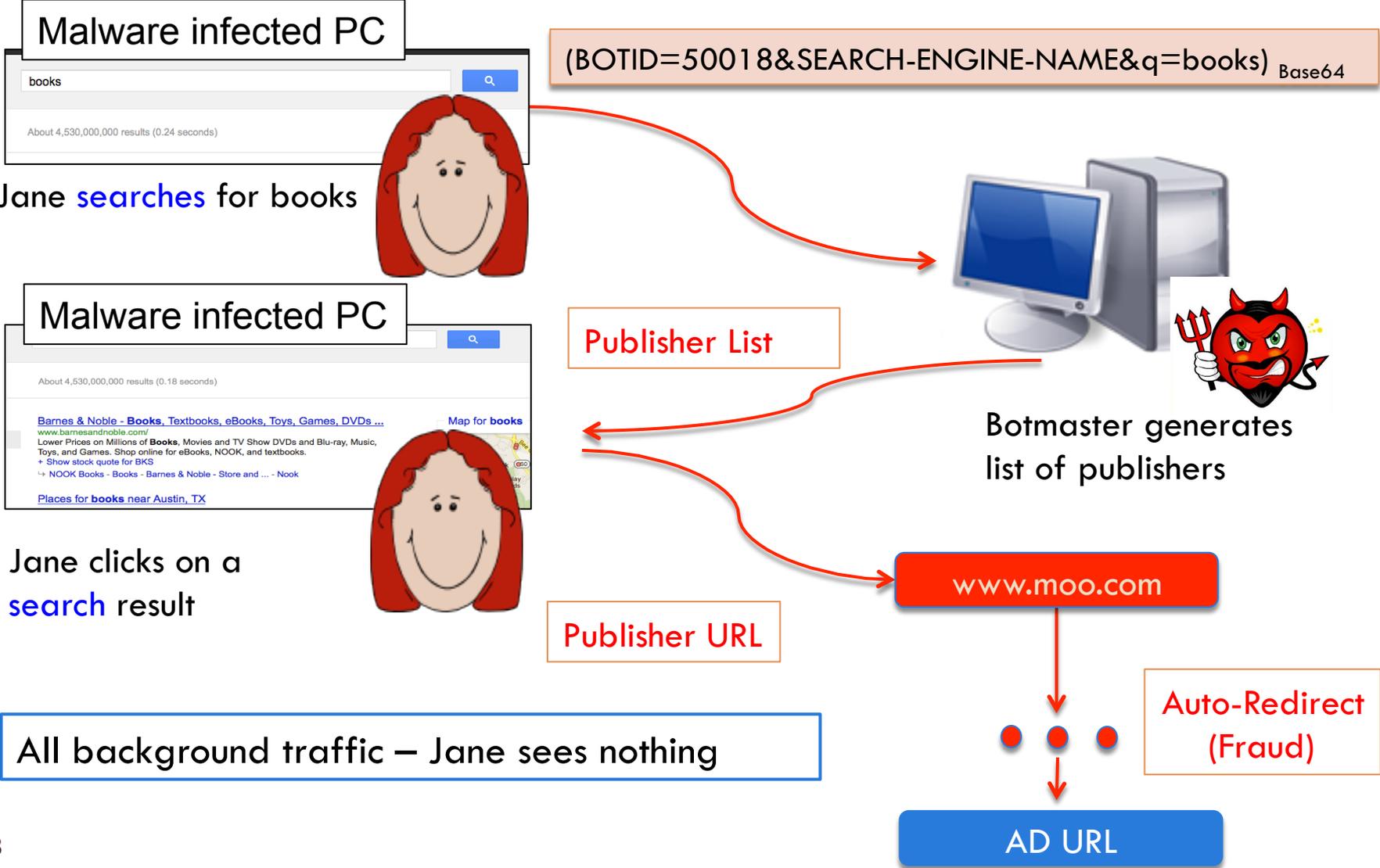
CoolSearchNOW

123 **BOUNCE**

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: —

Case Study 1 - Malware driven click fraud



Case Study 1 - Malware driven Click fraud

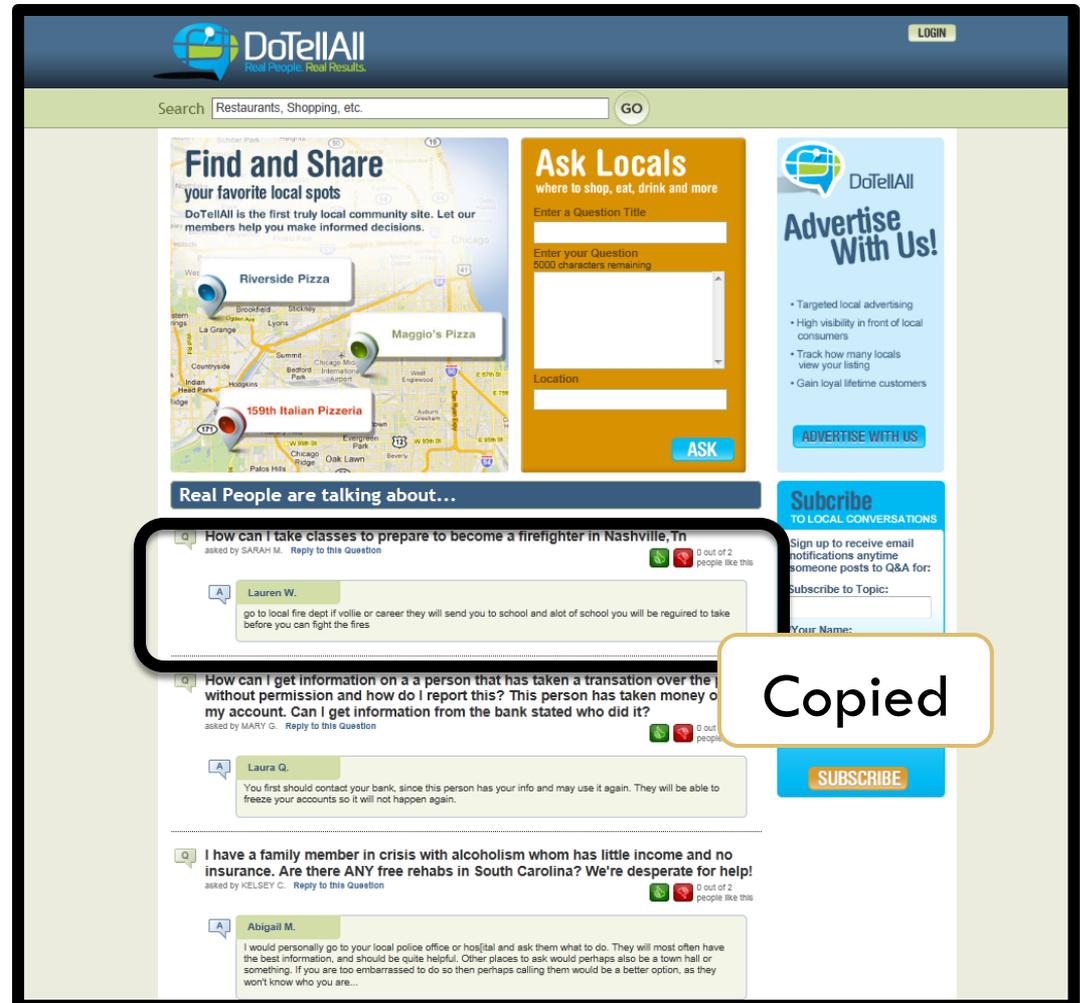
24

- ◆ 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

25

- ◆ 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

26

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**

The screenshot shows the DoTellAll website interface. At the top, there is a search bar with the text 'airport job' and a 'GO' button. Below the search bar, there is a list of job listings from various sources, including Indeed, Job Source, and Delta Air Lines. A blue box labeled 'Ads' is overlaid on the page, indicating that the page contains only advertisements. A red text box is also overlaid on the page, stating 'Site earns \$\$\$\$ from X'.

Job Title	Distance	Contact Info
West Coast Financial Jobs	7.82 Miles	8620 Silverstone Cir Austin, TX 78759 512-502-0168
Americas Job Network	0.28 Miles	816 Congress Ave Austin, TX 78701 512-493-3658
Wiser Construction Job Site	0.28 Miles	Austin, TX 78701 512-891-0000
Small Mold Jobs R US	0.28 Miles	Austin, TX 78701 512-282-8682
Job Training Institute	2.56 Miles	2211 S IH 35 Austin, TX 78741 512-445-5400
Job News Austin	3.25 Miles	2028 E Ben White Blvd Austin, TX 78741 512-693-9937
Job News USA	3.25 Miles	2028 E Ben White Blvd Ste 120 Austin, TX 78741 512-693-4750

Click-spam and Arbitrage

27

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

The screenshot shows a job search website with the following elements:

- Header: DoTellAll Real People, Real Results.
- Search bar: Search airport job GO
- Featured Listings section containing several job listings:
 - Airports jobs**: Search for Airports jobs. Find your new job today. Indeed™.
 - (Hiring) Airport Jobs**: New Airport Positions Daily. www.airport-jobs.org
 - Now Hiring Immediately**: Over 303 Local Jobs Found. Jobsources.info
 - Delta Air Lines - Apply**: Found: Hiring Positions Near You. \$10.50 - \$83.75/hr. Apply Online. Hiring.Jobs.WorkGrabber.net
 - Airport Job**: 2 Jobs Left. Apply Now! Airport Job. US.Jobrapido.com
 - Delta Air Lines - Hiring**: Positions available at Delta Air Lines. Apply now. Jobs.CareersLocal.net
- A blue callout box labeled "Ads" points to the job listings.
- A red text box says "Site earns \$\$\$\$ from X".
- Below the featured listings is a list of search results with columns for job title, distance, address, and phone number.

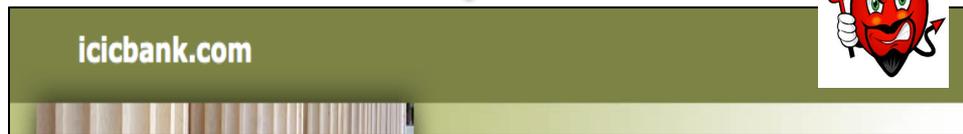
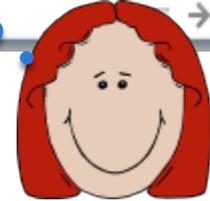
Job Title	Distance	Address	Phone Number
West Coast Financial Jobs	7.82 Miles	8620 Silverstone Cir Austin, TX 78759	512-502-0168
Americas Job Network	0.28 Miles	816 Congress Ave Austin, TX 78701	512-493-3658
Wiser Construction Job Site	0.28 Miles	Austin, TX 78701	512-891-0000
Small Mold Jobs R US	0.28 Miles	Austin, TX 78701	512-282-8682
Job Training Institute	2.56 Miles	2211 S IH 35 Austin, TX 78741	512-445-5400
Job News Austin	3.25 Miles	2028 E Ben White Blvd Austin, TX 78741	512-693-9937
Job News USA	3.25 Miles	2028 E Ben White Blvd Ste 120 Austin, TX 78741	512-693-4750

Case Study3 - Click Fraud using Parked Domains

Go to icicibank.com

www.icicibank.com

Jane mistypes **icicbank.com** in her browser and presses enter



Parked Domain

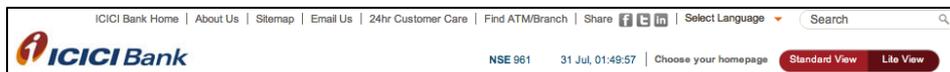
Auto-Redirect

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

29

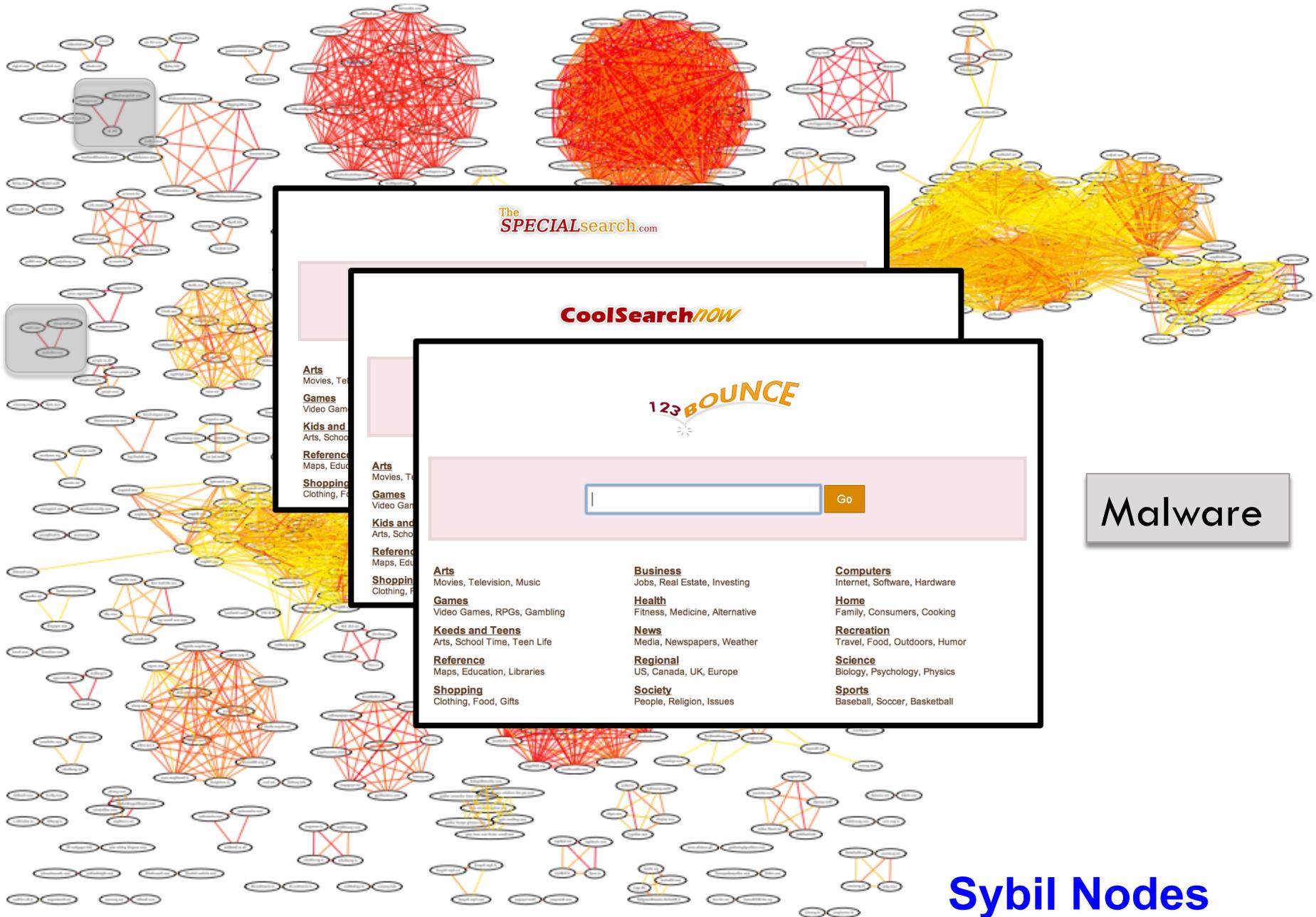
- ◆ 41 of 400 parked domains hosted on a single IP
 - ◆ Misspellings of common websites:
 - icicbank.com, [nsdi.com](#) 😊
 - ◆ Auto-redirect depends on Jane's geo-location
 - ◆ IP hosts 500,000 such domains
- ◆ User mistypes a URL
 - ◆ Advertiser must pay!
- ◆ User behavior indistinguishable from normal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

Case Study 4 – Mobile click-spam

30

- ◆ 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

CoolSearch*now*

123**BOUNCE**

Arts
Movies, Television, Music

Games
Video Games, RPGs, Gambling

Kids and Teens
Arts, School Time, Teen Life

Reference
Maps, Education, Libraries

Shopping
Clothing, Food, Gifts

Arts
Movies, Television, Music

Games
Video Games, RPGs, Gambling

Kids and Teens
Arts, School Time, Teen Life

Reference
Maps, Education, Libraries

Shopping
Clothing, Food, Gifts

Arts
Movies, Television, Music

Games
Video Games, RPGs, Gambling

Kids 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

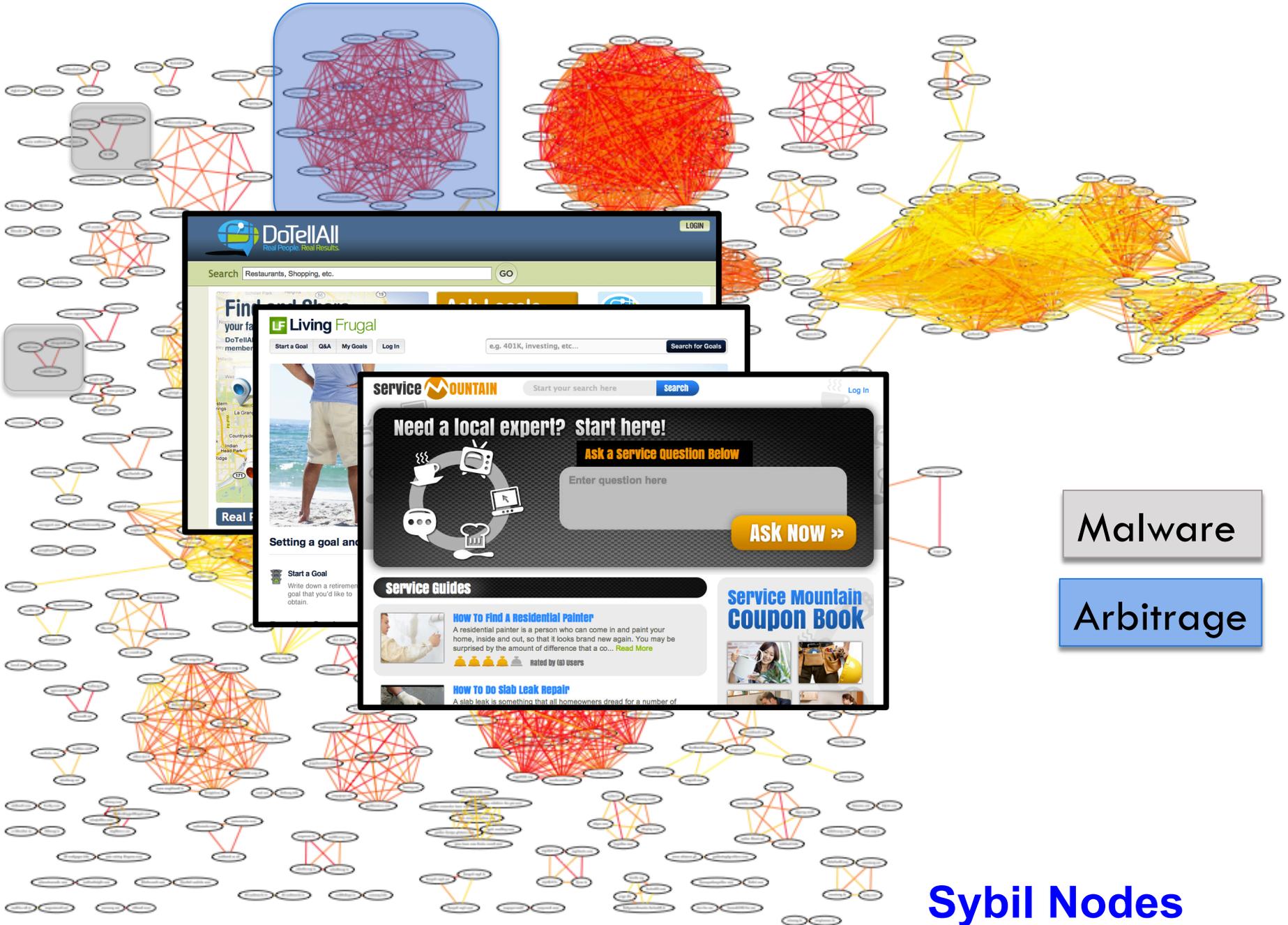
Recreation
Travel, Food, Outdoors, Humor

Science
Biology, Psychology, Physics

Sports
Baseball, Soccer, Basketball

Malware

Sybil Nodes



DoTellAI
Real People. Real Results.

Search Restaurants, Shopping, etc.

LOGIN

Living Frugal

Start a Goal Q&A My Goals Log In

e.g. 401K, investing, etc...

Service Mountain Start your search here 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 (3) users

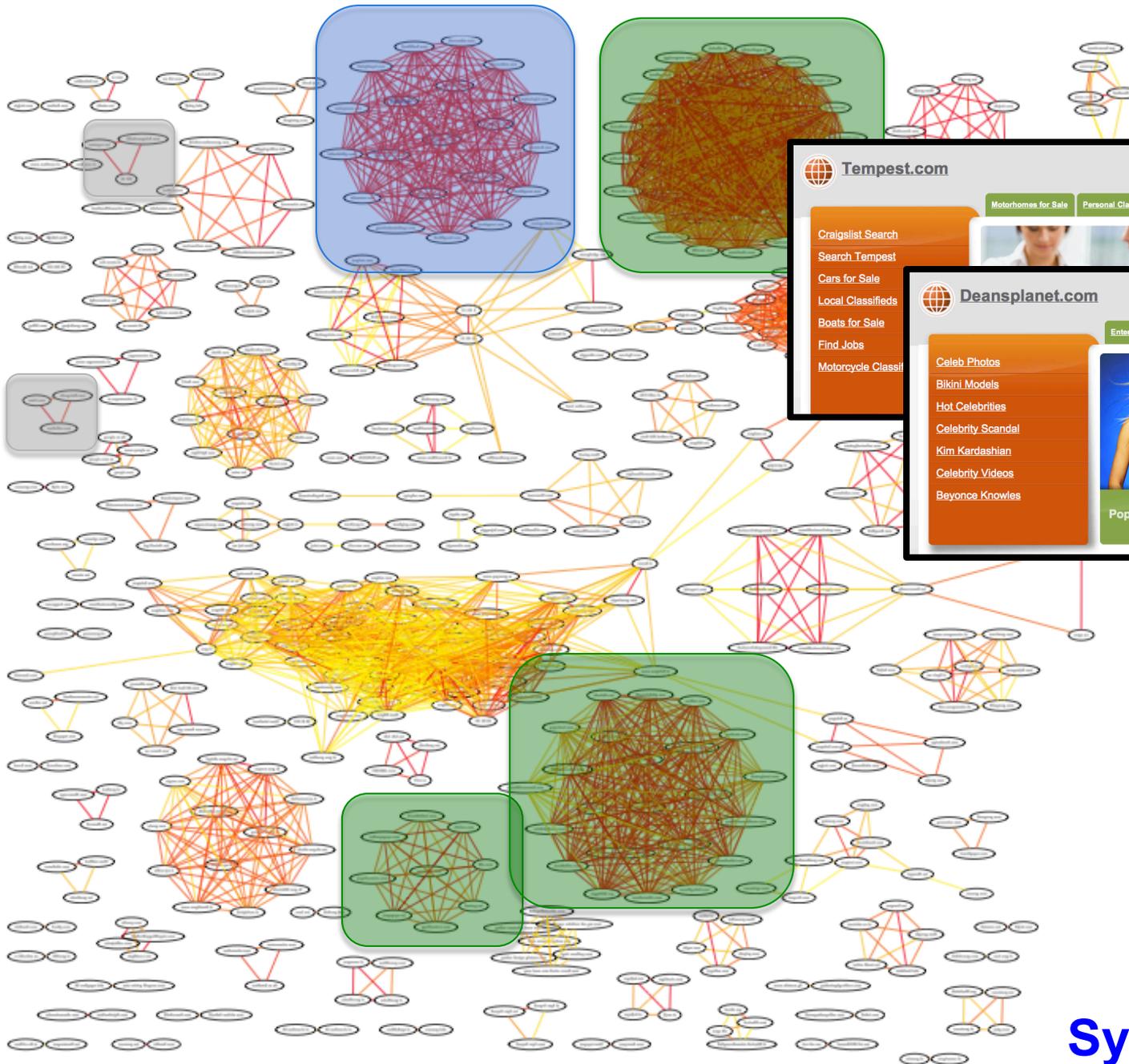
Service Mountain Coupon Book

HOW TO do slab Leak Repair
A slab leak is something that all homeowners dread for a number of

Malware

Arbitrage

Sybil Nodes



A screenshot of the Tempest.com website interface. The header includes the logo and navigation links for "Motorhomes for Sale", "Personal Classifieds", and "Used Cars". The main content area features a search bar and a list of categories: "Craigslst Search", "Search Tempest", "Cars for Sale", "Local Classifieds", "Boats for Sale", "Find Jobs", and "Motorcycle Classif". There are also images of people's faces.

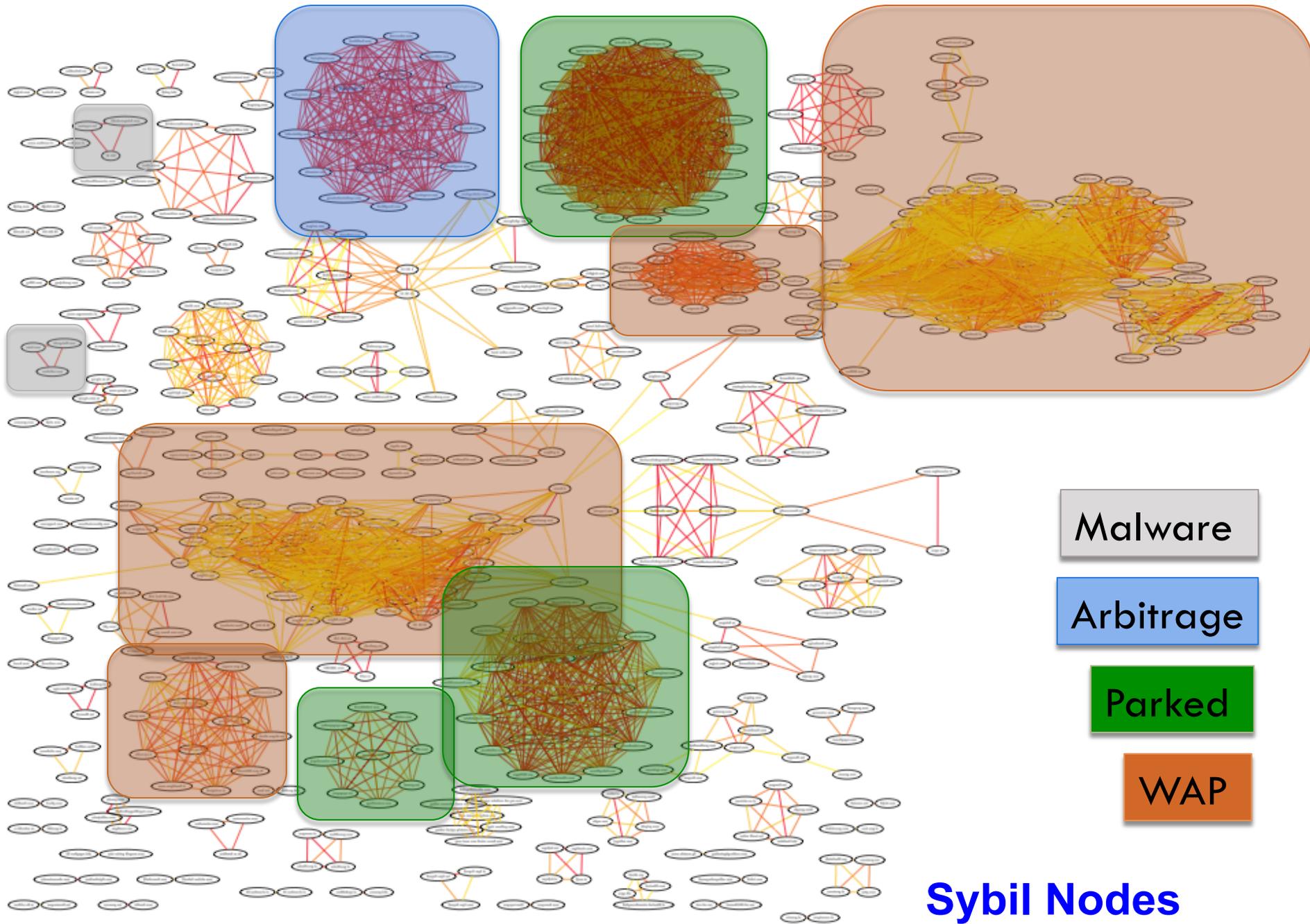
A screenshot of the Deansplanet.com website interface. The header includes the logo and navigation links for "Entertainment News", "Celebrity Style", and "Gay Personals". The main content area features a list of categories: "Celeb Photos", "Bikini Models", "Hot Celebrities", "Celebrity Scandal", "Kim Kardashian", "Celebrity Videos", and "Beyonce Knowles". There are also images of celebrities and a "What are you looking for?" section with "Popular Links" for "Hot Celebrities Kim Kardashian" and "Bikini Models Celeb Photos".

Malware

Arbitrage

Parked

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>

Data at:

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