The background of the slide features a large, faint watermark of the Georgetown University seal. The seal is circular and contains a central figure of an eagle with wings spread, perched on a shield. Above the eagle is a lyre, and below it is a banner with the motto 'VERITAS LIBERABIT VOS'. The seal is surrounded by a laurel wreath and the Latin motto 'SIGILLUM UNIVERSITATIS GEORGETOWNENSIS' is visible around the perimeter.

Privacy & Social Media

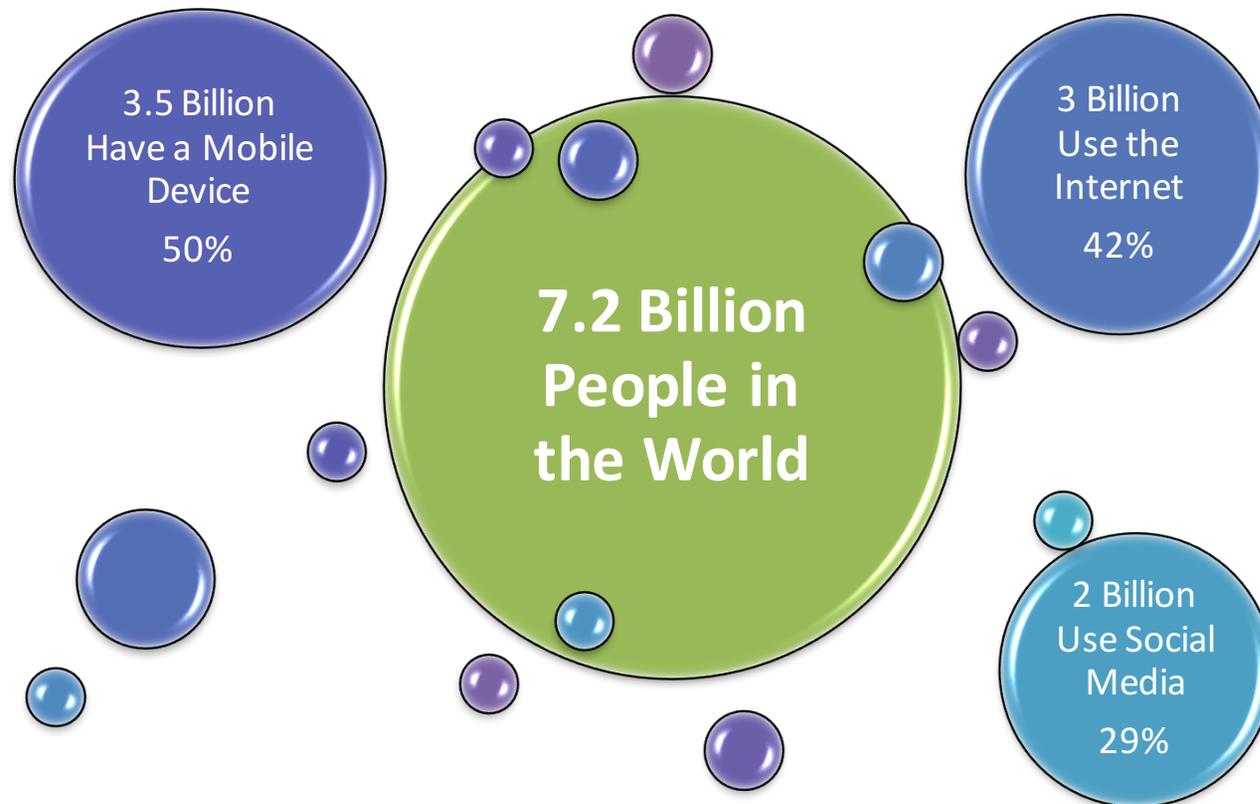
Lisa Singh, PhD
Department of Computer Science
Georgetown University

GEORGETOWN
UNIVERSITY

Outline

- Our world on the Internet
- Data privacy in a public profile world
- Methods for determining our web footprints
- Taking control of our web identities

Our presence on the Internet and social media



Data, so much data...

Users share 70 billion pieces of content each month on Facebook

190 million tweets are sent per day

65 hours of video are uploaded to YouTube every minute

Image from <http://www.playbuzz.com/jaylam10/which-social-media-fits-your-personality>

Privacy settings and social media

- 25% of Facebook users do not bother with any privacy settings (velocitydigital.co.uk, 2013)
- 37% of Facebook users have used the site's privacy tools to customize how much information apps are allowed to see (Consumer reports, 2012)
- 40% of teen Facebook users DO NOT set their Facebook profiles to private (friends only) (Pew Study 2013)
 - 71% post their school name
 - 71% post the city or town where they live
 - 53% post their email address
 - 20% post their cell phone number

Consequences of Over-sharing

- Identity theft
- Online and physical stalking
- Blackmailing
- Negative employment consequences
- Enabling of snoopers



Data Privacy Expectations

- We should expect data privacy
- We should expect freedom from unauthorized use of our data
- We should expect freedom from data intrusion.



How informative, linkable, or sensitive is your public profile – your web footprint?



Divorced

Gay

Department of Defense

Washington, DC

Spanish-speaking



Georgetown University

John Smith

Catholic

Software Developer



Republican

John Smith



Your name

You? Claim & edit »	Lisa Singh ?	Gusryan St Baltimore, MD	Associated people: unknown See full listing »
You? Claim & edit »	Lisa Singh	Lamberton Square Rd Silver Spring, MD	Associated people: unknown See full listing »
You? Claim & edit »	Lisa D Singh (Age 35-39) ?	Sharma St Capitol Heights, MD	Associated people: Liza Singh Elvis D Singh + more... See full listing »
You? Claim & edit »	Lisa D Singh (Age 50-54) ?	Ridge Rd Westminster, MD	Associated people: Paul Singh Amanda Singh + more... See full listing »
You? Claim & edit »	Lisa E Singh (Age 40-44)	Mount Aetna Rd Hagerstown, MD	Associated people: unknown See full listing »
You? Claim & edit »	Lisa J Singh (Age 50-54) ?	Manor Rd Glen Burnie, MD	Associated people: unknown See full listing »
You? Claim & edit »	Lisa O Singh (Age 40-44)	Brightstone Pl Ellicott City, MD	Associated people: Robert P Singh + more... See full listing »

It for Micah Sherr

h: Last name only

[Micah B Sherr](#)
(Age 30-34)

Associated people:
Lisa M Sherr
Lisa L Huang
+ more...
[See full listing »](#)

SPONSORED LINKS

[Micah B Sherr](#) (Age 33) Washington, DC Email Phone

Sponsored by [PeopleSmart](#)

nal Suggestions

[M Sherr](#)
(Age 30-34)

Royal Oaks Blvd
Franklin, TN

Associated people:
unknown
[See full listing »](#)

[Mickey C Sherr](#)
(Age 30-34)

N Miller Rd
Scottsdale, AZ

Associated people:
Miriam C Sherr
+ more...
[See full listing »](#)

Linking data

Facebook

First Name: Sally
Last Name: Smith
Gender: Female
Location: Georgetown
Hometown: Pittsburgh
Favorite Sports Team: Seahawks
Religion: Atheist

Google+

First Name: Sally
Last Name: Smith
Gender: Female
Location: Georgetown
Occupation: Dentist
Relationship Status: Married
Zip code: 22033

Linking data

Facebook

Adversary's Beliefs Google+

First Name: Sally

Last Name: Smith

Gender: Female

Location: Georgetown

Hometown: Pittsburgh

Favorite Sports Team: Seawhaws

Religion: Atheist

First Name: Sally

Last Name: Smith

Gender: Female

Location: Georgetown

Hometown: Pittsburgh

Occupation: Dentist

Favorite Sports Team: Redskins

Religion: Atheist

Relationship Status: Married

Zip Code: 22033

First Name: Sally

Last Name: Smith

Gender: Female

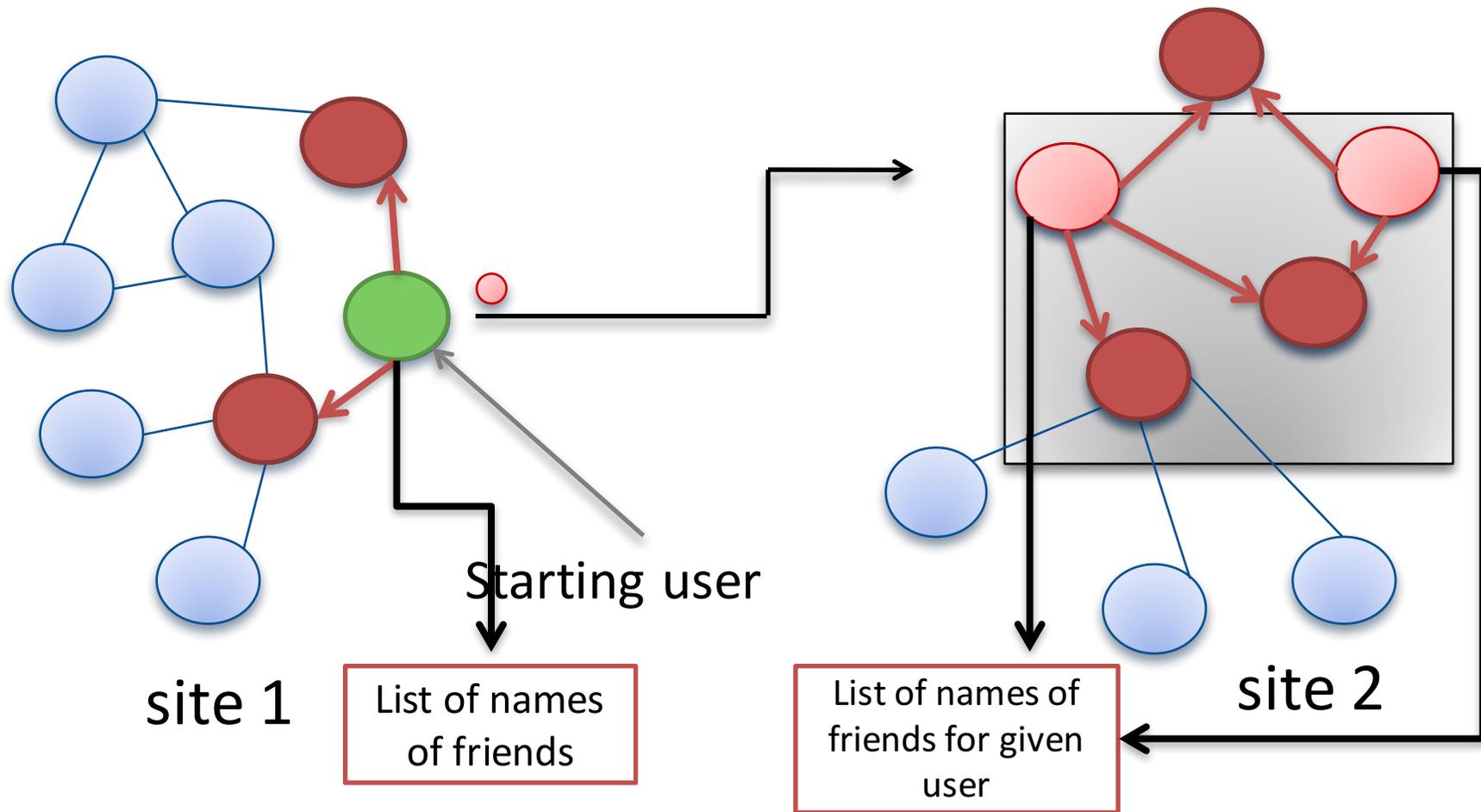
Location: Georgetown

Occupation: Dentist

Relationship Status: Married

Zip Code: 22033

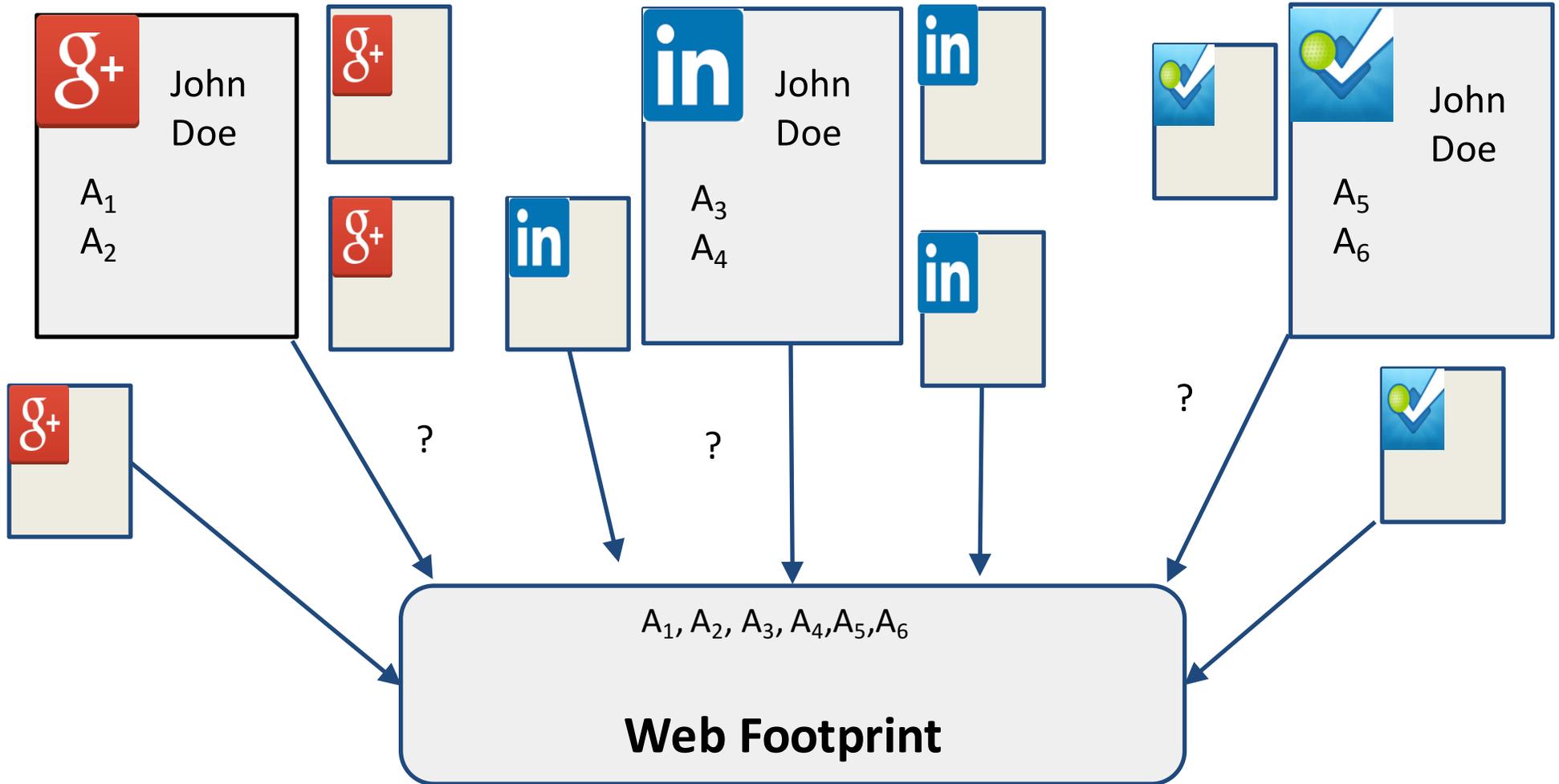
What about friends?



match = number overlapping friends between users

[Ramachandran et al., 2012]

Really linking data



Shared Public Attributes

Google+

- Company
- Occupation
- Education
- Location
- Birthdate
- Relationship status
- Gender
- Graduation Year

LinkedIn

- Company
- Location
- Education
- Email
- Occupation
- Skills
- Industry
- Website
- Languages

FourSquare

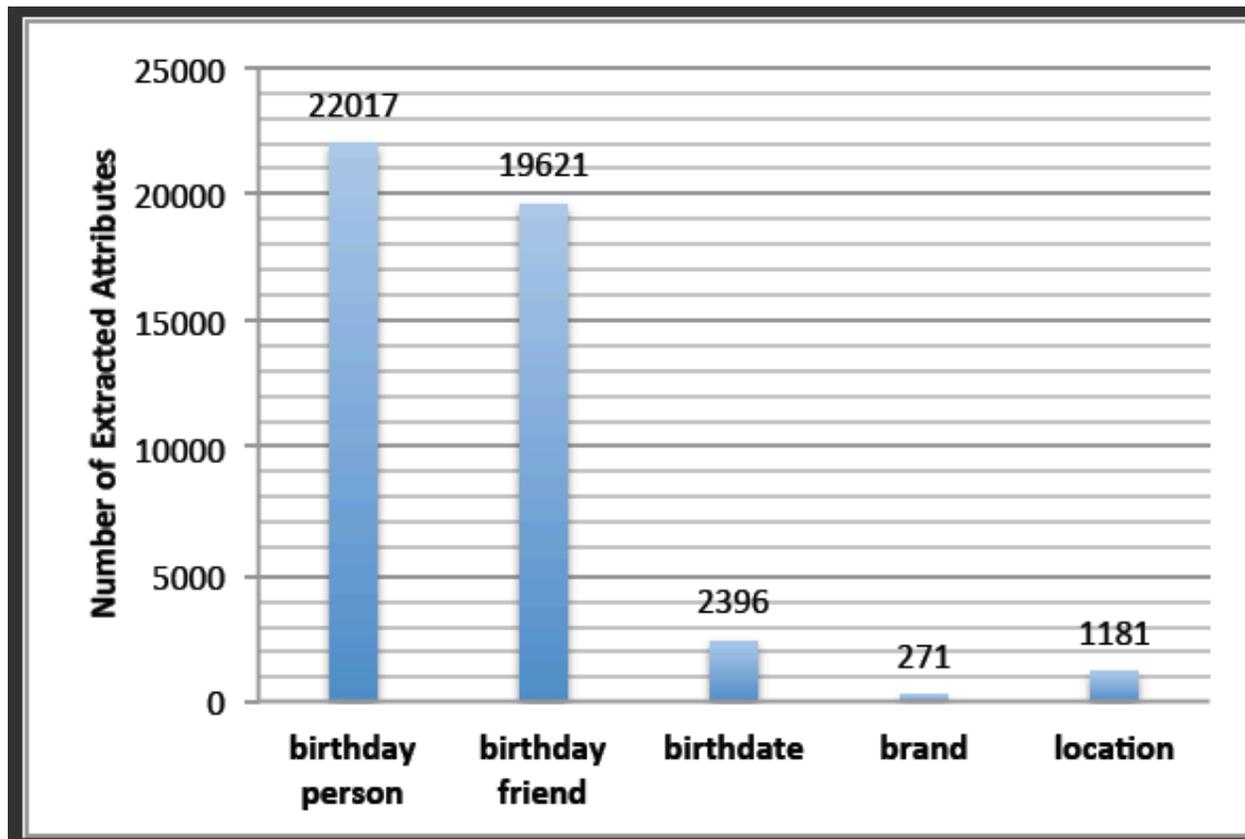
- Facebook id
- Twitter handle
- Email
- Gender
- Location
- Phone number
- Relationship status

What do group memberships tell us?

The screenshot shows a Facebook group page for 'College Democrats'. The header features a red banner with an equals sign and a 'Join Group' button. Below the banner are tabs for 'College Democrats', 'Members', and 'Photos', along with a search bar. The main content area is titled 'RECENT ACTIVITY' and shows a post by Grace Smith from February 24. The post includes a photo of a rainbow flag and the text 'Breaking Binaries' with event details: 'Tuesday, February 24 at 8:00pm', 'ICC 102', and '133 people went'. Below the post are 'Like' and 'Share' buttons, and a notification that 'Patrick Byllis likes this.' On the right side, the 'ABOUT' section shows '155 members' and 'Open Group within Georgetown'. Below this is a 'CREATE NEW GROUPS' section with a description and a '+ Create Group' button. The 'SUGGESTED GROUPS' section lists three groups: 'Georgetown Summer Sublet', 'Georgetown Pep Band', and 'Philodemic Society', each with a '+ Join' button.

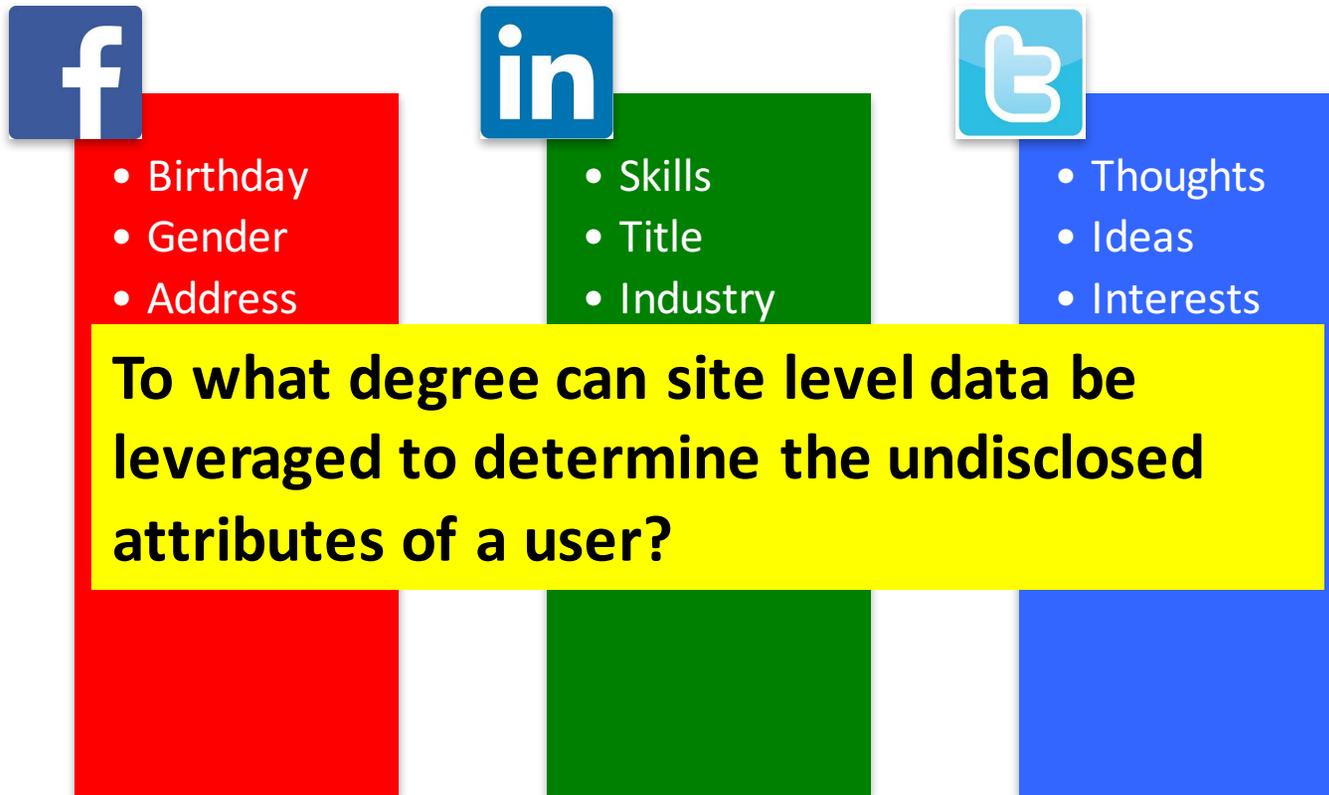
What about tweets?

- A special wish for a special girl #HappyBirthday
- I love #Starbuck #MangoTeaLemonade
- Go #Bears!!!!

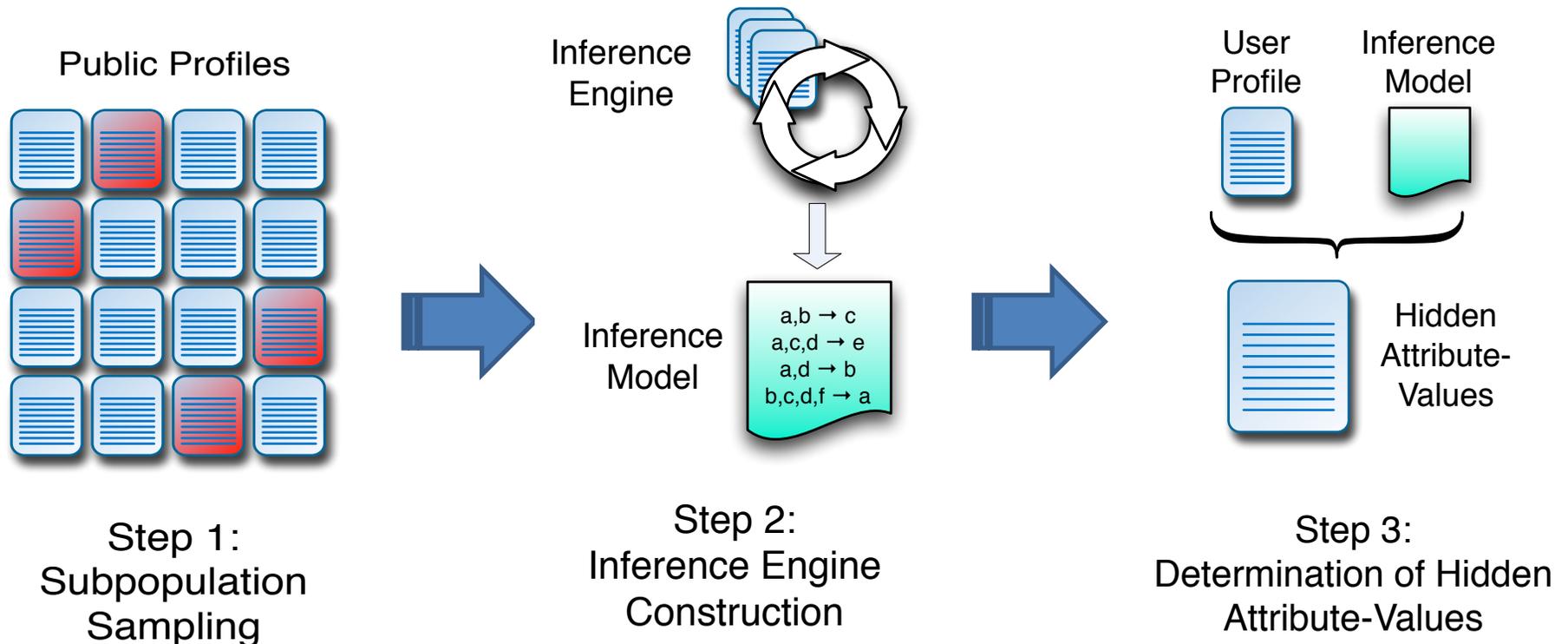


[Singh et al., 2015]

What about the population?



Methodology

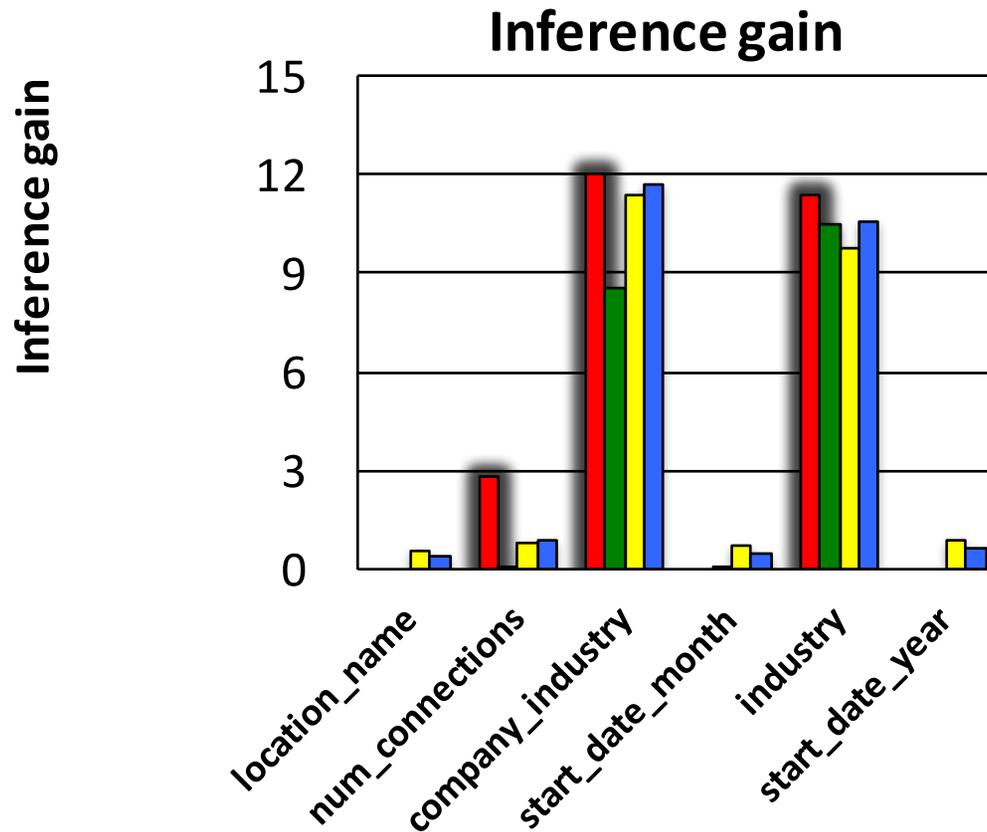


- Use sample profiles to construct an inference engine containing a set of inference rules.
- Make inferences using the inference engine.

LinkedIn dataset:

91,150 public profiles

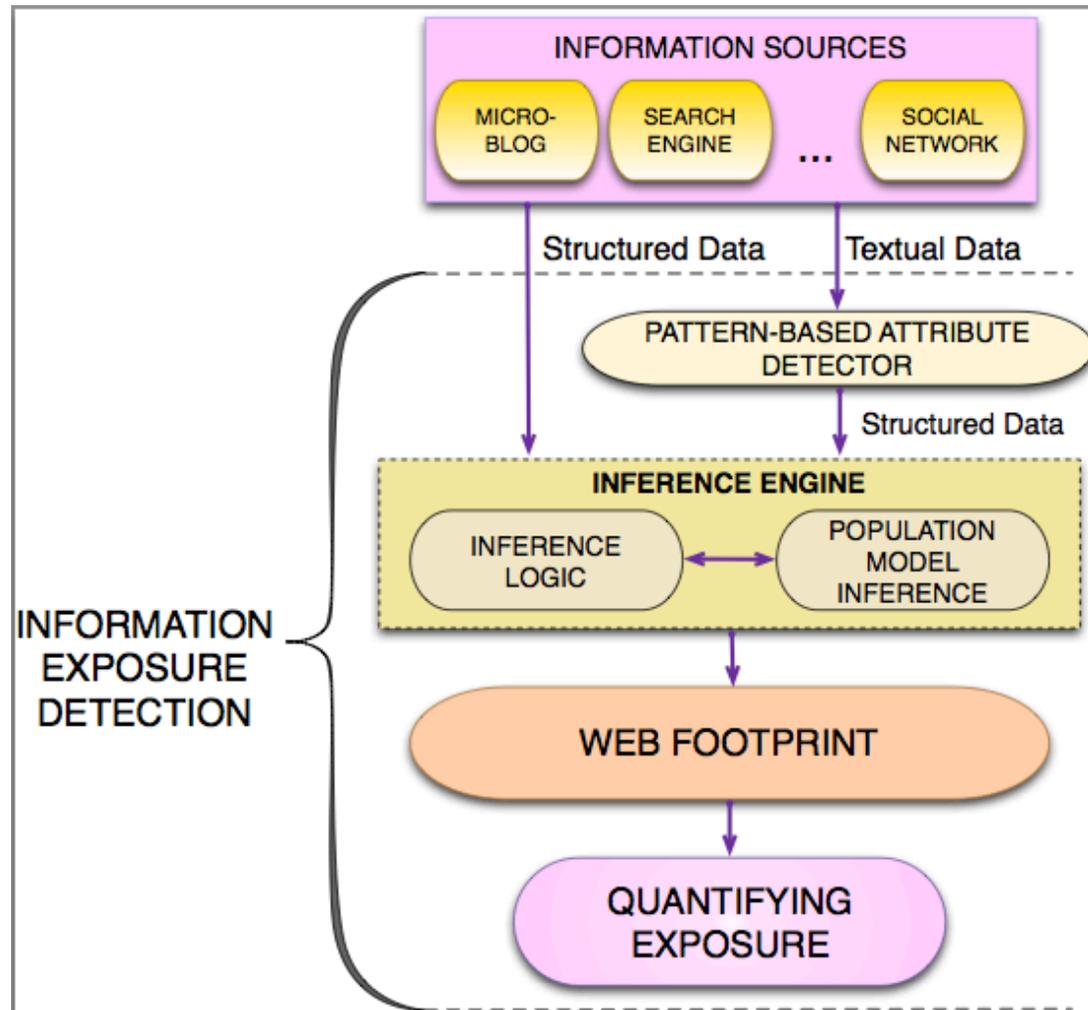
12 attributes per profile



- Association Rule
- Naïve Bayes
- LDA
- Ensemble

[Moore et al., 2013]

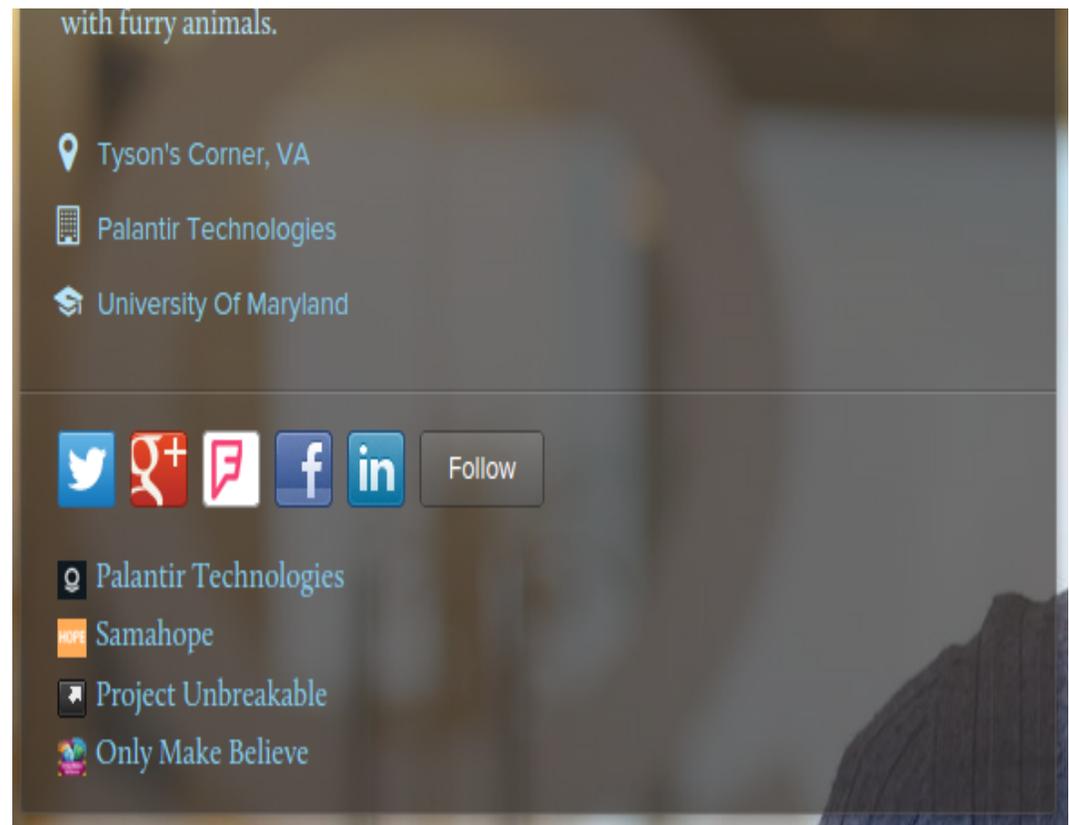
What can be inferred from the population?



Web Footprinting

Experiments for Understanding Public Profiles

- About.me - personal website hosting site
 - Each user can make a custom webpage about themselves
 - Can list links to their social media profiles on multiple websites
- Using their API, we collected 124,497 people's information -> Ground Truth



Creating Web Footprints Using Google+, Foursquare, LinkedIn Profiles

Initial Beliefs (B_{core})	Nbr of True Beliefs	Information Accessibility	Info. Exposure
First Name, Last Name	6	16	0.83
First Name, Last Name, Location	7	11	0.92
First Name, Last Name, Education	10	17	0.85
First Name, Last Name, City	11	16	0.87
First Name, Last Name, Relationship Status	27	38	0.88
First Name, Last Name, Birthday	13	20	0.86
First Name, Last Name, College	11	17	0.87
First Name, Last Name, Gender, Location	6	7	0.9
First Name, Last Name, Gender, Location, City	7	8	0.93
First Name, Last Name, Gender, Location, City, Education	10	11	0.96
F. Name, L. Name, Gender, Loc., City, Edu., Relationship Status	11	12	0.96

[Singh et al., 2015]

Synonyms can be found

Work

Occupation
Student, Teaching Assistant, Research Assistant

Skills
Java, Python, C++, Puns

Employment

Georgetown University
Teaching Assistant, 2012 - present

Google
Software Engineering Intern, 2014 - 2014

Education

Georgetown University
Computer Science, 8 - present

Contact Information

Home

Phone 2023403171

Places



Previously
Bloomington MN - Stamford, CT - Croissy-sur-Seine, France - Washington DC

Basic Information

Gender Male

Birthday September 30, 1992

Relationship Single

Andrew Hian-Cheong

185
connections

Engineering Practicum Intern at Google

Washington D.C. Metro Area | Computer Software

Current Georgetown University Department of Computer Science

Previous Google, Executive Office of the President, Anant

Education Georgetown University

Recommendations 1 person has recommended Andrew

Experience

Research Assistant

Georgetown University Department of Computer Science

January 2014 – Present (1 year 4 months) | Georgetown University

- Research lab focused on analyzing information exposure on public social networking websites
- Duties involved Database management (PSQL) and writing programs to do data collection through public APIs and web scraping
- Wrote algorithms to do probabilistic joins on information gained from social media

Computer Science Teaching Assistant

Georgetown University

June 2012 – Present (2 years 11 months) | Washington DC



GEORGETOWN
UNIVERSITY

Dbpedia

Meronym

Ontology Property Name

owl:sameAs

dbpedia-owl:wikiPageRedirects

dbpprop:parent

foaf:name

dbpprop:nickname

Property Value

http://fr.dbpedia.org/resource/UniversitÃ_de_Georgetown

dbpedia:Georgetown_University,_Washington,_D.C.

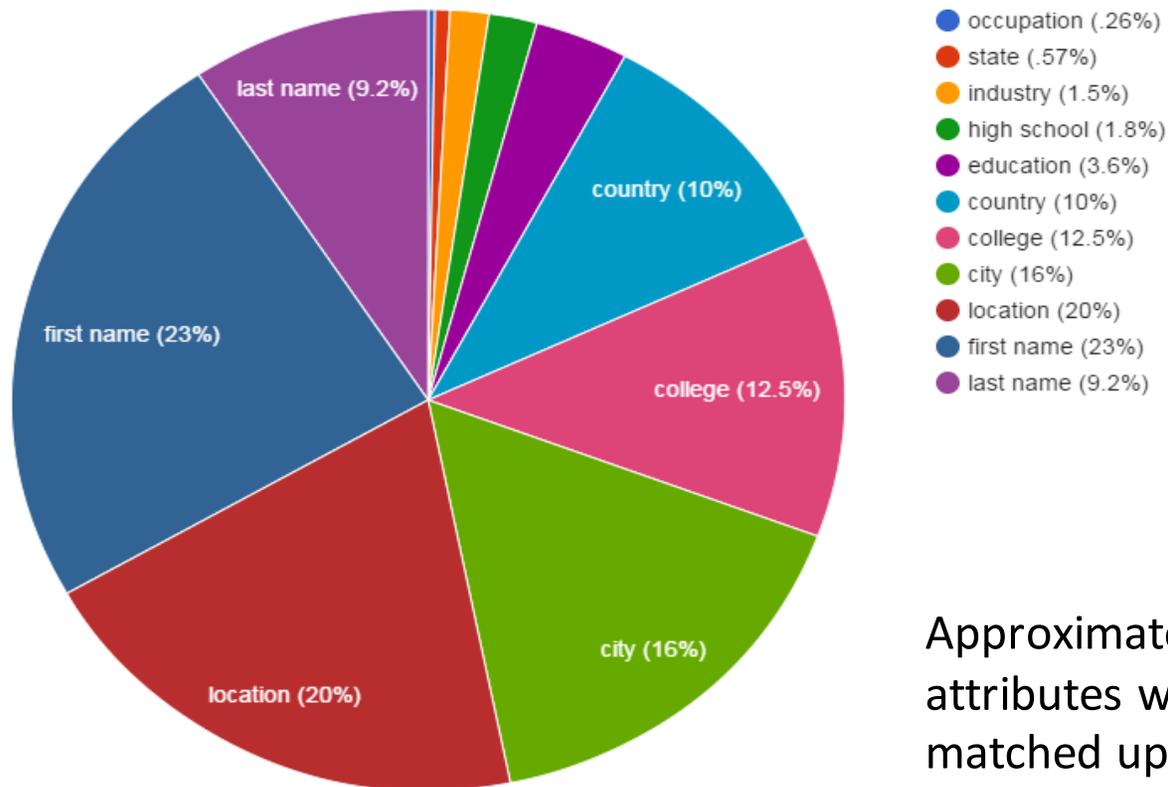
dbpedia:Georgetown_University_Law_Center

Collegium Georgiopolitanum Ad Ripas Potomaci in Marylandia

dbpedia:Georgetown_Hoyas

Synonyms

Using an Ontology



Approximately 8000 attributes were matched up from the ontology

Taking Control of Our Web Identity and Data

1. Keep your public profile professional.
2. Change all your social media account settings that have personal information on them from public to private.
3. Choose your friends wisely – add them selectively.
4. Join groups related to your professional interests.
5. Make it difficult for automated tools to link your accounts, e.g. use different account user names, share different information, etc.
6. Install ad blockers to reduce data about your click through habits.
7. Set your browser to not accept cookies from sites that you have not visited before.

The world around us



DATAFICATION



Data Ethics

- Regulation
 - We need to hold companies to higher standards.
- Data ethics standards
 - We need discussion, debate, and possibly a new discipline.
- Catalog of personal data
 - Individuals should be able to see, correct and/or remove data companies have about them.

[Singh, 2016]

Final Thoughts

- There is a cultural acceptance of sharing private data publicly.
- This is a problem - I have shown you different techniques for generating web footprints. It is too easy!!
- We need new ways to help users understand what data can be determined about them and help them take control of their information.
- We need to pause and debate online privacy and ethical uses of large-scale human behavioral data.
- We need to develop guidelines and regulations that protect users.

We need to take back control of our data.



References

J. Zhu, S. Zhang, L. Singh, H. Yang, and M. Sherr. "Generating Risk Reduction Recommendations to Decrease Identifiability of Public Online Profiles." under submission.

A. Hian-Cheong, L. Singh, M. Sherr, H. Yang. "Semantics and Public Information Exposure Detection." invited.

L. Singh, H. Yang, M. Sherr, A. Hian-Cheong, K. Tian, J. Zhu, and S. Zhang. "Public Information Exposure Detection: Helping Users Understand Their Web Footprints." *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Paris, France: IEEE/ACM, 2015.

L. Singh, H. Yang, M. Sherr, Y. Wei, A. Hian-Cheong, K. Tian, J. Zhu, S. Zhang, T. Vaidya, and E. Asgarli. Helping Users Understand Their Web Footprints. *International Conference on World Wide Web - Companion Proceedings. World Wide Web (WWW)*, Florence, Italy. Poster Paper, 2015.

W. B. Moore, Y. Wei, A. Orshefsky, M. Sherr, L. Singh, H. Yang. "Understanding Site-Based Inference Potential for Identifying Hidden Attributes." *International Conference on Privacy, Security, Risk and Trust*. Alexandria, VA: IEEE Computer Society, 2013.

J. Ferro, L. Singh, M. Sherr. "Identifying individual vulnerability based on public data." *International Conference on Privacy, Security and Trust*. Tarragona, Catalonia, Spain: IEEE Computer Society, 2013.

F. Nagle, L. Singh, and A. Gkoulalas-Divanis. "EWNI: Efficient Anonymization of Vulnerable Individuals in Social Networks." *Pacific Asian Conference on Knowledge Discovery and Data Mining (PAKDD)*. Kuala Lumpur, Malaysia: Springer, 2012.

A. Ramachandran, L. Singh, E. Porter, and F. Nagle. "Exploring re-identification risks in public domains." *Conference on Privacy, Security and Trust (PST)*. IEEE Computer Society, 2012.

The Team & Support

- Faculty:
 - Lisa Singh, Micah Sherr, Grace Hui Yan
- Students & Researchers:
 - Rob Churchill, Kristen Skillman, Kevin Tian, Sicong Zhang, Yanan Zhu
- Alumni:
 - Aditi Ramachandran, Frank Nagle, John Ferro, Yifang Wei, Brad Moore, Andrew Hian-Cheong, Janet Zhu

Support: National Science Foundation

5 Reasons to Join Our Program!

1. We are research active and provide full financial RA support for PhD students for 5 year.
2. We have full and partial scholarships for Master's students.
3. Our courses span applied and theoretical areas of computer science, as well as interdisciplinary areas like data science.
4. We have exceptional job placement in top tech firms, national labs, and government agencies.
5. We have a strong community among students and faculty.

Need more information:

Website: <http://cs.georgetown.edu/>

Graduate Director: Lisa Singh

(singh@cs.georgetown.edu)

Application deadline **April 1!**

