PERSONAL DATA AND PRIVACY IN A NETWORKED WORLD

Kristina Lerman
USC Information Sciences Institute

DSCI 552 – Spring 2021

March 31, 2021

“We’ve lost control of our personal data”

“The current business model for many websites offers free content in exchange for personal data. Many of us agree to this – albeit often by accepting long and confusing terms and conditions documents – but fundamentally we do not mind some information being collected in exchange for free services. But, we’re missing a trick. As our data is then held in proprietary silos, out of sight to us, we lose out on the benefits we could realise if we had direct control over this data, and chose when and with whom to share it. What’s more, we often do not have any way of feeding back to companies what data we’d rather *not* share – especially with third parties – the T&Cs are all or nothing...”

-- Sir Tim Berners-Lee in “Three challenges for the web, according to its inventor,” March 12, 2017

Keeping up with privacy policies

2005 — "only members of groups you belong to"*

2006 — "your school, your specified local area"*

2007 — "networks you allow… (e.g., school,…, friends of friends)"*

2008 — friends list is public by default, list of Pages you like has to be public

2009 — certain default privacy settings to Everyone, "subject to indexing by third party search engines… may be imported and exported by us and others without… limitations"*

2010 — applications you connect to have access to "your and friends' names, profile pictures, gender, connections, and any content shared" with Everyone*

…

*https://www.eff.org/deeplinks/2010/04/facebook-timeline

Personal information on the web

- People reveal private information online
  - Your online profile
    - Birthdate, home town, education, religion, marital status
  - Your Likes and tastes
    - Favorite tv shows, movies, music, food
  - Your status updates
- People reveal private information through their social connections
  - The friends who may reveal private information
  - The groups people join

Privacy in social networks

- Beyond identity and attribute disclosure
  - Link re-identification [Zheleva & Getoor’07]
- Anonymization by removing the profile attributes not enough
  - Structural re-identification [Backstrom et al.‘07]
  - Structural anonymization [Hay et al.’08]
  - Re-identification across networks [Narayanan & Shmatikov’09]
- Differentially-private analysis of graphs
  - [Task & Clifton‘14, Raskhodnikova & Smith‘15]

Privacy in social networks

- Understanding privacy risks
  - Inferring information in social networks [Zheleva'09]
  - Privacy scores [Liu & Terzi'10]
  - Information-sharing models [Kleinberg & Ligett'10]
  - Re-identification through online activity [Kosinski et al, 2014]
- Connecting privacy and fairness
  - FairTest: unwarranted associations framework [Tramer et al.'17]

What is private information?

- People may choose **not to reveal** certain pieces of information about their lives
  - Private information includes
    - Sexual orientation, age, income, …
    - Political affiliation
    - Intelligence
    - **Personality**
      - “the combination of characteristics or qualities that form an individual's distinctive character.”
  - This information might be **predicted**
    - From unrelated publicly disclosed information

Big 5 Personality Traits

- Openness to Experience/Intellect
  - High scorers tend to be original, creative, curious, complex; Low scorers tend to be conventional, down to earth, narrow interests, uncreative.
- Conscientiousness
  - High scorers tend to be reliable, well-organized, self-disciplined, careful; Low scorers tend to be disorganized, undependable, negligent.
- Extraversion
  - High scorers tend to be sociable, friendly, fun loving, talkative; Low scorers tend to be introverted, reserved, inhibited, quiet.

Big 5 Personality Traits

- Agreeableness
  - High scorers tend to be good natured, sympathetic, forgiving, courteous; Low scorers tend to be critical, rude, harsh, callous.
- Neuroticism
  - High scorers tend to be nervous, high-strung, insecure, worrying; Low scorers tend to be calm, relaxed, secure, hardy.

The Big 6 Personality Test

https://www.truity.com/test/big-five-personality-test

I see myself as someone who...

1. ...Is talkative
Strongly Disagree 1 2 3 4 5 Strongly Agree

2. ...Tends to find fault with others
Strongly Disagree 1 2 3 4 5 Strongly Agree

3. ...Does a thorough job
Strongly Disagree 1 2 3 4 5 Strongly Agree

4. ...Is depressed, blue
Strongly Disagree 1 2 3 4 5 Strongly Agree

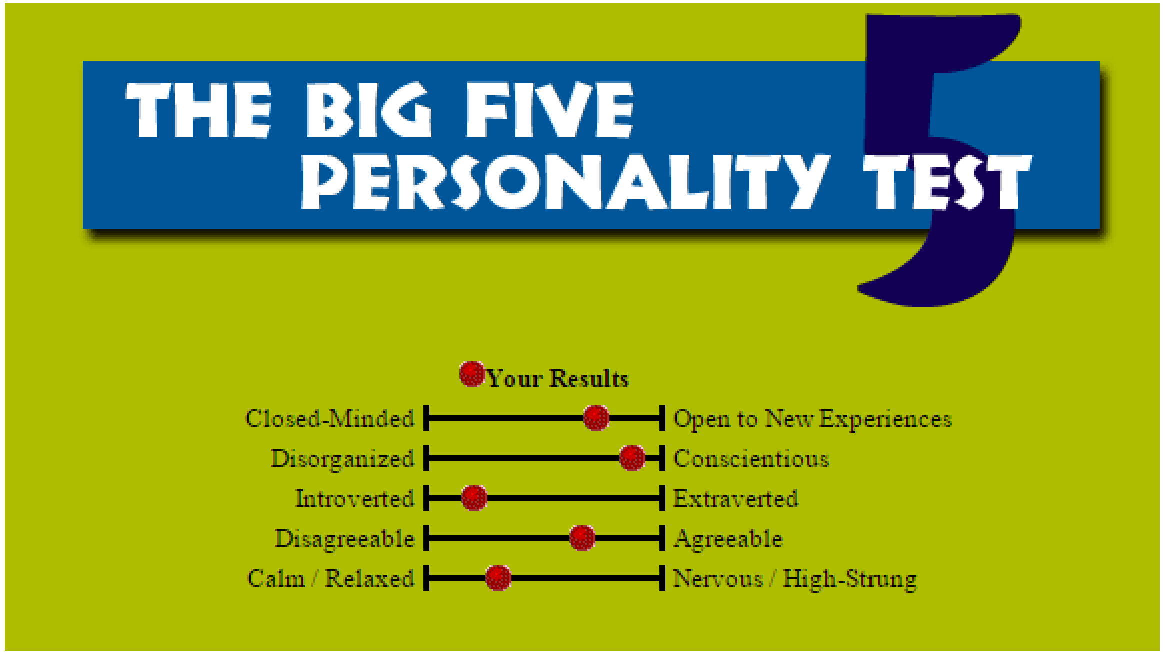
5. ...Is original, comes up with new ideas
Strongly Disagree 1 2 3 4 5 Strongly Agree

6. ...Is reserved
Strongly Disagree 1 2 3 4 5 Strongly Agree

7. ...Is helpful and unselfish with others

Big 5 Personality Traits

https://www.truity.com/test/big-five-personality-test



Predicting personality traits with social media

- Can a user's personality be predicted from his/her publicly available social media profile information?
  - Through self-description, status updates, interests etc. much of the user's personality is reflected
- This paper describes the type of data collected, their methods of analysis and the results of predicting personality traits through machine learning
- Implications to social media design, interface design and other applications such as online marketing and advertising

[Golbeck et al (2011) "Predicting Personality Traits with Social Media", in CHI.]

Data Collection

- Created a Facebook application to administer “Big Five” personality survey to 279 users. Collected their profile statistics:

1. Structural Features
   - User’s ego-centric network size, density
2. Personal Information
   - user name
   - gender
   - birthday
   - relationship status
   - religion
   - education
   - hometown

Data Collection (cont.)

1. Activities and Preferences
   – lists of favorite, such as movies, music etc.
2. Language Features
   – This all the text that the users shared through
     • “About Me”, “blurb” text and status updates.
   – Of the 279 users, only 167 had enough words to be actually used in this data set.
   – Used LIWC to analyze the text producing statistics on 81 different features of text in the 5 categories.

Personality-features correlation

Table 2: Pearson correlation values between feature scores and personality scores. Significant correlations are shown in bold fo $p < 0.05$. Only features that correlate significantly with at least one personality trait are shown.

	Open.	Consc.	Extra.	Agree.	Neuro.
Linguistic Features					
Swear Words	0.006	**-0.171**	0.032	-0.084	-0.120
Social Processes (e.g. Mate, talk, they, child)	0.010	**0.264**	0.091	-0.022	-0.142
Human Words (e.g. baby, man)	0.078	**0.203**	0.070	-0.050	-0.062
Affective Processes (e.g. Happy, cried, abandon)	0.105	-0.009	0.136	**0.203**	0.038
Positive Emotions (e.g. Love, nice, sweet)	0.052	0.045	0.117	**0.167**	-0.013
Anxiety Words (e.g. Worried, fearful, nervous)	0.044	-0.150	0.008	0.101	**0.192**
Perceptual Processes (e.g. Observing, heard, feeling)	-0.040	**-0.195**	**-0.163**	-0.027	0.096
Seeing Words (e.g. View, saw, seen)	0.060	**-0.227**	-0.112	0.013	0.067
Biological Processes (e.g. Eat, blood, pain)	-0.014	0.042	0.038	**0.154**	0.067
Ingestion Words (e.g. Dish, eat, pizza)	-0.098	-0.050	0.029	0.031	**0.207**
Work Words (e.g. Job, majors, xerox)	0.134	0.096	**0.154**	0.048	-0.044
Money Words(e.g. Audit, cash, owe)	**-0.161**	0.024	0.012	-0.006	0.029
Structural Features					
Number of Friends	-0.094	-0.078	**0.186**	0.013	-0.069
Egocentric Network Density	**-0.152**	0.050	**-0.224**	0.059	0.032
Activities and Preferences					
Activities (char length)	0.115	0.095	**0.188**	0.066	-0.145
Favorite Books (char length)	**0.158**	-0.093	0.019	0.082	0.028
Personal Information					
Relationship Status (none listed,single, not single)	0.093	0.071	**0.194**	0.040	-0.036
Last Name length in characters	0.012	-0.111	0.000	-0.044	**0.184**

Personality and Social Media Profile

Weak correlations between user’s profile and personality scores:

- *Conscientiousness*
  - negatively correlated with the frequency of swear words as well as words that described perceptual processes.
  - positively correlated with words concerning social processes.
- *Agreeableness* correlated positively with words describing feelings and
- *Neuroticism* correlated positively with words describing anxiety
- Length of the reported activities and interests correlated with *extraversion* and *openness*

Personality and Social Media Profile

- Extroverts
  - Larger networks (more friends)
  - Less dense network structure.

- Correlation between neuroticism and length of the user's last name.

- Woman were more conscientious, agreeable and neurotic than men.
- Users who chose to share a URL to an external website were more open than those who didn't.

Differences between subpopulations

Value	Open.	Consc.	Extra.	Agree.	Neur.
Male	3.841	**3.313**	3.145	**3.638**	**2.680**
Female	3.671	**3.582**	3.476	**3.806**	**2.996**
p	*0.101*	*0.018*	*0.018*	*0.095*	*0.018*
No Website	**3.710**	3.495	3.264	3.697	2.900
Website	**4.010**	3.498	3.508	3.773	2.770
p	*0.003*	*0.978*	*0.071*	*0.382*	*0.275*

Predicting Personality

- Predicted personality trait score based on 74 user features
  - using regression with 10-fold cross-validation
  - The Mean Absolute Prediction Error around 11%

Factor	M5′Rules	Gaussian	M5′Rules	Gaussian
	MAE		Correlation Coefficient	
Open.	0.099	0.117	0.653	0.179
Consc.	0.104	0.117	0.595	0.094
Extra.	0.138	0.124	0.553	0.050
Agree.	0.109	0.117	0.482	0.150
Neuro.	0.127	0.117	0.531	0.106

Predicting private traits from digital records of human behavior

[Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *PNAS*]

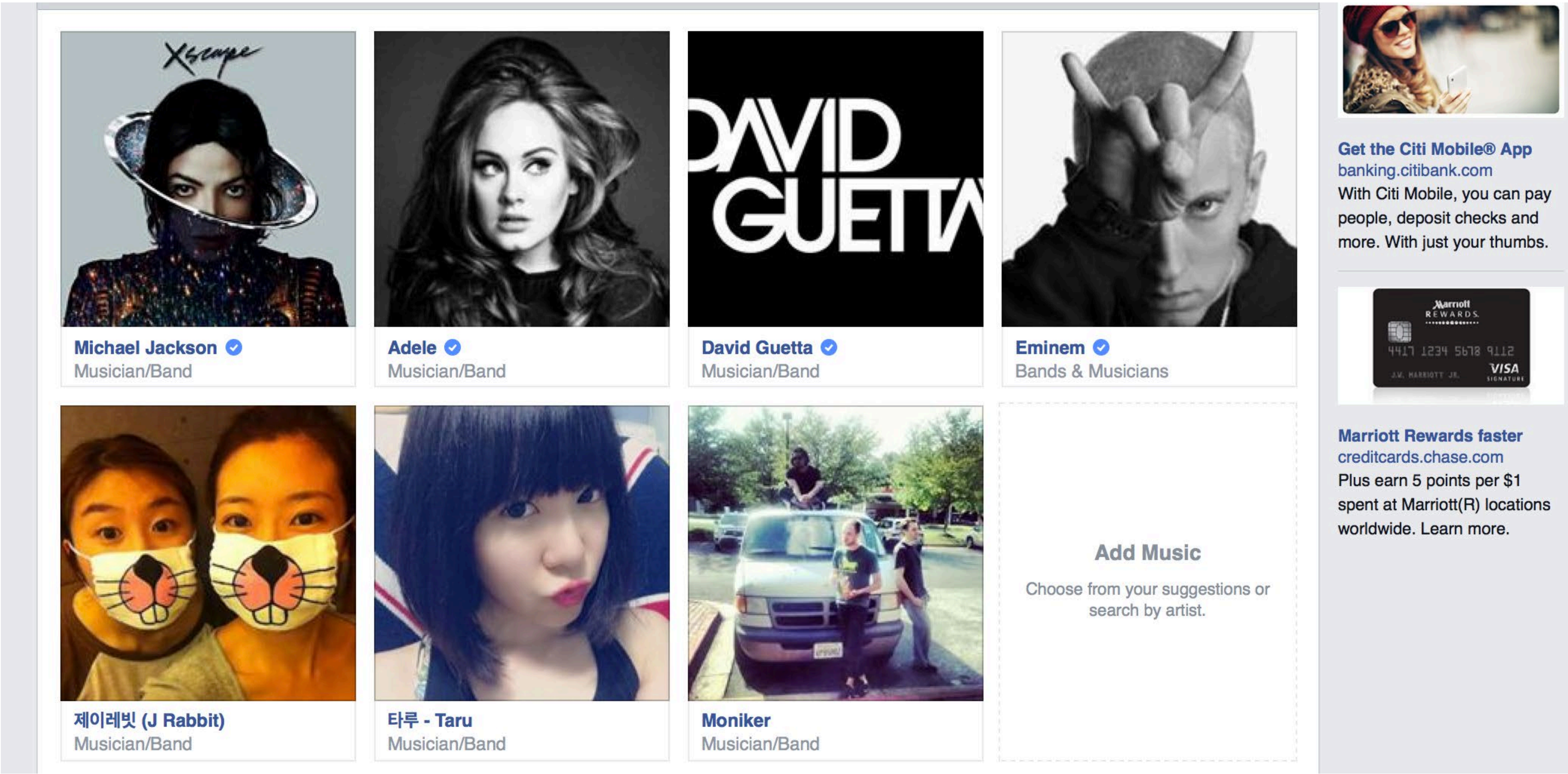
- Facebook users employ ‘likes’ to express their positive attitudes towards online content and entities
  - Publicly available
  - Similar to Web search queries, Web browsing history, and credit card purchases.
- Statistical method to automatically and accurately estimate personal attributes from Facebook ‘likes’

Simon's Likes on Sports

What kinds of sports does Simon like?

Simon's Likes on Music

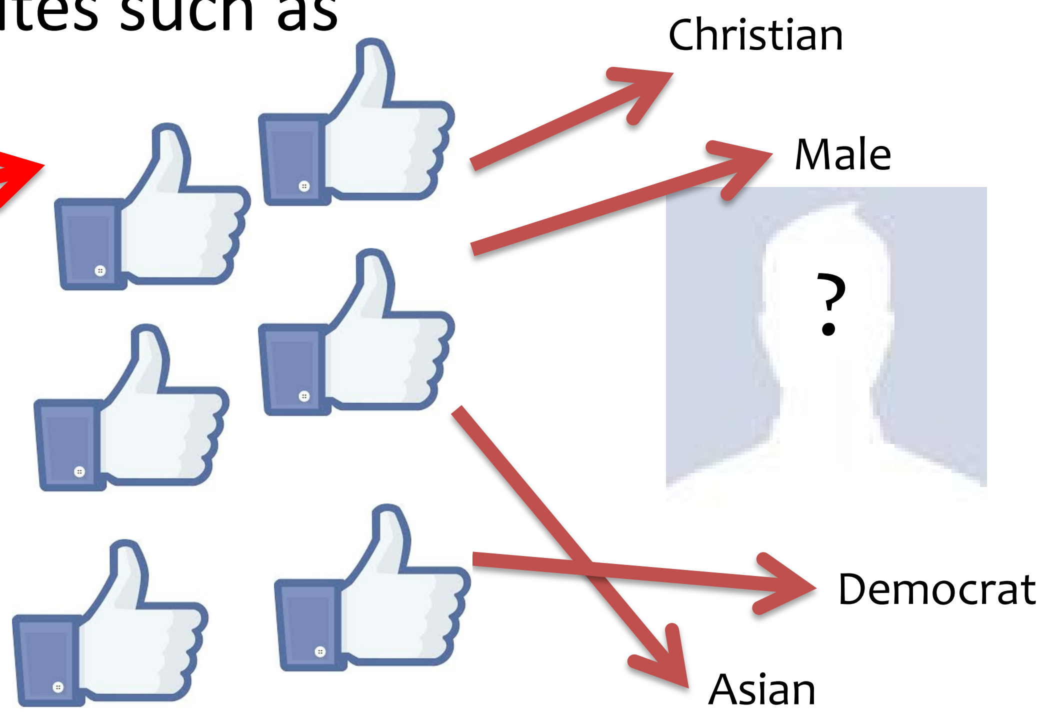
Can you guess Simon's music tastes ?



Facebook Likes →(?) Personal Traits (Digital Fingerprinting)

Likes in Facebook can be used to automatically and accurately predict with high probability for the highly sensitive personal attributes such as

- gender
- ethnicity
- religious
- political views
- personal traits
- intelligence
- happiness



Main Questions

- Can we ***automatically*** estimate a wide range of personal attributes that people would typically consider to be private ?
- Can we ***accurately*** estimate a wide range of personal attributes that people would typically assume to be private ?
- Can we use only **Facebook Likes** ?

Design of User Study

- **Step1.** Selected the private and intrusive attributes
  - sexual orientation, ethnic origin, political views, religion
  - personality, intelligence, satisfaction with life (SWL), substance use (alcohol, drugs, cigarettes), whether an individual's parents stayed together until the individual was 21 years old
  - basic demographic attributes such as age, gender, relationship status, and size and density of the friendship network

Design of User Study

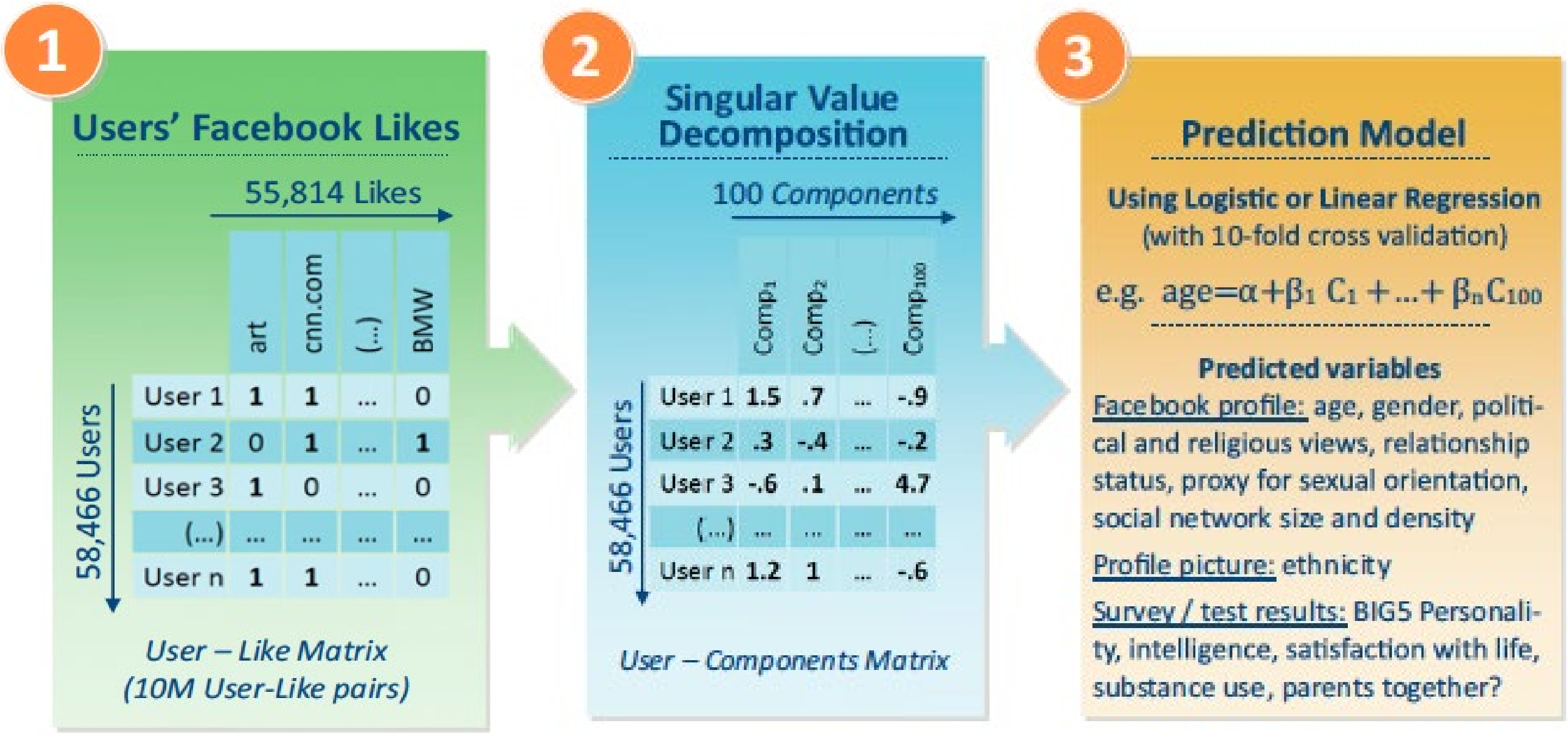
- **Step2.** Data Collections

They obtained data set from 58,000 volunteers who share their 10M Facebook Likes, profiles, and took psychometric surveys.

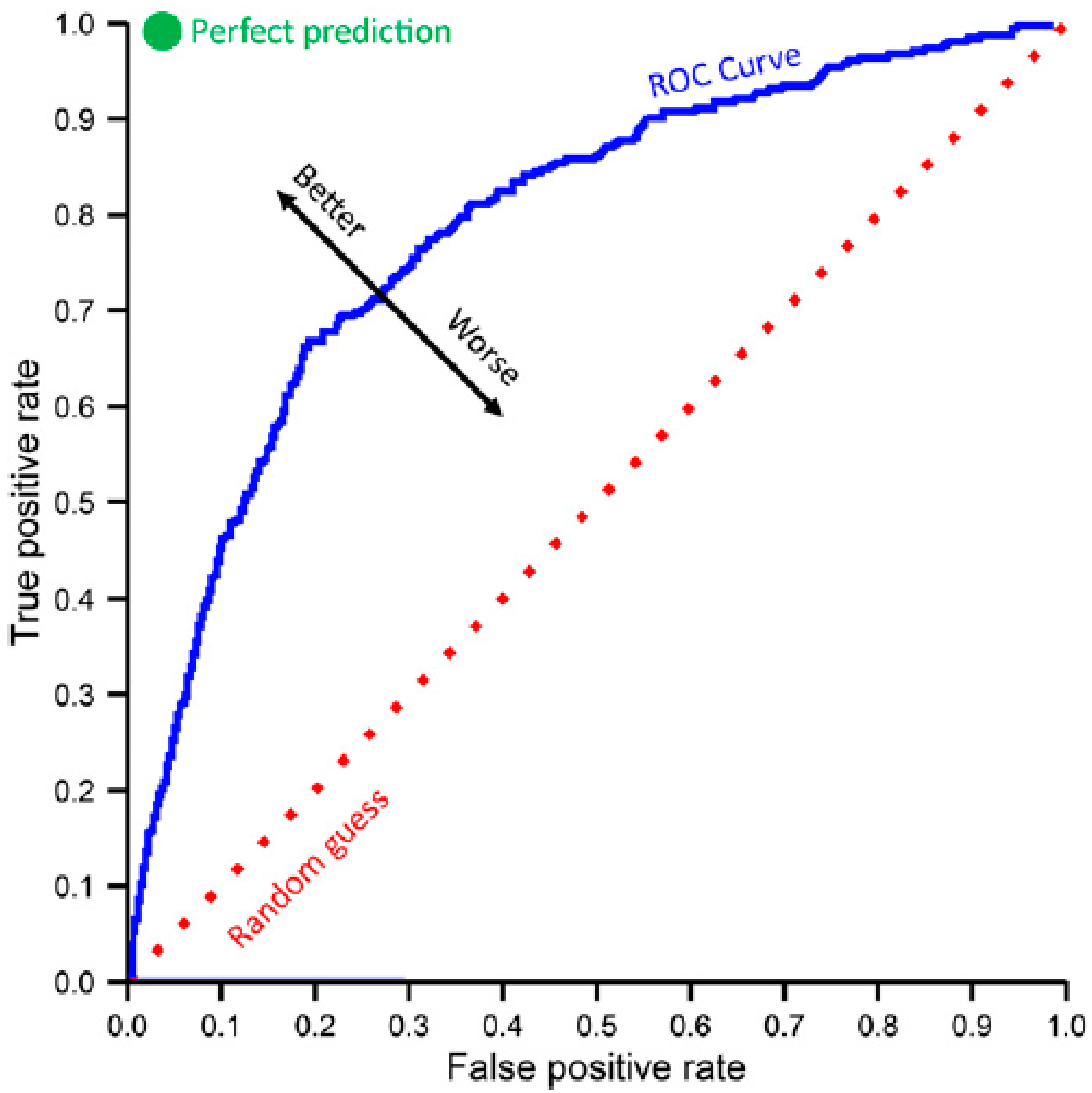
Methodology/Models

- **Truncated SVD for computation and data analysis**
  - Used top 30~100 SVD components, depending on the attributes.
- **Logistic/linear regression model for Prediction**
  - **Gender = $\alpha_{Gender} + \beta_1C_1+\beta_2C_2+\ldots+\beta_iC_i+\ldots\beta_{30}C_{30}$**
  - **Age = $\alpha_{Age} + \beta_1C_1+\beta_2C_2+\ldots+\beta_iC_i+\ldots\beta_{100}C_{100}$**
- **Evaluations**
  - Pearson product–moment correlation coefficient between the actual and predicted values to measure the accuracy of prediction.

Prediction with Facebook likes

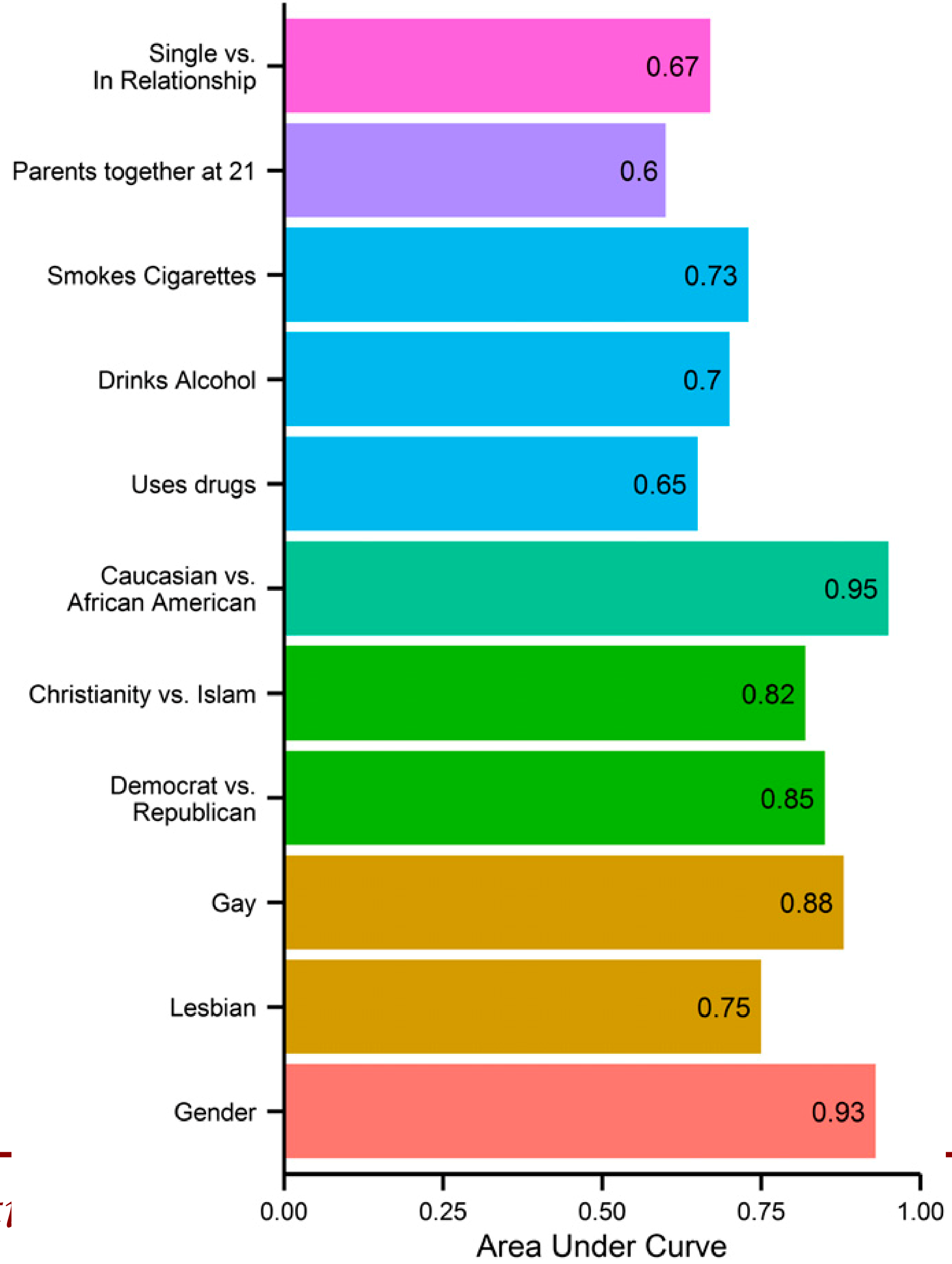


ROC Curve

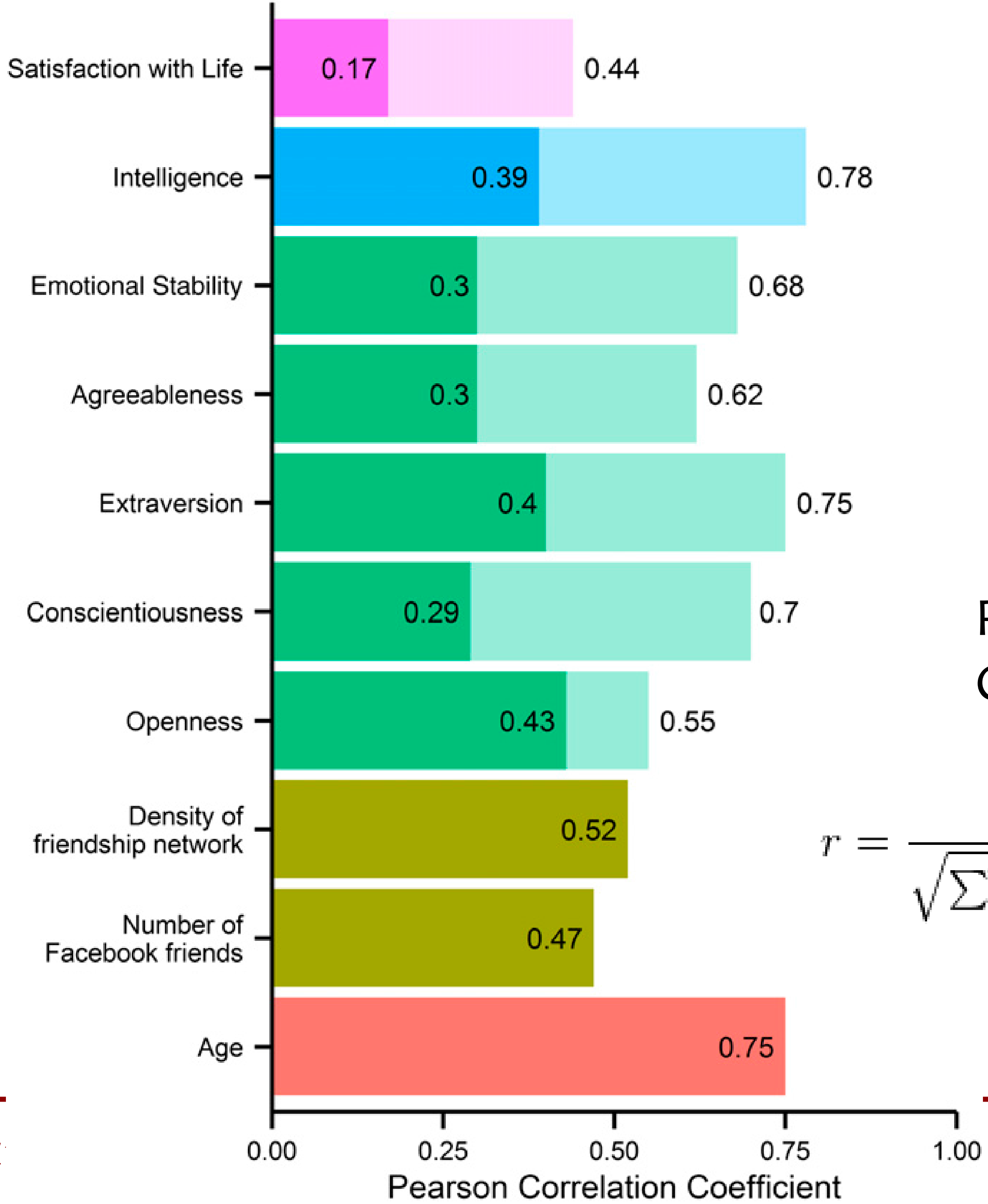


Perfect prediction
ROC Curve
Better
Worse
Random guess
True positive rate
False positive rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Classification accuracy for binary personal attributes



USC Viterbi School of Engineering

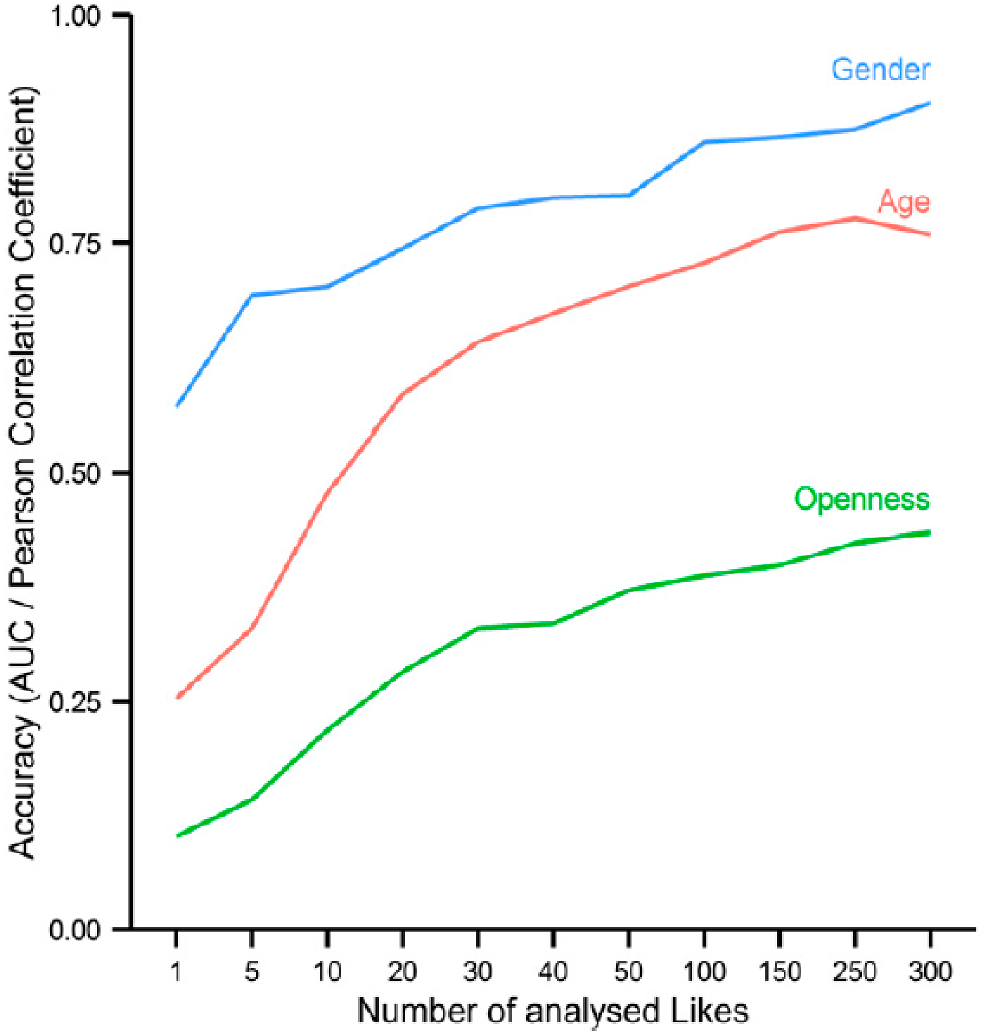


Accuracy of regression for numeric attributes

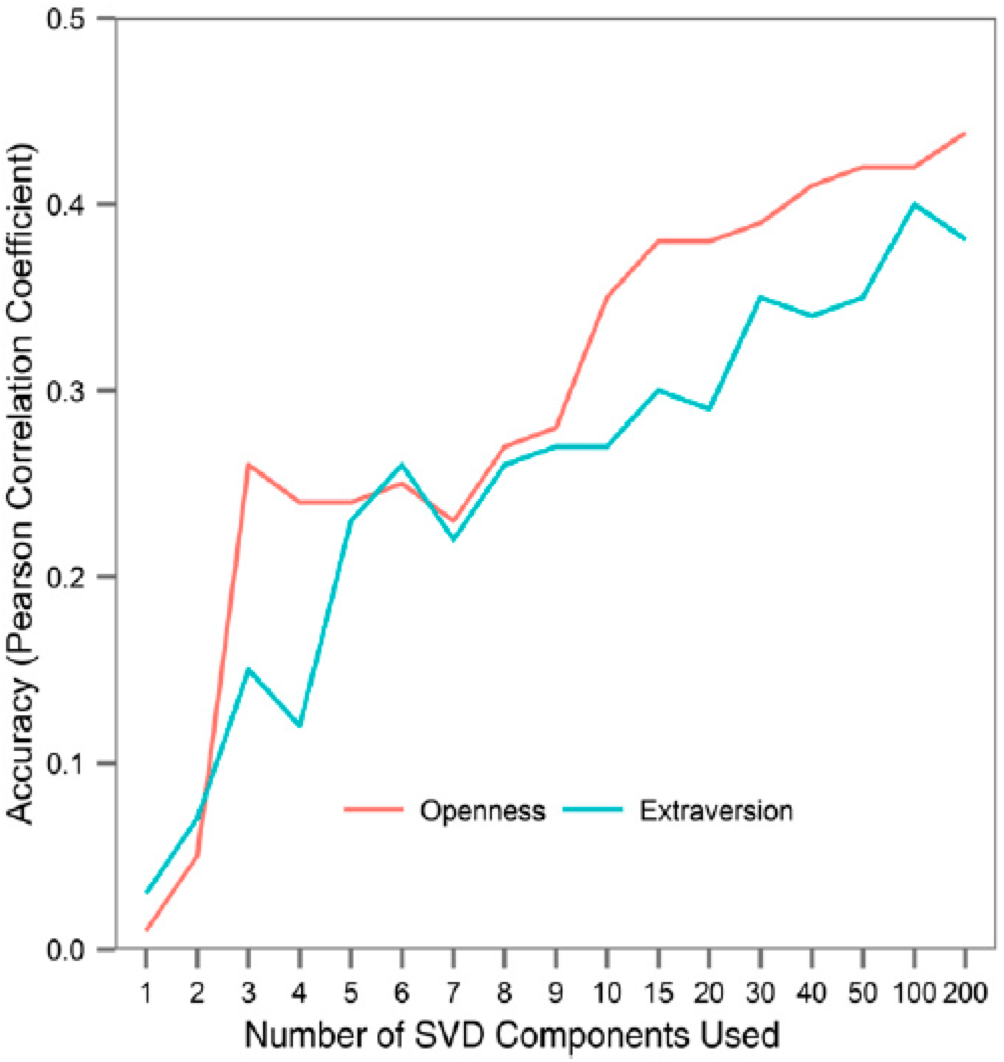
Pearson Correlation Coefficient r

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Prediction accuracy vs volume of data (#likes)



Prediction accuracy vs #SVD components



Most predictive likes

IQ	***High*** The Godfather Mozart Lord Of The Rings	***Low*** Sephora I Love Being A Mom Harley Davidson
Extraversion	***Outgoing & Active*** Beerpong Dancing Cheerleading Chris Tuker Theatre	***Shy & Reserved*** Video Games Programming Role Playing Games Manga Minecraft
Openness	***Liberal & Artistic*** Oscar Wilde Leonardo Da Vinci Leonard Cohen Bauhaus	***Conservative*** NASCAR ESPN2 The Bachelor Justin Moore

Smoking: Yes	Smoking: No
Cradle Of Filth	That Spider Is More Scared Than U Are Oh Really Did It Tell U That
Under Armour	Honda
Slayer Band	Move Out Of The Way Children I've Been Waiting 11 Years To See Toy Story 3
Inbox 1 Makes Me Nervous	FBI Open The Door No Its Cool When You Break In
Dimebag Darrell	How To Make A Girl Smile <3<3
Rob Zombie	The Desk Able To Protect You From Fire Earthquakes And Nuclear War
I Always Accept The Terms And Conditions Without Reading Them	When Your Fortune Cookie Knows What's Up
I Bottle Everything Up Until I Finally Snap	Rocky
Life Is Better In Summer	When Little Kids Are Chasing Me I Run Slow So They Think They're Fast
Screwing Around In Walmart	I Drop My I-pod Then My Headphones Save Its Life

Age	*Old*	Cup Of Joe For A Joe	Walt Disney Records	*Young*
		Coffee Party Movement	Body By Milk	
		Dr Mehmet Oz	Harperteen	
		Fixit And Forgetit	J Bigga	
		The Closer	Because I Am A Girl	
		Joyce Meyer Ministries	I Hate My Id Photo	
		Proud To Be A Mom	293 Things To Do In Class When You Are Bored	
		Freedomworks	Dude Wait What	
		Small Business Saturday	JCP Teen	
		Fly The American Flag		
Friends	*Many*	Mojo-Jojo	The Dark Knight	*Few*
		Biology	In'n'out Burger	
		Dollar General	Hard Rock	
		Hillary	Honey, Where Is My Supersuit	
		106 & Park	Hating ICP	
		Jennifer Lopez	Minecraft	
		Paid In Full	Iron Maiden	
		Yo Gotti	Walking With Your Friend & Randomly Pushing Them Into Someone/Something	
		The Dollar You Are Holding Could've Been In A Stripper's Butt Crack		

Agreeableness	*Cooperative*	Compassion International Logan Utah Jon Foreman Redeeming Love Pornography Harms The Book Of Mormon Circles Of Prayer Go To Church Christianity Marianne Williamson	I Hate Everyone I Hate You I Hate Police Friedrich Nietzsche Timmy South Park Atheism / Satanism Prada Sun Tzu Julius Caesar Knives	*Competitive*
Emotional Stability	*Neurotic*	Sometimes I Hate Myself Emo Girl Interrupted So So Happy The Addams Family Vocaloid Sixbillionsecrets.com Vampires Everywhere Kurt Donald Cobain Dot Dot Curve	Business Administration Getting Money Parkour Track & Field Skydiving Mountain Biking Soccer Climbing Physics / Engineering 48 Laws Of Power	*Calm & Relaxed*

Conclusions

- Many private attributes can be inferred from people's publicly accessible activity, such as Facebook likes
  - Some attributes are remarkably easy to predict… even a single like can lead to significant improvement over random prediction
- This can have benefits, but also risks with respect to privacy

The New York Times

How Trump Consultants Exploited the Facebook Data of Millions

By Matthew Rosenberg, Nicholas Confessore and Carole Cadwalladr

March 17, 2018

2081

Leer en español

(After this story was published, Facebook came under harsh criticism from lawmakers in the United States and Britain. Read the latest.)

LONDON — As the upstart voter-profiling company Cambridge Analytica prepared to wade into the 2014 American midterm elections, it had a problem.

Facebook Halts Ad Targeting Cited in Bias Complaints

By Noam Scheiber and Mike Isaac   March 19, 2019

After years of criticism, Facebook announced on Tuesday that it would stop allowing advertisers in key categories to show their messages only to people of a certain race, gender or age group.

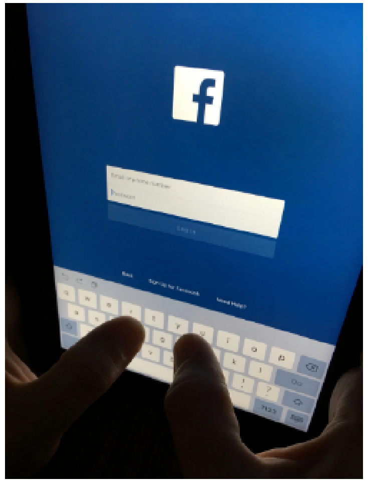
The company said that anyone advertising housing, jobs or credit — three areas where federal law prohibits discrimination in ads — would no longer have the option of explicitly aiming ads at people on the basis of those characteristics.

Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens

By Kevin Granville

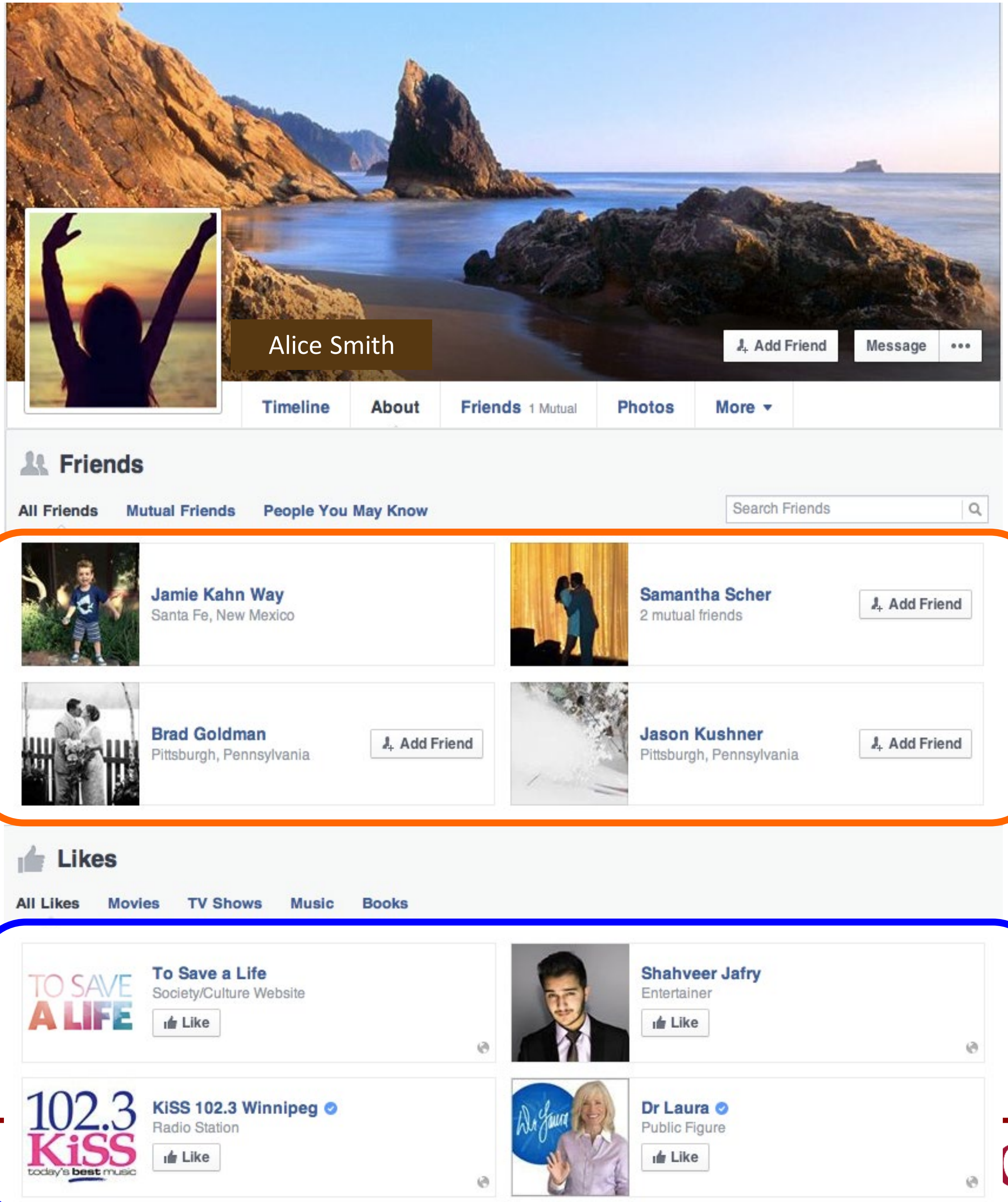
March 19, 2018

Leer en español



Cambridge Analytica, a political data firm hired by President Trump's 2016 election campaign, gained access to information on 50 million Facebook users as a way to identify the personalities of American voters and influence their behavior.
Elise Amendola/Associated Press

Privacy in social networks



Social links

Group affiliations

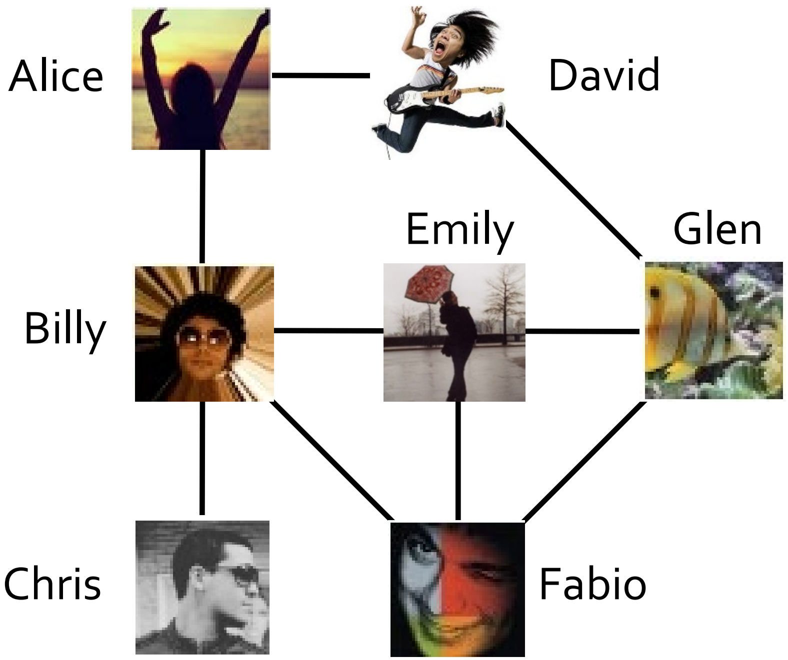
[slides courtesy of Elena Zheleva]

- Group affiliations often cannot be hidden!
- Hypothesis: you can predict private attributes based on public information
- Assumptions:
  - An online social network with public AND private profiles
  - Social links and group affiliations are always public
  - Adversary's goal: to predict attributes in private profiles
  - Adversary can build probabilistic models using public data

From networks to graphs

Social network

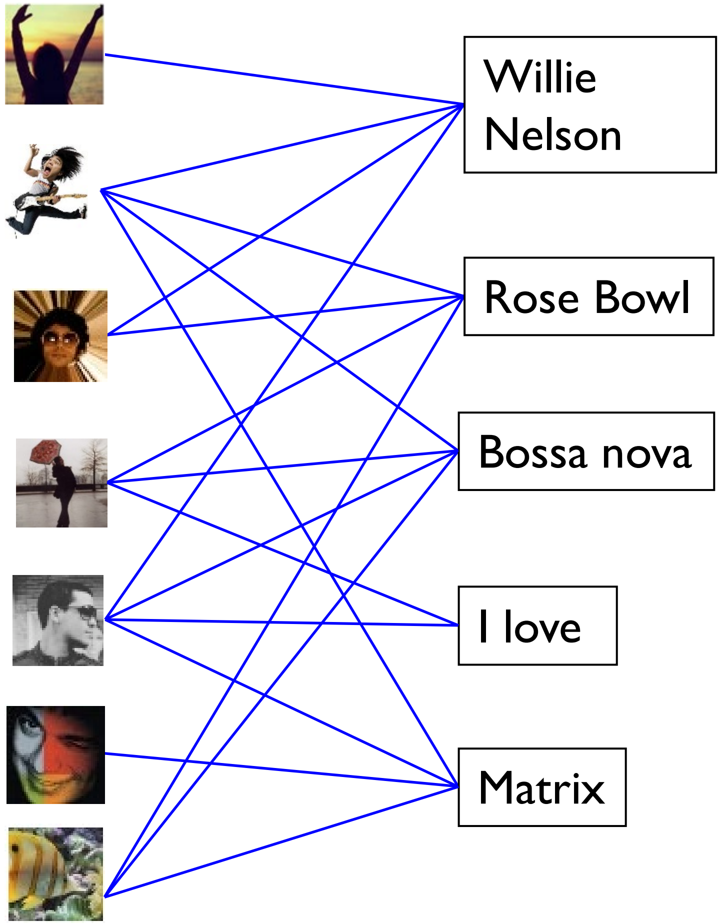
w/ friendship links



- Large, high-dimensional
- Multi-modal
- Multi-relational

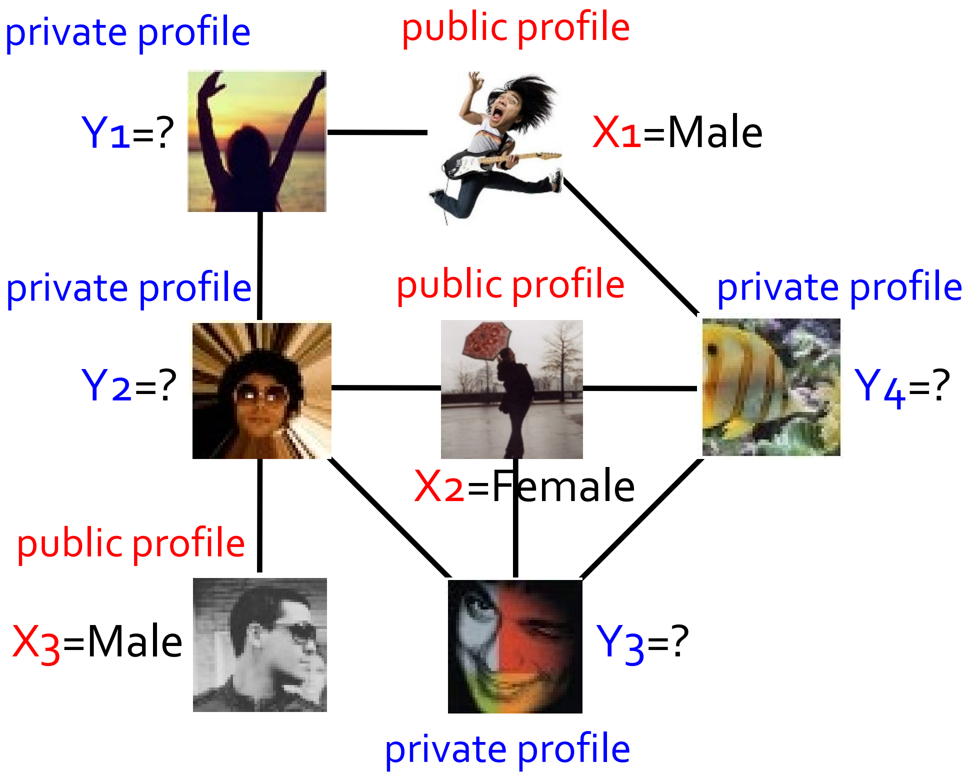
Affiliation network

w/ group links

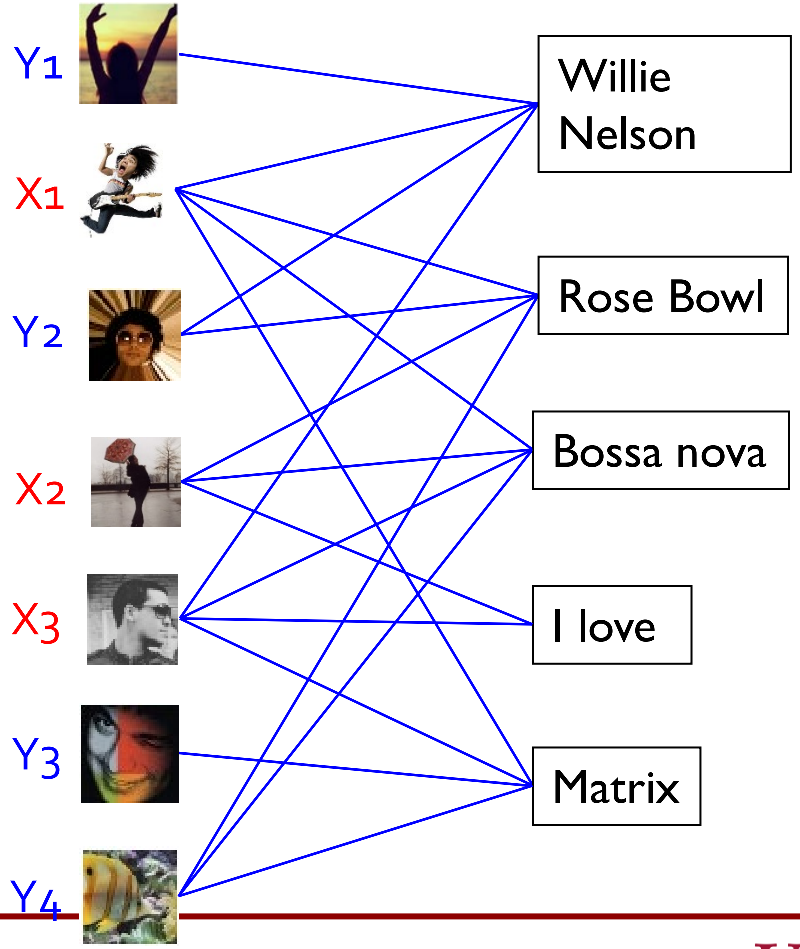


Sensitive attribute disclosure

- If an adversary is able to determine the value of a user attribute that the user intended to stay private



$$\hat{y}_M = \underset{y_i}{\operatorname{argmax}} P_M(y_i \mid X; G)$$



Collective classification*

- Given a set of trait labels (e.g., liberal or not), label the other objects in the network accurately

Is  liberal?

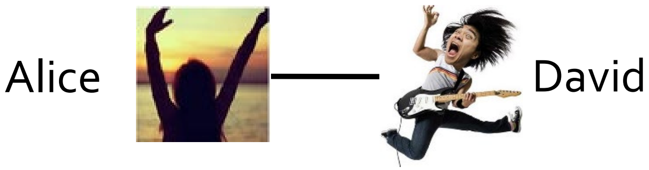
No links

Traditional ML: independent traits

Gender	Location	Liberal
Male	DC	Yes
Female	DC	?

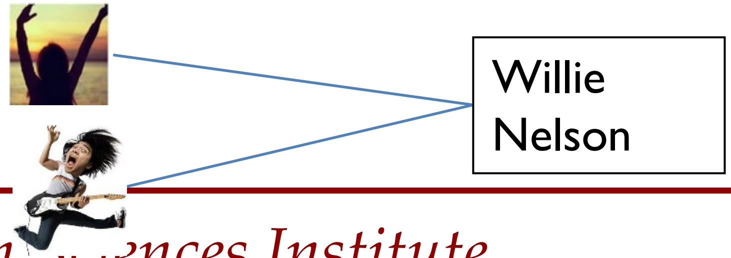
Social links

Friends <-> common trait



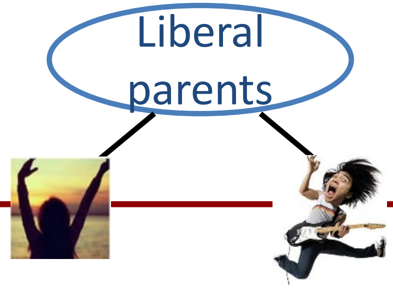
Interest group links

Affiliation <-> common trait



Latent links

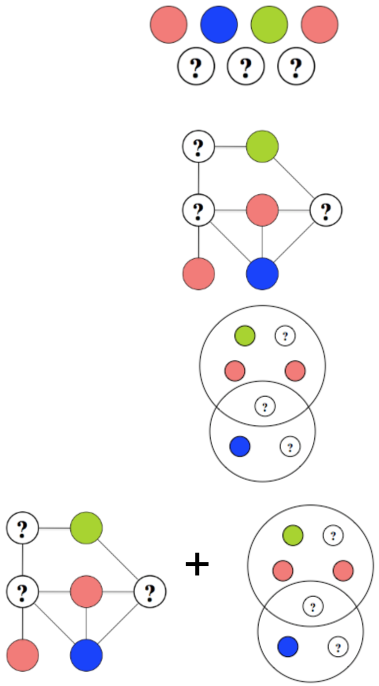
Trait due to common unobserved factor



Collective Classification in Network Data. Sen et al. AI Magazine 2008.

Collective classification

- Assume adversary can apply a probabilistic model M to predict the sensitive attribute

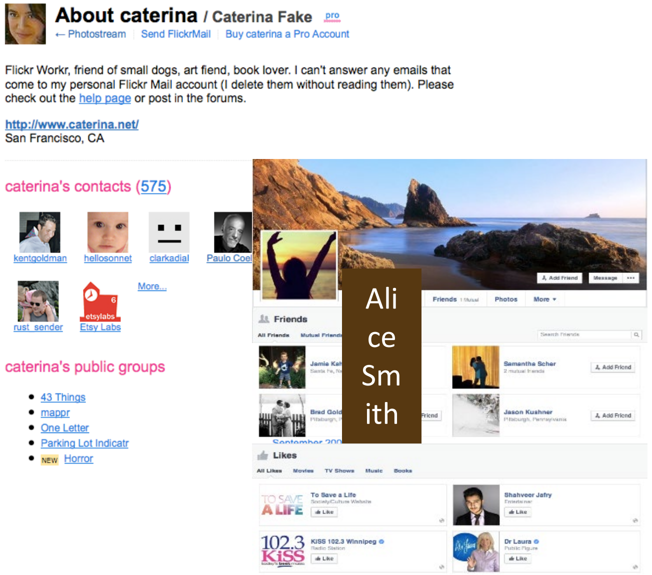


  - No links: BASIC
    - Majority label
  - Social links: AGG, CC, BLOCK, LINK
    - Feature space size: number of users
  - Group links: CLIQUE, GROUP, GROUP*
    - Feature space size: number of groups
  - Both link types: LINK-GROUP
    - Feature space size: number of users and groups

[Zheleva, Getoor. (2009) To Join or not to Join: the Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *WWW*.]

Data and experimental setup

- Flickr: snowball sample
  - ~9,000 profiles, 1 million links, 50,000 groups
  - sensitive: location (55 values)

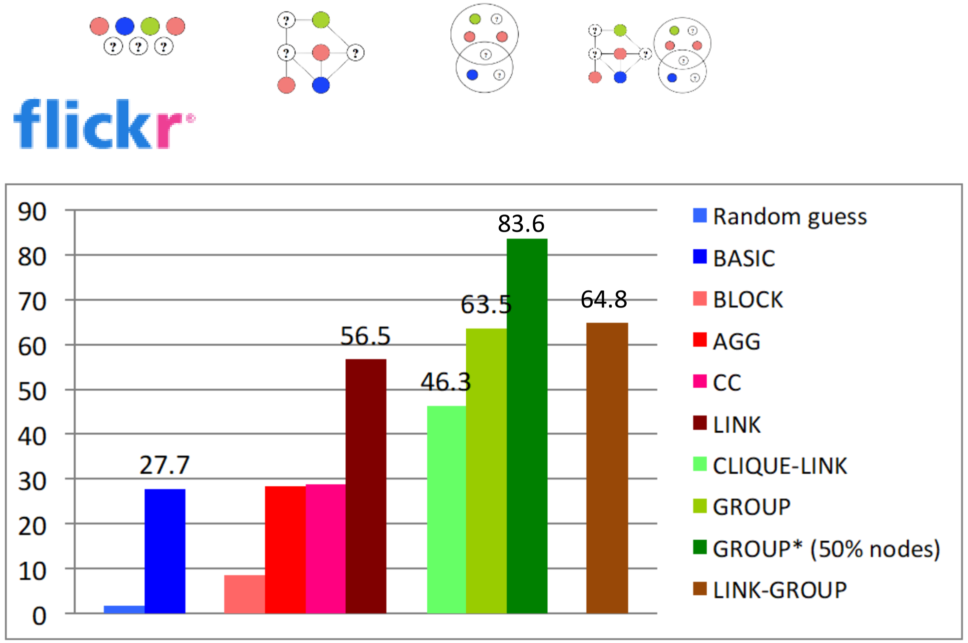


- Facebook: all freshmen (Harvard)
  - ~1,600 profiles, 86,000 links, 3,000 groups
  - sensitive: gender (2) and political views (6)

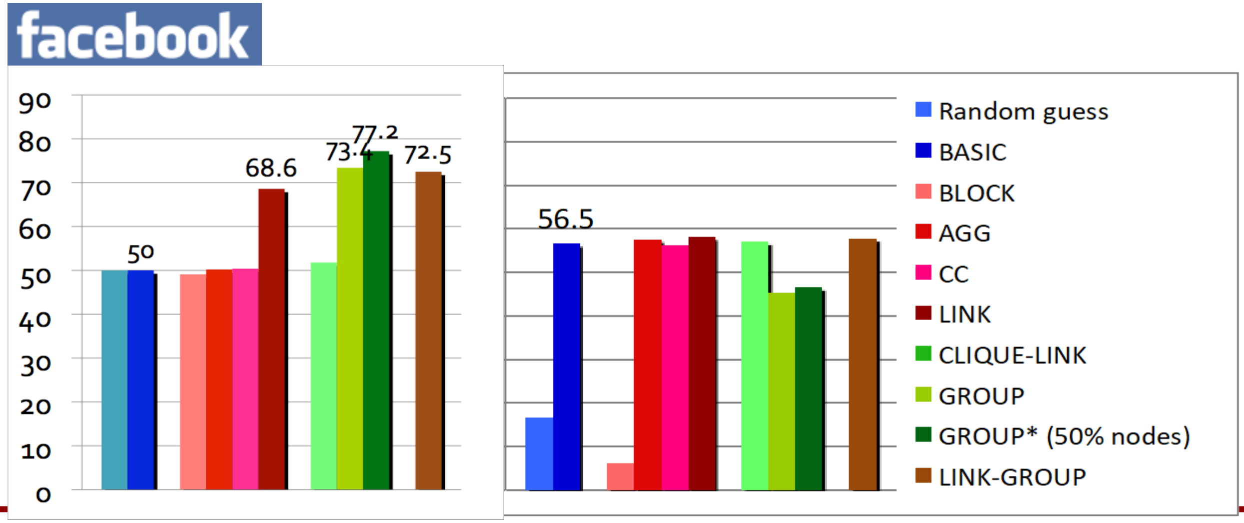
- Assign each profile to be public with probability=50%
  - Train model on public profile data (10 different models)
  - Predict on private profiles

Findings

- Affiliation networks have a high potential for leaking private information
  - More informative than friendship network
  - Homogeneous groups most informative



Country



Gender

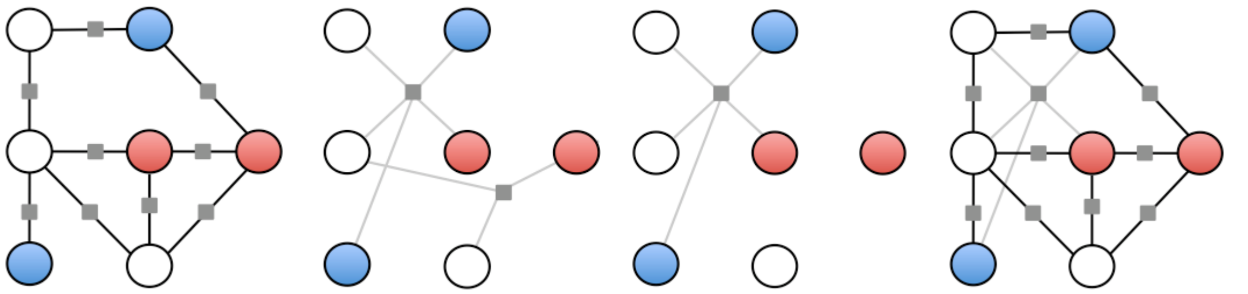
Political views

Increasing model complexity

- Predict "gender" in the Facebook dataset

Links as structural statistical dependencies

Links as features



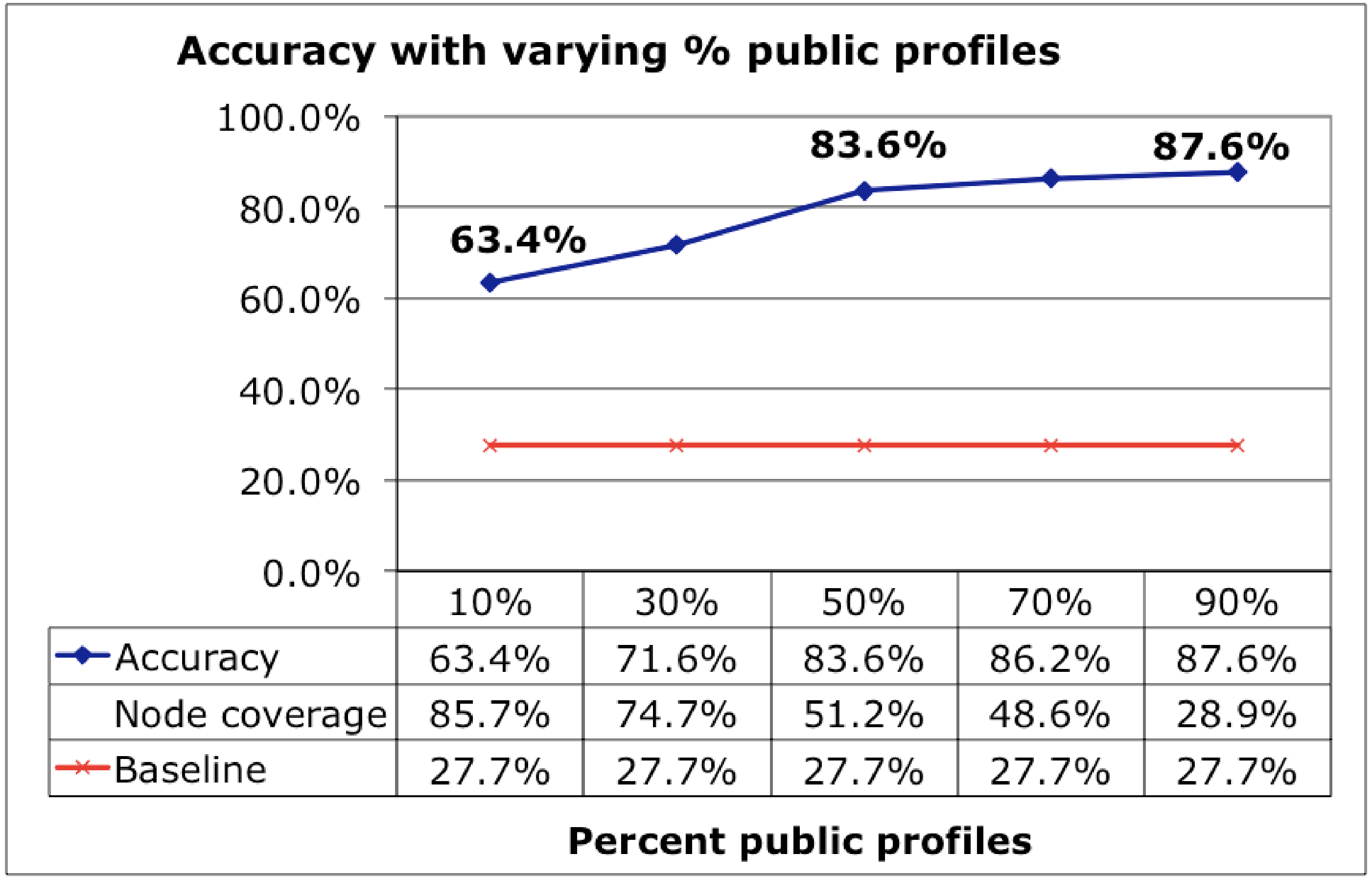
Features	LR	pMRF	hoMRF-AG	hoMRF-SG	hoMRF-SG-AL
Social links	64.06%	64.31%	69.13%	69.22%	69.22%
Groups	71.67%	71.83%	69.80%	74.12%	74.53%
Both	75.75%	75.84%	69.63%	77.39%	78.37%

[Zheleva, Getoor, Sarawagi (2010) *Higher-order Graphical Models for Classification in Social and Affiliation Networks*. NIPS Workshop on Networks Across Disciplines.]

Your privacy depends on others

- Higher % public profiles -> higher accuracy

flickr



	10%	30%	50%	70%	90%
Accuracy	63.4%	71.6%	83.6%	86.2%	87.6%
Node coverage	85.7%	74.7%	51.2%	48.6%	28.9%
Baseline	27.7%	27.7%	27.7%	27.7%	27.7%

Main takeaways

- Increasingly challenging to manage personal information online
  - Probabilistic models can infer hidden information and circumvent privacy preferences of users
  - Groups are most significant carriers of private information
- For privacy-concerned users
  - If possible, set group and friendship lists as private
- For social media businesses
  - Enable greater user control over release of information
- For policy makers
  - Set standards for personal data release practices

- Questions?

- Virtual office hour
- https://usc.zoom.us/j/95136500603?pwd=VEJhblhWK25lT2N3RC9FNWk3eTJKQT09