Beyond Networking

Krishna P. Gummadi
Max Planck Institute for Software Systems

Part I:

From Networking to Computational Social Science

Rise of online social networks

- Network-assisted human interactions

- New opportunities: Computational Social Science
  - Study digital traces of human activities in the network
  - Analyze data at unprecedented scale and granularity

Empirical studies of social interactions

- @Unprecedented scale
  - Structure of social graphs, 20M+ users
    - *[IMC '07]*
  - User influence in social networks, 50M+ users
    - *[ICWSM '10]*
- @Unprecedented granularity
  - Emergence of linguistic conventions in 50M+ social network
  - Analyzing all 2B+ messages exchanged over a 3-year period
    - *[ICWSM '12]*

Targeted Advertising on Social Media: Transparency, Control, Fairness, Privacy

Krishna P. Gummadi

Max Planck Institute for Software Systems

Summary: Ad Targeting on Social Media

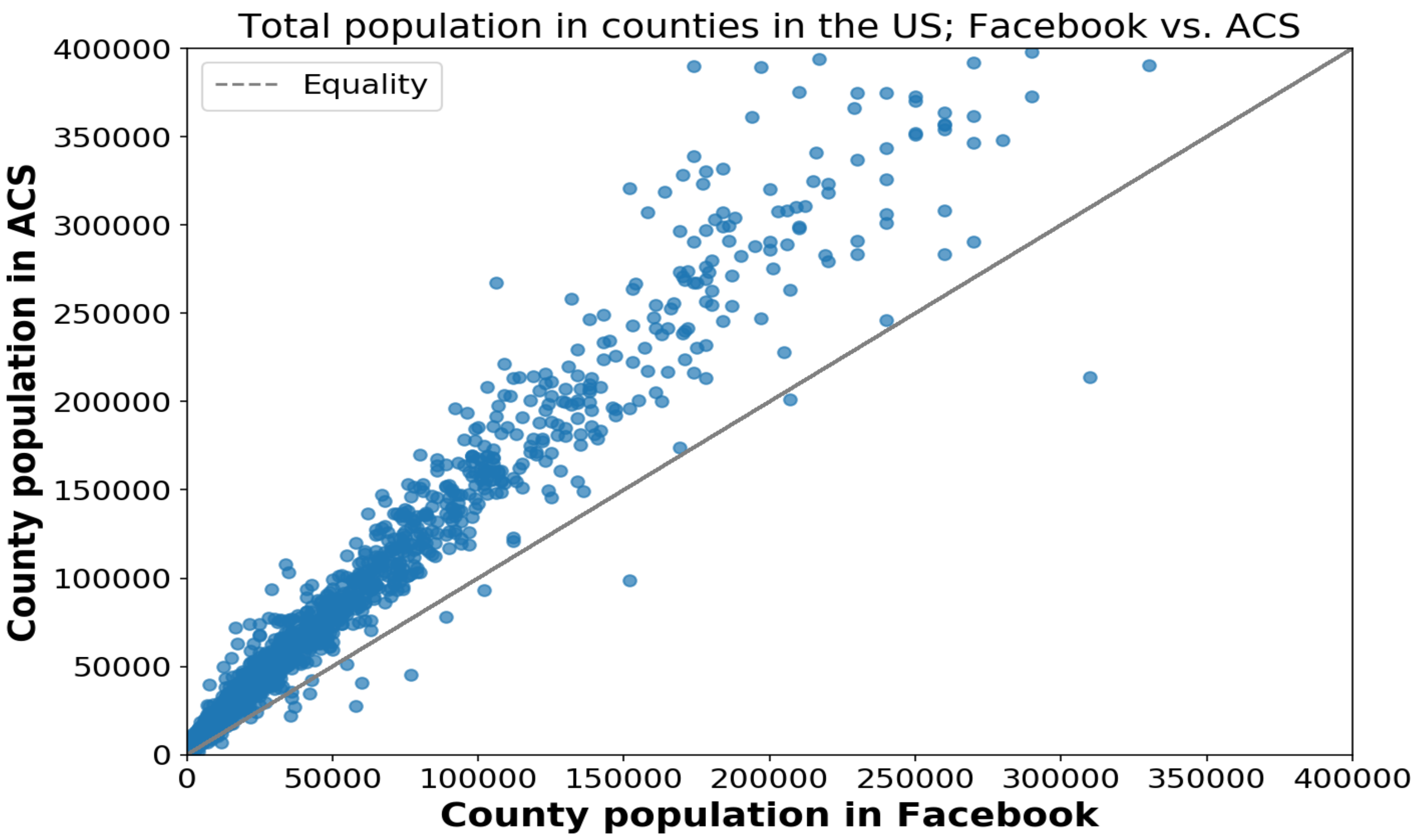
1. Transparency
   - Limited! Only some (not all) data used is revealed
2. Control
   - Limited! Providing satisfactory explanations is non-trivial
3. Fairness
   - Unclear! Need new measures & methods for fairness
4. Privacy:
   - Non-existent! Avoiding data leaks via ads is non-trivial!

- Lots of open challenges!

Census in the US

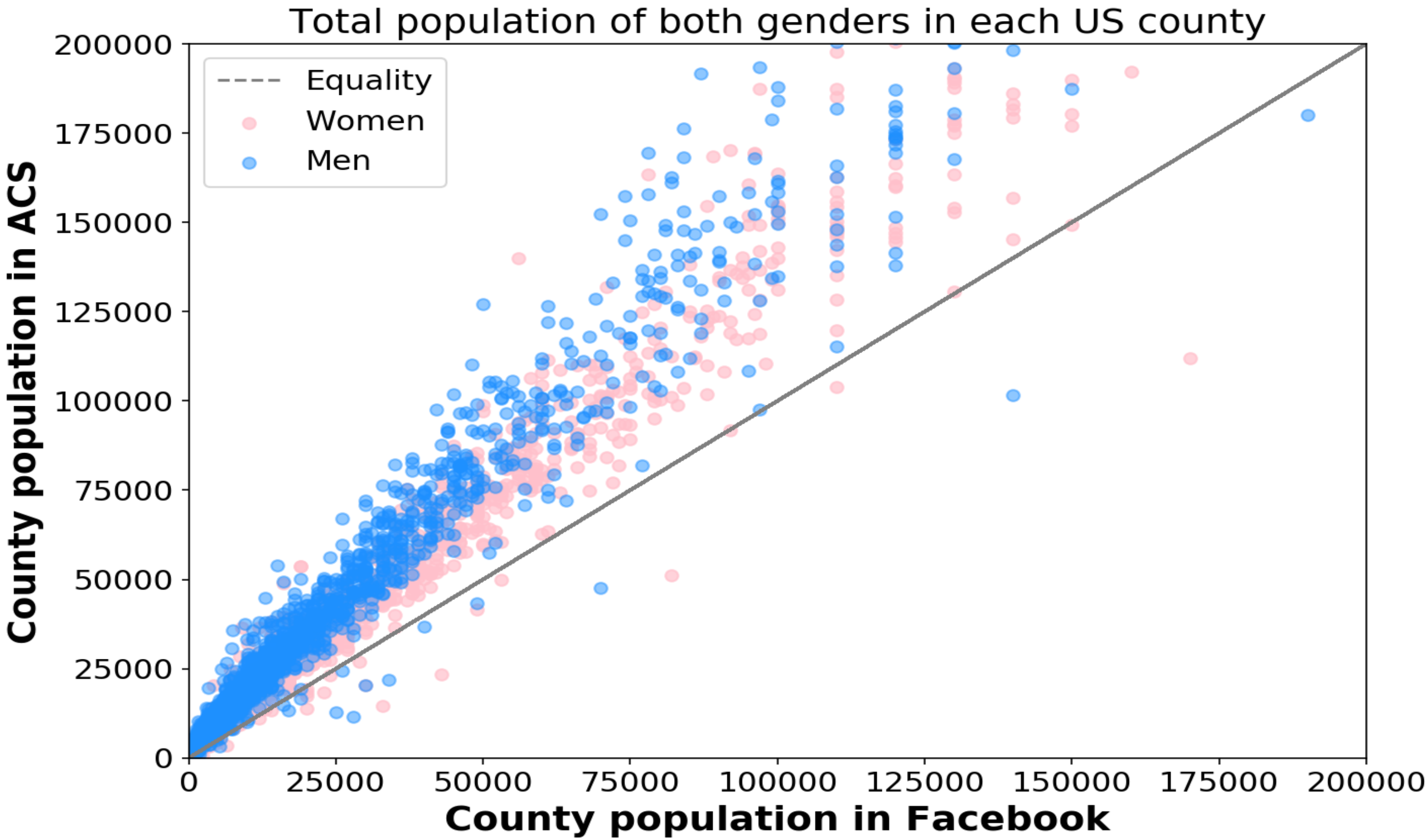
- Constitution requires enumeration of the population
  - Every 10 years since 1790
- Why? Gather factual data about impact of policies
- Innovations:
  - Ask more types of data about the population: Long-form
  - In 1940, sub-sample the population
  - In 1990, rolling census for more real-time data: ACS
- Real-time census using Facebook ad targeting API?
  - *[Journal of Population and Development Review '17]*

ACS vs. Facebook: Total Population



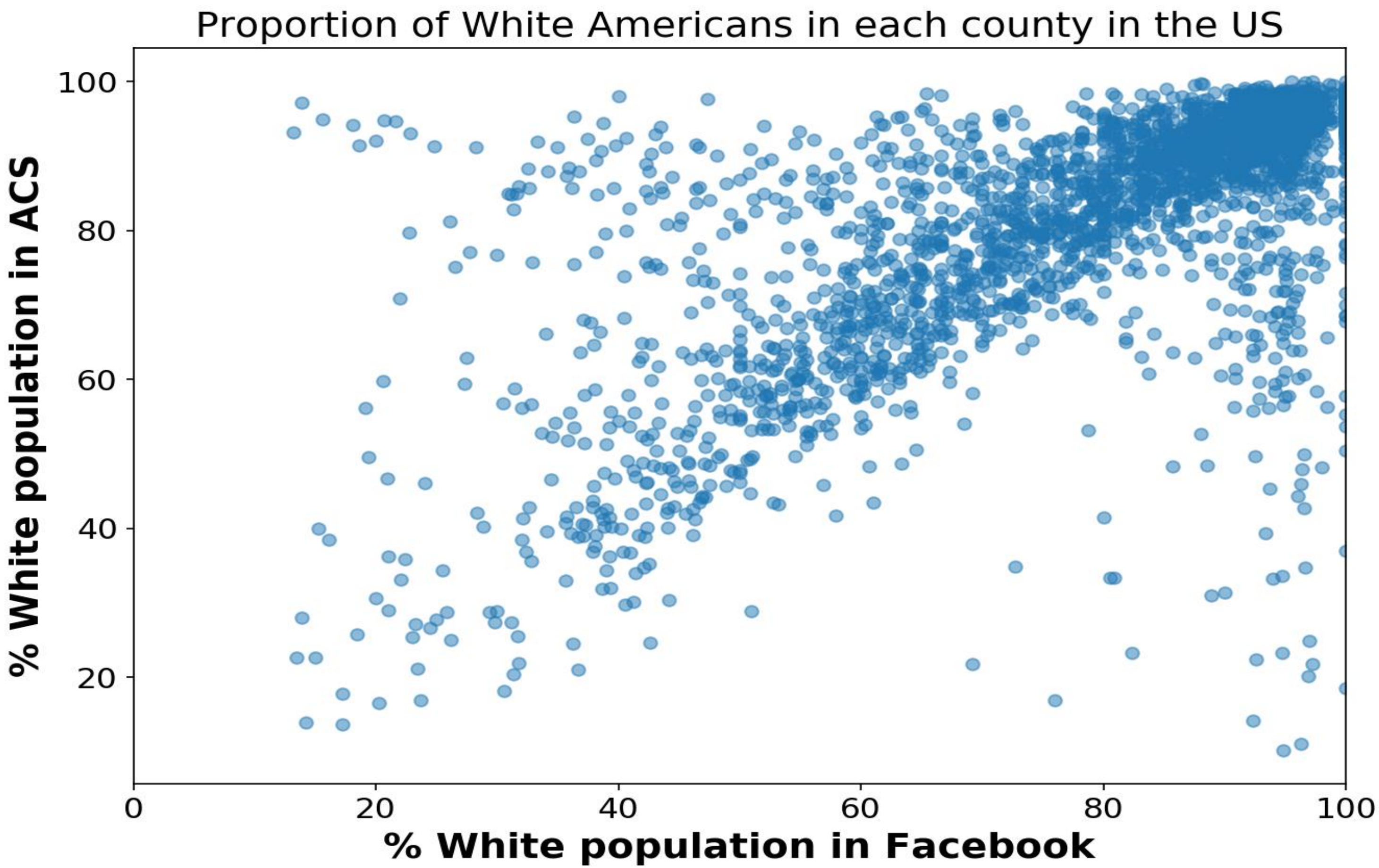
- r = 0.9768

ACS vs. Facebook: Men and Women

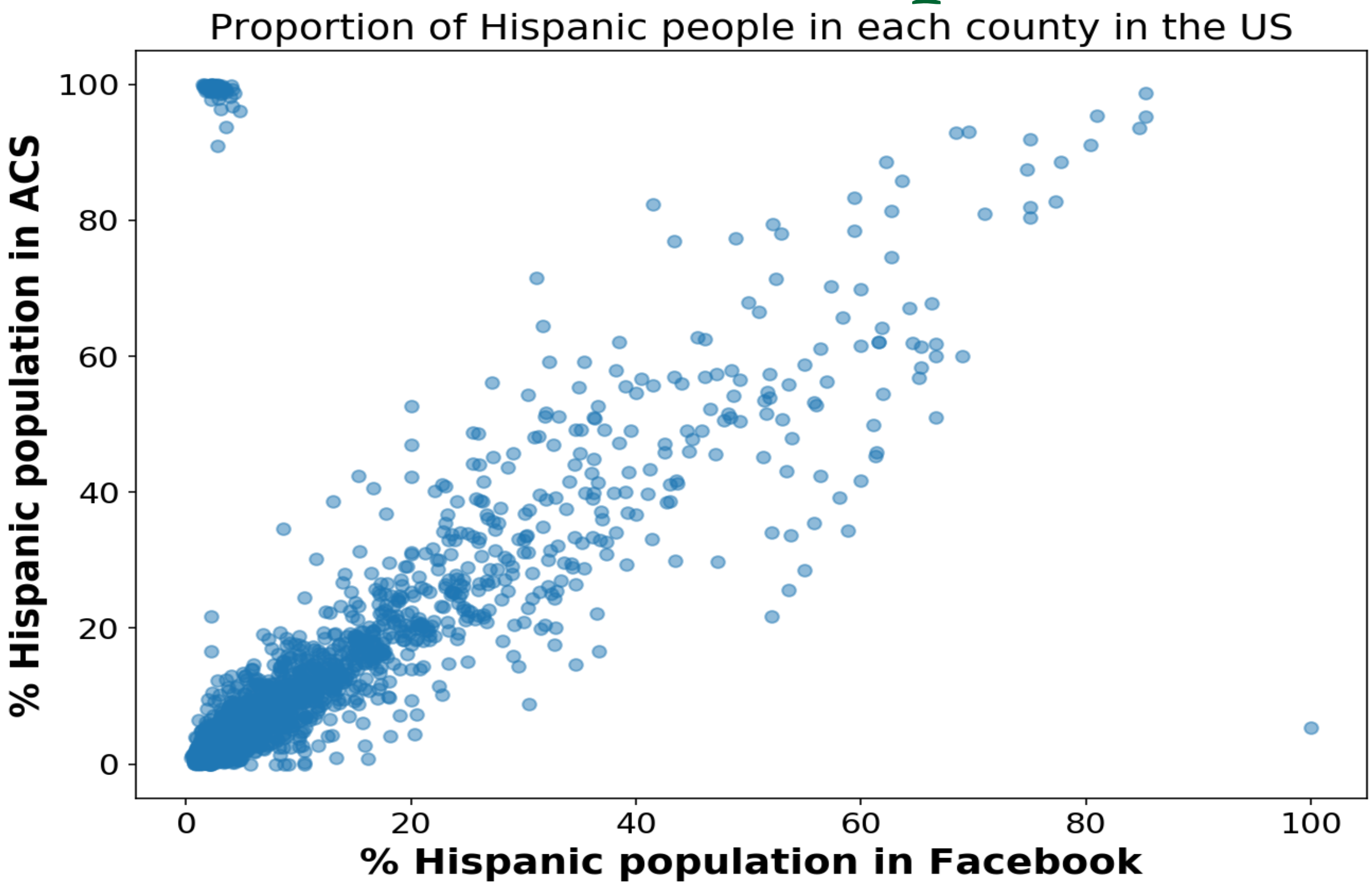


- $r_{men} = 0.9735$, $r_{women} = 0.9796$

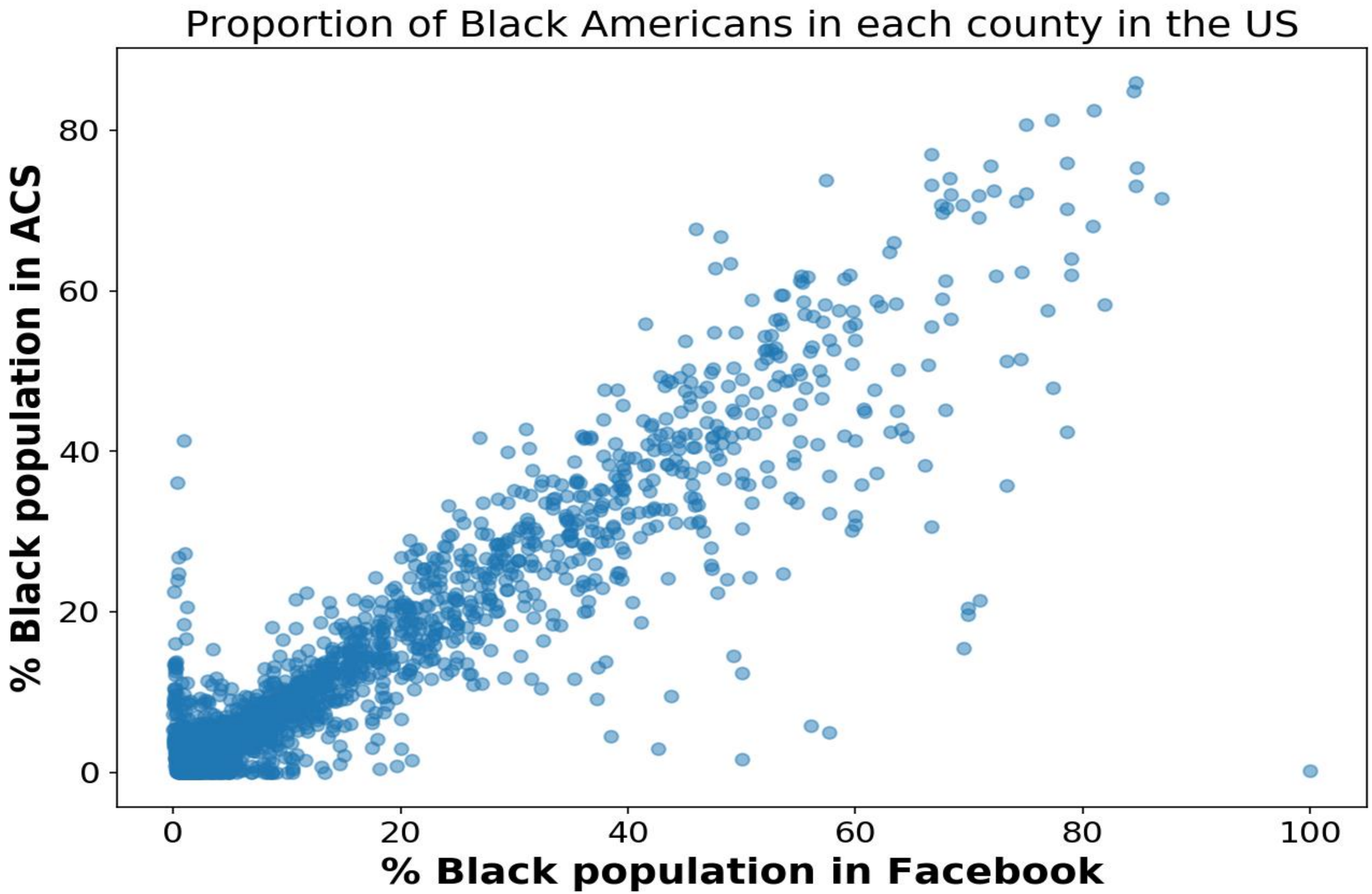
ACS vs. Facebook: Caucasian ethnicity



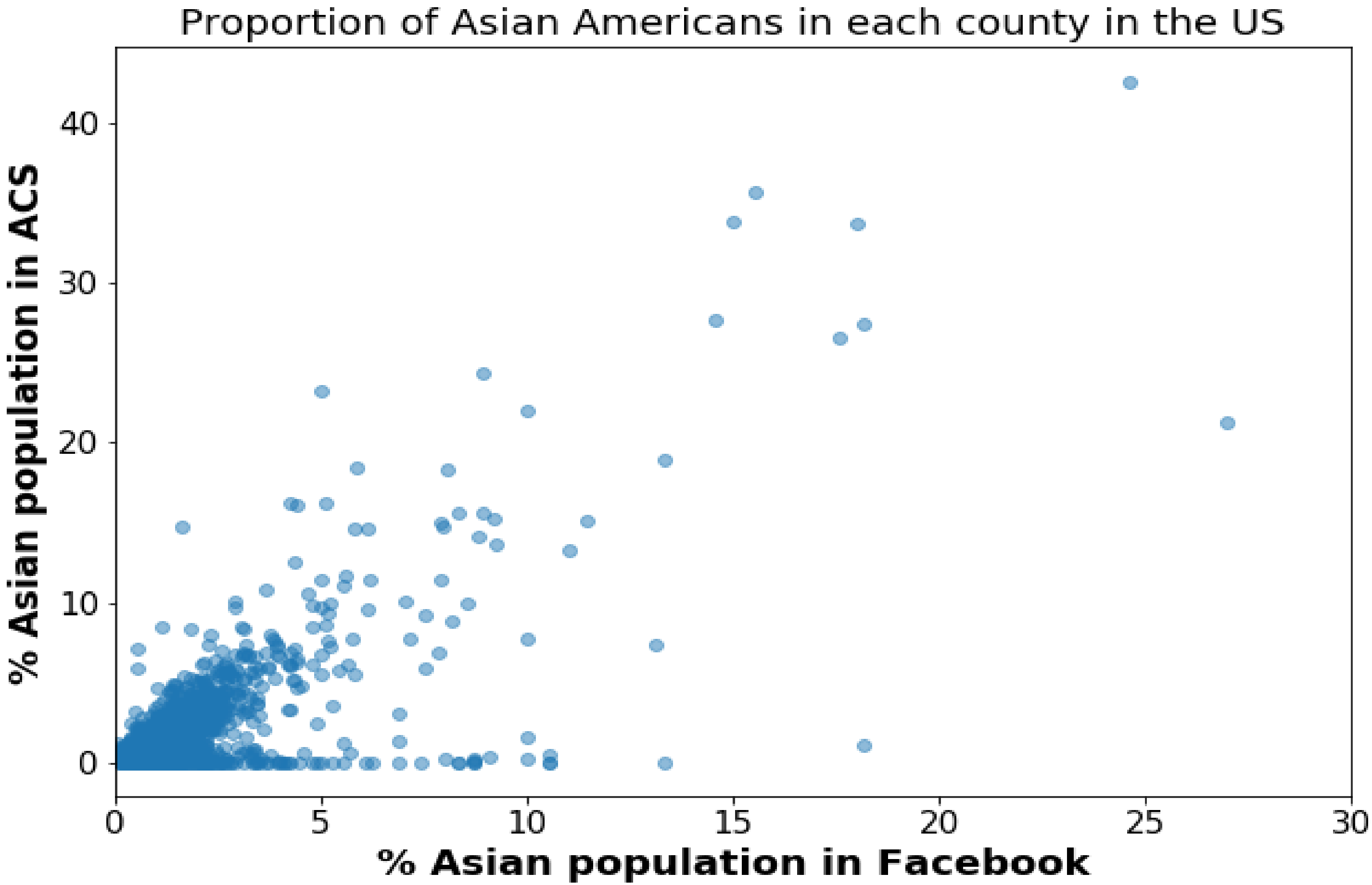
ACS vs. Facebook: Hispanic ethnicity



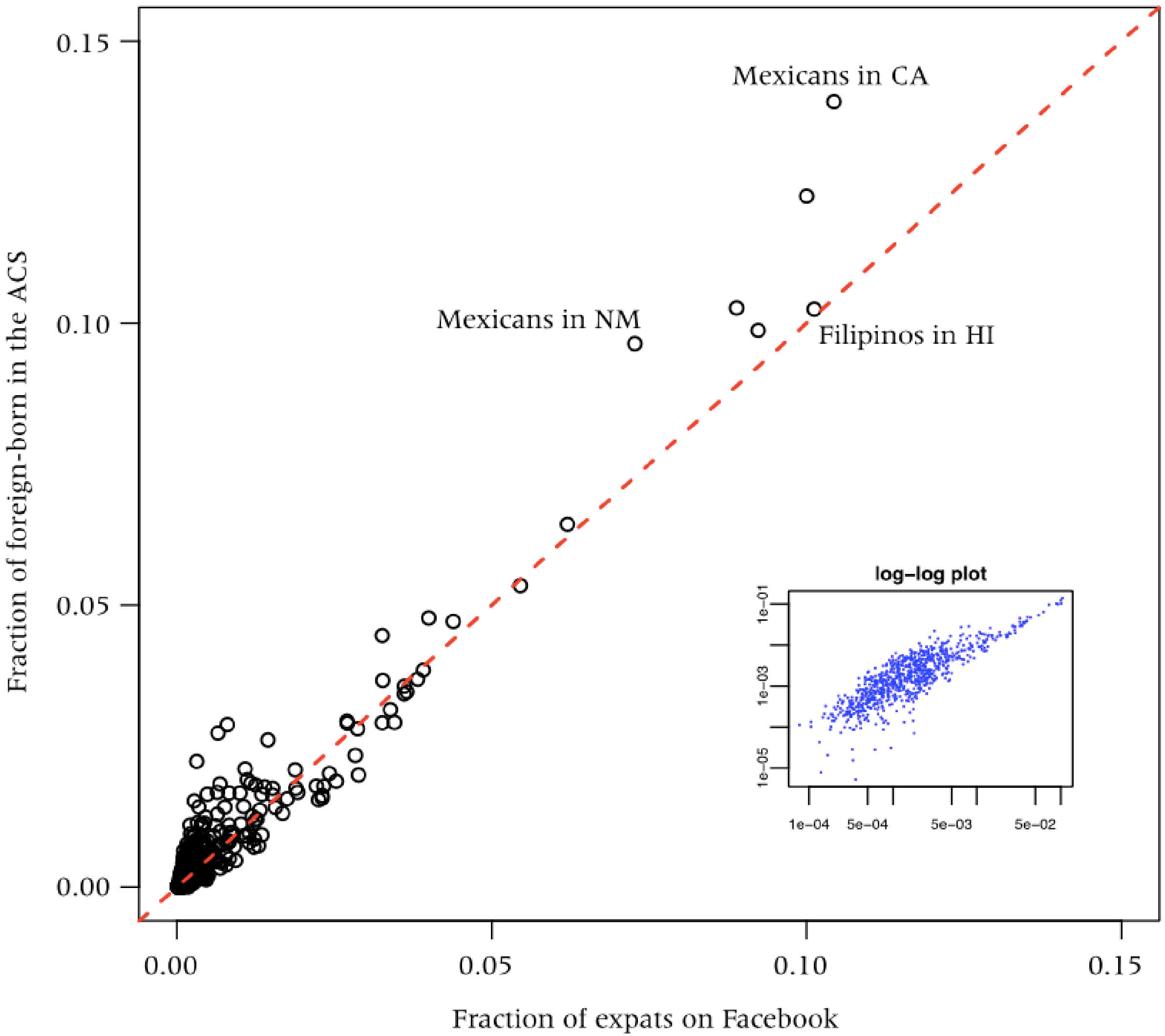
ACS vs. Facebook: African ethnicity



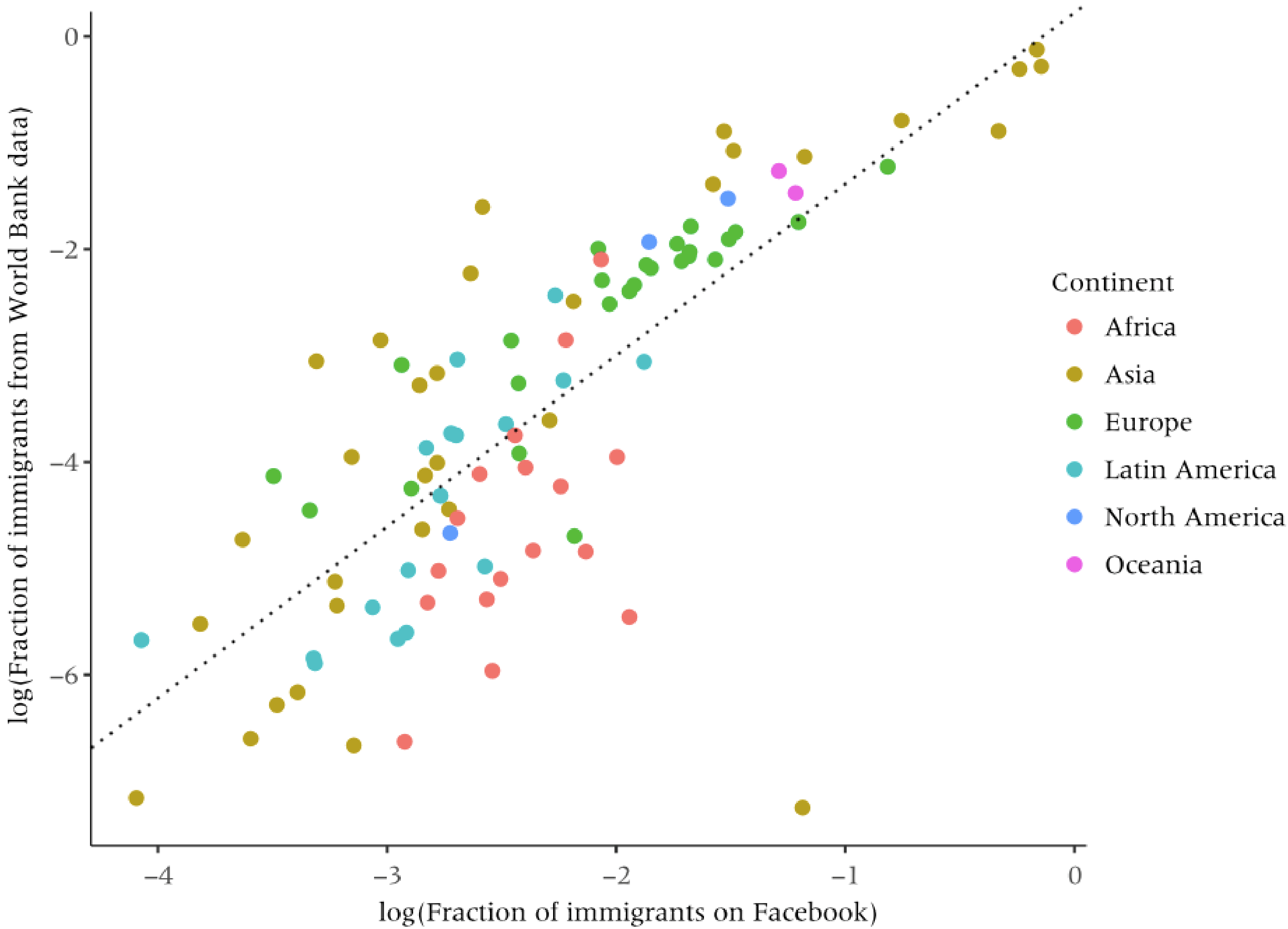
ACS vs. Facebook: Asian ethnicity



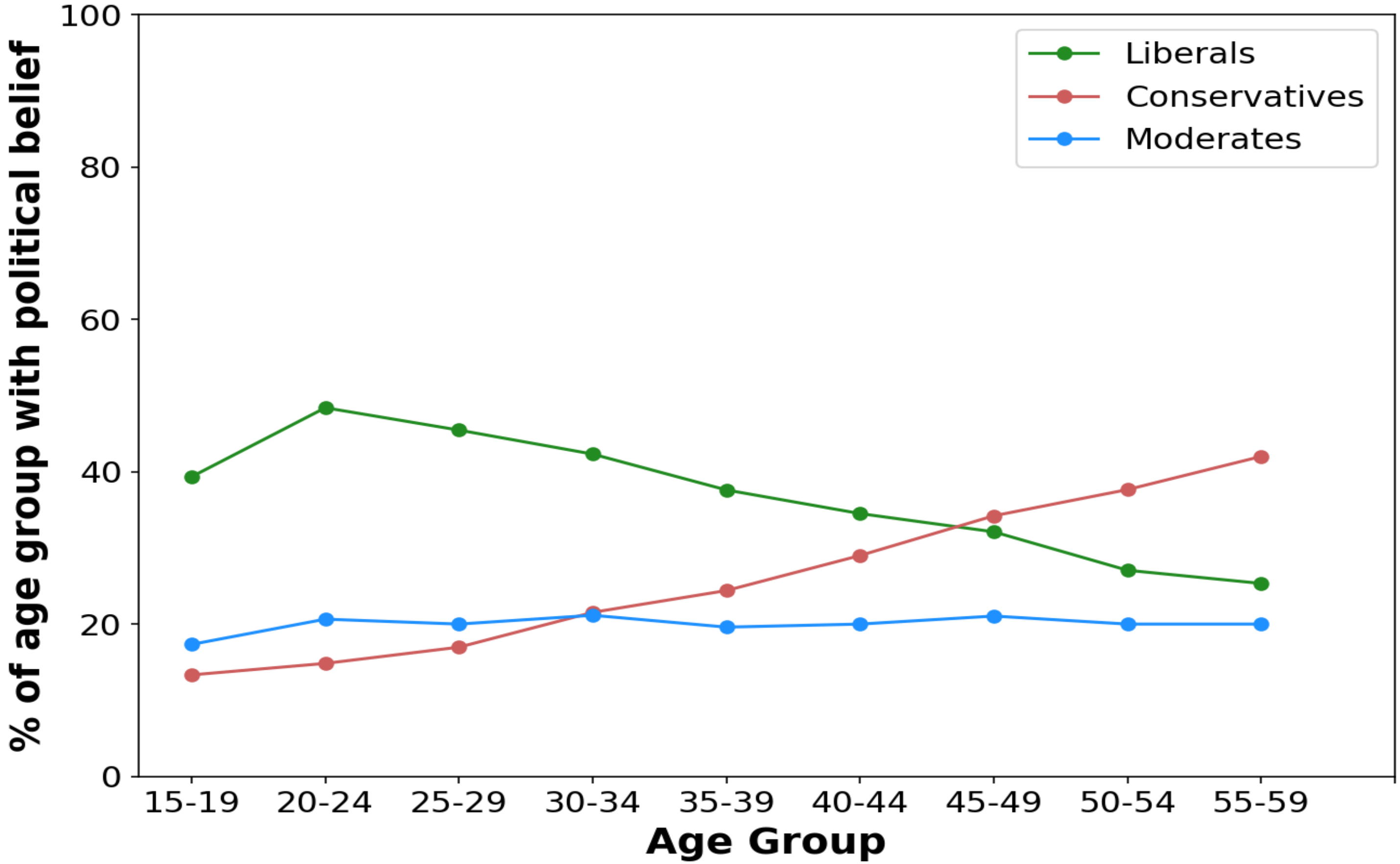
ACS vs. Facebook: Migrant population



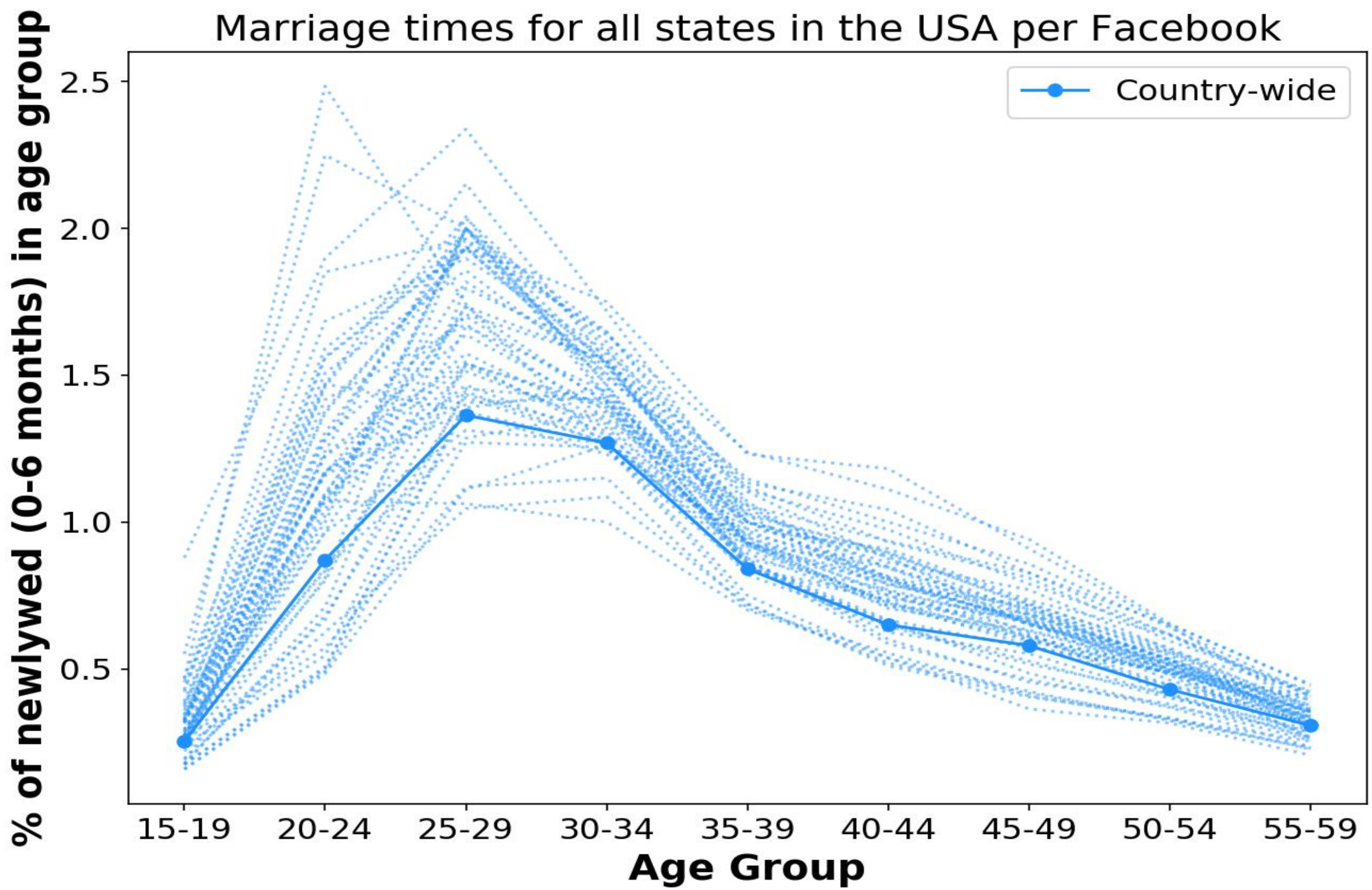
World bank vs. Facebook: Migrants



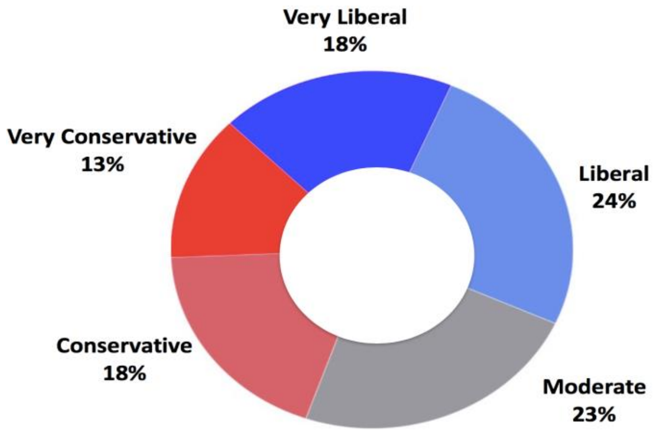
New uses: Political beliefs across ages



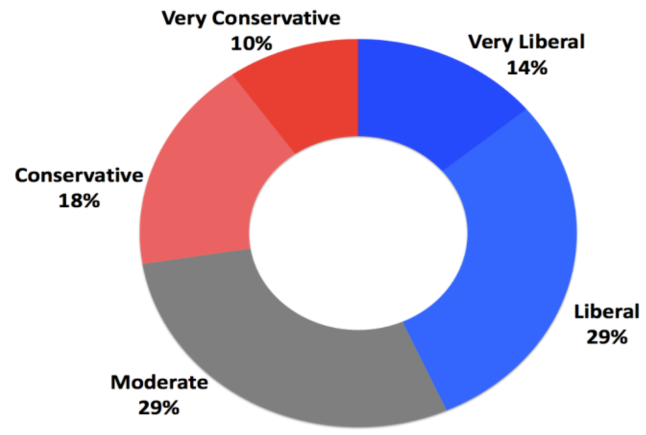
New uses: When people marry



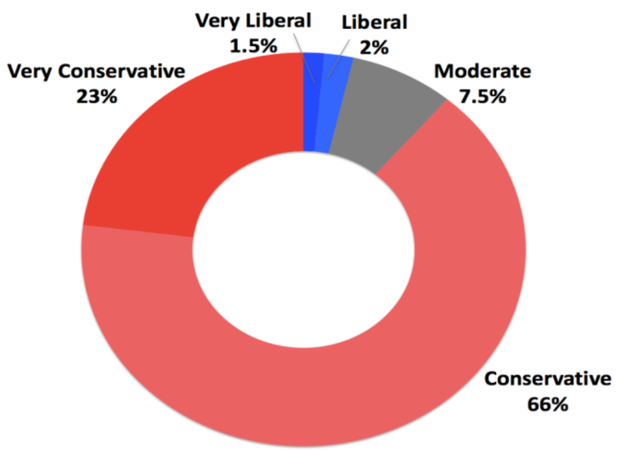
New use: Bias in news media *[ICWSM '18]*



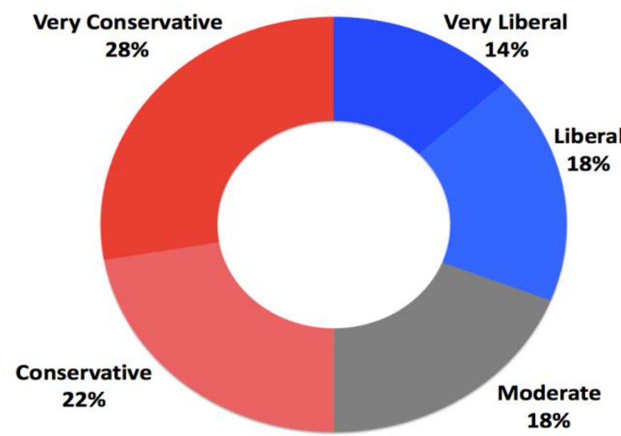
(a) US Population on FB



(b) NYTimes Readers on FB



(d) Brietbart Readers on FB



(c) FoxNews Readers on FB

Part II:

From Network Traffic Differentiation to Algorithmic Discrimination

Network transparency

Glasnost: Enabling End Users to Detect Traffic Differentiation

Marcel Dischinger
MPI-SWS

Massimiliano Marcon
MPI-SWS

Saikat Guha
MPI-SWS, Microsoft Research

Krishna P. Gummadi
MPI-SWS

Ratul Mahajan
Microsoft Research

Stefan Saroiu
Microsoft Research

Max
Planck
Institute
for
Software Systems

Glasnost: Test if your ISP is shaping your traffic

▪ Home ▪ Glasnost Tests ▪ Create your own test ▪ Run your own server ▪ Results ▪

From traffic differentiation to price discrimination

Measuring Price Discrimination and Steering on E-commerce Web Sites

Aniko Hannak
Northeastern University
Boston, MA
ancsaaa@ccs.neu.edu

Gary Soeller
Northeastern University
Boston, MA
soelgary@ccs.neu.edu

David Lazer
Northeastern University
Boston, MA
d.lazer@neu.edu

Alan Mislove
Northeastern University
Boston, MA
amislove@ccs.neu.edu

Christo Wilson
Northeastern University
Boston, MA
cbw@ccs.neu.edu

Detecting price and search discrimination on the Internet

Jakub Mikians†, László Gyarmati*, Vijay Erramilli*, Nikolaos Laoutaris*
Universitat Politecnica de Catalunya†, *Telefonica Research
jmikians@ac.upc.edu,{laszlo,vijay,nikos}@tid.es

Onto algorithmic discrimination

- Discrimination is a specific type of unfairness
- Well-studied in social sciences
  - Political science
  - Moral philosophy
  - Economics
  - Law
    - Majority of countries have anti-discrimination laws
    - Discrimination recognized in several international human rights laws

- But, less-studied from a computational perspective

What is a computational perspective? Why is it needed?

Defining discrimination

- A first approximate normative / moralized definition:

wrongfully impose a **relative disadvantage** on persons **based on** their membership in some **salient social group** e.g., race or gender

- Challenge: How to operationalize the definition?
  - How to make it clearly distinguishable, measurable, & understandable in terms of empirical observations

Need to operationalize 4 fuzzy notions

1. What constitutes a relative disadvantage?
2. What constitutes a wrongful imposition?
3. What constitutes based on?
4. What constitutes a salient social group?

Case study: Recidivism risk prediction

- **COMPAS** recidivism prediction tool
  - Built by a commercial company, Northpointe, Inc.

- Estimates likelihood of criminals re-offending in future
  - Inputs: Based on a long questionnaire
  - Outputs: Used across US by judges and parole officers

- Trained over big historical recidivism data across US
  - Excluding sensitive feature info like gender and race

COMPAS Goal: Criminal justice reform

- Many studies show racial biases in human judgments
- **Idea:** Nudge subjective human decision makers with objective algorithmic predictions
  - Algorithms have no pre-existing biases
  - They simply process information in a consistent manner
- Learn to make accurate predictions without race info.
  - Blacks & whites with same features get same outcomes
  - No disparate treatment & so non-discriminatory!

Is COMPAS non-discriminatory?

	Black Defendants		White Defendants	
	High Risk	Low Risk	High Risk	Low Risk
Recidivated	1369	532	505	461
Stayed Clean	805	990	349	1139

Is COMPAS non-discriminatory?

	Black Defendants		White Defendants	
	High Risk	Low Risk	High Risk	Low Risk
Recidivated	1369	532	505	461
Stayed Clean	805	990	349	1139

False Positive Rate: 805 / (805 + 990) = 0.45 349 / (349 + 1139) = 0.23

Is COMPAS non-discriminatory?

	Black Defendants		White Defendants	
	High Risk	Low Risk	High Risk	Low Risk
Recidivated	1369	532	505	461
Stayed Clean	805	990	349	1139

False Positive Rate: 805 / (805 + 990) = 0.45   349 / (349 + 1139) = 0.23

False Negative Rate: 532 / (532 + 1369) = 0.29   461 / (461 + 505) = 0.48

Is COMPAS non-discriminatory?

	Black Defendants		White Defendants	
	High Risk	Low Risk	High Risk	Low Risk
Recidivated	1369	532	505	461
Stayed Clean	805	990	349	1139

False Positive Rate: 805 / (805 + 990) = 0.45 >> 349 / (349 + 1139) = 0.23

False Negative Rate: 532 / (532 + 1369) = 0.29 << 461 / (461 + 505) = 0.48

- ProPublica: False positive & negative rates are considerably worse for blacks than whites!
  - Constitutes discriminatory **disparate mistreatment**

COMPAS study raises many questions

- Why does COMPAS show high racial FPR/FNR disparity?
  - Despite being trained without race information

- Can we train COMPAS to lower racial FPR/FNR disparity?

Analysis:

Why does COMPAS classifier show high racial FPR & FNR disparity?

How COMPAS learns who recidivates

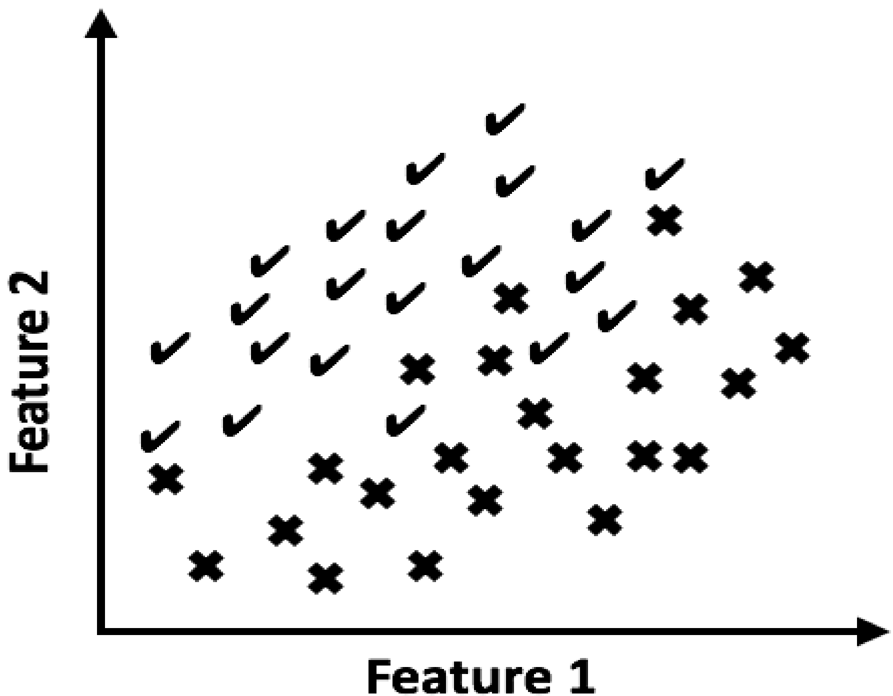
- By training over data about past outcomes

	F_1	F_2	…	F_m
Defendant$_1$	$x_{1,1}$	$x_{1,2}$	…	$x_{1,m}$
Defendant$_2$	$x_{2,1}$			$x_{2,m}$
Defendant$_3$	$x_{3,1}$			$x_{3,m}$
…	…			…
Defendant$_n$	$x_{n,1}$	$x_{n,2}$	…	$x_{n,m}$

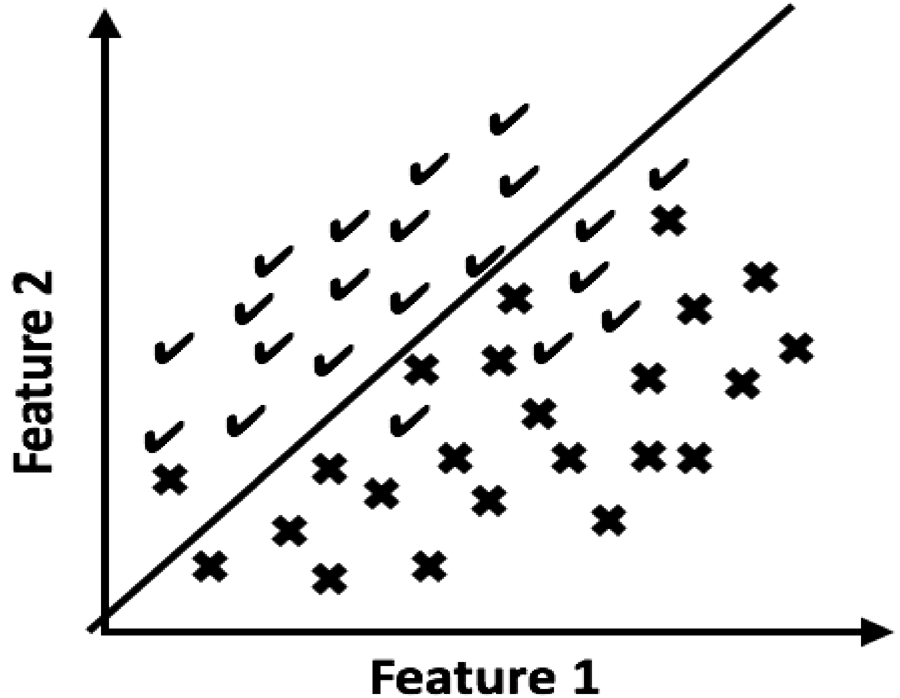
Past Outcomes
Recidivated
Stayed Clean
Stayed Clean
…
Recidivated

- Challenge: Learning a decision function over the features that separates the two classes of people

How COMPAS learns who recidivates



How COMPAS learns who recidivates



- By finding the optimal (most accurate / least loss) linear boundary separating the two classes
- How does COMPAS find (compute) it?

Learning (computing) the optimal boundary

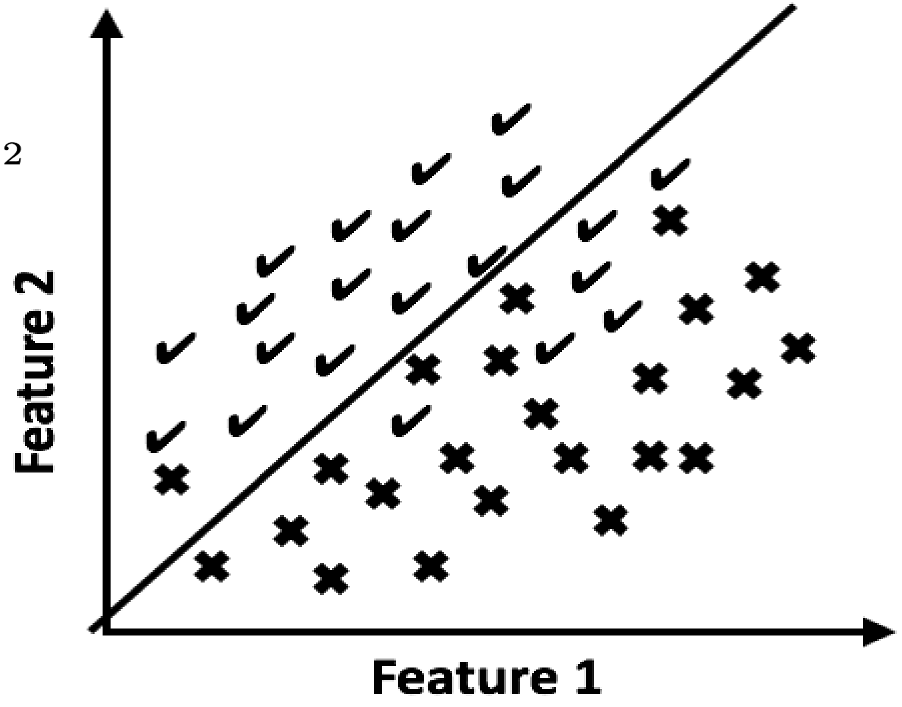
- Define & optimize a loss (accuracy) function
  - Capturing error (inaccuracy) in individual predictions

- Minimize loss over all examples in training data

$$L(\mathbf{w}) = \sum_{i=1}^{N} (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \qquad\qquad minimize\ L(\mathbf{w})$$

How COMPAS learns who recidivates

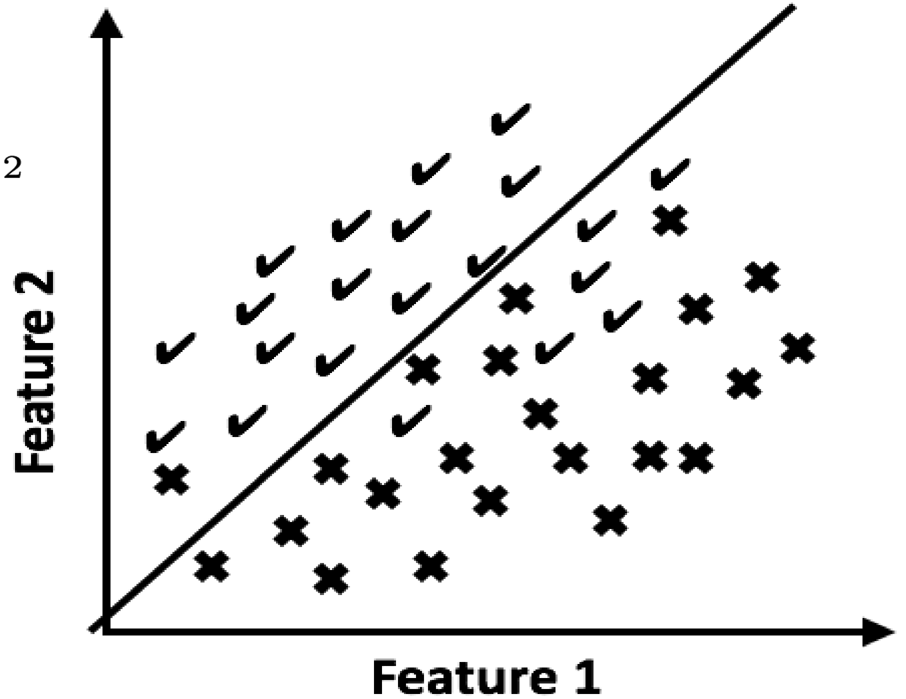
$$\min \quad \sum_{i=1}^{N} (y_i - d_{\mathbf{w}}(\mathbf{x}_i))^2$$



- How did COMPAS find most accurate linear boundary?

How COMPAS learns to discriminate

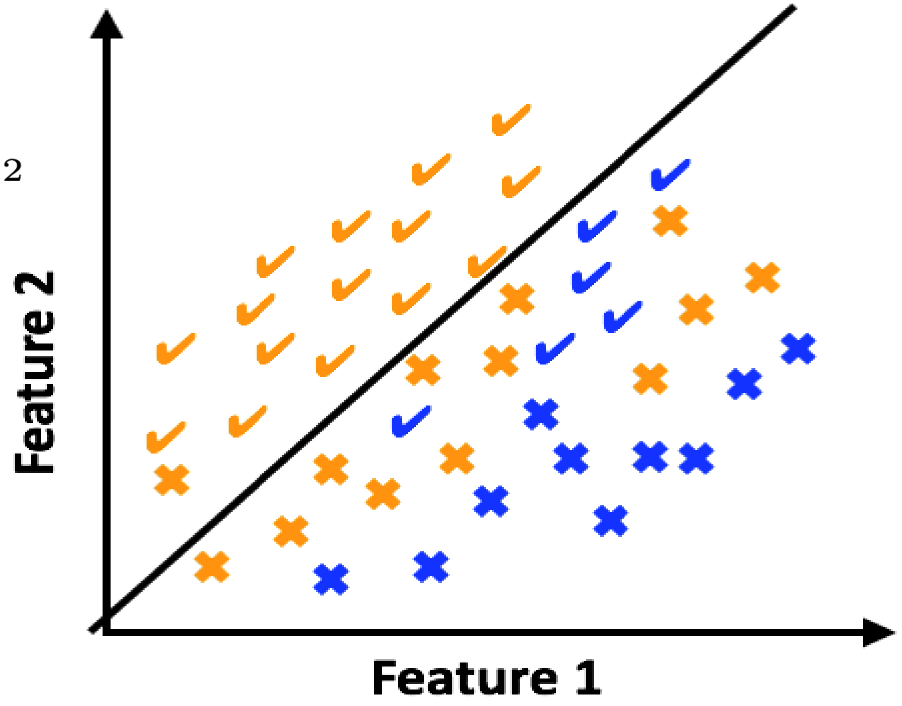
$$\min \sum_{i=1}^{N} (y_i - d_{\mathbf{w}}(\mathbf{x}_i))^2$$



- Observe the most accurate linear boundary

How COMPAS learns to discriminate

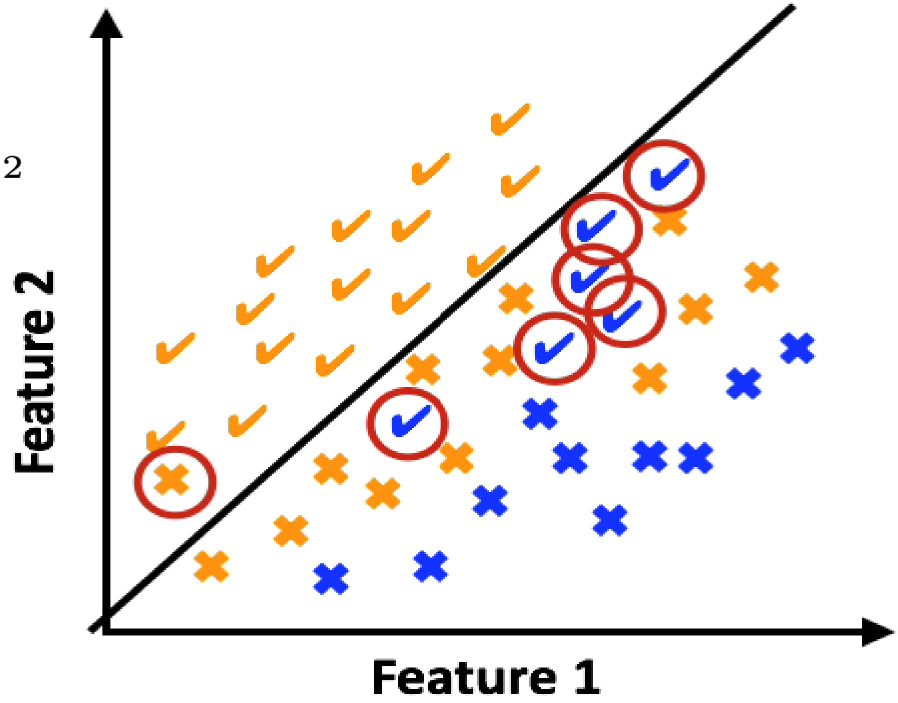
$$\min \sum_{i=1}^{N} (y_i - d_{\mathbf{w}}(\mathbf{x}_i))^2$$



- Observe the most accurate linear boundary

How COMPAS learns to discriminate

$$\min \sum_{i=1}^{N} (y_i - d_{\mathbf{w}}(\mathbf{x}_i))^2$$



- Observe the most accurate linear boundary
- Makes few errors for yellow, lots of errors for blue!
  - Causes disparate mistreatment – inequality in error rates

Synthesis:

How to train non-discriminatory classifiers? [WWW ‘17]

How to learn to avoid discrimination

- Specify discrimination measures as learning constraints
- Optimize for accuracy under those constraints

$$\min \quad P(y_{pred} \neq y_{true})$$

$$\text{s.t.} \quad P(y_{pred} \neq y_{true} \mid y_{true} = +1, \text{race}=B) = P(y_{pred} \neq y_{true} \mid y_{true} = +1, \text{race}=W)$$

$$P(y_{pred} \neq y_{true} \mid y_{true} = -1, \text{race}=B) = P(y_{pred} \neq y_{true} \mid y_{true} = -1, \text{race}=W)$$

- The constraints embed ethics & values when learning

- No free lunch: Additional constraints lower accuracy!
- Need race info in training to avoid disp. mistreatment!

The technical challenge

- How to learn efficiently under these constraints?

$$\min \ P(y_{pred} \neq y_{true}) \approx \min \sum_{i=1}^{N} (y_i - d_{\mathbf{w}}(\mathbf{x}_i))^2$$

s.t. $P(y_{pred} \neq y_{true} \mid y_{true} = +1, \text{race}=B) = P(y_{pred} \neq y_{true} \mid y_{true} = +1, \text{race}=W)$

$P(y_{pred} \neq y_{true} \mid y_{true} = -1, \text{race}=B) = P(y_{pred} \neq y_{true} \mid y_{true} = -1, \text{race}=W)$

- Problem: The above formulations are not convex!
  - Can't learn it efficiently

- Need to find a way to rewrite the constraints

Evaluation: Do our constraints work?

- Gathered a recidivism history dataset
  - Broward Country, FL for 2013-14
  - Features: arrest charge, #prior offenses, age,...
  - Class label: 2-year recidivism

- Traditional classifiers without constraints
  - Acc.: **67%** FPR Disparity: **+0.20** FNR Disparity: **-0.30**

- Training classifiers with fairness constraints
  - Acc.: **66%** FPR Disparity: **+0.03** FNR Disparity: **-0.11**

Lessons from the COMPAS story

Take-aways for ethical machine learning

High-level insight: Ethics & Learning

- Learning objectives implicitly embody ethics
  - By how they explicitly define trade-offs in decision errors
- Traditional objective accuracy reflects utilitarian ethics
  - The rightness of decisions is a function of individual utilities
  - The desired function is maximizing sum of individual utilities
- Lots of scenarios where utilitarian ethics fall short
  - Change learning objectives for other ethical considerations
    - E.g., non-discrimination requires equalizing group-level errors

Three challenges with ethical learning

- Operationalization:
  - How to formally interpret fairness principles in different algorithmic decision making scenarios?
- Synthesis:
  - How to design efficient learning mechanisms for different fairness interpretations?
- Analysis:
  - What are the trade-offs between the learning objectives?

Two operationalizations of discrimination:

disparate treatment & disparate mistreatment

Are they sufficient for all scenarios?

Discrimination in different scenarios

- What if training data labels were biased?
  - Require **equal group acceptance ~~error~~ rates** **[AISTATS '17]**
- Can requiring parity result in all groups being worse-off?
  - Yes! Parity outcomes are non pareto-optimal **[NIPS '17]**
  - Allow disparity when **no groups is worse-off than parity**
- Why not pick group-specific decision boundaries?
  - Need to avoid reverse-discrimination **[NIPS '17]**
  - Allow group-specific boundaries only **when they are envy-free**

Looking Forward:

From Non-Discrimination To Fair Algorithmic Decision Making

Social Welfare Theory

Social Choice Theory

Behavioral Economics

Moral Philosophy

Law

Journalism & Media Studies

Learning Non-Discriminatory Classification

Regression

Set Selection

Ranking

Matching

Clustering

Social Welfare Theory [KDD’18, NIPS‘18] [WWW’18, AAAI’18] **Moral Philosophy**

Social Choice Theory [FAT*19] [ICML’19, NDSS ‘18] **Law**

Behavioral Economics [AIES 19] [ICWSM ‘18] **Journalism & Media Studies**

Lea

erc Advanced Grant

on

[SIGIR’18] **Ranking**

[KDD ‘19] **Matching**

Clustering

Part III:

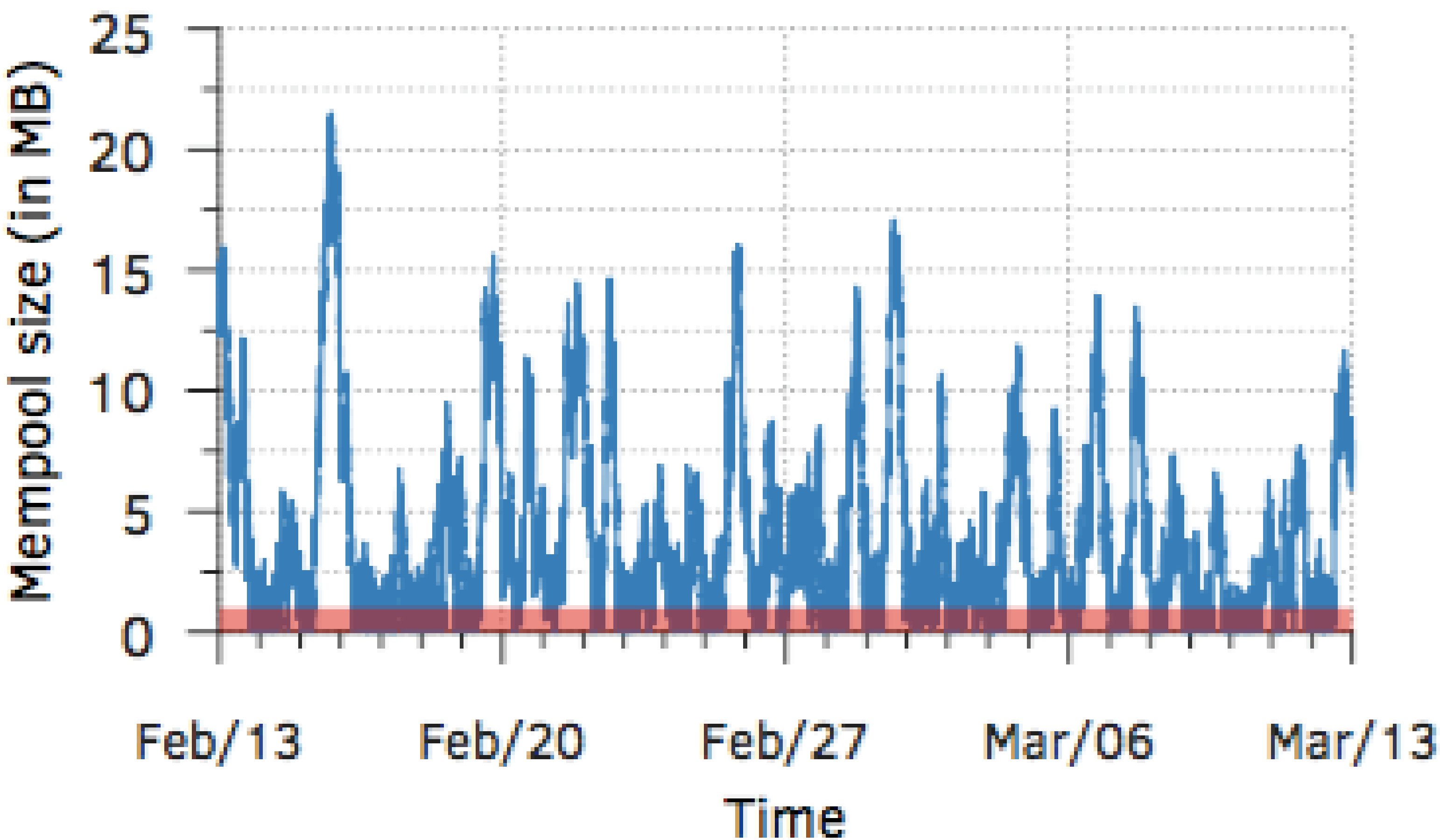
From Networking to Blockchains

Rewards for mining new blocks

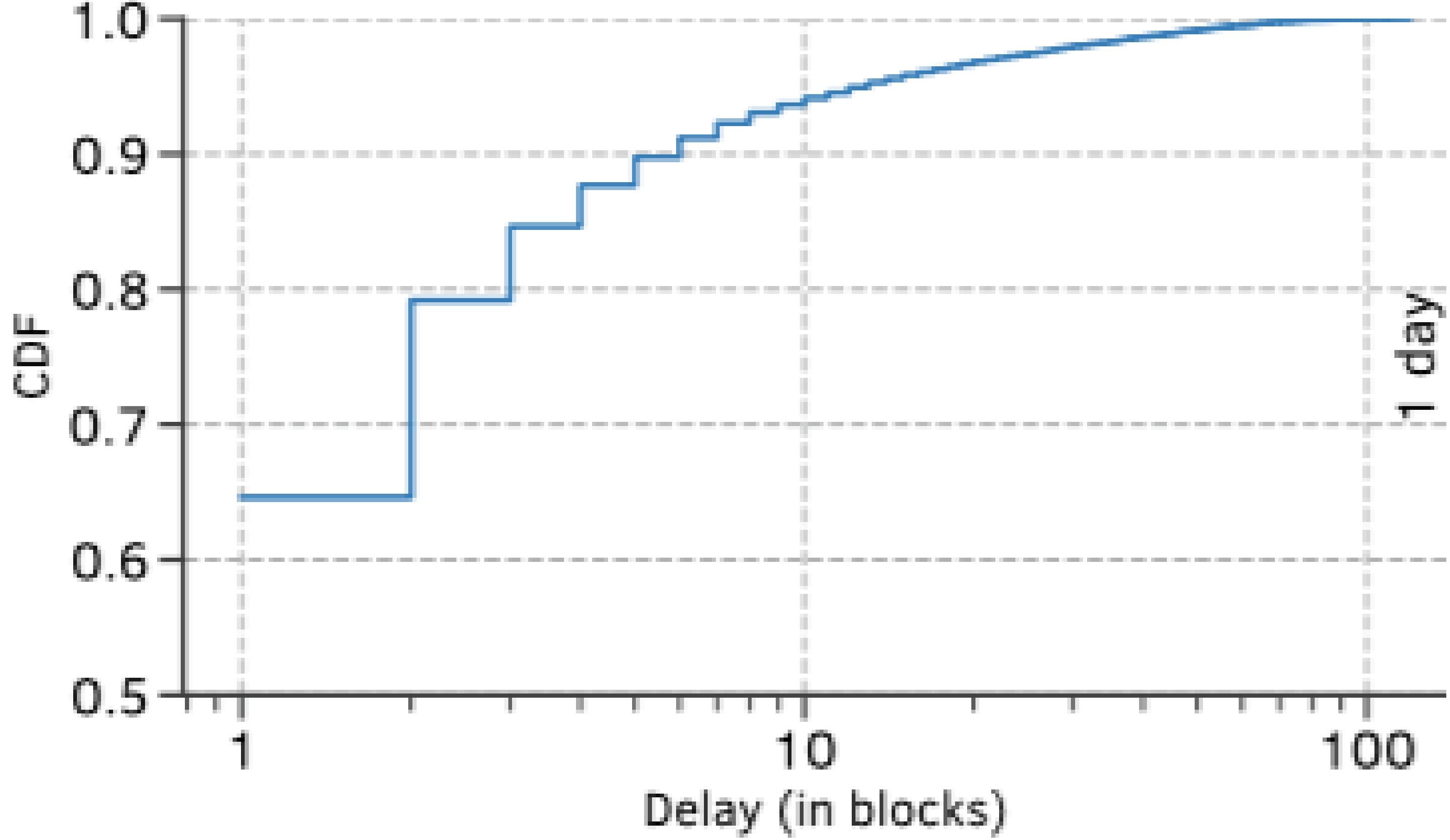
1. Each newly mined block generates new coins
2. Fee for including transactions in the block

- Fees incentivize miners to prioritize transactions for inclusion in next block
- Assumption: From the waiting transactions, miners pick transactions with highest fee/size
  - **Do they?**

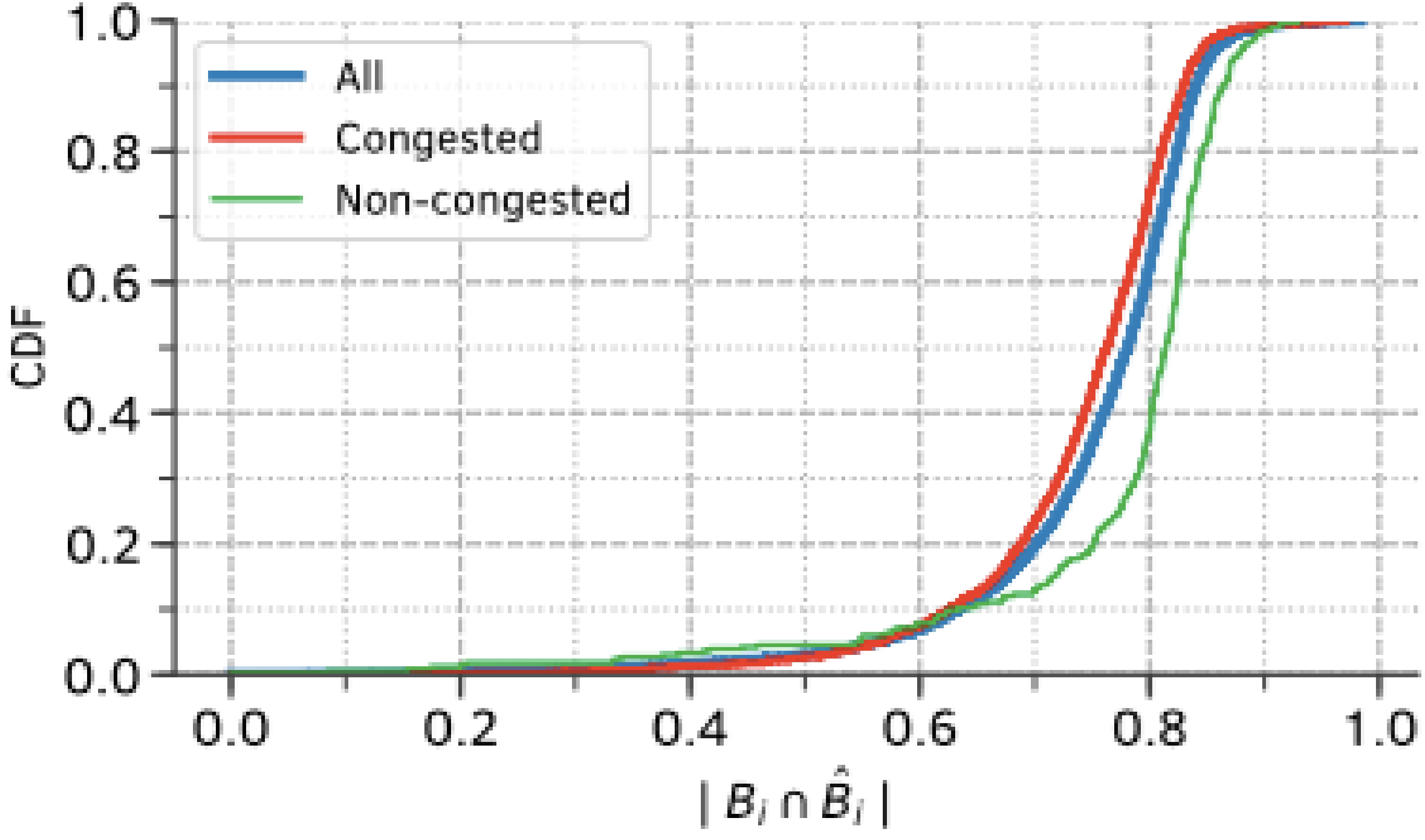
Are there periods of congestion?



Do transaction commit times vary?



Is queuing based on fee/size?



Research questions

- Why do miners deviate?
- Can we audit which miners are deviating?
- Can we re-design consensus protocols to ensure everyone follows agreed upon prioritization?

- What is the right prioritization policy?
  - Should transaction delay and value matter?
- Is the system vulnerable to DoS attacks?
- Is there a need for multiple priority queues?

Concluding thoughts / advice

- Work on growing problems, not shrinking problems
- Work on problems where you have unfair advantage
- Problems and opportunities don't align along community borders
  - Beware of community-centric group think and rise of populist strong men in the community