

Lecture 0: Welcome to the class!

Should you take it? How to succeed?

COMS 6998: Analysis of Networks and Crowds

Thursday September 5th, 2019

Lecture zero

- * Why networks and crowds matter? What's computational about it?
- * What is this class about?
- * Who should take it?
- * What do you need to succeed?
- * Getting to know your instructors
- * Roadmap

Can you recognize this picture?





When did we start organize and
automate our lives around data?

1st and 2nd industrial revolution



- * How to scale a manufactured society?
 - Commerce (Harrod's 1834): A fixed public retail price
 - Information (La Presse 1836): Pay quality content with ads
 - Politics (NY Tribune 1841): A nationwide address



- * In many ways, the Internet only perfected this success!

Early 2010s – the Data revolution

- * Data + personalization challenge our values

- Commerce (WSJ 2012):
Price discrimination creates new barriers



Valentino, Singer, Soltani (2012) Websites
Vary Prices, Deals Based on Users'
Information

- Information (C. ACM 2013):
Sponsored/Organic results biased by gender/race

Sweeney, L. (2013). Discrimination in
online ad delivery.



genome.gov
National Human Genome Research Institute
National Institutes of Health

Research Funding | Research at NIH | Health | Education | Issues in Genetics | Home | About | Office of the Director | Office of Population Genetics | Staff Resources | ebony.k.rosenblatt

DNPS Staff Biographies

Anatoli L. Wise, Ph.D.
Ebony B. Bookman, M.S.G.C., Ph.D.
Erica M. Ramey, Ph.D., M.P.H.
Heather Jones, M.S.
Linda A. Heslop, Ph.D., M.P.H.
Renee L. Hsiao, M.D., M.P.H.

EBONY B. BOOKMAN, M.S.G.C., Ph.D.
Epidemiologist
Office of Population Genetics

M.S. Howard University, 1999
Ph.D. Howard University, 2001

Ads by Google
Ebony Bookman Truth
Looking for Ebony Bookman? Check Ebony Bookman's Arrests.
www.instantcheckmate.com/

We Found Ebony Bookman
1) Get Ebony's Background Report 2) Contact Info & More
- Try Free!
www.peopleSMART.com/

We Found Ebony Bookman
Current Address, Phone and Age. Find Ebony bookman, Anywhere.
www.peoplefinders.com/

- Politics (NYT 2012):
Political websites vary based on cookies

Edsall, T. B. (2012). Let the Nanotargeting
Begin.



Don't Miss: Prescriber Checkup | Tobacco Bonds | Data
PRO PUBLICA Journalism in the
Home | Our Investigations | Data | MuckRaps | Get Involved | Message Machine | About | Contact | Privacy
Message Machine
Reverse Engineering the 2012 Campaign

CU

It could not get any worse, right?

- * 2016: Personalization runs gamut of moral hazards
 - Commerce (Wash. P, Oct. 31st)
“Sharing” economy exacerbates discrimination
 - Information (Propublica, Oct. 28th)
Ad-networks excludes recipient based races
 - Politics (Wash. P, Oct. 11th)
FB, TW & Instagram used by law enforcement



We need new science

* Growing concerns over ethical automated decisions

The collage includes the following elements:

- ProPublica Article:** "We Can't Trust Facebook to Regulate Itself" by Sandy Parak, January 15th, 2018. The article discusses machine bias and the use of older workers in employment law.
- The Washington Post Article:** "I Mentored Mark Zuckerberg. I Loved Facebook. But I Can't Stay Silent About What's Happening." by Sam Levi, November 17th, 2018. The article is about social media and technology.
- Facebook Image:** A screenshot of a Facebook page for "Being Patriotic Community" showing a post about the impact of China on the Internet.
- Airbnb Article:** "Airbnb Research: China's Impact on Its Social Data" by Sam Levi, November 17th, 2018. The article discusses the company's research and its impact on China.
- Mark Zuckerberg:** A video still of Mark Zuckerberg testifying at a House Commerce Committee hearing in Washington D.C. on April 11, 2018.
- GreatCall Advertisement:** An advertisement for the Jitterbug Smart2 smartphone, featuring the GreatCall logo and the text "The easy-to-use Jitterbug Smart2."
- CS@CU Logo:** The logo for the Center for Science and the Civic University (CS@CU).

Think outside the bubble

- * All previous bubble episode were retrospectively very important to deploy game changing techs.
- * Late 80s ... cheap microprocessors, no applications
 - **But** had brought millions of pcs to business/home
- * Late 90s ... end of the dot-com boom
 - **But** the Internet infrastructure was built for future
- * 2010s peak of the Big “Data” Bang, AI



Today

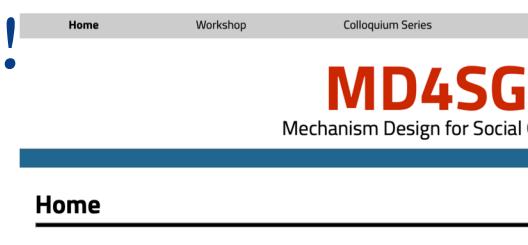
What are we building for the next generation?

“The best mind of my generation are thinking about
how to make people click ads.” J. Hammerbacher

“This Tech Bubble Is Different.”
A. Vance, Businessweek, 04/17/2011

Social Computing

- * The next generation could be the one with access to an unprecedented amount of **behavioral** data
- * This can solve **real** problems
 - ... not just finding a movie or a restaurant
 - ensuring energy efficiency
 - monitoring our environment
 - reduce inequality
 - informing social decision
- * And creates **new ones!**



Only convinced by numbers?

+40%

- * How much data production grows / year
 - Enough to double every 24months
(1h of videos upload YouTube each sec).

€260b

- * How much data can save on health care
 - In Europe [McKinsey] (U.S. save \$300b)

+300-1000%

- * How much lifts improve when ads are using behavioral targeting

What are Networks? Crowds?

- * Large set of **personal information** about users
 - History of Browsing, Purchasing, Rating
 - Sociological profile (age, gender, location, income)
 - Community of interests
- * Large set of **relational information** about users
 - Connections (friendship, collaboration, schoolmate)
 - Contacts (email IM phone calls etc., meeting)
- * Designing **mechanisms** between them
 - Contest, rankings, votings, estimating, influencing

A key principle

- * What **primarily** matters is your social environment!
 - For Business: how to best advertise a product?
 - For Media: how to find most relevant information?
 - For Engineers-CS: how to best design an application?
 - For Lawyers, Doctors, Public planners, etc.
 - For Science and Society at large: how to understand human behavior? Can (should?) we leverage it?
- ...
- ... 4 (classical) questions, being **reinvented** today

This “o” lecture

- * Why social networks matter?
- * What is this class about?
- * Who should take it?
- * What do you need to succeed?
- * Getting to know your instructors
- * Roadmap

Objectives of this class

- * Introduce **concepts** used in social networks
 - Connected to important scientific questions
 - and real systems, practical problems
- * Manipulate these concepts
 1. Make them familiar
Proof in class, Problem set to practice/experience
 2. Make them available for your critical eye
Case-studies with a goal: be satisfied and unsatisfied
 3. Understand their ethical implications (what?!)



Lecture Zero

- * Why social networks matter?
- * What is this class about?
- * Who should take it?
- * What do you need to succeed?
- * Getting to know your instructors
- * Roadmap

Why take this class: Reason #1

- * These concepts matter to the **companies** that you want a job from (including your startup!)
 - Social information is becoming the web's hottest commodity (Google, Facebook, IBM, Telcos, Media)
 - Users's data are company's key differentiating factor
- * You (not me) are the **social media generation!**
 - The game is just starting; it gets harder
 - CS deals with “complexity” deeply and elegantly
 - learning foundational concepts adds to your assets

Why take this class: Reason #2

- * This topic is **fertile for projects**, here and at large
 - Big data is everywhere, especially in public-funding
 - Many of these data are networks connecting people
- * Many **opportunities** within reach here at Columbia
 - Data science Institute: summer, fellows, post-docs
 - Brown Institute for Media Innovation
 - Multiple other departments
 - Overall, good topic for an academic job, eventually!

Why take this class: Reason #3

- * The topic offers a great **diversity** of experience
 - You can hack, make proof, interview people, make money, (perhaps) save lives
 - If you have broad interests (or wonder what you would like), this could be a way to figure out
- * Hence, this class **highly encourages** ...
 - Non-traditional majors/minors (especially enriching to our discussions)
 - CS, EE, Mech. E, Civil. E, OR, Med. School, Soc., Jlism

Before starting the trip



Before starting the trip

- * There is no textbook!



Why NOT to take this class?

- * Bad choice for maximum GPA / minimum workload
 - Switch between programming, maths, interpretation
 - Some topics are just out (immature), can manipulate unfamiliar concepts, a lot happens during the class!
Take this class if you'll enjoy it, it's not a requirement!
- * Bad choice if you don't tolerate contradictions:
 - (1) How come two models predict different results?
 - (2) What is THE model for network structure?
 - (3) Are data sufficient to represent a real social net?
 - (4) What is the ethical things to do?

How is your 1st social network class?



$$\begin{aligned}
& \text{cm} = 1 \times 10^8 \text{ N/m}^2 \quad R = 8.314 \text{ J/K} \quad k_B = 1.38 \times 10^{-23} \text{ J/K} = 8.62 \times 10^{-3} \text{ eV/K} \\
& N_A = 6.022 \times 10^{23} \quad \text{Water C} = 4.186 \frac{J}{g \cdot C^\circ} \quad L_f = 334 \text{ J/g} \quad L_v = 2260 \text{ J/g} \quad e = 1.6 \times 10^{-19} \text{ C} \\
& K_c = 9 \times 10^9 \text{ Nm}^2 / C^2 \quad c_0 = 8.85 \times 10^{-12} \frac{C^2}{Nm^2} \quad m_e = 9.1 \times 10^{-31} \text{ kg} \quad h = 6.63 \times 10^{-34} \text{ Js} = 2.03 \text{ eV} \cdot \text{s} \\
& c = 3 \times 10^8 \text{ m/s} \quad 1 \text{ eV} = 1.6 \times 10^{-19} \text{ J} \quad k_m = 1 \times 10^{-7} \frac{T \cdot m}{A} \quad \mu_0 = 4 \pi \mu_0 \\
& \text{Equations: } F = 9.81 \text{ N} + 32 \quad Q = mC\Delta T \quad Q = mL \quad Q_{\text{heat}} = Q_{\text{process}} \quad PV = nRT = Nk_B T \\
& P = F/A \quad E_{\text{internal}} = \frac{1}{2} (DOF) k_B T = \frac{1}{2} (DOF) RT \quad MC_p = \frac{1}{2} (DOF) R \quad Q = W + \Delta U \\
& W = \int P dV \quad MC_p - MC_v = R \quad \eta = 1 - \frac{Q_v}{Q_h} \leq 1 - \frac{T_v}{T_h} \quad \Delta S = \int \frac{dQ_{\text{rev}}}{T} = k_B \log \frac{W}{W_0} \\
& W = \frac{N!}{N_1! N_2! \dots} \quad N_i = C \exp \left[-\frac{E_i}{k_B T} \right] \quad C = \frac{N}{\sum_{i=1}^m \exp \left[-\frac{E_i}{k_B T} \right]} \quad N_1 = \frac{N}{1 + \exp \left[-\frac{E_1}{k_B T} \right]} \\
& P = \frac{1}{\exp \left[\left(\frac{E - \mu}{k_B T} \right) \right] + 1} \quad c_s = \frac{\hbar^2}{2m} (3\pi^2 N)^{2/3} \quad \hat{F} = \frac{k_B q_1 q_2}{r^2} \hat{r} \quad \hat{E} = \frac{k_B q}{r^2} \hat{r} \quad \hat{F} = q \hat{E} \\
& \Phi_E = \int \hat{E} \cdot d\hat{A} \quad \oint \hat{E} \cdot d\hat{A} = \frac{q}{\epsilon_0} \quad \Delta V = - \int_A^B \hat{E} \cdot d\hat{s} \quad \Delta U = q \Delta V \quad V = \frac{k_B q}{r} \quad U = \frac{k_B q_1 q_2}{r} \\
& E_x = - \frac{dV}{dx} \quad V = k_B \int \frac{dq}{r} \quad C = \frac{Q}{V} \quad C = \frac{q}{d} \quad U = \frac{1}{2} C V^2 \quad I = \frac{dq}{dt} \quad I = nqV A \\
& V = IR \quad R = \frac{\rho}{A} \quad \rho = \frac{m}{nq^2 r} \quad P = IV \quad \sum_{\text{loop}} \Delta V = 0 \quad \sum I_m = \sum I_{\text{out}} \quad q(t) = Q \left[1 - e^{-\frac{t}{\tau_{\text{rel}}}} \right] \\
& \hat{F}_m = q\hat{v} \times \hat{B} = \hat{I}l \times \hat{B} \quad r = \frac{mv}{qB} \quad \hat{r} = \hat{\mu} \times \hat{B} \quad U = -\hat{\mu} \cdot \hat{B} \quad \hat{d}\hat{B} = k_m I \frac{d\hat{l} \times \hat{r}}{r^3} \quad B = \frac{\mu_0 I}{2R} \\
& \oint \hat{B} \cdot d\hat{l} = \mu_0 I_{\text{out}} + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt} \quad B = \frac{\mu_0 I}{2\pi R} \quad B = \mu_0 \frac{N}{L} I \quad \oint \hat{E} \cdot d\hat{l} = 0 = -N \frac{d\Phi_E}{dt} = -L \frac{dI}{dt} \\
& \Phi_B = \int \hat{B} \cdot d\hat{A} \quad I = \frac{c}{\rho} \left[1 - e^{-\frac{r}{\tau_{\text{rel}}}} \right] \quad \lambda V = c \quad E_o = c B_o \quad \langle I \rangle = \frac{1}{2} c_o c E_o^2 \\
& \Phi_B = \left[\frac{B}{B_0} \frac{q_1 q_2}{2} \right] \quad 1 = \frac{1}{\tau_{\text{rel}}} \left[1 - e^{-\frac{r}{\tau_{\text{rel}}}} \right] \quad \lambda V = c \quad E = c B \quad \langle I \rangle = \frac{1}{2} c_o c E_o^2 \\
& \left(\frac{B}{B_0} + \frac{q_1 q_2}{2} \right) = \left(\frac{B}{B_0} \right)^{1.05} + \left(\frac{q_1 q_2}{2} \right)^{0.95} \quad B = \frac{5.96}{1.05} \quad B = 1.05 \frac{B_0}{1} \quad \left(\frac{B}{B_0} + \frac{q_1 q_2}{2} \right) = -N \frac{q_1 q_2}{2\pi R} = -\frac{q_1 q_2}{91} \\
& \frac{B}{B_0} = 1.05 \quad \frac{q_1 q_2}{2} = 0.95 \quad B = \frac{5.96}{1.05} \quad B = 1.05 \frac{B_0}{1} \quad \frac{B}{B_0} = 1.05 \quad \frac{q_1 q_2}{2} = 0.95 \quad B = \frac{5.96}{1.05}
\end{aligned}$$

Why NOT to take this class?

- * Social processes are (1) on graph, (2) messy
 - Not enough to understand definitions!
 - You need to work with definitions to prove results
- * No prerequisites per se, some important concepts:
 - Graph: vertices, edges, paths, shortest paths,
 - Probability: Expectation, Independence, Conditioning
 - Programming: one program. lang. (C, python, java)
- * This class is 6000 and heavy on analysis
- * Check HW1 (released soon) and come talk to me

More clarifications

- * This class is not “playing on FB and get credit for it”
 - although many of your friends will make fun of you
- * This class is not “how to build large web systems.”
 - although concepts we learn matter for that as well
 - take Prof. Geambasu’s class or Prof. Katz Bassets (EE) or another class on cloud computing
- * It relates but does not cover {Game Theory, ML, AI}
 - Very important topics, covered elsewhere,
 - We will only discuss these occasionally (again, no prerequisite)

How to succeed in this class?

- * Attend the lecture!
- * Come to our office hours

Wednesday 8:45-9:45am (Augustin), TBA (TAs)

Possible to attend remotely if requested before

- * Your grade: (there is no extra-credit)
 - ~3 homework assignments (proof/code/reason)
 - 1 midterm (short, in class)
 - 1 final “project” (similar to a call for proposal), 4~5p judging of your maturity in proposing a novel idea

A busy time for network

- * During this semester
 - Fairness of ML Symposium
 - MozFest in London
 - “Ethics of AI” College de France
 - NeurIPS
 - FAT* Program announced
 - Lots of deadline (WWW)
- * And more news that we can consume 😊

More on the course

- * On Piazza: Slides + Assignments + Readings + Additional References
- * Discussion.
- * To ask and answers all questions to the class and instructors
 - Please indicate before tomorrow Friday evening if you don't want to be enrolled automatically using your columbia.edu address
 - Unless otherwise specified you should plan for 24h maximum delay
- * **Integrity Rules:**
 - You will sign a cover-sheet for the first homework
 - Self respect to your future self, immediate respect for your comrades
- * **The “Apple” Policy:**



If you have to follow only 1 rule

- * Respect and support individuals in and out campus
 - As faculty, engineers, human beings, we have been in 2017 and the rise of the #MeToo movement. We know sexual harassment occurs (in campus and at conferences) in every discipline, including computer science.
 - We already spoke to support extraordinary courageous Columbia students (e.g., <http://on.wsj.com/2ENAQO2> NB: I understand that the material described in this article may be shocking and upsetting).
 - All of us received briefing, it is our duty to help all students and staff
 - Talk to anybody you feel comfortable to, we are prepared as a community it's our job too!



Suzanne Goldberg

Herbert and Doris Wechsler Clinical Professor of Law



Suzanne Goldberg, one of the country's foremost experts on gender and sexuality law and a leading advocate for the LGBTQ community, serves as the Herbert and Doris Wechsler Clinical Professor of Law. She also leads the Law School's Center for Gender and Sexuality Law and its Sexuality and Gender...
[READ MORE](#)



A bit about myself



- * Started research in 1999 in Bay area
... too young to get a beer!
- * @ ACM SIGCOMM/SIGMETRICS/WWW since 2000
- * Studied at ENS-INRIA in Paris (Ph.D in 2006)
 - Work @ Sprint (CA), Alcatel (Belg), IBM (NY), Intel (UK)
 - Worked 5 years for Technicolor (in Paris)
- * Interests: Fairness, Mobile, Social Networks, Privacy
 - Previously, models of TCP, Peer-to-peer, Human mobility
 - Emphasis on performance of networked algorithm

M O B I L E
S O C I A L
B

Not just me

- * Amazing students (typically from previous classes) applied as Teaching Assistant!
- * They will be announced and introduced in the next lecture!

Contents:

- Structure of Networks
 - * small world, weak tie, homophily, balance, ...
- Dynamics of crowds
 - * epidemics, influence, wisdom ...
- Applications
 - * crawling, ranking, crowdsourcing
- The 10 papers that will make you a social expert
- Some case studies, discussion within/outside lectures



Comic relief!

- * https://youtu.be/_MeQz8OWXQo
- * Also, recommended, on a similar topic
 1. the short but shrewd answers from Cambridge Analytica CEO
Exclusive: Alexander Nix - BBC Newsnight
<https://youtu.be/bTEJuLY1pWo>
 2. And the surprisingly under-reported article that had all these news in January 2017: H. Grassegger, M. Krogerus, The data that turned the world upside down, *Motherboard*, Jan. 28 2017
https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win