Lecture 0: Welcome to the class!

Who should take it? How to take it?

COMS 4995-1: Introduction to Social Networks
Tuesday September 4th
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
Can you recognize this picture?
“Only the paranoid survive”

What if …
- Current bubble explodes?
- Social media sold at its real market value?
- Social-networking not cool?

A. Grove (Intel cofounder)
All previous bubble episode were retrospectively very important to deploy game changer.

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

Early 2010s: peak of the social boom
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
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
Only convinced by numbers?

+40%  * How much data production grows / year
  o Enough to double every 24 months
    (72h of videos upload on YouTube in 1 min).

€260b  * How much data can save on health care
  o In Europe [McKinsey] (U.S. save $300b)

+300-1000%  * How much lifts improve when ads are using behavioral targeting
What are Social Networks?

* 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)
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 Science and Society at large: how to understand human behavior? Take advantage of it?

... 4 (classical) questions, being reinvented *today*
This “0” 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
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
   Interpretation case-studies
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!)
  o Social information is becoming the web’s hottest commodity (Google, Facebook, IBM, Telcos, Media)
  o Users’s data are company’s key differentiating factor
* You (not me) are the social media generation!
  o The game is just starting; it gets harder
  o CS deals with “complexity” deeply and elegantly
  o learning foundational concepts adds to your assets
“You think you are Google’s customer?
De facto, you are Google’s product!”
S. Vaidhyanathan (2011)

Who produces this oil?
Who makes money from it?
Why take this class: Reason #2

* This topic is fertile for research, here and at large
  o Big data is everywhere, especially in public-funding
  o Many of these data are networks connecting people
* Many opportunities within reach here at Columbia
  o Institute for Data science and Engineering (July 2012)
  o Brown Institute for Media Innovation (March 2012)
  o Lots of other departments
  o Overall, good topic to look for an academic job!
Computation/communication resources gradually become a commodity.

* But collecting/processing YOUR Data is not innocuous. *It may not pollute the air, but it may hurt your freedom, or your fair access to service/knowledge.*

Two issues need to be addressed:

1. The privacy tussle: How can a user keep track of what is known about her? And control it?
2. Is data a real currency? One you can own and use?
3. How to exploit them with better algorithms?
Why take this class: Reason #3

* The topic offers a great diversity of experience
  o You can hack, make proof, interview people, make money, (perhaps) save lives
  o If you have broad interests (or wonder what you would like), this could be a way to figure out
* Hence, diversity in this class is highly encouraged!
  o Special background, atypical majors/minors, are especially enriching to our discussions
  o Previously: 4 dept including Medical school
Before starting the trip
Before starting the trip

* There is no textbook!

Or maybe there are too much
Why NOT to take this class?

* Bad choice for maximum GPA / minimum workload
  o Combines programming, maths, interpretation
  o New, not a single textbook, may at time be unfamiliar
Take this class if you’ll enjoy it, it’s not a requirement!

* Bad choice if you don’t tolerate contradictions
  o How come two models predict different results?
    What is THE model for network structure?
At least, be reassured that the instructor sympathize
More clarifications

* This class is not “playing on FB and get credit for it”
  o although many of your friends may think it is

* This class is not “how to build large web systems.”
  o although some of what we see matter for it
  o Take Prof. Geambasu’s class or classes on cloud

* This class does not cover game theory / deep ML
  o Very important topics, covered elsewhere,
  o We will only discuss these occasionally (no prerequisite)
The topic is **broad**:

“CS-theory, Networking, Sociology, Physics”

- This is why the course focuses on algorithmic prop.

The topic seems (at times) **immature**:

“What is THE model? How to tell the cause?”

- Algorithmic research problems have an **impact**

Involves **some** mathematics & programming:

- Goal: **self-contained** (assuming elementary bases)
How to succeed in this class?

* Attend the 2 lectures (remotely, for CVN students)!
* Come to our office hours: easy, close to the course!
  Tuesday & Thursday 5:30-6:15pm (A. Chaintreau)
  Wednesday 3-4:30pm (C. Riederer)
  Possible to attend through Skype, Hang-out, if requested before

* Your grade: (no extra-credit)
  o 8~9 homework assignments
  o 1~2 blog posts (you can make teams) + participation
  o 1 (short) midterm (~Oct. 13th), 1 final (~Dec. 18-20th)
    Generally speaking: hard questions, generous grades
More on the course

  Slides + Assignments + Readings + Additional References
  To ask and answers all questions to the class and instructors
  Unless otherwise specified we promise you a 24h maximum delay
  Where you will write post and read the ones from your classmates

* Integrity Rules:
* The “Apple” Policy:
A bit about myself

* Started research in 1999 in Bay area
  
  ... I could not even order a beer then!

* Attended ACM SIGCOMM since 2000

* Studied at ENS-INRIA in Paris (Ph.D in 2006)
  - Interns at Sprint, Alcatel, IBM, Intel
  - Worked 5 years for Technicolor (formerly Thomson)

* Works on Mobile and Social Networks
  - Models of TCP, Peer-to-peer, Human mobility
  - Emphasis on performance of networked algorithm
Christopher J. Riederer
- 2nd year of Ph.D at Columbia CS
- Just came back from internship at MSR Redmond

Worked and works on a lot of cool stuff
- dancing robots, computational biology, gamification for economic development, average students’ faces
- these days, personal data and economics
- and computation journalism, within Brown Institute

BTW, you may meet more TAs (enrollment?)
Contents:

- Structure (September-October 15th)
  * small world, weak tie, power law, communities, …
  * pagerank, reinforcement, collaborative filtering, …
- Dynamics (October 15th-November)
  * epidemics, influence, crawl, wisdom of the crowds …
  * Gossip algorithms, submodular optimization …
- The 10 papers that will make you a social expert
- Some case studies, discussion within/outside lectures