

COMM-321

Big Data and Social Networks Research How Digital Technologies Shape Collective Behavior and Why it Matters

Spring 2015

Tuesday-Thursday, 1:30-3:00 pm, room 224

(Project discussions: Friday 3:00 to 4:00pm, Towne Building 311)

Instructor: Sandra González-Bailón, PhD
Annenberg School for Communication
sgonzalezbailon@asc.upenn.edu
www.asc.upenn.edu/sgonzalezbailon

Guest Lecturer: Victor M. Preciado, PhD
School of Engineering and Applied Science
preciado@seas.upenn.edu
<http://sites.google.com/site/victormpreciado>

Office hours: Tuesday-Thursday 3:00-4:00 pm (or by appointment)

Course Description

Online networks are prevalent in how we access and share information. They shape how we consume news, how we interact with friends, peers and strangers, and how we mobilize groups or try to gain the attention of large audiences. Online networks encourage dynamics of information exchange that can, potentially, be very consequential for the societies we live in or the societies we would like to build. At the same time, online networks are also creating new challenges for how we think about surveillance and censorship, freedom of speech and privacy, or mass collaboration and peer production. Networked technologies have, in brief, the potential to transform governance, collaboration, and organization, and radically change how we think about regulation and policy making in the digital age. Yet harnessing the potential of networked technologies requires a technical understanding of how networks operate, what dynamics they are more likely to encourage, and what features limit their ability to diffuse information or spread behavior. We need to understand the logic of networks before we can analyze their social impact or plan interventions –for instance, campaigns to spread news or encourage mobilization. This course aims to equip you with theoretical tools that help you unpack the black box of networks and assess the social implications of Big Data and online technologies.

Objectives

This research seminar aims to provide you with the theoretical and analytical tools to:

- Decode the logic of networks;
- Discuss the applications that derive from that knowledge;
- Engage in collaborative work across disciplinary divides.

The class discussions will introduce you to the building blocks of network science: we will cover both the social and technical aspects of network research and its applications in the context of digital technologies. Ultimately, the course aims to settle the foundations for multidisciplinary work: it will facilitate collaboration with engineering students, creating a space where you can learn from each other and engage in the design and delivery of joint projects. The ability to work in multidisciplinary teams is becoming increasingly important in the digital economy; this course will offer a training ground, and help you uncover the technological and social implications of networked technologies, especially as they intersect with policy and regulation. This course will give you the language and the tools to navigate that space comfortably, and it will allow you to think critically about current technologies, considering not only theoretical discussions but also the decisions behind the design and implementation of networked applications.

Assessment

Your final grade will be determined by your performance in four components:

1. Response essays (25 points each): you should write one page long, single-space essay before each session, summarizing and commenting on the corresponding reading. Essays should be submitted by 12 noon before each class.
2. Midterm paper (100 points): a 2000-2500 word essay (~5 pages) on the ethics of Big Data and network research is due by 1pm on Mar 17. The essay should answer the question: *What are the boundaries for ethical research in the networked age?* Your argument should consider the content discussed in class (including your reading summaries), as well as additional readings suggested through canvas.
3. Attendance and participation (100 points): this research seminar relies on your active participation in class, so your attendance is very important. For every session missed without proper justification (e.g. medical emergency), I will deduct 10 points.
4. Final research project (200 points): this component is designed to increase your ability to work in a multidisciplinary team. You will have to work with engineering students in the design and delivery of a research project around one of these five topics: (1) learning in social networks; (2) collective intelligence; (3) cascading behavior in networks; (4) information diffusion; and (5) network interventions. We will use the website <http://networks.asc.upenn.edu> to organize project management and discussions. In the evaluation of your work, two dimensions will be considered: first, the quality of your contributions to the project discussions, both online and in class (100 points); and, second, the quality of the final submitted project, co-authored with students in engineering (100 points).

The following tables summarize the evaluation criteria for this course and the final grade conversion:

T1. Evaluation Summary

Delivery	Points
Response essays	600
Midterm paper	100
Participation	100
Final project	200
Total	1000

T2. Grade Conversion

Points	Grade
951-1000	A+
901-950	A
851-900	A-
751-850	B+
701-750	B
651-700	B-
601-650	C+
551-600	C
501-550	C-
0-500	F

General Readings

The course discussions will rely on four books:

- Pentland, A. *Social physics: How good ideas spread—the lessons from a new science*. Penguin Press: New York: NY, 2014.
- Eagle, N.; Greene, K. *Reality mining: Using Big Data to engineer a better world*. MIT Press: Boston: MA, 2014.
- Watts, D.J. *Six degrees. The science of a connected age*. William Heinemann: London, 2003.
- Easley, D.; Kleinberg, J. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press: New York, NY, 2010.

Most sessions will consider a chapter in one of these books. Some sessions, however, will require other readings, as detailed in the list below.

Schedule

Weeks	Sessions	Topics	Deliveries
1	S1: Jan 15	• Introduction and course overview	
2	S2: Jan 20	• Social science and Big Data	RE-D1
	S3: Jan 22	• The connected age	RE-D2
3	S4: Jan 27	• From ideas to actions	RE-D3
	S5: Jan 29	• Graphs	RE-D4
4	S6: Feb 3	• Learning in networks	RE-D5
	S7: Feb 5	• Strong and weak ties	RE-D6
5	S8: Feb 10	• The flow of information and ideas	RE-D7
	S9: Feb 12	• Search in networks	RE-D8

6	S10: Feb 17	• Collective intelligence	RE-D9
	S11: Feb 19	• The madness of crowds	RE-D10
7	S12: Feb 24	• Interventions and change	RE-D11
	S13: Feb 26	• Networks in context	RE-D12
8	S14: Mar 3	• Data-driven societies	RE-D13
	S15: Mar 5	• The ethics of big data research	RE-D14
<i>Spring term break</i>			
9	S16: Mar 17	• Networks in small groups	MTP-D15 & RE-D16
	S17: Mar 19	• Virtual worlds, real behaviors	RE-D17
10	S18: Mar 24	• Urban analytics	RE-D18
	S19: Mar 26	• City science	RE-D19
11	S20: Mar 31	• Taking the pulse of society	RE-D20
	S21: Apr 2	• Designing technologies	RE-D21
12	S22: Apr 7	• Network interventions	RE-D22
	S23: Apr 9	• Network seeding and contagion	RE-D23
13	S24: Apr 14	• Risks of a data-driven society	RE-D24
	S25: Apr 16	• The messiness of human behavior	RE-D25
14	S26: Apr 21	• Project Presentations	
	S27: Apr 23	• Project Presentations	
15	S28: Apr 28	• Project Presentations	
16	<i>no classes</i>		FRP-D26

Legend: RE stands for 'response essay'; D stands for 'delivery'; MTP stands for 'midterm paper'; FRP stands for 'final research project'.

Weekly Readings

Week 1

S1 Course introduction

Week 2

S2 Social Science and Big Data

What is Big Data? Why is it changing research, decision making, and policy recommendations?

Reading:

González-Bailón, S. Social science in the era of Big Data. *Policy & Internet* 2013, 5, 147-160.

S3 The Connected Age

Why are networks important to understand collective dynamics? How do digital technologies help us improve our depiction and analysis of social networks?

Reading:

Watts, D.J. 2003. “The Connected Age”, chapter 1 of *Six Degrees*.

Week 3

S4 From Ideas to Actions

How can we use Big Data to analyze and intervene in groups and societies?

Reading:

Pentland, A. 2014. “From ideas to actions. Using Big Data to understand how human societies evolve”, chapter 1 of *Social Physics*.

S5 Graphs

How can we represent social networks for analysis? What are the building blocks of network structure?

Reading:

Easley, D.; Kleinberg, J. 2010. “Graphs”, chapter 2 of *Networks, Crowds and Markets*.

Week 4

S6 Learning in Networks

How is access to information mediated by networks? How can we harness the power of networks to improve decision making?

Reading:

Pentland, A. 2014. “Exploration. How can we find good ideas and make good decisions?”, chapter 2 of *Social Physics*.

S7 Strong and Weak Ties

Why are some network ties stronger than others? What are the implications for the flow of information?

Reading:

Easley, D.; Kleinberg, J. 2010. “Strong and weak ties”, chapter 3 of *Networks, Crowds and Markets*.

Week 5

S8 The Flow of Information and Ideas

Why does the flow of information in networks matter to understand individual and collective behavior? What type of network structure is more conducive to information flows?

Reading:

Pentland, A. 2014. “Idea flow. The building blocks of collective intelligence”, chapter 3 of *Social Physics*.

S9 Search in Networks

How do people navigate their networks? How easy is it to find shortcuts in a graph managing only local information?

Reading:

Watts, D.J. 2003. "Search in networks", chapter 5 of *Six Degrees*.

Week 6

S10 Collective Intelligence

Why are groups often smarter than individuals? Under what conditions does collective intelligence arise?

Reading:

Pentland, A. 2014. "Collective intelligence. How patterns of interaction translate into collective intelligence", chapter 5 of *Social Physics*.

S11 The Madness of Crowds

Are groups always smarter than individuals? What is 'herding' and how does it affect the quality of collective decision making?

Reading:

Watts, D.J. 2003. "Decisions, delusions, and the madness of crowds", chapter 7 of *Six Degrees*.

Week 7

S12 Interventions and Change

How can we tune networks to encourage change in behavior? Why should we nudge individuals to rewire their connections?

Reading:

Pentland, A. 2014. "Organizational change", chapter 7 of *Social Physics*.

S13 Networks in Context

How does social context shape the structure of networks? How does it help us understand how people form their connections?

Reading:

Easley, D.; Kleinberg, J. 2010. "Networks in their surrounding contexts", chapter 4 of *Networks, Crowds and Markets*.

Week 8

S14 Data-driven societies

How can the insights gained from network data be used for policy interventions and collective decision making? Do they allow more robust, evidence-based governance?

Reading:

Pentland, A. 2014. “Data-driven societies. What will a data-driven future look like?”, chapter 10 of *Social Physics*.

S15 The Ethics of Big Data Research

What are the ethical implications of Big Data research? What are the challenges that digital technologies present to researchers?

Reading:

boyd, d.; Crawford, K. 2012. “Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15, 662-679.

Spring term break

Week 9

S16 Networks in Small Groups

How can digital technologies help us gather Big Data about small group dynamics? What kind of insights can be gained from that data?

Reading:

Eagle, N.; Greene, K. 2014. “Gathering data from small heterogeneous groups”, chapter 3 of *Reality mining*.

S17 Virtual Worlds, Real Behaviors

How do internet technologies change the way in which we communicate and interact with each other? What do virtual networks tell us about the social world?

Reading:

Christakis, N.A.; Fowler, J.H. 2009. “Hyperconnected”, chapter 8 in *Connected. The surprising power of our social networks and how they shape our lives*. Little, Brown & Company: New York, NY.

Week 10

S18 Urban Analytics

What type of observational data is being collected at the level of cities? Does Big Data help govern and regulate urban life?

Reading:

Eagle, N.; Greene, K. 2014. “Traffic Data, Crime Stats, and Closed-Circuit Cameras: Accumulating Urban Analytics”, chapter 5 of *Reality mining*.

S19 City Science

Have cities become an object of scientific inquiry? What do we mean when we talk about smart cities?

Reading:

Pentland, A. 2014. "City Science. How social physics and Big Data are revolutionizing our understanding of cities and development", chapter 9 of *Social Physics*.

Week 11

S20 Taking the Pulse of Society

Are digital technologies creating a 'nervous system' for society? What does activity in online networks reveal about social life?

Reading:

Eagle, N.; Greene, K. 2014. "Taking the pulse of a nation: census, mobile phones, and Internet giants", chapter 7 of *Reality mining*.

S21 Designing Technologies

How does Big Data research feed back into the design of digital technologies? Can technology nudge users to behave in ways that benefit the collective?

Reading:

Pentland, A. 2014. "Design for harmony. How social physics can help us design a human-centric society", chapter 11 of *Social Physics*.

Week 12

S22 Network Interventions

Can we change the way in which networks operate or how people use those networks? In what scenarios is intervention justified?

Reading:

Valente, T.W. 2012. Network interventions. *Science*, 337, 49-53.

S23 Network Seeding and Contagion

Can we engineer social contagion in networks? What do we mean when we talk about "seeding" a network?

Reading:

Aral, S.; Muchnik, L.; Sundararajan, A. 2013. Engineering social contagions: Optimal network seeding in the presence of homophily. *Network Science*, 1, 125-153.

Week 13

S24 Risks of a Data-Driven Society

What are the risks of using Big Data for prediction and decision making? What can we do to control or limit those risks?

Reading:

Mayer-Schoenberger, V.; Cukier, K. 2013. "Risks", chapter 8 of *Big data. A revolution that will transform how we live, work and think*. John Murray: London.

S25 The Messiness of Human Behavior

What are the limitations of data driven approaches when it comes to understanding human behavior? What kind of problems are outside of reach for Big Data research?

Reading:

Watts, D.J. 2011. "The proper study of mankind", chapter 10 of *Everything is obvious once you know the answer. How common sense fails us*. Crown Business: New York, NY.

Week 14

S26 Project presentations

S27 Project presentations

Week 15

S28 Project presentations