ALGORITHMS & BIG DATA

Methods & Controversies

Prof Mike Ananny ananny@usc.edu mike.ananny.org @ananny

June 1 – June 12, 2015 10am – 1pm

Meeting Room: ANN-405
Office Hours: By appointment (ANN-310B)

Communication research is increasingly concerned with, and inseparable from, the technologies and cultures producing information algorithms and large-scale data sets. Whether *studying* or *using* algorithms or "big data", researchers need to understand how their research questions intersect with the logics of automation and scale underpinning networked, computational platforms. In this workshop, students will: analyze emerging, critical literature on the technologies and cultures of big data/algorithm research; get hands-on experience with big data/algorithmic tools, platforms, and activities (no programming skills required); discuss recent big data/algorithm controversies, and examine how researchers have responded; develop a short research proposal on some aspect of big data/algorithms they find personally valuable. Students should leave the workshop with an appreciation of the challenges of using and researching big data/algorithms, a framework for understanding future big data/algorithm controversies, and the beginnings of a research project to pursue beyond the workshop.

TEXTS & COURSE MATERIALS

Borgman, C. (2015). Big data, little data, no data: Scholarship in the networked world. Cambridge, MA: MIT Press.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Recommended:

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Lonodn, UK: SAGE.

All other texts will be provided as PDFs or URLs.

EXPECTATIONS

In an intensive workshop like this it's critical that you attend all sessions, arrive ready to participate actively, and commit to creating work products that are high-quality, personally meaningful, and communicated to the group. To that end, the following is expected of each student:

• Daily Discussion Questions

By 9am every day (except the first and last days), students should email ONE discussion question to the entire class. The question can be on virtually any aspect of the course but will ideally relate to that day's theme or compare/contrast the assigned materials. It should help set the tone for that day's meeting and offer a starting point for a class conversation.

Active Participation & Attentiveness

In <u>every</u> class, students should have their phones off, use laptops for course business only, help shape discussions and enrich conversations, and treat all class members with thoughtfulness and respect.

Opening a "Foundational" Reading

Twice (2x) each student will "open" a "foundational" readings. I'll say more about this in the first class, but opening a reading does NOT mean simply summarizing the text (you can assume everyone has read it). It means using the text to ground a discussion, surface interpretations, and develop the class's collective intelligence. You don't need to prepare a handout or a formal presentation, just lead a discussion and create a rich conversation.

Research Project Proposal

On the last day students will *individually* submit an extended abstract of a research proposal (1-2 pages) and give a 12-15 minute presentation proposing a research project. The proposal should succinctly state 5 things:

- 1. A research question;
- 2. The project's stakes (*i.e.*, conceptually and empirically grounded answers to questions like "why does this project matter?", "how are you a good person to study it?", and "who might fund it?");
- 3. A brief review of relevant bodies of literature (there is no time for a full literature review but you should identify key citations, areas of inquiry, and related projects);
- 4. A methodological approach, and a discussion of its power and challenges;
- 5. An interpretation of at least one of the project's anticipated or potential findings (i.e., "finding X would be significant in way Y").

I do not expect this to be a *fully* developed proposal of the sort you'd write for your dissertation or a polished piece of prose (you can even write in outline form if you like), but it should establish your project's viability, and be a starting point for future work that would make sense to someone who wasn't in the class. Since time is compressed, you should start thinking about this project and sketching ideas immediately. I might periodically ask students to give updates.

SCHEDULE

- 1. Depending on the course pace and student interests, I might change some of the below. If so, I'll give enough notice.
- 2. On some days more than one activity is listed, to accommodate students with different levels of technical ability.

Monday, June 1 :: INTRODUCTION ::

Turner, J. H. (2005). A new approach for theoretically integrating micro and macro analysis. In C. Calhoun, C. Rojek & B. S. Turner (Eds.), The Sage handbook of sociology (pp. 405-422). London, UK: Sage.

→ when reading this article, think about: (1) what kind of "theories of the middle range" (Merton) are being created or assumed by research on big data and algorithms; (2) what the "unit of analysis" is in your own work and your interest in big data / algorithms.

Wallach, H. (2014, December 19, 2014). Big data, machine learning, and the social sciences: Fairness, accountability, and transparency. *Medium*. Retrieved December 20, 2014, from https://medium.com/@hannawallach/big-data-machine-learning-and-the-social-sciences-927a8e20460d

Come with two questions about big data / algorithms.

In-Class Activity:

Build "If This Then That" feeds; reflect on process/assumptions of creating rule structures to traverse a data space of unknown size.

Turadau luna 2	
Tuesday, June 2	
Foundations (read all)	:: WHAT IS "BIG DATA" & WHY DOES IT MATTER? :: Applications (choose 1 or 2 and be ready to discuss)
•	
boyd, d., & Crawford, K. (2012). Critical	Anderson, C. (2008, June 23, 2008). The end of theory: The data deluge makes the scientific
questions for big data. <i>Information</i> ,	method obsolete. Wired Magazine. Retrieved March 4, 2010, from
Communication & Society, 15(5), 662-679.	http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory
Borgman, C. (2015). Big data, little data, no	Ford, H. (2014). Big Data and small: Collaborations between ethnographers and data scientists. Big
data: Scholarship in the networked world.	Data & Society, 1(2), 1-3. doi: 10.1177/205395171454433
Cambridge, MA: MIT Press. → Chapters 1-	
4 only	Burrell, J. (2012a, May 28, 2012). The ethnographer's complete guide to big data: Answers (part 2
,	of 3). Ethnography Matters. Retrieved December 2, 2014, from
Shah, D. V., Cappella, J. N., & Neuman, W.	http://ethnographymatters.net/blog/2012/06/11/the-ethnographers-complete-guide-to-big-data-
R. (2015). Big Data, Digital Media, and	part-ii-answers/
Computational Social Science: Possibilities	Bell, G. (2012, April 3, 2012). Anthropologist for Intel describes big data as a person <i>Tech Web TV</i> .
and Perils. The ANNALS of the American	Retrieved December 3, 2014, from https://www.youtube.com/watch?v=WVB6_QP_2s0
Academy of Political and Social Science,	
<i>659</i> (1), 6-13. doi:	Miller, G. (2014, December 8, 2014). The huge, unseen operation behind the accuracy of Google
10.1177/0002716215572084	Maps. Wired. Retrieved December 18, 2014, from http://www.wired.com/2014/12/google-maps-
	ground-truth/
	Healy Kieron (2012) June 0, 2012) Heing metadata to find Dayl Boyers, Betrioved August 20, 2014
	Healy, Kieran. (2013, June 9, 2013). Using metadata to find Paul Revere. Retrieved August 20, 2014,
	from http://kieranhealy.org/blog/archives/2013/06/09/using-metadata-to-find-paul-revere/
	The Economist. (2012, October 12, 2012). Big data and the democratisation of decisions. The
	Economist. Retrieved May 31, 2015, from http://www.economistinsights.com/technology-
	innovation/analysis/big-data-and-democratisation-decisions

In-Class Activities:

Build a web scraper and/or Twitter bot, and reflect on process/assumptions. OR

Use Overview project (https://www.overviewproject.org) to create / analyze a document data set (register for free account).

Wednesday, June 3		
:: WHAT ARE ALGORITHMS AND WHY DO THEY MATTER? ::		
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)	
	Baker, P., & Potts, A. (2013). 'Why do white people have thin lips?' Google and the perpetuation of stereotypes via auto-complete search forms. <i>Critical Discourse Studies, 10</i> (2), 187-204. doi: 10.1080/17405904.2012.744320	

Play the "silent game" (from MIT NSF project) and/or take an "algorithmic walk" (designed by Malte Ziewitz)

Thursday, June 4 :: SOCIOTECHNICAL PRACTICES: MAKING SENSE OF BIG DATA / ALGORITHM TRACES IN LIGHT OF WHAT PEOPLE DO & THINK ::	
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)
Hargittai, E. (2015). Is Bigger Always	Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & boyd, d. (2011). The revolutions were
Better? Potential Biases of Big Data	tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. International
Derived from Social Network Sites. The	Journal of Communication, 5, 1375–1405.
ANNALS of the American Academy of	
Political and Social Science, 659(1), 63-76.	Vanderbilt, T. (2013, August 7, 2013). The science behind the Netflix algorithms that decide what
doi: 10.1177/0002716215570866	you'll watch next. Wired. Retrieved August 20, 2014, from
	http://www.wired.com/2013/08/qq_netflix-algorithm/
Marres, N., & Weltevrede, E. (2013).	
Scraping the social? Issues in live social	Papacharissi, Z., & Oliveira, M. (2012). Affective news and networked publics: The rhythms of news
research. Journal of Cultural Economy,	storytelling on #Egypt. <i>Journal of Communication, 62</i> (2), 266-282.
6(3), 313-335. doi:	
10.1080/17530350.2013.772070	Brogan, J. (2015, May 13, 2015). The case of the ornamental anthropologist. <i>Slate</i> . Retrieved May
Falami M. Dialman A. Vasanna K	17, 2015, from
Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K.,	http://www.slate.com/articles/technology/future_tense/2015/05/netflix_tries_to_put_a_human_f
Sandvig, C. (2015). "I always assumed that	ace_on_big_data_with_its_own_anthropologist.single.html
I wasn't really that close to [her]":	Madrigal, A. (2014, April 14, 2014). Behind the machine's back: How social media users avoid
Reasoning about invisible algorithms in the	getting turned into big data. <i>The Atlantic</i> . Retrieved June 3, 2014, from
news feed. Paper presented at the CHI	http://www.theatlantic.com/technology/archive/2014/04/behind-the-machines-back-how-social-
2015, Seoul, Republic of Korea.	media-users-avoid-getting-turned-into-big-data/360416/
2013) Seodi, Republic of Roled.	intedia discrib divola getting tarried into sig data/500 (12)
	Geiger, R. S., & Ribes, D. (2010). The work of sustaining order in Wikipedia: The banning of a
	vandal. Paper presented at the Computer Supported Cooperative Work, Savannah, Georgia.
	http://www.pensivepuffin.com/dwmcphd/syllabi/info447_wi12/readings/wk05-
	ConflictInCollaborations/geiger.BanningAVandal.CSCW10.pdf
In-Class Activity:	

Exploring a "trace ethnography" approach to studying algorithms and big data.

Friday, June 5 :: WORKING WITH BIG(ISH) DATA – SAMPLING & CONNECTING CONTEXTS ::			
[Visitor: Prof Kjerstin Thorson, USC]			
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)		
Driscoll, K., & Thorson, K. (2015). Searching and Clustering Methodologies: Connecting	Thorson, K., Driscoll, K., Ekdale, B., Edgerly, S., Thompson, L. G., Schrock, A., Wells, C. (2013). YouTube, Twitter, and the Occupy movement. <i>Information, Communication & Society, 16</i> (3), 421-		
Political Communication Content across	451. doi: 10.1080/1369118X.2012.756051		
Platforms. The ANNALS of the American			
Academy of Political and Social Science,	Butler, D. (2013, February 13, 2013). When Google got flu wrong. <i>Nature</i> . Retrieved May 2, 2014,		
659(1), 134-148. doi: 10.1177/0002716215570570	from http://www.nature.com/news/when-google-got-flu-wrong-1.12413		
	Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big		
Driscoll, K., & Walker, S. (2014). Working within a black box: Transparency in the	data analysis. <i>Science, 343,</i> 1203-1205.		
collection and production of big Twitter	Ananny, M. (2011, April 14, 2011). The curious connection between apps for gay men and sex		
data. International Journal of	offenders. The Atlantic. Retrieved January 8, 2014, from		
Communication, 8, 1745–1764.	http://www.theatlantic.com/technology/archive/2011/04/the-curious-connection-between-apps-		
	for-gay-men-and-sex-offenders/237340/		
	Owen, T. (2015, May 25, 2015). The violence of algorithms: Why big data is only as smart as those who generate it. <i>Foreign Affairs</i> . Retrieved May 25, 2015, from		
	https://www.foreignaffairs.com/articles/2015-05-25/violence-algorithms		
	inteps, j www.ioreignatians.com/articles/2015/05/25/violence algorithms		

Working with DiscoverText, exploring its power and methodological challenges.

Monday, June 8 :: AUDITING ALGORITHMS: THE CHALLENGE & CONTESTED VALUE OF SYSTEM TRANSPARENCY ::		
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)	
Sandvig, C., Hamilton, K.,	Felten, E. (2012, September 12, 2012). Accountable algorithms. <i>Freedom to tinker</i> . Retrieved May 2, 2013,	
Karahalios, K., & Langbort, C.	from https://freedom-to-tinker.com/blog/felten/accountable-algorithms/	
(2014). Auditing algorithms:		
Research methods for detecting	Diakopoulos, N. (2015, April 27, 2015). Towards a standard for algorithmic transparency in the media. <i>Tow</i>	
discrimination on internet	Center for Digital Journalism. Retrieved April 29, 2015, from http://towcenter.org/towards-a-standard-for-	
platforms. Paper presented at the	algorithmic-transparency-in-the-media/	
Data and Discrimination:		
Converting Critical Concerns into	Schudson, M., & Sonnevend, J. (2010, February, 2010). Beyond transparency: Is more information always a	
Productive: A preconference at	good thing? Columbia Journalism Review. Retrieved March 30, 2015, from	
the 64th Annual Meeting of the	http://www.cjr.org/the_research_report/beyond_transparency.php	
International Communication		
Association, Seattle, WA.	Zara, C. (2015, April 9, 2015). FTC chief technologist Ashkan Soltani On algorithmic transparency and the	
	fight against biased bots. International Business Times. Retrieved May 1, 2015, from	
Hamilton, K., Sandvig, C.,	http://www.ibtimes.com/ftc-chief-technologist-ashkan-soltani-algorithmic-transparency-fight-against-	
Karahalios, K., & Eslami, M.	<u>biased-1876177</u>	
(2014). A path to understanding		
the effects of algorithm	Lessig, L. (2009, October 9, 2009). Against transparency: The perils of open government. <i>The New Republic</i> .	
awareness. Paper presented at	Retrieved August 17, 2010, from http://www.tnr.com/article/books-and-arts/against-transparency	
CHI 2014, Toronto, ON.		
	Brill, J. (2015, February 28, 2015). Scalable approaches to transparency and accountability in decisionmaking	
Seaver, N. (2014, nd). On reverse	algorithms: Remarks at the NYU conference on algorithms and accountability. Federal Trade Commission.	
engineering: Looking for the	Retrieved April 3, 2015, from	
cultural work of engineers.	https://www.ftc.gov/system/files/documents/public_statements/629681/150228nyualgorithms.pdf	
Medium. Retrieved April 3, 2014,		
from	Madrigal, A. C. (2015, March 27, 2015). Many, many Facebook users still don't know that their news feeds	
https://medium.com/anthropolo	are filtered by an algorithm. Fusion. Retrieved March 29, 2015, from http://fusion.net/story/110543/most-	
gy-and-algorithms/d9f5bae87812	facebook-users-still-dont-know-that-their-news-feeds-are-filtered-by-an-algorithm/	

Using one or more algorithmic auditing techniques outlined in foundation readings, design and/or conduct comparative algorithmic audits of at least two social media platforms.

Tuesday, June 9		
:: ETHNOGRAPHY OF DATA & ANALYTICS IN AN INSTITUTIONAL CONTEXT – NETWORKED NEWS ::		
[Visitor: Caitlin Petre, NYU]		
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)	
Petre, C. (2015, May 7, 2015). The	Ananny, M., & Crawford, K. (2014). A liminal press: Situating news app designers within a field of	
traffic factories: Metrics at	networked news production. <i>Digital Journalism, 3</i> (2), 192-208. doi: 10.1080/21670811.2014.922322	
Chartbeat, Gawker Media, and The		
New York Times. Tow Center for	O'Donovan, C. (2014, July 8, 2014). Q&A: Tarleton Gillespie says algorithms may be new, but editorial	
Digital Journalism. Retrieved May	calculations aren't. <i>Nieman Lab</i> . Retrieved July 9, 2014, from http://www.niemanlab.org/2014/07/qa-	
10, 2015, from	tarleton-gillespie-says-algorithms-may-be-new-but-editorial-calculations-arent/	
http://towcenter.org/research/traff		
<u>ic-factories/</u>	Somaiya, R. (2014, October 26, 2014). How Facebook is changing the way its users consume journalism.	
ONE OF	The New York Times. Retrieved December 27, 2014, from	
ONE OF:	http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-	
Parasie, S. (2014). Data-driven	<u>consume-journalism.html</u>	
revelation? Epistemological		
tensions in investigative journalism	Goel, V., & Somaiya, R. (2015, May 13, 2015). Facebook begins testing instant articles from news	
in the age of "big data". <i>Digital</i>	publishers. The New York Times. Retrieved May 14, 2015, from	
Journalism. doi:	http://www.nytimes.com/2015/05/13/technology/facebook-media-venture-to-include-nbc-buzzfeed-	
10.1080/21670811.2014.976408	and-new-york-times.html	
,		
OR	Christin, A. (2014, August 28, 2014). When it comes to chasing clicks, journalists say one thing but feel	
Co. Los - NA (2014) The collection	pressure to do another. <i>Nieman Lab</i> . Retrieved August 30, 2014, from	
Carlson, M. (2014). The robotic	http://www.niemanlab.org/2014/08/when-it-comes-to-chasing-clicks-journalists-say-one-thing-but-feel-	
reporter: Automated journalism	pressure-to-do-another/	
and the redefinition of labor,	La serve NA (2014 August 10, 2014) Tuittenus Feesback en anna Fersina de sur the	
compositional forms, and	Ingram, M. (2014, August 18, 2014). Twitter vs. Facebook as a news source: Ferguson shows the	
journalistic authority. Digital	downsides of an algorithmic filter. GigaOM. Retrieved January 5, 2015, from	
Journalism. doi: 10.1080/21670811.2014.976412	https://gigaom.com/2014/08/18/twitter-vs-facebook-as-a-news-source-ferguson-shows-the-downsides-of-an-algorithmic-filter/	
10.1000/210/0011.2014.9/0412	<u>Or-an-aigoritinnit-inter/</u>	

- 1. Caitlin Petre visits to talk about studying data analytics at Chartbeat, Gawker, and The New York Times.
- 2. Use Newsdiff (http://newsdiffs.org/), NPR API Query Generator (http://pageonex.com/), NPR API Query Generator (http://pageonex.com/)

Wednesday, June 10 :: MAKING SENSE OF BIG DATA HOLES & ALGORITHM RESISTANCE – AND SPECULATIVE DESIGN AS A RESEARCH METHOD ::		
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)	
Lerman, J. (2013). Big data and its exclusions. Stanford Law Review, 66, 55-63. Gangadharan, S. P. (Ed.). (2014). Data and discrimination: Collected essays: Open Technology Institute & New America. →	Maddock, J., Starbird, K., & Mason, R. (2015). Using historical Twitter data for research: Ethical challenges of tweet deletions. <i>Paper presented at the Computer Supported Cooperative Work workshop: Ethics for studying sociotechnical systems in a big data world</i> , Vancouver, BC. https://cscwethics2015.files.wordpress.com/2015/02/maddock.pdf Berg, N. (2014, June 25, 2014). <i>Predicting crime</i> , LAPD-style. <i>The Guardian</i> . Retrieved August 20, 2014. <i>Incomplete Management of the Management and Japanese Management</i> .	
skim and pick a few you're interested in (they're very short essays)	2014, from http://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report	
Brunton, F., & Nissenbaum, H. (2011). Vernacular resistance to data collection and analysis: A political theory of obfuscation. <i>First Monday, 16</i> (5).	Crawford, K. (2013, April 1, 2013). The hidden biases in big data. <i>Harvard Business Review</i> . Retrieved August 27, 2013, from http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html	
	Vertesi, J. (2014, May 1, 2014). My experiment opting out of big data made me look like a criminal. Time. Retrieved May 1, 2014, from http://time.com/83200/privacy-internet-big-data-opt-out/ Crawford, K., & Schultz, J. (2013). Big data and due process: Toward a framework to redress	
In Class Astinitus	predictive privacy harms. Boston College Law Review, 55(1).	

Using speculative design and design fiction techniques to create big data / algorithm concept prototypes that help surface meanings of resistance and opting out.

Thursday, June 11 :: STUDYING (WITH) BIG DATA ALGORITHMS – THE ETHICS OF "DATA SCIENCE" & AN EMERGING PROFESSION ::		
Foundations (read all)	Applications (choose 1 or 2 and be ready to discuss)	
Shilton, K., Butler, B., Goggins, S., & Winter, S.	Meyer, R. (2014, September 8, 2014). Everything we know about Facebook's secret mood	
(2015). Research ethics for open online community	manipulation experiment. <i>The Atlantic</i> . Retrieved December 3, 2014, from	
data: A case study of human subjects research	http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-	
online. Paper presented at the Computer	facebooks-secret-mood-manipulation-experiment/373648/	
Supported Cooperative Work workshop: Ethics for		
studying sociotechnical systems in a big data	Auerbach, D. (2015, May 18, 2015). The silicon tower. Salon. Retrieved May 20, 2015,	
world, Vancouver, BC.	from	
https://cscwethics2015.files.wordpress.com/2015/	http://www.slate.com/articles/technology/bitwise/2015/05/facebook_study_why_silicon_	
02/shilton.pdf	valley_s_incursion_into_academic_research_is.single.html	
Gray, M. L. (2014, August 19, 2014). Microsoft	Grimmelmann, J. (2015, May 27, 2015). Do you consent? Salon. Retrieved May 28, 2015,	
Research faculty summit 2014 ethics panel recap.	from	
Social Media Collective. Retrieved October 2,	http://www.slate.com/articles/technology/future_tense/2015/05/facebook_emotion_con	
2014, from	tagion_study_tech_companies_need_irb_review.html	
http://socialmediacollective.org/2014/08/19/msr-		
faculty-summit-2014-ethics-panel-recap/	Gillespie, T. (2014, July 4, 2014). Facebook's algorithm: Why our assumptions are wrong,	
	and our concerns are right. Culture Digitally. Retrieved July 5, 2014, from	
Markham, Annette, & Buchanan, Elizabeth. (2012).	http://culturedigitally.org/2014/07/facebooks-algorithm-why-our-assumptions-are-wrong-	
Ethical Decision-Making and Internet Research:	and-our-concerns-are-right/	
Recommendations from the AoIR Ethics Working		
Committee. Association of Internet Researchers.	Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, AL., Brewer, D., Van Alstyne, M.	
2 nd revision. Retrieved October 19, 2014, from	(2009). Computational Social Science. <i>Science</i> , 323(5915), 721-723. doi:	
http://aoir.org/reports/ethics2.pdf	10.1126/science.1167742	
In-Class Activity:		

Sketch interdisciplinary code of ethics for: studying big data and algorithms; building "data factories"; and creating a "data science" profession.

Friday, June 12 :: WRAP-UP & PROPOSAL PRESENTATIONS ::