

CRITICAL STUDIES OF NETWORKED INFRASTRUCTURES

A focus on methods and controversies in the study of “big data” & algorithms

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

May 23 – June 3, 2016
(no meeting on Memorial Day, May 30th)
10am – 1pm

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

This workshop introduces students to the critical analysis of networked communication infrastructures, focusing on the diverse methods and epistemologies used to study information algorithms and “big data”. Combining approaches from Communication and Science & Technology Studies, students will: critically read emerging literature on technologies and cultures of big data/algorithm research; examine and design alternations to case studies of big data/algorithms; get hands-on experience with big data/algorithms exercises (no programming experience necessary); examine methodological controversies and ethical tensions in big data/algorithms research; and develop a short pitch (research question, relevant concepts, required resources) to study an aspect of big data/algorithms that they might pursue beyond the workshop. (Note: the workshop is purposefully exploratory and interdisciplinary and is not designed to be tutorials in statistical or network analysis software.)

I don't think there's any single method for studying “big data” and algorithms (nor should there be). As Communication scholars working at interdisciplinary boundaries, our collective (and hopefully fun?) exercise in this workshop is to try applying infrastructural studies approaches developed by Science & Technology Studies to algorithms and data – as objects and sites of study that we will probably first see as *communication* media, institutions, and practices. These objects and sites are all very much in flux and resist being described through any single method; our meta-task is to recognize the epistemological trade-offs scholars make when studying them and resist prematurely collapsing complexities as we look at the work of others and plan our own future projects.

TEXTS & COURSE MATERIALS

All course materials are provided either as books purchased for you or as PDFs. We unfortunately won't have time to cover all of these resources (I've had to make some painful choices and exclude texts and authors that are fantastic), but the following are recent books we'll draw on either implicitly as background, or explicitly in the schedule:

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

Bouk, D. (2015). *How our days became numbered: Risk and the rise of the statistical individual*. Chicago, IL: University of Chicago Press.

Gillespie, T., Boczkowski, P. J., & Foot, K. A. (Eds.). (2014). *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press.

- Gitelman, L. (Ed.). (2013). *"Raw Data" is an oxymoron*. Cambridge, MA: MIT Press.
- Hargittai, E., & Sandvig, C. (Eds.). (2015). *Digital research confidential*. Cambridge, MA: MIT Press.
- Howard, P. (2015). *Pax Technica: How the Internet of Things may set us free or lock us up*. New Haven, CT: Yale University Press.
- Igo, S. (2007). *The averaged American: Surveys, citizens, and the making of a mass public*. Cambridge, MA: Harvard University Press.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London, UK: SAGE.
- Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Eamon Dolan/Mariner Books.
- Mosco, V. (2014). *To the cloud*. London, UK: Routledge.
- Nafus, D. (Ed.). (2016). *Quantified: Biosensing technologies in everyday life*. Cambridge, MA: MIT Press.
- Neff, G., & Nafus, D. (2016). *Self-tracking*. Cambridge, MA: MIT Press.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
- Peters, J. D. (2015). *The marvelous clouds*. Chicago, IL: University of Chicago Press.
- Sismondo, S. (2009). *An introduction to science and technology studies* (2nd ed.). London, UK: Wiley-Blackwell.

Finally, this is more of a meta-recommendation, but please see the syllabus for Prof Christian Sandvig's "Planning for First-Year Research Project", <http://www-personal.umich.edu/~csandvig/698F15/> for a particularly helpful overview of the challenges—and fun!—of creating a doctoral research project.

EXPECTATIONS

In an intensive workshop it's critical that you attend all sessions, arrive ready to participate, and create work products that are high-quality, personally meaningful, and shared with the group. A note on the readings: **It looks like there is way too much reading, but please, don't freak out.** I know that since we're meeting every day and you're likely taking another ASIMS workshop, there isn't a lot of time to read. To that end:

- Many of the readings are short, commentary style, and not full-length journal articles.
- I've generally listed the "foundational" readings in order of importance so if you run out of time, you'll likely have encountered the most critical ideas.
- Use this as a chance to practice reading relatively quickly – skim articles at first, get a sense of their shape and argument, and build some familiarity. It's okay to read deeply selectively since different people will focus on different parts of articles and we'll be able to teach each other.

The following is expected of each student:

- **Daily Discussion Questions**
By 9am every day, 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 and materials. It should help set the tone for that day's meeting and offer starting points for class conversations.

- **Active Participation & Attentiveness**

In every 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**

Once (1x) during the course, 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 Pitch**

On the last day students will individually give a 12-15 minute research pitch. By calling this a "pitch," I mean it to be less daunting than a formal research proposal, but the pitch should succinctly cover these 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?", "what would change if we had the knowledge you propose to create?", "why are you a good person to study it?", "what skills might you build as you study it?", "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 – giving an outline of how you *would* start to map literature if you were to do the project);
4. A bounded object of study (what you're going to research and how you distinguish it from something else) and a methodological approach (with acknowledgement of this approach's advantages and limitations);
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

Depending on the course pace and student interests, we might change some of the below.

Monday, May 23

:: INTRODUCTION: USING INFRASTRUCTURE STUDIES TO INVESTIGATE BIG DATA & ALGORITHMS ::

Sismondo, S. (2009). The prehistory of science and technology studies *An introduction to science and technology studies* (2nd ed., pp. 1-11). London, UK: Wiley-Blackwell.

Sismondo, S. (2009). Actor-network theory *An introduction to science and technology studies* (2nd ed., pp. 81-92). London, UK: Wiley-Blackwell.

Star, S. L. (1999). Ethnography of infrastructure. *American Behavioral Scientist*, 43(3), 377-391.

Recommended:

Gillespie, T., Boczkowski, P. J., & Foot, K. A. (Eds.). (2014). *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press.

Latour, B. (2004). On using ANT for studying information systems – a somewhat Socratic dialogue. In C. Avgerou, C. Ciborra & F. Land (Eds.), *The social study of information and communication technology: Innovation, actors, and contexts* (pp. 62-76). Oxford, UK: Oxford University Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford, UK: Oxford University Press.

Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology*, 24, 95-117.

Merton, R. K. (1968). On sociological theories of the middle range *Social theory and social structure* (pp. 39-72). Boston, MA: Free Press.

Monteiro, E. (2004). Actor network theory and cultural aspects of interpretative studies. In C. Avgerou, C. Ciborra & F. Land (Eds.), *The social study of information and communication technology: Innovation, actors, and contexts* (pp. 129-139). Oxford, UK: Oxford University Press.

Star, S. L., & Bowker, G. C. (2006). How to infrastructure. In L. A. Lievrouw & S. M. Livingstone (Eds.), *Handbook of new media: social shaping and social consequences of ICTs* (pp. 151-162). London, UK: Sage Publications.

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.

Come to class with TWO example objects of study in mind that you think somehow relate to big data and algorithms as you currently understand them. This can be a system, an artifact, an organization, a community, or any other “thing” you’re interested in studying.

Tuesday, May 24

:: WHAT IS "BIG DATA" & WHY DOES IT MATTER? ::

Foundations (read all)

Bowker, G. C., & Star, S. L. (1999). Some tricks of the trade in analyzing classification *Sorting things out: Classification and its consequences* (pp. 33-50). Cambridge, MA: The MIT Press.

Jurgenson, N. (2014, October 9, 2014). View from nowhere. *The New Inquiry*. Retrieved May 9, 2016, from <http://thenewinquiry.com/essays/view-from-nowhere/>

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big Data, Digital Media, and Computational Social Science: Possibilities and Perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 6-13. doi: 10.1177/0002716215572084

Recommended, not required:

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., . . . Van Alstyne, M. (2009). Computational Social Science. *Science*, 323(5915), 721-723. doi: 10.1126/science.1167742

Levy, K. (2013). Relational big data. *Stanford Law Review*, 66, 73-79.

Browse *Big Data & Society* (<http://bds.sagepub.com>) journal and a recent *International Journal of Communication* special issue (<http://ijoc.org/index.php/ijoc/article/view/2167/1164>) to see how the topic is framed:

Applications (pick one and be ready to discuss)

Anderson, C. (2008, June 23, 2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*. Retrieved March 4, 2010, from http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Bell, G. (2012, April 3, 2012). Anthropologist for Intel describes big data as a person *Tech Web TV*. Retrieved December 3, 2014, from https://www.youtube.com/watch?v=WVB6_QP_2s0

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 <http://ethnographymatters.net/blog/2012/06/11/the-ethnographers-complete-guide-to-big-data-part-ii-answers/>

Butler, D. (2013, February 13, 2013). When Google got flu wrong. *Nature*. Retrieved May 2, 2014, 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 data analysis. *Science*, 343, 1203-1205.

Miller, G. (2014, December 8, 2014). The huge, unseen operation behind the accuracy of Google Maps. *Wired*. Retrieved December 18, 2014, from <http://www.wired.com/2014/12/google-maps-ground-truth/>

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>

McDonald, S. M. (2016). *Ebola: A Big Data Disaster: Privacy, Property, and the Law of Disaster* Experimentation Centre for Internet and Society Series. Delhi, India: The Centre for Internet and Society.

Possible In-Class Activities:

1. Look at Google Flu Trends (<https://www.google.org/flutrends/us/#US>) in both graph and Google Earth form; reflect upon assumptions, dynamics, and claim-making supported by the tools.
2. Use Overview project (<https://www.overviewproject.org>) to create / analyze a document data set (register for free account).
3. Analyze Google NGram data sets: what are these data, where do they come from, how would you explain them to different audiences?
4. Use, NPR API Query Generator (<http://www.npr.org/api/queryGenerator.php>), PageOneX (<http://pageonex.com/>)

Wednesday, May 25

:: WHAT ARE ALGORITHMS AND WHY DO THEY MATTER? ::

Foundations (read all)

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-194). Cambridge, MA: MIT Press.

Recommended, not required:

Just, N., & Latzer, M. (2016). Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society*. doi: 10.1177/0163443716643157

Crawford, K. (2016). Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics. *Science, Technology & Human Values*, 41(1), 77-92. doi: 10.1177/0162243915589635

Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, 24(3). doi: 10.1111/comt.12039

Neyland, D., & Möllers, N. (2016). Algorithmic IF ... THEN rules and the conditions and consequences of power. *Information, Communication & Society*, 1-18. doi: 10.1080/1369118X.2016.1156141

Gillespie, T. (2014b, June 25). Algorithm [draft] [#digitalkeyword]. Culture Digitally. Retrieved October 16, 2014, from <http://culturedigitally.org/2014/06/algorithm-draft-digitalkeyword/>

Applications (pick one and be ready to discuss)

Ananny, M. (2011, April 14, 2011). The curious connection between apps for gay men and sex offenders. *The Atlantic*. Retrieved January 8, 2014, from <http://www.theatlantic.com/technology/archive/2011/04/the-curious-connection-between-apps-for-gay-men-and-sex-offenders/237340/>

Berg, N. (2014, June 25, 2014). Predicting crime, LAPD-style. *The Guardian*. Retrieved August 20, 2014, from <http://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report>

Gillespie, T. (2012, July 31, 2012). Is Twitter us or them? #twitterfail and living somewhere between public commitment and private investment. *Culture Digitally*. Retrieved January 6, 2014, from http://culturedigitally.org/2012/07/is_twitter_us_or_them/

Lohr, S. (2013, March 10, 2013). Algorithms get a human hand in steering web. *New York Times*. Retrieved April 2, 2014, from <http://www.nytimes.com/2013/03/11/technology/computer-algorithms-rely-increasingly-on-human-helpers.html>

Manonich, L. (2013, December 16, 2013). The algorithms of our lives. *The Chronicle of Higher Education*. Retrieved August 20, 2014, from <http://chronicle.com/article/The-Algorithms-of-Our-Lives-/143557/>

Priluck, J. (2015, April 25, 2015). When bots collude. *The New Yorker*. Retrieved April 28, 2015, from <http://www.newyorker.com/business/currency/when-bots-collude>

Owen, T. (2015, May 25, 2015). The violence of algorithms: Why big data is only as smart as those who generate it. *Foreign Affairs*. Retrieved May 25, 2015, from <https://www.foreignaffairs.com/articles/2015-05-25/violence-algorithms>

Stoller, M. (2014, April 9, 2014). Uber's algorithmic monopoly. *Observations on credit and surveillance*. Retrieved July 4, 2014, from <http://mattstoller.tumblr.com/post/82233202309/ubers-algorithmic-monopoly-we-are-not-setting-the>

Possible In-Class Activities:

1. Play "Silent Game" developed at MIT School of Architecture and then take an "algorithmic walk" (designed by Malte Ziewitz)
2. Re Bowker & Star's study of classification, redesign the "Critical Algorithms Studies" list categories: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>

Thursday, May 26

:: EMERGING TECHNIQUES FOR STUDYING & USING ALGORITHMS ::

Foundations (read all)

Kitchin, R. (2016). Thinking critically about and researching algorithms. *Information, Communication & Society*, 1-16. doi: 10.1080/1369118X.2016.1154087

Seaver, N. (2014, nd). On reverse engineering: Looking for the cultural work of engineers. *Medium*. Retrieved April 3, 2014, from <https://medium.com/anthropology-and-algorithms/d9f5bae87812>

Recommended, not required:

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., . . . Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in the news feed. Paper presented at the CHI 2015, Seoul, Republic of Korea.

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63-76. doi: 10.1177/0002716215570866

Lessig, L. (2009, October 9, 2009). Against transparency: The perils of open government. *The New Republic*. Retrieved August 17, 2010, from <http://www.tnr.com/article/books-and-arts/against-transparency>

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing algorithms: Research methods for detecting discrimination on internet platforms*. Paper presented at the Data and Discrimination: Converting Critical Concerns into Productive: A preconference at the 64th Annual Meeting of the International Communication Association, Seattle, WA.

Applications (pick one and be ready to discuss)

Vanderbilt, T. (2013, August 7, 2013). The science behind the Netflix algorithms that decide what you'll watch next. *Wired*. Retrieved August 20, 2014, from http://www.wired.com/2013/08/qq_netflix-algorithm/

Felten, E. (2012, September 12, 2012). Accountable algorithms. *Freedom to tinker*. Retrieved May 2, 2013, from <https://freedom-to-tinker.com/blog/felten/accountable-algorithms/>

Danaher, J. (2015, July 26, 2015). How to study algorithms: Challenges and methods. *Philosophical Disquisitions*. Retrieved May 3, 2016, from <http://philosophicaldisquisitions.blogspot.ca/2015/07/how-to-study-algorithms-challenges-and.html>

Diakopoulos, N. (2015, April 27, 2015). Towards a standard for algorithmic transparency in the media. *Tow Center for Digital Journalism*. Retrieved April 29, 2015, from <http://towcenter.org/towards-a-standard-for-algorithmic-transparency-in-the-media/>

Schudson, M., & Sonnevend, J. (2010, February, 2010). Beyond transparency: Is more information always a good thing? *Columbia Journalism Review*. Retrieved March 30, 2015, from http://www.cjr.org/the_research_report/beyond_transparency.php

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 <http://www.ibtimes.com/ftc-chief-technologist-ashkan-soltani-algorithmic-transparency-fight-against-biased-1876177>

Madrigal, A. C. (2015, March 27, 2015). Many, many Facebook users still don't know that their news feeds are filtered by an algorithm. *Fusion*. Retrieved March 29, 2015, from <http://fusion.net/story/110543/most-facebook-users-still-dont-know-that-their-news-feeds-are-filtered-by-an-algorithm/>

Possible In-Class Activities:

1. Build a Twitter bot using "Cheap Bots Done Quick" (<http://cheapbotsdonequick.com/>), connect it to "If This Then That" recipes (<https://ifttt.com/>) and engage in "reflective action" about process/assumptions of creating rule structures to traverse and predict data spaces and platforms.
2. Apply "Wordsmith" (<https://wordsmith.automatedinsights.com/>) to a "Data is Plural" data set (<http://tinyletter.com/data-is-plural/archive>)

Friday, May 27

:: HISTORICAL CONTEXTS OF BIG DATA & ALGORITHMS ::

[Visitor: Prof Megan Finn, University of Washington]

Foundations (read all)

Ankerson, M. S. (2015). Read/write the digital archive: Strategies for historical web research. In E. Hargittai & C. Sandvig (Eds.), *Digital research confidential* (pp. 29-54). Cambridge, MA: MIT Press.

Igo, S. (2007). Introduction: America in aggregate *The averaged American: Surveys, citizens, and the making of a mass public* (pp. 1-22). Cambridge, MA: Harvard University Press.

Recommended, not required:

Browse the history of data journalism in the archive of Scott Klein's mailing list *Above Chart*:
<http://tinyletter.com/abovechart/archive>

Bouk, D. (2015). *How our days became numbered: Risk and the rise of the statistical individual*. Chicago, IL: University of Chicago Press.

Garvey, E. G. (2013). "facts and FACTS": Abolitionists' database innovations. In L. Gitleman (Ed.), *"Raw data" is an oxymoron* (pp. 89-102). Cambridge, MA: MIT Press.

Applications (pick one and be ready to discuss)

Andersen, M. (2016, May 16, 2016). A guy just transcribed 30 years of for-rent ads. Here's what it taught us about housing prices. *Medium*. Retrieved May 19, 2016, from <https://medium.com/@andersem/a-guy-just-transcribed-30-years-of-for-rent-ads-heres-what-it-taught-us-about-sf-housing-prices-bd61fd0e4ef9#.jfd2fcvil>

Bedoya, A. M. (2014, November 7, 2014). Big data and the underground railroad. *Slate*. Retrieved November 7, 2014, from http://www.slate.com/articles/technology/future_tense/2014/11/big_data_underground_railroad_history_says_unfettered_collection_of_data.single.html

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/>

Houston, B. (2015, November 12, 2015). Fifty years of journalism and data: A brief history. *Global Investigative Journalism Network*. Retrieved November 23, 2015, from <http://gijn.org/2015/11/12/fiftyyears-of-journalism-and-data-a-brief-history/>

Klein, S. (2015, March 17, 2015). Antebellum data journalism: Or, how big data busted Abe Lincoln. *ProPublica*. Retrieved November 20, 2015, from <https://www.propublica.org/nerds/item/antebellum-data-journalism-busted-abe-lincoln>

Robertson, H., & Travaglia, J. (2015, October 13, 2015). Big data problems we face today can be traced to the social ordering practices of the 19th century. *LSE Impact Blog*. Retrieved January 4, 2016, from <http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/13/ideological-inheritances-in-the-data-revolution/>

Maddock, J., Starbird, K., & Mason, R. (2015). Using historical Twitter data for research: Ethical challenges of tweet deletions. *Paper presented at the Computer Supported Cooperative Work workshop: Ethics for studying sociotechnical systems in a big data world*, Vancouver, BC. <https://cscwethics2015.files.wordpress.com/2015/02/maddock.pdf>

Possible In-Class Activities:

1. Examine how scope and limitations of web archives afford/constrain research designs – e.g., Wayback Machine (<https://archive.org/web>) & Newsdiff (<http://newsdiffs.org/>)
2. Activity with Prof Finn.

Tuesday, May 31

:: EMERGING TECHNIQUES FOR STUDYING & USING BIG(ISH) DATA ::

Foundations (read all)

Dubois, E., & Ford, H. (2015). Trace interviews: An actor-centered approach. *International Journal of Communication*, 9, 2067–2091.

Geiger, R. S., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. *Paper presented at the 44th Annual Hawaii International Conference on Systems Sciences, Hawaii*.

Vertesi, J., Kaye, J., Jarosewski, S. N., Khovanskaya, V. D., & Song, J. (2016). Data Narratives: uncovering tensions in personal data management. *Paper presented at the Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, San Francisco, California, USA.

Recommended, not required:

Gangadharan, S. P. (Ed.). (2014). *Data and discrimination: Collected essays*: Open Technology Institute & New America.

Marres, N., & Weltevrede, E. (2013). Scraping the social? Issues in live social research. *Journal of Cultural Economy*, 6(3), 313-335. doi: 10.1080/17530350.2013.772070

Lerman, J. (2013). Big data and its exclusions. *Stanford Law Review*, 66, 55-63.

Papacharissi, Z., & Oliveira, M. (2012). Affective news and networked publics: The rhythms of news storytelling on #Egypt. *Journal of Communication*, 62(2), 266-282.

Shaw, A. (2015). Hired hands and dubious guesses: Adventures in crowdsourced data collection. In E. Hargittai & C. Sandvig (Eds.), *Digital research confidential* (pp. 155-172). Cambridge, MA: MIT Press.

Applications (pick one and be ready to discuss)

Brogan, J. (2015, May 13, 2015). The case of the ornamental anthropologist. *Slate*. Retrieved May 17, 2015, from http://www.slate.com/articles/technology/future_tense/2015/05/netflix_tries_to_put_a_human_face_on_big_data_with_its_own_anthropologist.single.html

Brunton, F., & Nissenbaum, H. (2011). Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday*, 16(5).

Chammah, M. (2016, February 3, 2016). Policing the future. *The Marshall Project*. Retrieved May 3, 2016, from <https://www.themarshallproject.org/2016/02/03/policing-the-future>

Ford, H. (2014). Big Data and small: Collaborations between ethnographers and data scientists. *Big Data & Society*, 1(2), 1-3. doi: 10.1177/205395171454433

Madrigal, A. (2014, April 14, 2014). Behind the machine's back: How social media users avoid getting turned into big data. *The Atlantic*. Retrieved June 3, 2014, from <http://www.theatlantic.com/technology/archive/2014/04/behind-the-machines-back-how-social-media-users-avoid-getting-turned-into-big-data/360416/>

Petre, C. (2015, May 7, 2015). The traffic factories: Metrics at Chartbeat, Gawker Media, and The New York Times. *Tow Center for Digital Journalism*. Retrieved May 10, 2015, from <http://towcenter.org/research/traffic-factories/>

Wang, T. (2016, January 20, 2016). Why big data needs thick data. *Medium*. Retrieved May 6, 2016, from <https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7#.pldzp16bc>

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/>

Possible In-Class Activities:

1. Design an interview framework for soliciting “data narratives” of a social media platform of your choice.
2. Design a “trace ethnography” of a “Data is Plural” data set (<http://tinyletter.com/data-is-plural/archive>)
3. Design a research approach for tracing big data resistance as infrastructural work.

Wednesday, June 1
:: WORKING WITH BIG(ISH) DATA – SAMPLING & CONNECTING CONTEXTS ::
[Visitor: Prof Kjerstin Thorson, USC]

Foundations (read all)

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>

Driscoll, K., & Walker, S. (2014). Working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication, 8*, 1745–1764.

Thorson, K., & Wells, C. (2015). Curated Flows: A Framework for Mapping Media Exposure in the Digital Age. *Communication Theory*, doi: 10.1111/comt.12087

Recommended, not required:

Driscoll, K., & Thorson, K. (2015). Searching and Clustering Methodologies: Connecting Political Communication Content across Platforms. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 134-148. doi: 10.1177/0002716215570570

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & boyd, d. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication, 5*, 1375–1405.

Zimmer, M., & Proferes, N. J. (2014). A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management, 66*(3), 250-261. doi: doi:10.1108/AJIM-09-2013-0083

Applications (pick one and be ready to discuss)

Tell us about one of the chapters of the “Responsible Data Handbook”:
<https://responsibledata.io/resources/handbook/>

Experiment with one of the tools described here: <https://www.databasic.io/en/>

Clifton, B., Lotan, G., & Pierson, E. (2015, October 9, 2015). How to tell whether a Twitter user is pro-choice or pro-life without reading any of their tweets. *Quartz*. Retrieved May 6, 2016, from <http://qz.com/520309/how-to-tell-whether-a-twitter-user-is-pro-choice-or-pro-life-without-reading-any-of-their-tweets/>

Crawford, K. (2013, April 1, 2013). The hidden biases in big data. *Harvard Business Review*. Retrieved August 27, 2013, from http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html

Mier, B. (2014, October 13, 2014). The mailman mapping Brazil's largest favela by hand. *Vice*. Retrieved October 24, 2014, from <http://motherboard.vice.com/read/the-mailman-mapping-brazils-largest-favela-by-hand>

Onuoha, M. (2016, February 10, 2016). The point of collection. *Data & Society, Points*. Retrieved May 4, 2016, from <https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa#.erdhuq20y>

Thorson, K., Driscoll, K., Ekdale, B., Edgerly, S., Thompson, L. G., Schrock, A., . . . Wells, C. (2013). YouTube, Twitter, and the Occupy movement. *Information, Communication & Society, 16*(3), 421-451. doi: 10.1080/1369118X.2012.756051

In-Class Activity:

Work with Newswhip data dashboard to frame research questions around multi-platform news propagation.

Thursday, June 2

:: ETHICS, VALUES, & CONTROVERSIES IN (STUDYING) BIG DATA & ALGORITHMS ::

Foundations (read all)

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society*. doi: 10.1177/2053951716650211

Sismondo, S. (2009). *Controversies An introduction to science and technology studies* (2nd ed., pp. 120-135). London, UK: Wiley-Blackwell.

Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313-325. doi: 10.1007/s10676-010-9227-5

Recommended, not required:

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology & Human Values*, 41(1), 93-117. doi: 10.1177/0162243915606523

Crawford, K., & Schultz, J. (2013). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, 55(1).

Dwork, C. (2011). A firm foundation for private data analysis. *Commun. ACM*, 54(1), 86-95. doi: 10.1145/1866739.1866758

Markham, Annette, & Buchanan, Elizabeth. (2012). Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee. *Association of Internet Researchers*. 2nd revision. Retrieved October 19, 2014, from <http://aoir.org/reports/ethics2.pdf>

Applications (pick one and be ready to discuss)

Auerbach, D. (2015, May 18, 2015). The silicon tower. *Salon*. Retrieved May 20, 2015, from http://www.slate.com/articles/technology/bitwise/2015/05/facebook_study_why_silicon_valley_s_incursion_into_academic_research_is.single.html

Grimmelmann, J. (2015, May 27, 2015). Do you consent? *Salon*. Retrieved May 28, 2015, from http://www.slate.com/articles/technology/future_tense/2015/05/facebook_emotion_contagion_study_tech_companies_need_irb_review.html

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 <http://culturedigitally.org/2014/07/facebooks-algorithm-why-our-assumptions-are-wrong-and-our-concerns-are-right/>

Gillespie, T. (2016, May 18, 2016). Algorithms, clickworkers, and the befuddled fury around Facebook Trends. *Social Media Collective*. Retrieved May 18, 2016, from <https://socialmediacollective.org/2016/05/18/facebook-trends/>

Gray, M. L. (2014, August 19, 2014). Microsoft Research faculty summit 2014 ethics panel recap. *Social Media Collective*. Retrieved October 2, 2014, from <http://socialmediacollective.org/2014/08/19/msr-faculty-summit-2014-ethics-panel-recap/>

Markham, A. (2016, May 18, 2016). The OKCupid data release fiasco: It's time to rethink ethics education. *Social Media Collective*. Retrieved May 19, 2016, from <https://socialmediacollective.org/2016/05/18/the-okcupid-data-release-fiasco-its-time-to-rethink-ethics-education/>

Meyer, R. (2014, September 8, 2014). Everything we know about Facebook's secret mood manipulation experiment. *The Atlantic*. Retrieved December 3, 2014, from <http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/>

O'Donovan, C. (2014, July 8, 2014). Q&A: Tarleton Gillespie says algorithms may be new, but editorial calculations aren't. *Nieman Lab*. Retrieved July 9, 2014, from <http://www.niemanlab.org/2014/07/qa-tarleton-gillespie-says-algorithms-may-be-new-but-editorial-calculations-arent/>

Shilton, K., Butler, B., Goggins, S., & Winter, S. (2015). Research ethics for open online community data: A case study of human subjects research online. *Paper presented at the Computer Supported Cooperative Work workshop: Ethics for studying sociotechnical systems in a big data world*, Vancouver, BC.

<https://cscwethics2015.files.wordpress.com/2015/02/shilton.pdf>

Tufekci, Z. (2016, May 19, 2016). The real bias built in at Facebook. *New York Times*.

Retrieved May 19, 2016, from <http://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html>

Possible In-Class Activities:

1. I lead discussion of Facebook Contagion case study.
2. Students lead discussion of Facebook Trends bias controversy and/or OK Cupid data release case
3. Sketch interdisciplinary code of ethics for: studying big data and algorithms; building “data factories”; and creating a “data science” profession.

Friday, June 3

:: WRAP-UP, REVISIT ANYTHING WE RAN OF TIME WITH, STUDENT PITCH PRESENTATIONS ::