Media Programming in an Era of Big Data

Timothy Havens¹

University of Iowa

timothy-havens [AT] uiowa.edu

Abstract:

While scholars are producing a growing body of work that speculates on the theoretical significance of “algorithmic culture,” where the role of human agents in recommending and sharing cultural objects are replaced by machines, very little of this work examines the algorithms themselves and how the different kinds of knowledge they produce affect existing cultural practices, including production practices. In this essay, I use the original Netflix series *House of Cards* to think through some of the ways that Big Data is influencing the television industry and some of the questions that it raises for both the study and operation of the media industries. Ultimately, I argue that media industry scholarship, which is rooted in methods of observation, interview, and ethnography, could add a great deal to ongoing scholarly discussions of algorithmic culture.

Keywords: Digitization, Management, Production, Television, Technology

Whether we call it “Big Data,” “algorithmic culture,” or “the addressable audience,” the integration of digital recording, distribution, and data analysis technologies with the legacy media industries has generated a good deal of hype, concern, and speculation. To take a recent, prominent example: proponents of the idea that Netflix’s algorithm leads to better understanding and targeting of viewers’ tastes celebrated how effectively the company filtered its glut of subscriber data to create and distribute its series *House of Cards* (2013-). Detractors, for their part, worried that knowing every detail about how tens of millions of subscribers watch television would make creative workers beholden less to their own creative visions and more to the demands of the algorithm.

In this essay, I use the *House of Cards* example—the most advanced application to date of Big Data to programming decisions—to think through some of the ways that Big Data is influencing the television industry and some of the questions that it raises for both the study and operation of the media industries. In particular, I want to focus on the Netflix algorithm itself, the forms of knowledge it produces, and how this knowledge has and has not altered how programming decisions have historically been made.

The audience remains the central obsession of the commercial media industries. Because of the high and increasing costs of creating and producing popular television and film—especially the

high-ticket “quality” television that subscription-paying customers expect—the need to know and anticipate audience tastes is crucial for predicting successful returns on investment. Supporters and detractors alike believe that the massive amount of data on viewership that Netflix collects, including more than fifty data points about the consumption of every film and television show, give the company substantial capacity to predict what viewers want, far more than traditional Nielsen ratings ever gave broadcasters. As Andrew Leonard states, writing for the online magazine *Salon*, “The interesting and potentially troubling question is how a reliance on Big Data might funnel craftsmanship in particular directions. What happens when directors approach the editing room armed with the knowledge that a certain subset of subscribers are opposed to jump cuts or get off on gruesome torture scenes or just want to see blow jobs. Is that all we’ll be offered?”²

There’s good reason to doubt how much Netflix actually knows about viewer preferences for particular types of content, much less specific production or post-production techniques. Perhaps the best piece of evidence is the *House of Cards* series itself. As widely publicized, the decision to produce a remake of the 1990s BBC series starring Kevin Spacey and directed by David Fincher resulted from a proprietary algorithm that analyzed trillions of Netflix data points. To me, however, what was most astonishing about the series was how straight between the “quality TV” goal posts it flew. Yes, the music score was fantastic, the casting and acting impeccable, and the storyline complex and riveting. But, as one of a slew of quality dramas that fit this exact description, the only mildly innovative element in the series was Kevin Spacey’s interruption of the narrative flow to directly address the camera.

How could a television series supposedly conceived in such a radically new manner be so aesthetically uninventive? The answer, I think, lies in the details of the algorithms that Netflix uses and the role that the algorithm plays in creative and corporate decision-making. While scholars are producing a growing body of work that speculates on the theoretical significance of “algorithmic culture,”³ where the role of human agents in recommending and sharing cultural objects are replaced by machines, very little of this work examines the algorithms themselves and how the different kinds of knowledge they produce affect existing cultural practices, including production practices. Media industry scholarship, which is rooted in methods of observation, interview, and ethnography could add a great deal to these ongoing scholarly discussions.

Netflix deploys what is known as a “recommendation algorithm” to suggest content to its subscribers. The company made big news in 2006 when it announced an open competition to try to improve the predictive quality of its recommendation algorithm, promising one million dollars to anyone who achieved a 10 percent improvement over Netflix’s own algorithm, Cinewatch. The winning algorithm, designed by a collaboration known as the BellKor Pragmatic Chaos team, achieved the required 10 percent improvement over Cinewatch by employing parallel/distributed calculations across multiple computers in October 2009.⁴ Less than two years later, Netflix announced its plan to enter into the original programming market, citing its new algorithm as a central component of its marketing strategy.⁵

While Netflix could and did use its recommendation algorithm to try to market *House of Cards* to subscribers, much of the hype surrounding the series suggested that it was the algorithm itself that selected the series, the star, and the director.⁶ By way of inference, audiences were led to believe that the decision to remake *House of Cards* was the result of the company’s new and

improved algorithm. Jonathan Friedland, the communications director for Netflix, claimed with regard to *House of Cards*, “We know what people watch on Netflix and we’re able with a high degree of confidence to understand how big a likely audience is for a given show based on people’s viewing habits.”⁷

The winning BellKor Pragmatic Chaos algorithm did not, however, recommend producing *House of Cards*. The Netflix prize was built around improving what’s known as a “collaborative filtering” algorithm, which uses such things as viewer ratings, history, and behavior to recommend content that *already exists* on the service. The algorithm is “content-neutral,” i.e., not taking into consideration popular content categories—such as genres or stars—in making recommendations, but only what other subscribers with similar profiles have watched and rated positively.

Since *House of Cards* was an unknown property, and since collaborative filtering algorithms are especially poor at dealing with unknowns,⁸ Netflix had to use what’s known as a “content-based” algorithm to suggest the series. However, content-based algorithms have serious problems of their own, not least that “only a very shallow analysis of certain kinds of content can be supplied. In some domains the items are not amenable to any useful feature extraction methods with current technology (such as movies, music, restaurants).”⁹ While the capacity of content algorithms to extract data from films and television shows has most certainly improved in the fifteen years since this statement was published, the fact that Netflix extracted only three content variables from its terabytes of user data—director, star, and genre—to make its decision suggests that the depth of use of content-based algorithms has not increased substantially with regard to television and film in recent years.

It’s also important to keep in mind that *House of Cards* was acquired by Netflix from Media Rights Capital, not self-produced or commissioned. Hence, the series was not, in fact, a product of the algorithm, but rather the closest approximation to what the algorithm predicted, *based upon the interpretations of Netflix programmers*. In other words, much in the way programmers of commercial media have done for decades, Netflix executives took the data available to them and made decisions about what television content to acquire based upon their “gut” feelings.¹⁰ Examined in this light, it would seem that television programming decisions in the digital age might not be all that different than those in the analog age.

Programmers themselves insist that they are still relevant as interpreters of cultural tastes and trends. “Data is not the whole story,” according to Linda Ong, president and brand strategist for the online channel Truth TV. “Data tells you what; it doesn’t tell you why.”¹¹ And while we should take such comments with a grain of salt, as programmers strive to remain relevant in an age of Big Data, they remain on the job, even at Netflix. Reports of their death (or irrelevance) may have been greatly exaggerated.

At the same time, the programmer’s job, as well as the industries that programmers work within, are being changed by new forms of data gathering and analysis. Laura Staro, senior vice president of research and strategy for channels TLC and Discovery Fit & Health, tells *Daily Variety* that, “There’s more of a trend for people to embrace research and information.”¹² Just how programmers’ jobs have changed and how those changes influence the autonomy of creative workers in the media industries requires grounded investigation. Do programmers still rely on their gut instincts as much as they used to? How often, and in what contexts, do data gleaned through recommendation algorithms affect the autonomy of television writers,

producers, actors, set designers, and so forth? How exactly do the algorithms operate, how effectively, and what is their status as predictors of audience behavior in various industry settings? These are crucial questions to tackle before we can begin to ask large theoretical questions about the state of media industries in an algorithmic culture.

At minimum, one major change that has taken place for media industry workers at all levels is a shift from an era of scarcity of audience data to an era of overabundance. Because of earlier research in the media industries, we have some information regarding how executives and creators managed a *paucity* of data, including a reliance on gut instincts, industry lore, and complicated power plays among creators and gatekeepers that often deployed different conceptualizations of the audience. In an age of data overabundance, however, different individuals and organizations—different “power roles,” to use Joseph Turow’s generative concept—use data to try and gain acquiescence and advantage over other players.¹³ At the same time, an increasing amount of information is private and restricted. While Nielsen data was always private, almost everyone in the television industry had access to the same information. What is the status of privatized information today and how is it distributed among different power roles? We know, for instance, that one of the revolutionary aspects of the Netflix algorithm prize was the company’s willing release of a large amount of private data. Naturally, this action delighted scholars working on recommendation algorithms because they are so rarely allowed access to such information. Ultimately, this question of privatization and access to Big Data may prove to be a more important social question than the predictive power of the algorithm itself.

¹ Timothy Havens is Associate Professor of Communication Studies, African American Studies, and International Studies at the University of Iowa. He is the author of *Black Television Travels: African American Media Around the Globe* (NYU Press, 2013) and *Global Television Marketplace* (BFI Publishing, 2006); the co-author with Amanda D. Lotz of *Understanding Media Industries* (Oxford University Press, 2011); the co-editor with Aniko Imre and Katalin Lustyik of *Popular Television in Eastern Europe During and Since Socialism* (Routledge, 2012); and a former Senior Fulbright Scholar to Hungary. His research has also appeared in numerous scholarly journals in media and cultural studies and anthologies on media globalization and television studies.

² Andrew Leonard, [“How Netflix is Turning Viewers into Puppets,”](#) *Salon.com*, February 1, 2013.

³ Alexander R. Galloway, *Gaming: Essays on Algorithmic Culture* (Minneapolis: University of Minnesota Press, 2007); Ted Striphas, [“Algorithms are Decision-Making Machines,”](#) *The Late Age of Print* (blog), November 18, 2012.

⁴ Yehuda Koren, [“The BellKor Solution to the Netflix Grand Prize,”](#) *Netflix Prize Documentation*, August 2009,

⁵ Andrew Wallenstein, “Netflix Shows Hand with ‘Cards’ Deal,” *Daily Variety* 310, no. 55 (March 21, 2011): 4.

⁶ Leonard, “How Netflix is Turning Viewers into Puppets.”

⁷ *Ibid.*

⁸ Marco Balbanovic and Yoav Shoham, “Fab: Content-Based, Collaborative Recommendation,” *Communications of the ACM* 40, no. 3 (1997): 66-72.

⁹ *Ibid.*, 67.

-
- ¹⁰ Todd Gitlin, *Inside Prime Time* (London: Routledge, 1983).
- ¹¹ Kate Kaye, [“Welcome to the Era of the Data-Driven Programmer,”](#) *Advertising Age*, April 8, 2013.
- ¹² Kaye, “Welcome to the Era of the Data-Driven Programmer.”
- ¹³ Joseph Turow, *Media Systems in Society: Understanding Industries, Strategies, and Power*, 2nd ed. (White Plains: Longman 1997).

Bibliography

- Balbanovic, Marco and Yoav Shoham. “Fab: Content-Based, Collaborative Recommendation.” *Communications of the ACM* 40, 3 (1997): 66-72.
- Galloway, Alexander R. *Gaming: Essays on Algorithmic Culture*. Minneapolis: University of Minnesota Press, 2007.
- Gitlin, Todd. *Inside Prime Time*. London: Routledge, 1983.
- Kaye, Kate. [“Welcome to the Era of the Data-Driven Programmer.”](#) *Advertising Age*, April 8, 2013.
- Koren, Yehuda. [“The BellKor Solution to the Netflix Grand Prize.”](#) *Netflix Prize Documentation*. August 2009.
- Leonard, Andrew. [“How Netflix is Turning Viewers into Puppets.”](#) *Salon*, February 1, 2013.
- Striphas, Ted. [“Algorithms are Decision-Making Machines.”](#) *The Late Age of Print* (blog). November 18, 2012.
- Turow, Joseph. *Media Systems in Society: Understanding Industries, Strategies, and Power*. 2nd ed. White Plains: Longman 1997.
- Wallenstein, Andrew. “Netflix Shows Hand with ‘Cards’ Deal.” *Daily Variety*, March 21, 2011.

Copyright © 2014 (Timothy Havens). *Media Industries* is an open-access, peer-reviewed, online academic journal. As such, we aim to participate in the open exchange of information. This work is licensed under a [Creative Commons Attribution Noncommercial No Derivatives \(by-nc-nd\) License](#). Under this license, this work is available for sharing and noncommercial distribution provided the appropriate attribution is given.