

The Algorithmic Self: Layered Accounts of Life and Identity in the 21st Century

Annette N. Markham

Department of Aesthetics & Communication
Aarhus University
Denmark
amarkham@gmail.com

Abstract

This paper takes an actor network theory approach to explore some of the ways that algorithms co-construct identity and relational meaning in contemporary use of social media. Based on intensive interviews with participants as well as activity logging and data tracking, the author presents a richly layered set of accounts to help build our understanding of how individuals relate to their devices, search systems, and social network sites. This work extends critical analyses of the power of algorithms in implicating the social self by offering narrative accounts from multiple perspectives. It also contributes an innovative method for blending actor network theory with symbolic interaction to grapple with the complexity of everyday sensemaking practices within networked global information flows.

Keywords

Identity; actor network theory; algorithms; agency; critical; remix

Introduction

I know I shouldn't care that Facebook's timeline forces my personal history into a chronology of the jobs I've held. I know that in the scale of important things to worry about in the world, I shouldn't focus on the fact that my Facebook timeline encourages me to add materials to my timeline, until I've reached the category "born."

As if my history started then.

(research participant notes, 22 August 2012)

Latour (2012) suggests that we are not just part of networks, but wholly defined by what can be found in databases. Senft (2012) refers to this as 'the grab,' an action of taking something we want, but also as a temporary fixing, as we might accomplish through a screengrab, or by making sense of something or someone very rapidly and then moving on. We might do this deliberately and consciously, or it might be a decision made through a complex and largely invisible interaction with the algorithms that mediate the interface.

A couple of years ago, I read an article that said we are being trapped by the technologies we're using now, and that while most of us believe we still have freedom of choice, this just isn't the case. He argued that our reliance on particular interfaces grows more powerful every time we upload another photo or document moments of our lives, because we're less and less likely to want to start all over again.

So Facebook is based more on inertia. Even if we want to leave, we just don't.

(research participant notes, 22 August 2012)

A growing body of scholarship contributes to how we conceptualize our relationships with our platforms, devices, and technologies at the conceptual level. Deuze, Banks, & Spears (2012), for example, talk about the ways in which we're more than just surrounded by media, but experiencing life *in* media. This aligns with and in many ways pushes past previous scholars such as Gergen (1991) or Turkle (1995, 2011), who have long contended that the individual (and our understanding of

identity) is impacted by ever-increasing connection with computers, technologies, and information. As Bolter (2012) notes, we interact not only with others but also “with the algorithm, the code that lies beneath the surface of the application” (p. 39). He continues, “Good digital design today encourages its users to proceduralize their behavior in order to enter into the interactions, and a large portion of those in developed countries have accepted this as the path to participation in digital media culture” (p. 45). These ‘event loops’ are designed to appear seamless, as an essential part of everyday sociality. As Gillespie (2012) notes, these algorithms provide a particular sort of “knowledge logic” that function powerfully and invisibly. Hayles contends that we need to more directly address autonomous algorithms as agents, citing her recent study of the way rapid and automated market trading algorithms can cause mini-crashes in the stock market before a human literally has time to even notice anything has happened (2012).

This paper focuses on these issues at the level of everyday sensemaking. In what ways do we feel ‘grabbed?’ How do people talk about the meaning of their identity as bound up with everyday digital platforms for interaction? By exploring the way people move through these information spaces and flows, and talking with them about their everyday activities, we can add a rich voice to our conversations about meaning, identity, and algorithms.

I resent it because I feel trapped. Bad enough that we’ve become so accustomed to using facebook for personal communication among our networks of friends that we can’t just turn it off. Now, I have to adjust to the parameters yet again, forcing my expression of identity into a still narrower box of options. This time, my timeline. My history.

The sad, pathetic part of this story is that when I saw my profile in a timeline today, tears came, unbidden, rolling down my face. Silly to think I could continue to avoid timeline. Sillier still to realize how much it feels like a small violation of my independence. Ironic, when I know I participate willingly in Facebook and could shut down my account at any moment.

It’s just not that simple.

(research participant notes, 22 August 2012)

Methodology

This paper takes a symbolic interactionist approach to explore how people make sense of their relationships with their own information and the technologies that mediate their everyday activities. To get at this sort of information, I borrow Markham’s (1998) techniques for participant observation and in-depth ethnographic interviewing; Latour’s (2005) actor network theory concepts for following the data wherever it might go and for considering the agency of non-human agents; and Markham’s lens of remix (2013) as a way to grapple with the complexity of data emerging from the interviews and observations.

In addition to intensive engagement with individuals who describe themselves as heavily saturated in social media, this project includes auto-phenomenological or -ethnographic data as a close read of multiple agencies at play in everyday interactions with and in digital media.

Discussion

As if my history could be so encapsulated by time, linear at that.

I feel betrayed by the interface. Betrayed by an interface that appears to give so many choices on the surface, while limiting almost every bit of our creative endeavor to the pre-defined and pre-packaged boxes and categories within which we’re supposed to find a place. It hurts us all, in different, small ways. Sure, I feel fine clicking the ‘female’ category, but I know at least two-dozen friends who wouldn’t be able to choose a box. I’m supposed to declare a ‘hometown’ but I’ve not had one for more than 20 years, so that’s not useful at all. I now have to mark places on a map, or accept the default map

that appeared on my profile just this afternoon. I have to work to figure out how to shut down options. Modify the interface to make sure it's not doing something else invisible to me.

The idea of being locked into a history that Facebook creates? What if I don't want to be defined by time or any other moment that Facebook has determined is "relevant" in my life?

Why am I still crying?

I mean, seriously, why does this affect me so much? Why would I care so much about Facebook?

(research participant notes, 22 August 2012)

We live increasingly public lives. We spend a lot of time managing our identities in ways we never did before the internet made our every move potentially public. We also experience the world as it is mediated by algorithms functioning as "deep structures" (Deetz, 1992; Mumby, 1998). This study focuses on the everyday experience of people who would consider themselves to be a technologically "saturated self," in the sense that Gergen (1991) articulates. This particular paper will present a pastiche of narrative accounts that take a very close look at some of the ways people are making sense of their relationships with their platforms, technologies, and devices.

Truth be told, Google owns me much more than Facebook ever could. Google Plus uploaded all my photos from my computer and phone for sharing before I could figure out how to 'opt out.' At least I expect it from Facebook. But Google? It seemed so benign.

(research participant notes, 24 August 2012)

References

- Bolter, J. D. (2012). Procedure and performance in an era of digital media. In Lind, R. (Ed.). *Producing theory in a digital world: The intersection of audiences and production in contemporary theory* (pp. 33-50). New York: Peter Lang.
- Deetz, S. (1992). *Democracy in an age of corporate colonization: Developments in communication and the politics of everyday life*. New York: SUNY Press.
- Dueze, M., Blank, P., & Speers, L. (2012). A life lived in media. *Digital Humanities Quarterly*, 6(1). Available at: <http://digitalhumanities.org/dhq/vol/6/1/000110/000110.html>
- Gergen, K. (1991). *The saturated self: Dilemmas of identity in contemporary society*. New York: Basic Books.
- Gillespie, T. (2012). The relevance of algorithms. Forthcoming in Gillespie, T., Boczkowski, P., & Foot, K. (Eds.). *Media Technologies*. Cambridge, MA: MIT Press. Pre-publication draft available at: <http://www.tarletongillespie.org/essays/Gillespie%20-%20The%20Relevance%20of%20Algorithms.pdf>
- Hayles, K. (2012). Economic infrastructure and artificial intelligences: The case of automated trading programs. Keynote address at the Media Places: Infrastructure, Space, Media conference. Umea University, December 6, 2012. Abstract available at: <http://mediaplaces2012.humlab.umu.se/program.html>
- Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullie, D. (2012). The whole is always smaller than the sum of its parts: A digital test of Gabriel Tarde's monads. Available at: <http://www.bruno-latour.fr/sites/default/files/123-WHOLE-PART-FINAL.pdf>
- Latour, B. (2005). *Reassembling the social: An introduction to actor network theory*. Oxford, UK: Oxford University Press.

Markham, A. (2013 in press). Remix culture, remix methods: Reframing qualitative inquiry for social media contexts. In Denzin, N., & Giardina, M. (Eds.). *Global Dimensions of Qualitative Inquiry*. Walnut Creek, CA: Left Coast Press.

Markham, A. (1998). *Life online: Researching real experience in virtual space*. Walnut Creek, CA: AltaMira Press.

Mumby, D. (1988). *Communication and power in organizations: Discourse, ideology, and domination*. Norwood, NJ: Axley Press.

Senft, T. (2012). Microcelebrity and the branded self. In Burgess, J. & Bruns, A. (Eds.). *Blackwell Companion to New Media Dynamics* (pp 346-354). Boston, MA: Blackwell.

Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*, New York: Touchstone.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

License

This article is ©2013 Annette N. Markham, and licensed under CC BY-NC-ND.

Identity and Musical Genre in the Age of Algorithms

Holly Kruse

Department of Communications
Rogers State University
United States
Holly.Kruse@gmail.com

Abstract

Traditionally, genre has been an important factor in the construction of music users' and creators' identities. Musical taste and practices, as defined by genre, play an important role in the articulation of identity, especially when subjects position their tastes and practices in opposition to a perceived mainstream. As musical practices move online, including to video-sharing sites like YouTube, algorithms determine music recommendations without little or no consideration of genre. This paper examines the implications of algorithmic suggestions for notions of identity among popular music creators and users.

Keywords

Identity; genre; popular music; algorithms; YouTube.

Identity, genre, and algorithms

In “traditional” popular music culture, genre is central in helping to construct participants' senses of identity. “Genre” refers to a kind or class of things, and in discussions of cultural artifacts, definitions of genres are perhaps most importantly created by the users who identify with and through those genres (Chandler, 2000). In ethnographic research of indie music scenes, genre has proven important in how participants frame their relationships to scenes, and to others. For instance, one musician, in attempting to define the genre in which he works by meaningfully differentiating it from others, states, “I would call it pop music. I would call it wimpy pop music. I would not call it power pop. Power pop has now become a term that means a hard guitar band that uses pop as the format...” (qtd. in Kruse, 2003, p. 116). Indeed, among participants in many music scenes, positioning one's self and one's taste as part of a tradition of expression that values social and cultural practice that are somewhat oppositional to a perceived mainstream is crucial. And one's taste in music (or any cultural product) helps determine how one sees one's self in this relationship. Pierre Bourdieu famously notes that “taste is the basis of all that one has – people and things – and all that one is for others, whereby one classifies oneself and is classified by others” (1984, p. 56).

This terrain of practices, which helps construct identities, is continually shifting, and it leads one to wonder what happens when music cultures move, at least partly, online and are infiltrated by non-human actors: namely, computer code in the form of algorithms?

In looking at relevant notions of identity, one sees that Laclau and Mouffe (2001) argue that identities are not fixed, but rather are articulated within a structure of social relations that causes every social agent to occupy multiple social positions at once, through identifications of race, gender, class, ethnicity, occupation, educational level, tastes, and so on. Further, as Stuart Hall observes, identification does not happen once and for all (1989, p. 73). Identities are produced within an ideological field where signs “can be discursively re-articulated to construct new meanings, connect with different social practices, and position social subjects differently” (Hall, 1988, p. 9). Central to this way of thinking about identity is the idea that a particular practice is articulated within a specific discursive terrain, and it is within this terrain that one constructs identities, including oppositional identities. Indeed, as much as the word 'identification' seems to imply a sense of belonging, perhaps even more it describes a process of differentiation. As Laclau and Mouffe state, “all values are values of opposition and are defined only by their difference” (2001, p. 106). Senses of shared identity are alliances formed out of oppositional stances.

YouTube, and the recommendations that its algorithm generates, is one site where music users may assert oppositional identities. This paper focuses particularly on the ways in which YouTube helps construct identity through its recommendations – recommendations that are often unrelated to the conventional definitions of genre valued by music users and producers – and commenters’ uses of oppositional discourses and invocations of taste and genre in the face of YouTube’s “suggestions” and targeted advertising.

YouTube is obviously not only a music video channel, and its algorithms are used across content type. Founded in 2005, it provides a platform for video-sharing by users, and in that capacity it fills multiple roles: as archive, broadcast medium, and even social network. It doesn’t provide content; users provide content (Burgess and Green, 2009, p. 5). Among its users are commercial clients, with whom YouTube enters into revenue-sharing agreements and for whom its algorithms provide data. These clients include major music industry players, like Vevo. Vevo is a consortium started by Universal Music, Sony Music, and Abu Dhabi Media in late 2009 in an effort to monetize music video content. Although it has its own website, which is among the most visited sites online, Vevo’s most visible presence is on YouTube, where it attracts most of its viewers and where it has accounted for up to 80 percent of the music videos watched (Graham, 2010). With the 2012 break-up of EMI – which licensed its content to Vevo – and EMI’s absorption into Universal Music Group, Warner Brothers is the only other major label player on YouTube, and it is the second most streamed YouTube partner after Vevo (Mlot, 2012). Vevo artists include Justin Bieber, Lady Gaga, Eminem, and Rihanna; Warner Brothers artists include Josh Groban, Muse, and The Black Keys

To look at Vevo in particular, while its name may not be well known among members of the public, its audience is now 25 times as large as MTV’s was during the height of the popularization of televised music videos (Graham, 2010). Vevo clearly brands its videos and is in charge of selling space on its YouTube channel to other advertisers, then sharing that ad revenue with YouTube’s owner, Google (Graham, 2010). Vevo is thus multiply invested in the algorithms that drive viewers to its videos within YouTube. Interestingly, YouTube recently switched from its original recommendation system to the one used by Amazon, which is based on collaborative filtering and suggests items to users based on perceived shared tastes with other users (Linden, 2011). YouTube has also begun creating automated channels created not by users or partners, but by algorithms that track user activity. As described in *New Scientist* (2012), in this process software analyzes how users navigate search results, notes the videos on which users click and how they move from video to video, and gathers information from user comments. The site then recommends channels. For example, “A user who searches for videos about the US Open tennis tournament, then proceeds to watch nothing but Roger Federer clips, for example, might get a recommendation to subscribe to the Roger Federer channel” (Hodson, 2012, p. 25).

In this paper, I examine whether recommendations and advertising on YouTube’s music video partner pages are congruent with traditional notions of popular music genre, or if they in fact help construct a different kind of popular music user (and even creator) identity. Further, I look at the degree to which commenters on these pages appear to acquiesce to or oppose their positioning as potential viewers of videos (or purchasers of music) by algorithmically selected artists. How many users, when confronted with a Vevo advertisement for Justin Timberlake’s new release, are tempted to respond (as one user did) with “Fuck off, Justin Timberfake! :P ”?

References

- Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. Cambridge, MA: Harvard University Press.
- Burgess, J., & Green, J. (2009). *YouTube*. Cambridge: Polity Press.
- Chandler, D. (2000). An introduction to genre theory. Retrieved from <http://www.aber.ac.uk/media/Documents/intgenre/intgenre1.html>
- Graham, J. (2010, Dec. 17). Vevo turned free music videos into revenue. *USA Today*, p. 4B.

Hall, S. (1989). Cultural identity and cinematic representation. *Framework* 36, 68-81.

Hall, S. (1988). *The hard road to renewal: Thatcherism and the crisis of the left*. London: Verso.

Hodson, H. (2012, Dec. 12). Now, watch this. *New Scientist*. 216(2895), 25.

Kruse, H. (2003). *Site and sound: Understanding independent music scenes*. New York: Peter Lang.

Laclau, E., & Mouffe, C. (2001). *Hegemony and socialist strategy: Towards a radical democratic politics*. London: Verso.

Linden, G. (2011, Feb. 11). YouTube uses Amazon's recommendation algorithm.

Retrieved from <http://glinden.blogspot.com/2011/02/youtube-uses-amazons-recommendation.html>

Mlot, S. (2012, June 26). Top 5 YouTube partners nab 1.5 billion monthly streams. *PC Magazine*. Retrieved from <http://www.pcmag.com/article2/0,2817,2406324,00.asp>

License

This article is ©2013 Holly Kruse, and licensed under CC BY-NC-ND.

SNPs to Sheepy: Genetic Genealogy & Place-Identity

Molly Wright Steenson

School of Journalism & Mass Communication
University of Wisconsin-Madison
United States
msteenson@wisc.edu

Abstract

This paper investigates notions of identity and place as a result of genetic genealogy. One's personal relationship with genetic genealogical data, such as from DNA testing services such as 23andMe, offers another angle on digital identity. Using a critical cultural studies lens, the author presents juxtapositions of traditional knowledge logics for explaining one's ancestry with contemporary interactions with genetic data. The results raise important questions, such as: What does it mean when DNA becomes a place, a location of historical significance? How does this contribute to the notion of a meaningful relationship derived through engagement with one's own genetic data? And how does that change notions of personal identity and place-identity?

Keywords

23andme; algorithms; identity; place; genetics

Mouth and place: An introduction of sorts

Perusing my family tree, I came across a tinted photograph from about 1890 of some relatives on my mother's side. Instantly, the girl in the left-hand corner drew my eye, or more specifically, her mouth. It was because it was my mouth. The girl in the picture is my great-grandmother, Ellen Blower. I compared it with the framed picture of my mother at age 14, with family pictures of my aunt and her daughter. From 1890 to today, that mouth repeats in the female descendants of my great-grandmother.

Ellen was born in Sheepy Magna, England. She moved to nearby Hinckley, then emigrated to Moline, Illinois when she was 30. Typing "Sheepy Magna" into Google Earth yields a green quiltwork of fields. A Google image search serves the sign to Sheepy Magna, the church, the main street, a car submerged in a flood, the graveyard of people to whom I am related. When Ellen emigrated to the United States she could be connected to a new network of places, all of which I could navigate from a screen far away in space and time. All of them could be places that contribute to her sense of identity; all of them represented by contemporary data streams that are meaningful to me.

U5a1b1

When I joined the 23andMe genetic testing service and received my results, I learned I belonged to the U5a1b1 haplogroup, which is determined by my mitochondrial DNA (mtDNA). mtDNA passes from mother to daughter to daughter, marked by a single nucleotide polymorphism (SNP) mutation that took place 7000 years ago. Its umbrella haplogroup U5 originated 30–50,000 years ago as humans migrated from the Near East toward Eurasia. While U5 reaches across Africa and as far as Mongolia, its center is northern Europe.

The mutation that created U5a, the branch to which I belong, occurred 20,000 years ago during the ice age, keeping people in southern Europe; 15,000 years ago, they migrated north to Norway and northern Germany, or east toward Turkey and Iran in smaller numbers. The most famous U5a ancestor is "Cheddar Man," a 9,000 year old skeleton discovered in a cave in Cheddar, England, whose haplogroup was determined through mtDNA testing of a tooth, and 1–1.5% of the population of the UK is likely to share the U5a haplogroup (Lyall, 1997). The mutations in mtDNA take place slowly over time: an ancestor from further back in time is more likely to manifest a mutation along the way than someone a few generations back (Sykes, 2002).

Just as I could trace my mouth through what I could see in a photograph, a testing service that made tangible my mtDNA and maternal haplogroup allowed me to trace that same flow, my mother's mother's mother back 7000, 20,000, 40,000 years.

The 23andMe database identifies similarities in runs of SNPs, suggesting family relationships beyond coincidence. Maureen, a possible 4th to 7th cousin, contacted me on 23andMe:

We have a run of I think around 1800 snps that are identical on Chromosome 2. That probably suggests that we are 5th or maybe 4th cousins which suggests we probably have a great great great grandparent that is the same.

(Maureen, personal correspondence with the author)

Whereas I looked at the shape of a mouth in a photograph and identified my great-grandmother, Maureen looked at data, comparing them in an Excel spreadsheet she keeps in order to compare her chromosomal similarities to people that 23andMe identifies as potential relatives.

In Maureen's characterization, chromosomes, too, are a place. They are themselves a location with the data for certain traits; a location that points to the potential geography from which those traits derive; but also a locus of personal meaning:

I think our connection may be on the Anderson side either from the group that stayed in Sweden or the branch that went to Norway. I say that because on the place we share on Chromosome 2 I also share some overlapping genes with a 3rd cousin who still lives in Sweden whose family came from Stafsing.

(Maureen, personal correspondence, emphasis mine)

In this message, Maureen conflates three elements: Chromosome 2 is a "place" where we connect; this same place on Chromosome 2 connects both of us to Sweden and another relative; and Chromosome 2 is a place that she can access, visualize and navigate, just as I accessed, visualized and navigated the visual similarities across my family, as mapped on a tree.

Discussion

This paper seeks to investigate notions of identity and place as a result of genetic genealogy. An individual's sense of place that conventionally corresponds to a hometown or ancestral land can also be found in how people talk about their chromosomes when they engage with genetic genealogy tools. For such individuals, a part of Chromosome 2 may have as much meaning as a shared hometown or physical characteristic for another.

One's personal relationship with genetic genealogical data, such as from DNA testing sites, offers another angle on digital identity. Throughout the early 1990s, Turkle noted the growing multiplicity of identities as "a set of roles that can be mixed and matched, whose diverse demands need to be negotiated" (Turkle, 1997, p. 180). Martin's (1992) study of people's identifications with their immune systems demonstrate how these experiences are marked by distorting and scale-shifting notions of time and space. She writes, "The self has retreated inside the body, is a witness to itself, a tiny figure in a cosmic landscape, which is the body" (p. 125). More than a "distributed presence" (Turkle, 1997), these retreated selves turn outward. They can visualize the methods of their own transmission across space and time, across trees, as parts of networks. Through their visualization, they are meaningful.

Chromosomal place is aligned with traditional ways of talking about place that in turn exercise an impact on personal identity. As "an organized world of meaning" (Tuan, 2001, p. 6), place is about the familiar and the comfortable, rather than the anonymous and systematized. Tuan characterizes "place" as different than "space:" it has an "aura," an "identity" that develops in the process of becoming acquainted with a space, thus becoming a place (p. 5). Prohansky terms the identification an individual has with place as "place-identity" (1978). A component of self-identity, place-identity

develops and changes over time and is informed by different physical settings and roles. Place-identity is multidimensional, a “potpourri of memories, conceptions, interpretations, ideas, and related feelings about specific physical settings, as well as types of settings” (Proshansky, Fabian & Kaminoff, 1983, p. 60). As someone becomes acquainted with the genetic picture of herself, reading SNPs like someone might read cities or physical traits, her place-identity extends to her data. Her data--in this case, a run of SNPs on the second chromosome--becomes a place that shapes how she sees herself.

If, as Martin (1992) writes, the self retreats into the body, then how does our own engagement with genetic data externalize that relationship? What does it mean when DNA becomes a place, a Sheepy to a run of SNPs? How does this contribute to the notion of a meaningful relationship derived through engagement with one’s own genetic data? And how does that change notions of personal identity and place-identity?

References

- Lyll, S. (1997, March 24). Tracing Your Family Tree to Cheddar Man's Mum, New York Times. Retrieved from <http://www.nytimes.com/1997/03/24/world/tracing-your-family-tree-to-cheddar-man-s-mum.html>
- Martin, E. (1992). The End of the body? *American Ethnologist*, 19(1), 121-140.
- Maureen (2011, February 5). [Message on 23andMe].
- Proshansky, H. (1978). The City and Self-Identity. *Environment and behavior*, 10(147).
- Proshansky, H., Fabian, A., & Kaminoff, R. (1982). Place-identity: Physical world socialization of the self. *Journal of Environmental Psychology*, 3, 57-83.
- Sykes, B. (2002). *The Seven Daughters of Eve: The Science That Reveals Our Genetic Ancestry*. New York: W.W. Norton & Company.
- Tuan, Yi-Fu. (2001). *Space and Place: The Perspective of Experience*. Minneapolis: University of Minnesota Press.
- Turkle, S. (1997). *Life on the Screen*. New York: Simon & Schuster.

License

This article is ©2013 Molly Wright Steenson, and licensed under CC BY-NC-ND.

Algorithmically supported sense-making of network visualizations: a call for reflexivity and transparency

Jeff Hemsley
University of Washington
Information School
United States
jhemsley@uw.edu

Abstract

Based on abundant social media data, researchers employ sophisticated algorithms to represent social networks visually. Researchers rely on these visualizations as a sensemaking tool in a similar way that scatter plots are used to check for relationships between discrete variables. As such, these visualizations played an important role in knowledge production related to social network analysis. This paper demonstrates how the design choices in three frequently-used network layout algorithms can alter the ‘reality’ presented by the data. The paper demonstrates that algorithmic design choices build bias into the presentation of the data. Thus, while network visualization can facilitate recognition of some patterns, it can obscure others, and lead to faulty assumptions about the nature of the data. Researchers must demonstrate an understanding of, and report on, the role of algorithms in their sensemaking and data analysis. Such an understanding can also lead to deeper insights about the underlying data.

Keywords

Algorithms; social network analysis; data visualization; methodology

Introduction

A growing body of literature attempts to understand human behavior from massive collections of data drawn from the public application programming interfaces (API) of social media websites. Using these data, researchers employ sophisticated algorithms to represent social networks visually. And just as researchers use scatter plots to check for relationships between discrete variables, network researchers rely on these visualizations to make sense of complex network data. As such, these visualizations have played an important role in knowledge production related to network research for decades (Bender-deMoll & McFarland, 2006). These efforts have produced insights about information sharing on social networks (Leskovec, Backstrom, & Kleinberg, 2009) and how blogs drive viral content (Nahon, Hemsley, Walker, & Hussain, 2011; Nahon & Hemsley, 2011).

While there is undeniable value in social network analysis and its accompanying visualizations, it is rarely acknowledged that network visualization algorithms mediate the relationship between the data and the researcher and thereby influence sensemaking. In this paper I explore this mediating effect using three different algorithms for network data visualization. Specifically, I ask: How can different network visualization algorithms highlight or obscure specific network features?

Method

This paper demonstrates how three frequently-used network layout algorithms can alter the ‘reality’ presented by the data. The first two are the *force based* algorithms of Kamada and Kawai (1989) and Fruchterman and Reingold (1991). The third, Martin et al.’s (2011) OpenOrd layout, is a multi-level approach that employs an average link clustering technique coupled with a force-based algorithm. It is noted that other layout algorithms (circular, 3D, spring-force, tree, etc.) and graphing programs (e.g. Gephi, Pajek, UCINET, and NodeXL) were compared before the final selection of these three algorithms, each of which was selected for their similarities, differences and frequent use by network

analysis experts. The statistical software R¹, with the iGraph package, was used for all data manipulation and graphing.

The data were collected from October 31st, 2001 to November 28th, 2011 using the Twitter streaming API². Only tweets containing the hashtag #OccupyHouston are used for this analysis. Work with other types of data (e.g. hyperlinks, term co-occurrence, forum interactions) inform this analysis, but Twitter retweets were selected to highlight certain features and limitations of the layouts and to illustrate that algorithms are developed under specific assumptions about the data to which they are applied. In today's media environment, some of these assumptions may not be valid.

I instantiated the network by selecting only retweets from the larger set of tweets. We assume that retweets can be used as human communication trace data that provide clues about information diffusion on social networks. The Twitter API delivers a generous amount of meta data for each tweet, part of which is retweet information. In the case of a retweet, the original tweet is wholly embedded in the meta data (excepting "old style" or manual retweets, not used in this analysis). Cleaning the data involved selecting out all user names and the names of the users who they were retweeting for each tweet. The result was a link list: two column data set with retweeters in the first column and retweeted in the second. Redundancies are ignored for this analysis, which would be problematic in an actual study, but not relevant for the illustrative analysis of this short paper. The network has 768 nodes (users or Twitter accounts) and 1,340 links (cases where some user B retweeted some other user A).

Discussion of findings

For each graphic I discuss some of the more easily made observations about the network from the visualization. Where relevant I relate these observations to the inner workings of the algorithm.

Figure 1 was created using the Kamada and Kawai (1989) layout. From it we can see a large single component, which is a subset of the network where each node connects to at least one other node in the component, surrounded by an unconnected ring. Within the ring it is obvious that some of the dyads, two nodes connected by a link, form their own isolated component, but because of the link overlap it is difficult to determine if it is made up primarily of isolated dyads or other formations. The link overlap is a byproduct of the algorithm, which is designed to situate linked nodes an ideal distance apart. This allows us to visually determine the number of links that information would need to flow from one node to another, which results in concentric circles radiating out from the dense core of the network. The groupings of nodes with a single link all connected to the same node represent cases where many people (end nodes) retweeted a single person's tweet.

¹ <http://www.r-project.org/>

² <https://dev.twitter.com/docs/streaming-apis>

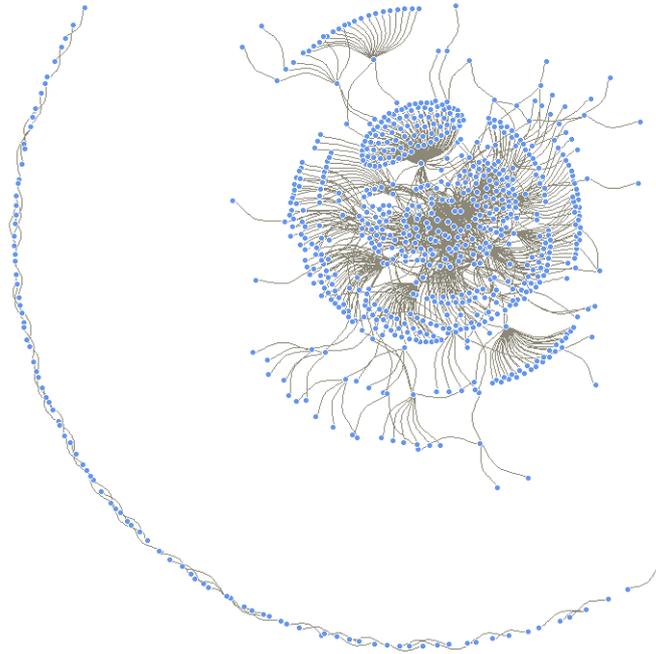


Figure 1: Kamada and Kawai force based network layout.

Fruchterman and Reingold's force directed layout is plotted on the same dataset in figure 2. The idea behind force directed node placement is that linked nodes exert attractive forces on each other, and, depending on the algorithm, repulsive forces. Where Kamada and Kawai attempt to repulse nodes at least as far as the ideal distance, Fruchterman and Reingold's approach is to have all nodes exert a repulsive force on all other nodes. These forces can be calibrated via researcher-chosen parameters. The intent behind the software is for the resulting layouts to have link overlap than the previous layout. A researcher can quickly see that most of the disconnected components are dyads, but a number of them form different configurations. With this layout we can also see some of the linking patterns in the core of the network (heaver linking between nodes).

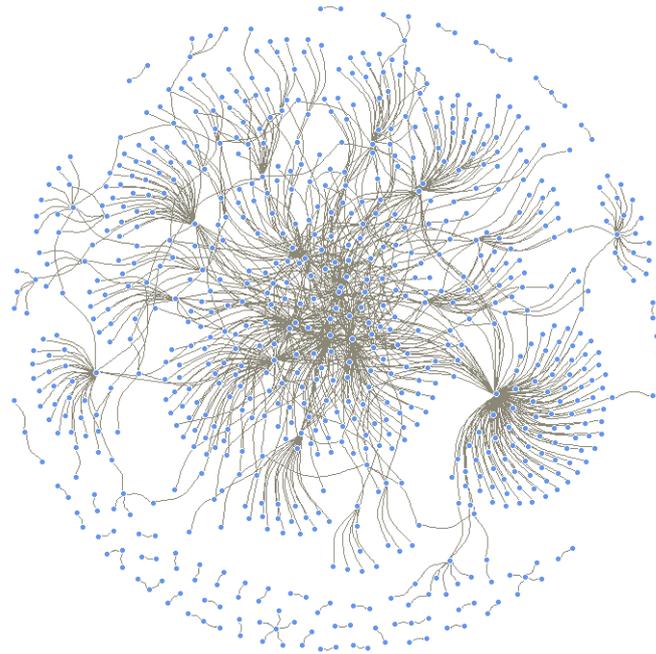


Figure 2: Fruchterman and Reingold force based network layout.

Figure 3 is the OpenOrd layout (Martin et al., 2011). Notice we no longer can discern the sizes of retweet events. Indeed, a researcher looking at this graph may interpret clumped nodes as network clusters, yet they instead may be users bunched together because they all retweeted a single user at the heart of the clump. We also lose any resolution into our disconnected dyads. The algorithm intentionally clumps nodes together when they are tightly connected to each other but loosely connected to distant parts of the graph. Such clustering is useful when the dominant structures of the network are being studied.

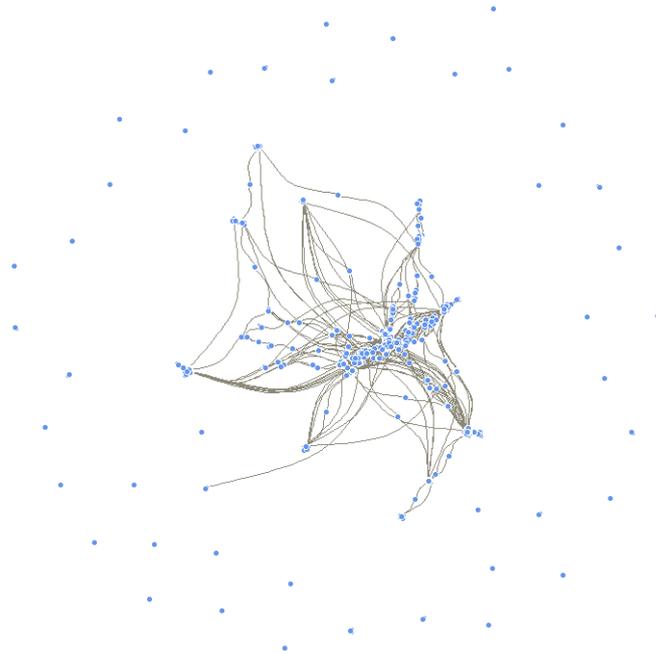


Figure 3: Martin et al.'s OpenOrd layout.

Conclusions

As I demonstrate in this short paper, the design choices of network layout algorithms can produce visualizations that highlight some patterns, obscure others, and may lead to faulty assumptions about the nature of the data. When network visualizations are employed, researchers must demonstrate an understanding of, and reflexively report on, the underlying algorithms that play a role in facilitating sensemaking and data analysis. However, an opportunity exists for researchers with such knowledge in that they can treat visualizations as a specific view into the data. Indeed, research in visual analytics suggests that visualizations can lower the cognitive cost of sensemaking and interpreting large datasets (Thomas & Cook, 2005). Thus, understanding the algorithms implies that researchers can select multiple, complimentary views of the data that, when compared, can yield deeper insights and create opportunities for qualitative interpretation of large datasets.

Acknowledgments

Data used in this work was collected by the Social Media Lab at UW and funded by the National Science Foundation.

References

- Bender-deMoll, S., & McFarland, D. A. (2006). The art and science of dynamic network visualization. *Journal of Social Structure*, 7(2), 1–38.
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11), 1129–1164.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1), 7–15.

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 497–506). Citeseer.

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2011). OpenOrd: an open-source toolbox for large graph layout. In IS&T/SPIE Electronic Imaging (pp. 786806–786806–11). International Society for Optics and Photonics.

Nahon, K., & Hemsley, J. (2011). Democracy.com: A Tale of Political Blogs and Content. Presented at the HICSS-44 (Hawaii International Conference on System Sciences).

Nahon, K., Hemsley, J., Walker, S., & Hussain, M. (2011). Fifteen Minutes of Fame: The place of blogs in the life cycle of viral political information. *Policy & Internet*, 3(1). doi:10.2202/1944-2866.1108

Thomas, J. J., & Cook, K. A. (Eds.). (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr. Retrieved from http://vis.pnnl.gov/pdf/RD_Agenda_VisualAnalytics.pdf

License

This article is ©2013 Jeff Hemsley, and licensed under CC BY-NC-ND.