



Racial formation, inequality and the political economy of web traffic

Charlton Mcllwain

To cite this article: Charlton Mcllwain (2017) Racial formation, inequality and the political economy of web traffic, *Information, Communication & Society*, 20:7, 1073-1089, DOI: [10.1080/1369118X.2016.1206137](https://doi.org/10.1080/1369118X.2016.1206137)

To link to this article: <https://doi.org/10.1080/1369118X.2016.1206137>



Published online: 11 Jul 2016.



Submit your article to this journal [↗](#)



Article views: 2016



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 6 View citing articles [↗](#)



Racial formation, inequality and the political economy of web traffic

Charlton McIlwain

Department of Media, Culture, and Communication, New York University, New York, NY, USA

ABSTRACT

Few studies attempt to demonstrate whether and how systemic racial inequality might form on the web. I use racial formation theory to conceptualize how race is represented, and systematically reproduced on the web, and how both may reveal forms of racial inequality. Using an original dataset and network graph, I document the architecture of web traffic, and the actual traffic patterns among and between race-based websites. Results demonstrate that web producers create hyperlink networks that steer audiences to websites without respect to racial or nonracial content. However, user navigation reflects a racially segregated traffic pattern; users navigate to racialized versus nonracialized websites (and vice versa) more than what would be expected by chance. These results, along with disparities in website traffic rankings, provide evidence of, and demonstrates how a race-based hierarchy might systematically emerge on the web in ways that exemplify disparate forms of value, influence and power that exist within the web environment.

ARTICLE HISTORY

Received 28 December 2015
Accepted 22 June 2016

KEYWORDS

Internet; racial segregation; inequality; social networks; algorithms; racial formation

Conversations about racial inequality on the web are longstanding. However, since the mid-1990s, these conversations proceeded along disparate paths, without a central connecting thread, theoretical model, or substantial evidence to suggest whether and how racial inequality gets produced on the web. Three such conversations stand out. First, differential access to the Internet in the mid-1990s sparked conversation about a racialized digital divide. (Hoffman & Novak, 1998). Despite narrowed access gaps, access questions related to differential digital knowledge, preparation and participation persist (Hargittai, 2010; Schradie, 2012).

A second early conversation focused on problematic racial formations such as identity tourism (Nakamura, 2002), White's domination of racial discourse in online spaces (Burkhalter, 1999; Kendall, 1998), digital Othering (Chow-White, 2006) and the proliferation of White Supremacist organizations online (Daniels, 2009; Futrell & Simi, 2004). Other work focused on various diasporic communities forming online that used Internet affordances to connect with other users of color, and to control racial group representations in

online spaces (Mitra, 1997, 2001). More recent scholarship about racial segregation and social networking sites (boyd, 2013), race-based web navigation (2004) and displaying explicit racism in online spaces (Daniels, 2008, 2013) do not venture far from early conversations about online representation politics germinating in the late 1990s.

A third conversation is least developed, but most significant for our ability to conceptualize and document racial inequality online. Two significant peaks in an otherwise flat scholarly literature mark attempts to address race, technological systems and systematically produced inequality. Friedman and Nissenbaum (1996) proposed that computer systems might propagate a variety of biases, racial and cultural biases included. The authors built a comprehensive framework to identify, assess and potentially remediate such biases; however, the next scholarly work connecting the technical infrastructure of computer systems to racial productions does not arrive until Tara McPherson (2011). She links two narratives that emerged simultaneously, yet independently in the 1960s. One first focused on the development of the UNIX operating system, and the other, post-Civil Rights racial discourse. McPherson argued that the post-civil rights racial logic resembles the lenticular logic underpinning UNIX. Why did it take nearly two decades (between Friedman and Nissenbaum and McPherson's work) to return to questions about the link between digital technology and the systematic production of race? Why did it take so long to imagine a specific relationship between the two in ways that might direct our thinking about how digital technology might shape not only racial identity or racial discourse, but also potentially reproduce racial inequality?

I outline in the following pages a theoretical foundation for exploring how race is represented, and systematically reproduced on the web, and how both may reveal forms of racial inequality. First, I argue that we must conceptualize racial inequality online as a racial – more than purely technological – formation. Second, I use a spatial analytic to briefly outline the dominant political and economic logic of web traffic that underpins today's web. Third, I draw on an original dataset and network graph to document the architecture of web traffic, and the actual traffic patterns among and between race-based websites. These data demonstrate the ways in which a race-based hierarchy might systematically emerge on the web in ways that help us to identify disparate forms of value, influence and power that exist within the web environment.

Racial formation and web inequality

I attempt to develop a framework for conceptualizing and demonstrating how racial inequality might be produced and systematically propagated on the web. The myriad ways that people use online platforms and tools to accomplish race-based or racist goals is not my concern here. Rather, I aim to understand how racial advantage and disadvantage might get systematically produced within the web's structural edifice – one built to assist users in navigating the web and shuttling groups of users in and out of various web spaces.

Recent attention to algorithmic power (Beer, 2009; Gillespie, 2014), the values encoded in them, and the outcomes they produce, is central to identifying how racial inequality might get produced online. However, I argue that racial formation theory provides a broader, and more appropriate framework for understanding how race works on the web. To be sure, the algorithmic lies at the heart of what I think is most significant

here. But algorithms are agnostic tools, used to engineer solutions for which there are antecedent problems or interests – both of which emerge from a set of historical contexts and circumstances. To fully understand how a technological system such as the Internet might produce tangible forms of race-based inequality, we must consider how the Internet developed as a part of a longstanding history and process of racial formation – the complex, racialized historical contexts, circumstances, interests and problems that predate, but may either be exacerbated or corrected by the web's technological environment.

Technological formation as racial formation

Racial formation reflects the ways that individuals and communities conceptualize, represent and articulate the meaning of race, and its significance. Racial formation also encompasses the means by which racial meanings are systemically incorporated, circulated and appropriated by institutions that automatically produce and reproduce, sustain or challenge an existing social and political order that confers access, privilege and power along racial lines. Racial formation theory (Omi & Winant, 1994) helps us to understand the existence, development and persistence of racial inequality by detailing the ways that representational systems (Hall, 1997) and societal institutions, draw on prevailing racial common sense to produce systematic differential advantages and disadvantages that intentionally or unintentionally flow to racial group members.

Identifying a guiding logic animating both racial and technological systems is a useful starting point to understand the Internet as a technological formation that is part of a broader historical trajectory of racial formation. Broadly speaking, a 'logic' is a coherent system of principled reasoning leading to specific conclusions. In computing, logics function to engineer how machines operate. To suggest that both racial formations and technological formations share a central logic is to say that each are, by definition, technological. Technologies – whether we're referring to race or computers – are a means to an end, a tool to produce specific outcomes.

McPherson (2011) argues that a lenticular logic which privileges concealment and modularity animated narratives that birthed both personal computing, and prevailing post-civil rights narratives emerging in the mid-sixties. Lenticular logics mask core operations, be they computational or political. 'UNIX's intense modularity and information-hiding capacity were reinforced by its design, that is, in the ways in which it segregated the kernel from the shell,' McPherson writes, adding that 'the second half [of the twentieth century] increasingly hides its racial 'kernel,' burying it below a shell of neoliberal pluralism' (McPherson, 2011, p. 29).

My point here is twofold. First, while the lenticular was central to computing applications, the logic itself was already socially established. Two significant racial projects were already underway at the time of the computing debates that McPherson cites. First, there was a normative shift toward racial equality. Mendelberg (2001) argues that this shift did not necessarily signal widespread racial attitude change so much as it rendered explicit racial prejudice and discrimination less publicly tolerable. The transformation did not diminish the technological power of racialization. It simply necessitated effective means to conceal the ways that race animated institutional systems.

The second racial project manifesting this lenticular logic is the neoconservative racial project (Hall, 2005), which rearticulated landmark civil rights legislation in the 1960s as

the culmination – rather than the beginning or continuation – of the civil rights struggle. This reframed narrative accomplished two goals: it truncated the movement and singled out legislative action as its primary form of political action. This narrative covertly erases the reality that the civil rights struggle persisted since Reconstruction – what Hall and others refer to as the *long* civil rights movement; obliterates the complex and integral connection between racial struggles, labor struggles, and the freedom struggles of other racial groups and women; minimizes racial inequality's institutional breadth that extends to voting, housing, education, criminal justice, employment, banking, media and beyond; and finally, helps to establish the ideology of colorblindness as the new racial common sense. This extended far beyond simply providing the veneer of neoliberal pluralism; the narrative made it possible to steadily erode the need, desire and ability for government to provide public policy interventions to mitigate race-based inequalities. As such, the neo-conservative racial project normalizes racial silence, such that racial discourse in policy and political contexts is seen as aberrant and deleterious.

I reiterate, but also extend McPherson's central argument. The lenticular logic that pervades both computing and the sociopolitical field in the U.S. is not merely coincidental. The concrete neoconservative racial project that elevates colorblindness and negates public motivation to identify or recognize racial inequality and disparities, and intervene to mitigate it, extends specifically and directly to today's digital technologies, particularly the online environment. Understanding whether and how racial inequality might emerge in the contemporary online environments necessitated viewing the development and contemporary workings of the Internet as part and parcel of this neoconservative racial project.

Racial formation, spatial logics and the political economy of web traffic

The neoconservative racial project extends to the web, which – in addition to the lenticular – mobilizes a spatial logic online that mirrors the spatial/geographical production of racial inequality offline. Historically, space has been the tool by which racialization, and the dominant system of White Supremacy has been both created and sustained. White elites, and White commoners who benefited from whiteness, configured and used space to transform the fantasy of racial distinction into a meaningful reality. By structuring the spatial relationships between people and valuable resources (land, property, infrastructure and institutions), geography formed the foundation for persistent racial problems: segregation, urbanization, ghettoization, race migration, racial zoning, redlining, blockbusting, bussing, integration, gentrification, steering, property tax funding of education, gerrymandering, racially exclusive social networks and the like. Racial inequality was created and has been maintained by defining, assigning value to, and strictly policing inclusion and exclusion into and from racially marked spaces that confer power, privilege and access (Briggs, 2005).

Race and space have been so intimately intertwined in the U.S. that they are essentially co-constitutive (Delaney, 2002; Lipsitz, 2007). Space makes race, and race – its meaning in both concept and everyday life – expresses itself in particular kinds of spaces and spatial configurations. As Lipsitz highlights through what he dubs a White vs. Black spatial imaginary, the spatial alignment with race has resulted in two broad outcomes. First, space is racially marked for the purposes of identity and identification, in ways that simultaneously

serve both maintaining white supremacy and building racial solidarity among people of color. Second, the racial marking of space is used to create networks, consisting of the geographical arrangement of people, power and resources. These spatialized networks systematically control the advantages gained by being granted access to certain networks, and disadvantage through systematic exclusion or unequal access.

I focus on these racial-spatial relationships for two reasons. First, both the web's nomenclature and interface is largely spatial (Graham, 2013; Graham & Zook, 2011). Second, if space structures the web environment in ways that it structures social, political, racial life offline, then it makes sense to ask: how do configurations of web space produce racial meaning? Furthermore, how might these configurations systematically produce race-based inequality that is specifically tied to the web's spatial and economic structure?

Racializing 'sites'

Scholars have long critiqued the web's presumed racelessness as an assumed Whiteness (Burkhalter, 1999; Kendall, 1998; Nakamura, 1995), and have demonstrated how racial representations align with longstanding racial stereotypes (Daniels, 2009, 2013). But such anecdotal examples can still be dismissed as the work of individual web users acting out personal prejudices in online spaces. If we want to know how race and power, advantage, disadvantage and inequality are systematically embedded in the web's operating structure; however, we must first determine how race itself is systematically – rather than anecdotally – represented online. We must ask, what significance and use-value does race have in the political economy of the web, and how does racial meaning and value get systematically produced and circulated throughout the web?

There is a distinction to be made between the social, political and economic structure of the web, as opposed to the structures that have historically governed everyday social life. Historically, racial formations presume the existence of both individual human beings and racial group categories. Individuals exist by virtue of our corporeal being. The individual also exists through legal inscription and political authority that grants political subjects the rights and responsibilities afforded citizens of geopolitical entities. Individuals vote, own property and are held accountable for legal violations, for example. Race, and racial categories historically exist much the same way. Rights, power, privilege, advantage, access and disadvantage: all – more or less – are social characteristics afforded raced individuals and is accounted for through bureaucratic systems that have an interest in tracking both (Morning, 2011; Morning & Sabbagh, 2005). I belabor the point to highlight the contrasting reality that, for all intents and purposes, neither the individual, nor racial categories are similarly salient within the basic structure of the web.

Understanding how race systematically figures into the structure of the web begins with the 'site,' not with the individual, though there is a bit of a paradox here. In one respect, sites that thrive on individual users as the source of content used to produce audiences seem averse to identifying, or even allowing users to systematically identify themselves based on race and/or ethnicity. Web users know that trafficking the web almost always includes opportunities or requirements to reveal personal information. In either case, rarely, if ever are users asked to identify themselves by race/ethnicity. The widespread lack of opportunities to identify by race/ethnicity reveals the unstated presumption that race either does not exist and/or is not useful. To the degree that it is seen as both real

and useful – the lack of opportunity to identify may reveal a belief that such information should not be used to structure the web environment.

Sites replace individuals in the web's central organizing framework; neither you nor I is of interest to those who traffic in it, except in the aggregate. The web's central function is to facilitate the movement of audiences through a network of sites where a secondary commodity gets produced – content (Van Couvering, 2011). Content attracts audiences, but sites are what are most visible on the web. Users are counted and accounted for only as members of an aggregated audience being shuttled around from one source of content to another. As such, two demographics are principally measured in the political economy of the web – audience traffic, and the ranking of the individual sites that are the sources and destinations for audience traffic.

How then does racial identity and identification operate at the level of the site? How do we know whether a site is racialized, and what racial group it is associated with? What are the means by which sites create racial identity and/or are identified as being racially associated with some group or another? Most importantly, how does racial identity and identification at the level of the site get systematically (and automatically) produced and circulated?

Two examples of how racialization functions in everyday life (offline and on) serve as a point of comparison here. First, in societies where race is salient and accounted for, the process of identity and identification is, more or less, reciprocal. Bureaucratic agencies require citizens to identify. The agency sets the categories, citizens choose from the given options and the bureaucracy collects, counts, aggregates and uses the data for various political purposes. Aside from bureaucratic forms of racialization, spaces become racialized based on the flow of bodies in, out and occupying particular spaces. Black spaces are created when a critical mass of black people, for instance, occupy a space. Their occupation, and their codification as 'black space' establish a particular real or assumed character onto a space. It also brings with it a certain valuation (or devaluation). This process of creating racial space with particular cultural associations and valuations underlies things such as neighborhood segregation, gentrification, steering (Briggs, 2005).

A second example demonstrating how websites sites become racially marked in systematic ways can be encapsulated in the website categorization project conducted at DMOZ.org. Categorizing both sites and content are central to how the web works, particularly for search engines, which dominantly mediate web traffic. However, only one attempt has been made to systematically account for, and categorize sites according to racial/ethnic identity. DMOZ.org has, since 1999, relied on a vast team of editors to categorize the web's vast array of websites. The DMOZ project comes closest to approximating the way that racial identity and identification has worked historically, in the more bureaucratic sense. On the one hand, DMOZ editors have the power to create the categories and provide instructions to other editors about how to determine what fits in the category. On the other hand, it is a site's content – or information gleaned or inferred from the site – that is the primary basis for categorization. Sites (more specifically their owners, and/or agents) determine content, not DMOZ. To be sure, racial and ethnic categories are not primary categories in DMOZ's system. One can arrive at varying race and ethnicity site categories from several different starting points in DMOZ's path structure. The most direct route to the greatest number of sites classified by race/ethnicity is: Society > Ethnicity, which leads to several categories of mostly white, European-based 'ethnic' group categories such as

Albanian, French and Romani. One can also get to Hispanic/Latino and African-American on this path: Society > Ethnicity > African > African-American, where one finds 308 websites. A brief description of the sites categorized here states that,

Websites maintained by African Americans, or focusing on African Americans. African Americans are generally people who are classified as citizens of the United States, but also have aboriginal African ancestry. Sites about people born in Africa but living in the United States may also be listed here.¹

This example is instructive in several ways. First, while editors are given some freedom to categorize within certain parameters, there are some clear characterizations and valuations being imposed in describing/defining the category's boundaries. One can readily see how such a description might be problematic or controversial. Second, the description provides too much flexibility to be of much reliable use. Is a site owned and maintained by African-Americans an African-American site, even if its content does not focus on African-Americans? What about U.S. Citizens who have some 'aboriginal African Ancestry,' but would not otherwise classify themselves as African-Americans? Are native Africans living in America African-American? White South Africans living in the U. S.? U.S. citizens with Algerian roots?

These highlight just a few limitations of DMOZ's model for racially categorizing sites. But they are enough to see why it is a failed project in terms of providing a systematic way to account for race within the web's structure. It also demonstrates, perhaps, why various constituents of the web appear agnostic about, or even hostile to such a project altogether. Despite this, and even though DMOZ may not be the best way to account for race on the web, it is significant in one other particular and novel respect. It demonstrates how race could be systematically accounted for, and relied on, within the web's site-content-traffic structure. The metrics used to rank sites and measure traffic to, from and between them rely on sites whose content is identified and categorized. At least one (and likely more) of the prominent Internet traffic measurement enterprises – Alexa.com – is built on top of DMOZ's categorical architecture. To understand how a systematic reproduction of sites categorized by race might work, take, for example, the popular African-American focused news site, *theroot.com*. Alexa ranks the site at 2474.² This ranking includes in its assessment the category to which the site belongs, as traffic is measured in comparison to similarly categorized sites. Furthermore, Alexa uses the top-level category as the primary basis for comparison. In DMOZ's category scheme, *theroot.com* – despite the fact that its content is clearly targeted toward African-Americans and Black-themed issues – is not categorized as a race-based site. Its category tree structure is news > magazines and e-zines > *theroot.com*. Thus, the site's 2474 ranking is based on sites within the news category.

Two consequences of this layered classification/ranking scheme are worth pointing out. First, a site's traffic rank is related to the category to which it belongs, and a site's prominence within a ranked search engine is a result of both its traffic rank, and the sites in the category to which it compares. If one searches the keyword 'news' in Google, for instance, it may come as no surprise that *theroot.com* does not rank in the total number of 368 sites returned in its results. In this case, a site like *theroot.com* is rendered virtually invisible and ghettoized when searched according to its identified category (as a nonracial, news site). The second thing to point out is that somewhere in the web's systemic architecture, race and blackness is being associated, and attributed to the site. How do we know? If

you alter the search just slightly using the phrase ‘black’ or ‘African-American’ news (for which there is no category), then theroot.com falls somewhere within the top three sites returned in Google’s search results.

So why does theroot.com rank at the top of a search for black news when no such category exists, and does not rank at all within the category to which it has been assigned? One likely – though not definitive – answer is that the categorization process is taking place through a different and less transparent process than with DMOZ and Alexa. For one, Google draws on a different source for characterizing and categorizing site content. DMOZ, and by extension, Alexa, relies on human categorization into discreet (and contradictory, and messy) categories. A search engine such as Google relies on coded contextual information – rather than human beings – to render a site’s identity. This code – referred to as meta-tags, are pieces of html code that site producers write into their webpages. These meta-tags – such as title tags, description tags, image tags, snippet tags, and the like – can communicate many things to Google. Such tags give site producers the power to brand sites using descriptive text and imagery they feel best communicates the site’s identity.

The second part of the explanation for why and how the site finds its way to the top of the Google search results for ‘black news,’ rather than just ‘news’ is that Google categorizes sites automatically, not only by collecting and associating content from a page’s various tags, but by associating tagged content with those found on other websites, and by associating this content with search queries. While we do not know how the algorithms specifically operationalize this content to arrive at a particular result, Google’s instructions about how to maximize visibility within its search results provide a clear enough indication that these and other forms of metadata are what it uses to both categorize and rank the web pages it indexes.³

So what does all of this tell us about racial identity and identification on the web? First, the relationship between website categories, traffic rank and search results/rank demonstrate how information about race/ethnicity can be systematically passed along (or withheld) in the web environment, in ways that approximate historical methods of doing so. Second, the inconsistencies between racial or nonracial content categories, and the ability to search for and find catalogued sites based on the presence of racialized content demonstrates a kind of ambivalence about race on the web. Race may not be a significant variable for cataloguing sites, but at the same time, the ability to locate race-based content is clearly valued. This is to say that one feature of the online environment cloaks the significance of race in a veil of supposed colorblindness. Race is apparent, but not transparent.

Finally, my brief discussion thus far about web traffic, search engine rankings and the variety of metrics that figure into those calculations demonstrate the ways that racial identity and identification on the web may have more significant implications beyond personal or collective preferences about representation.

Racial inequality and the political economy of the web

The complicated, confusing, contradictory and haphazard manner in which racial identity and identification functions on the web necessitates a specific racial project. To understand whether and how racial inequality might play out on the web, then we must construe race in particular ways to even render answering the question possible. To identify racial

inequality online requires: a strategic way to define and systematically identify sites by race; identify how audience traffic flows within and between racial and nonracial sites; means to measure traffic between varying racial site categories and identify disparities between them; identify what accounts for the traffic disparities; and, finally, describe what implications follow from these disparities. In sum, we must identify and determine how racial disparities in site traffic might significantly impact the primary constituents of the web – both site ‘owners’ and site audiences in specific material and, perhaps, nonmaterial ways. In the remainder of this paper, I focus primarily on the first two items above, and, to a limited degree, the third.

Race and traffic flow

Consistent with Wu and Ackland (2014), website networks can be conceptualized in two ways: navigational (hyperlink) networks or clickstream networks. Navigational networks consist of outlinks and inlinks to websites and seek to direct user’s navigation patterns within and between websites, based on web producers’ strategic motivations. Clickstream networks, however, consist of users’ actual navigation patterns. When we consider these types of networks in the context of race, the critical question is this: do producers or traffickers of race-based websites seek to create, and/or actually navigate among racially homophilous networks? Homophily pervades social life across multiple contexts, both offline (McPherson, Smith-Lovin, & Cook, 2001), and online (Appiah, 2004; Nowak & Rauh, 2005; Thelwall, 2009), making it reasonable to suspect that both the directed networks created by hyperlinks and the clickstream patterns of actual users would reflect this tendency toward segregation that continues to characterize contemporary residential patterns in the U.S.

I generated two networks to test this hypothesis. I used VOSON,⁴ and beginning with Alexa’s list of ‘Top Black Sites’ as seed sites, I produced a network based on inlinks and outlinks to each. Fifty-six initial seed sites produced a network consisting of 5596 nodes (sites). As a data quality control measure, I reduced the number of nodes significantly by filtering out those with fewer than two connections to other nodes in the network. What remained was a network consisting of 961 nodes, with 3034 links between them.

Beginning with the same seed sites, I also developed a clickstream network using data from Similarweb,⁵ a web analytics service that provides Internet traffic data that track both the amount of traffic websites generate, as well as the directional flow of traffic from and to websites. The clickstream network consists of the seed sites, the sites people arrive at the seed sites from, and sites where people go after visiting the seed sites. The total network consisted of 12,524 Nodes and 25,029 edges. Again, for ease of analysis, I reduced the size of the network into a sample consisting of nodes with two degrees or more. The resulting network consisted of 3233 nodes, and 15,884 links between them.

Given my general hypothesis, the linking structure in these networks may reflect the desire to steer traffic, or navigate in a homophilous/inbreeding pattern based on one of three characteristics: race, content category or traffic rank. The first would reflect motivations to affiliate based on racial similarities characteristic of site owners and/or content, the second solely on content similarity, and the third based on what we might call capital gains – the ability to increase reputational capital (search visibility) by affiliating with others high value sites. To test the above hypothesis, I constructed three separate variables.

Race is a binary variable constructed by assessing a combination of primarily linguistic factors that identified whether race related terminology was found either in the title, description or keyword meta-tags. Sites where race was found in either of those data categories were categorized as racial, as opposed to nonracial. Content category variables were constructed as a binary variable corresponding to whether the site either belonged or did not belong to the following highest represented site categories: arts and entertainment, news media, people and society, Internet/telecomm or business and industry. Finally, I collected data on each site's traffic rank (MozRank), which assesses website prominence based on a number of factors.⁶

I used UCINET, to calculate homophily. The program uses a spatial, autocorrelation algorithm that compares observed versus actual links found among and between racial and nonracial sites. That is, it calculates what we might call an opportunity structure (expected number of connections within and between groups based on chance) and then compares whether the actual connections significantly exceed or fall below what was expected based on chance. A fully homophilous linking pattern would look like this: each group would have a significantly greater number of connections within the group than what would be expected by chance, and the links between groups would be significantly fewer than what would be expected by chance. This is what I designate as a segregated traffic pattern. The networks could exhibit two other patterns. One is an integrative pattern, where nonracial and racial sites have significantly fewer links within, and significantly greater number of links between them. The third potential pattern is what I designate as a status quo traffic pattern. In this scenario, sites do not link in any significant way beyond what would be expected by chance.

Racial homophily

Using the above criteria, the hyperlink network is not racially homophilous. Seen in [Table 1](#), links between nonracial sites in the network are fewer than what would be expected by chance. These sites connect more frequently to racial sites, though not in a significantly greater way. Differently than the nonracial sites, racial sites connect more frequently to other racial sites. But again, the degree to which they connect with other racial sites is not significantly greater than what would be expected by chance. Thus, the hyperlink network exhibits the status quo linking pattern. That is, content producers on these sites do not go out of their way to steer audience traffic to other sites based on the racial or nonracial nature of the site. In a network consisting of racial and nonracial sites, links form randomly between them with little effort to intervene. Again, this is a hyperlink network, meaning that the links function to direct users' navigation to other sites. The presence of a link does not guarantee that a user will actually click on a link and follow it to the specified site.

While content producers of the sites in this network do not direct traffic flow to other sites on the basis of race, the opposite pattern is apparent in the clickstream network that

Table 1. Navigation network racial homophily.

Navigation network	Expected	Observed	Difference	<i>P</i> <
Nonracial/nonracial	1129	977	-152	.20
Nonracial/racial	1385	1467	82	.13
Racial/racial	423	493	70	.25

indicates the incoming and outgoing flow of actual users to and from sites in the network. This network exhibits a segregated pattern. Seen in [Table 2](#), users visiting nonracial sites visit other nonracial sites with greater frequency than what would be expected by chance, and visit racial sites less frequently than what would be expected by chance. Visitors to racial sites, more than what would be randomly expected, visit other racial sites. In each case, visitors to racial and nonracial sites visit similar websites in ways that exceed the opportunity structure.

Content homophily

Given that users navigate to racial and nonracial sites in homophilous ways, it is important to determine whether such patterns exist when other reasons for group formation exist. While users might navigate the web within and across racial boundaries, content preferences also motivate navigational patterns. Content categories in my analysis include: arts and entertainment, people and society, Internet/telecommunication, and business and industry.

From the standpoint of the hyperlink network, the question is whether web producers direct web traffic in ways that steer users toward sites with similar or different categories of content? The data in [Table 3](#) show mixed results. Business and Internet categories reflect status quo patterns, where connections with and between sites in each category do not depart from random expectations. Arts and entertainment sites reflect an integrative pattern, where links to sites outside the category prevail. However, both the news media and people and society categories reflect much more of a segregated pattern, where within-group linking patterns prevail. The homophilic tendencies in these two categories might reflect the broader patterns of polarization found in news and political information consumption patterns (Adamic & Glance, 2005; Lawrence, Sides & Farrell, 2010). While I only test five of many content categories, the results here suggest that web producers, by and large, freely direct users across content categories. Two particular categories, however, reflect the opposite pattern.

The clickstream network reflects no ambivalence. The results in [Table 4](#) are completely uniform; each category reflects a status quo pattern in which users navigate between sites in expected ways, unbound by stronger or weaker connections to sites based on the content category to which it belongs. So, while users do navigate between sites based on their racial characteristics, they, in this instance, do not create such clickstream patterns based on content.

Ranking homophily

The final possibility I test here is whether users connect to sites based on their perceived or actual prominence, especially when a search engine is mediating that navigation. In that

Table 2. Clickstream network racial homophily.

Clickstream network	Expected	Observed	Difference	<i>P</i> <
Nonracial/nonracial	11,127	12,475	1347	.01*
Nonracial-racial	3306	2118	-1187	.01*
Racial-racial	245	85	-160	.05*

*Indicates a significant difference between expected and observed frequencies.

Table 3. Navigation network content homophily.

Navigation network	Expected	Observed	Difference	<i>P</i> <
Not A&E/A&E	1470	470	-1000	.00*
Not A&E/A&E	1217	1735	518	.00*
A&E/A&E	250	732	482	.00*
Not news/not news	2143	2511	368	.00*
Not news/news	732	426	-305	.00*
News/news	62	0	-62	.00*
Not PS/not PS	2547	2756	209	.00*
Not PS/PS	376	181	-195	.00*
PS/PS	14	0	-14	.07*
Not Internet/not Internet	2530	2675	145	.17
Not Internet/Internet	392	262	-130	.18
Internet/Internet	15	0	-15	.06
Not business/not business	2697	2800	103	.20
Not business/business	235	137	-98	.20
Business/business	5	0	-5	.24

*Indicates a significant difference between expected and observed frequencies.

case, the sites most visible and most likely to be visited by users are those at the top of search ranking results. If this economy of visibility is prominently at work, we might expect it to be reflected in two opposing patterns among hyperlink and clickstream networks. In the former, a nonhomophilous pattern may be most desirable, reflecting a web producer's motivation to link to higher ranking sites for the purpose of gaining the capital it might receive from a reciprocal link. In the latter, a segregated traffic pattern would indicate the tendency on the part of users to traffic within neighborhoods of low- or high-ranking sites – creating and maintaining a kind of class-based division between highly and less prominent sites.

Contrary to the possible scenario outlined above for the hyperlink network, the homophily results reflect a status quo pattern, meaning that site producers are not directing users to other sites based on the rank or visibility of the site. Producers link to low- and high-ranking sites in random pattern without purposeful attention to whether linked sites are consistent with the linking site's own rank.

However, as seen in [Table 6](#), the clickstream network tends toward segregation. While this tendency is present, the in-group connection patterns do not, in each instance, depart

Table 4. Clickstream network content homophily.

Navigation network	Expected	Observed	Difference	<i>P</i> <
Not A&E/A&E	7570	6961	-610	.22
Not A&E/A&E	5943	6354	411	.20
A&E/A&E	1164	1363	198	.26
Not news/not news	11,486	11,764	278	.36
Not news/news	2997	2751	-246	.35
News/news	195	163	-40	.38
Not PS/not PS	13,209	13,638	429	.21
Not PS/PS	1430	1022	-408	.21
PS/PS	38	18	-20	.26
Not Internet/Not Internet	12,278	11,938	-340	.28
Not Internet/Internet	2292	2584	291	.28
Internet/Internet	107	156	49	.20
Not business/not business	12,303	12,149	-154	.37
Not business/business	2270	2399	129	.38
Business/business	104	130	26	.30

Table 5. Navigation network prominence rank.

Navigation network	Expected	Observed	Difference	<i>P</i> <
Low Moz/low Moz	741	905	164	.13
Low Moz/high Moz	1470	1467	-3	.45
High Moz/high Moz	726	565	-161	.13

Table 6. Clickstream network prominence rank.

Clickstream network	Expected	Observed	Difference	<i>P</i> <
Low Moz/low Moz	3721	4031	310	.31
Low Moz/high Moz	7341	7263	-78	.08*
High Moz/high Moz	3616	3384	-232	.36

*Indicates a significant difference between expected and observed frequencies.

significantly from chance. I would argue that the one significant departure here – that those navigating low-ranking sites have fewer connections with high-ranking sites – provides at least ample reason to suspect that a kind of segregated, class-based clickstream pattern may exist within the racial web. That is the general segregated pattern exists here, with the qualification that the number of cross-race connections is the only dimension where the observed number of connections differs significantly from chance (Tables 5 and 6).

Discussion

The idea that racism and racial inequality may pervade the online environment is not a novel concept. But until recently, evidence supporting this contention has been anecdotal, largely limited in its focus on interpersonal and representational forms of bigotry that now routinely circulate throughout the web. By contrast, a century worth of attention and documentation lays out the ways in which the concept of race originated, was imbued with social and political significance, and the ways that racial meanings have and continue to be systematically produced and reproduced in ways that deliver advantage and disadvantage along racial lines. To the degree that everyday contemporary social life increasingly tethers us to the online environment, I argue that it behooves us to more fully explicate how race pervades the technological system that is the web – a system that increasingly mediates commerce, politics, education and interpersonal social relations.

Why has such a large scale racial project not been undertaken before now, when we have known from the very beginning that race is a salient feature of life on the web? I would simply argue that our technological formations matured to the point of existing racial formations. When Friedman and Nissenbaum wrote about bias in technological systems in 1996, the commercial Internet was in its relative infancy – so much so that the examples they appealed to revolved around things such as airline reservation systems or bureaucratic computer systems used to deliver services or rewards to specific constituencies. It took time for the web to grow into the expansively connected and fully interactive medium with its own economy and a critical mass of users to make it a place where race may be critically significant. Race was apparent on the web as it existed in 1992.

What I attempted to do in this paper is to take a small step toward outlining where and how we might go about trying to understand how race gets systematically incorporated into the web's environment, and in what way(s) this system might produce advantages and disadvantages across racial lines. The site is central to the web's economy. It is the place where content resides. It is a primary 'place' where race is encoded, apparent and, more or less, visible. Sites and search engines are the two dominant arbiters of web traffic – shuttling, directing, steering or otherwise enabling the movement of people and their data through the web's vast network. It is the place I argue we should look to investigate whether and how racial inequality might get produced on the web in ways that might comport with historical and geographical racial formations.

The first question I attempted to address was essentially, how do we account for race in this environment. My approach to this question is to look to the principal way that the system provides individuals/entities the opportunity to identify – at the level of the site, using self-described tags to characterize the site and its content in ways that are most consistent with site producer's own interests and identity. This makes it possible to accomplish one critical component of identifying and measuring racial inequality: racial classification.

The second question I attempted to address is: what is the spatial relationship between these now racialized sites (racialized, to some degree by the sites themselves or by me through this process of categorization). This question is consistent with the historical fact that racial inequality was largely produced and systematically reproduced through spatial relations. The evidence in these data is by no means conclusive, but they suggest a tendency toward racially segregated site navigation. The model evident in these data – with some qualifications – suggests that two things are going on at once in terms of what we might consider the 'built' environment of the web that seeks to steer audience traffic, and the actual traffic patterns of users as they move to and from sites around the web. Web producers seem to build pathways providing equitable access to sites, without concern for the racial nature of the site. This might produce truly equitable traffic patterns if users only – or even primarily – relied on site links to direct the flow of traffic. But other things intervene – namely, individuals' own prerogatives, search engines or a combination of both. The lesson is, just because people build a road to get from point A to B does not mean people will choose to drive on it, or use it to go from point A to C, when C is a destination that comports more with their individual preferences.

In the networks I analyzed here, such diversions most likely occur when the dividing factor is race, more than the content category to which a site belongs. Layered on top of these racially segregated traffic patterns is a tendency toward a class-based segregation as well, where users tend to navigate to sites with the same, rather than differing status. These two characteristics together – segregated traffic patterns divided by race, and the ranking of sites – form a foundation where inequality along racial lines might exist. And it sets the stage for inequality to exist in its truest form, where disparate outcomes along racial lines are coupled with specific advantages or disadvantages. In the data used for my analysis, it is quite easy to demonstrate that disparate outcomes exist with respect to racial and nonracial sites and their rankings: The mean rank of nonracial sites is 4.5, compared to 4.2 for racial sites. Despite what seems like a miniscule numerical difference in mean scores, the difference is statistically significant at the highest level

($p < .00$). There is a clear disparity between site rankings based on their racial or nonracial classification here.

So what do we do with these findings? What significance do they have for better understanding how racial inequality might be produced in the online environment? First, I argue that these findings demonstrate that two key variables that have historically contributed to racial inequality are present within the web's environment: segregated traffic patterns (that lead to segregated destinations); and disparate valuations of those destinations. This finding supports the notion that race may play a significant role in arbitrating systems of value, access an opportunity online in ways that they have historically done so offline. The findings in this paper are significant as well because once we know that segregation and disparate value exist in the online environment, we know what questions remain for research to ask and answer to fully determine whether and how racial inequality may get produced on the web. These remaining necessary, and most significant questions are these: what actual value do site rankings possess? What traffic advantage(s) are to be gained from having higher site rankings? What disadvantages(s) are there to having lower site rankings? Finally, what are the real implications of, or – differently stated – what is the tangible 'harm' for a site (and, presumably its human owner(s)) that is disadvantaged in this traffic network?

In the same way that addressing such questions are a necessary condition for legal action under the disparate impact theory of discrimination (Morning & Sabbagh, 2005), demonstrating disparate outcomes based on race within the online environment, and establishing the tangible implications/harms they produce will help us to determine more precisely how racial inequality may continue to be systematically produced within the web's political economy. Thus, future research should both replicate aspects of this study to determine whether similar patterns of racial segregation and disparate site valuations are present and develop new research designs aimed at determining the tangible gains and harms produced by such patterns of web traffic.

Notes

1. [http://www.dmoz.org/desc/Society/Ethnicity/African/African American](http://www.dmoz.org/desc/Society/Ethnicity/African/African%20American)
2. <http://www.alexa.com/siteinfo/theroot.com>
3. <https://support.google.com/webmasters/answer/79812?hl=en>
4. <http://uberslink.com/>
5. <http://www.similarweb.com/>
6. <https://moz.com/learn/seo/mozrank>

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by IDEAs Challenge Grant from NYU's Steinhardt School of Culture, Education and Human Development.

Notes on contributor

Charlton McIlwain is an Associate Professor of Media, Culture, and Communication at New York University. His work focuses on the intersection of race, media and politics [email: cdm1@nyu.edu].

References

- Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery* (pp. 36–43). New York: ACM.
- Appiah, O. (2004). Effects of ethnic identification on web browsers' attitudes toward and navigational patterns on race-targeted sites. *Communication Research*, 31(3), 312–337.
- Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, 11(6), 985–1002.
- boyd, d. (2013). White flight in networked publics?: How race and class shaped American teen engagement with Myspace and Facebook. In Lisa Nakamura & Peter A. Chow-White (Eds.), *How race and class shaped American teen engagement with Myspace and Facebook. Race after the Internet* (pp. 203–222). New York, NY: Routledge.
- Briggs, X. d. s. (2005). *The geography of opportunity*. Washington, DC: Brookings Institution Press.
- Burkhalter, M. (1999). Reading race online: Discovering racial identity in USENET discussions. In Smith, M. A., & Kollock, P. (Eds.), *Communities in cyberspace* (pp. 60–75). New York: Psychology Press.
- Chow-White, P. A. (2006). Race, gender and sex on the net: Semantic networks of selling and storytelling sex tourism. *Media, Culture & Society*, 28(6), 883–905.
- Daniels, J. (2008). Race, civil rights, and hate speech in the digital era. *Learning Race and Ethnicity: Youth and Digital Media*, 5, 129–154.
- Daniels, J. (2009). *Cyber racism: White supremacy online and the new attack on civil rights*. Lanham, MD: Rowman & Littlefield.
- Daniels, J. (2013). Race and racism in Internet studies: A review and critique. *New Media & Society*, 15(5), 695–719.
- Delaney, D. (2002). The space that race makes. *The Professional Geographer*, 54(1), 6–14.
- Futrell, R., & Simi, P. (2004). Free spaces, collective identity, and the persistence of US white power activism. *Social Problems*, 51(1), 16–42.
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347.
- Gillespie, T. L. (2014). The relevance of algorithms. In Gillespie, T. L., Bockzkowski, P. and Foot, K. (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). Cambridge, MA: MIT Press.
- Graham, M. (2013). Geography/internet: Ethereal alternate dimensions of cyberspace or grounded augmented realities? *The Geographical Journal*, 179(2), 177–182.
- Graham, M., & Zook, M. (2011). Visualizing global cyberscapes: Mapping user-generated place-marks. *Journal of Urban Technology*, 18(1), 115–132.
- Hall, J. D. (2005). The long civil rights movement and the political uses of the past. *Journal of American History*, 91(4), 1233–1263.
- Hall, S. (1997). *Representation: Cultural representations and signifying practices* (Vol. 2). Thousand Oaks, CA: Sage.
- Hargittai, E. (2010). Digital natives? Variation in internet skills and uses among members of the 'net generation'. *Sociological inquiry*, 80(1), 92–113.
- Hoffman, D. L., & Novak, T. P. (1998). Bridging the racial divide on the Internet. *Science*, 280(5362), 390–391.
- Kendall, L. (1998). Meaning and identity in "cyberspace": The performance of gender, class, and race online. *Symbolic interaction*, 21(2), 129–153.
- Lawrence, E., Sides, J. and Farrell, H. (2010). Self-segregation or deliberation? blog readership, participation, and polarization in American politics. *Perspectives on Politics*, 8, 141–157.

- Lipsitz, G. (2007). The racialization of space and the spatialization of race theorizing the hidden architecture of landscape. *Landscape Journal*, 26(1), 10–23.
- McPherson, T. (2011). US operating systems at mid-century: The intertwining of race and UNIX. In Nakamura L, Chow-White P (Eds.), *Race after the Internet* (pp. 21–37). New York: Routledge.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Mendelberg, T. (2001). *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton, NJ: Princeton University Press.
- Mitra, A. (1997). Diasporic web sites: Ingroup and outgroup discourse. *Critical Studies in Mass Communication*, 14(2), 158–181.
- Mitra, A. (2001). Marginal voices in cyberspace. *New Media & Society*, 3(1), 29–48.
- Morning, A. (2011). *The nature of race: How scientists think and teach about human difference*. Berkeley, CA: Univ of California Press.
- Morning, A., & Sabbagh, D. (2005). From sword to plowshare: Using race for discrimination and antidiscrimination in the United States. *International Social Science Journal*, 57(183), 57–73.
- Nakamura, L. (1995). Race in/for cyberspace: Identity tourism and racial passing on the Internet. *Works and Days*, 25(26), 1–8.
- Nakamura, L. (2002). *Cybertypes. Race, ethnicity, and identity on the Internet*. New York: Routledge.
- Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1), 153–178.
- Omi, M., & Winant, H. (1994). *Racial formation in the United States: From the 1960s to the 1990s*. New York, NY: Routledge.
- Schradie, J. (2012). The trend of class, race, and ethnicity in social media inequality: Who still cannot afford to blog? *Information, Communication & Society*, 15(4), 555–571.
- Thelwall, M. (2009). Homophily in myspace. *Journal of the American Society for Information Science and Technology*, 60(2), 219–231.
- Van Couvering, E. (2011). Navigational media: The political economy of online traffic. In D. Y. Jin & D. Winseck (Eds.), *The political economies of media: Hierarchies, markets and finance in the global media industries* (pp. 1–22). London: Bloomsbury.
- Wu, L., & Ackland, R. (2014). How Web 1.0 fails: The mismatch between hyperlinks and clickstreams. *Social Network Analysis and Mining*, 4(1), 1–7.