Redrawing the "Color Line": Examining Racial Homophily of Associative Networks in Social Media

Authors: Hedwig Lee[1,2], Nina Cesare[1], Tyler McCormick[1,2,3], Julia Morris [1], Ali Shojaie[4]

[1] Department of Sociology, University of Washington
[2] Center for Studies in Demography and Ecology, University of Washington
[3] Department of Statistics, University of Washington
[4] Department of Biostatistics, University of Washington

ABSTRACT

Online social spaces such as Twitter are becoming increasingly salient social contexts for friendship formation. However, demographers have yet to analyze friendship formation and racial segregation in this context. In this paper, we examine racial friendship segregation on Twitter to better understand whether online spaces mirror offline racial segregation trends. Acknowledging past work on the role of structure and agency in friendship and social networks, we argue that Twitter serves to blur the roles of these forces in influencing friendship segregation because users actively create and are influenced by their own "structure." We generate representative samples of Twitter users and their friends and estimate the racial composition of users' friendship networks. We use these data to search for evidence of racial segregation within Twitter as a means of understanding whether race impacts network formation differently on Twitter than it does within offline networks.

INTRODUCTION

Despite policymakers' best efforts to derail a legacy of racial inequality in the United States, neighborhoods, schools, and other social contexts continue to be defined along racial lines. In particular, most associative networks remain racially homogeneous even among cohorts of individuals born decades after the dismantling of Jim Crow and other discriminatory policies (Moody 2001; Quillian and Campbell 2003). A large body of sociological research identifies structural and individual factors associated with patterns of segregation across social contexts (McPherson, Smith-Lovin and Cook 2001; Moody 2001; Quillian and Campbell 2003; Wimmer and Lewis 2010). While no one causal explanation exists, research examining friendship segregation across a variety of contexts concludes that both individual preferences and structural barriers help to either foster or reduce the existence cross-race associative networks (McPherson, Smith-Lovin and Cook 2001; Moody 2001; Quillian and Campbell 2003; Wimmer and Lewis 2010).

Despite their thorough examination of segregation within offline contexts, demographers and social scientists have yet to explore whether similar patterns of network formation appear within social media sites – online spaces that allow users to "(1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (Boyd and Ellison 2007:211). It has been proposed that online spaces are unique social contexts that constitute a "habitus of the new" where agency and structure are perpetually co-evolving (Papacharissi and Easton 2012; Taylor-Smith 2012). Moreover,

structural influences on offline friendship formation, such as living in the same neighborhood or attending the same school, are not present online. Finally, the way in which friendship is defined varies between offline and online contexts. Overall, the unique structure of online spaces invites analysis of how online friendship networks form and if patterns of segregation are produced within these spaces similar to those of offline networks.

Some challenges associated with understanding segregation within online contexts result from the structure of the sites themselves. While some social media spaces, such as Facebook, feature networks that parallel users' offline connections, others, like Twitter, have more self-contained networks and may therefore be less affected by the spatial factors and social norms that impact patterns of friendship segregation. In addition, it is often challenging to gather information on social media users' racial characteristics without directly surveying the users themselves. This is particularly true of Twitter, which features sparse profiles that offer limited information about the user.

In this paper, we use Twitter to develop methods to better understand how online spaces promote or deemphasize racial segregation trends seen offline. We first review the literature examining the role of structure and agency in friendship and social networks and argue that Twitter serves to blur the roles of structure and agency in influencing friendship segregation because users interact in a social space where they are actively engaged in constructing their own "structure" (Papacharissi and Easton 2012; Taylor-Smith 2012). Furthermore, we highlight that Twitter differs from previously analyzed contexts in regard to racial segregation and friendship networks due to the lack of offline macro level and/or institutional restrictions, such as neighborhood level segregation (Massey and Denton 1988; Massey and Denton 1993; Mouw and Entwistle 2006). Given that Twitter's network environment does *not* parallel that of offline social networks, it is possible that these structural factors are less influential or nonexistent in friendship formation. Conversely, the racial compositions of friendship networks on Twitter may simply mirror patterns of racial segregation found in offline social contexts despite the unique structural and agency-based characteristics of the space.

To explore trends in racial segregation within friendship networks we generate representative samples of Twitter users and their networks, as well as utilizing verified techniques for estimating the racial composition of Twitter users' friendship networks (McCormick et al. 2013). In addition to this, we examine the racial composition of these sampled networks in order to determine if Twitter user networks are more or less racially diverse than would be expected if users chose their connections without regard to race (i.e., at random).

METHODS

Description of Twitter

Twitter is a microblogging platform that allows users to record their thoughts in 140 characters or less. The text-based content of these messages may include personal updates, humor, or thoughts on media and politics. This concise format allows users to update their blogs multiple times per day, rather than every few days, as is the case with traditional blogging platforms (Java et al 2007). Besides projecting their thoughts independently, users can communicate with

one another either through private messages, by re-tweeting one another's tweets, or by using the *@reply* command. They may also contribute to broader conversations by including a *hashtag* identifier in their tweet. Tweets from those whom the user follows are displayed as a sequential feed that is updated in real time.

Network structure on Twitter is different than that of other well-known social networking sites such as Facebook and MySpace. Whereas some sites - such as Facebook - are characterized by mutual friendship relations and are intended to parallel or supplement users' offline friendship networks, the "Twittersphere" allows users to maintain directed (i.e., unreciprocated) networks. In other words, a Twitter user may "follow" another user, but that does not require the other user to reciprocate the connection. It is not necessary for the user to know another user in order to follow them, nor must the user be a follower of another user to tweet "at" them (Marwick and Boyd 2010). The level of reciprocity in users' networks varies significantly according to how he or she intends to use the platform. This friendship network structure makes clear that Twitter users' networks often do not necessarily parallel their offline networks and therefore may not be subject to structural constraints present offline. In addition to this, it suggests significant conceptual differences between how friendship is defined offline and how it is defined within this space.

It is important to make a distinction between *space* and *place* when discussing patterns of racial segregation. As noted by the Harrison and Dourish (1996), *space* refers to the objective structure of an environment and *place* is what happens to spaces when users transform them into social settings with unique behavioral appropriateness, cultural expectations, and other normative constraints. Given this distinction, it is clear that Twitter not only differs from offline contexts in terms of *space* but also in terms of *place*. Within Twitter, the active, user-generated construction of "*place*" – an entity somewhat synonymous to Bourdieu's notion of the "habitus" - blurs the line between structure and agency and renders the influence of each on network formation ambiguous (Taylor-Smith 2012). Thus, aside from the structural differences in space between the offline social world and the Twitter community, the co-evolution of structure and agency within Twitter invites social researchers to reexamine the phenomenon of racial segregation and investigate whether it persists within this unique and evolving social place.

Description of Data

Data used for this study were collected using Twitter's streaming API - a quick but limited access means of gathering large amounts of incoming tweets from the website's users. One of the primary advantages of interfacing through the streaming API is the ability to collect data in real-time, as events unfold and as beliefs and opinions are changing across the Twitter feed. For general users, Twitter allows access to approximately 3,000 tweets per second from either a randomly selected sample of all incoming real-time tweets, or from a subset of all real-time tweets which match any of a set of queries (keywords, phrases, or hashtags) as specified by the collector. This method, besides collecting the text and time of the tweet, also easily allows

collection of metadata related to the user whose tweet is collected, including a link to his or her profile picture, location (if supplied by the user in their profile), and handle.

Measuring Friendship and Racial Segregation on Twitter

Given the exploratory nature of this study and the challenge of defining "friendship" on Twitter, we will examine racial segregation within users' follower networks, among those with whom they've engaged in one way exchanges (either by tweeting at or being tweeted at by another user), and those with whom they've engaged in two-way tweet exchanges. We will compare our estimates of these exchanges to the racial composition of friendship and acquaintance networks estimates that would be expected if users chose connections at random, irrespective of race.

PRELIMINARY RESULTS

We constructed egocentric networks from a set of about 150 randomly selected Twitter users. We begin with a randomly selected user, or ego, then collect information about each user the ego has a following relationship with (i.e., either follows or is followed by the ego). We then use the Mechanical Turker to code demographic information from both the ego and her/his network members. Figure 1 displays a boxplot of the proportion of each ego's network that is made up of individuals with the same race. In general Hispanic and Asian users have the smallest portion of their network comprised of individuals with the same race. Blacks display the greatest variability, with the middle half of the distribution ranging from about 20% to 65%. As described above, future work will explore different definitions of association.

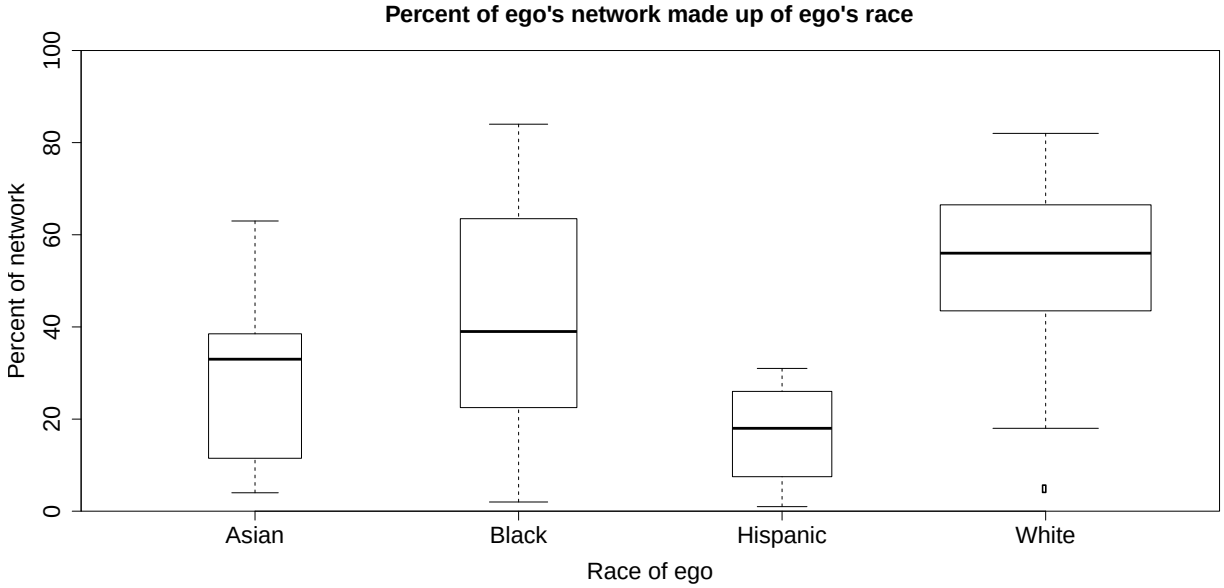

Figure 1. Percent of ego's network that is the same race as the ego.

The results in Figure 1 are limited because they do not account for the racial make-up of Twitter. To address this issue, our final paper will include a standardized, or residual, measure of connectivity. This measure will compare the racial composition of an ego's network to what

would be expected under an assumption of random mixing. Under random mixing, we assume that individuals form ties at random after accounting for differences in ego's network sizes and the prevalence of various groups in the population. If a user has 100 followers, for example, and 40 percent of Twitter users are white, then under random mixing we would expect this user to have about 40 white followers. We will estimate an up to date racial makeup of Twitter users by coding a random sample with the Mechanical Turks, and define preferential mixing based on the difference between the user's observed network and the composition expected under random mixing.

DISCUSSION

In this paper, we will analyze the existence of racial segregation in friendship networks in a previously unexplored online context – Twitter. As mentioned previously, Twitter networks are likely less influenced by offline structural factors that guide patterns of segregation and because Twitter may constitute a "habitus of the new" in which structure and agency are co-evolving and their influence on network formation is less clear. These unique characteristics of the "Twittersphere" invite researchers to reexamine what is known about friendship segregation within this context.

REFERENCES

1. Boyd, D.M, and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230.
2. Papacharissi, Z., & Easton, E. (2012). In the habitus of the new: Agency, structure, and the social media habitus. *Blackwell companion to new media dynamics. Oxford: Blackwell. http://tigger.uic.edu/~zizi/Site/Research_files/HabitusofNewZPEE.pdf.*
3. Pew Research Center, "The Demographics of Social Media Users — 2012" (Pew ResearchCenter, Philadelphia, 2013).
4. Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). ACM.
5. Marwick, Alice E. and danah boyd. 2010 I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media and Society* 13(1): 114 - 133
6. Wimmer, Andreas, and Kevin Lewis. (2010). "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116(2):583–642. Retrieved June 27, 2013 (http://www.jstor.org/stable/10.1086/653658).
7. Massey, DS, and NA Denton. 1988. "Dimensions of Residential Segregation." *Social forces*. Retrieved April 19, 2013 (http://sf.oxfordjournals.org/content/67/2/281.short).
8. Massey, DS, and NA Denton. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.
9. Mouw, Ted, and Barbara Entwisle. (2006). "Residential Segregation and Interracial Friendship in Schools." *American Journal of Sociology* 112(2):394–441. Retrieved May 3, 2013 (http://www.jstor.org/stable/10.1086/506415).
10. McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

11. Moody, James. (2001). “Race, School Integration and Friendship Segregation in America.” *American Journal of Sociology* 107(3):679–716.
12. Quillian, L, and ME Campbell. (2003). “Beyond Black and White: The Present and Future of Multiracial Friendship Segregation.” *American Sociological Review* 68(4):540–566. Retrieved May 3, 2013 (http://www.jstor.org/stable/10.2307/1519738).
13. Harrison, S., & Dourish, P. (1996, November). Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work* (pp. 67-76). ACM.
14. Taylor-Smith, E. (2012). *Participation spaces*. Pre-print for the Fourth international conference on eParticipation, Kristiansand, Norway.
15. McCormick TH, Lee H, Cesare, N, Shojaie A. (2013). Using twitter for social research: Tools for data analysis. *Working Paper No 127*. Center for Statistics and the Social Sciences. University of Washington, Seattle, WA.