Using Twitter to Enhance the Social Support of Hispanic and Black Dementia Caregivers

NCT03865498

IRB approved on 09/13/2021

Study Description

1.   Study Purpose and Rationale

The purpose of this study is to evaluate the impact of the Twitter-based intervention on Tweet content and social network structures of Hispanic and Black dementia caregiver networks. The prevalence of dementia is higher in Hispanics and Blacks than non-Hispanic Whites.[1, 2] Moreover, the caregiving demands in dementia exceed that of other chronic conditions.[3, 4] Consequently, dementia caregivers often experience social isolation and loneliness and decreased health status, and lower quality of life.[5-8] Social support interventions have been demonstrated to help mitigate these issues.[5, 9] The expansion of social media use, particularly Twitter, among Hispanics and Blacks,[10-15] offers great promise for improving social support for Hispanic and Black dementia caregivers.[16] In addition, the availability of large volumes of Tweets and associated metadata make it possible to apply innovative data science methods to characterize engagement in a social support intervention and to assess its efficacy over time.[17-27] However, analytic frameworks for analyzing content and social network structures in Twitter communities are relatively new and evolving rapidly, and there has been limited use of Twitter for intervention delivery.[16]

Our research team has conducted federally-funded research with Twitter data since 2009 and published a practical tutorial on Twitter mining in the *American Journal of Preventive Medicine* in 2013.[25] More recently, through the New York City Hispanic Dementia Caregiver Research Program (R01 NR014430) and a supplement (R01 NR014430-03S2) which focused on understanding mechanisms of social support in Twitter, we have developed a Tweet corpus of >2.6 million English and Spanish Tweets related to dementia caregiving. Moreover, we have applied data science-based cultural analytic techniques for Twitter data mining and developed innovation visualizations for Twitter data.[17, 19-27] Analyses of Tweet topics detected a high frequency of social support topics and revealed that the sentiments of Hispanic dementia caregivers were primarily negative.[17, 19-21, 24] Network analyses revealed a high proportion of isolates (i.e., Tweets without responses) and a low proportion of balanced micro-networks, suggesting sub-optimal social support in the networks.[17-22, 24] We also created a database of >1600 online resources for dementia caregiving.[17-22, 24] This substantial preliminary work provides evidence that our research team is uniquely qualified to operationalize, implement, and evaluate a Twitter-based social support intervention for Hispanic and Black dementia caregivers.[17-31]

2.   Study Design and Statistical Procedures

<u>Design:</u> The study purpose will be achieved through a quasi-experimental design (non-equivalent 2 group pre and post-test design). The intervention group is **anonymous followers** of two Twitter accounts established: the African American/Black dementia caregiver group and the Hispanic dementia caregiver group. **No personally identifiable information (PII)** under the definition used by the National Institute of Standards and Technology (NIST) such as names, addresses, date of birth, and email addresses will be collected. Instead, we will analyze publicly available Tweet data for network engagement (count of pressing 'like,' 'retweet,' 'reply' button), emotional valence (auto-calculation from Tweet text messages: "this is so helpful"-emotion score +4, "I face many challenges"- emotion score -6), topic changes, the volume of isolates, size and number of communities, and communication types for our outcome variables. To examine general confounders, **three demographic elements (self-reported age, gender, race/ethnicity, un-linkable to name, phone, email, or address)**.

The intervention is summarized in table 1 and comprises the posting of one Tweet per day, replying or directly messaging followers, liking someone else's Tweet, retweeting someone else's Tweet, and a monthly Twitter chat at an established time. The Tweet messages which the IRB will approve from our Aim 2 will be used as an intervention. In addition, Twitter who re-confirm Twitter terms of the agreement, including privacy policy and terms of service (Appendix attached), will be required for this study. If someone decides to follow one of our accounts (i.e., participate in the intervention) and subsequently changes their mind, they can easily change stop following our networks. We will retain our de-identified Twitter followers for our Twitter communities in multiple ways: 1) we will invite caregivers in the dementia caregiver registries (including the NHiRP project), who agreed to be contacted for the future study, developed by Drs. Luchsinger and Mittelman to follow the relevant Twitter accounts; 2) we will use online advertising function for marketing our Twitter communities using TwitterAds (15 days in Year 3 and 15 days in Year 4); 3) research flyers will be posted in relevant clinics; 4) we will send an invitation via Twitter to the publicly available user IDs (estimated Black user

IDs-2500, estimated Hispanic user IDs-1800); 4) using the @mention function, we will ask the existing organization and community leaders (e.g., @alzassociation, which has 107K followers) to retweet our invitation messages to reach out their followers; and 5) monthly, we will determine if there are any new key dementia/Alzheimer's Twitter community leaders and send out an invitation to the publicly available user I.D.s.

Table 1. Twitter-based Intervention

Concepts	Tweets	Time	Source	Function	Recipient
Instrumental support	Prepackaged	Daily**	library	Post*	All followers
Appraisal support	Prepackaged	Daily**	library	Post*	All followers
Informational support	Prepackaged	Daily**	library	Post*	All followers
Emotional support	Prepackaged	Daily**	library	Post*	All followers
Engage in social network	Participate in discussion	Daily	Tweeters	Reply, Direct message	All followers (Individual)
Engage in social network	Answer questions from a direct message	Daily	Tweeters	Direct message	An individual
Engage in social network	Answer question from a posted message	Daily	Tweeters	Reply	All followers
Engage in social network	Like followers' Tweets	Daily	Tweeters	Like	All followers
Engage in social network	Retweet relevant Tweets	Daily	Tweeters	Retweet	All followers
Social support	Host 2hr Twitter chat	Monthly	experts	#hashtag	All followers

*posting: open to all followers. **a message daily (e.g., Monday: Instrumental, Tuesday: Emotional, Wednesday: Informational, Thursday: Appraisal, Friday: Instrumental, repeat the sequence to avoid weekday and weekend effect.

The setting, Sample, and Power Analysis: The setting is the U.S. The Sample will be the de-identified flowers of one of the two dementia caregiver networks (Hispanic, Black). Given our recruitment strategies and preliminary work, we anticipate recruiting > 400 de-identified followers in each group. Based on similar Twitter communities (e.g., @ dementiatoday), population N is estimated at 153,300 for 365 days.

For statistical analyses, the necessary sample sizes for hypothesis 1 and hypothesis 2 based upon the power analysis[34] are described below. Hypothesis 1 focuses on within-group differences and Hypothesis 2 on between-group differences. The hypotheses include both continuous (emotion valence) and binary (% of isolates) outcome variables.

Hypothesis 1: Assuming a correlation of pre- and post-test conservatively: reliability of 300 followers will be required. If the correlation between pre- and post is larger, smaller sample sizes are required. Thus, a sample of 300 Blacks and 300 Hispanics is adequate to examine the within-subject change using a paired t-test. The pre-post-test for binary outcomes (% of isolates) is McNemar's test. The minimum required sample size is 208. Thus, the proposed sample sizes will be adequate to test within-group differences.

Hypothesis 2: The total minimum required sample size in each ethnic group (Black and Hispanic) will be 192 for a total of 385 with effect size 0.332 (Cohen's d). Assuming that α=0.05, 1-β=0.80 (power), the total minimum required sample size for the two groups (Black and Hispanic) with effect size 0.1525 (P_0-P_1) is **318 (159 per group).**

Data analysis: Extensive processing is required before the actual data analysis, which comprises data science methods and statistical tests for our hypotheses.

Data processing: We will process Twitter data for data analyses through several steps:1) merging of data files in our institutional high performance computing (HPC) environment (Habanero), 2) converting XML to txt file format into a human-readable format for integrity checking, 3) feature selection, and 4) indexing.[35]

Data science analyses for publicly available de-identified data: We will use a variety of software packages (R, Python, Nvivo, ORA, NodeXL, SAS) for the data science analyses. Since Twitter user demographics and personal data are not directly available, we will apply algorithms to users' Tweets and images to estimate their demographics, including those considered Social Determinants of Health.[36-39] We will stratify those variables into categories for data analysis procedures as relevant. To answer research question 1, we will descriptively characterize (e.g., #Tweets, # followers, #likes, #retweets) Twitter activity in the two networks. To address the second research question, we will apply Syuzhet sentiment analysis and LDA to characterize negative affect scores and topic models, respectively. For the third research question, we will compute the network measures listed in **Table 1**.

Emotional valence, expressing topics, and conversational styles in macro-,meso- and micro-level among **de-identified Twitter followers** will be computed from **publicly available Tweet data.** Unlike the traditional methods, data science methods such as natural language processing (NLP) and network analyses enable us to detect emotional valence score, changes in topics, volume of isolates, and communication styles **without asking the users any questionnaires**. The following table summarizes concepts, measurements, units of analysis, and data science analytic methods.

Table 1. Concepts, Measures, Data Science Methods and Units of Analysis

	Concept	Measures[§]	Data Science Methods	Unit
Content	Emotional valence[40]	Emotional valence score [18, 40-45] detected from text data (e.g., "This is so helpful" – emotion score +4, "I am sad"- emotion score -6)	Syuzhet packages [46]	***
	Social support (informational, instrumental, emotional, appraisal) [47-50]	Topic changes in social support [17, 19-21]	Topic modeling using hierarchical [51] and n-gram embedded LDA algorithms [51-54]	***
Structure	Social network: Macro-level [55]	The volume of isolates, Network type, Density, Centralization, Cohesiveness [26, 27, 55, 56]	Network analysis [17-20, 23, 24, 26, 27]	***
	Social network: Meso-level [57, 58]	Size and number of communities, the volume of the community activities, leaders of community [26, 27, 56, 57]	Clauset-Newman-Moore [59] Girvan-Newman[58]	**
	Social network: Micro-level [60]	16 communication types among triad, number of individual posting activities [26, 27, 56, 57, 60]	Triad census analysis [60]	*

* individual level, ** community level, *** global network level, [§]all measures in real-time)

Statistical Analyses: Pre-post comparisons will be examined from baseline (3 months pre-intervention) to 3 months (9-12) months after initiation of the intervention. Analyses of pre-post mean differences will be performed using SAS PROC TTEST, PROC Mixed, PROC GLIMMIX, and PROC CATMOD, with adjustments for the correlation within participants induced by repeated measures and for unreliability and missing data treatment. The continuous outcome will be analyzed using SAS PROC TTEST, PROC Mixed, and the binary outcome will

be analyzed using PROC GLIMMIX or PROC CATMOD (for the McNemar test), depending on the missing data. We will compare measures among the two groups using independent t-tests for continuous variables and Chi-square tests for proportions to test the second hypothesis. Machine learning-based trend analysis [61, 62] will be applied for the continuously detected emotional valence score.
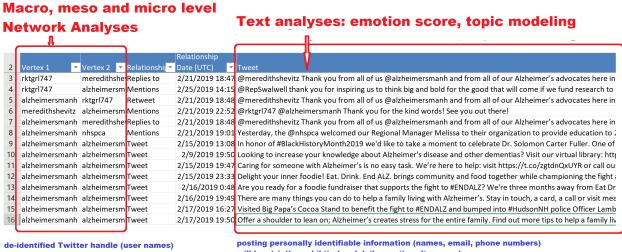
3. Study Procedures

Procedure: Publicly available Tweet data (e.g., Tweets, time stamp, self-reported location) from de-identified Twitter followers of our Twitter sites will be obtained from the public database. No personally identifiable information (PII) under the definition used by the National Institute of Standards and Technology (NIST) such as names, addresses, date of birth, and email addresses will be collected. We will analyze publicly available Tweet data for emotional valence (auto-estimation from Tweet messages), topic changes, the volume of isolates, size and number of communities, and communication types for our outcome variables. In addition, three demographic elements (i.e., self-reported age, gender, race/ethnicity, which are un-linkable to their PIIs) will be collected to examine general confounders. Again, no PII will be used or accessed.

Intervention: Two Tweeters, including a bilingual Hispanic, will strictly follow the intervention delivery protocol developed in Aim 2 with guidance from the Tweet-S2 Expert Panel. Daily for 365 days that cross Years 3 and 4, the Tweeters will post a Tweet from the Tweet library to each Twitter community alternating among those related to information, instrumental, emotional, and appraisal support. The Tweeter will also engage in social network activities by replying to, liking, and retweeting messages posted by others. This intervention will be consistently conducted (e.g., time, center brand themes, designs) to enhance intervention fidelity.[63] After a Tweet is posted, followers can respond using Twitter functions (e.g., retweet to quote and disseminate our Tweet message, posting a Tweet message to express their after thoughts and questions using @mention, like, and reply function). In addition, the P.I. will examine the archive files to monitor intervention fidelity and reinforce expected behaviors of the Tweeters as needed.

Publicly available Twitter data collection: During the baseline (3 months pre-intervention) and intervention (365 days crossing Years 3 and 4) periods, we will collect publicly available Tweet messages and associated metadata (e.g., userID, relationship, time stamp, Tweet, Hashtags, Twitter page, geo code, location, type) daily. We will also double-check the data integrity (e.g., incompletely captured data) daily. In terms of missing Twitter data,[64] we will make an effort to minimize missing data in 3 ways: our designated Tweeters and the P.I. will be available during data collection, two collection tools (NodeXL and NCapture) will be used, and collection of Tweets will be monitored daily to ensure that capture was adequate and with no PII. **Figure 1** shows an example of publicly available data collected.

Figure 1. An example of publicly available data collected via NodeXL: Vertex 1 and 2 will be used for macro, meso, and micro level network analyses. Tweet text with no personal identifiers will be used for text analyses, including emotion score calculation and topic modeling.

Macro, meso and micro level Network Analyses

Text analyses: emotion score, topic modeling

	Vertex 1	Vertex 2	Relationship	Relationship Date (UTC)	Tweet
2					
3	rktgrl747	meredithshev	Replies to	2/21/2019 18:47	@meredithshevitz Thank you from all of us @alzheimersmanh and from all of our Alzheimer's advocates here in
4	rktgrl747	alzheimersm	Mentions	2/25/2019 14:15	@RepSwalwell thank you for inspiring us to think big and bold for the good that will come if we fund research to
5	alzheimersmanh	rktgrl747	Retweet	2/21/2019 18:48	@meredithshevitz Thank you from all of us @alzheimersmanh and from all of our Alzheimer's advocates here in
6	meredithshevitz	alzheimersm	Mentions	2/21/2019 22:52	@rktgrl747 @alzheimersmanh Thank you for the kind words! See you out there!
7	alzheimersmanh	meredithshev	Replies to	2/21/2019 18:48	@meredithshevitz Thank you from all of us @alzheimersmanh and from all of our Alzheimer's advocates here in
8	alzheimersmanh	nhspca	Mentions	2/21/2019 19:01	Yesterday, the @nhspca welcomed our Regional Manager Melissa to their organization to provide education to ï
9	alzheimersmanh	alzheimersm	Tweet	2/15/2019 13:08	In honor of #BlackHistoryMonth2019 we'd like to take a moment to celebrate Dr. Solomon Carter Fuller. One of
10	alzheimersmanh	alzheimersm	Tweet	2/9/2019 19:50	Looking to increase your knowledge about Alzheimer's disease and other dementias? Visit our virtual library: htt
11	alzheimersmanh	alzheimersm	Tweet	2/15/2019 19:47	Caring for someone with Alzheimer's is no easy task. We're here to help: visit https://t.co/zgtdnQxUYR or call ou
12	alzheimersmanh	alzheimersm	Tweet	2/15/2019 23:33	Delight your inner foodie! Eat. Drink. End ALZ. brings community and food together while championing the fight ï
13	alzheimersmanh	alzheimersm	Tweet	2/16/2019 0:48	Are you ready for a foodie fundraiser that supports the fight to #ENDALZ? We're three months away from Eat Dr
14	alzheimersmanh	alzheimersm	Tweet	2/16/2019 19:49	There are many things you can do to help a family living with Alzheimer's. Stay in touch, a card, a call or visit mea
15	alzheimersmanh	alzheimersm	Tweet	2/17/2019 16:27	Visited Big Papa's Cocoa Stand to benefit the fight to #ENDALZ and bumped into #HudsonNH police Officer Lamb
16	alzheimersmanh	alzheimersm	Tweet	2/17/2019 19:50	Offer a shoulder to lean on; Alzheimer's creates stress for the entire family. Find out more tips to help a family liv

de-identified Twitter handle (user names) upon following Twitter networks

posting personally identifiable information (names, email, phone numbers) will be strictly prohibited and daily monitored/removed.

4. Study Drugs or Devices
Not applicable

5. Study Instruments (e.g., Questionnaires, Interview Outlines, Focus Group Guides)
No direct instrumentations such as questionnaires (i.e., survey, interview) will be applied for the outcome measurers. Instead, emotional valence, topics, and conversational styles in macro-,meso- and micro-level among **anonymous Twitter followers** will be computed from the publicly available Tweet data. Unlike the traditional methods, data science methods such as natural language processing (NLP) and network analyses enable us to detect emotional valence score, changes in topics, volume of isolates, and communication styles without asking the users. Table 1 above in Data Analyses summarizes concepts, measurements, units of analysis, and data science analytic methods.

6. Study Subjects

<u>Inclusion criteria:</u> The anonymous dementia caregivers must be 18 years of age or older, Black or Hispanic, living in the U.S. including the U.S. territories, a dementia caregiver with any duration, able to speak English or Spanish/bilingual. The participants must agree to terms of conditions of use and privacy policy and rules of one of the two dementia caregiver networks (Hispanic and Black), the Twitter user agreement of the terms of service, Twitter privacy policy, and Twitter rules, including intellectual property, violence, misconduct, abusive behavior, private information and spam and security, and use a smartphone or a feature phone (i.e., a cell phone with text messaging).

Screening/Determining Eligibility

Upon a Twitter user follows one of our Twitter networks after seeing our advertisement on Twitter or our flyers from our clinics and emails, our team will automatically receive a new follower notification from Twitter. Within 24 hours during the weekdays (72 hours during the weekend), our Tweeter/coordinator will send a screening message via the direct message function in Twitter (script attached) and respond if they are eligible to participate in our research Twitter networks (script attached).

When a Twitter user meets our inclusion criteria, the study staff will remind the user regarding our Twitter networks as a research study, voluntary participation, and unfollowing our Twitter networking sites at any time. Furthermore, using the direct message function via Twitter, our study staffs will ensure that the eligible user 1) creates an anonymous Twitter handle if the person uses his/her full name as a Twitter handle, and 2) hides their names from the profile setting. When a Twitter user does not meet our inclusion criteria, our study staff will request them unfollowing our Twitter network via a direct message function (script attached). Ineligible followers will be blocked within 72 hours after the unfollowing request.

7. Recruitment
Considering 60% attrition rate of anonymous online users, we anticipate to enroll >400 per group, to accrue up to a total of 800 (400 per group, total minimum required Sample are 318 users: 159 per group) and retain anonymous followers (no actual name or phone number/email address on the profile or Twitter handle) for our Twitter communities in multiple ways: 1) A flyer will be distributed up to 5 times (per a week) via an email to the caregivers, who agreed to be contacted for the future studies, in the dementia caregiver registries; 2) we will use an online advertising function for marketing our Twitter communities using TwitterAds (15 days in Year 3 and 15 days in Year 4); 3) research flyers will be posted in relevant clinics (i.e., AIM clinic and neurology units at CUIMC); 4) using the @mention function, we will ask the community leaders (e.g., @alzassociation, which has 107K followers) to post our flyer; and 5) every month, we will contact relevant community leaders within Twitter to post our flyer. This study does not offer any incentives (e.g., no online credits, no monetary compensation).

8. Informed Consent with waiver of documentation (45 CFR 46.116)

Informed consent will be obtained with a waiver of documentation (e.g., signature). Anonymous Twitter users without PII (no name, email address, phone number, or addresses are linkable) will participate in this study, analyzing publicly available de-identified data. Accidental PII exposure will be strictly prohibited (see four strategies in privacy protection) and monitored daily. Therefore, this research to post social

support messages to our Twitter network followers involves minimal risk to the subject. The signature on the informed consent document would be the only record linking the subject to the research.

9. Confidentiality of Study Data

Our study data will comprise 1) publicly available Twitter data and 2) three demographic elements, including self-reported age, gender, and race/ethnicity (un-linkable to name, email address, phone number). For both, our team strictly follows the institutional policy for computers and endpoint devices to store the Tweet corpora. Data preprocessing and manipulation will occur using the secure data repositories at the Columbia University Medical Center, such that no data we collect will be stored out of our control. We will use a secured network for the Department of General Medicine for temporary storage for daily data collection, as well as a university-provided fully encrypted laptop and a university-provided fully encrypted external drive for the long term backup purpose. Although our Twitter data does not contain personally identifiable information, all data will be managed and stored under our institutional policy which all endpoint devices of our research team should be protected with 256-bit encryption (at minimum), which severely restricts access to the data in a far superior manner than password protection. For data sharing and preparation, we will use a passcode-protected Columbia University Habanero HPC environment for the transient file storage while merging, converting, and computing the large volume of data

10. Privacy Protections

The anonymous followers of our Twitter sites will be informed that no identifiers such as names or email addresses will be stored in the data collection. Their PII, such as email addresses and names if the participants contacted the study team via email for the recruitment purpose, will be removed from the recruitment documentation on the day of data collection. All data will be de-identified before analysis, and no identifying information about the individual will be used in reporting the study results.We will implement the following protections against risks:
1) Warning: our two dementia caregiver Twitter networks will have a privacy protection reminder with a warning sign on their title pages. Our Twitter network sites will include short and relevant privacy statements (i.e., sharing aggregate data of age group and estimated gender, how to contact our organization's data protection officer.)
2) Reminding: a designated research team member will post a welcome video reminding the target audience about the protection of privacy using a Tweet direct message function upon a participant following (subscribing) our Twitter sites and post a reminder Tweet about privacy protection once a week (i.e., Posting self-identifiable photos, names, email address or phone number will be strictly prohibited).
3) Monitoring - Privacy: a designated research team member will monitor user-generated content daily for the risks of exposure to PII.
4) Monitoring – Content: A designated research team member will remove Misleading or inappropriate content as necessary daily.

11. Potential Risks

There are some minimal risks specific to this study. First, though we will not collect any direct PIIs nor personally identifying photographic images, and all analyses are done at an aggregate level, followers (i.e., site users/study participants) may accidentally post texts or images that contain PII. Second, there is a possibility that followers may post information that is inaccurate and potentially harmful.

12. Potential benefits of the proposed research to the subjects and others

Those who participate in the Twitter-based intervention study may perceived increased social support (informational, instrumental, emotional, appraisal) and may feel belongingness which is a human basic need.

13. Importance of the knowledge to be gained

The important contributions of this study to knowledge development are both substantive and methodological. The study will extend what is known about dementia caregiving in minority populations and the mechanisms of social support. Moreover, the study will expand what is known about using Twitter as a mode for intervention delivery and methods for creating culture-specific lexica and culturally-sensitive Tweet messages and analysis of Twitter data.

14. Alternatives

This research involves an intervention that presents no greater than minimal risk to participants. If participants change their minds, they can easily change a setting to stop following our Twitter networks.

15. Data and Safety Monitoring

Although our Tweet corpora only include publicly available Twitter data, our team strictly follows the institutional policy for computers and endpoint devices to store the Tweet corpora. Data preprocessing and manipulation will occur using the secure data repositories at the Columbia University Irving Medical Center. No data we collect will be stored out of our control. We will use a secured network for the General Medicine for temporary storage for daily data collection, as well as a university-provided fully encrypted laptop and a university-provided fully encrypted external drive for the long term backup purpose. Although our Twitter data does not contain personally identifiable information, all data will be managed and stored under our institutional policy which all endpoint devices of our research team should be protected with 256-bit encryption (at minimum), which severely restricts access to the data in a far superior manner than password protection. For data sharing and preparation, we will use a passcode-protected Columbia University Terremoto HPC environment for the transient file storage while merging, converting, and computing the large volume of data,

References

1. Gurland BJ, Wilder DE, Lantigua R, Stern Y, Chen J, Killeffer EH, et al. Rates of dementia in three ethnoracial groups. *International journal of geriatric psychiatry*. 1999;14:481-493
2. 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*.12:459-509
3. National Alliance for Caregiving. Dementia caregiving in the u.S. Available at http://www.Caregiving.Org/wp-content/uploads/2017/02/dementiacaregivingfinal_web.Pdf. 2017 Feb
4. Friedman EM, Shih RA, Langa KM, Hurd MD. Us prevalence and predictors of informal caregiving for dementia. *Health Aff (Millwood)*. 2015;34:1637-1641
5. Dam AE, de Vugt ME, Klinkenberg IP, Verhey FR, van Boxtel MP. A systematic review of social support interventions for caregivers of people with dementia: Are they doing what they promise? *Maturitas*. 2016;85:117-130
6. Sarkar S. Impact of caregiving role in the quality of life of family caregivers for persons with Alzheimer's disease. 2015
7. Kovaleva MA, Hepburn KW, Spangler SA. Chronic perceived stress, social isolation, and perceived loneliness as a symptom cluster among dementia caregivers. 2015
8. Chiao CY, Wu HS, Hsiao CY. Caregiver burden for informal caregivers of patients with dementia: A systematic review. *Int Nurs Rev*. 2015;62:340-350
9. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: A systematic review. *Am J Alzheimers Dis Other Demen*. 2013;28:216-222
10. Krogstad JM. Social media preferences vary by race and ethnicity, pew internet & american life project, feburary 2015, http://www.Pewresearch.Org/fact-tank/2015/02/03/social-media-preferences-vary-by-race-and-ethnicity/ accessed on april 25, 2017.
11. Greenwood S, Perrin A, Duggan M. Social media update 2016, pew internet & american life project, november, 2016 http://www.Pewinternet.Org/2016/11/11/social-media-update-2016/ accessed on april 25, 2017.
12. Florini S. Tweets, tweeps, and signifyin' communication and cultural performance on "black twitter". *Television & New Media*. 2014;15:223-237
13. Sharma S. Black twitter? Racial hashtags, networks and contagion. *New Formations*. 2013;78:46-64
14. Brock A. From the blackhand side: Twitter as a cultural conversation. *Journal of Broadcasting & Electronic Media*. 2012;56:529-549
15. Bosker B. Why are african americans more likly to join twitter? 8/3/2011 http://www.Huffingtonpost.Com/2011/08/02/african-americans-twitter-use_n_916411.Html accessed on april 25, 2017. *Huffingtonpost, Tech*. 2011

16. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: A systematic review. *Am J Public Health*. 2017;107:143

17. Yoon S, Suero-Tejeda N, Lucero R, Bakken S. Visualizing social support mechanisms in english and spanish tweets to meet self-management needs of family dementia caregivers (abstract). *AMIA Proc.* 2017:2240

18. Yoon S, Kyung D, Boradwell P, Lee H, Odlum M. What can we learn about the middle east respiratory syndrome (mers) outbreak from tweets? . *Big Data and Information Analytics*. 2017 2:1-5, doi: 10.3934/bdia.2017013

19. Yoon S, Suero-Tejeda N, Lucero R, Bakken S. Visualization of topics and social support mechanisms in english and spanish tweets to meet self-management needs of dementia caregivers (abstract). *Data Science Day 2017 Columbia Data Science Insisute, New York, NY*. 2017

20. Yoon S, Suero-Tejeda N, Lucero R, Bakken S. Visualization of topics and social support structures in twitter as the foundation for insights about dementia caregiving (abstract). *Rita Kobb Nursing Informatics Symposium, FL*. 2017

21. Yoon S, Suero-Tejeda N, Huter B, Bakken S. Visualization of topics from twitter and focus groups as the foundation for insights about dementia caregiving, . *AMIA Annual Symposium Proceedings (Abstract)*. 2016

22. Yoon S, Co MC, Jr., Bakken S. Network visualization of dementia tweets. *Stud Health Technol Inform*. 2016;225:925

23. Yoon S. Analytic strategies of streaming data for ehealth. *Stud Health Technol Inform*. 2016;225:675-677

24. Yoon S. What can we learn about mental health needs from tweets mentioning dementia on world alzheimer's day? *J Am Psychiatr Nurses Assoc*. 2016;22:498-503

25. Yoon S, Elhadad N, Bakken S. A practical approach for content mining of tweets. *Am J Prev Med*. 2013;45:122-129

26. Yoon S, Bakken S. Methods of knowledge discovery in tweets. *N.I. 2012 (2012)*. 2012;2012:463

27. Yoon S. Application of social network analysis and text mining to characterize network structures and contents of microblogging messages: An observational study of physical activity-related tweets. 2011

28. Odlum M, Yoon S. Health information needs and health seeking behavior during the 2014-2016 ebola outbreak: A twitter content analysis. *bioRxiv*. 2017:103515

29. Yoon S, Bakken S. Mining twitter to explore mechanisms of social support to meet self-management needs of hispanic dementia caregivers, . *2016 AMIA Summit on Clinical Research Informatics (Abstract)* 2016

30. Odlum M, Yoon S. Hiv/aids and the millennium development goals: A public sentiment analysis of world aids day twitter chat. *Int J AIDS Res*. 2016;3:129-132

31. Odlum M, Yoon S. What can we learn about the ebola outbreak from tweets? *Am J Infect Control*. 2015;43:563-571

32. Cochran WG. *Sampling techniques*. John Wiley & Sons; 2007.

33. Slonim MJ. Sampling in a nutshell. *New York*. 1960

34. Cohen J. Statistical power analysis for the behavioral sciences: Routledge. 1988

35. Rocklin M. Dask: Parallel computation with blocked algorithms and task scheduling. *Proceedings of the 14th Python in Science Conference*. 2015:130-136

36. Pennacchiotti M, Popescu A-M. A machine learning approach to twitter user classification. *Icwsm*. 2011;11:281-288

37. Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist JN. Understanding the demographics of twitter users. *ICWSM*. 2011;11:5th

38. Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on twitter. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2011:1301-1309

39. Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in twitter. *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. 2010:37-44

40. Lane RD, Chua PM, Dolan RJ. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*. 1999;37:989-997

41. Yoon S, Parsons F, Sundquist K, Julian J, Schwartz J, Burg M, et al. Comparison of different algorithms for sentiment analysis: Psychological stress notes. *Studies in health technology and informatics*. 2017;245:1292-1292

42. Valdivia A, Luzón MV, Herrera F. Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*. 2017;32:72-77

43. Smithies J. Software intensive humanities. *The digital humanities and the digital modern*. Springer; 2017:153-202.

44. Kuamri S, Babu CN. Real time analysis of social media data to understand people emotions towards national parties. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. 2017:1-6

45. Schmidt BM. Do digital humanists need to understand algorithms? 2016

46. Jockers M. Extracts sentiment and sentiment-derived plot arcs from text, package 'syuzhet' version 1.0.4. 2017

47. Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: A conceptual analysis. *Journal of advanced nursing*. 1997;25:95-100

48. Picot SJ. Choice and social exchange theory and the rewards of african american caregivers. *J Natl Black Nurses Assoc*. 1994;7:29-40

49. Hill E.W. A theological perspective on social exchange theory. *J Relig Health*. 1992;31:141-148

50. Nord WR. Social exchange theory: An integrative approach to social conformity. *Psychol Bull*. 1969;71:174-208

51. Lee DD, Seung H.S. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*. 2001:556-562

52. Blei D.M. Probabilistic topic models. *Communications of the ACM*. 2012;55:77-84

53. Mcauliffe J.D., Blei D.M. Supervised topic models. *Advances in neural information processing systems*. 2008:121-128

54. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research*. 2003;3:993-1022

55. West D.B. *Introduction to graph theory*. Prentice hall Upper Saddle River; 2001.

56. Scott J. *Social network analysis*. Sage; 2017.

57. Gottlieb BH. Social networks and social support in community mental health. *Social networks and social support*. 1981;4:11-42

58. Girvan M, Newman ME. Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002;99:7821-7826

59. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. *Physical review E*. 2004;70:066111

60. Wasserman SS. Random directed graph distributions and the triad census in social networks. *Journal of Mathematical Sociology*. 1977;5:61-86

61. Yoon S, Trembowelski S, Steinman RC, Bakken S, Weng C. Temporal analysis of the usage log of a research networking system. *AMIA Jt Summits Transl Sci Proc*. 2014;2014:116-122

62. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Improvements to the smo algorithm for svm regression. *IEEE transactions on neural networks*. 2000;11:1188-1193

63. Ibrahim S, Sidani S. Intervention fidelity in interventions: An integrative literature review. *Res Theory Nurs Pract*. 2016;30:258-271

64. Sullivan TR, Yelland LN, Lee KJ, Ryan P, Salter AB. Treatment of missing data in follow-up studies of randomised controlled trials: A systematic review of the literature. *Clin Trials*. 2017:1740774517703319