MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220172

Who Influences Cancer Conversations on Twitter?: A Comparative Surveillance of Cancer Communications

Nishant Jain [a, b, c], Iris Zachary [a, b, c], Suzanne Boren [a, b]

[a] MU Institute of Data Science and Informatics, University of Missouri, Columbia, Missouri, USA
[b] Department of Health Management and Informatics, University of Missouri, Columbia, Missouri, USA
[c] Missouri Cancer Registry& Research Center, University of Missouri, Columbia, Missouri, USA

Abstract

In this study, we used social network analysis to compare the Twitter social networks of top five cancers in the United States (as ranked by the CDC) to determine the key influencers in cancer-related conversations. We find that organizations and groups geared toward patients that provide patient support, promote cancer awareness, cancer prevention and cancer management comprised up to 40% of influencers. Researchers (24%) and physicians (14%) were also found to be influential participants; the extent of influence varying by each cancer, being as high as 40% research influence for colorectal cancer. Notably, scientific organizations (JAMA, CDC_cancer, AACR) played a key role in conversations about colorectal cancer whereas patient-focused organizations played a greater influencing role in conversations about prostate cancer and skin cancer. This study shows that Twitter data can be a valuable source of cancer surveillance data, and has potential to influence policies, strategies, and research directions around each cancer.

Keywords:

Social media, Cancer, Network Analysis

Introduction

Past studies on cancer conversations on the social media platform Twitter include how social media can be leveraged for cancer awareness [4,7] and cancer patient education and support [5]. To study Twitter-based health conversations, previous studies used various analyses such as cross-sectional reviews [4], sentiment analysis, and qualitative content analysis [6]. However, thus far, to the best of our knowledge, no studies have captured the visual maps of Twitter users' conversations about various types of cancers and there are no visualization studies comparing the sharing behaviors of users discussing cancer on this emerging platform with user-generated content.

Because of the self-reporting nature, Twitter is often vulnerable to misinformation-led campaigns [3]. Therefore, it is important to evaluate the credibility of cancer-related information on this platform. One way to evaluate this aspect is to gauge the characteristics of influencers. For instance, if more credible sources such as experts, professionals, professional scientific organizations, patient focused communities or entities are found influencing those networks, that may be one way to gauge the credibility of what is being said in those social networks. For instance, past studies have indicated that lung and breast cancer can have close-knit support communities but we do not know the kind of communities exist around other cancers. We use powerful visualization techniques to derive shapes of cancer information sharing networks and subnetworks on Twitter. We compare various parameters between these networks by searching for hashtags of CDC-defined top five cancers. We have attempted to find similarities and differences between the characteristics of the major subgroups derived from these online Twitter-based cancer networks. Our objective is to derive more meaningful information from the shapes, top hashtags and top influencers of online communities around search terms for cancer. Further, we have categorized the influencers identified in these conversations as an indicator of credibility of information being provided in those conversations.

Methods

Twitter data was collected and analyzed using NodeXL [2], a social network analysis tool developed by the Social Media Research Foundation. Search terms included the following hashtags: #BreastCancer, #SkinCancer, #ProstateCancer, #ColorectalCancer and #LungCancer. Additional words were added in the search terms as needed to carry out further relevant analysis. The tweets retrieved went back to up to a month from the day of the analysis and were extracted within csv files to use various analytics tools on the dataset. This analysis included Twitter data extracted from 7336 users, 14068 unique tweets and 19635 total tweets. Number of Tweets ranged from 506-8191 and number of users ranged from 324- 3386 users for each of the cancer networks. Figure 1 provides further details about the dataset. Quantitative and network analysis was done using NodeXL. This included retrieving top hashtags and top ten influencers for each search term and its overall network.

Results

We examined the dataset for the number of tweets and users (Figure 1). Further we summarized the network characteristics for each cancer. (Table1) Among the five networks derived, the largest overall network was that of #BreastCancer with 3386 users followed by 1384 users for #ProstateCancer and the smallest overall network was #SkinCancer with 324 users (Table 1). Based on unique edges, which represent unique tweets only (excluding retweets and replies), the largest networks from ascending to descending were in the following order: breast cancer, prostate cancer, lung cancer, colorectal cancer and skin cancer.

Network shapes and topics

The network shape for breast cancer conversation displayed tight crowds discussing the topic of breast reconstruction. The network shape for prostate cancer conversations involved tight crowds discussing online communities around "ending prostate cancer". Lung cancer conversation shape showed tight crowds

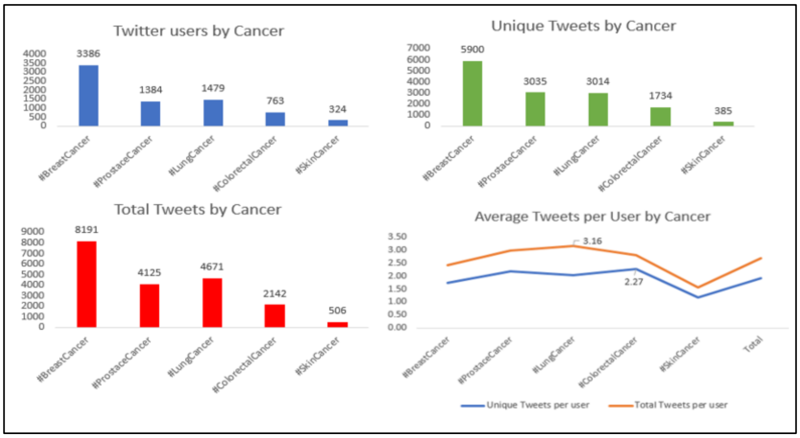

Figure 1 Characteristics of Twitter Dataset by Cancer

Table 1 Network characteristics

Type of Cancer	#BreastCancer	#ProstateCancer	#LungCancer	#ColorectalCancer	#SkinCancer
Entities (Vertices) in the communication	Network of 3,386 Twitter users	Network of 1,384 Twitter users	Network of 1,479 Twitter	Network of 763 Twitter users	Network of 324 Twitter users
Search Term Used	#BreastCancer	#ProstaceCancer	#LungCancer	#ColorectalCancer	#SkinCancer
Unique Edges	5900	3035	3014	1734	385
Total Edges	8191	4125	4671	2142	506
Shape(s) of network and topic	Tight Crowd about treatment and BreastReConstruction	Tight Crowd or Community Clusters #Endprostatecancer #ProstateUK (fundraising perhaps)	Tight Crowd about #AACR21 #LCSM	Tight crowds about screenings use new research early	Brand Clusters health loss risk Symptoms

Table 2 Influencer Category Definition

Category	Criteria: Words found in Twitter Profile Description
Physician/Physician organization	Individual Physician or Physician group
Patient Advocacy Group	If the entity used the word patient advocate or patient advocacy in their description
Patient focussed entities	Support, Awareness, Promotion, Prevention, Management, publicly and self-identified patient/cancer survivor
Public Health Promotion	General Health Promotion not specific to cancer
Cancer Research/Research News	Researcher, Research Organization, Public health research organization, journal editor
Cancer Treatment/App	Treatment
Unidentifiable/Citizen	No particular affiliation could be identified based on description
Celebrity	World renowned celebrity

discussing the latest research presented about lung cancer at the American Association of Cancer Research-2021 (AACR21) conference. Colorectal cancer conversation shape demonstrated Twitter users discussing the early use of new cancer research in a shape of tight crowds and included major organizations such as Journal of American Medical Association (JAMA), CDC_cancer and American Association of Cancer Research (AACR) as influencers. Skin cancer conversations and the network was shaped as branded clusters (disconnected participants) and revealed users discussing loss of health, risks and symptoms associated with skin cancer.

Top hashtags

We examined the top hashtags that were retrieved other than the search term itself. When ranking these hashtags, we excluded the name or the words included in the name of the cancer itself, if they showed up among the top hashtags. The top hashtags for each of the cancer were #bcsm, #pcsm, #lcsm, #crcsm and #melanoma. (Table 4).

Influencer analysis

NodeXL helps to retrieve the top influencers based on betweenness centrality, identifying the most important vertices in the graph. The top influencer results were anonymized and

Table 3 Characteristics of influencers

Type of Influencer for each cancer	Physician/ Physician org	Patient Advocacy Group	Patient focussed entities	Public Health Promotion	Cancer Research/ Research News	Cancer Treatment/ App	Unidentifiable/ Citizen	Celebrity	Total
#BreastCancer	2	2	1	1	1	2	1	0	10
#ProstateCancer	1	0	5	0	3	1	0	0	10
#LungCancer	1	3	2	1	2	0	1	0	10
#ColorectalCancer	1	0	3	2	4	0	0	0	10
#SkinCancer	2	0	4	1	2	0	0	1	10
Total	7	5	15	5	12	3	2	1	50

Table 4 Top Hashtags for each Cancer

Top Hashtags in the Network for each Cancer				
BREAST CANCER	**PROSTATE CANCER**	**LUNG CANCER**	**COLORECTAL CANCER**	**SKIN CANCER**
breastcancer(n=4468)	prostatecancer(n=1697)	lungcancer(n=1945)	colorectalcancer(n=852)	skincancer(n=344)
bcsm(n=530)	cancer(n=137)	**lcsm(n=479)**	**crcsm(n=107)**	**melanoma(n=73)**
cancer(n=507)	**pcsm(n=96)**	aacr21(n=279)	oncology(n=71)	cancer(n=25)
aacr21(n=192)	urology(n=72)	cancer(n=174)	aacr21(n=65)	skincare(n=20)
oncology(n=170)	oncology(n=65)	nsclc(n=160)	cancer(n=53)	sunscreen(n=20)
breastcancerawareness(n=161)	clinicaltrial(n=62)	oncology(n=124)	covid19(n=46)	skin(n=18)
ai(n=161)	aacr21(n=53)	egfr(n=120)	coloncancer(n=34)	dermatology(n=17)
health(n=150)	medtwitter(n=53)	immunotherapy(n=83)	ayacsm(n=29)	cervicalcancer(n=16)
machinelearning(n=135)	menshealth(n=44)	covid19(n=78)	sleevegastrectomy(n=29)	breastcancer(n=12)
breastreconstruction(n=130)	breastcancer(n=37)	danielsjourney(n=72)	bowelcancerawarenessmonth(n=24)	lungcancer(n=11)

tabulated by identifying the category they belonged to (Table 3). For the purpose of this study, we defined the various categories as described in Table 2. Overall, for all the five cancers, the role of influencers was the highest for patient focused entities and patient advocacy groups (40%) followed by researchers (24%), Physicians (10%), treatment focused services (6%) and lowest for celebrities (2%).

Discussion

We used the volume of tweets and users retrieved during a specified period using a certain search term to measure the popularity of the search term within that period. Based on unique edges, which represent unique tweets only (excluding retweets and replies), the largest networks from ascending to descending were in the following order: breast cancer, prostate cancer, lung cancer, colorectal cancer and skin cancer. This order parallels the CDC's ranking of top cancers by new cancer cases from 2017. Considering that Twitter is a relevant platform to check for recent trends and conversations around a topic, it was interesting to see that the most popular networks were in the same order as the order of new cancer cases. Additionally, we observed that Twitter handle, @CDC_cancer was among the top influencers for colorectal cancer networks.

The results for the top hashtags in each network revealed an interesting trend (Table 4, Figure 3). Four of the five cancer communities seemed to rely on an abbreviation derived by joining the "name of the cancer" with "social media". e.g., #BCSM stands for breast cancer social media. #LCSM stands for lung cancer social media. #PCSM and #CRCSM stand for prostate cancer social media and colorectal cancer social media. There seem to be more meaningful and useful discussions and tight crowds around these Twitter hashtags. Interestingly, for skin cancer, as somewhat expected, the hashtag #melanoma seemed to attract more meaningful online communities. Further, there seems to be varying degree of knowhow among the users of Twitter about how to use hashtags for cancer as it seems that lung and skin cancer tweeters had greater use of the abbreviation hashtags that were indicative of their respective cancer communities.

Who drives the discussion?

A key finding was that the discussion within the studied networks was driven by patient-focused or patient-focused entities or patient advocacy-based entities organizations (40%) followed by research-based entities (24%). Only 14% of the conversation was influenced by self-identified physicians. Some influencers could not be categorized because of ambiguous profile information but even for these accounts, a quick survey of their tweets indicated some aspect of patient support as their interests.

What do the shapes tell us?

The five cancers showed different shapes of conversations: closely knit Twitter communities around breast, prostate, lung and colorectal cancers and brand cluster types of networks for skin cancer. The closely knit communities reflect meaningful conversations around topics of interest, while the brand clusters reflect disconnected participants, especially when the topic is related to consumer goods aimed at cancer patients (e.g., skin cancer). Past studies have shown that large hub and spoke shapes of networks on health topics, such as diabetes, can be tied to negative influencers who are polarizing political figures on Twitter, which can sometimes overshadow the more useful conversations [1]. In this study, we found that the conversations were potentially of much useful nature as many closely knit subgroups discussed and shared information between each other for each of the cancers. We attribute this partly to the "hash tagged" search term used in this study.

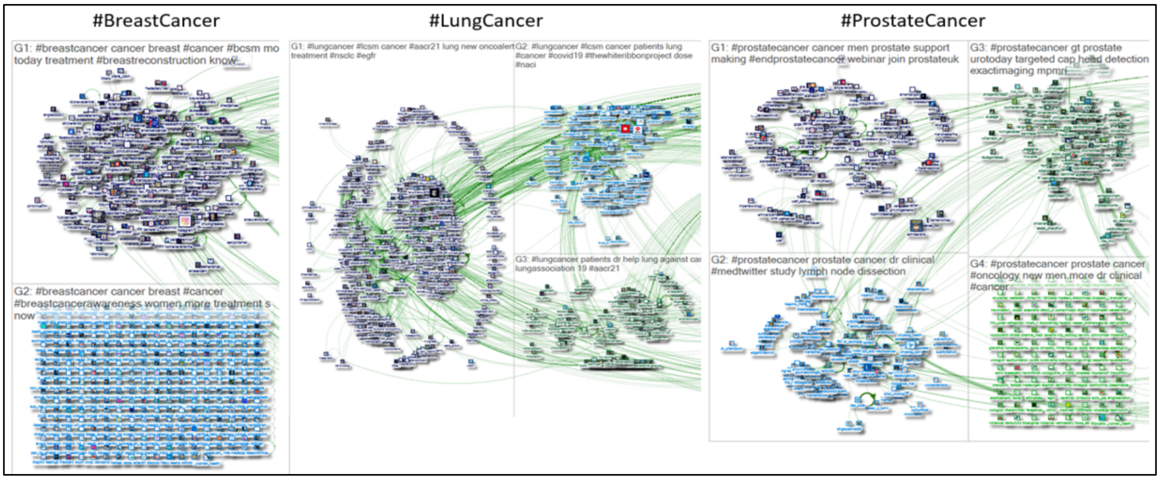

Figure 2a Network Shapes of Top Subgroups (Networks with >1000 Users)

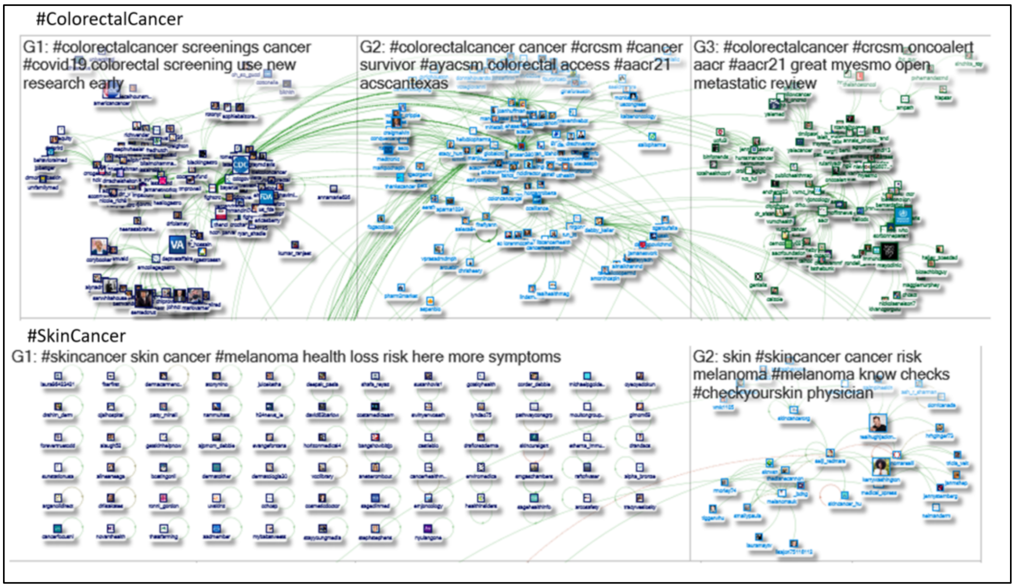

Figure 2b Network Shapes of Top Subgroups (Networks with <1000 Users)

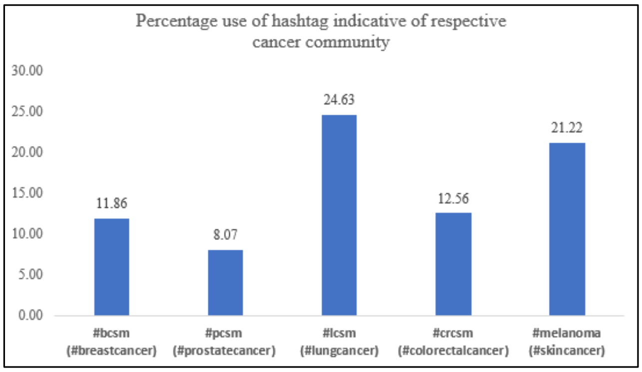

Figure 3 Strategic use of hashtag by cancer community

Conclusions

This study demonstrates that social network analysis can be a useful tool to predict the comparative importance of topics in real time among the public as well as who drives conversations around chronic diseases on Twitter. The study suggests that how public health organizations rank the salience of health topics is mirrored in social media conversations about the same topics. Further, strategic use of hashtags could potentially help researchers and healthcare professionals take the lead in increasingly being able to influence disease-related conversations. Education and training for researchers and health care professionals in strategic social media communication could potentially have a role in improving the social media milieu for important health care topics such as cancer.

References

[1] I. Desideri, S. Pilleron, N.M.L. Battisti, F. Gomes, N. de Glas, N.R. Neuendorff, G. Liposits, I. Paredero-Pérez, W.C.W. Lok, K.P. Loh, C. DuMontier, H. Mian, and E. Soto-Perez-de-Celis, Caring for older patients with cancer during the COVID-19 pandemic: A Young International Society of Geriatric Oncology (SIOG) global perspective, *Journal of Geriatric Oncology*. **11** (2020) 1175–1181. doi:10.1016/j.jgo.2020.05.001.

[2] D. Hansen, B. Shneiderman, and M.A. Smith, Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2010.

[3] S.O. Oyeyemi, E. Gabarron, and R. Wynn, Ebola, Twitter, and misinformation: a dangerous combination?, *BMJ*. **349** (2014) g6178. doi:10.1136/bmj.g6178.

[4] D. Teoh, R. Shaikh, R.I. Vogel, T. Zoellner, L. Carson, S. Kulasingam, and E. Lou, A Cross-Sectional Review of Cervical Cancer Messages on Twitter during Cervical Cancer Awareness Month, *J Low Genit Tract Dis*. **22** (2018) 8–12. doi:10.1097/LGT.0000000000000363.

[5] C. Ure, A. Galpin, A.M. Cooper-Ryan, and J. Condie, Charities' use of Twitter: exploring social support for women living with and beyond breast cancer, *Information, Communication & Society*. **22** (2019) 1062–1079. doi:10.1080/1369118X.2017.1402943.

[6] C. Vasconcelos Silva, D. Jayasinghe, and M. Janda, What Can Twitter Tell Us about Skin Cancer Communication and Prevention on Social Media?, *DRM*. **236** (2020) 81–89. doi:10.1159/000506458.

[7] M. Yoosefi Nejad, M.S. Delghandi, A.O. Bali, and M. Hosseinzadeh, Using Twitter to raise the profile of childhood cancer awareness month, *Netw Model Anal Health Inform Bioinforma*. **9** (2019) 3. doi:10.1007/s13721-019-0206-4.

Address for correspondence

Nishant Jain jainn@health.missouri.edu