

Citizen science on twitter: Using data analytics to understand conversations and networks

MAZUMDAR, Suvodeep <<http://orcid.org/0000-0002-0748-7638>> and THAKKER, D.

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/27838/>

This document is the Published Version [VoR]

Citation:

MAZUMDAR, Suvodeep and THAKKER, D. (2020). Citizen science on twitter: Using data analytics to understand conversations and networks. *Future Internet*, 12 (12), 1-22. [Article]

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Article

Citizen Science on Twitter: Using Data Analytics to Understand Conversations and Networks

Suvodeep Mazumdar ^{1,*†} and Dhavalkumar Thakker ^{2,*†}

¹ Information School, The University of Sheffield, Sheffield S10 2TN, UK

² Department of Computer Science, University of Bradford, Bradford BD7 1DP, UK

* Correspondence: s.mazumdar@sheffield.ac.uk (S.M.); d.thakker@bradford.ac.uk (D.T.)

† These authors contributed equally to this work.

Received: 15 October 2020; Accepted: 22 November 2020; Published: 26 November 2020



Abstract: This paper presents a long-term study on how the public engage with discussions around citizen science and crowdsourcing topics. With progress in sensor technologies and IoT, our cities and neighbourhoods are increasingly sensed, measured and observed. While such data are often used to inform citizen science projects, it is still difficult to understand how citizens and communities discuss citizen science activities and engage with citizen science projects. Understanding these engagements in greater depth will provide citizen scientists, project owners, practitioners and the generic public with insights around how social media can be used to share citizen science related topics, particularly to help increase visibility, influence change and in general and raise awareness on topics. To the knowledge of the authors, this is the first large-scale study on understanding how such information is discussed on Twitter, particularly outside the scope of individual projects. The paper reports on the wide variety of topics (e.g., politics, news, ecological observations) being discussed on social media and a wide variety of network types and the varied roles played by users in sharing information in Twitter. Based on these findings, the paper highlights recommendations for stakeholders for engaging with citizen science topics.

Keywords: citizen science; social network; twitter; crowdsourcing; network analysis

1. Introduction

Citizen science is a research technique that enlists the public in gathering scientific information [1]. Over the last few years, citizen science has seen an increasing interest in adoption by authorities, researchers and organisations. The ubiquitous nature of emerging technologies (mobile devices, IoT, etc.) has facilitated large numbers of volunteers to engage in projects across a wide range of spatial and temporal levels of granularity. As a part of citizen science projects, much of these technologies are employed in collecting data from cities and neighbourhoods, to help inform how our urban environments are evolving and shaping. However, citizen science projects and activities are not only limited to passive collection of data, and often involve citizens and communities. It is important to understand how citizen science is discussed in a conversational manner among the general public as it can help project owners understand their communities better. In addition to access to newer interaction mechanisms (e.g., wearables, smartphones and tablets), the very nature of the Internet has evolved. While traditionally, the Internet was designed for large-scale consumption of information, the modern (social) web has now revolutionised how we interact with the wider world, essentially enabling us to share and express ourselves in ways not possible even a few years back. It is this very conversational, dialogical and user-centric focus of the modern web that has contributed immensely to developing, establishing and nurturing communities online.

Social media, defined as a set of online tools designed to enable and promote social interactions [2] has revolutionised how we communicate with others and share our daily experiences and opinions, providing opportunities for citizens to present their views without being subject to the influence of economic and political pressure [3]. Social media has also been defined as as “a group of Internet-based applications” based on Web 2.0 that “allow the creation and exchange of User Generated Content” [4]. Over the past several years, while multiple terms (“social network sites”, “social networking sites”, “online social networks”, “social networks”) and definitions [5] have been used interchangeably [6], our reference to social media aligns most with the definition offered by [7]. In essence, participants in social network sites have unique identifiable profiles, who consume, interact with and contribute to streams of user generated content and publicly articulate connections they create. Social media, ubiquitous computing and mobile technology have become essential parts of our daily lives and significantly contribute to how we sense our cultures and engage as citizens [8,9]. While social media have been extensively used in a variety of fields for gathering real-time understanding of real-life phenomena in a variety of applications such as emergency response [10], event monitoring [11], social activism [12], crime [13], etc., there is little understanding of how social media contributes to citizen science—particularly within the context of Open data sources such as microblogging sites like Twitter. Many projects and initiatives already have their own online social environments where participants share and discuss ideas. However, such interactions are difficult to study and, often inaccessible to non-project participants. Twitter, on the other hand is an (open) social media platform that supports such a conversational nature of discussion, with easy to access means for collecting data. Users can interact with one another, thereby creating connections (by tagging other users, replying to posts or retweeting user posts) out of their interactions. These connections can form networks of interaction, based on conversations, thereby helping researchers better understand the purpose and behaviour of communications (e.g., broad-/live-casting, information seeking, promotions, etc.).

The work reported here aimed to gather an understanding of the citizen science community and the subsequent networks that are formed as a result of interactions on Twitter about citizen science causes. To our knowledge, this is the first large-scale long-term study of Twitter communities discussing citizen science related topics that is beyond the scope of individual projects or specific domains (e.g., [14–16]). Specifically, we aimed to answer the following research questions:

- RQ1. *What are the predominant topical discussions around the citizen science in microblogging communities?*
- RQ2. *What networks facilitate the spread of citizen science information within microblogging communities?*
- RQ3. *What behaviours do microblogging citizen science communities exhibit?*
- RQ4. *How can citizen science project owners leverage this understanding to increase their presence on microblogging platforms?*

Studying the information available in the public domain can help understand what are the different types of topic the general public are engaging with and what are the different networks that facilitate the dissemination of such topics. This understanding is immensely valuable to many stakeholders such as project owners, practitioners and citizen scientists as this can provide unique insights into the structure of online communities to help better design strategies for communication and dissemination of results. Furthermore, analysing these networks and the conversations on Twitter is key to understand how such platforms are being used by the citizen science community and how they in-turn influence the citizen science community.

We discuss the role of social media in citizen science (Section 1), followed by how social networks have been analysed so far (Section 2); we then discuss how we collected Twitter data on the topic and provide an initial analysis of the data; Section 5 presents more detailed analysis on Twitter communities and Section 6 presents the topical discussions that emerge from automatically analysing our dataset; we finally conclude the paper with discussions and future work (Section 7).

2. Citizen Science and Social Media

Since the growth and popularity of Social Networking sites, there has been an increasing interest in engaging citizens and communities via social media [16,17]. Applications of social media have seen various incarnations within citizen science—for e.g., group-pages, profiles, hashtags, mobile applications, plug-ins, blogs (for example, iSeeChange (<https://www.iseechange.org/>)) or discussion pages (e.g., Zooniverse Talk (<https://www.zooniverse.org/talk>)) and projects often use multiple means to connect with their networks for, e.g., mobile applications often embed social media feeds and plug-ins to share information, observations and reports via social media.

In their study, Ref. [18] identified social media to be one of the primary opportunities for social interactions among participants after face-to-face interactions. In fact, participants are more likely to volunteer if they are more social [19]. It is this inherently interactive nature that platforms and services such as Facebook, Flickr, Twitter, etc. do so successfully, engaging millions of users worldwide. Such platforms provide excellent means for citizen science projects to reach and engage larger audiences [20,21] and the potential of such tools in facilitating connections, developing and organising projects are immense [22]. Several applications of social media have been explored [17] ranging from understanding biodiversity, population trends, landscape [23], park and forest visitor numbers [24], municipal satisfaction [25], cultural services [26], monitoring and predicting events [27], etc. and many projects already employ social media [28] to recruit and disseminate results of the projects, thereby helping in retaining citizen scientists [29]. Most citizen science projects tend to establish an online presence on such platforms [30], aiming to develop and establish a strong community and a dissemination medium [28,31]. While this presence is essential for developing and curating strong networks, the process of science communication itself can be significantly enriched via social media and several guidelines are available that provide recommendations and suggestions on communication strategies [29,32,33].

Much research has been conducted to understand the motivations, quality and nature of participation in (online) citizen science projects so far, albeit within the context of isolated use cases [21,28,34], fields of study [16,17] or integrated frameworks and platforms [35–37]. Several factors such as improvement of skills, enjoyability, reciprocity [38], enhancement of status, social network properties [39], group membership sizes [40] and identification within communities [41–43] are important incentives promoting stronger engagement by participants. [44] present findings from their survey on why people share information on social media: social interaction, information seeking, pass time, entertainment, relaxation, communicatory utility, convenience utility, expression of opinion, information sharing and surveillance/knowledge about others. The studies conducted by [45] identified factors such as collective, intrinsic, norm-oriented motives and reputation that contribute to the quality and quantity of contributions by citizen science volunteers. Many of these findings relate to the very nature of citizen science—the collaborative process of data collection, curation and analysis by a network of individuals to contribute towards a scientific process [28].

3. Understanding Social Networks

While it is important to understand content shared and discussed on social media, the structure and nature of communities on social media themselves offer significant insights into how communities and individuals behave. This is particularly relevant as online social networks can potentially heavily influence opinions in various areas [46]. Very limited studies exist in understanding social networks within the citizen science community and while the present study aimed to fill this knowledge gap, we look at the relevant literature on related topics. At the core of the need to study social networks is the notion that Twitter acts as an information network as well as a social network [47]. Understanding information shared on social media can give insights on how topics and ideas are shared and discussed within communities. In addition to studying content shared on social media, many researchers have also explored the very nature of social connections. Understanding networks in social media can provide key insights into the growth and evolution of communities, support

structures, social structures, hierarchies and sub-networks. There have been several studies aimed at understanding different characteristics of social networks, within a variety of contexts. Ref. [48] studied large social networks to identify user intents and understand communities and their interactions. One of the more studied topics involve understanding political discussions and information spread across social networks. Ref. [49] analysed political Twitter discussions during the 2012 US elections to understand how political user communities are exposed to information based on their interactions and social structures. Ref. [50] showed religiously polarised communications especially during periods of violence following analysis of large scale Twitter data collected in 2013. Ref. [51] studied network structures of data collected during the Occupy Wall Street movement to identify central players in early stages of social discourse and understand how the movement spread online rapidly. Emergency response is another application area where social media analysis can provide critical insights into community behaviour relevant to emergencies. Ref. [52] tracked the network evolution of Twitter communities during the Paris Attacks in 2015. Perhaps closest to one of our aims of understanding online networks is the work done by [15], albeit in a specialised domain of palaeontology. The authors used hashtags relevant to palaeontology that are common in citizen science, in addition to hashtags of popular conferences to collect Twitter data, subsequently investigating users and the flow of scientific information. The study reports on the diversity of people and practices involved in the field. The results suggest a greater breadth and magnitude of the role of the public in sharing scientific information. Our study compliments this study and broadens our understanding to a wider citizen science discourse, as well as investigating multiple dimensions of analysing Twitter communications on citizen science.

4. Methodology

In order to understand social interactions among the citizen science and crowdsourcing (A considerable number of projects use crowdsourcing to collect information from citizens for citizen science activities. As a result, crowdsourcing is a term which is used closely with citizen science.) community, we analysed Twitter data before, during and following the European Citizen Science Association (ECSA) Conference 2018, held in Geneva, Switzerland between 3–5 June (Figure 1). We chose our data collection timeframe to coincide with a popular citizen science event since this would ensure the collection of relevant tweets (and we would not miss out on tweets that are relevant but not using the same keywords specified). In addition, this would also ensure that we would be able to study conversations on citizen science not just during the event, but before and after.

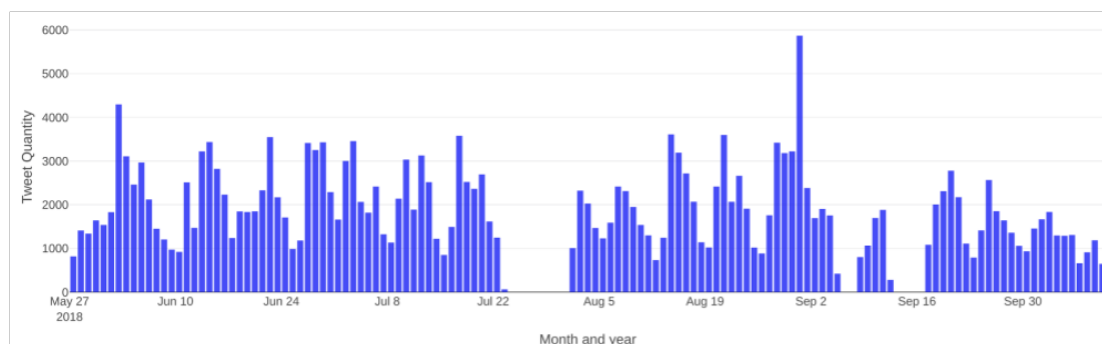


Figure 1. A timeline of the collected Twitter data from 27 May 2018 to 10 October 2018. Missing data indicate technical outages.

It is important to note that with a popular citizen science event, we would expect discussions picked up during the event would be more engaged than usual. As such, data collection after the event was continued to ensure any impact of the discussion during the ECSA event would be visible. However, as can be visible in the tagcloud presented in Figure 2, ECSA does not appear to be a

Java and Solr, using standard data formats such as JSON and was chosen as the tool of choice due to the convenience in easily exporting in formats acceptable in other tools (Gephi). Further to the data collection, more computation-intensive processes analyse the collected data in batches. We analysed Tweets collected using a combination of methods:

1. An initial exploratory visual analysis (using trendlines, timelines, geographic and wordcloud visualisations) of the data collected via the SocialMiner UI (to understand different hashtags and keywords being used, temporal patterns, geographical locations tweeted from). We use findings from this study to partially answer our RQ1.
2. A network analysis of citizen science information shared to understand community structures using an Opensource network visualisation tool, Gephi [53]. We use findings from this study to answer our RQ2 and RQ3.
3. A final analysis using R to conduct topic modelling for thematic analysis [54]. We use findings from this study to answer our RQ1.

We finally reflect upon findings from these methods to build a higher level understanding of how citizen science project owners can benefit from scientific communication on Twitter to answer our RQ4.

5. Exploratory Data Analysis

The period of data collection was from 27 May 2018. The collection was initiated a few days before the ECSA conference to allow capturing any conversations leading up to the event. Data collection continued beyond the ECSA conference until 10 October 2018, allowing for a long-term observation of the citizen science community over four months, except for a few technical outages where the data could not be collected. In total, 238,063 tweets were collected, spread over 124 days, collecting an average of around 2000 Tweets per day. It is to be noted that Twitter only releases 1% of their entire Twitter data via their free online streaming service [55]—as a result, the actual number of tweets generated could have been higher. Since the Tweets made available are capped based on the total tweets generated at the time, we see this random nature of the data collection an expected limitation. However, for less popular topics (as were the ones chosen), this limitation was not expected to significantly impact the data being collected [56]. Figure 1 shows the number of Tweets collected throughout the period of data collection.

As can be seen from the graph, there have been three gaps in the data collection process due to outages. There are two primary peaks—an initial peak immediately before (2 June) the start of the ECSA 2018 conference (3–5 June) and a further peak on the 31 August. Both of these peaks were due to specific tweets that were widely shared by many people in very short amounts of time. Figure 2 presents a tagcloud of hashtags shared on Twitter during the period of data collection.

Overall, several keywords come across as dominant topics of discussion over the period of time. The top ten shared hashtags are as follows: #citizenscience and #citsci (23401 + 1494), #crowdsourcing and #crowdsource (13534 + 1208), #data (5367), #blockchain (2816), #crowdfunding (2663), #innovation (1959), #cswglobal18 (1705), #ai (1225), #decentralization (1158) and #crypto (1144). A large majority of tweets however have not been associated with hashtags (156,221). Among the tweets without hashtags, the most common keywords (without considering stopwords) were crowdsourcing/crowdsource (56408), data (21242), platform (9930), medical (7831), child (7827). Following the distribution of the most tweeted hashtags, we observe (Figure 3) that there are three major peaks of conversation around the most discussed hashtags—an initial peak around the 3–5 June during the ECSA conference (purple colour, #citizenscience), and two peaks mid-way during the data collection (green colour, #data)



Figure 3. Trendline showing distributions of the top hashtags in the Twitter dataset. #citizenscience—purple, #data—green.

Overall, very few tweets (0.04%) have been shared with their geolocations (as seen in Figure 4)—as the map shows, a majority of these tweets have been shared from the US, with a few shared from Europe.

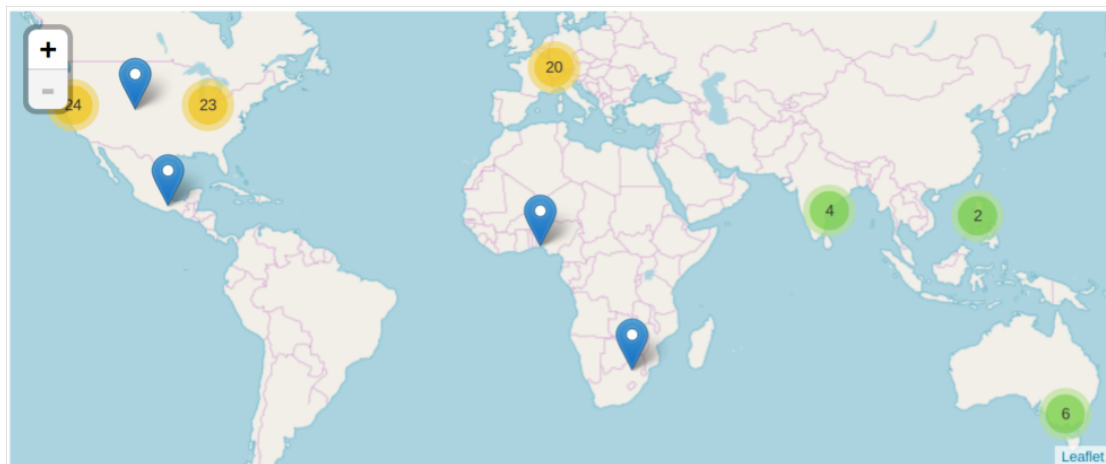


Figure 4. Geographical plot of all the geo-located tweets of the Twitter data, indicating very few tweets have location information.

While this narrative indicates a heavy use of several terms such as crowdsourcing, citizen science, data, blockchain, medical, child and so on, it is important to note the high amount of variation in the topics covered within the citizen science and crowdsourcing community. It is also important to note that a vast majority of the discussions in the space have been retweets (173,568). Retweets are important ways for information to propagate across networks [57] and have been recognised as a key mechanism of information diffusion on Twitter [58]. As such, retweets can potentially indicate the discussions, ideas, articles and news the community engages with and better gauge the ideas the community reads and recognise with [59]. Tweets that were not retweets, on the other hand, can help better understand new content generated by the community.

6. Network Analysis

The citizen science and crowdsourcing community were observed forming several smaller topical communities, primarily developing around specific users. The data collected were processed to identify individual links between users using the '@' mentions, where each tweet can have multiple users being mentioned. Individual links were collected and analysed, to be represented as a node-link graph, where each user is represented by a node, and each interaction (mention) is represented by a directed link between users. Figure 5 illustrates an example interaction, where the users @A and @B have mentioned @C in two different tweets.

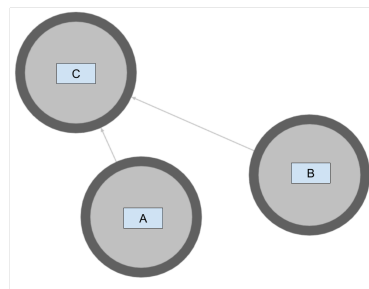


Figure 5. Illustration of a Twitter interaction where nodes indicate Twitter users and links indicate a tweet from a user, tagging another user.

With this representation, a directed graph of the entire network was developed and visualised on a standard visualisation and network analysis tool, Gephi. The network was then analysed to estimate global metrics of the final graph to have a quantitative understanding of its structural properties. The network had 37,452 nodes with 57,424 edges. On an average, the average minimum number of steps (avg. path length) between any two nodes in the network was 5.941. The average of the number of edges adjacent to all nodes (degree) in the graph was 2.001, indicating each node had on an average connections with two other nodes. The network had a high modularity value of 0.892, which indicates how well a network can be decomposed into modular components (e.g., hubs, communities, etc.). The average clustering coefficient, which measures on an average, how complete the neighbourhood of a node is, was a low value of 0.130. We also note a high number of strongly connected components (35,263) and weakly connected components (2025).

Figure 6a indicates the entire network of 37k users, aligned based on a weighted force-directed layout (using the Force Atlas 2, [60]), set to prevent node overlaps. A force-directed layout is a standard approach for visualising graphs and networks, where users are represented as nodes, and interactions as edges. The primary principle of such layouts is that nodes repel each other, while edges act as attractive forces between two nodes. The force of attraction (F_a) is defined as a linear relation to the distance between two connected nodes $n1$ and $n2$ as per

$$F_a(n1, n2) = d(n1, n2)$$

The force of repulsion is defined by considering the degree values of the two nodes as per

$$F_r(n1, n2) = k_r \frac{(deg(n1) + 1)(deg(n2) + 1)}{d(n1, n2)}$$

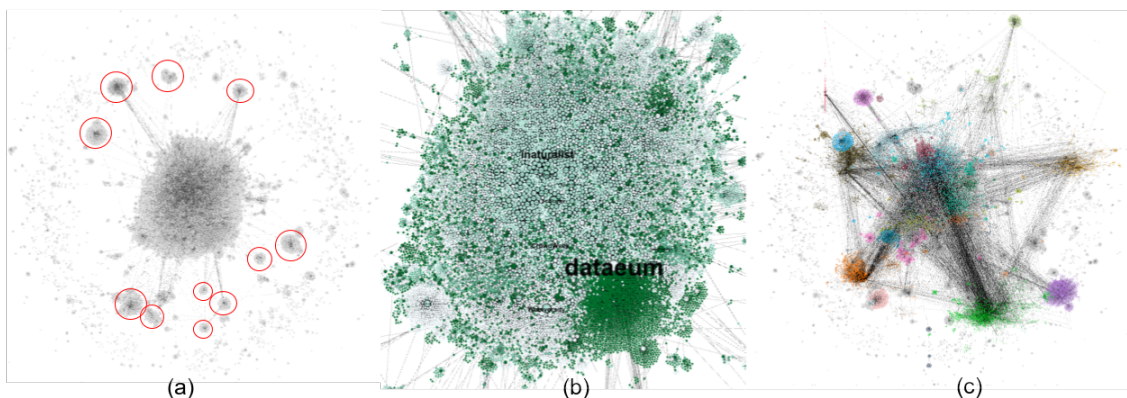


Figure 6. (a) Network visualisation of Twitter discussions on citizen science and crowdsourcing; (b) visualisation of the inner core; (c) visualising hubs within the inner core of the network.

This is particularly helpful as communities appear as groups of nodes [61]. In the figure, we can note three characteristics: (i) a large number of clusters (red circles), where a central node interacts with a high number of nodes in its vicinity (this, in addition to (ii) explains the high modularity scores); (ii) a heavy core of the network with the largest number of nodes at the heart of the network, with strong connections (this explains the high number of strongly connected components); and (iii) a very large number of users in the periphery of the network with limited interactions with other users (this explains the low average clustering coefficient score of 0.130). The formation of clear clusters is interesting and users in these clusters appear to interact only with central nodes who have strong links with the core of the network or other clusters. Let us consider some examples of such clusters (Figure 7).

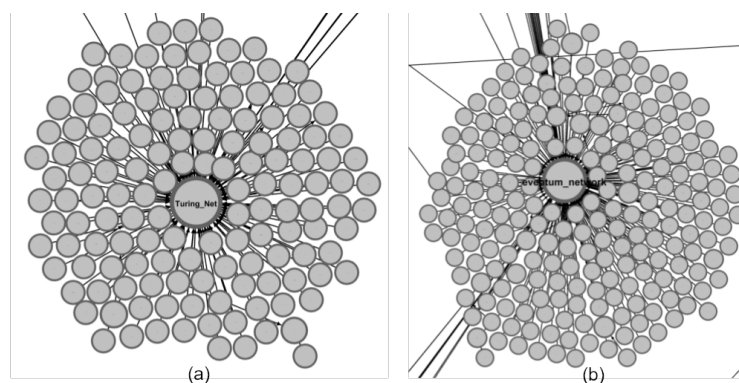


Figure 7. Examples of small clusters of users—(a) Turing Net AI Platform and (b) Eventum Network

The two clusters have the central user @Turing_Net (TuringNet AI Platform, Figure 7a) and @eventum_network (Eventum Network, discussing cryptocurrency news, Figure 7a). Both these users can be observed to have in-bound interactions with large number of users. Upon analysing the tweets, it is observed that these interactions are mainly retweets of initial tweets posted by these users.

Turing_Net (496 conversations) for example, had tweeted (on 29 August 2018):

```
‘‘One company won’t do. #Decentralization of #AI is the way to go!
#TuringNet will crowdsourcethe improvement of AI, foster a healthy
ecosystem where all participants will receive a fair share of rewards.
@APompliano @BTCTN @crypto @AINewsletter @techreview @ForbesTech
@techradar https://t.co/NuRJ9iEdv1’’
```

which was favoured 469 times. A response to this tweet (on 15 September 2018),

```
‘‘Great product, hopt to see turing being use in the future @kurulpier
@Turing_Net @kmlveo @HerbertRSim @6BillionPeople @CryptoBoomNews
@YudhaPu88192935 @IrisY_Jura https://t.co/rlz1iyIXBd’’
```

was retweeted (55) and favoured (115) multiple times during the period of data collection, which explains the cluster developing around Turing_Net. A similar interaction was observed in the @eventum_network cluster, where large numbers of users retweeted the following tweet:

```
‘‘RT @eventum_network: Read more about the OXgame betting experiment and how Eventum helped power it
#Blockchain #crowdsourcing... ’’
```

In fact, examining several such clusters (red circles) in Figure 6a shows the majority of the interactions are in-bound, and primarily retweets of popular users/tweets. The following illustrates the different activities:

- **Crowdsourcing/funding initiatives to:** find stolen laptop; support a defamation case against a politician; seek stories on sharing/hiring/borrowing assets or persons; seek lawyers willing to help the Philippine Long Distance Telephone Company (PLDT) employees for their contract termination;

- **Promotion of:** a hackathon to crowdsource mobility solutions; a homebuyer's app; a marine mammal surveyor course on citizen science; maternity t-shirts; a non-profit organisation to promote Democratic candidates;
- **News/Opinions regarding:** how crowdsourcing (WhatsApp) was used to disrupt anti-terrorist operations in Kashmir; the release of a book on crowdsourcing for filmmakers; US DOJ's actions on election integrity.

Observing a variety of such clusters (groups) indicates the primary function of them as support clusters, based on specific topics and purposes. As observed in Figure 6a, the large number of clusters surround a heavy core at the centre. This core (Figure 6b), consisting of around 23.8 k nodes and 41.5 k edges was isolated from the wider network, and studied for different characteristics (Figure 8).

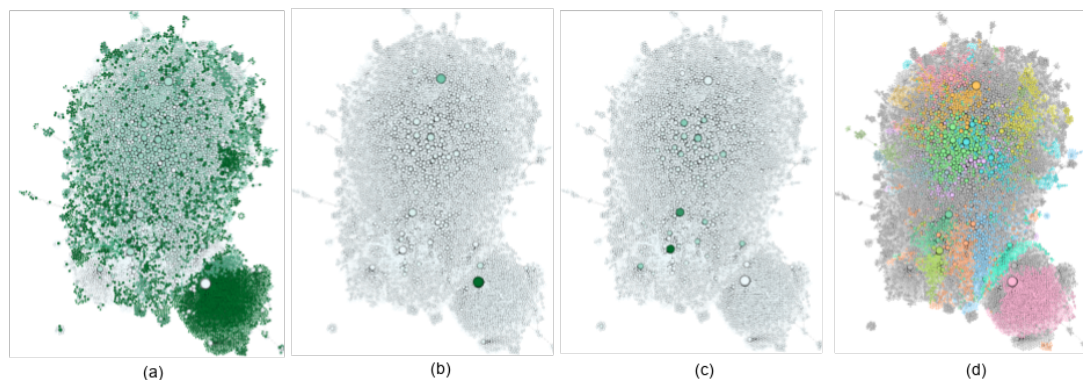


Figure 8. Isolating the inner core and colour-coding based on different measures—(a) closeness centrality, (b) in-degree, (c) out-degree, (d) modularity

Figure 8 shows the inner core network visualised based on a force-directed layout from a variety of perspectives. Figure 8a visually encodes the nodes based on their closeness centrality, which indicates the average distance from a node to all other nodes in the network. As can be seen, the cluster of the nodes on the bottom right appears to have the highest values. These nodes are primarily connected (53 k retweets) to the network via only @dataeum, a collaborative and decentralised platform for data generation, based on blockchain and crowdsourcing. Much of the information shared via these nodes are hence related to cryptocurrency, decentralization, blockchain and crowdsourcing. These interactions help explain the relative distance of these nodes to the others in the network. Figure 8b visualises the same network, nodes colour-coded based on weighted in-degree values. The image shows that there are only a few nodes which have a high amount of inward communication (@mentions). Four of the highest of these users are @dataeum (53,009), @inaturalist (3118), @SciStarter (935) and @CitSciOZ (409). Beyond the discussions of dataeum, the other users are linked with a wide variety of topics, particularly related to citizen science. In exploring the hashtags (and discounting the obvious use of #citizenscience and #citsci) of the tweets mentioning inaturalist, SciStarter and CitSciOZ, we observe the following: #snail, #malaysia, #spider, #australia, #moth, #biodiversity, #mexico, #india, #hongkong, #slime, #frog, #brazil, #cobra, #eagle, #salamander, #taiwan, #hiring. Exploring these top hashtags, it appears the daily tweets by inaturalist generate quite a lot of public interest on ecological observations of animals found in different countries. For example,

‘‘RT @inaturalist: Our Observation of the Day is this *Rhinocochlis nasuta* #snail, seen in #Malaysia by danolsen! <https://t.co/F3Lph2jEVs> #malacology #mollusk #snails #nature #citizenscience #biodiversity’’

‘‘RT @inaturalist: iNat user rajibmaulick photographed a *Tachypompilus analis* #wasp dragging a paralyzed #spider in #India, and it's our Observation of the Day! <https://t.co/ezr3iHece2> #insects #citizenscience #nature #motherhood #hymenoptera #wildlife <https://t.co/rgg5MXgION>’’

Figure 8c presents the core of the network, visually encoding the measures of weighted out-degree tweets (higher number of tweets generated by these nodes are coloured darker). The figure shows

that several users have been actively tweeting. Four of the highest of these users are as follows: @Webiversity, @CrowdWeek, @SciStarter and @CitSciOZ. Tweets from Webiversity (236) appear automated and exhibit bot-behaviour, promoting crowdsource and crowdfund awards at regular patterns (Figure 9 compares the automated tweets with @Crowdweek’s tweeting patterns).

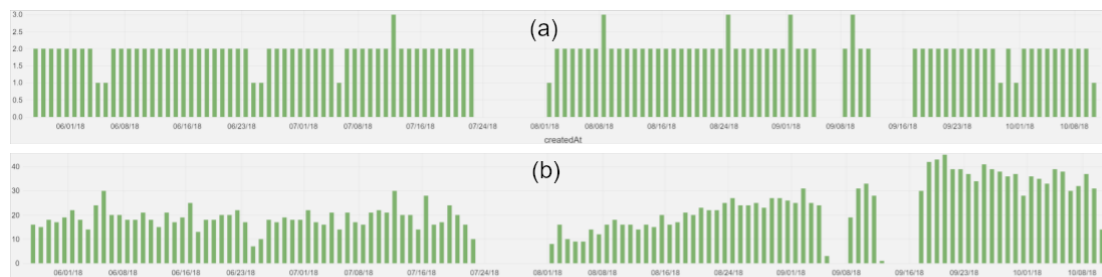


Figure 9. Comparing bot-behaviour observed within the Twitter dataset with typical manual behaviour. (a) Webiversity, (b) Crowdweek.

Crowdweek’s tweets appear very relevant to crowdsourcing topics, primarily focussed around the CSWGlobal18 (Crowdsourcing Week Global 2018 event, themed at decentralization through crowdsourcing), capturing topics such as innovation, decentralization, crowdfunding, natureofwork, blockchain, crypto, remotework. The tweets posted by SciStarter (509) and CitSciOz (504) on the other hand primarily discuss topics around citizen science events (e.g., ecsa2018, science week, Eureka prizes, city nature challenge, pollinator week, shark week, etc.). Figure 8d presents the network, visually encoded by modularity scores, to indicate the capacity of a network to be divided into modules. Structurally, the figure shows that the dense network, based on its inherent properties can be further isolated into smaller modular elements, indicating the appearance of sub-communities. Figure 6a further indicates that there are a large number of individual users who have discussed relevant topics with small number of users and hence, are presented in the periphery of the network. Several observations can be made—a large number of small groups are formed, comprising of 2 or 3 users upto 15 or 20. Most of these users are isolated from the wider network, without much interaction, although discussing topics relevant to crowdsourcing and citizen science.

Figure 10 presents a more detailed (zoomed-in) view of the heavy core in the network, where nodes (users) are well-connected with each other via common contacts. As such, there exists key influencers with high number of outgoing and incoming links which, when weighted with the amount of connections (weighted degree) position themselves within the core as key influencers within the community. Users such as dataeum, Webiversity, CrowdWeek, SciStarter, CitSciOZ, inaturalist, OpenLitterMap, FedCitSci, CoopSciScoop, WeObserveEU, CrowdPrecision are some examples of such social media accounts. Such accounts are primarily organisations, research projects and citizen science portals.

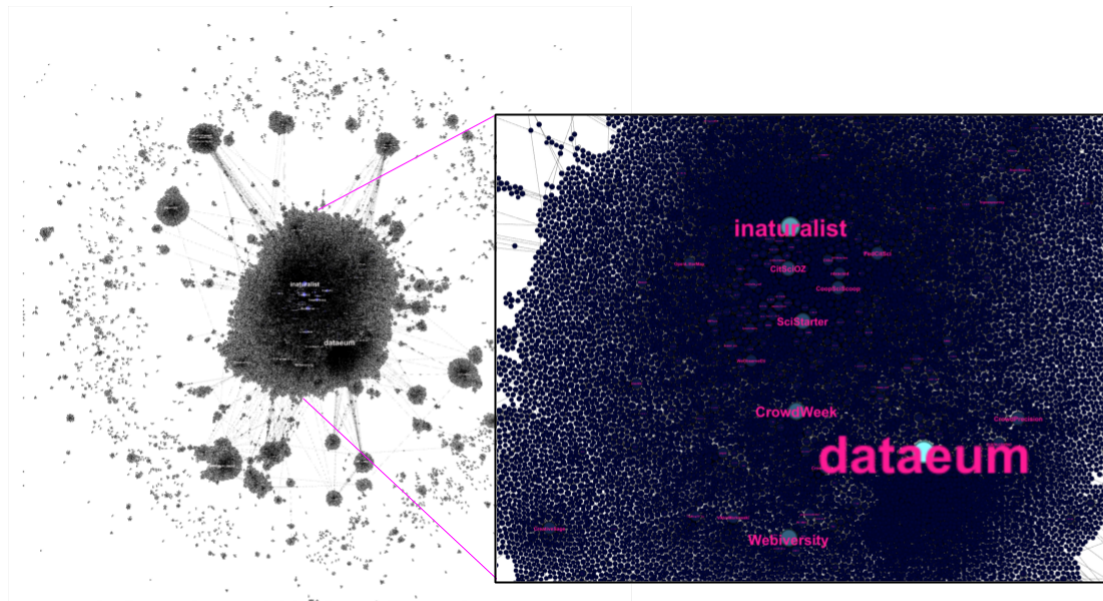


Figure 10. Visualising a more zoomed-in view of the inner core showing the strong network connections with several organisations acting as key influencers.

7. Understanding Topical Discussions

Using keywords and hashtags as a means to understand social media messages is helpful while considering a smaller social media dataset with a limited number of tweets that use standardised hashtags and keywords. However, for larger datasets as the one collected and very generic topics such as citizen science and crowdsourcing, it is challenging to employ a manual process. Instead, it was decided to use a topic model approach to automatically interpret the topics of discussions. Topic modelling is an unsupervised learning technique that aims to identify patterns and relationships among text documents [62,63], and a widely adopted technique used within a variety of domains such as bioinformatics [64], politics [65], transportation [66], etc. For this research, we have used Latent Dirichlet Allocation (LDA) as topic modelling method [67] for understanding topics of discussion. Techniques offer the possibility of scaling up analyses to hundreds of thousands of tweets.

The process involved initially taking the entire corpus and cleaning the text to remove unnecessary elements such as stopwords (NLTK's (<https://www.nltk.org/>) English stopwords corpus) and standard punctuations. Furthermore, the process required the removal of specific keywords and hashtags that were too frequent within the corpus to be distributed in most topics—e.g., 'RT', 'rt', '#crowdsourcing', 'crowdsource', '#crowdsource', 'crowdsourcing', 'citizenscience' and '#citizenscience'. Upon cleaning the tweets, the tweets were split into words, to be created into a dictionary. The process then created a term frequency inverse document frequency (TFIDF) model based on the corpus, which was further used to create an LDA model based on 7 topics. The number of topics were based on initial manual exploration of the tweets and several experiments to identify a good fit for the number of topics. The final model was then visualised using pyLDAvis which is a Python based implementation of LDA visualisation [68]. Figure 11 presents a visualisation of the distances between the 7 topics identified across the two principle component axes. The barchart on the right indicates the most salient terms across the entire dataset, where saliency indicates the distribution of topics weighted by the terms overall frequency [68]. Clicking on the topic bubbles on the left provides updated visualisations of the terms relevant to the respective topics. While many of the salient terms are fairly generic and common (e.g., alternative, international, fold, voted, etc.), analysing uncommon terms provides an indication of the different themes that emerge from the analysis. Salient

terms for each topic were identified and manually queried on the dataset to understand what were the primary discussion themes.

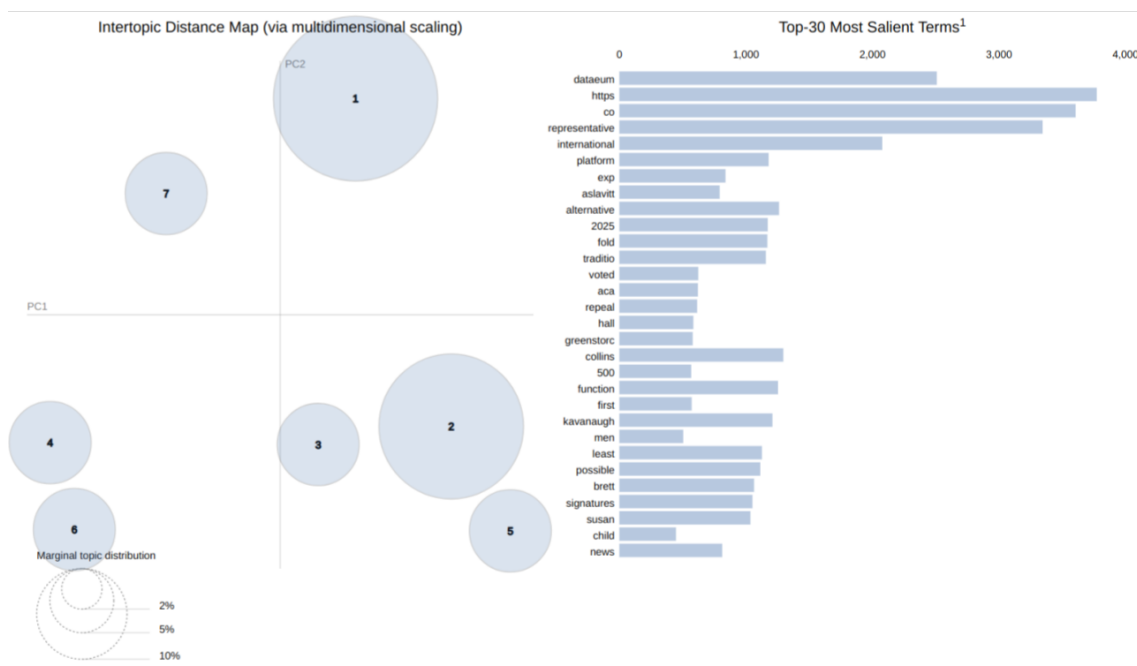


Figure 11. Visualisation of topic modelling using Latent Dirichlet Allocation (LDA) showing the 7 emerging topics on an intertopic distance map (left), with the top 30 most salient terms (right).

Exploration of the topics show the following distribution of the key terms, together with a summary of the themes emerging (terms in italics were considered for manual exploration):

Topic 1 keywords: *'function', 'kavanaugh', 'least', 'collins', 'brett', 'signatures', 'possible', 'supplies', 'susan', 'national', 'trying', 'faq'*.

Themes: politics (news on crowdsourcing funds for political opponent), promotion of citizen science platform on a national level, frequently asked questions (FAQ).

Topic 2 keywords: *'alternative', '2025', 'fold', 'tradition', 'uses', 'representative', 'sc', 'evening', 'newyorkstateag', 'watch', 'bird'*.

Themes: news on (i) a child crowdsourcing money for medical care, estimations of crowdsourcing to grow 20 fold by 2025 as an alternative to traditional labour market; (ii) voters for appealing the Affordable Care Act; (iii) an investment award to develop a service to support workers in reporting labour violations; (iv) a felony indictment of a global crowdsourcing company's CEO; (v) an elementary school located in radioactive and chemically contaminated land.

Topic 3 keywords: *'greenstorc', 'care', 'registered', 'token', 'das33', 'group', 'every', 'trial', 'ico', 'cryptocurrency', 'dascoinofficial'*.

Themes: discussions on a crowdsourcing platform on green energy solutions supported by blockchain and crypto, news on a new cryptocurrency crowdsourcing real estate development platform.

Topic 4 keywords: *'town', 'pay', '5k', 'remote', 'day', 'comes', 'busy', 'told', 'sending', 'several', 'crazy', 'helping', 'pre', 'inaturalist'*.

Themes: citizen science observations of animals and plants in various locations, crowdsourcing for funds to support university education, discussion on (i) crowdsourcing to afford medical care and supplies; (ii) towns and town-centres.

Topic 5 keywords: *'aslavitt', 'aca', 'hall', '500', 'men', 'child', 'sentenced', 'repeal', 'voted', 'conspiracy', 'series', 'followers', 'exploitation'*.

Themes: political discussions on protecting healthcare for people with pre-existing conditions, political discussions on repealing the Affordable Care Act, news on seven men sentenced for their roles in an international child exploitation crowdsourcing conspiracy

Topic 6 keywords: ‘*dataeum*’, ‘*cswglobal18*’, ‘*october*’, ‘*seven*’, ‘*decentralization*’, ‘*talked*’, ‘*got*’, ‘*medium*’, ‘*growth*’, ‘*washington*’.

Themes: promotion of a blockchain-based platform using crowdsourcing, variety of generic discussions on the impact of crowdsourcing and innovation, promotion of various events organised in October, generic discussions on crowdsourcing and decentralization, political discussions and related events in Washington, news about Washington Post crowdsourcing the location of migrant children separated from their parents.

Topic 7: ‘*look*’, ‘*best*’, ‘*scale*’, ‘*final*’, ‘*100k*’, ‘*many*’, ‘*give*’, ‘*another*’, ‘*take*’, ‘*city*’, ‘*vote*’, ‘*solution*’, ‘*manila*’, ‘*voter*’, ‘*supporter*’, ‘*1k*’.

Themes: crowdsourcing for funds to support university education in 5k, 2k, 1k, 500 denominations, generic information requests related to Manila, requests to identify registered voters in Manila City, information on FAQ sessions to supporters.

8. Discussions

Public participation in science and scientific topics has seen a significant increase in recent years, thanks to the increasing use of social media [69], with platforms such as Twitter and Facebook serving an important role in dissemination of scientific information [70]. While high rates of adoption of social media offer enormous opportunities for a wide spectrum of users (e.g., content creators, local businesses, etc.), a high number of people now consume social media as their source of news [71] (Statista, as of February 2020, <https://www.statista.com/statistics/718019/social-media-news-source/>). Whilst the opportunities offered by social media in dissemination of scientific discourse, particularly in reaching across large user bases are significant, in recent years, social media has observed immense challenges. Politicised narratives around ‘fake news’, misinformation, disinformation and the potential of social network sites in influencing public opinions, and even election outcomes have raised deeper questions about the role of social network sites. The 2016 US elections [72,73], Brexit [73] for example, had demonstrated the potential of social network sites in polarising large populations. The nature of social media, where users follow topics by engaging with specific conversations and following individuals or sources, leading to consciously/subconsciously re-inforcing beliefs through “filter bubbles” [74] and “echo chambers” [75] brings further interesting questions on how can users make more informed decisions based on a diverse range of viewpoints. At the time of writing this paper, the recent COVID-19 pandemic and 2020 US elections have also observed the active role social network sites like Facebook (<https://www.facebook.com/formedia/blog/working-to-stop-misinformation-and-false-news>) and Twitter (https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html) have played in suppressing or flagging misinformation [76], albeit with varying levels of success and criticisms (Politico’s analysis of Facebook content indicates the spread of misinformation despite Facebook’s efforts in countering false information, <https://www.politico.eu/article/facebook-misinformation-fake-news-coronavirus-covid19/>).

While our research focusses on discussions observed on Twitter around citizen science topics, it is worthwhile to recognise the challenges around social media and the increasing relevance of discussions around misinformation, fake news and polarisation. At the same time, citizen science as a field is not impervious to “fake science” and “infiltration” to support agendas [77]. Although the context of misinformation is beyond the scope of this paper, we note the critical role citizen scientists and citizen science can play a significant role in countering the spread of misinformation [78]. As such, understanding discussions and narratives around citizen science in social media is a critical first step. In this paper, we have presented an overview of the variety of discussions that emerge on social media within the citizen science community. Using a variety of methods (exploratory analysis, social network analysis and topic modelling) on a large scale social dataset, we have analysed what content is shared

on social media and how social media communities are structured around topics related to citizen science. In particular, we answer our research questions as follows:

RQ1. What are the predominant topical discussions around the citizen science microblogging communities?

From the study conducted in Sections 4 and 6, we have observed a wide variety of topics that fall within the scope of citizen science. Wider topics such as blockchain, cryptocurrency, decentralisation indicate many conversations around emerging technologies. We have also encountered much discussion around current news, politics, citizen science activities, events as well as crowdfunding. We summarise our findings as a way of addressing the RQ1 as our observations of the citizen science community discussions largely around the following themes:

- Discussions on citizen science projects, platforms, organisations, personal appeals and courses.
- Citizens and organisations sharing (ecological) observations.
- Sharing news and current affairs on politics and public policy.
- Sharing examples where crowdsourcing has made an impact.
- Discussions on emerging topics such as bitcoin, blockchain, decentralization, data.

We also note that there are several studies where citizens are engaged on social media to share their observations around different topics [16]. We also note that the role of social media as an information and news dissemination medium is well-established, and a large segment of different populations rely on social media as a source for their daily news [71]. Our observations showcasing how social media is used for storytelling and share positive examples of applications of citizen science and crowdsourcing could also help capture new audiences and increase engagement [79]. While topical analysis of social media content can help identify emerging themes and topics [80] (a common application of research being in event detection [81]), it was interesting to observe the citizen science community discuss topics that are highly topical. Our observations also indicate the use of social media as a mechanism for broadcasting and advertising projects, platforms, personal appeals and so on [82].

RQ2. What networks facilitate the spread of citizen science information within microblogging communities?

Visualising the network in graph layouts helped us observe three types of communities that had emerged:

- Large number of users connected to a handful of users via few connections discussing topics in isolation;
- large number of larger clusters that appear in isolation and not engaging with other clusters;
- a heavy core with very strong internal connections.

As our study is the first of its nature on citizen science conversations on Twitter, we believe it is interesting to observe different networks created as a result of communication exchange in conversations. The role of influencers in broadcasting information to large numbers of users, thereby increasing engagement with wider networks is already established [83]. Within our study, we noticed this role being played by organisations, research projects and citizen science portals. Several networks are built around retweets that support the promotion of a central idea. While this insight into the different types of networks may not be that surprising given the nature of social media, it is important to note that there are a number of users who constitute a very strong central network. Such users can be immensely valuable in sharing citizen science results, eventually helping engage with a wide range of audiences [84].

RQ3. What behaviours do microblogging citizen science communities exhibit?

We have also observed different types of behaviours when citizens interact with citizen science and crowdsourcing topics on Twitter. A majority of the discussions centred around retweeting and posting additional comments, replying or adding supporting statements to original tweets. Our observations also indicate users often acting primarily as broadcasters or receivers, with some few organisations engaging in broadcasting and receiving high number of tweets. We observe these users as influencers, who help relay information to wider audiences. Much of the citizen science related discussions are spread via these influencers, who have strong networks of users following them. Interestingly,

we have also observed bot-behaviour where Tweets seem automatically generated and disseminated to networks.

In addition to studying themes and user behaviour, a broader exploration of the data identified several purposes why users engaged on Twitter within the topic of citizen science and crowdsourcing. While we have not encountered a study specific to the field of citizen science, we note other studies who have identified these behaviours, albeit within other contexts:

- **Information/Resource seeking**—we observed examples which highlighted crowdsourcing in action—where users share requests for information regarding specific areas, looking for recommendations, suggestions or relevant people. Other studies have also highlighted information and resource seeking as a motivator for users to link up on Twitter [85]. Twitter was also observed to be used as a platform for raising funds [86] to fulfil personal, social or financial ambitions.
- **Promotions and Advertisements**—the use of Twitter as a way to promote destinations [87], organisations [88], software platforms, individuals or even citizen science platforms was interesting. Particularly interesting was the sharing of information that related with topics of public interest such as politics or societal issues.
- **Dissemination of news or personal opinions**—a large number users on Twitter were observed to share current topics and news items, particularly ones where crowdsourcing was used/being used to address issues of public interest. It was also interesting to see technologies (e.g., blockchain, decentralized networks, data platforms) being widely discussed and shared [89].
- **Broadcasting/Livecasting**—some users (particularly, influencers) were observed to be broadcasting information such as ecological/natural observations submitted by users. Given that the data collection was seeded around a public event, with some other events being picked up, it was observed that several users were live tweeting [90] about public events they would be attending/engaging with.
- **Engagement with members of the public**—some users were observed to host live sessions to engage with the public to provide expert analysis or answers to technical topics [91]. This was an interesting use of Twitter to help engage with a large number of users who may want to have a better understanding of some topics.

RQ4. How can citizen science project owners leverage this understanding to increase their presence on microblogging platforms?

While this paper does not attempt to cover the depth of the material already available on social media engagement strategies and best practices [21] (<https://www.deirdrebreakenridge.com/the-science-of-the-tweet-the-dos-and-donts-for-scicomm-on-twitter/>), we believe the observations can help stakeholders (primarily project owners and managers) understand how to use social media, particularly within the context of citizen science and crowdsourcing. We also believe that these recommendations could be applicable to other domains and contexts.

- **Identify networks of support and influencers to connect with:** it could be helpful to understand who are the users who are influencers in the citizen science domain, particularly within the context of the domain of study. Developing strong connections with other networks and influencers can project owners connect with other strong communities who could help disseminate information to wider audiences. This necessitates the need for strong collaborations and communities of practice [92] and the seven principles of cultivating a community of practice proposed by [21] are a helpful step in this direction.
- **Develop a strong network of followers and a good understanding of followers:** while having a large number of followers is helpful, it could be more helpful to understand the needs of the followers. This could help connect with larger audiences and help facilitate dissemination of information, particularly when this aligns with the needs of followers. It may also be helpful to align with other topics that are of public interest and showcase the value of the information within the context of wider societal issues.

- **Facilitate sharing of information that are valuable to wide audiences:** sharing and helping the dissemination of information from other networks can be a very useful way to develop strong networks. Project owners could potentially reach other networks, while at the same time, help increase awareness and contribute to reaching wider audiences.
- **Conduct interactive sessions with the public:** providing pre-scheduled access to experts on relevant topics of public interest could be an interesting way to help connect with members of the public. At the same time, such exercises could help in dissemination activities and help share the findings of projects. Twitter chat sessions can be helpful to connect with users who are interested about learning specific topics [93].
- **Interactive online discussions during events:** It was interesting to observe the active use of Twitter during live events, sharing news and information of sessions as they progress. Twitter use appeared high during such events and they could be a good opportunity to engage with larger audiences.

9. Conclusions and Future Work

In summary, our research explored how citizen science topics are discussed by the general public on social media. A wide variety of topics were observed, ranging from news and current events to emerging topics like blockchain and decentralisation. We also explored the different types of networks that are created as a result of social media conversations, ranging from large topical clusters to small networks of handful of users. Our analysis also observed a range of social media users who are more active in sharing discussions and broadcasting information to their wider communities. We also observed the differing behaviours of social media users in sharing news or seeking information, broadcasting or even running public engagement sessions.

Whilst this analysis provides an insight into the long-term discussions on social media, it is important to note the limitations in the research. An obvious one is that Twitter, as a microblogging platform is only one of the ways the citizen science community communicates—a future qualitative study is planned which will explore other platforms like Facebook, Pinterest and Instagram. Another limitation is that given the large volume of data to be analysed, statistical approaches such as categorisation, topic modelling, hashtag and keyword analysis are helpful. However, a deeper manual process perhaps together with random sampling of tweets could provide complimentary ways of analysing the tweets. We will explore this approach as a part of future work. It is also important to note that our analysis did not look at comparing the reach/impact of individual tweets as it would be beyond the scope of our study—instead, we studied conversations as a whole and explored networks and the different themes emerging from the discussions. As a part of future work, we will also look at understanding how quantitative metrics of individual Tweets and Twitter users impact on the importance of topics. The study provides an insight into social media use over a long term, however, one limitation could be investigating the use of social media over a much longer duration—perhaps spanning multiple years. We would also like to explore how citizen science communities engage with and deal with increasing volumes of misinformation and fake news. This will be a part of a future study and could provide further insights into the evolution of online citizen science communities.

Author Contributions: Writing—original draft, S.M. and D.T.; Writing—review & editing, D.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The social media data collection and initial analysis of the collected data was conducted while the lead author was based in the Department of Computing, Sheffield Hallam University. The lead author wishes to acknowledge the support provided by Sheffield Hallam University, including computing resources provided by research support. The author would also like to acknowledge the advice received from colleagues at the Information School in the University of Sheffield, Ziqi Zhang, Paul Clough and Nigel Ford.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANON	(Anonymised)
RQ	Research Question
FAQ	Frequently Asked Question
NLTK	Natural Language Toolkit
LDA	Latent Dirichlet Allocation
TFIDF	Term Frequency Inverse Document Frequency
ECSA	European Citizen Science Association

References

- Bhattacharjee, Y. Citizen scientists supplement work of Cornell researchers. *Science* **2005**, *308*, 1402–1404. [[CrossRef](#)] [[PubMed](#)]
- Hansen, D.L.; Shneiderman, B.; Smith, M.A. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*; Morgan Kaufmann: Burlington, MA, USA, 2010.
- Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*; Yale University Press: London, UK, 2006.
- Kaplan, A.M.; Haenlein, M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horizons* **2010**, *53*, 59–68. [[CrossRef](#)]
- Tess, P.A. The role of social media in higher education classes (real and virtual)—A literature review. *Comput. Hum. Behav.* **2013**, *29*, A60–A68. [[CrossRef](#)]
- Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. *J. Comput.-Mediat. Commun.* **2007**, *13*, 210–230. [[CrossRef](#)]
- Ellison, N.B.; Boyd, D.M. Sociality through social network sites. In *The Oxford Handbook of Internet Studies*; Oxford University Press: Oxford, UK, 2013.
- Foth, M.; Forlano, L.; Satchell, C.; Gibbs, M. *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement*; MIT Press: Cambridge, MA, USA, 2011.
- Freeman, J. *Digital Civic Participation in Australian Local Governments: Everyday Practices and Opportunities for Engagement*; Springer: Cham, Switzerland, 2016; pp. 195–218.
- Middleton, S.E.; Middleton, L.; Modafferi, S. Real-time crisis mapping of natural disasters using social media. *IEEE Intell. Syst.* **2014**, *29*, 9–17. [[CrossRef](#)]
- Flew, T.; Mitchell, P.; Ben-Harush, O.; Vann, P. *Social Media Use in Events and Festivals—Two Australian Case Studies: Pilot Studies Report*; QUT: Brisbane, Australia, 2016.
- Gerbaudo, P. *Tweets and the Street: Social Media and Contemporary Activism*; Pluto Press: London, UK, 2012.
- Malleson, N.; Andresen, M.A. The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartogr. Geogr. Inf. Sci.* **2015**, *42*, 112–121. [[CrossRef](#)]
- Freihardt, J. Can Citizen Science using social media inform sanitation planning? *J. Environ. Manag.* **2020**, *259*, 110053. [[CrossRef](#)]
- Bex, R.T.; Lundgren, L.; Crippen, K.J. Scientific Twitter: The flow of paleontological communication across a topic network. *PLoS ONE* **2019**, *14*, e0219688. [[CrossRef](#)]
- Daume, S.; Galaz, V. “Anyone Know What Species This Is?”—Twitter Conversations as Embryonic Citizen Science Communities. *PLoS ONE* **2016**, *11*, e0151387. [[CrossRef](#)]
- Di Minin, E.; Tenkanen, H.; Toivonen, T. Prospects and challenges for social media data in conservation science. *Front. Environ. Sci.* **2015**, *3*, 63. [[CrossRef](#)]
- Wiggins, A.; Crowston, K. Surveying the citizen science landscape. *First Monday* **2015**, *20*, 1–5. [[CrossRef](#)]
- Wilson, J.; Musick, M. Who cares? Toward an integrated theory of volunteer work. *Am. Sociol. Rev.* **1997**, *62*, 694–713. [[CrossRef](#)]
- Silvertown, J. A new dawn for citizen science. *Trends Ecol. Evol.* **2009**, *24*, 467–471. [[CrossRef](#)] [[PubMed](#)]
- Liberatore, A.; Bowkett, E.; MacLeod, C.J.; Spurr, E.; Longnecker, N. Social media as a platform for a citizen science community of practice. *Citiz. Sci. Theory Pract.* **2018**, *3*, 3. [[CrossRef](#)]

22. Newman, G.; Wiggins, A.; Crall, A.; Graham, E.; Newman, S.; Crowston, K. The future of citizen science: emerging technologies and shifting paradigms. *Front. Ecol. Environ.* **2012**, *10*, 298–304. [[CrossRef](#)]
23. Tieskens, K.F.; Van Zanten, B.T.; Schulp, C.J.; Verburg, P.H. Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landsc. Urban Plan.* **2018**, *177*, 128–137. [[CrossRef](#)]
24. Tenkanen, H.; Di Minin, E.; Heikinheimo, V.; Hausmann, A.; Herbst, M.; Kajala, L.; Toivonen, T. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Sci. Rep.* **2017**, *7*, 17615. [[CrossRef](#)]
25. Zarei, F.; Nik-Bakht, M.; Hammad, A. Visualisation of Local Municipal Satisfaction by Twitter Data Analysis. In *CSCE Conference Proceedings*; Canadian Society for Civil Engineering: Greater Montreal, QC, Canada, 2019.
26. Richards, D.R.; Tunçer, B. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosyst. Serv.* **2018**, *31*, 318–325. [[CrossRef](#)]
27. Albayrak, M.D.; Gray-Roncal, W. Data Mining and Sentiment Analysis of Real-Time Twitter Messages for Monitoring and Predicting Events. In *Proceedings of the 2019 IEEE Integrated STEM Education Conference (ISEC)*, Princeton, NJ, USA, 16 March 2019.
28. Robson, C.; Hearst, M.; Kau, C.; Pierce, J. Comparing the use of social networking and traditional media channels for promoting citizen science. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, Las Vegas, NV, USA, 21–26 July 2013.
29. Ambrose-Oji, B.; van der Jagt, A.P.; O’Neil, S. Citizen Science: Social Media as a Supporting Tool, 2014. Available online: https://www.researchgate.net/publication/265887193_Citizen_Science_Social_Media_as_a_supporting_tool (accessed on 24 November 2020).
30. Wiggins, A.; Crowston, K. From conservation to crowdsourcing: A typology of citizen science. In *2011 44th Hawaii International Conference on System Sciences*; IEEE: Piscataway, NJ, USA 2011; pp. 1–10.
31. Nov, O.; Arazy, O.; Anderson, D. Dusting for science: Motivation and participation of digital citizen science volunteers. In *Proceedings of the 2011 iConference*; ACM: New York, NY, USA, 2011; pp. 68–74.
32. Volunteers, T.C. *Guide to Using Social Media for Citizen Science Projects*; The Conservation Volunteers: London, UK, 2015.
33. Bik, H.M.; Goldstein, M.C. An introduction to social media for scientists. *PLoS Biol.* **2013**, *11*, e1001535. [[CrossRef](#)]
34. Holohan, A.; Garg, A. Collaboration online: The example of distributed computing. *J. Comput.-Mediat. Commun.* **2005**, *10*, JCMC10415. [[CrossRef](#)]
35. Simpson, R.; Page, K.R.; De Roure, D. Zooniverse: Observing the world’s largest citizen science platform. In *Proceedings of the 23rd International Conference on World Wide Web*; ACM: New York, NY, USA, 2014; pp. 1049–1054.
36. Raddick, M.J.; Bracey, G.; Gay, P.L.; Lintott, C.J.; Murray, P.; Schawinski, K.; Szalay, A.S.; Vandenberg, J. Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv* **2009**, arXiv:0909.2925.
37. Graham, G.; Cox, J.; Simmons, B.; Lintott, C.; Masters, K.; Greenhill, A.; Holmes, K. How is Success Defined and Measured in Online Citizen Science: A Case Study of Zooniverse Projects, 2015. Available online: https://www.researchgate.net/publication/277349570_Defining_and_Measuring_Success_in_Online_Citizen_Science_A_Case_Study_of_Zooniverse_Projects (accessed on 24 November 2020).
38. Wasko, M.M.; Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Q.* **2005**, *29*, 35–57. [[CrossRef](#)]
39. Kim, Y.; Choi, S.M.; Sohn, D. Understanding Motives For Using Social Network Site. In *American Academy of Advertising. Conference. Proceedings (Online)*; American Academy of Advertising: San Mateo, CA, USA, 2008; p. 235.
40. Butler, B.S. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Inf. Syst. Res.* **2001**, *12*, 346–362. [[CrossRef](#)]
41. Schroer, J.; Hertel, G. Voluntary engagement in an open web-based encyclopedia: Wikipedians and why they do it. *Media Psychol.* **2009**, *12*, 96–120. [[CrossRef](#)]
42. Hertel, G.; Niedner, S.; Herrmann, S. Motivation of software developers in Open Source projects: An Internet-based survey of contributors to the Linux kernel. *Res. Policy* **2003**, *32*, 1159–1177. [[CrossRef](#)]

43. Rotman, D.; Preece, J.; Hammock, J.; Procita, K.; Hansen, D.; Parr, C.; Lewis, D.; Jacobs, D. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*; ACM: New York, NY, USA, 2012; pp. 217–226.
44. Whiting, A.; Williams, D. Why people use social media: A uses and gratifications approach. *Qual. Mark. Res. Int. J.* **2013**, *16*, 362–369. [[CrossRef](#)]
45. Nov, O.; Arazy, O.; Anderson, D. Scientists@ Home: What drives the quantity and quality of online citizen science participation? *PLoS ONE* **2014**, *9*, e90375. [[CrossRef](#)]
46. Centola, D. The spread of behavior in an online social network experiment. *Science* **2010**, *329*, 1194–1197. [[CrossRef](#)]
47. Myers, S.A.; Sharma, A.; Gupta, P.; Lin, J. Information network or social network?: The structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web*; ACM: New York, NY, USA, 2014; pp. 493–498.
48. Java, A.; Song, X.; Finin, T.; Tseng, B. 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*; ACM: New York, NY, USA, 2007; pp. 56–65.
49. Halberstam, Y.; Knight, B. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *J. Public Econ.* **2016**, *143*, 73–88. [[CrossRef](#)]
50. Weber, I.; Garimella, V.R.K.; Batayneh, A. Secular vs. islamist polarization in egypt on twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*; ACM: New York, NY, USA, 2013; pp. 290–297.
51. Tremayne, M. Anatomy of protest in the digital era: A network analysis of Twitter and Occupy Wall Street. *Soc. Mov. Stud.* **2014**, *13*, 110–126. [[CrossRef](#)]
52. Berest, M.; Gera, R.; Lukens, Z.; Martinez, N.; McCaleb, B. Predicting network evolution through temporal Twitter snapshots for paris attacks of 2015. In *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, Washington, DC, USA, 28 June – 1 July 2016.
53. Hobbs, M.; Della Bosca, H.; Schlosberg, D.; Sun, C. Turf wars: Using social media network analysis to examine the suspected astroturfing campaign for the Adani Carmichael Coal mine on Twitter. *J. Public Aff.* **2019**, *20*, e2057. [[CrossRef](#)]
54. Negara, E.S.; Triadi, D.; Andryani, R. Topic Modelling Twitter Data with Latent Dirichlet Allocation Method. In *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*; IEEE: Piscataway, NJ, USA, 2019; pp. 386–390.
55. Reuter, C.; Scholl, S. Technical Limitations for Designing Applications for Social Media. In *Mensch & Computer Workshopband*; Oldenbourg Wissenschaftsverlag: Munich, Germany, 2014; pp. 131–139.
56. Campan, A.; Atnafu, T.; Truta, T.M.; Nolan, J. Is Data Collection through Twitter Streaming API Useful for Academic Research? In *2018 IEEE International Conference on Big Data (Big Data)*; IEEE: Piscataway, NJ, USA, 2018; pp. 3638–3643.
57. Chen, J.; Hossain, M.S.; Zhang, H. Analyzing the sentiment correlation between regular tweets and retweets. *Soc. Netw. Anal. Min.* **2020**, *10*, 13. [[CrossRef](#)]
58. Xu, Z.; Yang, Q. Analyzing user retweet behavior on twitter. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*; IEEE: Piscataway, NJ, USA, 2012; pp. 46–50.
59. Jalali, N.Y.; Papatla, P. Composing tweets to increase retweets. *Int. J. Res. Mark.* **2019**, *36*, 647–668. [[CrossRef](#)]
60. Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **2014**, *9*, e98679. [[CrossRef](#)] [[PubMed](#)]
61. Noack, A. Modularity clustering is force-directed layout. *Phys. Rev. E* **2009**, *79*, 026102. [[CrossRef](#)]
62. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; Zhao, L. Latent Dirichlet Allocation (LDA) and Topic modeling: Models, applications, a survey. *Multimed. Tools Appl.* **2019**, *78*, 15169–15211. [[CrossRef](#)]
63. Bolelli, L.; Ertekin, Ş.; Giles, C.L. Topic and trend detection in text collections using latent dirichlet allocation. In *European Conference on Information Retrieval*; Springer: Berlin, Germany, 2009; pp. 776–780.
64. Liu, L.; Tang, L.; Dong, W.; Yao, S.; Zhou, W. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus* **2016**, *5*, 1608. [[CrossRef](#)]

65. Yoon, H.G.; Kim, H.; Kim, C.O.; Song, M. Opinion polarity detection in Twitter data combining shrinkage regression and topic modeling. *J. Inf.* **2016**, *10*, 634–644. [[CrossRef](#)]
66. Sun, L.; Yin, Y. Discovering themes and trends in transportation research using topic modeling. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 49–66. [[CrossRef](#)]
67. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
68. Sievert, C.; Shirley, K. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014; pp. 63–70.
69. Ryan, L. Social media and popularising space: Philae Lander (@ Philae2014) and the journey to comet 67P/Churyumov-Gerasimenko. *Space Policy* **2017**, *41*, 20–26. [[CrossRef](#)]
70. Tsubokura, M.; Onoue, Y.; Torii, H.A.; Suda, S.; Mori, K.; Nishikawa, Y.; Ozaki, A.; Uno, K. Twitter use in scientific communication revealed by visualization of information spreading by influencers within half a year after the Fukushima Daiichi nuclear power plant accident. *PLoS ONE* **2018**, *13*, e0203594. [[CrossRef](#)] [[PubMed](#)]
71. Westerman, D.; Spence, P.R.; Van Der Heide, B. Social media as information source: Recency of updates and credibility of information. *J. Comput.-Mediat. Commun.* **2014**, *19*, 171–183. [[CrossRef](#)]
72. Allcott, H.; Gentzkow, M. Social media and fake news in the 2016 election. *J. Econ. Perspect.* **2017**, *31*, 211–36. [[CrossRef](#)]
73. Maher, P.J.; Igou, E.R.; van Tilburg, W.A. Brexit, Trump, and the polarizing effect of disillusionment. *Soc. Psychol. Personal. Sci.* **2018**, *9*, 205–213. [[CrossRef](#)]
74. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You*; Penguin UK: London, UK, 2011.
75. Gillani, N.; Yuan, A.; Saveski, M.; Vosoughi, S.; Roy, D. Me, my echo chamber, and I: introspection on social media polarization. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 823–831.
76. Van der Linden, S.; Roozenbeek, J.; Compton, J. Inoculating Against Fake News About COVID-19. *Front. Psychol.* **2020**, *11*, 2928. [[CrossRef](#)]
77. Rasmussen, L.M. Confronting research misconduct in citizen science. *Citiz. Sci. Theory Pract.* **2019**, *4*, 10. [[CrossRef](#)]
78. Katapally, T.R. A Global Digital Citizen Science Policy to Tackle Pandemics Like COVID-19. *J. Med. Internet Res.* **2020**, *22*, e19357. [[CrossRef](#)]
79. Martinez-Conde, S.; Macknik, S.L. Opinion: Finding the plot in science storytelling in hopes of enhancing science communication. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8127–8129. [[CrossRef](#)]
80. Alkhodair, S.A.; Ding, S.H.; Fung, B.C.; Liu, J. Detecting breaking news rumors of emerging topics in social media. *Inf. Process. Manag.* **2020**, *57*, 102018. [[CrossRef](#)]
81. Aldhaferi, A.; Lee, J. Event detection on large social media using temporal analysis. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
82. Davis, E.; Caffrey, J.M.; Coughlan, N.E.; Dick, J.T.; Lucy, F.E. Communications, outreach and citizen science: Spreading the word about invasive alien species. *Manag. Biol. Invasions* **2018**, *9*, 415. [[CrossRef](#)]
83. Darnton, A.; Horne, J. Influencing Behaviours: Moving beyond the Individual: A User Guide to the ISM Tool, 2013. Available online: <https://www2.gov.scot/Resource/0042/00423436.pdf> (accessed on 24 November 2020).
84. MacLeod, C.J.; Scott, K. Mechanisms for Enhancing Public Engagement with Citizen Science Results, 2020. Available online: <https://besjournals.onlinelibrary.wiley.com/doi/10.1002/pan3.10152> (accessed on 24 November 2020).
85. Holton, A.E.; Baek, K.; Coddington, M.; Yaschur, C. Seeking and sharing: Motivations for linking on Twitter. *Commun. Res. Rep.* **2014**, *31*, 33–40. [[CrossRef](#)]
86. Qomfo, S.; Chilya, N.; Chuchu, T.; Maziriri, E.T.; Nodoro, T. Perceptions of the effectiveness of Twitter as a crowdfunding communication tool for raising university fees. *Communitas* **2019**, *24*, 1–17. [[CrossRef](#)]
87. Bonsón, E.; Perea, D.; Bednárová, M. Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities. *Gov. Inf. Q.* **2019**, *36*, 480–489. [[CrossRef](#)]
88. Langa, L.A. Does Twitter Help Museums Engage with Visitors? 2014. Available online: <https://www.ideals.illinois.edu/handle/2142/47289> (accessed on 24 November 2020).

89. Pai, P.; Arnott, D.C. User adoption of social networking sites: Eliciting uses and gratifications through a means–end approach. *Comput. Hum. Behav.* **2013**, *29*, 1039–1053. [[CrossRef](#)]
90. Shi, Z.; Rui, H.; Whinston, A.B. Content sharing in a social broadcasting environment: evidence from twitter. *MIS Q.* **2014**, *38*, 123–142. [[CrossRef](#)]
91. Smith, A. “Wow, I didn’t know that before; thank you”: How scientists use Twitter for public engagement. *J. Promot. Commun.* **2015**, *3*, 320–339.
92. Wenger, E.; McDermott, R.A.; Snyder, W. *Cultivating Communities of Practice: A Guide to Managing Knowledge*; Harvard Business Press: Brighton, MA, USA, 2002.
93. Gilbert, S. Learning in a Twitter-based community of practice: An exploration of knowledge exchange as a motivation for participation in# hcsma. *Inf. Commun. Soc.* **2016**, *19*, 1214–1232.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).