# NETWORKING AGGIE: BROADCASTING INFORMATION TO TOPICAL COMMUNITIES

Marco T. Bastos
University of California, Davis

Mark Lubell
University of California, Davis

## ABSTRACT

Modern agricultural systems are experiencing a revolution in how information is disseminated and exchanged among networks of outreach professionals, farmers, consumers, and community stakeholders. The traditional approach to agricultural extension relied on a top-down, continuum model that went from university researchers to cooperative extension and finally to growers. With internet penetration rates rising in rural communities, stakeholders are increasingly experimenting with social media and agricultural information has been widely shared across local, national, and global networks. This paper addresses such networks by deploying a supervised snowball census of the agricultural social web to investigate the formation of communities dedicated to sharing agricultural information. We identified a cohort of 153 individuals responsible for the outreach initiatives at the University of California and mapped the first and second level network of followers connected to this community, thus rendering a population of 59K users that tweeted 250M tweets since signing up to Twitter. We processed the content of each tweet posted in 2014 using multiple classifiers and determined that the boundaries of the network comprise 32K nodes and 4M edges. The resulting graph shows that this community is clustered in areas of agricultural expertise with limited overlap across cliques. We also found increasing patterns of core-periphery dynamics associated with the level of expertise attached to each topical subnet. The paper concludes by discussing related literature and the policy implications of our research.

## 1. Introduction

Modern agricultural systems are experiencing a revolution in how knowledge is disseminated and exchanged among networks of outreach professionals, farmers, consumers, and community stakeholders. The traditional approach to agricultural extension relied on a top-down, continuum model that went from university researchers to cooperative extension and finally to farmers. With internet penetration rates growing in rural communities, stakeholders are increasingly experimenting with social media and

other online forms of communication to share agricultural information across local, national, and global networks. Yet, systematic scholarship investigating the dynamics of online networks to the agriculture and food sectors, and the potential effects of social media at spreading expert knowledge, have remained largely unexplored.

Social media affordances have been remarkably efficient in connecting nationwide and global chains of information diffusion to otherwise local activities (Waldman, 2011). This transition was particularly salient in topics and problems relying on a small number of specialists who engage a large number of highly diverse and continuously expanding body of potential stakeholders. Recent studies have shown that one way to disseminate knowledge efficiently across broader communities is through specialized knowledge networks (Bidwell et al., 2013, Kalafatis et al., 2015). These networks are primarily knowledge extension and include organizations, actors, and communication infrastructure through which information is disseminated, often transforming or supporting decision-making policies that impact local and national communities.

Within our field of inquire, knowledge networks allow for specialized information to travel within and among communities of interest (Henri and Pudelko, 2003)—that is, from Twitter accounts managed by government agencies to Twitter accounts of outreach individuals and finally to the locally grounded community of users. Agricultural extension builds up on such communities and capitalizes on the network structure of local agricultural knowledge systems comprising distributed actors with a diversity of specializations and expertise (Lubell et al., 2014). To this end, this paper contributes to the literature on specialized knowledge networks by exploring how a network of Twitter users invested in the dissemination of agricultural information provide higher exposure to specialized knowledge.

In the following we report on an investigation of a community of users dedicated to the dissemination of agriculture information in California and beyond. The seed users comprise 153 Twitter accounts associated with the Agriculture and Natural Resources of the University of California. We identified the followers and followees connected to the seed users and retrieved the timelines of each Twitter account, thus rendering a population of 59K users and 68M tweets. We designed two classifiers to process the content of each tweet posted in 2014 and determined the boundaries of the network. As Twitter interactions consist of @-mentions, a publicly-visible message targeting other accounts, and retweets, the action of rebroadcasting a message to the users' followers, we mined the timelines and profiles of each of the individuals in the population to identify instances of information sharing. In what follows, we seek to advance foregoing research by studying the communication patterns within this community.

## 2. Previous Work

Previous studies have investigated the use of social media in establishing a green virtual sphere by supporting access to environmental information and providing a space for debate. This literature connects the long-standing debate on the emergence of a public sphere in modern society (Calhoun, 1992, Habermas, 1991) with the potential changes associated with the use of social media platforms to engage the public in open discussions (Papacharissi, 2002, Bastos, 2011). Subsequent studies have explored the

elapsed effects of social media on the green public sphere (Torgerson, 2006, Torgerson, 1999, Yang and Calhoun, 2007), particularly the user of online social networks to air grievances, develop coordinated actions for natural or man-made disasters, and mediate partisan battles over environmental regulation, energy policy, and climate change.

The notion of a green public sphere was originally advanced by Torgerson (1999) and refers to a subsection of the public sphere focused on environmental issues. Following the early definition proposed by Habermas (1991), the green public sphere would form a space for discussion detached from individual or private interest and displaying an interest in plurality of opinions however inconvenient and troubling they can be. Yet, according to Torgerson (1999), the green public sphere is not without its boundaries, as meaningful disagreements require clear and agreed limits to move forward the discussion. According to Yang and Calhoun (2007), the emergence of a green public sphere resulted from users resorting to legacy and social media to voice their grievances about environmental issues which otherwise would not have been heard.

Within this field of inquire, Segerberg and Bennett (2011) analyzed the content of tweets associated with protests against the 2009 United Nations Climate Summit in Copenhagen and White et al. (2014) explored the motivations behind social media use during the Alberta oil sands. This body of scholarship reported that Twitter was used to access news, particularly alternative sources covering the unfolding events. White et al. (2014) reported that even though users resorted to Twitter to engage in debates over the Northern Gateway pipeline, interviewees underplayed the importance of Twitter in the formation of a green virtual sphere, as access to the space was not equal and the discussions most likely monitored. Similarly, Bortree and Seltzer (2009) examined communication strategies employed by environmental advocacy groups on Facebook and Cheong and Lee (2010) investigated how the use of Twitter in Australia was connected to the Earth Hour 2009 campaign.

We draw from this scholarship on social media and the green public sphere and hypothesize that the ensuing networks used to convey agriculture information follows a multi-step model of information diffusion centered on core users and surrounded by highly active brokers exporting local knowledge to other agricultural systems and the broader public. This framework has been applied in Twitter literature to describe a process of information diffusion that often deviates from patterns observed in social networks (Wu et al., 2011), as information exchange relies on hubs and authorities (Dubois and Gaffney, 2014) and the network topology behaves much like an information network with pronounced amplification effect for information dissemination (Myers et al., 2014). Broadly speaking this framework foregrounds the diffusion of information from elite towards ordinary users and is consistent with the two-step flow theory of communication originally proposed by Katz (1957).

## 3. Rationale

Due to technical requirements and stringent rules to retrieve Twitter data, scholarship on Twitter has often focused on a set of hashtagged tweets as a marker to unfolding events around the world. With the exception of a few macroscopic studies of Twitter network (Kwak et al., 2010, Gabielkov et al., 2014, Myers et al., 2014), the high computational and

financial costs involved in retrieving a complete set of tweets compelled researchers to focus on subsets of data revolving around events and/or contextual markers. As a result, Twitter scholarship has dealt with a particularly set of data mostly characterized by the phatic function of social media communication (Miller, 2015). This *de facto* standard approach to Twitter research is centered on messages motivated by ongoing events and often crafted with the potential for viral cascades.

While datasets retrieved using markers such as keywords and hashtags are useful to track the diffusion of information about a given topic, calls for awareness about political issues, and the coordination of social movements (Bastos and Mercea, 2015), this procedure of data collection emphasizes phatic messages against the dialogic exchanges between social media users. In fact, Twitter users often begin interacting within a hashtagged stream only to subsequently drop the hashtag as their one-to-one interaction unfolds. As a result, researchers mining Twitter Streaming API[1] are often left with data lacking interactions initiated with the hashtag but dropped afterwards. Another significant shortcoming of this approach is that data can only be retrieved prospectively, as access to historical Twitter data is prohibitively expensively.

As a result, social media research has involuntarily emphasized and scrutinized a set of messages purposely tailored to be phatic and designed for maximum effect, all the while ignoring the larger conversation in the background that is potentially connected to the hashtag stream. Moreover, the representation of a public sphere by means of an *ad-hoc* public deserves careful consideration as the *raison-d'etre* of hashtagged messages is to perform a one-to-many broadcast. By losing track of myriad interactions between Twitter users, and frequently relying on hashtags, social media scholarship has inspected a reduced version of Twitter's agora (Miller, 2015) that departs considerably from the engaged, content-driven, dialogic public sphere envisioned by Habermas (1987); a public sphere marked by the lively exchange carried out on a daily basis.

Data collection based on hashtags or other textual markers are thus inadequate for tracking the dynamics of communities over extended periods of time. The research design described in the following sections addresses these shortcomings by moving the focus from the streaming of evens to long and sustained interactions within a community of users. While historical Twitter is difficult and often expensive to retrieve, Twitter Rest API can be used to retrieve the timeline of individual users with no temporal restriction or extemporaneous costs beyond the computational resources required to cover a given population of users. By focusing on the organic interactions between users that evolved over several months or years, this research design can shed light on community formation, users' interaction on social media, and the structure and dynamics of communities of interest.

## 4. Objectives

The aims of this study are threefold. Firstly, we present a census-based, viable, cost-effective, and replicable research design to track the information exchange within a community of Twitter users. Secondly, we report on the affordances of social media to

---

[1] Twitter Streaming API documentation is available at https://dev.twitter.com/streaming/overview

outreach initiatives focused on the dissemination of agriculture-related information. Thirdly, we investigate whether the network connecting outreach centers, farmers, and government agencies follows a multi-step flow of communication that structurally displays core-periphery dynamics. We hypothesized that the network follows a two-step flow of communication with a core social network centered on a particular set of nodes surrounded by highly active brokers exporting local knowledge to other agricultural systems and the broader public. Specifically, we anticipated that the network might enforce different paths to disseminate agriculture and general information to the periphery of the graph. To this end, and considering the rationale for developing a "snowballing census" approach to internet communities reviewed hitherto, we pursue the following research objectives:

RO1.  Describe a replicable research design based on seed nodes and successive network levels by snowballing the follower-followee network;

RO2.  Report on the diffusion of agriculture information across network cliques and subcommunities;

RO3.  Identify whether the network follows a multi-step flow of communication structured with strong core-periphery dynamics.

## 5.  Research Design

Due to restrictions imposed by Twitter's REST API, data collection began with a relatively small subset of users listed in the UC Ag and UC ANR Twitter lists, from which we snowballed data collection to users they follow or are followed. All data collection occurred between May 5 and June 27, 2015 via the Twitter REST API, thus providing an estimate for the time required for this research approach. The ensuing dataset includes the complete set of messages posted by this population since signing up to Twitter, up to a limit of 3200. Therefore, the population studied in this paper includes the core 153 seed nodes extracted from the UC Twitter lists and their followers and followees, totaling 59,761 users. Figure 1 shows the resulting two levels of this network, from the 153 seed nodes (colored back) to the 59K users interconnected as follower and followees (colored blue).

There are differences between @-mentions and retweets that need to be accounted for separately. Although @-mentions are arguably more social, they are also used for drawing attention of celebrities, politicians, and media pundits and are often an instrument for sockpuppeting (Bastos et al., 2013). As we are interested in the diffusion of topical information within agriculture communities, we relied on @-mentions and retweets as relevant edges connecting Twitter accounts. In both cases we draw an edge connecting two accounts that have posted at least one message with agriculture-relevant content at some point in the year of 2014. Although the network object includes both @-mentions and retweets, the analyses reported in the next section were performed on @-mentions and retweets separately. Figure 2 details the process of data collection, processing, and analysis and informs RO1.
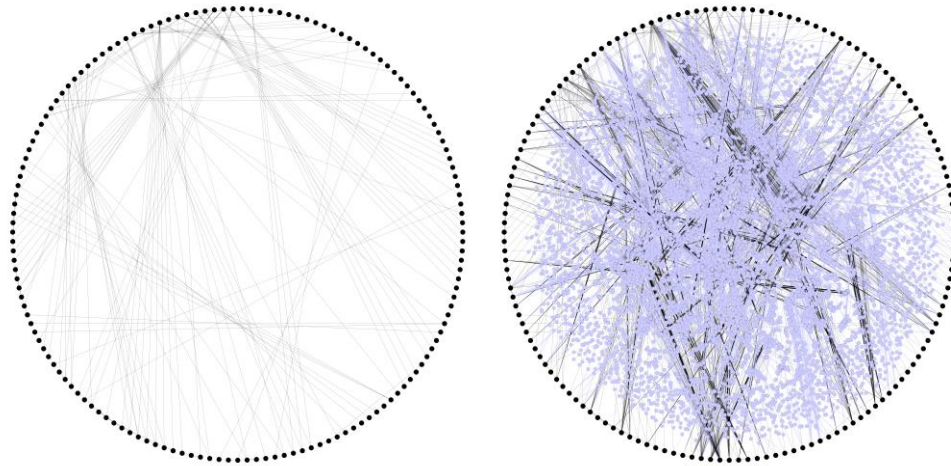
**Figure 1: Initial group of 153 seed users in black and extended network in blue (2nd level network)**

The complete set of messages tweeted by these users since joining Twitter totals 285M tweets (285,628,862). From the 59,761 accounts in the population we managed to retrieve the timelines of 54,422 accounts (91% of users in the target population). From the total of 5352 users left out of the network, we found that 501 accounts have yet to tweet a single message and 4894 accounts were protected or have been deactivated since data collection started. Finally, a total of 43 accounts were both protected and had not tweeted any message at the time of data collection.

In addition to these silent accounts, Twitter Rest API limits access to a maximum of 3200 statuses (tweets) per user. Our population of 59,761 users includes 13,112 accounts that tweeted over this limit. This technical limitation imposes considerable challenges to retrieving the complete set of tweets of a population over extended periods of time. As a result, we managed to retrieve only 65M (65,294,710) tweets from 54,422 users, as opposed to a potential total of 285M. Moreover, the temporal series retrieve for this population is likely to vary considerably as only a portion of the timelines retrieved was objected to Twitter restriction of 3200 tweets per user.

We addressed this problem by identifying the average cut-off date for users that posted over 3199 tweets and removed messages posted prior to this date. As a result, we only analyzed messages tweeted by the target population after this point in time. We resorted to this procedure to filter the 65M tweets collected from Twitter API and ended up with a total of 43M messages. As our research is focused on information exchange, we subsequently removed messages that were not retweets nor included @-mentions to other users in the target population. The procedure further reduced the dataset to 26M messages posted between August 1, 2013 and May 15, 2015, thus covering a period of roughly 2 years (652 days) with conspicuous dips corresponding to summer vacations and regular oscillations associated with monthly cycles of activity.
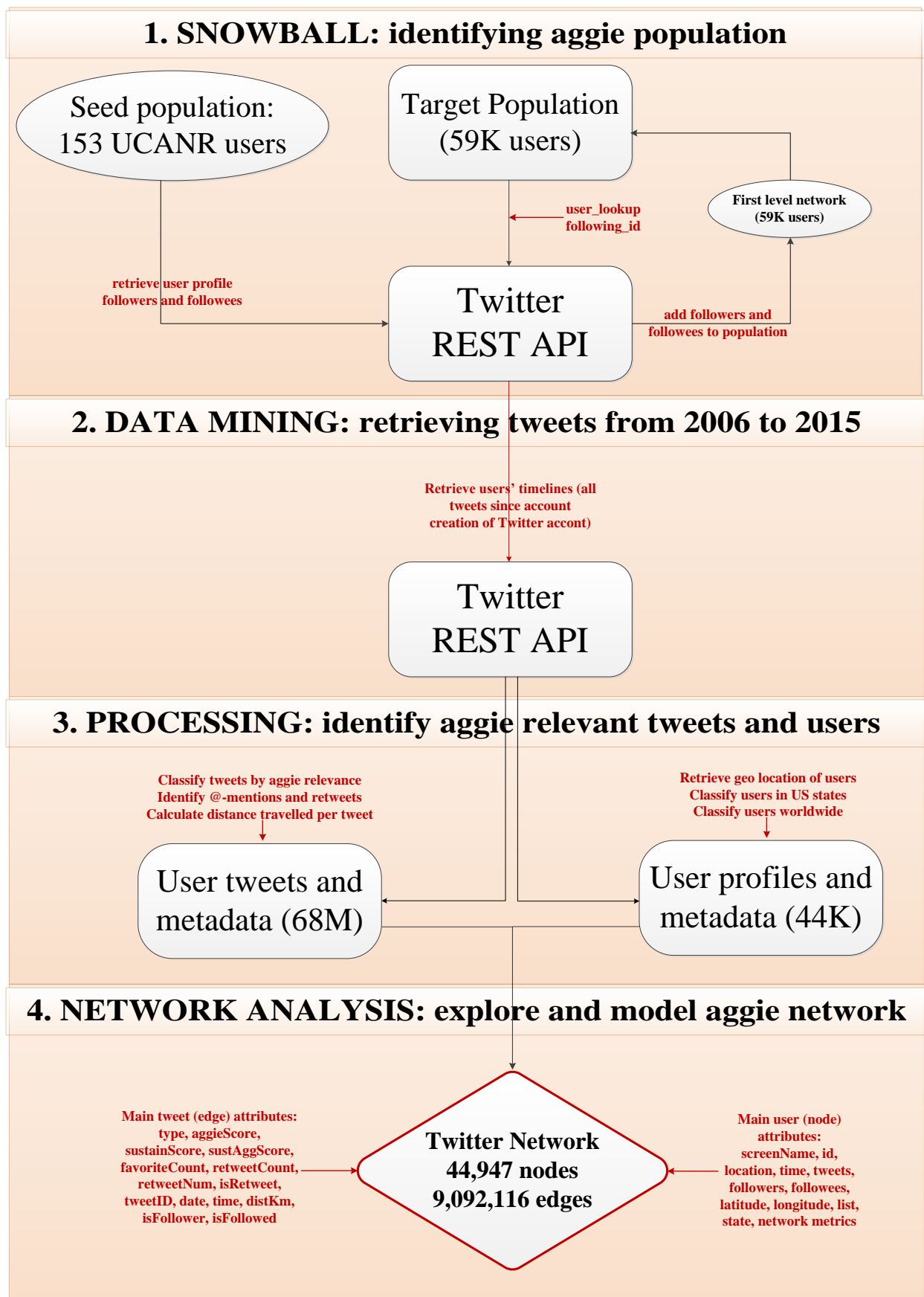
# 1. SNOWBALL: identifying aggie population

Seed population: 153 UCANR users

Target Population (59K users)

user_lookup
following_id

First level network (59K users)

retrieve user profile
followers and followees

Twitter REST API

add followers and
followees to population

# 2. DATA MINING: retrieving tweets from 2006 to 2015

Retrieve users' timelines (all tweets since account creation of Twitter accont)

Twitter REST API

# 3. PROCESSING: identify aggie relevant tweets and users

Classify tweets by aggie relevance
Identify @-mentions and retweets
Calculate distance travelled per tweet

Retrieve geo location of users
Classify users in US states
Classify users worldwide

User tweets and metadata (68M)

User profiles and metadata (44K)

# 4. NETWORK ANALYSIS: explore and model aggie network

Main tweet (edge) attributes:
type, aggieScore,
sustainScore, sustAggScore,
favoriteCount, retweetCount,
retweetNum, isRetweet,
tweetID, date, time, distKm,
isFollower, isFollowed

Twitter Network
44,947 nodes
9,092,116 edges

Main user (node) attributes:
screenName, id,
location, time, tweets,
followers, followees,
latitude, longitude, list,
state, network metrics

**Figure 2: Data mining workflow and resulting network graph**

Figure 3a shows a histogram of messages binned by month, with a cut-off date around mid-2003 that includes the complete set of messages tweeted both by filtered users (>3200 tweets) and unfiltered users (<3200 tweets). Although this period of 652 days includes a comprehensive set of tweets posted by this community, we found inconsistencies in the temporal distribution of tweets. Kernel estimation shows that the data drops off artificially at the upper end of the time series. This is likely a result of user's timelines being collected sequentially and thus at different points in time. In addition to that, we anticipated that older messages would fare relatively better in terms of retweets compared to newer messages, as they would have benefitted from a larger period of time to spread throughout the network.
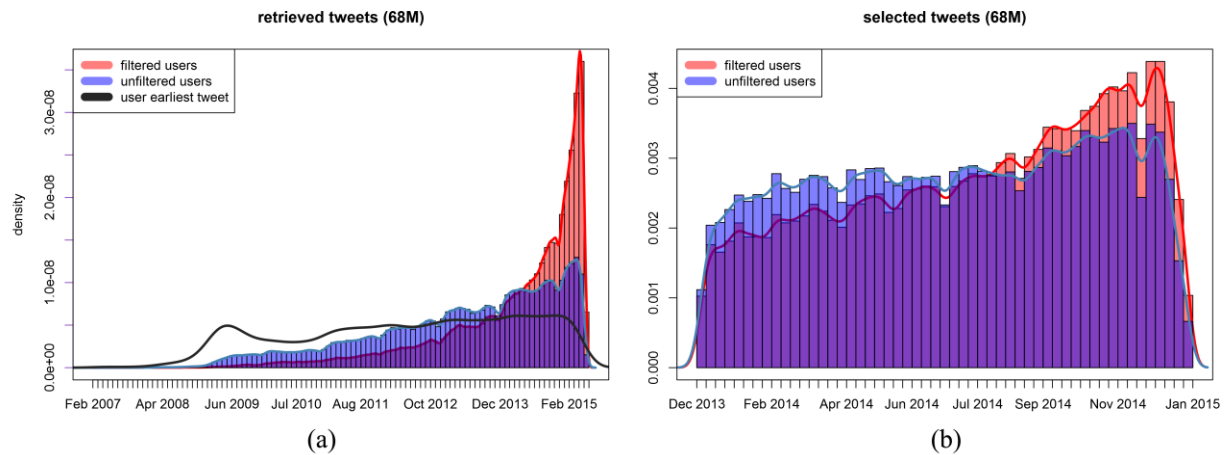


Figure 3: Complete set of messages retrieved binned by month (a) and sampled data binned by week (b)

We addressed these issues by selecting an intermediate period of time that was unaffected by any variation resulting from data collection. This period of time comprehends the entire year of 2014 and includes a total of 3.7M (3,691,342) tweets. Therefore, our resulting dataset includes messages posted in 2014 and the analysis reported in this paper refers to this subset of retrieved tweets. The frequency of tweets binned by week is shown in Figure 3b, with similar frequency distributions across filtered and unfiltered users. We expect these procedures to have addressed the restrictions imposed by Twitter REST API and to have provided a comprehensive set of messages posted by our population in 2014.

In summary, our research design draws from Liang and Fu (2015) by identifying a representative set of Twitter users (egos), collecting all the egos' alters (i.e., followers and followees), and the following relationships among the alters. Finally, we obtained the profiles and the timelines of the selected users (egos and alters) and processed the data to generated a graph with various edge and node properties. We expect this approach based on sampling of users, rather than sampling of tweets, to provide a more reliable and appropriate approach to analyzing individual and community-level behaviors.

## 6. Data Analysis

For the purposes of this investigation we trained two classifiers to identify messages dedicated to agriculture and messages associated with sustainability issues. Tweets often include URL links pointing to the actual content under discussion and without which is not possible to determine the topic addressed by each tweet. To account for this supplementary body of text, we retrieved the webpage title of each URL in the dataset and ran the classifiers over the combined corpus of tweet and webpage title (when available). The agriculture classifier is based on a set of 37 items, while the sustainability classifier relies on a set of 30 keywords, bigrams, and tokens.

Each classifier returns a score based on the concentration of such terms, bigrams, keywords, and tokens relative to the number of words in the tweet or the number of words in the tweet plus the webpage title (when URL link was available). We ran the classifiers on set of 9,627,146 tweets and found that only 12.09% (1,164,014 tweets) of the dataset scored above zero for agriculture content and only 5.48% (527,167 tweets) for sustainability content. The average score was .009 and .005 for agriculture and sustainability messages, respectively.

We relied on the scores calculated by the classifiers (aggieScore and sustainScore, henceforth) to process the text and identify relevant messages. Users that failed to post any agriculture or sustainability related message during the entire period were further removed from the data. In other words, the final dataset is restricted to tweets posted in 2014 and it does not include users that did not tweet any agricultural-relevant message. Yet, the data does include non-agricultural content posted by users who tweeted relevant content at some point.



**Figure 4: Gall-Peters projection of @-mentions and retweets in the sampled network**

We subsequently retrieved geographic information from our population and identified the location of 73% of users (39,858 from a total of 54,422 profiles). Twenty percent of users with geographic information were based in California (12,058 users), which is

unsurprising given the location of the seed nodes. Next we processed the text corpora and identified senders and receivers of retweet messages in the form of 'rt @' and 'via @' (including retweets resulting from Twitter's retweet button) that appeared in the text corpora. We subsequently extracted the arcs between users that mentioned other users, creating a link between the author of the tweet and every other account mentioned in a message that was not a retweet (thus allowing for multiple arcs within the same tweet). As a result, the number of @-mentions is slight higher than the number of retweets. Figure 4 shows the geographic location (Gall-Peters projection) of sender and recipient of @mentions (in red) and retweets (in blue) with hotspots of activity in North America and a clear cross-Atlantic highways of retweet activity (the directionality of the message is bent clockwise).

We removed self-loops and messages directed to or originated from users outside the target population of 59,761, which defines the boundaries of our community. As a result of the multiple sampling procedures, we ended up with a network of 32,152 nodes (users) and 4,418,390 edges (2,502,107 @-mentions and 1,916,283 retweets) posted between the first and the last day of 2014. As our research design is focused on information diffusion, we consider $A{\rightarrow}B$ when $B$ retweets $A$ and $A{\rightarrow}B$ when $A$ mentions $B$ (thus following the directionality of the information flow). Given the different rationales involved in retweeting and mentioning, we labelled the network edges to separate retweets from mentions. Figure 5 depicts the directionality of network edges considered in this study.
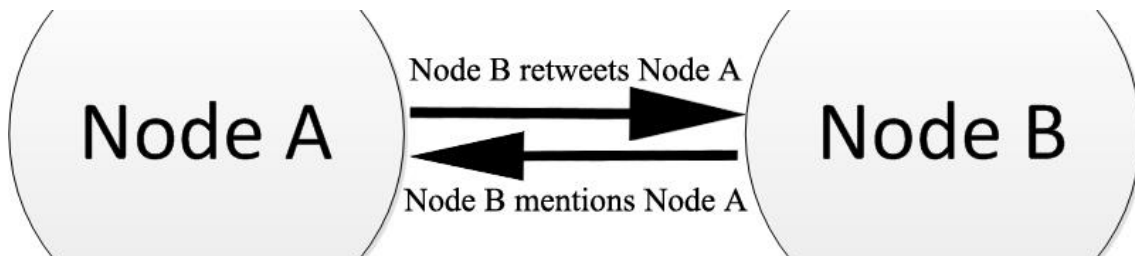


**Figure 5: Direction of edges considered in this study, with the author of a retweeted message as the sender**

The last step of data processing consisted of calculating the geographic distance travelled by each retweet and @-mention message. We retrieved the location and geographical coordinates of users in our population and deployed a function to calculate the Euclidean distance travelled by each arc connecting two users. The calculation relies on an estimate of the earth radius with the canonical mean equatorial radius of 6378.145 kilometers. Together with other information collected by processing the data, as well as the information about each message offered by Twitter REST API, the resulting graph includes detailed information about each of the nodes and each of the edges in the network (Table 1).

**Table 1: Table1: Node and edge attributes in the processed network data**

| NODES | | EDGES | |
|---|---|---|---|
| user | Usernames or Twitter handles | type | Type of exchange (@-mention or retweet) |
| id | Unique ID of Twitter users | aggieScore | Aggie score calculated by aggie classifier |
| location | Geographic location | sustainScore | Sustainability score calculated by sustain classifier |
| time | Account creation date | sustAggScore | Aggregated aggie and sustain scores |
| tweets | Number of tweets posted by user | text | Tweeted message |
| followers | Number of users following the account | favoriteCount | Number of times the message was favorited |

| followees | Number of accounts followed by user | retweetCount | Number of retweets globally (all Twitter users) |
|-----------|-------------------------------------|--------------|-------------------------------------------------|
| description | User's profile description | retweetNum | Number of retweets within the population |
| lastTweet | Last retrieved tweet | isRetweet | Whether the message is a retweet (Retweet |
| latitude | Geographic location (coordinate) | tweetID | Unique ID of the message |
| longitude | Geographic location (coordinate) | dateChar | Timestamp when the message was posted (UTC) |
| list | UC ANR Twitter list | timeChar | Timestamp when the message was posted (UTC) |
| california | Geographic location (if in California) | timeNum | Timestamp when the message was posted (UTC) |
| state | Geographic location (US states only) | distKm | Geographic distance between sender and receiver |
| state.abb | Geographic location abbreviated) | color | Edge color (mentions in red and retweets in blue) |
| membership | Modularity community assigned to | isFollower | Whether the sender follows the receiver |

## 7. Results

We relied on the fastgreedy community algorithm implemented in igraph (Csardi and Nepusz, 2006) to classify the nodes into subcommunities or modules. The network was thus divided into 10 large modules that account for 80% of the graph (32,152 users), with the 11th module including the remaining, more sparsely connected nodes in the network. We subsequently mined the content tweeted by users in each module and found it to be consistent with topical subcommunities broadly associated with agriculture. The modules detected in the network consistently presented different averages for indegree and outdegree, both for retweets and @-mentions, with influential accounts displaying high indegree for @-mentions and high outdegree for retweets, thus confirming our assertion in RO2. Such accounts provide most of the content to the community and represent important stakeholders to whom the community directs their questions and expectations. As an example, the UC ANR account has a @-mention indegree of 2,446 and a retweet outdegree of 1,820 but the @-mention outdegree and the retweet indegree are considerably lower at 370 and 458, respectively. Figure 6 gives the breakdown of inbound and outbound connections across modularity communities compared to their average number of followers, followees, and tweets. Modules #1 and #6 present a higher average and a few extraordinarily active and connected nodes.
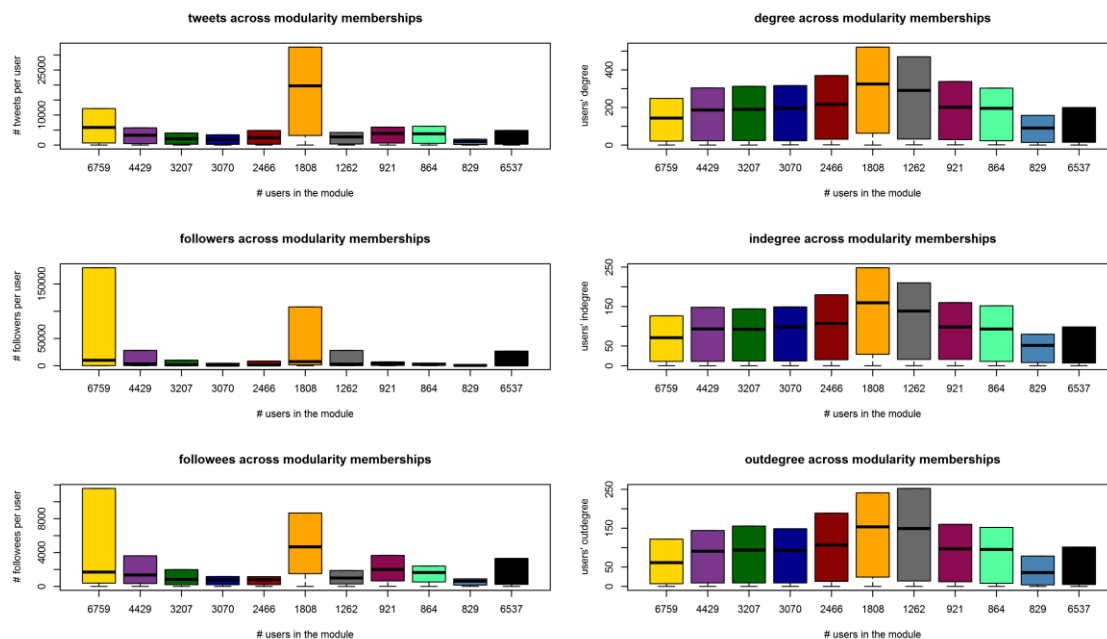


Figure 6: Follower, followee, and tweet averages compared to degree, indegree, and outdegree per module

The nonreciprocal and directed nature of retweet and @-mention interactions indicates the existence of an informational relationship between users, as the information flows only one way from one user to another, with reciprocal relationships—either as mutual @-mentions or retweets—being particularly uncommon in the community. This is also consistent with the hypothesis that the network is structured around a core (Borgatti and Everett, 2000, Holme, 2005) and a periphery and sheds particular light on RO3, with retweets flowing from core members of the community towards the periphery, and @-mentions flowing from the periphery towards the core. We subsequently tested the network for core-periphery structure and found significant correlations between coreness and degree for retweet indegree and outdegree (.69 and .40, respectively, $p<.001$) and @-mention indegree and outdegree (.52 and .77, respectively, $p<.001$). The test for the aggregated network of @-mentions and retweets also reported a significant correlation between coreness and degree with $r=.65$ ($p<.001$). Most remarkably, @-mentions mostly flow from California towards Washington, DC, likely a marker of messages politically laden and aimed at policy makers. Figure 7 shows the interactions between East and West coast, with the direction of retweets following clockwise from California towards Washington, DC and New York, and @-mentions flowing from East to the West coast.
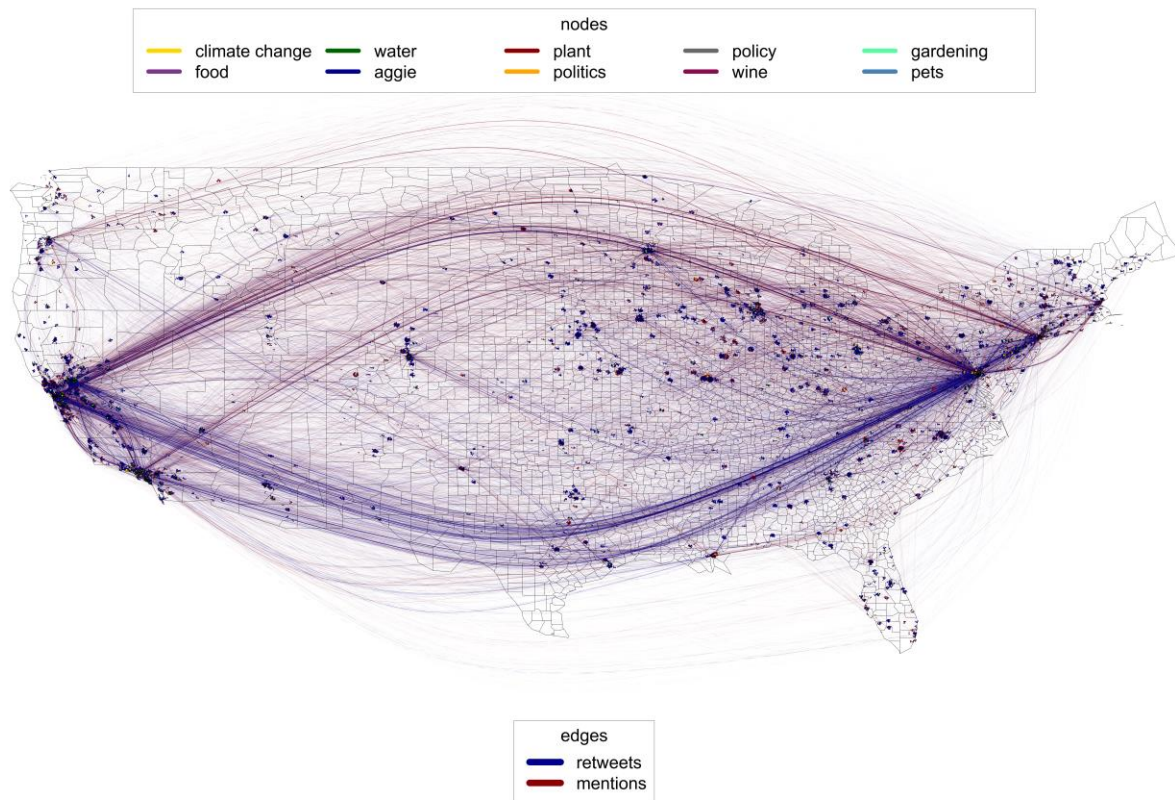


Figure 7: Directionality of @-mentions and retweets from West to East coast and back again

Core-periphery patterns are yet stronger when subsetting the network for messages pertinent to agriculture and sustainability. The correlations between coreness and degree for retweets is significantly higher for indegree and outdegree ($r=.77$ and .64, respectively, $p<.001$). The same difference is observed with agriculture and sustainability relevant @-mention messages, with yet higher correlations between coreness and @-mention indegree and outdegree (.56 and .80, respectively, $p<.001$). Although the estimates for

product-moment correlation between coreness and degree remain significant for non-agriculture-relevant messages, it is significantly lower compared to the subnetwork of agriculture-relevant messages ($r$=.77 and .57, respectively, $p$<.001). Comparatively, a random Erdös-Rényi graph with the same number of nodes and edges presents a much lower correlation between coreness and degree compared to our network ($r$=.30 and .65, respectively, $p$<.001), and unsurprisingly, a random small-world network with the same dimensions reports a correlation between coreness and degree of only .07 ($p$<.001).

After establishing that the network was structured as a core-periphery (RO3) we further investigated the network properties of users in each of the modules. We found that communities #1 and #6 present much higher average number of inbound @-mention and outbound retweet, a marker of their position as informational centers of the network that include highly-active and highly-followed accounts (RO2). We subsequently fit an ERGM model (Goodreau, 2007, Robins et al., 2007) to a single day of message exchanges and confirmed that users located in California, as well as tweet activity and following size, were significant predictors or tie formation ($p$<.0001, $p$<.005, and $p$<.0001, respectively). We understand these results to indicate the presence of hubs in the network feeding information to the community, an asymmetry detailed in Figure 8.
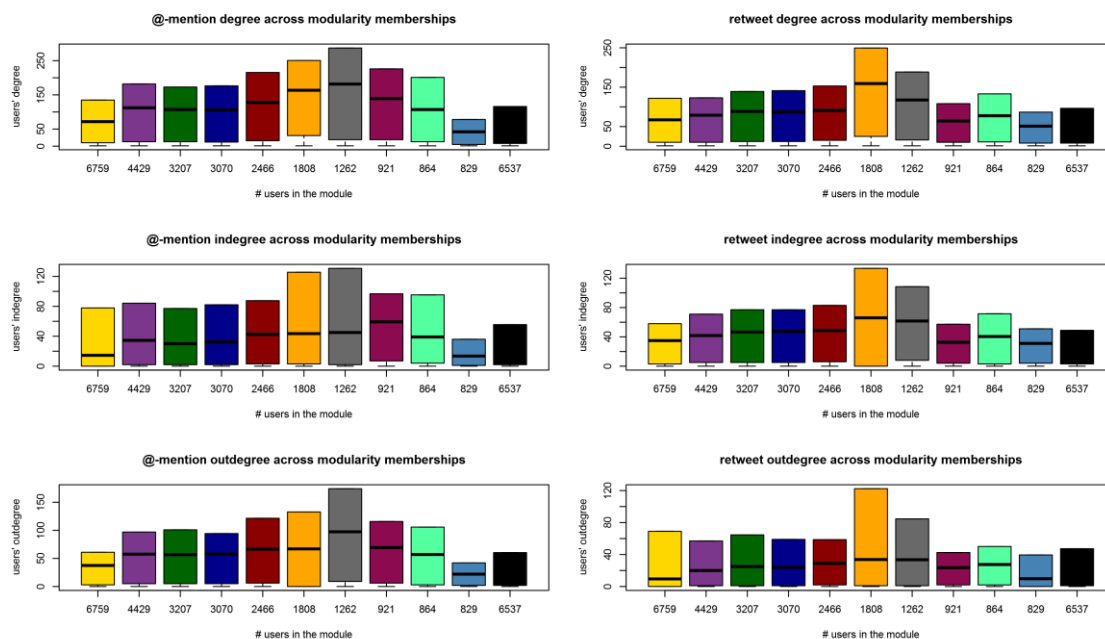


**Figure 8: Average network metrics for users in each of the 10 modules**

The degree distribution of the Twitter network is characterized by a very long tail (Kwak et al., 2010) that makes indegree and outdegre averages impractical for network analysis. However, by subsetting degree averages per module, we found that the 10 subcommunities present relatively comparable number of users and similar averages for the number of retweet and mentions. Moreover, the subcommunities display considerable skewed averages for @-mentions and retweets when indegree and outdegree are considered. Figure 8 provides detailed information about the 10 subcommunities with prevailing higher retweet indegree averages compared to retweet outdegree, particularly in communities #1, #6, and #10. The same difference is observed in the averages for @-

mentions, with considerable differences depicting subcommunities of users that are the object of messages and subcommunities of users that constantly direct their messages to other accounts. Most remarkably, and again consistent with the core-periphery structure of this network, we found that central subcommunities were both the object of @-mentions and the source of retweets, another marker of their centrality in the network. As shown in Figure 8, this is particularly the case of communities #1, #4, and #6, which include a large number of highly-tweeted and highly-mentioned accounts.
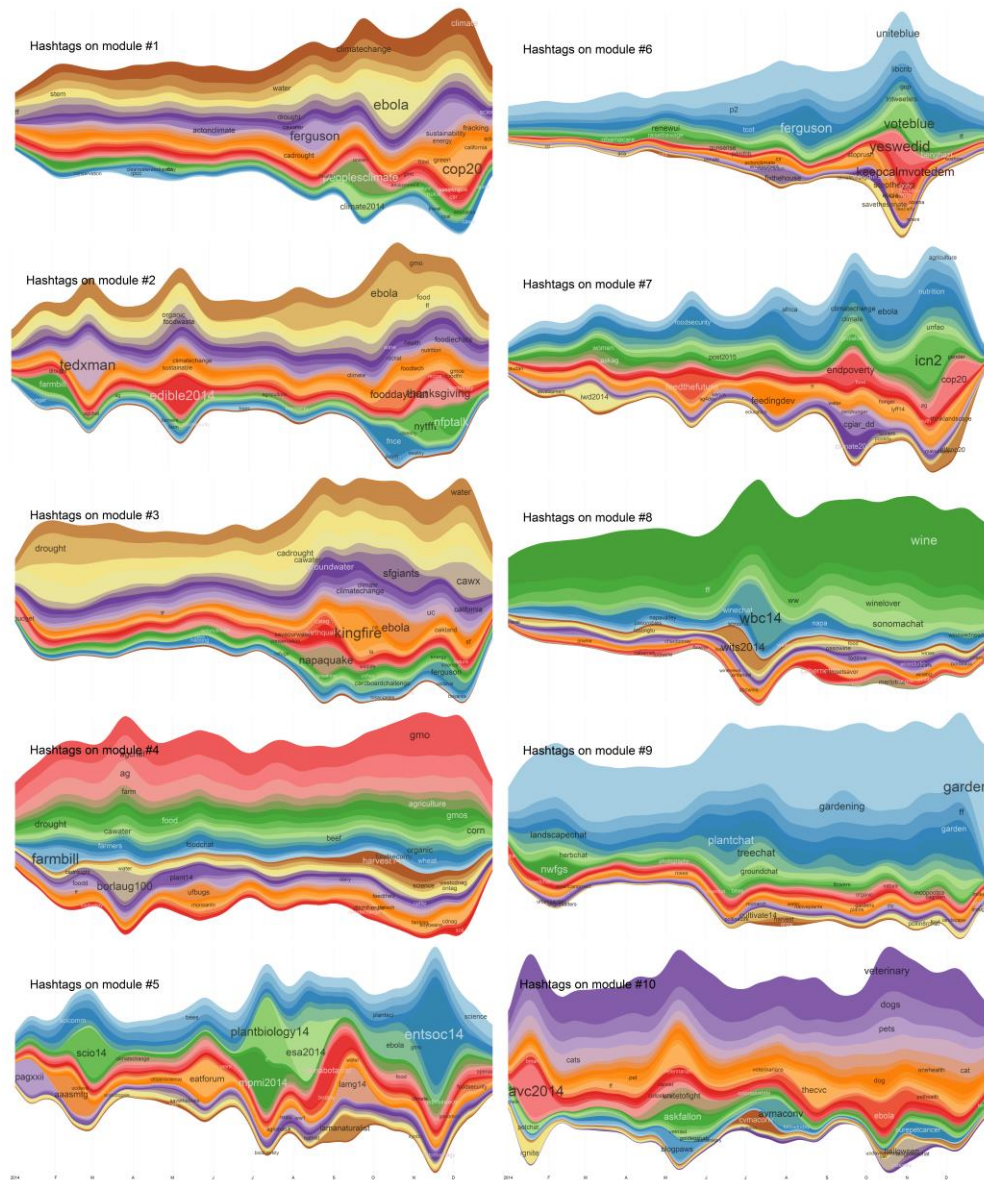


**Figure 9: Hashtag frequency over time in each of the 10 subcommunities**

Taken together the network comprises a long tail of the indegree and outdegree distributions, both for retweets and @-mentions. Accounts with large retweet outdegree refer to users whose tweets were retweeted many times by different users, often at a very high rate. This is the set of users spreading information that we refer to as broadcasters. The results of the aggieScore show that these users tend to either generate agriculture-relevant content or introduce it to the network. Accounts with a high retweet indegree, on

the other hand, refer to users that retweet a large number of messages. We refer to these users as receivers or the information, albeit from a user-centered perspective one can also assert that they are moving information throughout the network.

Next we modelled the topics discussed across the network to understand the overarching issues considered by this community. We relied on hashtags as unifying textual markers and modelled the most common terms for each of the 10 modules or groups. Perhaps unsurprisingly, the communities showed clear thematic focus, particularly around issues such as water, food, agriculture, wine, climate change, pets, gardening, policy, and politics. We subsequently analyzed the hashtag frequency and the topics discussed by users in each of the subcommunities. Figure 9 shows the frequency of hashtags over time for each of the 10 modules and highlights the thematic consistency within each subcommunity.

In order to provide a conclusive answer to RO3, we sampled the network in 10 graphs restricted to the users identified in each of the subcommunitues. As topical modules are necessarily a subset of the network, we generated Erdös-Rényi random graphs of equal dimensions (equal number of nodes and edges) and compared the correlation of coreness and degree between the modularity groups and the random graphs. By iteratively retrieving core-periphery estimates for the observed subcommunities and the random network, we tested the hypothesis that the subcommunities present cumulative core-periphery dynamics as the network becomes increasingly specialized. The estimators for randomly selected subnets consistently presented lower correlation between coreness and degree and a significantly more cohesive core was found in all subnetworks compared to the random networks.
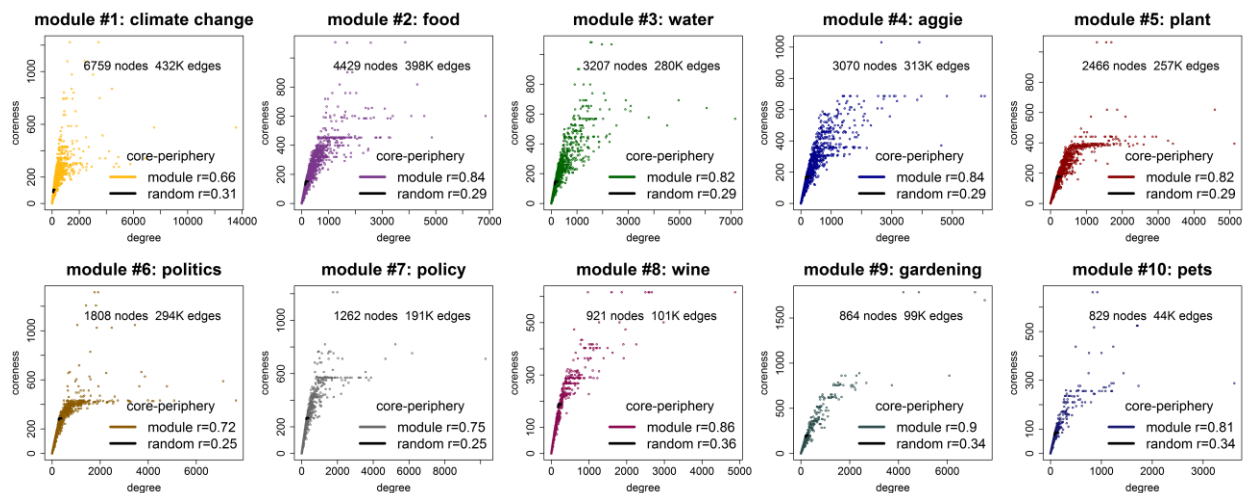


**Figure 10: Core-periphery estimates for observed communities and Erdös-Rényi random graphs of equal size**

The only exception to these results is module #11, which again confirms the results as this module it is not a subcommunity but an aggregated of nodes detached from any organic subcommunity. We understand these results to indicate that the core-periphery topology of Twitter subnetworks dedicated to agriculture is associated with the type of content shared and discussed by users, with more specialized content structuring the network around cores and peripheries broadly consistent with the separations between

topical experts and general public. Figure 10 shows the tests statistics for the observed subnetworks and the random graphs.

## 8. Conclusion

In this paper we reported on an investigation on the affordances of social media to the diffusion of agricultural information and described a replicable research design that can be applied to other studies focused on social media communities. We identified, mined, graphed, and retrieved the profile of 54,422 users and their following invested in spreading agricultural information and assigned users to cliques and topical subcommunities. We found that the network structure shows distinct patterns of core-periphery shifts due to a dense, cohesive core and a sparse, unconnected periphery (Borgatti and Everett, 2000) extending beyond the geographic borders of the immediate clique of tightly knit users. We found that retweets cascade from a few accounts to a large crowd of peripheral, but highly active users, and that @-mentions originate from or are directed to a few users that perform the role of hubs and authorities in the network. As a result, we posit that outreach initiatives resort to social networks to broadcast information from relevant government agencies to a tightly interconnected community of users invested in specific subtopics of agriculture, a theoretical framework consistent with the two-step flow theory of communication originally proposed by Katz (1957).

The intuitive understanding of the core-periphery structure stems from the layout of a network that cannot be subdivided into exclusive cohesive subgroups or factions, even though some nodes appear much better connected than others. According to Pattison (1993) and Borgatti and Everett (2000), core-periphery networks consist of just one group to which all actors belong to a greater or lesser extent. Therefore, the core community of agricultural influential users occupy the center of the graph and are proximate not only to each other but to all nodes in the network, while the remaining nodes are located in the outskirts and relatively close only to the center. In addition to the core-periphery patterns found in these graphs, the structural properties of the networks are neither completely regular nor completely random, as they display small average path lengths typical of random graphs.

This pattern of information exchange suggests that a few accounts source information to the community which subsequently retweet this information to their subcommunities of interest. This snapshot of the network provides an alternative view of the dynamics of a green virtual sphere. Although it includes a considerable portion of contentious topics, particularly climate change and issues related to draught and water resources, it also includes a large set of messages and the ensuing user interaction focused on specialized agriculture information, particularly wine and plant sciences. This confluence of subtopics mirrors the diversity of stakeholders involved in the agriculture and food sectors, including outreach centers, farmers, and government agencies. However, the highly skewed distribution of @-mention and retweet indegree and outdegree, together with the concentration of such user accounts in metropolitan areas, indicates that the core of the network is centralized around government agencies and news outlets, as opposed to farmers and growers who can benefit from having easy and direct access to new sources of agriculture knowledge. Despite the efforts of outreach professionals, the Twitter subcommunities associated with agriculture seemingly replicate the top-down, continuum

model in which information flows from government agencies and news organizations towards growers, with little reciprocal interaction between users in the periphery of the network.

## 9. References

BASTOS, M. T. 2011. Public Opinion Revisited: The propagation of opinions in digital networks. *Journal of Arab & Muslim Media Research,* 4**,** 179–195.

BASTOS, M. T. & MERCEA, D. 2015. Serial Activists: Political Twitter beyond Influentials and the Twittertariat. *New Media & Society*.

BASTOS, M. T., PUSCHMANN, C. & TRAVITZKI, R. 2013. Tweeting across hashtags: overlapping users and the importance of language, topics, and politics. *Proceedings of the 24th ACM Conference on Hypertext and Social Media.* Paris, France: ACM.

BIDWELL, D., DIETZ, T. & SCAVIA, D. 2013. Fostering knowledge networks for climate adaptation. *Nature Climate Change,* 3**,** 610-611.

BORGATTI, S. P. & EVERETT, M. G. 2000. Models of core/periphery structures. *Social Networks,* 21**,** 375-395.

BORTREE, D. S. & SELTZER, T. 2009. Dialogic strategies and outcomes: An analysis of environmental advocacy groups' Facebook profiles. *Public Relations Review,* 35**,** 317-319.

CALHOUN, C. J. 1992. *Habermas and the public sphere,* Cambridge, MIT Press.

CHEONG, M. & LEE, V. 2010. Twittering for earth: A study on the impact of microblogging activism on Earth Hour 2009 in Australia. *Intelligent Information and Database Systems.* Springer.

CSARDI, G. & NEPUSZ, T. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems,* 1695.

DUBOIS, E. & GAFFNEY, D. 2014. The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter. *American Behavioral Scientist*.

GABIELKOV, M., RAO, A. & LEGOUT, A. 2014. Studying Social Networks at Scale: Macroscopic Anatomy of the Twitter Social Graph. *In:* ACM Sigmetrics 2014, 2014 Austin, United States.

GOODREAU, S. M. 2007. Advances in exponential random graph (p*) models applied to a large social network. *Social Networks,* 29**,** 231-248.

HABERMAS, J. 1987. *The Philosophical Discourse of Modernity,* Cambridge, MIT Press.

HABERMAS, J. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society,* Cambridge, MIT Press.

HENRI, F. & PUDELKO, B. 2003. Understanding and analysing activity and learning in virtual communities. *Journal of Computer Assisted Learning,* 19**,** 474-487.

HOLME, P. 2005. Core-periphery organization of complex networks. *Physical Review E,* 72**,** 046111.

KALAFATIS, S. E., LEMOS, M. C., LO, Y.-J. & FRANK, K. A. 2015. Increasing information usability for climate adaptation: The role of knowledge networks and communities of practice. *Global Environmental Change,* 32**,** 30-39.

KATZ, E. 1957. The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis. *Public Opinion Quarterly,* 21**,** 61-78.

KWAK, H., LEE, C., PARK, H. & MOON, S. 2010. What is Twitter, a social network or a news media? *In:* 19th International Conference on World Wide Web, 2010 New York, NY, USA. ACM, 591-600.

LIANG, H. & FU, K.-W. 2015. Testing Propositions Derived from Twitter Studies: Generalization and Replication in Computational Social Science. *PLoS ONE,* 10**,** e0134270.

LUBELL, M., NILES, M. & HOFFMAN, M. 2014. Extension 3.0: Managing Agricultural Knowledge Systems in the Network Age. *Society & Natural Resources,* 27**,** 1089-1103.

MILLER, V. 2015. Phatic culture and the status quo: Reconsidering the purpose of social media activism. *Convergence: The International Journal of Research into New Media Technologies*.

MYERS, S. A., SHARMA, A., GUPTA, P. & LIN, J. 2014. Information network or social network?: the structure of the twitter follow graph. *Proceedings of the 23rd International Conference on World Wide Web.* Seoul, Korea: International World Wide Web Conferences Steering Committee.

PAPACHARISSI, Z. 2002. The virtual sphere. *New Media & Society,* 4**,** 9-27.

PATTISON, P. 1993. *Algebraic models for social networks,* Cambridge, Cambridge University Press.

ROBINS, G., SNIJDERS, T., WANG, P., HANDCOCK, M. & PATTISON, P. 2007. Recent developments in exponential random graph (p*) models for social networks. *Social Networks,* 29**,** 192-215.

SEGERBERG, A. & BENNETT, W. L. 2011. Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *The Communication Review,* 14**,** 197-215.

TORGERSON, D. 1999. *The promise of green politics: Environmentalism and the public sphere*, Duke University Press.

TORGERSON, D. 2006. Expanding the green public sphere: Post-colonial connections. *Environmental Politics,* 15**,** 713-730.

WALDMAN, S. 2011. Information Needs of Communities: The Changing Media Landscape in a Broadband Age. Washington D.C.: Diane Publishing Company.

WHITE, B., CASTLEDEN, H. & GRUZD, A. 2014. Talking to Twitter users: Motivations behind Twitter use on the Alberta oil sands and the Northern Gateway Pipeline. *First Monday,* 20.

WU, S., HOFMAN, J. M., MASON, W. A. & WATTS, D. J. 2011. Who Says What to Whom on Twitter. *In:* 20th international conference on World Wide Web, 2011 New York. ACM, 705-714.

YANG, G. & CALHOUN, C. 2007. Media, civil society, and the rise of a green public sphere in China. *China Information,* 21**,** 211-236.