Core-periphery or decentralized? Topological shifts of specialized information on Twitter

Marco Bastos (City, University of London)
Carlo Piccardi (Politecnico di Milano)
Michael Levy (University of California at Davis)
Neil McRoberts (University of California at Davis)
Mark Lubell (University of California at Davis)

Accepted for publication in Social Networks
(pre-publication version: some changes still possible)

Abstract

In this paper we investigate shifts in Twitter network topology resulting from the type of information being shared. We identified communities matching areas of agricultural expertise and measured the core-periphery centralization of network formations resulting from users sharing generic versus specialized information. We found that centralization increases when specialized information is shared and that the network adopts decentralized formations as conversations become more generic. The results are consistent with classical diffusion models positing that specialized information comes with greater centralization, but they also show that users favor decentralized formations, which can foster community cohesion, when spreading specialized information is secondary.

Keywords: Core-periphery; Social networks; Centralization; Twitter; Information diffusion; Sustainable agriculture

Highlights

- Twitter agriculture social web is modelled with the core-periphery profile approach
- Network centralization increases when Twitter users share specialized information
- Network shifts from centralized to decentralized as conversations turn more generic
- Results identify when Twitter is an information diffusion system or a social network
- The agriculture social web replicates the top-down model from government to growers

Corresponding author
Marco Bastos. City, University of London, Department of Sociology. Northampton Square, London EC1V 0HB, UK. Email: marco.bastos@city.ac.uk

Core-periphery or decentralized? Topological shifts of specialized information on Twitter

In this paper we investigate how Twitter networks can shift from a centralized topology, characterized by a high core-periphery profile, to a decentralized topology characterized by low core-periphery estimates. Classical diffusion models (Rogers, 2010; Schon, 1971) posit that centralized networks are more efficient in spreading specialized information to specific communities of interest. On the other, recent studies have foregrounded the role of decentralized networks in disseminating behavior and facilitating the development of social norms that reinforce learning in local networks (Centola, 2010; Centola and Baronchelli, 2015). Centralized networks are particularly salient in sectors relying on a small number of specialists who engage a highly diverse and continuously expanding body of potential stakeholders, a diffusion system in which experts constitute the network core feeding information to the peripheral audience. Decentralized systems, on the other hand, facilitate the emergence of new ideas growing out of practical experience. Such systems lack a clear core or periphery as the information is more widely sourced and shared by all members of the network.

Twitter is an atypical social network in which the topological characteristics of both centralized and decentralized diffusion systems are present (Gabielkov et al., 2014; Kwak et al., 2010). The basic proposition of this study is that communities of interest assume different network formations that optimize the information diffusion from an active core to a relatively passive periphery; or inversely, allow the horizontal sharing of information that can be tailored to fit with users' needs where individual decisions on which source to seek information from are relatively free, thus facilitating adaptation and implementation by local users. We explore this proposition by isolating subsets of generic and specialized tweets posted by several communities of users involved in agriculture and subsequently measuring the core-periphery profile of their multiple, comparable subgraphs. For the purposes of this study, we refer to subgraphs as a defined set of nodes and arcs of the original Twitter graph selected on the basis of specific characteristics of the message.

Agriculture and the more specialized field of sustainable agriculture are an important and useful setting in which to study the diffusion of specialized information. 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 is highly centralized and relies on a top-down, continuum model going from university researchers to cooperative extension agents and finally to farmers (Rogers, 2010; Van den Ban and Hawkins, 1996). With internet penetration rates growing in rural communities (USDA, 2015), stakeholders are increasingly adopting social media and other online forms of communication to share agricultural information across local, national, and global networks. Notwithstanding these major developments, the impact of network technologies to the diffusion of specialized information remains relatively uncharted, with only a handful studies exploring the use of social media within the agriculture and food sectors (Chowdhury et al., 2013; Rhoades and Aue, 2010).

Although agricultural extension services in the United States are historically associated with centralization (Rogers, 2010), sustainable agriculture comprises a subset of agricultural extension that can benefit from decentralized diffusion systems, with stakeholders increasingly adopting digital strategies to complement more traditional outreach systems (Lubell et al., 2014). Agricultural extension and outreach remains rooted in specialized information about agricultural practices, economic conditions, and other relevant decision-making parameters. This specialized information must be applicable at the local level to individual farms and agricultural communities, but more general ideas need to be developed at the global level by upscaling multiple local experiences and then downscaling information to catalyze local learning. Thus, the diffusion of specialized information about sustainable agriculture requires a capacity to continuously facilitate a recursive flow of local and global information, a dynamic that can benefit from both centralized and decentralized diffusion systems (Valente and Rogers, 1995).

As a consequence of this duality in communicating specialized agricultural information, the different strategies surrounding agricultural extension and sustainable agriculture outreach offer an ideal case study to investigate the diffusion of specialized information on social media. Sustainable agriculture is a quintessential example of a community where knowledge networks must spread information across specialized sub-communities that are concerned with different aspects of the complex global food system (Klerkx et al., 2015; Klerkx and Proctor, 2013). The overall knowledge network not only has to deal with internal components of the system, for example understanding climate change and water management, but must also link the specialized system to the broader global culture represented by social media platforms like Twitter. Sustainable agriculture is not unique in this way—we expect similar dynamics may apply to other broad epistemic communities, e.g. social media users discussing “energy independence,” “national defense,” and other similarly specialized topics (Lubell et al., 2011; Lubell et al., 2014).

However, sustainable agriculture is a particularly useful domain in which to study the dynamics between network structure and knowledge specialization because there is an important tradition of knowledge extension among the education and outreach professionals involved with agriculture (Clark et al., 2016). The traditional approach to knowledge extension was to deliver research findings from universities to farmers and other interested stakeholders via personal communication and networks of local extension agents. With the advent of new information and communication technology (ICT) and social media, extension professionals involved with agricultural knowledge systems (Hermans et al., 2015) are increasingly experimenting with online forms of communication and continue to contend with general ideas such as network centralization and knowledge specialization that may apply to the specific topics of interest for agriculture.

In the following, we briefly review the literature on diffusion of innovations and detail an approach to core-periphery analysis that returns a continuous measurement of the centralization observed in the network. In the later sections of the paper we present the results of this study and discuss the more general policy implications of our findings.

Specialization and Network Centralization

Classical diffusion models posit that innovation originates from expert sources and then diffuses uniformly to potential adopters who either accept or reject the innovation. The source of information is situated at the center of the communication network and adoption is mostly a passive act of imitation of the source behavior. This classical model was successfully applied to agricultural extension services and the underlying model is derived from Ryan and Gross (1943) seminal study that tracked the diffusion of hybrid corn throughout the Midwest. The original study identified diffusion agencies, commercial channels, and neighbors as key actors that informed farmers of the new seed and affected their rate of adoption. Much agricultural diffusion in the United States emerged from this centralized model, in that key decisions about how to diffuse them, and to whom, were left to a small number of technical experts (Rogers, 2010).

Schon (1971) called into question this seminal model by exploring the reality of emerging diffusion systems and criticizing the classical diffusion theory, which he referred to as the “center-periphery model.” According to Schon (1971), the assumption that innovations originate from a centralized source and then diffuse to users fails to capture the complexity of decentralized diffusion systems in which innovations originate from numerous sources, are shared among individuals, and evolve as they diffuse via horizontal networks. In such decentralized systems, innovations pop up from users at the operational levels (as opposed to the core) and new ideas can spread horizontally via peer networks, with a high degree of re-invention occurring as innovations are modified by users to fit their conditions. The topology of decentralized systems shares a remarkable resemblance with social networks, which allow information diffusion to be widely shared by adopters who also serve as their own change agents (Centola, 2010; Gibbons, 2007).

Diffusion of innovation theories thus comprehend a spectrum from centralized, information diffusion systems to decentralized, horizontal networks. Rogers (2010) argued that centralized diffusion systems were defined by a top-down diffusion from governmental agencies and technical experts to local users and often displayed a low degree of local adaptation and sharing of innovation among adopters, whereas decentralized diffusion systems were characterized by peer diffusion through horizontal networks and a high degree of local adaptation and sharing among adopters. These models of diffusion of innovations were subsequently revised and applied to the diffusion of new communication technologies, with Valente (1996) presenting a threshold concept to provide a social network formulation to the diffusion of innovations and Rice (1987) arguing that computer networks facilitated the diffusion of information to organizations’ environments.

Based on this history, it is apparent that literature exploring the link between network structure and knowledge distribution is centered on the extent to which decentralized networks are more effective at distributing information, specialized or otherwise. The relationship between network structure and task performance was found to be dependent on the type of task performed within organizations (Ahuja and Carley, 1999; Cummings and Cross, 2003), with non-routine tasks performing better in less hierarchical networks compared with more routine or simpler tasks which benefit from hierarchy, in line with the postulates of classical diffusion of information theory. Transposed to our empirical study, we hypothesize that as the proportion of specialized information being shared increases, the more likely Twitter communities will be to display centralized network formations.

Twitter: Centralized Information System or Decentralized Social Network?

Diffusion of innovation theories have offered a fertile ground for the study of “influentials” and the spread of novel information on social media, with a range of studies exploring potential metrics to assess users’ influence and passivity based on their information-forwarding activities (Bakshy et al., 2011; Romero et al., 2011; Wu et al., 2011). These seminal theories also echoed the literature of social networks and the formal definition of small-world networks. Watts and Strogatz (1998) and Newman (2000) designed a mechanism to investigate interpersonal influences through high clustering coefficient and small path length. Such a network topology deviates from centralized networks that are mostly optimized for information diffusion from a clearly-defined core to a large periphery of nodes. Compared with decentralized, well-structured small-world networks, diffusion of innovations was found to be slow in regular networks and fast but sporadic in random networks (Delre et al., 2007).

Decentralized networks proved particularly useful in the propagation of rumor and the spread of diseases (Valente, 1995). Albrecht and Ropp (1984) found that workers were more likely to report talking about new ideas with colleagues with whom they also discussed personal matters, as opposed to following prescribed channels based on hierarchical role relationships. This body of scholarship led to developments such as targeted advertising directed at cohesive subgroups who were next in the line of innovation adoption. It also contributed to the theoretical debate by suggesting that social network characteristics which influence the diffusion of innovations include centrality, density, and particularly reciprocity; a feature markedly absent of Twitter social networks, in which network topology presents characteristics typical both of social networks and of highly centralized diffusion systems (Wu et al., 2011).

Social media literature has long debated whether Twitter is an information diffusion system, characterized by a skewed distribution of links and low rate of reciprocal ties (Bakshy et al., 2011; Wu et al., 2011), or a social network, structured around social relations, with a higher incidence of reciprocal ties and a distribution of outgoing links similar to that of incoming links (Newman and Park, 2003). The debate hinges on the overall network structure observed on Twitter and is relevant to organizations and users seeking to optimize the reach of their message in the social network. If an outreach organization uses Twitter primarily as an information diffusion system, then to be effective, it is imperative for the community to identify "influentials"—i.e., users that belong to a central core and perform the role of hubs relaying information to the periphery of the network. On the other hand, if it is being used primarily as a social network, then outreach strategies should involve the development of many local relationships and dense network structures with reciprocal ties and transitive triangles will be beneficial.

Metrics of Centralization: Core-Periphery Analysis

The definition of core-periphery is intuitive and comprehends the union of a dense core with a sparsely connected periphery. More nuanced approaches tend to contrast, at one extreme, one homogeneous group with a large set of undifferentiated actors, and at the other, a two-class partition of nodes with one class being the core and the other being the periphery (Boyd et al., 2006). The core-periphery structure was only relatively recently given a formal definition by Borgatti and Everett (2000) and the bulk of the scholarship remains rooted on economics research, social inequality, and power dynamics between elites and non-elites (Csermely et al., 2013; Holme, 2005; Rombach et al., 2014). Core-periphery analysis has also been applied to the structural patterns of Usenet groups (Choi and Danowski, 2002), dynamics of protest movements (Barberá et al., 2015), the spatial distribution of ties in social networks (Volkovich et al., 2012), and knowledge networks of wine producers (Giuliani, 2005). While these studies rely on blockmodeling and k-shell decomposition (Dorogovtsev et al., 2006; Žnidaršič et al., 2017), we rely on the core-periphery profile approach (Della Rossa et al., 2013) which returns a global network measure of "core-peripheriness"—i.e., an indicator of network centralization.

Drawing from this body of scholarship, we hypothesize that the observed networks will exhibit increasingly higher estimates of core-periphery structure as users relay more specialized information (i.e., agriculture-related information) compared with the baseline of generic, non-agriculture-relevant information shared by the same set of users. In short, we expect the network topology to display structural shifts depending on the type of information that is shared among its users. Specifically, we suggest that the more specialized the information, the more diffusion will rely on core users surrounded by brokers who export information to the broader public. This framework describes a process of information diffusion that often deviates from patterns observed in social networks, as the network behaves much like a broadcast system with pronounced amplification effects for information dissemination (Myers et al., 2014). Compared with decentralized systems, it stresses the diffusion of information from experts and elite users towards ordinary users, foregrounding classical diffusion models and underplaying the potential for horizontal information sharing allowed by social networks.

Lastly, a growing body of scholarship has explored the role of brokers in spreading innovation between Twitter communities (Frahm and Shepelyansky, 2012; Mantzaris, 2014), particularly in reference to nodes that do not belong to any community but that may bridge different groups (Shore et al., 2016; Takaguchi et al., 2014). Within this line of inquiry, Grabowicz et al. (2012) explored the relationship between weak, intermediary, and strong ties on Twitter and found it to be largely structured around groups, with personal interactions more likely to occur on internal links to the groups, and new information more likely to be channeled through links connecting different groups. While these studies have advanced the understanding of information diffusion across Twitter communities, they are largely focused on information diffusion between non-specialized communities. With this new analysis, we seek to understand how communities of interest interact with specialized and generic information. We explore how network formations within these communities are affected by the type of information being shared. As such, we do not consider brokerage across communities or inter-community information diffusion, but only the relationship between information diffusion and the type of content transmitted within communities.

Hypotheses

With this study we test the hypothesis that the topology of Twitter network becomes increasingly more centralized as communities share more specialized information. We start by calculating the core-periphery score of the entire network and move forward towards the ten largest communities of interest within this network identified with the Walktrap community detection algorithm implemented in igraph, a max-modularity method based on random walks to find communities of densely connected vertices (Csardi and Nepusz, 2006; Pons and Latapy, 2005). These endogenous communities reflect areas of agricultural expertise with limited overlap across each topical subnet. Each community also presents a common group of users that tweeted both agriculture and non-agriculture-relevant tweets, thus allowing for estimating the core-periphery structure of the network in the absence or presence of specialized information.

Our substantive research questions are informed by the literature on social media and diffusion of innovations and inquire about the process by which specialized information spreads on Twitter. Does it spread from user to user in a decentralized fashion? Does it depend on influential users positioned at the center of the network? To observe these effects, we modelled and estimated the core-periphery profile of multiple, comparable subgraphs of the network that included the same set of users but resulted from messages that were either specialized (i.e., agriculture-relevant) or generic (i.e., non-agriculture-relevant). For the purposes of this study, the variation in core-periphery estimates are only meaningful if they refer to the same set of users but are generated by substantively different sets of information. By calculating the core-periphery profile of multiple subgraphs resulting from the exchange of specialized and generic information, and testing the significance of each test against a set of 100 randomized comparable networks, we test the following hypotheses to advance our understanding of how sharing specialized information on Twitter alters its network topology:

H1. The network structure of the Twitter agricultural social web is centralized and displays strong patterns of core-periphery;

H2. The network structure of Twitter agricultural social web becomes more centralized when users share specialized as opposed to generic information;

H3. Specialized, hashtag-based subnetworks are more centralized than generic, hashtag-based subnetworks.

Data Sampling, Representation, and Generalizability

The goal of our research design was to identify the agricultural network centered on the University of California Division of Agricultural and Natural Resources (UCANR), which is the organizational unit that coordinates Cooperative Extension in California comprising extension faculty, extension specialists, and extension agents dedicated to outreach activities related to agricultural and natural resources. UCANR and the University of California system comprise one of the largest, most sophisticated, and most experienced agricultural outreach systems in the world. On that basis, the UCANR provides an excellent case study to investigate the diffusion of specialized information.

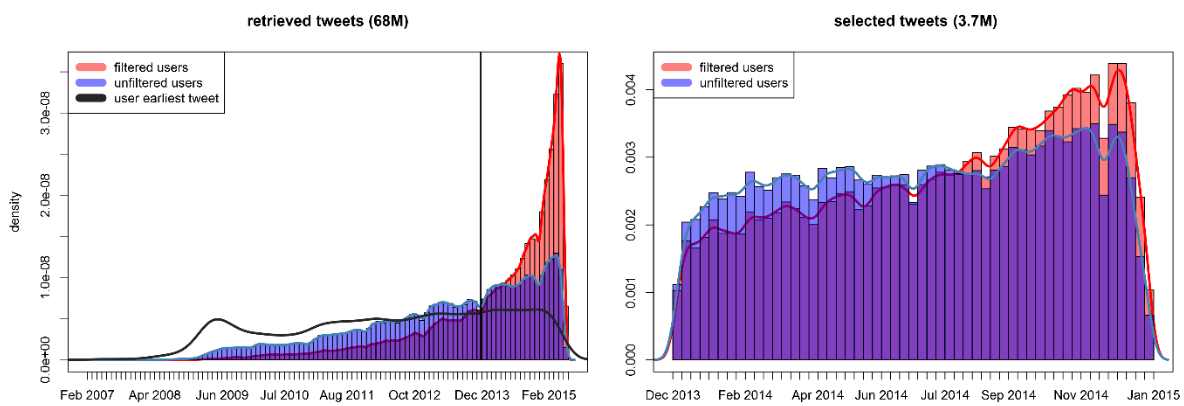
We relied on a census-based approach to collect data, starting with 153 Twitter users identified by UCANR as important sources of information on the topics of agriculture and environment which are central to the mission of the organization. This purposeful research design seeks to begin with an exogenously determined community focused on specialized agricultural topics and related subfields of agricultural expertise. We believe this approach is appropriate to isolate an initial segment of the Twitter user base that is of interest for the purposes of this study. This initial set of users formed the seed nodes from which we snowballed data collection to the larger group of users following or being followed by these users, rendering a population of 59,761 Twitter accounts that tweeted a total of 285M (285,628,862) messages since first joining Twitter.

This sample is intended to capture a specific group focused on California Agriculture where notions of expertise and communities of interest play a pivotal role. From the 59,761 accounts in the population, we managed to retrieve the timelines of 91% of the population (54,422 accounts) and geographic information from 73%. These users denote the networks nodes we explore in this study. The results reported in the next sections are limited to our sampling strategy, focused on the ANR-centered community of 153 users and their immediate network of followers and followees. Although limited to the University of California Division of Agricultural and Natural Resources, we believe this approach provides a defensible and relevant sampling of a knowledge network dedicated to the distribution of specialized information.

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. Finally, 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. The target population of 59,761 users includes 13,112 accounts that tweeted over this limit. This technical limitation imposes considerable challenges to retrieving complete sets of messages posted within a specified time-frame. As a result, we managed to retrieve only 65M (65,294,710) tweets from 54,422 users, as opposed to 285M potential tweets. These limitations are not only to be expected, but also experienced by most Twitter research and our conclusions are conditional on these constraints (Liang and Fu, 2015).

The temporal series is subject to variations as only a portion of the timelines violated 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. We resorted to this procedure to filter the 65M tweets collected from Twitter API and ended up with a total of 43M messages. We subsequently removed messages that were not retweets or @-mentions—messages from which network edges are later drawn as proxies for information diffusion—and further reduced the dataset to 26M messages.

Figure 1 shows a histogram of messages binned by month, with a cut-off date at the end of 2013 (Figure 1a) that includes the complete set of messages tweeted both by filtered users (>3200 tweets) and unfiltered users (<3200 tweets). Although the period from the end of 2013 to 2015 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 with newer messages, as they would have benefitted from a longer period to spread throughout the network.



(a) (b)

Figure 1: Messages retrieved binned by month (a) and sampled data binned by week (b)

We addressed this issue by selecting an intermediate period that was unaffected by variations resulting from data collection (Figure 1b). This period comprises the entire year of 2014 and includes a total of 3.7M (3,691,342) tweets. Our resulting dataset includes messages posted in 2014 and the analysis reported in this paper refers to this subset of tweets. The frequency of tweets binned by week is shown in Figure 1b, 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 data collection is informed by Liang and Fu (2015) and relies on a purposive sample of Twitter users (egos) extended to accounts listed as their followers and/or followees.

The last step in data collection consists of obtaining the profiles and the timelines of the selected users (egos and alters) and processing the data to generate a network 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 replicable approach to analyzing individual and community-level social media behavior. We relied on the sampled data to graph a network of @-mentions and retweets connecting users, with $A \rightarrow B$ when B retweets A and $A \rightarrow B$ when A mentions B (thus following the directionality of the information flow). In both cases, we draw an @-mention or retweet edge connecting two accounts that have posted at least one message with specialized information at some point in the year of 2014. These sampling processes rendered a network of 4.4M edges and 32K nodes. Figure 2 details the process of data collection, processing, and analysis, and describes the resulting network.

This approach to data collection is thus based on a specifically selected community of users, with the subgraphs generated for hypothesis testing being not only of comparable size, but resulting from patterns of interaction across regular communities of users. Specialized, agriculture related subgraphs are often less dense compared with their generic counterparts, but they result from patterns of information exchanged between the same set of users and observed during the same timeframe. This represents a considerable departure from hashtag-based studies, which necessarily filter information on contextual markers and thus include entirely different populations. In short, we expect this approach to data collection to allow us to better understand how Twitter network structure changes as users discuss different sets of topics. This is only possible once we can draw from the same population, with network nodes remaining unchanged over the multiple iterations of hypothesis testing.

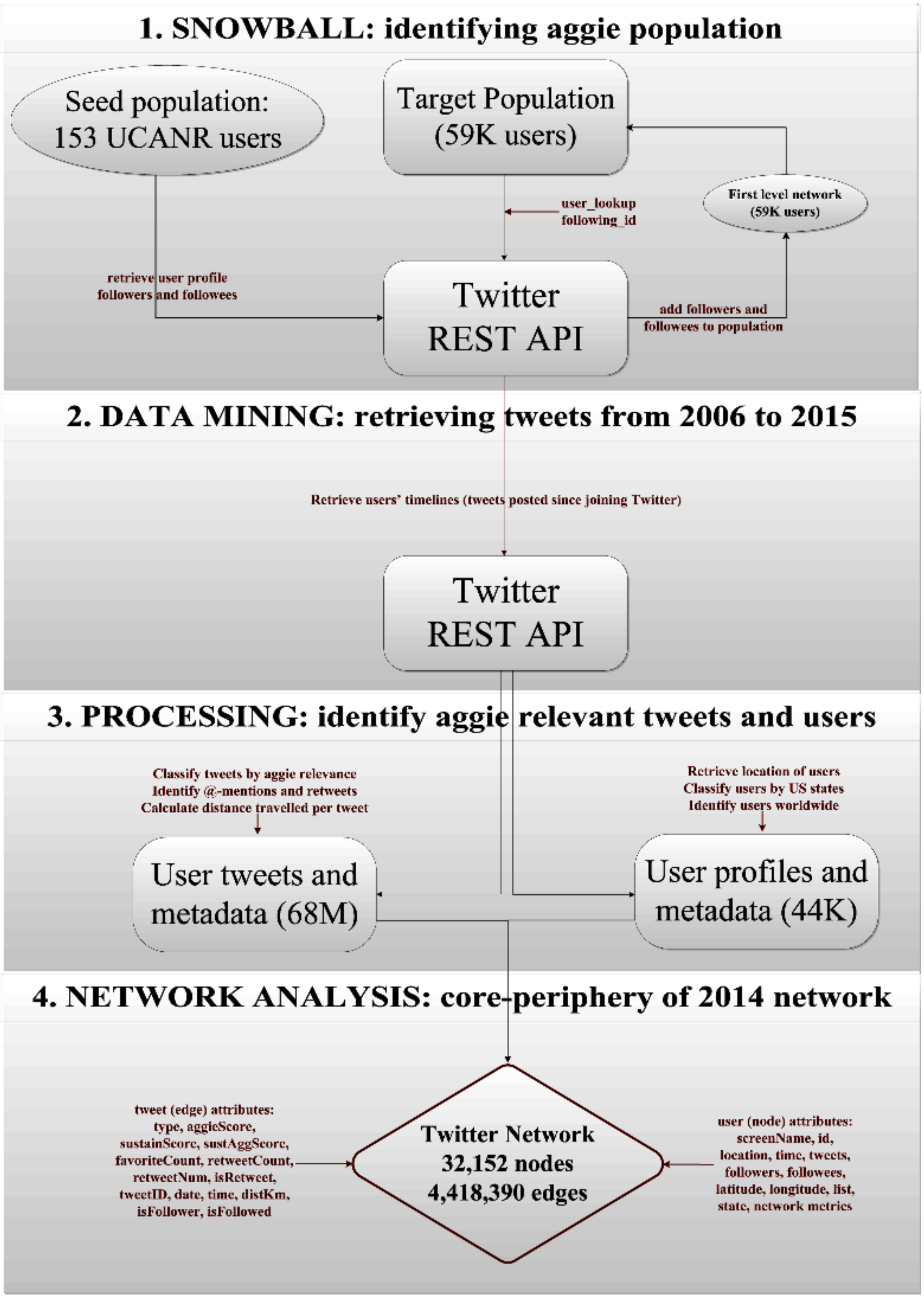


Figure 2: Data collection, mining, and network analysis

Methods: The Core-Periphery Profile

In this work we rely on the approach introduced by Della Rossa et al. (2013) to estimate the core-periphery structure of a network. By elaborating the dynamics of a random walker, a curve (the core-periphery profile) and a numerical indicator (the core-periphery score) are derived. The approach measures the extent to which the network is organized in core and periphery or, inversely, in an homogeneous structure. Simultaneously, a coreness value is attributed to each node, qualifying its position and role. This approach is fully applicable to directed and weighted networks and provides improvements over methods that define a global indicator of core-periphery based on core nodes having high closeness centrality. It also avoids explicit and artificial partitions in subnetworks (e.g., blockmodeling or k-shell decomposition) and issues related to arbitrarily defining a notion of shortest-path distance in weighted networks.

Compared with other methods that also explore core-periphery in networks quantitatively (Verma et al., 2016; Zhang et al., 2015), the core-periphery profile approach provides an estimate of core-periphery at the global network level—the "cp-score"—which can be leveraged to study shifts in the network topology. This approach relies on the notion of *persistence probability* α_S of a subnetwork S, which is the probability that a random walker currently in any of the nodes of S remains in S at the next time step (Della Rossa et al., 2013; Piccardi, 2011). Naturally, α_S=0 when S is a single node (provided no self-loop exists) and α_S=1 when S is the entire network. In an ideal core-periphery structure (Borgatti and Everett, 2000), peripheral nodes are not connected to each other but are only linked to core nodes. Thus, α_S=0 for any S composed of peripheral nodes only. However, in real-world networks such as the one investigated in this study a weak (but non-zero) connectivity exists among the peripheral nodes. This suggests a heuristic strategy for ranking the nodes from the periphery to the core of the network.

We start by the node i with minimal strength, since typically peripheral nodes are the least connected ones. Then we generate a sequence $S_1 \subset S_2 \subset \cdots \subset S_n$ of subnetworks, where $S_1=\{i\}$ is the initial node and $S_n = \{1, 2, \ldots, n\}$ is the whole network, by adding at each step k the node attaining the minimal possible value of the persistence probability α_k of the subnetwork S_k. The obtained sequence $0=\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n=1$ is the core-periphery profile of the network and α_k is a coreness measure of the node inserted at step k. By this procedure, we grow the peripheral set by adding one node at a time, trying to keep it as weakly connected as possible as the periphery should be. By keeping the persistence probability α_k as small as possible, we leave the inclusion of the highly-connected nodes to the last steps, since they would otherwise sharply enhance connectivity. As a result, highly connected nodes are typically to be found at the core of the network.[1]

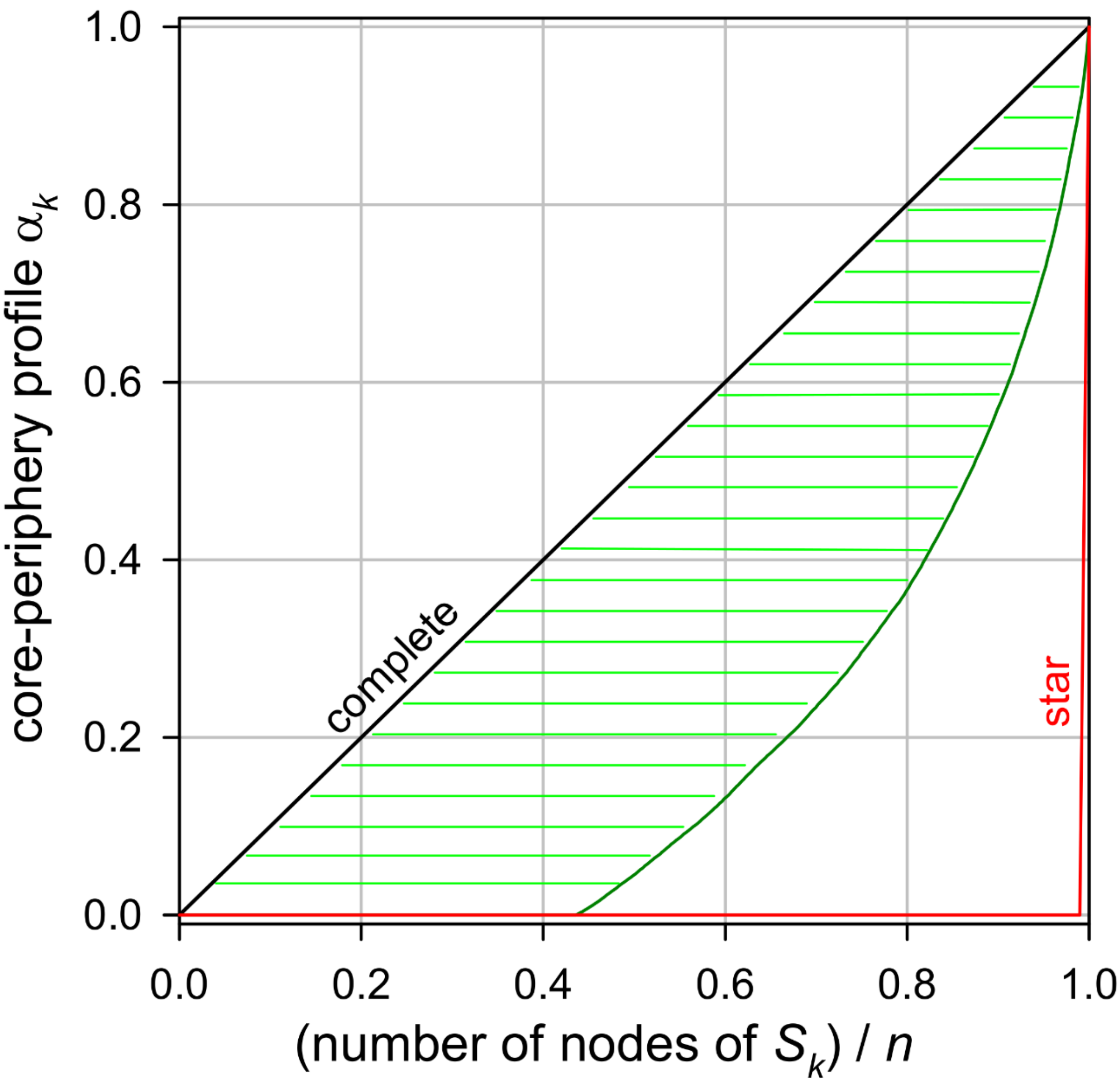


Figure 3: The core-periphery score C is the area between the core-periphery profile of a given network and that of the complete (all-to-all) network, as shown in the shaded area. The value is normalized to $0 \leq C \leq 1$, so that C=0 for the complete network and C=1 for the star network.

---

[1] Typically the measures α_k and k-coreness are positively correlated and consistent for most nodes, but anomalous nodes may exist which reveal peculiar features: low k-coreness with high α_k denote peripheral nodes acting as bridges among different network regions; high k-coreness with low α_k denote central nodes (in the k-core sense) which however fail in connecting core to periphery.

The complete (all-to-all) network and the star network are extreme cases for the core-periphery profile. The former has no core-periphery structure as all nodes are equivalent, so that $\alpha_k = \frac{(k-1)}{(n-1)}$ grows linearly from 0 to 1, while the latter is the most centralized network and has $\alpha_1 = \alpha_2 = \ldots = \alpha_{n-1} = 0$, $\alpha_n = 1$. Any other network falls somewhere between these extremes, and with this procedure we quantify the extent to which the network is centralized by the core-periphery score *C* (cp-score henceforth). *C* is the normalized distance of the core-periphery profile from that of the complete network, so that *C*=0 for the complete, all-to-all network, and *C*=1 for the star network. As shown in Figure 3, *C* becomes larger as we consider networks with more pronounced core-periphery structure and stronger centralization.

Core-periphery computations were performed in MATLAB and an implementation in R (2014) is being developed. The results of the core-periphery profile were paired with unsupervised content analyses of the text corpora. We introduced two lexicon-based classifiers to identify messages dedicated to agriculture and sustainable agriculture. As tweets often include URL links without which it is difficult to determine the topic addressed by each tweet, we retrieved the webpage title of each URL in the dataset and ran the dictionary-based classifiers over the combined corpus of tweet and webpage title (when available). The agriculture classifier is based on a set of 37 terms, 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 a URL link was available).

Results of the randomized cross validation of the classifier (Powers, 2011; Sing et al., 2005) yielded a mean accuracy of 80 and an area under the ROC curve of .81 in distinguishing between agriculture and non-agriculture relevant tweets. We ran the classifiers on the set of 9,627,146 tweets and found that only 12.09% (1,164,014 tweets) of the data was positively associated with agriculture and 5.48% (527,167 tweets) with sustainability. The classifier scores were subsequently combined and transformed into a logical (binary) vector. The resulting vector identifies agriculture relevant messages (specialized information) and allows for modelling subgraphs of the network based on specialized and generic information tweeted by the community. Even within such highly-specialized communities, communication about ordinary topics remains profuse, with sports being a common subject across communities. The tests are thus designed to identify shifts in the topology of the network as discussions within each community move from generic to specialized.

The tests reported in the following section refer to sets of networks comprising specialized/non-specialized information in addition to 100 randomized networks for each subgraph. This step was necessary to evaluate the significance of the cp-scores measured. For each one of the 10×2 (specialized/non-specialized) or 10×4 (hashtags) networks, we first check the statistical significance of the cp-score with respect to randomization. For each test, we create and write a sample of 100 random networks with the same strength of each individual node (i.e., identical sum of the weights of the incident links). This null model, a standard approach in the literature and the basis of the definition of modularity (Newman, 2010), allows us to assess the significance of the cp-score with respect to randomization. We quantify the significance of the cp-score *C* by computing $z = \frac{C - \mu C}{zC}$ where μ^{C} and z^{C} are mean and standard deviation of the cp-scores of the 100 random networks. A large *z* (e.g., $z>2$) indicates that the network has a significant, non-random core-periphery structure (i.e., it is self-organized in a more centralized form than its random counterparts).

Results

We started by categorizing the communities identified by the community detection algorithm into ten specialized topical subnetworks. These ten large modules account for 80% of the graph (32,152 users) and the remaining, more sparsely connected nodes, are not considered in the following analyses. Consistent with previous results (Bastos et al., 2013), we found that the communities tweeted dominant hashtags that could be leveraged to distinguish substantive thematic communities. We used our expert judgment, and the hashtags tweeted within each community, to identify their substantive topic of expertise subsequently labelled as the following: climate change, food policy, water management, agriculture, plant sciences, politics, international development, viticulture, gardening, and animal welfare.

Next we computed the core-periphery profile of the entire network (32K users) and found that it presents a clear pattern of core-periphery, with cp-score *C*=.79. These results reject the null hypothesis of a decentralized network marked by strong peer-to-peer dynamic and support hypothesis H1: the network structure is centralized, displays clear patterns of core-periphery, and strongly departs from that of a complete (all-to-all) network, tending instead towards a star-like topology. To further test hypothesis H1, we subsequently ran the core-periphery test on each of the ten communities and found significant variations across groups. The variation is not dependent on community size and it is significant for each of the ten topical communities, which present a higher core-periphery estimate compared with a similar, randomly generated network.

Maximum and minimum core-periphery scores *C* across communities were .73 and .87 (mean $\bar{x}$=.78, median $\tilde{x}$=.77, and standard deviation σ=.04), compared with .60 and .72 for the randomized subgraphs ($\bar{x}$=.67, $\tilde{x}$=.68, and σ=.03). While the average *C* core-periphery score observed across communities is similar to that observed on the entire network (*C*=.78 and *C*=.79, respectively), the variance observed across communities suggest that higher core-periphery estimates are associated with the level of expertise attached to each of the communities, particularly in view of the high estimates observed in highly specialized communities like water management and viticulture (.80 and .82, respectively). This is in sharp contrast to more generic networks such as politics and gardening, in which core-periphery centralization scores range from .73 to .74. Although the difference between *C* core-periphery score in specialized topics and generic information appear small, they are significant with respect to randomization. Figure 4 summarizes these findings and shows the temporal distribution of hashtags across communities, with an indication of the number of nodes, the observed core-periphery estimate, and the cp-score of a comparable, randomized network.

We subsequently ran multiple core-periphery analyses to test the hypothesis that higher estimates of core-periphery structure are to be observed when, within each of the 10 communities, the discussion shifts from specialized, agriculture-driven versus generic, non-agriculture-related information (H2). For this test, we relied on the unsupervised text classifiers described above which ranked messages from 0 (generic) to 1 (specialized). Any message scoring above zero was identified as agriculture relevant, thus transforming the continuous scale to a binary one. Based on such classification, we generate two subgraphs of roughly comparable size (henceforth referred to as “aggie” and “non-aggie”) and calculate the core-periphery centralization for each of the communities and their comparable, randomly generated network.

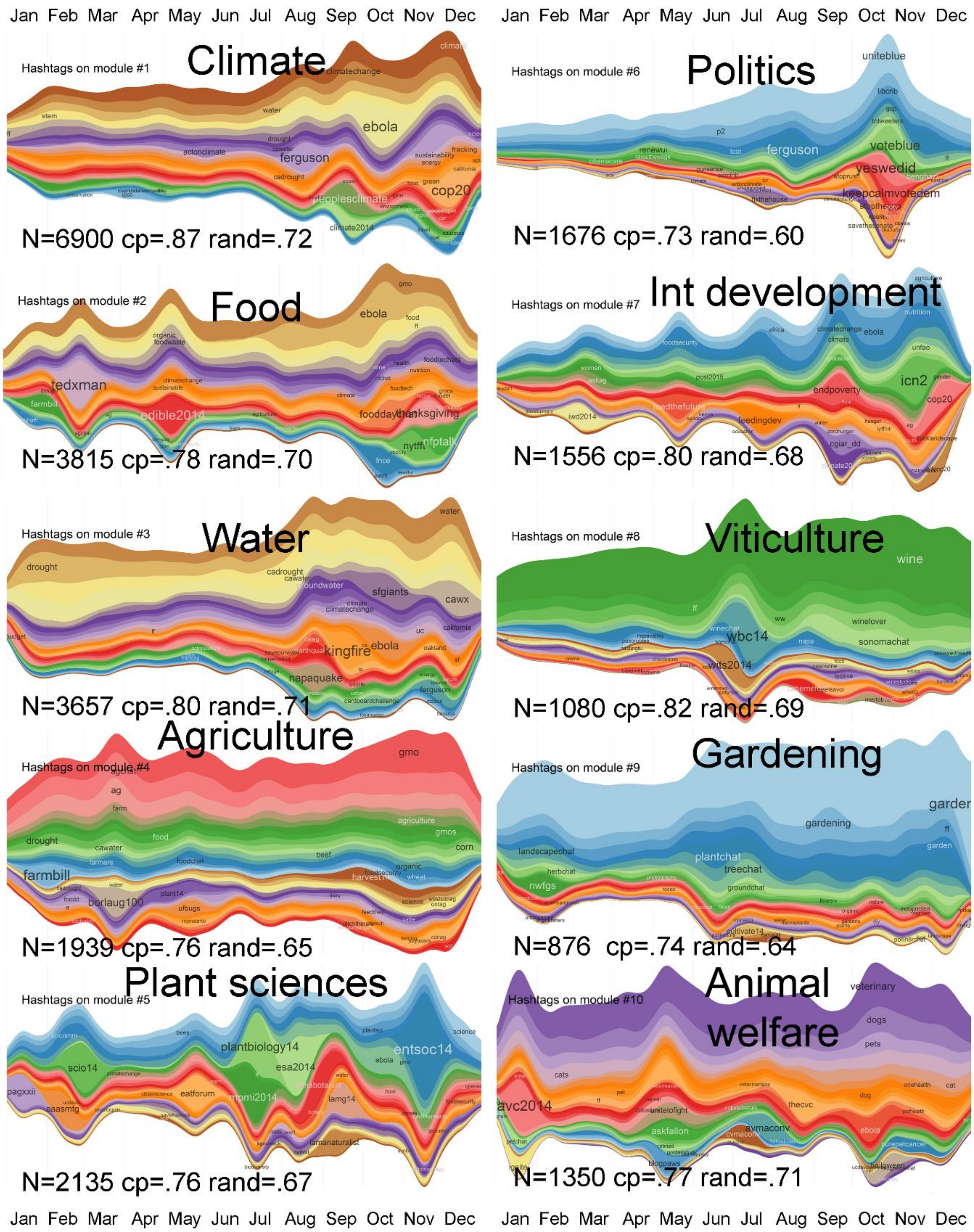


Figure 4: Temporal distribution of hashtags across communities. Labels show number of nodes, observed core-periphery estimate, and the randomized estimate for each subgraph

The rationale of this procedure is to test hypothesis H2 that more specialized topics are likely to present more pronounced core-periphery centralization estimates. As mentioned above, for each iteration of our tests we first compare the observed values of core-periphery against the observed values of a random simulation. In the 10×2 specialized/non-specialized networks (aggie vs. non-aggie), the results are consistently significant with respect to randomization, with *z* never falling below 5.2. On the basis that the twenty subgraphs generated during this iteration have presented significantly higher core-periphery estimates compared with their random counterparts, the results support Hypothesis H2 as each specialized, agriculture-defined subgraph also presented a higher estimate of core-periphery profile compared with its generic, non-agriculture defined subgraph.[2]

While specialized subgraphs presented a mean *C* estimate of .81 ($\tilde{x}$=.80, σ=.04), the generic counterpart reported a mean *C* estimate of .78 ($\tilde{x}$=.77, σ=.04). The hypothesis that the two means are equal can only be rejected with mild significance (p=0.138, two-sample t-test), which is unsurprising given that the two distributions broadly overlap. However, when performing a cp-score pairwise comparison community by community (i.e., aggie vs. non-aggie), we found that the former is larger in 10 cases out of 10. In addition to that, we performed a binomial test to check the hypothesis that aggie and non-aggie subnetworks have the same probability of having a larger score. We believe the binomial test is a more appropriate for this case, chiefly because the pairwise comparison is performed over networks of comparable size. The results of the binomial test strongly rejected the hypothesis that the two means are equal (p=0.5^10).

The difference in core-periphery estimates is indicative of the overall impact of specialized information to the network structure formed by the interaction of Twitter users. Figure 5 presents a detailed account of these results, with a higher core-periphery estimate for all ten aggie-relevant subgraphs compared with their generic subgraphs. Each subgraph is plotted over the geographic grid of continental USA for easy comparison between specialized, agriculture subgraphs and generic, non-agriculture bounded subgraphs. Many communities also featured substantial international linkages, which highlights the potential for social media to

[2] The cp-scores of the 10 aggie and the 10 non-aggie networks are, respectively, A=[0.89, 0.80, 0.83, 0.78, 0.79, 0.78, 0.83, 0.83, 0.74, and 0.80] and B=[0.87, 0.78, 0.80, 0.76, 0.76, 0.73, 0.79, 0.81, 0.73, 0.76].

increase the geographic range of specialized communication relative to traditional interpersonal outreach strategies.

Lastly we test hypothesis H3 by generating a total of forty subgraphs. For each of the ten communities, we selected four hashtags that are particular to that community, two of which were judged to be very specialized and two very generic.[3] Therefore, two of the hashtag-based subgraphs refer to specialized conversations and two refer to generic topics of conversation within that community. All hashtags selected for generating subgraphs were selected from the list of ten most frequently used hashtags within each community. We repeated the procedure for each community, hence rendering 10×4 subgraphs—twenty subgraphs of specialized conversations and twenty subgraphs of generic interactions. For each iteration of this test we also calculate the core-periphery profile of a set of 100 randomized, comparable network. Once again, a large z (e.g., $z>2$) indicates that the network has a significant, non-random core-periphery structure. In the 10×4 hashtag-based subgraphs, z is larger than 2 in 38 cases out of 40. This shares a resemblance with the previous reported experiment, but it was specifically designed to test hypothesis H3: that the network structure of specialized, hashtag-based subnetworks is more centralized and that the network is increasingly structured around a core and a periphery as the topic of conversation becomes more specialized.

The results of this last experiment confirmed hypothesis H3, that increasing specialization of topics is associated with star-shaped network formation, with the specialized hashtag-based subgraphs exhibiting significantly higher centralization than their generic counterparts. Figure 6 unpacks these results, with the reference line in green indicating the core-periphery estimate for the entire network. The core-periphery estimates for subgraphs of specialized conversation are shown in blue and present consistently higher core-periphery scores C compared with the subgraphs of generic conversations shown in magenta. Average core-periphery estimates for generic subgraphs are equal to the core-periphery estimate of the entire network ($\bar{x}$=.80, $\tilde{x}$=.79, σ=.05). In sharp contrast, this baseline value (C=.79) is the absolute minimum core-periphery estimate observed for specialized subgraphs, with much higher

[3] We selected the following forty hashtags: *Climate change*: california, climatechange, ff, peoplesclimate; *Food policy*: edible2014, ff, gmo, recipe; *Water management*: cadrought, oakland, saveourwater, sfgiants; *Agriculture*: farmbill, ff, harvest14, organic; *Plant sciences*: ebola, entsoc14, ff, lamg14; *Politics*: environment, ferguson, ff, obamacare; *International development*: africa, climatechange, ff, globalag; *Viticulture*: ff, sonomachat, winelover, ww; *Gardening*: americangrown, ff, photography, plantchat; *Animal welfare*: dogs, ff, onehealth, veterinary.

quartiles (Q_1=.84 and Q_3=.90, respectively) and an average of C=.87 across all communities ($\bar{x}$=.88, $\tilde{x}$=.87, σ=.04).[4] The pairwise comparison, community by community, of the mean of the two specialized cp-scores against the mean of the two generic cp-scores yields a larger value for the former in all 10 cases.

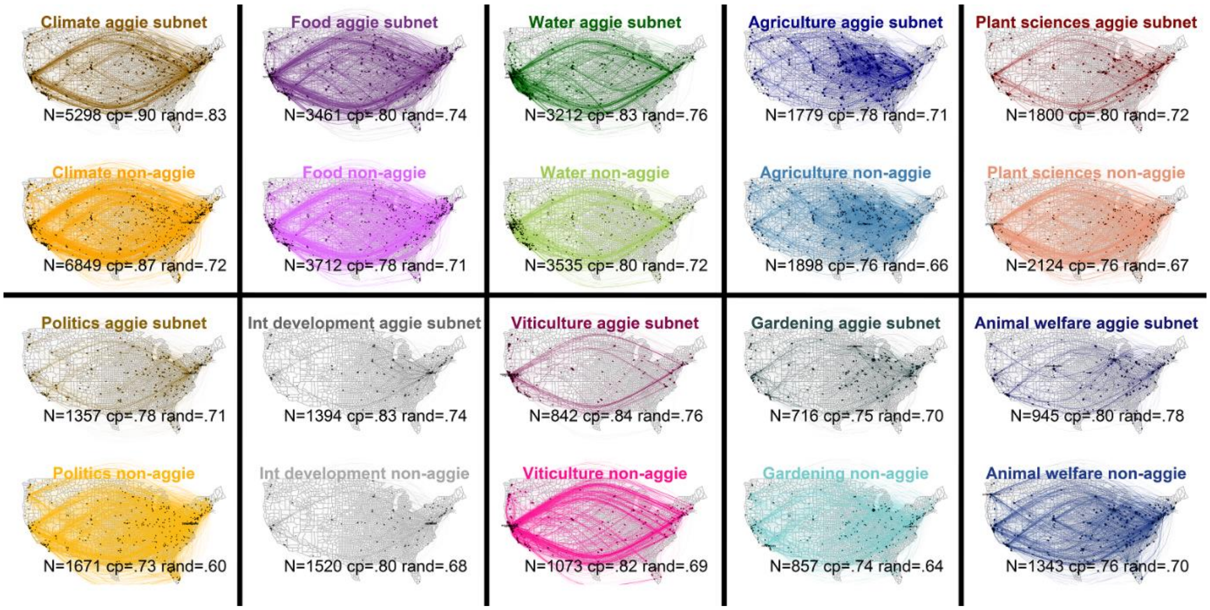


Figure 5: Core-periphery estimates for specialized (aggie) vs. generic (non-aggie) subgraphs. All values are significant with respect to randomization

The few exceptions to this trend were found in the communities water management and gardening, where one of the selected generic hashtags (#sfgiants and #photography, respectively) scored higher than the one of the specialized hashtags (#cadrought and #americangrown, respectively). Other negative results include the hashtag #ff, which was selected only due to the lack of other large generic hashtags in the communities. Although #ff is in no way a specialized topic of conversation (it stands for “Follow Friday”), it is potentially problematic because it is designed to tell other users whom to follow every Friday, and therefore can constrain the network to a core-periphery structure. Yet, subgraphs based on #ff scored lower than specialized

[4] The cp-scores of the 20 specialized networks are A=[0.83, 0.81, 0.89, 0.79, 0.97, 0.87, 0.90, 0.85, 0.84, 0.79, 0.87, 0.86, 0.88, 0.83, 0.94, 0.92, 0.94, 0.90, 0.88, 0.83]. The cp-scores of the 20 generic networks are B=[0.78, 0.71, 0.78, 0.78, 0.94, 0.82, 0.82, 0.73, 0.80, 0.78, 0.86, 0.81, 0.83, 0.78, 0.85, 0.75, 0.79, 0.78, 0.84, 0.78]. The two distributions are more differentiated compared to the previous test (H2) and the statistical hypothesis that the two means are equal can be rejected even at less than 1% significance level (p=8e-5). The difference of the two means is 0.0690 and the 95% confidence interval of this difference is completely above zero [0.0374, 0.1006].

hashtags in seven out of the nine tests. Within the communities agriculture and politics, #ff outperformed one of the specialized hashtags (#harvest14 and #ferguson, respectively).

Despite these caveats, from the forty tests based on subgraphs of generic (20) and specialized (20) hashtags, only four generic subgraphs exhibited a higher core-periphery score than their corresponding specialized subgraph. Some specialized subgraphs based on agriculture-relevant subtopics presented extraordinarily high core-periphery estimates, particularly #saveourwater within the water management community, which presented a core-periphery profile of .97 and thus very close to a perfect star network. The results therefore confirm that 1) these networks are significantly structured around a core and a periphery; 2) specialized, agriculture-relevant subnetworks are more centralized than non-agricultural, generic subnetworks; and 3) specialized, hashtag-based subnetworks tend to give shape to more centralized networks compared with generic, hashtag-based subgraphs.

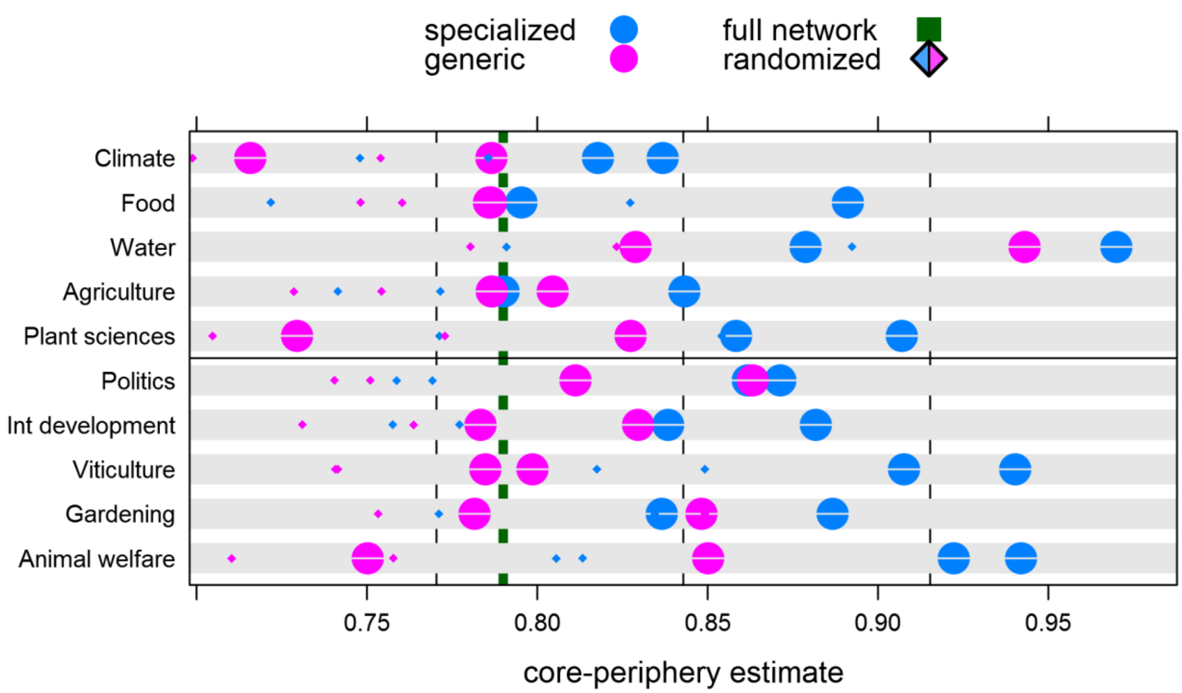


Figure 6: Core-periphery *C* scores of specialized (blue) and generic (pink) hashtag-based subnetworks and the observed *C* scores of random simulations. The green dotted line shows the core-periphery *C* score of the entire network. All values are significant with respect to randomization

Discussion

The results reported herein indicate that the entire network shows strong patterns of core-periphery shifts due to a dense, cohesive core and a large sparsely connected periphery (Borgatti and Everett, 2000). The communities providing specialized agricultural information become significantly more star-shaped as users tweet, retweet, or comment on messages relaying specialized information. The shift from more horizontal, decentralized topologies in which users interact and discuss generic topics (e.g., #sfgiants) towards a hierarchical, star-like structures in which information cascades from a few accounts to a large crowd of peripheral users is consistent with classical diffusion models, which posit that decentralized diffusion system are more likely to emerge when the innovations being diffused does not require high levels of technological expertise.

As such, Twitter networks become more centralized and structured as an efficient information broadcast system when users change their conversation from generic to specialized topics. Consistent with classical diffusion models, this latter type of information originates mostly from mainstream media, individual specialists, government and non-government agencies, and research groups that appear prominently in the core of the network, but whose interaction is limited and mostly broadcast their information to large groups of users in the periphery of the network. The results also lend support to current attempts to establish whether Twitter is an information diffusion system with a skewed distribution of links and low rate of reciprocal ties or a social network structured around social relations. The structure of Twitter communities dedicated to agriculture seems remarkably flexible with the results indicating that patterns both associated with information diffusion system and with social networks may emerge as a function of the type of content transmitted within communities.

This study has important implications for policymakers and outreach professionals dedicated to the real-world diffusion of specialized information on Twitter. Firstly, our results indicate that Twitter is not much of an equalizer when it comes to the diffusion of specialized information, as the communities are not decentralized and instead tend towards a star-like, centralized network typical of broadcasting systems. Secondly, social media professionals managing Twitter accounts should consider the highly influential core through which most of the network flow passes and the hashtags used by these sources. Such professionals are faced with the challenge of monitoring and engaging with users from different sectors of society that play a critical role in shaping this diffuse, but effective core comprising both locally influential users and globally dominant players. Thirdly, although the seed nodes are prominently positioned in the core of the network, they are superseded by media outlets that are effectively the central actors driving the information flow. In view of that, social media-based outreach would benefit from pragmatically engaging with the mainstream media embedded to the network core.

Outreach professionals might find these results unsurprising given the historically high level of centralization found in the agricultural extension services in the United States (Rogers, 2010; Valente and Rogers, 1995) and the difficult task of establishing horizontal channels of information exchange against the backdrop of increasingly specialized information. This explains not only the high cp-score observed at the top level of the network, but the significant increase of core-periphery scores C observed in each of the communities as they turn their attention to specialized information. Nonetheless, these results offer important insights for agriculture outreach and extension seeking to optimize the reach of their message in the social network. Organizations resorting to Twitter primarily as an information diffusion system are likely to benefit from engaging with the central core that relays information to the periphery of the network. On the other hand, outreach professionals seeking community cohesion are likely to benefit from outreach strategies that seek to create new connections among their network of followers.

These professionals can develop different strategies to engage with individual cores pertaining to each of the specialized communities. Given the context of specialization, core users are likely to require tailor-made tweets that suit their policy agenda to engage with the content. Outreach professionals can also minimize or refrain from using too many hashtags which can tire the core audience. As shown in Figure 4, conversations across communities are mostly fragmented across various hashtags and the divided public attention is a deterrent in actively engaging the core. A higher rate of success cases is found in communities using hashtags that have gained traction with the community (e.g., #wine for viticulture and #cawater for water management). Consequently, outreach and social media professionals are likely to benefit from using recognizable hashtags that route specialized conversations to and from the core, tailoring their messages to reach larger audiences by @-mentioning central users, including web links into their posts, and attaching images and multimedia features to the message. These simple practices can minimize the likelihood of the network falling back to highly centralized formations in which users interact little and contribute even less.

Conclusion

In this paper we analyzed the relationship between Twitter network structure and the diffusion of specialized information. We relied on a snowball-based census of the California agriculture social web to identify the boundaries of multiple communities invested in agriculture and retrieved a sample of tweets comprising both specialized and generic information shared by this population. After identifying ten endogenous communities organized around topical themes, we controlled for the type of information being shared by users and tested the hypothesis that the diffusion of specialized information leads to stronger patterns of core-periphery. The significant estimates of core-periphery have interesting implications on how the information flows among users, and when inducing the communities to 2×10 subgraphs of specialized versus non-specialized networks, we found that in all instances a more pronounced core-periphery structure emerges from specialized, agriculture-related subgraphs.

We subsequently identified a set of 4×10 hashtags associated with specialized and generic conversations and defined forty subnetworks. In most instances, specialized subgraphs gave rise to a more pronounced core-periphery structure than generic subgraphs. This second experiment confirmed again the set of hypotheses laid out in this study, which proposes that specialization is linked to higher centralization, even in horizontal networks such as Twitter. These analyses were possible due to a census-based approach to Twitter communities that has both strengths and limitations. On the one hand, it is fully reproducible, free of cost, free from Twitter data reseller constraints, retrieves relevant Twitter @-mention and retweet networks beyond hashtag-based samples, and perhaps more importantly, it allows researchers to retrieve historical data. On the other hand, and given the stringent limits imposed by Twitter REST API, the process of data collection can be time consuming and the challenges in scaling up this approach can only to a certain extent be addressed with distributed computing.

Consistent with classical diffusion models, the results show that Twitter network structure presents significantly higher degree of centralization when users are sharing specialized as opposed to generic information. The highly-skewed distribution of @-mentions and retweets, mostly concentrated on user accounts located 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 could benefit from sharing information and having direct access to new sources of agriculture information. Conversely, Figure 5 shows that messages on Twitter have the potential to reach a more diverse and geographically dispersed audience of peripherical followers, much in line with the advantages of decentralized networks to the diffusion of sustainable agriculture (Lubell et al., 2014). Despite these promising opportunities, the agricultural community on Twitter continues to 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 when specialized information comes into play.

These results are perhaps surprising given the multiple possibilities of horizontal propagation allowed by Twitter. They also highlight potential strategies that the agriculture extension could implement to improve their reach and effectiveness: if they can position themselves as brokers in online social networks, they may be able to facilitate a more distributed, horizontal, and efficient flow of information among the periphery which comprises the large majority of users in the network. The findings also shed light on the long-standing debate about whether Twitter is a centralized systems of information diffusion or a social network. The social media platform can rapidly shift between information diffusion and social network formations as users move from specialized to generic topics of conversation. In other words, Twitter communities are likely to adopt a centralized formation when spreading specialized information, but these communities will also favor more decentralized formations, which can foster community cohesion, when the diffusion of specialized information is not critical.

Another conclusion of this study is that the strong patterns of core-periphery indicate that a few accounts source information to users which subsequently retweet this information to their communities of interest. This network topology of specialized communities is broadly consistent with the separations between topical experts and general public and speaks to the core of theories of two-step flow of information diffusion, a conceptualization that was originated prior to most diffusion research but that anticipates the central assumptions of diffusion theories. We found these results relatively surprising given that each iteration of our tests was applied to an organic community of users simultaneously discussing generic (e.g., #sfgiants) and specialized (e.g., #cawater) topics. In short, the tests were designed to ensure that the generic and specialized subgraphs were drawn from the same population and rendered graphs of comparable size, as shown in Figures 3 and 4. This control mechanism allowed us to run multiple iterations of the core-periphery calculations over subgraphs that vary only in terms of the type of information being shared at a given point in time.

With this study, we sought to advance this scholarship by exploring the structure of network communities as they move to different topics of discussion. We believe this is an important contribution to the literature, as differences found at this level are not confounded by aggregate user behavior from substantively different subgraphs. In fact, by controlling for the pool of users comprising the network we can model dependencies between the information being shared and the network structures emerging from it beyond broad comparison of substantively dissimilar networks. The results also highlight the conditions under which Twitter behaves like a centralized systems of information diffusion or a social network. The social media platform continuously shifts back and forth as conversations move between the specialized-generic polarities. When a community of users is discussing ordinary topics, we can expect a more horizontal, all-to-all network typical of social conversations. As the topic of the conversation becomes more specialized, the network structure becomes more hierarchical with the large majority of users listening to experts, who are more likely to be @-mentioned and retweeted. At this point, the network behaves much like an information diffusion system with pronounced amplification effect for information, a topology broadly consistent with the original articulation found in classical diffusion theory.

References

Ahuja, M.K., Carley, K.M., 1999. Network Structure in Virtual Organizations. Organization Science 10, 741-757.

Albrecht, T.L., Ropp, V.A., 1984. Communicating about Innovation in Networks of Three U.S. Organizations. Journal of Communication 34, 78-91.

Bakshy, E., Hofman, J., Mason, W., Watts, D., 2011. Everyone's an Influencer: Quantifying Influence on Twitter, Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. ACM, Hong Kong, China, pp. 65-74.

Barberá, P., Wang, N., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J., González-Bailón, S., 2015. The Critical Periphery in the Growth of Social Protests. PLoS ONE 10, e0143611.

Bastos, M.T., Puschmann, C., Travitzki, R., 2013. Tweeting across hashtags: overlapping users and the importance of language, topics, and politics, 24th ACM Conference on Hypertext and Social Media. ACM, Paris, France, pp. 164-168.
Borgatti, S.P., Everett, M.G., 2000. Models of core/periphery structures. Social Networks 21, 375-395.
Boyd, J.P., Fitzgerald, W.J., Beck, R.J., 2006. Computing core/periphery structures and permutation tests for social relations data. Social Networks 28, 165-178.
Centola, D., 2010. The Spread of Behavior in an Online Social Network Experiment. Science 329, 1194-1197.
Centola, D., Baronchelli, A., 2015. The spontaneous emergence of conventions: An experimental study of cultural evolution. Proceedings of the National Academy of Sciences 112, 1989-1994.
Choi, J.H., Danowski, J.A., 2002. Making a Global Community on the Net – Global Village or Global Metropolis?: A Network Analysis of Usenet Newsgroups. Journal of Computer-Mediated Communication 7, 0-0.
Chowdhury, A., Odame, H.H., Ramsey, D., 2013. Social Media for Enhancing Innovation in Agri-food and Rural Development: Current Dynamics in Ontario, Canada. The Journal of Rural and Community Development 8, 97-119.
Clark, W.C., van Kerkhoff, L., Lebel, L., Gallopin, G.C., 2016. Crafting usable knowledge for sustainable development. Proceedings of the National Academy of Sciences 113, 4570-4578.
Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695.
Csermely, P., London, A., Wu, L.-Y., Uzzi, B., 2013. Structure and dynamics of core/periphery networks. Journal of Complex Networks 1, 93-123.
Cummings, J.N., Cross, R., 2003. Structural properties of work groups and their consequences for performance. Social Networks 25, 197-210.
Della Rossa, F., Dercole, F., Piccardi, C., 2013. Profiling core-periphery network structure by random walkers. Scientific Reports 3, 1467.
Delre, S.A., Jager, W., Janssen, M.A., 2007. Diffusion dynamics in small-world networks with heterogeneous consumers. Computational and Mathematical Organization Theory 13, 185-202.
Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F., 2006. K-core organization of complex networks. Physical review letters 96, 040601.

Frahm, K.M., Shepelyansky, D.L., 2012. Google matrix of Twitter. The European physical journal B 85, 355.
Gabielkov, M., Rao, A., Legout, A., 2014. Studying Social Networks at Scale: Macroscopic Anatomy of the Twitter Social Graph, ACM Sigmetrics 2014, Austin, United States.
Gibbons, D.E., 2007. Interorganizational Network Structures and Diffusion of Information Through a Health System. American journal of public health 97, 1684-1692.
Giuliani, E., 2005. The structure of cluster knowledge networks: uneven and selective, not pervasive and collective, DRUID Tenth Anniversary Summer Conference, Copenhagen, pp. 27-29.
Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M., Eguiluz, V.M., 2012. Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media. PLoS ONE 7, e29358.
Hermans, F., Klerkx, L., Roep, D., 2015. Structural Conditions for Collaboration and Learning in Innovation Networks: Using an Innovation System Performance Lens to Analyse Agricultural Knowledge Systems. The Journal of Agricultural Education and Extension 21, 35-54.
Holme, P., 2005. Core-periphery organization of complex networks. Physical Review E 72, 046111.
Klerkx, L., Álvarez, R., Campusano, R., 2015. The emergence and functioning of innovation intermediaries in maturing innovation systems: the case of Chile. Innovation and Development 5, 73-91.
Klerkx, L., Proctor, A., 2013. Beyond fragmentation and disconnect: Networks for knowledge exchange in the English land management advisory system. Land Use Policy 30, 13-24.
Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter, a social network or a news media?, 19th International Conference on World Wide Web. ACM, New York, NY, USA, pp. 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., Hillis, V., Hoffman, M., 2011. Innovation, cooperation, and the perceived benefits and costs of sustainable agriculture practices. Ecology and Society 16, 23.
Lubell, M., Niles, M., Hoffman, M., 2014. Extension 3.0: Managing Agricultural Knowledge Systems in the Network Age. Society & Natural Resources 27, 1089-1103.

Mantzaris, A.V., 2014. Uncovering nodes that spread information between communities in social networks. EPJ Data Science 3, 26.

Myers, S.A., Sharma, A., Gupta, P., Lin, J., 2014. Information network or social network? The structure of the twitter follow graph, 23rd International Conference on World Wide Web. ACM, Seoul, Korea, pp. 493-498.

Newman, M., 2000. Models of the Small World. Journal of Statistical Physics 101, 819-841.

Newman, M., 2010. Networks: an introduction. Oxford University Press, Oxford.

Newman, M., Park, J., 2003. Why social networks are different from other types of networks. Physical Review E 68, 036122.

Piccardi, C., 2011. Finding and testing network communities by lumped Markov chains. PLoS ONE 6, e27028.

Pons, P., Latapy, M., 2005. Computing communities in large networks using random walks, in: Yolum, P., Gungor, T., Gurgen, F., Ozturan, C. (Eds.), International Symposium on Computer and Information Sciences. Springer, Berlin, pp. 284-293.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technology 2, 37-63.

R Core Team, 2014. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 3.0.3 ed. CRAN, Vienna, Austria.

Rhoades, E., Aue, K., 2010. Social agriculture: Adoption of social media by agricultural editors and broadcasters, Proceedings of the SAAS Agricultural Communications Section Annual Meeting.

Rice, R.E., 1987. Computer-Mediated Communication and Organizational Innovation. Journal of Communication 37, 65-94.

Rogers, E.M., 2010. Diffusion of innovations. Simon and Schuster, New York.

Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J., 2014. Core-periphery structure in networks. SIAM Journal on Applied mathematics 74, 167-190.

Romero, D., Meeder, B., Kleinberg, J., 2011. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter, 20th International Conference on World Wide Web, Hyderabad, India, pp. 695-704.

Ryan, B., Gross, N.C., 1943. The diffusion of hybrid seed corn in two Iowa communities. Rural sociology 8, 15.

Schon, D.A., 1971. Beyond the stable state: Public and private learning in a changing society. W. W. Norton & Company, New York.
Shore, J., Baek, J., Dellarocas, C., 2016. Network structure and patterns of information diversity on Twitter, SSRN.
Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 3940-3941.
Takaguchi, T., Maehara, T., Toyoda, M., Kawarabayashi, K.-i., 2014. Existence of outsiders as a characteristic of online communication networks, arXiv.
USDA, 2015. Farm Computer Usage and Ownership, in: Agriculture, U.S.D.o. (Ed.), National Agricultural Statistics Service.
Valente, T.W., 1995. Network models of the diffusion of innovations. Hampton Press, Cresskill, NJ.
Valente, T.W., 1996. Social network thresholds in the diffusion of innovations. Social Networks 18, 69-89.
Valente, T.W., Rogers, E., 1995. The Origins and Development of the Diffusion of Innovations Paradigm as an Example of Scientific Growth. Science Communication 16, 242-273.
Van den Ban, A.W., Hawkins, H.S., 1996. Agricultural extension. Blackwell Science, Oxford.
Verma, T., Russmann, F., Araújo, N.A.M., Nagler, J., Herrmann, H.J., 2016. Emergence of core–peripheries in networks. Nature communications 7, 10441.
Volkovich, Y., Scellato, S., Laniado, D., Mascolo, C., Kaltenbrunner, A., 2012. The Length of Bridge Ties: Structural and Geographic Properties of Online Social Interactions, 6th International AAAI Conference on Weblogs and Social Media. AAAI, Dublin, pp. 346-353.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. Nature 393, 440-442.
Wu, S., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Who Says What to Whom on Twitter, 20th international conference on World Wide Web. ACM, New York, pp. 705-714.
Zhang, X., Martin, T., Newman, M., 2015. Identification of core-periphery structure in networks. Physical Review E 91, 032803.
Žnidaršič, A., Ferligoj, A., Doreian, P., 2017. Actor non-response in valued social networks: The impact of different non-response treatments on the stability of blockmodels. Social Networks 48, 46-56.