UC DAVIS
UNIVERSITY OF CALIFORNIA

Center for
Environmental Policy and Behavior
CEPB

Research Brief

The Structure of Twitter Networks for California Agriculture

Written by Marco Bastos, Mark Lubell, Neil McRoberts, and Michael Levy

Funded by the University of California at Davis

Issue

Modern agricultural systems are experiencing a revolution in how information is disseminated and exchanged among networks of outreach professionals, farmers, consumers, and community groups. These stakeholders are increasingly experimenting with social media and other new information communication technologies to share agricultural information across local, national, and global networks. The possibilities opened by these new channels of information diffusion potentially complement traditional agricultural extension programs that rely on in-person communication and established publication outlets. Social media can potentially reach more individuals from diverse perspectives and at more remote geographic locations. For a local context like California, social media can help import new knowledge from other areas, but also export ideas and expertise from California to other parts of the world.

To provide insight into these ideas, this research brief focuses on a group of Twitter users centered on California agriculture. Twitter is one of the most common social media platforms used by agricultural outreach and extension professionals, and is conceptualized as an information distribution network. Starting with a seed list of 153 users curated by the University of California's Division of Agriculture and Natural Resources (UCANR), we constructed a database the followers and followees of these accounts and amassed a population of 59K users and 68 million tweets. These users represent the Twitter users centered on UCANR, including advisors, specialists, staff, and programs, as well as the University of California's agriculture faculty, programs, centers, and staff. The 59k nodes in our network include anyone who was retweeted or mentioned by any of these individuals.

We subsequently sampled this large dataset and mapped users' interactions by drawing a link from each node (user) that retweeted or mentioned another user. When a user retweets another user, the link represents information flowing from the author of the original tweet and being forwarded by the user that performs the retweet. When a user mentions another user, the link represents the author of the tweet sending information to the user that is targeted by the @-mention. The resulting network was analyzed using social network analysis and "big data" computational methods to calculate the geographic reach of this population and the topical sub-groups around which communication occurs.

Key Findings

Based on the structure of their interactions, the network graph is divided into 10 self-organized, smaller communities. The communities are relatively independent and focused on subfields of agricultural knowledge with sparse interaction between communities. Based on the most common hashtags in each community, we assigned them the following labels: climate, food, water,

agriculture, plant sciences, politics, international development, viticulture, gardening, and animal welfare. Each community has a set of frequently used hashtags, which defined the topical ideas and can be included in social media messages tailored to particular audiences. Each community features a diverse set of actors—government agencies, NGOs, major media outlets, individuals, and research institutions—that play central roles in the network in terms of being the target or the source of most @-mentions and retweets.

Geography plays a crucial role in the entire network as well as within each community. For the overall network, there appears to be a one-way flow of information with West Coast users sending mentions to important East Coast users like government agencies and major media outlets, and East Coast users' tweets being retweeted by West Coast users. Inversely, there is not as much activity when West Coast users are the source of messages retweeted by East Coast users, nor when East Coast users mention West Coast users. This suggests that West Coast users mainly trying to get the attention (via @mentions) of East Coast users, and then forwarding information (via retweets) that originates in the East Coast. The communities differ in the geographic patterns of their activity ranging from statewide to global geographic coverage. Some communities such as water are focused within particular states, while others cross multi-state regions. Some communities feature bi-coastal communication (East and West coast), covering distances around 2,500 miles, while the community focused on international development has a distinctively global reach with interaction ranging from 2,500 to 9,300 miles.

Policy and Management Implications

The topical and geographic structure of the Twitter network associated with California agriculture offers some insights on how social media can potentially foster more effective outreach and education strategies. Together with social engineering aimed at behavior change, the core goal of agricultural extension is to deliver information to the right people, in the right place, and at the right time therefore assisting with the decision-making process. It is clear that Twitter and other online social networks are substantially increasing the social and geographic reach of agricultural and environmental research and other information. Outreach professionals should feel comfortable using social media strategies, and the value of an effective social media strategy should be recognized in personnel decisions.

The topical focus of Twitter communities adds depth to some of the standard strategies for increasing the reach of information delivered in social media. Social media optimization companies, based on findings of social media research, have typically asserted that tweets are more effective when they contain recognized hashtags, mention central users, include a web link (URL), and a picture. Frequently used hashtags and central users allow communities to customize their focus on specific topics. Outreach professionals should pay attention to these strategies. As shown in Figure 1, we found that in some communities the use of hashtags could be more strategic in terms of consistently using a smaller set of hashtags rather than resorting to a large set of tags that end up dividing the public attention among many different combinations.

The communication flows between the East and West Coast may be particularly important for California agriculture and environment. On the positive side, it shows that social media does provide a pipeline for information to flow from California to important policy-makers and media outlets in the East Coast, where policy decisions are made, that together with market forces, shape the overall food system at a national level and with important consequences for California agriculture. On a more negative side, but perhaps because of our sampling methods based on California users, Twitter accounts in the West Coast are mainly mentioning the East Coast users in an attempt to get their attention, and the retweet flow of information is mainly from East to

West Coast users. Developing more "conversations" by means of interactions, where East Coast users mention and retweet West Coast users, can potentially establish a two-way flow of information and catalyze higher integration into individual decisions.

Methodology

The seed users comprise 153 Twitter users monitored by the UCANR Twitter account. We identified users following or being followed by the seed users and retrieved the messages tweeted by each Twitter account from 2007-2015, thus rendering a population of 59,761 users and over 68M tweets. We subsequently removed messages directed to or originated from users outside the target population of 59K users, which defines the boundaries of our community, and restricted our analysis to tweets posted in 2014. After implementing these sampling procedures, we ended up with a network of 32,152 nodes (users) and 4,418,390 links (2,502,107 @-mentions and 1,916,283 retweets) posted between the first and the last day of 2014.

For the purposes of this study, we designed two classifiers to identify and rank all tweets posted in 2014 based on their relative importance to agriculture and sustainability. 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 messages and profiles of each of the individuals in the population to identify instances of information sharing. As a result, we processed the text of each tweet and identified senders and receivers by means of retweet or @-mentions, thus creating a link between the author of the tweet and every other account mentioned or retweeted.

Communities were identified by relying on the walktrap algorithm that classifies users into modules or communities according to the links they share. The network was thus divided into 10 large communities that account for 80% of the graph, 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 communities broadly associated with agriculture. Lastly we analyzed the information on users' profiles and the tweets posted in the period to identify the geographic location of our population. We managed to identify the location of 73% of the complete population of users (39,858 from 54,422 profiles), with twenty percent of them based in California (12,058 users), which is unsurprising given the location of the seed nodes.

Results

Figure 1 shows the most frequent hashtags used for each community in 2014. The x-axis shows the periods associated with the use of the hashtags, ranging from January to December 2014. The y-axis provides a visual description of the ebb and flow of the volume of tweets containing each hashtag. There are many infrequently used hashtags not shown on the figure. Examining the hashtags provides the basis for an overall label for each community, which can be thought of like a "channel" on television. Each community is defined by a small set of hashtags that concentrate most of the discussion within that community. Some of the communities, viticulture and gardening, have a very dominant hashtag that provides a central coordination point for communication.

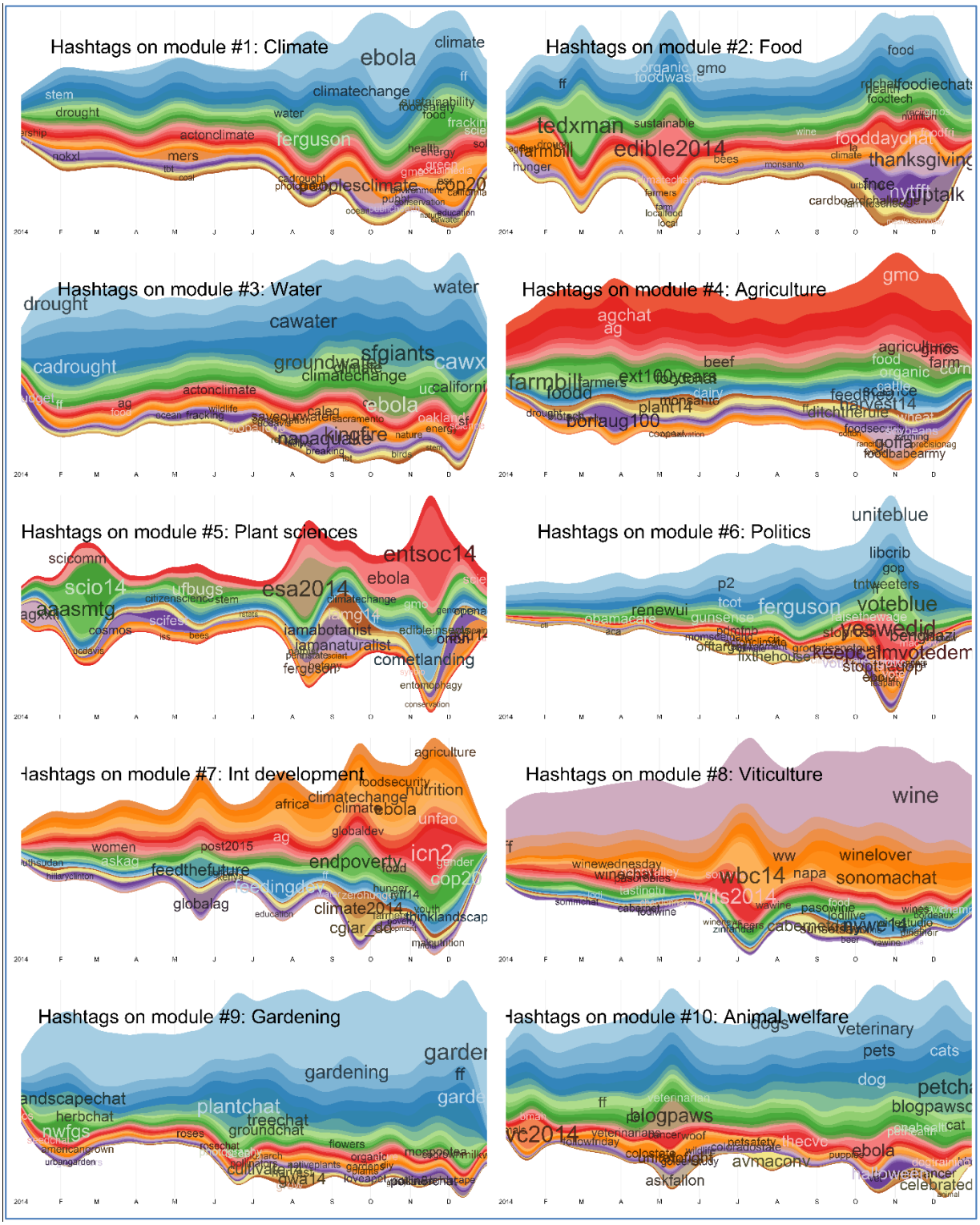

Figure 1: Top hashtags used in community from January to December 2014

Figure 2 shows the most central US users within each community based on their centrality within the network, as measured by the number of mentions and tweets to which they are linked (x-axis). Formally, this is called "total degree" centrality in the language of network analysis. The central users are a diverse mix of major news media (e.g. New York Times and Los Angeles Times); individual experts (e.g. Peter Gleick and Tracy Blevins); government agencies (e.g. EPA Water and USDA); non-governmental agencies (e.g.; American Rivers and Napa Valley Grape Growers), and research groups (e.g. Princeton Water and UCANR). From the perspective of UCANR, it is heartening to see those accounts with a high position in the water and agriculture communities. It is also important to note that despite dominant hashtags within these communities, they also have important intersections with the broader culture in terms of both hashtags and users. For example, Jerry Brown is part of the water community but obviously has links to many to many other issues.

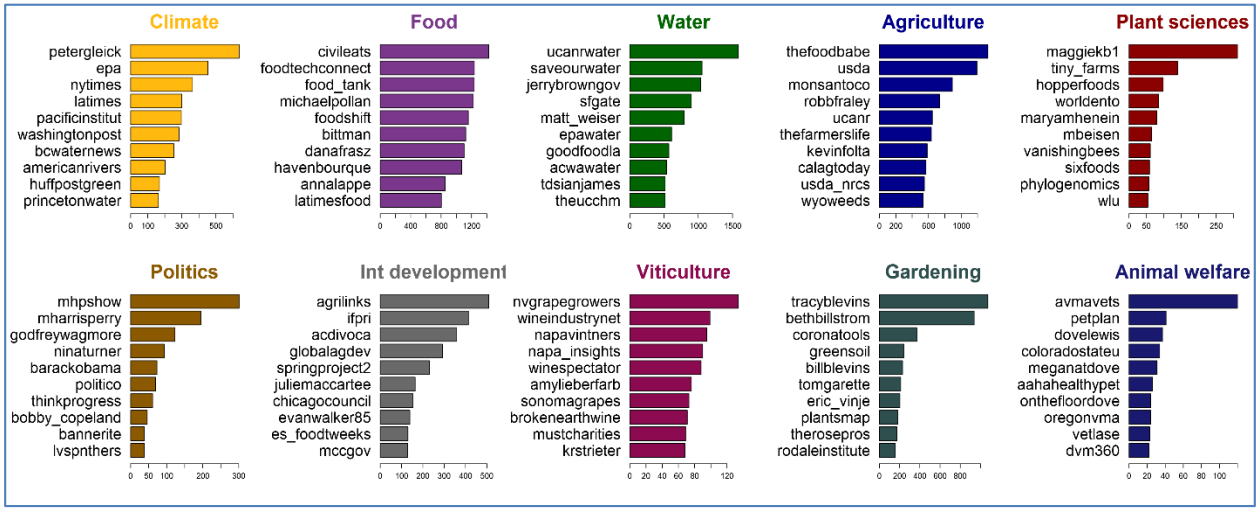

Figure 2: Most central users across channels (measured by the number of tweets to which they are linked)

Figure 3 shows the directionality of @-mentions and retweets between the East and West Coast for all messages in 2014. The retweets (in blue) follow a clockwise direction, indicating West Coast users retweeting posts originating in East Coast metropolitan areas. The mentions (in red) indicate tweets originating in the West Coast that are directed to users in the East Coast. Overall, this figure suggests that West Coast users are actively exporting information to the East Coast, with East Coast users being the object of @-mentions from users in the West Coast. There are not as many two-way conversations. Influential accounts are usually the target of many @-mentions and do not @-mention other accounts themselves. The same imbalance is observed with the retweet distribution: influential accounts display high outdegree and low indegree for retweets (meaning they are retweeted but do not retweet many users themselves). 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.

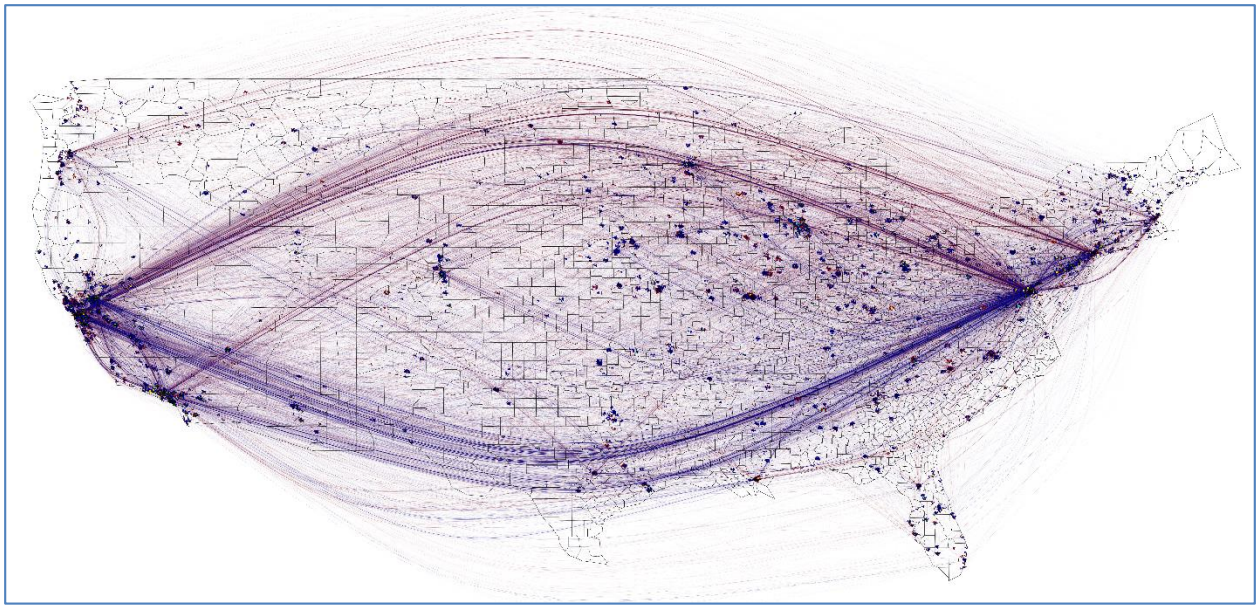

Figure 3: Directionality of @-mentions and retweets from West to East Coast and back again

The maps shown in Figure 4 present in detail the different geographic patterns for each community. All of the communities have a somewhat of a bi-coastal pattern reflecting the distribution of major metropolitan areas in the country. Some communities have unique geographic patterns related to other economic, social and environmental factors. For example, the agriculture community has high density of activity in the Midwest, which is consistent with the high levels of agricultural production and crop yields in the area. Lastly, Figure 5 shows the geographic patterns observed in each of the communities during 2014, both in the United States and beyond, with distinctive patterns of information exchange associated with international development, plant sciences, and climate in terms of local and global reach. By identifying the location of users sending and receiving information, we managed to classify the information exchange in geographic areas according to the regional divisions used by the United States Census Bureau. For example, water and viticulture are mostly concentrated locally within California given the importance of water management and the wine industry in the state. In short, the geographic patterns observed in each of the communities are linked to the topics under discussion; with water being an intrinsically local issue and international development operating globally (see Figure 5).

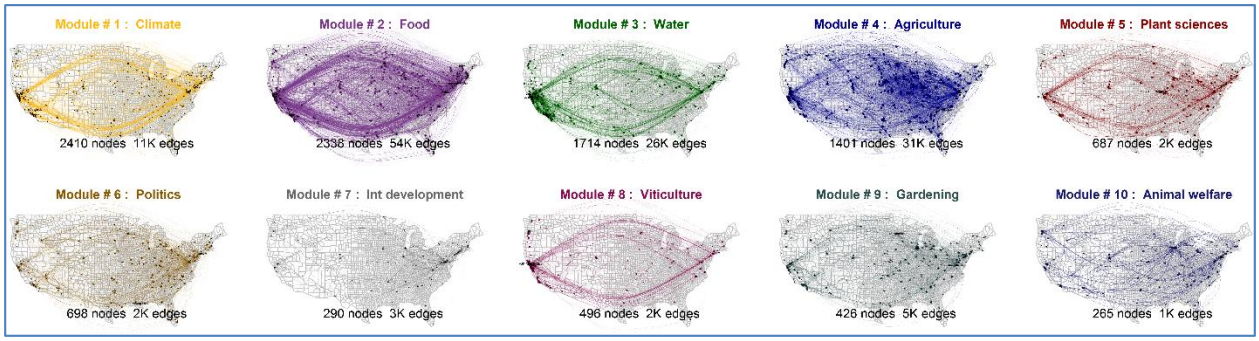

Figure 4: Retweets and @-mentions associated with agriculture across the communities in the US

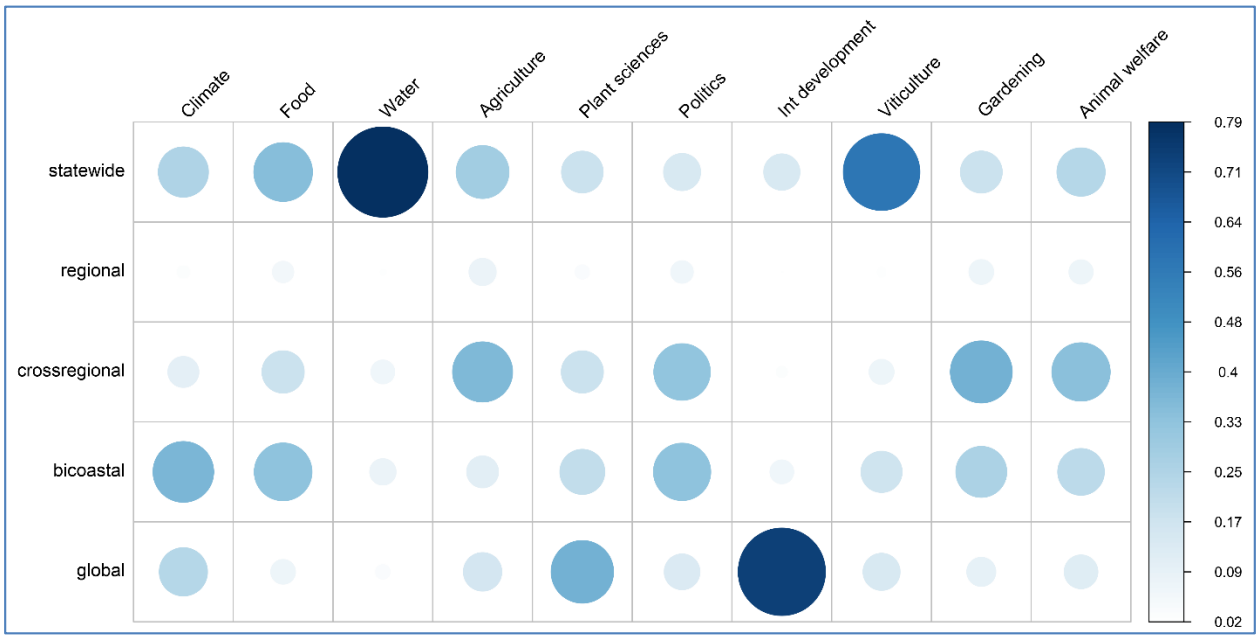

Figure 5: Proportion of Twitter messages occurring in different geographic scales per community

New Research Directions

This brief provides a basic description of the topics and geography of the Twitter network associated with agriculture in California. The next round of analysis will more deeply examine the structure, function, and processes of information flow that are occurring on Twitter. We expect that different hashtags will have distinct temporal dynamics within each channel, representing the tendency of a hashtag topic to maintain a constant rate of discussion versus having a tendency for more "bursty" behavior swinging between attention and inattention. We hypothesize that as communities become more specialized around particular topics, they will exhibit a "core-periphery" structure where a few clear thought leaders dominate information exchange. We also intend to provide a deeper analysis of the national and international geographic structure of the networks to predict if links are more likely to occur between metropolitan areas, to follow trade patterns, or to match regions with similar agricultural and economic activities (homophily). From an applied perspective, we hope that by understanding the information processes that occur in these networks, we can provide recommendations to outreach and education professionals about how to use information and communication technology to deliver knowledge to the right people, in the right place, and at the right time.

Publications and Presentations

Bastos, M. T., Piccardi, C. & Lubell, M. (2016). *Core-periphery dynamics in thematic subnetworks: a case study of social media adoption by outreach professionals.* XXXVI Sunbelt Conference of the International Network for Social Network Analysis (INSNA), Newport Beach, CA.

Bastos, M. T., & Lubell, M. (2015). *Networking Aggie: Broadcasting Information to Topical Communities.* Internet Research 16: Imagining Audiences, Phoenix, AZ.

For more information contact:
Center for Environmental Policy and Behavior
University of California, Davis
http://environmentalpolicy.ucdavis.edu/, 530-752-5880