Semantic Analysis of Energy-Related Conversations in Social Media:

A Twitter Case Study*

Martha G. Russell[1, 2], June Flora[1], Markus Strohmaier[3], Jan Pöschko[3], Rafael Perez[4], and Neil Rubens[4]

ABSTRACT

This paper describes the initial phase of a study of ecolinguistic-based social media analytics aimed at understanding the frequency, semantics, context and potential persuasive influence of social media conversations about energy issues, metaphors, frames and behaviors. Our broad research question asked, "How does the online conversation about energy efficiency behavior change overtime?" We operationalized conversations to be overtime mentions of ecolinguistic terms in Twitter. We conducted a preliminary analysis of the messages, users and content hashtags in tweets over 4 months. Illustrative results demonstrate new tools for data acquisition, curation and analysis. They demonstrate an initial concept proof of the tweetonomy construct and provide preliminary network analysis of energy terms in twitter streams. Results indicate opportunities for using time series analysis to understand the rhythm of the social conversation to provide insights about when people are reachable for persuasive communications . They indicate opportunities for text analysis on Twitter content to understand how to improve relevance of energy efficiency communications. They demonstrate user-friendly tools to visualize the semantics of vast quantities of user-generated content about energy and sustainability in order to better understand contextual issues that define issue publics and frame positions. Preliminary results from this exploratory study confirm the feasibility of using Twitter streams to detect awareness, describe attitudes and infer influence about communication campaigns intended to persuade changes in energy behavior.

General Terms

energy behavior, semantic analysis, Twitter, data mining, user generated content, network analysis, sustainability, energy efficiency, persuasion, social media, ecolinguistic terms, tweetonomy, time series

1. INTRODUCTION

This paper describes a large scale social media data collection and research program focused on energy efficiency and climate change. The ultimate goal of this two year program of work is to refine an energy and climate change taxonomy of terms; develop analytics that are robust in the face of an extensive data-base; conduct iterative tests of simple and complex analyses (frequency of mentions overtime; connections to societal events, social networks and building, and maintenance of online energy related issue communities). The final goal of the program is to make the analytic tools and data base available to energy-related scholars as well as others for the advancement of a social movement of environmental sustainability.

People influence others and are, in turn, influenced with respect to the actions, brands, products and issues they love and hate. This influence promotes awareness, motivates trial behaviors, reinforces purchase decisions, and sustains product or action loyalty. Large scale social movements such as energy efficiency and sustainability require an initial articulate community with a shared belief in order to create virality. Word of mouth, online or in person, is essential to break down barriers, to spread ideas and actions, and to develop innovations. In fact, the ripple effect of word of mouth is estimated to have three times the reach and impact of traditional advertising (Hogan, Lemon, Libai, 2004). Diffusion of information underpins the success of campaigns to persuade change in behavioral actions and adoption of a sustainable lifestyle to reduce greenhouse gas emissions.

The term social media describes the online tools and platforms that people use to share opinions, insights, experiences, and perspectives with each other. Social media can take many different forms, including text, images, audio, and video. Understanding conversations in online social media

*This research was funded in part by the Department of Energy (DOE), through an ARPA-e research cooperative agreement to Stanford University, Byron Reeves principal investigator, focused on human behavior change via feedback from an array of energy monitoring sensors (both high and low resolution sensors) on online coverage of targeted behavioral descriptions, technologies, and polices. Supported in part by Special Coordination Funds for Promoting Science and Technology, the Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT), Knowledge Systems (Okamoto) Laboratory, the Austrian FWF research grant P20269, and the Know-Center Graz.

[1] Stanford University
[2] Media X
[3] Graz University of Technology, Austria
[4] University of Electro-Communications, Japan

has the potential of providing program planners and communication campaign mangers unique insights into individuals' thoughts and verbal productions about energy efficiency and climate change. With the increasing adoption of social media (73% of teens and 72% of young adults - Pew, 2010), new opportunities are available for studying the role of online conversation in persuasion.

In 140 characters or less, the concerns, interests and public narratives about energy efficiency and climate change are easily identified. Twitter conversations also reflect the social aspects of information diffusion through conventions such as retweets and other conversational responses, through the membership and social distance of these conversations, and in overtime changes in frequency and form of the networks. With Twitter, a particularly popular type of social media that has proven relevant in a number of societal challenges and conversations recently (such as the 2010/2011 revolutions in the middle east), the social response to societal or national events, as well as to media coverage of these events, and persuasive communication campaigns can be observed. Twitter is both a medium and the message (Savage, 2011).

This paper outlines the broad aims of an ongoing research program to track and analyze social media conversations of over one hundred terms related to energy efficiency and climate change for insights about consumer attitudes and behavior. It describes in detail the first assessment of the project, in which we have begun with an exploratory analysis to investigate the extent to which Twitter conversations about energy consumption can be tracked and analyzed using a social media analysis framework for studying energy consumption descriptors, concerns, actions, and sources of information. Preliminary results from this initial phase of research confirmed the existence of energy-related conversations on Twitter, and the usefulness of Twitter streams to assess awareness of communication campaigns, brands, products and policies related to creating communities of awareness, interest, and action regarding energy behavior. We examined the periodicity, content and context of the this conversation.

This arena of scholarship is essentially interdisciplinary. Our research team leverages expertise from multiple fields of scholarship; climate science, communication, systems analysis regarding marketing and advertising, and computer science (more specifically machine learning, natural language processing and network analysis). In our view such collaborative work adds value to individual disciplines as well as provides actionable insights into the international concern of advancing energy efficiency and slowing climate change. While we find this interdisciplinary approach to studying energy-related conversations on Twitter fruitful and necessary, we are well aware of the challenges related to gauging and evaluating the research output of such an endeavor.

This paper does not aim to provide any definitive answers to the questions of whether, how and to what extent public conversations on energy-related topics take place on Twitter. It does, however, initiate a research agenda on analyzing the semantics of conversations in social media regarding energy and climate change - hoping to spur further research in this domain.

2. WHO TWEETS, ABOUT WHAT AND WITH WHAT IMPACT?

Twitter like other social media has been assumed to start with teens and move to other audiences. But Twitter has taken a different turn:

"Another reason that teenagers do not use Twitter may be that their lives tend to revolve around their friends. Though Twitter's founders originally conceived of the site as a way to stay in touch with acquaintances, it turns out that it is better for broadcasting ideas or questions and answers to the outside world or for marketing a product. It is also useful for marketing the person doing the tweeting, a need few teenagers are attuned to." Claire Can Miller, New York Times, Aug 2009.

Content analyses and user interviews reveal that Twitter is not focused on personal social issues but has a stronger emphasis on ideas. Recent data from Pew's ongoing media tracking studies (Pew 2010) reinforces that Twitter users are are demographically diverse. Of the 72% of adult online media users, 8% are twitter users. More specifically:

- Young adults – Internet users ages 18-29 are significantly more likely to use Twitter than older adults. Also, these voting age adults can significantly influence community and societal policy as well as opinion regarding climate change.

- African-Americans and Latinos – Minority Internet users are more than twice as likely to use Twitter as are caucasian Internet users. The unique broad reach into this diverse audience allows the rapid diffusion of educational and vocational opportunities thought to bring about economic recovery in the U.S.

- Urbanites – Urban residents are roughly twice as likely to use Twitter as rural dwellers.

Further, Twitter users are nearly equally divided between those who check the site on a daily basis (or multiple times per day) and those who check the site infrequently or never. Just over one-third of Twitter users (36%) check for material posted by others on a daily basis or multiple times per day—this is roughly comparable to users' engagement frequency. Two in five (41%) say they check the site less than every few weeks, or never do so at all. The remaining one-quarter of users say they check the site for updates a few days each week or every few weeks (Pew, 2010).

Social media have been widely used for consumer goods and services. Market leaders such as Coca-Cola and Proctor & Gamble, specialty product companies such as Snap-on Tools, and broadcast channels such as CNN and MSNBC have incorporated Twitter into the social media into the media channels they use proactively – by creating communication campaigns that include social media for their consumer communities. In fact the 2011 Superbowl game featured the first ever Twitter hashtag to be included in a television commercial as a call-to-action (Audi, 2011). A distinguishing feature of social media is that, while it may be initiated or catalyzed by communications in other channels, it is inherently consumer-generated content and its virality depends on engagement Even those companies that do not actively sponsor social media have tapped into social media to monitor the nature and strength of consumer sentiment about their brands, products, and services (Wasserman, 2011).

Business metrics for social media measurement include subscriptions, traffic, and content and user profiles. The

speed, breadth, depth, and rhythm of social media communications create a frame, or window, through which the volume, sentiment and dispersion of consumer sentiment can be tracked These metrics are considered early indicators and provide quick feedback on consumer response. Media effectiveness analysis typically focuses on a single metric – the number of click throughs to view on YouTube, the number of friends on Facebook, the number of connections on LinkedIn, the number of followers on Twitter, the number of click throughs on blogs, the number of registrations on podcast downloads. Newer metrics that more accurately reflect the multidimensional nature of engagement are needed.

For example, early metrics based on buzz have proven useful for non-invasively monitoring consumer opinions and behavior through social media. Marketers have used metrics to guide their direct and indirect deployment of communications for awareness, trial, conversion, and loyalty campaigns (Russell, 2010). However, the known persuasive influences of communities and key opinion leaders have led to the development of revolutionary assessment techniques for diagnosing the health of a social media community and for identifying social roles in social media (Wu, 2009.) The competitive value of metrics lies in being able to make early diagnosis for swift interventions in either sponsored or ad hoc social media – in order to achieve the most favorable results in influencing consumers' attention and behaviors – and in being able to marry traditional marketing, advertising and promotional strategies and tactics with the viral potential of word-of-mouth influence through social media.

Twitter use is growing in prevalence among market leaders and innovators, and it reaches audiences essential to a social movement of sustainable lifestyles. Its use encompasses strong and weak personal and social ties; and it reaches into previously untapped but affected audiences. Perhaps most importantly Twitter content is issue laden. Thus, systematic access to data and conversations on Twitter allows us to study public conversations on a new scale.

3. RESEARCH APPROACH AND SETUP

Based on these observations, we selected public conversations and dialog on Twitter as a relevant phenomenon for our research on energy behavior change. To validate the decision, we asked whether, how and to what extent public conversations on energy-related topics take place on Twitter.

In the first phase of our research program our broad research question asks, "How do online conversations about energy efficiency behavior change overtime?" We developed an ecolinguistic taxonomy of terms to describe energy use opinions, energy efficiency behaviors, frames, metaphors, technologies and to determine standard sources of energy information, such as the Department of Energy (DOE) and Environmental Protection Agency (EPA). We captured Tweets containing those terms, parsed identified elements of the communications, curated the data, and archived the data for access. We conducted a preliminary analysis of the messages, users and content of the 18,338 hashtags that occurred at least 3 times in the 2.47 million tweets collected between September 3, 2010 and January 4, 2011. The analysis of this ecolinguistic-filtered social media was aimed at understanding the frequency, context and potential persuasive influence of social media conversations about changing energy behaviors. We operationalized conversations to be overtime mentions of our predetermined ecolinguistic terms and co-occurrence of related hashtags in Twitter. Three techniques were used to cull discernible patterns from the large quantities of data and portray the data visually: content analysis of the full Tweets, network analysis of co-occurring hashtags, and semantic analysis of the co-occurring hashtags and their authors.

3.1 The role of social media conversations in social movements

There is a growing academic literature on the definition and measurement of persuasion, and it includes social media conversations. Studies of argumentation pathways acknowledge the influence of source, attitudes, information processing, behavior, mental shortcuts, and sustaining reinforcers (Kaptein et al., 2010). New insights are continuing to advance our understanding of the complex interactions between the cognitive (resonance – "get it" – speaks to me), emotional (totally immersed, absorbed, the opposite of indifference), social (interactive, participative and involved), and temporal (longevity, a commitment to the future, seeing a long term relationship) factors in interactive persuasion. Continuous partial attention characterizes the shift of attention from one media to another in consumers' day-to-day multitasking and multimedia environments (Block and Schultz, 2009; Ophir et al., 2009.) Programs focused on persuasive interactions address the management of attention to obtain total engagement (Reeves and Read, 2009) in persuasive interactions.

The predominant one-to-many model of mass communication campaigns has undergone significant change with the proliferation of new channels and the advent of unstructured communication (Daft, Lengel & Trevino, 1987), the declining audience of traditional channels (Livingstone, 2004) and the broad participation of users in generating content for single and mass audiences (Fogg, 2008). Effective messages must be increasingly targeted to smaller audiences including the audience of one. The personalization of communications to an expanding number of individuals has depended on advances in several fields, including artificial intelligence, machine learning, psychology, sociology and communication (Nass and Yen, 2010).

Further, the expansion of undirected and user controlled communication forums, such as Twitter, offer the possibility of a new understanding of "grass roots" social change. Thus, both directed programs (such as socially motivated games) as well as less well defined and even spontaneous programs (such as a new YouTube video or a new climate calculator mounted on a credible website) have the potential to stimulate user-generated persuasion, in the context of technologically-enhanced word-of-mouth. Along the continuum of intentions to influence how people think and act, persuasion depends on attention, engagement and argumentation. Because it is user-generated, social media includes all three of these elements. Either directly or indirectly, it evokes the presence of, or reference to, other people as social agents (Bailenson et al., 2005).

3.2 Tools for data acquisition, curation and analysis

Data was acquired on a daily basis by utilizing the NodeXL Twitter Importer module (Smith et al., 2009), which captured the latest messages containing energy related keywords (Table 1). In addition to the message itself, acquired

data includes username of the message sender, along with the time and date of the message. The dataset was then parsed and analyzed by utilizing Hadoop Map Reduce distributed processing on the Amazon's EC2 computing cloud. By utilizing 20 standard computing instances (~1 GHz, 1.7 GB memory), the complete dataset could be processed in under 1 hour. The data processing selected Tweets collected between 2010.09.03 and 2011.01.04, containing hashtags that occurred in at least 3 of the 2.47 million tweets. Data for these Tweets was then passed to other applications - Excel, NodeXL, Gephi and TwitterExplorer - for further analysis and visualization.

3.3 Using energy-related terms to filter Twitter messages

A critical preliminary task in this research program was to create a taxonomy that describes the target behaviors, technologies, programs, information sources, issues, issue frames metaphors, and visualizations related to a social movement for the development of sustainable lifestyles. Based on literature on lexical framing of climate change, metaphors from analyses of climate change journalism, risk communication and public discourse, we constructed a taxonomy that included positive and negative valanced frames, compound metaphors, frames that encompass moral imperative, values, and social movement frames, as well as energy reduction, conservation, and energy wasting behaviors (Lakoff, 2010). Tweets that contained these terms were selected into our data set.

While we used the conceptual underpinnings from the growing field of ecolinguistics and home fields of rhetoric and linguistics, the actual wordings of behaviors and topics were standardized to several credible sources. For our current investigation, standardization meant that the search terms for behaviors and topical issues were "exactly" as worded in source documents. If for example, the behavior "take shorter showers" is worded differently in different source documents, all are included. Because there are a myriad of web, NGO, government, and other sources for energy behaviors and terms, we are continuing to standardize to those most closely aligned to USA national policy. For example, all energy related phases are derived from energy efficiency and renewable energy website and booklets from the United States Department of Energy (DOE), the Environmental Protection Agency (EPA), the National Academy of Science, and the Environmental Defense Fund. In the future food issues and behaviors and transportation terms will be added and standardized to credible sources.

Table 1 shows the broad set of energy efficiency terms selected for our research program. Included are metaphors of energy efficiency, behavior and feedback as well as issues and four classifications of behavior.

In addition to ARPA-e salient terms, we also included words that were indirectly related on close contextualized issues such as global warming and climate change as well as broader but related social movements such as "simpler lifestyle," "reducing consumerism." etc. Finally, not included fully in this table, but considered essential to the complete set of terms relevant to the energy efficiency behavior change, are behaviors related to broader issues; advocacy behaviors, food and transportation behaviors. Because many of the terms in the ecolinguistic taxonomy were already in use in common parlance, the creation of this taxonomy will allow an observation over time of changes in the use of the terms. The complete taxonomy of terms, their selection, collection, and specificity as well as standardization to know sources are described in more detail in a working white paper for Media X (Flora and Russell, in preparation.)

Term Category	Number	Percent of List
1. Energy technologies/hardware &software (e.g. plug load sensors, smart meters, Power strips, brands such as TED)	19	13%
2. Communication behaviors, e.g. monitor energy bills	4	3%
3. Energy & climate change frames, metaphors, & visualizations	41	28%
4. Energy efficiency and climate change innovative programs (e.g. Carrotmob, 350, Earth hour, Earth day)	6	4%
5. Issues such as renewable energy, global warming, energy insecurity	11	8%
6.. Utilities, Venture firms and companies (all California)	9	6%
7. . Behavior: Low cost high impact	20	14%
8. Behavior: Low cost lower impact	21	15%
9. Behavior: High cost high impact	9	6%
10. Behavior: Higher cost lower impact	4	3%
Subtotal behaviors	**54**	**38%**
Current total	144	

Table 1: Categorization of energy-related terms.

3.4 Analyzing the Conversations

Several analytical lenses were tested in this exploratory study: frequency, periodicity, valence, co-occurrence, and context. Through several methods, we were able to describe snapshots and detect changes over time in the conversation as well as identify conversation stimulating events, such as national policy, new technology launches, and media events.

3.4.1 Frequency of mentions

The frequency of term mentions was counted as the number of Tweets collected by filtering on the ecolinguistic terms. The frequency of September 2010 mentions of a sample of energy terms is shown in Table 2.

3.4.2 Periodicity

To understand the "drivers" of shifts in the conversation, changes in frequency over time were of particular interest. A sampling of tweets on all ecolinguistic terms over a two month period revealed two rhythmic patterns. The social conversation showed a day-of-week rhythm, as well as consistently lower energy Tweets on weekends, during the two month period.

Figure 1 show the daily rhythm of all energy tweets collected in months of September and October 2010. More analysis over longer periods of time is needed to determine if Figure 3 is a baseline of "energy tweets" a growing trend or response to large media coverage of energy related issues. The relatively lower frequency of Tweets over the weekend was noted and prompted an inquiry in the daily patterns.

Figure 2 shows the proportion of Tweets containing the ecolinguistic terms occurring by day-of-week during Septem-

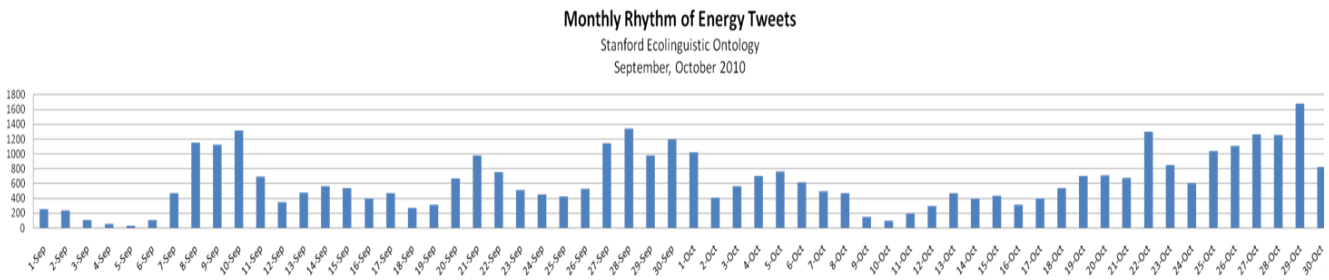


Figure 1: Daily Rhythm of 44,847 Energy Tweets Over Two Months.

Number of Tweets including term September 2010	Energy Term
11,629	"power strips"
21,600	"smart homes"
26,405	"layer up"
39,205	"powermeter" (Google)
40,145	"thermostats" (programmable)
52,287	"remote + sensing"
61,187	"energy + Audit"
81,023	"cfls + lowlight"
81,029	"cfls + bad"
81,232	"cfls + ugly"
83,097	"cfls + mercury"
83,128	"alert devices"
83,617	"black balloons" (visualization from energy campaign in Australia)
113,145	"energy conservation"
233,791	"teds" (the energy detective)
336,741	"Energy Star" (energy efficient government approved appliances)
340,533	"smartmeters" (utility installed hourly signal energy meters)
352,199	"windpower"
374,487	"solarpower"
395,495	"cleantech"
468,647	"led"
615,919	"energy efficiency"

Table 2: Energy Term Test Sample, September, 2010.

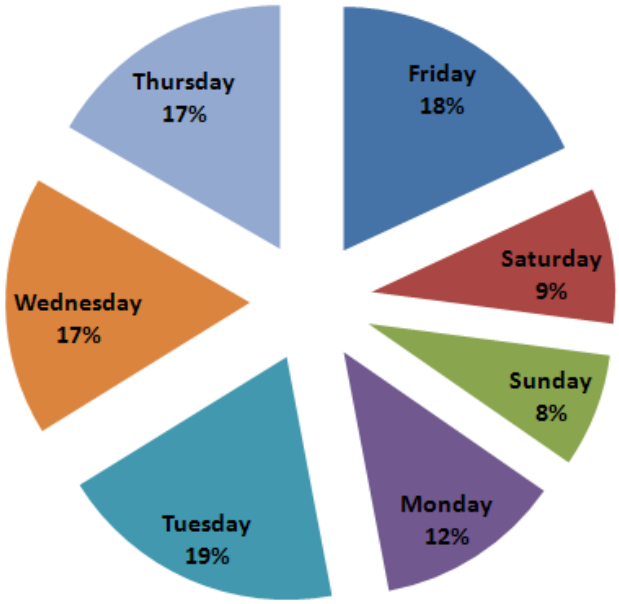


Figure 2: Proportion of 44,000+ Energy Tweets Over Two Months (Sep, Oct, 2010) by Day of Week..

ber and October 2010. In general, fewer Tweets about energy occurred over the weekends and on Mondays, while a higher proportion of tweets occurred between Tuesdays and Fridays. Further analysis would be necessary to see if the content of workday tweets are qualitatively different from weekend tweets.

Closer analysis of one particular term over the two-month period revealed bursts of conversation during two periods. An analysis of the September-October 2010 tweets containing the term "smart meter" shows two periods of time during which a greater number of Tweets about smart meters took place the social media conversation took place.

Word visualizations of the content of "smart + meter" Tweets during the two periods of heavy use revealed the influence of public media on the Twitter conversations. Further analysis of the content of these Tweets indicated reference to a political event and a media-promoted issue. During early September, the content in the "smart + meter" Tweets included many mentions of the announcement of a new rebate program for the installation of "smart + meters," shown in a word cloud in Figure 4. In the 107 Tweets, selected for their mention of "smart + meters" during the period September 8 through 21, 2010, the relative frequency of "Smart," "smart," and "energy" allow visibility of a wide diversity of other terms used in the Tweets. "RT," the convention for retweeting a message, is also visible. An examination of the content of the Tweets, including urls, revealed interest in the announcement of a new Federal rebate program for installation of smart meters and the release of a study about energy efficiency.

The content analysis of the full text of 210 Tweets between October 25 and October 29, 2010 including "smart + meters", is shown in a word cloud Figure 5. Relative to the number of words in the Tweets, the high proportion of the words, "grid," "power," energy," and "security" revealed consumer concerns about the security of personal data and protection from hackers; urls referred to several news stories on these issues. The strength of "RT" indicates the word-of-mouth value of some of the Tweets. The selection criteria for this same included the words, "smart," "meters," and"meter;" and the prominence of those terms is visible.

3.4.3 Valence and Sentiment

Frequent positive or negative word associations with terms can reflect perceived and actual sentiment associated with the terms. Referring back to Table 2 and the September 2010 test of a sample of the data, words with negative as-

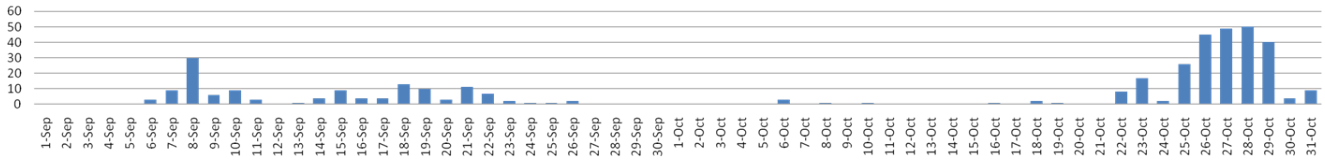


Figure 3: Daily Rhythm of 427 "Smart Meter" Tweets Over Two Months (Sep, Oct, 2010).

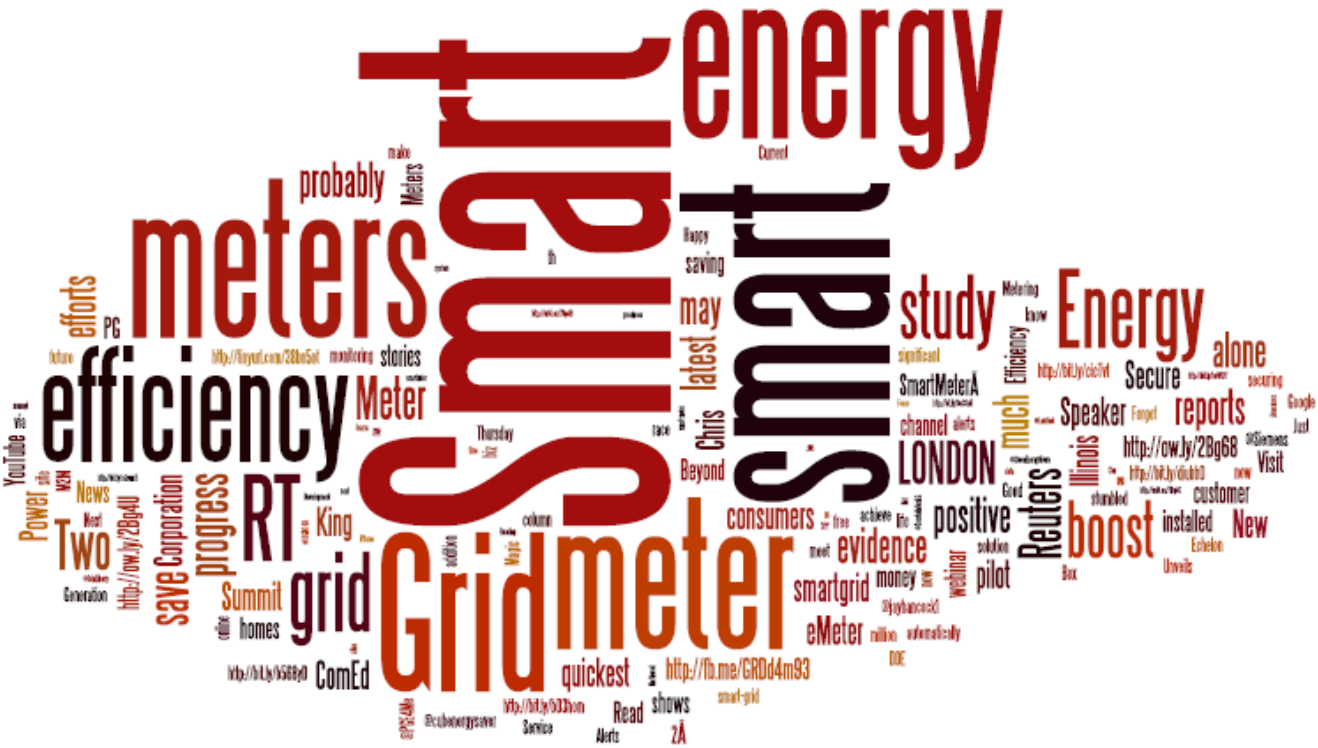


Figure 4: Word Cloud of Full Text of 107 Tweets During September 8 - 21, 2010.

sociations ("ugly" and "bad") were observed in combination with the "cfls", forming compound terms (cfls + . . .) These frequently occurring compound terms reflected time-specific negative attributes (perceived and actual) of compact fluorescent lights (CFLs), a highly touted energy saving replacement for incandescent bulbs.

Sentiment analysis requires highly nuanced interpretations, represents a complex set of evaluations and in other forms of social media indicators such as pronoun use (shared reality and social connection), verb use (particularly emotionally-valenced or recommendation verbs), syntactical references, and style (immediacy – short words, first person pronouns, low number of articles, positive emotion, verbs more than nouns). Sentiment analysis can provides important indicators of engagement, intimacy and social connection to brands, products or services. Linguistic markers and sentiment analysis can be used to diagnose engagement in online conversations and evaluate the customer context (Pennebaker & Niederhoffer, 2003). Because of the contracted lexicon and linguistic conventions of Twitter, new classification schema are needed for sentiment analysis of Twitter messages. At the time of this writing several proprietary, unpublished tools for this are known to exist and are in use by businesses, but none are known to be available as open source for scholarly use.

3.4.4 Context

Context is important in understanding how energy efficient behaviors and actions, such as purchases, align with climate change issues or with comprehensive lifestyle movements such as "low impact lifestyles." This includes the network patterns created by co-occurrence of ARPA-e related energy terms as well as the nature and extent of terms describing related issues such as energy insecurity, climate change, and environmental sustainability. In this preliminary study, context was visualized in the network patterns created by directed communication between Tweeters, using the @Twittername conventions. Context was also operationalized as networks of networks of hashtags - occurring in Tweets selected because they included one or more of the selected ecolinguistic terms. Additionally, context was visualized as a Tweetonomy (Wagner and Strohmaier 2010), using Twitter Explorer (Poeschke and Strohmaier, 2010.)

3.4.5 Directed Communication Between Tweeters

In this paper Twitter messages are viewed as an element

Figure 5: Word Cloud of Full Text of 210 Tweets during October 25-29, 2010.

of conversation in one of three ways. The person sending the message, the Tweeter, is intentionally addressing the message to a known individual, using the convention of @personname (@ replies). The Tweeter is referring a message or partial message that has been received from a known individual (retweets, indicated with the "RT" convention). Or, the Tweeter is broadcasting a message to the broader social media community, as to an imagined individual or group of individuals, sometimes using key words or tags, such as #hashtag. Most Twitter users associate an image with their Twittername, and Figure 8 also uses the Twitter image as the node to visualize the social conversation.

A network of directed communication between Tweeters mentioning smartmeters or "smart + meters" is shown in Figure 8. Each @Twittername creates a directional edge from the author to the intended recipient. Each RTtwittername creates a directional edge from the retweeter to the author. Roughly half of the 107 mentions in this sample of September 2010 Tweets containing #smartmeter hashtags are not directed or connected; the Tweets simply include the hashtag. In the remainder of Tweets, one large network focused on the Twittername, "SFGate," dominates, with several other smaller clusters also evident., all showing a "reply magnet" social role (Gleave et al., 2009). The network patterns show a simple star pattern, in which one node forms the center of the network, without interconnectivity across the other nodes.

3.4.6 Hashtag network visualization

Hashtag is a Twitter linguistic convention constructed by placing a "#" in front of a keyword, abbreviation or acronym. Hashtags provide a means of tagging, labeling, adding context and metadata to Tweets. The use of multiple hashtags in a single Tweet can be interpreted as a connection between the thoughts, terms or issues on the part of the author. We examined the co-occurrence of hashtags in the network analysis of 18,323 hashtags to evaluate the visibility of issues such as climate change, energy efficiency and environmentalism.

A network visualization of the co-occurrence of 18,323 hashtags is shown in Figure 7. Each hashtag is represented by a node; the size of the node is proportional to the connectedness of that node to other nodes. In addition to #energy, other salient nodes with their own clusters of networks are #green, #solar, #climate, #eco and #news. A more distal node with a significant network is the term #jobs. The spatial distance on the network indicates the conversations about energy efficiency is separate from the conversation about cleantech or energy jobs.

The use of graphic images to represent network configurations is important because it allows investigators to gain new insights into the patterning of connections (Freeman, 2009). Information visualization amplifies human comprehension through an expressive view that can stimulate insight on a given phenomena represented by the data.

The network visualization in Figure 7 was generated utilizing the Network Explorer module (Rubens et al., 2011)

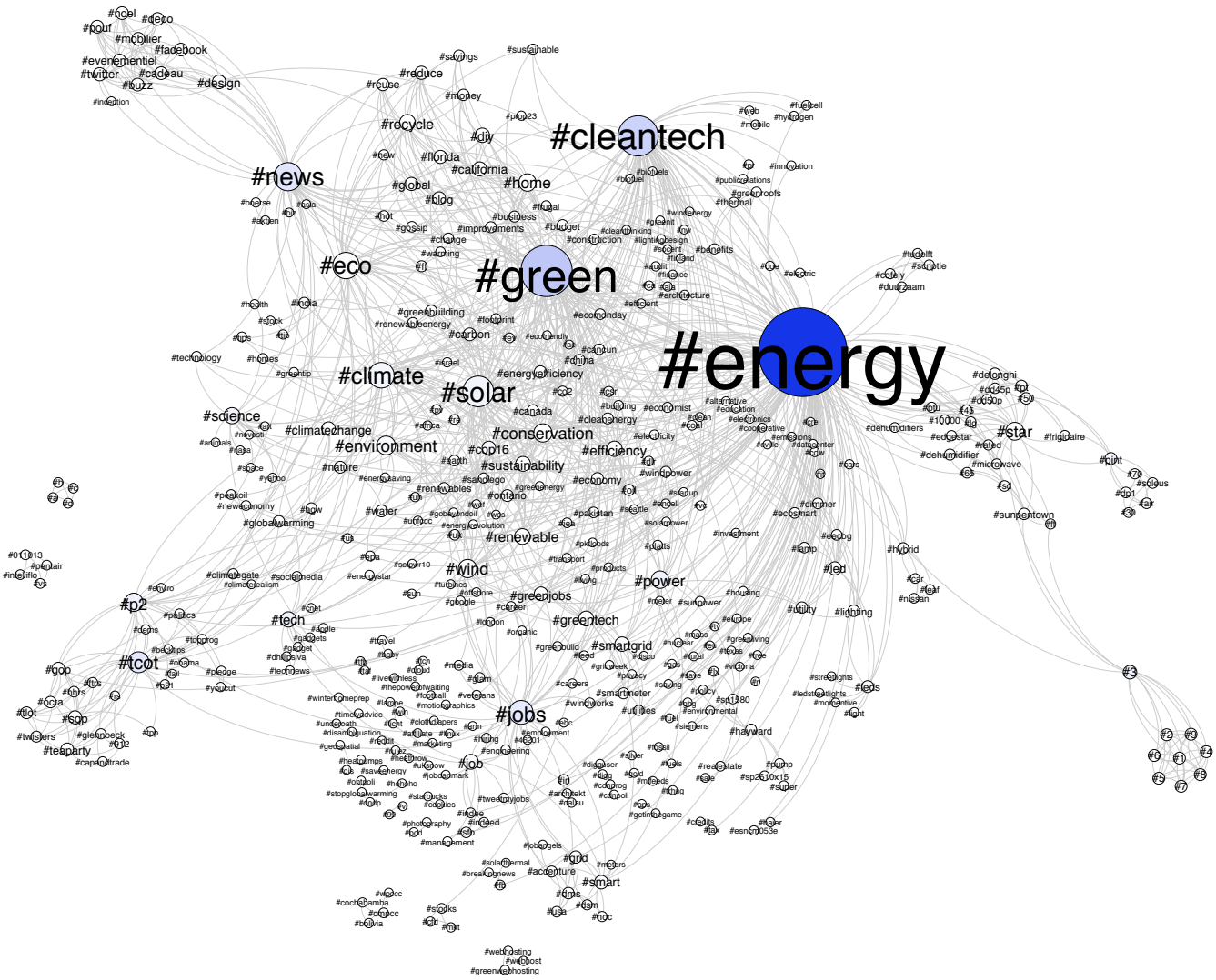


Figure 7: Visualization (zoomable) of 416 Hashtags in Energy-Related Tweets, September 2010 Through January 2011; created with Network Explorer (Rubens et al., 2011).

for Gephi (Bastian et al., 2009). The size of the node corresponds to the node's degree (number of edges), and the gradation of nodes color corresponds to the node's betweenness centrality (how well it connects the nodes, i.e. a measure proportional to the number of paths in the network that pass through a given hashtag). Thus, the term #energy has a high degree of centrality to all terms in the visualization. The position of nodes (network layout) was determined in two stages: (1) cluster-based stage, (2) relation-based compacting stage.

In the cluster-based stage we use OpenOrd layout algorithm (Martin, 2011), since it produces layout that allows to better distinguish clusters based on the interconnections between the nodes. We then applied ForceAtlas (Bastian et al., 2009), to lay out the subgraphs (nodes repel each other, while edges pull nodes closer together). The network figure is embedded in the pdf document by using scalable vector graphics, so it is possible zoom in to look at finer details of the network without the loss of resolution. The data for the network structure was obtained by applying the tweetonomy concept (Wagner and Strohmaier 2010) to the dataset. To produce a smaller subnetwork of higher 'quality' we have kept only the hashtags that were connected to each other through edges with weight (number of co-occurrences between hashtags) of 100 or larger. The resulting network consists of 416 nodes and 1,103 edges.

3.4.7 *Constructing Tweetonomies: Acquiring Latent Semantic Structures*

"Tweetonomy" is a concept (envisioned and studied by Wagner and Strohmaier 2010) that denotes the latent semantic structures that can emerge in social streams such as Twitter as a byproduct of online conversations. Based on analysis of messages and their content, as well as URLs, and other user-based syntax such as hashtags, slashtags or @ replies, Tweetonomies can be constructed in a variety of ways. Emerging communication conventions on Twitter yield a large number of ways of in which such taxonomies

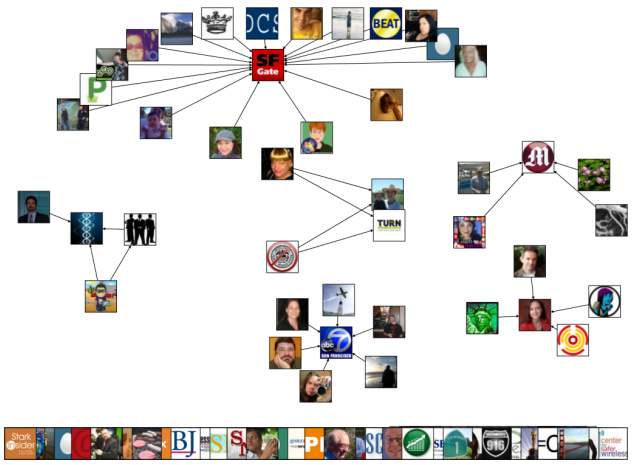


Figure 6: Networks of Tweeters Mentioning "smartmeters" or "smart + meters" during September, 2010. Data for Figure 8 were visualized using NodeXL and its functionality for linking and revealing Twitter images, using the NodeXL Twitter Importer module (Hansen et al., 2010). Visualization is created by Marc A. Smith.

can be constructed by considering for example, (i) the author(s) of messages, (ii) recipients of messages, (iii) retweets or diffusion of information, (iv) contents of a message such as links embedded in messages, keywords or hashtags, and (v) the time of messages.

The term Tweetonomy was inspired by research on social tagging systems (cf. for example Koerner et al. 2010, Helic et al. 2011) where the collective act of tagging resources (such as assigning the tag 'news' to the resource www.cnn.com) yields so-called folksonomies# – from folk-generated taxonomy. The idea of folksonomies is that a common structure emerges from the individual classification acts of a large number of users. Similarly, Tweetonomies denote the emerging structures that can be observed when large numbers of users collectively organize and structure conversations in online social media such as Twitter.

In the Figure 8, TwitterExplorer - a partial implementation of the concept of tweetonomies - is used to to analyze and visualize the latent semantic structures embedded in energy-related conversations on Twitter. TwitterExplorer is a user-friendly social streams analytics tool that was implemented based on the Django web development framework, using Whoosh for indexing, NetworkX for network layouting, and KMETIS for graph partitioning. TwitterExplorer visualizes semantic relations between terms used in Twitter messages based on different aggregation and similarity measures. In the following examples, semantic similarity between terms is calculated based on co-occurrence of terms within messages (Tweets).

Illustrative results and analysis.

This analysis was performed on a sample of the 2.47 million Tweets that contained any of the ecolinguistic terms included in this study. Tweets were selected into this sample based hashtags that were present in at least 3 Tweets. Illustrative results demonstrate new tools for data acquisition, curation and analysis. In Figure 8 three findings can be observed; the entry term #smartmeter is most often co-occurs with terms, #energy, #utilities, #cleantech, #greenjobs, #greentech, #smartgrid and two locations, #Illnois and #California, and one brand, #deloitte. This first layer of co-occurring terms is then connected to a larger layer of co-occurring terms at the edge of the network circle. Interestingly, few of these networked terms convey the increased ability of the customer to understand their energy use. Additional insight on the context of the conversation and the issues it includes can be explored through examination of the content of the Tweets.

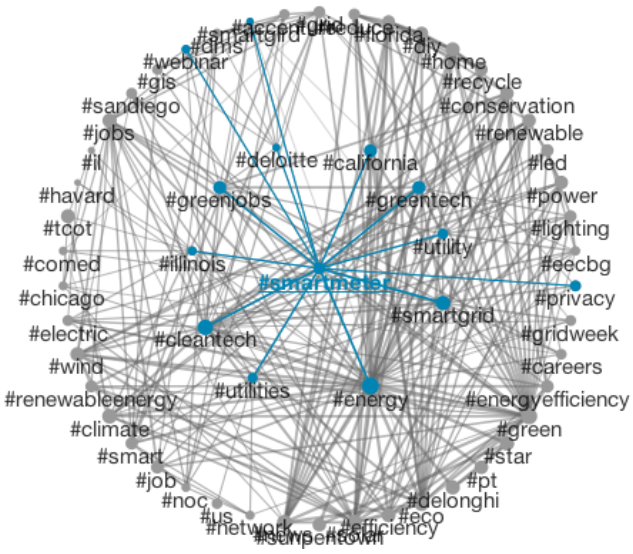


Figure 8: Two-hop filtered ego-centric network for the hashtag #smartmeter.

3.4.8 *TwitterExplorer: An interactive tool to explore tweetonomies on Twitter*

Figure 9 provides a screenshot of the Energy-TwitterExplorer showing details for the hashtag #energy. In the upper left, TwitterExplorer shows a list of tags related to #energy ranked by co-occurrence in Tweets. In the upper mid section, TwitterExplorer shows a 2-hop ego-network of terms related to #energy based on co-occurrence in tweets. On the upper right, current tweets that are including the hashtag #energy are pulled in real time from the Twitter API. The chart on the left ('statistics') visualizes different properties for the #energy hashtag stream over time, including for example, social diversity of the stream which characterizes the extent to which Tweets in the stream were authored by different users. The chart on the right ('environment statistics') plots the same properties for the hashtags that are semantically related to the #energy hashtag. While this screenshot depicts semantic statistics for a single hashtag (here: #energy), TwitterExplorer provides an easy to use user interface to support analysts in exploring all other topics (hashtags) in the dataset in a similar manner. On the bottom, a list of all tweets that include the hashtag #energy in the dataset is presented.

4. SUMMARY

The data, tools and initial analysis presented in this paper represent first steps towards more refined analytical approaches that help understand the large scale conversations taking place on Twitter and elsewhere. This project has demonstrated the feasibility of using data mining techniques to gather and analyze vast amounts of data from ongoing social media conversations and of analyzing the data for meaningful metrics that describe conversations about energy consumption behavior. The methods for this preliminary investigation included: development of a list of energy related metaphors, terms and general descriptors as well as a list of energy conservation and reduction behaviors, monitoring term usage (location, frequency, context, clustering) on the Internet at frequent intervals, analyzing data for frequency and location of communication, clustering of terms over time – such as changes in their proximity and occurrence, introduction of new terms and fading of others. Our initial exploration confirmed that conversations about energy-related issues are, indeed, taking place in social media, specifically Twitter and that these communications can be studied to better understand how to use technologically-enhanced word-of-mouth to stimulate user-generated persuasion. Using content analysis of full Tweets, network analysis of co-occurring hashtags, and semantic analysis of the co-occurring hashtags and their authors, this preliminary investigation identified descriptors, concerns, actions, and issues. We confirm that studying Twitter communications can provide actionable means for assessing engagement, identifying influencers, and identifying word-of-mouth communities that can accelerate change in energy efficiency behaviors.

An ecolinguistic taxonomy of over one hundred terms was established and included terms for: energy technologies/hardware &software; communication behaviors; energy & climate change frames, metaphors, & visualizations; energy efficiency and climate change innovative programs; issues such as renewable energy, global warning, energy insecurity; utilities, venture firms and companies; and behaviors (high and low cost and impact.)

By example, we have demonstrated that it is possible to capture an issue-based sample of the Tweetstream and curate it as a dataset. During the period from September 3, 2010 to January 4, 2011, over 2.47 million Tweets that included one or more of the terms in our ecolinguistic taxonomy were gathered. These Tweets contained 18,338 hashtags that were mentioned three or more times. Data analysis methods were applied to explore the aggregate of energy consumption-related conversations in vast quantities of Twitterdata.

We demonstrated preliminary steps in tracking the extent to which energy-related terms are used by Tweeters, the periodicity and the content of their communications. Periodicity was observed by day or week and by topic in the Twitter conversation, as a whole and for hashtags of specific interest. The presence of emerging social communities was observed in a network of directional Tweets. Content analysis of the Tweets revealed a relationship between the volume of Tweets on a topic (using a particular hashtag) and the visibility of social interest and concern in issues of broad scale significance, and sources of information of social media conversations about changing energy behaviors.

Using network analysis of hashtags we analyzed and visualized contextual relationships of among the salient terms used in social conversations and identified several clusters of related issues, revealed by the co-occurrence of hashtags. Through Twitter Explorer, we demonstrated preliminary and partial application of the Tweetonomy construct, with a network analysis of energy terms in twitter streams, time series analysis to track the evolution of consumer attitudes about energy efficiency, and metrics to characterize the conversations. We used social and linguistic structures of communication (repeat communications and varied types of Twitter communication conventions such as pictures, hashtags, retweets, and urls) to analyze the establishment of self-organizing communities of consumers. These communities share many of the characteristics of issue publics, and further research on similarities and differences to other issue publics is needed in order to understand how to create, grow and sustain word-of-mouth persuasion for energy behavior change.

We initiated exploration into the social aspects of information diffusion through conventions such as retweets and other conversational responses, through the membership and social distance of these conversations, and in overtime changes in frequency and form of the networks. demonstration of tools that permit visualization of vast quantities of user-generated content about energy and sustainability.

5. RECOMMENDATIONS

It is not yet known how social media will be best integrated into the portfolio of communications that will achieve the tipping point of consumer behavior change in energy use. History has repeatedly shown us that new media join rather than replace traditional media (Reeves and Nass, 1996). To achieve optimum results in minimum time, continual feedback is needed for rapid iteration of innovative approaches. Openly available, rapid response metrics that reflect consumer engagement with changes are vital. Social media metrics offer this opportunity. The identification and refinement of tools for analyzing conversations in social media is vital for future societal issues and challenges. Understanding energy-related conversations on Twitter can provide a stepping stone for detecting and potentially triggering changes in energy consumption behavior through social media. The social diversity of the tweets, for example, can indicate opportunities to explore specific hashtags for deeper understanding of social roles in the conversation.

The results of this project have potential application in identifying new ways of tracking public opinion and behavior change related to sustainability and energy consumption and for analyzing domain-specific, user generated content on social media platforms. Our ongoing research will continue to monitor the extent to which ecolinguistic terms appear in social media used by the general population. We will track the presence, participation and meaning of energy consumption-related conversations in social media and, with these insights, support the development of initiatives to validly detect and potentially trigger changes in energy consumption behavior through social media. Additionally, we believe there are significant opportunities to monitor social media to examine the influence of larger social issues and contextual factors (environment, climate, local and global events) on the conversation about energy, environment and climate change.

Based on our initial results, we recommend continued collection of data and development of analytical methods and tools that can: track public opinion related to energy

consumption; analyze domain-specific, user generated content on social media platforms; identify and track indicators such as semantics and social roles; identify and explore patterns and disruptions; identify and benchmark grassroots resources such as author networks; characterize opportunities for resource transformation; and build semantic models to understand the aggregations of conversation streams.

We have identified four major questions for persuasive technologies lie at the intersection between people and information technology and require continued interdisciplinary exploration:

- What (and how much, and what kind of) information is needed by consumers in order to inform changes in their energy use?
- What balance of technology-driven personalization, disambiguation, and privacy are optimum for building communities of awareness, interest and action?
- How can the impact of best practice framing, empowerment and compliance speed the adoption of technologies that encourage reduced energy consumption?
- What combination of persuasive technologies – both old and new – will make changing energy behavior contagious and persistent?

Analytical tools are needed for observation and interpretation of the millions of personal conversations already occurring online. Reliable and valid metrics are needed to measure the effectiveness of persuasive communications and enable meaningful interpretation of the impact. The semantic analysis of tweets containing ecolinguistic content in this paper is a preliminary example of the data, tools, and metrics that can inform the timeliness, relevance and positioning of persuasive communications.

To accelerate exploration of these important issues, we are making the ecolinguistic taxonomy, the Twitter Energy data, the TwitterExplorer and the Network Explorer available for other researchers. Against the urgency of climate change and the need to mobilize widespread changes in energy consumption, we encourage other researchers to join us in a research agenda that includes: analyzing and characterizing energy consumption behavior; tracking public opinion related to energy consumption; analyzing domain-specific, user generated content on social media platforms; identifying and tracking leading indicators of attitude and behavior change; and identifying patterns and disruptions that can accelerate change and provide alerts for emergency response.

References

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.

Biascovich, J. and Bailenson, J. (2011). *Infinite Reality.* Harper Collin Publishers, New York.

Block, M. and Schultz, D. E. (2009). *Media Generations: Media Allocation in consumer-Controlled Marketplace.* BIG Research.

Daft, R. L., Lengel, R. H., and Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly*, 11:3:320.

Flora, J. A. and Russell, M. G. (2011: to be published). A taxomony of energy efficiency terms: issues, frames, metaphors, sources and behaviors. Technical report, Media X, Stanford University.

Fogg, B. (2003). *Persuasive Technology.* Morgan Kaufman Publishers, Amsterdam.

Fogg, B. (2008). Mass interpersonal persuasion: An early view of a new phenomenon. In Oinas-Kukkonen, H., Hasle, P., Harjumaa, M., Segerstahl, K., and Ohrstrom, P., editors, *Persuasive Technology*, volume 5033 of *Lecture Notes in Computer Science*, pages 23–34. Springer Berlin / Heidelberg.

Freeman, L. C. (2009). *Encyclopedia of Complexity and Systems Science.* Springer, New York.

Gleave, E., Welser, H. T., Lento, T. M., and Smith, M. A. (2009). A conceptual and operational definition of social role in online community. In *Proceedings of the 42nd Hawaii International Conference on System Sciences.*

Hansen, D., Shneiderman, B., and Smith, M. (2010). *Analyzing Social Networks with NodeXL: Insights from a Connected World.* Morgan Kaufmann.

Helic, D., Strohmaier, M., Trattner, C., Muhr, M., and Lerman, K. (2011). Pragmatic evaluation of folksonomies. In *20th International World Wide Web Conference (WWW2011).* ACM.

Hogan, J. E., Lemon, K. N., and Libai, B. (2004). Quantifying the ripple: Word-of-mouth and advertising effectiveness. *Journal of Advertising Research*, pages 271–280.

Kaptein, M., Markopoulos, P., de Ruyter, B., and Arts, E. (2010). Persuasion in ambient intelligence. *Journal of Ambient Intelligence and Humanized Computing*, 1:1:43–56.

Lenhart, A., Purcell, K., Smith, A., and Zickuhr, K. (2010). Social media and young adults. http://www.pewinternet.org/Reports/2010/Social-Media-and-Young-Adults/Summary-of-Findings.

Nass, C. and Yen, C. (2011). *The Man Who Lied to His Laptop.* Penguin Group, New York.

Ophir, E., Nass, C., and Wagner, A. D. (2009). Cognitive control in media multitaskers. In *PNAS.* http://www.pnas.org/content/early/2009/08/21/0903620106.full.pdf+h

Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54:547–577.

Poeschke, J. and Strohmaier, M. (2011). Twitterexplorer-energy. http://mediax.stanford.edu/changeeb.html.

Reeves, B. and Read, L. (2009). *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete*. Harvard Business School Publishing, Boston, MA.

Rubens, N., Russell, M. G., and Perez, R. (2011: to be published). Network Explorer. Technical report, Media X, Stanford University.

Russell, M. G. (2009). A call for creativity in new metrics for liquid media. *Journal of Interactive Advertising*, 9:2.

Russell, M. G. (2011). *Handbook of Advertising Research*, chapter Evolving Media Metrics from Assumed Attention to Earned Engagement, pages 125–144. Information Science Reference, Hershey, New York.

S. Martin, W. M. Brown, R. Klavans and Boyack, K. (2011). OpenOrd: An Open-Source Toolbox for Large Graph Layout. In *SPIE Conference on Visualization and Data Analysis (VDA)*.

Savage, N. (2011). Twitter as medium and message. *Communications of the ACM*, 54:3:18–20.

Slater, M. D. (2007). Reinforcing spirals: The mutual influences of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory*, 17:281–303.

Smith, A. and Rainie, L. (2010). 8% of online Americans use Twitter. http://www.pewinternet.org/Reports/2010/Twitter-Update-2010/Findings.aspx.

The National Academy of Science (2011). Climate change. http://dels-old.nas.edu/climatechange.

US Department of Energy (2011). U.S. DOE Energy Efficiency and Renewable Energy (EERE). http://www.eere.energy.gov.

U.S. Environmental Protection Agency (EPA) (2011). Energy Portal. http://www.epa.gov/energy/energy.html.

Wagner, C. and Strohmaier, M. (2010). The wisdom in tweetonomies: acquiring latent conceptual structures from social awareness streams. In *Proceedings of the 3rd International Semantic Search Workshop*, pages 1–10. ACM.

Wasserman, T. (2011). Audi superbowl dd claims first use of twitter hashtag. http://mashable.com/2011/02/02/audi-super-bowl-twitter-hashtag/.

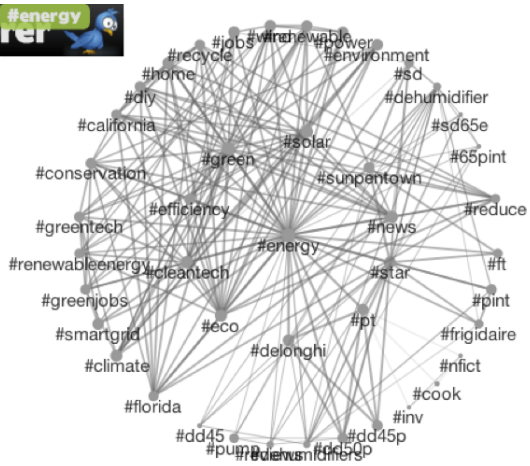
Wu, M. (2008). Community health index for online communities. Technical report, Lithium Technologies, Inc.

#energy

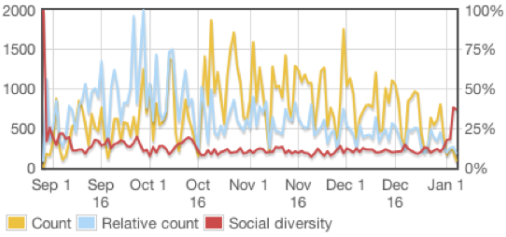
in 95611 tweets | #energy on twitter

Co-occurrences

#green	21316
#star	12966
#delonghi	9477
#eco	9366
#efficiency	7859
#solar	7842
#news	6568
#pt	6323
#cleantech	6260
#sunpentown	5950



Statistics



Tagness:	12.561%
Relative position:	66.022%
Only hashtags following:	52.776%
With link:	93.175%
Social diversity:	4.592%

Environment statistics

Tag	Count	Tagness	Relative position	Only hashtags following	With link	Social diversity
0 hops						hide
#energy	95611	12.561%	66.022%	52.776%	93.175%	4.592%
1 hop						hide
#star	12995	10.561%	72.401%	73.428%	99.854%	3.009%
#pt	6339	89.106%	59.364%	99.763%	99.779%	4.070%
#delonghi	10104	96.661%	15.502%	62.579%	100.000%	3.157%
#eco	16900	77.698%	69.211%	43.639%	95.635%	7.623%
#cleantech	27407	29.251%	74.813%	66.427%	88.489%	10.620%
#efficiency	8207	3.890%	63.835%	29.353%	89.302%	6.543%
#green	49655	22.214%	68.656%	50.660%	95.114%	8.051%
#solar	23025	8.118%	57.158%	35.170%	94.280%	11.551%
#sunpentown	5950	95.598%	18.956%	48.824%	100.000%	4.655%
#news	15860	21.377%	67.335%	56.900%	98.366%	7.450%
avg		45.447%	56.723%	56.674%	96.082%	6.673%
weighted avg		34.137%	63.044%	54.047%	95.009%	7.832%
2 hops						show
3 hops						show

Tweets in dataset

barineauac 2010-12-21 15:25:13
Don't miss out! The Energy Efficient AC and Heat Pump Rebate is available until December 31. You can save $ and #energy: http://ow.ly/3rWCf

alps7894 2010-10-19 02:19:44
Cooking with a low carbon footprint:#energy http://factoidz.com/how-to-cook-with-a-low-carbon-footprint-select-and-prepare-eco-friendly-meal

bryantdulin 2010-10-18 21:39:16
RT @ksdulin: "#Energy production accounts for over two-fifths of all #US #water use, meaning a big #carbon footprint creates a big water footprint."

machinedoctor 2010-10-18 13:32:26
Reduce your #carbon footprint and save $. We'll teach you how to do #Air Leak Survey or we'll do it for you. #energy http://bit.ly/cbOGM6

ecowarriorandco 2010-10-15 10:20:04
Interested in knowing your carbon #footprint & how to reduce it. Save money too. Free Home #Energy Check-up from us :-) http://ow.ly/2KUAz

*Note: for compactness the layout of elements differs from the actual one.

Figure 9: Screenshot of the Energy-TwitterExplorer.