Twitris: A System for Collective Social Intelligence

Amit Sheth, Ashutosh Jadhav, Pavan Kapanipathi, Chen Lu, Hemant Purohit, Gary Alan Smith, and Wenbo Wang
Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis), Wright State University, Dayton, OH, USA

Synonyms
Citizen sensing; Community evolution; Event analysis on social media; Interaction network; People-content-network analysis; Real-time social media analysis; Semantic perception; Semantic Social Web; Sentiment-emotion-intent analysis; Social media analysis; Spatio-temporal-thematic analysis; Web 3.0

Glossary
Citizen Sensing Humans as citizens on the ubiquitous Web, acting as sensors and sharing their observations and views using mobile devices and Web 2.0 Services
Citizen-Sensor Network An interconnected network of people who actively observe, report, analyze, and disseminate information via text, links to other resources, and various media including audio, images, and videos
Social Media Analytics Social media analytics is the practice of gathering data from social media websites and analyzing that data to gain new insights from social media and to make informed decisions
Semantic Web Semantic Web is a group of methods and technologies to help machines and humans understand the meaning – or “semantics” – of the data on the World Wide Web
Spatio-Temporal-Thematic (STT) Analysis Social media analytics taking into account what is being said about an event (theme) and where (spatial) and when (temporal) it is being said
People-Content-Network Analysis (PCNA)
Social media analytics taking into account social media user (People), data shared on social media websites (Content), and network of social media users (Network)

Sentiment-Emotion-Intent (SEI) Extraction
Analyzing social media content to extract insights about social media users’ sentiment (positive, negative, and neutral), emotion (happy, angry, upset, etc.), and the user’s intention (seeking information, sharing information, etc.)

Introduction
Well over a billion people have become “citizens” of an Internet- or Web-enabled social community. Web 2.0 fostered the open environment and applications for tagging, blogging, wikis, and social networking sites that have made information consumption, production, and sharing so incredibly easy. With over five billion mobile connections, over a billion with data connections (smartphones) and with many more having the ability to communicate using SMS, digital media can be shared with the rest of the humanity instantly. As a result, humanity is interconnected as never before. This interconnected network of people who actively observe, report, collect, analyze, and disseminate information via text, audio, or video messages, increasingly through pervasively connected mobile devices, has led to what we term citizen sensing (Sheth 2009a, b). This phenomenon is different from the traditional centralized information dissemination and consumption environments where citizens primarily act as consumers of reported information from several authoritative sources.

This citizen sensing is complemented by the growing ability to access, integrate, dissect, and analyze individual and collective thinking of humanity, giving us a capability that is recognized as collective intelligence. Citizen sensing involves humans in the loop and with it all the complexities associated with and intelligence captured in human communication. As citizen sensing has gained momentum, it’s generating millions of observations and creating significant information overload. In many cases it becomes nearly impossible to make sense of the information around a topic of interest. Given this data deluge, analyzing the numerous social signals can be extremely challenging. In response to this growing citizen sensing data deluge, Twitris has been developed with the vision of performing semantics-empowered analysis of a broad variety of social media exchanges.

Twitris, named by combining Twitter with Tetris, a tile-matching puzzle game, has incorporated increasingly sophisticated analysis of social data and associated metadata, combining it with background knowledge, and more recently (albeit not discussed here) machine sensor or data captured from sensors and devices that make up the Internet of Things (IoT). Twitris’ evolution can be characterized in three phases (and corresponding versions of the system). Figure 1 outlines the corresponding dimensions Twitris considers.

Twitris is a comprehensive platform for analyzing social content along multiple dimensions leading to in-depth insights into various aspects of an event or a situation. The central thesis behind this work is that citizen sensor observations are inherently multidimensional in nature, and taking these dimensions into account while processing, aggregating, connecting, and visualizing data will provide useful organization and consumption principles. Twitris evolved in
three phases, characterized by the versions of the systems:

- **Twitris v1:** Spatio-temporal-thematic (STT) processing of Twitter and associated news, multimedia, and Wikipedia content (Sheth 2009b; Nagarajan et al. 2009a; Jadhav et al. 2010)

- **Twitris v2:** People-content-network analysis (PCNA) (Purohit et al. 2011) with use of background knowledge and semantic metadata extraction and querying/exploration

- **Twitris v3:** Sentiment-emotion-intent (SEI) extraction (Chen et al. 2012; Wang et al. 2012; Nagarajan et al. 2009b) along with personalization (Kapanipathi et al. 2011a) and emerging continuous semantics (Sheth et al. 2010) capability involving semantic streaming social stream (i.e., real-time) processing using dynamically generated and updated domain models for semantics and context

The above versions, or phases, of Twitris’ development are not as granular as painted above, that is, the issues identified above are not explicitly segregated by the version of the Twitris which has been in continuous development with senior students graduating and new students picking up the work. Four talks including a tutorial cover many of the issues covered by Twitris (Sheth 2009a; Nagarajan 2010; Nagarajan et al. 2011; Sheth 2011).

### Key Points

Throughout this article we investigate the role and benefits of using semantic approach, especially by metadata extraction and enrichments and contextually applying relevant background knowledge, along with demonstrating examples on real-world data using system (Twitris) developed at Kno.e.sis. The article focuses on the following key points:

1. Event-specific analysis of citizen sensing and discussion on opportunities and challenges in understanding temporal, spatial, and thematic cues.
2. Facets of people-content-network analysis with focus on user-community engagement analysis.
3. Real-time social media data analysis and the concept of continuous semantics supported by dynamic model creation.
4. Sentiment and emotion identification from citizen sensing data.
5. Recent advances in developing semantic abstracts or semantic perception to convert massive amounts of raw observational data into nuggets of information and insights that can aid in human decision making.

### Historical Background

The idea of research leading to Twitris occurred on November 26, 2008. Terrorists struck Mumbai, India, and over the next 3 days, they proceeded to make mayhem in nine locations. Each of the nine sub-events of this overall event separated by time and location (space) had distinct thematic elements or topical content. The importance of Twitter, especially in terms of citizen sensing – the ability of a regular person to use his or her mobile device to share his or her personal observation, thoughts, and belief, well before a traditional news media has a chance to do reporting and to shape opinions – was extensively discussed in the immediate aftermath of this momentous event. This event also gave us a clear case for the needs and benefits of analyzing social media content such as tweets and Flickr posts and related news stories along the three dimensions of spatial (location of observation), where; temporal (time of observation), when; and thematic (the event in question), what (Battle 2009; Impact Lab 2008; Keralaravind 2008).

### Twitris Platform and Three Stages of Its Evolution

#### Twitris v1: Spatio-Temporal-Thematic (STT) Processing of Twitter and Associated News, Multimedia, and Wikipedia Content

Twitris v1 (Jadhav et al. 2010) was designed with the following three major steps:

1. Data collection: collect user posted tweets pertaining to an event from Twitter,
Twitris: A System for Collective Social Intelligence, Fig. 2 A snapshot of spatio-temporal-thematic slice of citizen sensing showing content related to Mumbai terrorism (thematic) related to Taj hotel (spatial, thematic), during a period of interest (temporal)

Twitris: A System for Collective Social Intelligence

2. Data analysis: (a) process obtained tweets to extract strong event descriptors considering spatial, temporal, and thematic event attributes and (b) process event related news, multimedia, and Wikipedia content to get event context and gain a better understanding.

3. Visualization: present extracted summaries on Twitris v1 user interface.

Twitris v1 performs a two-step processing to extract strong event descriptors from tweets. First, it creates the spatio-temporal clusters of the tweet corpus surrounding an event, since every event is different and we want to preserve the social perceptions that generated this data. TFIDF computation is performed to fetch the n-grams from this set. The second step involves the association of spatial, temporal, and thematic bias to these n-grams by means of enhancing the weights while preserving the contextual relevance of these event descriptors to the event. Further details of the text-processing algorithm are available in (Nagarajan et al. 2009a). Twitris v1 user interface (Fig. 3) facilitates effective browsing of when (temporal/time), where (space/location), and what (thematic/context) slices of social perceptions behind an event.

The objective of the Twitris v1 user interface is to integrate the results of the data analysis (extracted descriptors and surrounding discussions) with emerging visualization paradigms to facilitate sensemaking. To start browsing, users are required to select an event. Once the user chooses a theme, the date is set to the earliest date of recorded observations for an event and the map is overlaid with markers indicating the spatial locations from which observations were made on that date. We call this the spatio-temporal slice (Figs. 4 and 5).

Users can further explore activity in a particular space by clicking on the overlay marker. The event descriptors extracted from observations in this spatio-temporal setting are displayed as an event descriptor cloud. The spatio-temporal-thematic (STT) scores determine the size of the descriptor in the tag cloud. In order to get event context and better understanding of the event, we enhanced Twitris, by integrating event related news, multimedia (images and videos), and Wikipedia articles. We leveraged explicit semantic information from DBpedia to identify relevant news and Wikipedia articles.
Twitris: A System for Collective Social Intelligence, Fig. 3  STT biased scoring mechanism of Twitris v1 for relevance and ranking of keyphrases compared to traditional TFIDF-based ranking: “mumbai” ranked highest based on TFIDF is far less informative compared to “foreign relations perspectives”

Twitris: A System for Collective Social Intelligence, Fig. 4  Early version of Twitris v1 user interfaces for displaying thematic component (using STT biasing) on right (b) based on spatial and temporal selection on left (a)

Twitris: A System for Collective Social Intelligence, Fig. 5  Twitris v1 user interface with spatio-temporal slice and multimedia widgets
Twitris: A System for Collective Social Intelligence, Fig. 6  Twitris v1 user interface with event descriptor cloud, related tweets, news, and Wikipedia articles for event “Austin plane attack.” Joe Stack, the man responsible for the Austin suicide plane attack on the IRS office, put up his suicide note online about the attack. He was a former bass player for the Billy Eli band. Here Twitris captures STT event descriptors summarizing the important facets

When a user clicks on a particular descriptor, we display tweets containing the event descriptors and the top current news items as well as related Wikipedia articles (Fig. 6).

**Twitris v2: People-Content-Network Analysis (PCNA) with Use of Background Knowledge and Semantic Metadata Extraction and Querying/Exploration**

The Mumbai terrorism event of 2008 gave the impetus to study the event from STT dimensions and focus on connecting with relevant news content. Social media continues to grow and revolutionize the way users interact with each other and information. Social network users are not only creators and recipients of the information but also critical relays to propagate information. This powerful ability of sharing has played an important role in events with varied social significance, audience, and duration, such as political movements (e.g., the Jasmine Revolution in Tunisia), brand management and marketing, and perhaps most visibly, crisis and disaster management (e.g., Haitian and Japanese earthquakes). The Twitris team started to look at the issues such as the role of content nature for high vs. low attributed information diffusion (a phenomenon of propagating messages via friendship/follower connections among users of social network) (Nagarajan et al. 2010b) and user engagement (given a discussion topic on social media, what motivates a user to engage in the discussion for his/her first interaction) (Purohit et al. 2011; Ruan et al. 2012). Consequently, Twitris v2 embarked on a more comprehensive analysis along the three pillars of what makes anything social: who is engaging in the social activity, what is being communicated, and how does this communication flow between those engaged in the social activity. The idea is to gain insights into how permanent and transient networks arise
and what and why information flows across such networks. Twitris v2 developed the significant capability to extract more types of metadata, and the infrastructure became more semantic with the use of Semantic Web standard RDF as well as relevant background knowledge. The latter enabled Twitris v2 to support the deep exploration capability with use of DBpedia and SPARQL over metadata extracted from the tweets. Twitris v2 research that focus on coordination during disasters also led to integrate Twitris with Ushahidi’s SwiftRiver open source platform and support ingestion of SMS which were used for events such as Pakistan Floods in 2010.

Let’s look at some examples of Twitris v2 capabilities:

• **Evolving ad hoc nature of social media communities:**
Event-centric communities with varied nature (Purohit et al. 2011) often bring together users from different parts of the social network, especially in Twitter where we keep switching discussions of our interests, and we may not already be connected to other participants of those communities. Therefore, in such ad hoc communities, it is difficult to depend on just follower graphs for understanding the dynamics. Twitris v2 introduced analysis of user interaction networks so that human dynamics in the evolving communities can be understood at granular levels – influencer analysis, contextually important people with roles to engage with, community evolution, etc. Twitris v2 built this feature by extending our research in the user interaction network analysis on brand-page communities (Purohit et al. 2012a).

• **Contrast in the structure of interaction networks:**
Figure 7 shows the networks of influencers in two topical communities during the Occupy Wall Street (OWS) movement, “OccupyChicago” on the left and “OccupyLA” on the right. Such an analysis provides insights to understand not only the real dynamics of the actors (e.g., what organizations supporters belong to and to whom are they strongly connected) but also the potential of the influencers to drive actions in the communities (tightly connected influencers are likely to drive effective “call for action” propagation in the communities). In this figure, the influencer network of OccupyLA is highly connected and self-organized as compared to sparsely connected...
Twitris: A System for Collective Social Intelligence, Fig. 9 Interaction network evolution for topical community surrounding Mitt Romney, US Presidential Election 2012

one for OccupyChicago and, therefore, likely to reach masses effectively for any call for action. Even the Facebook page for OccupyLA reflected such activism.

- **Slicing and dicing the networks by user features:**
  To glean insights about actionable information in the ad hoc communities, we need to understand the participants better. Therefore, Twitris v2 introduced slicing and dicing analysis of the interaction networks by providing user/node centric features. For example, professional or organizational affiliation of users provides clues to understand the cause for dynamics – e.g., who are the people behind the organized network of OccupyLA? Are such users from the same type of organizations lead to coordinated actions? Similarly, Twitris v2 introduced the content-centric analysis, thus realizing the full potential of PCNA. Users are clustered by grouping them into sentiment segments of the target topic, thus answering questions like which candidate is going stronger in the influencer network from a sentiment perspective (Fig. 9) between Mitt Romney and Ron Paul and for what issues?

- **Understanding group dynamics by community evolution:**
  Twitris v2 focused on the larger goal of predictive ability for group dynamics, and the people-content-network analysis (PCNA) framework was the key to the untapped potential of group dynamics. Therefore, Twitris v2 created clusters in the ad hoc communities based on the sentiment of the users for a targeted topic over time and associated events on the timeline for causal analytics. Figure 9 shows an example of community evolution centered around Republican presidential nominee Mitt Romney during March 1–31, 2012. It shows three snapshots taken over a 10-day period, and we observed an extremely modularized community in the end of the analysis, which was not really the case for the closest competition, Rick Santorum. And as we know, Santorum exited the race on April 9. Thus, the analysis of community evolution made Twitris v2 capable of understanding group dynamics of ad hoc communities by not limiting the output to just understand users but also the group behavior.

  Twitris v2 leverages Semantic Web technologies by the use of background knowledge such as DBpedia to provide deeper insights about the event. Background knowledge changes the way you can look at the information, as it puts the information in context. This is especially important for tweets because they are short and therefore individually lack the volume of information that provides an informative context. For example, in the above Fig. 10, questions such as “Who are the dead people that are mentioned in the context of OWS movement” can be answered using the background knowledge, whereas simple keyword search cannot put the information of tweets in context. Further, to answer the questions in the figure and generate answers such as Rosa Parks, the system has to have the background knowledge about this named entity as a Person and also that she is dead. Going deeper into the
background knowledge provides information that Rosa Parks was famous for the Montgomery Bus Boycott during the US civil rights movement in 1955–1956.

**Twitris v3: Emotion-Sentiment-Intent, Real-Time View and Other Advancements**

Behind every (well, most of the important) tweet, there is a human. And a human is complex. Through a tweet, a person expresses emotion, sentiment, and intent. Understanding this dimension is a key to unlock the true potential of social media. This is especially true for monetization of social media. Understanding an underlying intent can tell us if a user is expressing a transactional (potentially for buying a product) intent, seeking information, or just sharing information (Nagarajan et al. 2009b). Sentiment is perhaps the most sought after type of analysis of social data. Currently, it is the primary basis of social media analysis to predict whether a product or a movie will succeed, who is more likely to win an election, or to attempt to identify consumer interest and hence use it for targeting the advertisement. Analysis of or identification of emotion is likely the dark horse of the three – while techniques for its analysis are not yet as mature as sentiment analysis, it is likely to be combined with the other two to give far more signal than without it.

A key innovation in sentiment analysis, employed in Twitris v3, is topic-specific sentiment analysis – to associate sentiment with an entity (Chen et al. 2012). This enables us to identify two different sentiments associated with different entities in a single tweet. For example, in tweet “The King’s Speech was bloody brilliant. Colin Firth and Geoffrey Rush were fantastic!” we can identify both the sentiment (i.e., bloody brilliant) associated with the movie “The King’s Speech” and the sentiment (fantastic) associated
More recently, we are associating sentiments with events – when there is a significant change in sentiment, we attempt to associate that with real-world events. For example, by tracking both the event- and entity-specific sentiments, Twitris v3 is able to capture a substantial increase of positive sentiment towards President Obama on the immigration issue on June 15, 2012 (the day on which President Obama outlined a new immigration policy), and associate it with the event descriptors such as “dream act,” “Obama’s immigration move,” and “new immigration policy.” Figure 10 shows that Twitter users have the opposite sentiments towards two candidates: Obama (green/positive) and Romney (red/negative) on the same topic “final debate.” The reason is that Obama received more positive feedback from Twitter users than Romney did, which is in line with the impression from news media. This example demonstrates Twitris’s power in identifying topic-specific sentiments.

Compared with sentiment, emotion is more implicit. For example, “I will have a calculus test in two hours, but I’m not prepared at all.” We
Twitris: A System for Collective Social Intelligence, Fig. 12  Peek patterns of the emotion joy due to excitement of Twitter users caused by three debates and one TV program (the Daily Show) in the 2012 presidential election.

can infer that the person is nervous about the test, though there are no explicit emotion words, such as “nervous” or “panic.” It is very difficult and time consuming to label sentences with emotions, considering the implicitness of emotion. In Twitris v3, we are able to automatically create a large emotion-labeled dataset (of about 2.5 million tweets) covering: joy, sadness, anger, love, fear, thankfulness, and surprise, by harnessing emotion-related hashtags available in the tweets (Wang et al. 2012). Machine learning classifiers are trained on the large dataset to learn how to identify people’s emotions behind their tweets. And, as another key innovation, Twitris v3 can analyze people’s emotional responses in different events. For example, Fig. 11 shows the volume of joyful tweets, reaching peaks on Oct. 3, 2012 (first debate), Oct. 16, 2012 (second debate), Oct. 22, 2012 (third debated), and Oct. 19, 2012 (Obama went to the Daily Show). The reason is that Twitter users are very enthusiastic about all three presidential debates and Obama’s presence in the Daily Show TV program. Other than analyzing emotions out of tweets, Twitris v3 is also able to identify emotions from blogs, news headlines, etc. The reason is that we adapt the classifiers trained on Twitter data to other domains with a relatively small amount of labeled emotion data in other domains.

Besides detecting users’ emotional states, we also explore how to automatically identify users’ intents from posts so that monetization can be more targeted on users’ needs (Nagarajan et al. 2009b). The highlight of our study is that we discover and differentiate three types of posts: (a) transactional posts, e.g., “I am looking for a 32 GB iTouch”; (b) information sharing posts, e.g., “I like my new 32 GB iTouch”; and (c) information seeking posts, e.g., “what do you think about 32 GB iTouch?” For monetization purposes, transactional posts and information seeking posts are more valuable than information sharing posts because users are looking for information that advertisers can exploit. By extracting intent/keywords/cues from transactional and information seeking posts, our system achieved an accuracy of 52% on ad impressions using MySpace and Facebook data, while the baseline, without using our system, only achieved an accuracy of 30%.

All the above-mentioned precious assets (sentiment, emotion, and intent) of content exist due to an actionable purpose of humans. When such individual level purposes start to bring higher engagement in the groups, they become source of group-level actions, apparently, leading to the evolution of human dynamics in the social network. Therefore, we are exploring the integral
Twitris: A System for Collective Social Intelligence, Fig. 13 Some of the capabilities of Twitris v3: (1) show popular Topics, also called social signals (weighted n-grams) related to the chosen event for today and any day of the past since the event began to be tracked; (2) search from among the event related tweets with autocomplete, popular event hashtags, and active users and explore content for deep analysis (e.g., who are the dead people mentioned most often in the “occupy wall street event”) using background knowledge (default source is Wikipedia/DBpedia) and Semantic Web technologies (RDF/SPARQL); (3) show key topics of discussions by locations/regions, states, and country (e.g., see the differences in social signals from Mississippi (a “red state”) vs. Massachusetts (a “blue state”) related to President Obama’s Nobel Prize); (4) see event relevant tweets in real time on a world map or any region; (5) analyze topic-/people-/region-specific sentiment (e.g., for the US election, sentiment on candidates by states, and by topics identified by election specific topics); (6) see the networks with insights from static (e.g., followers) and dynamic features (e.g., retweet) and people/demographics (e.g., with knowledge of profession of each person); (7) display tweets, recent news, and Wikipedia pages related to selected events and social signals; (8) show event-specific multimedia (images and video); (9) see tweet traffic; (10) change date of video/analysis; (11) select location of interest – each pin shows a collection of social signals emanating from a location; and (13) select an event of interest (e.g., US Election, Occupy Wall Street, Japanese Tsunami)

role of intent with sentiment and emotions for purposeful actions in the groups. Specifically, we are focusing on intent and sentiments behind group coordination because coordinated activity has the potential to make or break the system (Fig. 13).

Advanced Research
Detailed research on social data analysis encompasses social intelligence in real time (Gruhl et al. 2010) which involved a Kno.e.sis-IBM collaboration leading to the operationally deployed BBC Sound Index system, prediction of topic volume on Twitter (Ruan et al. 2012), emotion identification using Twitter “big data” (Wang et al. 2012), brand tracking (Purohit et al. 2012a), psycholinguistic analysis during emerging coordination (Purohit et al. 2012b), privacy-aware content dissemination (Kapanipathi et al. 2011b), user-community engagement (Purohit et al. 2011), information diffusion (Nagarajan et al. 2010b), trust in social
media (Thirunarayan and Anantharam 2011), and monetization of social activities (Nagarajan et al. 2009b) reported in over 30 publications and summarized in a comprehensive tutorial (Nagarajan et al. 2011).

Key Applications

Twitris has been used in a research context for studying and analyzing social sensing and perception of a broad variety of events: politics and elections, social movements and uprisings, crisis and disasters, entertainment, environment, etc. We are now investigating more commercial applications including brand tracking and advertisement campaign effectiveness, and empowering professional users.

Future Directions

The next evolution of Twitris will be in incorporating social media content along with data from sensors and Web of Things, as well as in advanced applications for crisis and disaster support.

Acknowledgments

Alumni, Karthik Gomadam, Meena Nagarajan, and Ajith Ranabahu, have made substantial contributions to Twitris. Twitris continues to benefit from current collaborators, including Prof. Krishnaprasad Thirunarayan and Prof. Valerie Shalin, and team members Pramod Anantharam and Shreyansh Bhatt. This work is partially supported by NSF funded grants “SoCS: Social Media Enhanced Organizational Sensemaking in Emergency Response” (IIS-1111182) and “EAGER: Expressive Scalable Queries over Linked Open Data” (IIS-1143717).

References


Cross-References

- Collective Intelligence, Overview
- Dynamic Community Detection
- Futures of Social Networks: Where Are Trends Heading?
- Semantic Social Networks Analysis
- Spatiotemporal Footprints in Social Networks
- User Sentiment and Opinion Analysis