EventGraphs: Charting Collections of Conference Connections

Derek Hansen[1], Marc A. Smith[2], Ben Shneiderman[3]

[1]University of Maryland
College of Information Studies
Center for the Advanced Study of Communities & Information
College Park, MD, USA
dlhansen@umd.edu

[2]Connected Action Consulting Group
Silicon Valley, CA, USA
marc@connectedaction.net

[3]University of Maryland
Dept. of Computer Science
Human-Computer Interaction Lab
College Park, MD, USA
ben@cs.umd.edu

Abstract

EventGraphs are social media network diagrams of conversations related to events, such as conferences. Many conferences now communicate a common "hashtag" or keyword to identify messages related to the event. EventGraphs help make sense of the collections of connections that form when people follow, reply or mention one another and a keyword. This paper defines EventGraphs, characterizes different types, and shows how the social media network analysis add-in NodeXL supports their creation and analysis. The structural patterns to look for in EventGraphs are highlighted and design ideas for their improvement are discussed.

1. Introduction

Social network maps of the connections created through social media can reveal important patterns and insights about social life. While many aspects of the social world have been affected by widespread adoption of social media, physical gatherings for events, conferences, and conventions have changed in particularly interesting ways. These rituals for social networking have become networked.

Networks and mobile devices have changed the process of conferring and convening in many ways. Not long ago, people at conferences gathered around corkboards covered with slips of paper with names and short messages in them to try to make face-to-face appointments or reschedule missed ones. Today, comments from potentially many thousands of individuals participating directly or remotely in an event can be made immediately visible and searchable. Many conferences support official and unofficial "back channels" of conversations in chat rooms, via SMS, or on Twitter or Facebook.

Networked conferences and other events now generate and leave public archives of a wide range of digital objects. Tweets, blog posts, photos, status updates, slide decks, videos, and more are routinely created during most public gatherings. Inherent in most of these digital objects is data about relationships between the people who create content and those who comment and link to it. These relationships tie people to people and people to topics and other digital objects.

People worldwide now use tools like Twitter to spontaneously converge around unexpected events such as oil spills and ash clouds. Social media tools are becoming important venues and archives for political, social, cultural, and economic activists who provide a running commentary on current happenings. Increasingly, mobile-based social media is supporting collective action through "smart mobs" that form to protest or entertain [1].

Specialized search engines now make it possible to weed through the 50 million daily tweets to collect messages containing a common string, term, keyword, or "hashtag" (which use the "#" to prefix a topic grouping). This has reduced, although not eliminated, the costs of coordination for converging on commonly agreed upon terms to link content together. The question "What's the hashtag?" that often accompanies an event illustrates the interest and effort people take in order to adopt a common term in exchange for making their content visible to a particular audience.

But what do these new crowds, communities and populations look like? How do they vary from one another and over time? EventGraphs are an effort to build meaningful network graphs of the collections of connections created by those participating and discussing events.

The goal of this paper is to introduce and examine EventGraphs – a specific genre of network graph that shows the underlying social structure of people discussing an event in real-time. In particular, we focus on conference events as a sub-genre of time or topic bounded maps. After providing a definition and taxonomy of EventGraphs, we discuss how to create meaningful EventGraphs of Twitter conversations using NodeXL. We also discuss how to "read" them by identifying structural properties of interest and how to customize them by adjusting visual properties of the graphs such as node size and color. Finally, we discuss design ideas to better support the creation of meaningful EventGraphs.

2. Literature Review

While the core infrastructure now exists to rapidly collect, disseminate, and search masses of real-time messages, many insights into the structure and dynamics of these message collections remain out of reach. A battery of approaches including sentiment analysis, visualizations, and text summarization have been applied to distill insights from volumes of social media data. These are presented in traditional computer science conferences, mathematically inclined social science conferences, business conferences, human-computer interaction conferences, and a host of new conferences like the International AAAI Conference on Weblogs and Social Media (ICWSM) and the IEEE International Conference on Social Computing (SocialCom). Meanwhile, independent hackers and corporations produce new tools almost daily to make sense of social data, particularly large-scale, open data such as found on Twitter. Most of the existing approaches focus on the content and volume of the messages, delivering a trend tracking service. Systems like TweetStats and TweetVolume provide reports on a single user's activity over time.

In contrast, we focus on the relationships between those discussing in order to gain insights from the social structures associated with events. Related systems, for example MentionMap and Neuro Production's Twitter Browser, also explore the social network space in these systems but do so in a more limited manner. Our approach is to build on the robust metrics and visualization techniques developed by social network analysis (SNA) researchers.

Social network analysis encompasses the mathematics of graph theory, empirically-based social science studies, and computational models and algorithms of networks (e.g., [2]). These methods have been used to study computer-mediated communication since the advent of networked computing (e.g., [3-4]). However, until recently, network data collection, analysis, and visualization tools were only accessible to those with advanced training and often only after-the-fact. More recently, tools like NodeXL have made it possible for non-technical users with minimal training to make sense of network data captured from social media platforms in real-time [5-7]. It is now time to systematically consider specific genres of network visualizations, such as EventGraphs, that can provide insights to a much wider audience than academics, as well as how to design tools that effectively create them.

3. EventGraphs

3.1. Definition of EventGraphs

EventGraphs are social network diagrams that illustrate the structure of connections among people discussing an event via social media services like Twitter. They are a specific genre of the more general network graph, defined by connection to a real world event such as a conference. The example EventGraph in Figure 1 shows the social structure of those discussing the ten-day Washington DC festival called Digital Capital Week focused on "technology, innovation, and all things digital." It was created by collecting all tweets that use the hashtag #dcweek on Twitter on June 14th, 2010.

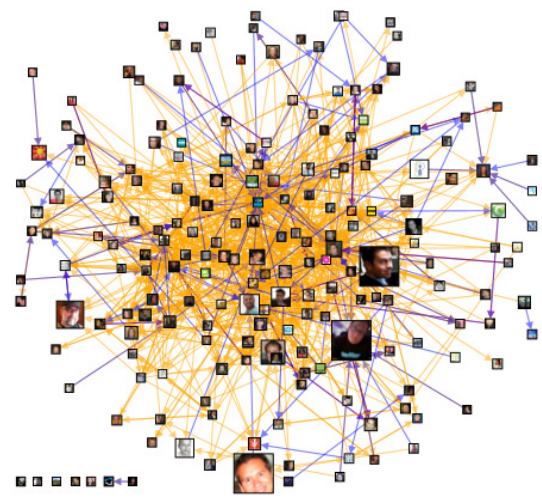


Figure 1. An EventGraph of "#DCWEEK" Twitter data on June 14, 2010 with image size mapped to total Twitter followers

Like all networks, the primary building blocks of EventGraphs are vertices (i.e., nodes) and edges (i.e., ties or connections). Vertices typically represent those discussing an event, and in Figure 1 are shown as Twitter profile images. They may also include organizations discussing an event, such as a news outlet or a professional association with an official voice (e.g., Twitter account).

There are two main types of connections that can be represented in EventGraphs, each of which can be instantiated in different ways:

- *Conversational connections.* These ties link people together based on conversational acts such as replying to another person, forwarding another person's message to others, or mentioning another person in a post. For example, using Twitter people can send tweets that are direct 'replies to' or more indirect 'mentions' of a person. These connections are vital components of the dynamic structure of event-based discussions. They are directed (i.e., asymmetric) since they flow in one direction. They are weighted (i.e., valued) because people may reply to someone multiple times, which can be captured as a single edge that varies in width in proportion to the volume of exchange. In Figure 1, Twitter 'mentions' are blue edges with arrows pointing toward the person being mentioned, while 'replies to' are red lines pointing toward the person being replied to. Edge colors from multiple edges are blended, so that purple edges have both a 'reply to' and a 'mentions' relationship between the individuals.
- *Structural connections.* These ties link people together based on explicitly created Friend or Follow relationships. Many social networking sites allow users to create an explicit and durable link to another user such as when two people become "Friends" on Facebook or when a Twitter user "Follows" another user. Even in the absence of any other connection or communication, these ties suggest a desire to read another user's content and/or an awareness and interest of one person in another. These ties may be undirected (i.e., symmetric) as with mutual Facebook Friend connections or directed as with Twitter Follow relationships which may not be reciprocated. These ties are unweighted (i.e., dichotomous or binary), since they either exist or do not exist. In Figure 1, Follow relationships are orange edges.

3.2. Taxonomy of EventGraphs

A variety of EventGraphs can be created. Different types of EventGraphs lend themselves to different analyses and structural patterns. We identify 4 key dimensions on which EventGraphs differ including the duration, frequency, spontaneity of the event, and the geographical dispersion of event discussants.

3.2.1 Duration of Event. The duration of the event is the first key dimension. While duration is technically a continuous variable, we distinguish between the following types:

- *Point Events.* These events happen at a single point in time. Examples include births, deaths, declarations of peace or war, announcements of awards, and so forth. Taken to the extreme, all action happens in time and is thus a point event, but only a handful of point events are significant enough to prompt real-time discussions by large numbers of people. Although point events happen at a single moment in time, the conversations about them may linger on much longer as evidenced by events such as the death of Michael Jackson or a major earthquake.
- *Hours-long Events*. Examples include a baseball game, concert, press conference, and workshop.
- *Days-long Events.* Examples include a conference, protest rally, space shuttle flight, and initial disaster response.
- *Weeks-long Events.* Examples include disaster responses, music festivals, election campaigns, and sports seasons.

3.2.2 Frequency of Event. Events may vary in their frequency. Some are repeated, while others are one-time occurrences. Many conferences, award ceremonies, sporting events, and elections are repeated events. The specific locations and dates may change, but a direct tie exists between repeated events. In contrast, many events such as a one-time workshop, a disaster response and recovery in a particular location, or a press release on a rarely occurring topic may lack much if any frequency.

Analysts of repeated events can compare the conversational and structural connections across events. Doing so for a series of annual conferences could allow researchers to track the progress of an academic field over time from a structural perspective.

3.2.3 Spontaneity of Event. Events can be planned ahead of time or occur spontaneously as a result of unexpected occurrences. For example, conferences, award ceremonies, sporting events, press releases, and protests are typically planned events. In contrast, natural disasters, celebrity deaths, and oil spills are unexpected.

Planned events often choose an explicit tag to act as a collector of conversations about the event. For example, many conferences encourage those discussing the conference to add a pre-specified Twitter hashtag (e.g., #CHI2010) to their tweets, blog posts, photos, videos, slides and related materials to help people find and follow all the conference relevant content. Spontaneous events may develop a shared tag over time, but don't have the luxury of having an "official" tag from the start, which can lead to multiple tags and more disjointed conversations. For example, Twitter users discussing the impact of the volcanic ash cloud over Europe adopted multiple hashtags including #ashcloud and #ashtag (a play on the word hashtag).

3.2.4 Geographical Dispersion of Event Discussants. Events also have a geographic dimension. EventGraphs vary in terms of the geographic proximity and density of the participants. In some events, such as specialized workshops or conferences, nearly all of the discussants are co-located. They may use Twitter to submit questions to presenters who can view them on a large display or have a moderator choose from them. They may also send announcements about the event itself, discuss presentations, and share links to resources discussed in talks with other attendees and a small group of remote observers. In contrast, popular conferences of interest to more than just the attendees allow people from across the globe to participate by posting questions, links, and commentary. This potentially provides them with 2-way access to the conference, with varying levels of integration into the conference itself. In this way, large groups of remote individuals can participate in events otherwise closed to them, at least in real time.

As more people geotag their messages, it will be increasingly easy to automatically assess the geographical distribution of the members in EventGraphs. Currently, even without geotags it is possible to identify the home time zones of participants, which provides a coarse, but insightful measure of geographic dispersion. Events that are international in scope may show up in many languages under varying terms posing significant challenges to accurately and exhaustively map a coherent global conversation. Analysts can use geographical data to understand the international reach of events and identify geographical clusters of interest.

3.3. EventGraph Data

EventGraphs can theoretically be created from many sources of data, although in practice technical hurdles or privacy considerations limit access. Rich EventGraphs are created by systems that support large-scale, real-time, public conversations. Streams of status updates from systems like Twitter, Facebook, LinkedIn, and Buzz provide potential sources for EventGraph data. Of these services, Twitter is arguably the most public and has an interface layer designed for software applications to collect data from the service. In contrast, many social media systems favor the exchange of private content that grant access to data only to pre-established friends or contacts. Social sharing sites like Flickr, YouTube, SlideShare and Delicious allow the use of tags that support the aggregation of multimedia content related to common events or topics. Threaded conversations are also sources of data for an EventGraph. While it is increasingly possible to automatically collect data in real-time from Application Programming Interfaces (APIs), many systems including Twitter limit access and grant larger quotas of queries against their service to those who have been “rate limit lifted”.

All sources of the EventGraph data have important qualifications and limitations. EventGraphs are only as accurate as the underlying data. Analysts must remember that EventGraphs only capture the conversations and social structure that is represented in the particular platforms studied, not the full range of human communication possibilities. Not everyone uses Twitter, and those who do not tweet will not appear in Twitter-based EventGraphs. Twitter EventGraphs of tech savvy conferences like South by Southwest attended by social media wonks create a denser network than conferences for late adopting populations. When there are systematic differences in who uses a tool and who discusses or attends an event, EventGraphs will be biased representations of the event. Furthermore, each person’s usage pattern will determine his or her prominence in EventGraphs, with power users showing up more prominently than passive users. If, however, the focus is on the Twitter discussion itself, then EventGraphs are accurate insofar as proper sampling techniques are used.

Sampling methods for selecting messages and/or authors related to a particular event are crucial. As mentioned, some conferences have a predetermined keyword or hashtag (e.g., #CHI2010) that helps identify relevant content. For more spontaneous events, sampling must be based on keyword searches of entire messages, which can pose problems. Poor precision can occur, for example, when a search for the visualization tool called “Tableau Software” pulls up French messages on completely different topics. Poor recall can occur, for example, when a search for “#ashcloud” misses the significant volume of Twitter messages about the ash cloud over Europe that used the alternative “#ashtag” instead.

In place of keywords, it will soon become possible to identify all messages from attendees who are co-located at events, which will increase recall but reduce precision since some of their messages may not be about the event. When there are too many messages to capture (e.g., given API rate limits and the vast volumes of content linked to popular or general interest topics), analysts must make due with a snowball or random sample of messages and authors. While these techniques may be sufficient to show key structural properties (e.g., subgroups, degree distribution), they do not provide a complete overview of the conversation network or identify all key individuals.

Despite these limitations, EventGraphs provide important initial insights and hypotheses, and can act as artifacts that spur self-reflection and conversation among event participants and observers. Analysts should be careful to know the limitations and not overstate claims without further analysis. Systematic comparison of conference attendee lists with data from EventGraphs can help characterize any biases.

4. Creating EventGraphs with NodeXL

We describe the workflow needed to create EventGraphs using the general-purpose social media network analysis and visualization tool NodeXL.

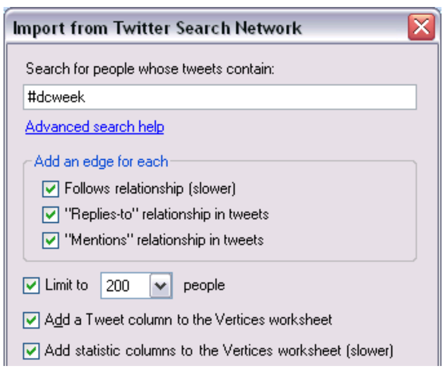


Figure 2. NodeXL 'Import from Twitter Search Network' Dialog

NodeXL is a free and open add-in for Excel that supports network analysis and visualization [5]. It couples the familiar spreadsheet layout with a graph visualization pane as shown in Figure 3. See http://www.codeplex.com/nodexl for the code and application. In addition to importing network data from edge lists, matricies, graphML, UCINet, and Pajek files, NodeXL can import data from various social media platforms such as Twitter, YouTube, email, WWW hyperlink networks, and Flickr via their APIs. NodeXL allows non-programmers to quickly generate useful network statistics and metrics and create visualizations of network graphs [6-7].

To create an EventGraph with NodeXL, data are collected using the 'Import from Twitter Search Network' dialog shown in Figure 2. To capture data on an event, a user must specify keywords or hashtags of interest (e.g., #dcweek). Twitter search syntax (e.g., Boolean operators) can be used. NodeXL then identifies the first 1,000 messages made available via the Twitter API (the time frame of messages returned depends on the popularity and volume of messages on the topic). It creates a network with 3 types of edges (Follows, Replies-to, and Mentions) as discussed in Section 3.1. It can also capture the most recent Tweet (i.e., message) and user statistics and information such as number of follow/following ties, profile image, time zone, last tweet date, and number of tweets. Finally, results can be limited to only a sample of individuals (e.g., 200 people in Figure 1 and Figure 2) to reduce the data extraction time.

Once data are collected, NodeXL can calculate overall graph metrics (e.g., density, number of components and isolates) and node-specific metrics (e.g., in-degree, out-degree, betweenness, eigenvector, and closeness centrality). Other metrics can be pasted in from other network programs if desired. NodeXL can also automatically calculate clusters to identify subgroups based on structural connections [8]. These data can then be mapped to visual attributes. For example, in Figure 1 size is mapped to number of Twitter followers and the profile images are used as the "shape". Advanced NodeXL features such as Dynamic Filters allow users to "play back" the conversation using the latest Tweet timestamps, filter out peripheral members, or switch easily between the 3 different edge types (follows, mentions, replies). Data in the spreadsheet can be sorted on graph metrics to quickly identify important individuals as shown in Figure 3.

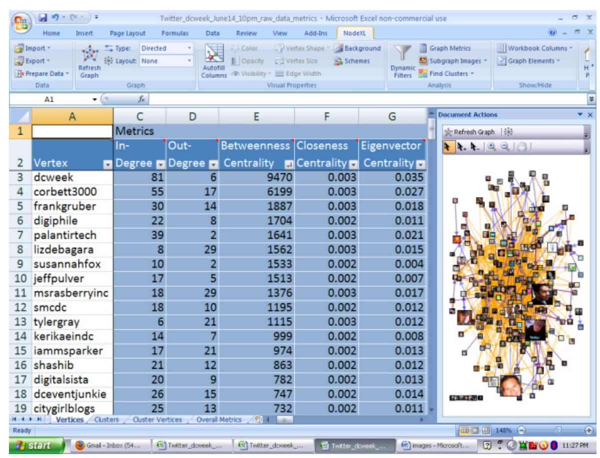


Figure 3. Connections among users who included #dcweek in a tweet, sorted from highest to lowest Betweenness Centrality

5. Analyzing EventGraphs

Gaining insights from EventGraphs requires knowledge of what social features to look for and what network metrics to consider. This section identifies how to "read" EventGraphs created from Twitter data and manipulate them in NodeXL.

5.1. What is the Social Structure of an Event Related Discussion?

The population of people who discuss an event may be a tight-knit group or community, a collection of separate subgroups, or a set of disconnected individuals. These collections of users may form a community centered on a few powerful individuals or form a more egalitarian and distributed structure. Several network metrics help quantify differences and similarities between social media networks, such as density, reciprocity, transitivity, connectedness, hierarchy, efficiency, and least upper bound [2]. EventGraphs created in NodeXL can visually represent many of these including:

- Size of the main component and its edge density. This measures how strongly connected the core community is by seeing how many of the possible connections are realized. The tangled ball of orange ties (high density) in Figure 1 suggests that #dcweek event participants know each other or have started following each other during the week.

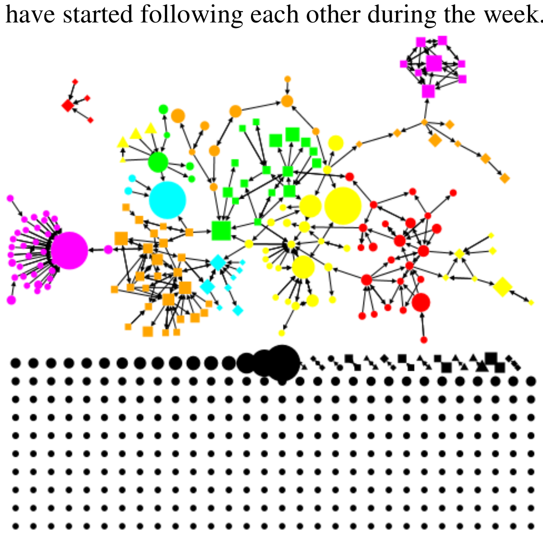


Figure 4. EventGraph of "oil spill" Twitter data from May 4, 2010 with clusters colored differently and size based on Twitter followers

- Fraction of the discussants that are not part of the core component(s). For example, the number of isolates, dyads, and triads who discuss an event but don't know or reply to any others who discuss the topic. NodeXL groups the isolates and small components together at the bottom of the graph in an ordered set of rows. Notice how many more isolates are shown in Figure 4 discussing the BP oil spill than Figure 1 on #dcweek.
- Number of subgroups in main component. For example, Figure 4 shows several distinct clusters of respondents to the "oil spill" event. Some include environmentalists, others include followers of a particular news outlet or celebrity, and others represent skeptics of environmentalists. NodeXL can automatically identify clusters of people using several "community detection" algorithms (e.g., [8]), which can be mapped to unique color/shape combinations (see Figure 4).
- Percent of follow relationships that are reciprocated (i.e., if I follow you, you follow me). Ties that are reciprocated suggest a mutual relationship that may be more stable. With the use of Excel formulas in NodeXL, reciprocated edges can be identified with different colors if desired.

5.2. Who are Important Event Discussants?

Social network analysis also provides a set of person-specific metrics that identify who is the most "central" to the community and/or conversation including in-degree, out-degree, betweenness, closeness, and eigenvector centrality. These help analysts determine whose comments reach the most people, who is most active in a conversation, who are peripheral members with high influence elsewhere, and who spans across subgroups if they exist.

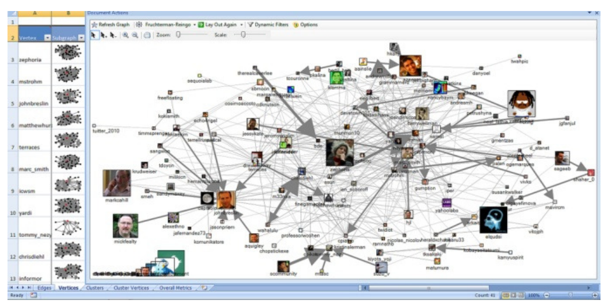


Figure 5. EventGraph of connections among Twitter users who mentioned ICWSM on May 25, 2010 scaled by number of followers, edge weights vary by multiplexity, colored frames map to betweenness.

Measures of centrality, as well as Twitter usage metrics (# of tweets or followers) can be mapped to visual properties such as size, opacity, and color to identify people with unique roles. NodeXL also supports the creation of subgraph images (i.e, egonetworks) of each individual and those who connect to them directly (see left-hand side of Figure 5). Different flavors of EventGraphs can be customized to identify different types of important people. E.g.:

- Node size can be mapped to total Twitter followers, as a measure of a person's reach. Large nodes on the periphery of a given EventGraph (e.g., bottom of Figure 1 and left-hand side of Figure 5) identify people likely to take news of the event to a broader audience. Large nodes in the center of the graph suggest well-known, frequent Twitter users who are central to the event. Small nodes near the center of a graph suggest people who are central to the community, but not well-known beyond the group.

- Node color can be mapped to automatically identified clusters (Figure 4) and node size to betweenness centrality. This can help identify subgroups and those who span those subgroups.
- Composite metrics can be created using formulas in NodeXL and mapped to size or opacity. Such metrics can identify people who score highly on multiple metrics (e.g., a Twitter power-user score can identify people who tweet often, have many follower/following ties, and joined Twitter early).
- Filters can be applied to remove people who are more peripheral and focus on the key discussants.
- The X and Y coordinates can be used to plot important individuals based on their participation in the conversations and/or their network centrality. As seen in Figure 6, a correlation exists between tweets and followers, but not everyone converts tweets to followers at the same rate. Below the diagonal are those who over convert tweets to followers, those above the diagonal under convert tweets to followers.

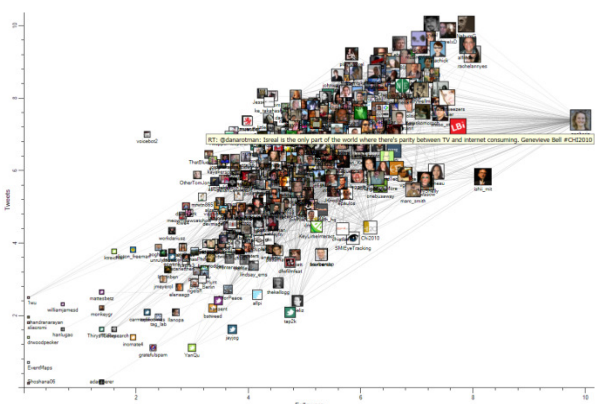


Figure 6. EventGraph of Twitter users who mentioned "CHI2010" on April 12, 2010. X-axis = Log(Followers), y-axis = Log(tweets) scaled by number of followers.

5.3. What is the Nature of the Event Conversation?

Does the conversation include many people? Is it a true conversation or a collection of isolated comments by people that use a shared tag? What people and subgroups are most active in the conversation? Who initiates conversations and who replies to them?

Understanding the conversation around an event on Twitter requires looking at EventGraphs that emphasize reply and mention networks rather than the follow network (see Section 3.1). Doing so allows you to see how many people are conversing, if there is a shared conversation or collection of isolated comments by people using a shared tag, which groups and people are most active, and who initiates conversations versus replies to others. These networks are less dense than follow networks, often including more separate components. This makes calculating complete-graph metrics (such as betweenness centrality) or clusters inappropriate in many cases. In- and out-degree are always appropriate as they are simply indications of the number of messages received/sent and forwarded to others. Edge weight (typically represented as edge thickness and/or opacity) can be used to indicate the number of messages exchanged between people.

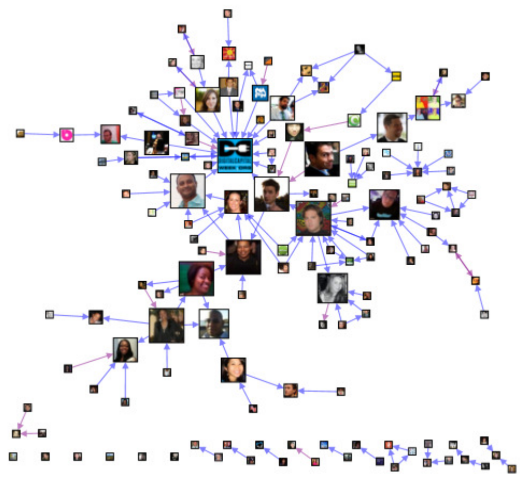


Figure 7. EventGraph of discussion network from #dcweek Twitter data with node size mapped to betweenness centrality

Figure 7 shows a discussion network that is a filtered view of Figure 1, including reply and mentions relationships among Twitter users who tweeted with the hashtag #dcweek. Graphs like this illustrate:

- Multiple components indicate sub-conversations that make up the larger event discussion. Although Figure 7 includes only one large component, there are several smaller components or "wings" of the large component, suggesting that the conversation is not a shared one among all participants. Instead it is a collection of smaller conversations that are loosely joined by a few bridge spanners (larger nodes with high betweenness centrality scores).
- Individuals with high in-degree indicate that many people mention them or reply to them. For example, in Figure 7 the Twitter account "dcweek"

has the highest in-degree since many people posted messages mentioning or replying to it.

- Individuals with high out-degree indicate that these authors mention or reply to many others.
- Thick edges can indicate active exchanges between pairs of individuals.

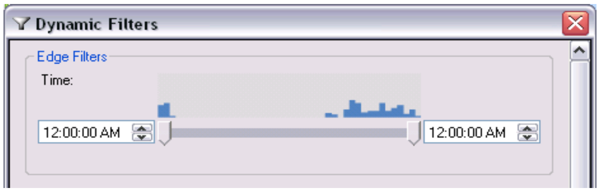


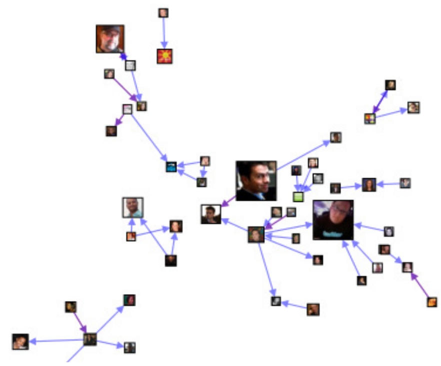
Figure 8. Dynamic filters with sliders on each end and a histogram of each value above the sliders

5.4. Tracking EventGraphs over Time

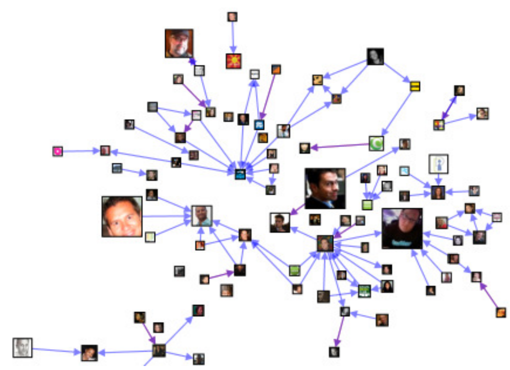
Examining EventGraphs over time can improve our understanding of how a conversation unfolds or the effect of an event on the shape of friend and follow networks. Conversations occur over time, with messages sent in reply to other messages or messages and ideas getting forwarded to others in a distributed network. EventGraphs can help evaluate the effect of an event on structural connections between event participants (e.g., does the network density of a Twitter follow graph increase substantially after a conference? Does it forge connections between formerly separated subgroups?). EventGraphs can also identify important conversation starters and help characterize how ideas propagate through the network.

Visually representing network changes over time is an active area of research with great opportunity for improvements. Here are a few techniques possible using NodeXL for understanding network graphs:

- Edges representing follow relationships or the exchange of messages are time-stamped in NodeXL. Dynamic filters can then be used to move through time by sliding the beginning and ending time periods (see Figures 8 and 9). This allows analysts to see a conversation or a structural network play out over time, while fixing the location of the nodes. Graphical distributions are shown above the dynamic filter (see Figure 8) to indicate the prevalence of different values.
- The opacity (or width) of edges can be modified based on the timestamps so that older edges are darker, while newer ones are lighter. This helps capture time in smaller, less dense networks, but does not scale well to large, dense networks.
- A series of EventGraphs can be shown, each capturing a different time slice. If the emphasis is on understanding individuals, then fixing their location is preferable (NodeXL allows analysts to specify the X-Y coordinates of all nodes). However, if the emphasis is on larger structural patterns identified by network layouts such as Fruchterman-Reingold, then fixing node location is not preferable.
- NodeXL allows analysts to schedule automatic downloads of data from social media platforms.



6pm EST



9pm EST

Figure 9. Two EventGraphs of "#dcweek" after using Dynamic Filters to only show edges occurring before 6pm and before 9pm. Size maps to Twitter followers.

5.5. Comparing Related EventGraphs

Looking at a single conference-based EventGraph (e.g., Figure 1) doesn't fully make sense until compared against other conference-based EventGraphs that show higher density, more distinct subgroups, a different distribution of active Twitter users, etc. For example, comparing EventGraphs of an annual conference in an emerging interdisciplinary field would likely show an increase in density and reduction in subgroups as the field becomes more established. Alternatively, EventGraphs of the same event based on different hashtags (e.g., #ashtag vs. #ashcloud) help identify and characterize different subcommunities.

Supporting a more rigorous comparison of networks that underlie EventGraphs is an active area of research as evidenced by ManyNets and related efforts [9].

6. Designing for EventGraphs

Creating and analyzing EventGraphs is hardly trivial, particularly if designing them for a non-technical audience for whom network analysis concepts and metrics are new. While the current version of NodeXL (version 1.0.1.126) supports the creation and analysis of a variety of EventGraphs, there are exciting possible designs that will enable significant improvements. This section provides some design ideas and challenges based on our experience in creating and sharing numerous EventGraphs.

6.1. Data Collection & Filtering

EventGraphs are only as good as the data they are based on. There is need for tools that import and integrate data from multiple online sources. While NodeXL imports data from a several sources, there is no current support for integrating data across these, even when they share a common tag (e.g., #CHI2010). Grabbing data from social media aggregators or new cross-platform standards like Activity Streams is a promising approach. However, many of these tools do not yet natively support the creation of network data structures. Furthermore, designing usable import interfaces for complex data structures is challenging.

Another data collection challenge is to increase recall and precision of event data. When a hashtag is announced for planned events, not everyone knows about it, suggesting a need for automatic query expansion. Query refinement may be needed as well if distinct groups unwittingly use the same hashtag.

As discussed, spontaneous events often do not have a pre-specified tag to capture relevant conversations. There is a need for event-detection services that automatically identify bursts of activity among a group of connected and/or co-located people, or around a particular topic (as identified by Natural Language Processing techniques that use message content rather than pre-established tags). A proactive version of an event-detection service would send a message to all likely event participants proposing a hashtag, thereby speeding up the process of community formation.

Dealing with large-scale events (e.g., the world cup) poses new data collection and filtering challenges. There are obvious solutions like centralizing data storage. However, there are also less obvious solutions that would help create and evaluate "random" sampling from the network without distorting the overall network properties. Filtering techniques may also be useful, which allow data to be collected on only select individuals or subgroups of interest. While NodeXL allows this to occur after an entire network is downloaded, it is feasible to allow this to occur at an earlier stage in the data collection process to reduce unnecessary data draws.

6.2. Merging Network & Attribute Data

One useful technique is to combine network data that shows connections between event participants and attribute data that describes those individuals or the messages they exchange. NodeXL currently does this by capturing networks as well as usage data on Twitter users (e.g., number of tweets, join date, number of followers). However, natural language processing and sentiment analysis of message content can significantly enrich EventGraphs. For example, nodes and edges could be colored or sized based on the sentiment of their messages (green = positive messages; red = negative messages). Looking at an EventGraph of a corporate merger announcement could then allow you to easily see how each company perceives the merger alongside the relationships of corporate employees. Alternatively, NLP could help identify messages with conflict, helping draw attention to disagreements and the larger social context in which they occur.

Text analysis techniques can also automatically identify important attribute data such as people, places, and organizations. Network transformation could make these entities nodes and connect them together based on their co-occurrence in conversations.

6.3. EventGraph Exploration and Analysis

EventGraphs should not be thought of as static, even though those presented in this paper are necessarily so. Exploring EventGraphs in NodeXL using dynamic filters (see Section 5.4), mapping different visual attributes to nodes, and sorting on the underlying spreadsheet data, allow users to get much more insight out of interacting with these networks. This is particularly true of large EventGraphs that can become overly cluttered and lose their detail. Supporting effective exploration and analysis of networks is an active area of research that has many opportunities [10].

Many network tools, including those that tightly integrate network visualizations and data metrics (e.g., NodeXL, SocialAction [10]), require the use of a desktop client application. More recently, sites have begun to embed network exploration tools inside of web-based tools (see [11] for an early example). Network browsers allow people who don't want to create an EventGraph to still benefit from them.

6.4. Integrating EventGraphs and Events

One promising opportunity is to more closely integrate EventGraphs with live events such as conferences. The authors have presented EventGraphs of a specific conferences or workshops to attendees on many occasions. They are met with great interest from participants and are likely to have increased the level of participation and/or inspired new connections between attendees, although systematic evaluation remains to be done. Audiences quickly identify key people in the EventGraphs, which usually include prominent conference organizers and researchers mixed with less familiar individuals who frequently Tweet and thus play an important role in the Twitter community. EventGraphs can be a powerful tool to promote self-reflection about a particular event (e.g., iConference) or community around that event (e.g., iSchools). They can also be used to suggest and evaluate social interventions. For example, the top 20 influencers (identified by network and Twitter metrics) could be asked to Tweet about a conference session and the growth in size of the main component could be measured to evaluate its effect.

Local views of EventGraphs, such as an ego-network for a single event participant could be attached to an individual's event profile or nametag to spur further conversations and make new connections. Systems like nTAG, SpotMe, Poken, or Minglesticks that help people connect with others that have shared interests could pull data from social media sources that underlie EventGraphs. Integrating EventGraphs with distributed events is more challenging, but still possible. An online EventGraph infrastructure could be created that would allow event participants to explore, discuss, annotate, and compare EventGraphs.

7. Conclusions

An EventGraph is a specific genre of network graph that shows the social structure of people discussing an event. They can show structural connections (e.g., follow relationships on Twitter) or conversational connections (e.g., replies or mentions on Twitter) between co-located or dispersed event discussants. They may discuss events of different duration or type (spontaneous vs. planned; repeated vs. one-time).

EventGraphs can be used to understand the social structure underlying an event, identify key people related to an event, map the conversation around the event and track it over time, and compare related events. Although EventGraphs are useful in their current instantiation as created in NodeXL, advances in data collection, integration with NLP techniques, and better support for interactive exploration and analysis will allow even more useful EventGraphs in the future. We hope this characterization of EventGraphs will inspire designers, analysts, and event planners to find better ways of mapping the social worlds that surround the prominent events in our lives.

8. References

[1] Rheingold, H., *Smart Mobs: The Next Social Revolution (First Edition.)*, Perseus Publishing, Cambridge, MA, 2002.

[2] Wasserman, S. and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, MA, 1998.

[3] Freeman, L. C., The Development of Social Network Analysis: A Study in the Sociology of Science, North Charleston, SC: BookSurge, LLC, 2004.

[4] Garton, L., C. Haythornthwaite, and B. Wellman, "Studying Online Social Networks", Journal of Computer-Mediated Communication, 3(1), 1997.

[5] Hansen, D. L., B. Shneiderman, and M. A. Smith, Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann, Burlington, MA, 2010.

[6] Bonsignore, E. M., C. Dunne, D. Rotman, M. Smith, T. Capone, D. L. Hansen, and B. Shneiderman, "First steps to NetViz Nirvana: evaluating social network analysis with NodeXL", SIN '09: Proc. IEEE Computer Society Press, Vancouver, Canada, 2009.

[7] Hansen, D. L., D. Rotman, E. Bonsignore, N. Milic-Frayling, E. Rodrigues, M. Smith, and B. Shneiderman, "Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Data", U. of Maryland Tech Report: HCIL-2009-17, July 2009.

[8] Newman, M. E. J. and M. Girvan, "Finding and evaluating community structure in networks", Phys. Rev. E 69 026113, 2004.

[9] Freire, M., C. Plaisant, B. Shneiderman, and J. Golbeck. "ManyNets: an interface for multiple network analysis and visualization", ACM- CHI '10: Proc. ACM, New York, NY, 213-222, 2010.

[10] Perer, A. and B. Shneiderman, "Balancing Systematic and Flexible Exploration of Social Networks", IEEE Transactions on Visualization and Computer Graphics 12, 5, 693-700, 2006.

[11] Heer, J., and D. Boyd, "Vizster: Visualizing Online Social Networks", Proc. IEEE Symposium on Information Visualization, 2005.