

TOPSY

Coalescing of Communication

In Social Media

140 Second Street
Second Floor
San Francisco, CA 94105

www.topsy.com

Contents

- i Introduction
- ii Key Questions
- iii Analysis
- iv Summary and Conclusions
- v About Topsy

Introduction

Social media has become the de-facto medium for humans to communicate about events as they are occurring. Whether it's a natural disaster such as a tornado or earthquake, a social movement such as the Arab Spring or Occupy Wall Street, or a popular event such as a concert or the Olympics, social media has become the mainstream method for all people involved to share their thoughts with each other.

In many instances, it's not immediately obvious what people are talking about as an event unfolds, since those discussing the event tend to refer to it in a myriad of ways. As a result, it is initially rather difficult to fully aggregate the conversation. But, what occurs over time (minutes, hours or days, depending upon the event's longevity) is that people discover the terms others are using to describe the same thing they are experiencing. Within Twitter these are typically expressed as hashtags, and people begin to reference these terms in their own communication about the event, making it easier to aggregate and quantify the communication describing the same thing.

Through the analysis of hundreds of billions of tweets over the past couple of years, we've observed this distinct cycle time and time again: initial fragmentation is followed by a later unification of terms or hashtags within the Twittersphere, and we refer to this as the coalescing of communication within social media. For widespread, persistent events --- such as the Arab Spring or Occupy Wall Street - we also consistently see a secondary fragmentation after coalescence. In this final step in the cycle, the coalesced terms break off into sub-threads that refer to local or derived applications of the initial event.

This white paper presents findings showing some of the coalescing patterns we've observed using real-life data points.

Key Questions

- *How do the words people use to describe events and topics on the social Web evolve over time?*

The implications of answering this question are clear: if you don't know how people are describing an event, or if you happen to be tracking terms which have subsequently changed, your analyses will be missing some crucial information.

- *What are the stages of communication throughout the duration of these events?*

If you know that people are just beginning to speak to each other about an event, you should track many different terms referencing it in order to get an accurate portrait of the conversation. If, however, discussions have already coalesced around a common term or set of terms, you can be confident that you're analyzing most of the communication. If expressions have evolved into sub-threads of the initial reference, you should also be aware of these derivations, as they are essential for a complete analysis.

- *Can you track when and where certain hashtags become popular?*

The ability to identify where and when certain terms start to become popular puts you ahead of the communication curve. Furthermore, if you can isolate the users creating or amplifying popular terms, you can track their posts for clues about future discussions.

From an organization's perspective, it's clear that being able to address each of these questions in realtime by analyzing the full corpus of conversational activity from Twitter provides benefits from a business, social and political perspective.

"If... discussions have already coalesced around a common term or set of terms, you can be confident that you're analyzing most of the communication."

"The ability to identify where and when certain terms start to become popular puts you ahead of the communication curve. "

Analysis

The phases of communication we consistently see as events unfold can be categorized into three areas:

- Phase I: Fragmented Buzz
- Phase II: Coalescence of Terms
- Phase III: Subsequent Fragmentation & Sub-Threads

There are many benefits in being able to listen to online conversation. However, these benefits exponentially increase when you can also discover and predict when that conversation will coalesce, and identifying further fragmentation allows for a uniquely powerful grasp on the fluidity of social chatter.

To demonstrate these benefits in action, we've shared some of our findings related to the communication phases of the Occupy Wall Street movement.

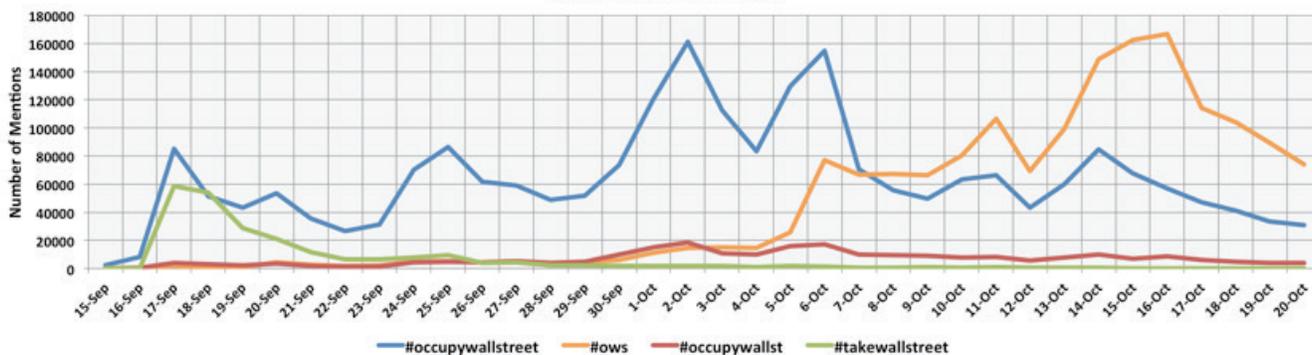
Phase I: Fragmented Buzz

In the initial stages of a discussion, both participants and spectators tend to use a wide range of terms to refer to the same overall developments. When the discussion has begun at the level of the people—such as the grassroots movement fueling Occupy Wall Street (OWS) did—the initial fragmented buzz of communication is even more noticeable, since 'official' terms often do not exist yet. The social web is a perfect place for such young grassroots movements to gain momentum amongst potential supporters, and, consequently, is also a convenient place to begin analysis of the way people around the world are discussing it.

When it comes to OWS in particular, the first month of protests in New York City sent a flood of new terms and hashtags into the Twittersphere. Using Topsy's historical tracking capabilities, the following graph (Figure 1) presents the earliest set of popular hashtags used to refer to the Occupy Wall Street protests. Initially popular terms such as #takewallstreet and #occupywallstreet gradually become replaced by the more easily-typed #ows as the first few weeks of the movement passed.

“In the initial stages of a long-standing discussion, both participants and spectators tend to use a wide range of terms to refer to the same overall developments “

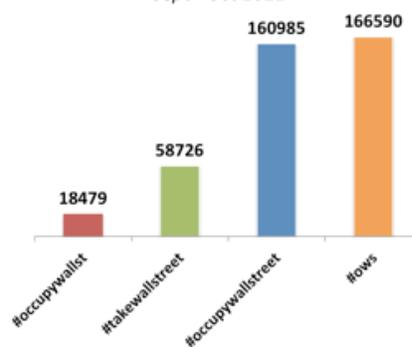
Figure 1: Number of Mentions of Popular Occupy Wall Street hashtags
September-October 2011



In Figure 2, we observe the peak daily values of the most popular hashtags referring specifically to the OWS protests during the first month of major activity.

Though the #ows hashtag appeared in discussions much later relative to the others, its daily peak quickly rose to the highest position at 166,590 mentions—right around the same date as the Worldwide ‘Day of Rage’ to support OWS on October 15, 2011.

Figure 2: Peak Daily Mentions
Sept - Oct 2011



Phase II: Coalescence of Terms

After the initially fragmented phase of communication discussed above, we can often observe precisely when most users on the social web simultaneously decide to use the “best” of the set of popular hashtags. Whether it is due to a completely new addition of meaning, a catchy abbreviation, or simply ease of typing (as was the case with #ows), when one of the hashtags takes off in peak popularity, we very suddenly observe a distinct shift in the way most users refer to the movement.

“...we can often observe precisely when users simultaneously decide to use the ‘best’ of popular hashtags.”

In the case of #ows, we have pinpointed the first instance of the term in a suggestion from an Occupy follower on August 13, 2011:



Heather Marsh
@GeorgieBC



#OCCUPYWALLSTREET needs to become
#OWS @USDDayofRage #USDOR

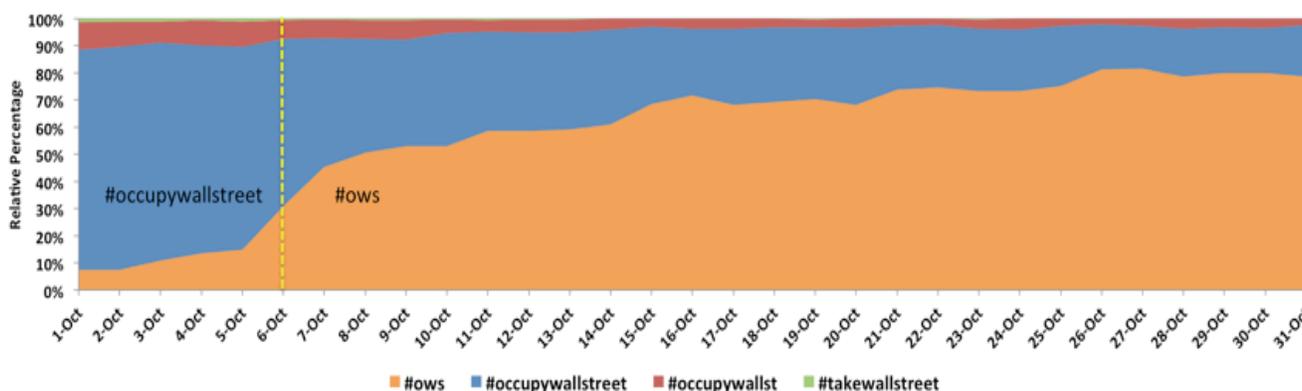
The term did not take off, however, until October 6, 2011, when the Occupiers themselves asked Twitter followers directly to use the more easily-recognizable #ows:



Let's stop the hash tag soup and use **#ows** for OccupyWallStreet

This characteristically honest and open form of online communication to supporters has proven to be an extremely effective tactic at propagating the use of unique terms. We can see in the share-of-voice analysis below that after October 6th, the conversation surrounding the Wall Street protests became almost completely dominated by the #ows hashtag:

Figure 3: Occupy Wall Street #hashtag Coalescence
October 2011



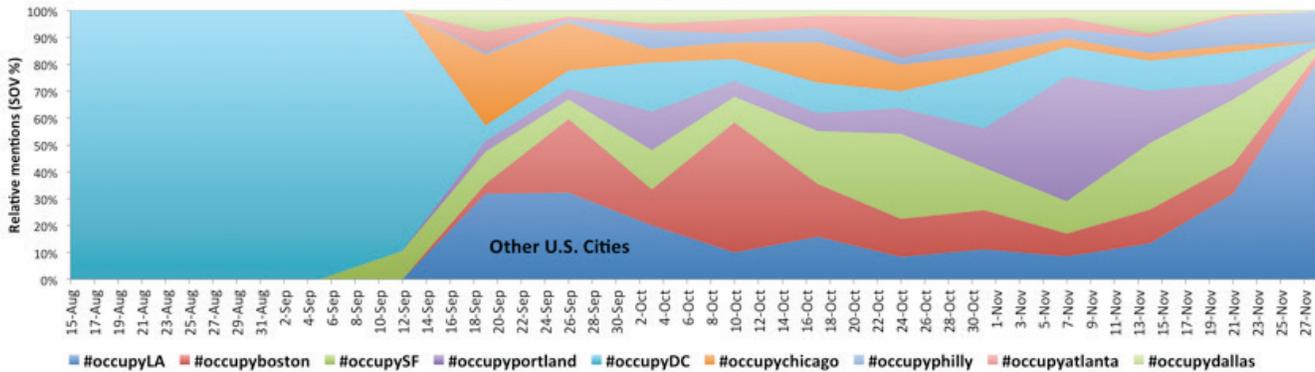
Phase III: Subsequent Fragmentation and Sub-Threads

Once the most popular terms take off, the conversation returns to a fragmented chatter similar to that seen in Phase I. However, this time around, there is one crucial difference: the fragmentation itself comes from applications of the most widely-used terms, which simultaneously refer to both the original event and related or derived events in other situations.

In the case of the Occupy Wall Street protests, supporters of the movement's general premise (that the highest concentration of the country's wealth should not fall upon only 1% of the population) began applying the "Occupy" term to their own cities scattered throughout the country in solidarity. As seen in Figure 4, the number of mentions of #occupy applied to other US cities began very early on in the movement with our nation's capital, Washington, D.C.

Once the protests on Wall Street began receiving national press coverage in September, we see rises in the relative percentages of more outside cities using the #occupy hashtag:

Figure 4: #Occupy-ing other U.S. Cities
Share-of-Voice through November, 2011



In light of observing the applications of hashtags to other cities, it is also useful to pinpoint exactly where original terms become popular. Table 1 below presents the top five countries in which the #occupy hashtag was used in the period of 15 September – 1 November 2011, using Topsy’s geo-inference method. After the U.S., high numbers of mentions in the UK, The Netherlands, Canada, and Germany were very much expected, especially considering the success of the London, Amsterdam, Toronto, Vancouver, and Berlin protests. Subsequently, the hashtags #occupylondon, #occupyamsterdam, #occupytoronto/ vancouver, and #occupyberlin all displayed hikes in worldwide mentions.

“...it is also useful to pinpoint exactly where original terms become popular”

rank	country	mentions
1	USA	301,436
2	UK	26,140
3	Netherlands	17,676
4	Canada	12,891
5	Germany	8,822

Summary and Conclusions

This analysis has used the conversation activity accessed from the full Twitter firehose surrounding the worldwide #Occupy movement to examine three phases in communication we typically observe when people describe the unfolding of an event:

Phase I: Fragmented Buzz

During this stage, people refer to the event or topic by many different names and hashtags.

Phase II: Coalescence of Terms

People discover what term(s) others are using to describe the same event and reference this term in their own communication.

Phase III: Subsequent Fragmentation and Sub-Threads

If the event or topic resonates with a wide audience over time, we begin to see versions of the original term applied to different realms (referred to as sub-threads); while these new terms do technically refer to derivations of the initial topic, they also inherently refer to the initial topic itself.

These communication stages are mostly observed in topics which begin at the level of the people. We've observed the same patterns with other recent events: for example, elements of the chatter surrounding the tragic shooting of Trayvon Martin, the Spring 2012 Canadian student protests, and the London Olympics all follow the same fragmentation-coalescence-fragmentation arc. The conclusions and implications from these analyses impact any organization that cares about what people are saying within social media outlets. Specifically, this paper has highlighted the following suggestions:

- Realtime analyses of all social communication surrounding an unfolding event are needed to even consider tapping into and quantifying the amount and type of communication occurring. Batched analyses utilizing samples of data will be outdated and inaccurate.
- Trend detection is essential. Specifically, measurement systems which take into account tweet volume, momentum, and acceleration of terms are invaluable in this realm, as they allow for early detection of significant patterns in resonating terms.

“The conclusions and implications from these analyses impact any organization that cares about what people are saying within social media”

- Discovery of evolving terms must be part of social event analyses. A person studying the social communication around an event will typically know a handful of terms they think are important. Human communication evolves and eventually those initial terms become blinders if the user is not able to discover and learn what new terms people are using to describe an event over time.
- Obtaining the source of an amplified conversation over time is important for understanding the credibility of shared information. If an important post about an event has been widely retweeted, but the content originator has consistently tweeted only jokes in the past, then credibility should be called into question.

It is important to note here that large events such as the Occupy Wall Street movement are not only type of events these conclusions and implications apply to; product, political, brand and entertainment related conversations similarly begin at the level of the people and evolve over time.

Topsy provides instant access to realtime and extensive historical analyses for any keyword, term or link using the world's largest public index of Twitter data. Users are able to discover new patterns in communication as they are occurring, quantify the volume of communication and measure how people are expressing themselves within Tweets around the globe, isolating communication by language, geography and influence.

About Topsy

Topsy provides deep, comprehensive analyses of hundreds of billions of Tweets and web pages gathered from millions of unique websites, blogs and social media services. Topsy's platform leverages these social conversations to index, analyze and rank content and trends. With instant access to conversations from moments ago or years ago, Topsy enables users to discover, quantify, predict and make decisions using the world's most powerful social analytics products.

"... product, political, brand and entertainment related conversations similarly begin at the level of the people and evolve over time"

"Topsy provides instant access to realtime and extensive historical analyses"

Topsy's solutions are used by some of the largest global marketing, news, entertainment, and financial organizations to stay on top of breaking events, design and optimize content and communication, quantify market trends and customer feedback, measure competitor mindshare and predict what's about to happen.

Headquartered in San Francisco, Topsy is backed by BlueRun Ventures, Ignition Partners, Founders Fund and Scott Banister.

For more information, contact us at bizdev@topsy.com about.topsy.com