Project Titles

- **G1. Witty Title**
  - Stephanie, Seean, Olina

- **G4. Mine Influential Users of Brands on Twitter for Product Sponsorship**
  - Suparna, Amanda, Wali

- **G5. Information Propagation on Twitter**
  - Muftiat, Claire, Nina

- **G6. Power of Popularity**
  - Chris, Macodou, Josue

- **G15. The Infectiousness of Complaining on Twitter**
  - Joseph, Troy, Kevin
Group 1

Meaningful Memes

Stephanie Niedoba, Seean Kim, and Olina Xie
CORPORATE TWITTER ACCOUNTS AND MEMES

Why should we care about how corporations portray themselves on Twitter?
Twitter as a Social Platform
Twitter is the Fastest Growing Social Platform

Why Does Twitter Matter?
Twitter proves to be the fastest growing Social Media Platform today. It has a very dynamic network structure and the network relations are always changing. We have already learned a lot about the popularity, brevity, and live action nature of tweets. So why do companies care about their twitter presence? The twitter newsfeed truly acts as a stream of consciousness of its users, and to flood this stream of consciousness with advertising is an incredibly valuable marketing asset.

Some Info
- Over 350K Tweets are posted each minute
- 288M Users
- Content Sharing

65M
SOCIAL MEDIA COMPUTING

How do the topics presented in this course (CMSC489J) relate to our project?
Taking Advantage of Weak Ties
How Twitter is Making Reaching the Entire Graph a Breeze

- Strong and Weak Ties
  Are two easy ways to categorize the links in our social graph

- Strong Ties
  Are strong links between users.

- Weak Ties
  Represent users you may follow, but do not DM or retweet from often

- Hard To Reach Spots
  Can be accessed from links and jumps using weak ties.

- Six Degrees of Separation
  The distinctly separated clusters can be reached through weak ties.

- 288M Active Users
  Is a hefty and desirable audience. However, if a company manages to make their tweet into a meme, even more can be gained as the tweet travels through other social media outlets such as Reddit, Tumblr, or Facebook!
Degree Centrality

The Company Twitter Account Acts as a Central Node

Diagram showing clusters and connections between Verizon, AT&T, T-Mobile, and other clusters of users and followers.
Power Laws & Rich Get Richer Phenomena
The Uneven Distribution of Where Most of Our Twitter Content Comes From

Power Laws
Power Laws and Rich-Get-Richer Phenomena are incredibly important.

Content
Even though Verizon, AT&T, and T-Mobile all share similar content, their follower count is different.

Unpopular
An unpopular account can be a difficult feat to overcome on Twitter.

Verizon
Actually proves to be the most popular account with over 1.7M followers. (This is actually more than AT&T and T-Mobile combined)
Cascading Behaviors in Networks

How a Company is able to Gain More Followers

- **Following An Account**
  When considering our lessons involving cascading behaviors in a network we can think of following a new twitter account as a desired new behavior.

- **Importance of Memes**
  While a couple of popular tweets may seem fleeting, they have the potential to trigger a behavior cascade.

- **Homophily**
  We understand that links connect with similar characteristics. These characteristics can be immutable via selection or mutable via social influence (or socialization).

- **Membership Closure**
  If user A follows Verizon as a “focus” and user A also has a strong tie with user B, user B is highly likely to exemplify triadic closure and also become interested in Verizon’s account.

**Graph Theory & Game Theory**

The Lessons we have learned in Social Media Computing can help us understand how corporations gain a large twitter following. Popular tweets have the potential to trigger a behavioral cascade of users following the cooperate account. We can also understand the persistence of behavioral cascades using other previously learned concepts like homophilly, triadic closure, and membership closure.
OUR FINAL THESIS

We aim to investigate the effect of informal vs. formal tweets posted by a brand on the number of retweets that it receives.
Examples and Inspiration

Some Tweets that inspired our Thesis!

**Jessica Jones**
In response to a tweet from @billyhammock who tweeted "Trying to finish watching Jessica Jones. Keyword is trying. This is hands down my least favorite Marvel tv/film property. So boooooooring."

**Taco Bell**
In response to a tweet from @MensHumor reading "If Taco Bell delivered, they would make so much money today. #420".

**Tesco Mobile**
Tesco sasses a twitter user who tweets "Immediate turn off if a girl’s mobile network is tesco mobile" and becomes an internet sensation for a bit!
Our Dataset
How have we evaluated our thesis?

Based on Professor Hadi’s recommendation, we decided to limit our dataset to the three powerhouse telecommunication companies: Verizon, AT&T, and T-Mobile. Originally we planned to use Hadi’s “Churn Prediction in Social Media” dataset, but these tweets tended to be from users. In order to accurately evaluate our thesis we pulled over 3,000 of the most recent tweets from each company, and then sorted them based on the amount of favorites and retweets the tweets receive.
FORMAL VS INFORMAL

One of the most difficult questions that we needed to address was how do we distinguish formal and informal tweets?
The Language Of Twitter

The Language of Twitter is surprisingly different than most people expect. Given our research and understanding of how Twitter users communicate, we came to the conclusion that it would be best to closely analyze the top 20 tweets from each company for the purposes of our research. [NOTE: This if we were to pursue this project further, it would be valuable to thoroughly research the computation quantitative ways tweets can be classified. The current method of evaluation is much more qualitative and likely more objective because of that].
THE POTENTIAL PROS AND CONS

We spent a large amount of time investigating the alternative aspects of our thesis and analyzing the research and perspectives of others.
Potential Benefits

There are a lot of advantages in creating a popular tweet!

The most obvious advantage of creating a viral tweet is reaching a large audience. On top of being seen by a wider group of people, it also creates a lasting impression, especially with the younger generations. The concept of going viral is more relevant than ever, and the effects of creating such a tweet can help change the image of a company to be hip and fun to the younger demographic.
While using Twitter for corporate social media can be an extremely powerful tool, it can also backfire quickly and cause more harm than good. A frequent and unfortunate instance of this backfiring can be seen in companies who try to jump in on trending hashtags and topics without truly understanding the background or the origin of a tag. Twitter should be used deliberately and with caution to avoid becoming offensive or even infamous.
CONCLUSION

Based on our analysis of the top 20 retweeted tweets of all time from Verizon, AT&T, and T-Mobile we have concluded the following...
Final Results and Observations

Cumulative Retweets From Top 20 Dataset

It takes balls to admit you’re wrong—so we called in @IAMSteveHarvey to school @Verizon. #BALLOGIZE

RETTWEETS LIKES
28,430 43,124

4:38 PM - 7 Feb 2016
Closing Presentation

As previously mentioned, there are a handful of ways this project can be expanded and improved upon. The dataset extracted uses only three corporations, and the use of more companies spanning a wider variety of products and practices could definitely improve the global relevance of any discoveries made. Additionally a larger dataset that pulls more than 3,000+ tweets and qualitatively evaluates more than 20 tweets per company could also greatly improve the accuracy and confidence in our conclusions.

Furthermore, retweets are not the only metric to measure a popular tweet. We could look at favorites and other things like @replies which generate conversation about the tweet. It would also be wonderful if we could find a way to measure how far the tweet goes when it extends beyond twitter and begins to pervade the content of other social platforms like Facebook, Reddit, and Tumblr.

Finally, computational linguistics and quantitatively evaluating human language is an incredibly challenging and impressive field! The imaginative applications of CL allow for unbelievable innovations, and this field lays the groundwork for an unreal future possibilities! [go see the movie Her (dir. Spike Jones)] A stronger expertise in this field could have infinitely improved the accuracy and focus of our research.

Stephanie Niedoba
Seean Kim
Olina Xie
THANK YOU FOR WATCHING
Mining Influential Brand Users for Product Sponsorship

C. Suparna Barua, Amanda Lotwin, Waliullah Rifai
Advertising is expensive, and often inefficient, however most companies rely on effective advertising for company growth.
Solution

- Identify happy customers of a certain brand
- Compensate users for advertising brand to followers
- Factor each user’s compensation based on user reach
- Ensure that user is not advertising for competitor brands
Evaluation Method

**Twitter**
We determined Twitter to be the most useful network to create a solution.

**Datasets**
We used a social graph, Twitter database, as well as already formatted CSV files.

**Rule Based Classifier**
We used a classifier to identify the customers of a certain brand.

**Twitter Crawler**
To help us calculate influence, our crawler finds essential tweet popularity statistics.

**Sentiment Analysis**
We weeded out any users with negative brand sentiment using Java's Stanford Core NLP.

**Calculation of Reach**
The reach of each user is stored in a max heap, each node containing a twitter ID and a reach score.
Datasets

- Twitter
Twitter

Tweets

User ids

Followers
Rule Based Classifier
Classifier Criteria

**Follower Tweets**
If > 30% of a user’s followers tweet about a brand, and less than 10% of their followers tweet about each competitor, they are a customer.

**Twitter ID**
The twitter ID and screen name of any potential customer is stored.

**Target Brand**
Users must have less than 10% of their followers tweeting about competitor brands to ensure minimal overlap.
Twitter Crawler
Crawler

Retweets
The crawler finds the number of times a user’s tweet about a brand was retweeted and favorited which is factored into the calculation of the reach.

Brand Loyalty
We also use the crawler to find the percentage of a user’s tweets that are about a certain brand. This data is fed into the classifier to identify the real customers.
Graph

Followers
We used the social graph to collect a list of a user’s followers to be utilized in our calculations of customer criteria.

Identify what percentage of followers tweet about the target brand versus the rival brands.
Stanford Core NLP

Natural language analysis tool
Sentiment Analysis

Single sentence positive/negative classification accuracy rate: 85.4%

(http://nlp.stanford.edu/sentiment/)
Sentiment Analysis

Positive sentiments
Classify customer’s tweets as positive by processing and testing for the majority weight of key positive words

Negative sentiments
Classify customer’s tweets as negative by processing and testing for the majority weight of key negative words
Language Specifics and Storage
## Language Use

<table>
<thead>
<tr>
<th></th>
<th>Python</th>
<th>Java</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSV and JSON Parser</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Rule-Based Classifier</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Twitter Crawler</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>
Data Storage

HashMap

i.e.

AT&T @Theon_reek @joffrey
Verizon @nedAlive @rob_north
Data Storage

Max Heap for each brand

i.e.

- **@obama**: 143
- **@sanders**: 105
- **@trump**: 89
Final Result

Output list of all brand customers (name and screen_name) who have tweeted positively about the brand, in descending order based on the reach of their tweet.
Thanks!
Resources and References

- https://business.twitter.com/help/what-are-promoted-tweets
- http://simplymeasured.com/blog/twitter-metrics-defined-reach/#sm.00000ttwvydnqff2jzyfovn8kgb9y
- http://www.anthology.aclweb.org/P/P11/P11-1016.pdf
- http://rahular.com/twitter-sentiment-analysis/
- http://stanfordnlp.github.io/CoreNLP/sentiment.html
- J. Yang, J. Leskovec. Temporal Variation in Online Media. ACM International Conference on Web Search and Data Mining (WSDM '11), 2011.
SOCIAL PROPAGATION ON TWITTER

muftiat ogunsanya
claire pitman
nina dinh
PROBLEM

How do sentiment and number of followers predict the number of followers who will also tweet with a specific hashtag?

This could allow real time prediction of the number of followers who will tweet with the same hashtag when you use a hashtag (campaigning/ advertising strategy)
PROPOSAL

original goals:

analyze spread of positivity and negativity in social movements

if a person were to tweet something negative, how likely is it that the followers will do the same?
use to Twitter Archiver to obtain sets of tweets based on different social movement hashtags

#FeelTheBern
#MakeAmericaGreatAgain
#BlackLivesMatter

data: tweet, user, time of tweet, number of retweets, number of followers
APPROACH: PREPROCESSING

Sentiwordnet and NLTK: calculate sentiment score $[-1, 1]$

Twitter API and Tweepy

- get followers
- get user IDs for each tweet
final processed data includes:

tweet, user, time, # followers, # retweets

sentiment, # followers who tweeted the same hashtag
EVALUATION

Run prediction on a different hashtag

Check error
DIFFICULTIES

Twitter API rate limits: wait it out (15 minutes)

dataset size vs. time

to obtain # followers who tweet with the same hashtag, need Twitter API to get IDs from screen names and list of follower IDs

compromise:

get list of follower IDs for first 100 tweets

get IDs for users who tweet over span of 12 hours (~ 600)
RESULTS + INSIGHTS

Weak correlation between sentiment and number of followers who tweet with the same hashtag
IMPROVEMENTS

acquire a larger processed data set for more accurate testing and training

test for ability of triadic closure presence to predict number of followers who tweet with the same hashtag
WILL THEY RE(TWEET)?

PRESENTERS: MAC DIOUF, JOSUE GOMEZ, CHRISTOPHER HOANG NGO
SOCIAL MEDIA ON THE RISE

• Social medias play a key role in today’s society

• Some of the uses:
  • Communication
  • Sharing information on large scale
  • Advertising
  • Business management
THE TWITTER NETWORK

• Follower/followed relationship:
  • Follower gets a feed of posts by individual who’s followed.
  • Nodes (users) linked by edges (follow relationship).
  • The more followers one has, the more people they can reach.

• Users and tweet, retweet, favorite, and reply tweets
  • A retweet will share a tweet from a user one follows to their own followers.
WHERE BUSINESSES COME IN

• Aside from an initial tweet by the company’s account, retweets play an important role in advertising.

Meek Mill take it from us- if you gonna serve beef serve it high quality
11:34 PM - 30 Jul 2015

104,308 replies 78,699 likes
SO WHAT’S THE ISSUES?

With the propagation of numerous amounts of information across a social network, what inclines a user to continue the flow of information or data?

Problem:

Given a hashtag, will it be retweeted by a said user?
WHY IT MATTERS!

Determining influential factors in a network and how users react to these elements provides variable information in multiple fields.

• Targeted Marketing
• Political Action campaigns
• Product Branding
WHY IT MATTERS CONT.

• **Cost of Advertisement**
  • Companies spend a LOT of money on advertising.
  • The bigger the target audience, the higher the cost.

• **Propagation**
  • Clever way to target a specific audience which will then cascade information onto a wider audience without any extra work or cost.
Given a hashtag, will it be retweeted by a central Node?

- Within our analysis we found that the attributes of a central Node I.E maintaining edges to a majority within a network, did not provide any new influential factors in which we could create a metric.
THESIS

• Our algorithm will predict whether someone will tweet or retweet something pertaining to an original keyword (hashtag in our case).

• Using that information, one can tweet at a user, who will then retweet it that to their followers.
WordNet
A lexical database for English

TOOLS AND APPROACH

gceopy

API
We used geolocation as a mutable variable and an influential factor.

Location of tweets provides surrounding context which we incorporated in our predictability algorithm.

Homophily:

The principle that we tend to be similar to our friends
“History repeats itself”

User behavior, we investigated whether a user held any history of using the hashtag.

Sourced WordNet ntlk to also seek for synonyms of hashtags to parse tweet text.
APPROPRIATE METRICS STANDARD

Followee (friend) behavior:

we measured how many friends held a tweet with given hashtag. As the number of friends that tweet with the given hashtag (and synonyms) increased, the chances of studied user will retweet will also increase.

Conforming to a collective outcome.
ALGORITHM

Threshold = 1/3
Score = 0;

For each user tweet
    if user tweet text contains hashtag or synonyms
        score += 3;
    End for

For each friend adjacent to user
    if friend tweet text contains hashtag or synonyms
        score += 2;
    if friend share any geotag with user
        score += 1;
End for
ALGORITHM

If score \( \geq \) threshold return true user will retweet/tweet hashtag

If score = 0 return false user will not retweet/tweet hashtag

If 0 < score < threshold inconclusive.
THOUGHTS GOING IN.

• User behavior would be the most impactful metric

• Threshold should be $1/2$ of possible score.

• User tweet would be predictable in span of one week.

• Wanted to run algorithm on max friend of 50, max 600 tweets

  looking over 30,000 tweets.
RESULTS

Running our program in three day intervals,

We found that our algorithm held a 70 percent accuracy rate

“Take with a grain of salt”

• Short time interval, what about month intervals?

• What about larger data sets?
RESULTS

Will the hashtag “Syria” be tweeted

<table>
<thead>
<tr>
<th>User</th>
<th>Score</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>BarackObama</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
<tr>
<td>jms_dot_py</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
<tr>
<td>CloseDanger</td>
<td>0.190050251</td>
<td>Inconclusive</td>
</tr>
<tr>
<td>bellingcat</td>
<td>0.40782776349</td>
<td>Will Tweet</td>
</tr>
<tr>
<td>YouTube</td>
<td>0.02</td>
<td>Inconclusive</td>
</tr>
<tr>
<td>PalantirTech</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
<tr>
<td>TheEllenShow</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
<tr>
<td>Grifter801</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
<tr>
<td>CNN</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
<tr>
<td>jtimberlake</td>
<td>0.0</td>
<td>Will NOT tweet</td>
</tr>
</tbody>
</table>
Retrieved Tweets

- bellingcat
- CloseDanger
- jtimmerlake

User: Green
friend 2: Orange
friend 3: Red
friend 4: Blue
friend 5: Purple
friend 6: Light green
friend 7: Light green
friend 8: Brown
friend 9: Brown
friend 10: Purple
INSIGHT

What we found ...

1) Followee behavior on average held higher tweet action with given hashtag. In comparison to studied user.

2) Geotag theoretically useful, but proved useless. Geotag isn’t commonly used while tweeting.

3) Threshold of ½ was too optimistic

4) Time span of a month or year greatly changes results.
REFLECTION

• Calculate a more accurate threshold
• Run algorithm on definitive data set instead of live stream
• Hashtags are time sensitive
• Results does not incorporate all user data Tweepy request rate limit caused us to set max friend to 10, max tweet to 200 for time efficiency.
• Algorithm could potentially be used on central node for advertisement purposes.
Measuring Engagement of Negative Brand Sentiments on Twitter

Troy Edleman
Kevin McCloskey
Joe Dyer
Number of people (in millions) who plan to switch cell phone carriers within the next 12 months in the United States

User A expresses a negative sentiment towards their cell phone carrier on Twitter. That carrier's (or a competitor's) support account recognizes this negative tweet and responds. $G =$ gap in time between the complaint and the carrier's response. **What is the effect of G?**
Why analyze the response time of cellular carriers to negative tweets?

- New integrative form of customer service
  - People don’t expect a result from complaining on the internet. “[only] 18.7% (215/1152) questions received at least one response” -Paul et al

- Large social media presence may lead to elevated sentiment in consumer pool

- Quick response time may lessen the negative sentiment in one customer

- Unconfronted negative customer sentiment will spread memetically, similar to word-of-mouth branding (Jansen et al., 2009)
Social Media presence in General

- **Social Capital**
  - Burt (2000, p.3) argues, "better connected people enjoy higher returns".
  - For companies, more connections means higher customer satisfaction rates and more sales.
  - Social media > advertising
- **Access**
  - "receiving a valuable piece of information and knowing who can use it"
  - Service providers have the most access (Raquel Recuero et al)
    - Already have the valuable information
What happens when a tweet goes unresponded?

- “we observed that the craftiness of a saying influences its “retweetability”. This can be reflected in tweets having parallel syntactic structure, which can be captured by sentence similarity within a tweet” – (Suh et al)
- Homophily in play
  - People follow and retweet those whose opinions are similar. So much so that they tweet similarly
- Social influence at play
  - Opinions change based on tweets from those they follow.
  - Cascade of negative opinion
- Companies closing window on rebranding that cascade
  - Best case scenario cascade is positive
  - Average case is they stop cascade
- Bringing it all back
  - How fast do they have to get there?
  - How long can a tweet go unresponded?
Opportunities for new customers

@BrigidH27

@ATT I called back a 3rd time, and the rep told me I was eligible for a free upgrade (!) ... then he hung up on me :(  

@BrigidH27 @ATT

Hate u verizon :(  

@cassygarciaaa

Sprint Forward @sprintforward · Apr 29
@cassygarciaaa It's time for something different Cassy. You can switch to Sprint for free at: sprint.co1WYsGaH ^CR

Verizon Wireless CS @VZWSupport · Apr 29
@cassygarciaaa We want you to love us. What's going on? How may we further assist you?  

*KB

Verizon @verizon · Apr 12
@BrigidH27 Quality customer service is just one part of the better network. DM us for deals to help you switch. spr.ly/6014BY17q  

*MAB
- **Data Entry:** Negative tweet and first official response from support account

- **Timeline:** April - May 2016

- **Companies analyzed:**
  - AT&T
  - Sprint
  - T-Mobile
  - Verizon

- **Key Tweet Fields:**

  - `created_at`: ‘Tue Jul 02 11:43:18 +0000 2013’
  - `id`: 351897277478076417
  - `user{id}`: 528140042
  - `user{screen_name}`: ‘giwaang’
  - `text`: ‘i hate verizon’
  - `retweet_count`: 2
  - `favorite_count`: 1

---

**Tweet:**

@Dukegirlforlife

I HATE Verizon!!! CAN NOT wait til September when my contract ends and I can find something else

5:12 AM - 22 Apr 2016

---

Verizon Support @VerizonSupport · Apr 22

@Dukegirlforlife We don’t want you to feel that way! Can you tell us more about what’s going on? ^CCR
Dataset Collection Methodology

Users complain about brands on Twitter

Data Collection Script
TwitterSearch API

AWS / PostgreSQL instance
with tables per brand

Custom configuration
for each brand

config.json files

Data Analysis Scripts
Search Keywords

- Seeded searches with 29 negative keywords
- Chosen by SentiWordNet rating & by manually examining negative tweets
- Examined tweets with positive SentiWordNet score > 1.0 and removed if they were deemed non-negative
Challenges

Collecting Live Twitter Data

Limited Twitter API

- Rate Limits

- No intuitive method for gathering responses to a tweet

Small sample size

- Just over 3000 tweets between 4 brands

False Negatives

@MarioMalvado

@verizon My mom won't make me toaster strudels :(  
6:36 AM · 13 Apr 2016

@BAPfromtheHart

really hate people that buy tickets just to resell for an egregious amount; drake & future tix at Verizon center for $50 reselling for $1030  
3:00 PM · 29 Apr 2016
Results: AT&T

- **Total Tweets/Responses Collected:** 605
- **Mean Response Time:** 31 mins 23 sec
  - 85.02% Responded To Before Mean
  - 14.98% Responded To After Mean
- **Median Response Time:** 11 mins 03 sec
- **Response Time Standard Deviation:** Approx. 96.23 mins
Results: Sprint

- **Total Tweets/Responses Collected:** 1313
- **Mean Response Time:** 1 hr 38 min 58 sec
  - 75.15% Responded To Before Mean
  - 24.85% Responded To After Mean
- **Median Response Time:** 14 mins 37 sec
- **Response Time Standard Deviation:** Approx. 201.97 mins
Results: T-Mobile

- Total Tweets/Responses Collected: 646
- Mean Response Time: 9\text{mins }56\text{sec}
  - 84.06% Responded To Before Mean
  - 15.94% Responded To After Mean
- Median Response Time: 3\text{mins }15\text{sec}
- Response Time Standard Deviation:
  - Approx. 32.29\text{mins}
Results: Verizon

- Total Tweets/Responses Collected: 786
- Mean Response Time: 8 mins 54 sec
  - 70.98% Responded To Before Mean
  - 29.02% Responded To After Mean
- Median Response Time: 5 mins 36 sec
- Response Time Standard Deviation: Approx. 21.19 mins
- Companies employ different strategies for engaging disgruntled customers on Twitter
- The company with the lowest average G has the largest subscriber base (Verizon)
- Twitter and other social media services are becoming important customer service resources
- Using tools like ours, companies can analyze their customer service engagement on social media
- Apply to larger datasets
- Analyze different types of companies such as airlines, automobile manufactures, restaurant chains, etc...
- Determine ultimate outcome of customer and company support interaction
- Jansen, B. J., Zhang, M., Sobel, K., and Chowdhury, A. Twitter Power: Tweets as Electronic Word of Mouth. Journal of American Society for Information Science & Technology, 60(11), 2169-2188
Questions?