

INTRODUCTION

Twitter

In the past few years there has been an enormous spike in the amount of time people spend on social networks. From MySpace and Facebook to LinkedIn and Twitter, social networks have been increasing in size to the point where there are several hundred million users that connect and interact with each other on these networks. Each network has its own purpose and is defined by its own unique interactions. Some networks have mutual relationships (e.g., Facebook and LinkedIn) while others have only one-way relationships (MySpace and Twitter). And some networks are geared towards professional relationships, while others are geared toward personal relationships.

Due to the popularity of social networks businesses have started interacting and engaging with users on these networks. The purpose of this study was to gauge how effective various interaction methods are for companies based on the network. Specifically, to measure how effective certain types of tweets are for businesses hoping to spread their message, promotions, and news as far as possible.

As of April 2010, Twitter has more than 100 million users with a daily sign-up rate of 300,000. Twitter also generates more than a billion tweets a day. Twitter is the ideal network to study because of its news-broadcasting characteristics. Facebook limits the reach of information to particular users based on their friendship and other social networks like MySpace and LinkedIn – which aren't as popular as Twitter – don't have as many status/news updates as Twitter.

Business

Businesses and organizations have an invested interest in understanding how their information is being shared both within and outside of their network of followers. They also want to know why information is shared in certain circumstances and not others to ultimately understand the value of their network is (i.e., what is the value of single follower, what is the value of their collective followers, what is the marketing value of their network, etc.).

The business that we teamed up with to measure the effectiveness of different tweets was an Ann Arbor-based clothing retailer Underground Printing. Underground Printing is a clothing retailer that focuses on the collegiate apparel segment. Underground Printing also specializes in custom screen-printed apparel.

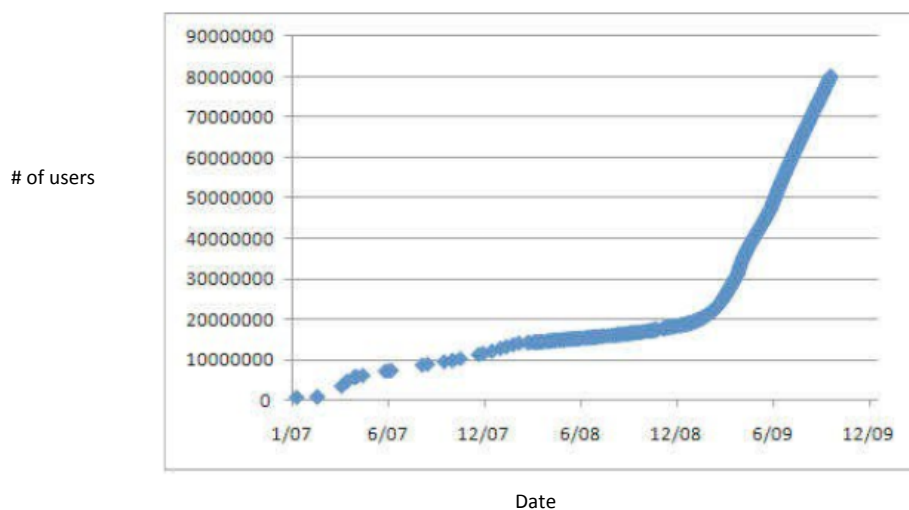
In the early part of 2010, Underground Printing began to expand their social networking presence. Part of their social networking efforts included a weekly promotion on Twitter each Tuesday. As of December 2010, Underground Printing has more than 2,000 followers for their Twitter network with a 2-4% weekly increase. Their tweets are often retweeted (RTed) as part of

their weekly promotions, as well as for their general tweets. We were allowed to collect and analyze their twitter data over a 2-week period.

AIM

Our motivation to conduct this analysis was to provide a method for any organization to analyze their social network. In a very short time, Twitter's growth curve has shown a sharp increase from an obscure segment of the web to a must-have for business identities. Twitter proves itself highly useful for individuals and organizations to connect and exchange information. While there are plenty of concerns (e.g., security, management, and interruptions issues) there are also benefits for organizations using the social networking site. Many of these advantages deal with the real-time spreading and gathering of information. Other benefits include gaining competitive intelligence, brand awareness, feedback mechanisms, and human resources assistance.

Figure 1. Increase in usage of Twitter from May 2007 to April 2009



Currently, many organizations use Twitter for broadcasting; however, not much focus have been directed towards the social network analysis of actual Twitter activities. In order for a business to make use of Twitter, a critical analysis of its network activity between itself and its followers was needed. With the approval of Underground Printing (UGP), we were able to access their actual business account and send out experimental tweets.

Goals

The two primary analyses of this study were to visualize the effectiveness of certain tweets and how these tweets flowed through Underground Printing's (@UGPCustomShirts) network of followers. These would be essential for developing best practices in regards to broadcasting and diffusing information throughout Twitter. Our aim was to provide an analysis of what type(s) of content are most effective for retweets (RTs).

Underground Printing currently broadcasts two types of tweets: promotional and informational tweets. **Promotional tweets** are those that are used solely for marketing purposes (e.g., to promote discounts, offers, free giveaways, etc.). **Informational tweets** are used solely for providing general information related to the organization (e.g., new product info, press releases, and upcoming events).

We wanted to analyze whether promotional or informational tweets drew the most RTs among Twitter users. We also wanted to analyze whether including a link would increase the amount of RTs for both promotional and informational tweets. Our initial hypothesis was that promotional tweets using an attachment would draw the most RTs among users.

METHODOLOGY

To validate our hypothesis, we conducted an experiment. We sent out weekly tweets through the @UGPCustomShirts Twitter account. These tweets were either promotional or informational tweets (depending on the week), that were sent with and without links. There were four types of tweets: 1) promotional with link, 2) promotional without link, 3) informational with link, and 4) informational without link.

We then analyzed followers based on whether they RTed or not in order to identify networks of active followers who promoted content posted by the @UGPCustomShirts Twitter account. This analysis also included segregating the results of the four tweet types in order to determine the best type of tweet for information diffusion on the Twitter network.

Figure 2. A sample output of the JAVA Code using Twitter4J API

```

Open the following URL and grant access to your account:
http://api.twitter.com/oauth/authorize?oauth_token=ry98wThba0LmeAqFsSonkuajmQnqjTyafzn0NTrX2vM
Enter the PIN(if available) or just hit enter.[PIN]:9749151
48781437
token : 48781437-dff0HcwpNc6vV7BZVh2Qj03auw0QY6Dj10oFJrGgFF
tokenSecret : KK0Mcmk1LubiZC4ZrIjjA7AmZ4BGyXaMSKNqY3cXQdZs
Enter Search Query:@UGPCustomShirts
#DrogoBoffin - Got Shirts? #FF @saucewear @sogeshirts @UGPCustomshirts @GritFX
@Sassy_Sadie - @ugpcustomshirts Grandma Got Run Over By a Reindeer is my favourite! http://ow.ly/3qV09
@durstslovepens - #FF @iGo @AbolitionistJB @Cowbells @UdderlySmooth @cafePress @PghZoo @McDonalds @themanregistry @NewYorkLife @EdgeShaveZone @
@gingergeri - @UGPCustomshirts http://www.youtube.com/watch?v=83uUFxHivFc&feature=fvst Let It Snow!
@ginette4 - @UGPCustomshirts We are the World http://www.youtube.com/watch?v=TZdn7NKNQqQ
@elphokitty - @UGPCustomshirts http://www.amazon.com/Pimp-My-Sleigh/dp/B001H5W1K4 <Pimp my Sleigh. It's awesome.
@MityDad - #FF @pinkblush99 @UGPCustomshirts @BlackBerryHelp @pcmag @durstslovepens @cheesecubes523 @luv_mydachsund @dinaj @interstatebatts
@thunderbirdsGO - #FF @GardeningExpres @UGPCustomshirts @zllondonzoo @empiremagazine @dolphinmusic @MissHighStreet @GurgleUK @PrimaMag @NatMa
@twinklenicci - #FollowFriday @WicknaxCandles @RealVimto @blackberrycampus @lifestylish @Rituals @UGPCustomshirts @MyKiddiStore (via @ffhelper
@eyewonit - RT & Follow @UGPCustomshirts for a chance to win an #AWESOME HOLIDAY SURPRISE from us #TwitterTuesday #Sweeps #free #14L http://
@wildflower00 - RT & Follow @UGPCustomshirts for a chance to win an #AWESOME HOLIDAY SURPRISE from us #TwitterTuesday #Sweeps #free #14L http://
@eyewonit - RT & Follow @UGPCustomshirts for a chance to win an #AWESOME HOLIDAY SURPRISE from us #TwitterTuesday #Sweeps #free #14L http://ow
@SnoWhite37 - Thank you @UGPCustomshirts!!! I won their Twitter Tuesday Sweepstakes! Happy Holidays indeed. Need a custom printed shirt? Give U
@eyewonit - RT & Follow @UGPCustomshirts for a chance to win an #AWESOME HOLIDAY SURPRISE from us #TwitterTuesday #Sweeps #free #14L http://ow
User 1: DrogoBoffin---User 2: Sassy_Sadie---Friendship: false
User 1: DrogoBoffin---User 3: durstslovepens---Friendship: false
User 1: DrogoBoffin---User 4: gingergeri---Friendship: false
User 1: DrogoBoffin---User 5: ginette4---Friendship: false
User 1: DrogoBoffin---User 6: elphokitty---Friendship: false
User 1: DrogoBoffin---User 7: MityDad---Friendship: false

```

We made a java program using the twitter4j API to collect data. A major challenge in using the API was its recent transition to OAuth (Open Standard for Authorization), which requires users to hand out tokens instead of credentials. This allows users to share their resources with any application without having to hand out their credentials. After authentication with OAuth the

program performed a search query for users that retweeted @UGPCustomShirts' tweets. We were able to perform search queries because we initially assigned every tweet with unique hashtag.

After the search results generated a list of users who had retweeted @UGPCustomShirts' tweets the program then compared each user with one another to determine any "relationships" (i.e., did they follow that person and/or did that person follow them). The twitter4j API provides the existFriendship(user1, user2) function that returns the value "True" if there is a relationship or "False" if there exists no ties between the followers.

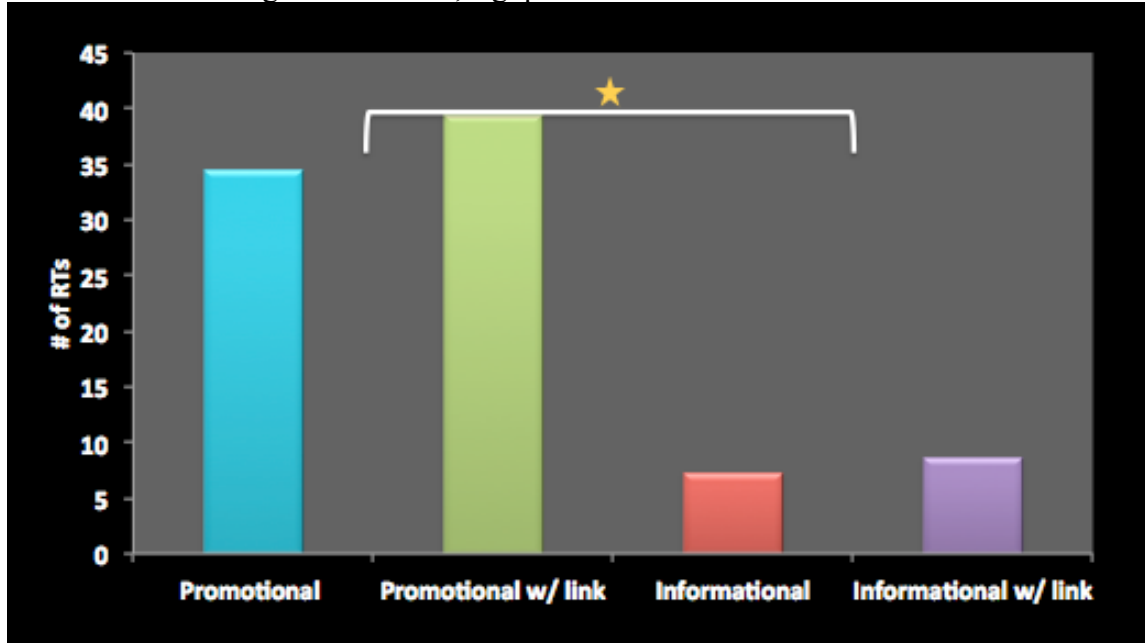
[Note: There were constraints that limited our analysis for this section. The twitter API only allows for 350 requests from a single account. For sending more queries the account has to be given "white-listed" access that can only be done by filling out and sending a web-based request form to Twitter. Although we did fill out the form we never heard back from Twitter. We proceeded with running the code multiple times with different sets of users and checking that the number of requests was below 350 for each hour. Also, because of a popular bug with twitter4j API, we were only able to find whether the two followers were following one another. We were unable to fetch information about who was actually following whom. Hence, the sub-graph that we used for analysis was an undirected graph whereas the real Twitter network is directed. Once the twitter API becomes more mature and servers become more efficient for developers our study could be extended for providing more profound results.]

We have attached our Java Code in the appendix. In the next section we will present the data analysis that we formulated with the help of this code.

DATA ANALYSIS

A one-way analysis of variance (ANOVA) was calculated on mean number of RTs for promotional without link, promotional with link, informational without link, and informational with link. The analysis was significant, $F(3, 40) = 97.252, p = .000$ (see Figure 3.). We found participants RTed promotional tweets more than informational tweets. Participants also RTed promotional tweets with a link ($M = 39.364, SD = 7.09$) more than promotional tweets without a link ($M = 34.273, SD = 7.38$). Participants RTed informational tweets with a link ($M = 7.456, SD = 3.69$) more than informational tweets without a link. The differences between promotional tweets with and without a link and informational tweets with and without a link were not significant, $t(20) = -1.650, p = .114$; $t(20) = -0.857, p = .401$, respectively.

Figure 3. There is a significant difference between promotional and informational RTs, where Promotional tweets get RTed more, sig. $p < .01$.



Looking at the propagation of RTs over time, we were able to determine which tweets seemed to last longer over time. By graphing the number of RTs by the number of hours the tweet was RTed, we found that promotional and informational tweets with a link seem to be RTed the most over time. We did notice towards the tenth and eleventh hour of the tweets being RTed that both kinds of tweets were being RTed the same amount. See Figure 4 and 5.

Overall, in this study we found that promotional with link tweets were RTed the most (only significant when lumped with promotional tweets without link and compared with informational). Clients are better off using promotional tweets, but if a client wanted to use informational tweets they would be better off using a tweet with a link. Tweets with links get RTed the most over time, but as time gets longer the use of links does not have an effect on number of RTs.

Figure 4. Informational tweets over time. RTs are being RTed the same amount at the tenth and eleventh hours.

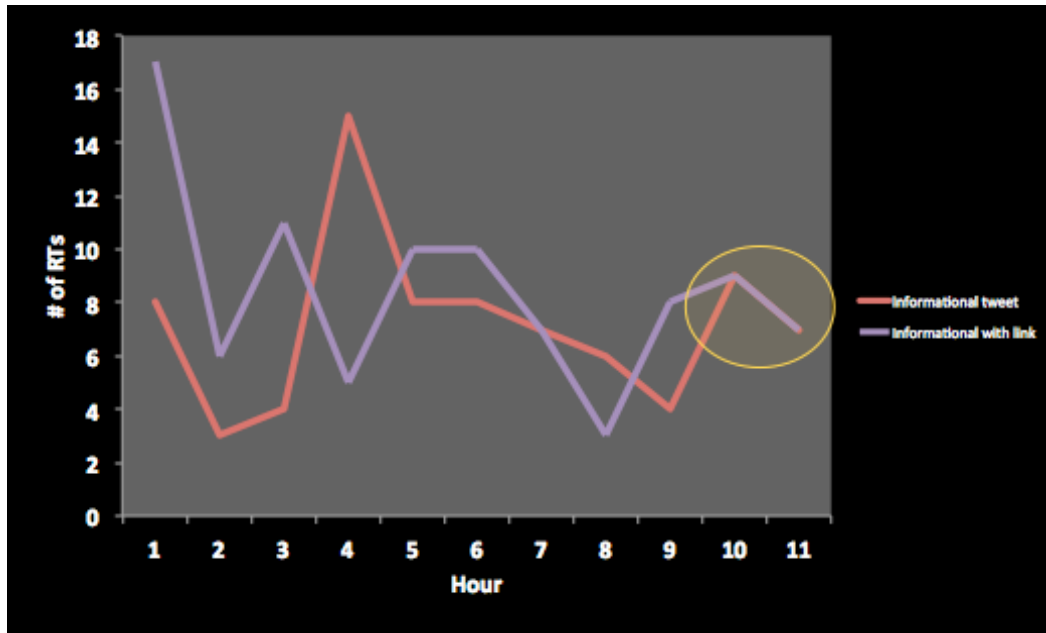
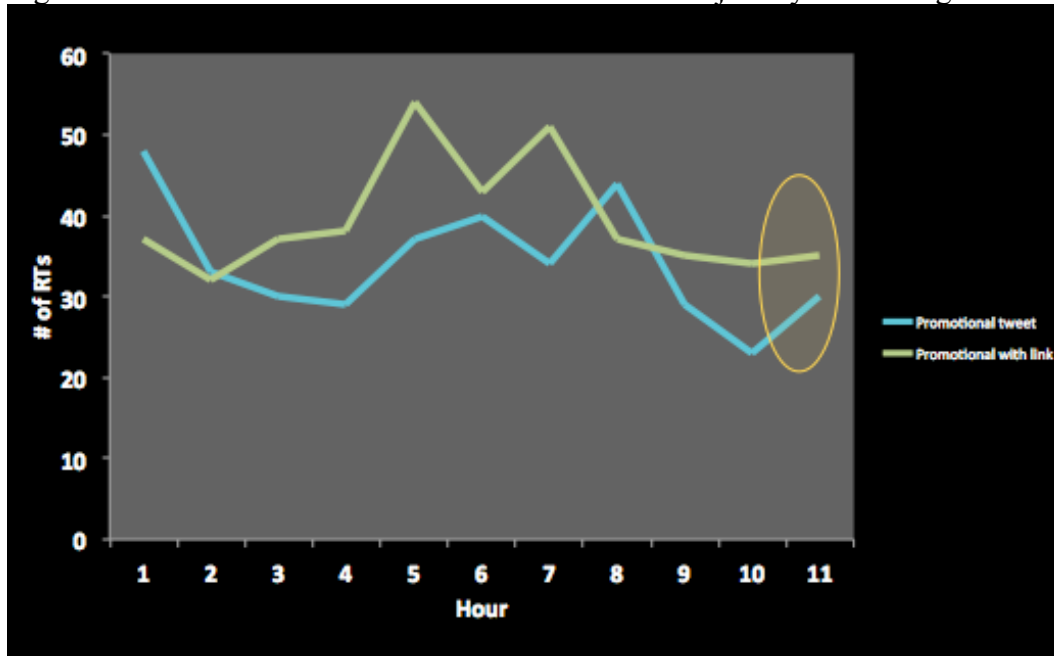


Figure 5. Promotional tweets over time. RTs show a trajectory of convergence.



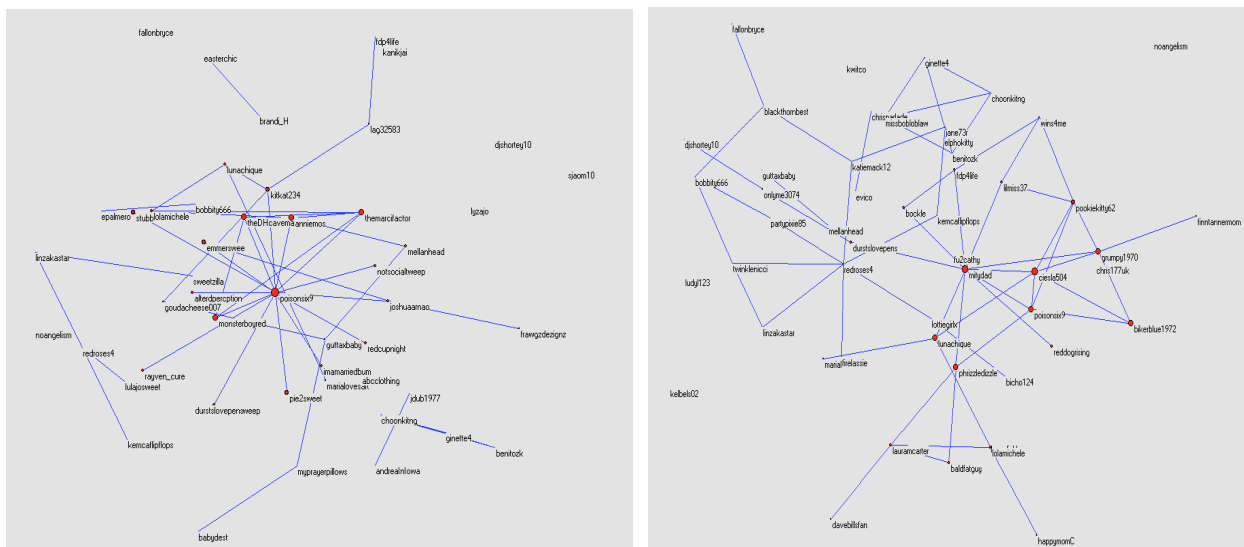
NETWORK ANALYSIS

In addition to comparing whether promotional and informational tweets with and without a link were more effective for RTs by followers, we also wanted to look at how these RTs were being diffused within Twitter. The diffusion of information is critical for companies using Twitter as a

marketing or PR tool because every business wants to maximize their flow of information such that each tweet reaches as many viewers as possible and each RT also has the potential to reach further audiences outside of the direct network of followers.

Using Pajek, we focused specifically on the diffusion of informational tweets, comparing the networks for followers who RTed informational tweets with a link and for followers who RTed informational tweets without a link. With these networks visualized, we observed that the network for informational tweets with a link was more disconnected than the network for informational tweets without a link. Without a link, there was a clear giant component in which almost all of the nodes were connected. However, with a link, there were several disconnected components and the overall network was not as connected.

Figure 6. Left: Network for Informational with Link. Right: Informational without Link



We analyzed these findings and came to the conclusion that the network that was almost fully connected would not be as beneficial in terms of diffusing information. In this densely connected network, although users are RTing UGP's tweets, the RTs are merely getting bounced around within this specific network, so users who have already seen the tweeted information keep re-seeing the same tweets. Contrastingly, the network with more separated components shows that users are not all followers of each other, thus their RTs have a higher probability of getting diffused into their own Twitter networks. In this network, the nodes completely separate from the giant component are especially important because we hypothesized that these users would act as the hubs diffusing UGP's informational tweets to further users outside of the original followers network.

Likewise, after doing the same visualization and analysis for the promotional tweets with and without links, we found very similar results in these two networks. The network for promotional tweets with a link was more separated than the network for tweets without a link, thus helping us reassure that tweets with a link are more beneficial for diffusion of information to other groups of users.

In addition, we needed to find what types of networks our four RT networks (informational with a link, informational without a link, promotional with a link, and promotional without a link) represented. We hypothesized that each would have properties of being a small world network. For each of our four networks, we found the average clustering coefficient and the average shortest path. In addition, we also created random networks that corresponded to each of our four networks by using the same number of nodes and found the same two measures for those random networks. When comparing each of the networks with their corresponding random networks, we found that all four of our RT networks had higher average clustering coefficients and lower average shortest paths than the random networks. This confirmed that our informational and promotional networks were, indeed, small world networks.

Figure 7. Small world testing for network of informational tweets.

Network	Average Clustering Coefficient	Average Shortest Path
Informational w/ link	0.1226919	2.39003
Random Network	0.0466045	3.79565
Information w/o link	0.1839286	2.40961
Random Network	0.0683565	3.02920

With all of these findings, we came to the conclusion that although all four of the RT networks were small world networks, that we wanted to distinguish promotional tweets with a link to be the most effective method for getting followers to RT and that tweets with a link were typically more useful for diffusing information out to the general public.

LIMITATIONS

There were several limitations that arose during our study. The limitations resulted in a restriction of data as well as a reduction in the overall reach of our data analysis. The limitations are as follows:

1. Twitter's Search API does collect search results, but not all search results are collected

This led to some missing data (i.e., missing RTs) though we do not think it would have had a significant effect on our results.

2. Twitter's REST API limits users to 300 requests per hour

In order to analyze the relationships between users a request had to be made to check whether a user was a follower of another user. Twitter allows for directed relationships (i.e., one-way relationships) there would also need to be a request made to find out whether the inverse relationship were true as well.

The number of request that needed to be made was equal to $n!/(n-2)!*(2!)$, since we are looking at the relationship between two users. Due to this limitation, it was strictly unfeasible given our

computing resources and time to process the relationship mapping our Underground Printing's entire follower network of 2000+ users. Given this limitation we focused our efforts on analyzing only the network of followers who RTed.

3. Twitter's search results are not indicative of the time in which the tweets were tweeted

If a user RTed a tweet at 12:01 pm the search results could either show 12:01 pm or some other time throughout the day (e.g., 3:00 pm). There isn't a direct relationship between when users RTed a message and when the search results for that RT appear. Therefore it was impossible to take time into consideration to map out how the tweets propagated throughout the system according to which user RTed and when they sent their RT. Knowing this limitation, we recorded how many users RTed tweets over an eleven hour time span.

4. Length of project

Because the length of our project was limited to a few weeks we could only afford a few weeks of actual data collection because the other weeks needed to be used to analyze and parse through our collected information in order to make informed findings.

FUTURE AREAS OF STUDY

Due the limitations and time constraints of our study several areas could be further studied for more clarification. Those areas are as follows:

1. A look at the entire network compared to the network of retweeters to see if there are any structural differences amongst the different networks

Different structures could affect the way in which information is propagated throughout the network.

2. How tweets propagate both inside and outside of Underground Printing's follower network

If time could be accounted for then a visualization of how RTs appeared in relationship to who was able to see them (i.e., who was a follower of the retweeter) could shed light on who was influencing who within the network. This would be a valuable indicator for network analysis as well as for businesses as it would provide information about in-links that could be considered when determining 'authority'.

3. More types of messages

Our study was limited to promotional tweets (i.e., tweets about giving something away) and informational tweets (i.e., tweets about news and information that did not involve promotions). Examining additional types of messages could also provide more insight into how to best effectively leverage twitter and the networks within.

4. Tweet impressions

From a measuring aspect it would be helpful to determine the number of impressions certain tweets have in comparison to the number of RTs they generate. This would allow us to determine a probability p for which tweets will be RTed based on the type of message. This could help develop new models for information diffusion within twitter using the probability p found.

5. Additional ways to visually represent the network

Our visualization of Underground Printing's network was limited to the relationship between users who RTed. Additional ways of representing the network include representing the weight of vertices (e.g., the number of RTs, the size of their network, etc.) and representing the weight of edges (e.g., how often a person RTs another user's tweets, how often a person interacts with another user, how long a person has been following a user, etc.).

6. Larger data sets

Our study was limited to one organization. Including other organizations within a singular study along with more types of messages during more times of the day would give us a larger data set to analyze across a broader spectrum. Determining whether our initial results hold true against larger data sets is something that needs to be done to determine whether our results are limited to just Underground Printing.

REFERENCES

1. “There’s a List for That”. blog.twitter.com. 30 Oct. 2009.
<<http://blog.twitter.com/2009/10/theres-list-for-that.html>>
2. Lake, Laura. "Twitter: A Valuable Marketing Tool?." *Marketing - Marketing Strategy - Marketing Plan - Marketing Careers and Jobs*. N.p., n.d. Web. 14 Oct. 2010.
<http://marketing.about.com/od>
3. “Twitter snags over 100 million users, eyes money-making”. *The Economic Times*. 15 Apr. 2010. <http://economictimes.indiatimes.com/infotech/internet/Twitter-snags-over-100-million-users-eyes-money-making/articleshow/5808927.cms>

APPENDIX

JAVA CODE USING TWITTER4J API

```

package twitter4j.examples.oauth;
import java.io.BufferedReader;
import java.io.InputStreamReader;

import twitter4j.IDs;
import twitter4j.Status;
import twitter4j.Tweet;
import twitter4j.Twitter;
import twitter4j.TwitterException;
import twitter4j.TwitterFactory;
import twitter4j.http.AccessToken;
import twitter4j.http.RequestToken;
import twitter4j.User;
import twitter4j.Query;
import twitter4j.QueryResult;
import java.util.List;
import java.io.*;

public class Welcome {
    /**
     * @param args
     */
    public static void main(String args[]) throws Exception {
        // The factory instance is re-useable and thread safe.
        Twitter twitter = new TwitterFactory().getInstance();

//insert the appropriate consumer key and consumer secret here

        twitter.setOAuthConsumer("ZGTGgFDb9az2Oh0Qekdlw",
            "IropTprVMJcYLgYWkjiONVaM8mBnNq5Qy0gtbfY2Yf4");
        RequestToken requestToken = twitter.getOAuthRequestToken();
        AccessToken accessToken = null;
        BufferedReader br = new BufferedReader(new InputStreamReader(System.in));
        while (null == accessToken) {
            System.out.println("Open the following URL and grant access to your account:");
            System.out.println(requestToken.getAuthorizationURL());
            System.out.print("Enter the PIN(if available) or just hit enter.[PIN]:");
            String pin = br.readLine();
            try {

                if(pin.length() > 0){
                    accessToken = twitter.getOAuthAccessToken(requestToken, pin);
                }
            }
        }
    }
}

```

```

        }else{
        accessToken = twitter.getOAuthAccessToken();
        }
        } catch (TwitterException te) {
        if(401 == te.getStatusCode()){
        System.out.println("Unable to get the access token.");
        }else{
        te.printStackTrace();
        }
        }
    }
    //persist to the accessToken for future reference.
    System.out.println(twitter.verifyCredentials().getId());
    System.out.println("token : " + accessToken.getToken());
    System.out.println("tokenSecret : " + accessToken.getTokenSecret());
    //storeAccessToken(twitter.verifyCredentials().getId() , accessToken);
    long lCursorIDs = -1;
    System.out.print("Enter Search Query:");
    String querystr = br.readLine();
    Query query=new Query(querystr) ;
    QueryResult result=twitter.search(query);
    List<Tweet> tweets=result.getTweets();
    // String[] userlist=new String[1000];
    String[] userlist={ "benitozk", "ginette4", "notsocialtweep", "emmerswee",
    "frawgzdezigz", "imabumtweep", "imamarriedbum", "monsterboyred", "alterdpercption",
    "rayven_cure", "redcupnight", "stubbornbrclt", "theDHCaveman", "themarcifactor",
    "kemcaflipflops", "durstslovenpens", "sweetzilla", "myprayerpillows", "marialovesart",
    "djshortey10", "choonkitng", "mellanhead", "noangelism", "guttaxbaby", "lyzajo", "lulajosweet",
    "bobbity666", "lag32583", "fallonbryce", "andreaInIowa", "poisonsix9", "joshuaarnao",
    "jdub1977", "brandi_H", "lolamichele", "lunachique", "babydest", "redroses4", "sjaom10",
    "goudacheese007", "easterchic", "kitkat234", "fdp4life", "linzakastar", "abcclothing",
    "epalmero", "pie2sweet", "annimos", "kanikjai"};
    int i=0;
    int checkexceed=0;
    for(int k=42;k<userlist.length;k++)
    {
        for(int p=0;p<userlist.length;p++)
        {
            if(p!=k&&userlist[k]!=null&&userlist[p]!=null)
            {
                if(twitter.existsFriendship(userlist[k], userlist[p])==true)
                {
                    System.out.println("User "+k+": " + userlist[k] + "---User
"+p+ ": "+userlist[p]);
                }
            }
        }
    }

```

```
        if(checkexceed++==325)
        {
            System.out.println("I am waiting");
            Thread.sleep(3600000);
            checkexceed=0;
        }
    }
}
System.out.println("HI"+i+ userlist[0]+userlist[1]+userlist[2]);
IDs followerIDs = twitter.getFollowersIDs(twitter.getId(), lCursorIDs);
System.exit(0);
}
```