

Information flow on Twitter surrounding regional events

Jason Koepke

D. Sc. Student

Applied Information Technology

Towson University

Outline

- Introduction
- Literature Review
- Proposed Research Questions
- Proposed Research Design
 - Design
 - Architecture
 - Data Normalization
 - Feature Set and Training Set
- Initial Experimental Results
- Potential Direction

Introduction

- We are currently in a period of rapid growth in the use of smart phones which provide the ability to spread information across mediums like Twitter quickly (Bertel, Stald, 2011).
- We identify several events in a regional area and profile users Twitter activity before and during the event.
- We also attempt to understand what decisions users make with information they receive.

Literature Review

Study	Type
Twitter: Social Communication in the Twitter Age	Information Flow
Pass It On?: Retweeting in Mass Emergencies	Information Validity
Sharing News, Making Sense, and Saying Thanks	Information Flow
Twitter Under Crisis: Can we trust what we RT?	Information Validity
A Relational Approach to Reducing Uncertainty During a Crisis through the use of Mobile Technology	Information Validity and Information Flow
Finding True and Credible Information on Twitter	Information Validity

Proposed Research Questions

- Compared to their normal twitter activity, do individuals communicate more surrounding an event?
- Do users who receive communication from the initial conveyer then share that information with their followers?
- Are there instances where information surrounding an events fails to spread?

Identification & Labeling of Events

- Obtained over 9,000,000 tweets from Twitter comprising the U.S.
- Leveraged Google News and local newspapers to identify 4 events during a 6 month window
- Narrowed scope of study to the state of Maryland due to the vast geography of the U.S.

Regionalized Events (MD)

	Start Date	Duration	Tweets
Derecho	6/30/2012	2 hours	109
Hurricane Sandy	10/29/2012	> 2 days	1,185
Baltimore Orioles Playoffs	10/11/2012	1 day	618
Proposition 8	11/06/2012	1 day	1,144
Totals			3056

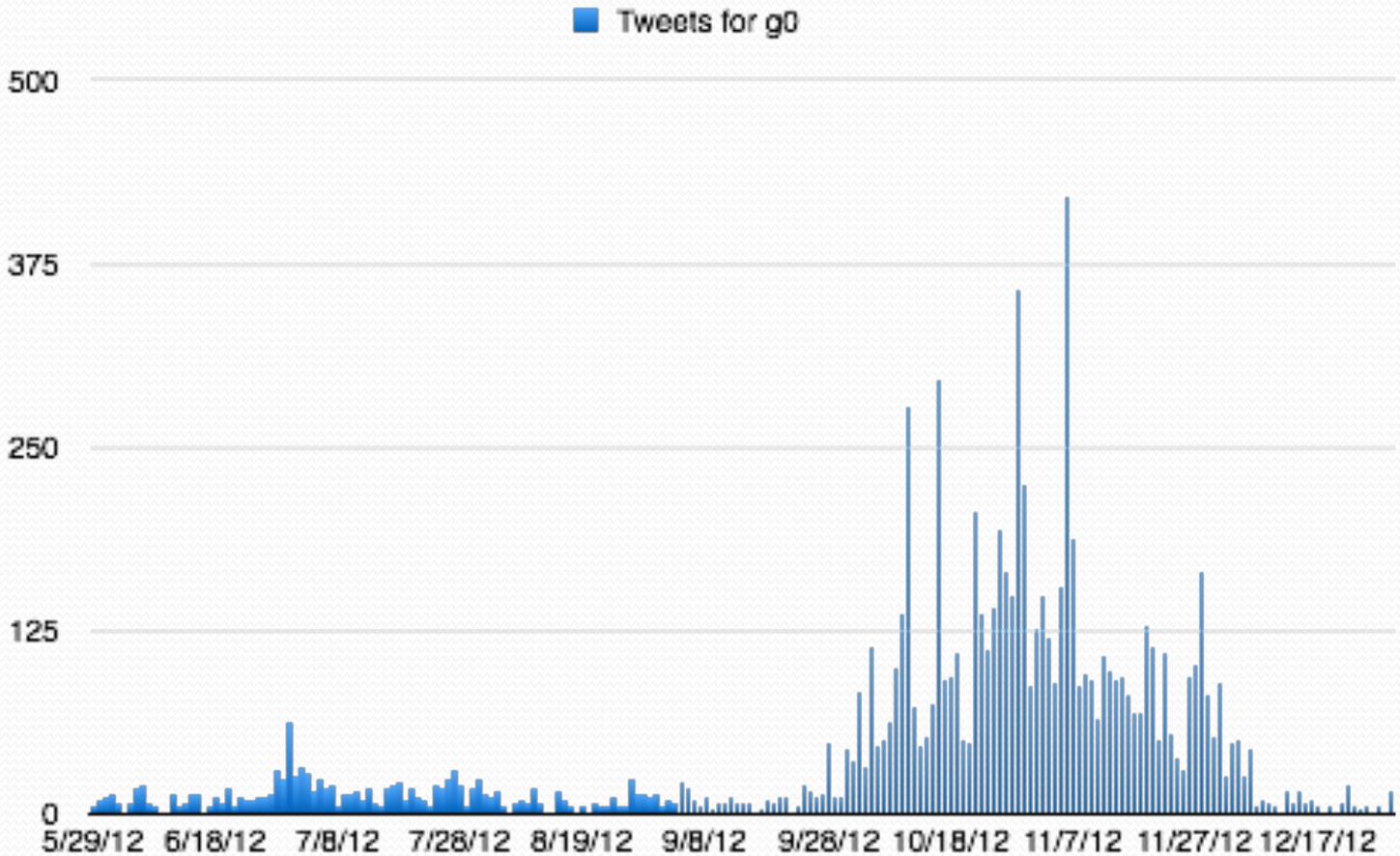
Users groupings

- Identified a set of 66 users encompassing 8,654 tweets and 1,837 retweets from the state of Maryland. Labeled this as group g_0
 - Defined group g_0 as the initiator of a conversation
- Labeled user group g_1 as a recipient of the communication initiated by g_0 .
 - Consisted of 134 users, 10,429 tweets, and 3,863 retweets

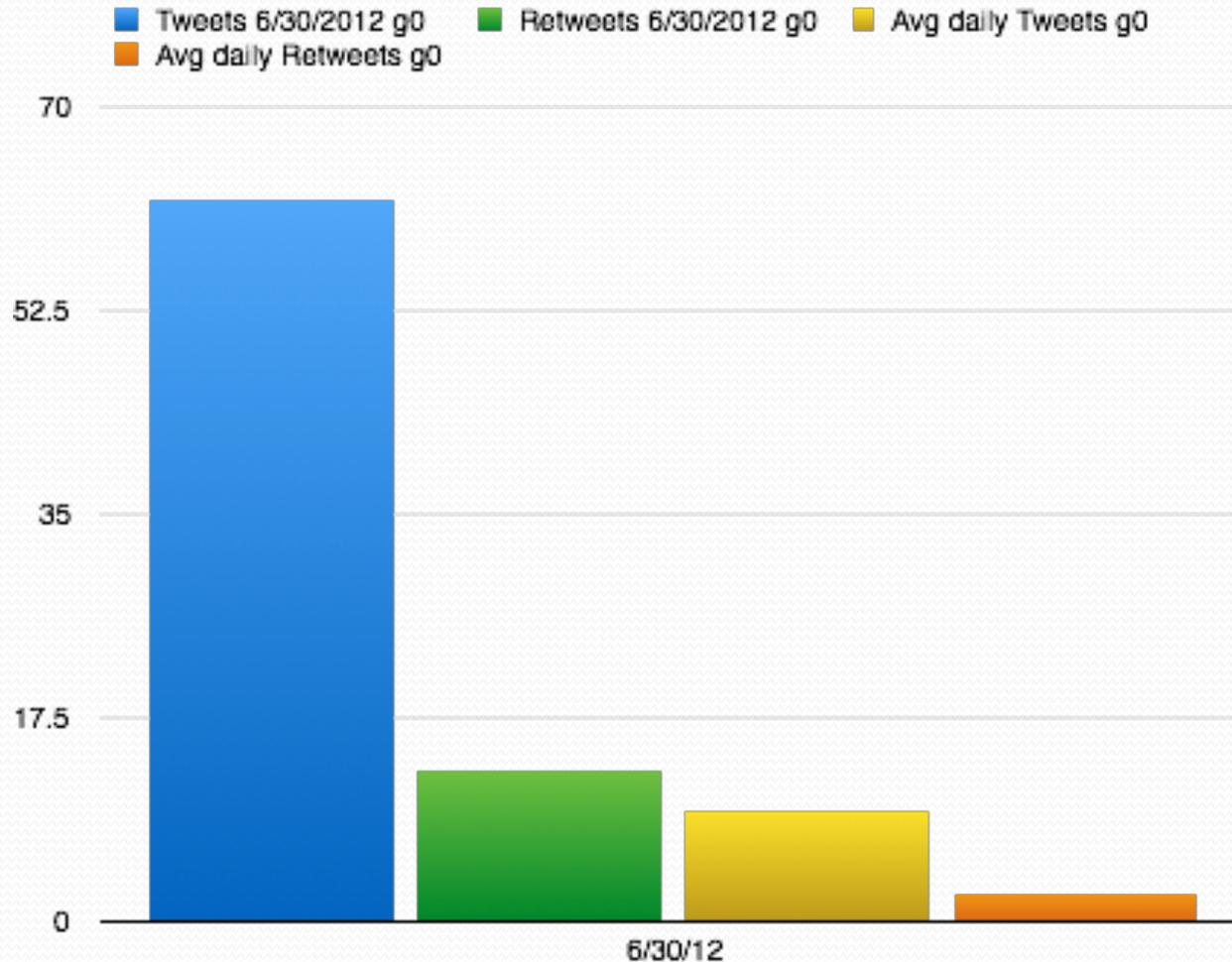
Communication breakout

	Tweets	Retweets	Users
g0	8,654	1,837	66
g1	10,429	2,026	134
Totals	19,083	3,863	200

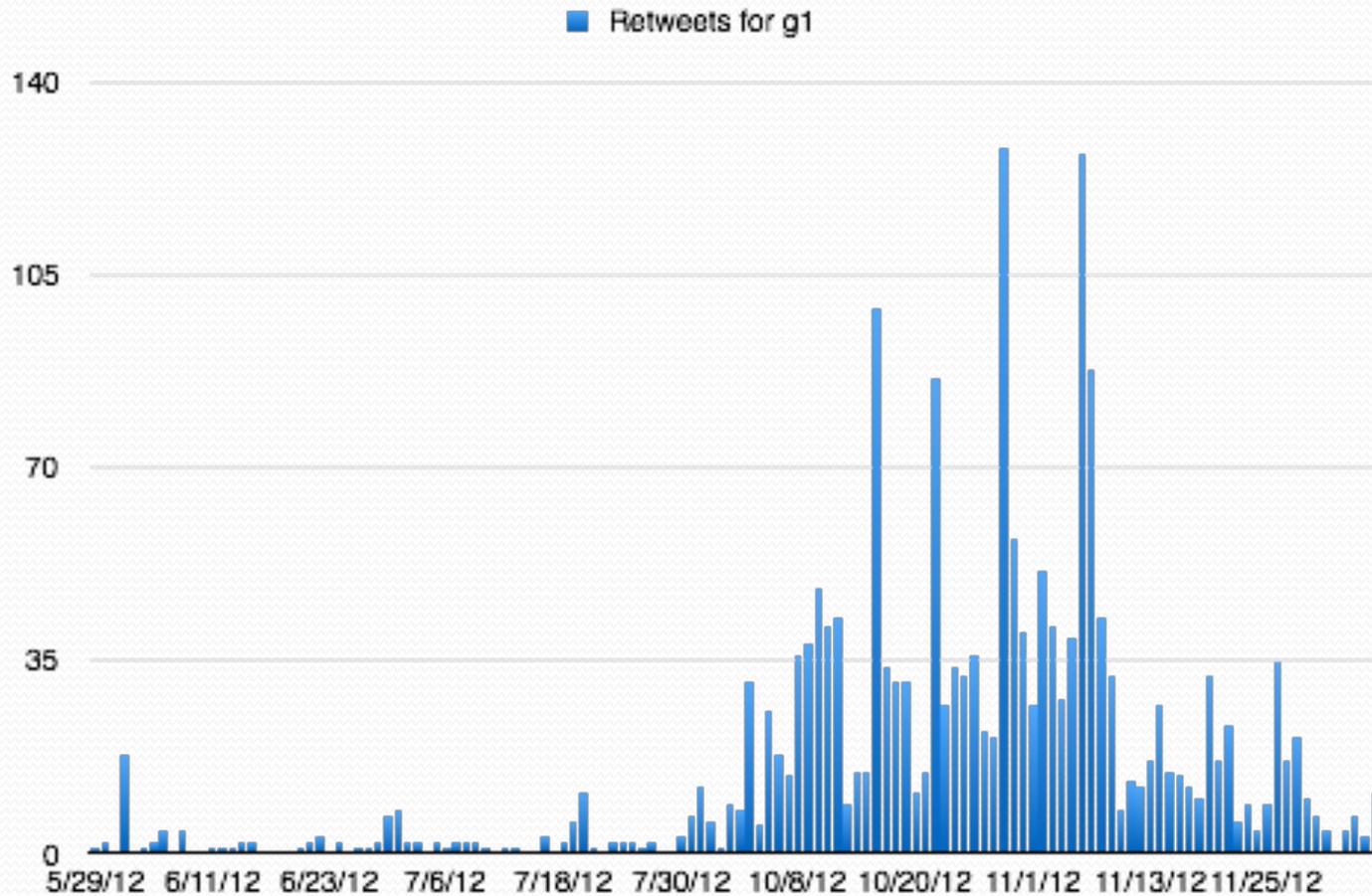
Tweets for g0



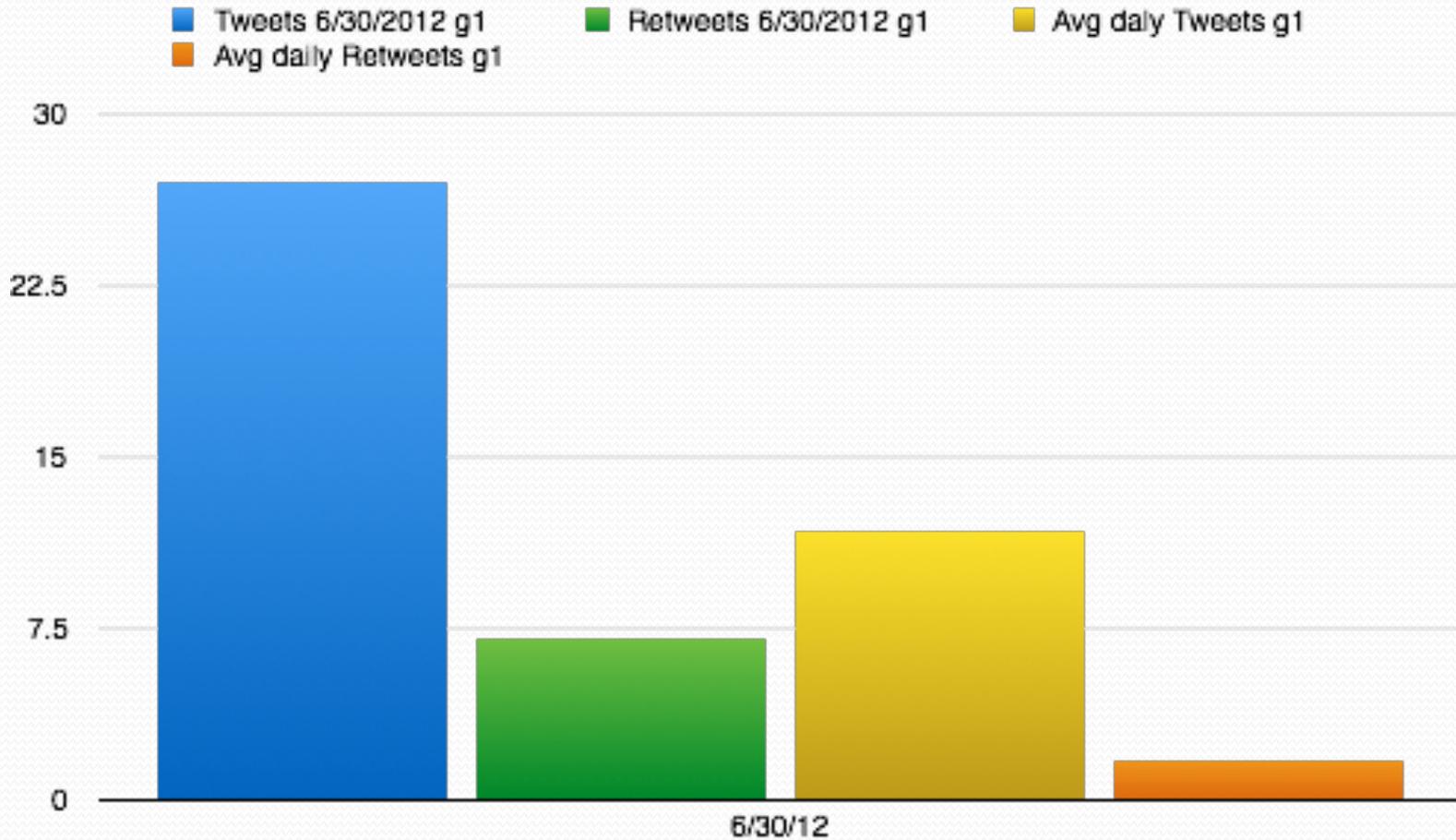
Tweets for derecho on 6/30/2012 compared to average over previous 30 days



Retweets for g1



Tweets for derecho on 6/30/2012 compared to average over previous 30 days



g0 and g1 results

- g0 users tend to communicate less during periods where widespread events do not occur.
- g0 users both tweet and retweet at up to 6 times their normal level of activity during events.
- Results for group g1 almost exactly mirrors results from group g0 indicating that the increased levels of activity from group g0 is not unique.

Social Network

- Identified 4 key users who stood out in the study as “super users”
 - Most interconnected amongst their peers
- A social network dynamic like this increases the spread of information

Future Work

- Understand the impact of the information being shared between users in their social circles
- Automating the identification of core trustworthy communicators
- Attempt to create the social network and environment similar to the one described in this paper to facilitate the spread of false information

Conclusions

- Users display heightened activity on Twitter during events compared to their historical averages
- Users who have experienced an event or are in a social circle with someone who has experienced an event continue to spread valid information to a wider audience
- Confirmed previous research from Mendoza, et al, indicating that the majority of the information shared amongst our social network is valid.