9-25-2013

Resurrecting My Revolution: Using Social Link Neighborhood in Bringing Context to the Disappearing Web

Hany M. SalahEldeen
Old Dominion University

Michael L. Nelson
Old Dominion University, mnelson@odu.edu

Recommended Citation
SalahEldeen, Hany M. and Nelson, Michael L., "Resurrecting My Revolution: Using Social Link Neighborhood in Bringing Context to the Disappearing Web" (2013). *Computer Science Presentations*. 13.
https://digitalcommons.odu.edu/computerscience_presentations/13

Resurrecting My Revolution

Using Social Link Neighborhood in Bringing Context to the Disappearing Web

Hany M. SalahEldeen & Michael L. Nelson

Old Dominion University

Department of Computer Science
Web Science and Digital Libraries Lab.

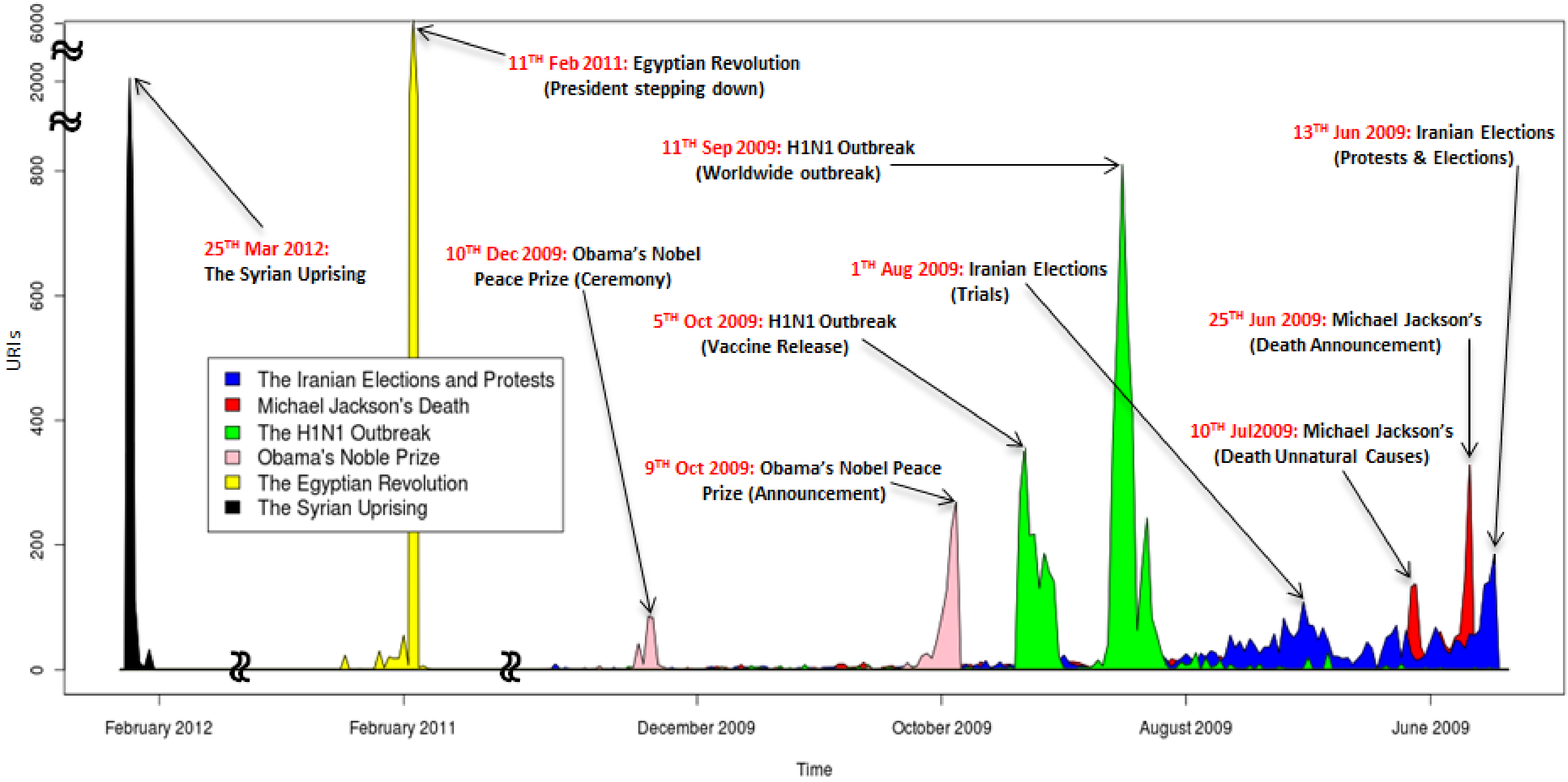
TPDL 2013

Six Socially Significant Events

- From Twitter, Websites, Books:
  - The Egyptian revolution.

- From Twitter Only:
  - Stanford's SNAP dataset:
    - Iranian elections.
    - H1N1 virus outbreak.
    - Michael Jackson's death.
    - Obama's Nobel Peace Prize.
  - Twitter API:
    - The Syrian uprising.

Social Events Having a Bimodal Time Distribution



11TH Feb 2011: Egyptian Revolution
(President stepping down)

25TH Mar 2012:
The Syrian Uprising

11TH Sep 2009: H1N1 Outbreak
(Worldwide outbreak)

13TH Jun 2009: Iranian Elections
(Protests & Elections)

10TH Dec 2009: Obama's Nobel
Peace Prize (Ceremony)

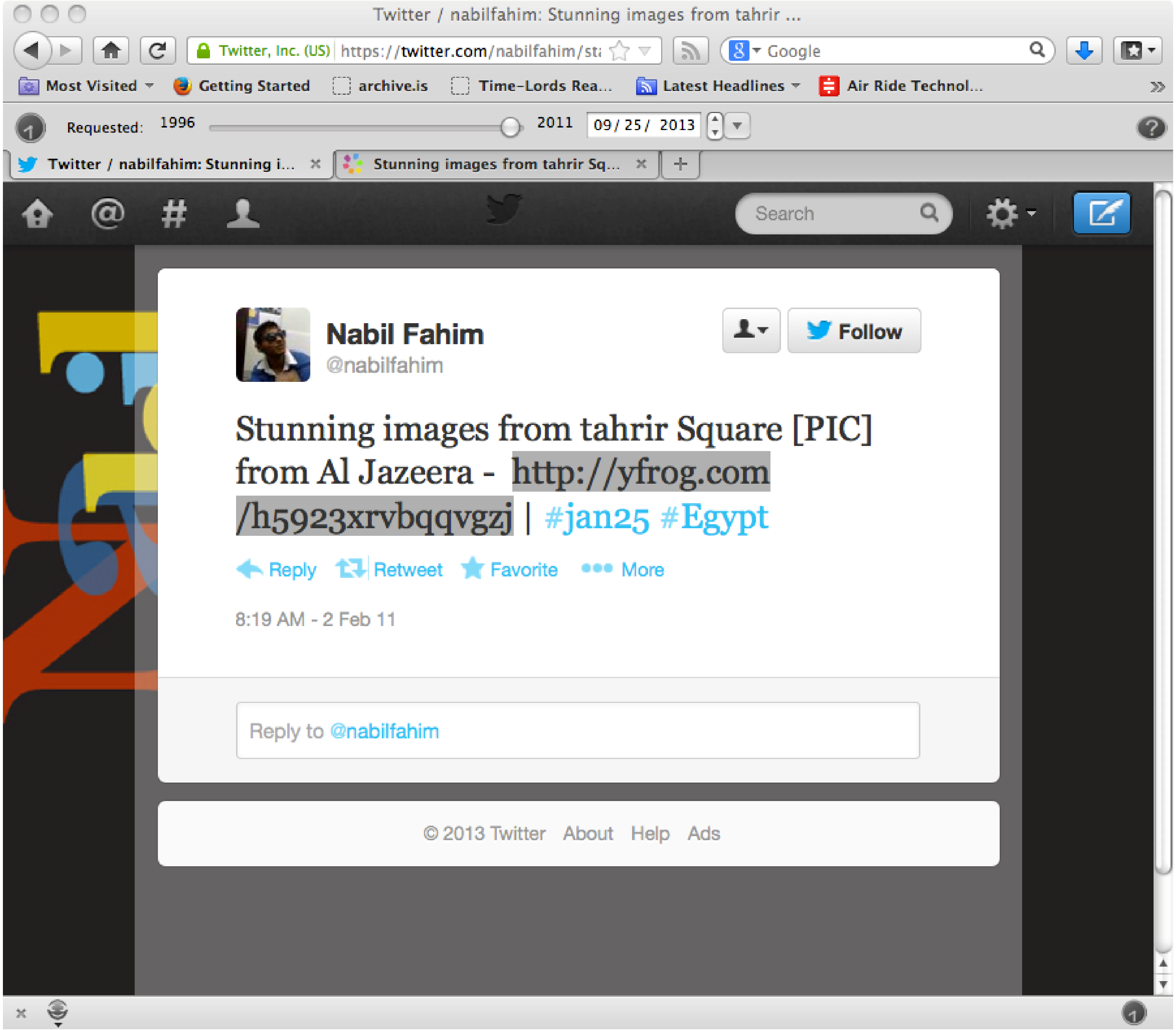
1TH Aug 2009: Iranian Elections
(Trials)

5TH Oct 2009: H1N1 Outbreak
(Vaccine Release)

25TH Jun 2009: Michael Jackson's
(Death Announcement)

10TH Jul2009: Michael Jackson's
(Death Unnatural Causes)

9TH Oct 2009: Obama's Nobel Peace
Prize (Announcement)



Twitter / nabilfahim: Stunning images from tahrir ...
Twitter, Inc. (US) https://twitter.com/nabilfahim/st
Google
Most Visited
Getting Started
archive.is
Time-Lords Rea...
Latest Headlines
Air Ride Technol...
Requested: 1996
2011
09/25/2013
Twitter / nabilfahim: Stunning i...
Stunning images from tahrir Sq...
Search
Follow
Reply
Retweet
Favorite
More

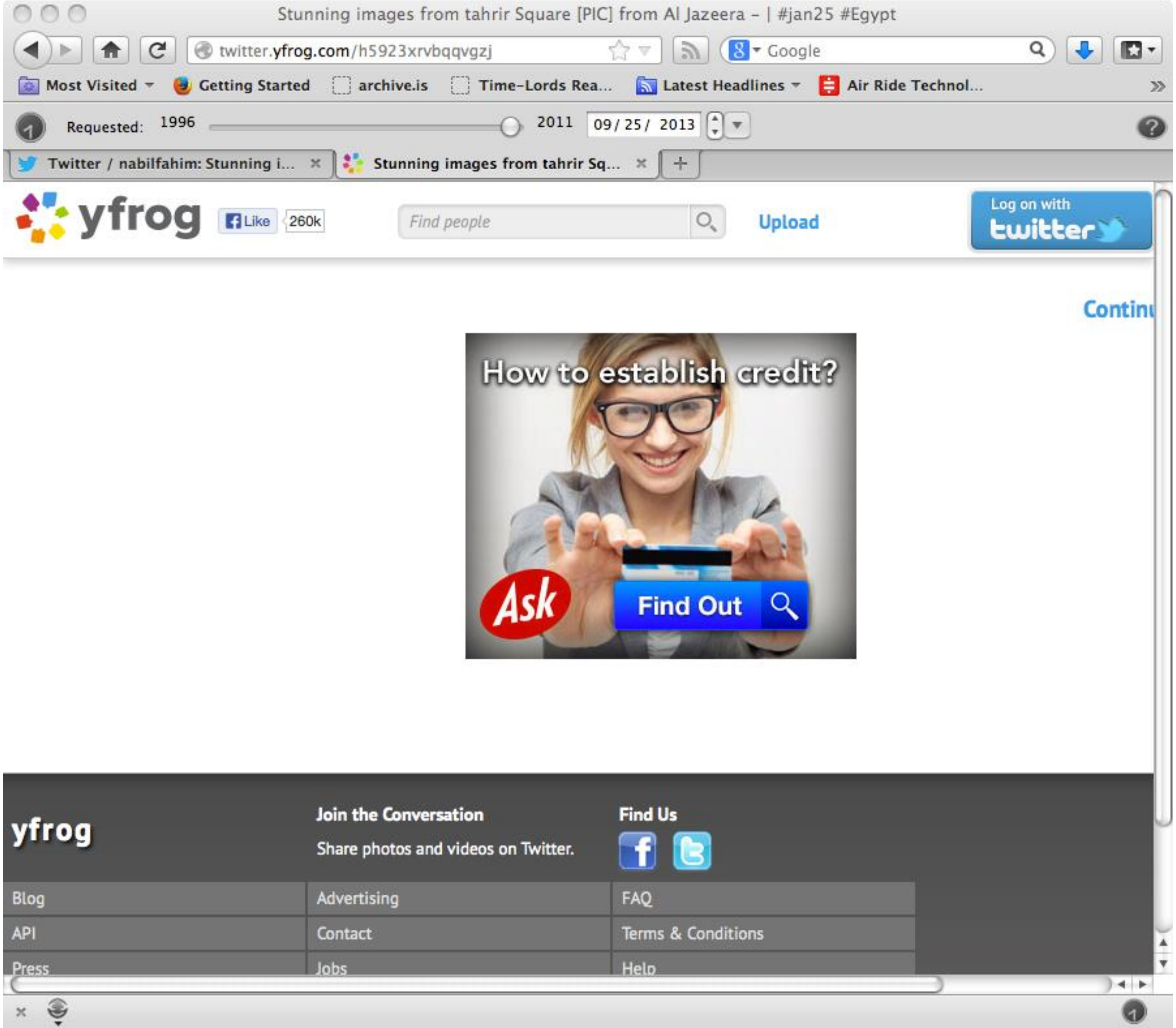
Nabil Fahim
@nabilfahim

Stunning images from tahrir Square [PIC] from Al Jazeera - http://yfrog.com/h5923xrvbqqvgzj | #jan25 #Egypt

8:19 AM - 2 Feb 11

Reply to @nabilfahim

© 2013 Twitter About Help Ads



Stunning images from tahrir Square [PIC] from Al Jazeera – | #jan25 #Egypt

twitter.yfrog.com/h5923xrvbqqvgzj

yfrog

Join the Conversation

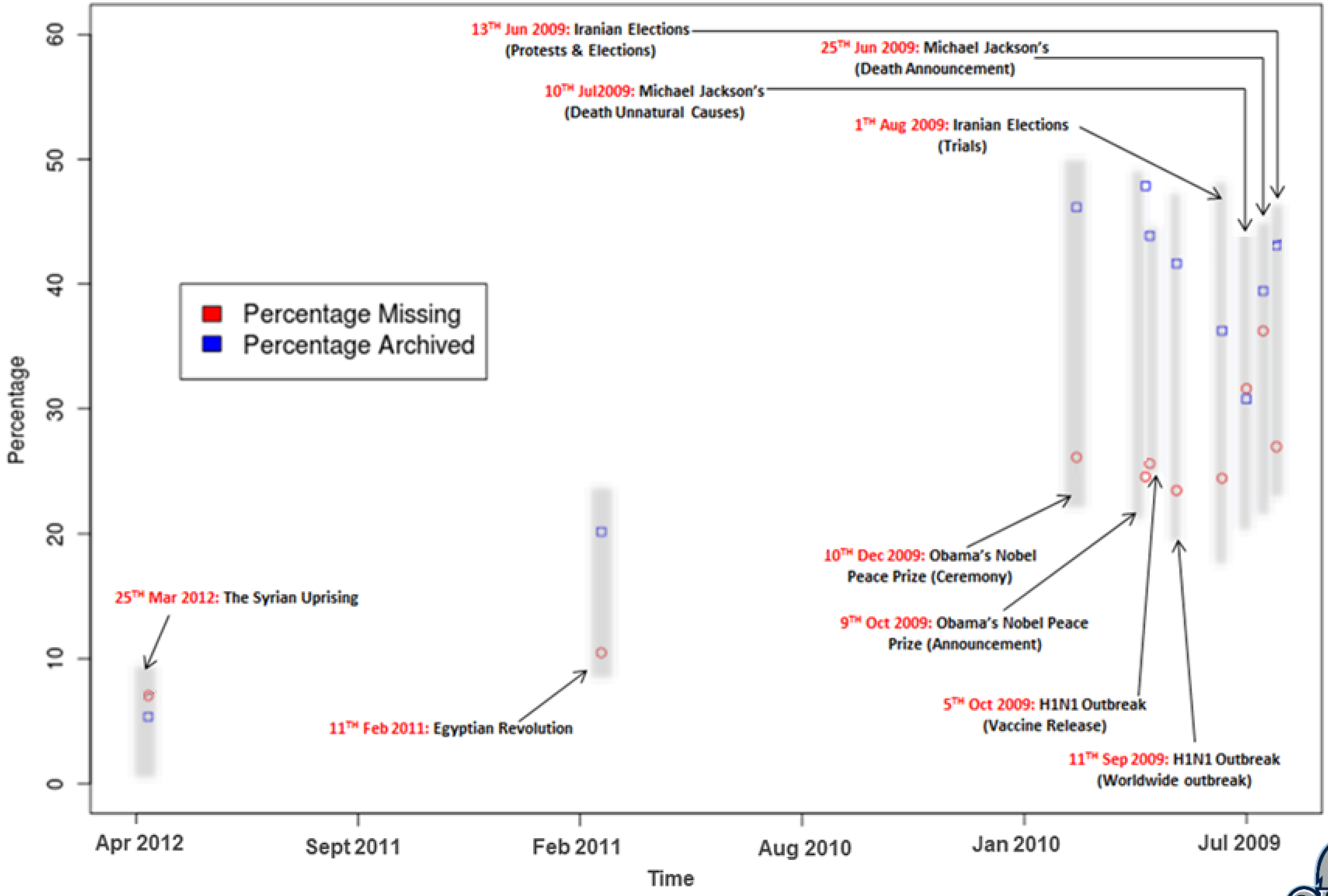
Share photos and videos on Twitter.

Find Us

Resources Missing & Archived

Collection	Percentage Missing	Percentage Archived
Michael Jackson	36.24% 31.62%	39.45% 30.78%
Iran	26.98% 24.47%	43.08% 36.26%
H1N1 Outbreak	23.49% 25.64%	41.65% 43.87%
Obama	24.59% 26.15%	47.87% 46.15%
Egypt	10.48%	20.18%
Syria	7.04%	5.35%

Missing and Archived Percentages Across Time



13TH Jun 2009: Iranian Elections (Protests & Elections)

25TH Jun 2009: Michael Jackson's (Death Announcement)

10TH Jul2009: Michael Jackson's (Death Unnatural Causes)

1TH Aug 2009: Iranian Elections (Trials)

10TH Dec 2009: Obama's Nobel Peace Prize (Ceremony)

9TH Oct 2009: Obama's Nobel Peace Prize (Announcement)

5TH Oct 2009: H1N1 Outbreak (Vaccine Release)

11TH Sep 2009: H1N1 Outbreak (Worldwide outbreak)

25TH Mar 2012: The Syrian Uprising

11TH Feb 2011: Egyptian Revolution

Previous Conclusions

$$Content\ Lost\ Percentage = 0.02(Age\ in\ days) + 4.20$$

$$Content\ Archived\ Percentage = 0.04(Age\ in\ days) + 6.74$$

- Measured 21,625 resources from 6 data sets in archives & live web.
- After a year from publishing about ***11%*** of content shared on social media will be gone.
- After this we are losing roughly ***0.02%*** daily.

New Research Questions

- **Validity:** is our estimation model still valid?
- **Existence Stability:** do resources on the live web remain missing?
- **Archival Stability:** do resources in public archives persist?
- **Social Context:** how can we extract social context of missing resources and potential replacements?

Revisiting Existence

- From previous study:

$$Content\ Lost\ Percentage = 0.02(Age\ in\ days) + 4.20 \tag{1}$$

- Rerunning after a year:

	MJ		Iran		H1N1		Obama		Egypt	Syria
Measured	37.10%	37.50%	28.17%	30.56%	26.29%	31.62%	32.47%	24.64%	7.55%	12.68%
Predicted	31.72%	31.42%	31.96%	30.98%	30.16%	29.68%	29.60%	28.36%	19.80%	11.54%
Error	5.38%	6.08%	3.79%	0.42%	3.87%	1.94%	2.87%	3.72%	12.25%	1.14%

Average Prediction Error = 4.15%

Revisiting Archival

- From previous study:

$$Content\ Archived\ Percentage = 0.04(Age\ in\ days) + 6.74 \tag{2}$$

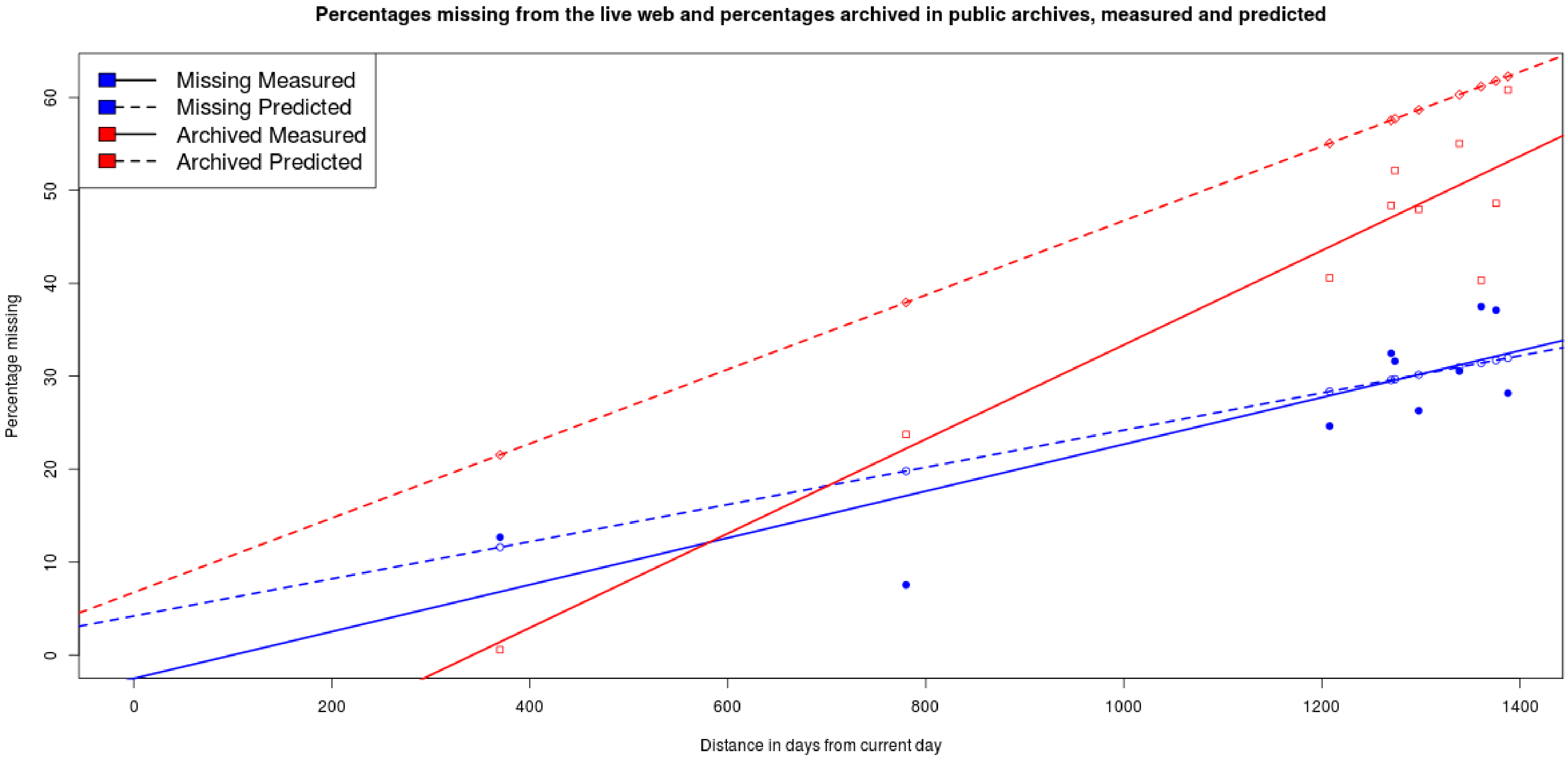
- Rerunning after a year:

	MJ		Iran		H1N1		Obama		Egypt	Syria
Measured	48.61%	40.32%	60.80%	55.04%	47.97%	52.14%	48.38%	40.58%	23.73%	0.56%
Predicted	61.78%	61.18%	62.26%	60.30%	58.66%	57.70%	57.54%	55.06%	37.94%	21.42%
Error	13.17%	20.86%	1.46%	5.26%	10.69%	5.56%	9.16%	14.48%	14.21%	20.86%

Average Prediction Error = 11.57%

in all cases, our archival predictions were too optimistic

Measured Vs. Predicted



Interesting Phenomenon: Reappearance On The Live Web And Disappearance From The Archives

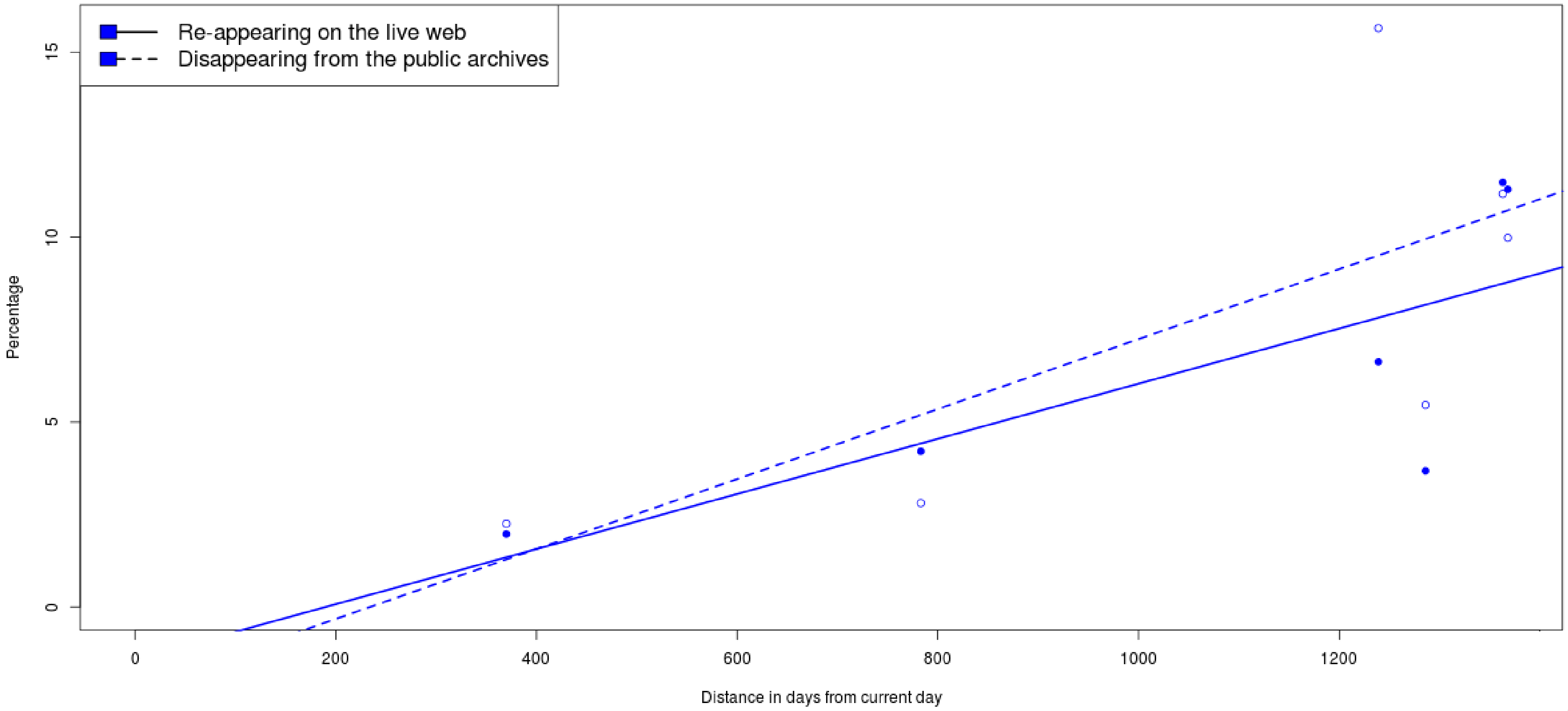
Event	MJ	Iran	Obama	H1N1	Egypt	Syria	Average
% Re-appearing on the web	11.29%	11.48%	6.63%	3.68%	4.21%	1.97%	6.54%
% Disappearing from archives	9.98%	11.17%	15.65%	5.46%	2.81%	2.25%	7.89%
% Going from 1 memento to 0	2.72%	2.89%	4.24%	1.96%	0.23%	0.28%	2.05%

$$Missing = Disappearance - Reappearance$$

Reappearing And Disappearance Predictions

$$LiveContent\ Reappearing = 0.01(Age\ in\ days) - 1.42$$

$$Mementos\ Disappearing = 0.01(Age\ in\ days) - 2.22$$



Tweet Existence

Problem:

We don’t have the URIs of most of the tweets

Solution:

compute loss of other tweets linking to the URI:

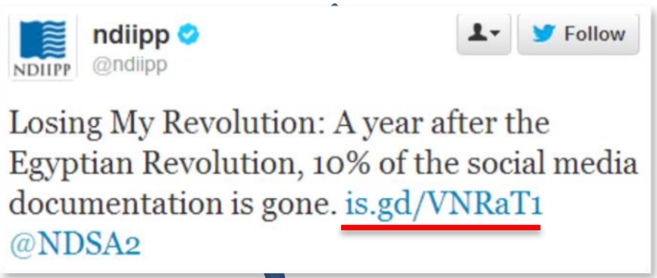
for each resource in the datasets:

-Extract all the tweets that link to the resource using Topsy API (up to 500 tweets)

-Check existence of each tweet.

- (Topsy (mostly) does not delete indexed tweets)

Using Topsy API

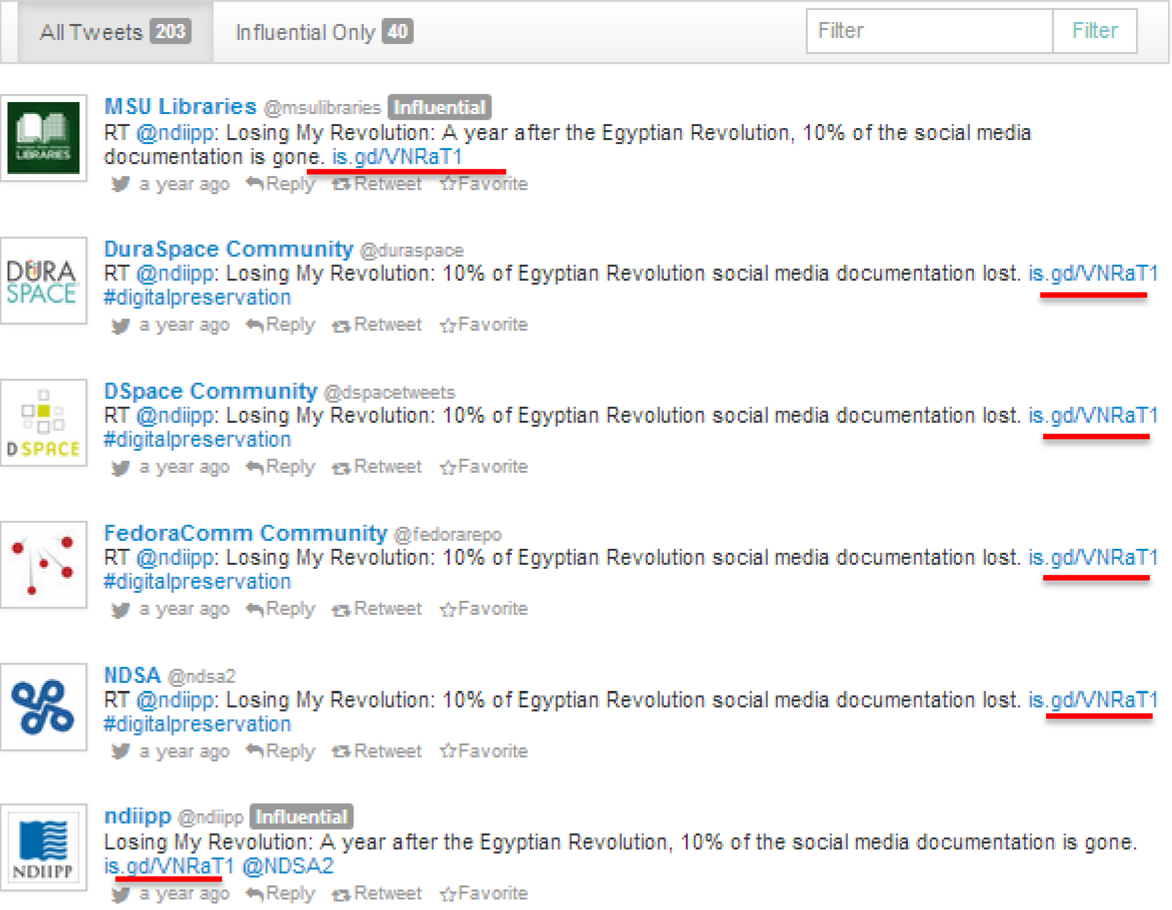


ndiipp @ndiipp

Follow

Losing My Revolution: A year after the Egyptian Revolution, 10% of the social media documentation is gone. is.gd/VNRaT1 @NDSA2

Get all the tweets having the same URI



All Tweets 203 | Influential Only 40 | Filter | Filter

MSU Libraries @msulibraries Influential
RT @ndiipp: Losing My Revolution: A year after the Egyptian Revolution, 10% of the social media documentation is gone. is.gd/VNRaT1
a year ago Reply Retweet Favorite

DuraSpace Community @duraspace
RT @ndiipp: Losing My Revolution: 10% of Egyptian Revolution social media documentation lost. is.gd/VNRaT1 #digitalpreservation
a year ago Reply Retweet Favorite

DSpace Community @dspacetweets
RT @ndiipp: Losing My Revolution: 10% of Egyptian Revolution social media documentation lost. is.gd/VNRaT1 #digitalpreservation
a year ago Reply Retweet Favorite

FedoraComm Community @fedorarepo
RT @ndiipp: Losing My Revolution: 10% of Egyptian Revolution social media documentation lost. is.gd/VNRaT1 #digitalpreservation
a year ago Reply Retweet Favorite

NDSA @ndsa2
RT @ndiipp: Losing My Revolution: 10% of Egyptian Revolution social media documentation lost. is.gd/VNRaT1 #digitalpreservation
a year ago Reply Retweet Favorite

ndiipp @ndiipp Influential
Losing My Revolution: A year after the Egyptian Revolution, 10% of the social media documentation is gone. is.gd/VNRaT1 @NDSA2
a year ago Reply Retweet Favorite

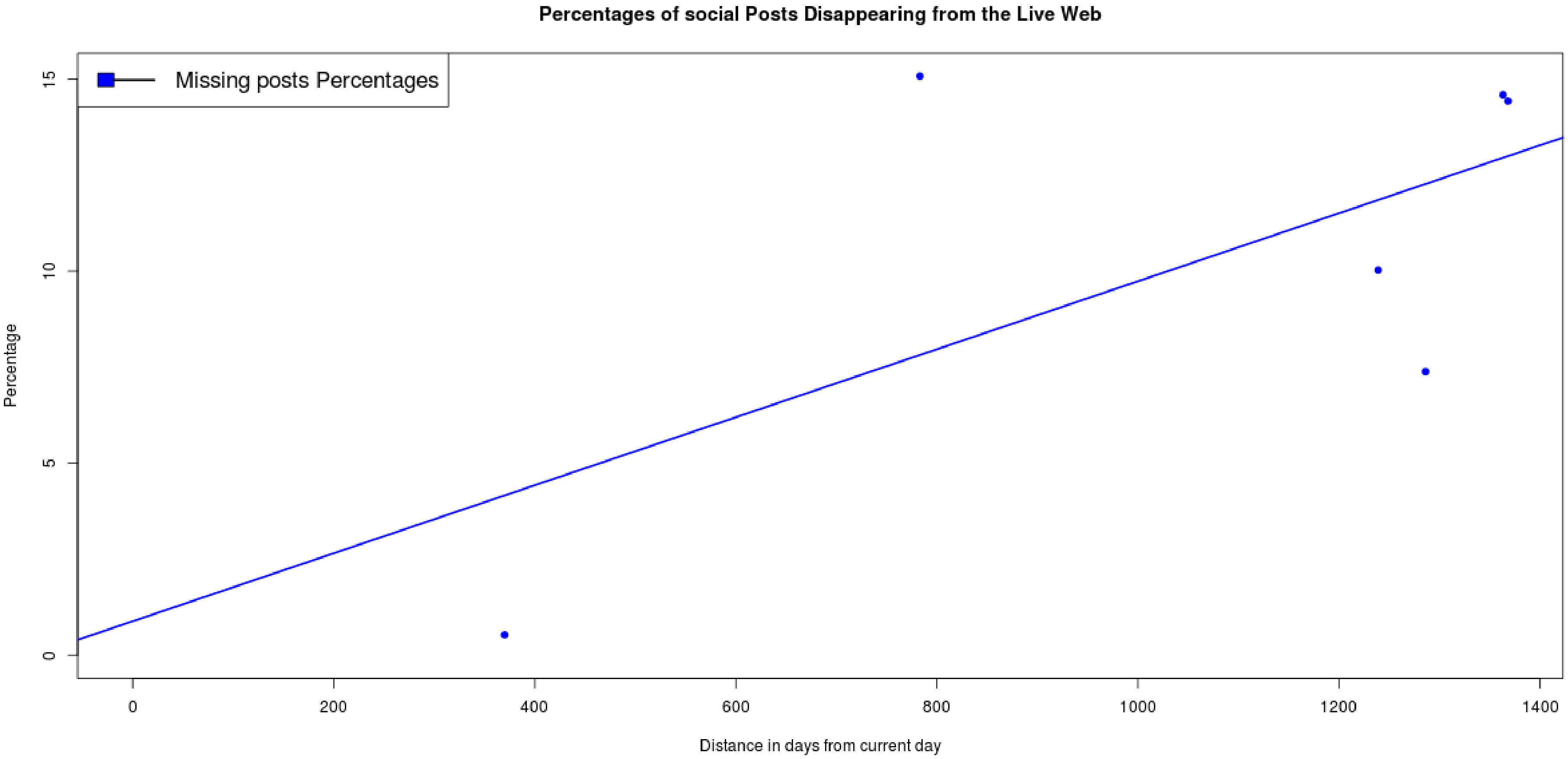
Check how many still exist on the live web

Tweet Existence

Event	MJ	Iran	Obama	H1N1	Egypt	Syria	Average
Average % of missing posts	14.43%	14.59%	10.03%	7.38%	15.08%	0.53%	10.34%

$$SocialPosts\ Missing = 0.01(Age\ in\ days) + 0.88$$

Tweets Disappearing Across Time



Context Discovery And Shared Resource Replacement

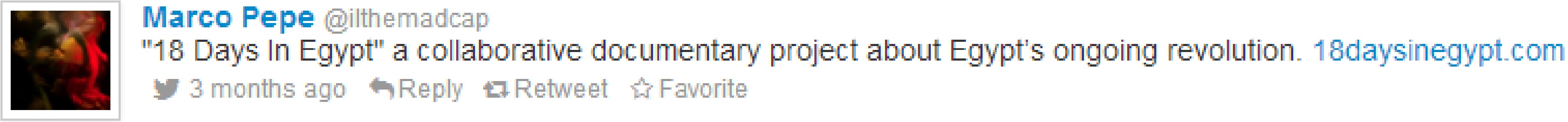
Problem:

140 characters limits the description of the linked resource. If it went missing, can we get the next best thing?

Solution:

- Shared links typically have several tweets, responses, and retweets
- We can mine these traces for context and viable replacements

Context Discovery

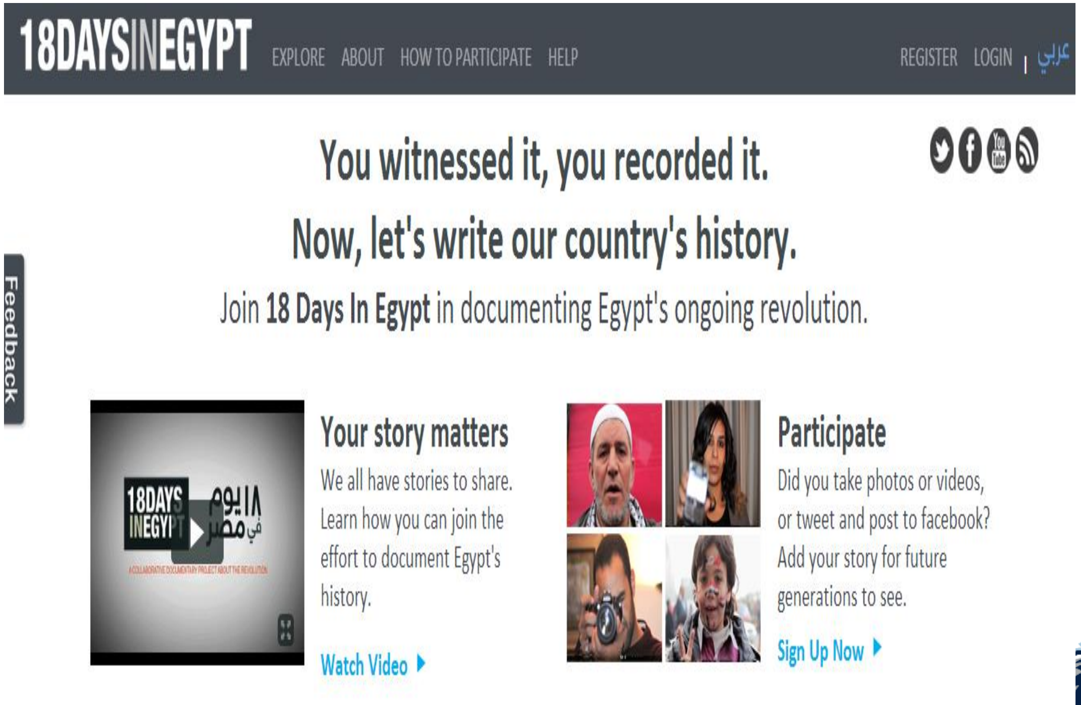


Marco Pepe @ilthemadcap

"18 Days In Egypt" a collaborative documentary project about Egypt's ongoing revolution. 18daysinegypt.com

Linking to:

http://beta.18daysinegypt.com/



You witnessed it, you recorded it.

Now, let's write our country's history.

Join 18 Days In Egypt in documenting Egypt's ongoing revolution.

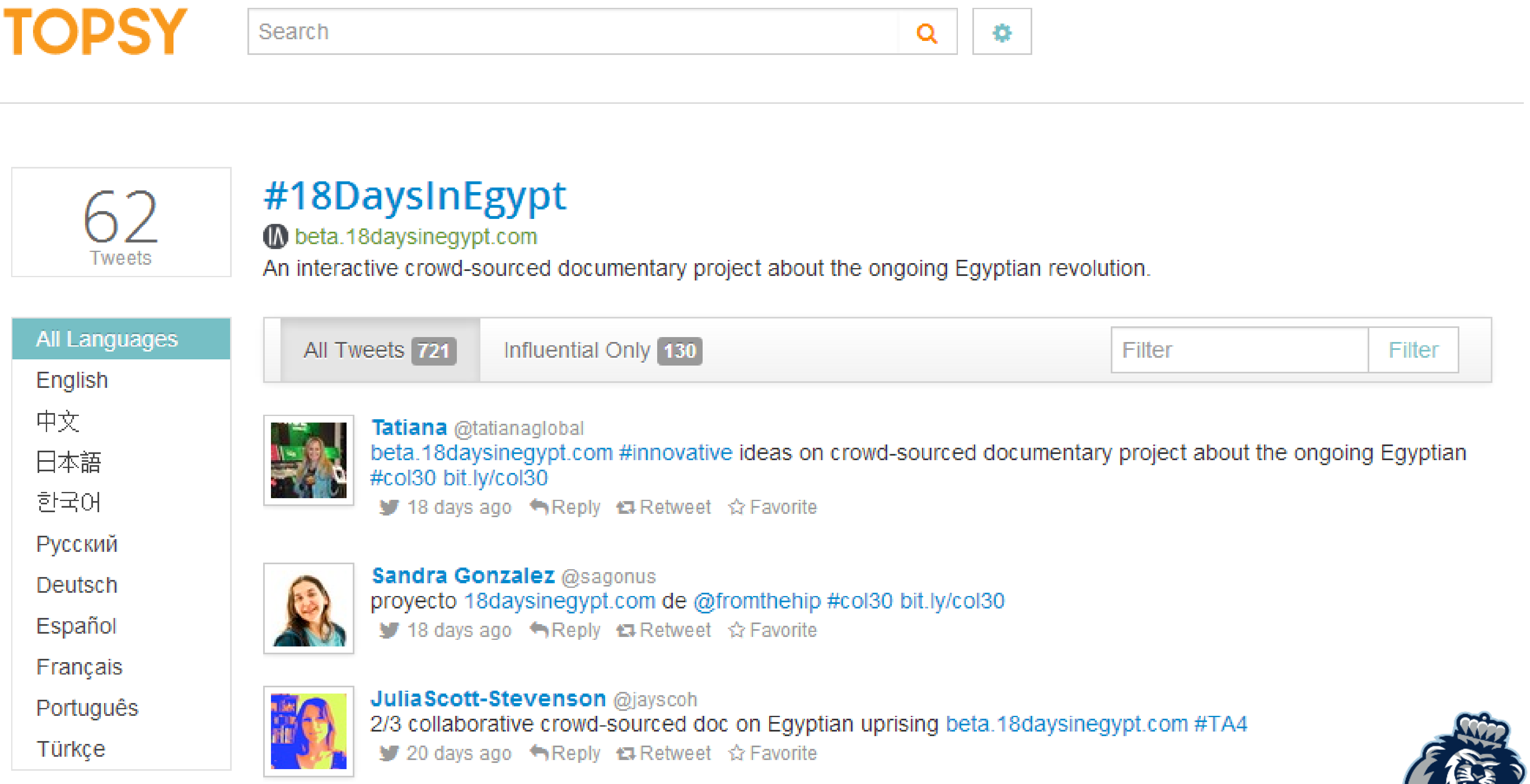
Your story matters

We all have stories to share. Learn how you can join the effort to document Egypt's history.

Participate

Did you take photos or videos, or tweet and post to facebook? Add your story for future generations to see.

Use Topsy to Discover Tweets with the Same Link



62 Tweets

#18DaysInEgypt

beta.18daysinegypt.com

An interactive crowd-sourced documentary project about the ongoing Egyptian revolution.

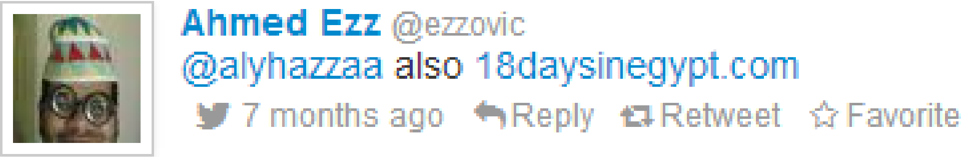
Tatiana @tatianaglobal
beta.18daysinegypt.com #innovative ideas on crowd-sourced documentary project about the ongoing Egyptian #col30 bit.ly/col30
18 days ago Reply Retweet ☆ Favorite

Sandra Gonzalez @sagonus
proyecto 18daysinegypt.com de @fromthehip #col30 bit.ly/col30
18 days ago Reply Retweet ☆ Favorite

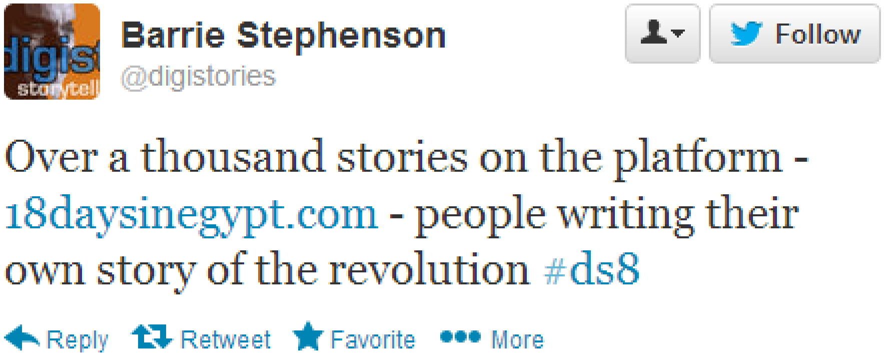
JuliaScott-Stevenson @jayscoh
2/3 collaborative crowd-sourced doc on Egyptian uprising beta.18daysinegypt.com #TA4
20 days ago Reply Retweet ☆ Favorite

Tweet Text Replacement

- From all extracted tweets, extract the best replacement tweet having the longest common N-gram


Ahmed Ezz @ezzovic

@alyhazzaa also 18daysinegypt.com


Barrie Stephenson @digistories

Over a thousand stories on the platform - 18daysinegypt.com - people writing their own story of the revolution #ds8

Replace with more descriptive one

Resource Replacement

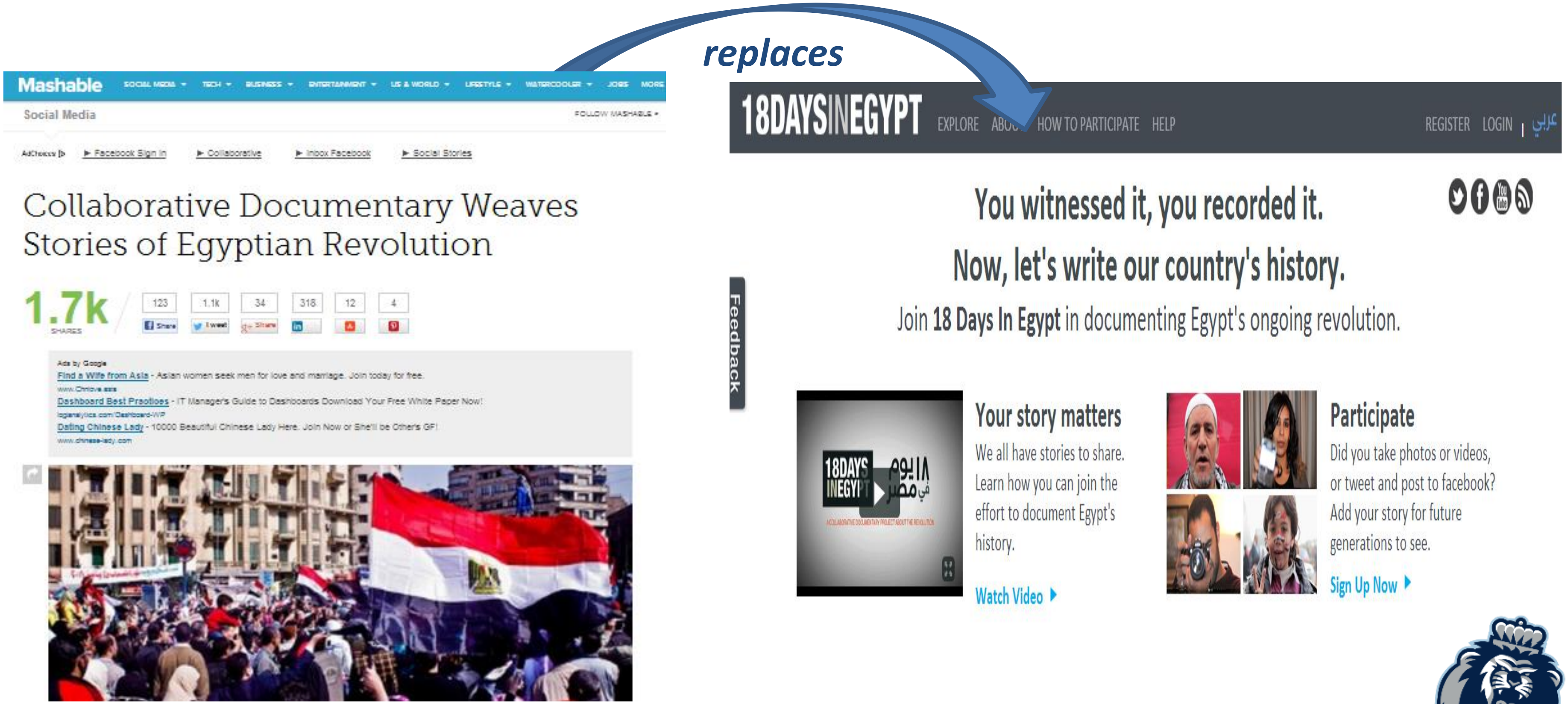
Assume that the resource linked in the tweet disappeared.

We mine the list of tweets for:

- Hashtags
- User mentions
- *Co-tweeted URIs*

Co-Tweeted Resources

A missing resource could be described or replaced by another resource that have been shared within the same tweet.



replaces

Mashable
SOCIAL MEDIA ▾ TECH ▾ BUSINESS ▾ ENTERTAINMENT ▾ US & WORLD ▾ LIFESTYLE ▾ WATERCOOLER ▾ JOBS MORE
Social Media
FOLLOW MASHABLE ▸
AdChoices ▷ ▶ Facebook Sign In ▶ Collaborative ▶ Inbox Facebook ▶ Social Stories

Collaborative Documentary Weaves Stories of Egyptian Revolution

1.7k
SHARES
123 Share
1.1k Tweet
34 Share
318
12
4

You witnessed it, you recorded it.

Now, let's write our country's history.

Join 18 Days In Egypt in documenting Egypt's ongoing revolution.

Your story matters

We all have stories to share. Learn how you can join the effort to document Egypt's history.

Watch Video ▸

Participate

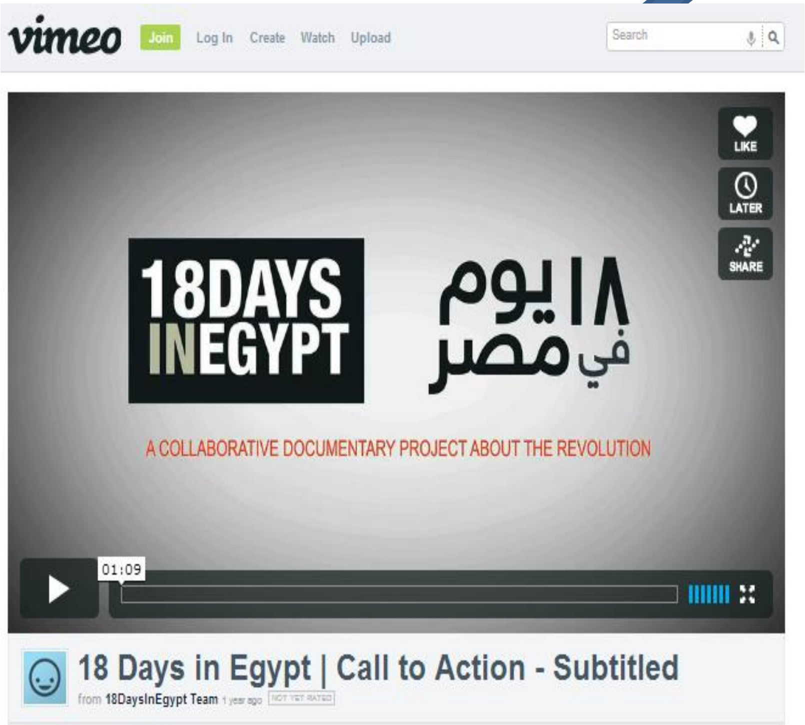
Did you take photos or videos, or tweet and post to facebook? Add your story for future generations to see.

Sign Up Now ▸

Co-Tweeted Resources

A missing resource could be described or replaced by another resource that have been shared within the same tweet.

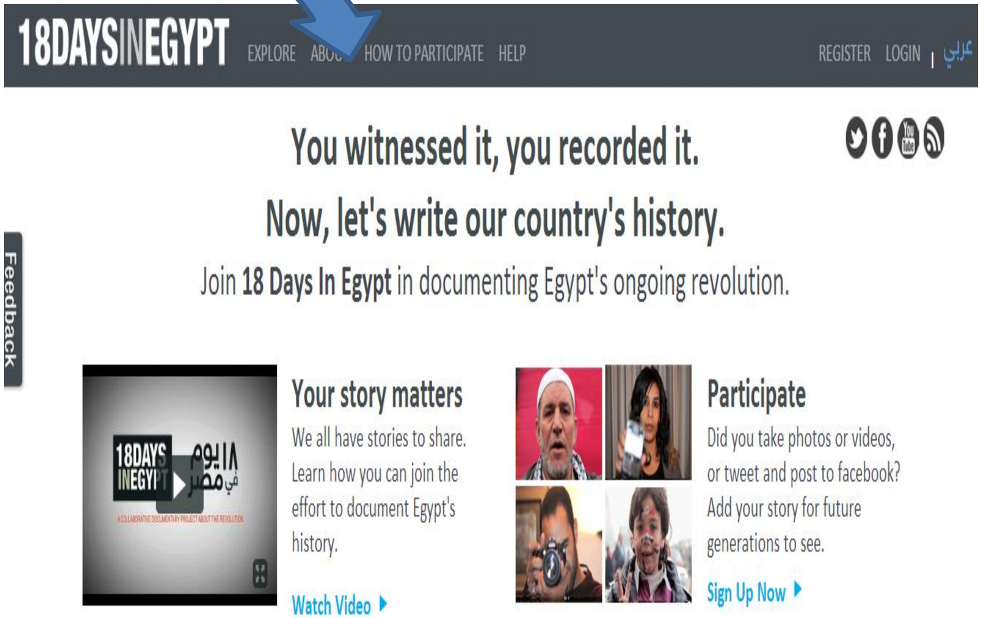
Or



18 Days in Egypt | Call to Action - Subtitled

from 18DaysInEgypt Team 1 year ago NOT YET RATED

replaces



You witnessed it, you recorded it.

Now, let's write our country's history.

Join **18 Days In Egypt** in documenting Egypt's ongoing revolution.

Your story matters

We all have stories to share. Learn how you can join the effort to document Egypt's history.

Watch Video ▸

Participate

Did you take photos or videos, or tweet and post to facebook? Add your story for future generations to see.

Sign Up Now ▸

Build a *Tweet Document*

A *tweet document* represents the concatenation of all extracted tweets:

> “*do you have a story to tell about your 18 days of revolution? share it or contact sara 18days brand new interactive storytelling project on egyptian revolution a very creative platform to tell your story daysinegypt marches heading to tahrir square now from all over cairo it's all over again use the website to document your revolutionary stories and share them with the world! check out awesome documentary project   crowdsourcing a people's narrative of the egyptian revolution …* ”

Tweet Signatures

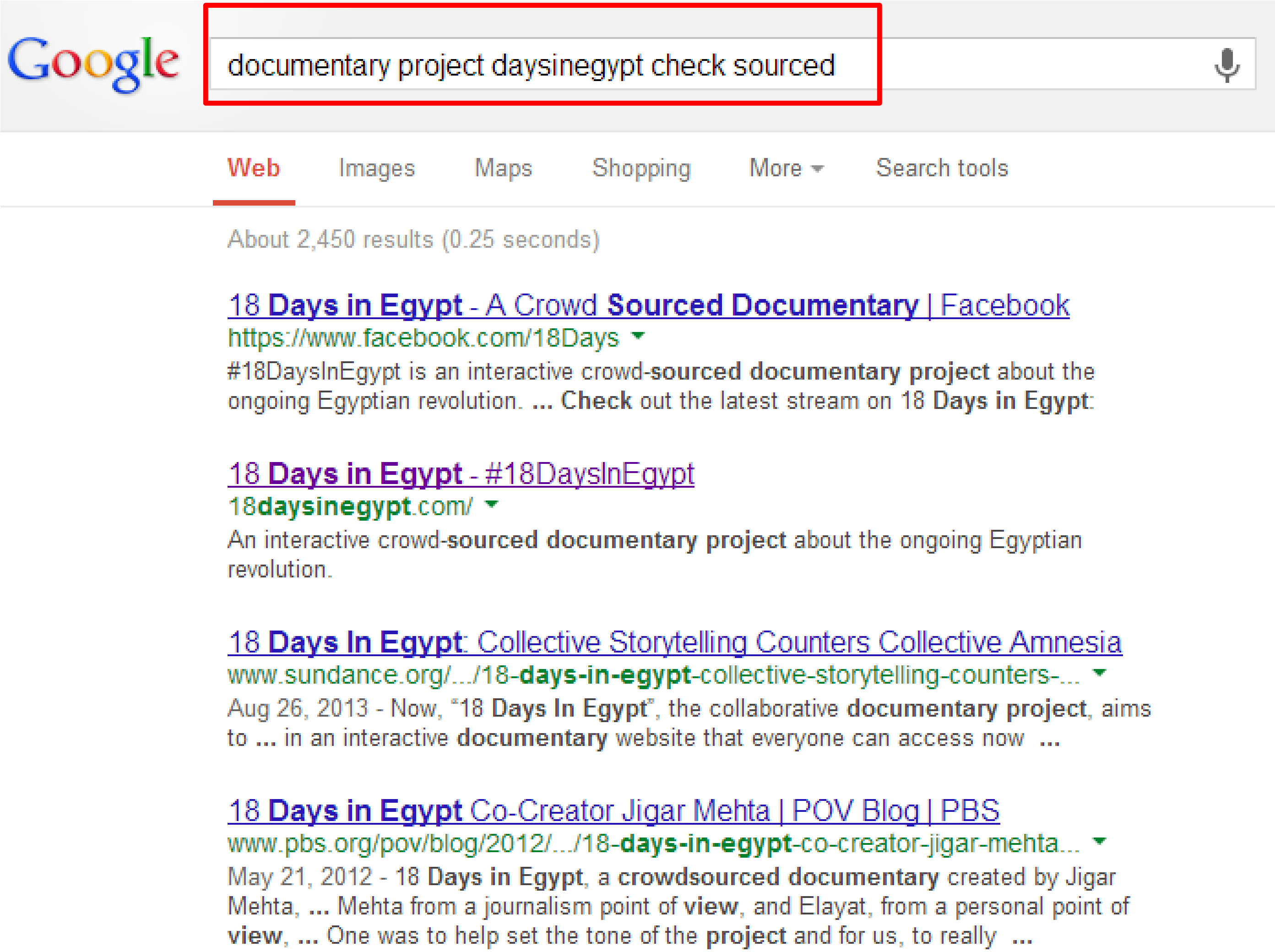
Tweet Document:

> “*do you have a story to tell about your 18 days of revolution? share it or contact sara 18days brand new interactive storytelling project on egyptian revolution a very creative platform to tell your story daysinegypt marches heading to tahrir square now from all over cairo it's all over again use the website to document your revolutionary stories and share them with the world! check out awesome documentary project crowdsourcing a people's narrative of the egyptian revolution …* ”

Tweet Signature = top 5 most frequent terms from Tweet Document

documentary project daysinegypt check sourced

Query Google w/ Tweet Signature



documentary project daysinegypt check sourced

About 2,450 results (0.25 seconds)

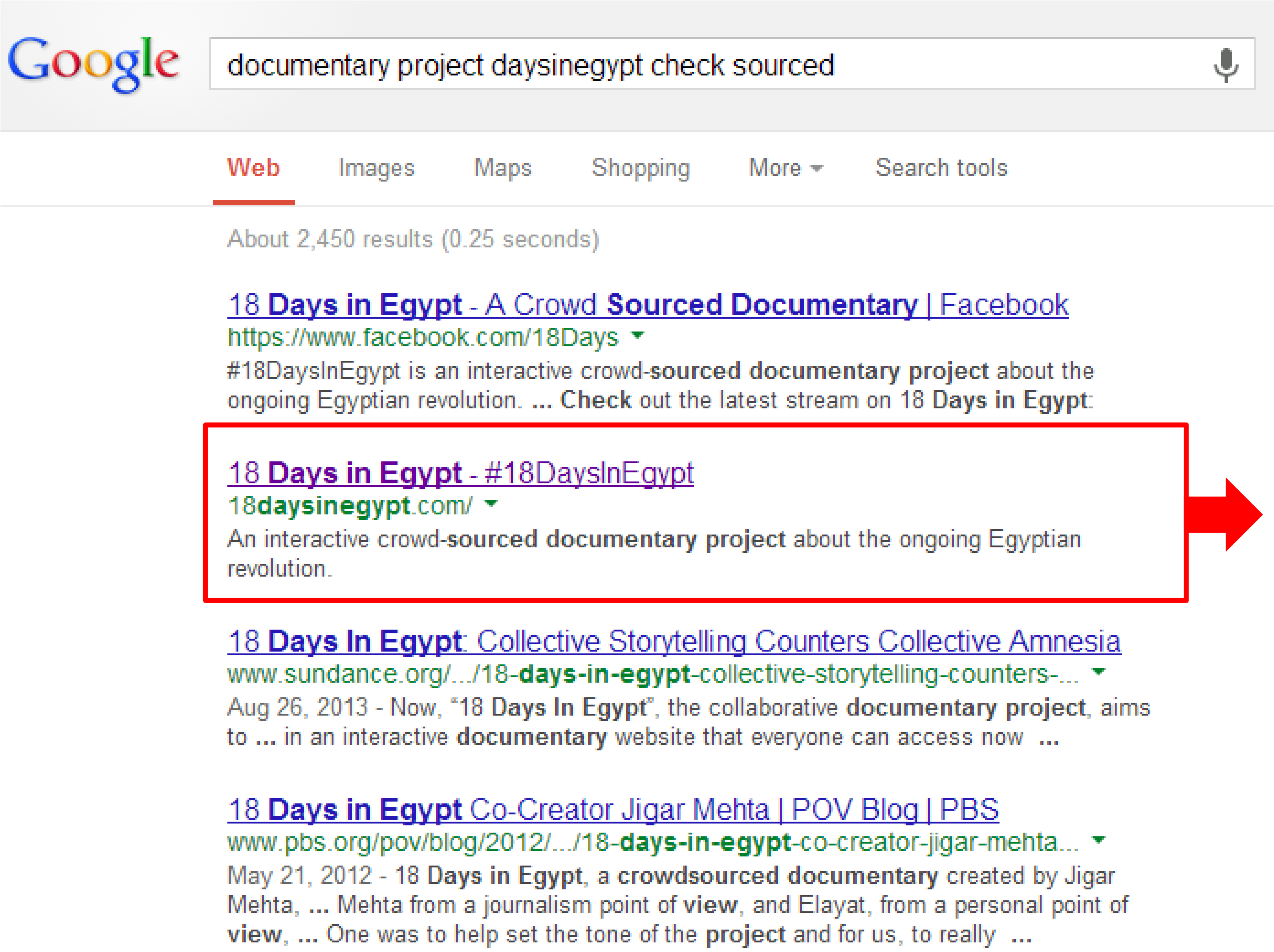
18 Days in Egypt - A Crowd Sourced Documentary | Facebook
https://www.facebook.com/18Days
#18DaysInEgypt is an interactive crowd-sourced documentary project about the ongoing Egyptian revolution. ... Check out the latest stream on 18 Days in Egypt:

18 Days in Egypt - #18DaysInEgypt
18daysinegypt.com/
An interactive crowd-sourced documentary project about the ongoing Egyptian revolution.

18 Days In Egypt: Collective Storytelling Counters Collective Amnesia
www.sundance.org/.../18-days-in-egypt-collective-storytelling-counters-...
Aug 26, 2013 - Now, "18 Days In Egypt", the collaborative documentary project, aims to ... in an interactive documentary website that everyone can access now ...

18 Days in Egypt Co-Creator Jigar Mehta | POV Blog | PBS
www.pbs.org/pov/blog/2012/.../18-days-in-egypt-co-creator-jigar-mehta...
May 21, 2012 - 18 Days in Egypt, a crowdsourced documentary created by Jigar Mehta, ... Mehta from a journalism point of view, and Elayat, from a personal point of view, ... One was to help set the tone of the project and for us, to really ...

Search Engine Results



documentary project daysinegypt check sourced

About 2,450 results (0.25 seconds)

18 Days in Egypt - A Crowd **Sourced Documentary** | Facebook
https://www.facebook.com/18Days
#18DaysInEgypt is an interactive crowd-**sourced documentary project** about the ongoing Egyptian revolution. ... **Check** out the latest stream on 18 **Days in Egypt**:

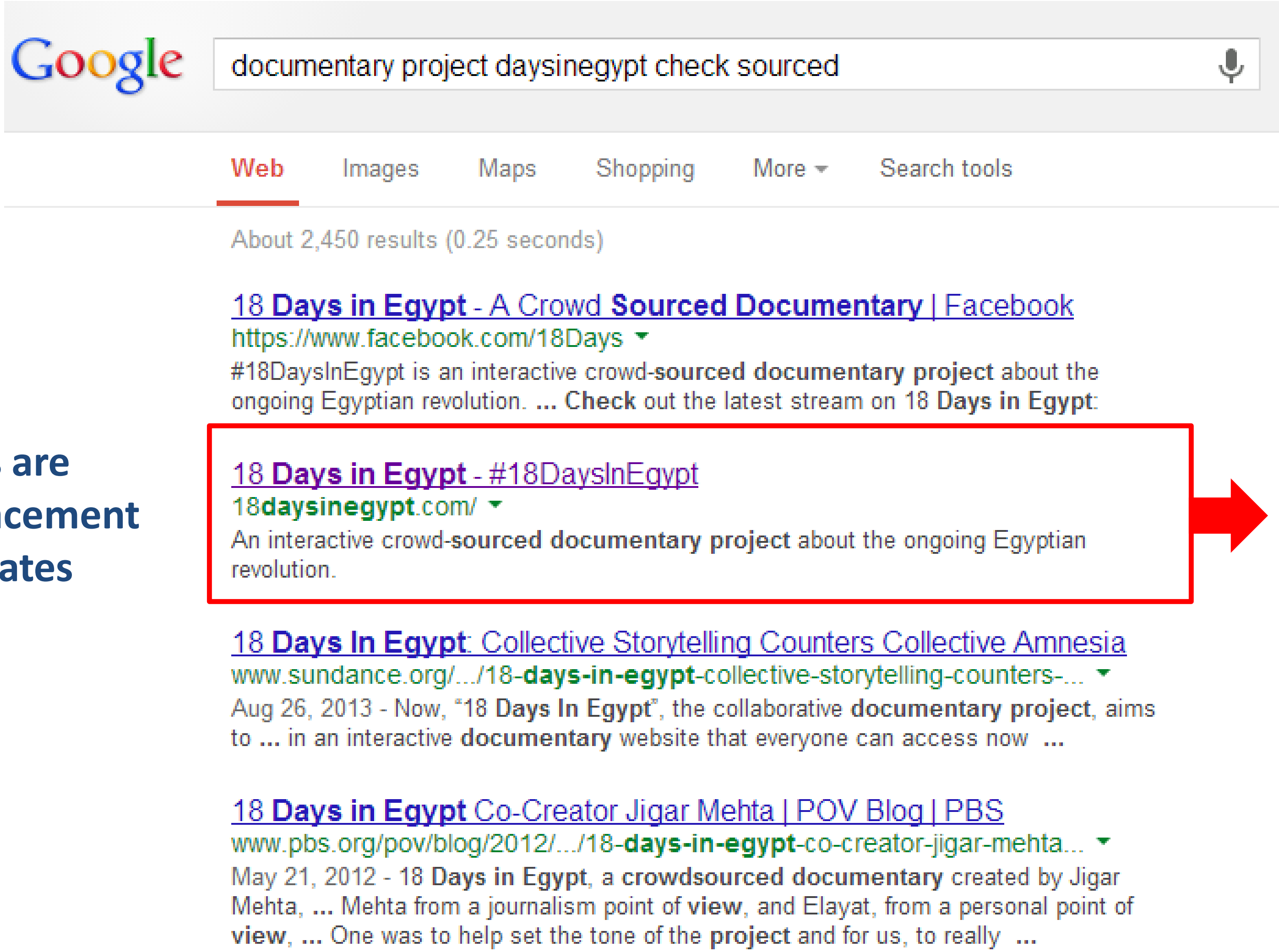
18 Days in Egypt - #18DaysInEgypt
18**daysinegypt**.com/
An interactive crowd-**sourced documentary project** about the ongoing Egyptian revolution.

18 Days In Egypt: Collective Storytelling Counters Collective Amnesia
www.sundance.org/.../18-**days-in-egypt**-collective-storytelling-counters-...
Aug 26, 2013 - Now, "18 **Days In Egypt**", the collaborative **documentary project**, aims to ... in an interactive **documentary** website that everyone can access now ...

18 Days in Egypt Co-Creator Jigar Mehta | POV Blog | PBS
www.pbs.org/pov/blog/2012/.../18-**days-in-egypt**-co-creator-jigar-mehta...
May 21, 2012 - 18 **Days in Egypt**, a **crowdsourced documentary** created by Jigar Mehta, ... Mehta from a journalism point of **view**, and Elayat, from a personal point of **view**, ... One was to help set the tone of the **project** and for us, to really ...

The original resource

Search Engine Results



documentary project daysinegypt check sourced

About 2,450 results (0.25 seconds)

18 **Days in Egypt** - A Crowd **Sourced Documentary** | Facebook
https://www.facebook.com/18Days
#18DaysInEgypt is an interactive crowd-**sourced documentary project** about the ongoing Egyptian revolution. ... **Check** out the latest stream on 18 **Days in Egypt**:

18 **Days in Egypt** - #18DaysInEgypt
18**daysinegypt**.com/
An interactive crowd-**sourced documentary project** about the ongoing Egyptian revolution.

18 **Days In Egypt**: Collective Storytelling Counters Collective Amnesia
www.sundance.org/.../18-**days-in-egypt**-collective-storytelling-counters-...
Aug 26, 2013 - Now, "18 **Days In Egypt**", the collaborative **documentary project**, aims to ... in an interactive **documentary** website that everyone can access now ...

18 **Days in Egypt** Co-Creator Jigar Mehta | POV Blog | PBS
www.pbs.org/pov/blog/2012/.../18-**days-in-egypt**-co-creator-jigar-mehta...
May 21, 2012 - 18 **Days in Egypt**, a **crowdsourced documentary** created by Jigar Mehta, ... Mehta from a journalism point of **view**, and Elayat, from a personal point of **view**, ... One was to help set the tone of the **project** and for us, to really ...

Others are good replacement candidates

The original resource

Recommendation Evaluation

We extract a dataset of resources that are currently available:

- Pretend these resources no longer exist (for a baseline)
- Each of the resources are textual based
- Each resource have at least 30 retrievable tweets.

→ Extracted 731 unique resources

Recommendation Evaluation

We use boiler plate removal library* to remove the template from the:

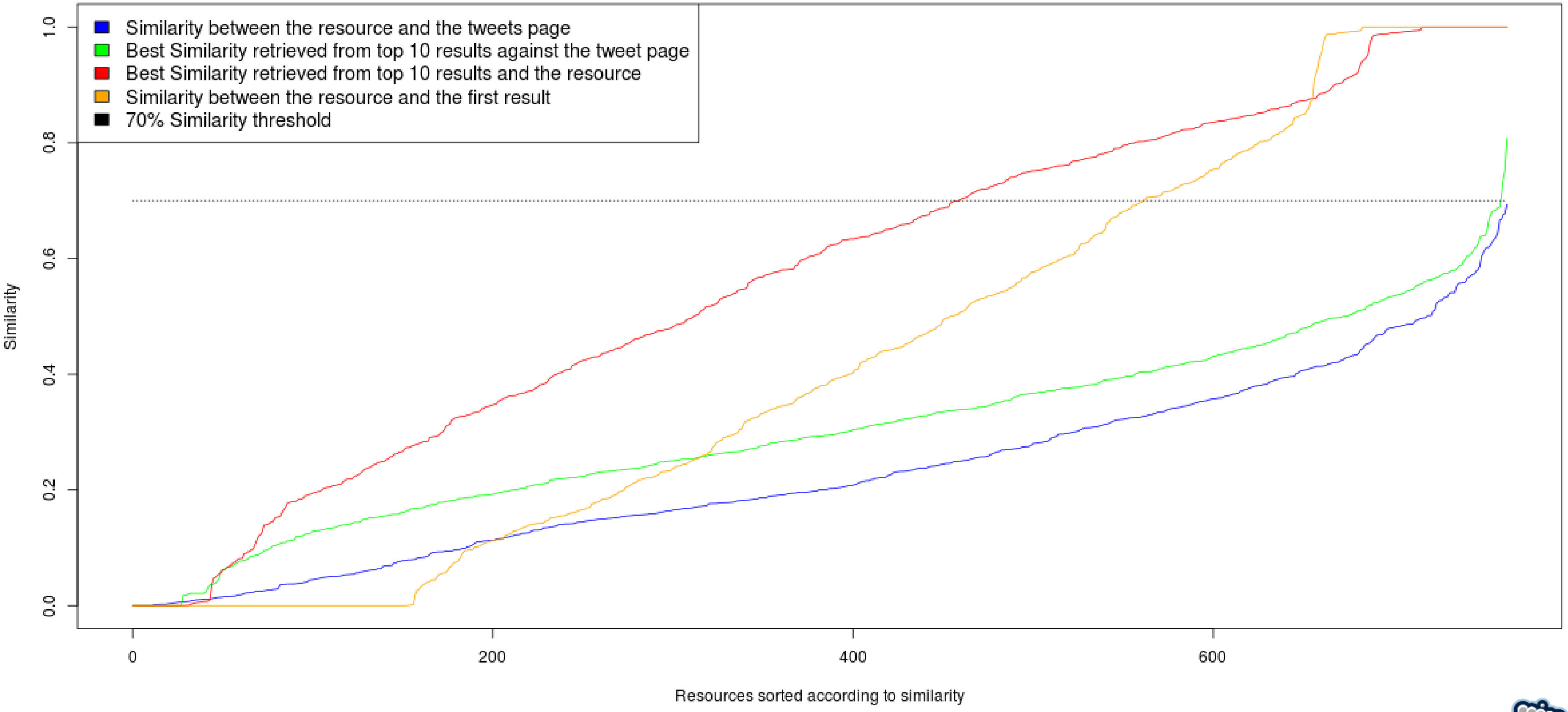
•linked resources

•top 10 retrieved results from Google

→We use *cosine similarity* to compare the documents

* https://github.com/misja/python-boilerpipe

Similarity Measures In Resource Replacements



Results

→41% of the test cases we can find a replacement page with at least 70% similarity to the original missing resource

→The search results provide a mean reciprocal rank of 0.43

Conclusions

- We validated our model in predicting the resource existence on the current web with ~4% error after one year.
- The archival prediction on the other hand produced a large error ~11.5%
- We explored a phenomenon of reappearing on the web after disappearing (6.54%), and disappearing from the archives too (7.89%).
- The removal of search engine caches in the most recent Memento revision could be a possible explanation of the disappearance from the archives.

Cont. Conclusions

- Measured the estimated percentage missing from the tweets ~10.5%.
- Utilized Topsy API to extract context information about tweets with missing resources.
- Investigated the viability of finding a replacement resource to the missing one.
- In 41% of the cases we were able to extract a replacement resource that is 70% or more similar to the missing resource.