690V-DHARIT

VAST CHALLENGE - 2011

By: Divyesh Harit

College of Information and Computer Sciences, University of Massachusetts Amherst

CHALLENGE INTRODUCTION

Vastopolis is a city where a mysterious disease seems to be moving through the population. There are some local activist groups that may be suspicious.

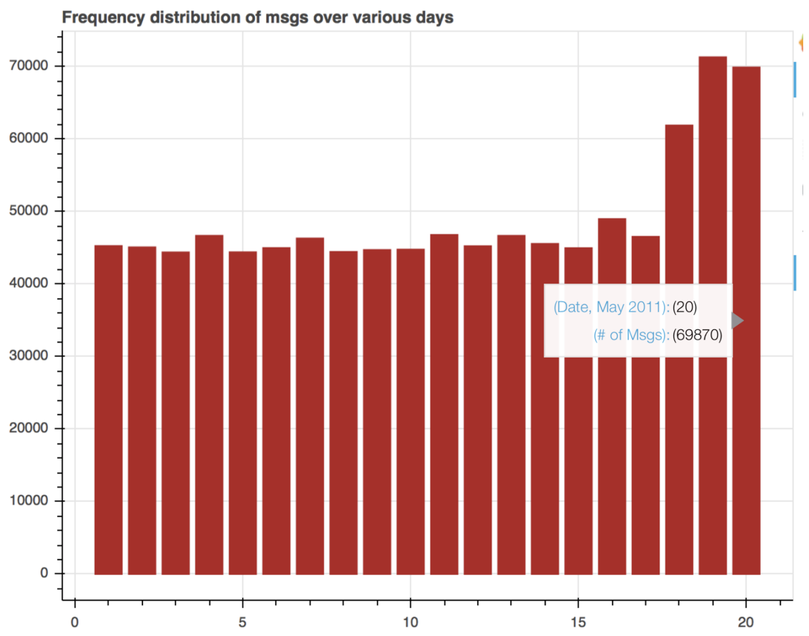
MC1: Provided with a large set of geo-coded messages, explain how the epidemic is spreading.

MC3: Use provided news reports to determine if there was potentially terrorist action.

APPROACH TO MC1

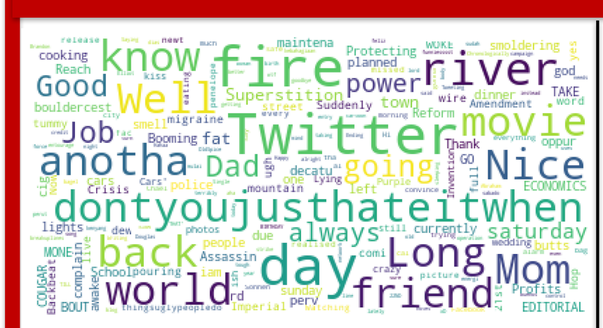
1. See frequency of messages by date
2. See most frequent words of busy days
3. Cluster by location per day
4. Cluster by symptom *and* location per day

MC1: MESSAGE FREQUENCY



Sudden bump in number of messages over last 3 days. Possible start of the epidemic? Looking deeper into messages over these days should help.

MC1: BUSY DAYS' TOP WORDS



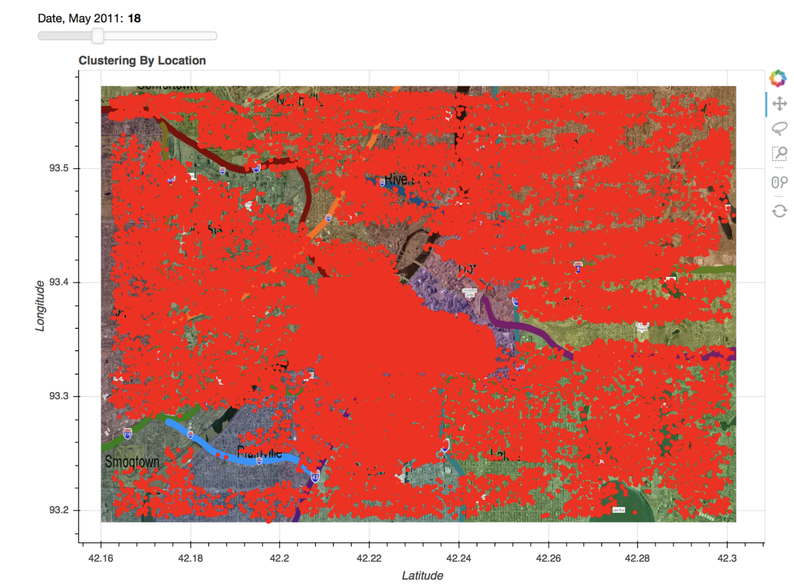
17th May

18th May

19th May

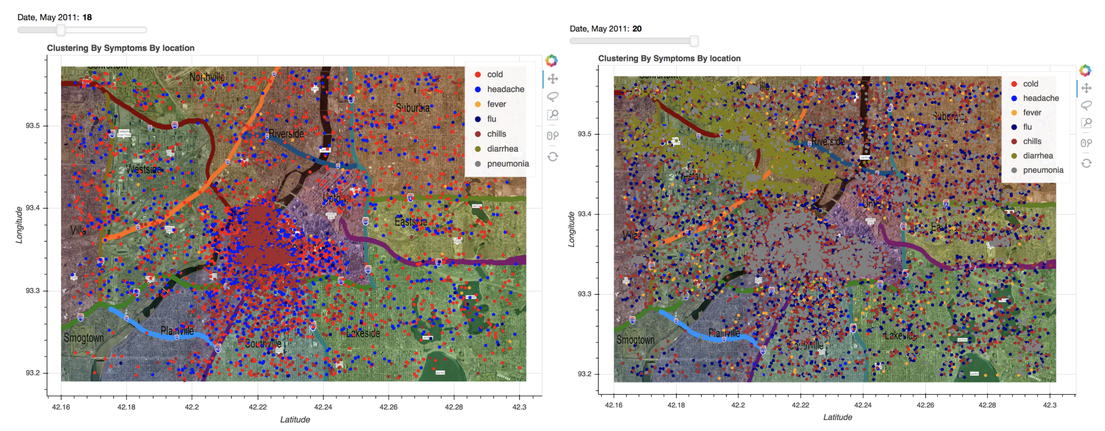
20th May

MC1: CLUSTERING MSGS BY LOCATION

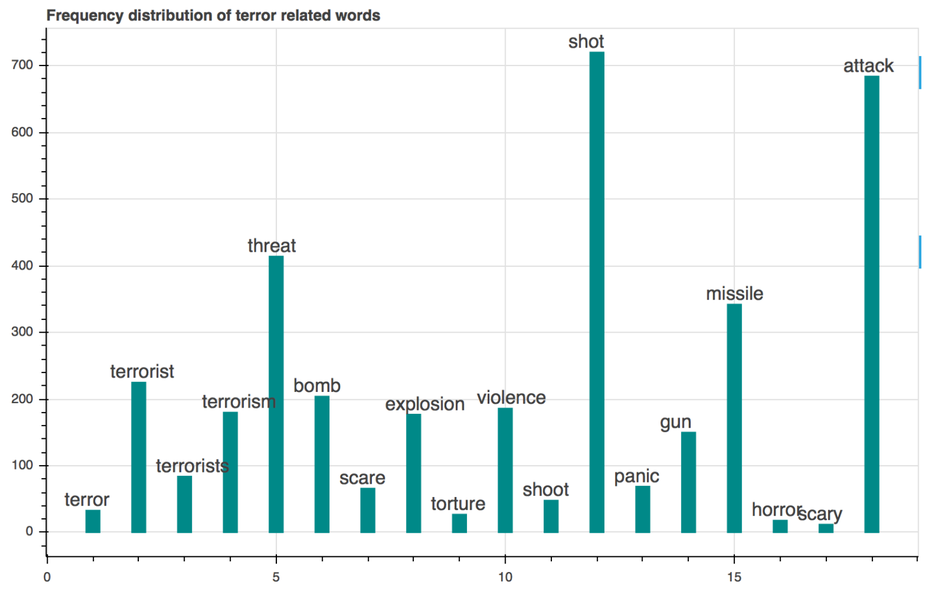


Increased concentration of messages in downtown, uptown and other central areas as the days progress

MC1: CLUSTERING BY SYMPTOMS + LOCATION



MC3: LOOKING FOR TERROR-RELATED WORDS



WHAT DIDN'T WORK: MC3 TOPIC MODELING

```
Topic 0: exchange tuesday political tax derryberry going rates late chicago policy
Topic 1: group index expected higher use early products board director states
Topic 2: world national general unit city results set current costs large
Topic 3: shares codi york net chairman funds hong economy yen francs
Topic 4: mr state make lost issues bonds communications value dow employees
Topic 5: market like trading far month fund white game losses family
Topic 6: billion investors fell industry games thursday end including line pay
Topic 7: year law lower increase volume th european offer told today
Topic 8: sales week house news work deal major high number trade
Topic 9: company analysts chief officials financial right firm department point july
Topic 10: share companies money friday international securities kong small better used
Topic 11: new time cents day months campaign income power buy strong
Topic 12: bank federal growth home profit investment technology network rate ended
Topic 13: said million quarter plans long help local hit japan press
Topic 14: government corp american way good public country want services ms
Topic 15: stock stocks big second past issue ago reported team little
Topic 16: president years rose report economic revenue wednesday union come need
Topic 17: says people business price earlier plan markets foreign monday internet
Topic 18: say party according convention left sell university vice loss job
Topic 19: prices china executive earnings recent service march close school problems
```

No sign of terror words in the top 20 topics