

Addressing the Sparsity of Location Information on Twitter

Dimitris Kotzias^{*◊}, Ted Lappas⁺, Dimitrios Gunopulos^{*}



^{*}University of Athens

⁺Stevens Institute of Technology

[◊]University of Oxford



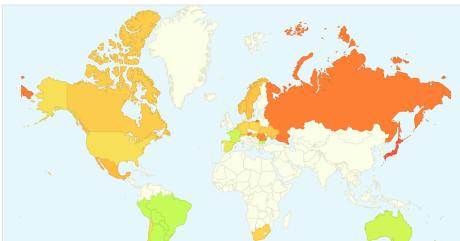
STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY

Outline

- Motivation
- Problem Definition
- Graph Analysis
- Max Edge Algorithm - Users
- QL, QLU - Tweets
- Results
 - Users, Tweets, LDA

Twitter Location Information is Important

google.org Flu Trends



Twitter India
@twi

We now have local trends for 6 cities in India!
Ahmadabad, Chennai, Hyderabad, Bangalore,
Delhi, and Mumbai: pic.twitter.com/MXjreNjZ

 **Big Chorizo**
@chilemasgrande 

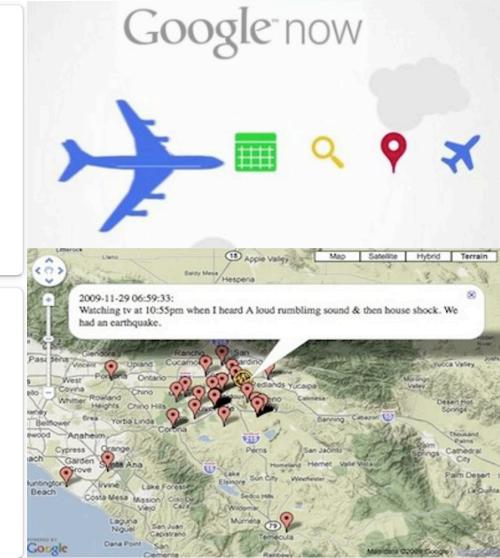
My sources say Whitney Houston found dead in Beverly hills hotel.. Not in the news yet!!
2:30 AM - 12 Feb 2012

39 RETWEETS 45 FAVORITES 

 **The Associated Press** 
@AP

BREAKING: Publicist Kristen Foster says singer Whitney Houston has died at age 48
2:57 AM - 12 Feb 2012

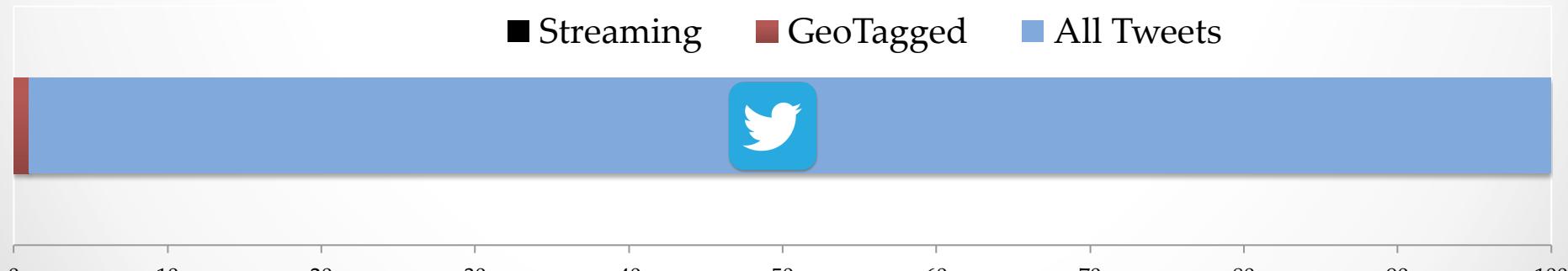
11,745 RETWEETS 210 FAVORITES 



- Event - Emergency Detection, Spread of Disease, News Propagation, Advertisement...
- Search, Trend Detection
-

Twitter Location Information is Important

- ...but also very rare!
 - 1% of users provide geographical location (Valkanas, ICDM'12)
 - 1% of tweets can be crawled from Twitter Stream
 - 48% provide City level accuracy in profile (Hecht, CHI'11)



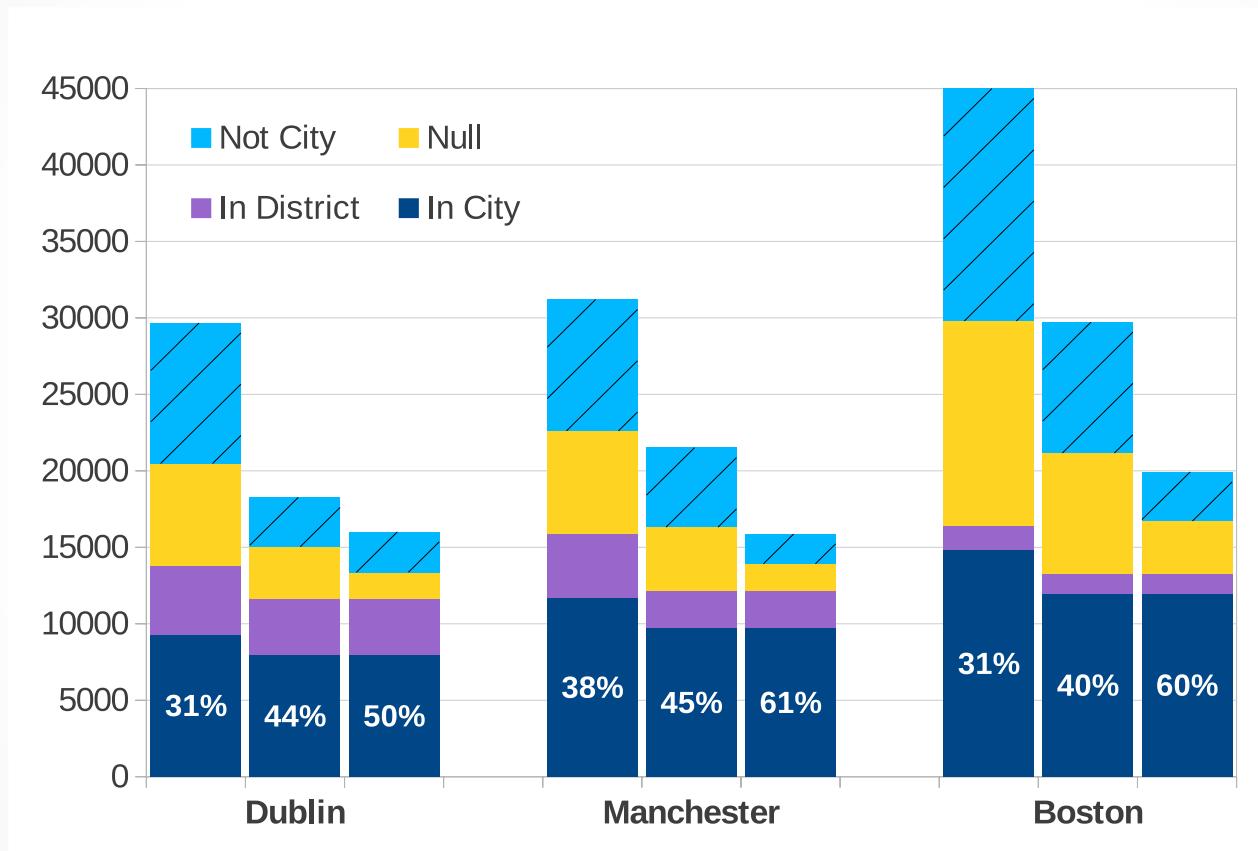
Related Work

- Locate Users
 - Local phrases(Cheng et. al., CIKM'10)
 - Named Local Phrases (Mahmud et. al., ICWSM'12)
 - Local Topics (Ahmed et. al., WWW'13)
 - *Location of all friends (Backstrom et. al., WWW'10)
 - Disambiguation on the Map (Valkanas et. al., ICDM'12)
- Locate Tweets
 - Points of interest (Li et. al., CIKM'11)
 - Query Likelihood (Kinsela et. al., SMUC'11)
 - Match Expression w. Location (Ikawa et. al., WWW'12)

Problem Definition

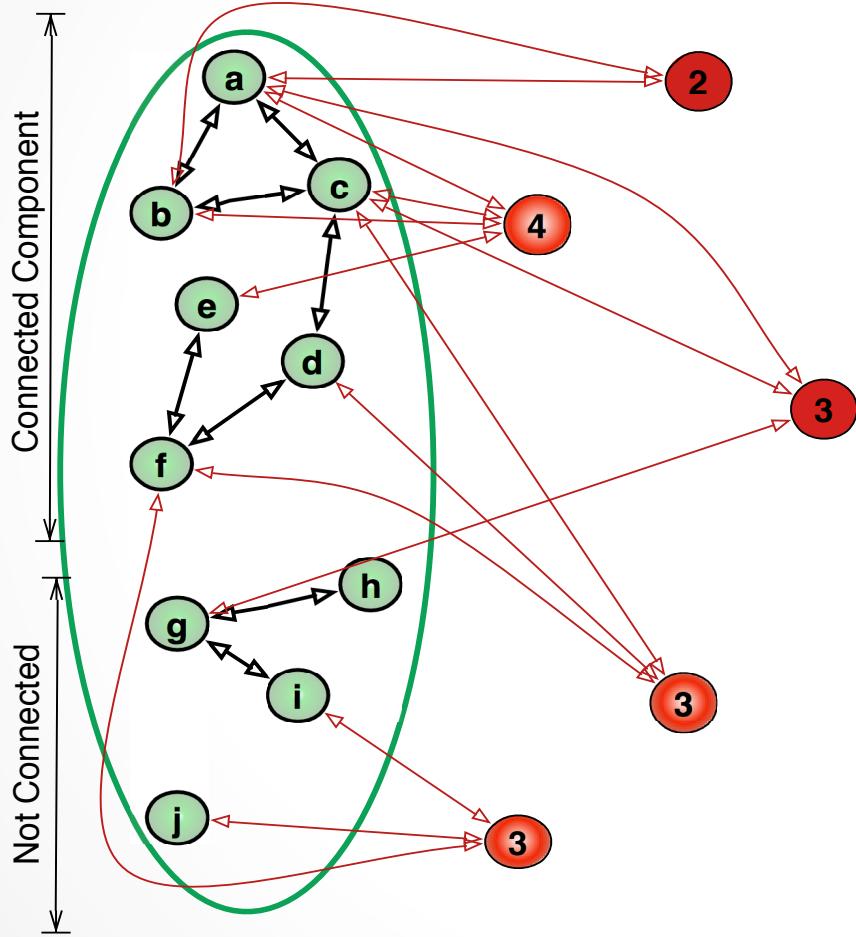
- Reverse the problem
 - More users from a specific city
- Consider Twitter limitations
 - Minimize Number of Queries
- Common framework for both users and tweets
 - Find Users, Accurately locate their tweets within the city

Graph Analysis



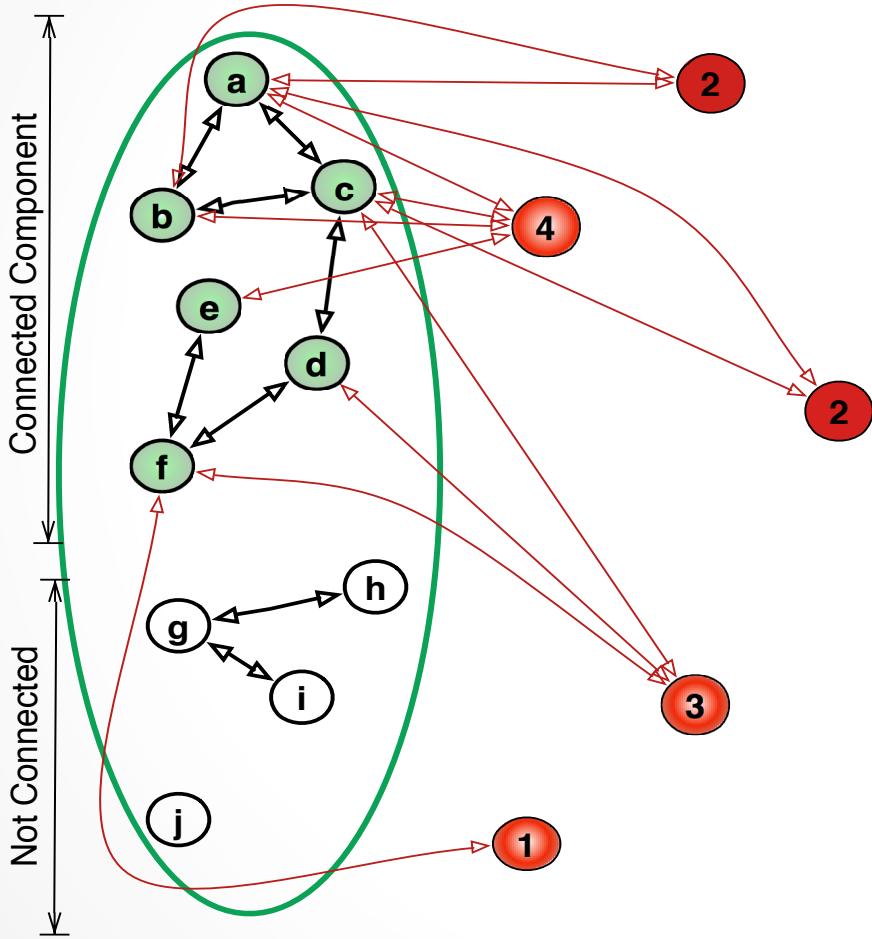
Location of Twitter Users, from Streaming API
(raw, largest connected, tweet 5 times)

Max Edge Algorithm - Users



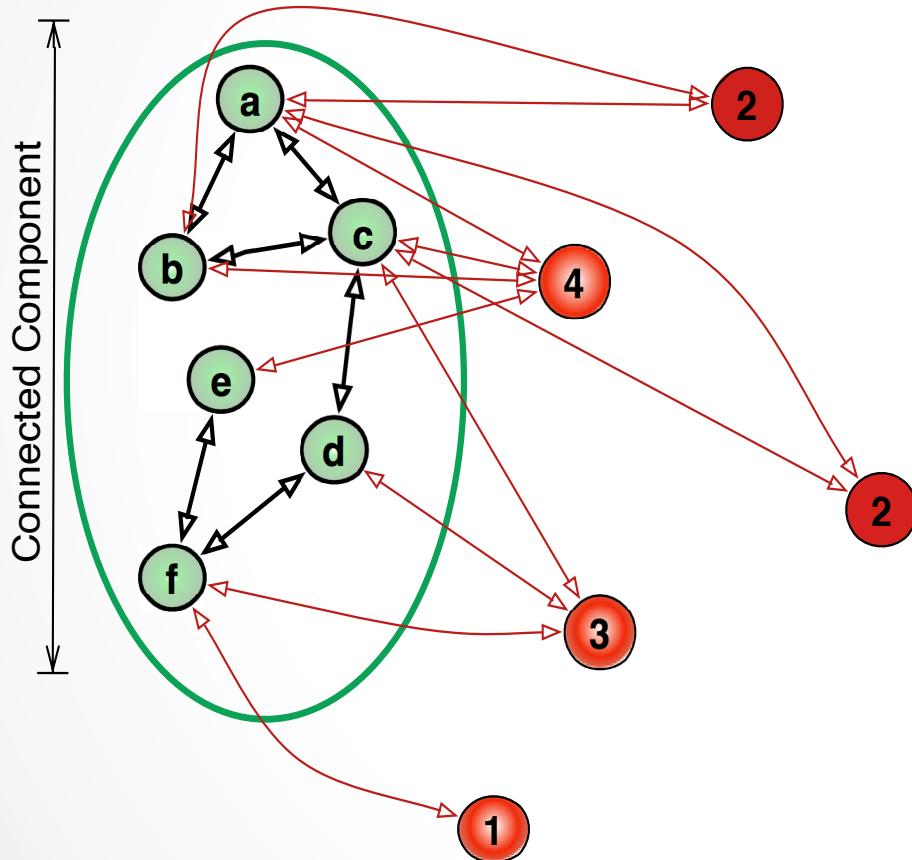
- Create User Graph
- Find largest Connect Component

Max Edge Algorithm - Users



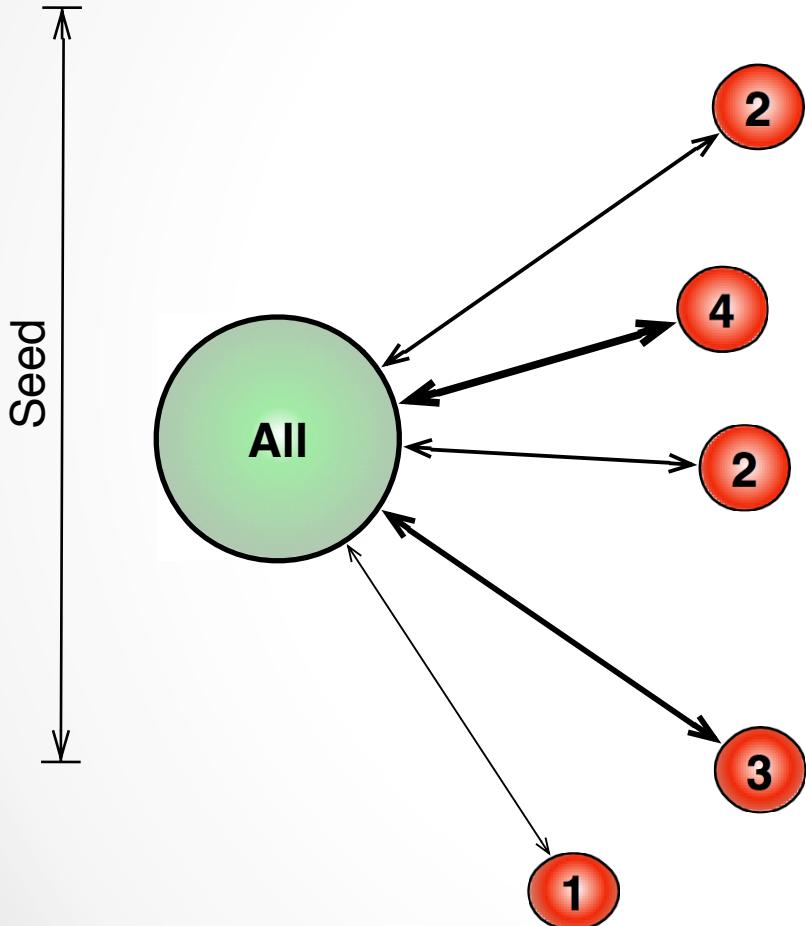
- Choose seed

Max Edge Algorithm - Users



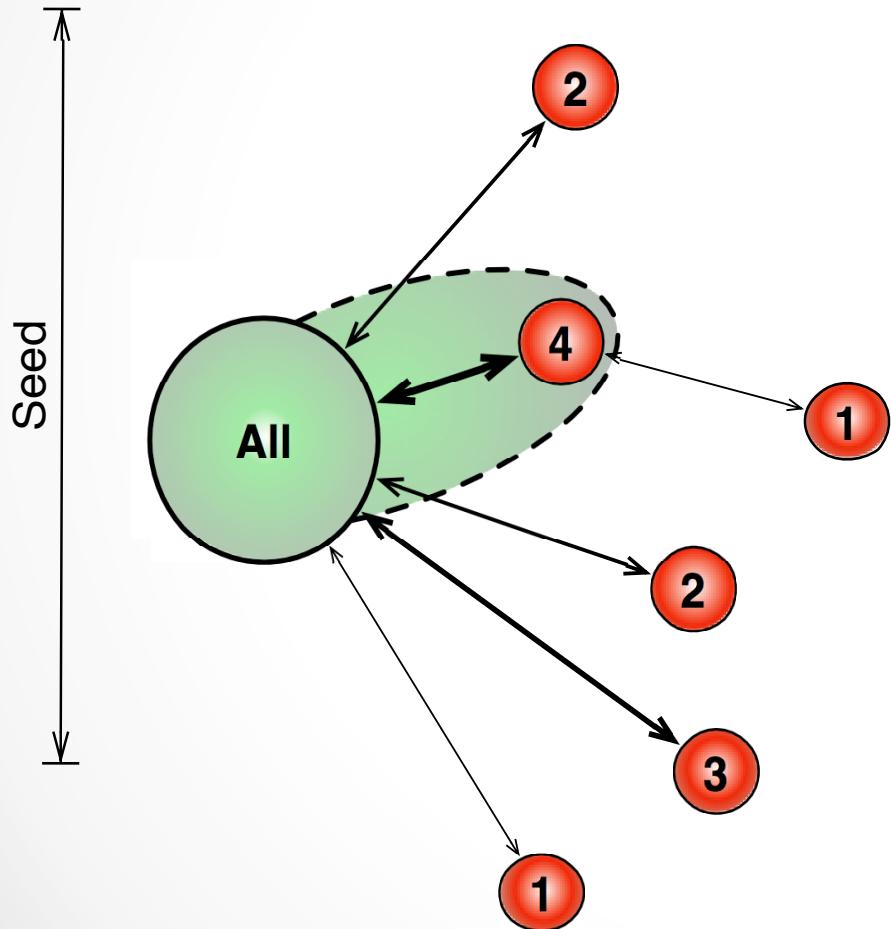
- Prune the rest
- Update weight of external edges

Max Edge Algorithm - Users



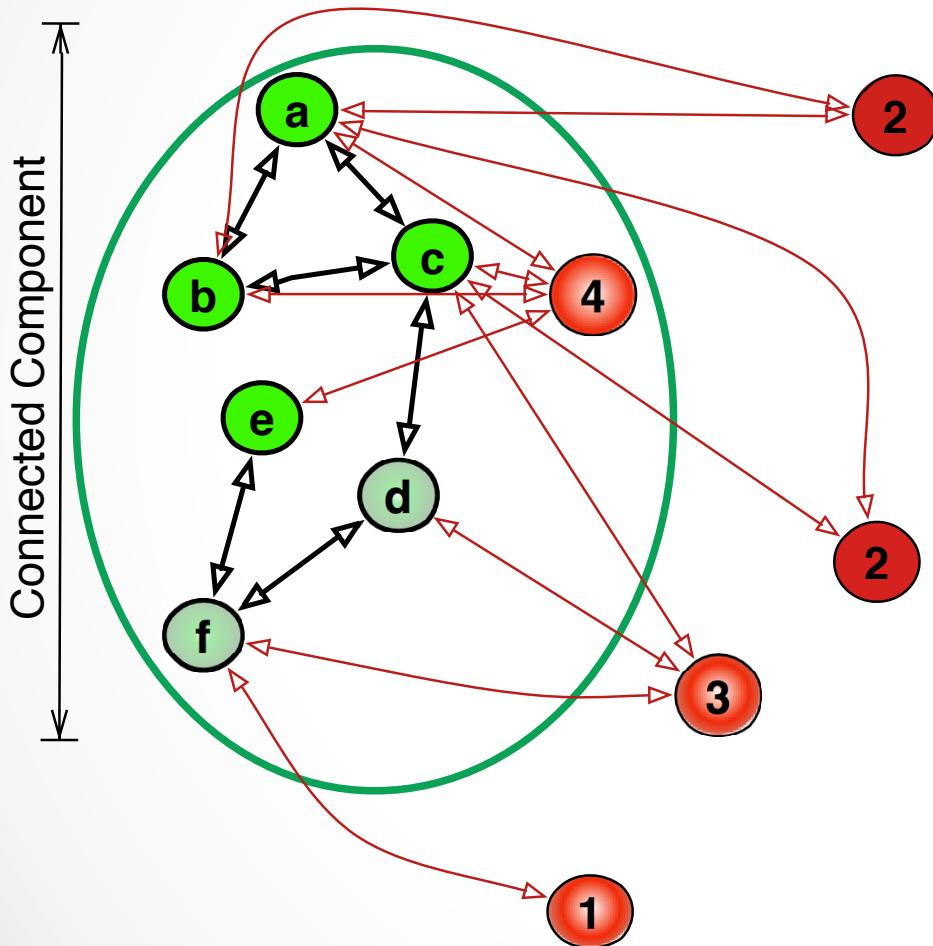
- Consider all as one Node
- Query Node with Max Edge weight

Max Edge Algorithm - Users



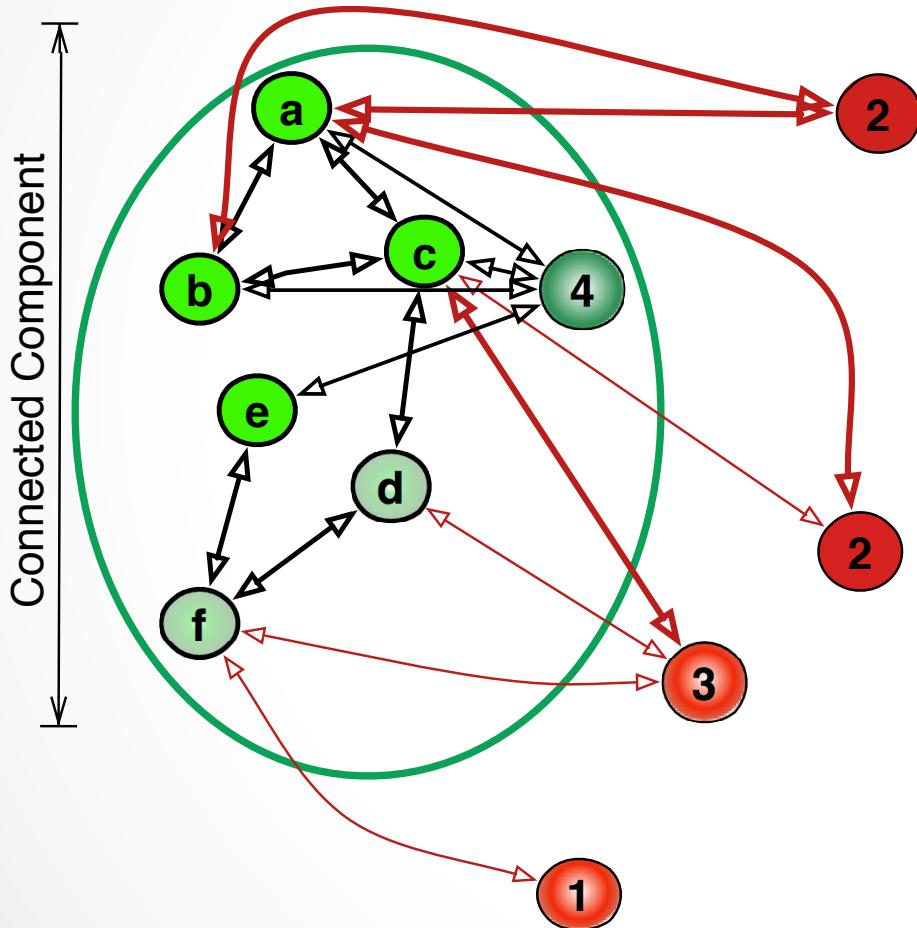
- If in city, add him to the seed
- Maximize conductance
(Community level, Leskovec et. al, WWW'08)

Max Edge Algorithm - Users



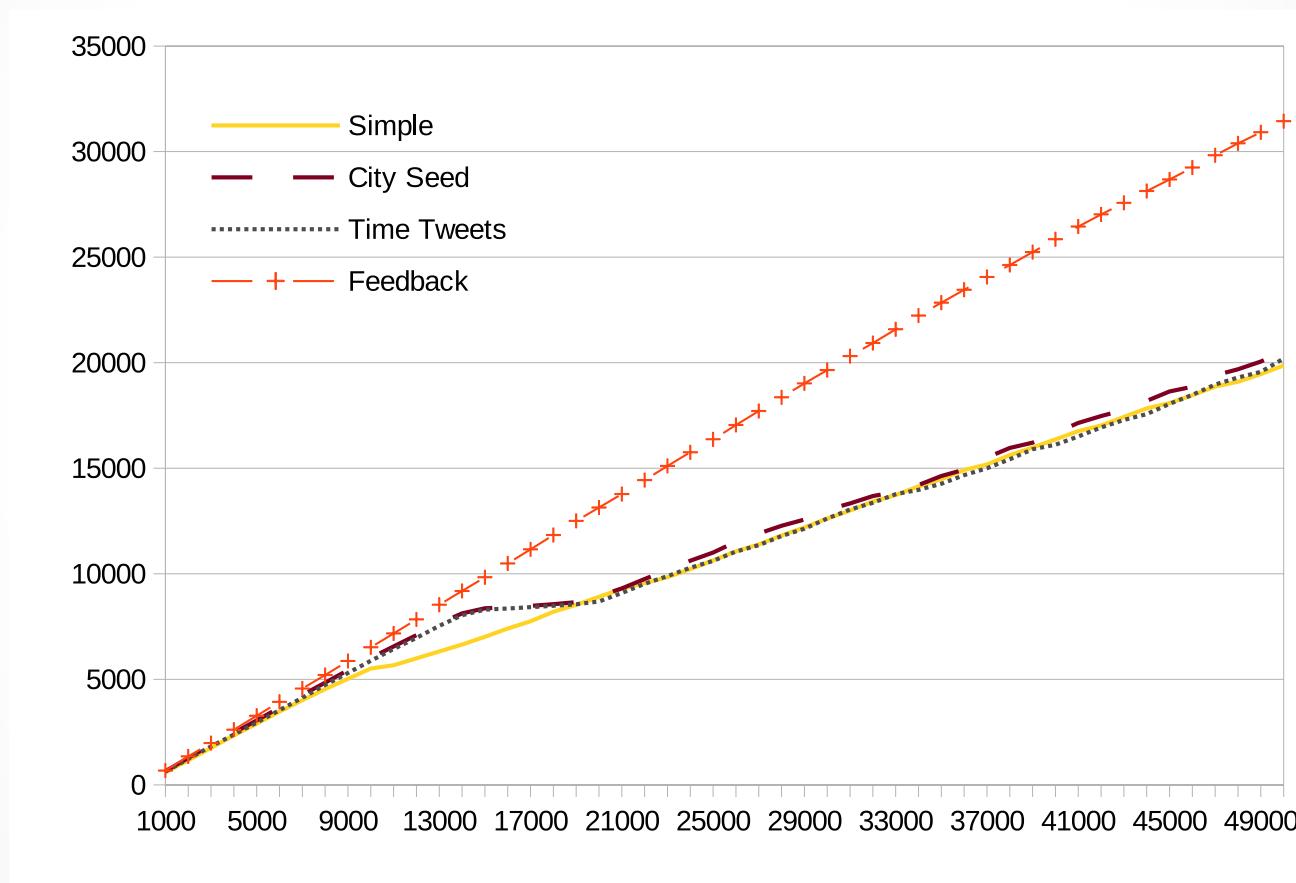
- Essentially, users vote for next node
- Find users who voted correctly

Max Edge Algorithm - Users



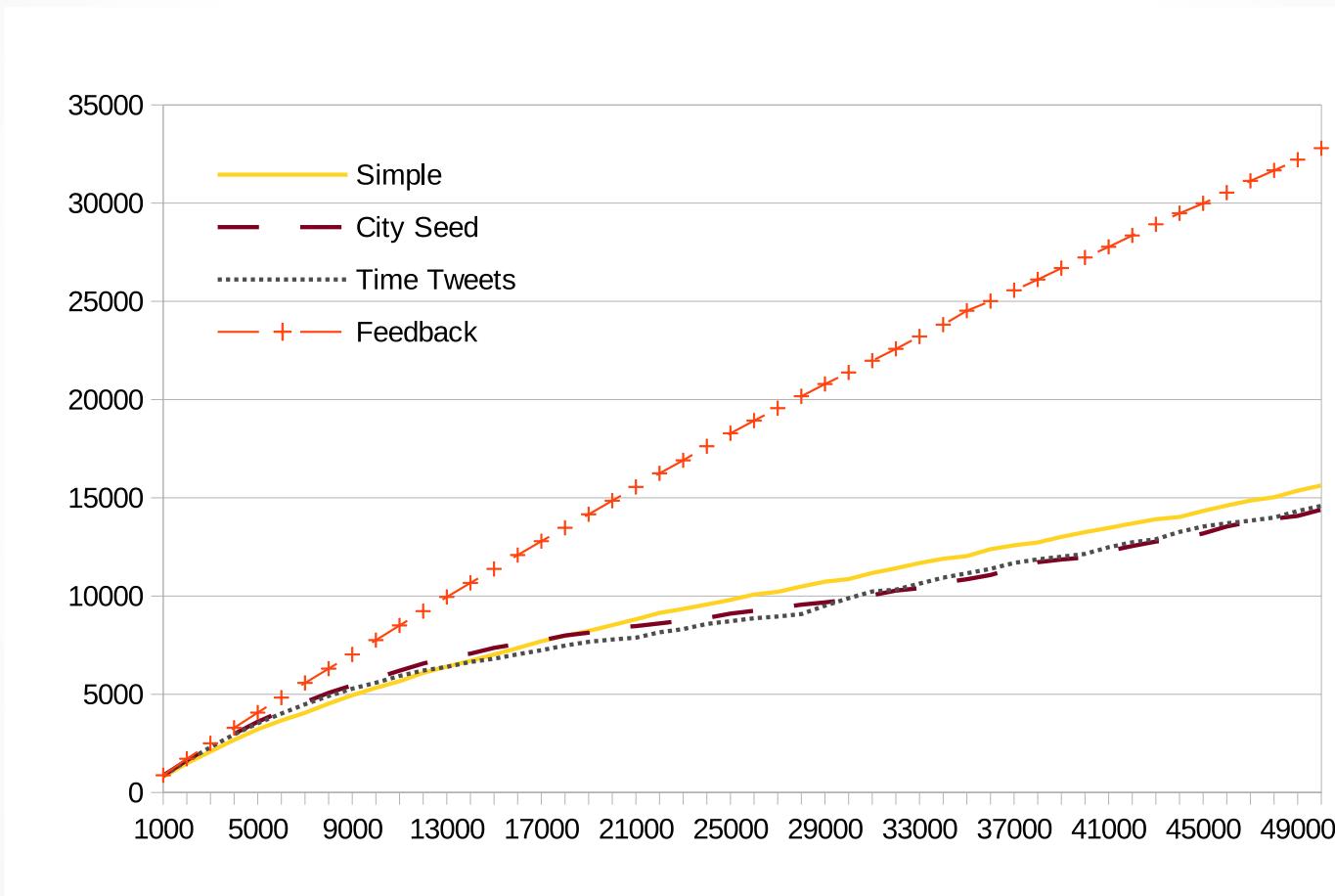
- Feedback
- Increase the weight of “correct” users

Users per Query- Dublin



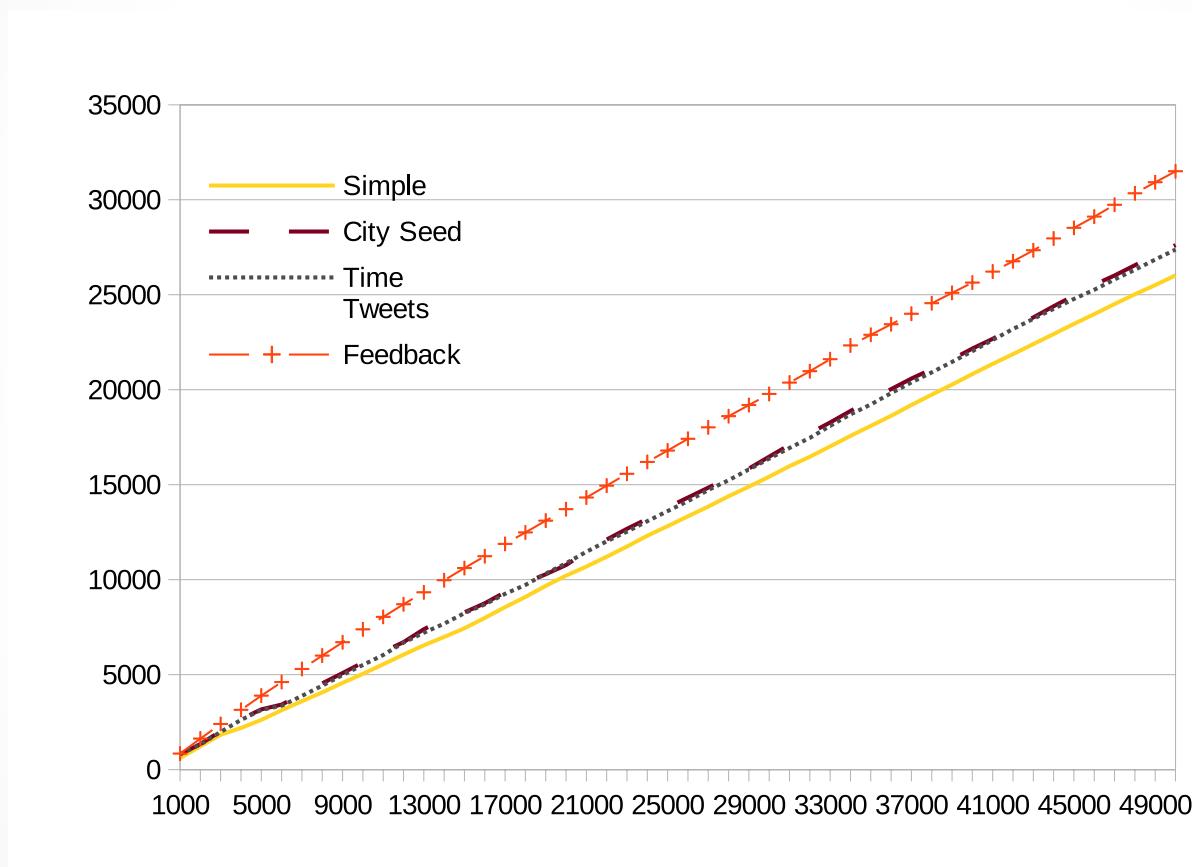
Number of Queries v Number of Users in Dublin

Users per Query- Boston



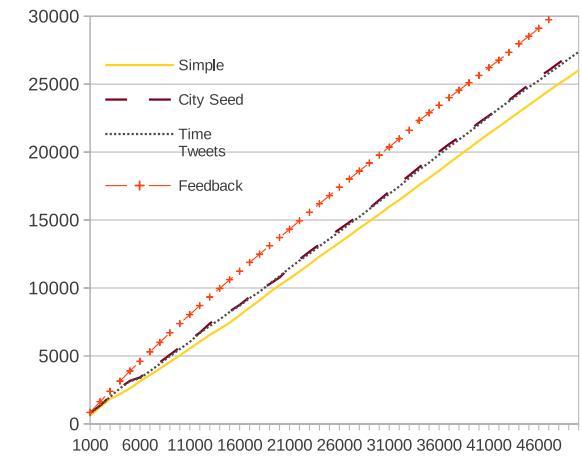
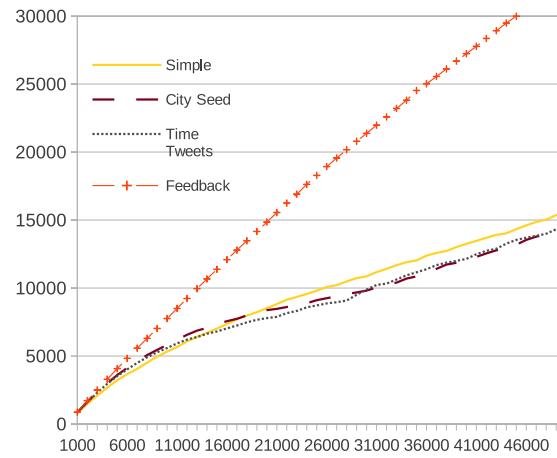
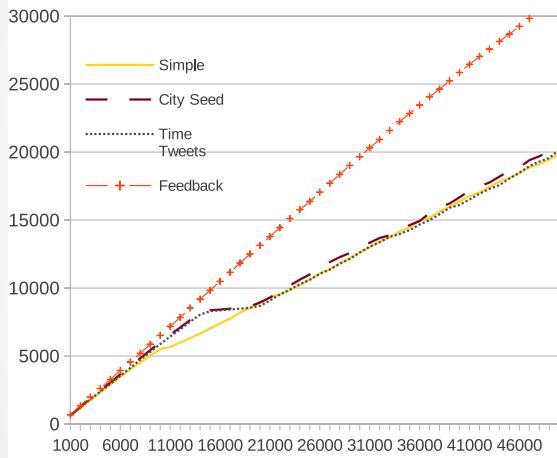
Number of Queries v Number of Users in Boston

Users per Query- Manchester



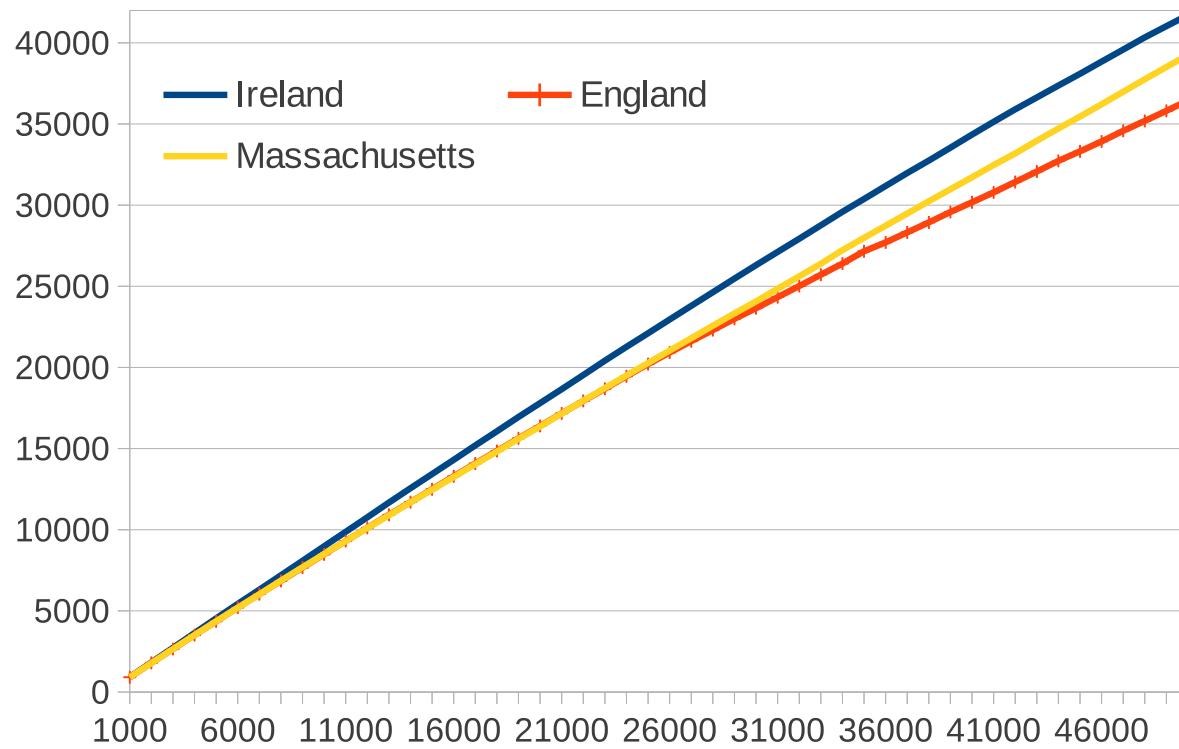
Number of Queries v Number of Users in Manchester

Results for Users



Number of Queries v Number of Users for All cities

Results for Region

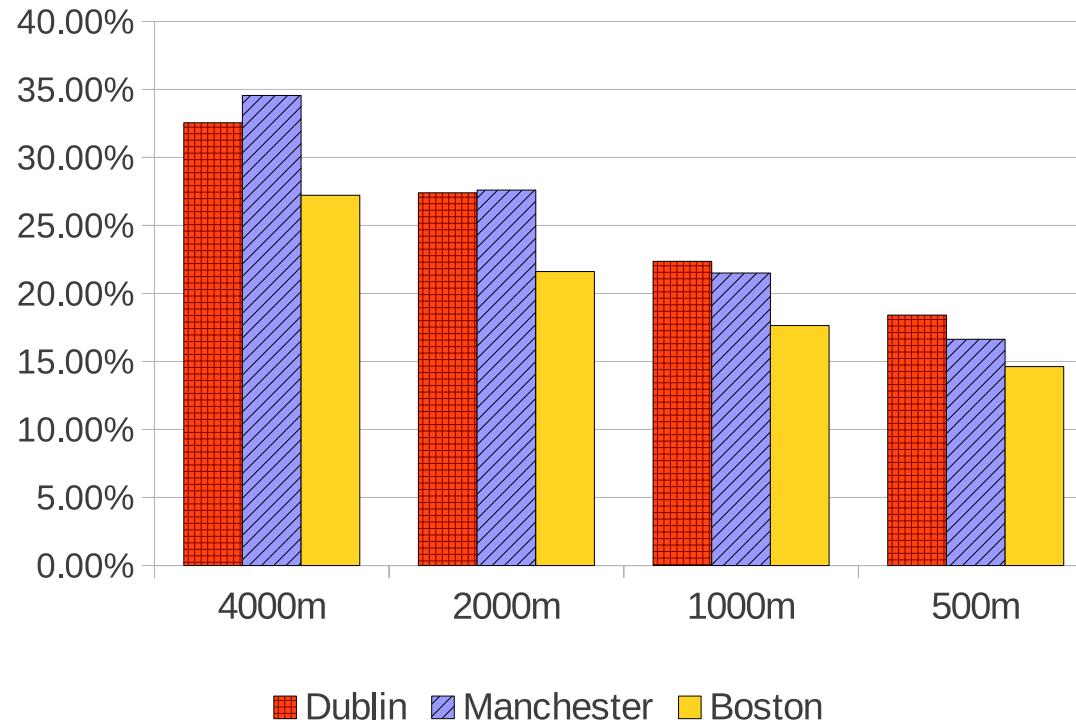


Number of Queries v Number of Users inside its general Region

QL, QLU – Tweets

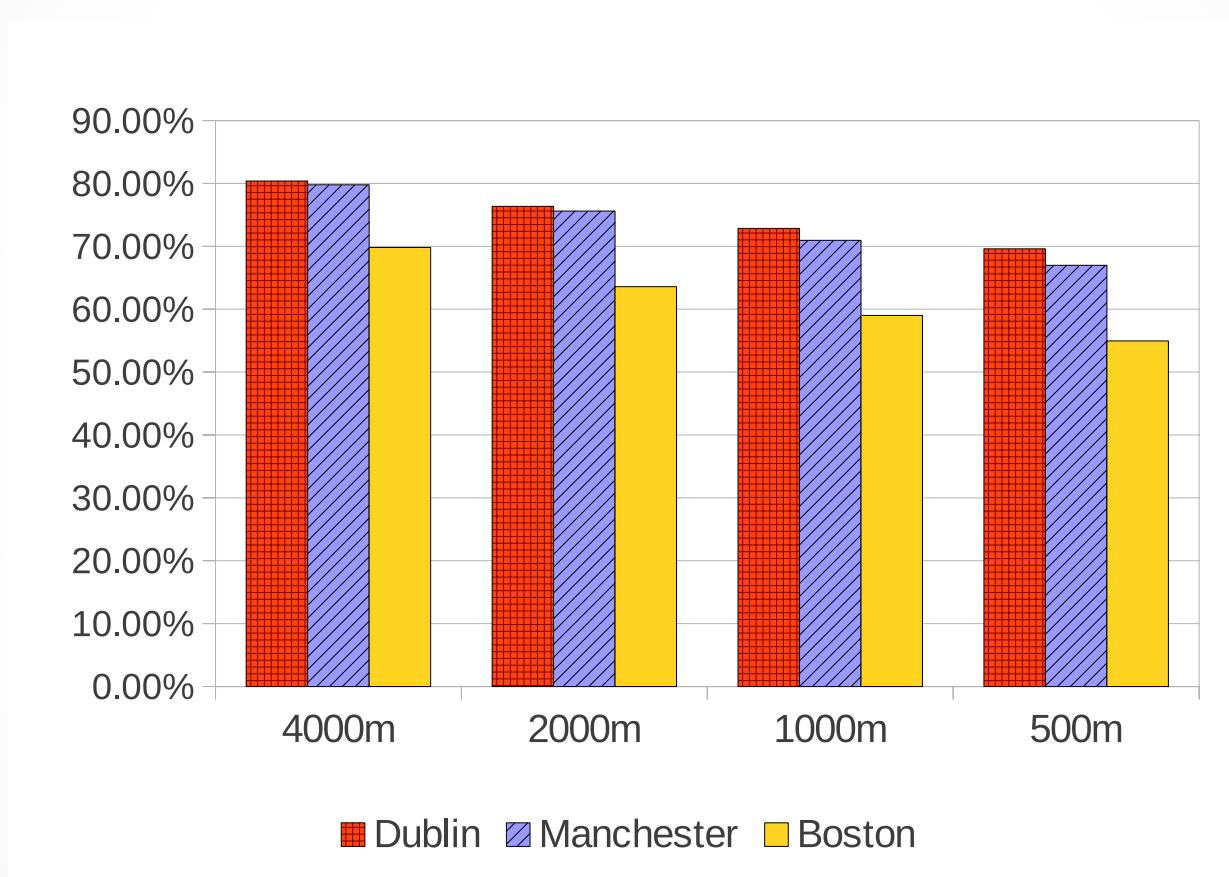
- Make Grid
- Consider each box as a document
- Maximum Query Likelihood for documents
- QLU: Most frequent box of user

Locating Tweets with QL



Precision for QL method for different accuracy levels

Locating Tweets with QLU

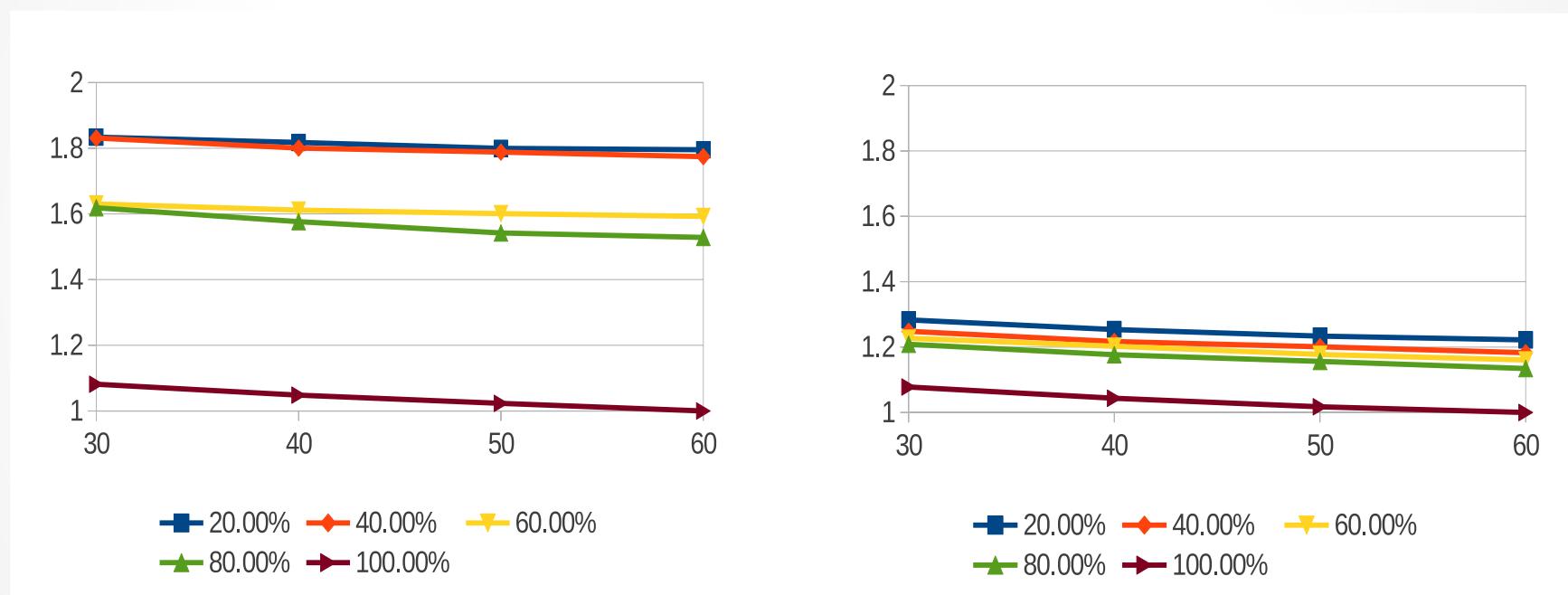


Precision for QLU method for different accuracy levels

Dataset Improvement

- x4 Users (20k to 80k)
- x8 Tweets (5M to 40M)
- LDA Perplexity (Wallach, ICML'09)
- LDA Topics

LDA Perplexity



Tweets from Streaming API

New Tweets

LDA Topics - Geo

		ireland women abortion vote france vinb law issue politics country bill pro europe reading point government tax people bank
20%	abortion	dublin ireland bar park photo guin dub street post tha cliath howth aerfort st bhaile square hotel temple green
20%	tourism	united mufc great castle de enjoy day pal weekend love mate massive yea ffriteback matchday stanakatic caskett hun fabfridayfollow
20%	football	lion leinster rugby ihl fitness lions coybib gatland trx bod fitfam pm ym seaofre final match anglotape fella heinekencup
20%	rugby	donegal final christina tom bloomsday roscommon fe lebron casa kenny patrick theapprentice hill floor princess rackard ding alley min
100%	?	dublin ireland bar guin dub tha cliath aerfort bhaile pic hotel airport temple storehouse pub st pint trinity college
100%	tourism	player united season game play arsenal football sign team haha league fan goal win suarez man mate mufc chelsea
100%	football: uk	check play rt music gig album video film listen show eventsindublin live song awesome watch band festival game cat
100%	music show	lion game great win match play final gaa leinster team rugby golf ireland today wimbledon player murray congrats dub
100%	rugby	watch man life show call god car made hit kid years laugh mr tom men face jesus dream break
100%	?	

LDA Topics - Full

		business ireland job dublin social jobfairy marketing digital media hire
20%	irish jobs	great tip irishjob startup network tech sales online company
20%	Politics:EC	ireland bank tax eu people europe pay report government uk protest minister year state home court news service job
20%	food	food restaurant wine lunch taste special beer coffee free cocktail dinner menu delicious chef recipe bar eat lovely yum
20%	abortion	abortion vinb life women ireland bill vote seanad pro people baby politics law party dail fg debate prolife labour
20%	?	ur ya ye im goin yea dont wat tho pal ha gettin il wit ill ah nite jus bout
100%	Politics: EC	ireland bank tax news eu britain report uk government police protest europe belfast human gold syria attack year minister
100%	research	ireland great health support student research eu event conference school day people children education week work launch europe
100%	Abortion+	abortion ireland vinb women vote life people bill seanad law party support politics pro dail debate woman gay marriage
100%	irish jobs	job dublin ireland business jobfairy irishbizparty hire sales recruit company bizhub engineer developer senior service client
100%	?	ur ya ye ha goin haha wat tho gettin il wit im dat fuckin de jus day yea nite

Conclusions

- Correlation between connectivity in the social graph and location in the real world
- Improve dataset size and quality

Thank You

- Questions?