

# TOPIC MODELING FOR INFRASTRUCTURE-RELATED DISCUSSIONS IN ONLINE SOCIAL MEDIA

MAZDAK NIK-BAKHT (Post Doctoral Research Associate)  
TAMER EL-DIRABY (Associate Professor)

University of Toronto



The CSCE International Construction Specialty Conference  
UBC, Vancouver, BC  
June, 2015



# Background

- Stakeholders of Urban infrastructure projects:
  - Internal (directly involved in decision making process)
    - official decision makers (Government Dept.s & Private Sectors)
    - Technical decision makers (Planners, Contractors, suppliers, etc.)
  - External (affected by the decisions made and related operations)
    - Local Community
    - General Public
    - Project Affected Groups
    - Local Neighborhood Members
    - Pressure Groups Such As NGOs.
    - News Media



*[Atkin & Skitmore 2008]*

# Background

- [External] Stakeholders Engagement:  
**Public Involvement (PI Programs)**

- **OFFLINE**

- Public Meetings
- Open Houses
- Workshops
- Surveys
- ....

- **ONLINE**

- Web 2.0

**“Micro-participation”**

*[Evans-Cowley & Griffin 2012]*



# Twitter – a good example



**No YYC Tunnel**  
@NoYYCTunnel

Join Citizens Opposed to the Airport Tunnel. The \$823-million tunnel is too expensive for taxpayers and offers only limited transportation benefit for Calgary.

Calgary, AB · noairporttunnel.ca

392 TWEETS 134 FOLLOWING 95 FOLLOWERS

**No YYC Tunnel** @NoYYCTunnel 8 Nov 10  
Kudos Farrell, Hodges, Lowe, MacLeod, Pincott, et al for making #yyccc beef up \$250K YYC study to consider other transportation issues #yyc  
Details

**No YYC Tunnel** @NoYYCTunnel 8 Nov 10  
Bad day for taxpayers. Mayor and 8 aldermen vote down comprehensive business case study of \$500M YYC tunnel. <http://tiny.cc/c7bv7> #yyccc  
Details



**Yes Tunnel**  
@yestunnel

Welcome to YESTunnel, the Twitter page dedicated to saving our link to YYC. Here find ongoing info on the progress of our mission!

Calgary, AB · yestunnel.com

547 TWEETS 131 FOLLOWING 76 FOLLOWERS

**Yes Tunnel** @yestunnel 20 Mar  
@SheratonCalgary  
Details

**Yes Tunnel** @yestunnel 17 Mar  
The  
Details



**Enbridge NGP**  
@NorthernGateway

This is the official Twitter account for Enbridge Northern Gateway Pipelines. Please join the conversation. Tweets by Northern Gateway communications team.

Canada · NorthernGateway.ca

1,687 TWEETS 1,486 FOLLOWING 3,623 FOLLOWERS

Followed by [Sustweetability](#).

**Enbridge NGP** @NorthernGateway 3h  
RT @suncorenergy: New #OSQAR blog post: Is there a carbon bubble? [owl.lilyjrx.com/oilsands](http://owl.lilyjrx.com/oilsands)  
Details

**Enbridge NGP** @NorthernGateway 4h  
@ghoberg It's not reasonable to tell BCers that "Enbridge refused to provide even basic information about environmental safety"  
Details

the 100 strategic infrastructure projects\* in NA have had accounts since 20

in Strategic Infrastructure Leadership For



**Central Subway**  
@Central\_Subway

Extending Muni's T Third Line to connect San Francisco's southeast neighborhoods with SoMa, downtown and Chinatown. Follow us for project news and updates.

San Francisco · centralsubwaysf.com

662 TWEETS 398 FOLLOWING 2,026 FOLLOWERS

**Central Subway** @Central\_Subway 29 May  
Check it out! RT @sfmuni: Get the info you need to get going--anytime, anywhere--on the new [sfmuni.com](#)!  
Details

**Central Subway** @Central\_Subway 29 May  
Great pics! RT @7x7: We went underground to document the progress of the @sfmuni's central subway construction <http://t.ly/lu85J>  
Details



**Pont Jacques-Cartier**  
@PontJCBridge

Le pont Jacques-Cartier est un pont à cinq voies qui relie Montréal au littoral sud du Saint-Laurent à la hauteur de Longueuil.

Montréal · pontjacquescartierbridge.ca

1,044 TWEETS 273 FOLLOWING 3,596 FOLLOWERS

**Pont Jacques-Cartier** @PontJCBridge 10h  
Pont Jacques-Cartier : aucuns travaux majeurs ce week-end sur le pont (1er et 2 juin) [goo.gl/fb/rh4Ng](http://goo.gl/fb/rh4Ng)  
Details

**Pont Jacques-Cartier** @PontJCBridge 10h  
Jacques Cartier Bridge: no major road work this weekend on the bridge (June 1st & 2nd) [goo.gl/fb/XLMkn](http://goo.gl/fb/XLMkn)  
Details

# Overview

- An Infrastructure Discussion Network (IDN):

WHO?

IDEAS

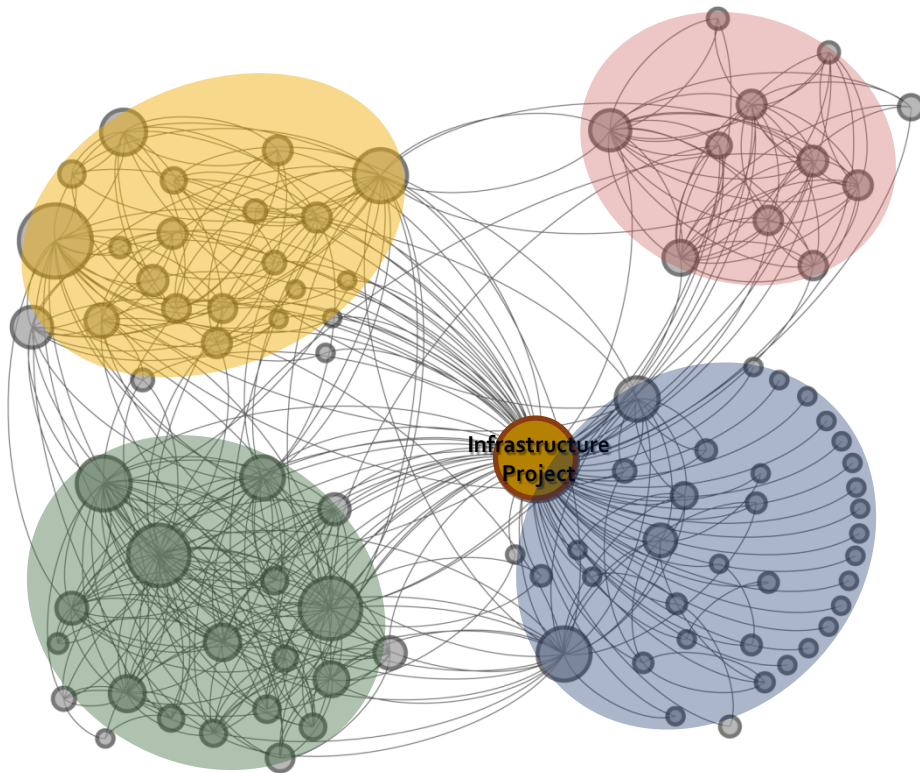
SAYS

WHAT?



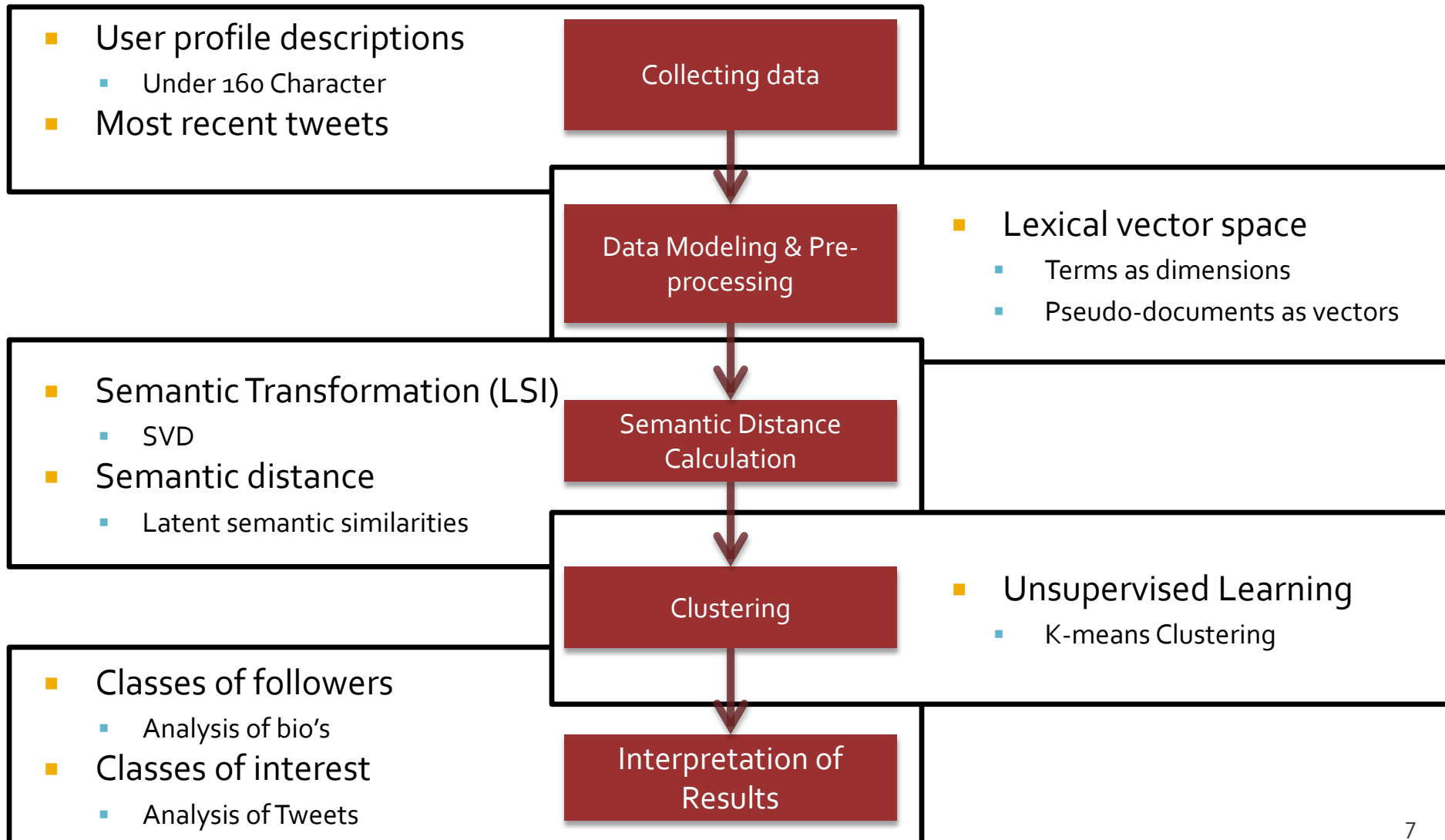
# Objective

## ■ Profiling Communities & Core Interests



- Finding groups with **Common Interests** (Stakeholders Typology)
- Detecting the **main Ideas discussed** (Stakeholders' Vested Interests)

# Methodology



# Case-study project

- **Eglinton Cross Town LRT (TO, ON)**
  - CAD8.4B project
  - Connects east and west of Toronto
  - Major parts underground (10 out of 19 km)
  - Construction began in 2011
  - Now under construction





# Case-study project



## Crosstown

@CrosstownTO FOLLOWS YOU

The Crosstown is a 19-kilometre light rail transit line that will run under and along Eglinton Avenue through the heart of Toronto. Under construction today.  
Toronto, Ontario <http://www.thecrosstown.ca>

177 TWEETS  
279 FOLLOWING  
584 FOLLOWERS

### Tweet to Crosstown

### Tweets

- Following
- Followers
- Favorites
- Lists
- Recent images

### Similar to Crosstown

-  **GoToMyPC** @GoToMyPC  
Promoted · Follow
-  **Fiona Crean** @TO\_Ombuds  
Follow
-  **Shelley Carroll** @shelleycarroll  
Follow

### Tweets

**Crosstown** @CrosstownTO 5 Oct  
Project Update: The #TTC will operate @CrosstownTO & @Metrolinx will build & maintain it when it opens to the public: [bit.ly/OawkfV](http://bit.ly/OawkfV)  
Expand

**Evergreen** @EvergreenCanada 1 Oct  
Join us tomorrow night for #InnovationTalks 4: Environment with @wrobertsfood, @cristoforo\_douc, @Simon\_DoUC and more!  
[ow.ly/e8bIO](http://ow.ly/e8bIO)  
Retweeted by Crosstown  
Expand

**Crosstown** @CrosstownTO 28 Sep  
Curious about the #LRT vehicle? Staff on hand this weekend @EvergreenCanada Brickworks [pic.twitter.com/636Bdj8Y](http://pic.twitter.com/636Bdj8Y)  
View photo

**Crosstown** @CrosstownTO 21 Sep  
The AFP model and what it means for #Toronto #transit riders: [bit.ly/Uyw5MG](http://bit.ly/Uyw5MG)  
Expand

**Crosstown** @CrosstownTO 21 Sep  
[@joedrew](#) The final design for the LRT vehicles hasn't been decided yet, but we will keep you posted! Thanks for your question.  
View conversation

**Crosstown** @CrosstownTO 13 Sep  
Let the digging begin! The first of two tunnelling contracts has been approved and is set to begin later this year: [bit.ly/ROMf3W](http://bit.ly/ROMf3W)  
Expand

© 2012 Twitter About Help Terms Privacy  
Blog Status Apps Resources Jobs  
Advertisers Businesses Media Developers

# Methodology

## Modeling followers through their Profile descriptions



**Kristyn Wong**  
 @kristynwongtam  
 Councillor, Ward 27 Toronto Centre-Rosedale.  
 Toronto, Canada <http://www.ward27news.ca>

BIO



**Josh Colle**  
 @JoshColle  
 City of Toronto Councillor, Ward 15 Eglinton Lawrence  
 Toronto <http://www.joshcolle.ca>

BIO



**Josh Matlow**  
 @JoshMatlow  
 Community Advocate & Toronto City Councillor for Ward 22-St. Paul's. Bringing a New Approach to City Hall.  
 Toronto <http://www.joshmatlow.ca>

BIO

### Crosstown

@CrosstownTO FOLLOWS YOU

The Crosstown is a 19-kilometre light rail transit line that will run under and along Eglinton Avenue through the heart of Toronto. Under construction today.

Toronto, Ontario <http://www.thecrosstown.ca>

Following

177 TWEETS

279 FOLLOWING

584 FOLLOWERS



**Karen Stintz**  
 @TTCchair  
 TTC #OneCity governance tweets by @karenstintz & @jpboutros. For #TTC operational info/answers, follow @ttchelps @ttchris @braite @ttcnotices.  
 Toronto <http://is.gd/59mXQU>

BIO



**Chris Dill**  
 @TTCchris  
 Chief Customer Officer of the Toronto Transit Commission. @ttchelps for queries, comments and complaints. @ttcnotices for service updates.  
 Toronto <http://www.ttc.ca>

BIO



**Brad TTC**  
 @bradTTC  
 Executive Director, Corporate Communications, Toronto Transit Commission. Follow @ttchelps for queries, kudos, complaints and @ttcnotices for service updates.  
 Toronto <http://www.ttc.ca>

BIO

# Methodology

- Each user as a pseudo-document (psd)
- A dictionary of all (say M) terms
- Each psd as a vector:

$$\begin{array}{c} \text{Dictionary terms} \\ (\text{size: } M) \end{array} \quad \begin{array}{c} \text{IDN Nodes (size: } N) \\ \left[ \begin{array}{ccc} [d(X_1)]_{M \times 1} & \dots & [d(X_N)]_{M \times 1} \end{array} \right]_{M \times N} \end{array}$$

- [Semantic] Similarity between vectors

$$\mathbf{A} = \mathbf{U} \times \mathbf{S} \times \mathbf{V}^T = \hat{\mathbf{A}}$$

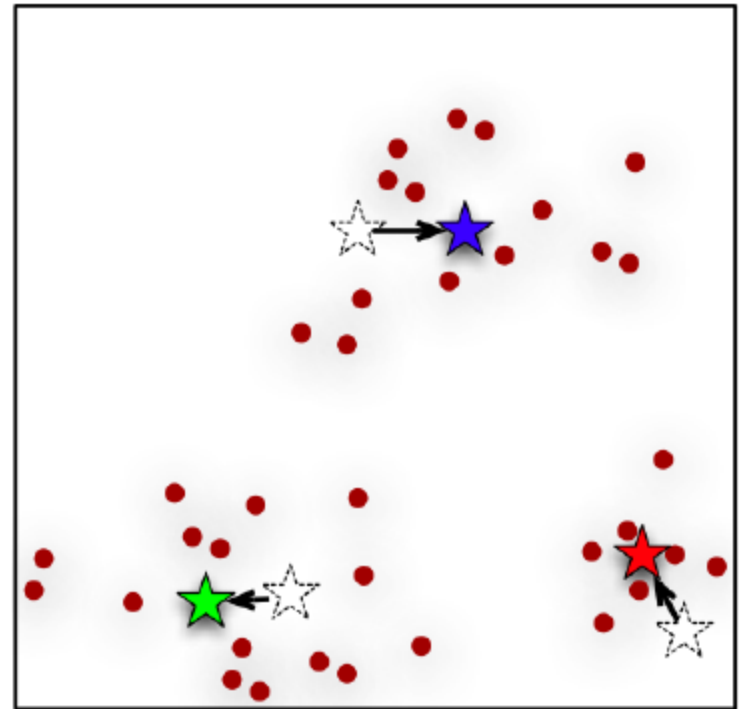
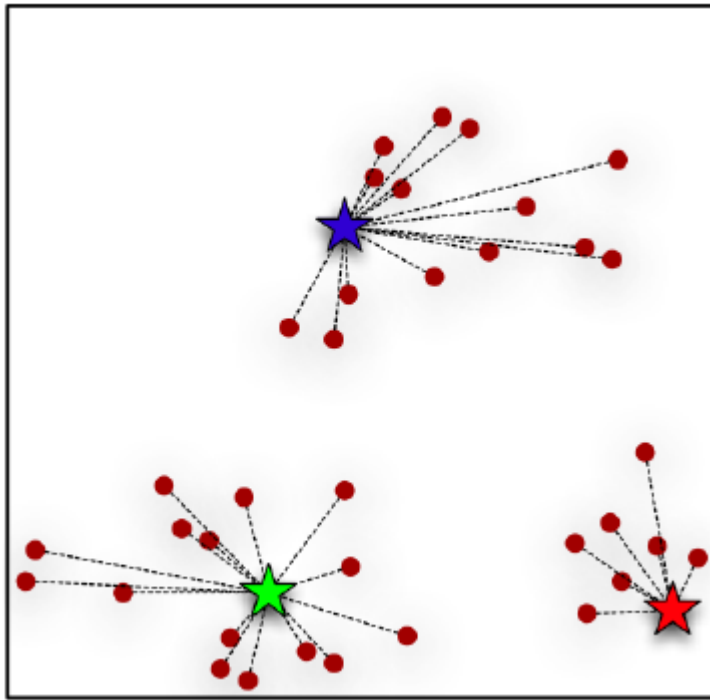
**D:**



*N followers (users) –  
N Pseudo- Documents*

# Methodology

- K-means Clustering



# Methodology

## ■ Selection of K:

k	Clusters	Clusters sizes	Average centroid distance for each cluster	Average inter-community distance	Performance (Davies-Bouldin Index)
2	C1	203	0.008	0.007	1.920
	C2	204	0.006		
3	C1	174	0.004	0.006	1.653
	C2	96	0.012		
	C3	137	0.005		
4	C1	53	0.012	0.005	1.425
	C2	78	0.006		
	C3	119	0.004		
	C4	157	0.004		
5	C1	110	0.004	0.008	1.730
	C2	72	0.005		
	C3	84	0.004		
	C4	97	0.003		
	C5	44	0.012		
6	C1	73	0.005	0.007	1.579
	C2	91	0.004		
	C3	91	0.003		
	C4	45	0.012		
	C5	7	0.000		
	C6	100	0.004		



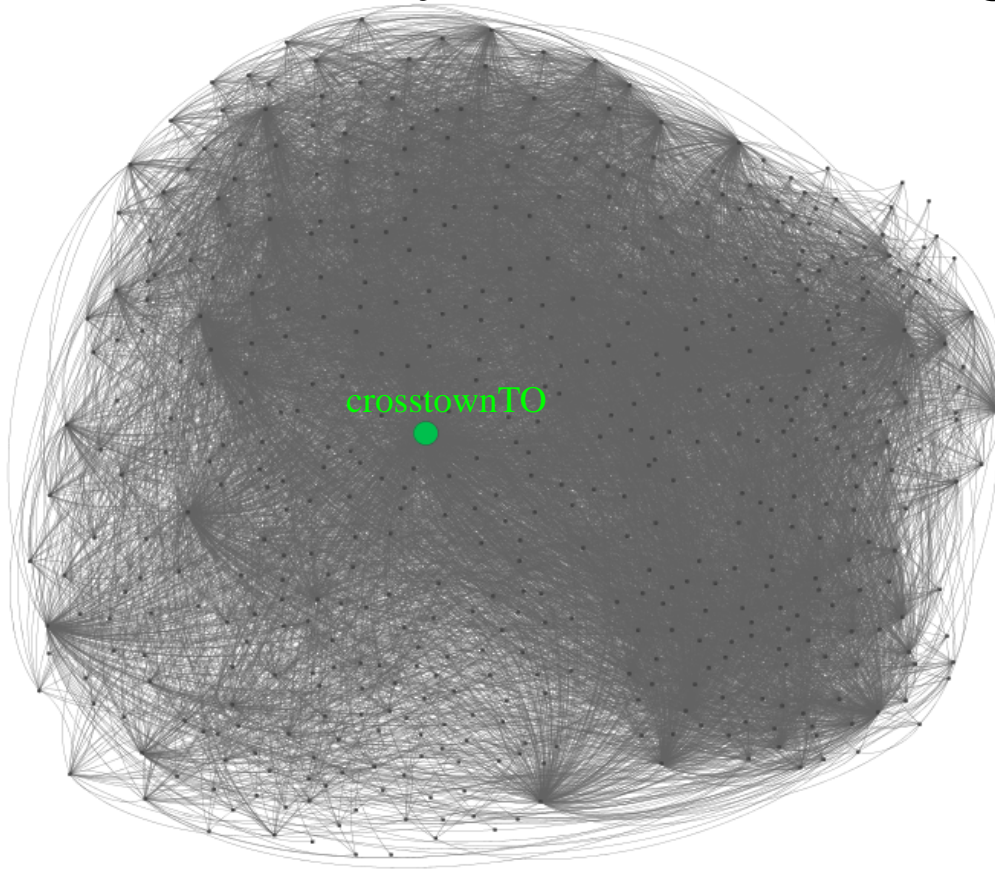
# Results (Main groups of followers)

	Cluster1	Cluster2	Cluster3	Cluster4
Size	157	119	78	53
	fan	transit	<b>city</b>	culture
	enthusiast	York	<b>news</b>	interested
	junkie	<b>Lrt</b>	civic	transit
	coffee	Mo`m	account	cities
	nerd	Work	world	housing
	<b>political</b>	<b>TTC</b>	awesome	enthusiast
	beer	<b>metrolinx</b>	business	<b>photographer</b>
	father	updates	<b>mayor</b>	design
	music	association	<b>Town_hall</b>	music
	views	Area	ward	<b>planner</b>
	member	village	senior	<b>transport</b>
	<b>cyclist</b>	good	twitter	issues
	<b>news</b>	enjoy	things	economics
	<b>writer</b>	<b>project</b>	active	interests
	avid	theatre	citizen	place
	addict	<b>construction</b>	resident	fan
	<b>designer</b>	leaside	affairs	avid
	technology	works	district	baseball
	grad	<b>transport</b>	job	human
	runner	<b>environment</b>	advisory	pop_culture
	<b>resident</b>	<b>chair</b>	<b>progressive</b>	policy

# Results (Main groups of followers) – Comparison

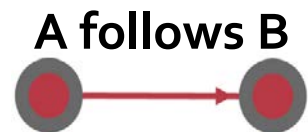
- Community-detection through SNA (bottom-up)

[Nik-Bakh & El-diraby 2012~13]



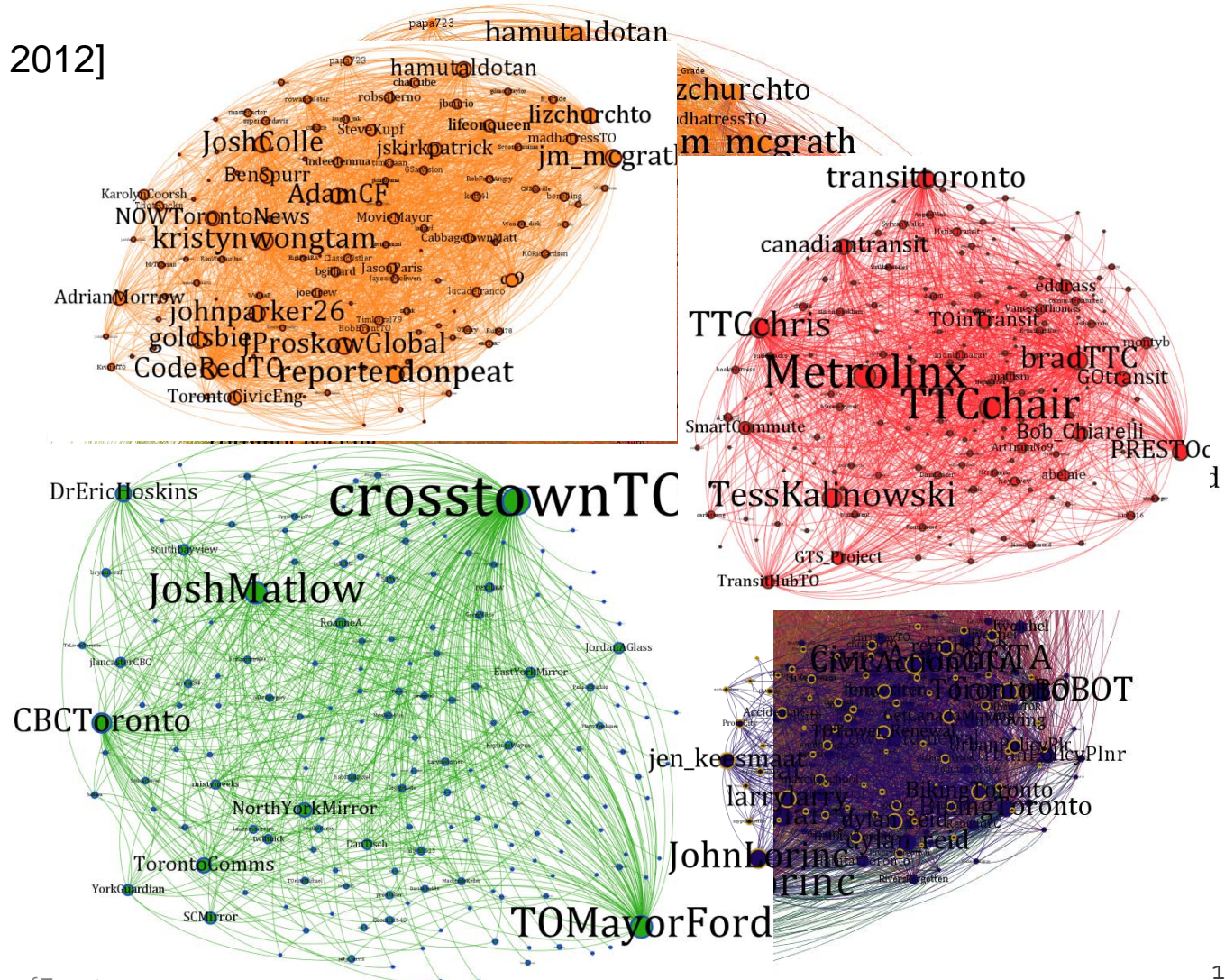
*Each Node shows:*  
**One project follower**

*Each Edge is the notion of a  
form of friendship/followership:*



# Results (Main groups of followers) – Comparison

[Nik-Bakh & El-diraby, 2012]



# Results (Main groups of followers) – Comparison

[Nik-Bakh & El-diraby, 2013]

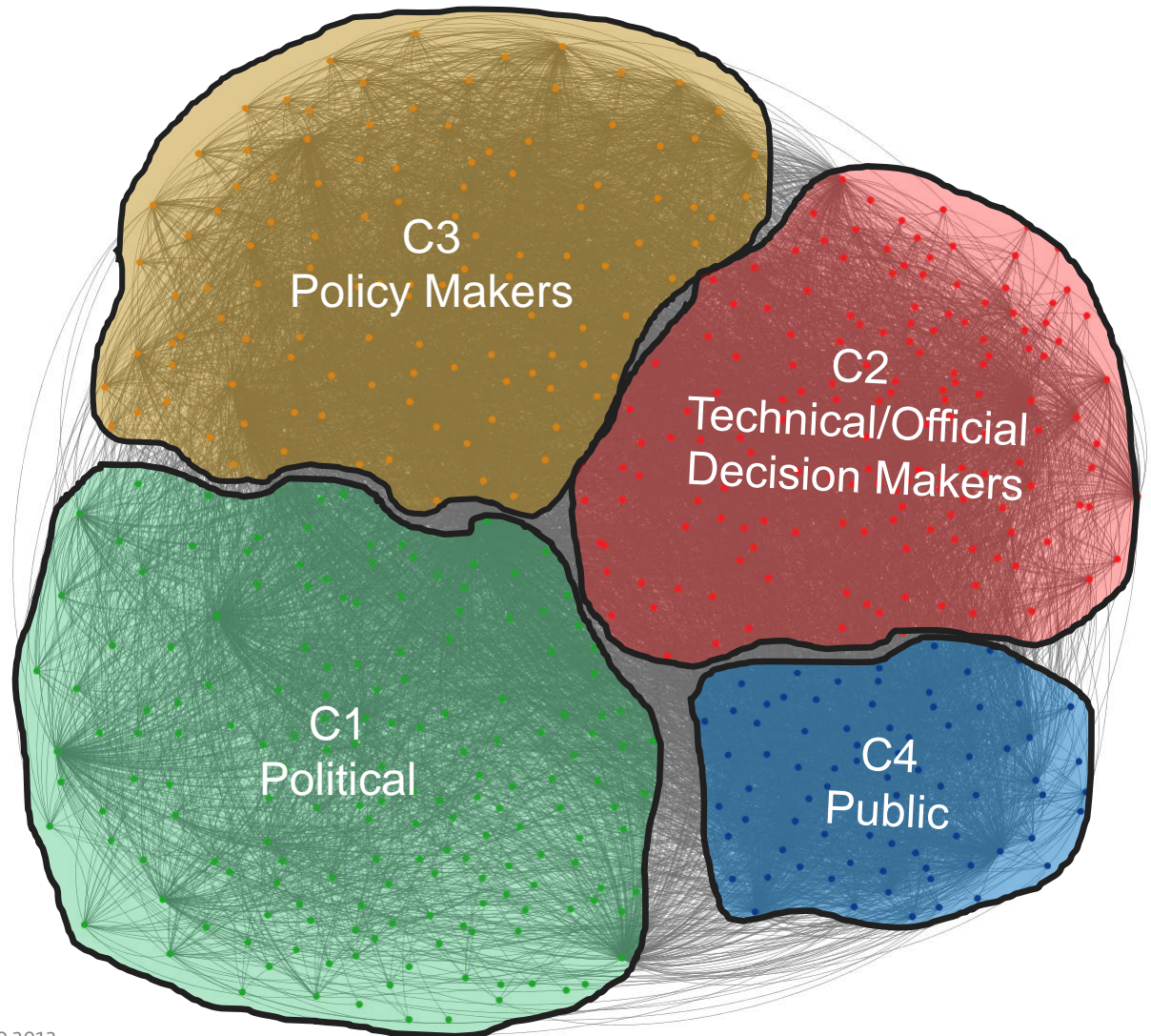
C1	C2	C3	C4
university	go	waste	<b>journalist</b>
<b>eglinton</b>	public	comms	cityhall
<b>davisville</b>	area	energy	freelance
indusry	construction	<b>strategist</b>	torontoist
town	hamilton	sustainable	<b>author</b>
updates	dedicated	commuter	civic
air	<b>#ttc</b>	economics	magazine
adventure	gta	ryersonu	write
business	<b>#ttchelps</b>	ceo	neighborhood
tv	<b>#ttcnotices</b>	<b>councillor</b>	<b>columnist</b>

**Analysis of Bio's in communities detected  
through SNA**



# Results (Main groups of followers)

Results of  
Top-down  
&  
Bottom-up  
Analyses  
Coincide!





# Results (Main ideas discussed)

## ■ Economy

- *XXX (representing monetary values), Dollars, Dollars\_Risk, Investing, Millions, Oaarchitects, Construction\_Safety, Mega\_Contract*

## ■ Technical features

- *Construction, Safety, Contract, Consultant, Contractor, Procurement, Procurement\_Process, LRT\_Project, and Tunnel\_Boring*

## ■ Community related issues, local/regional aspects

- *Neighbourhood, Community, Neighbourhood\_Revitalization, Eglinton, Yonge\_Bus, Finch, and Finch\_LRT, Realestate*

## ■ Political issues

- *Councilor, City\_Staff, Provincial\_Government, Harper*

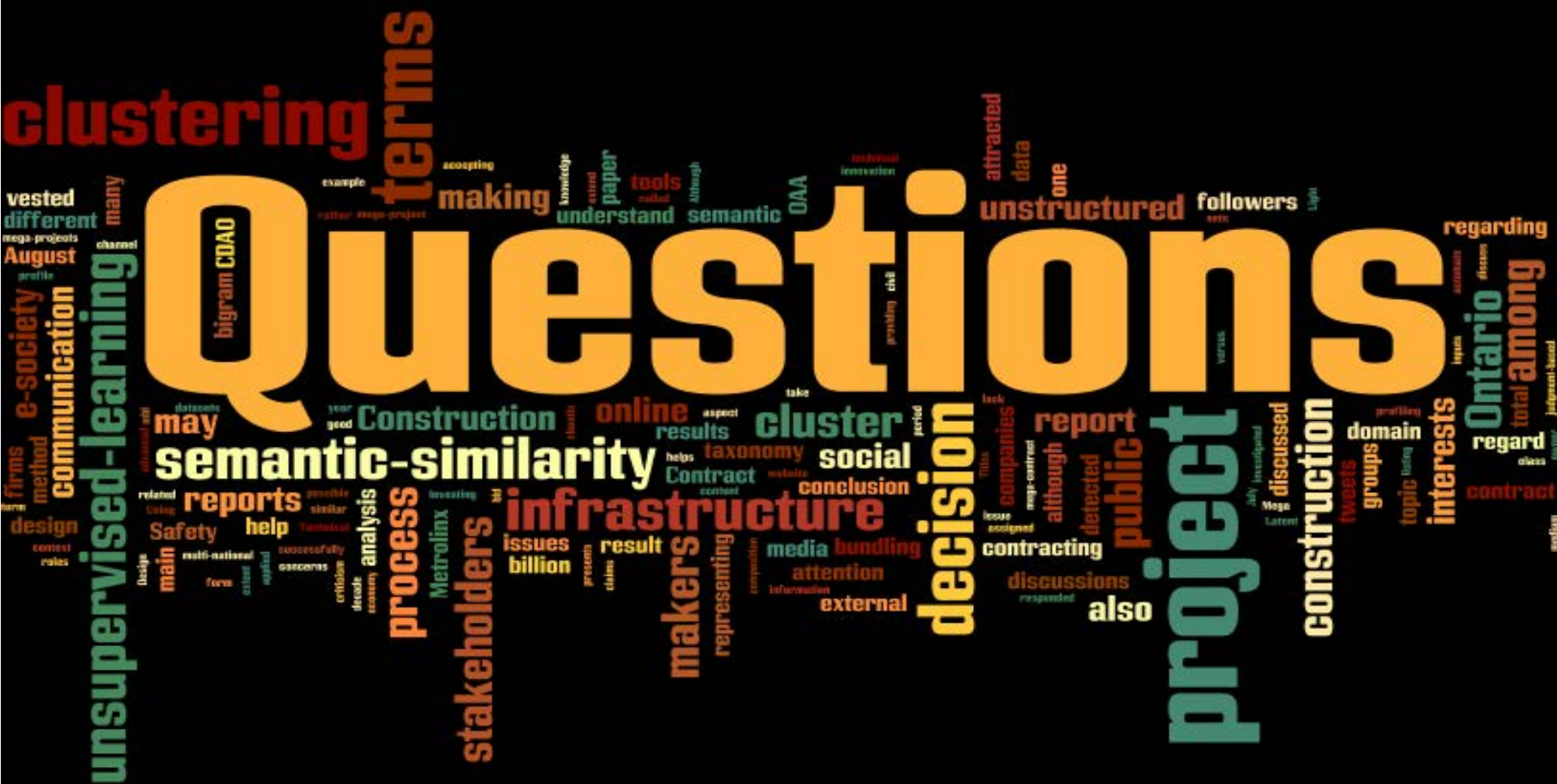
## ■ A small cluster with no specific theme

- *Manual screening: News and updates about the project, its schedule and improvement, and public meetings*

# Conclusion

- Stakeholders typology
  - Bottom-up (Through SNA)
  - Top-down (Through semantic clustering)
- Topic detection
  - Subjectivity of interpretation
  - Existing anomalies
- Future work
  - Semantic distance through an external knowledge-base/taxonomy
  - Higher levels of automation(?)

# THANKS – Questions?

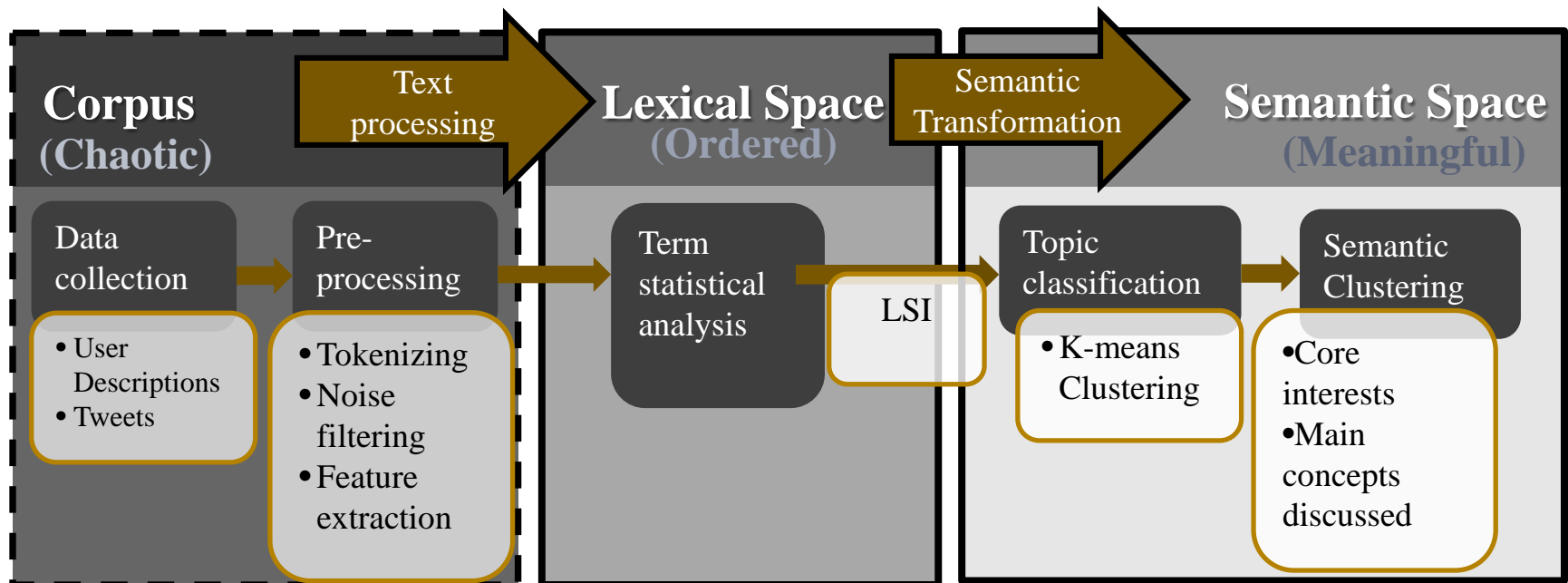


# Literature

Author	Data-set Domain	Goal	Features	Feature values	Distance measure
Kennedy et al. (2007)	Flickr	Detection of events and places from clusters of social annotation	Terms	TF-IDF	Geographic distance
Farahat & Kamel (2010)	Various benchmark data-sets	Document clustering	Terms with stemming	TF-IDF	Variants of corpus-based semantic similarity
Antai et al. (2011)	Pdf text files	Subject classification and topic detection	Terms		LSI semantic distance
Rangrej et al. (2011)	Twitter	Clustering tweets	Terms	TF-IDF	Cosine similarity and Jaccard coefficient
Xu & Orad, (2012)	Twitter	Topical clustering and detecting representativeness	Wikipedia terms	TF-IDF	Semantic distance through an external knowledgebase (Wikipedia)
Muntean et al. (2012)	Twitter	Semantics behind hashtags	Terms	TF-IDF	Jaccard distance measure
Kireyev et al. (2009)	Twitter	Crisis and disaster detection	Domain and event-specific terms	Bag of words	Cosine similarity
Avanija & Ramar (2013)	Usenet group dataset	Web documents clustering	Terms	TF-IDF	Semantic distance over an external ontology
Suganya & Srinivasan (2013)	Web search queries	To infer query's goal from feedback sessions	Pseudo Terms	TF-IDF	Semantic distance over an external knowledge-base (WordNet)

# Methodology

## ■ Analysis





# Methodology

## ■ K-means Clustering

