Dynamic Query Expansion for Event Detection on Twitter Data

Liang Zhao
Twitter Heterogeneous Network Modeling

- Term-to-Term Co-Occurrence Graph
- Spatial Neighborhood Graph
- User Follower Graph
- Tweet Retweet & Reply Graph

Nodes:
- Country
- State
- City
- Term
- Hashtag
- Owner or User Mention
- Link

Edges:
- Location relationship

Graphs:
- Spatial Neighborhood Graph
- User Follower Graph
See you on Saturday at 15:00 in the #MegaMarcha

"#MexicoExigeDemocracia"http://t.co/MdG5T3z0 Twitterers help me with a RT?. See you on Saturday at 15:00 in the #MegaMarcha.

@epigmenioibarra good morning!!! excuse from where the #MegaMarcha ??

Veracruz, Jalapa, Mérida, Tepotzotlan add to the #MegaMarcha vs imposición. Tambien Los Ángeles. Who else says

Ready to march, tweeting or filming tomorrow #MegaMarcha vs imposición. Hopefully many say #Vamon

"#MexicoExigeDemocracia""http://t.co/MdG5T3z0 Twitterers help me with a RT?. See you on Saturday at 15:00
Content Filtering
Content Filtering via Dynamic Query Expansion

- **Traditional Solutions:**
  1. Keyword Match based on a predefined dictionary
  2. Train a classifier (e.g., SVM) [WWW-2010]
  3. Rules or decision tree based

- **Challenges for Evaluating Twitter Data**
  - Heavily informal and ungrammatical
  - Huge volume (more than 1 million per day)
  - Dynamic
  - Noisy and Unbalanced (e.g., less than 5% related to civil unrest)
  - Rich Social Network Structure
Content Filtering via Dynamic Query Expansion

- Traditional Solutions:
  - 1. Keyword Match based on a predefined dictionary

We need:
Unsupervised and adaptive solution

Our Approach:
Dynamic Query Expansion

- Dynamic
- Noisy and Unbalanced (e.g., less than 5% related to civil unrest)
- Rich Social Network Structure
Content Filtering via Dynamic Query Expansion

A case on July 14, 2012 in Mexico

GSR event description

“The movement #yosoy132 march against "imposition" of the candidate in election.”
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion

Initialization: Seed Query

- march
- protest
Content Filtering via Dynamic Query Expansion

Initialization: Seed Query

- **march**
- **protest**

List the marches anti PEÑA NIETO by city, Locate your MARCHA #MegaMarcha http://t.co/hwGTl3FH

Ready to march, tweeting or filming tomorrow #MegaMarcha vs imposición. Hopefully many say #Vamon

RT @politicosmex: Mega Marcha in 32 cities in México against Peña Nieto. http://t.co/fQ95w5Yo #Megamarcha #YoSoy132

People march to protest Reform by what they called the imposition of @epn

"RT @TavoDeLosSantos: @anonctrl List the marches anti PEÑA NIETO BY CITY, LOCATE YOUR MARCHA #MegaMarcha http://t.co/hwGTl3FH"

List of the cities for marcha on July 14 #MexicoExigeDemocracia http://t.co/RpnNWZNh #YoSoy132 via @RedPuenteSur
Content Filtering via Dynamic Query Expansion

Initialization: Seed Query

- March
- Protest

List the marches anti PEÑA NIETO by city,
Locate your MARCHA #MegaMarcha http://t.co/hwGTl3FH

"RT @TavoDeLosSantos: @anonctrl List the marches anti PEÑA NIETO BY CITY,
LOCATE YOUR MARCHA #MegaMarcha http://t.co/hwGTl3FH"

Ready to march, tweeting or filming tomorrow #MegaMarcha
vs imposición. Hopefully many say #Yamon

RT @politicosmex: Mega Marcha in 32 cities in México against Peña Nieto http://t.co/fQ95w5Yo
#Megamarcha #YoSoy132

List of the cities for marcha on July 14
#MexicoExigeDemocracia http://t.co/RpnNWZNh
#YoSoy132 via @RedPuenteSur

People march to protest Reform by what they called the imposition of @epn
Content Filtering via Dynamic Query Expansion

Initialization: Seed Query

- **Initialization: Seed Query**
  - **march**
  - **protest**

- List the marches anti PEÑA NIETO by city, Locate your MARCHA #MegaMarcha http://t.co/hwGTl3FH

- Ready to march, tweeting or filming tomorrow #MegaMarcha vs imposición. Hopefully many say #Yamon

- RT @politicosmex: Mega Marcha in 32 cities in México against Peña Nieto http://t.co/fQ95w5Yo #Megamarcha #YoSoy132

- People march to protest Reform by what they called the imposition of @epn

- http://t.co/hwGTl3FH

- "RT @TavoDeLosSantos: @anonctrl List the marches anti PEÑA NIETO by CITY, LOCATE YOUR MARCHA #MegaMarcha http://t.co/hwGTl3FH"

- List of the cities for marcha on July 14 #MexicoExigeDemocracia http://t.co/RpnNWZNh #YoSoy132 via @RedPuenteSur

- epigmenio ibarra

- #MegaMarcha

- #MexicoExigeDemocracia

- imposición
Content Filtering via Dynamic Query Expansion

Initialization: Seed Query

- march
- protest

List the marches anti PEÑA NIETO by city, Locate your MARCHA #MegaMarcha http://t.co/hwGTl3FH

Ready to march, tweeting or filming tomorrow #MegaMarcha vs imposición. Hopefully many say #Yamon

RT @politicosmex: Mega Marcha in 32 cities in México against Peña Nieto http://t.co/fQ95w5Yo #Megamarcha #YoSoy132

People march to protest Reform by what they called the imposition of @epn

"RT @TavoDeLosSantos: @anonctrl List the marches anti PEÑA NIETO BY CITY, LOCATE YOUR MARCHA #MegaMarcha http://t.co/hwGTl3FH"

List of the cities for marcha on July 14 #MexicoExigeDemocracia http://t.co/RpnNWZNh #YoSoy132 via @RedPuenteSur

New key nodes of 1st Iteration

- http://t.co/hwGTl3FH
- imposición
- #MegaMarcha
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion

Key nodes of 1st Iteration

- epigmenio ibarra
- imposición
- # MegaMarcha
- http://t.co/MdG5T3z0
- #MexicoExigeDemocracia
Content Filtering via Dynamic Query Expansion

Key nodes of 1st Iteration

- **epigmenio ibarra**
- **imposición**
- **# MegaMarcha**
- **#MexicoExigeDemocracia**

RT @BiobeatsEidos: Good article @epigmenioibarra #MexicoExigeDemocracia"" http://t.co/LhSZENxs

Neither legislators swallow proposals anticorruption of PEÑA NIETO http://t.co/n1CychU #AMLO #Fraude2012 #MegaMarcha

Where will leave the #MegaMarcha Today? via @BlogdelPueblo #ExigimosDemocracia #YoSoy132 http://t.co/RkXh4ZoD #Unidad132

Mexico is no longer for cowards #megamarcha 14julio Plazas Públicas Fraude Electoral AMLO

Agreement against the imposition and March against Peña Nieta #YoSoy132.

#MegaMarcha against Peña Nieto; and posters http://t.co/0GqEVC7h #Veracruz #Edomex #Tamaulipas #Matamoros #Reynosa #Xalapa #15M

#MexicoExigeDemocracia"" http://t.co/MdG5T3z0 help me tweeters with a RT?. See you on Saturday at 3pm
Content Filtering via Dynamic Query Expansion

Key nodes of 1st Iteration

- epigmenio ibarra
- imposición
- # MegaMarcha
- # MexicoExigeDemocracia
- http://t.co/MdG5T3z0

- Neither legislators swallow proposals anticorruption of PEÑA NIETO http://t.co/n1CychUm # AMLO # Fraude2012 # MegaMarcha

- Where will leave the # MegaMarcha Today? via @BlogdelPueblo # ExigimosDemocracia # YoSoy132 http://t.co/RkXh4ZoD # Unidad132

- Mexico is no longer for cowards # megamarcha 14julio Plazas Publicas Fraude Electoral AMLO

- Agreement against the imposition and March against Peña Nieta # YoSoy132.

New key nodes of 2nd Iteration

- # yosoy132
- # yosoyMX
- # MexicoExigeDemocracia
- http://t.co/MdG5T3z0

- Peña Nieto
- 3pm
- http://t.co/MdG5T3z0

- help me tweeters with a RT?. See you on Saturday at 3pm
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion

Against the president
Event location

The ongoing movement
Protest organization
Event description
Key hyperlink

Key player
Event time
Event location
Content Filtering via Dynamic Query Expansion
Content Filtering via Dynamic Query Expansion

Converged!
"The movement #yosoy132 march against "imposition" of the candidate in election."
Content Filtering via Dynamic Query Expansion

**INPUT:** Seeds

**Tweet Space**

Sub-Space Search

**Ranking**

<table>
<thead>
<tr>
<th>tweet</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>......</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Reverse Ranking**

<table>
<thead>
<tr>
<th>tweet</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>......</td>
<td>0.1</td>
</tr>
</tbody>
</table>

---

Exchange Tweets, Update Weights
Content Filtering via Dynamic Query Expansion

\[ X_k = \phi_2 \left( (A \cdot \sqrt{S})M(A \cdot \sqrt{S})^T \cdot X_{k-1} + D \cdot \right) \]

**INPUT:** Seeds

**Sub-Space Search**

**Reverse Ranking**

<table>
<thead>
<tr>
<th>tweet</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>……….</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Exchange Tweets, Update Weights**
Content Filtering via Dynamic Query Expansion

\[
X_k = \phi_2 \left( (A \cdot \sqrt{S}M \cdot (A \cdot \sqrt{S})^T \cdot X_{k-1} + D \right)
\]

INPUT: Seeds

Sub-Space Search

Reverse Ranking

<table>
<thead>
<tr>
<th>tweet</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>......</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Exchange Tweets, Update Weights
## Demonstrations

<table>
<thead>
<tr>
<th>Detection Date</th>
<th>Key words detected</th>
<th>GSR Description</th>
<th>Happening Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun 29</td>
<td>eleccion, transparent, #yosoy132, yosoy13, march, movimient</td>
<td>With candles, YoSoy132 BUAP protest against the IFE-Puebla for transparent elections</td>
<td>Jun 30</td>
</tr>
<tr>
<td>Jul 01</td>
<td>#yosoy132, #granmarcha132, patrull Comaneren, PRI (the party who won election)</td>
<td>Youth movement # YoSoy132 staged a sit-in outside the local board of Federal Electoral Institute</td>
<td>Jul 01</td>
</tr>
<tr>
<td>Jul 04</td>
<td>#epnnunceramipresidente, #yosoy132, movimient, progres, contig, march, fraud</td>
<td>The student movement # YoSoy132 protesting alleged fraud in the elections</td>
<td>Jul 03</td>
</tr>
<tr>
<td>Jul 07</td>
<td>#megamarcha, #exigimosdemocracia, eugenio derbez, limpi, eleccion, march</td>
<td>&quot;Protesters unite to call for mega march&quot;  &quot;YoSoy132 go and concentrate on the Esplanade of Heroes&quot; Protesters unite to call for mega march against virtual presidential election</td>
<td>Jul 07</td>
</tr>
<tr>
<td>Jul 08</td>
<td>#megamarcha, #megamarcha, eugenio derbez, against election, march</td>
<td>A march was in protest of the imposition of the PRI candidate.</td>
<td>Jul 14</td>
</tr>
<tr>
<td>Jul 13</td>
<td>imposicion, #megamarcha, 15hrs principal, march</td>
<td># virtual Megamarch against the winner of the presidential election, Enrique Peña Nieto, left the Angel de Independencia to el Zocalo of Mexico City.</td>
<td>Jul 14</td>
</tr>
<tr>
<td>Jul 14</td>
<td>#megamarcha, #megamarcha, 14juli public, zocal, angel, march</td>
<td>Protesting for alleged fraud in the election of July 1</td>
<td>Jul 19</td>
</tr>
<tr>
<td>Jul 22</td>
<td>#megamarcha, #yosoy132, contingent Zocal, march, imposicion</td>
<td>a mega march against the alleged imposition of the PRI. YoSoy132 march arrives at El Zocalo and goes to the Monument to the Revolution</td>
<td>Jul 22</td>
</tr>
<tr>
<td>Jul 27</td>
<td>#ocupatelevisa, #yosoy132, televis, chapultepec, instal, march</td>
<td>Students symbolically take over facilities of Hidalgo Radio and TV; And fence outside Televisa Chapultepec in Mexico City</td>
<td>Jul 27</td>
</tr>
</tbody>
</table>
### Demonstrations

<table>
<thead>
<tr>
<th>Detection Date</th>
<th>Key words detected</th>
<th>GSR Description</th>
<th>Happening Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun 29</td>
<td>eleccion, transparent, #yosoy132, yosoy13, march, movimient</td>
<td>With candles, YoSoy132 BUAP protest against the IFE-Puebla for transparent elections</td>
<td>Jun 30</td>
</tr>
<tr>
<td>Jul 01</td>
<td>#yosoy132, #granmarcha132, patrull Comaner, PRI (the party who won elections)</td>
<td>YoSoy132 staged a sit-in outside the Federal Electoral Institute</td>
<td>Jul 01</td>
</tr>
<tr>
<td>Jul 04</td>
<td>#epnnuncasermipresidente, #yosoy132, movimient, progres, contig, march</td>
<td>YoSoy132 protesting alleged fraud in the elections</td>
<td>Jul 03</td>
</tr>
<tr>
<td>Jul 07</td>
<td>#megamarcha, Eugenio Derbez, limpi, eleccion, march</td>
<td>Protesters unite to call for mega march against virtual presidential election. Protests unite to call for mega march against the winner of the presidential election.</td>
<td>Jul 07</td>
</tr>
<tr>
<td>Jul 08</td>
<td>#megamarcha, Eugenio Derbez, against, march</td>
<td>Protests unite to call for mega march against the imposition of the PRI</td>
<td>Jul 14</td>
</tr>
<tr>
<td>Jul 13</td>
<td>imposicion, #megamarcha, 13juli principal, march</td>
<td>Protests unite to call for mega march against the imposition of the PRI. YoSoy132 March arrives at El Zocalo and goes to the Monument to the Revolution.</td>
<td>Jul 14</td>
</tr>
<tr>
<td>Jul 14</td>
<td>#megamarcha, 14juli public, zocal, angel, march</td>
<td>Protesting for alleged fraud in the election of July 1</td>
<td>Jul 19</td>
</tr>
<tr>
<td>Jul 22</td>
<td>#megamarcha, #yosoy132, contingent Zocal, march, imposicion</td>
<td>YoSoy132 march arrives at El Zocalo and goes to the Monument to the Revolution. Students symbolically take over facilities of Hidalgo Radio and TV; And fence outside Televisa Chapultepec in Mexico City.</td>
<td>Jul 22</td>
</tr>
<tr>
<td>Jul 27</td>
<td>#ocupatelevisa, #yosoy132, televis, chapultepec, instal, march</td>
<td>Students symbolically take over facilities of Hidalgo Radio and TV; And fence outside Televisa Chapultepec in Mexico City.</td>
<td>Jul 27</td>
</tr>
</tbody>
</table>
Step 2: Tweet Network

"RT @TavoDeLosSantos: LISTA DE MARCHAS ANTI PEÑA NIETO POR CIUDAD, UBICA TU MARCHA #MegaMarcha http://t.co/hwGTl3FH"

I'm broadcasting #MegaMarcha 14Julio plazas publicas – http://t.co/d3vqDRsw (3:59p.m.)`

"Algo que es verdad: el PRI jamás hubiera "ganado" sin trampas 14julio Plazas Públicas"

"LISTA DE MARCHAS ANTI PEÑA NIETO POR CIUDAD, UBICA TU MARCHA #MegaMarcha"

Pido a tuiteros convoquen, en distintos idiomas, a 2a. #MegaMarcha vs imposicion.

"Segunda #MegaMarcha #MexicoExigeDemocracia - nickops http://t.co/oDeDo53i #yosoym132"

Me informan que #YoSoy132 esta estableciendo un plan vs la "imposición" de @EPN"
Step 3: Tweet Network Clustering
Step 3: Tweet Network Clustering
Step 3: Tweet Network Clustering
Step 3: Tweet Network Clustering
Step 3: Tweet Network Clustering

Event:2034
Event:2031
Event:2035
Event:2036
Event:2027
No warning
False Negative
Event:2028
Event:2029
No event
False Positive
Event:2033
Event:2032
Event:2030
## Experimental Results

<table>
<thead>
<tr>
<th></th>
<th>GSR alerts</th>
<th>Warnings</th>
<th>Matches</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Countries</td>
<td>1544</td>
<td>1038</td>
<td>817</td>
<td>0.787091</td>
<td>0.529145</td>
</tr>
<tr>
<td>brazil</td>
<td>117</td>
<td>80</td>
<td>42</td>
<td>0.525</td>
<td>0.358974</td>
</tr>
<tr>
<td>colombia</td>
<td>100</td>
<td>113</td>
<td>81</td>
<td>0.716814</td>
<td>0.81</td>
</tr>
<tr>
<td>el salvador</td>
<td>58</td>
<td>50</td>
<td>31</td>
<td>0.62</td>
<td>0.534483</td>
</tr>
<tr>
<td>uruguay</td>
<td>37</td>
<td>74</td>
<td>25</td>
<td>0.337838</td>
<td>0.675676</td>
</tr>
<tr>
<td>chile</td>
<td>113</td>
<td>90</td>
<td>70</td>
<td>0.777778</td>
<td>0.619469</td>
</tr>
<tr>
<td>paraguay</td>
<td>177</td>
<td>92</td>
<td>79</td>
<td>0.858696</td>
<td>0.446328</td>
</tr>
<tr>
<td>argentina</td>
<td>162</td>
<td>102</td>
<td>93</td>
<td>0.911765</td>
<td>0.574074</td>
</tr>
<tr>
<td>venezuela</td>
<td>255</td>
<td>174</td>
<td>142</td>
<td>0.816092</td>
<td>0.556863</td>
</tr>
<tr>
<td>ecuador</td>
<td>50</td>
<td>13</td>
<td>10</td>
<td>0.769231</td>
<td>0.2</td>
</tr>
<tr>
<td>mexico</td>
<td>475</td>
<td>250</td>
<td>244</td>
<td>0.976</td>
<td>0.513684</td>
</tr>
</tbody>
</table>
Experimental Results

![Graph showing precision-recall curves for various countries.](image)
Appendix: Case Studies of DQE

- DQE for Latin America
- DQE for MENA region
Examples of DQE

Liang Zhao
[Warning ID: 22666], Mexico City

Students nationwide strike against 43 lives missing in normal school of Ayotzinapa.

Predicted date: Nov 1
Quality score: 3.55
En la CIDH se mantuvo un minuto de silencio por los seis estudiantes de Ayotzinapa que murieron el pasado 26 de septiembre.

(Translated in English) The IACHR one minute silence was held for six students Ayotzinapa who died last September 26.
Protestan alumnos de la Normal Superior frente a PGR

Predicted date: Nov 4
Example: Mar 16, 2013
Warning Id: 2009, Event Id: 5544

Initial query: protest, march, huelga, strike, greve.

GSR headline(Event Id: 5544):
SP tem maisum protesto contra pastor que preside Comissão de Direitos Humanos
DQE for MENA

Liang Zhao, Chang-Tien Lu
NVC
## Seed query for MENA

We used 6 Arabian words

<table>
<thead>
<tr>
<th>#</th>
<th>Arabian words</th>
<th>English meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>جَمَع</td>
<td>gathering</td>
</tr>
<tr>
<td>2</td>
<td>مُظاهَرَة</td>
<td>demonstration</td>
</tr>
<tr>
<td>3</td>
<td>اِنْتِفاضَة</td>
<td>insurrection</td>
</tr>
<tr>
<td>4</td>
<td>شَغَب</td>
<td>riot</td>
</tr>
<tr>
<td>5</td>
<td>ثَوْرَة</td>
<td>revolution</td>
</tr>
<tr>
<td>6</td>
<td>اِحْتِجاج</td>
<td>protest</td>
</tr>
</tbody>
</table>

Plus 3 English words:

Unrest, protest, strike
<table>
<thead>
<tr>
<th>Detection Date</th>
<th>Keywords detected</th>
<th>News description</th>
</tr>
</thead>
<tbody>
<tr>
<td>05-18</td>
<td>Revolution, demonstration, #Sisi, swear, presidency, thaer</td>
<td>Youth January revolution swear allegiance to al-Sisi.</td>
</tr>
<tr>
<td>05-19</td>
<td>Revolution, a student, shows, collector, collectors, rule</td>
<td>Students at Cairo University, stoking the flames in the Office of Administrative Security</td>
</tr>
<tr>
<td>05-21</td>
<td>Revolution, student, calamity, a summer, the timing, the memory</td>
<td>Security forces attack the engineering students of Alexandria and launches them live bullets and cartridge</td>
</tr>
<tr>
<td>05-23</td>
<td>Revolution, demonstration, Libya, Libya, march, Tripoli</td>
<td>Hundreds demonstrate in Tripoli’s Martyrs Square in “Friday of Dignity” protests</td>
</tr>
<tr>
<td>05-24</td>
<td>Revolution, demonstration, hungry, #Sisi, brother, Jabbar</td>
<td>aisl compatriots chasing a demonstration of the Brotherhood after chanting slogans against Sisi</td>
</tr>
<tr>
<td>05-26</td>
<td>Revolution, young, demonstration, election, Thaer, continuous</td>
<td>Ulna boycotters of the elections is haram call them citizens who are a burden on society and do not deserve to live in Egypt</td>
</tr>
<tr>
<td>05-31</td>
<td>Revolution, Revolution, sisi, Thaer, Jan, brother</td>
<td>Egyptians protest after Sisi election victory</td>
</tr>
</tbody>
</table>
Students at Cairo University, stoking the flames in the Office of Administrative Security.