Identifying Terrorist Affiliations Through Social Network Analysis Using Data Mining Techniques

Govand A. Ali
Valparaiso University, govand.ali@valpo.edu

Follow this and additional works at: http://scholar.valpo.edu/gas

Recommended Citation
http://scholar.valpo.edu/gas/2

This Poster Presentation is brought to you for free and open access by the Graduate School at ValpoScholar. It has been accepted for inclusion in Graduate Academic Symposium by an authorized administrator of ValpoScholar. For more information, please contact a ValpoScholar staff member at scholar@valpo.edu.
Identifying Terrorist Affiliations through Social Network Analysis Using Data Mining Techniques

MASTER’S THESIS By GOVAND A. ALI, Advisor: Sonja H. Streuber

INTRODUCTION
This study employs data mining tools to mine Twitter for terrorist ‘organizing’ vocabulary and to pinpoint, through the analysis of (admittedly sparse) tweet metadata, the most likely geographical location and connected identities behind the user accounts from which organizing or post-event information is disseminated.

Tools:
- Twitter
- R and R Studio
- twitteR package in R Studio
- NodeXL
- Gephi for visualization

SEARCH VOCABULARY
<table>
<thead>
<tr>
<th>English</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islamic State</td>
<td>الخلافة</td>
</tr>
<tr>
<td>Dabiq</td>
<td>البغدادي</td>
</tr>
<tr>
<td>Jihadi</td>
<td>المجاهدين</td>
</tr>
<tr>
<td>Ummah</td>
<td>خليفة المسلمين</td>
</tr>
<tr>
<td>Kuffar</td>
<td>باليكية وتعدى</td>
</tr>
<tr>
<td>Wilayat</td>
<td>داعي</td>
</tr>
<tr>
<td>Khawarij</td>
<td>الدولة الإسلامية ولاية</td>
</tr>
</tbody>
</table>

Search run: 02/20/2016-02/21/2016 (24 hrs)
ResultsReturned:
- English – 6,451
- Arabic – 4,353
After Cleaning:
- English – 6,321
- Arabic – 3,300

SENTIMENT ANALYSIS

K-means produces clusters that allow us to predict user accounts with certain vocabulary use characteristics.

STEP 1: Finding k (= number of clusters)

The optimal number of clusters is determined using the Elbow method, which runs the K-means algorithm with various numbers of clusters and displays the outputs. The location of the elbow in the graphic shows the optimal number of clusters.

Results:
- English – 5 clusters
- Arabic – 3 clusters

STEP 2: Determining linguistic groupings based on k

The Arabic cluster confirms the highest confidence in its ISIS affiliation, while the English cluster shows the highest confidence for Al Qaeda affiliation.

SOCIAL NETWORK VISUALIZATION

We selected three relevant users from Arabic cluster 3 to show affiliations. In this model, @alarhabi42 is shown in the lower right-hand corner, @nbnnm242 is shown in the upper right-hand corner, and @drdhchfcty appears in the upper left-hand corner. As the model shows, of the three users @drdhchfcty has the greatest number of followers, with the other two accounts exhibiting similar numbers. The interesting part of this graphic is that these accounts exist in the space between the user accounts, especially in the center of the triangle: These are followers subscribing to all three user accounts. Further snapshot inspection of the accounts in the center suggests that these users post with the identities of mujaheddin.

REFERENCES