Detection, Analysis, and Disruption of Online Extremist Communities

Matthew Benigni

mbenigni@cs.cmu.edu

Carnegie Mellon

Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

Agenda

- Admin
- Background and Problem Statement
- Detection
  - Overview
  - Hands-On: R
- Analysis
  - Overview
  - Hands-On: ORA
- Future Research
  - Overview
  - Discussion

Practical Exercise Materials

- Download the data and R code via

http://www.casos.cs.cmu.edu/projects/covert groups/OEC_Detection_Tutorial.zip

- ORA-PRO 64-bit Windows trial version:
  - Expires:7/15/2016

https://www.dropbox.com/s/axkpqkc57cxz9q5/ORA-NetScenes-l-iw-64.exe?dl=0

- Your trial license key is:64-JGAJJ8AJ8BF

Problem Statement

Goal

- Given a set of known **online extremist community** members, detect the greater community of interest at scale with high precision

Challenges

- Practical: Achieving high precision
- Technical: Incorporating multimode, multiplex graph data

Definition of Terms

- **Online Extremist Communities (OEC)** - A social network of users who interact within social media in support of an extremist group or groups.
- **Extremist Group** – a group advocating actions that pose a threat to national security or human rights.
- **Online Extremist Community Member** - Social media users who unambiguously affirm the leadership, ideology, or fighters of an extremist organization.
- The role of "**passive sympathizers**" who merely share or re-post content has been shown to be in increasingly important component of extremist propaganda dissemination in social networks. Veilleux-Lepage (2015)

Note: It is important to emphasize that a member's "support" is relative and in many cases not in violation of local law or a social media platform's terms of use.

Relevance

“So we need to, candidly, stop tweeting at terrorists. I think we need to focus on exposing the true nature of what Daesh is.”

Mr. Michael Lumpkin
Director, Global Engagement Center

Organizational Charter: coordinate, integrate and synchronize messaging to foreign audiences that undermines the disinformation espoused by violent extremist groups, including ISIL and al-Qaeda, and that offers positive alternatives.

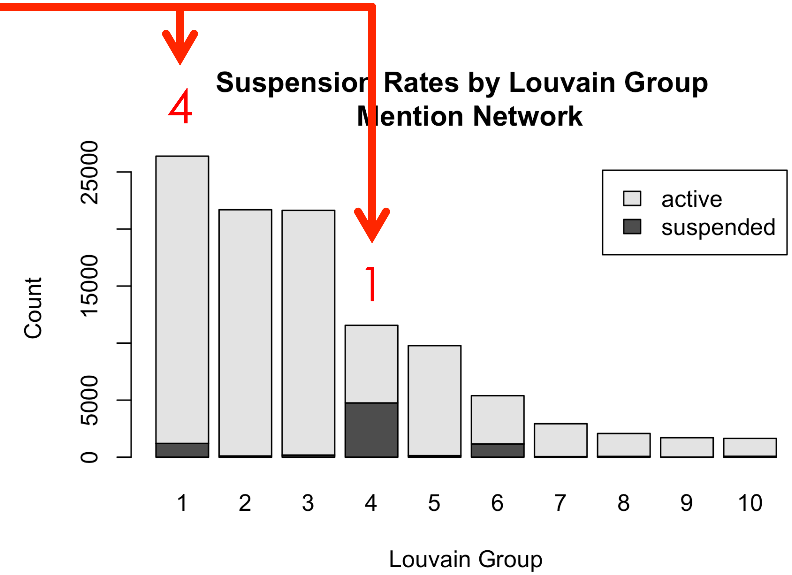

Pentagon: We don't have enough people tweeting against Isis to win online propaganda war

By Mary-Ann Russon
October 23, 2015 19:00 BST

Michael Lumpkin, Assistant Defense Secretary for Special Operations/Low-Intensity Conflict, has told the US Congress that there are not enough people working to combat Isis messages on social media (Reuters)

- Countering extremist social media campaigns will require
  - An understanding of the populations most susceptible to radicalization
  - An understanding of the online community’s topology
- Current practices attempts to identify OECs through
  - Researchers
    - Community detection (poor precision)
  - Practitioners
    - Bounded searches (poor precision)
    - Manual Construction (poor recall)

Background



Joseph A. Carter, Shiraz Maher, and Peter R. Neumann, "#Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks," *International Centre for the Study of Radicalization Report*, April 1, 2014.

- In NOV 2014 Collected 119k Twitter User Accounts using a 2-step snowball search of 5 prominent ISIS supporters'$_1$ following ties.
- Standard community detection methods fail to identify extremist clusters with adequate precision
- Twitter suspension rates indicate ISIS membership within specific Louvain groups and could be used to detect members as a classification problem

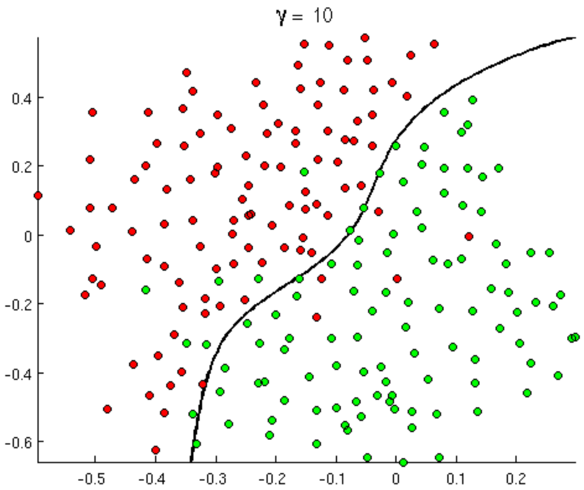
Detection / Classification

Implementation

- Data is usually large (70-200k users)

Training Data. How big is big enough?

- Our best results usually have thousands of labeled instances.
- Clustering methods can be used to find labels
- Active learning – select the most informative instances to label and re-train the classifier

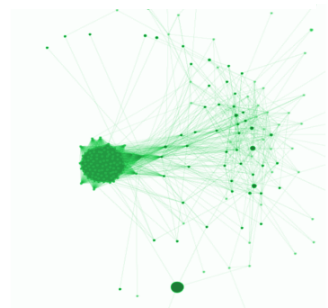


Data Representation

Social Media Feature Integration

Following Graph

Mention Graph



Hash Tag Graph

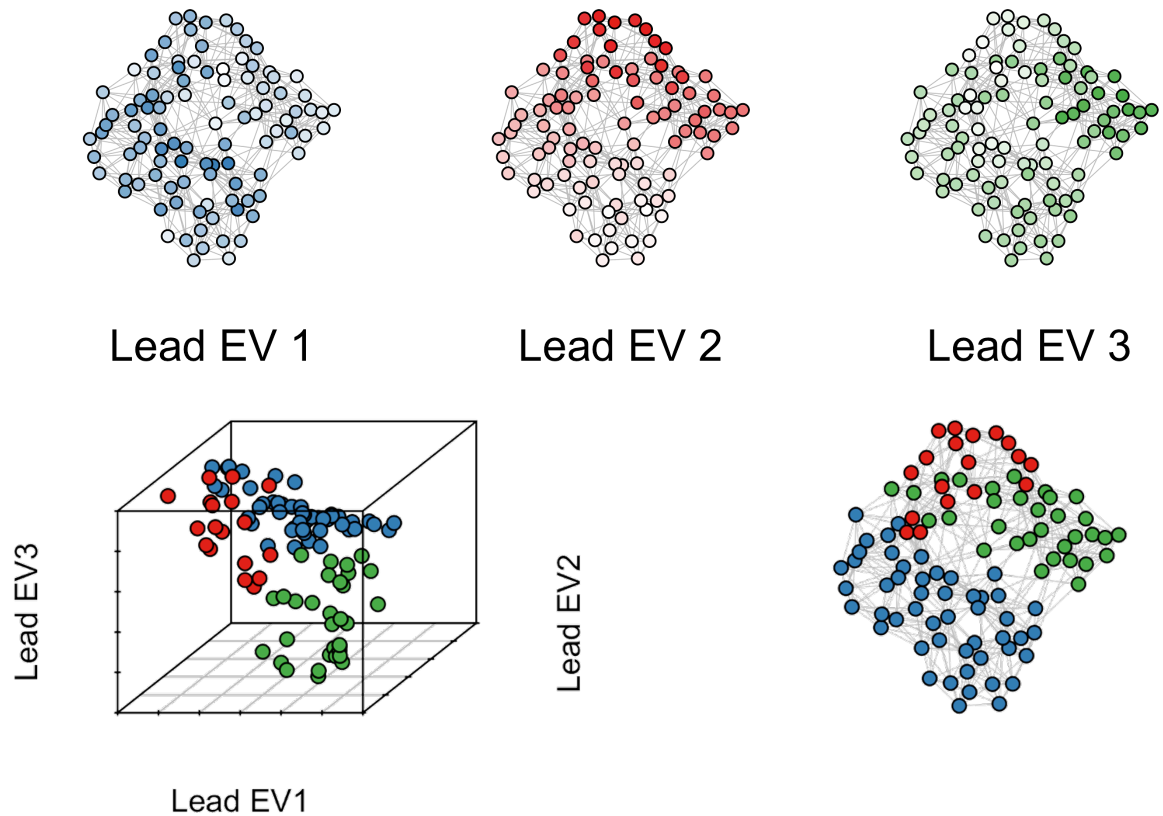
Dimensional reduction through extraction of the **k** lead Eigenvectors of each graph

$$\underset{n \times (m+4k)}{Z} = \underset{n \times m}{A} \;\cdots\; \underset{n \times k}{F} \;\cdots\; \underset{n \times k}{M} \;\cdots\; \underset{n \times k}{H}$$

Data Representation

1. Extract K lead eigenvectors
2. Form Matrix U by concatenating eigenvectors column-wise
3. Cluster (usually with k-means)

Simple Example



Normalized Spectral Clustering: Ng, Jordan, and Weiss (2002)

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct

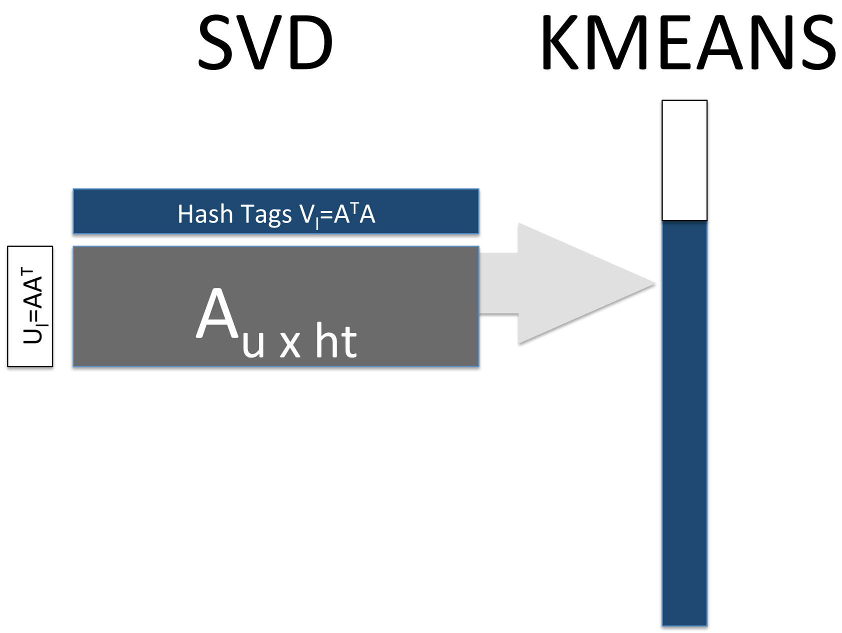
1. Let W be a weighted adjacency matrix
2. Compute the unnormalized Laplacian L_{sym}.
3. **Compute the first k eigenvectors $v_1, \ldots, v_k$ of L_{sym}.**
4. Let $V \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $v_1, \ldots, v_k$ as columns.
5. **Form the matrix $U \in \mathbb{R}^{n \times k}$ from V by normalizing the row sums to have norm 1,** that is $u_{ij} = \frac{v_{ij}}{\sum_k v_{ik}^2{}^{\frac{1}{2}}}$
6. For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the $i - th$ row of U.
7. Cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with k-means algorithm into clusters $C_1, \ldots, C_k$.

output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j | y_j \in C_i\}$

Data Representation

Co-clustering of Bipartite Graphs

- Dhillon 2001 projects both the users and hash tags into the same vector space
- Advantage: interpretable clusters
- Disadvantage: challenging to add other user features
- The user needs an efficient way to analyze clusters



Extracting lead Eigen vectors from the SVD decomposition of bipartite graphs allows us to incorporate other node classes to cluster users (i.e. hash tags)

Related Work

Principal Modularity Maximization (PMM) Uses dimensionality reduction to identify multiplex group structures

- Lei Tang, Xufei Wang, and Huan Liu, "Uncovering Groups via Heterogeneous Interaction Analysis," in *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on* (IEEE, 2009), 503–12.

SocioDim Uses dimensionality reduction and user account information to cluster/classify vertices

- Xufei Wang et al., "Discovering Overlapping Groups in Social Media," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (IEEE, 2010), 569–78.

- Lei Tang and Huan Liu, "Leveraging Social Media Networks for Classification," *Data Mining and Knowledge Discovery* 23, no. 3 (2011): 447–78

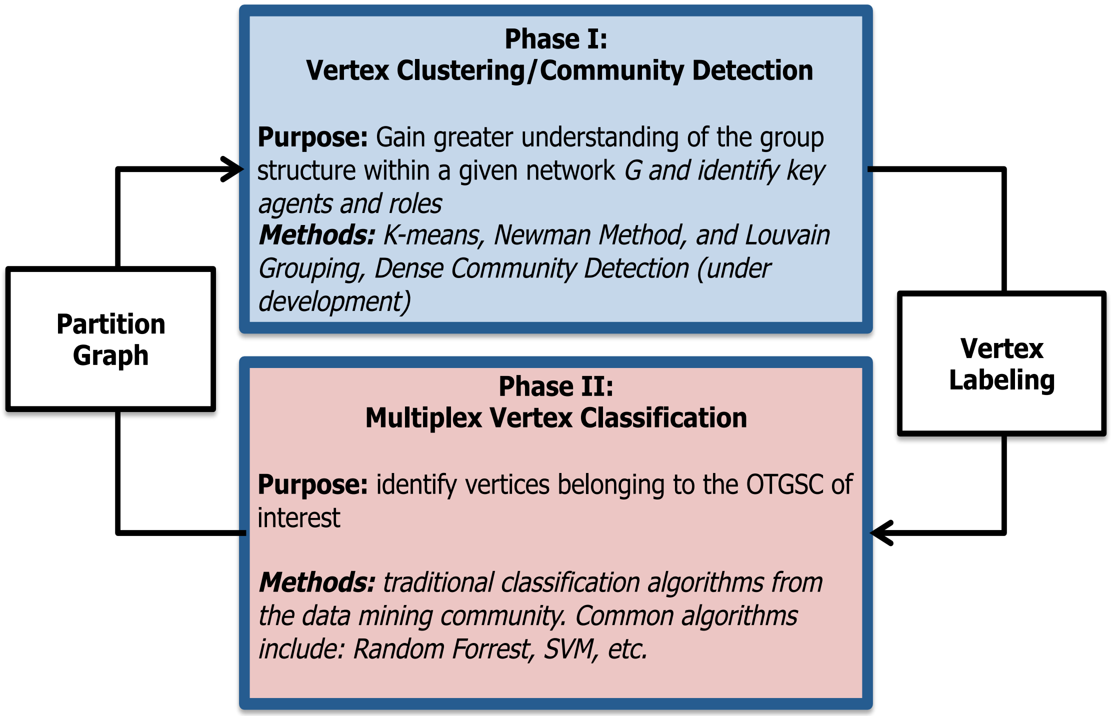
Methodology

Iterative Vertex Clustering and Classification

- Clustering methods are used to develop training sets
- Supervised learning detects the community at scale
- The density of OSNs like Twitter drive the methodology to be conducted iteratively

Methodology

Iterative Vertex Clustering and Classification (IVCC)



Detection/Classification

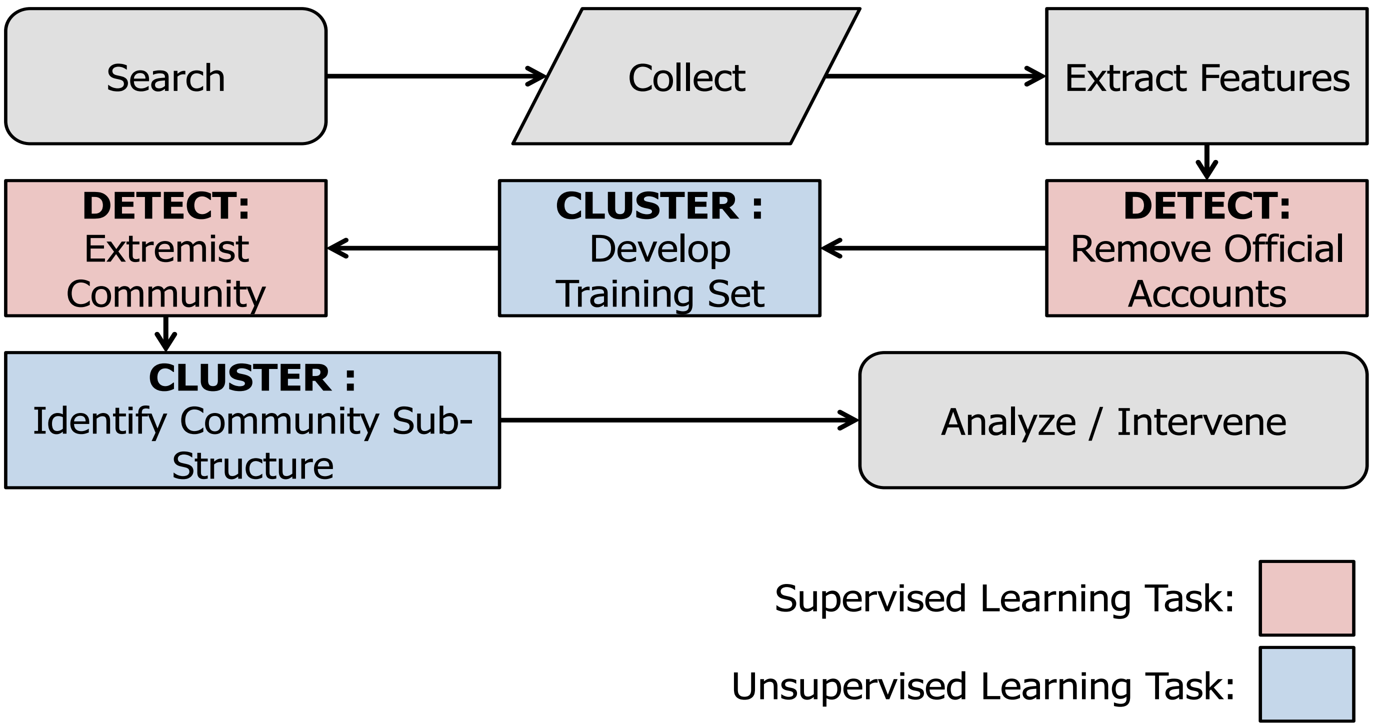
Iterative Classification

- Discriminating between accounts that are central within the extremist community of interest and globally central accounts is challenging
- Celebrity, news media, political leader and corporate accounts must be systematically removed through classification

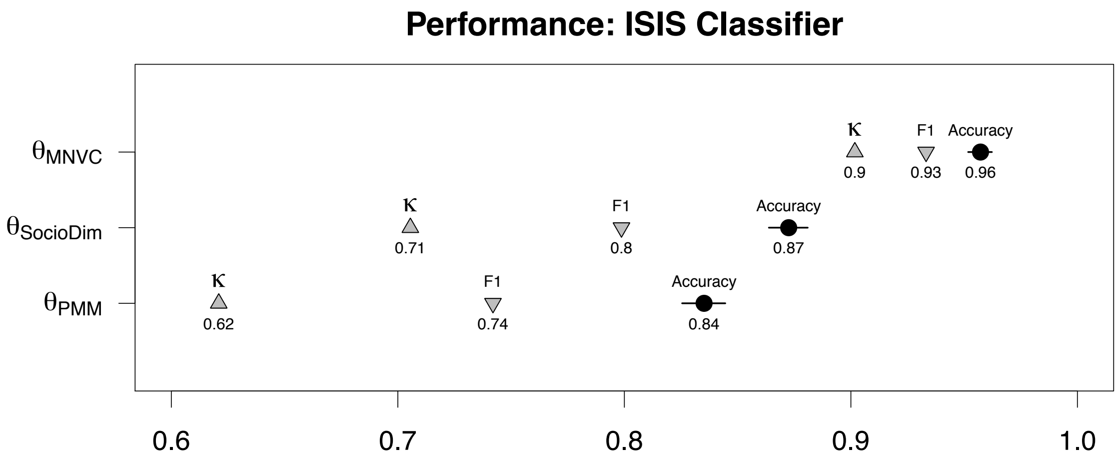
Methodology

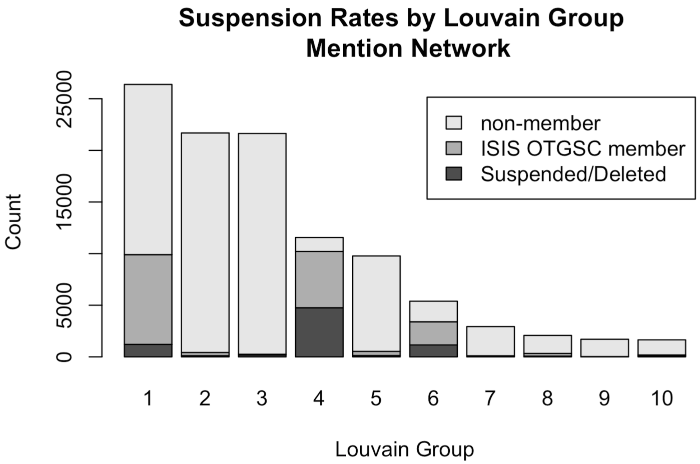
Iterative Vertex Clustering and Classification

Case Study Work Flow



Performance





Strengths

- High accuracy (with large training set)
- Scales well
- Computationally efficient

Limitations

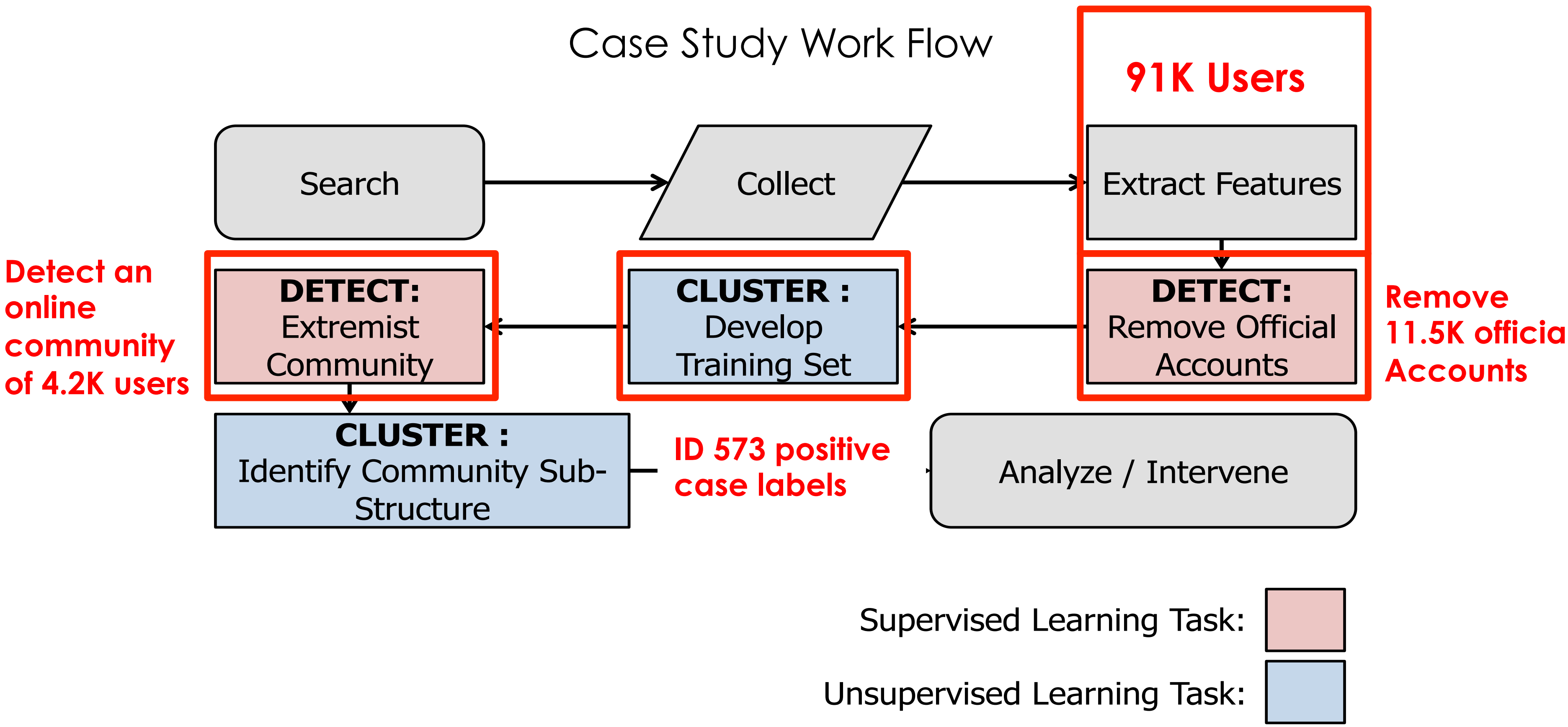
- Evaluation metrics are often positively biased
- Training sets are sensitive to false positives

PE: Detecting The Euromaidan Twitter Community

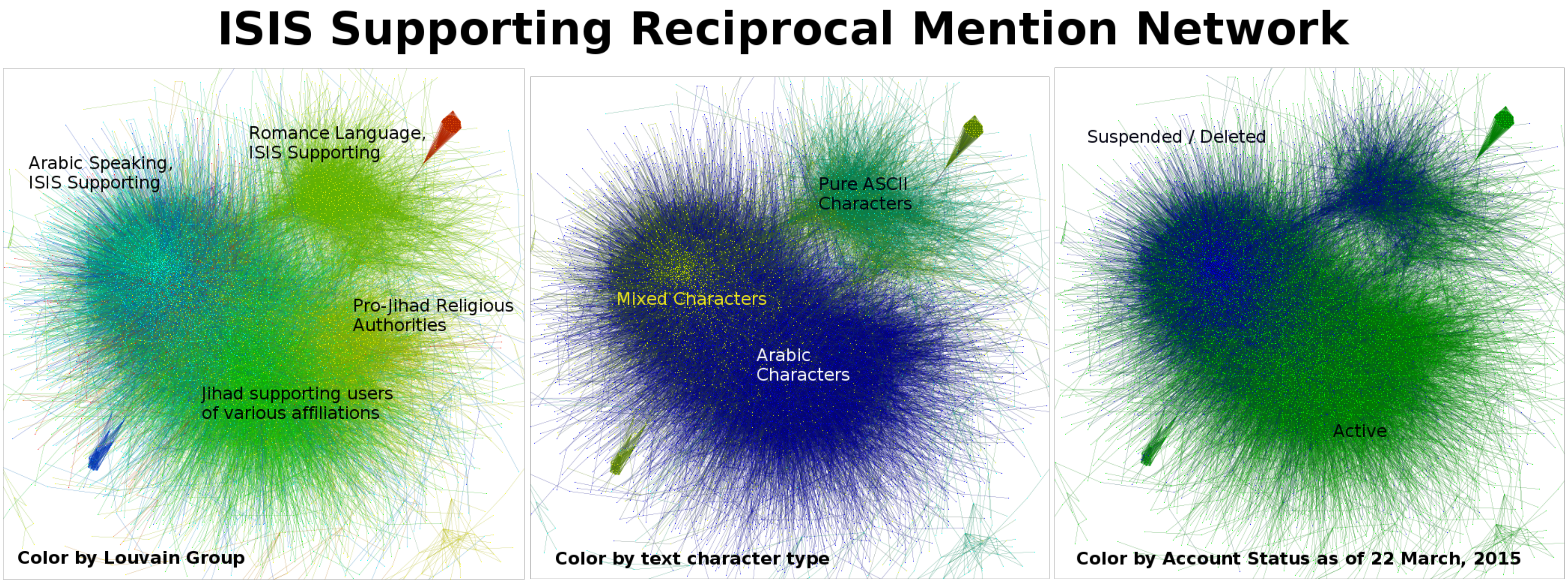
- 2-hop snowball sample of 8 accounts associated with the 2014 Euromaidan Revolution in Ukraine
- ~92,000 users collected from the Twitter Search API in OCT 2015
- Limited Training Data

Euromaidan PE

Iterative Vertex Clustering and Classification

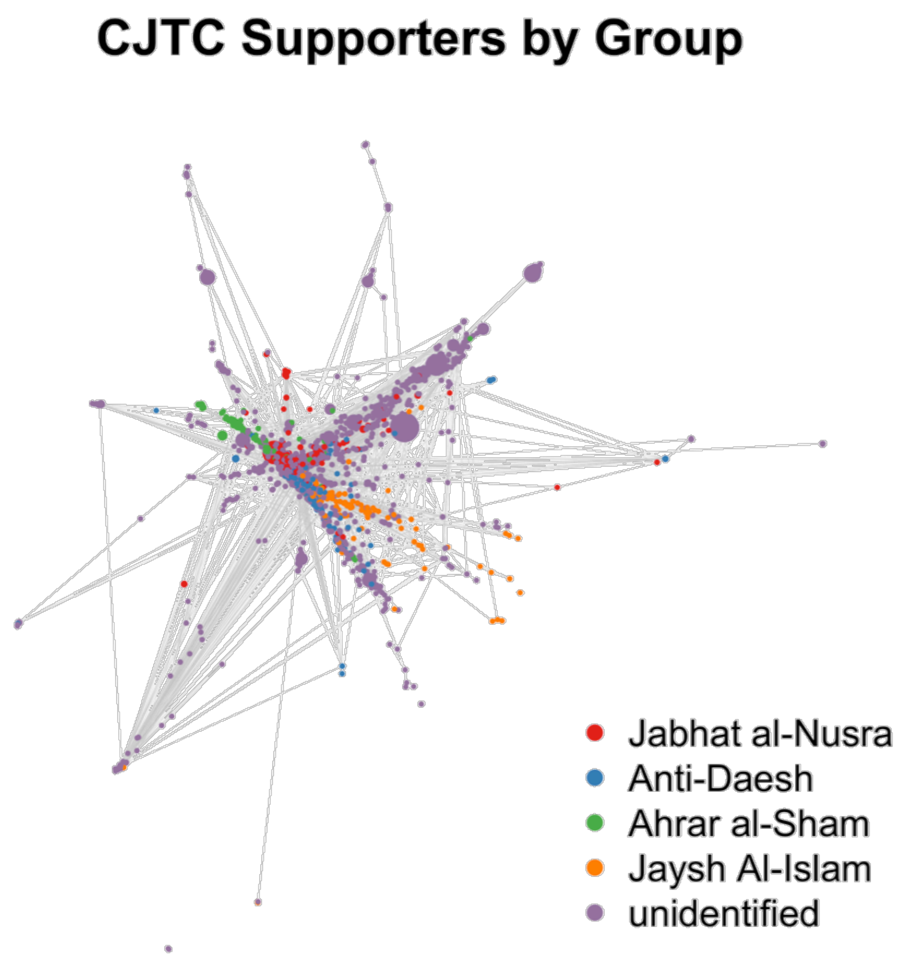


Analysis: Sub-Structure



Distinct communities would likely be interpretable by analysts

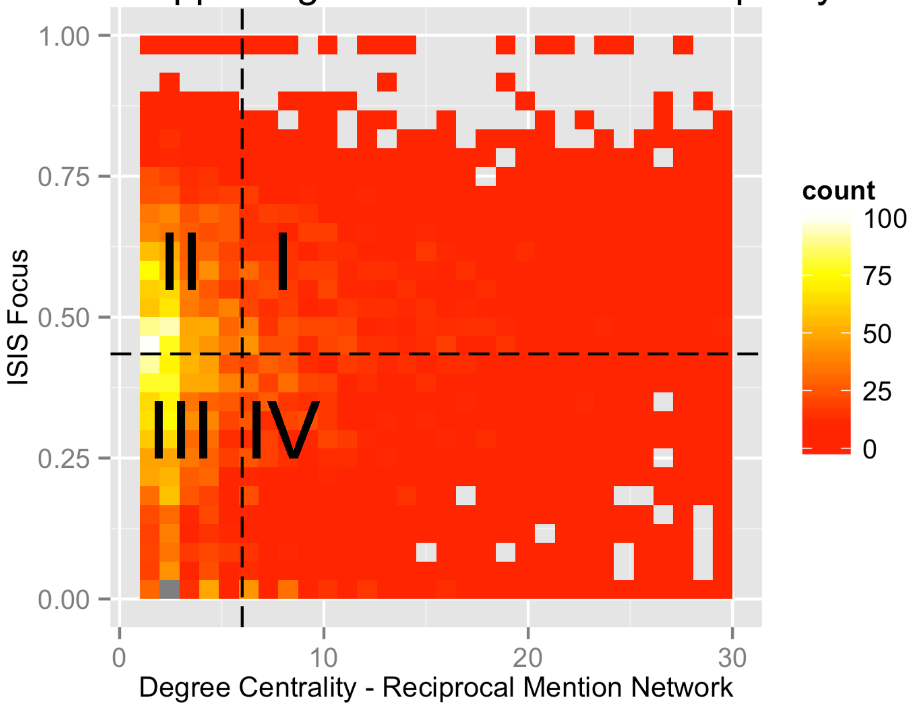
Analysis: Sub-Structure



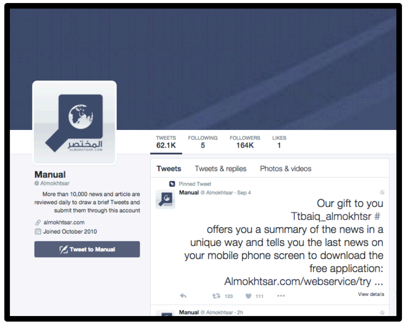
Distinct communities would likely be interpretable by analysts

Analysis: Key Individuals

ISIS Supporting Network: Core vs. Periphery



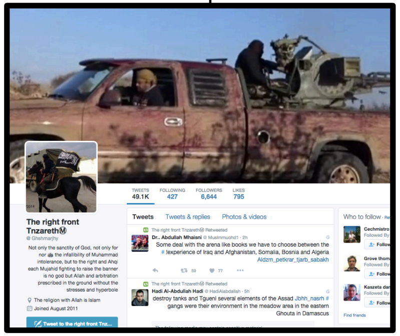
Pseudo News



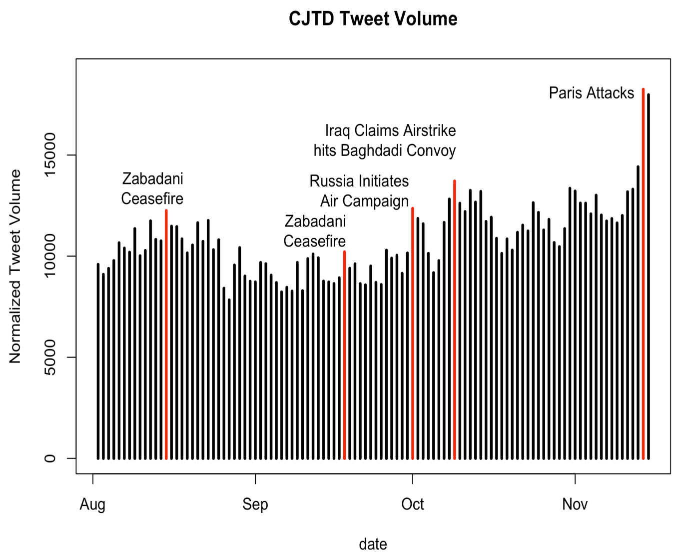
Islamic Teaching

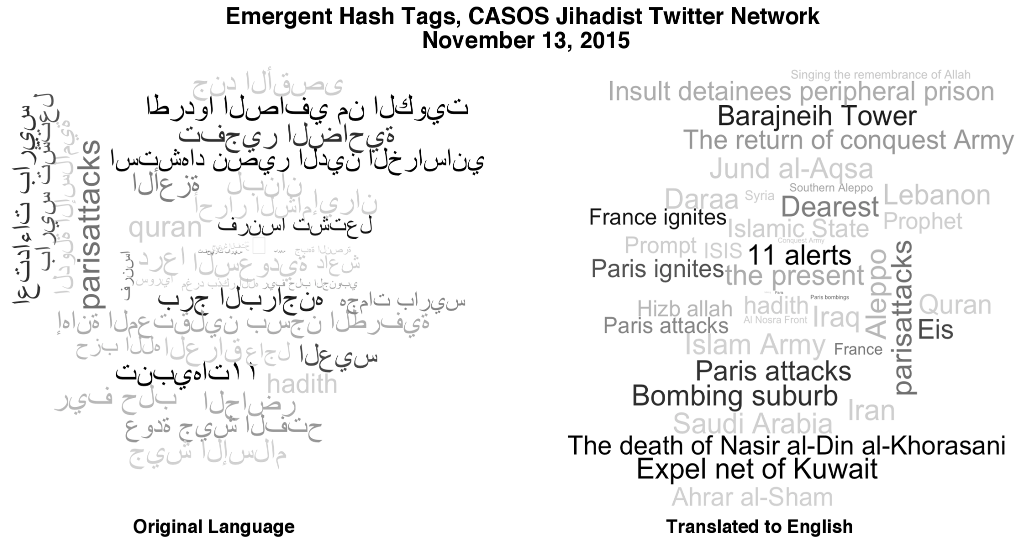
Anti-ISIS Propaganda

al-Nusra Battle Updates



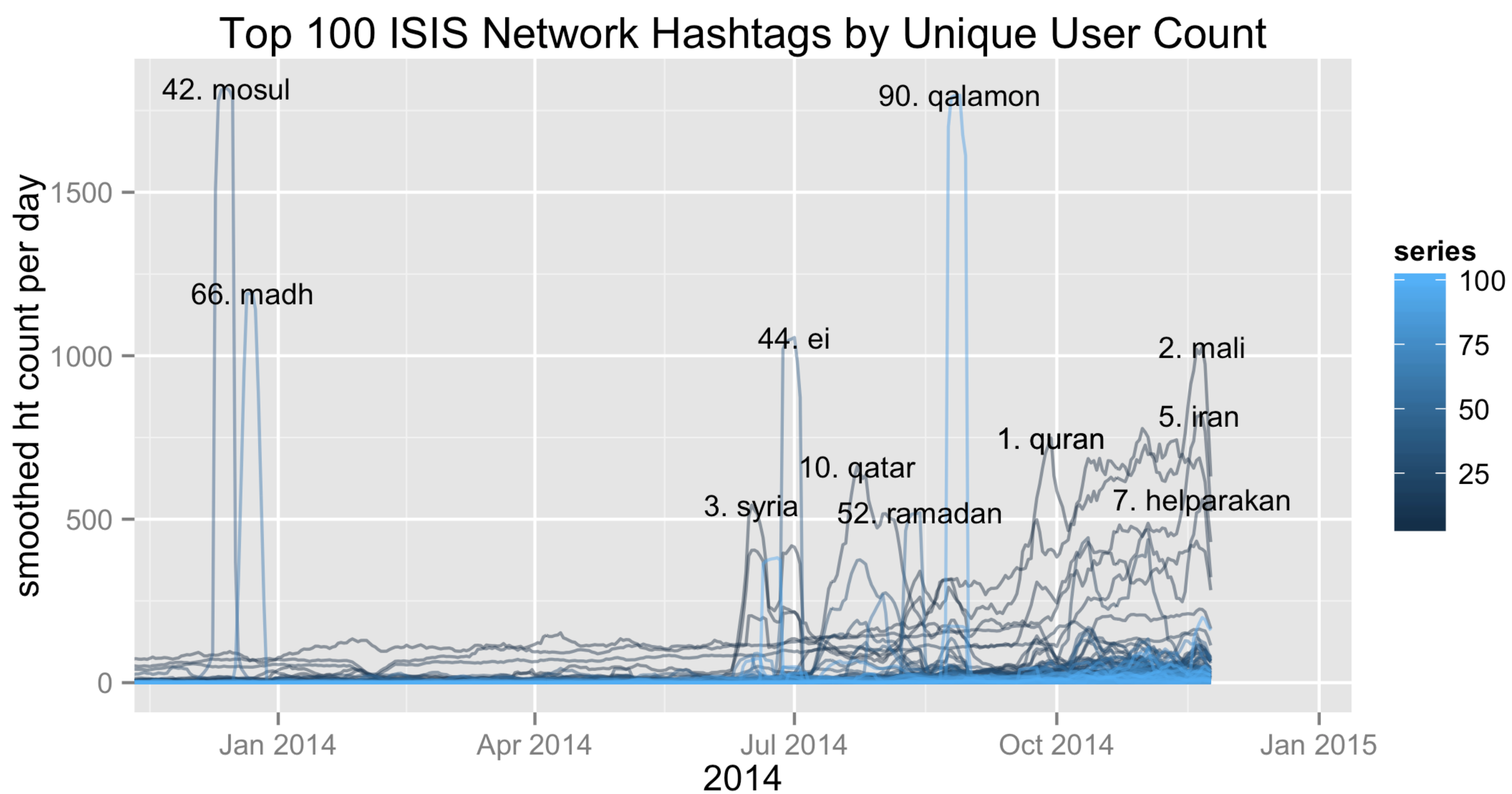
Analysis: Community Activity



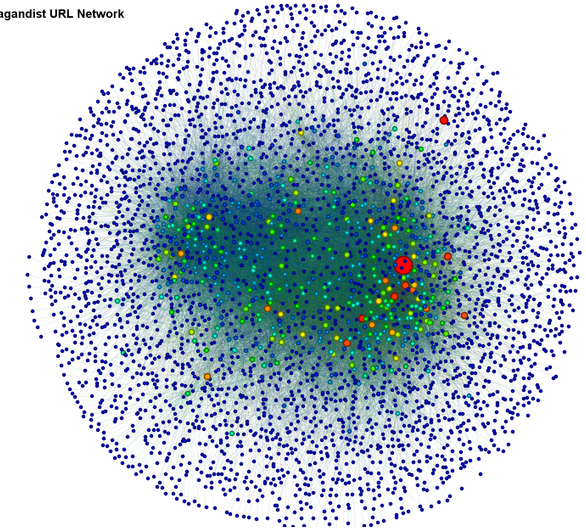


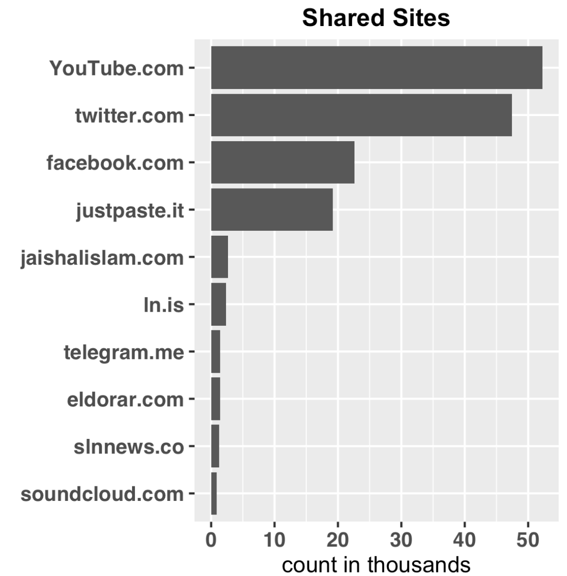
- CASOS Jihadist Twitter Network (CJTN):
  - 16,000 active tweeters promoting one or more of the major Sunni Jihadist groups engaged in Syria, Northern Iraq and Yemen.
- Could provide insight into:
  - demographics most likely to provide active support
  - The ongoing propaganda war for passive supporters

Analysis: Community Activity



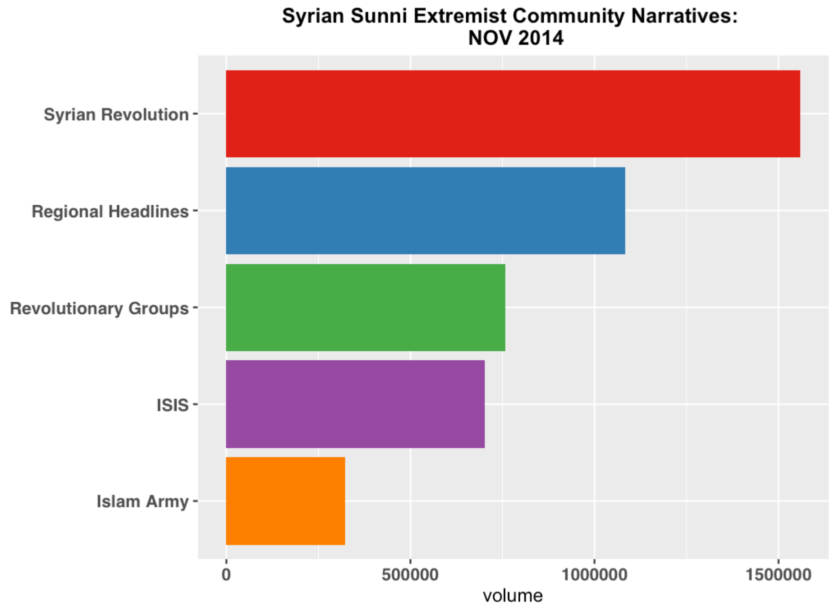
Analysis: Radicalization

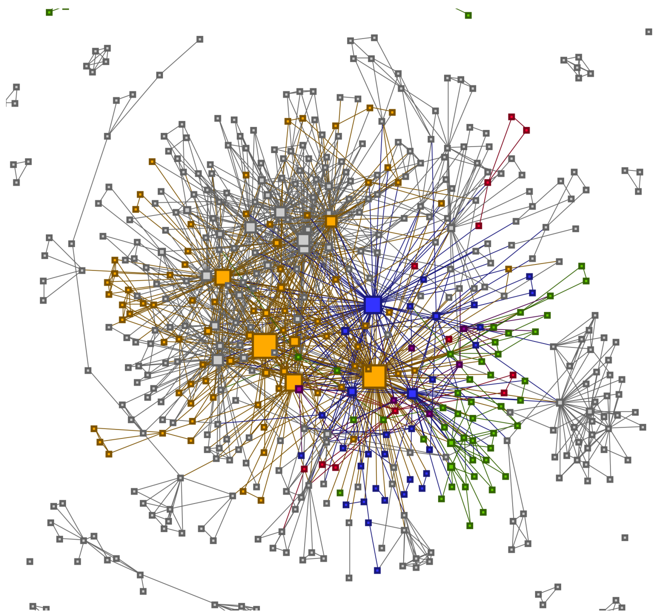




- Targeted grooming and recruiting require peer to peer messaging via the @mention and often use URLs to highlight propaganda or move conversations to a more secure site.
- Vertex color and size on the left plot indicates a the volume of tweets mentioning other members of the community and sharing URLs. We hypothesize blue vertices are recruitment targets. The right panel highlights shortened URLs most commonly used.
- Tweets sharing a peer to peer URL typically contain an userID as well.

Analysis: Mining Narratives





- The hash-tag co-occurrence network can be clustered to find trending narratives within the community
- Network structure of hash tag co-occurrence graph enables collection of related, lower frequency hash tags within the narrative

Analysis: Mining Narratives

Narrative: Syrian Revolution

translated

- Color depicts whether the term was 'trending' within the analyzed time period by comparing the daily rate to a 12 month average. The more gray the term, the less anomalous the daily rate during the analyzed time period.
- Size depicts raw count during the analyzed time period.
- The right panel depicts Google translations of each hash tag to English

Analysis: Mining Narratives

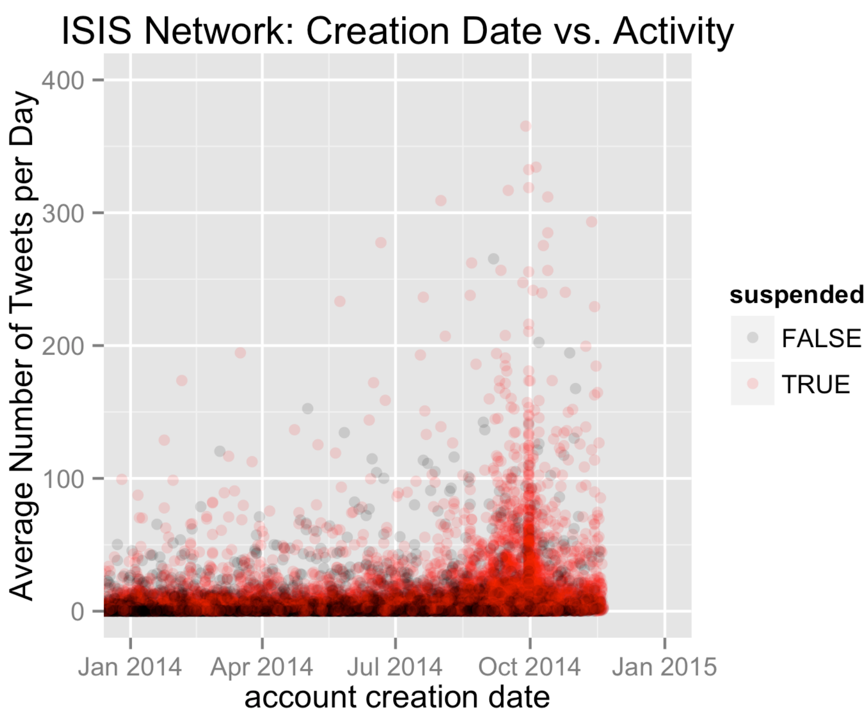
Narrative: ISIS

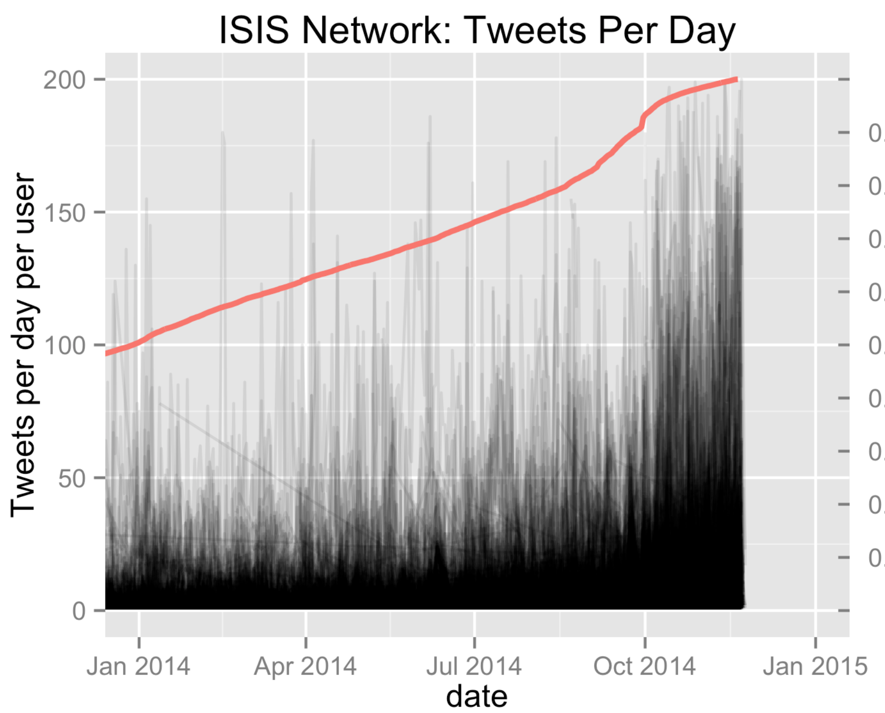
translated

anti-Daesh sentiment clearly present

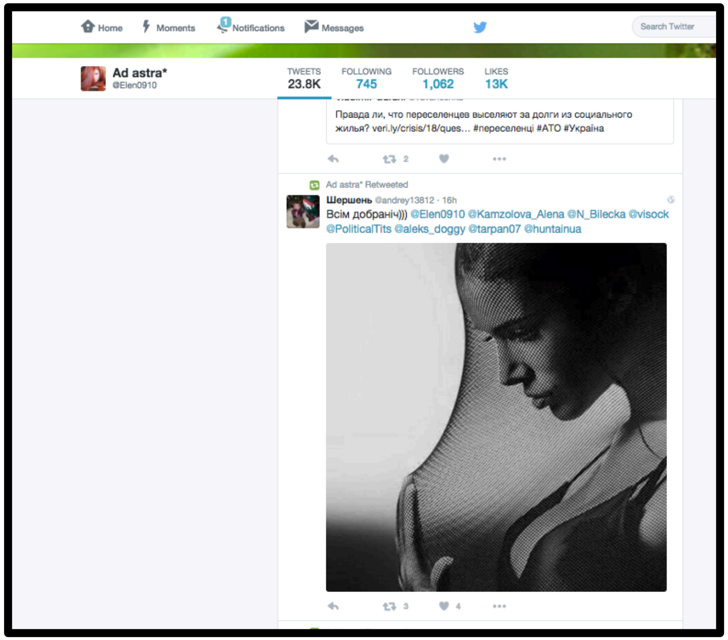
- Color depicts whether the term was ‘trending’ within the analyzed time period by comparing the daily rate to a 12 month average. The more gray the term, the less anomalous the daily rate during the analyzed time period.
- Size depicts raw count during the analyzed time period.
- The right panel depicts Google translations of each hash tag to English

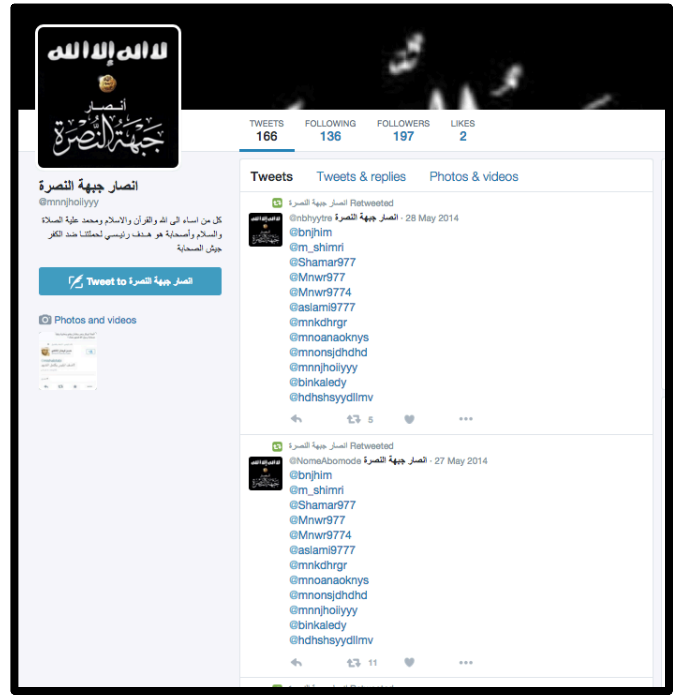
Analysis: Anomaly Detection





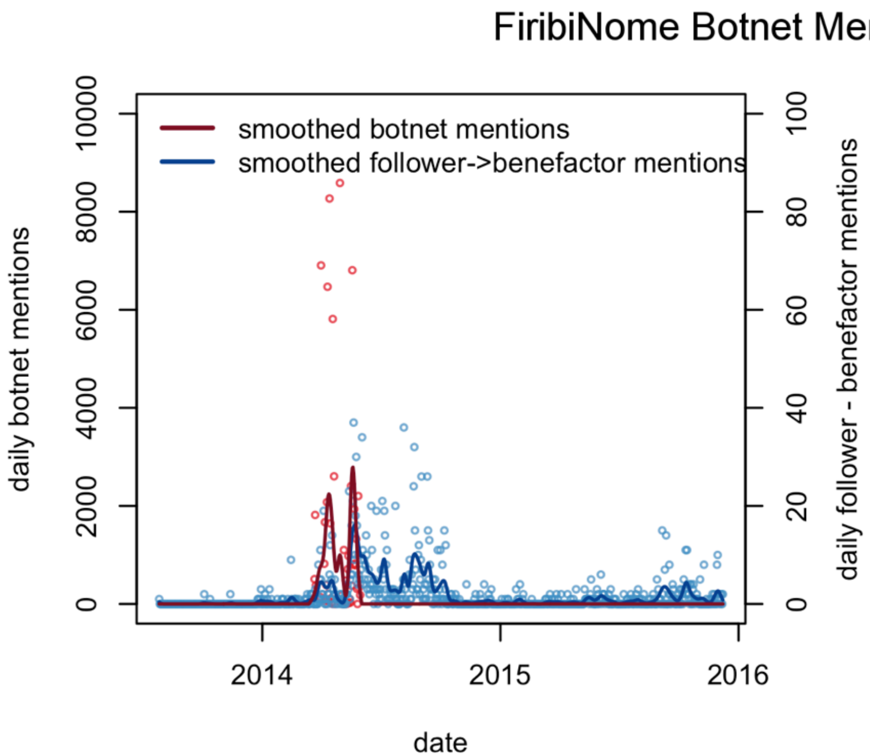
Analysis: Topology and Social Influence

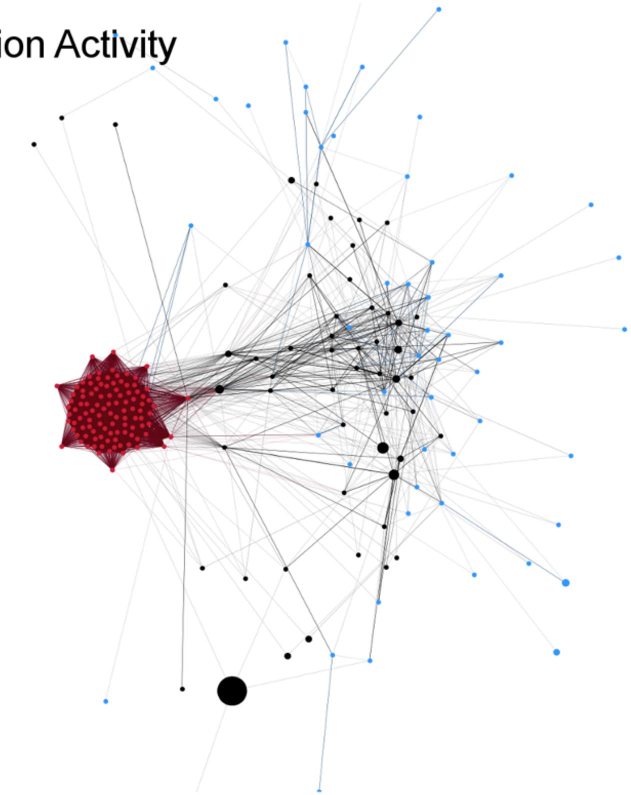




Sophisticated use of @mentions can be used to increase size and interconnections within communities

Analysis: Topology and Social Influence





- Social Botnet uses strings of mentions to gain followers and promote accounts
- If you understand the network topology, you can gain a large following and influence

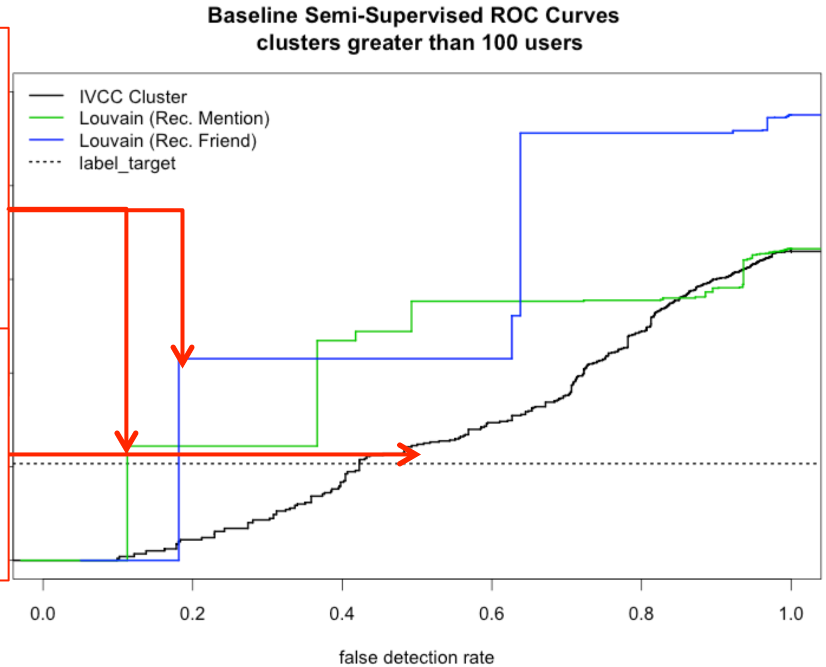
Analysis: PE

- We are unable to share tweets due to the terms and conditions of the Twitter API services, however ORA can:
  - Import and construct networks directly from Twitter JSON output
  - Conduct topic modeling on large Twitter corpuses

Limitations

existing methods return Large clusters with poor precision, or many small clusters with high

IVCC performs poorly as a clustering feature space



Evaluation

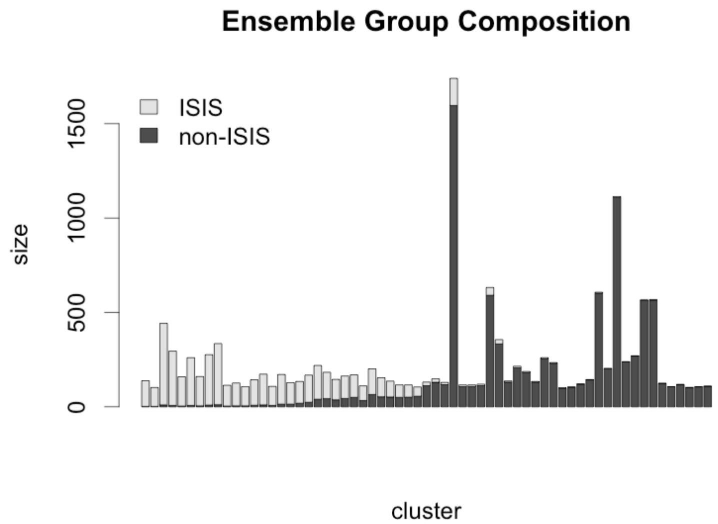
- Assume supervised results as ground truth
- Remove 'small' clusters
- Order from highest percentage of + case to lowest

- Large training sets needed to achieve high recall
- Performance evaluation non-trivial
- IVCC feature space works poorly for clustering
- Analysis of detected OTGSCs requires tailored methods as well

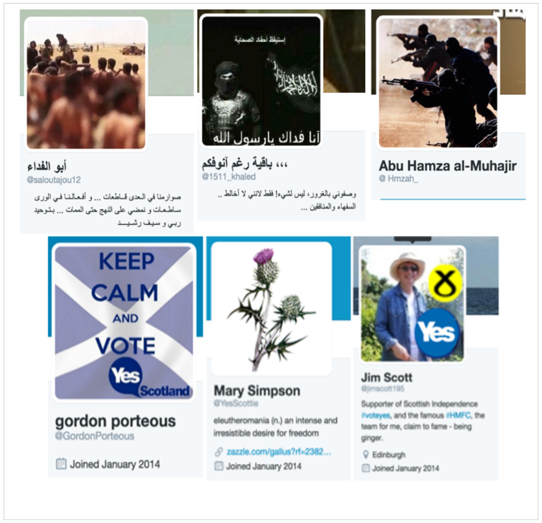
Future Work

- Unsupervised Dense Community Detection
- Active Learning to Incorporate Regional Expertise
- A research framework and extended literature review of theory and methods for detection, analysis, and disruption of **online extremist communities**

Dense Community Detection in Heterogeneous Graphs



Example ISIS Community



Example non-ISIS Community

Dense Community: clusters within a social network that are highly connected across multiple edge and vertex types.

- Ensemble returned 62 large dense communities containing ~13,000 users. The result was a large (over 3K), precise set of positive case labels
- Used to update the CJTD in MAR16 and MAY16, and CCMD with similar results