Modeling and inferencing for activity profile of terrorist groups

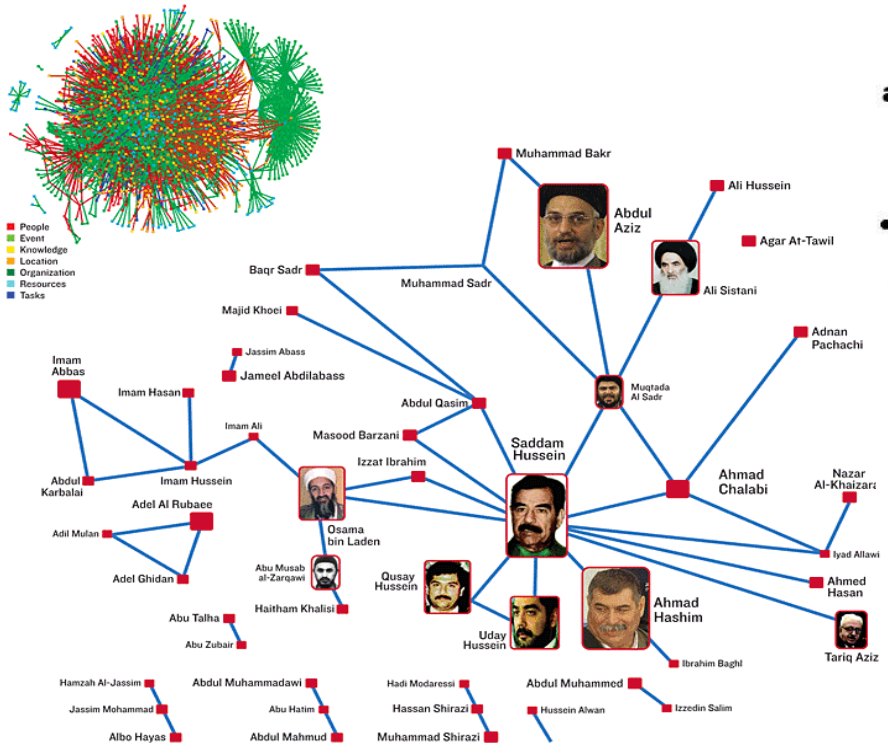
Vasanthan Raghavan

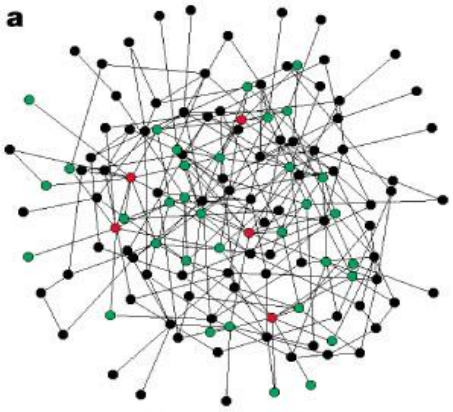
Qualcomm Flarion Technologies, Inc.

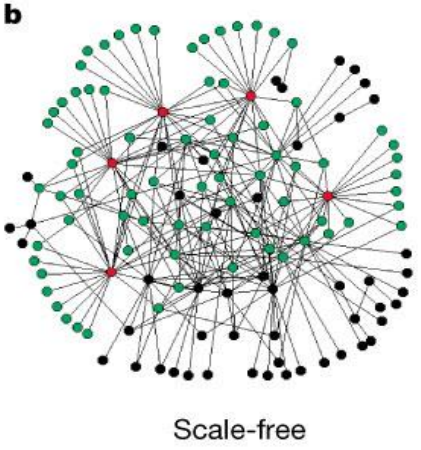
Bridgewater, NJ

October 24, 2014

TERRORIST NETWORKS







- Terrorism has been around and has been studied for a long time
- Ongoing radicalization of different interest groups
- Rise of social media has made tracking terrorist activity a harder task

- "Data science" problems: Network dynamics and evolution, user classification, information dissemination, missing links, anomaly detection

FUNDAMENTAL CHALLENGES

- **Challenge 0:** How to incorporate the network into the model?

- **Challenge 1:** Multivariate observations are of mixed type
  - Time and location of attack
  - Intensity of attack (injured, dead, "walking dead")
  - Impact of attack (economic damage, political damage, loss of confidence of any kind)
  - Localized vs. globalized impact, e.g., 9/11 vs. Oklahoma City bombings

  Not all the data can be quantified

  Not all the attacks are comparable

- **Challenge 2:** Temporal modeling issues

$$\mathcal{H}_{i-1} = \{M_1, \cdots, M_{i-1}\} \Longrightarrow \mathrm{P}(M_i = r | \mathcal{H}_{i-1}),\ r = 0, 1, 2, \cdots,\ i = 1, \cdots, \mathcal{N}$$

  - Point process model (Poisson, renewal, etc.)
  - Correlation/clustering of attacks in time

EXISTING MODELS FOR TERRORISM- I

- **Type 1:** Classical time-series techniques
  - Transform, fit trend, seasonality and stationary components to time-series [Brophy-Baermann & Coneybeare, Cauley & Im, Enders & Sandler]
  - Fit lagged value of endogenous variables, and other variables [Barros]
  - Quadratic or cubic trend = 4 parameters, seasonality = 3, stationary part = 1, often 8 or more model parameters
- **Key Theme:**
  - Study of impact of interventions (airport sec. checks, Reagan-era laws)

$$M_{1,i} = a_1 M_{1,i-1} + b_1 M_{2,i-1} + c_1 p_1 + \text{Other comps.}$$
$$M_{2,i} = a_2 M_{2,i-1} + b_2 M_{1,i-1} + c_2 p_1 + \text{Other comps.}$$

Two attack types

Impact of intervention

- Good-to-acceptable fit for time-series at the cost of large number of parameters in a model with complicated dependencies
- Some interventions have no apparent long-term effect

EXISTING MODELS FOR TERRORISM- II

- **Type 2:** Group-based trajectory analysis
  - Identify cases with similar development trends [Nagin]
  - Cox proportional hazards model + logistic regression methods for model selection [LaFree, Dugan & co-workers]

- **Key Themes:**
  - Focussed on worldwide terrorism trends instead of specific groups
  - Contagion theoretic viewpoint → Current activity of group is influenced by past history of group → Attacks are clustered

EXISTING MODELS FOR TERRORISM- III

- **Type 3:** Self-exciting hurdle model (SEHM)
- Puts the contagion point-of-view on a theoretical footing
- Motivated by similar model development in
  - Earthquake models – Aftershocks are function of current shock
  - Inter-gang violence – Action-reaction violence between gangs
  - Epidemiology – immigrants + offsprings in a cell colony

$$\mathrm{P}\left(M_i = r | \mathcal{H}_{i-1}\right) = \begin{cases} e^{-(B_i + SE_i(\mathcal{H}_{i-1}))}, & r = 0 \\ \frac{r^{-s}}{\zeta(s)} \cdot \left(1 - e^{-(B_i + SE_i(\mathcal{H}_{i-1}))}\right), & r \geq 1 \end{cases}$$

- Hurdle probability component: Accounts for few attacks
- Self-exciting component: Accounts for clustering of attacks
- **Key Theme:**
  - Excellent model-fit
  - Explains clustering of attacks from a theoretical perspective
  - Self-exciting component can be complicated → more parameters

[Mohler et al. 2011, Porter & White 2012, White, Porter & Mazerolle 2012, Lewis 2013]

MOTIVATING ASSUMPTIONS - I

- **Assumption 1:** Current activity of the group depends on past history only through k dominant states $\mathbf{S}_i = [S_{1,i}, \cdots, S_{k,i}]$ (that remain hidden)

$$\mathrm{P}(M_i|\mathcal{H}_{i-1}, \mathbf{S}_i) = \mathrm{P}(M_i|\mathbf{S}_i),\ i = 1, 2, \cdots$$

- **Assumption 2:** Of these k states, the two most dominant are
  - Its Intentions ($S_{1,i}$) – Guiding ideology/philosophy (e.g., Marxist-Leninist-Maoist thought, political Islam), designated enemy group, nature of high profile attacks, nature of propaganda warfare, etc.
  - Its Capabilities ($S_{2,i}$) – Manpower assets, special skills (bomb-making, IED), propaganda warfare skills, logistics skills, coordination with other groups, ability to raise finances, etc.
  - Capabilities are tempered by Strategies/Tactics (repeated/multiple attacks over time – group resilience, multiple attacks over space – coordination)

$$\mathrm{P}(M_i|\mathbf{S}_i) = \mathrm{P}(M_i|\{S_{1,i}, S_{2,i}\})$$

[Cragin and Daly, "The dynamic terrorist threat: An assessment of group motivations and capabilities in a changing world"]

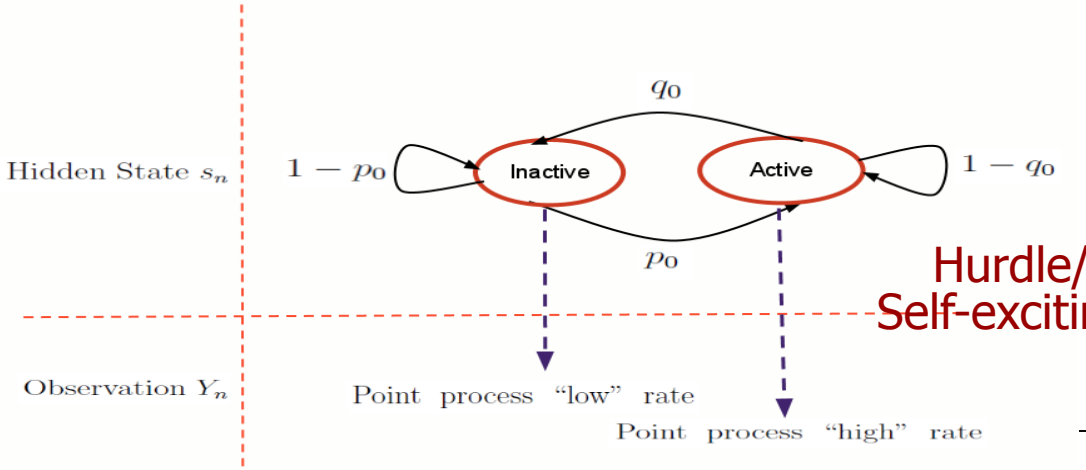
MOTIVATING ASSUMPTIONS - II

- **Assumption 3:**
  - Mature group ➔ Intentions are to attack (more or less)
  - Change in capabilities is primarily responsible for change in attack patterns

$$\mathrm{P}(M_i|\{S_{1,i}, S_{2,i}\}) = \mathrm{P}(M_i|S_{2,i})$$

$$S_{1,i} = \text{Always intend to attack}$$

  - A d-state model for Capabilities
    - d = 2: → Active state (high capability/strong), Inactive state (low capability/weak)
    - Observation density: Different possibilities (Poisson, shifted Zipf, geometric, etc.)



Hurdle/State transitions → Data rarity
Self-exciting comp./Diff. rates → Clustering

COMPARING MODEL FRAMEWORKS

- All three models (TAR, SEHM and HMM) provide a framework for explaining clustering of attacks
  - TAR: Current observation is explicitly dependent on past observations

$$M_{1,i} = \mathsf{a}_1 M_{1,i-1} + \mathsf{b}_1 M_{2,i-1} + \mathsf{c}_1 \mathsf{p}_1 + \text{Other comps.}$$
$$M_{2,i} = \mathsf{a}_2 M_{2,i-1} + \mathsf{b}_2 M_{1,i-1} + \mathsf{c}_2 \mathsf{p}_1 + \text{Other comps.}$$

  - SEHM: Prob. of attack is enhanced by history of group

$$\frac{\mathsf{P}(M_i > 0 | \mathcal{H}_{i-1})\Big|_{\text{SEHM}}}{\mathsf{P}(M_i > 0 | \mathcal{H}_{i-1})\Big|_{\text{Non-SEHM}}} = 1 + \frac{e^{-B_i}}{1 - e^{-B_i}} \cdot \left(1 - e^{-SE_i(\mathcal{H}_{i-1})}\right) \geq 1.$$

  - HMM: Combines facets of both TAR and SEHM
    - Observation depends on state
    - Current state depends on past state
    - Prob. of attack is enhanced based on state realization

$$\mathsf{P}\left(M_i = r | \mathcal{H}_{i-1}\right) = \sum_{j=0}^{d-1} \sum_{k=0}^{d-1} \mathsf{P}\left(M_i = r | S_i = j\right) \cdot \mathsf{P}\left(S_i = j, S_{i-1} = k\right)$$

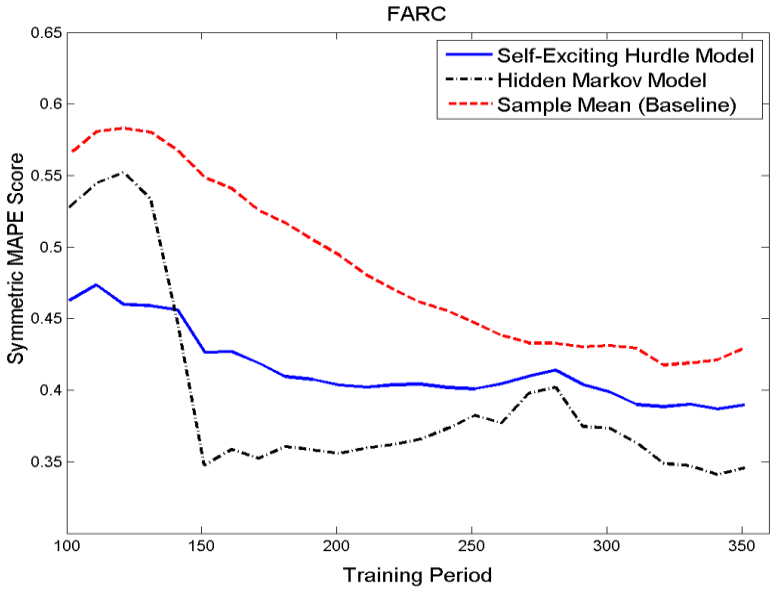
HMM vs. SEHM

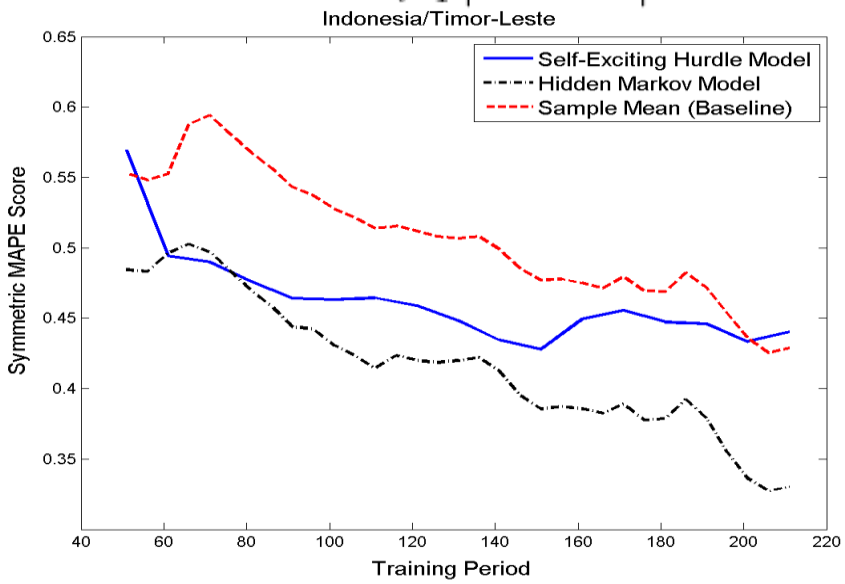
- Explanatory power: FARC and Indonesia datasets (AIC as metric)

$$\texttt{AIC}(n)\Big|_{\text{HMM}} \triangleq 2k_{\text{HMM}} - 2\log\left(\mathrm{P}(\Delta_1^N|\lambda)\right)$$

FARC			Indonesia/Timor – Leste		
n	SEHM	HMM	n	SEHM	HMM
100	671.68	671.06	100	723.78	729.47
200	1117.40	1112.07	165	1091.78	1116.92
300	1521.93	1521.36	200	1283.08	1305.27
400	2127.55	2121.81	250	1589.43	1615.87
450	2333.88	2327.02	300	2018.92	2041.35

- Predictive power (SMAPE as metric): $\texttt{SMAPE}(N) \triangleq \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\Delta_i - \widetilde{\Delta}_i}{\Delta_i + \widetilde{\Delta}_i}\right| \cdot \mathbb{1}(Q_i = 1)$





See [R, Galstyan & Tartakovsky, AOAS 2014] for more details

LESSONS FROM MODEL LEARNING

- HMM: If parsimony is critical, a geometric obs. model is good

$$\mathrm{P}(M_i = k | S_{2,i} = j) = (1 - \gamma_j) \cdot (\gamma_j)^k$$

  - Group has a short-term objective
  - Every new attack contributes equally to the success of this objective
  - As long as obj. is not met, group remains oblivious (memoryless) of past activity
- Otherwise, a hurdle-based geometric is a good fit

$$\mathrm{P}(M_i = k | S_{2,i} = j) = \begin{cases} 1 - \gamma_j & \text{if } k = 0 \\ \gamma_j \cdot (1 - \mu_j) \cdot (\mu_j)^{k-1} & \text{if } k \geq 1 \end{cases}$$

- Several extreme values: SEHM with shifted Zipf

- HMM and SEHM are competitive on explanatory power
- HMM outperforms SEHM in predictive power
- HMM approach is robust to missing data

ABRUPT CHANGES

- Organizational/Strategy changes in terrorist group
  - Group resilience
  - Level of coordination in group
- Increase in either leads to spurts in no. of attacks, but with different signatures in terms of activity profile
- Goal: Can such abrupt changes be detected and classified quickly?
- Two natural approaches for spurt detection
  - Exp. weighted moving average (EWMA)-based
  - State estimation using Viterbi algorithm

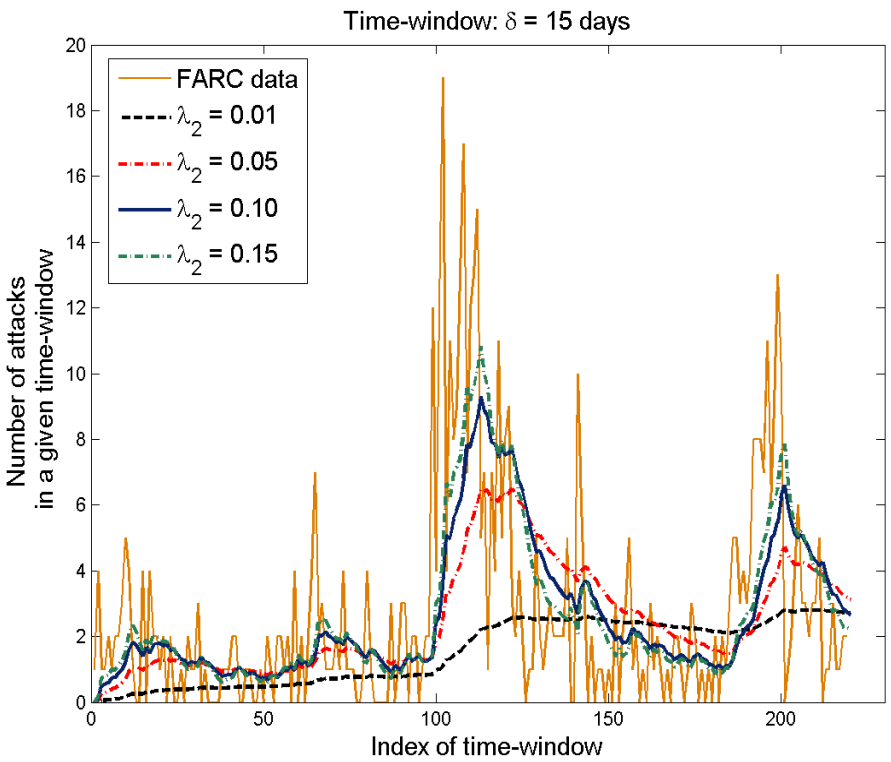
EWMA-BASED SPURT DETECTION

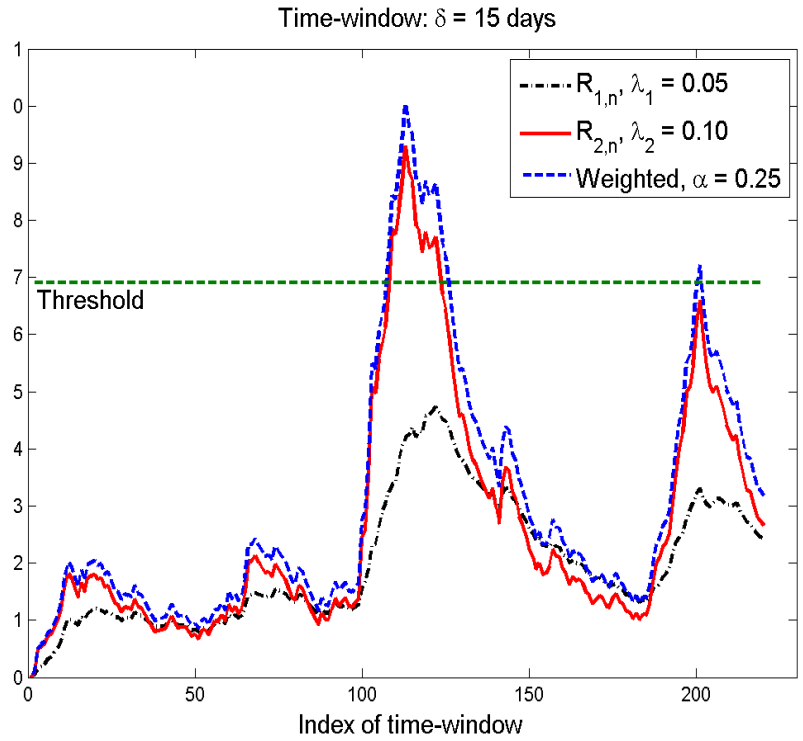
- Consider a time-window of δ days: $[(n-1)\delta + 1, \cdots, n\delta]\,,\ n = 1, 2, \cdots$
- Use no. of days of activity (X_n) and no. of attacks (Y_n)

$$R_{1,n} = (1-\lambda_1)R_{1,n-1} + \lambda_1 X_n$$
$$R_{2,n} = (1-\lambda_2)R_{2,n-1} + \lambda_2 Y_n$$

- $\lambda_1,\ \lambda_2$ are experimentally chosen to meet small FAR (typically small)

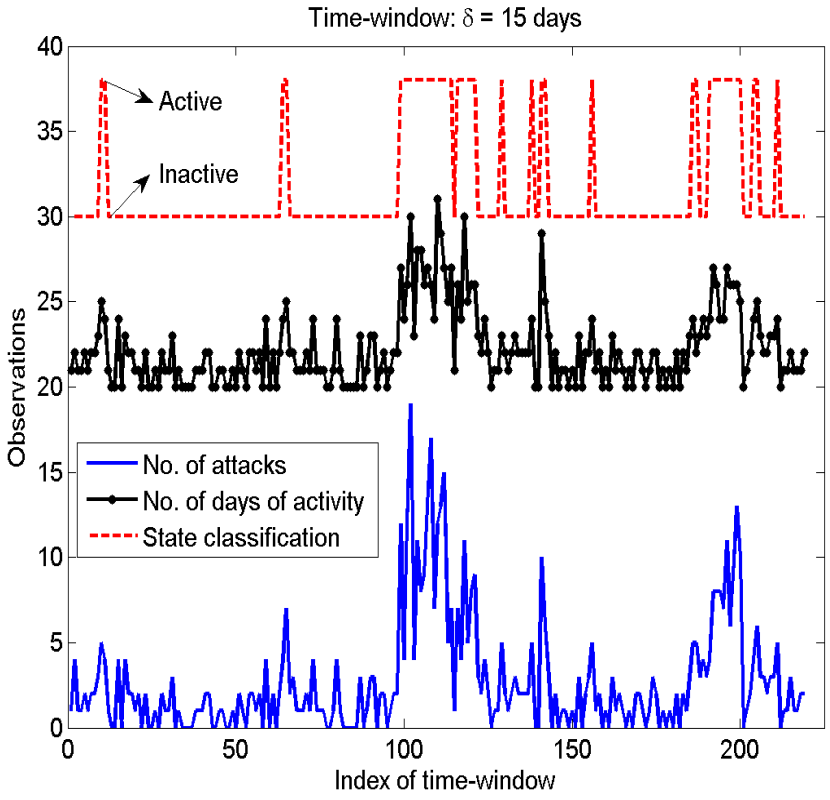
$$\tau_{\text{weighted}} = \inf\left\{n \geq 1\ :\ \alpha R_{1,n} + \sqrt{1-\alpha^2}R_{2,n} \geq A\right\}$$

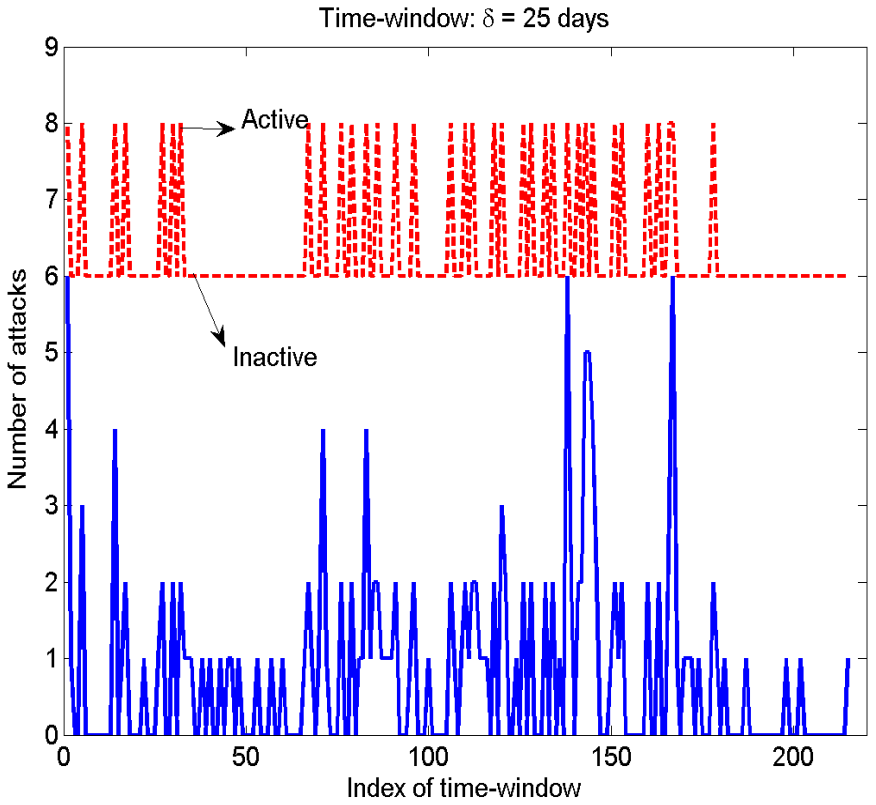




HMM STATE ESTIMATION

- Train: Learn HMM parameters (p_0, q_0, Active rate and Inactive rate) – Baum-Welch/EM algorithm
- Classify States: As Active/Inactive using Viterbi algorithm





EWMA vs. VA

- EWMA
  - Is oblivious of underlying distribution and robust
  - Detects persistent changes and tracks underlying process
  - But short moderate changes are not tracked
- Viterbi Algorithm
  - Is model-based and non-causal (both for training and state estimation)
  - Has good performance for state classification
- But neither approach can associate/link spurt with organizational changes/changes in strategy

- Resilience: Ability of group to launch repeated attacks over time
- Coordination: Ability of group to launch repeated attacks over geography
- IOW, if there are 25 attacks over 5 days with two different attack profiles ($\mathcal{A}_1 = [5, 5, 5, 5, 5]$ and $\mathcal{A}_2 = [25, 0, 0, 0, 0]$)
  - $\mathcal{A}_1$ suggests that the group is more resilient
  - $\mathcal{A}_2$ suggests that the group is better in coordination

MAJORIZATION THEORY

- Use majorization theory for event probability/attack frequency vector
- Majorization is a partial ordering on vectors with pos. entries and same sum
- Measures how one vector is more 'spread out' than the other
- Popular example: Gini index/income inequality
- Illustration: $\mathbf{M} \prec \mathbf{N}$ where

$$\mathbf{M} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right] \text{ and } \mathbf{N} = [1,\ 0,\ 0,\ 0]$$

- Caveat: Majorization is a partial order, not a complete order!
- Way out: Use the idea of Schur-concavity
- A function f(.) from R_N^+ to R is Schur-convex if

$$\mathbf{M} \prec \mathbf{N} \Longrightarrow f(\mathbf{M}) \leq f(\mathbf{N})$$

- If f(.) is Schur-convex, -f(.) is Schur-concave
- Examples:
  - Max function $f(\mathbf{M}) = \max_i \mathbf{M}_i$ is Schur-convex
  - Shannon entropy is Schur-concave: $f(\mathbf{M}) = -\sum_i \mathbf{M}_i \log(\mathbf{M}_i)$
- Under weak assumptions, certain Schur-convex functions can be used as a proxy for complete ordering

A PROXY FOR ORDERING

- Catalytic majorization (trumping): Let **M** and **N** be probability vectors. **M** is catalytically majorized by **N** if there exists **P** such that

$$\mathbf{M} \otimes \mathbf{P} \prec \mathbf{N} \otimes \mathbf{P}$$

$$\mathbf{M} \otimes \mathbf{P} = [\mathbf{M}_1\mathbf{P}_1, \cdots, \mathbf{M}_1\mathbf{P}_p, \mathbf{M}_2\mathbf{P}_1, \cdots, \mathbf{M}_2\mathbf{P}_p, \cdots, \mathbf{M}_m\mathbf{P}_1, \cdots, \mathbf{M}_m\mathbf{P}_p]$$

- Fact 1: The set of all majorizable prob. vectors is strictly contained in the set of all catalytically majorizable prob. vectors
- Fact 2 (Reverse catalytic majorization): Need just three functionals to "characterize" all catalytically majorizable vectors

Functional	**Schur-convexity**
$\mathrm{PM}(\mathbf{M}, \alpha) = \left(\sum_i \mathbf{M}_i^\alpha\right)^{1/\alpha}$	Schur-concave if $\alpha \in [0, 1]$ Schur-convex if $\alpha \leq 0$ or $\alpha \geq 1$
$\mathrm{SE}(\mathbf{M}) = -\sum_i \mathbf{M}_i \log(\mathbf{M}_i)$	Schur-concave
$\mathrm{GM}(\mathbf{M}) = \left(\prod_i \mathbf{M}_i\right)^{1/m}$	Schur-concave

- If two vectors satisfy ALL the correct inequalities corr. to the above functionals, the underlying vectors are catalytically majorizable

PROPOSED TEST

- No. of attacks over a time-window (Z_n)
- Shannon entropy: $X_n = \dfrac{\mathrm{SE}\left(\mathbf{M}|_{\Delta_n}\right)}{\frac{1}{\Delta}\sum_{i=1}^{\Delta}\mathrm{SE}\left(\mathbf{M}|_{\Delta_{n-i}}\right)}$
- Normalized power mean (NPM) = Power mean/No. of days of activity

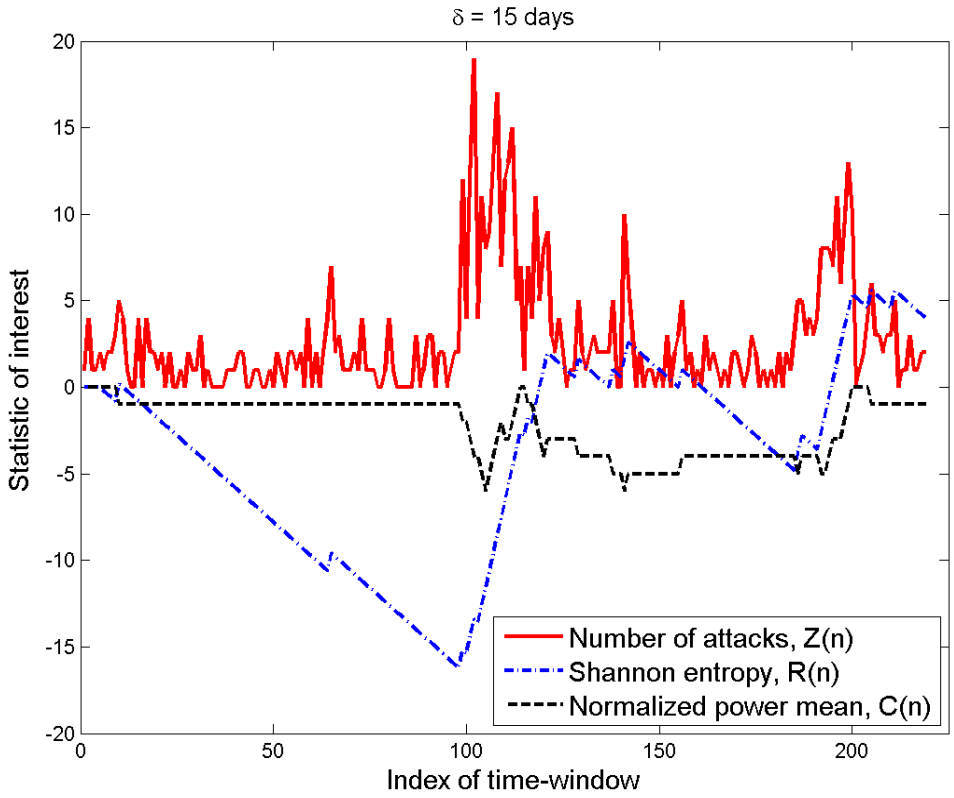
$$Y_n = \frac{\mathrm{NPM}\left(\mathbf{M}|_{\Delta_n}, \alpha\right)}{\frac{1}{\Delta}\sum_{i=1}^{\Delta}\mathrm{NPM}\left(\mathbf{M}|_{\Delta_{n-i}}, \alpha\right)}$$

$$R_n = R_{n-1} + \tau_{\mathcal{R}},\ n \geq 1,\ R_0 = 0$$

$$\begin{aligned}\tau_{\mathcal{R}} &= \mathbb{1}\left(X_n > \overline{\gamma}_{\mathcal{R}},\ Z_n > \tau\right) \\ &\quad -\mathbb{1}\left(X_n < \underline{\gamma}_{\mathcal{R}},\ Z_n > \tau\right) \\ &\quad -p_{\mathcal{R}} \cdot \mathbb{1}\left(Z_n \leq \tau\right)\end{aligned}$$

$$C_n = C_{n-1} + \tau_{\mathcal{C}},\ n \geq 1,\ C_0 = 0$$

$$\begin{aligned}\tau_{\mathcal{C}} &= \mathbb{1}\left(Y_n > \overline{\gamma}_{\mathcal{C}},\ Z_n > \tau\right) \\ &\quad -\mathbb{1}\left(Y_n < \underline{\gamma}_{\mathcal{C}},\ Z_n > \tau\right) \\ &\quad -p_{\mathcal{C}} \cdot \mathbb{1}\left(Z_n \leq \tau\right)\end{aligned}$$



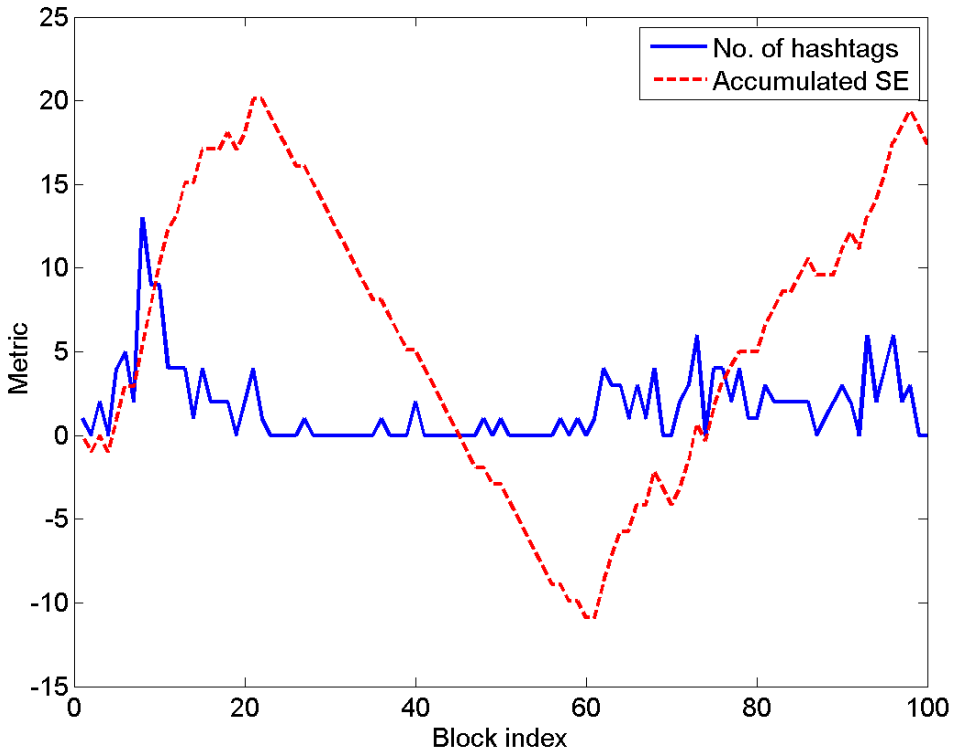
BURSTINESS DETECTION IN TWITTER

- Observations: No. of hashtags on a certain topic in a certain block of time $[(i-1)\delta + 1,\ i\delta]$ (say, 10 minutes) from all/relevant users
- Accumulated Shannon Entropy (SE)

$$\text{ASE}(i) = \text{ASE}(i-1) + \text{SE}(i)$$

$$\text{SE}(i) = \begin{cases} -1 & \text{if no hashtags} \\ -\sum_i p_i \log_2(p_i) & \text{if hashtags} \end{cases}$$

where $[p_1, \cdots, p_\delta]$ is the event probability vector



- Burst in topic interest is detected by change in slope from negative to positive
- Higher slope → more burstiness
- Shannon entropy metric is quick in detecting bursts, but more importantly non-parametric
- Useful in other applications also

CONCLUSIONS

- HMM-based model for terrorist activity is a good alternative modeling framework that is computationally advantageous
- Simple EWMA-based approach for spurt detection does not detect minor spurts
- VA classifies Active and Inactive states, but is non-causal and difficult to implement
- Proposed a simple majorization theory based framework that helps in detecting spurts as well as classifying them (resilience vs. coordination)
- Parametric approaches have fundamental difficulties in implementation