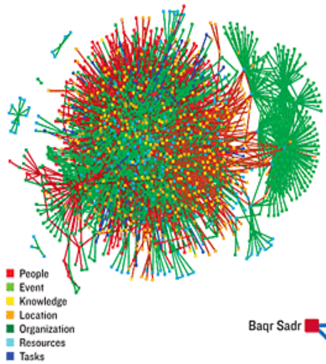
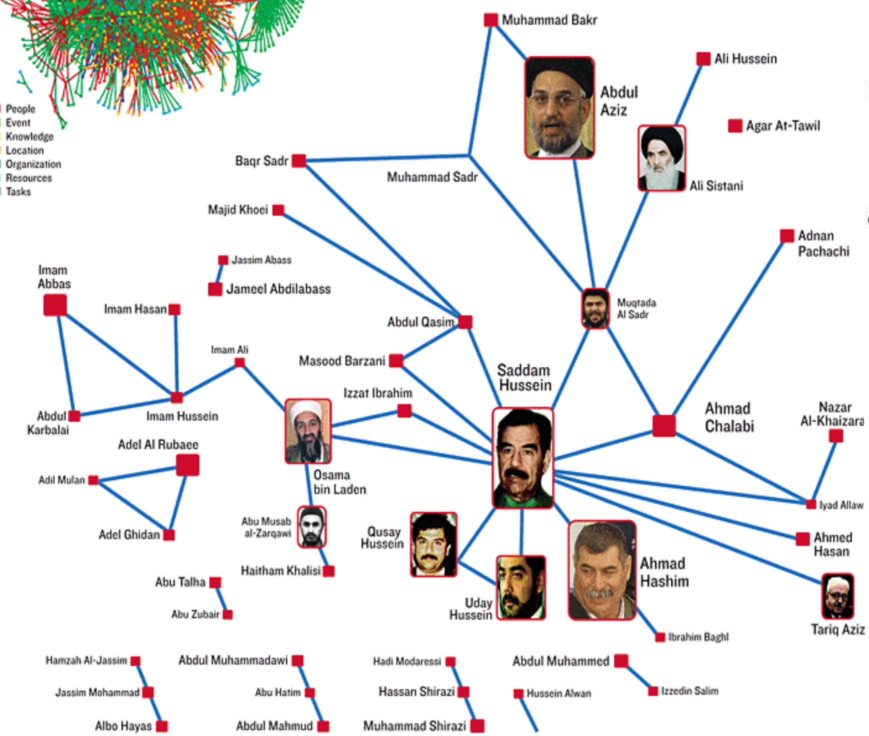
Terrorism trends via model learning and non-parametric approaches

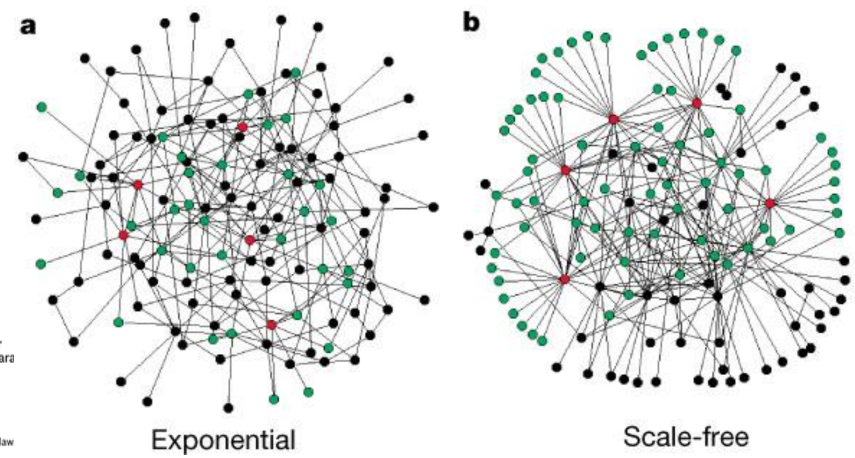
Vasanthan Raghavan
Qualcomm Flarion Technologies, Inc.
Bridgewater, NJ

October 26, 2016

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

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

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

  - 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 security checks, Reagan-era laws)

$$
\begin{aligned}
M_{1,i} &= \mathrm{a}_1 M_{1,i-1} + \mathrm{b}_1 M_{2,i-1} + \mathrm{c}_1 \mathrm{p}_1 + \text{Other comps.} \\
M_{2,i} &= \mathrm{a}_2 M_{2,i-1} + \mathrm{b}_2 M_{1,i-1} + \mathrm{c}_2 \mathrm{p}_1 + \text{Other comps.}
\end{aligned}
$$

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 from UMD START Center]

- **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

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

- 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]

A HMM FRAMEWORK FOR TERRORIST ACTIVITY

- **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)

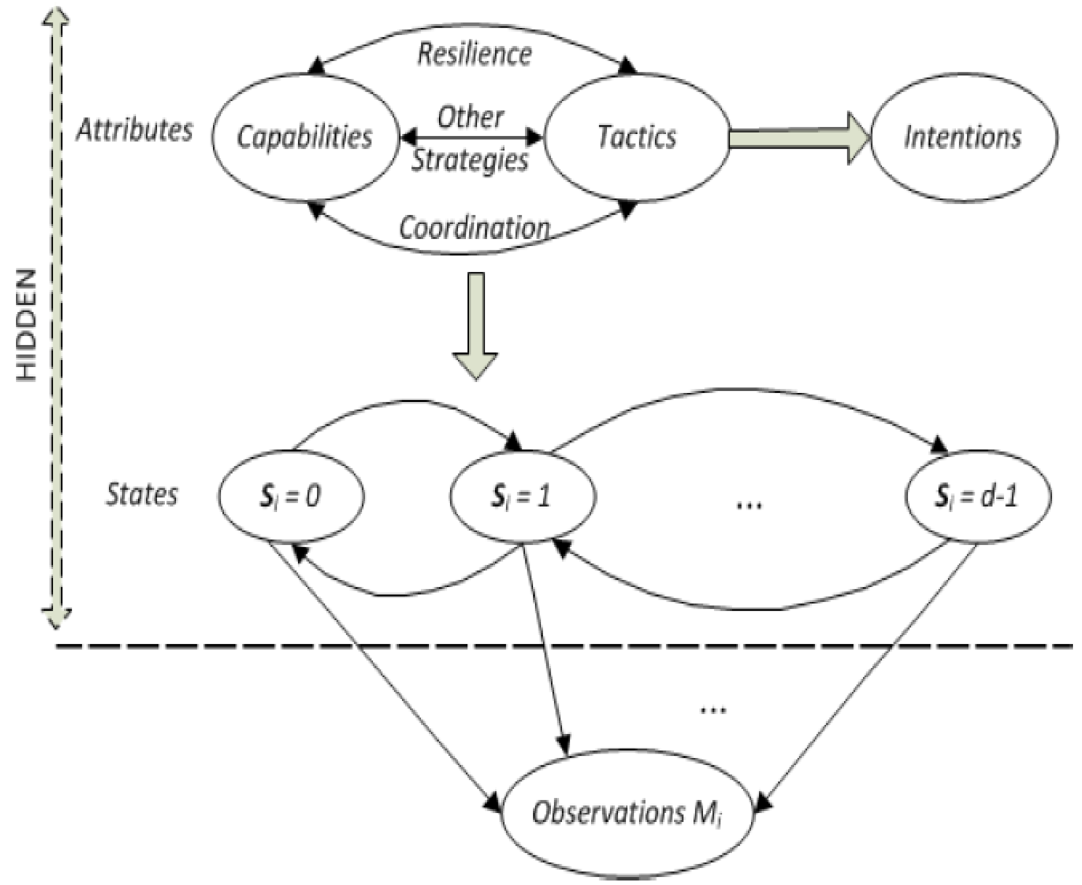
$$\mathsf{P}(M_i | \mathcal{H}_{i-1}, \mathbf{S}_i) = \mathsf{P}(M_i | \mathbf{S}_i), \; i = 1, 2, \cdots$$

- **Assumption 2:** These k dominant states include
  - The group's Intentions – Guiding ideology/philosophy (e.g., Marxist-Leninist-Maoist thought, political Islam), designated enemy group, nature of high profile attacks, nature of propaganda warfare, etc.
  - The group's Capabilities – 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)

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

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

A HMM FRAMEWORK FOR TERRORIST ACTIVITY

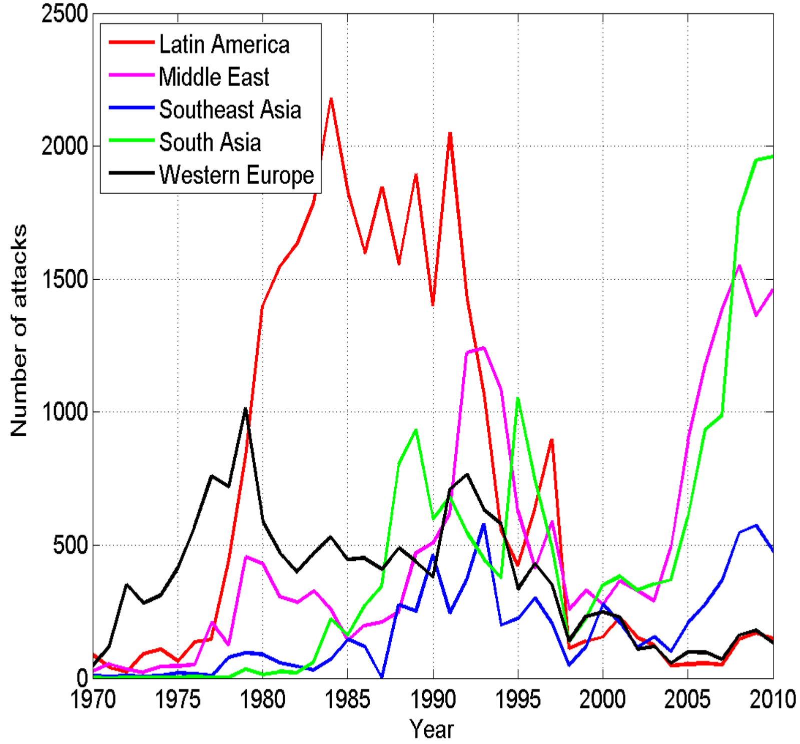


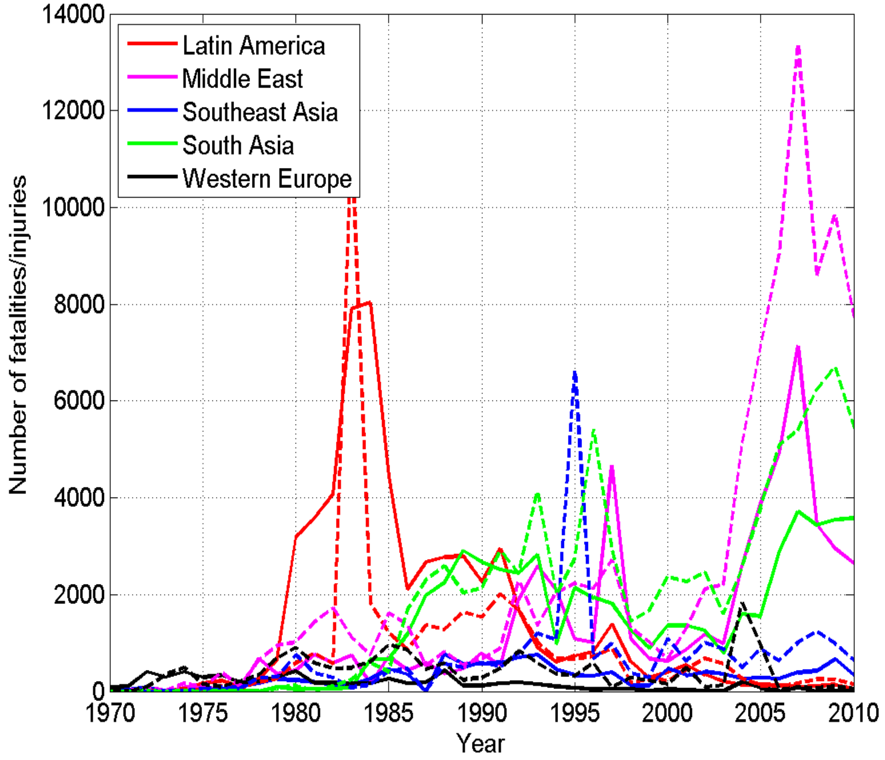
DATASET DESCRIPTION

- Data from 1970-2010 period from GTD/UMD START Center
- Missing data from 1993 substituted with data summary from GTD

- Data corresponding to five regions
  - Latin and South America – 28209 attacks
  - West Asia, North Africa and Central Asia – 19166 attacks
  - Southeast Asia, East Asia and Australasia – 6802 attacks
  - South Asia – 17727 attacks
  - Western Europe – 14701 attacks

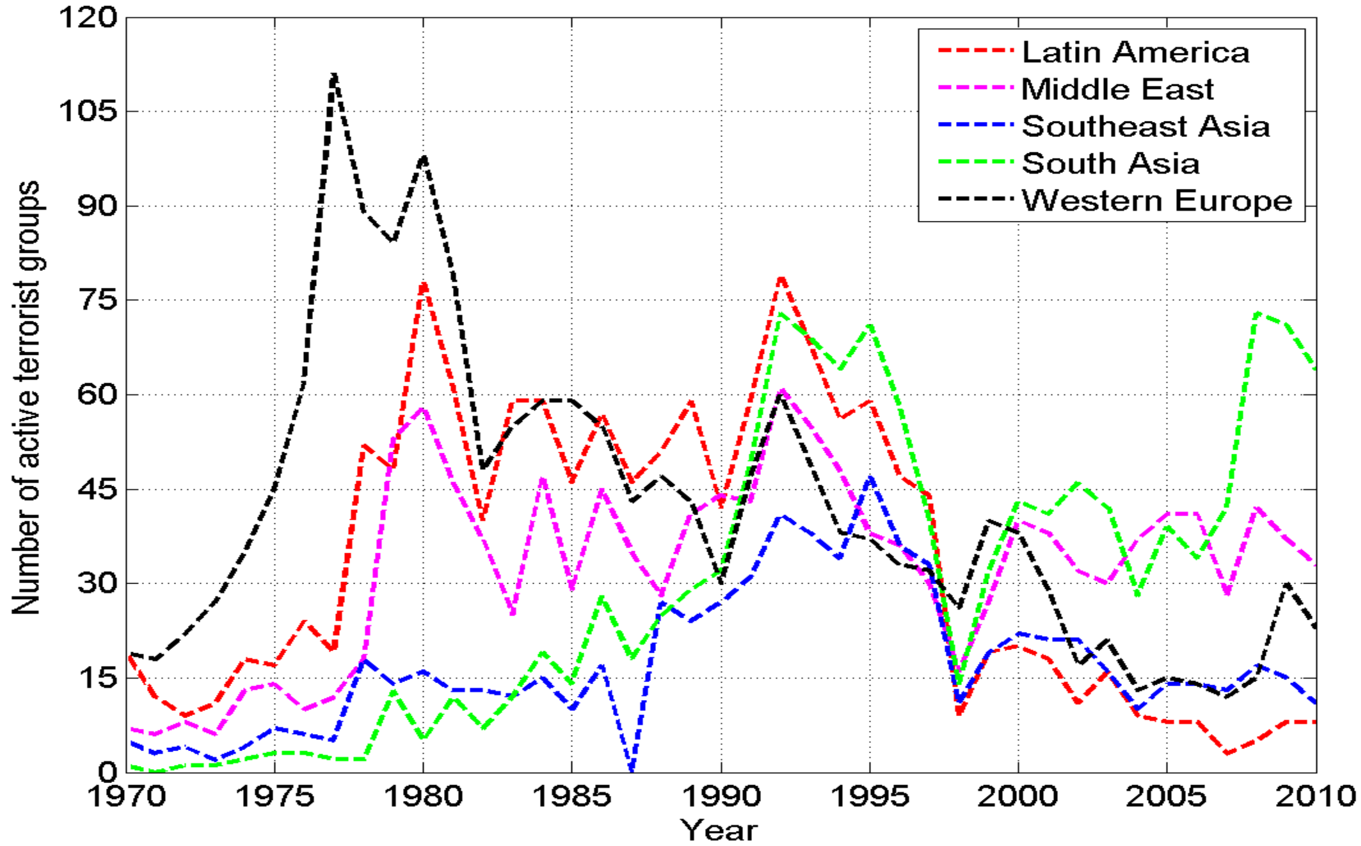
HOTSPOTS – I





- Broad correlation between no. of attacks and fatalities/injuries
  - WEU peaked in late 70s, LA in early 90s
  - SEA peaked in mid 90s and late 2000s
  - ME peaked in late 70s, mid 90s and mid 2000s
  - SA peaked in late 80s, mid 90s and late 2000s

HOTSPOTS – II



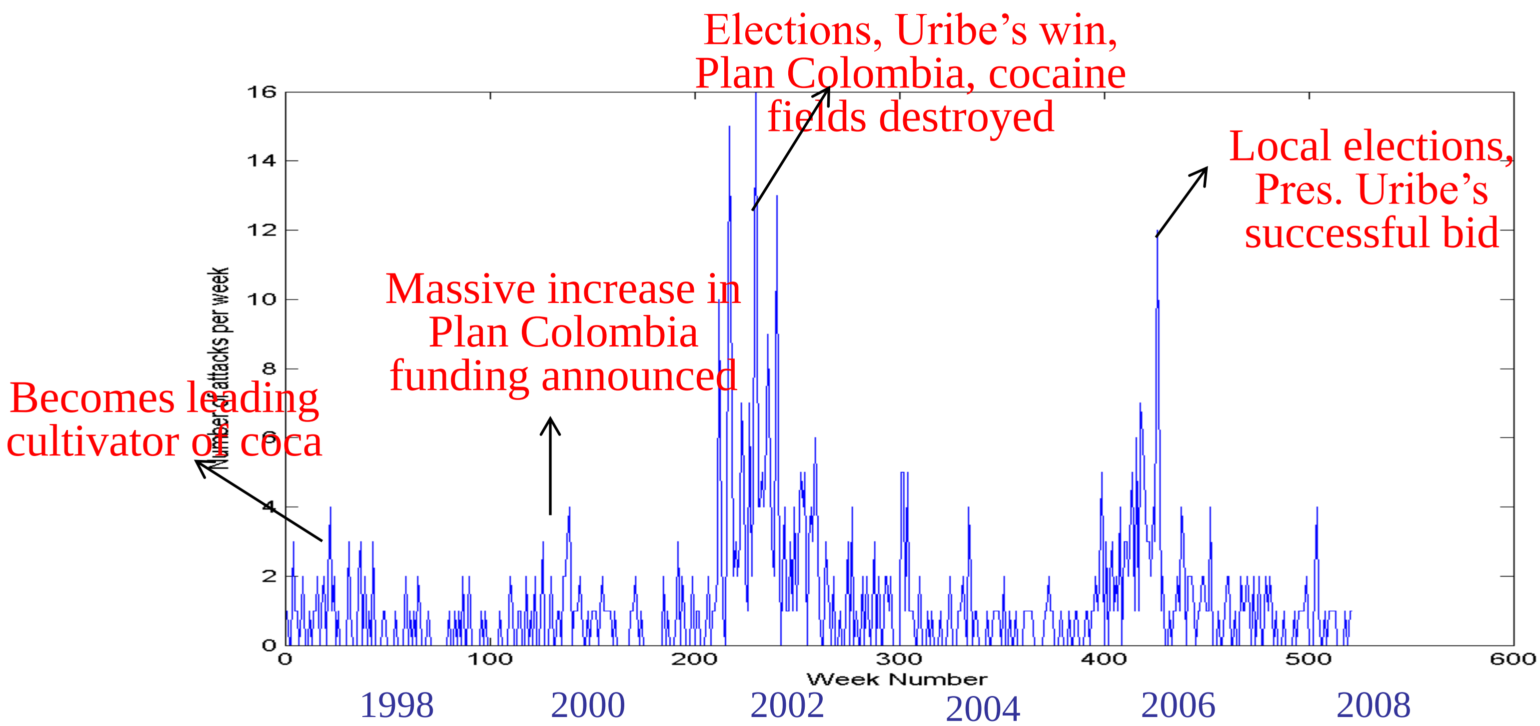
- Hotspots
  - WEU peaked in late 70s, LA in early 90s
  - SEA peaked in mid 90s and late 2000s
  - ME peaked in late 70s, mid 90s and mid 2000s
  - SA peaked in late 80s, mid 90s and late 2000s

A MORE DETAILED CASE STUDY: FARC

- Revolutionary Armed Forces of Colombia (FARC)
  - Oldest and largest terrorist group in the Americas, based in Colombia
  - Marxist-Leninist ideology, anti-establishmentist, uses guerilla warfare
  - Actively involved in cocaine cultivation and trans-shipment to U.S. and W. Europe, kidnapping rings, …

- Why FARC?
  - Dominant in Colombia → Less ambiguity in terms of other groups' attacks
  - Anti-establishment group → Strong signature in attack profile → Easy to differentiate FARC from non-FARC attacks in case of ambiguity

WHY FARC?



- Time-period of interest: 1998 – 2007, Why? Two key geo-pol events
  - Spurt 1
    - 1997: Colombia becomes leading cultivator of coca
    - 1999–2000: Plan Colombia with U.S. aid
    - 2001–2002: President Uribe's election on anti-FARC plank
  - Spurt 2
    - 2003–2004: Anti-FARC efforts bear fruit
    - 2005 – 2006: President Uribe's re-election bid and local elections

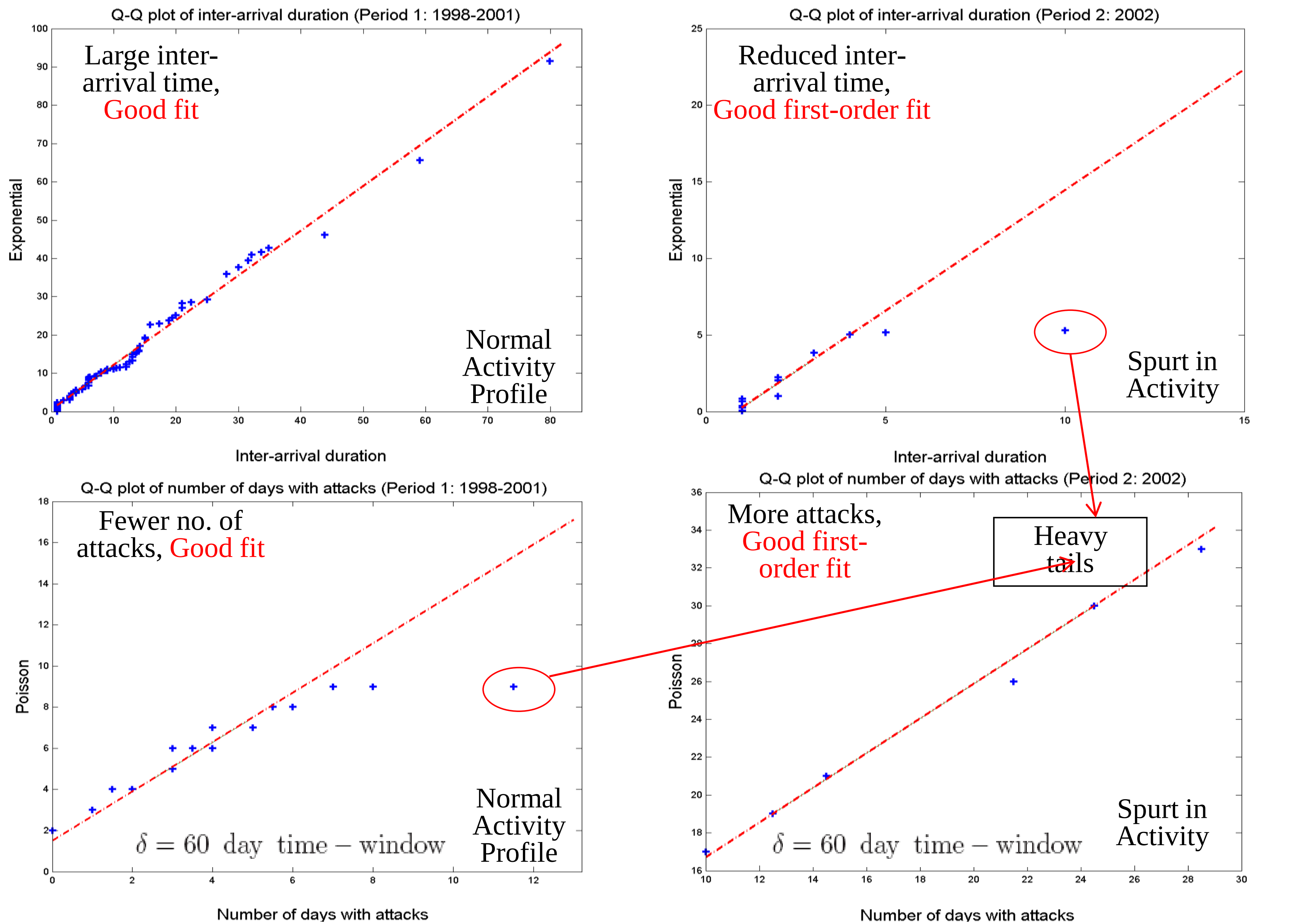
MODELS FOR FARC

Histogram of observed number of attacks per day for FARC data with different model-fits, $\delta = 15$ days

No. attacks (*Inactive* State)	Obs.	Poisson	Shifted Zipf	Geomet.	Pòlya	Hurdle-Based Zipf	Hurdle-Based Geomet.
0	2420	2421	2470	2430	2421	2420	2421
1	227	225	144	207	225	229	226
2	9	11	27	18	11	7	10
3	1	0	8	2	0	1	0
4	0	0	4	0	0	0	0
> 4	0	0	4	0	0	0	0
AIC		**1690.34**	1772.81	1696.74	1692.32	1692.58	1691.86
Parameter Estimate		0.0933	4.105	0.0854	$\widehat{r}_0 = 24.4749$, $\widehat{y}_0 = 0.0038$	$\widehat{\gamma}_0 = 0.0892$, $\widehat{y}_0 = 5.10$	$\widehat{\mu}_0 = 0.0444$, $\widehat{\gamma}_0 = 0.0892$

No. attacks (*Active* State)	Obs.	Poisson	Shifted Zipf	Geomet.	Pòlya	Hurdle-Based Zipf	Hurdle-Based Geomet.
0	384	359	455	404	389	384	384
1	174	202	87	144	160	189	171
2	46	57	33	52	56	31	52
3	19	11	16	19	17	11	16
4	4	1	9	7	6	5	5
> 4	3	0	30	4	2	10	2
AIC		1313.88	1416.88	1291.73	1288.85	1308.09	**1287.11**
Parameter Estimate		0.5651	2.40	0.3611	$\widehat{r}_1 = 1.4834$, $\widehat{y}_1 = 0.2759$	$\widehat{\gamma}_1 = 0.3905$, $\widehat{y}_1 = 2.61$	$\widehat{\mu}_1 = 0.3090$, $\widehat{\gamma}_1 = 0.3905$

MODEL VERIFICATION



LESSONS FROM MODEL LEARNING

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

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

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

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

- Several extreme values: SEHM with shifted Zipf is a better fit

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

TYPICAL ABRUPT CHANGES

- Organizational changes in terrorist group
  - Resilience of group
  - Level of coordination in group
- Different signatures in terms of activity profile
- Resilience has a less bursty signature, coordination has a more bursty signature

- Other applications
  - Sudden burstiness in a topic/hashtag on Twitter
  - Why is burstiness detection important?
    - Natural calamities (earthquakes)
    - Unexpected events (fire, snowstorm, armed person in campus/mall)
    - Epidemics (Google Flutrends, H5N1, meningitis)
    - Spread of panic (stock market crash, riots)
    - "Sense of social media" – Impact of political events/speech, election campaigns, policy announcements, etc.

- Goal: Can such abrupt changes be detected quickly?

SOME ASSUMPTIONS

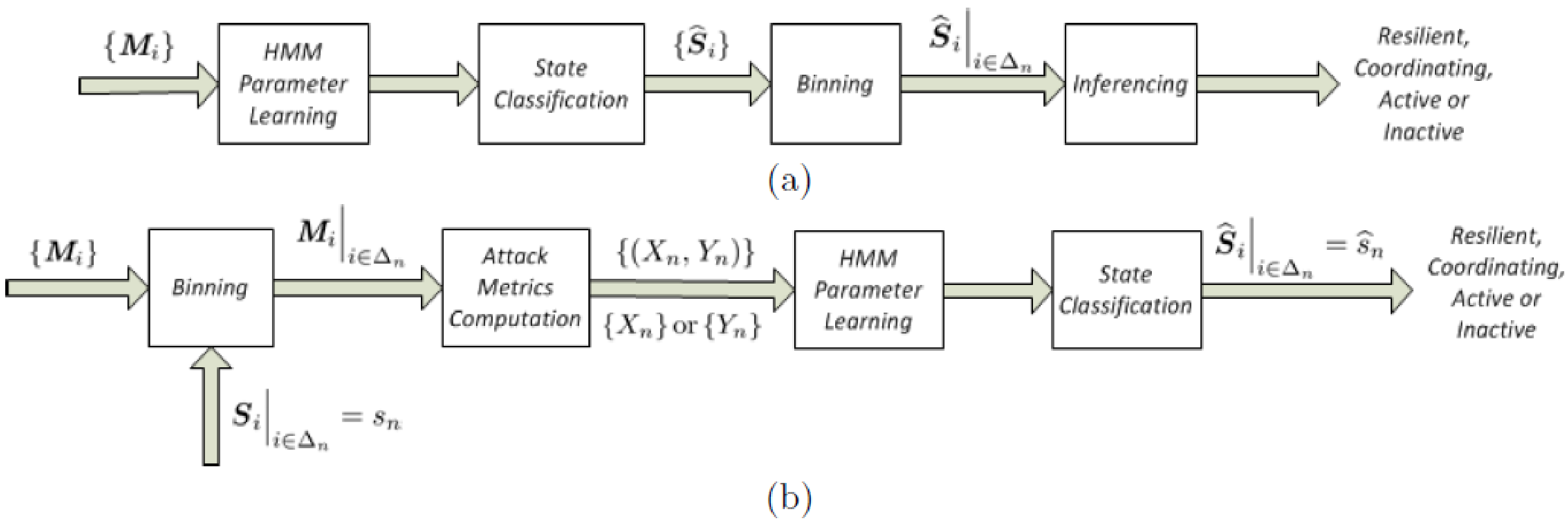
- Organizational changes in terrorist group
  - Resilience of group
  - Level of coordination in group
- Want to classify organizational behavior over a time-window Δ_n (week/fortnight/month etc., but not every day)
- An attack metric proxy for resilience is the number of days of attacks over Δ_n

$$X_n = \sum_{i \in \Delta_n} \mathbb{1}\left(\boldsymbol{M}_i > 0\right)$$

- An attack metric proxy for coordination is the number of attacks over Δ_n

$$Y_n = \sum_{i \in \Delta_n} \boldsymbol{M}_i$$

PARAMETRIC APPROACHES TO CLASSIFICATION



- Approach a:
  - Learn parameters with observations
  - Binary state classification
  - Binning and mapping to resilience and coordinating states
- Approach b:
  - Bin observations to form attack metrics
  - Learn parameters with attack metrics
  - Binary state classification and mapping to resilience and coordinating states

PROBLEMS WITH PARAMETRIC APPROACHES

- Terrorism is “rare” from a model learning perspective
  - For FARC, 641 incidents over a 10 year period ~ 1.23 incidents per week
  - Similar trends across almost all the groups in GTD
- Learning a 4 parameter HMM could need approx. 4 * 100/1.23 ~ 325 weeks ~ 6 ¼ years

- Models capture some underlying dynamic of group
  - Model stability issues
  - Inferencing on the short time-horizon?

- HMM learning and state classification is non-causal/retrospective
  - Applications in online decision-making?

NON-PARAMETRIC APPROACH TO CLASSIFICATION

- Approach based on majorization theory
- Majorization provides a partial ordering for probability vectors
- We use a reverse majorization theory for better than partial ordering

THEOREM 4.1. *Let $\{\underline{\boldsymbol{P}}, \underline{\boldsymbol{Q}}\} \in \mathbb{P}_\delta$. In one of two possibilities, $\underline{\boldsymbol{P}}$ and $\underline{\boldsymbol{Q}}$ are not comparable with each other in the form of a catalytic majorization relationship. In the other possibility, their comparability is verified by checking an equivalent set of conditions over only two types of functions:*

$$
\begin{array}{rl}
\text{i)} & \mathsf{PM}(\underline{\boldsymbol{P}}, \alpha) < \mathsf{PM}(\underline{\boldsymbol{Q}}, \alpha) \ \text{ if } \alpha > 1, \\
\text{ii)} & \mathsf{PM}(\underline{\boldsymbol{P}}, \alpha) > \mathsf{PM}(\underline{\boldsymbol{Q}}, \alpha) \ \text{ if } \alpha < 1, \ \text{and} \\
\text{iii)} & \mathsf{SE}(\underline{\boldsymbol{P}}) > \mathsf{SE}(\underline{\boldsymbol{Q}}).
\end{array}
$$

In the above equations, $\mathsf{SE}(\cdot)$ *and* $\mathsf{PM}(\cdot, \alpha)$ *stand for the Shannon entropy function and the power mean function corresponding to an index α, and are defined as,*

$$
\mathsf{SE}(\underline{\boldsymbol{P}}) \triangleq -\sum_{i=1}^{\delta} \boldsymbol{P}(i) \log\left(\boldsymbol{P}(i)\right), \qquad \mathsf{PM}(\underline{\boldsymbol{P}}, \alpha) \triangleq \left(\frac{\sum_{i=1}^{\delta} \boldsymbol{P}(i)^\alpha}{\sum_{i=1}^{\delta} \mathbb{1}(\boldsymbol{P}(i) > 0)} \right)^{1/\alpha}.
$$

$\square$

APPLICATION TO BURSTINESS DETECTION

- Define an attack frequency vector

$$\boldsymbol{P}_n(i) = \begin{cases} \frac{\boldsymbol{M}_{(n-1)\delta+i}}{\sum_{j\in\Delta_n} \boldsymbol{M}_j} & \text{if } \sum_{j\in\Delta_n} \boldsymbol{M}_j > 0, \\ 0 & \text{otherwise} \end{cases}$$

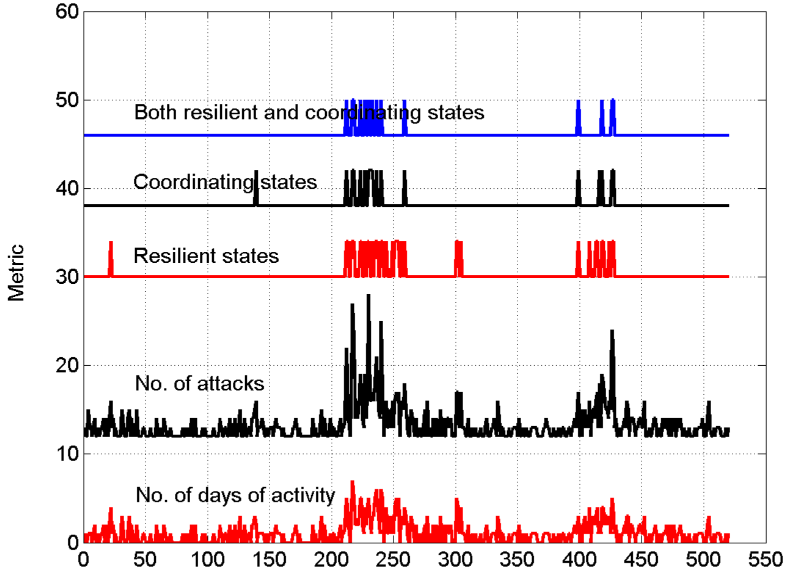
- Define two metrics
  - Shannon entropy
  - Normalized power mean with a fixed power index

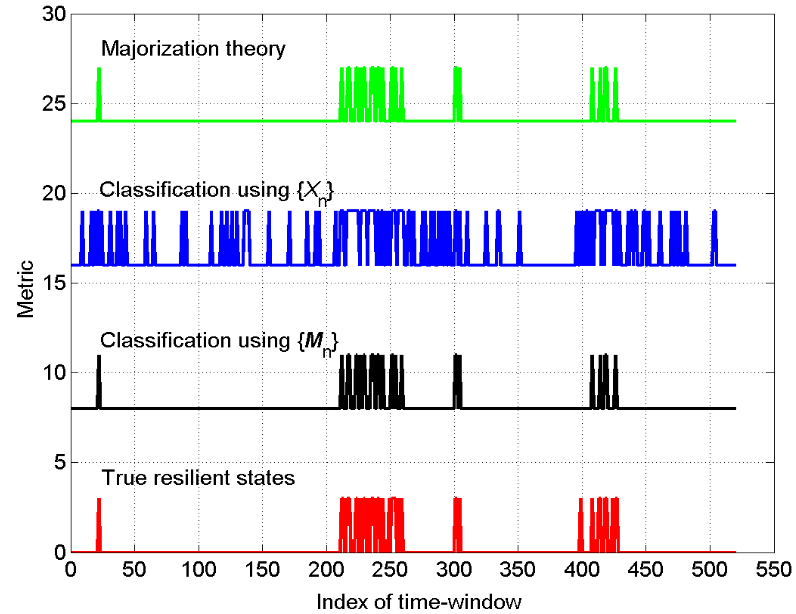
$$\begin{aligned} \mathsf{SE}(\underline{\boldsymbol{P}}_n) &= \log\left(\sum_{i\in\Delta_n} \boldsymbol{M}_i\right) - \frac{\sum_{i\in\Delta_n} \boldsymbol{M}_i \log(\boldsymbol{M}_i)}{\sum_{i\in\Delta_n} \boldsymbol{M}_i} \\ \mathsf{NPM}(\underline{\boldsymbol{P}}_n, \alpha^\star) &= \frac{\left(\sum_{i\in\Delta_n} (\boldsymbol{M}_i)^{\alpha^\star}\right)^{1/\alpha^\star}}{\left(\sum_{i\in\Delta_n} \boldsymbol{M}_i\right) \cdot \left(\sum_{i\in\Delta_n} \mathbb{1}\left(\boldsymbol{M}_i > 0\right)\right)^{1+1/\alpha^\star}}, \end{aligned}$$

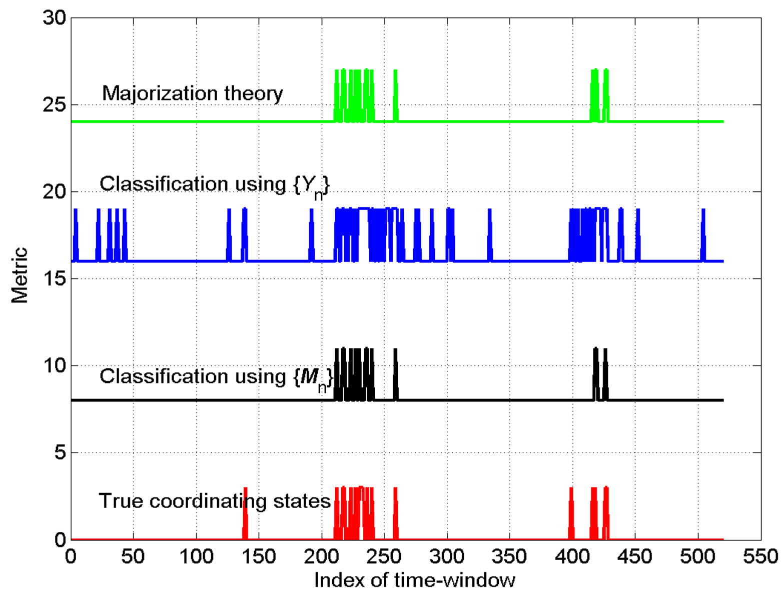
- Resilience and coordination classification

$$\begin{aligned} \text{Resilient} &\iff \mathsf{SE}(\underline{\boldsymbol{P}}_n) > \underline{\mathsf{SE}} \text{ and } X_n > \widetilde{\eta}_{\mathsf{X}} \\ \text{Coordinating} &\iff \mathsf{NPM}(\underline{\boldsymbol{P}}_n, \alpha^\star) > \underline{\mathsf{NPM}} \text{ and } Y_n > \widetilde{\eta}_{\mathsf{Y}} \end{aligned}$$

FARC EXAMPLE







FARC data

Setting	Parameters	Number of states classified and (P_{MD}, P_{FA})		
		Resilient	Coordinating	Both
True Observations	–	37	18	14
Learning with $\{M_i\}$	$\hat{\gamma}_0 = 0.0953, \hat{\mu}_0 = 0.0762$ $\hat{\gamma}_1 = 0.3988, \hat{\mu}_1 = 0.3087$	27 $(0.2703, 0)$	13 $(0.3889, 0.1538)$	13 $(0.2143, 0.1538)$
Learning with $\{X_n\}$	$\hat{\gamma}_0 = 0.0933, \hat{\mu}_0 = 0.3505$ $\hat{\gamma}_1 = 0.3921, \hat{\mu}_1 = 0.3505$	125 $(0, 0.7040)$	–	–
Learning with $\{Y_n\}$	$\hat{\gamma}_0 = 0.0951, \hat{\mu}_0 = 0.1232$ $\hat{\gamma}_1 = 0.2500, \hat{\mu}_1 = 0.5745$	–	73 $(0, 0.7534)$	–
Learning with $\{(X_n, Y_n)\}$	$\hat{\gamma}_0 = 0.0949, \hat{\mu}_0 = 0.0752$ $\hat{\gamma}_1 = 0.3958, \hat{\mu}_1 = 0.3082$	–	–	73 $(0, 0.8082)$
Majorization theory	–	27 $(0.2703, 0)$	15 $(0.2778, 0.1333)$	13 $(0.2143, 0.1538)$

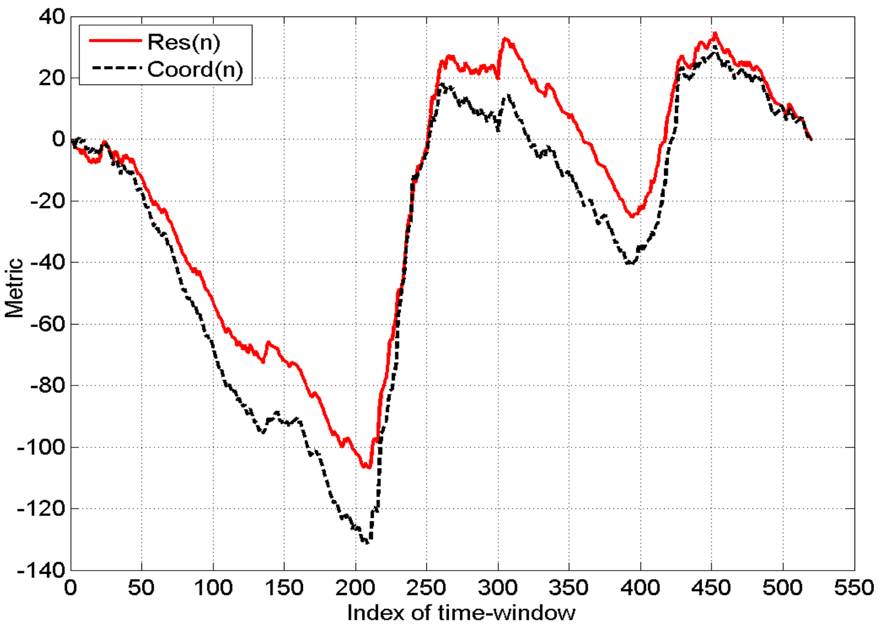
TRACKING RESILIENCE/COORDINATION

- Resilience and coordination classification

$$\begin{aligned}\text{Resilient} &\iff \mathsf{SE}(\underline{\boldsymbol{P}}_n) > \underline{\mathsf{SE}} \text{ and } X_n > \widetilde{\eta}_{\mathsf{X}} \\ \text{Coordinating} &\iff \mathsf{NPM}(\underline{\boldsymbol{P}}_n, \alpha^\star) > \underline{\mathsf{NPM}} \text{ and } Y_n > \widetilde{\eta}_{\mathsf{Y}}\end{aligned}$$

- Tracking functions

$$\begin{aligned}\mathsf{Res}(n) &= \mathsf{Res}(n-1) + \mathsf{SE}(\underline{\boldsymbol{P}}_n) + X_n - \frac{\sum_{n'=1}^{N_{\mathsf{max}}} \left(\mathsf{SE}(\underline{\boldsymbol{P}}_{n'}) + X_{n'}\right)}{N_{\mathsf{max}}} \\ \mathsf{Coord}(n) &= \mathsf{Coord}(n-1) + \mathsf{NPM}(\underline{\boldsymbol{P}}_n, \alpha^\star) + Y_n - \frac{\sum_{n'=1}^{N_{\mathsf{max}}} \left(\mathsf{NPM}(\underline{\boldsymbol{P}}_{n'}, \alpha^\star) + Y_{n'}\right)}{N_{\mathsf{max}}}\end{aligned}$$



KEY CONCLUSIONS

- Model learning is good to learn about what the group's behavior looks like in a very broad sense
- But it is a poor way forward for online/short-term detection/classification etc.
- Non-parametric approaches can be better if the metric is appropriately chosen for tracking
  - Low miss detection and low false alarm
  - Parametric approaches often result in high false alarms

[R, Galstyan & Tartakovsky, Annals of Applied Statistics, 2014]

[R & Tartakovsky, ArXiv 1604.02051]