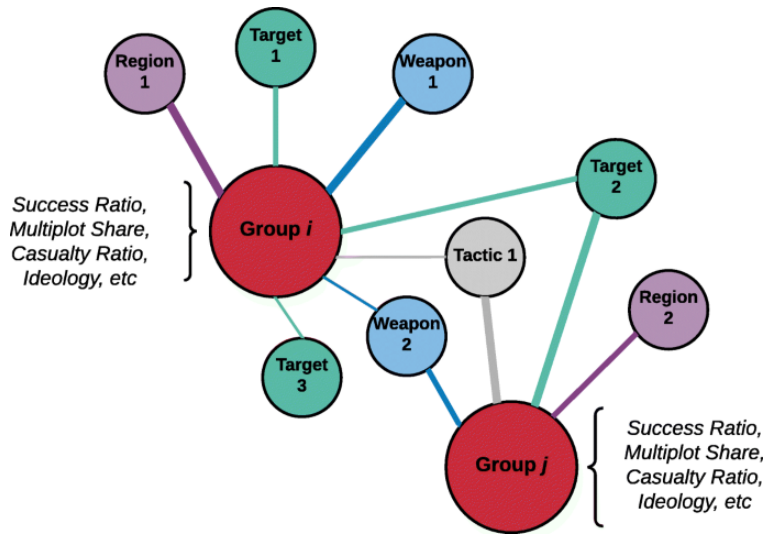
Statistical Analysis of Terrorist Attacks

Burak Varıcı

Rensselaer Polytechnic Institute

November 11, 2021

- Purpose of data/network science?
- Predicting lethality
- Predicting targets
- Investigating causes: immigration



Quantifying the Future Lethality of Terror Organizations

Yang, Y., Pah, A. R., & Uzzi, B. (PNAS, 2019)

- Estimating future **lethality** of individual terror groups. Why?
- Non-lethal groups: lower activity levels (2 attacks and 8 months lifespan on average).
- Lethal groups: up to 15 years and 150 attacks on average + psychological trauma.

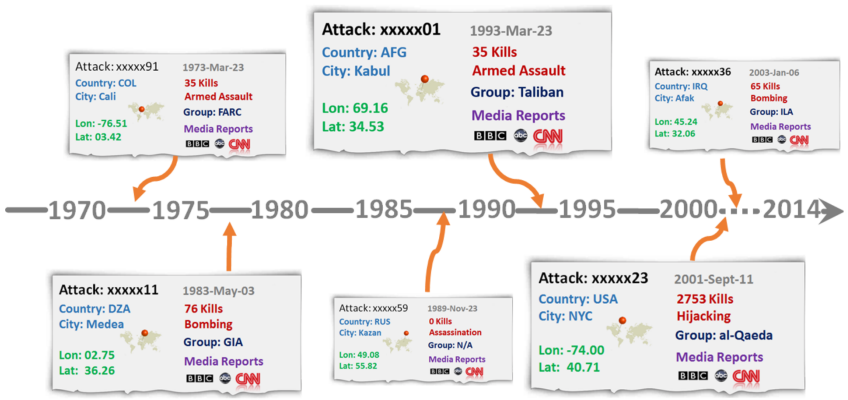
Quantifying the Future Lethality of Terror Organizations

- Prior research findings:
  - Negative correlation: locale's GDP, democratic governance
  - Positive correlation: population size, proximity to war zone
  - Attacks distribution: power law but unrelated to group size
- **Useful**: sociopolitical context and attack severity for a region.
- **More lethal groups within a region?**
- Iraq, Syria, Afghanistan: limited resources, many terror groups.

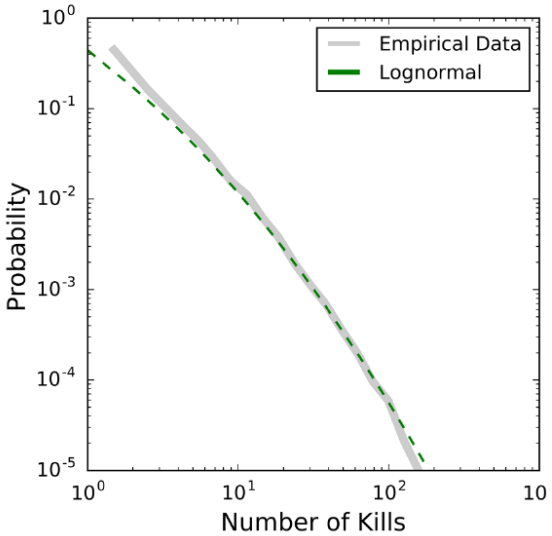
- Organization's performance: capabilities and resources.
- Covert organizations: similar patterns with hidden features.
- **Hypothesis:** Observed variables can be used to estimate capabilities and resources for lethality via MLE.
- Develop two parameters, data reliability, model robustness, and sensitivity tests.

Data

- Global Terrorism Database (GTD): Terrorist events around the world (1970-2014), $>$ 140.000 incidents.
- $\geq$ 10 attacks, $\geq$ 1 year $\rightarrow$ 342 groups (157 attacks and 14.8y average).
- Rand Database on Worldwide Terrorism Incidents (RDWTI): Similar events (1968-2009). 84 groups.
- GDP, population, Gini index, Democracy index, per capita income, government capability and constraint.



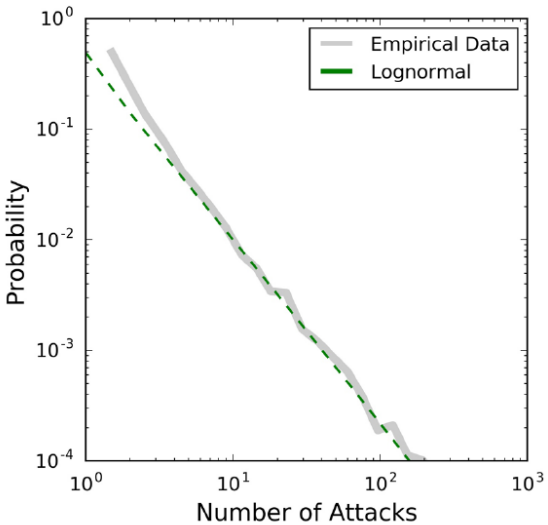
- ".. organizational assets relatively stable .."
- Sustain lethality: Know-how in armed attacks, cultural-religious principles ..
- Estimate relationship between Capabilities for group i, and number of fatalities in attack α.



Distribution of lethality is lognormal $\rightarrow$ multiplicative $\rightarrow$ construct

$$d_{i,\alpha} = Q_{5,i} p_\alpha$$

- If capabilities drive lethality $\rightarrow$ drives number of attacks
- Estimate joint distribution of Q_5, p(noise), N.
- Q_5 uncorrelated with lifespan
- p_α invariant across different decades and continents.

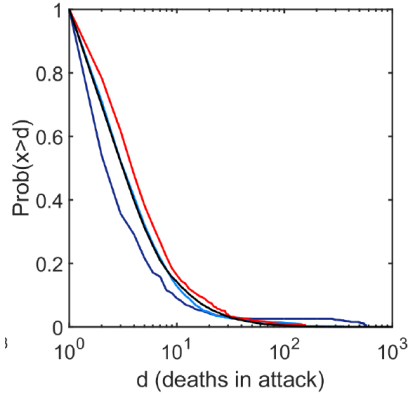


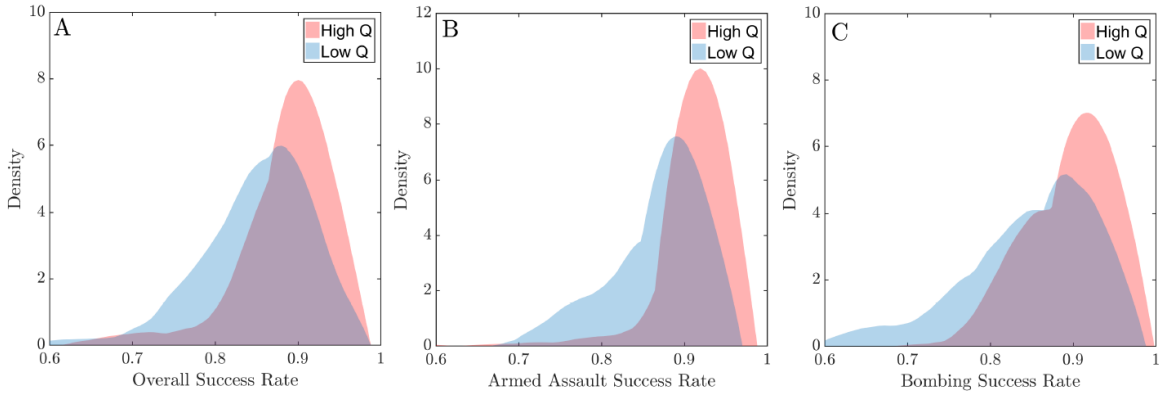
Maximum likelihood estimation to fit and population mean to adjust

$$Q_{5,i} = e^{\langle \log(d_{i,\alpha}) - \mu_{\hat{p}} \rangle}$$

$$\mu_{\hat{p}} = \langle \log(d) \rangle$$

- How do we know Q_5 actually relates to *capabilities?*
- Rescaling the distribution of lethality.
- Higher $Q_5 \leftrightarrow$ higher success





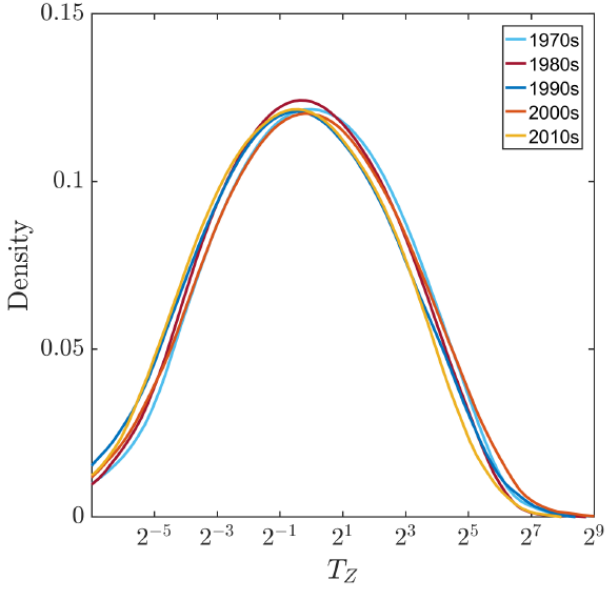
- Fluctuates over time (funding, leadership, intelligence etc.)
- **Timing pattern of attacks!** More resources $\rightarrow$ attacks become temporarily concentrated $\rightarrow$ distinguishable from randomized.
- Temporal normalization: first attack 0.0 - last attack 1.0

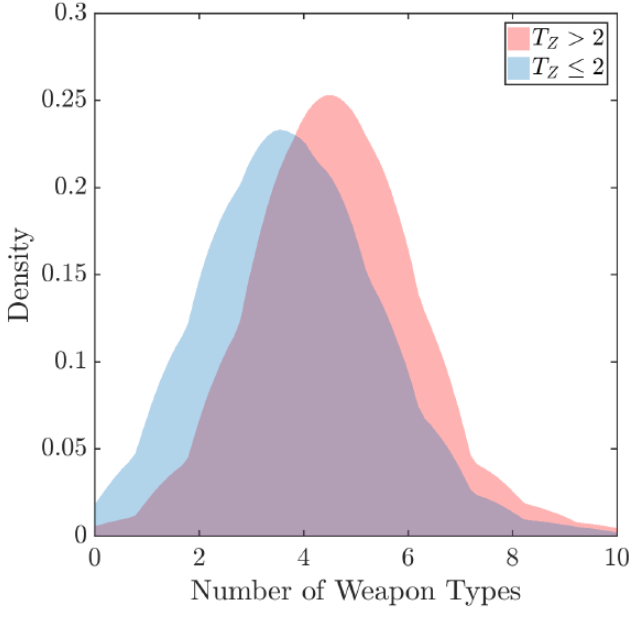
$$\mathrm{t} = \frac{\sigma_\tau - \mu_\tau}{\sigma_\tau + \mu_\tau} : \quad \text{how erratic a group's interevent timing}$$

- Generate 10^5 synthetic attack patterns. *Whether the real observed timing is different than the expected by chance?*

$$\mathrm{T_Z} = \left| \frac{t_{obs} - \bar{t}_{sim}}{std(t_{sim})} \right|$$

- Strong correlation with diverse weaponry. **Sign of resources.**
- Similar distributions across decades.
- No correlation between timing frequency of attacks across groups.





- Regression with sociopolitical variables, baseline (country and decade fixed effects, kills in first 5 attacks), Q_5, and T_Z.

- R^2: explained variance or coefficient of determination.

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$

	SocPol	Base	Q_5	Q_5+	T_Z	T_Z+	$Q_5\,T_Z$	$Q_5\,T_Z+$
R^2	0.11	0.37	0.11	0.41	0.10	0.45	0.19	0.47

- **Early-warning predictions.**
- Construct T_Z: first (i) 10, 20, 30, (ii) $10\%, 20\%, 30\%$ attacks.
- Recall full data: $R^2 = 0.47$.

	10%	20%	30%	10	20	30
R^2	0.44	0.46	0.40	0.34	0.29	0.34

- Parameters can be derived to reflect level of hidden capabilities and resources. Robust across location and decades.
- Additive and complementary: improve baseline models.
- Successful early signaling. Proactive approach.
- How other factors systemically correlate with Q_5 and T_Z? opposing forces, locale-based factors, media etc.
- Certain examples would have helped the discussion (e.g., comparison of two groups with similar early lethality).
- How to estimate for the next attack?

Learning future terrorist targets through temporal meta-graphs

Campedelli, G. M., Bartulovic, M., & Carley, K. M. (Nature Scientific Reports 2021)

- Terrorism: high level of uncertainty and unpredictability.
- Shine light on **high risk targets** $\rightarrow$ prevention mechanisms.
- **Prior work:** over the years/decades, how terrorist behaviors/strategies change?
- **Preceding paper:** Complex networks for terrorist target prediction (2018).
- Temporal meta-networks: (such as event x event) compute weighted adjacency matrices
- Network-based inference algorithm: rather poor results, but novel approach.

Learning future terrorist targets through temporal meta-graphs

- Terrorism: self-excitability and self-propagation.
- Operational recurring patterns.
- Treating events as uni-dimensional $\rightarrow$ understates the complexity.
- Exploit the **temporal multi-dimensionality** of the hidden connections.
- 3 dimensions of an attack: tactics, weapons, and targets.
- GTD data for Afghanistan and Iraq (2001-2018).

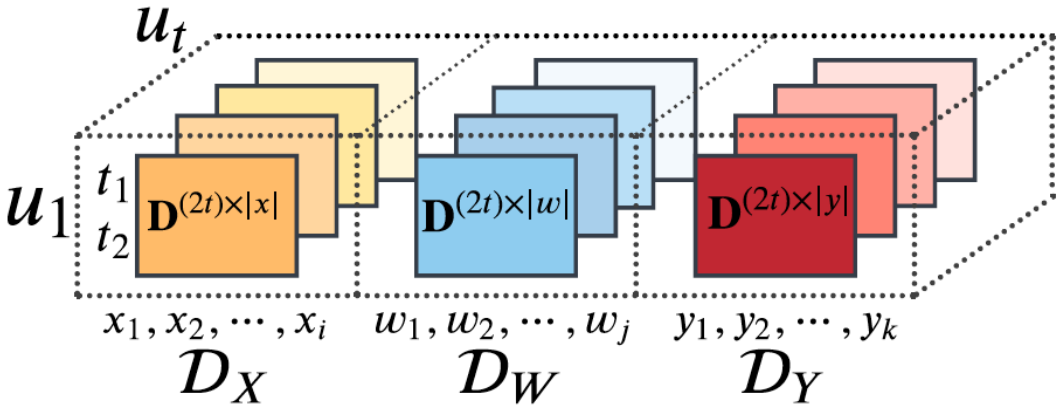
Country/ dataset	Original N of attacks	Filtered N of attacks	Daily average	Daily St. Dev.	Daily Min.	Daily Max.	N of targets	N of weapons	N of tactics
Afghanistan	14,371	12,106	1.841	2.722	0	68	18	5	9
Iraq	25,886	22,764	3.462	4.670	0	107	20	5	8

- Create U temporal slices for each dimension with 2-day units.

$$D^C = \{D_X, D_W, D_Y\}$$

- Construct meta-graph: number of times every pair of features has been connected

$$G_I[u] = D_I[u]^\top D_I[u], \quad I \in \{X, Y, Z\}$$



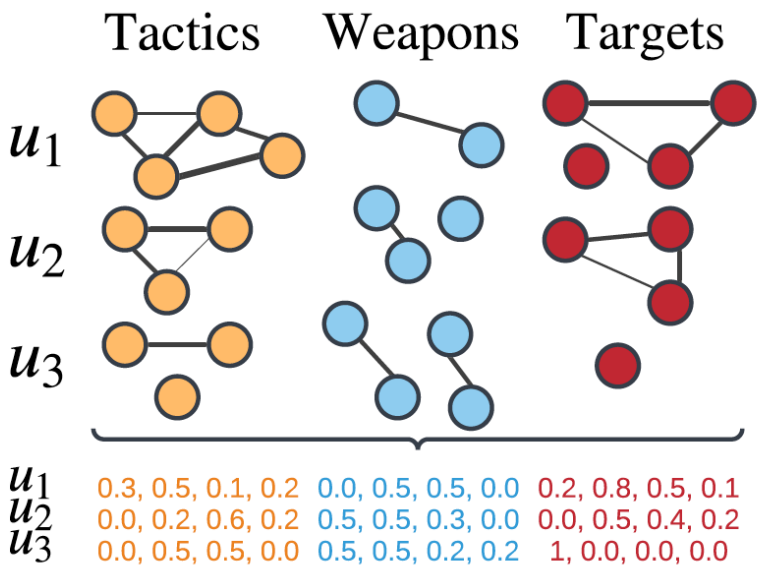
- Weighted degree centrality: How prevalent was a specific feature compared to the others.

- *Normalization:* Compare features across time units with high variation in activity.

$$\psi_i[u] = \sum_{i,\, j \neq i}^{|I|} G_{i,j}[u]$$

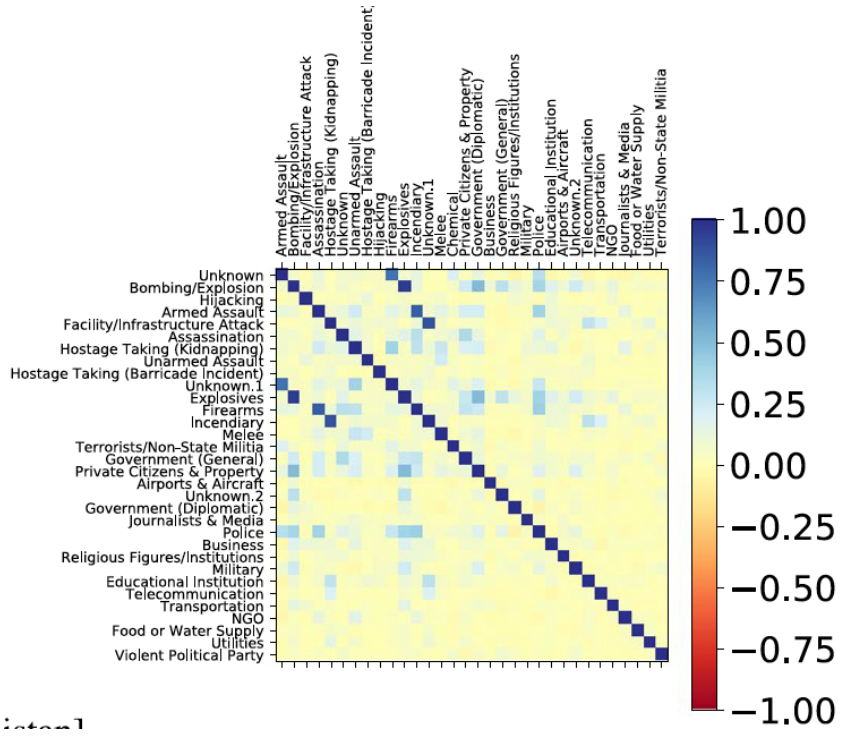
$$\psi_{i,\text{norm}}[u] = \frac{\psi_i[u]}{\max_{i \in I} \psi_i[u]}$$

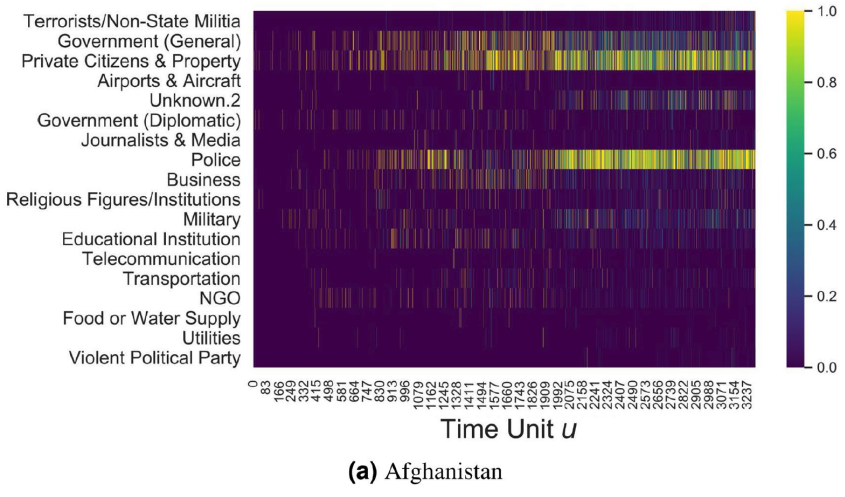
$$\Psi[U] = \{\psi_{i,\text{norm}}[u]\}_{u=0}^{|U|}, \quad i \in [|I|].$$



Learning future terrorist targets through temporal meta-graphs

- Correlations and centrality for targets: Afghanistan example

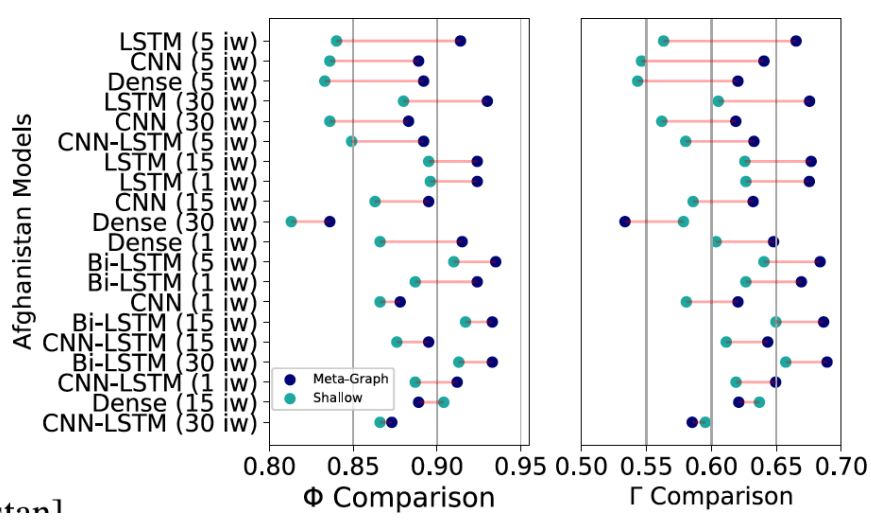


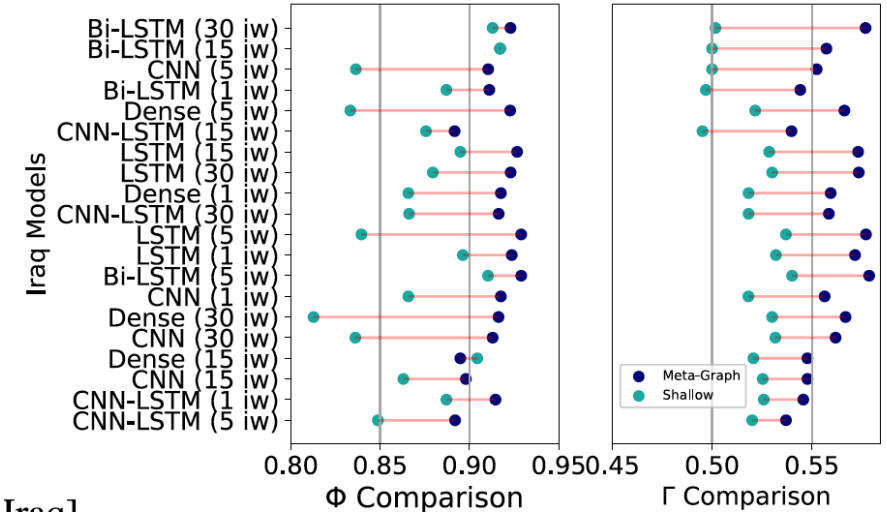


(a) Afghanistan

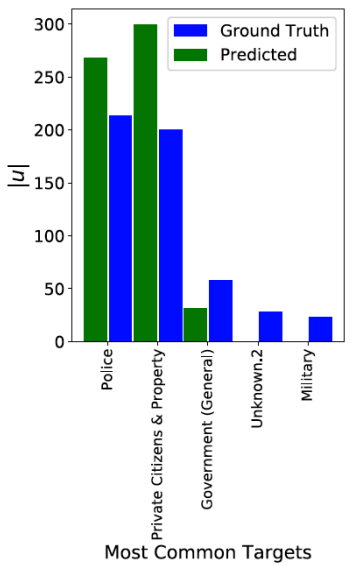
- Baseline: no change, i.e., $\psi_{i,\text{norm}}[u+1] = \psi_{i,\text{norm}}[u]$
- Shallow learning: use aggregate counts of weapon and tactics
- Meta-graphs: train various deep learning models
- Metrics:
  - MSE: $(\psi_{i,\text{norm}}[u] - \hat{\psi}_{i,\text{norm}}[u])^2$
  - Element-wise: $\phi[u] = 1$ if $\hat{S}[u] \cap S[u] \neq \emptyset$, 0 otherwise
  - Set-wise: $\gamma[u] = \frac{|\hat{S}[u] \cap S[u]|}{|S[u]|}$

- Meta-graph features >> shallow features
- Bi-LSTM: fitting a model with both backward and forward inputs → superior performance
- Baseline without learning ↓↓. Need for temporal element ✓✓

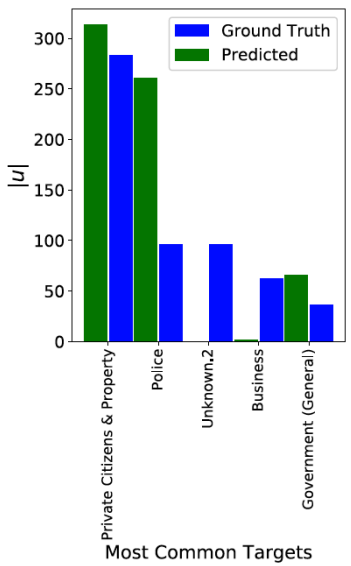




- Novel usage of meta-graphs, time-series and forecasting algorithms for terrorist targets.
- Validating the proposed methods.
- **Weakness:** Over-estimation of frequent targets - limited dataset.
- Lacking geo-spatial information? Regional level dynamics are more important in practice.



Afghanistan



Iraq

Undocumented immigration and terrorism: Is there a connection?

Light, M. T., & Thomas, J. T. (2021).

- Unauthorized immigration: divisive and controversial
- Changing course with 9/11: economic/crime $\rightarrow$ national security
- Federal government spends big $$
- Number of undocumented immigrants: 3.5M to 11.3M
- Systematic investigation of undocumented immigration & terrorism

- Immigrant networks spread terrorism and radicalization
  - Terrorism: overall immigration $\downarrow$, terror-prone countries $\uparrow$
  - Adverse socio-political and economic conditions $\rightarrow$ radicalization.
- Virtually no relationship
  - Only six percent of active terrorists were unauthorized.
  - None from Mexico. "We have terrorists coming through the southern border.. " (Trump, 2019).
  - Unauthorized immigration to US: not from terror-prone regions.

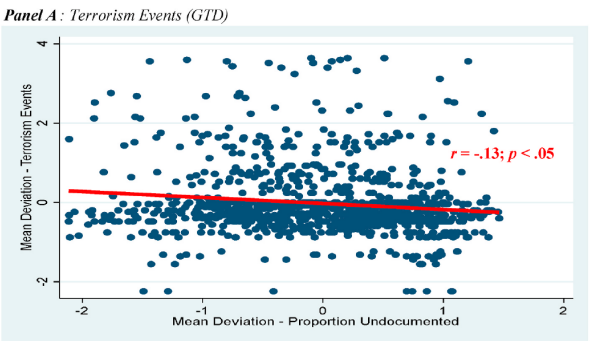
- **GTD:** Relies on public sources. Missing critical components.
  - Thwarted in planning stage, not linked to specific events etc.
- **PIRUS:** Profiles of Individual Radicalization in the US.
  - Detailed look into radicalization and broader definition of terrorism, e.g., ideology.
- **USSC:** U.S. Sentencing Commission.
  - Terrorism and national defense charges, even without act of violence.
- Center for Migration Studies: proportion of the unauthorized population.

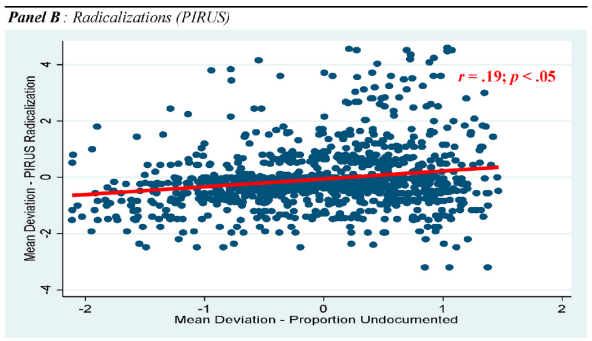
- State-year investigation. Rare events.
  fixed-effects negative binomial regression
- State and year effects are fixed.
- State-year investigation. Control variables:
  - Political ideology: NOMINATE measure from legislative votes.
  - Religious context: Christian fundamentalism etc.
  - Poverty, unemployment, low education, violent crime, lawful immigrants, demographic compositions..

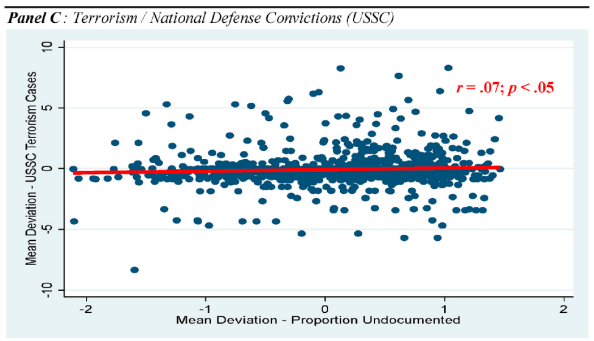
Undocumented immigration and terrorism: Is there a connection?

Results

- Mean deviations in proportion of undocumented vs. terrorism/radicalization
- Slightly negative (-0.13), slightly positive (0.19), negligibly positive (0.07).







- Baseline model / added lawful immigration / added full covariates
- All results are statistically insignificant
- Adding control variables: improved model fit, no change in R^2

Terrorism (GTD)	**Model 1**	**Model 2**	**Model 3**
Undocumented	0.23 (0.17)	0.17 (0.18)	0.21 (0.21)
Lawful	N/A	0.18 (0.15)	0.18 (0.16)
Radicalization (PIRUS)	**Model 1**	**Model 2**	**Model 3**
Undocumented	0.16 (0.14)	0.13 (0.13)	0.08 (0.11)
Lawful	N/A	0.09 (0.10)	-0.00 (0.10)

- Addressing a significant blind spot at the core of public attention.
- Reliably fail to reject the null hypothesis: no effect.
- Similar results for radicalization.
- How terrorism cases track terrorist activity? (e.g., arrest, sentencing)
- State-level analysis may be imperfect (e.g., internal mobility)
- International scope: similar methods may be unsuitable.