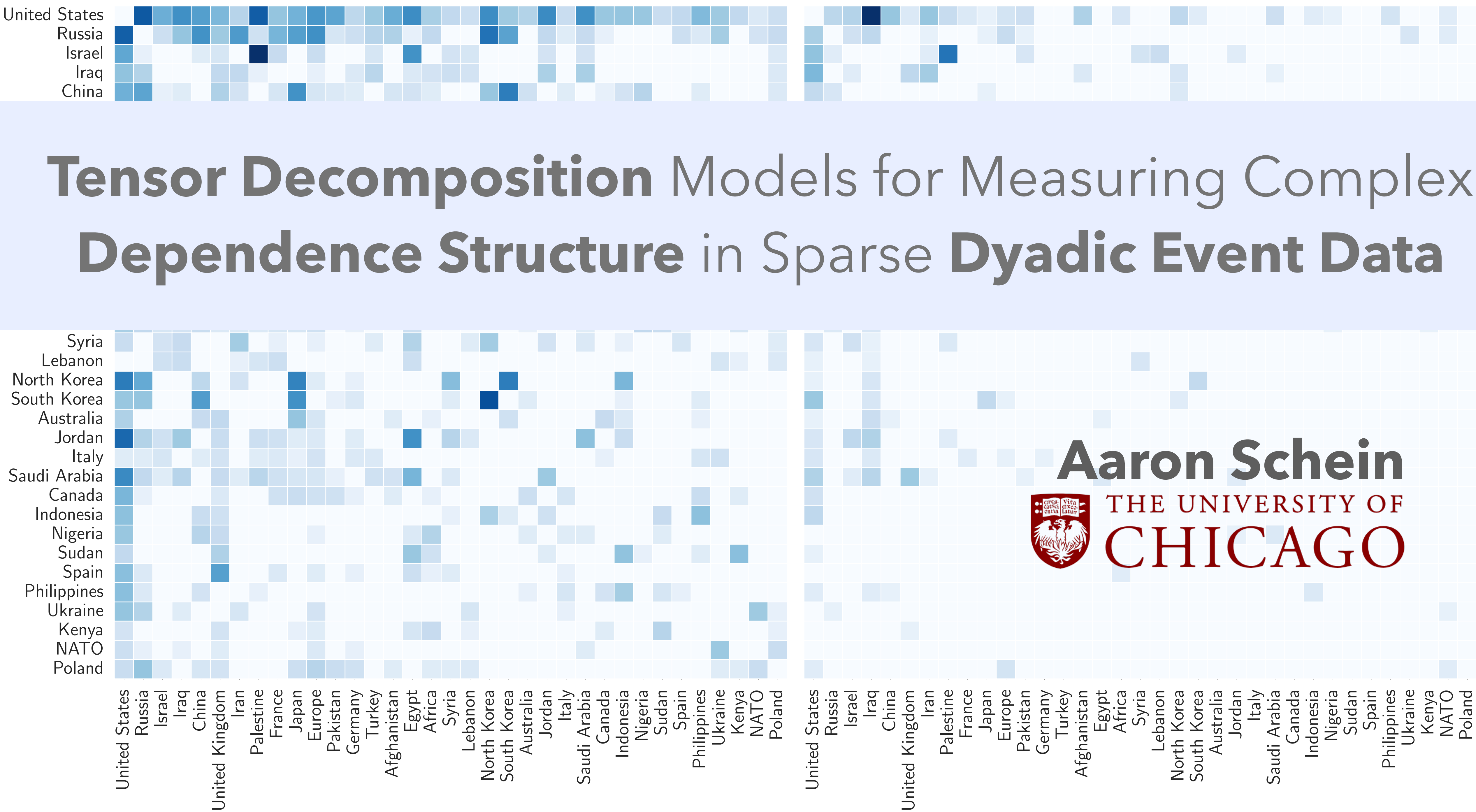
Tensor Decomposition Models for Measuring Complex **Dependence Structure** in Sparse **Dyadic Event Data**

Aaron Schein

THE UNIVERSITY OF CHICAGO



Dyadic event data

Example from Schrodt et al. [1995]:

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

Dyadic event data

Example from Schrodt et al. [1995]:

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	**IRQ**	**KUW**	**MOBILIZATION**
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

Dyadic event data

Example from Schrodt et al. [1995]:

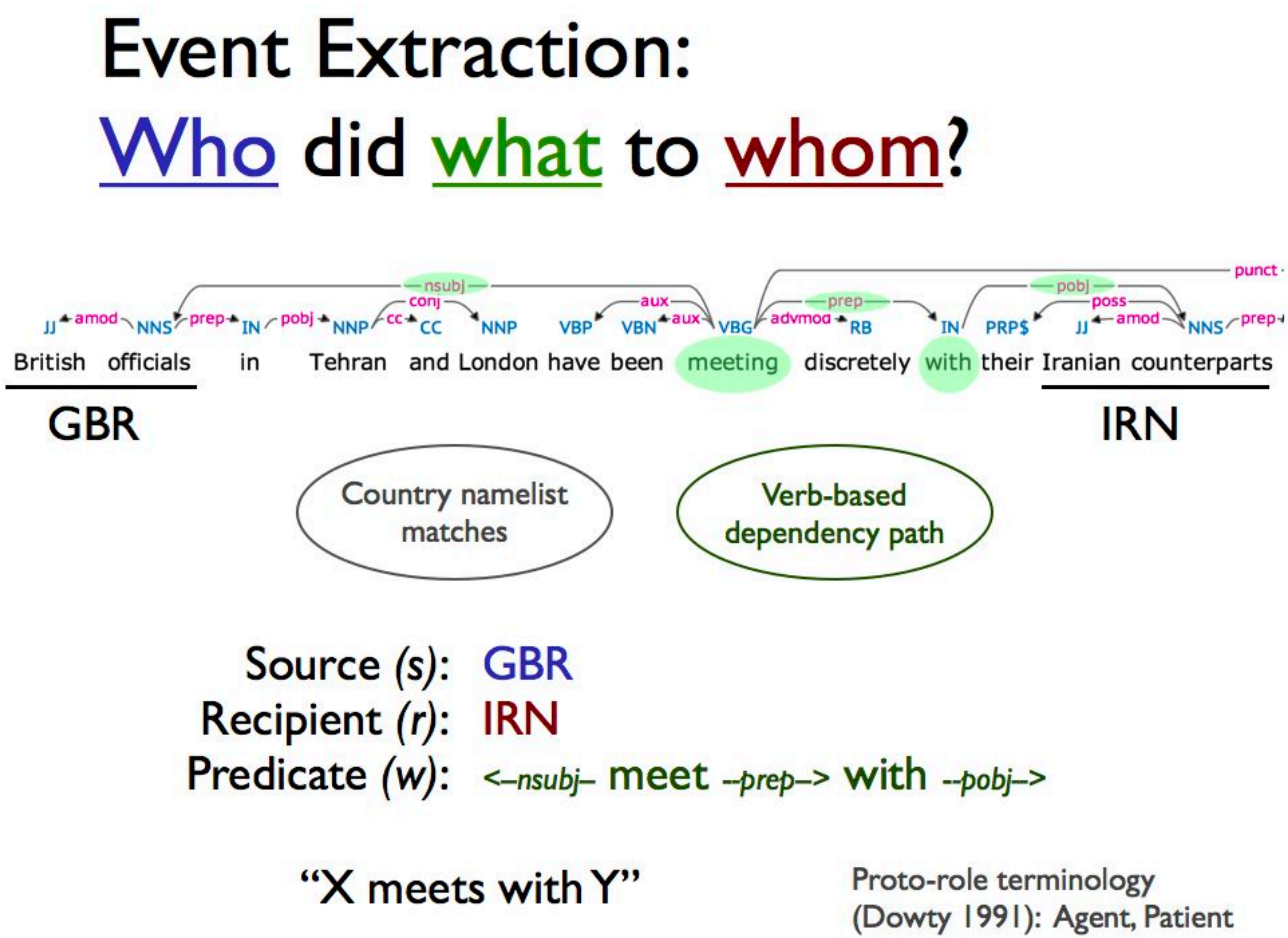
Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	**IRQ**	**KUW**	**MOBILIZATION**
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

July 31, 1990: IRAQ INCREASES TROOP LEVELS ON KUWAIT BORDER
Iraq has concentrated nearly 100,000 troops close to the Kuwaiti border, more than triple the number reported a week ago, the Washington Post said in its Tuesday editions.

MACHINE-CODED EVENT DATA

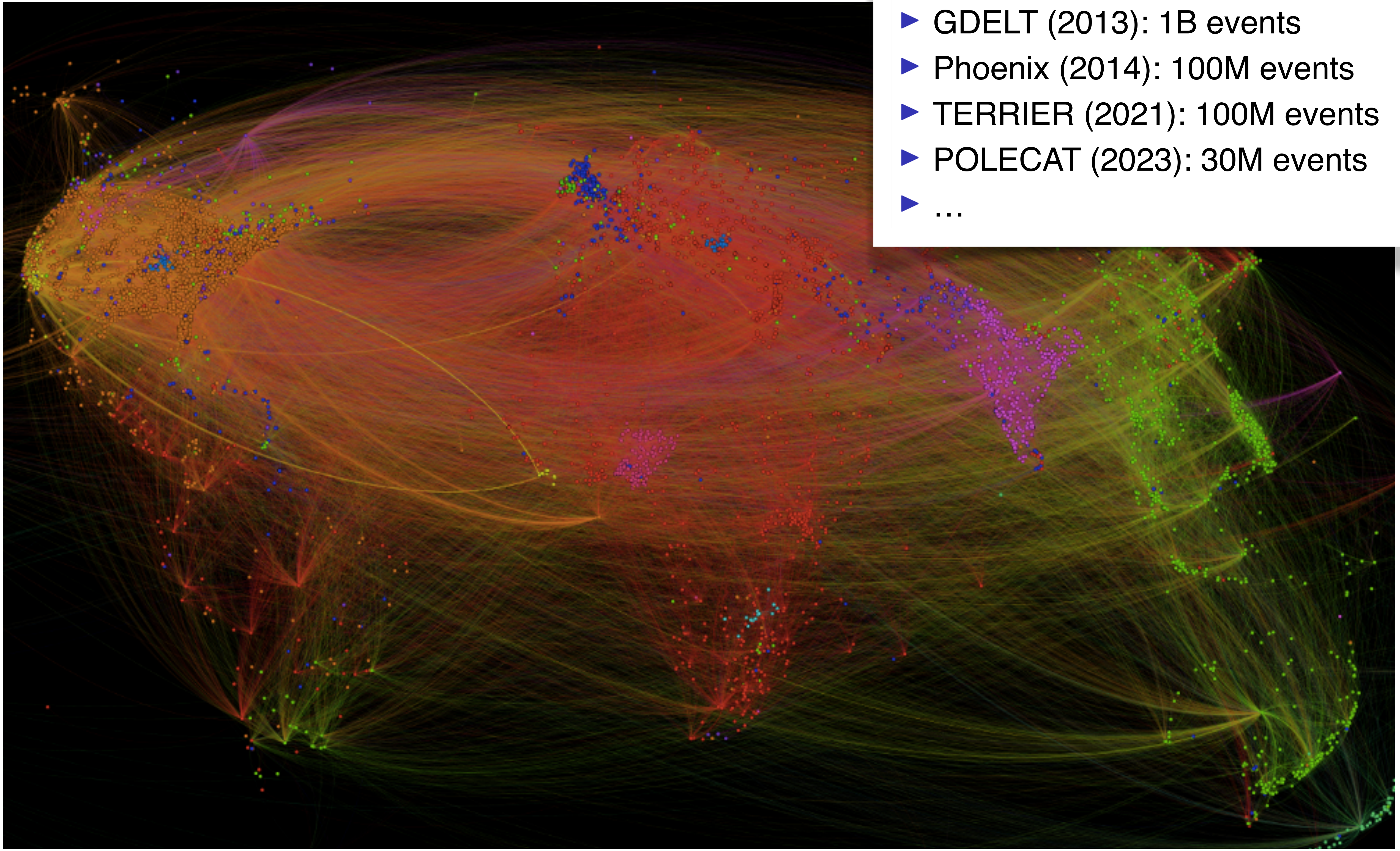
Figure from O’Connor [2013]:



MODERN DYADIC EVENT DATA SETS

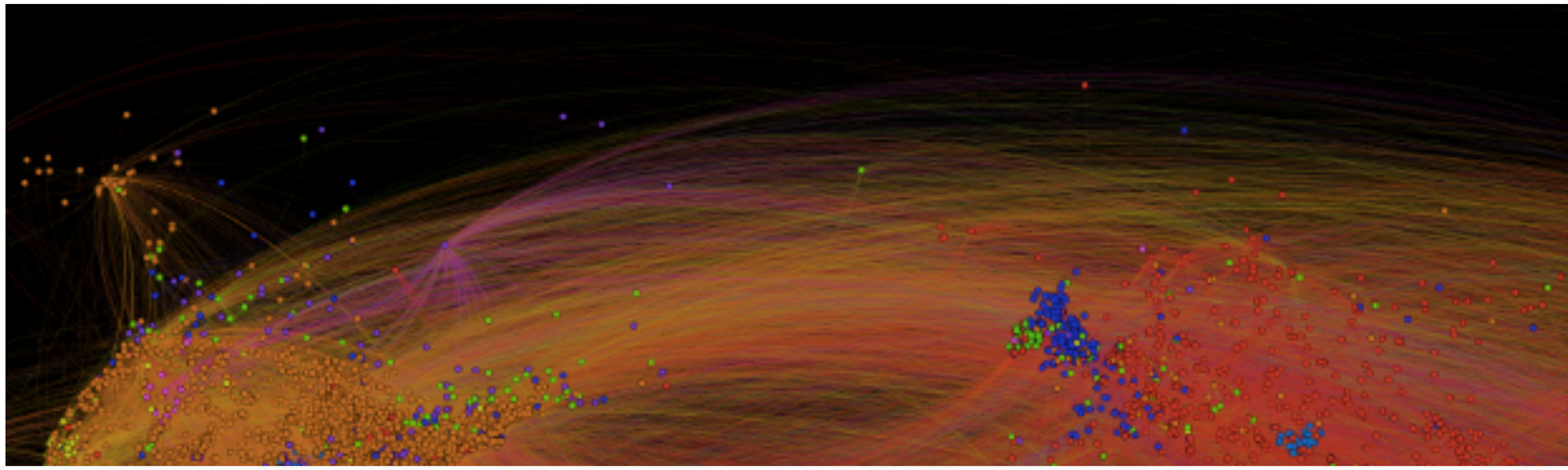
Lots of progress on event **extraction**…

- ICEWS (2011): 10M events
- GDELT (2013): 1B events
- Phoenix (2014): 100M events
- TERRIER (2021): 100M events
- POLECAT (2023): 30M events
- …



MODERN DYADIC EVENT DATA SETS

Lots of progress on event **extraction**…

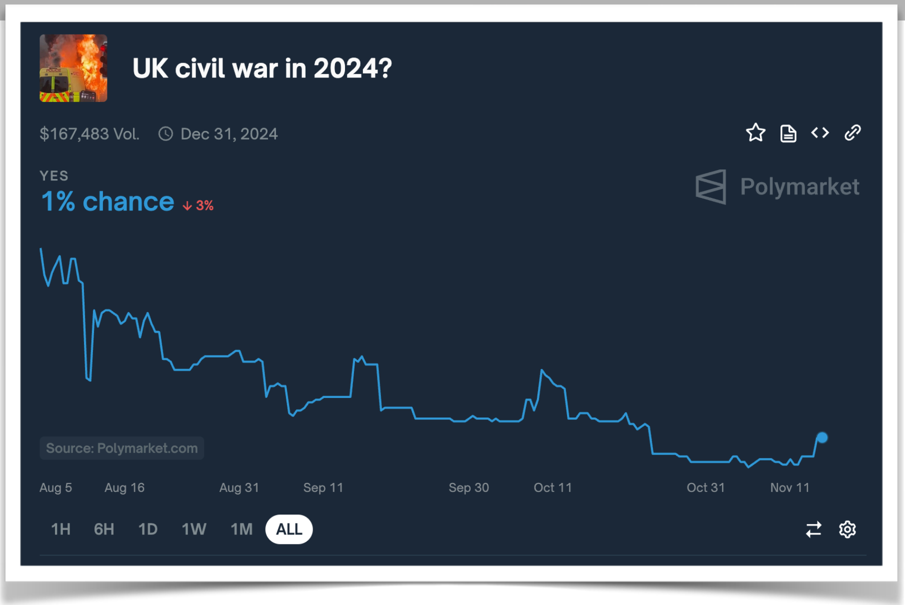


- ICEWS (2011): 10M events
- GDELT (2013): 1B events
- Phoenix (2014): 100M events
- TERRIER (2021): 100M events
- POLECAT (2023): 30M events
- …

Event_Date	Source_Name	Source_Sectors	Source_Country	Event_Text	CAME(	Target_Name	Target_Sectors	Target_Country
2022-06-14	Foreign Affairs (Japan)	Executive,Forei...	Japan	Consult	40	Abul Kalam Abdul ...	Executive,Forei...	Bangladesh
2002-04-10	Business (Iraq)	Social,Business	Iraq	Sign formal agre...	57	Industry (Bahrain)	Social,Business	Bahrain
2005-01-13	Israeli Defense Forces	Military,Govern...	Israel	fight with small ...	193	Citizen (Palestinia...	General Popul...	Occupied Pal...
2001-03-01	Royal Administration (...	Government	Nepal	Engage in negot...	46	Jiang Zemin	Executive Offic...	China
2005-06-27	Foreign Affairs (Irelan...	Executive,Forei...	Ireland	Consult	40	Secretary of State ...	Government,E...	United Kingd...
2006-03-13	Civil Service (Kazakhs...	Government	Kazakhstan	Consult	40	Foreign Affairs (C...	Foreign Minist...	China
1999-08-30	Ilya Iosifovich Klebanov	Executive Offic...	Russian Fed...	Host a visit	43	Jiang Zemin	Executive,Gov...	China
2022-04-20	Nikol Pashinyan	Elite,Executive,...	Armenia	Express intent t...	36	Federal Assembly	Government,L...	Switzerland
2012-09-03	Legislature (Syria)	Government,Le...	Syria	Consult	40	Foreign Affairs (R...	Executive,Gov...	Russian Fed...
2005-07-27	Head of Government (...	Executive,Gove...	Guyana	Engage in negot...	46	Tang Jiaxuan	Elite,Governme...	China
2009-10-02	Defense / Security Mi...	Government,De...	Belarus	Consult	40	Defense / Security...	Defense / Sec...	Russian Fed...
2004-12-17	Jean-Claude Juncker	Elite,Executive ...	Luxembourg	Criticize or deno...	111	Citizen (Belgium)	Social,General ...	Belgium
2004-04-23	Vladimir Putin	Executive,Exec...	Russian Fed...	Consult	40	Leonid Kuchma	Elite,Executive,...	Ukraine
2018-07-26	Qasem Soleimani	Government,Mi...	Iran	Threaten	130	Donald Trump	Government,E...	United States

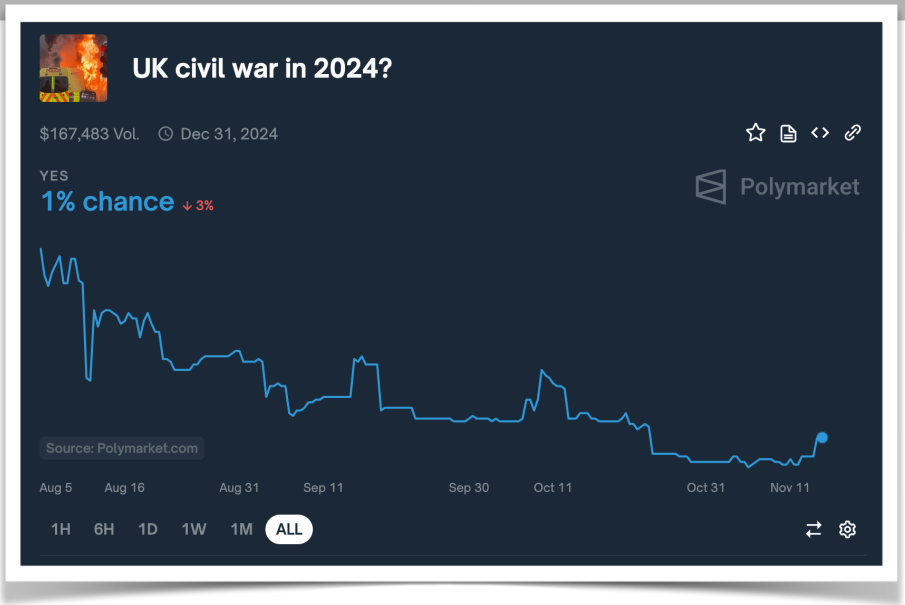
Results per page: 50

(POTENTIAL) APPLICATIONS OF EVENT DATA

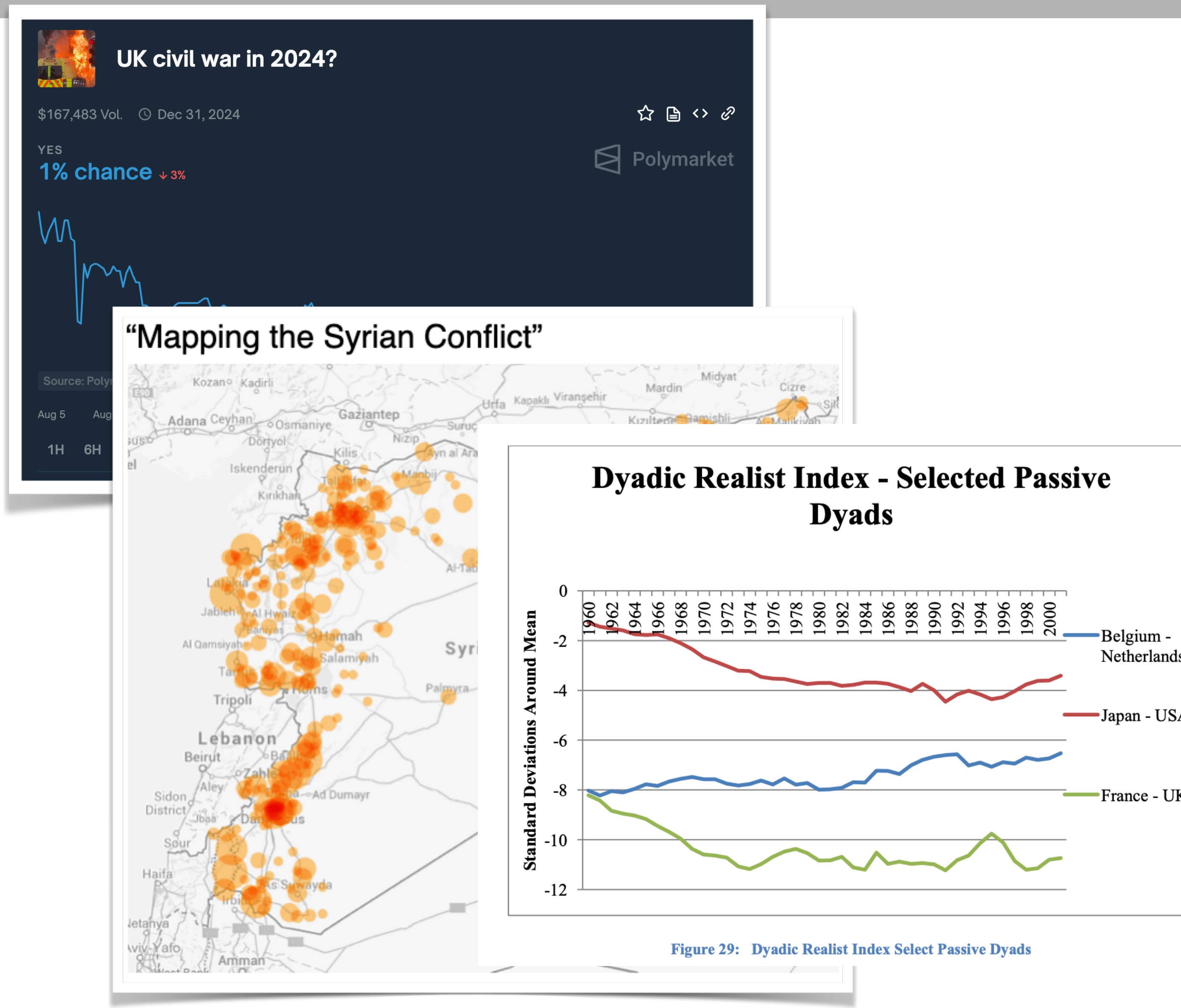


(POTENTIAL) APPLICATIONS OF EVENT DATA

- **Predictive:**
  - e.g., forecasting civil war

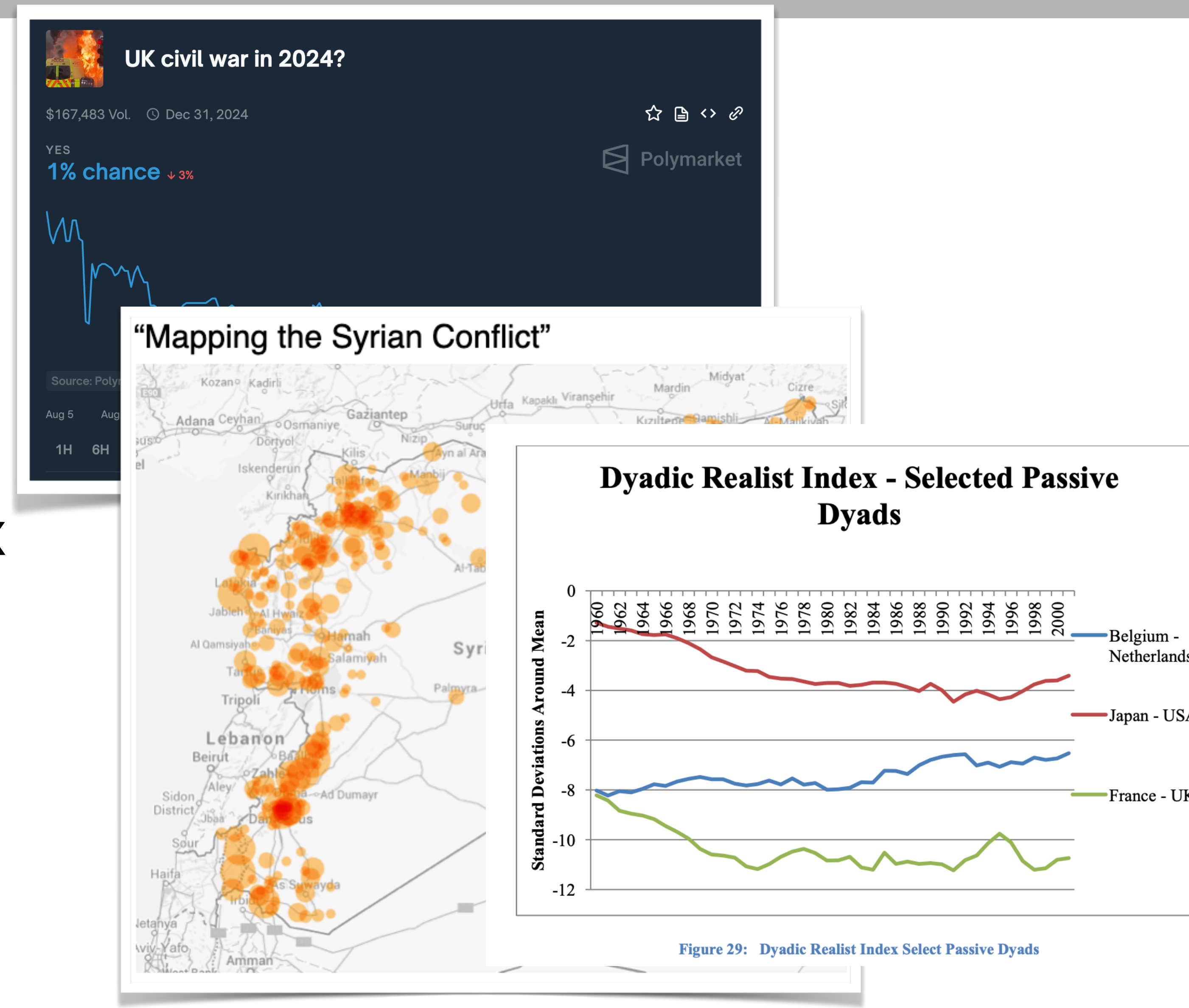


(POTENTIAL) APPLICATIONS OF EVENT DATA

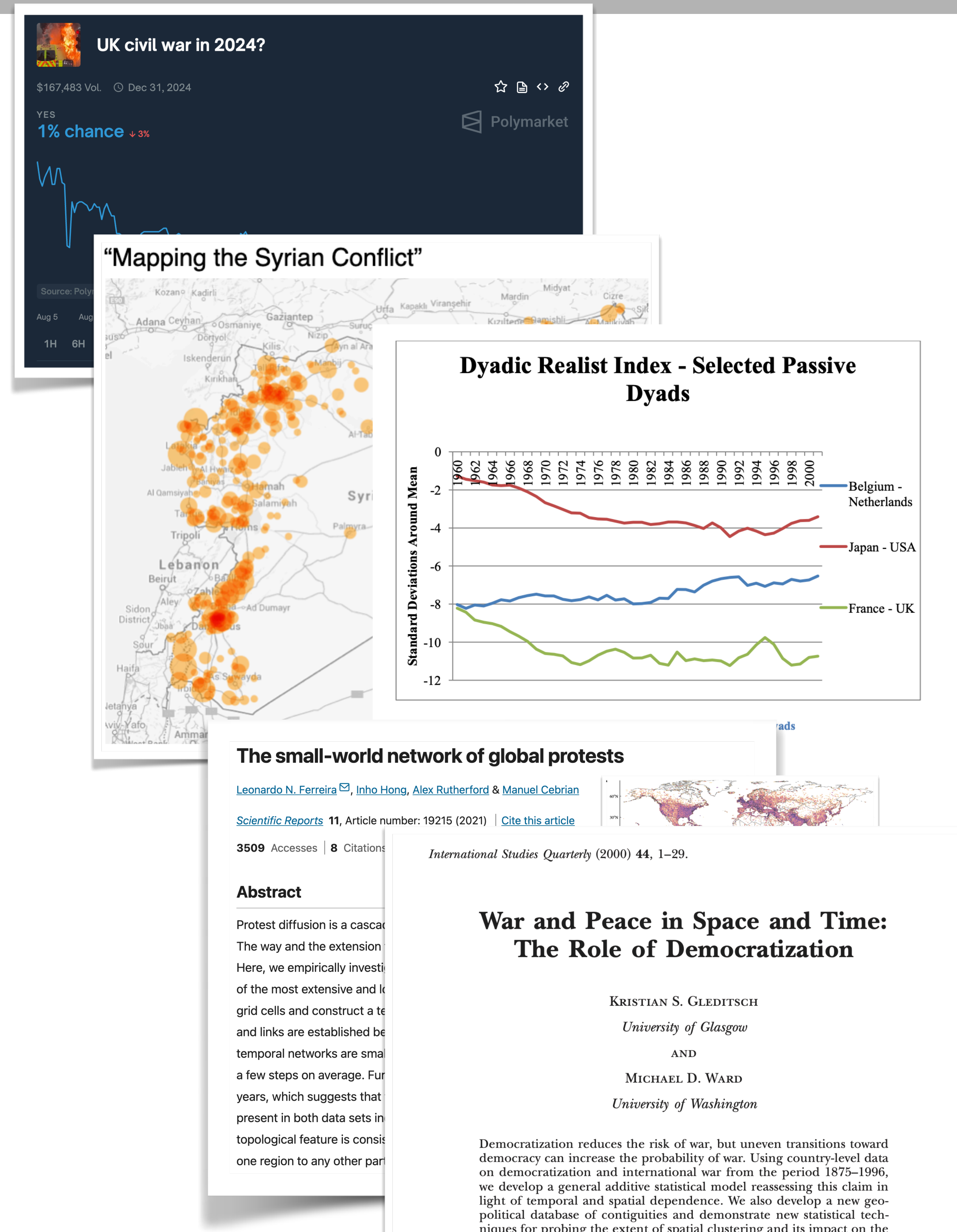


(POTENTIAL) APPLICATIONS OF EVENT DATA

- **Predictive:**
  - e.g., forecasting civil war
- **Descriptive**:
  - e.g., “power imbalance” index
  - e.g., conflict mapping
- **Exploratory:**
  - hypothesis generation

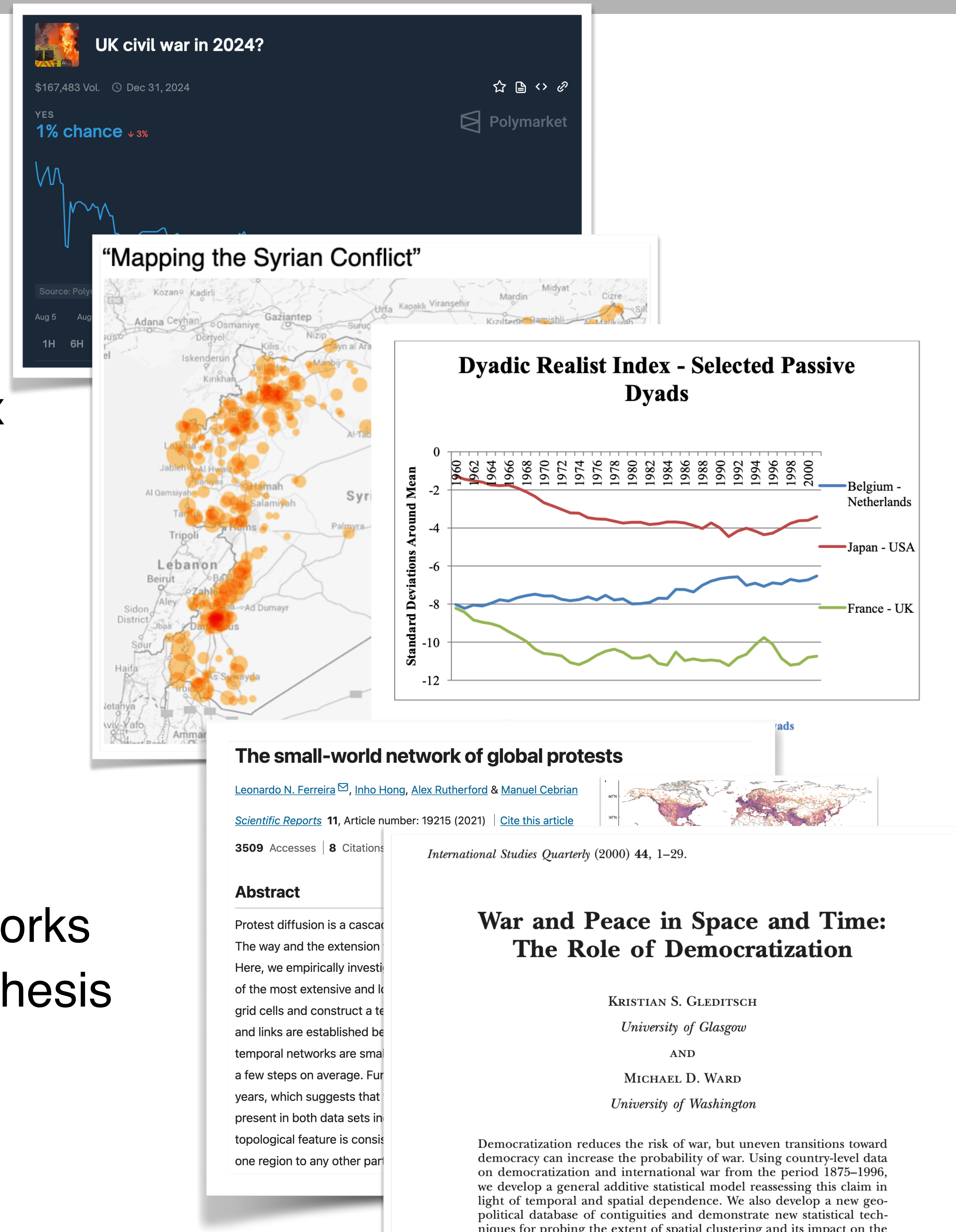


(POTENTIAL) APPLICATIONS OF EVENT DATA



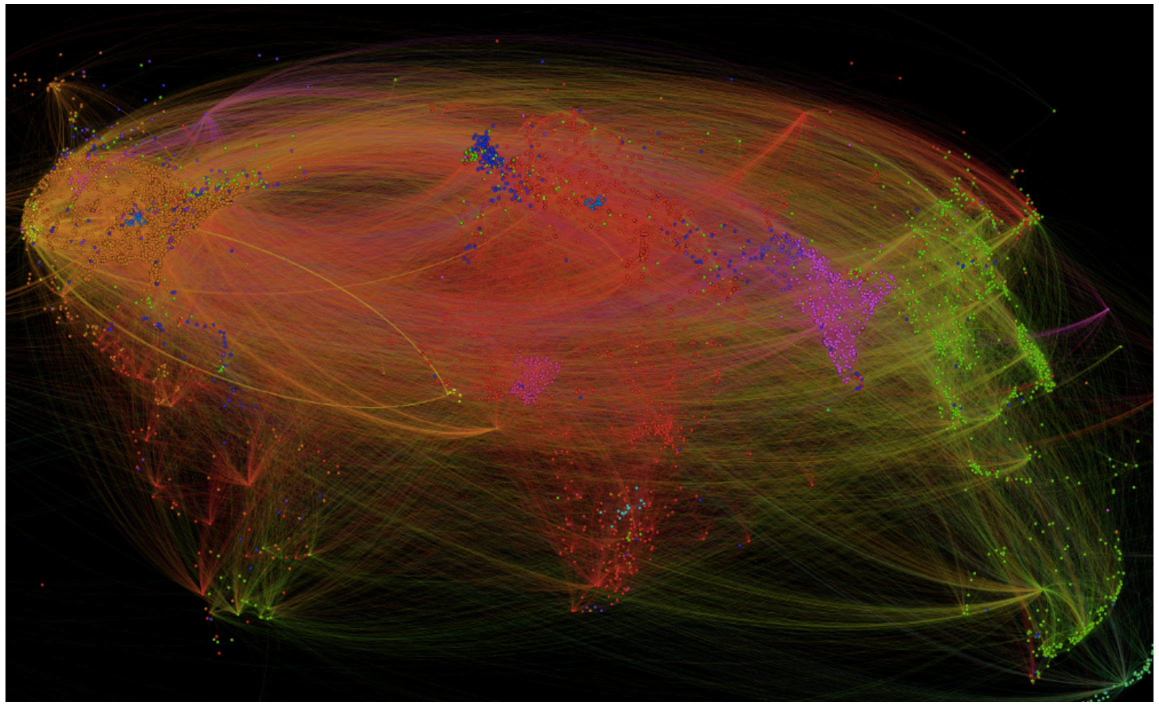
(POTENTIAL) APPLICATIONS OF EVENT DATA

- **Predictive:**
  - e.g., forecasting civil war
- **Descriptive**:
  - e.g., “power imbalance” index
  - e.g., conflict mapping
- **Exploratory:**
  - hypothesis generation
- **Explanatory**:
  - test theories / hypotheses
  - e.g., small-world protest networks
  - e.g., Democratic peace hypothesis



CHALLENGES

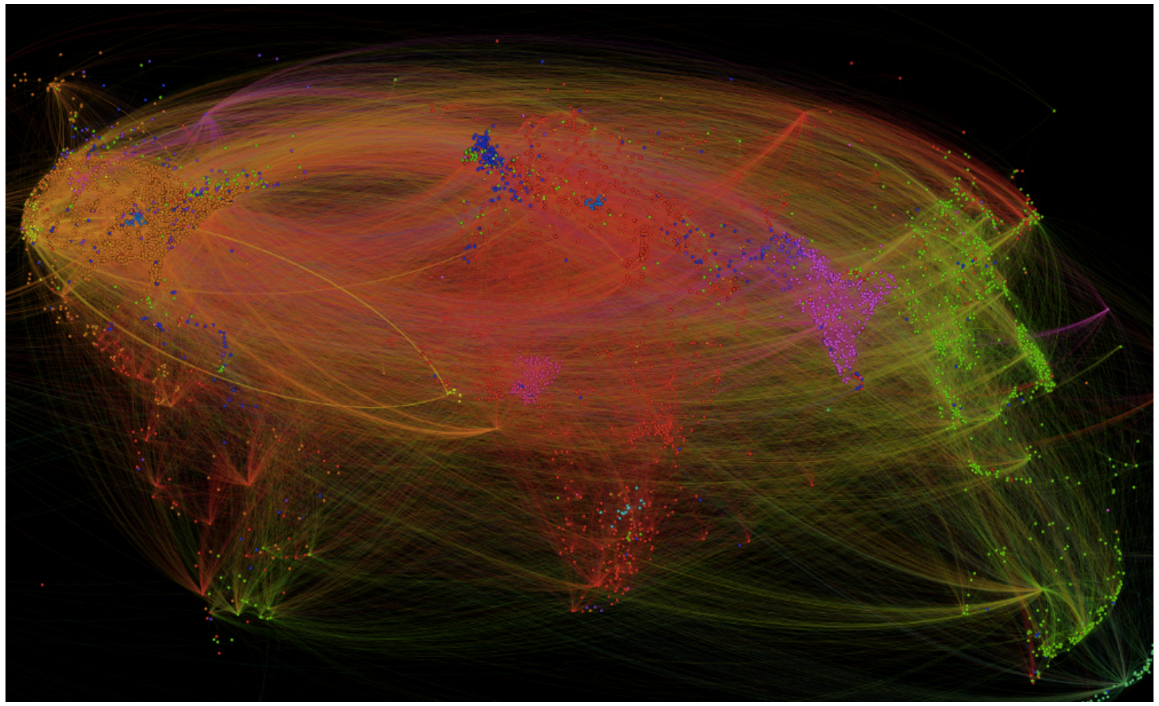
- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases



CHALLENGES

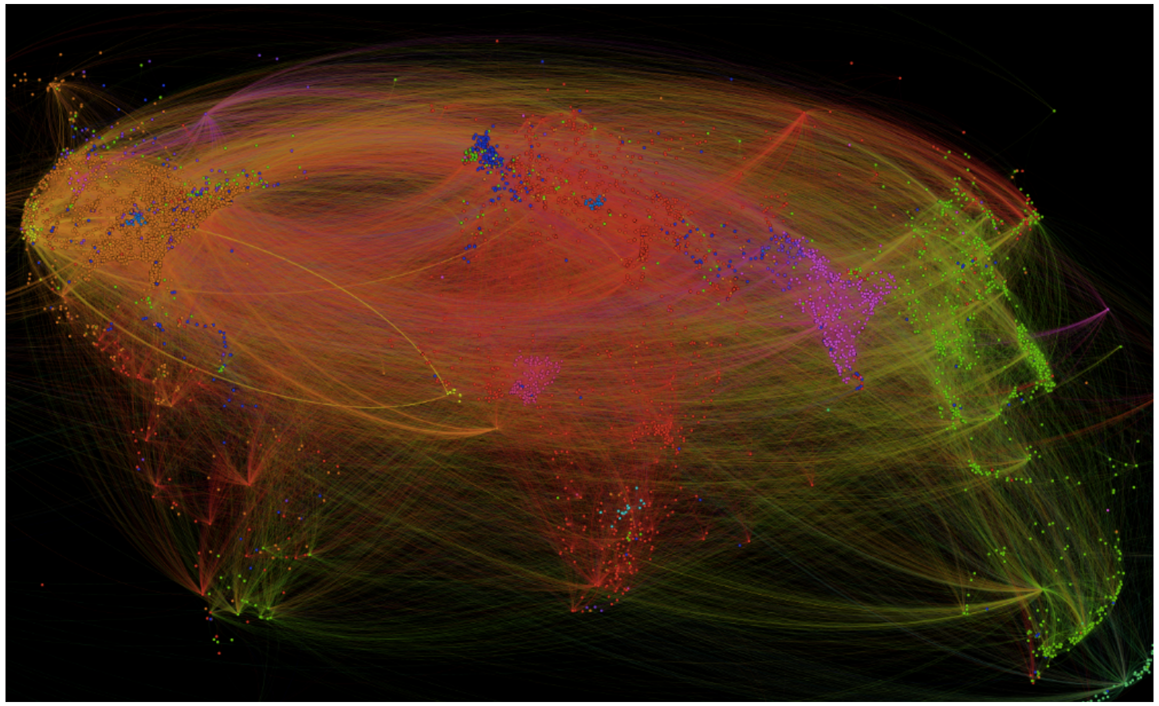
- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases

- **High-dimensional, big, structured:**
  - e.g., 100M events
  - e.g., 300K unique “source sectors”
  - e.g., hierarchical coding schemes



CHALLENGES

- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases
- **High-dimensional, big, structured:**
  - e.g., 100M events
  - e.g., 300K unique "source sectors"
  - e.g., hierarchical coding schemes
- **Complex dependencies**
  - network dependence
  - temporal dependence
  - …

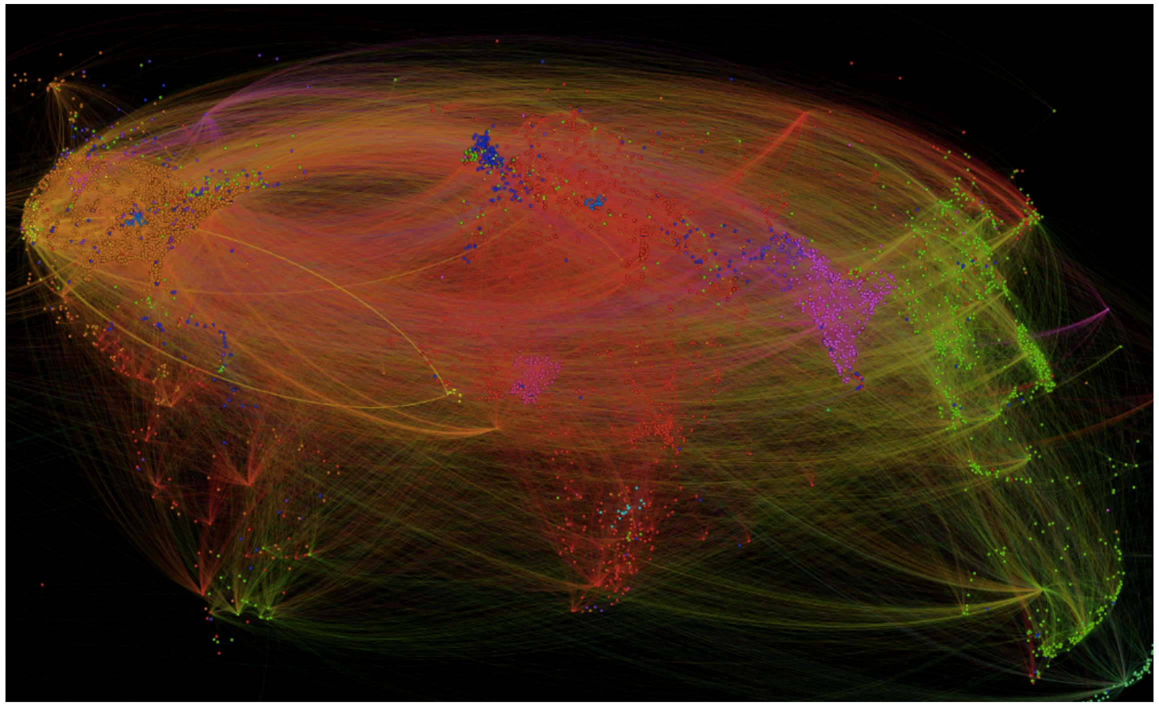


King [2001]:

> *[...] dyadic observations in international conflict data have* **complex dependence structures** *[...]*

CHALLENGES

- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases
- **High-dimensional, big, structured:**
  - e.g., 100M events
  - e.g., 300K unique "source sectors"
  - e.g., hierarchical coding schemes
- **Complex dependencies**
  - network dependence
  - temporal dependence
  - ...
- **Dyadic** only
  - e.g., summits not recorded
  - e.g., alliances not recorded



King [2001]:

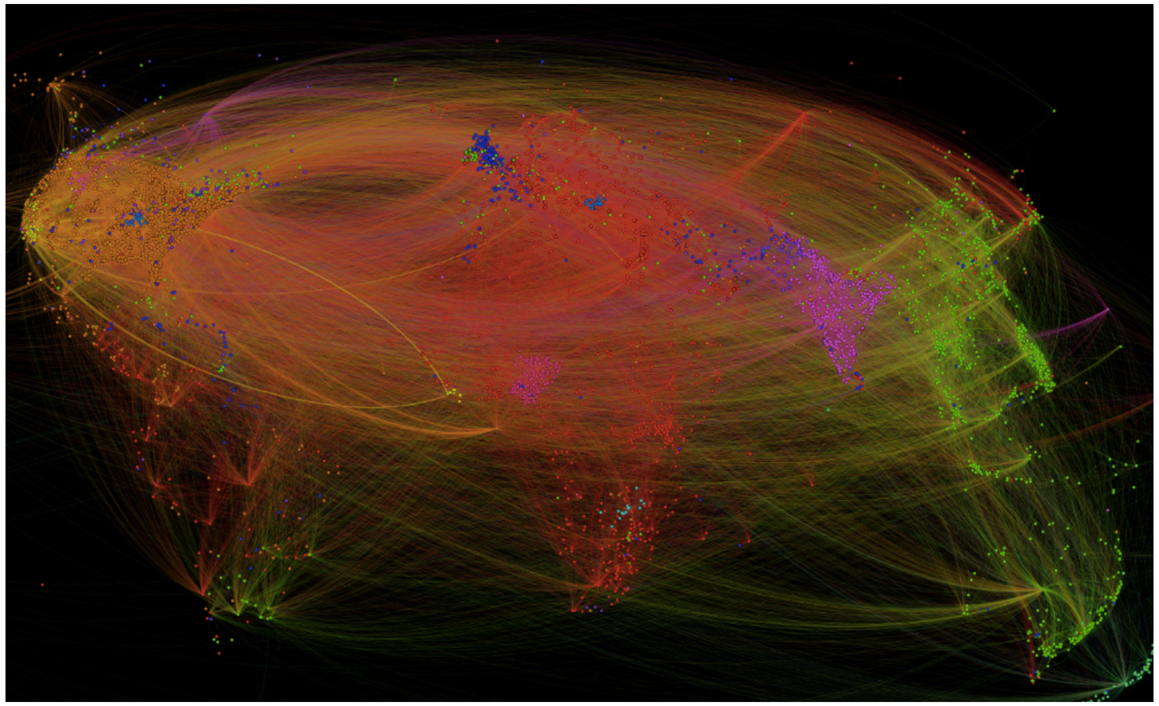
> *[...] dyadic observations in international conflict data have* **complex dependence structures** *[...]*

Poast [2010]:

> *Dyadic (state-pair) data is* **completely inappropriate** *for analyzing multilateral events (such as large alliances and major wars).*

CHALLENGES

- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases



- **High-dimensional, big, structured:**
  - e.g., 100M events
  - e.g., 300K unique "source sectors"
  - e.g., hierarchical coding schemes

- **Complex dependencies**
  - network dependence
  - temporal dependence
  - …

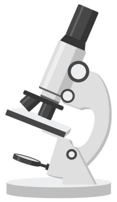
King [2001]:

> *[...] dyadic observations in international conflict data have* **complex dependence structures** *[...]*

- **Dyadic** only
  - e.g., summits not recorded
  - e.g., alliances not recorded

Poast [2010]:

> *Dyadic (state-pair) data is* **completely inappropriate** *for analyzing multilateral events (such as large alliances and major wars).*



⟹ **Need: models to extract meaningful signal**

CHALLENGES

Relax, Tensors Are Here
Dependencies in International Processes☆

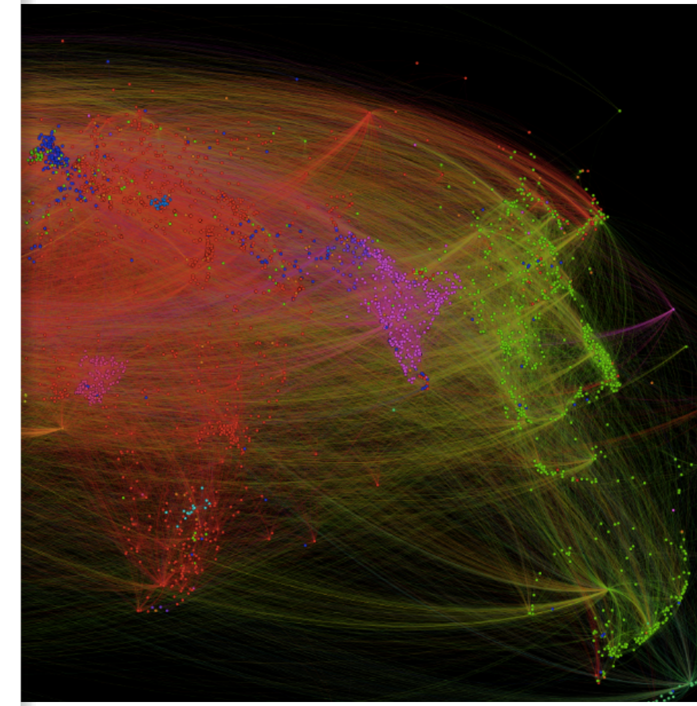
Shahryar Minhas[a], Peter D. Hoff[b], Michael D. Ward[a,*]

[a]*Department of Political Science, Duke University, Durham, NC 27701, USA*
[b]*Departments of Biostatistics & Statistics, University of Washington, Seattle, WA, USA*

Abstract

Previous models of international conflict have suffered two shortfalls. They tended not to embody dynamic changes, focusing rather on static slices of behavior over time. These models have also been empirically evaluated in ways that assumed the independence of each country, when in reality they are searching for the interdependence among all countries. We illustrate a solution to these two hurdles and evaluate this new, dynamic, network based approach to the dependencies among the ebb and flow of daily international interactions using a newly developed, and openly available, database of events among nations.

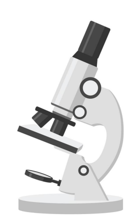
Keywords: Dynamic networks, time series, international crises, event data, tensor products



ns in interna-
complex de-
]

Poast [2010].

- **Dyadic** only
  - e.g., summits not recorded
  - e.g., alliances not recorded

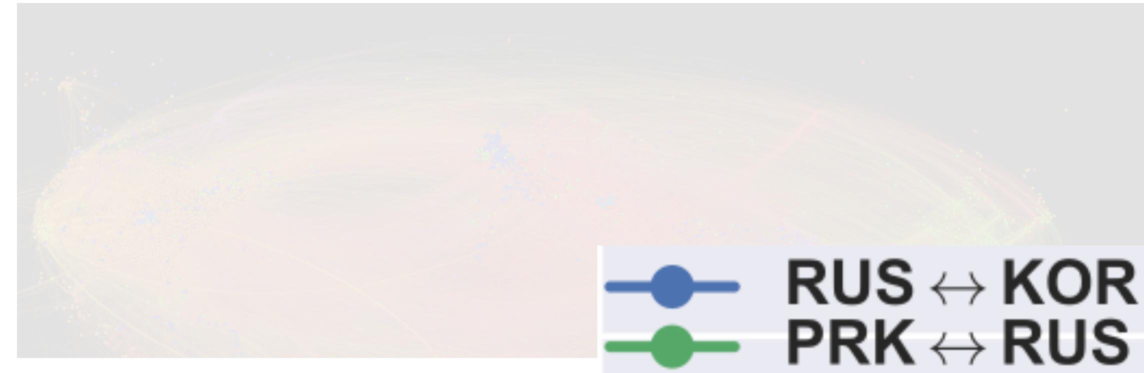


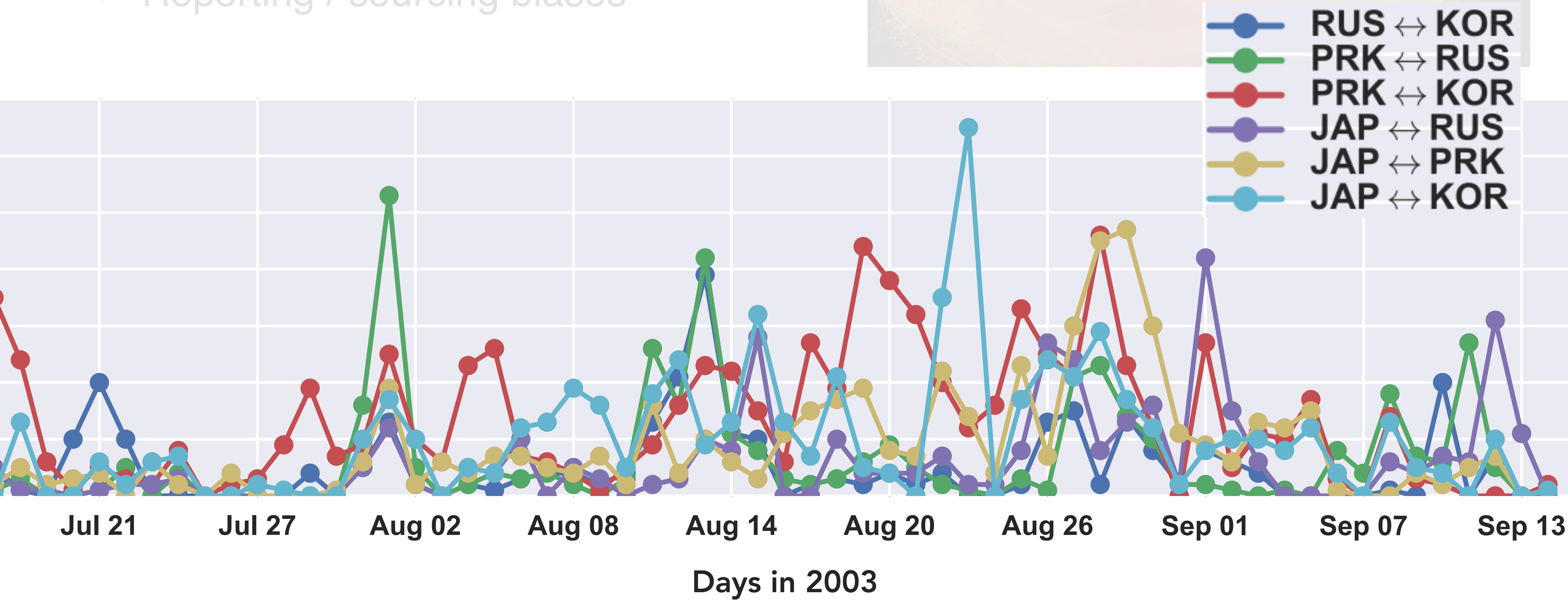
Dyadic (state-pair) data is **completely inappropriate** *for analyzing multilateral events (such as large alliances and major wars).*

⟹ **Need: models to extract meaningful signal**

CHALLENGES

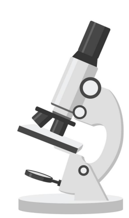
- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases





- **Dyadic** only
  - e.g., summits not recorded
  - e.g., alliances not recorded

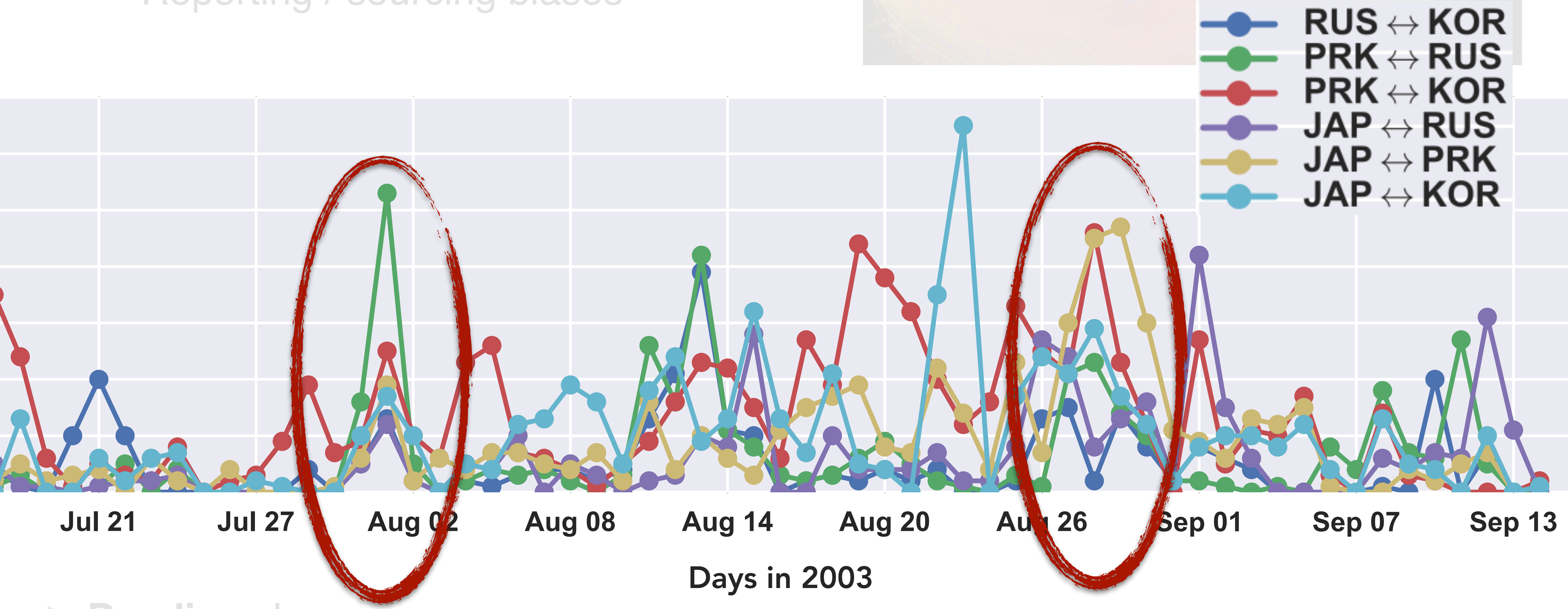
Dyadic (state-pair) data is ***completely inappropriate*** *for analyzing multilateral events (such as large alliances and major wars).*



⟹ **Need: models to extract meaningful signal**

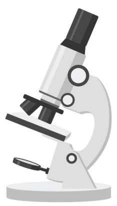
CHALLENGES

- **Noisy, error-prone, biased:**
  - Extraction / coding errors
  - Reporting / sourcing biases



- **Dyadic** only
  - e.g., summits not recorded
  - e.g., alliances not recorded

Dyadic (state-pair) data is ***completely inappropriate*** *for analyzing multilateral events (such as large alliances and major wars).*



⟹ **Need: models to extract meaningful signal**

CHALLENGES

- Noisy, error-prone, b
  - Extraction / coding
  - Reporting / sourcin

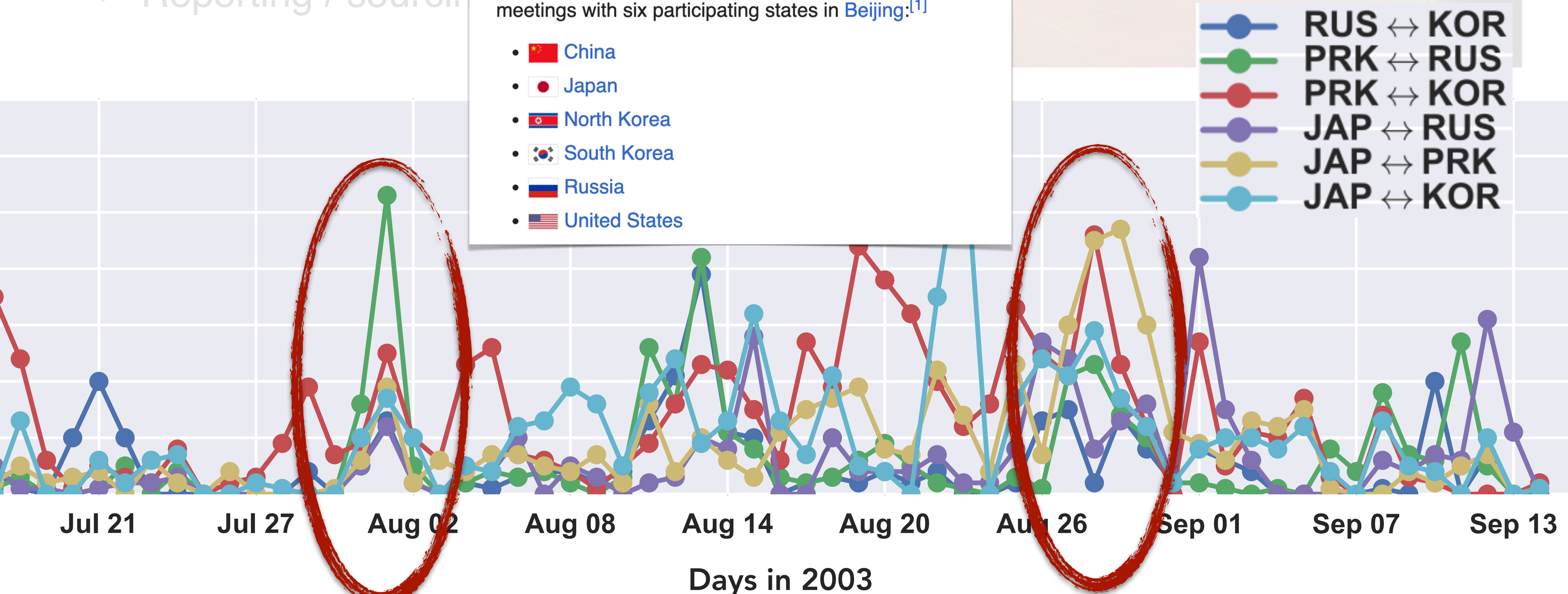

Six-party talks

Article Talk

From Wikipedia, the free encyclopedia

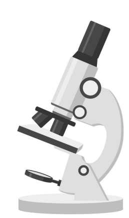
The **six-party talks** aimed to find a peaceful resolution to the security concerns as a result of the North Korean nuclear weapons program. There was a series of meetings with six participating states in Beijing:[1]

- China
- Japan
- North Korea
- South Korea
- Russia
- United States



- **Dyadic** only
  - e.g., summits not recorded
  - e.g., alliances not recorded

Dyadic (state-pair) data is ***completely inappropriate*** *for analyzing multilateral events (such as large alliances and major wars).*



⇨ **Need: models to extract meaningful signal** **(there *is* signal)**

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

Event **tokens**

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

[Schrodt, '93]

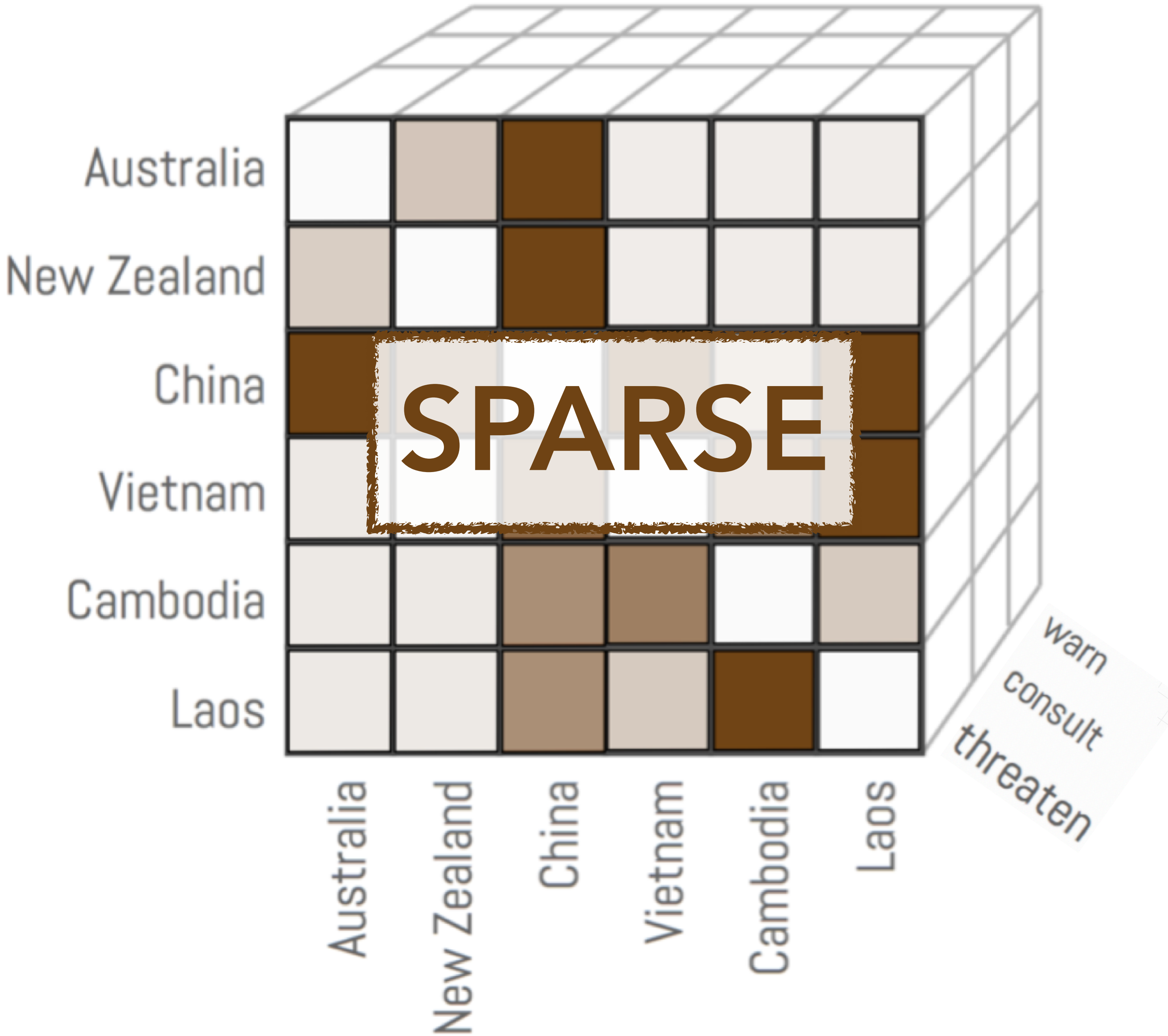
Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	**IRQ**	**KUW**	**MOBILIZATION**
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

Event **type counts**

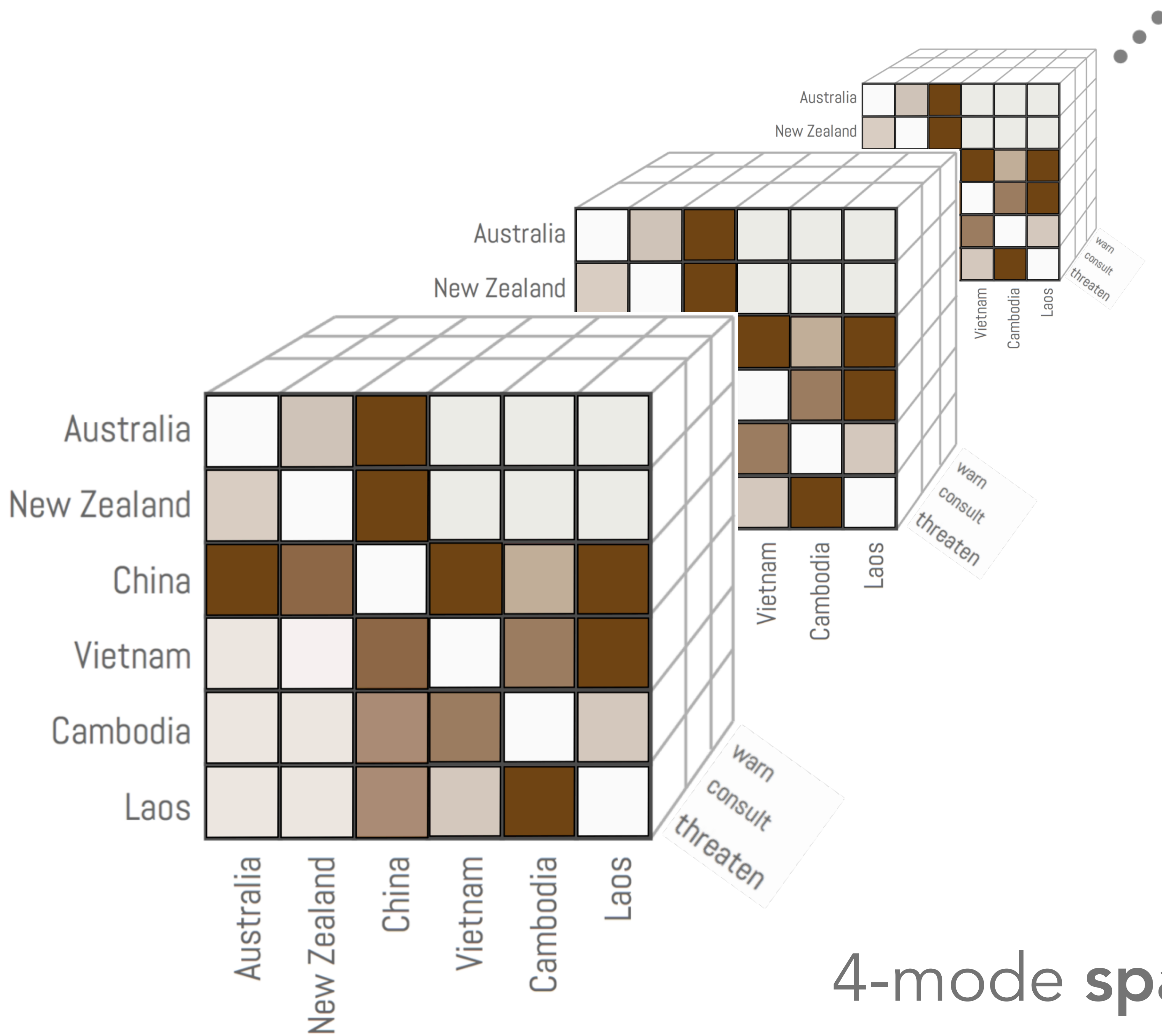
$$y^{(t)}_{i \xrightarrow{a} j} =$$

"the number of tokens of type **country *i*** took **action *a*** to **country *j*** during **time *t***"

Event **type counts**

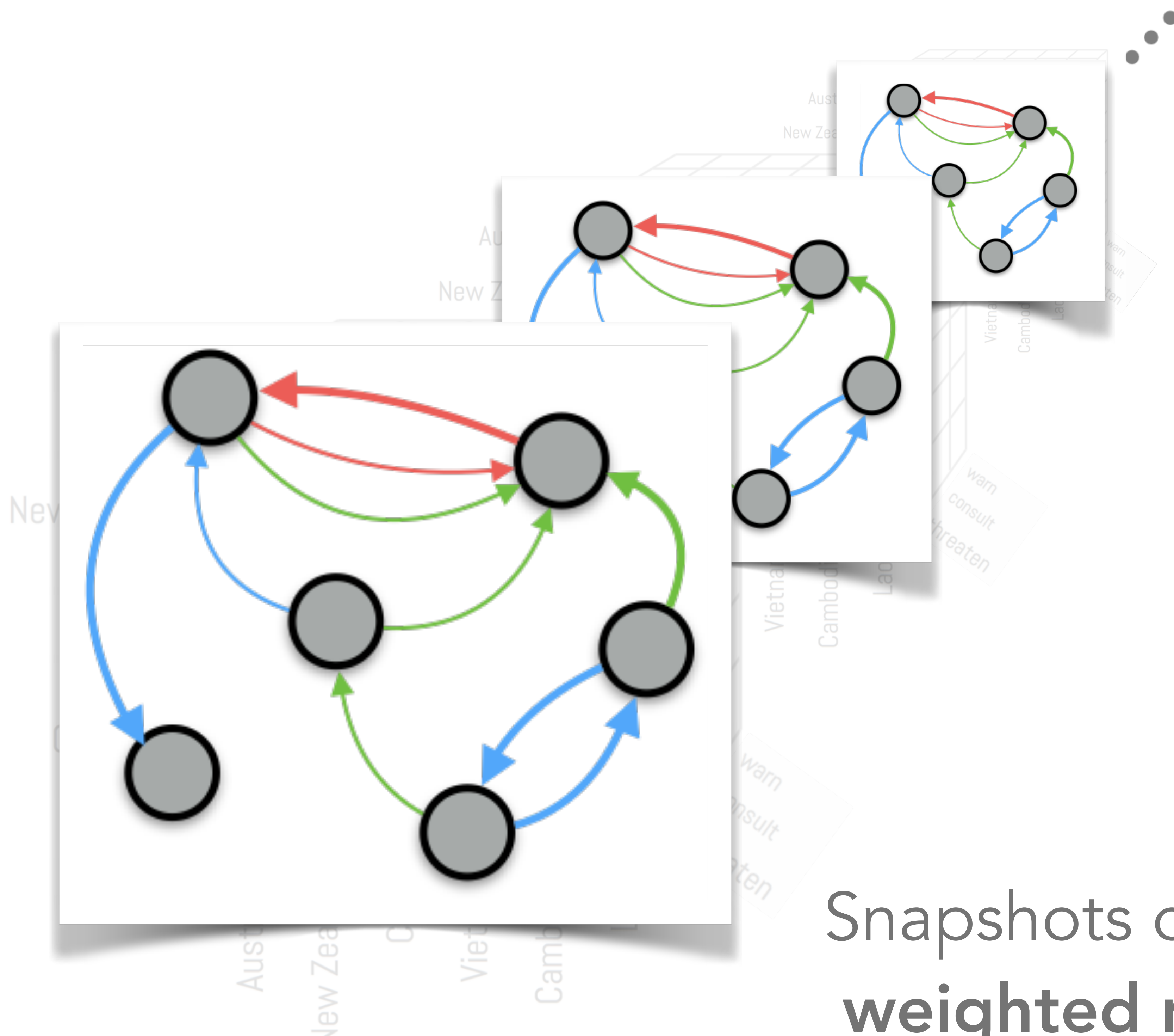


Event **type counts**



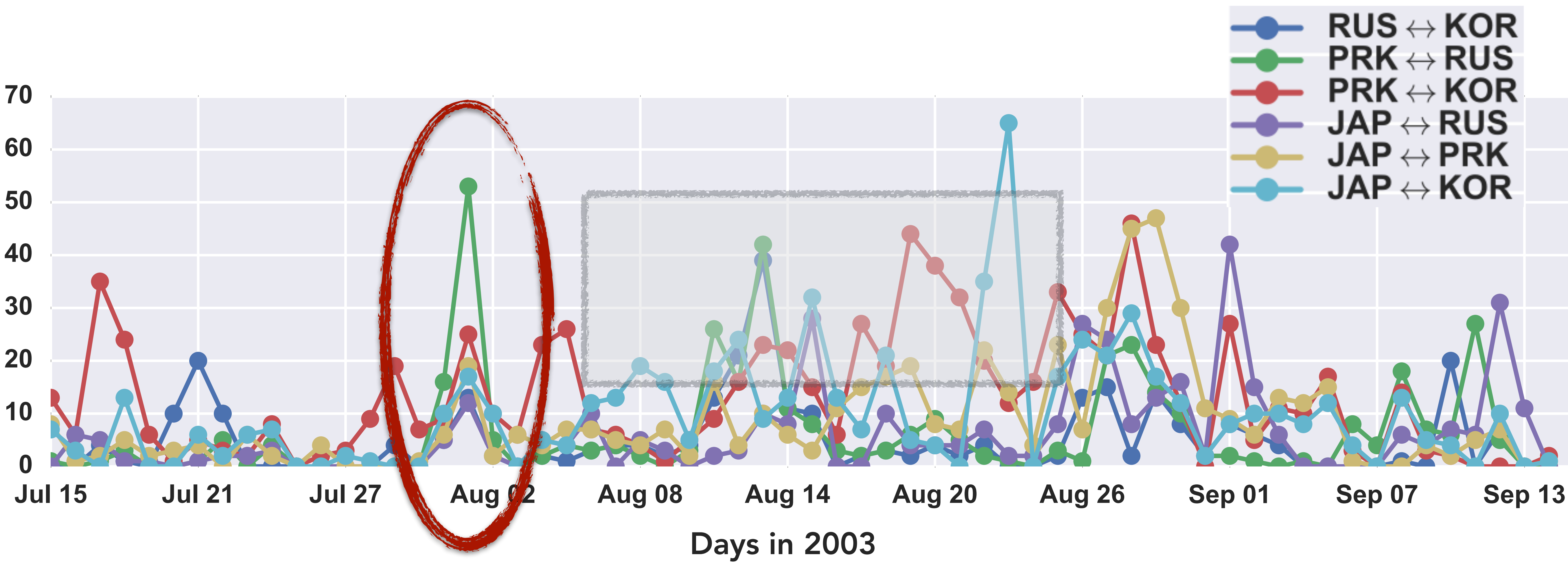
4-mode **sparse count tensor**

Event **type counts**



Snapshots of a **weighted multiplex network**

Event **type counts**

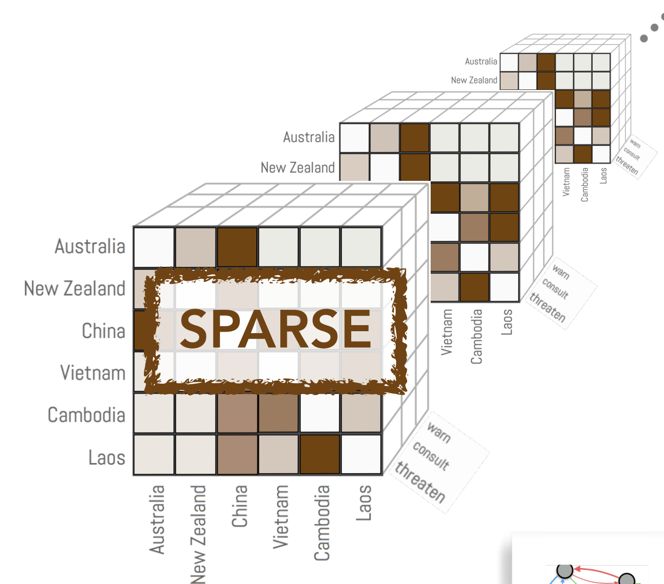


REPRESENTATIONS OF DYADIC DATA

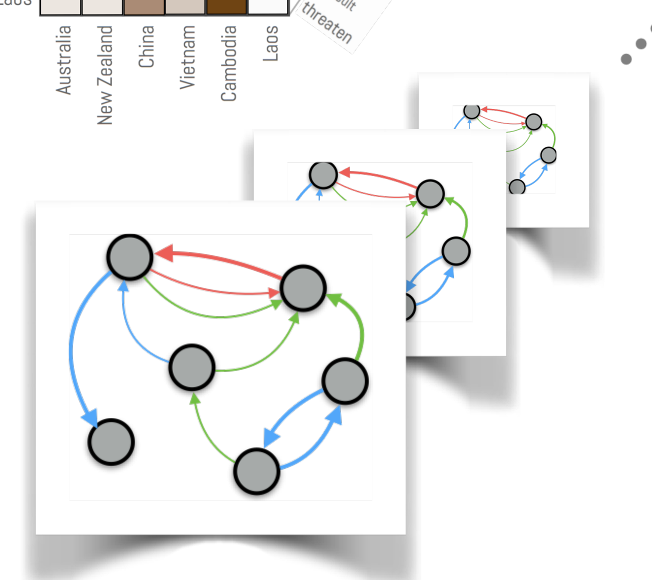
tokens

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

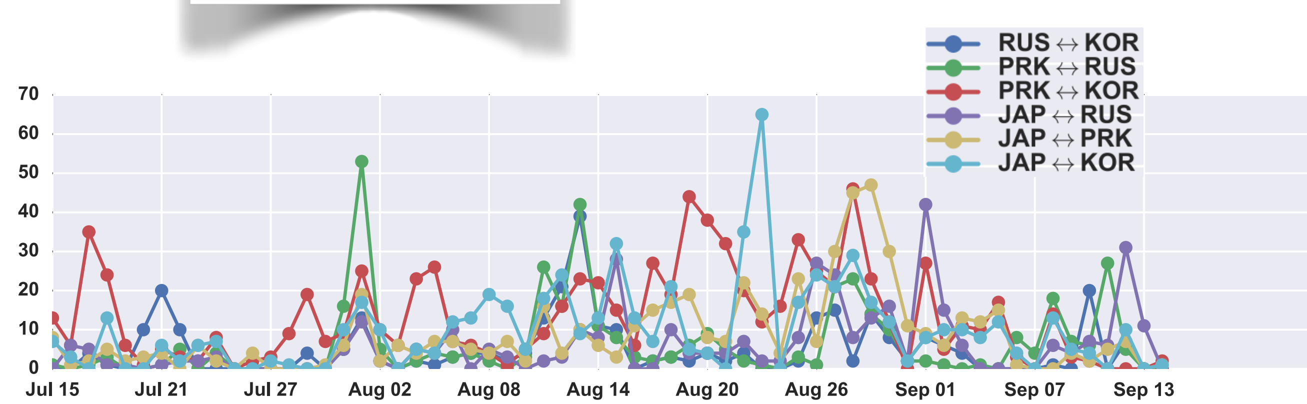
count tensor



networks



time series



OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

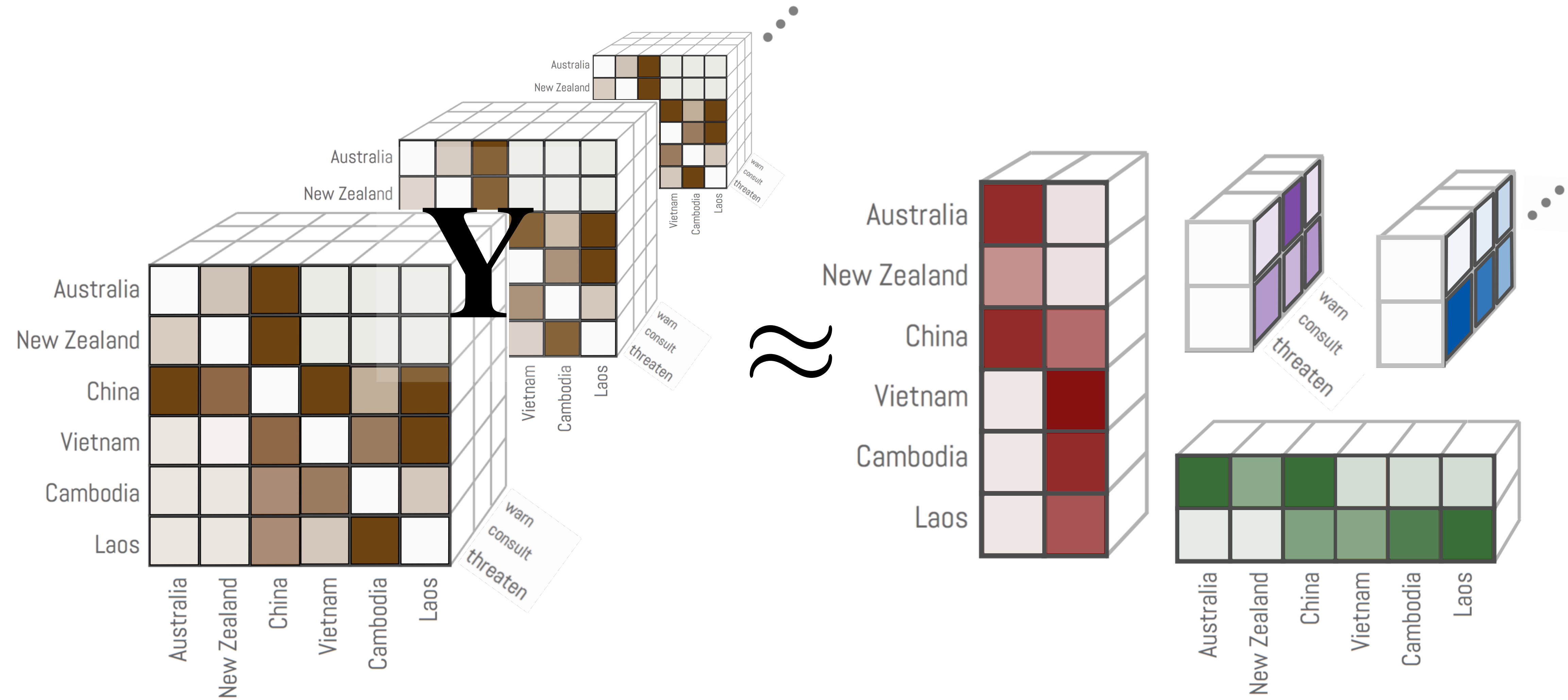
- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

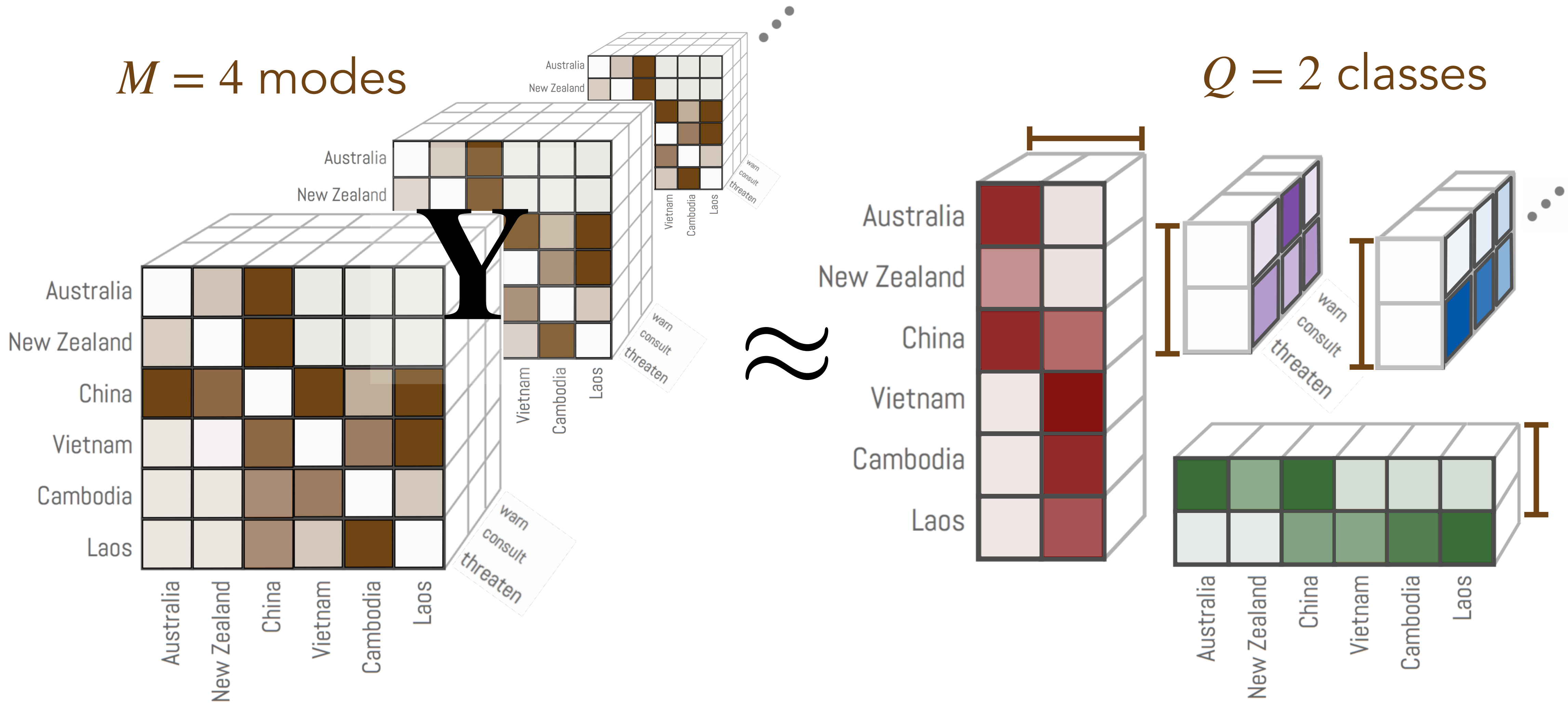
Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

CANONICAL POLYADIC (CP) DECOMPOSITION

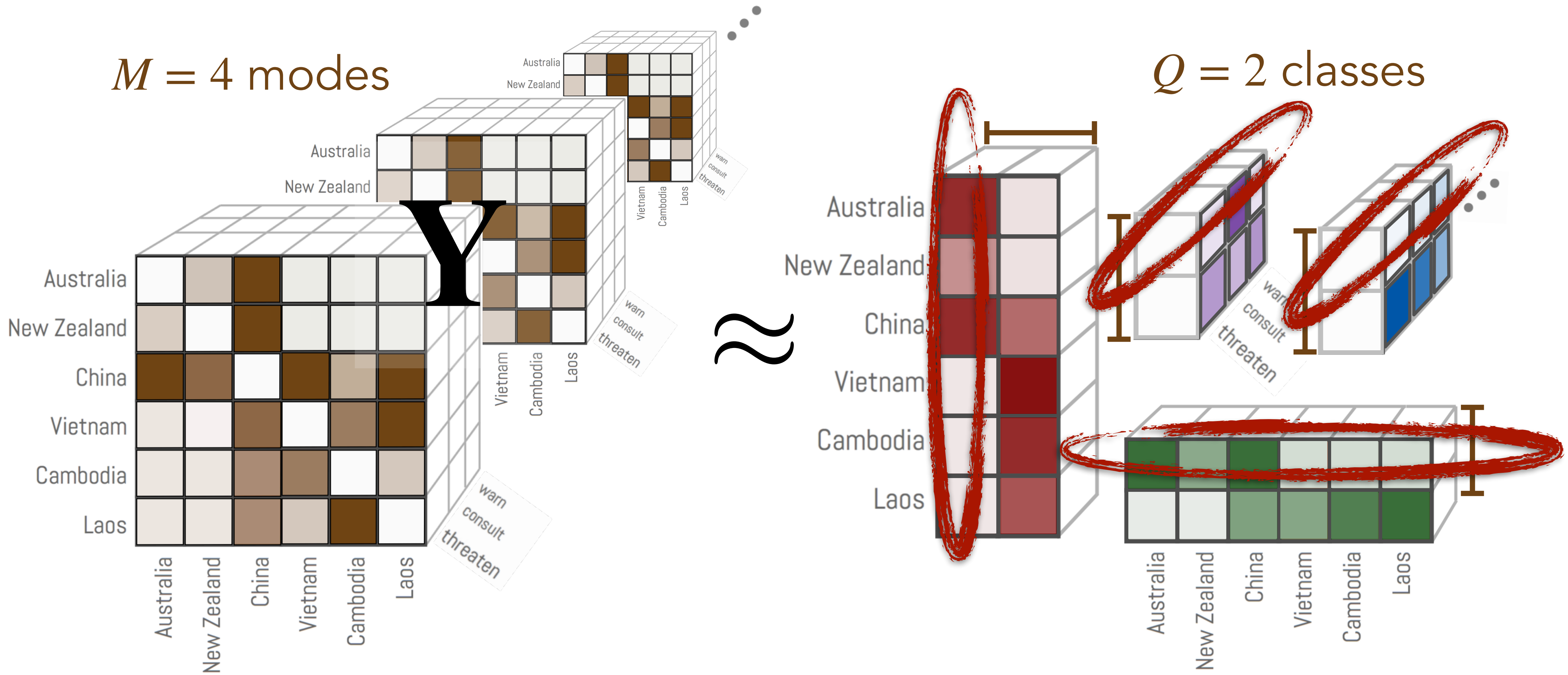


CANONICAL POLYADIC (CP) DECOMPOSITION



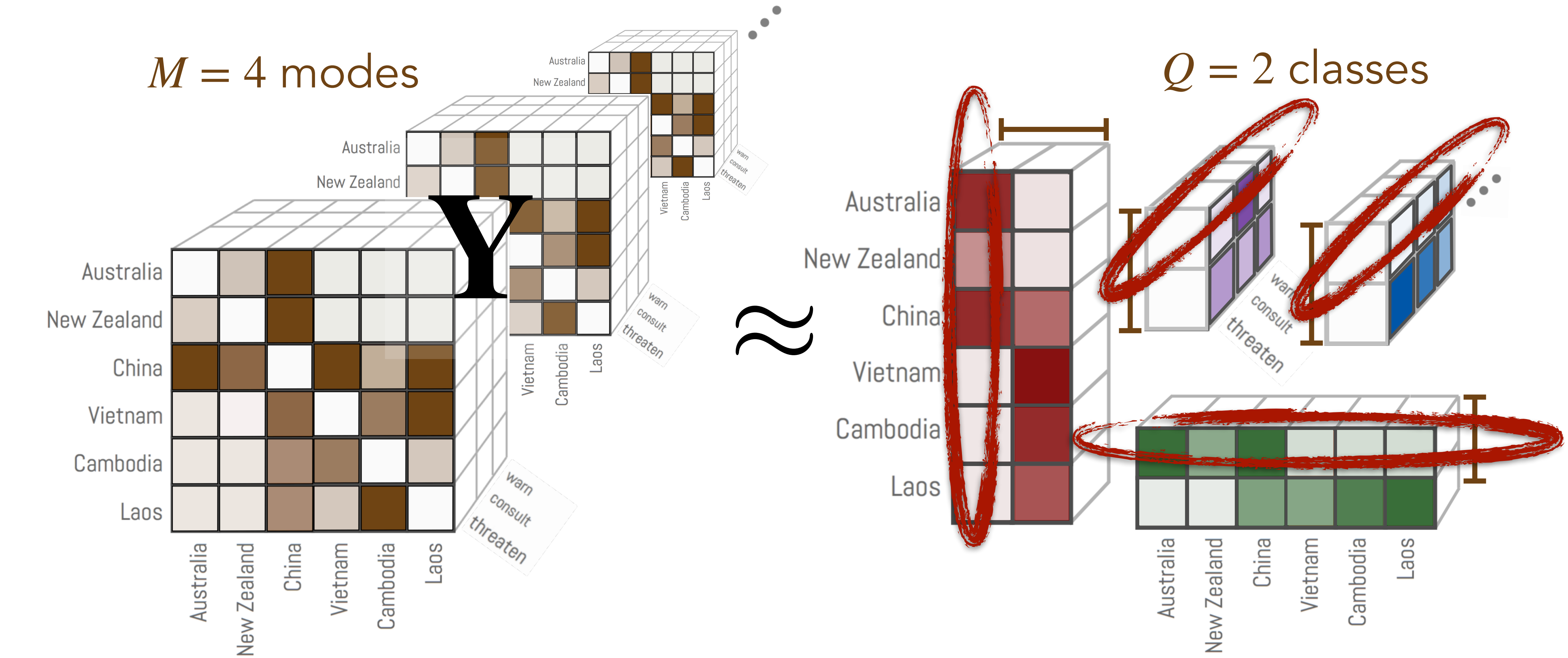
- M-mode tensor $\rightarrow$ M factor matrices
- Q latent "classes" (or "components")

CANONICAL POLYADIC (CP) DECOMPOSITION

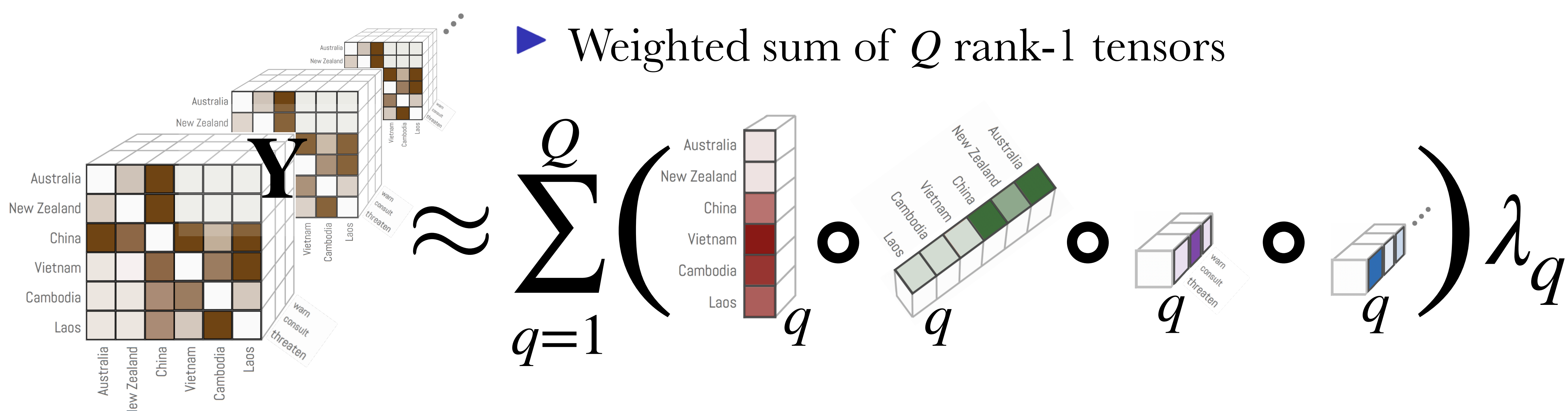


- Each "class" q associated with M "factors"

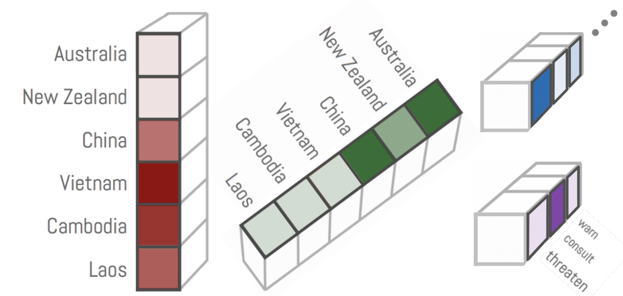
CANONICAL POLYADIC (CP) DECOMPOSITION

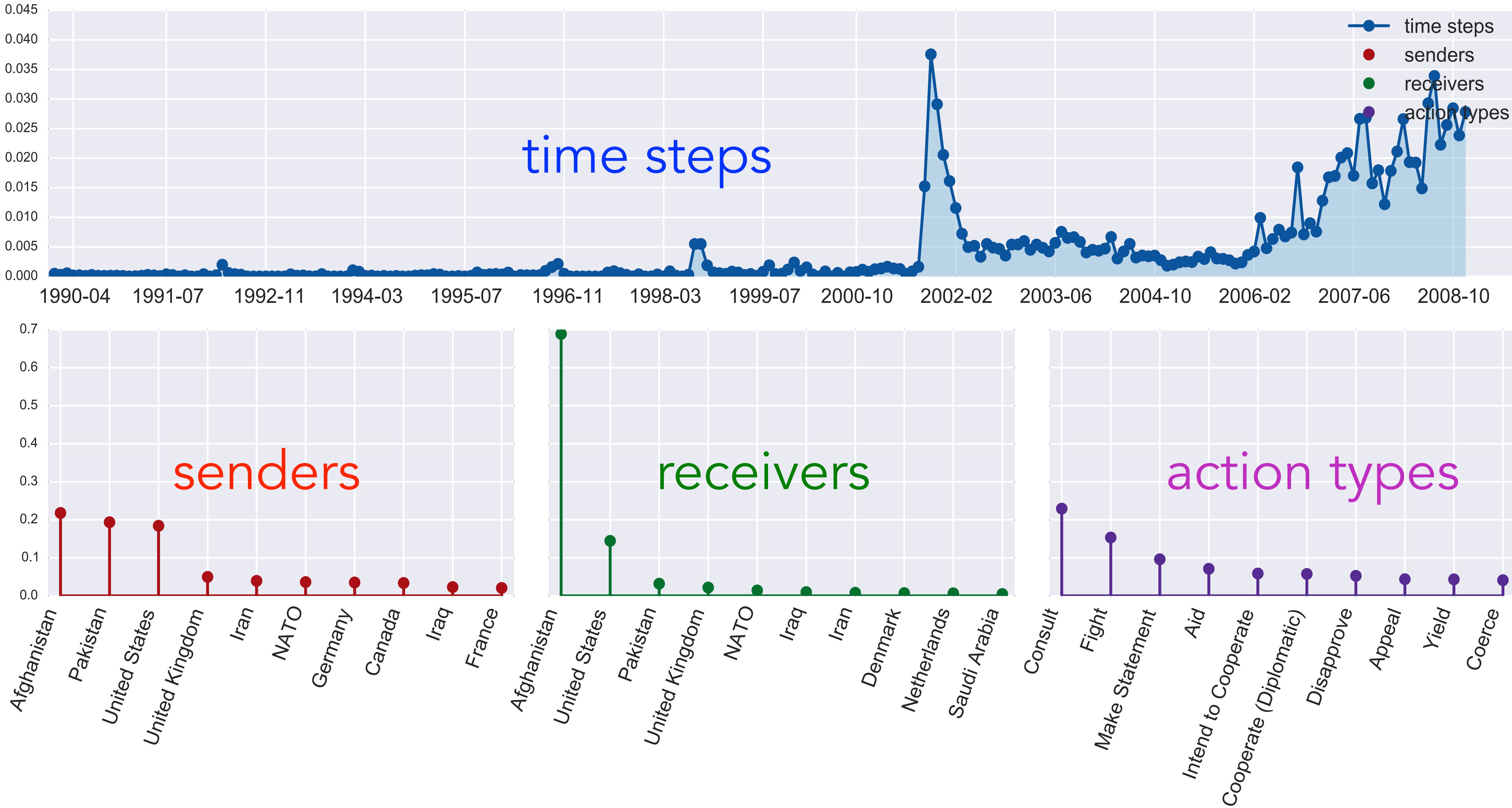


- Each "class" q associated with M "factors"
- Weighted sum of Q rank-1 tensors

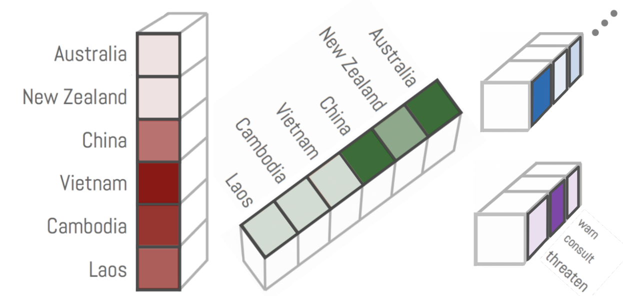


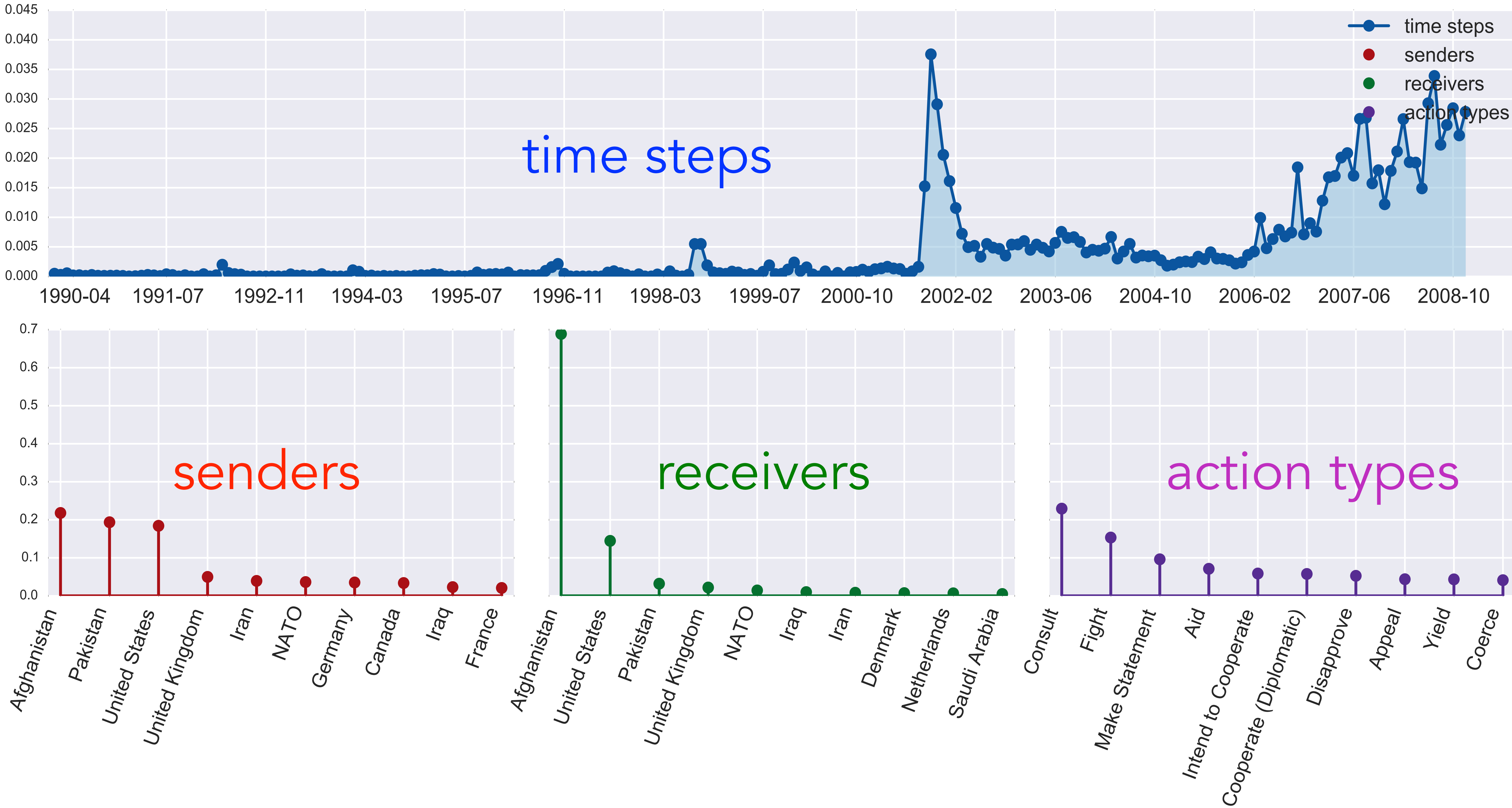
EXAMPLE CLASS



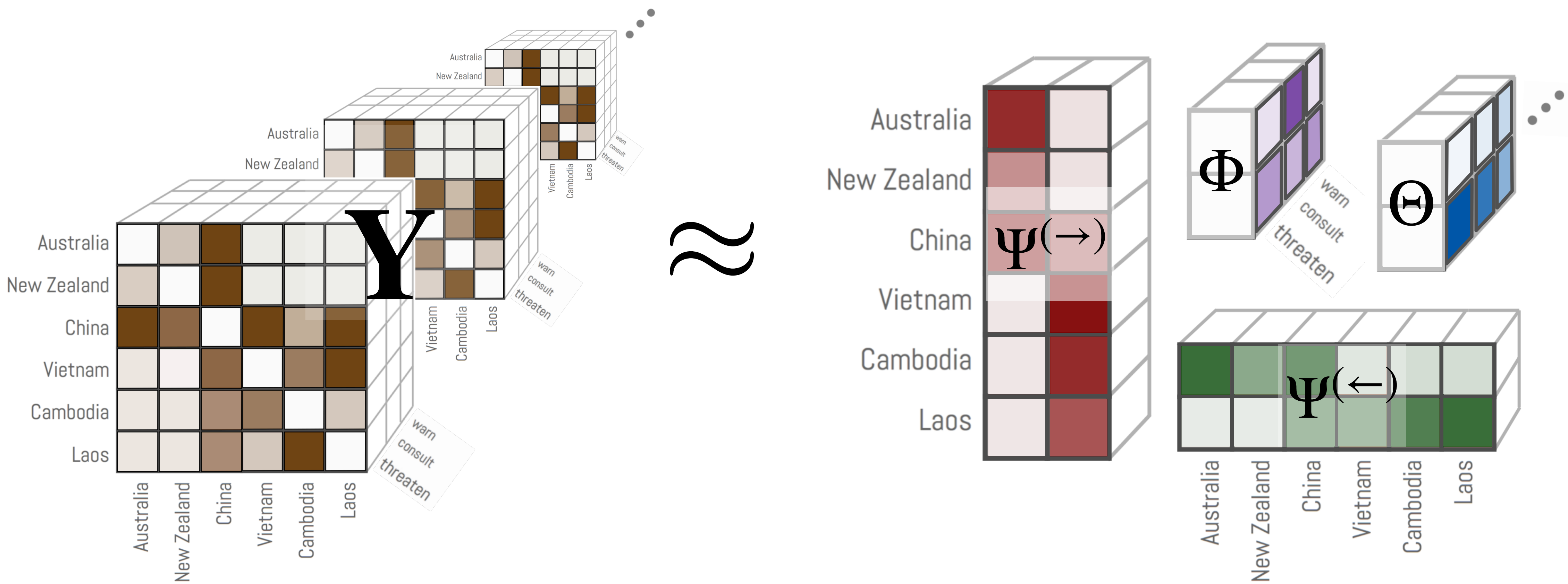


EXAMPLE CLASS: War in Afghanistan

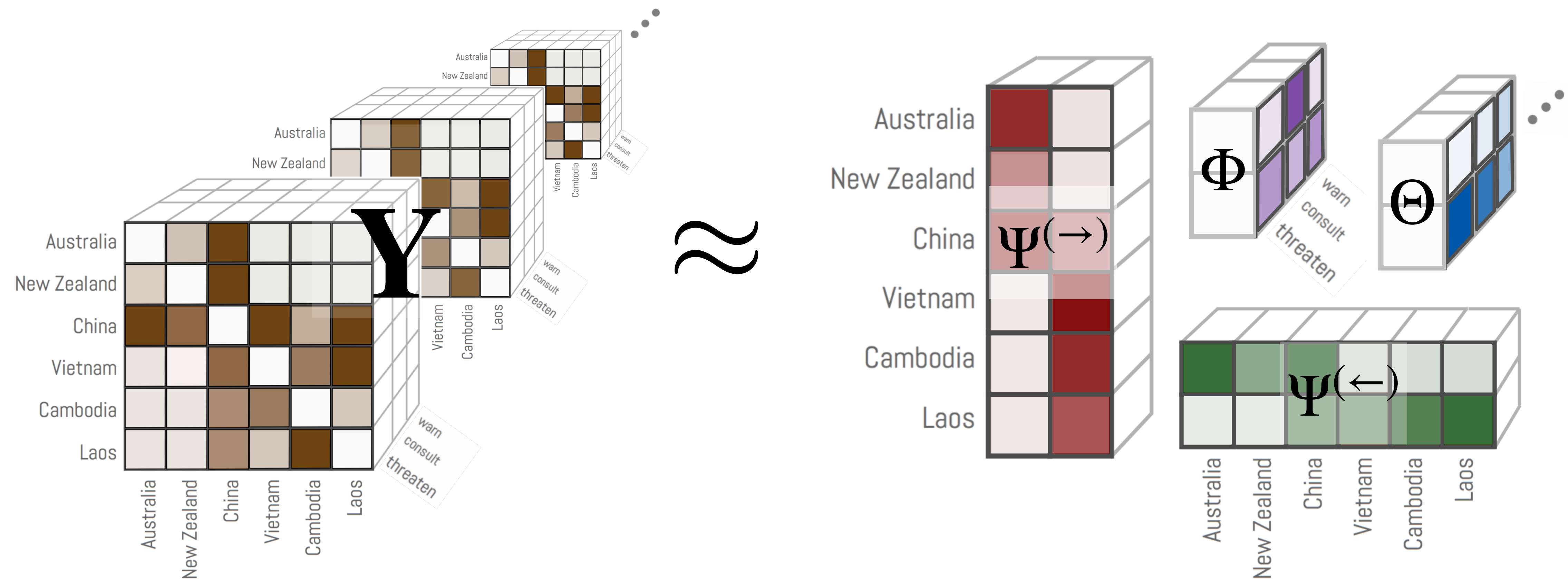




CANONICAL POLYADIC (CP) DECOMPOSITION

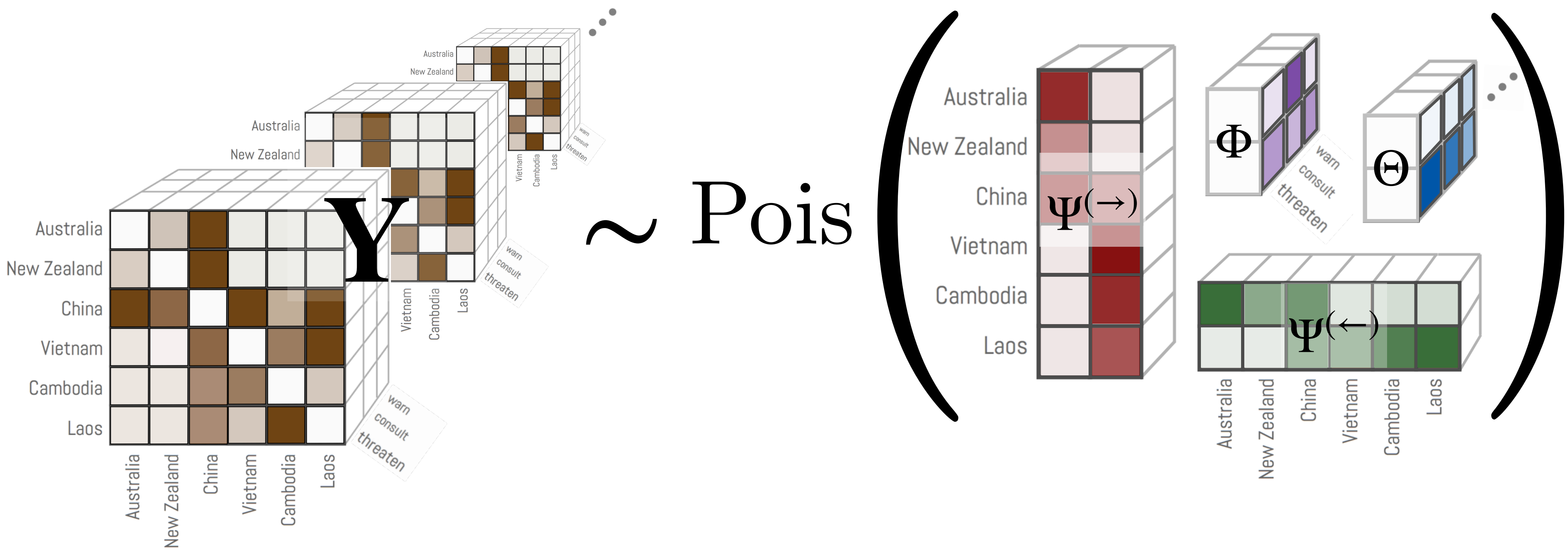


CANONICAL POLYADIC (CP) DECOMPOSITION



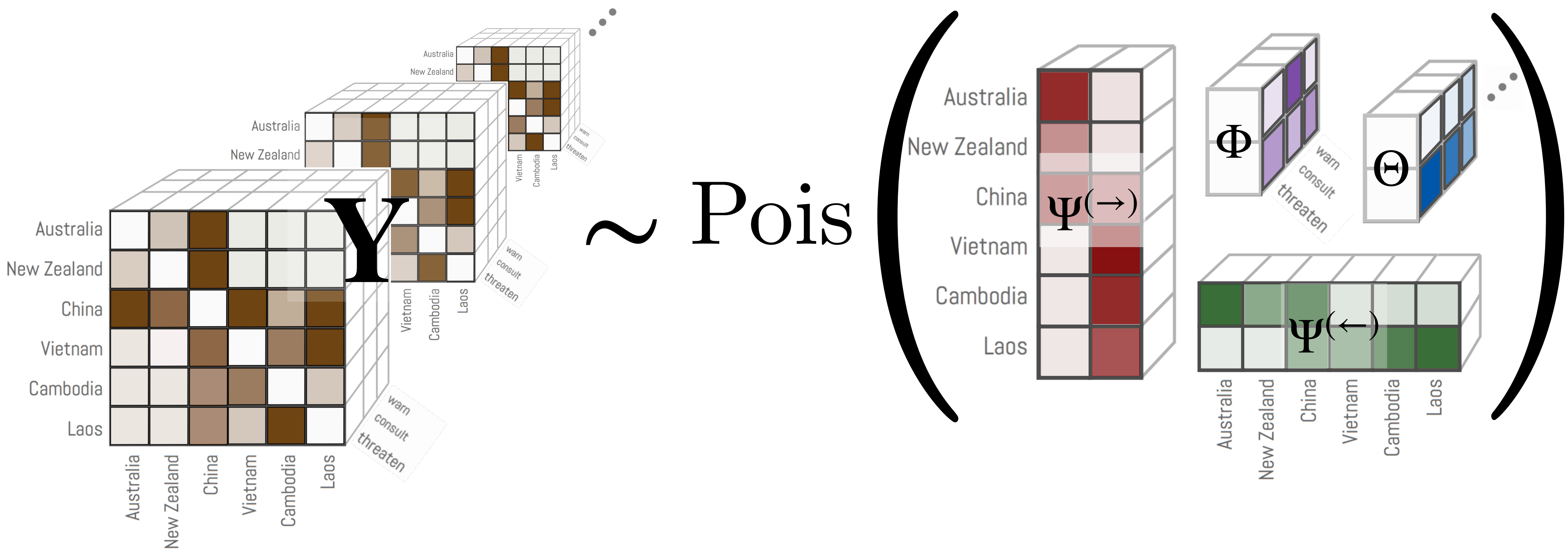
$$y^{(t)}_{i \xrightarrow{a} j} \approx \sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\, \psi^{(\leftarrow)}_{jq}\, \phi_{aq}\, \theta^{(t)}_{q}\, \lambda_q$$

POISSON CP DECOMPOSITION



$$y_{i \xrightarrow{a} j}^{(t)} \sim \mathrm{Pois}\left(\sum_{q=1}^{Q} \psi_{iq}^{(\rightarrow)}\, \psi_{jq}^{(\leftarrow)}\, \phi_{aq}\, \theta_q^{(t)}\, \lambda_q\right)$$

POISSON CP DECOMPOSITION

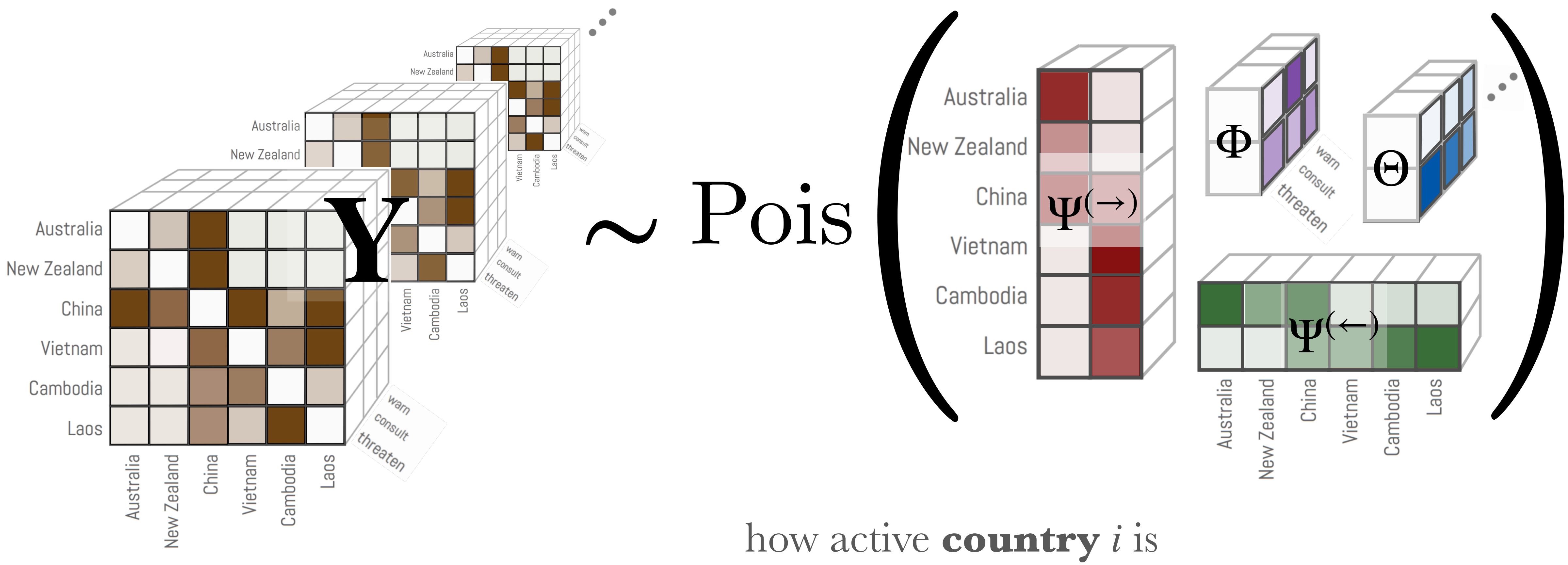


$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\sum_{q=1}^{Q} \psi_{iq}^{(\rightarrow)}\, \psi_{jq}^{(\leftarrow)}\, \phi_{aq}\, \theta_q^{(t)}\, \lambda_q\right)$$

non-negative

"parts-based" representation [Lee and Seung, 1999]

POISSON CP DECOMPOSITION



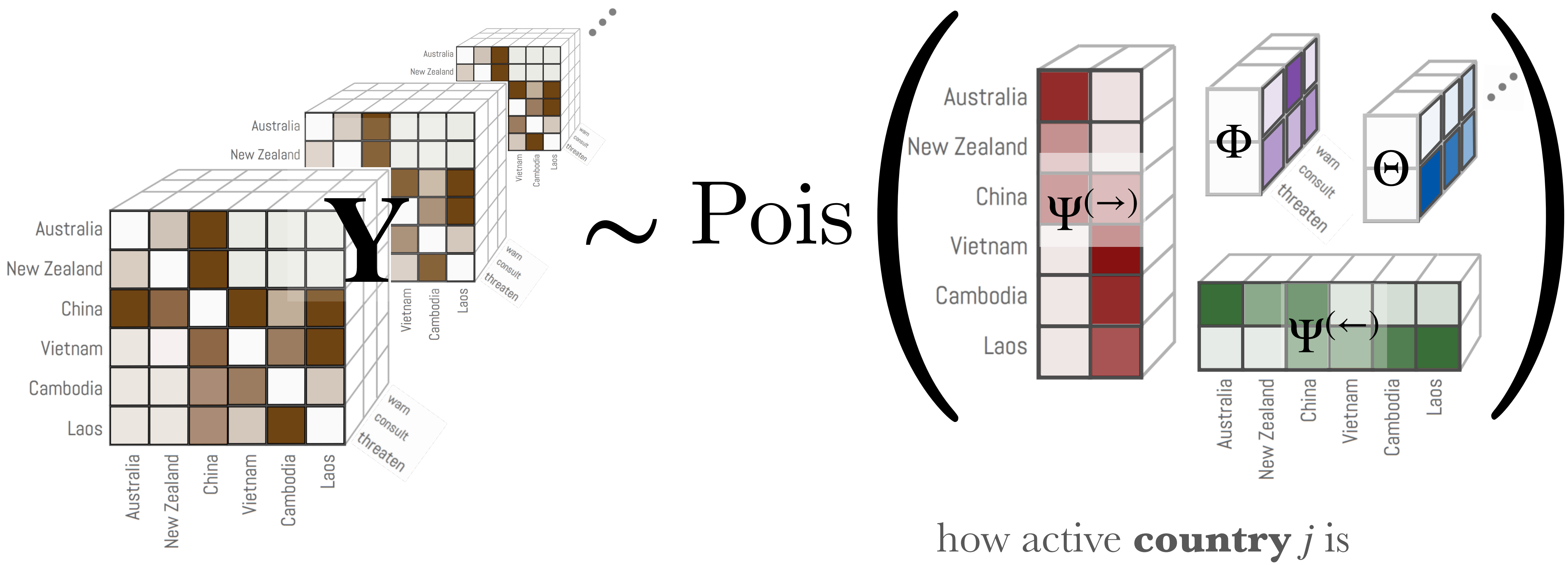
how active **country** i is *as a sender* in **class** q

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\, \psi^{(\leftarrow)}_{jq}\, \phi_{aq}\, \theta^{(t)}_{q}\, \lambda_q\right)$$

non-negative

“parts-based” representation [Lee and Seung, 1999]

POISSON CP DECOMPOSITION



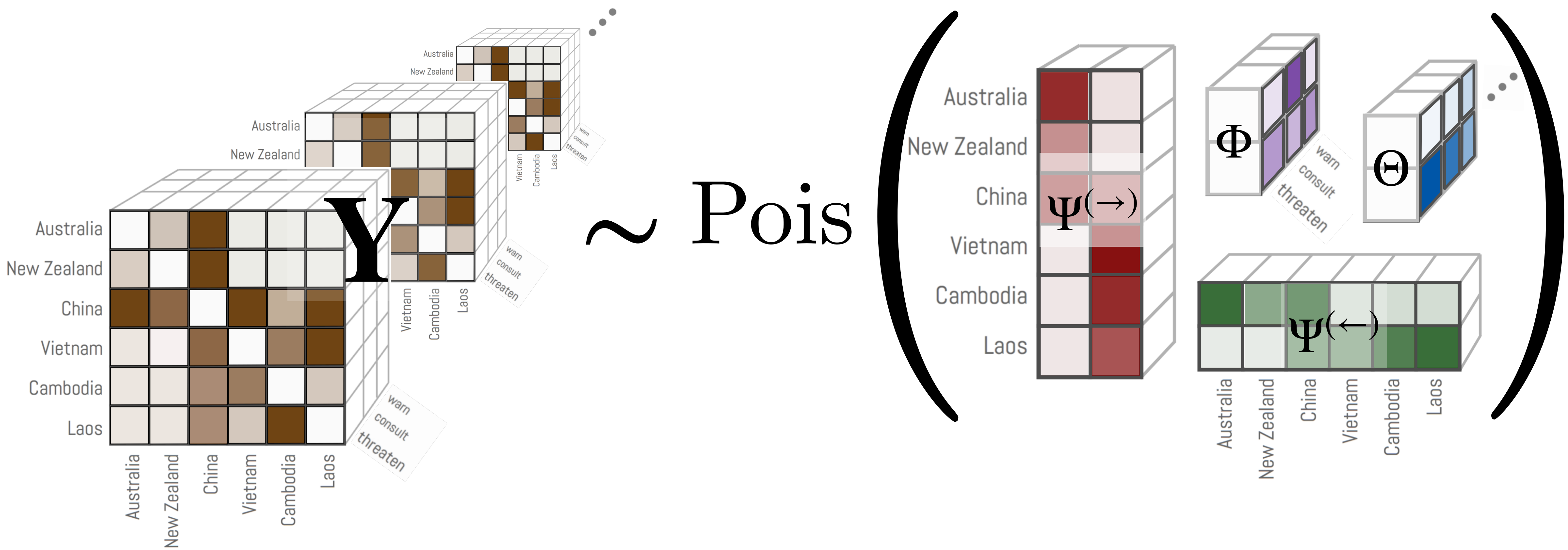
how active **country** j is
as a receiver in **class** q

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\; \psi^{(\leftarrow)}_{jq}\; \phi_{aq}\, \theta^{(t)}_{q}\; \lambda_q\right)$$

non-negative

"parts-based" representation [Lee and Seung, 1999]

POISSON CP DECOMPOSITION



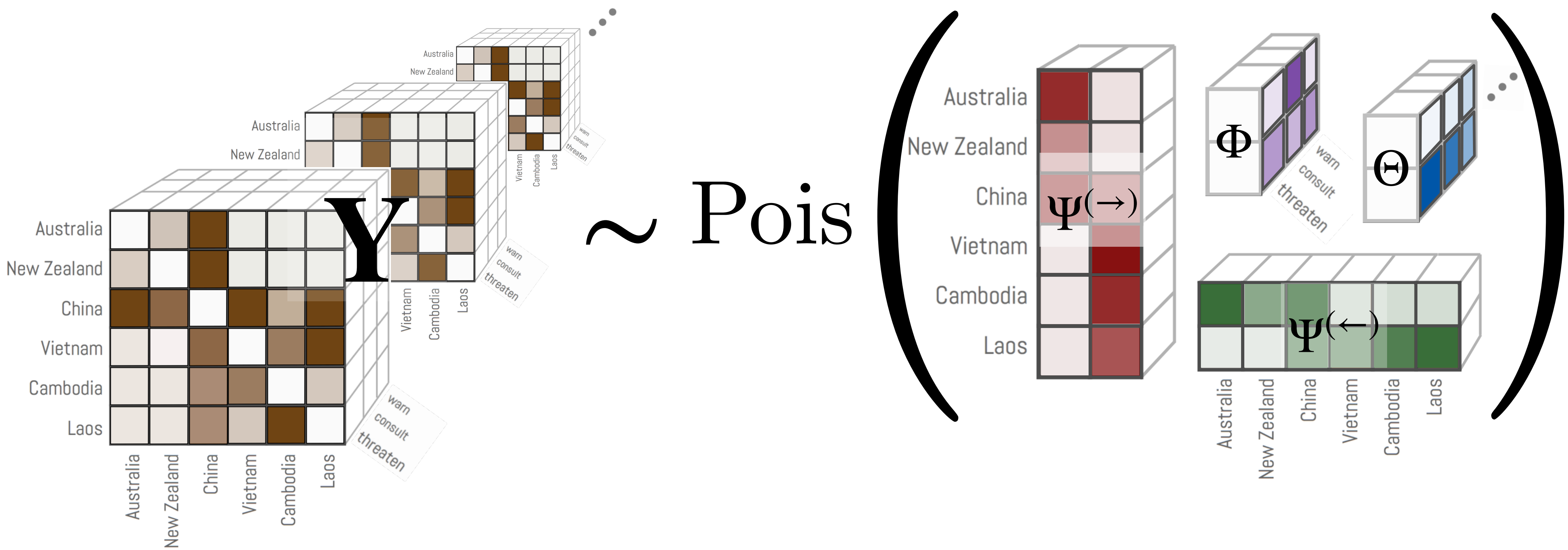
how prevalent **action** a is in **class** q

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\, \psi^{(\leftarrow)}_{jq}\, \phi_{aq}\, \theta^{(t)}_{q}\, \lambda_q\right)$$

non-negative

“parts-based” representation [Lee and Seung, 1999]

POISSON CP DECOMPOSITION



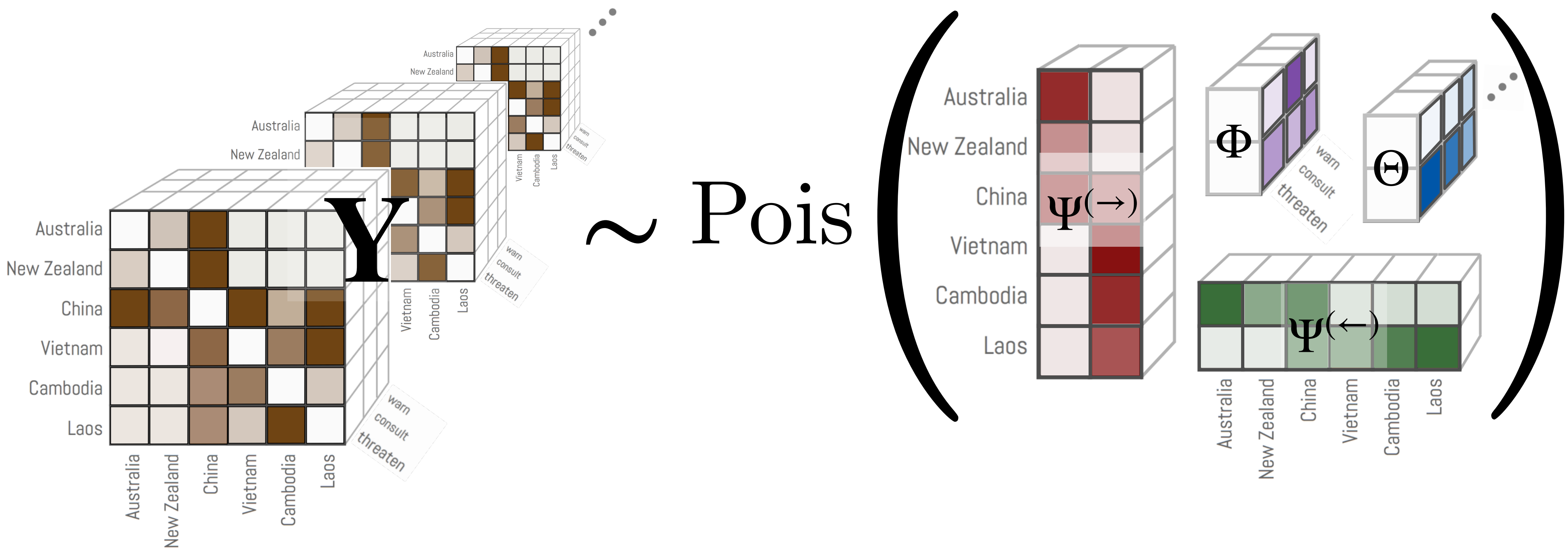
how active **class** q is
at **time** t

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\, \psi^{(\leftarrow)}_{jq}\, \phi_{aq}\, \theta^{(t)}_{q}\, \lambda_q\right)$$

non-negative

"parts-based" representation [Lee and Seung, 1999]

POISSON CP DECOMPOSITION



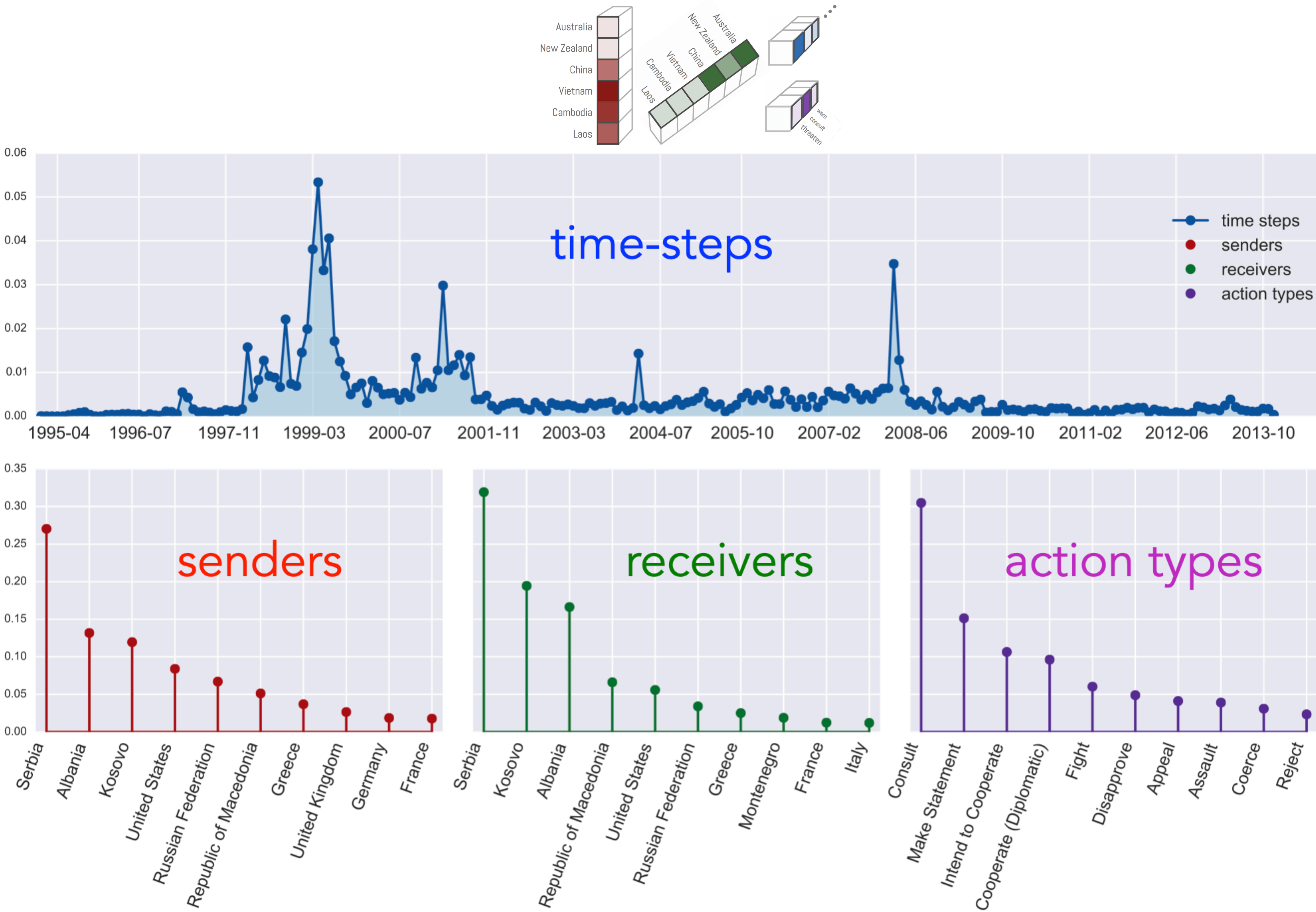
how active **class** q is overall

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\, \psi^{(\leftarrow)}_{jq}\, \phi_{aq}\, \theta^{(t)}_{q}\, \lambda_q\right)$$

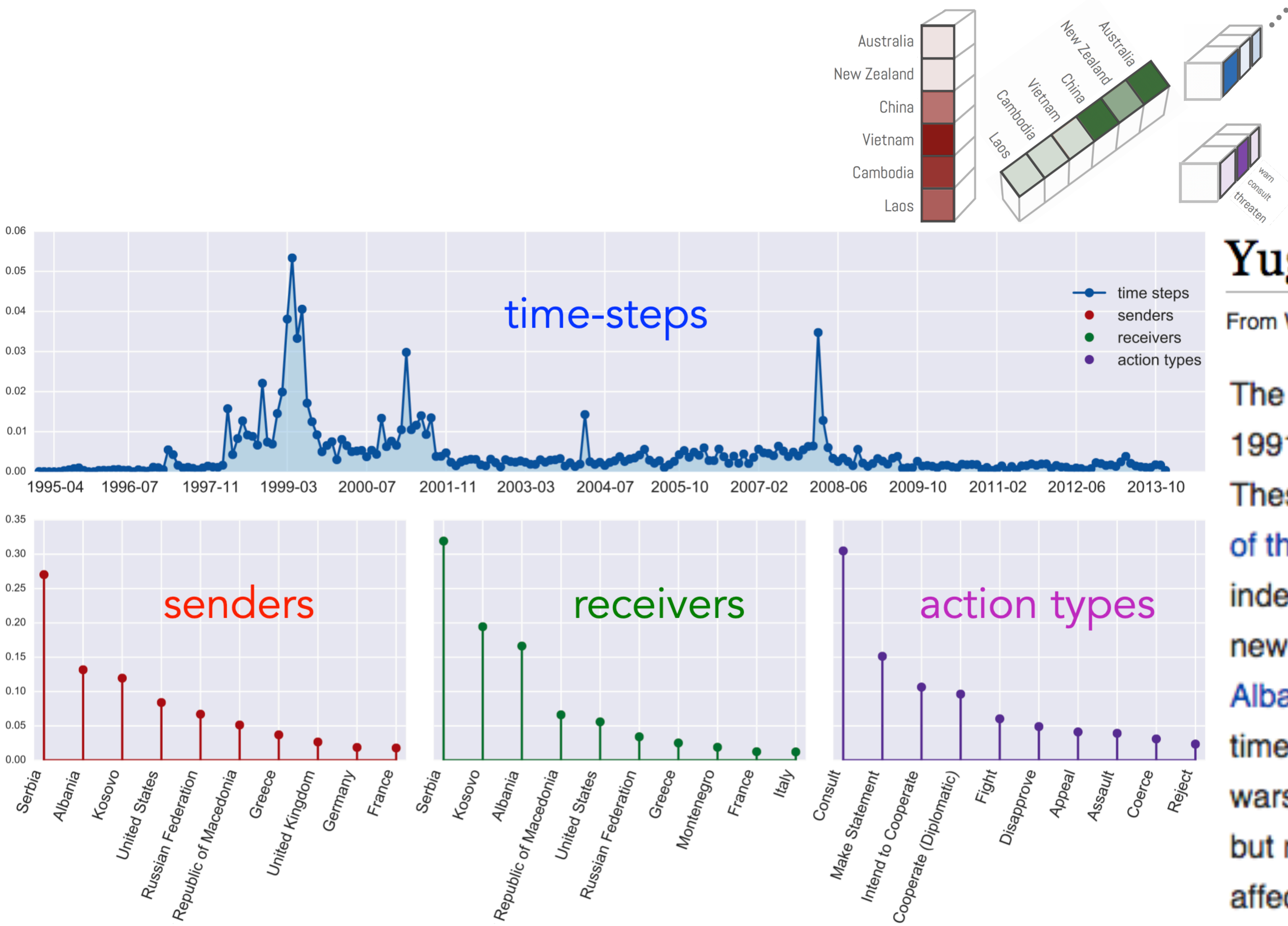
non-negative

“parts-based” representation [Lee and Seung, 1999]

EXAMPLE CLASS:



EXAMPLE CLASS: Yugoslav Wars



Yugoslav Wars

From Wikipedia, the free encyclopedia

The **Yugoslav Wars** were ethnic conflicts fought from 1991 to 2001 inside the territory of the former Yugoslavia. These wars accompanied and/or facilitated the breakup of the country, when its constituent republics declared independence, but the issues of ethnic minorities in the new countries (chiefly Serbs in central parts and Albanians in the southeast) were still unresolved at the time the republics were recognized internationally. The wars are generally considered to be a series of separate but related military conflicts which, occurred in, and affected most of the former Yugoslav republics:[2][3][4]

- War in Slovenia (1991)
- Croatian War of Independence (1991–1995)
- Bosnian War (1992–1995)
- Kosovo War (1998–1999), including the NATO bombing of Yugoslavia
- Insurgency in the Preševo Valley (1999–2001)
- Insurgency in the Republic of Macedonia (2001)

Classes almost always describe easily recognizable **multilateral** relations / events

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

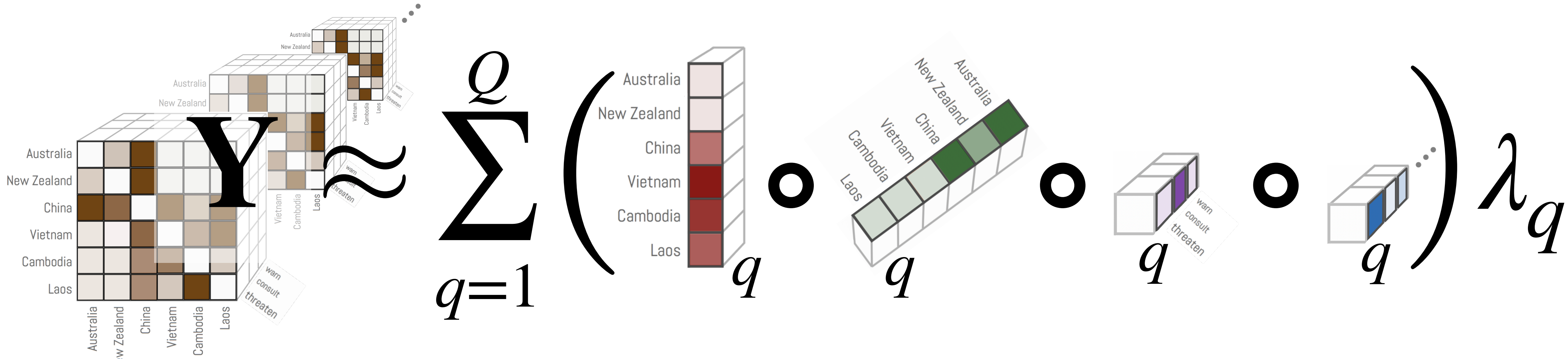
Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

TUCKER DECOMPOSITION

- CP is unappealing if mode dimensions are very different
- e.g., (country x country x action x month) for 1995-2013 ICEWS data

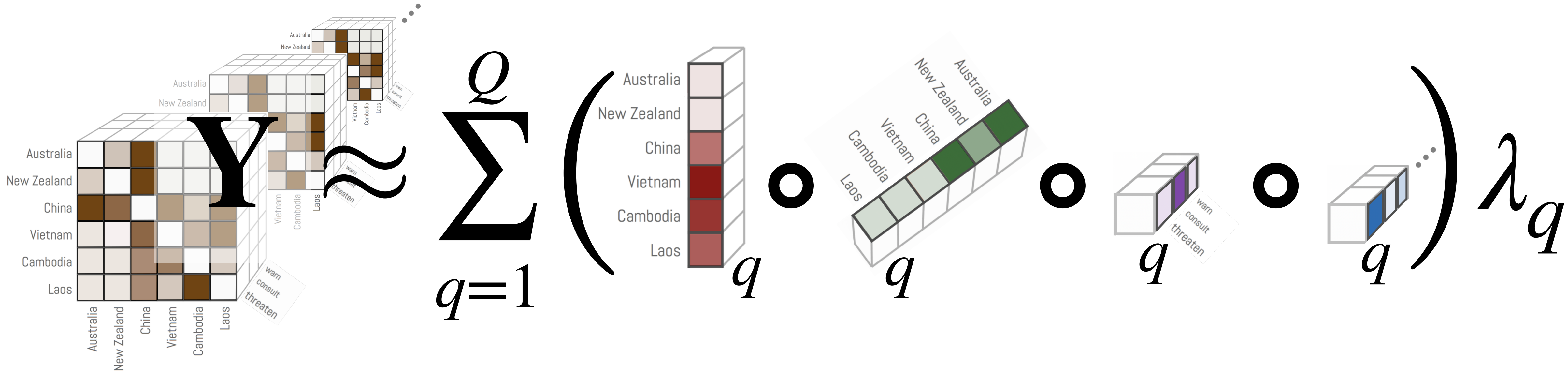
  = **(250 x 250 x 20 x 216)**
- Q = 100 classes; each must learn a different 20-dimensional action factor



TUCKER DECOMPOSITION

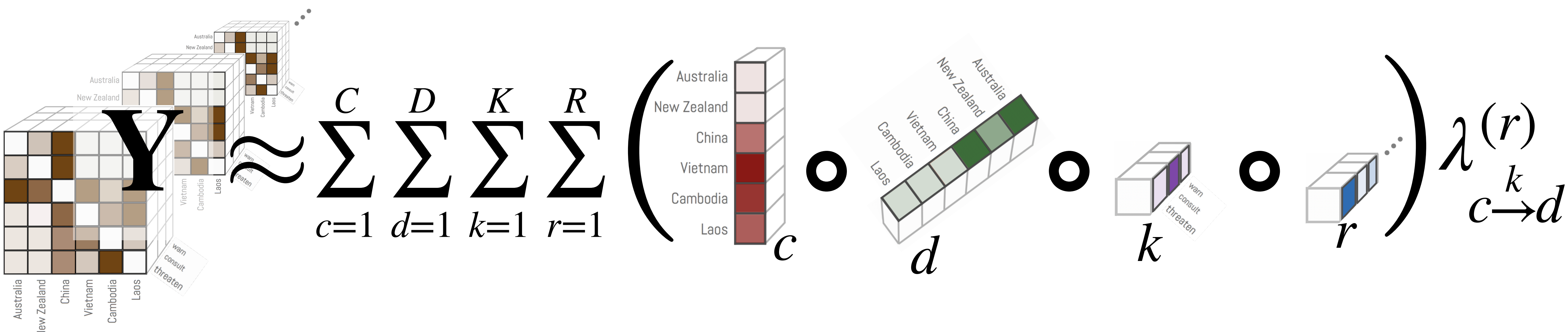
- CP is unappealing if mode dimensions are very different
- e.g., (country x country x action x month) for 1995-2013 ICEWS data

  = **(250 x 250 x 20 x 216)**
- Q = 100 classes; each must learn a different 20-dimensional action factor



$$\mathbf{Y} \approx \sum_{q=1}^{Q} \left(q \circ q \circ q \circ q \right) \lambda_q$$

- Tucker decomposition fixes this…

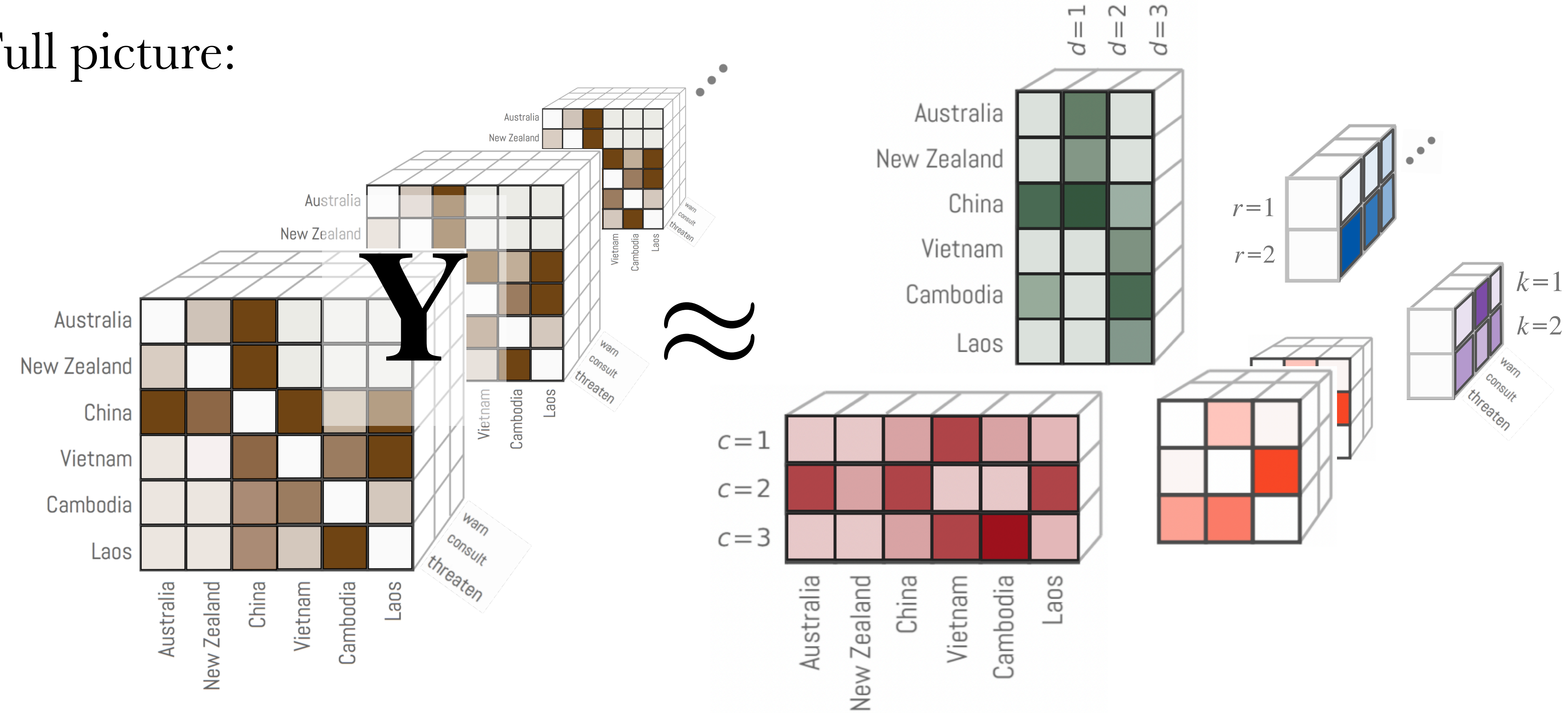


$$\mathbf{Y} \approx \sum_{c=1}^{C} \sum_{d=1}^{D} \sum_{k=1}^{K} \sum_{r=1}^{R} \left(c \circ d \circ k \circ r \right) \lambda^{(r)}_{c \xrightarrow{k} d}$$

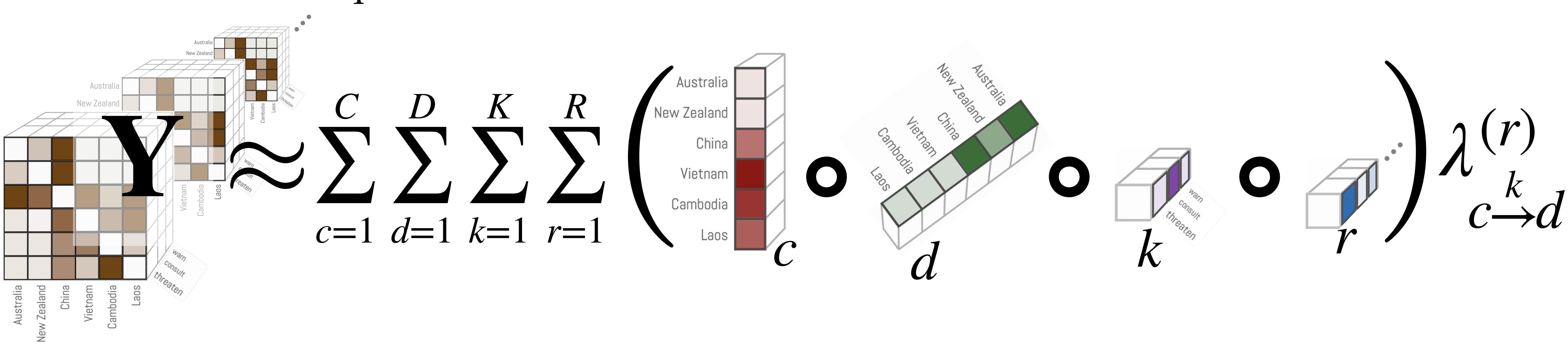
- Every class is now a unique combination (c, d, k, r) of **shared factors**

TUCKER DECOMPOSITION

- Full picture:



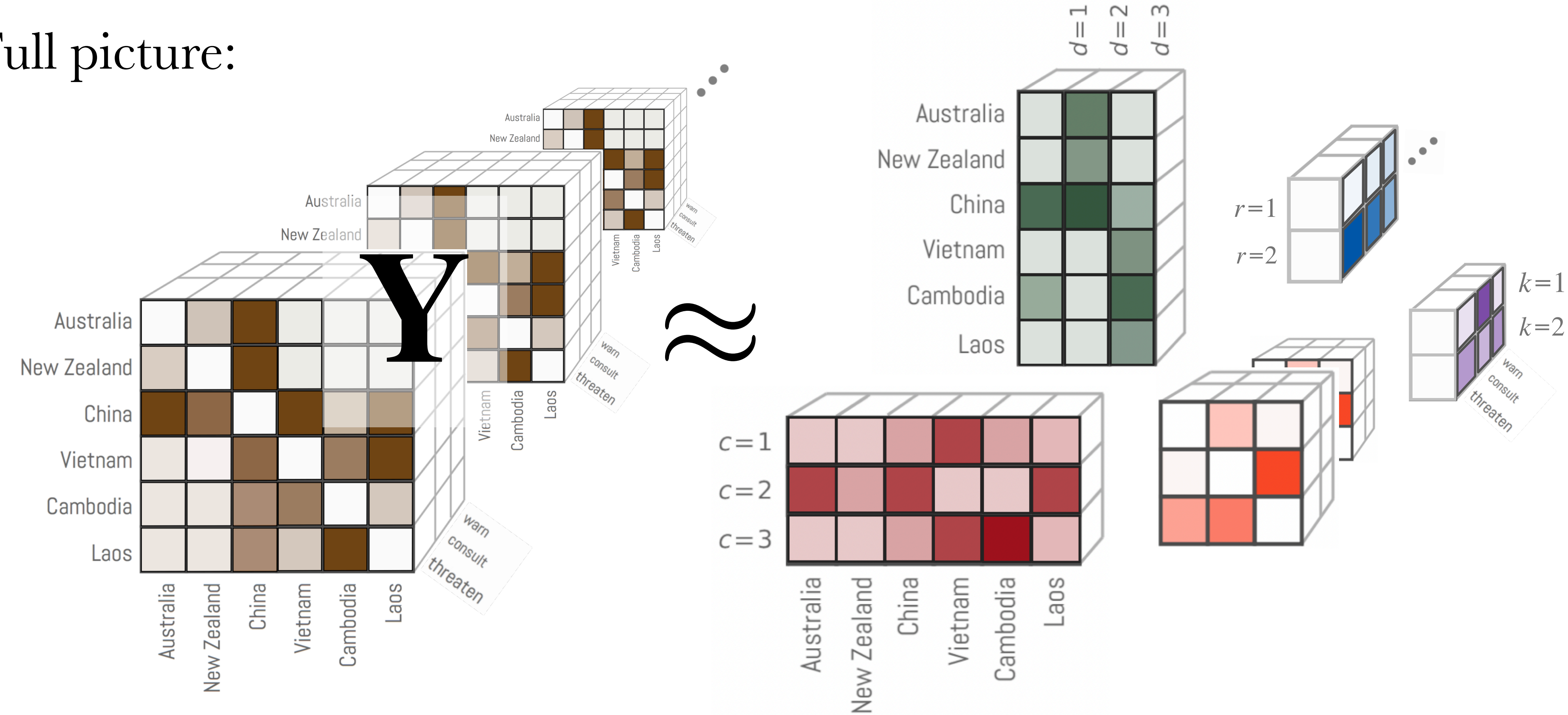
- Tucker decomposition fixes this…



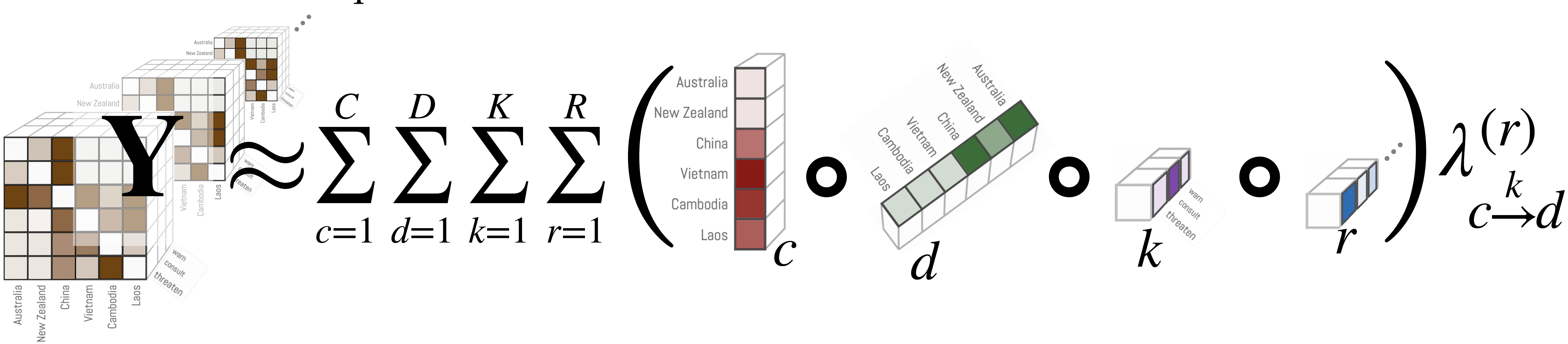
- Every class is now a unique combination (c, d, k, r) of **shared factors**

TUCKER DECOMPOSITION

- Full picture:



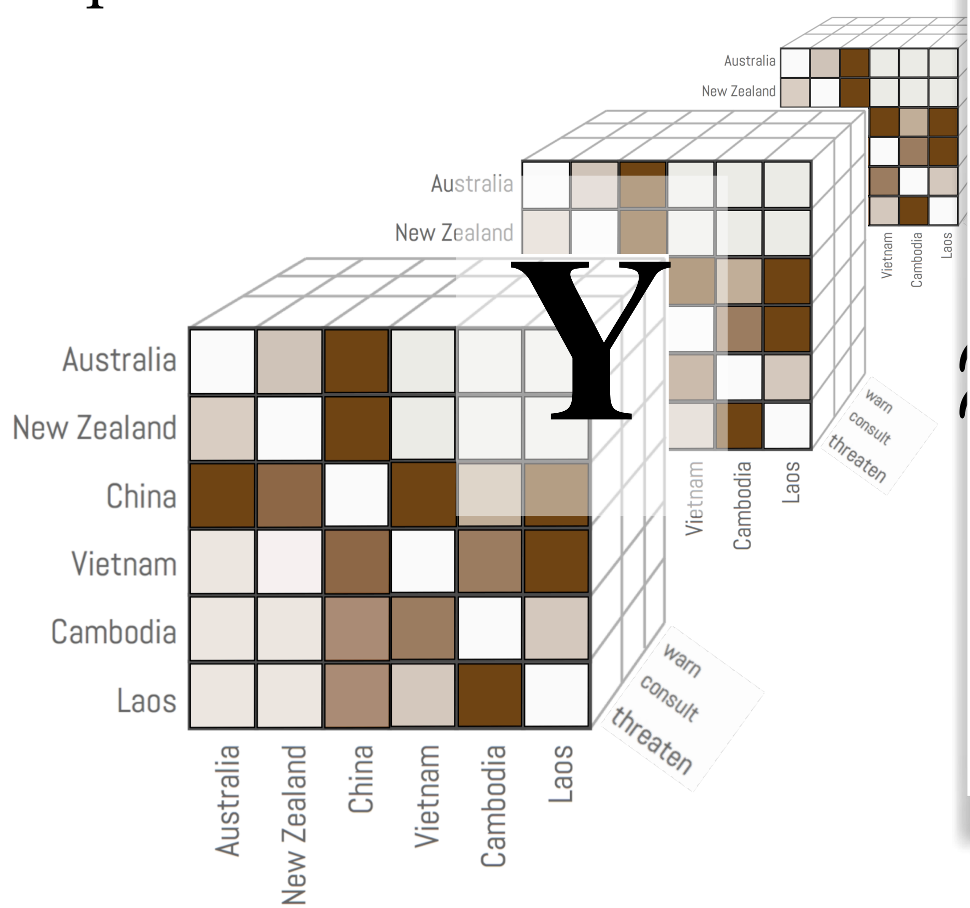
- Tucker decomposition fixes this… different number of factors in each mode



- Every class is now a unique combination (c, d, k, r) of **shared factors**

TUCKER DECOMPOSITION

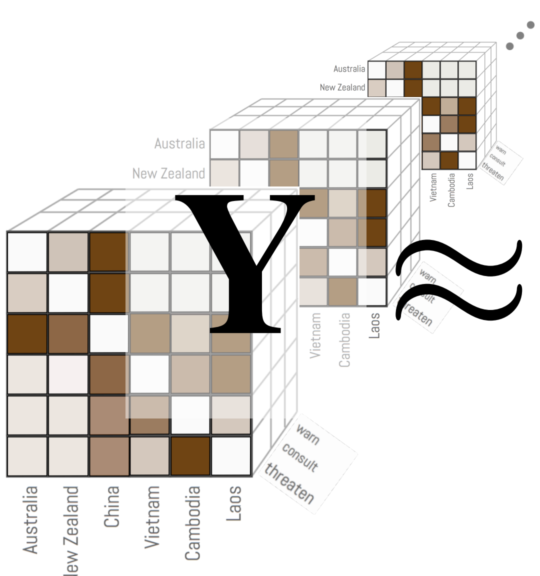
- Full picture:



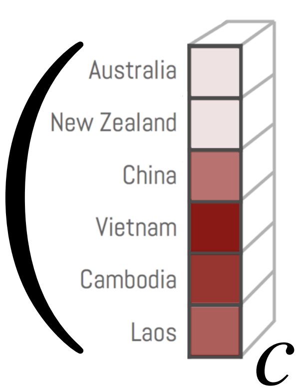
Convergent lines of research on Tucker for multilayer networks / dyadic event data:

- Tucker (regression) for dyadic event data: […] Minhas, Hoff, & Ward (2015), Hoff (2015), Hoff (2016), […]
- Non-negative Tucker for dyadic events: Schein et al. (2016)
- Non-negative Tucker for multilayer networks: De Bacco et al. (2017), Aguiar et al. (2023)

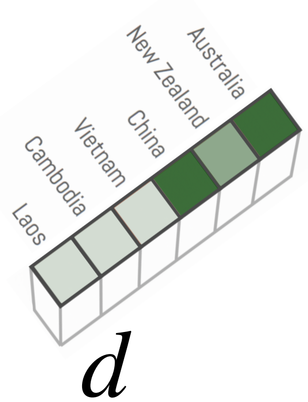
- Tucker decomposition fixes this… different number of factors in each mode



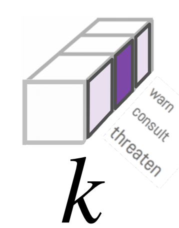
$$\approx \sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R}\Big($$



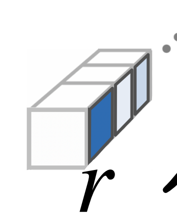
$\circ$



$\circ$



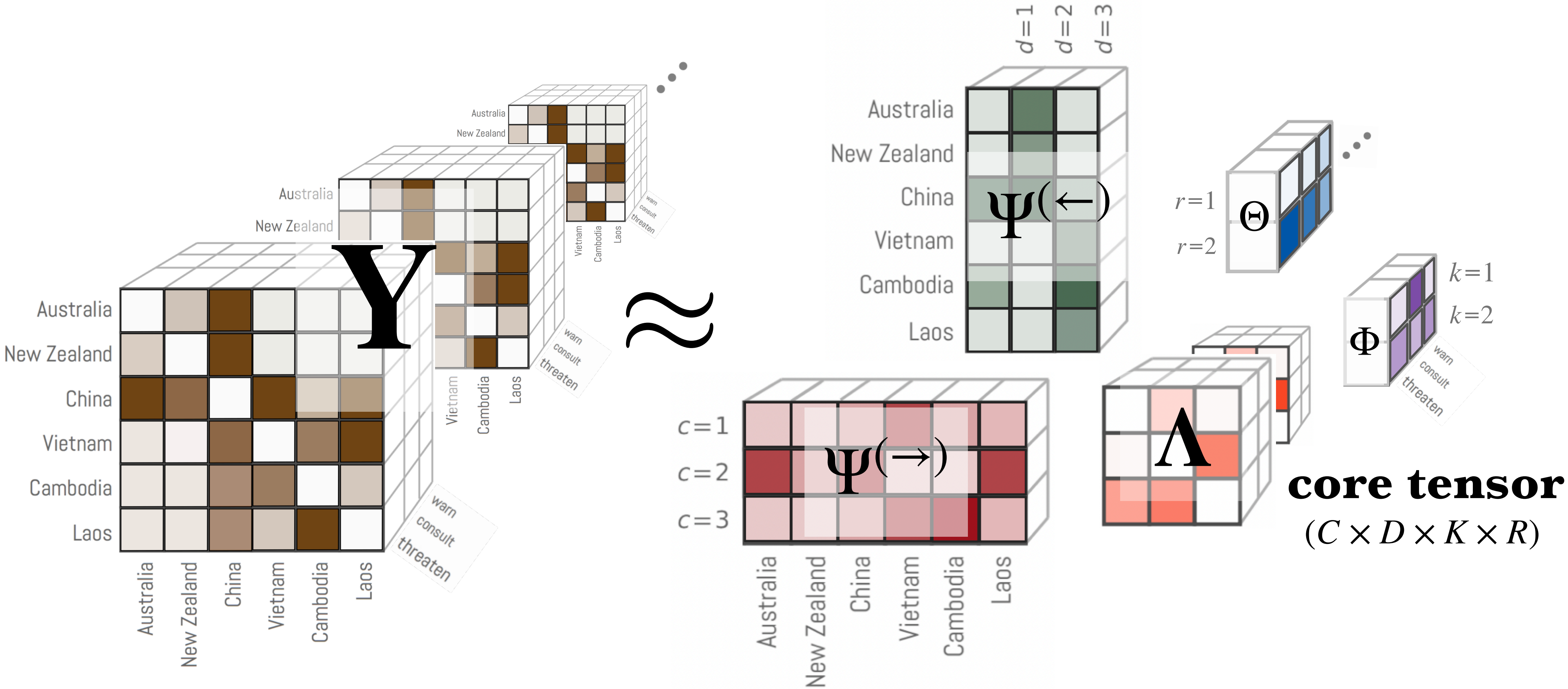
$\circ$



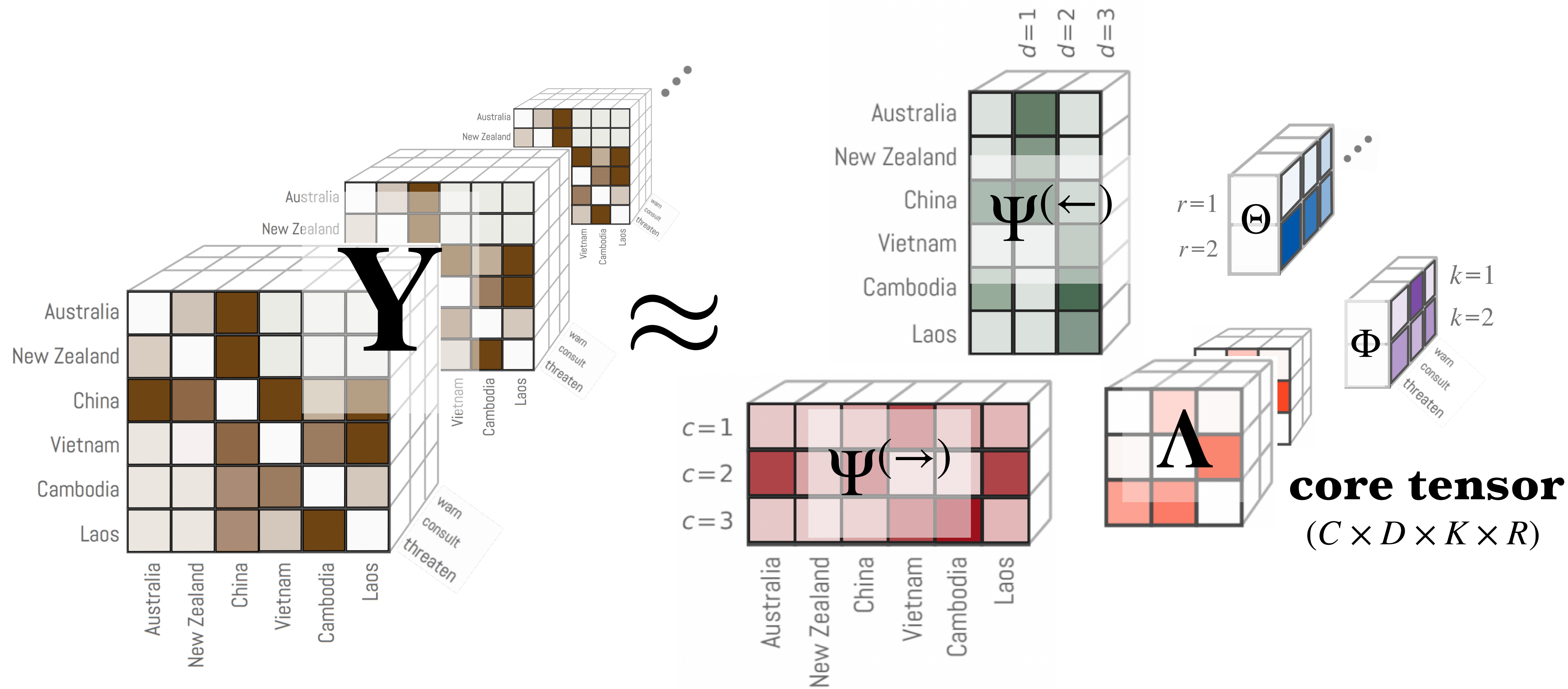
$$\Big)\lambda^{(r)}_{c\xrightarrow{k}d}$$

- Every class is now a unique combination (c, d, k, r) of **shared factors**

POISSON TUCKER DECOMPOSITION

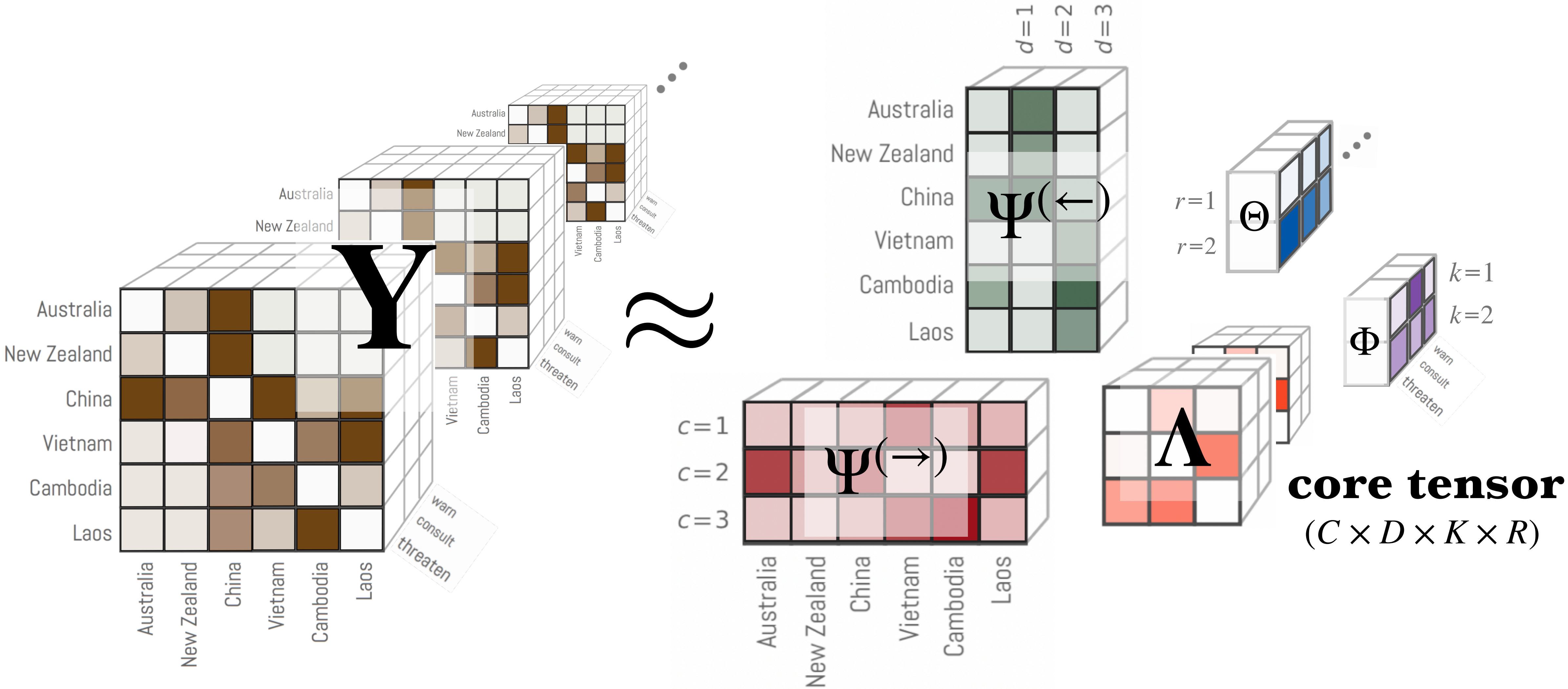


POISSON TUCKER DECOMPOSITION



- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

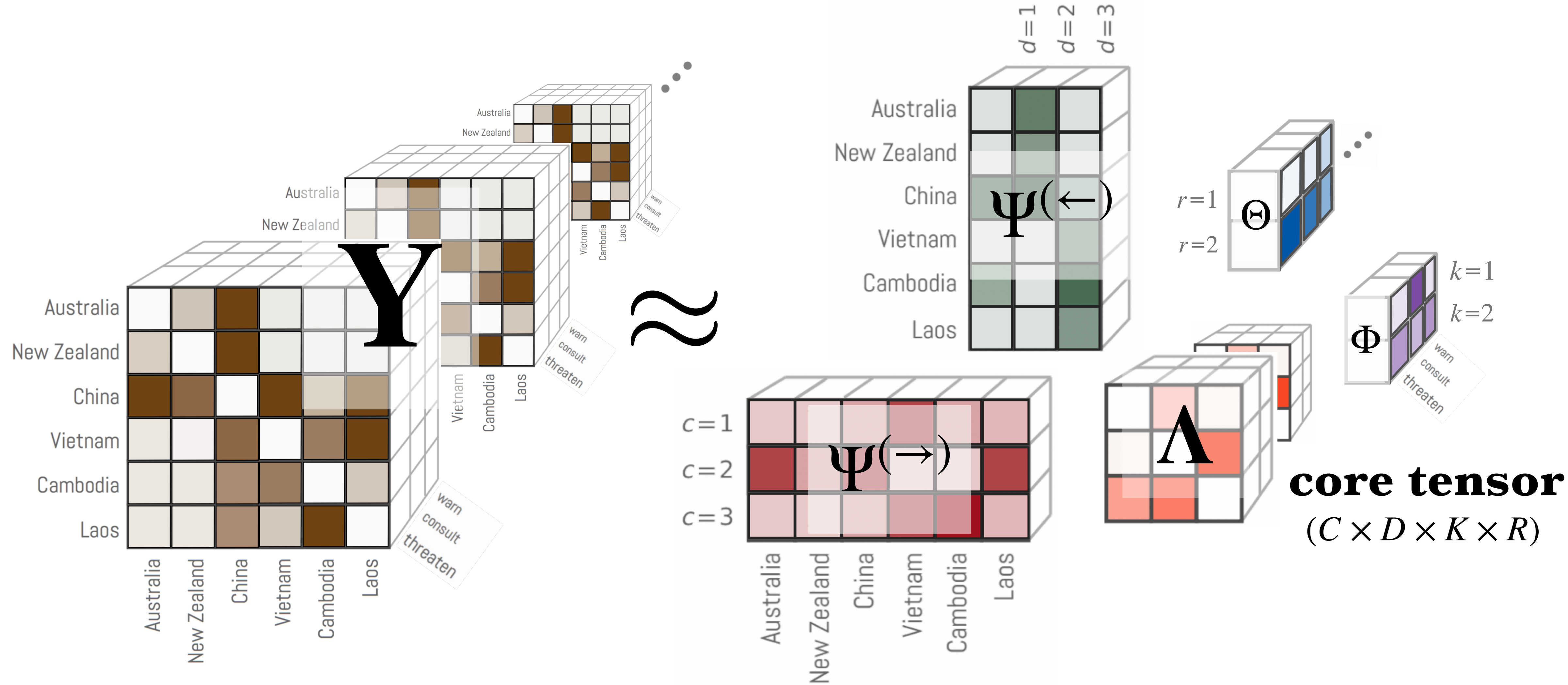
POISSON TUCKER DECOMPOSITION



- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

POISSON TUCKER DECOMPOSITION

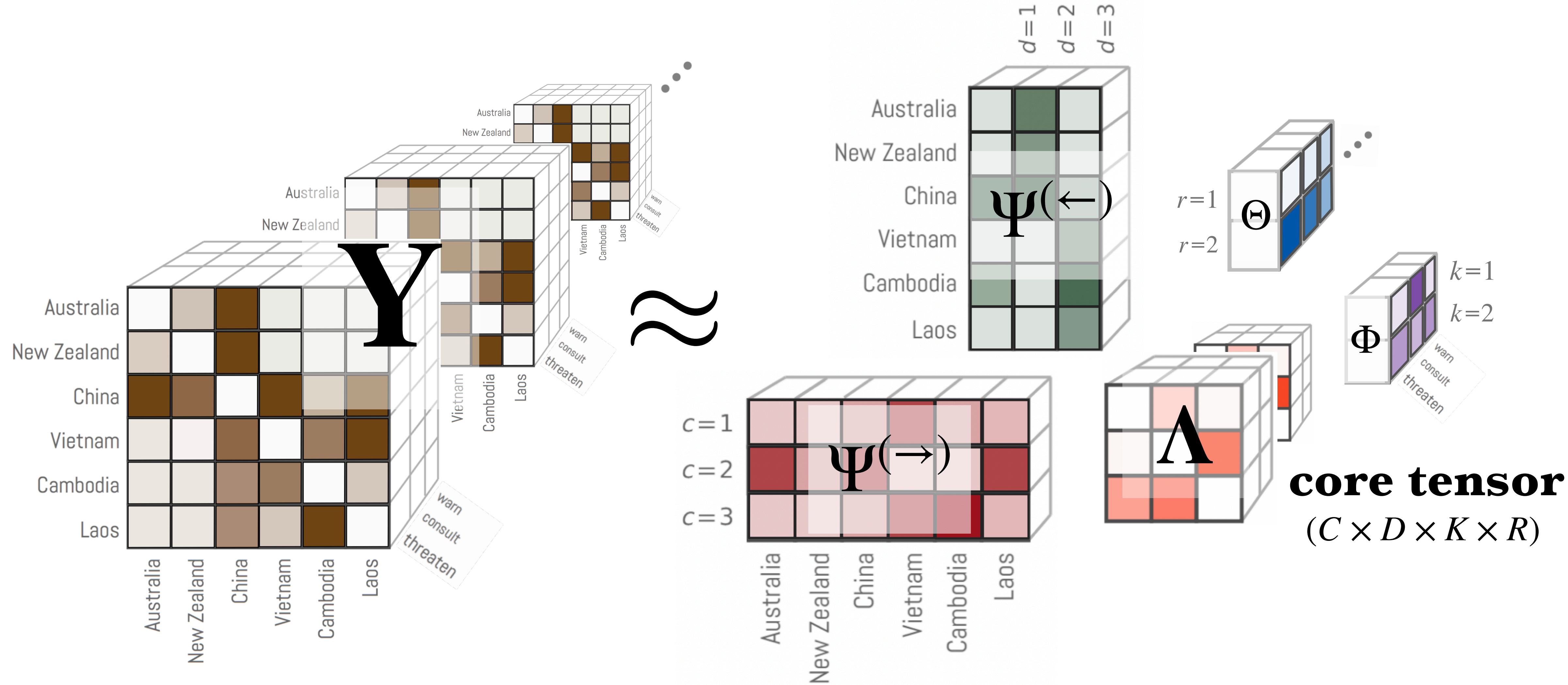


▶ **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

non-negative

POISSON TUCKER DECOMPOSITION

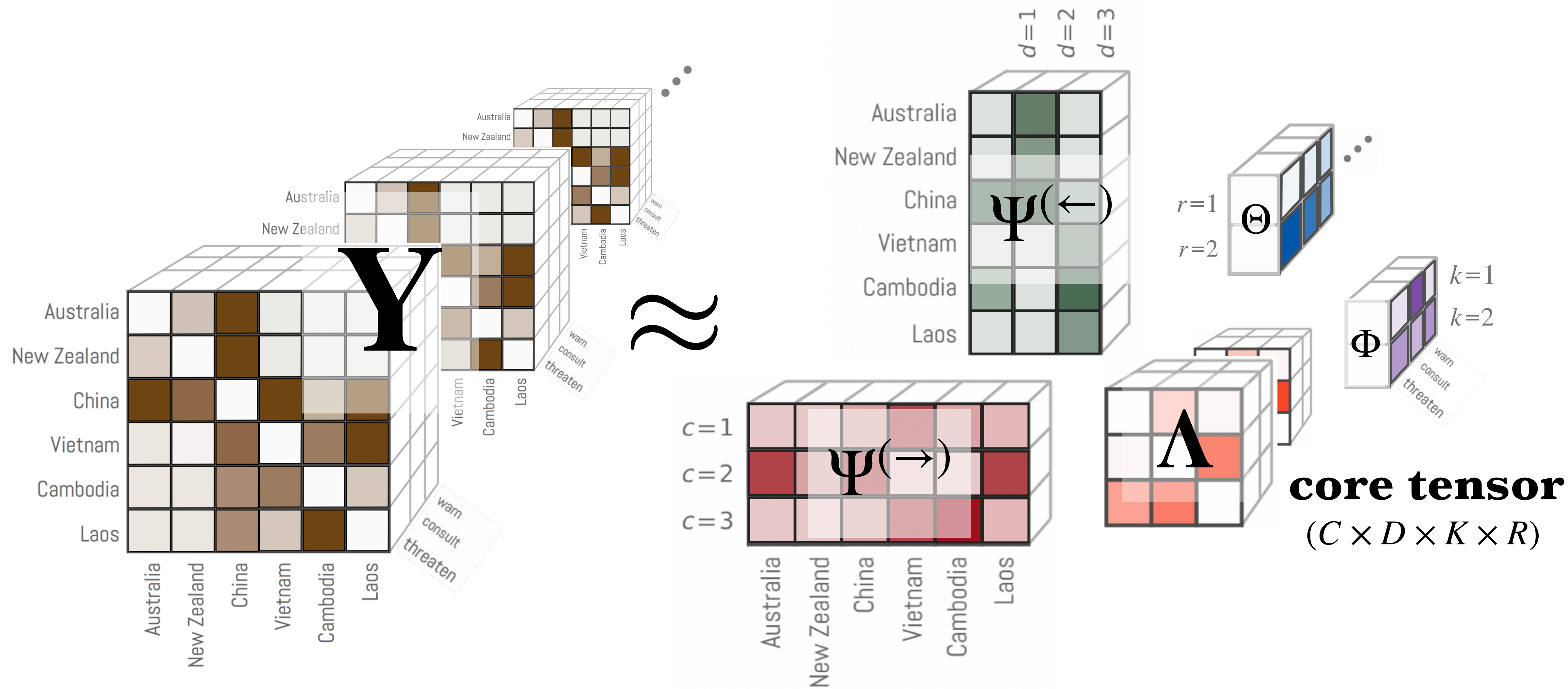


- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

how active **country** i is
in (sender) **community** c

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

POISSON TUCKER DECOMPOSITION

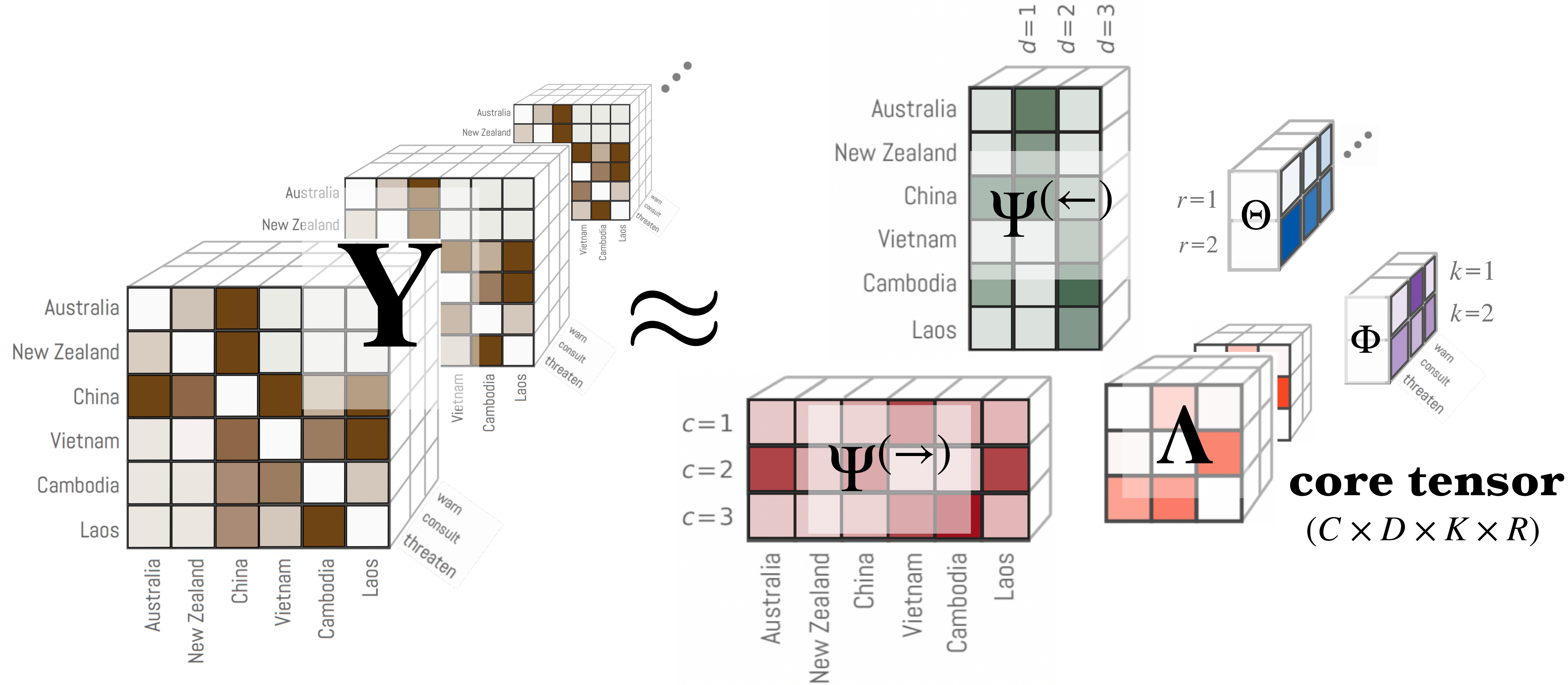


- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

how active **country** j is
in (receiver) **community** d

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

POISSON TUCKER DECOMPOSITION

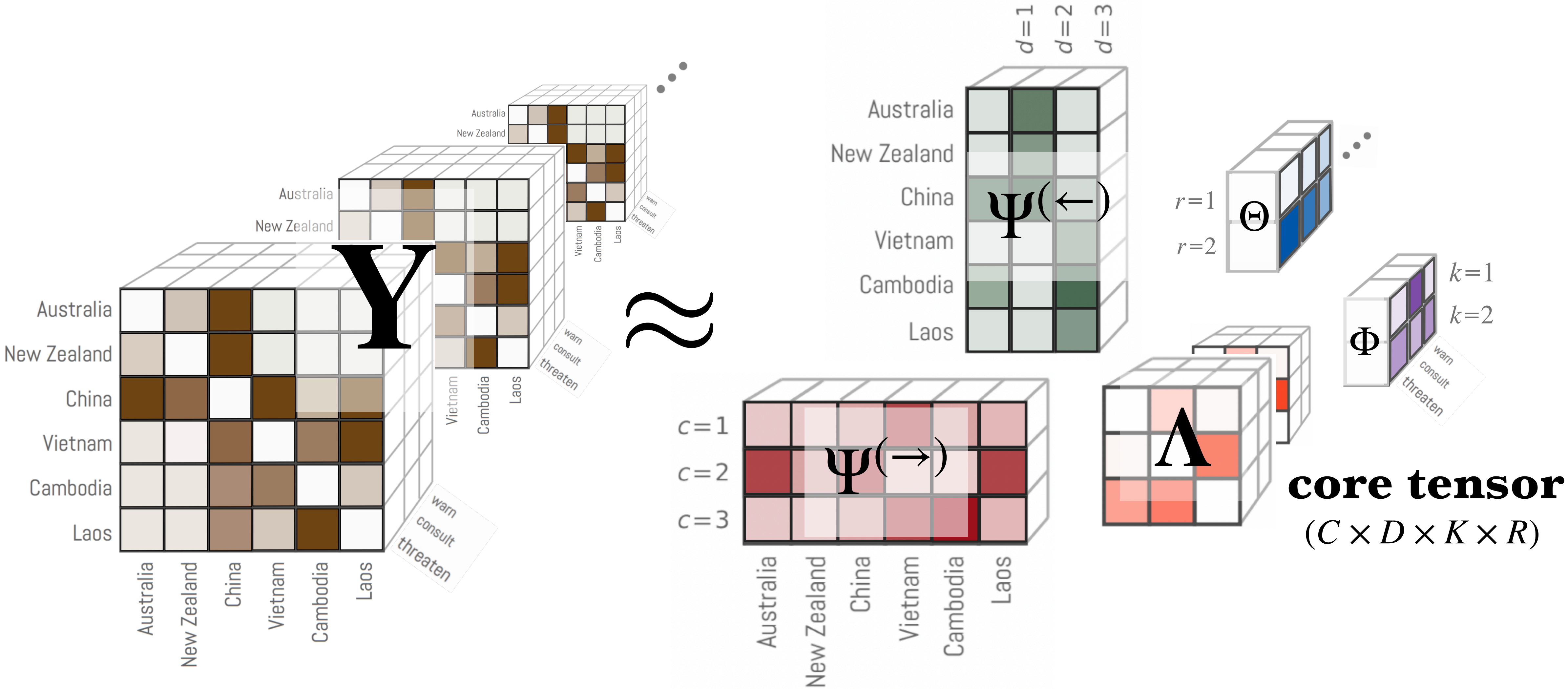


- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

how prevalent **action** a is in **topic** k

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

POISSON TUCKER DECOMPOSITION

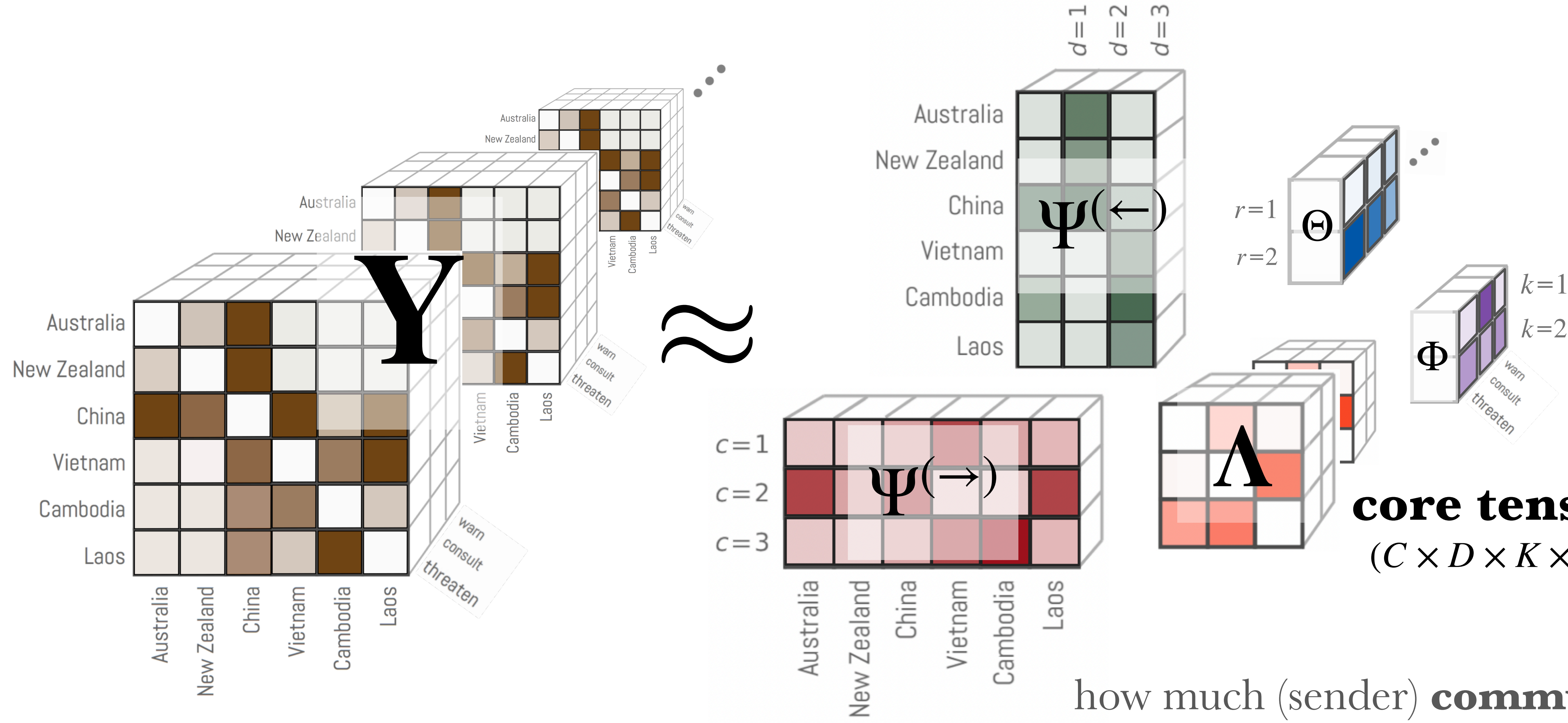


- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

how active **regime** r is at **time** t

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

POISSON TUCKER DECOMPOSITION



core tensor
$(C \times D \times K \times R)$

how much (sender) **community** c
takes **topic** k
toward (receiver) **community** d
under **regime** r

- **Optional:** $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

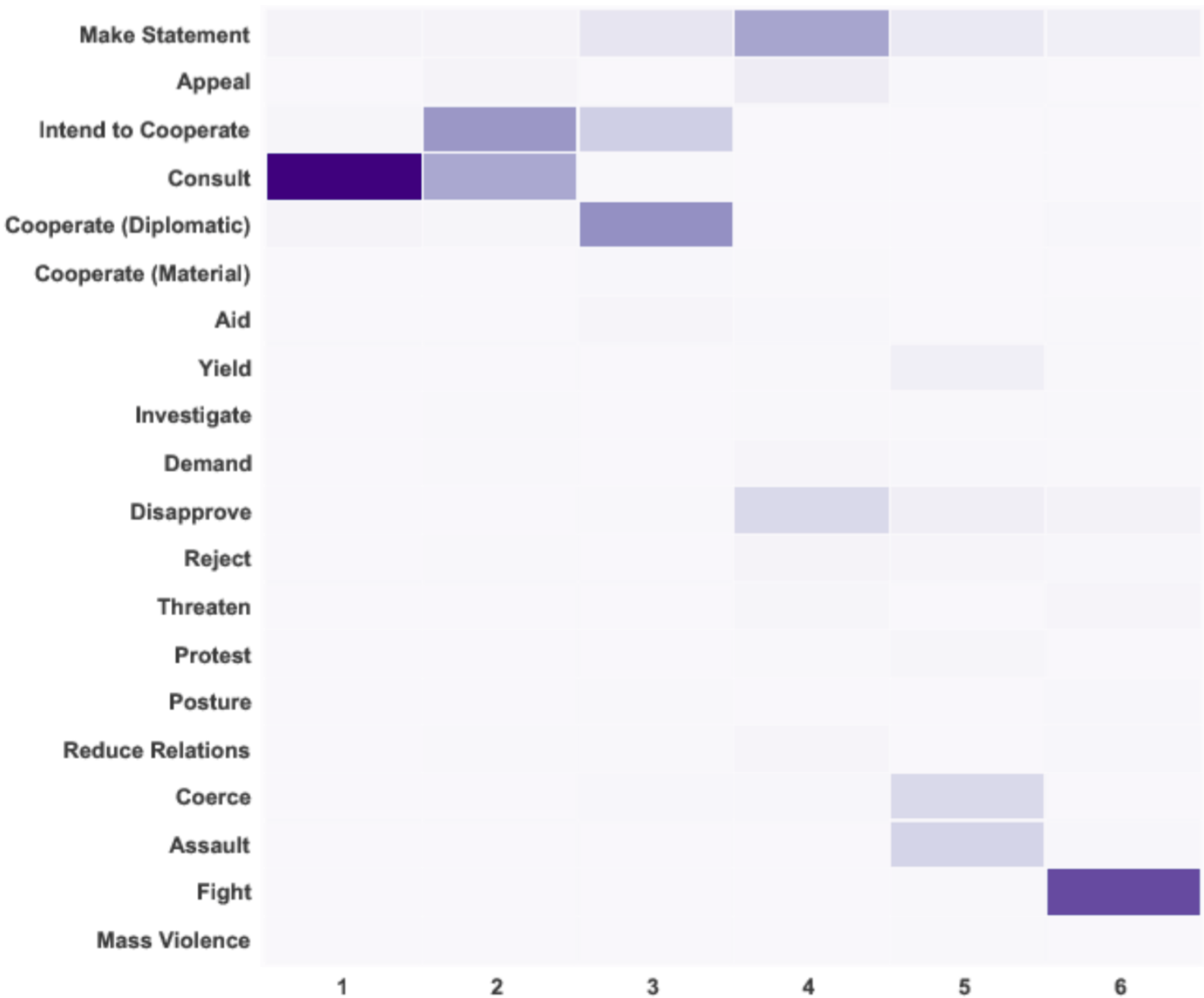
$$y_{i \xrightarrow{a} j}^{(t)} \sim \text{Pois}\left(\sum_{c=1}^{C}\sum_{d=1}^{D}\sum_{k=1}^{K}\sum_{r=1}^{R} \psi_{ic}^{(\rightarrow)}\, \psi_{jd}^{(\leftarrow)}\, \phi_{ak}\, \theta_r^{(t)}\, \lambda_{c \xrightarrow{k} d}^{(r)}\right)$$

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

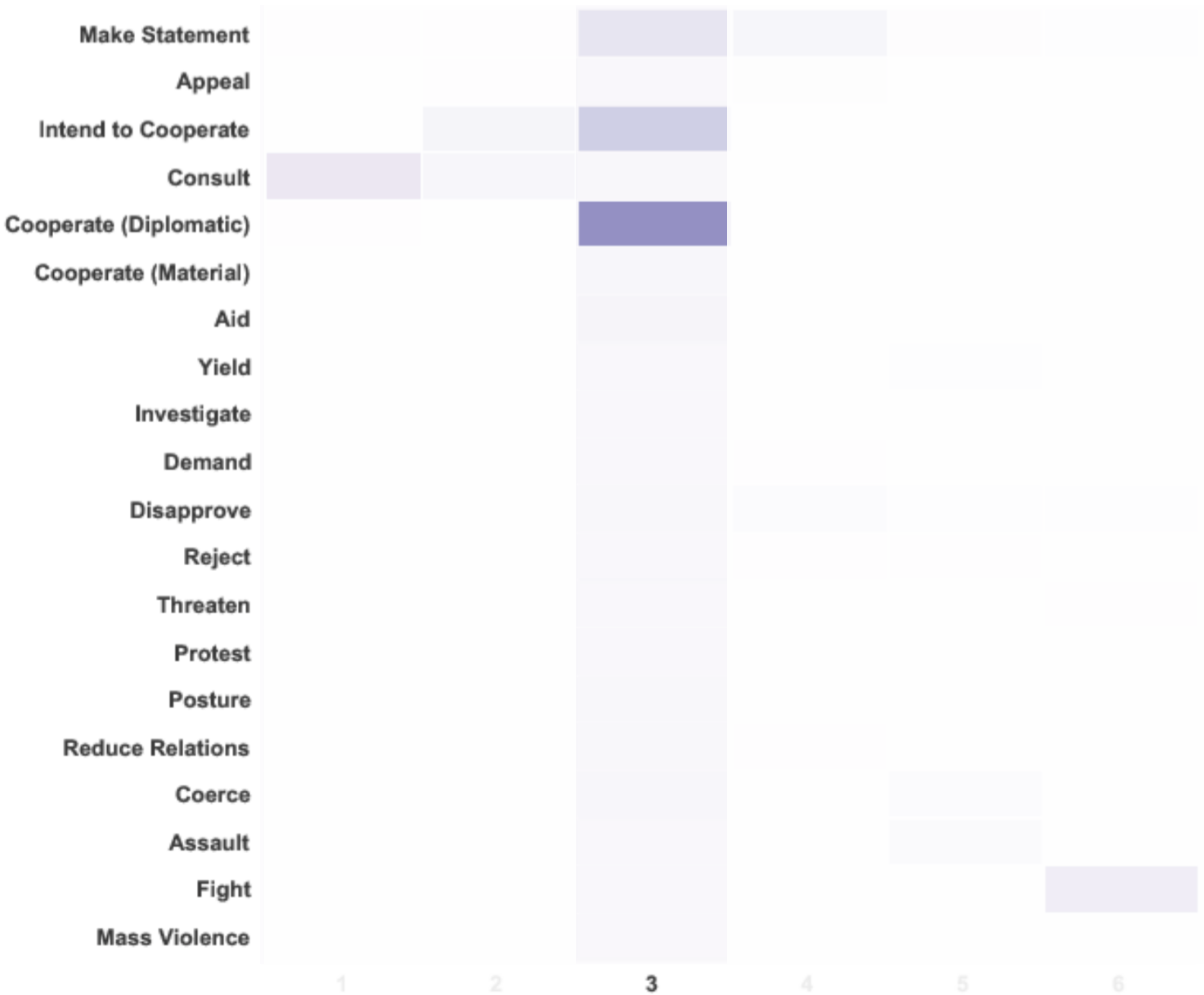


ϕ_{ak}

topics of **actions**

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



ϕ_{ak}

topics of **actions**

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

$$\phi_{ak}$$

topics of **actions**

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

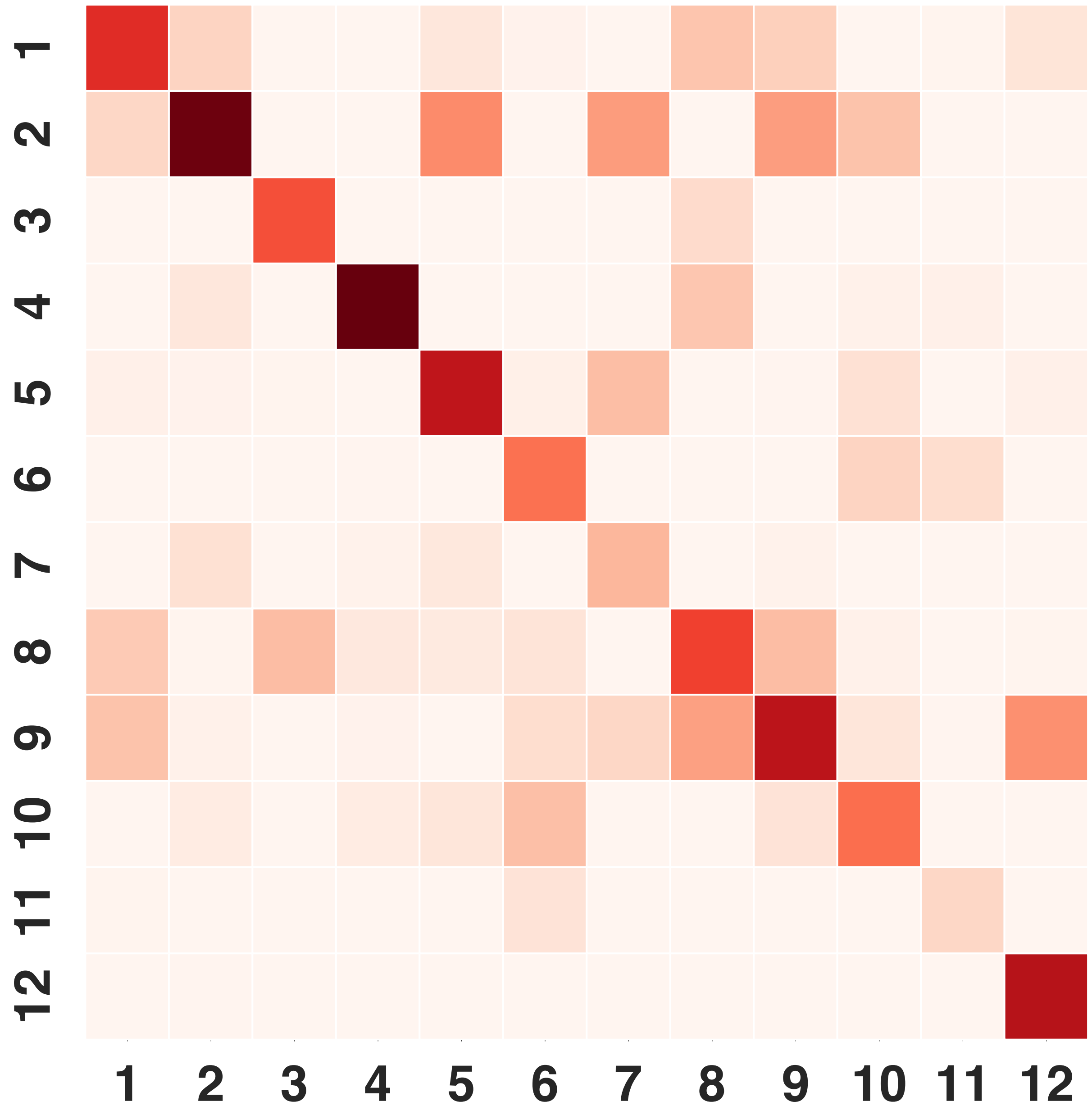
ϕ_{ak}

topics of **actions**

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

$$\lambda^{k}_{c \rightarrow d}$$

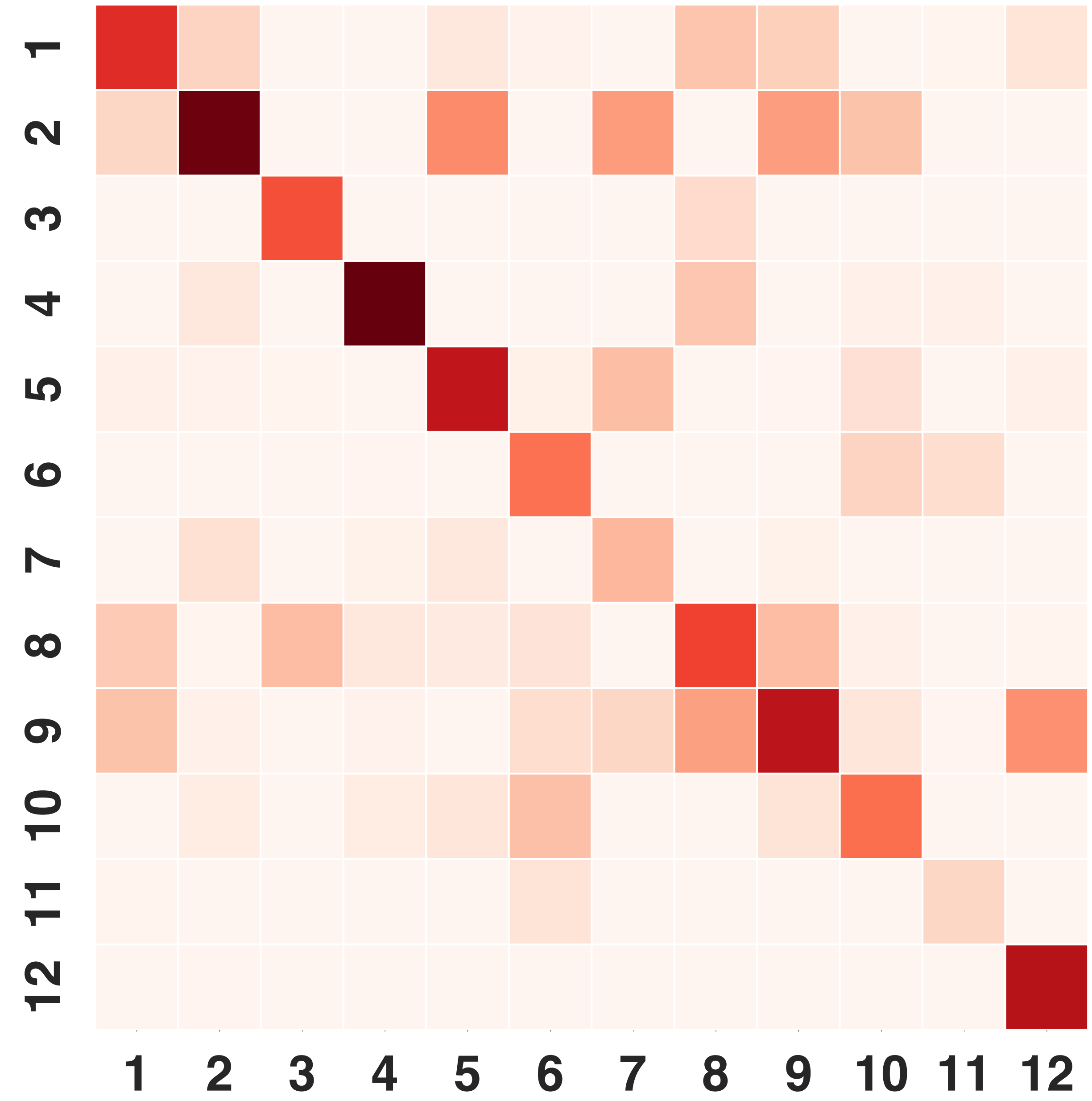


EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

$$\lambda_{c \rightarrow d}^{k}$$

Slice of core tensor for
"Lip service" topic k

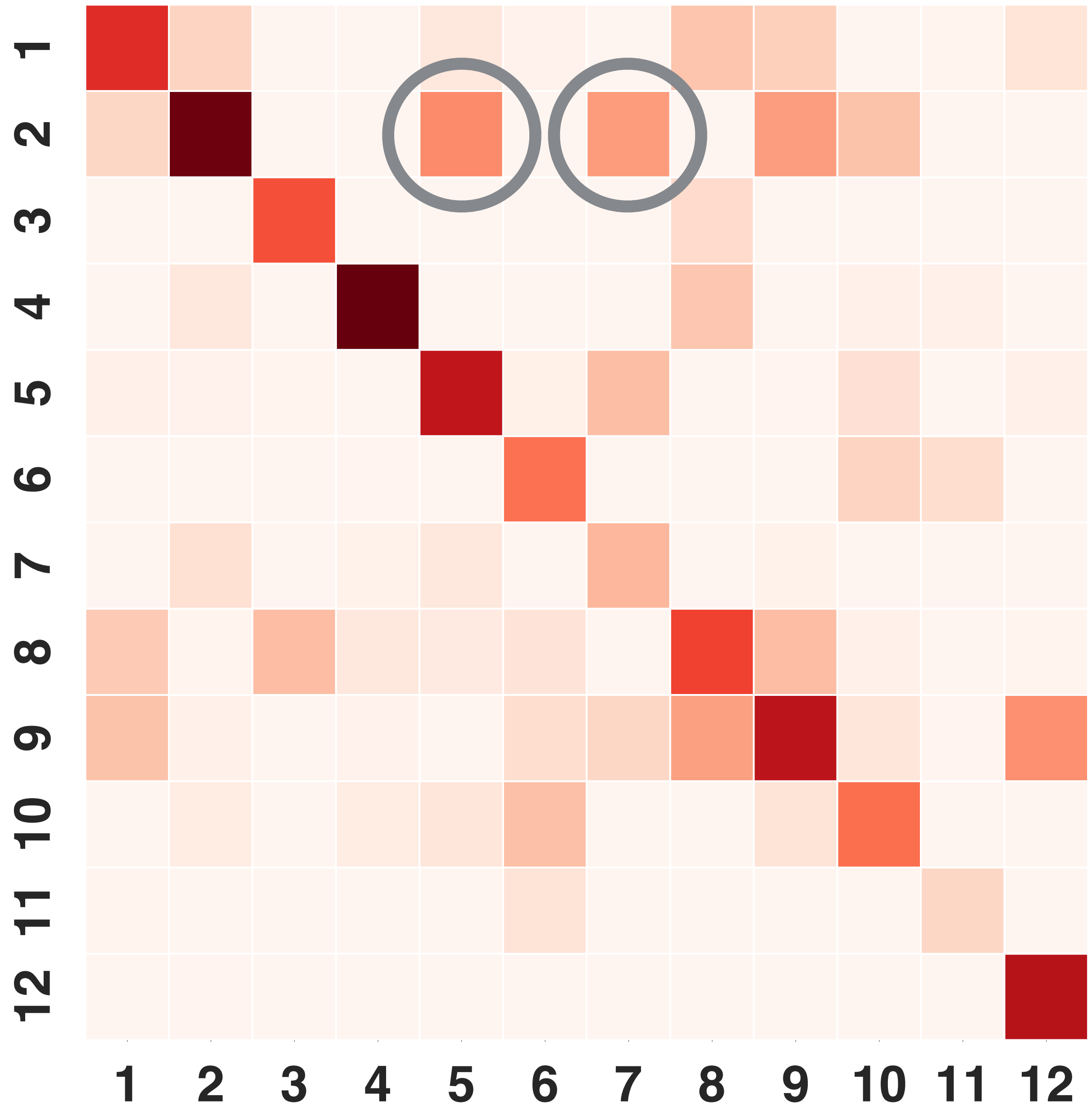


EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

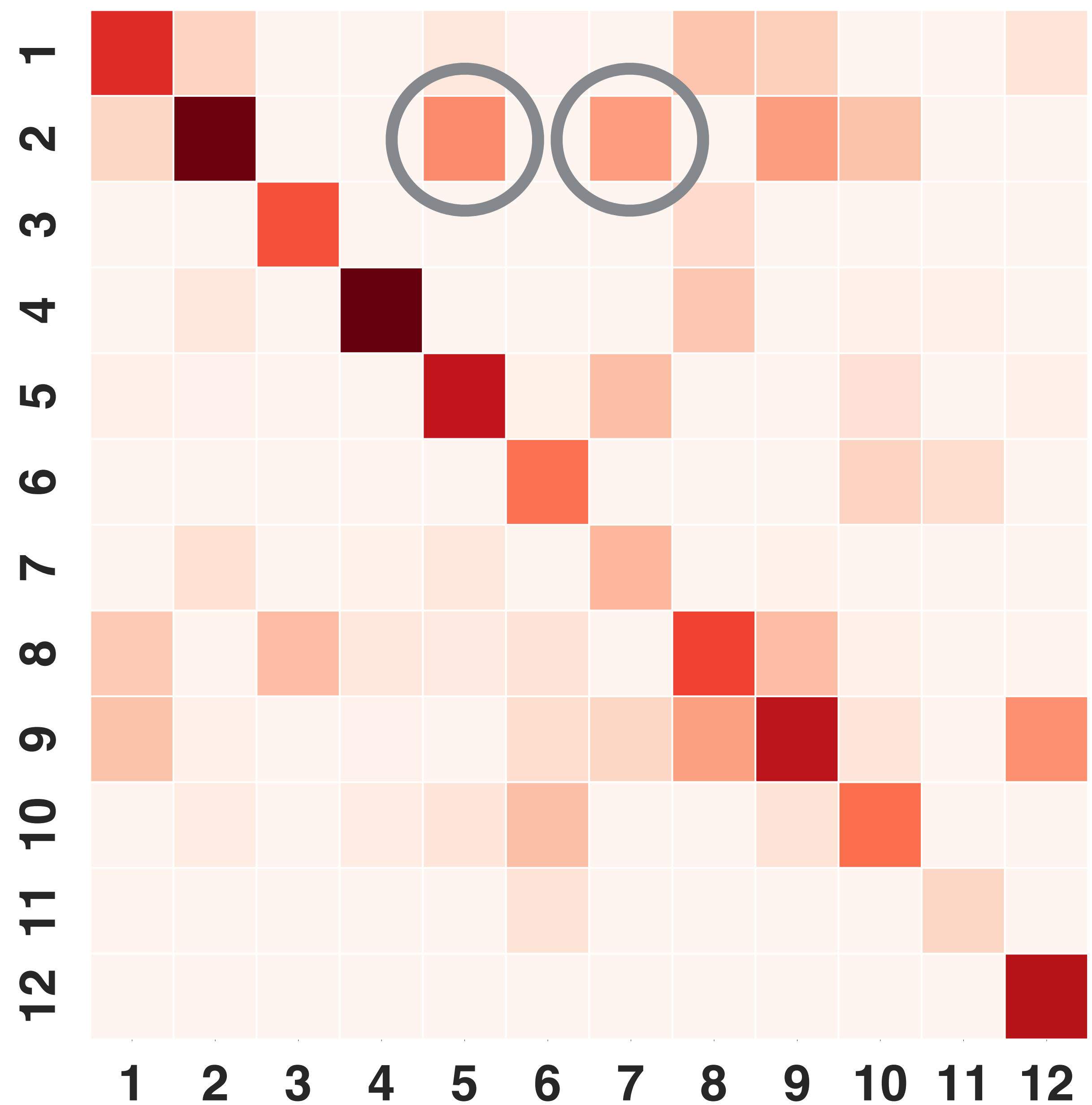
$$\lambda^{k}_{c \rightarrow d}$$

Slice of core tensor for "Lip service" topic k



EXAMPLE RESULTS

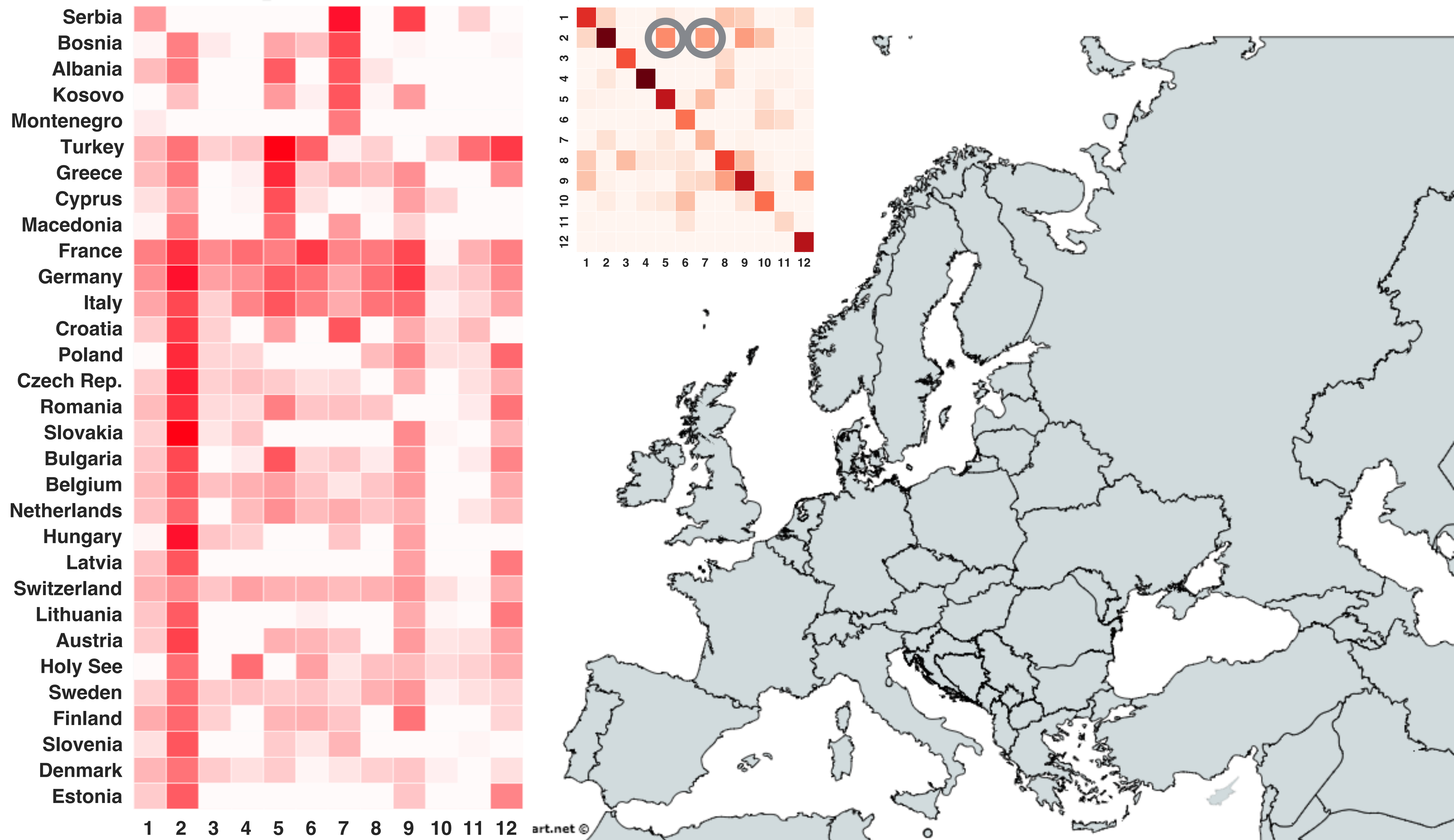
- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



Country-community matrix Ψ

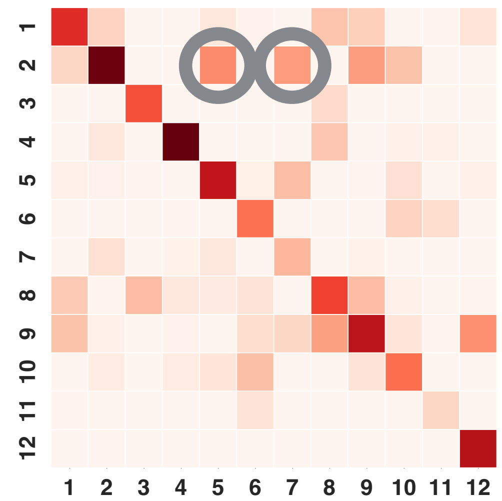
EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



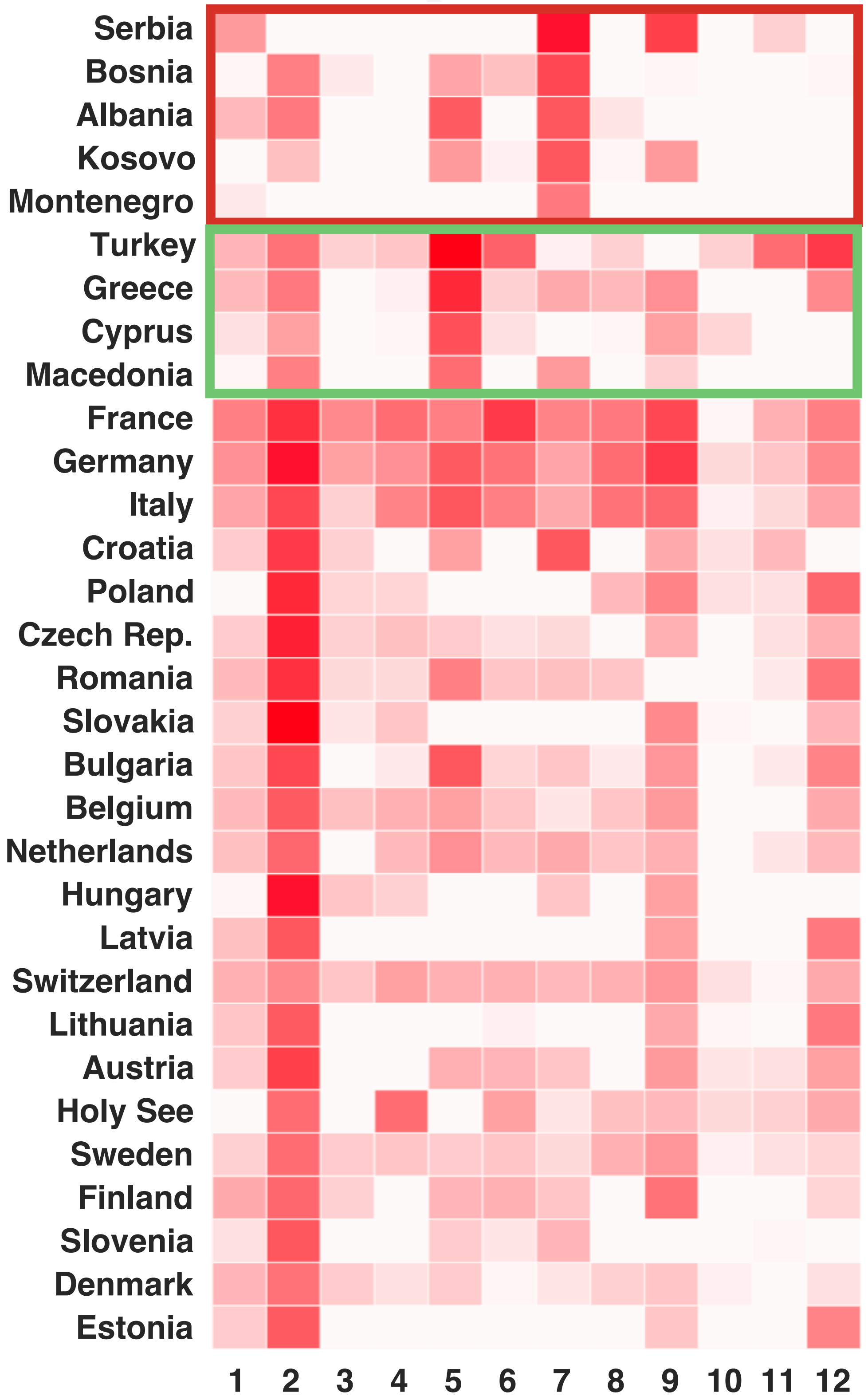
EXAMPLE RESULTS

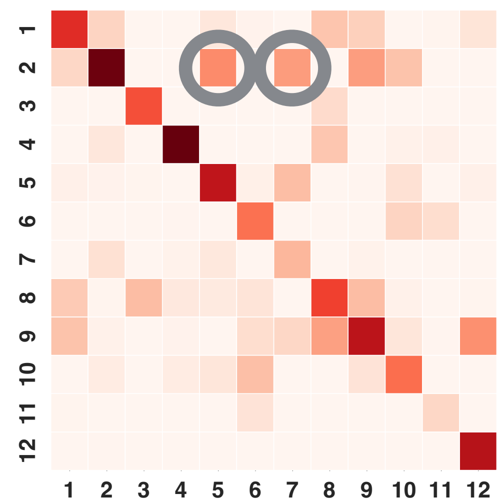
- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



EXAMPLE RESULTS

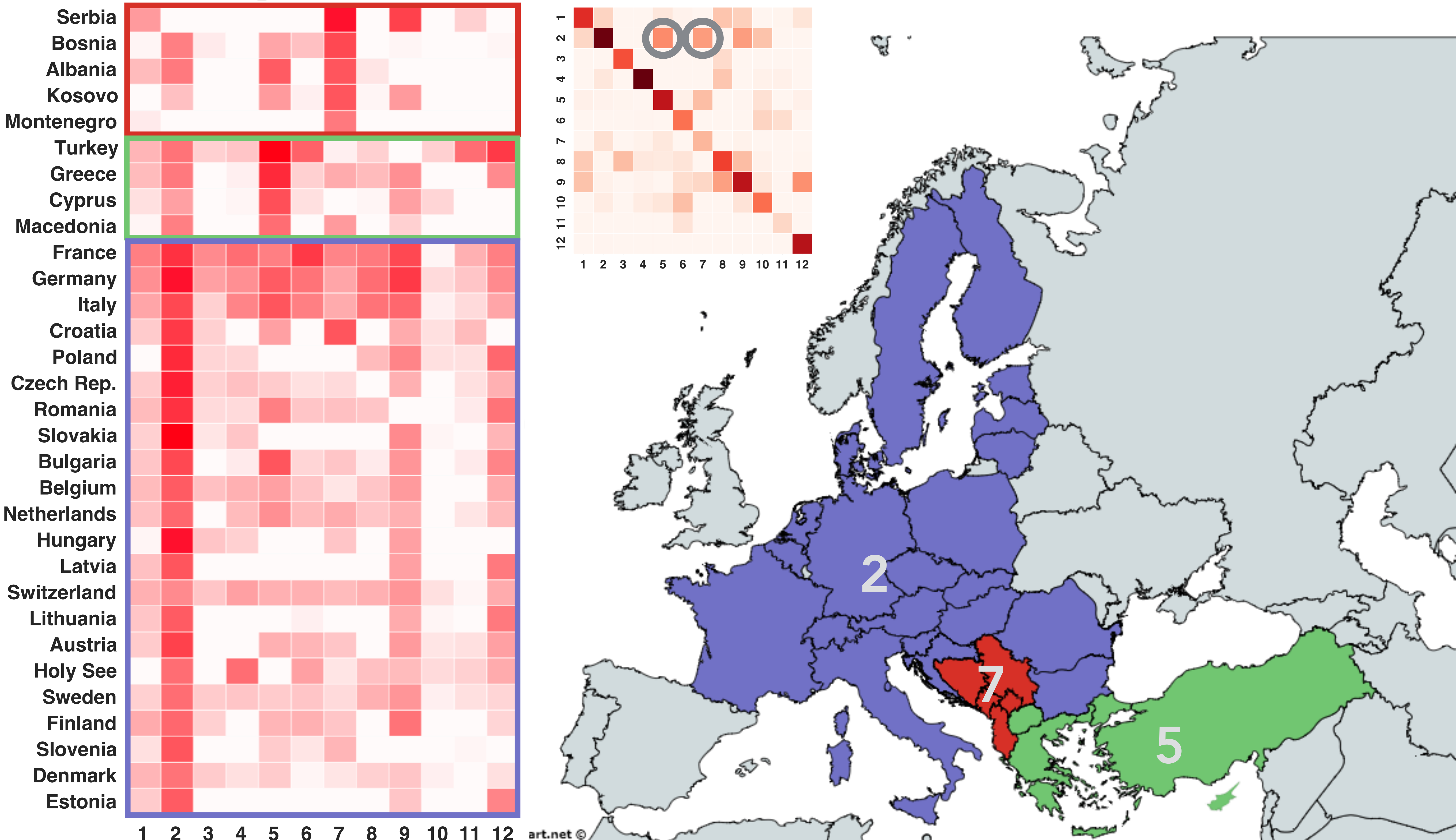
- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$





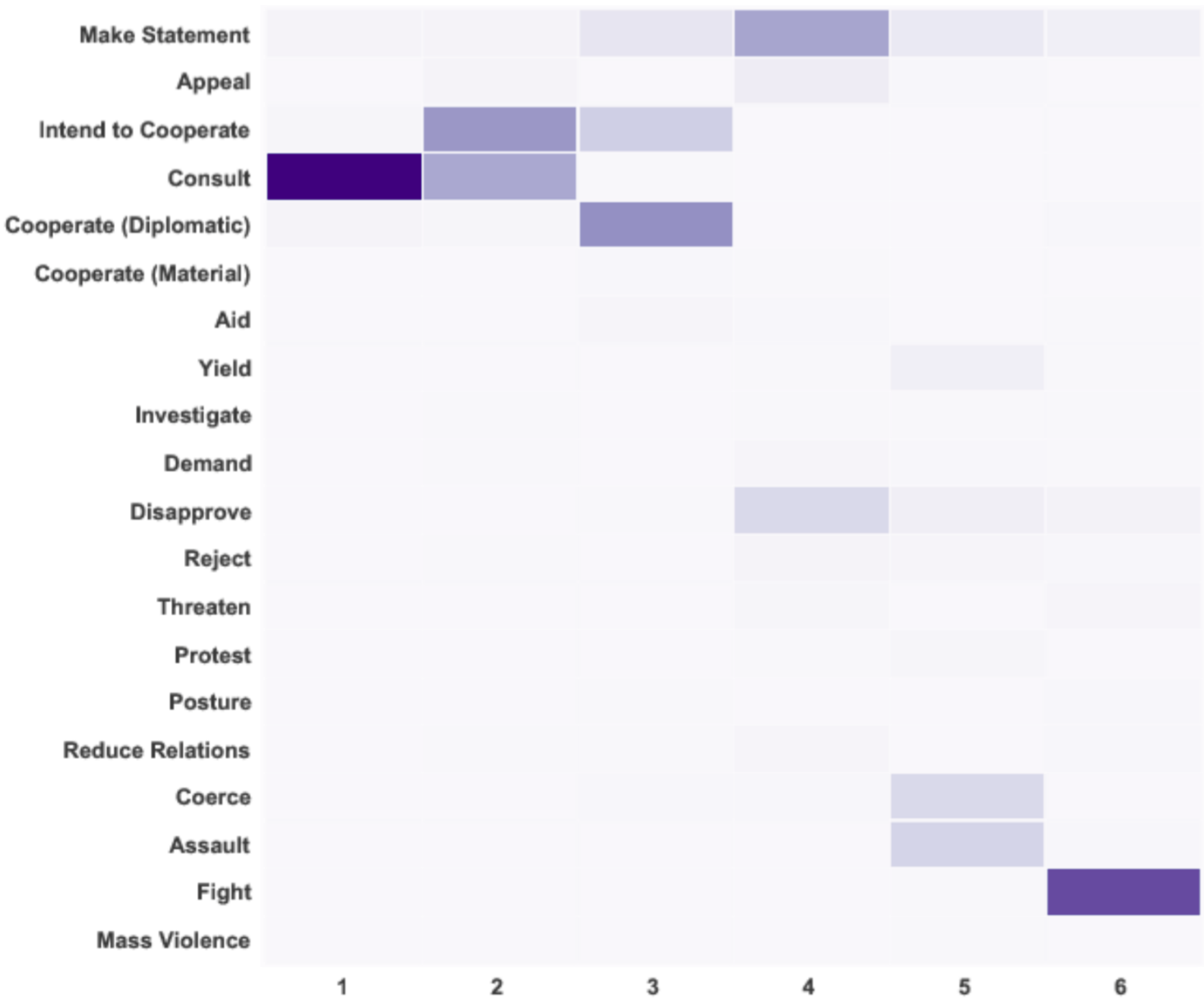
EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



ϕ_{ak}

topics of **actions**

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

ϕ_{ak}

topics of actions

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

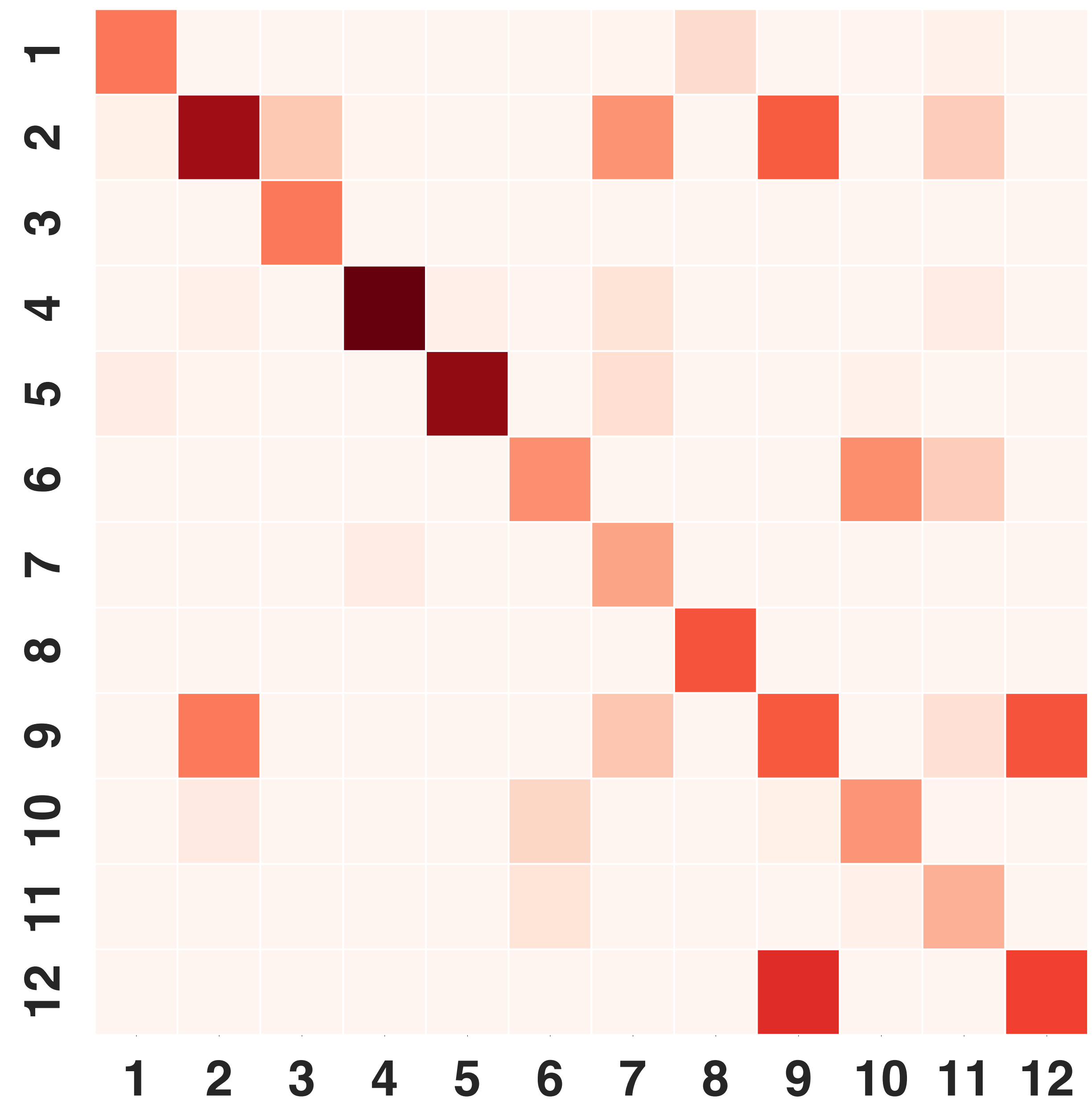
ϕ_{ak}

topics of **actions**

EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

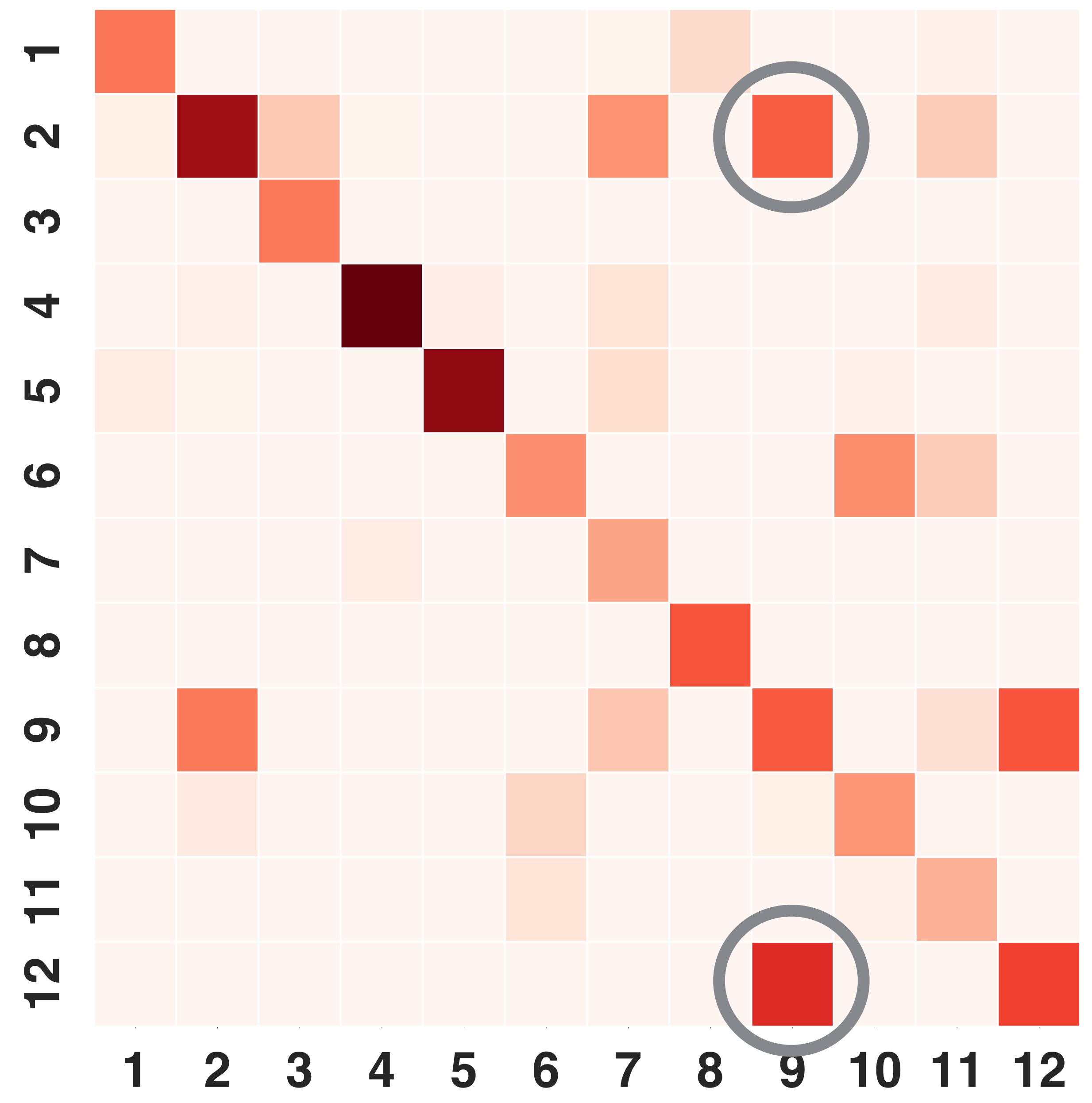
$$\lambda_{c \xrightarrow{k} d} \longrightarrow$$



EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

$$\lambda_{c \xrightarrow{k} d} \longrightarrow$$

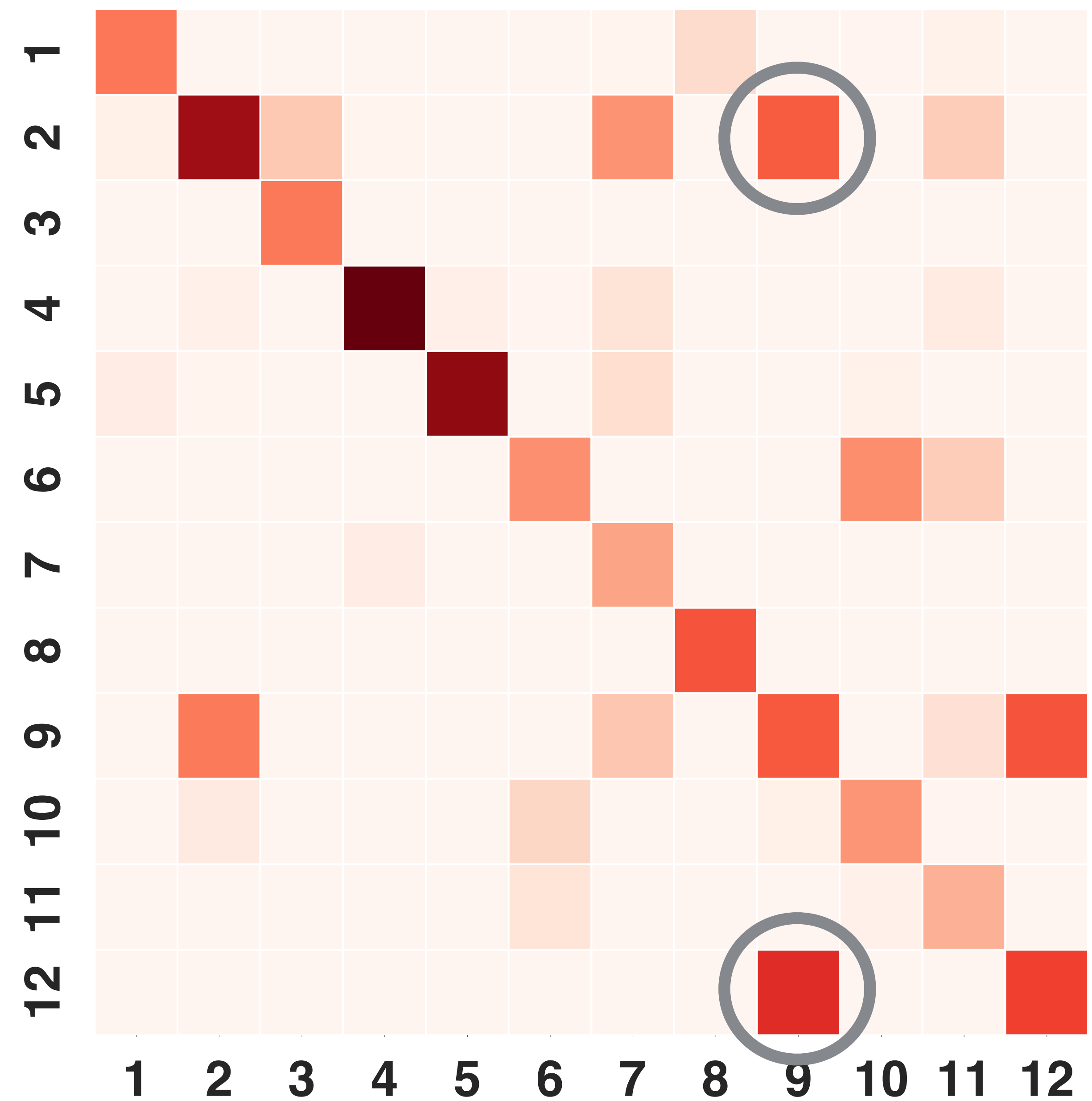


EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

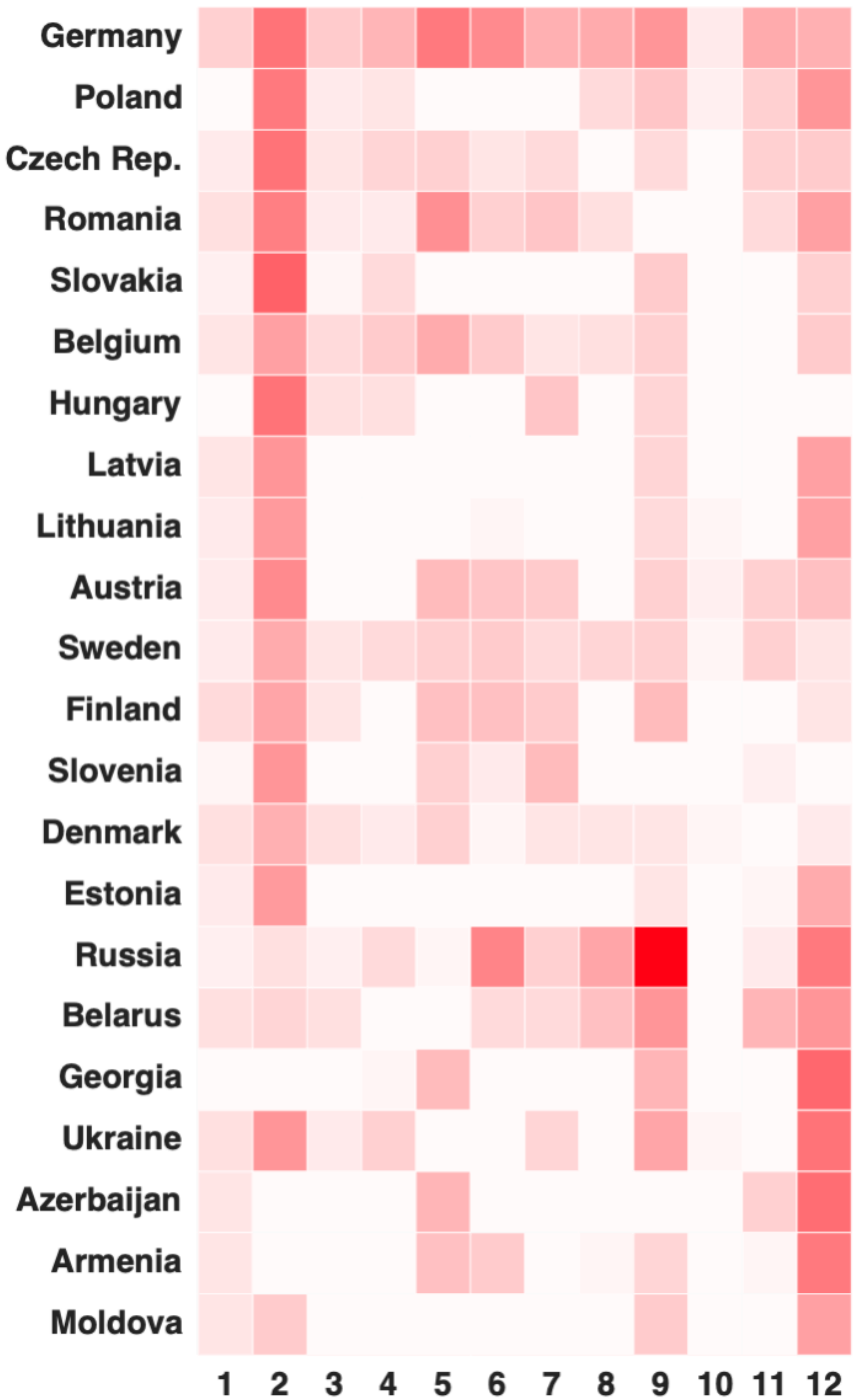
$$\lambda_{c \xrightarrow{k} d}$$

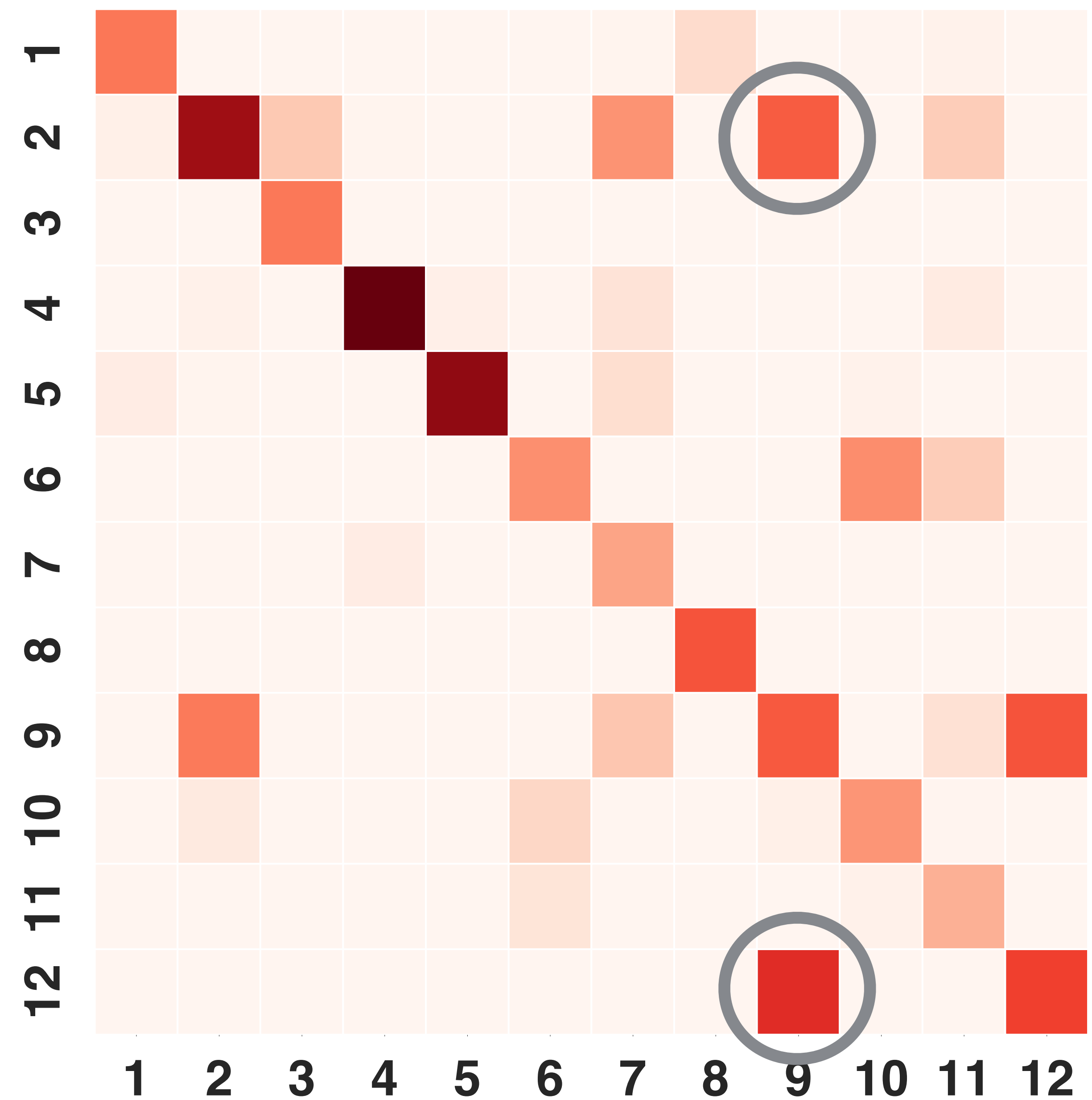
Slice of core tensor for
"Protest" topic k



EXAMPLE RESULTS

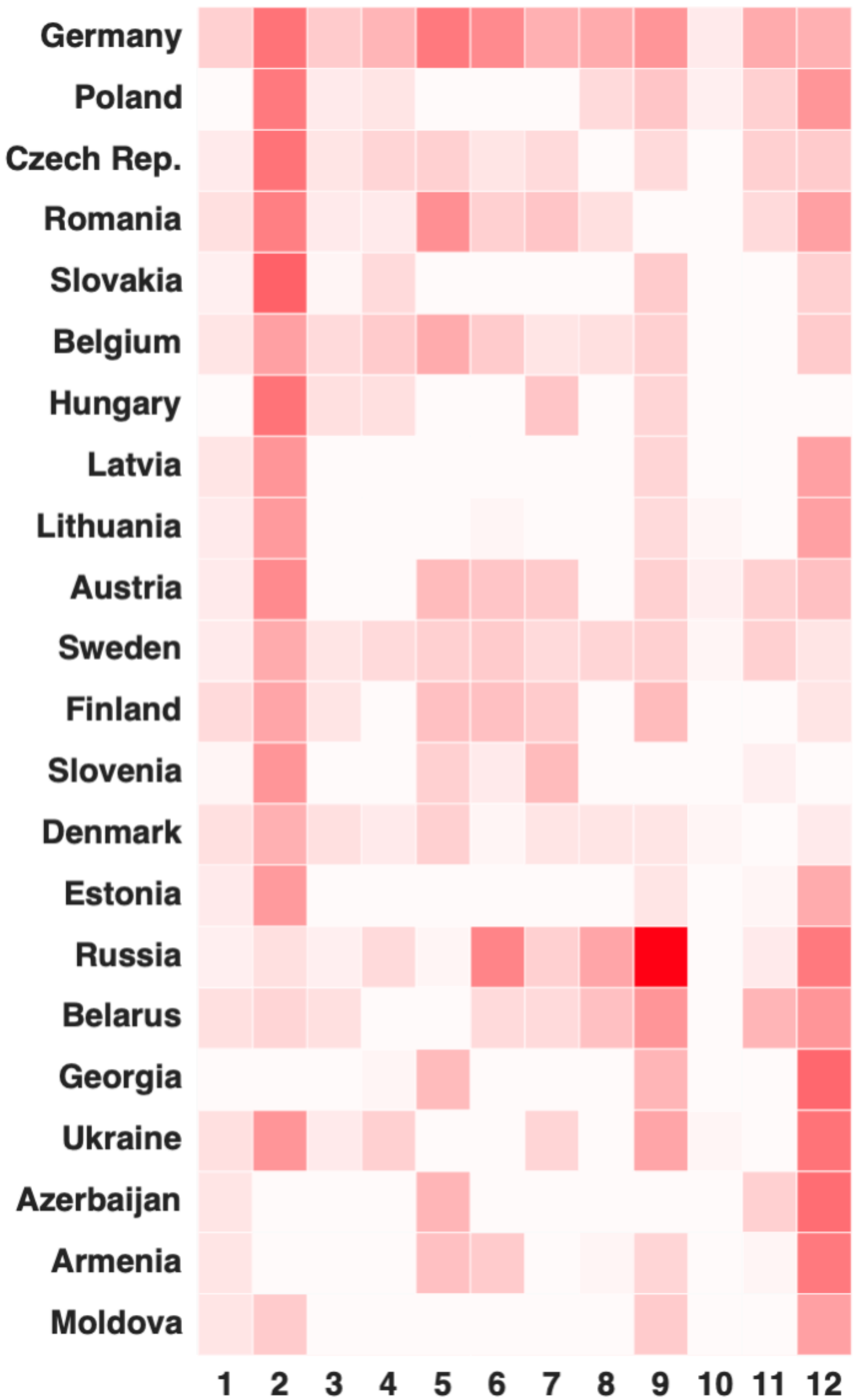
- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$

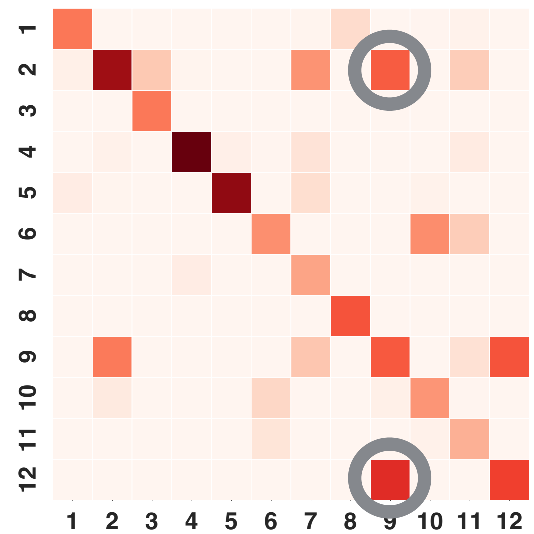




EXAMPLE RESULTS

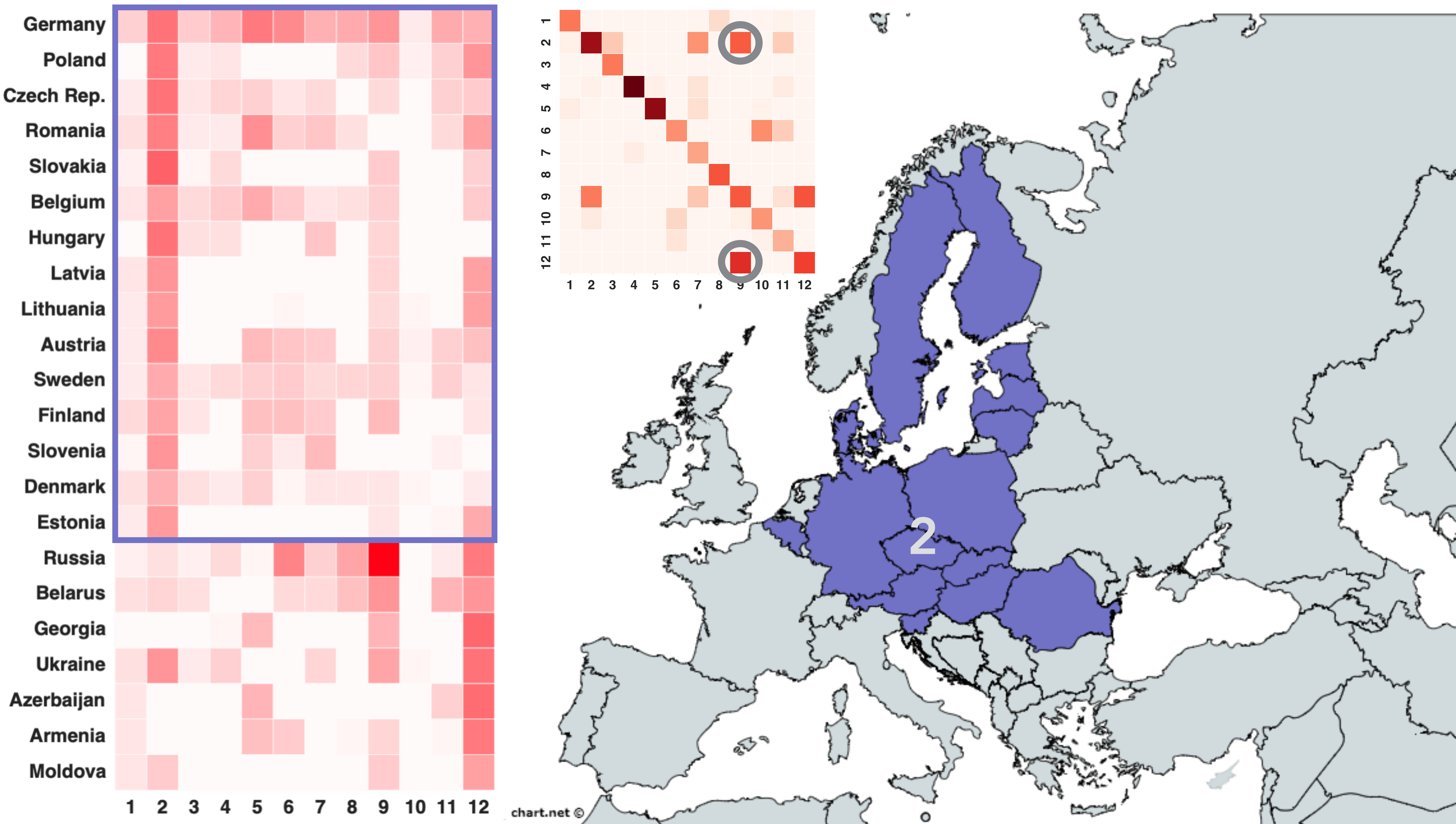
- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$





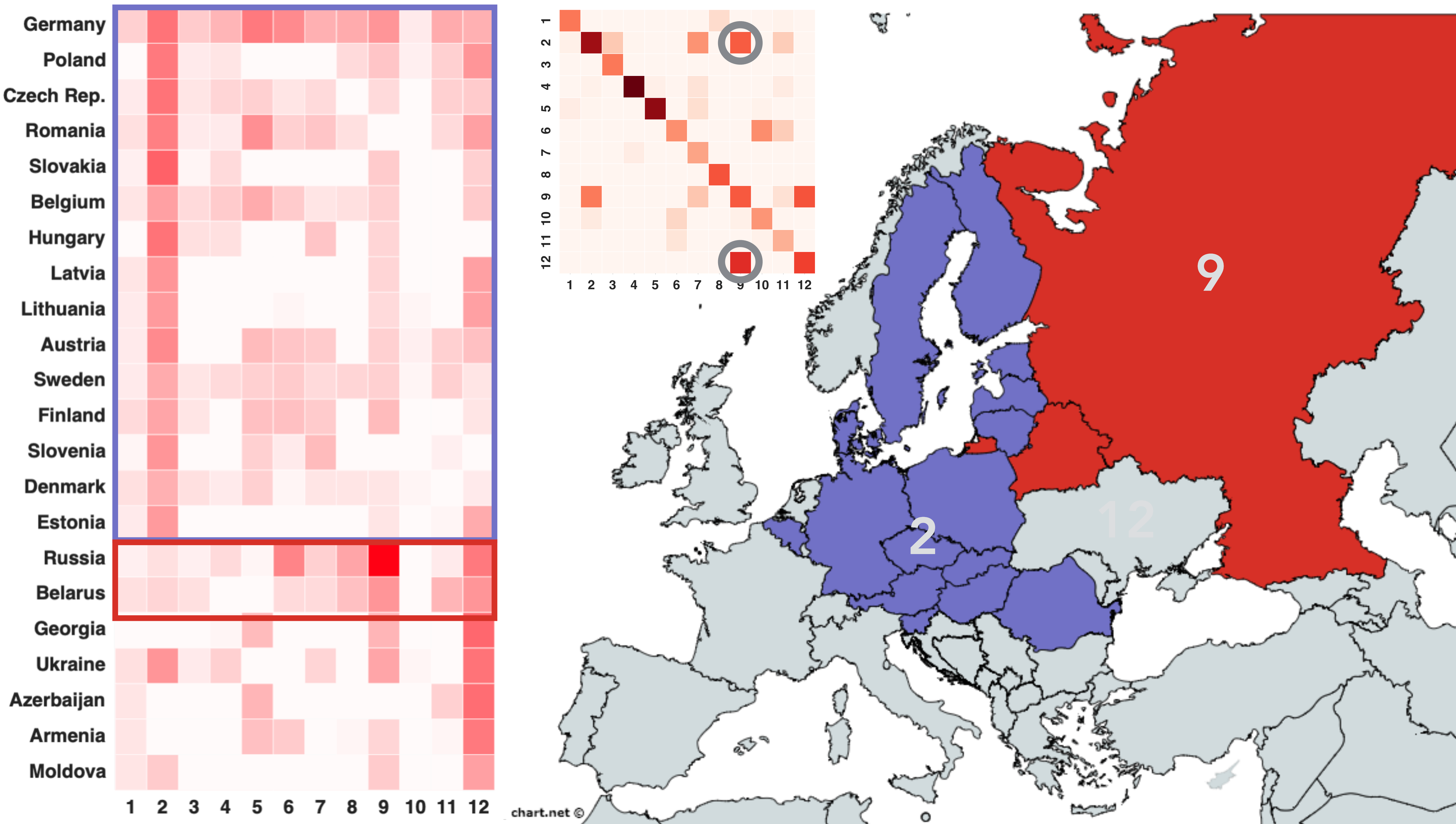
EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



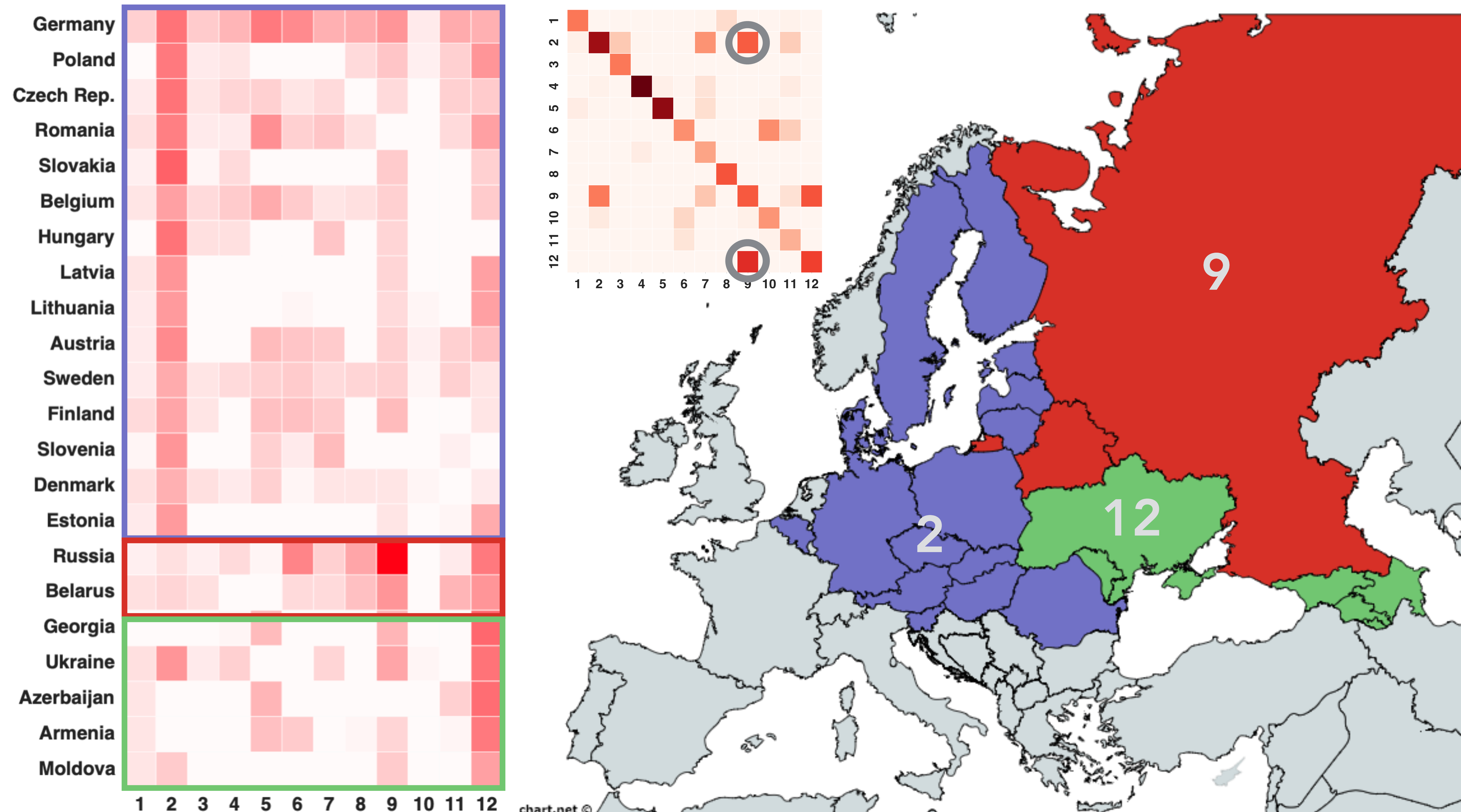
EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



EXAMPLE RESULTS

- GDELT tensor (1995-2000), monthly: (150 x 150 x 20 x 60)
- Core size: (12 x 12 x 6 x 1)
- $\Psi^{(\rightarrow)} = \Psi^{(\leftarrow)} = \Psi$



OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

ALLOCATIVE POISSON FACTORIZATION

All previous models were an instance of:

$$y_{\mathbf{i}} \sim \mathrm{Pois}(\mu_{\mathbf{i}}) \qquad \mathbf{i} \equiv (\mathbf{i}_1, \ldots, \mathbf{i}_M) \qquad \mu_{\mathbf{i}} = \sum_{\boldsymbol{\kappa}} \mu_{\mathbf{i},\boldsymbol{\kappa}}$$

Poisson counts | **multi-index** | **linear** function of parameters

ALLOCATIVE POISSON FACTORIZATION

All previous models were an instance of:

$$y_{\mathbf{i}} \sim \mathrm{Pois}(\mu_{\mathbf{i}}) \qquad \mathbf{i} \equiv (\mathbf{i}_1, \ldots, \mathbf{i}_M) \qquad \mu_{\mathbf{i}} = \sum_{\boldsymbol{\kappa}} \mu_{\mathbf{i},\boldsymbol{\kappa}}$$

Poisson counts — **multi-index** — **linear** function of parameters

e.g., Poisson CP decomposition for 4-mode dyadic event data

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\left(\mu^{(t)}_{i \xrightarrow{a} j}\right) \qquad \mathbf{i} \equiv (i, j, a, t) \qquad \mu^{(t)}_{i \xrightarrow{a} j} = \sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq} \, \psi^{(\leftarrow)}_{jq} \, \phi_{aq} \, \theta^{(t)}_{q} \, \lambda_q$$

ALLOCATIVE POISSON FACTORIZATION

All previous models were an instance of:

$$y_{\mathbf{i}} \sim \mathrm{Pois}(\mu_{\mathbf{i}}) \qquad \mathbf{i} \equiv (\mathbf{i}_1, \ldots, \mathbf{i}_M) \qquad \mu_{\mathbf{i}} = \sum_{\boldsymbol{\kappa}} \mu_{\mathbf{i},\boldsymbol{\kappa}}$$

Poisson counts **multi-index** **linear** function of parameters

e.g., Poisson CP decomposition for 4-mode dyadic event data

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\Big(\mu^{(t)}_{i \xrightarrow{a} j}\Big) \qquad \mathbf{i} \equiv (i, j, a, t) \qquad \mu^{(t)}_{i \xrightarrow{a} j} = \sum_{q=1}^{Q} \psi^{(\rightarrow)}_{iq}\, \psi^{(\leftarrow)}_{jq}\, \phi_{aq}\, \theta^{(t)}_{q}\, \lambda_q$$

e.g., Poisson Tucker decomposition for 4-mode dyadic event data

$$y^{(t)}_{i \xrightarrow{a} j} \sim \mathrm{Pois}\Big(\mu^{(t)}_{i \xrightarrow{a} j}\Big) \qquad \mathbf{i} \equiv (i, j, a, t) \qquad \mu^{(t)}_{i \xrightarrow{a} j} = \sum_{c,d,k,r} \psi^{(\rightarrow)}_{ic}\, \psi^{(\leftarrow)}_{jd}\, \phi_{ak}\, \theta^{(t)}_{r}\, \lambda^{(r)}_{c \xrightarrow{k} d}$$

ALLOCATIVE POISSON FACTORIZATION

"**Latent source**" representation [Dunson & Herring, (2005); Cemgil (2009)]:

$$y_{\mathbf{i}} \sim \textrm{Pois}\Big(\sum_{\boldsymbol{\kappa}} \mu_{\mathbf{i},\boldsymbol{\kappa}}\Big) \quad \leftrightarrow \quad y_{\mathbf{i}} = \sum_{\boldsymbol{\kappa}} y_{\mathbf{i},\boldsymbol{\kappa}} \ \text{ where } \ y_{\mathbf{i},\boldsymbol{\kappa}} \sim \textrm{Pois}\big(\mu_{\mathbf{i},\boldsymbol{\kappa}}\big)$$

Observed counts

Latent sub-counts ("sources")

ALLOCATIVE POISSON FACTORIZATION

"**Latent source**" representation [Dunson & Herring, (2005); Cemgil (2009)]:

$$y_{\mathbf{i}} \sim \text{Pois}\Big(\sum_{\kappa} \mu_{\mathbf{i},\kappa}\Big) \quad \leftrightarrow \quad y_{\mathbf{i}} = \sum_{\kappa} y_{\mathbf{i},\kappa} \;\text{ where }\; y_{\mathbf{i},\kappa} \sim \text{Pois}(\mu_{\mathbf{i},\kappa})$$

Observed counts | **Latent** sub-counts ("sources")

Inference (MCMC, VI, EM, …) in these models takes the following form:

Step 1: Given data $y_{\mathbf{i}}$, infer sources $(y_{\mathbf{i},\kappa})_{\kappa}$:

$$(y_{\mathbf{i},\kappa})_{\kappa} \mid y_{\mathbf{i}} \sim \text{Multinomial}\Bigg(y_{\mathbf{i}}, \Big(\frac{\mu_{\mathbf{i},\kappa}}{\sum_{\kappa'} \mu_{\mathbf{i},\kappa'}}\Big)_{\kappa}\Bigg)$$

ALLOCATIVE POISSON FACTORIZATION

"**Latent source**" representation [Dunson & Herring, (2005); Cemgil (2009)]:

$$y_{\mathbf{i}} \sim \mathrm{Pois}\Big(\sum_{\kappa} \mu_{\mathbf{i},\kappa}\Big) \quad \leftrightarrow \quad y_{\mathbf{i}} = \sum_{\kappa} y_{\mathbf{i},\kappa} \text{ where } y_{\mathbf{i},\kappa} \sim \mathrm{Pois}(\mu_{\mathbf{i},\kappa})$$

Observed counts

Latent sub-counts ("sources")

Inference (MCMC, VI, EM, …) in these models takes the following form:

Step 1: Given data $y_{\mathbf{i}}$, infer sources $(y_{\mathbf{i},\kappa})_{\kappa}$:

$$(y_{\mathbf{i},\kappa})_{\kappa} \mid y_{\mathbf{i}} \sim \mathrm{Multinomial}\left(y_{\mathbf{i}}, \left(\frac{\mu_{\mathbf{i},\kappa}}{\sum_{\kappa'} \mu_{\mathbf{i},\kappa'}}\right)_{\kappa}\right)$$

Step 2: Given sources $(y_{\mathbf{i},\kappa})_{\kappa}$, update parameters.

ALLOCATIVE POISSON FACTORIZATION

"**Latent source**" representation [Dunson & Herring, (2005); Cemgil (2009)]:

$$y_{\mathbf{i}} \sim \mathrm{Pois}\Big(\sum_{\kappa} \mu_{\mathbf{i},\kappa}\Big) \quad \leftrightarrow \quad y_{\mathbf{i}} = \sum_{\kappa} y_{\mathbf{i},\kappa} \;\text{ where }\; y_{\mathbf{i},\kappa} \sim \mathrm{Pois}(\mu_{\mathbf{i},\kappa})$$

Observed counts | **Latent** sub-counts ("sources")

Inference (MCMC, VI, EM, …) in these models takes the following form:

Step 1: Given data $y_{\mathbf{i}}$, infer sources $(y_{\mathbf{i},\kappa})_{\kappa}$:

$$(y_{\mathbf{i},\kappa})_{\kappa} \mid y_{\mathbf{i}} \sim \mathrm{Multinomial}\left(y_{\mathbf{i}}, \left(\frac{\mu_{\mathbf{i},\kappa}}{\sum_{\kappa'} \mu_{\mathbf{i},\kappa'}}\right)_{\kappa}\right)$$

Step 2: Given sources $(y_{\mathbf{i},\kappa})_{\kappa}$, update parameters.

(**Repeat**)

ALLOCATIVE POISSON FACTORIZATION

"**Latent source**" representation [Dunson & Herring, (2005); Cemgil (2009)]:

$$y_{\mathbf{i}} \sim \mathrm{Pois}\Big(\sum_{\kappa} \mu_{\mathbf{i},\kappa}\Big) \quad \leftrightarrow \quad y_{\mathbf{i}} = \sum_{\kappa} y_{\mathbf{i},\kappa} \;\text{ where }\; y_{\mathbf{i},\kappa} \sim \mathrm{Pois}\big(\mu_{\mathbf{i},\kappa}\big)$$

Observed counts

Latent sub-counts ("sources")

Inference (MCMC, VI, EM, …) in these models takes the following form:

Step 1: Given data $y_{\mathbf{i}}$, infer sources $\big(y_{\mathbf{i},\kappa}\big)_{\kappa}$:

$$\big(y_{\mathbf{i},\kappa}\big)_{\kappa} \mid y_{\mathbf{i}} \sim \mathrm{Multinomial}\Big(y_{\mathbf{i}}, \Big(\frac{\mu_{\mathbf{i},\kappa}}{\sum_{\kappa'} \mu_{\mathbf{i},\kappa'}}\Big)_{\kappa}\Big)$$

Step 2: Given sources $\big(y_{\mathbf{i},\kappa}\big)_{\kappa}$, update parameters.

Tends to be simple / inexpensive / parallelizable
(e.g., with conditionally conjugate priors).

(**Repeat**)

ALLOCATIVE POISSON FACTORIZATION

"**Latent source**" representation [Dunson & Herring, (2005); Cemgil (2009)]:

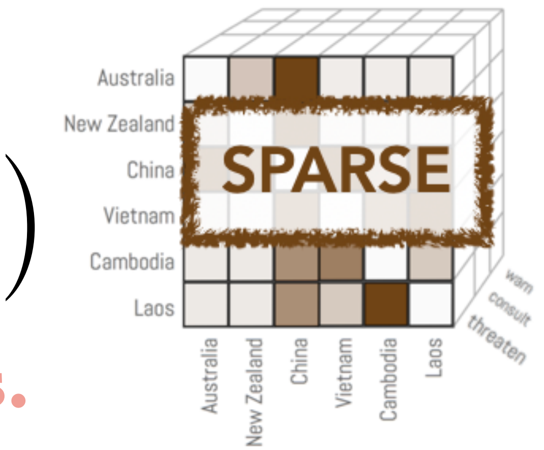
$$y_{\mathbf{i}} \sim \text{Pois}\Big(\sum_{\kappa} \mu_{\mathbf{i},\kappa}\Big) \quad \leftrightarrow \quad y_{\mathbf{i}} = \sum_{\kappa} y_{\mathbf{i},\kappa} \;\text{ where }\; y_{\mathbf{i},\kappa} \sim \text{Pois}(\mu_{\mathbf{i},\kappa})$$

Observed counts — **Latent** sub-counts ("sources")

Inference (MCMC, VI, EM, …) in these models takes the following form:

Step 1: Given data $y_{\mathbf{i}}$, infer sources $(y_{\mathbf{i},\kappa})_{\kappa}$:

$$(y_{\mathbf{i},\kappa})_{\kappa} \mid y_{\mathbf{i}} \sim \text{Multinomial}\Big(y_{\mathbf{i}}, \Big(\frac{\mu_{\mathbf{i},\kappa}}{\sum_{\kappa'} \mu_{\mathbf{i},\kappa'}}\Big)_{\kappa}\Big)$$



Scales with **only the non-zeros.**

Step 2: Given sources $(y_{\mathbf{i},\kappa})_{\kappa}$, update parameters.

Tends to be simple / inexpensive / parallelizable
(e.g., with conditionally conjugate priors).

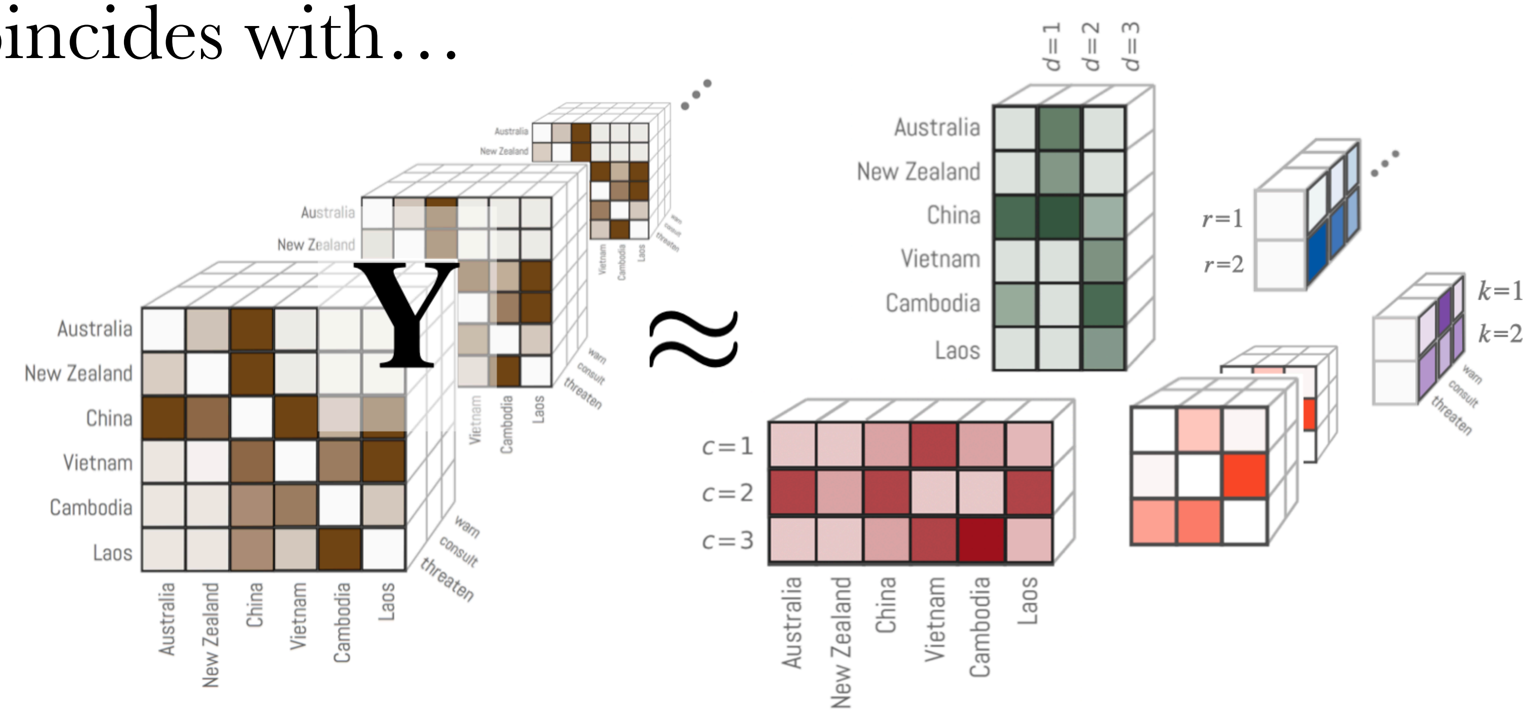
(Repeat)

ALLOCATIVE POISSON FACTORIZATION

“**Latent source**” representation [Dunson & Herring, (2005); Cemgil (2009)]:

Decomposition of a tensor

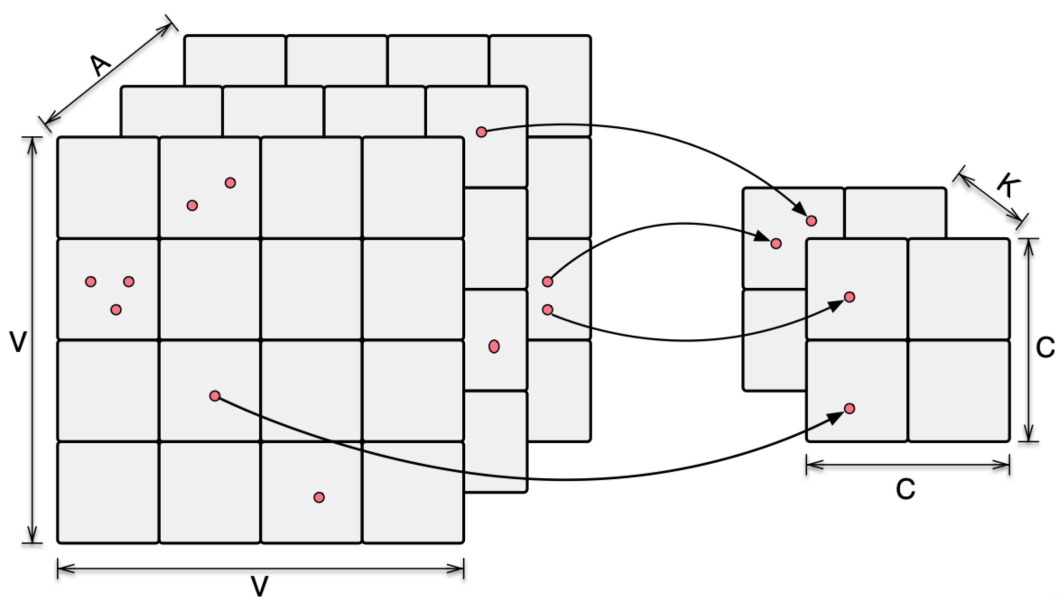
coincides with…



…**latent class allocation** of its tokens

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

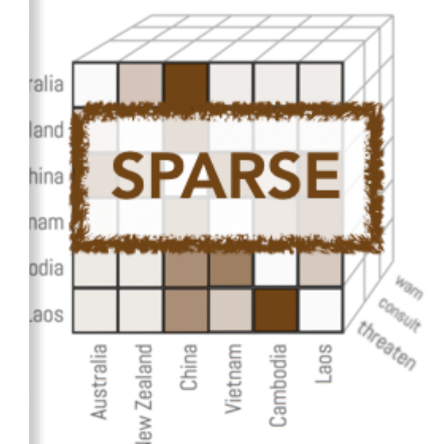
Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE



$\sim \mathrm{Pois}(\mu_{\mathbf{i},\kappa})$

(“sources”)

ing form:



ALLOCATIVE POISSON FACTORIZATION

Folk theorem of statistical computing [Gelman, 2008]:

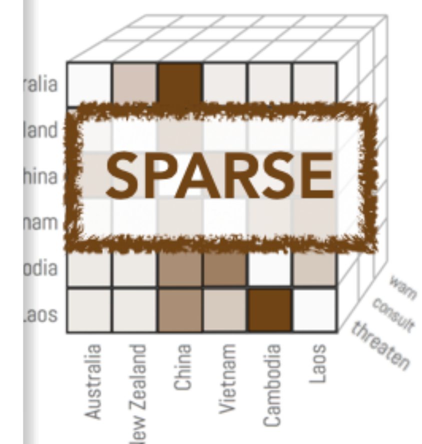
“When you have computational problems, often there’s a problem with your model.”

“**Late**

$\sim \text{Pois}(\mu_{\mathbf{i},\kappa})$

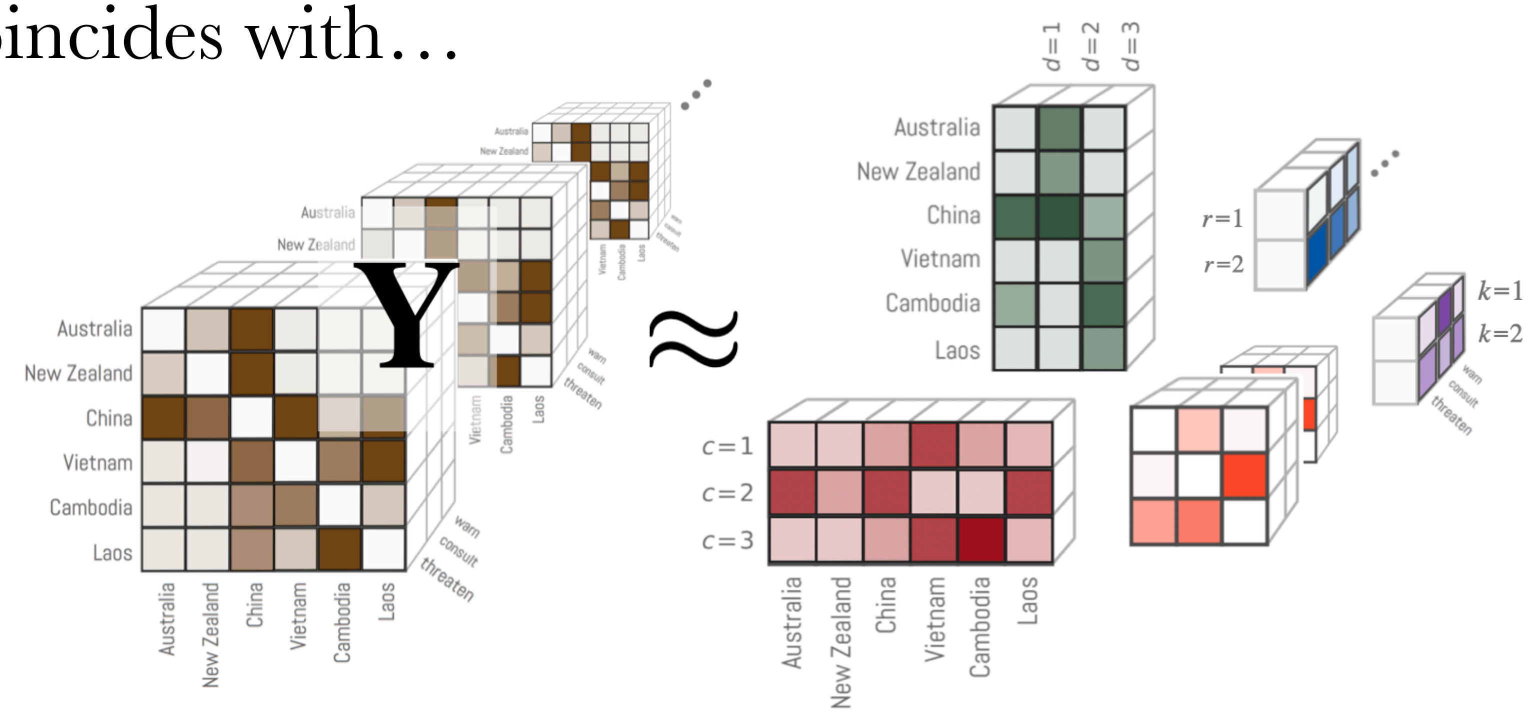
(“sources”)

ing form:



Decomposition of a tensor

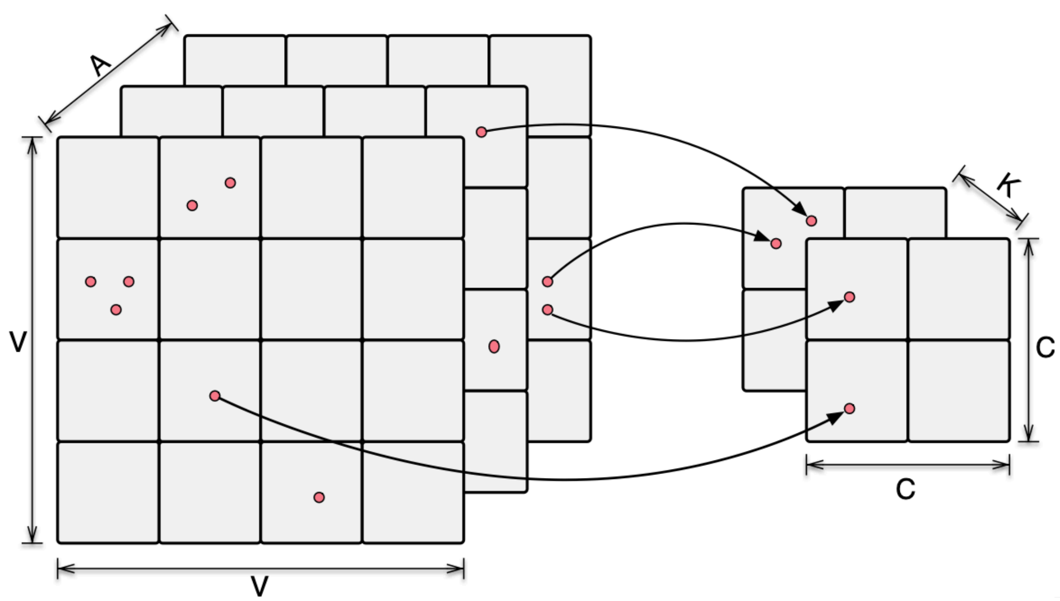
coincides with…



…**latent class allocation** of its tokens

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

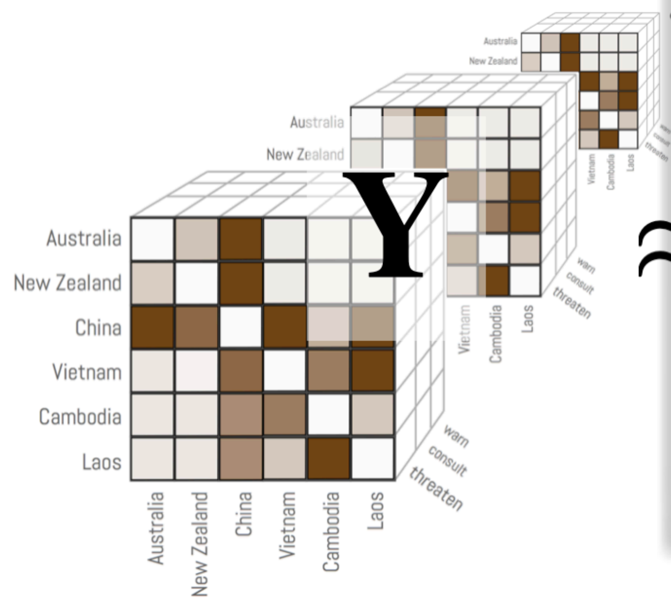


ALLOCATIVE POISSON FACTORIZATION

“Latent source” representation [Dunson & Herring, (2005); Cemgil (2009)]:

Decomposition of a t

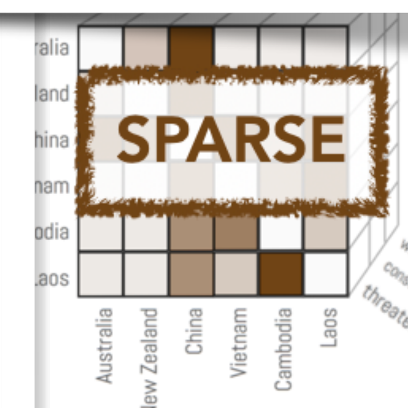
coincides with…



≈

Equivalence to models specified explicitly for tokens:

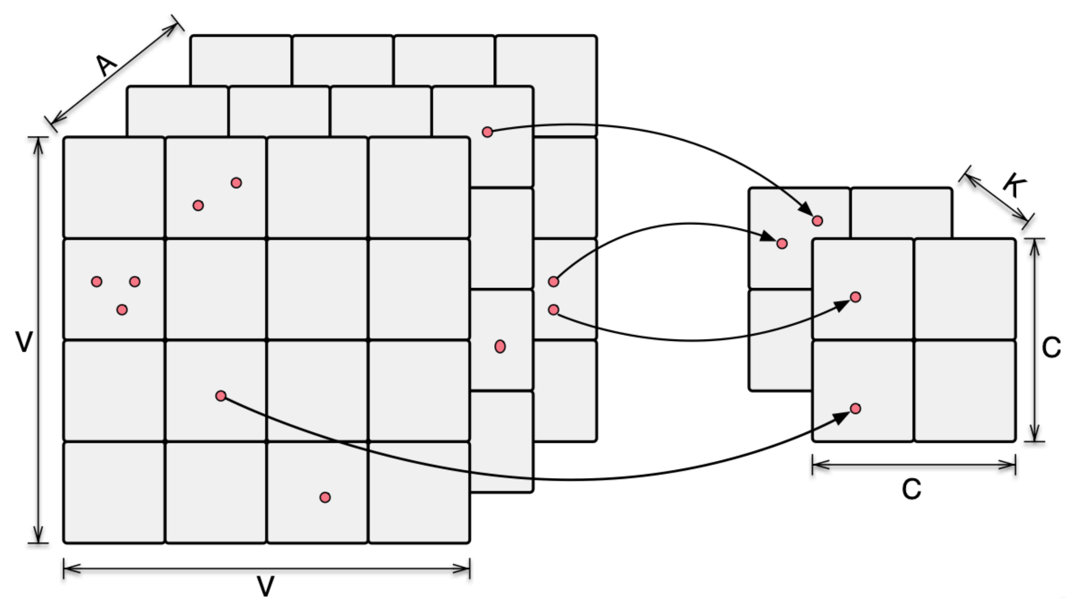
- Tensor decomposition for contingency table analysis: […] [Dunson & Xing, 2012], [Zhou et al., 2015], [Johndrow, Bhattacharya & Dunson, 2017] […]
- Latent class models: […] [Gu & Dunson, 2023] […]
- Topic models: e.g., [Blei et al., 2003]
- Graphical models: e.g., [Lauritzen, 1996]



…**latent class allocation** of its tokens

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE



ALLOCATIVE POISSON FACTORIZATION

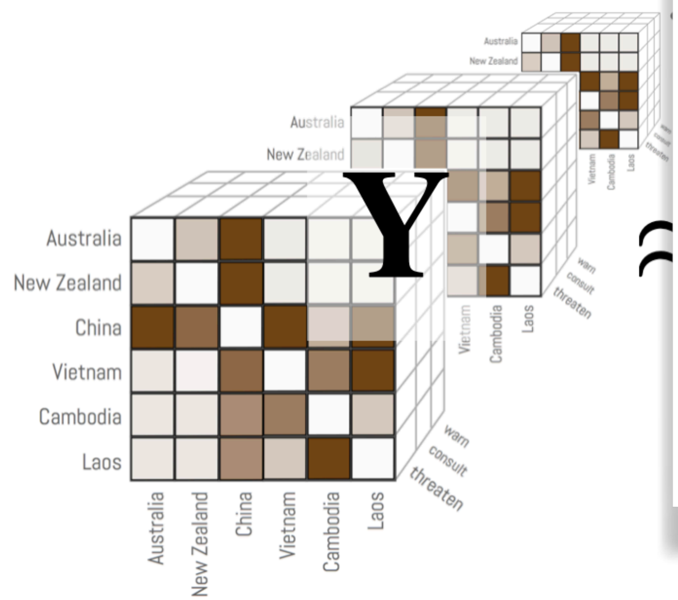
Folk theorem of statistical computing [Gelman, 2008]:

"When you have computational problems, often there's a problem with your model."

"**Late**

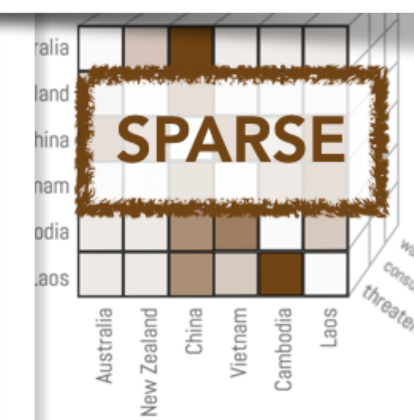
Decomposition of a t

coincides with…



Equivalence to models specified explicitly for tokens:

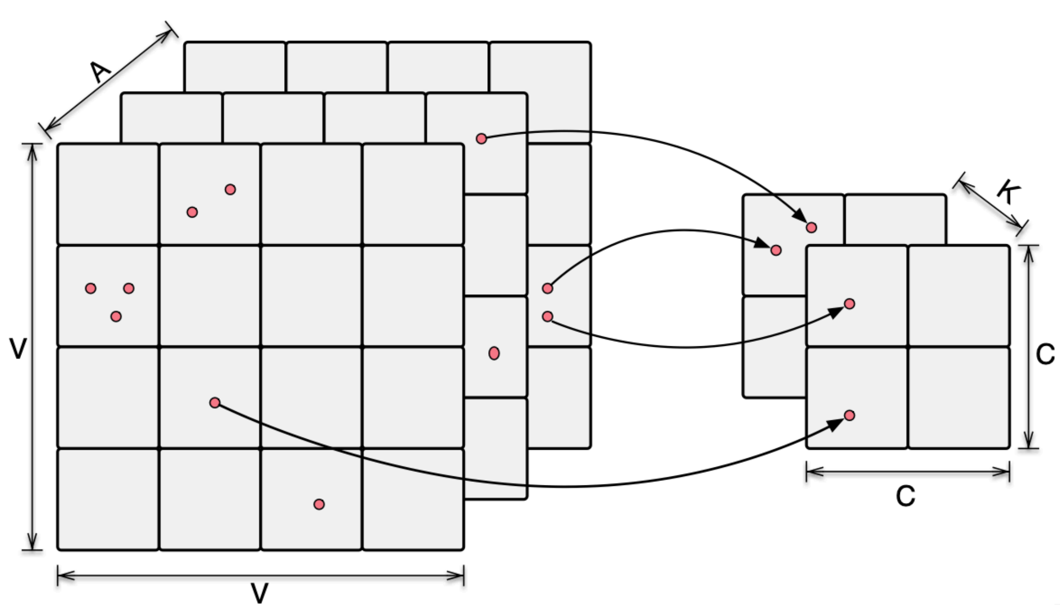
- Tensor decomposition for contingency table analysis: […] [Dunson & Xing, 2012], [Zhou et al., 2015], [Johndrow, Bhattacharya & Dunson, 2017] […]
- Latent class models: […] [Gu & Dunson, 2023] […]
- Topic models: e.g., [Blei et al., 2003]
- Graphical models: e.g., [Lauritzen, 1996]



…**latent class allocation** of its tokens

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

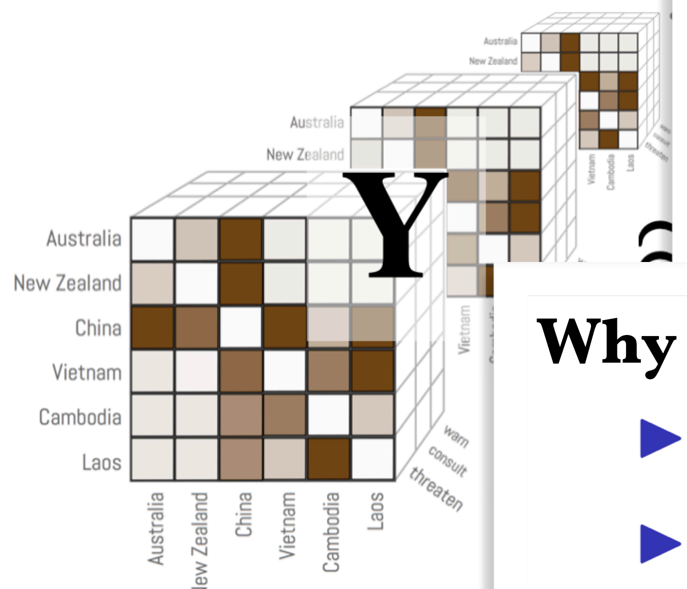


ALLOCATIVE POISSON FACTORIZATION

"Latent source" representation [Dunson & Herring, (2005); Cemgil (2009)]:

Decomposition of a t

coincides with…



…latent class all

Table 2
WEIS Coding of 1990 Iraq-Kuwait C

Date	Source	Target
900717	IRQ	KUW
900725	IRQ	EGY
900727	IRQ	KUW
900731	IRQ	KUW
900801	KUW	IRQ
900802	IRQ	KUW

Equivalence to models specified explicitly for tokens:

- Tensor decomposition for contingency table analysis: […] [Dunson & Xing, 2012], [Zhou et al., 2015], [Johndrow, Bhattacharya & Dunson, 2017] […]
- Latent class models: […] [Gu & Dunson, 2023] […]

Why this is particularly important in this setting:

- ICEWS (1995-2023): **20M tokens**
- (country x country x action x month)

  =(250 x 250 x 20 x 336)

  ≈ **450 million entries**
- (country x country x action x day x publisher)

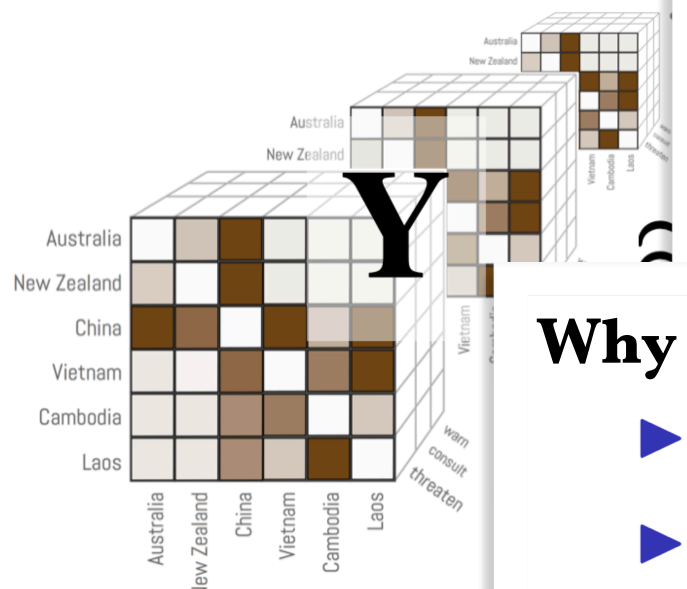
  =(250 x 250 x 20 x 10,220 x 200)

  ≈ **2.5 trillion entries**
- **When we add columns, or increase granularity, we get enormous tensors, but inference is still $\mathcal{O}$(#tokens)**

Folk theorem of statistical computing [Gelman, 2008]:
"When you have computational problems, often there's a problem with your model."

"Late

Decomposition of a t
coincides with…



…latent class all

Table 2
WEIS Coding of 1990 Iraq-Kuwait C

Date	Source	Target
900717	IRQ	KUW
900725	IRQ	EGY
900727	IRQ	KUW
900731	IRQ	KUW
900801	KUW	IRQ
900802	IRQ	KUW

Equivalence to models specified explicitly for tokens:

- Tensor decomposition for contingency table analysis: […] [Dunson & Xing, 2012], [Zhou et al., 2015], [Johndrow, Bhattacharya & Dunson, 2017] […]
- Latent class models: […] [Gu & Dunson, 2023] […]

Why this is particularly important in this setting:

- ICEWS (1995-2023): **20M tokens**
- (country x country x action x month)

  =(250 x 250 x 20 x 336)

  ≈ **450 million entries**
- (country x country x action x day x publisher)

  =(250 x 250 x 20 x 10,220 x 200)

  ≈ **2.5 trillion entries**
- **When we add columns, or increase granularity, we get enormous tensors, but inference is still $\mathcal{O}$(#tokens)**

NO FREE LUNCH

Allocative Poisson factorization…

$$y_{\mathbf{i}} \sim \text{Pois}(\underbrace{\textstyle\sum \ldots}_{\text{non-negative}})$$

…means non-negative priors.

Fine for simple i.id priors (e.g., Gamma, Dirichlet).

NO FREE LUNCH

Allocative Poisson factorization…

$$y_{\mathbf{i}} \sim \mathrm{Pois}(\sum \ldots)$$

non-negative

…means non-negative priors.

Fine for simple i.id priors (e.g., Gamma, Dirichlet).

More challenging for building complex priors, e.g.,

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \,|\, \boldsymbol{\theta}^{(t-1)})$$

Often requires new tools (e.g., augmentation schemes).

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts “multilateral” structure [Schein et al., 2015]
  - **Tucker** extracts “communities” [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the “exponential blowup” [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling “escalation”** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

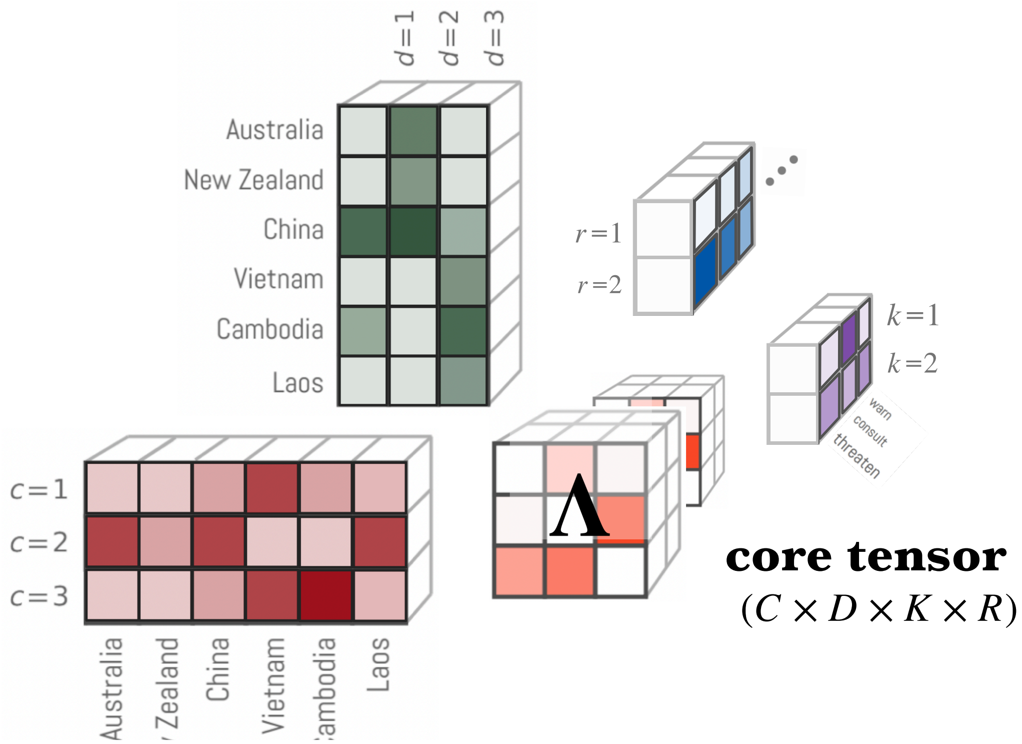
Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

TUCKER: APPEALING

Tucker's **conceptual appeal**

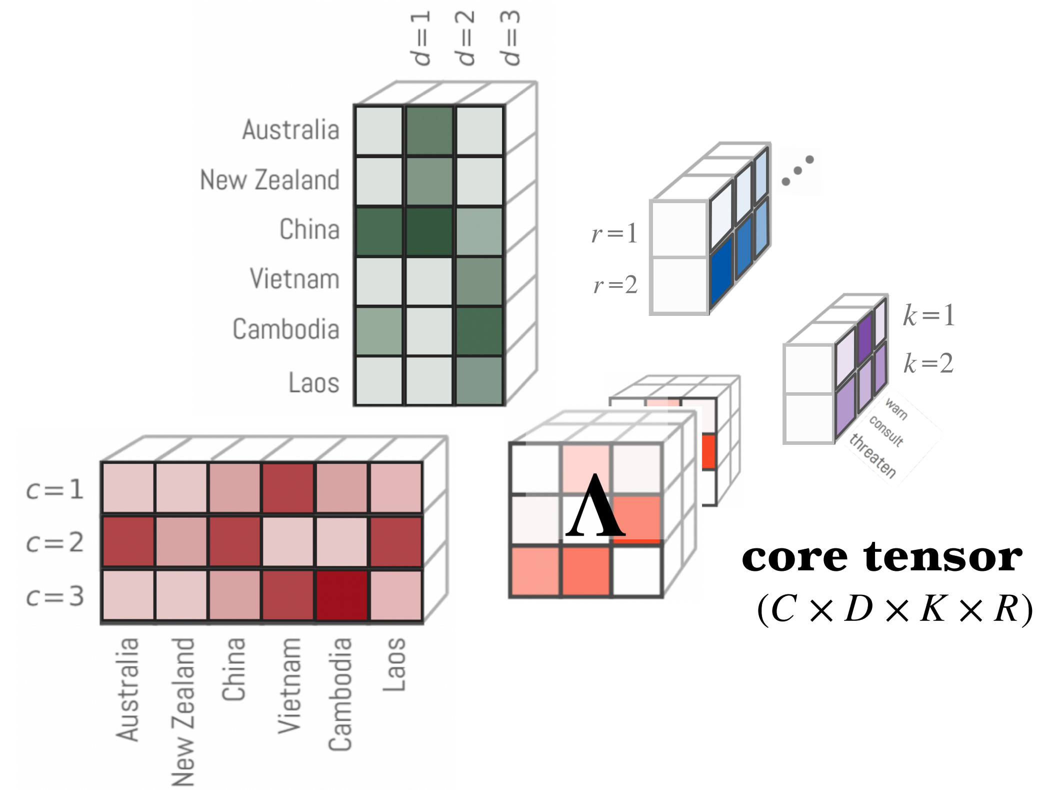
- Different number of factors in each mode
- Mode-specific interpretation of factors
  - e.g., "communities" of countries
  - e.g., "topics of actions
- Factor sharing across classes; less redundancy



TUCKER: APPEALING

Tucker's **conceptual appeal**

- Different number of factors in each mode
- Mode-specific interpretation of factors
  - e.g., "communities" of countries
  - e.g., "topics of actions
- Factor sharing across classes; less redundancy



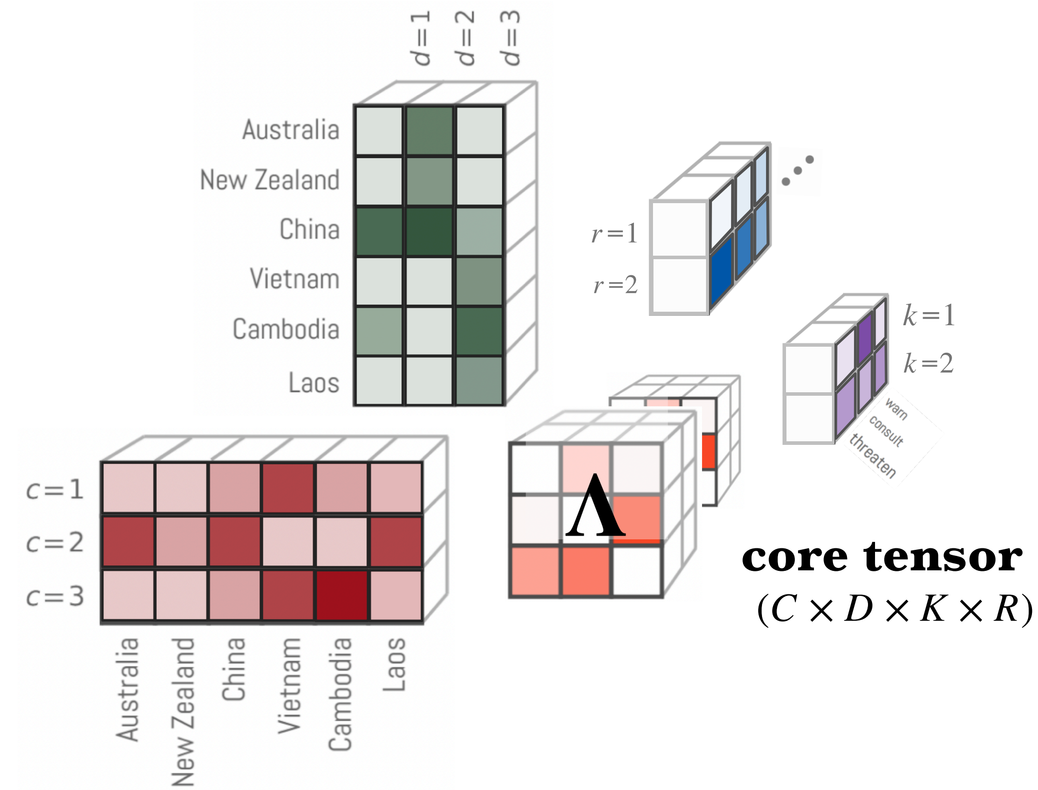
…is not **practically achievable**:

- Typically $\mathcal{O}(|\mathbf{Y}| \cdot |\boldsymbol{\Lambda}|)$, sometimes worse
- "Exponential blowup": $|\boldsymbol{\Lambda}|$ is **exponential** in M

TUCKER: APPEALING

Tucker's **conceptual appeal**

- Different number of factors in each mode
- Mode-specific interpretation of factors
  - e.g., "communities" of countries
  - e.g., "topics of actions
- Factor sharing across classes; less redundancy



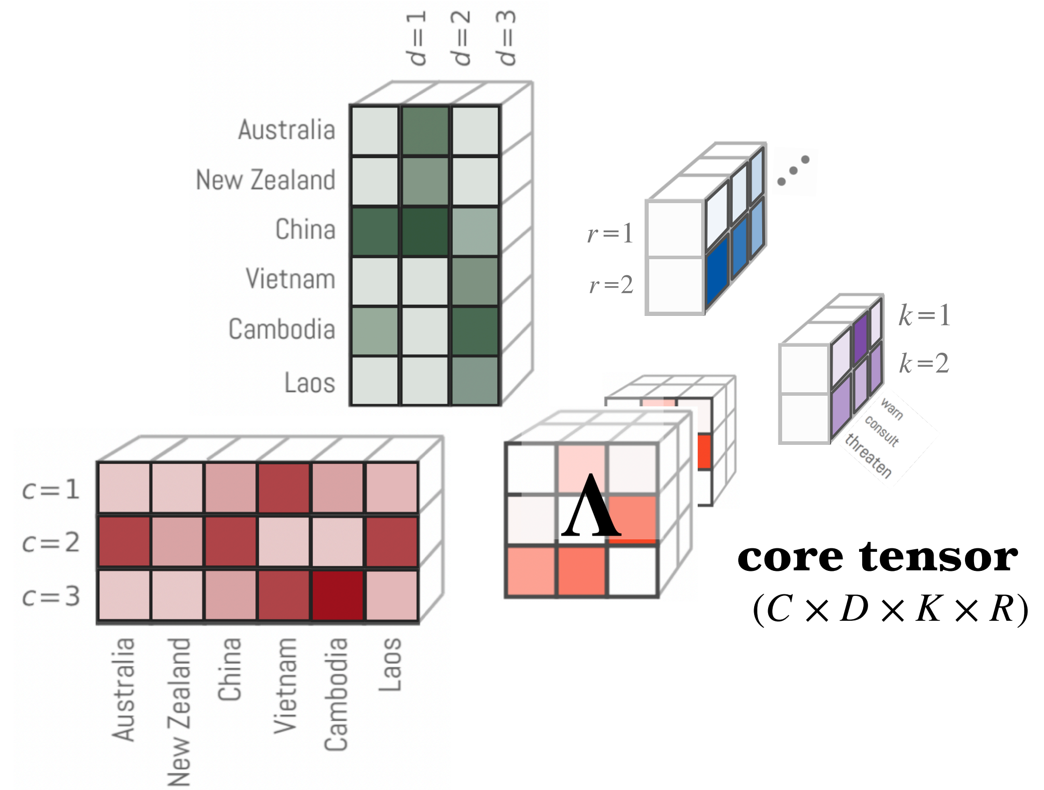
…is not **practically achievable**:

- Typically $\mathcal{O}(|\mathbf{Y}| \cdot |\boldsymbol{\Lambda}|)$, sometimes worse
- "Exponential blowup": $|\boldsymbol{\Lambda}|$ is **exponential** in M
- In practice: small cores e.g., (3 x 3 x 3)
- Not consistent with desired interpretation

TUCKER: APPEALING

Tucker's **conceptual appeal**

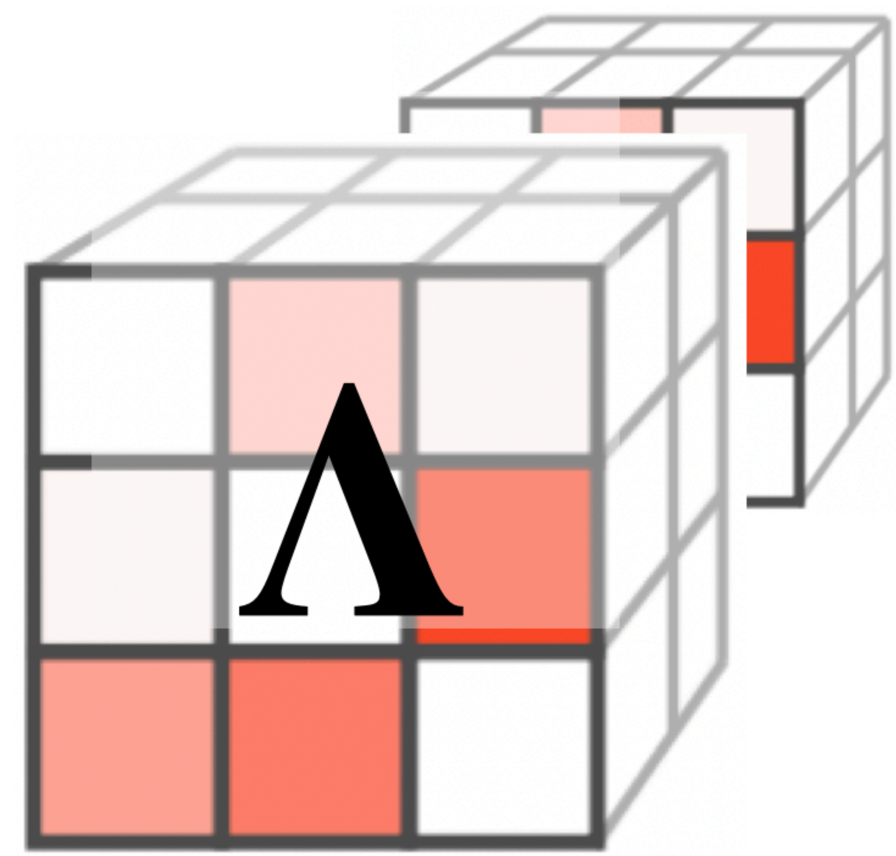
- Different number of factors in each mode
- Mode-specific interpretation of factors
  - e.g., "communities" of countries
  - e.g., "topics of actions
- Factor sharing across classes; less redundancy



…is not **practically achievable**:

- Typically $\mathcal{O}(|\mathbf{Y}| \cdot |\boldsymbol{\Lambda}|)$, sometimes worse
- "Exponential blowup": $|\boldsymbol{\Lambda}|$ is **exponential** in M
- In practice: small cores e.g., (3 x 3 x 3)
- Not consistent with desired interpretation
- e.g., 50 communities, 10 topics, 25 regimes =

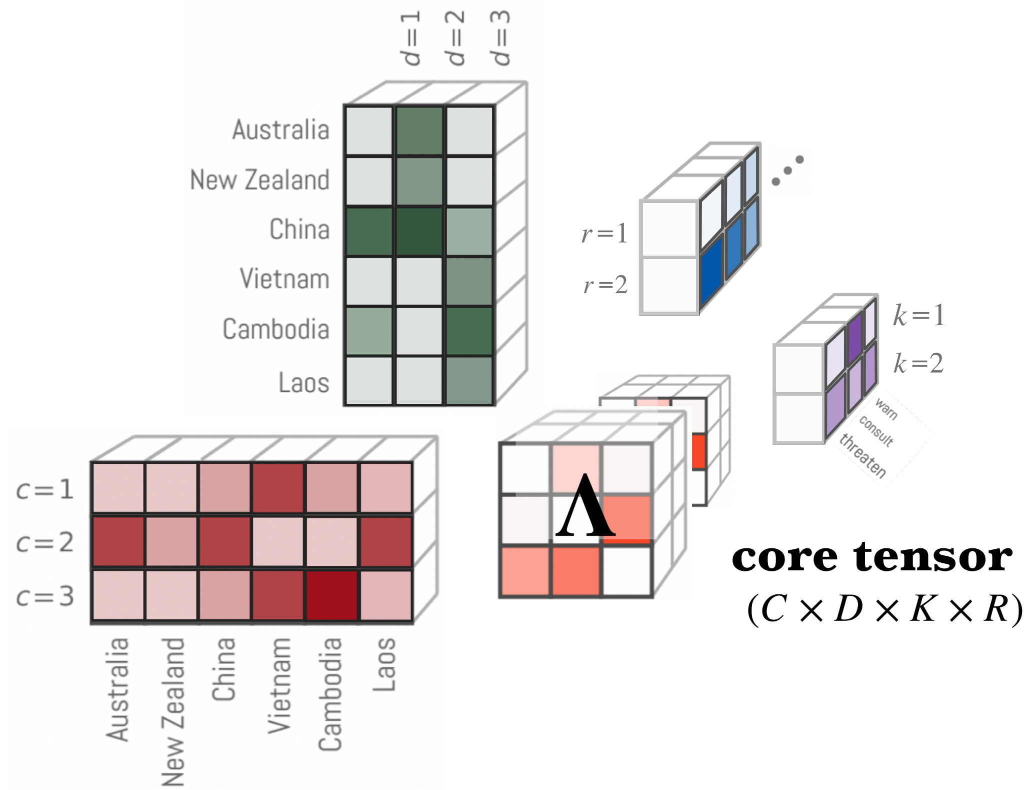
  625,000 classes



TUCKER: APPEALING

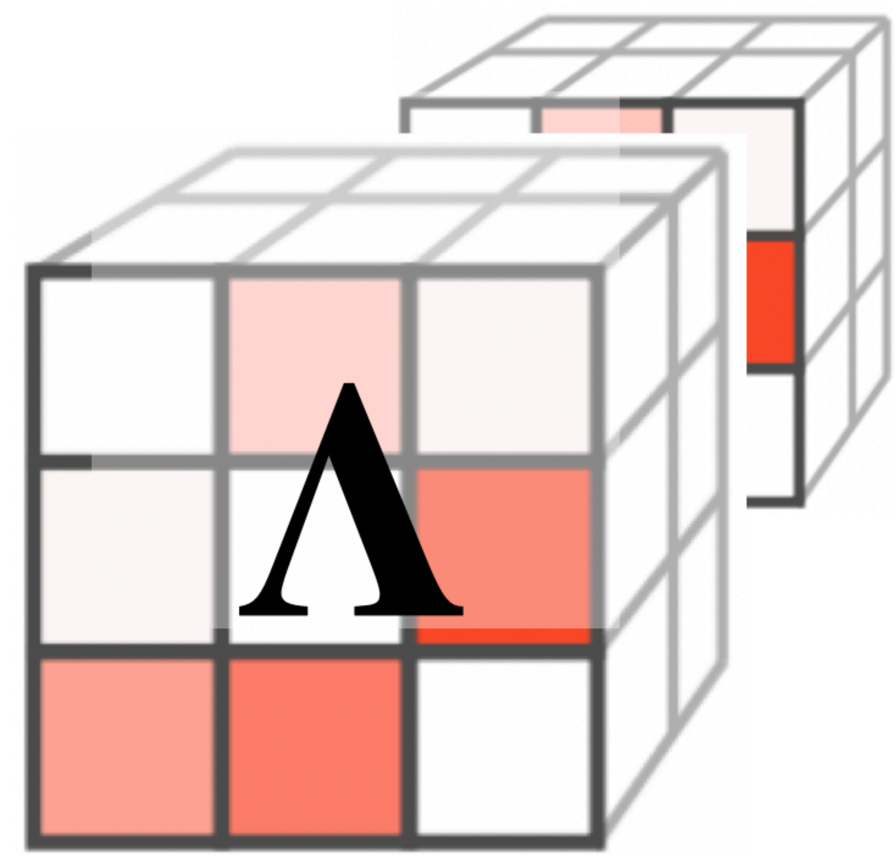
Tucker's **conceptual appeal**

- Different number of factors in each mode
- Mode-specific interpretation of factors
  - e.g., "communities" of countries
  - e.g., "topics of actions
- Factor sharing across classes; less redundancy



…is not **practically achievable**:

- Typically $\mathcal{O}(|\mathbf{Y}| \cdot |\mathbf{\Lambda}|)$, sometimes worse
- "Exponential blowup": $|\mathbf{\Lambda}|$ is **exponential** in M
- In practice: small cores e.g., (3 x 3 x 3)
- Not consistent with desired interpretation
- e.g., 50 communities, 10 topics, 25 regimes = 625,000 classes



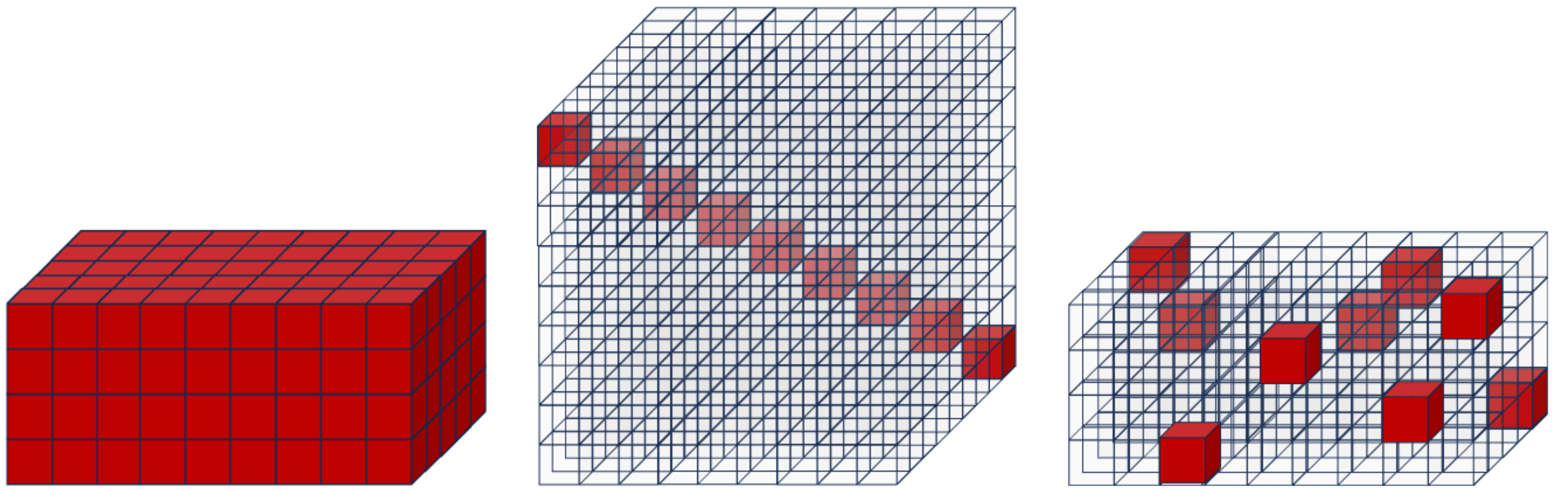
Poisson Tucker better, but still bad:

- $\mathcal{O}(||\mathbf{Y}||_0 \cdot |\mathbf{\Lambda}|)$

ALℓ_0CORE [Hood & Schein, 2024]

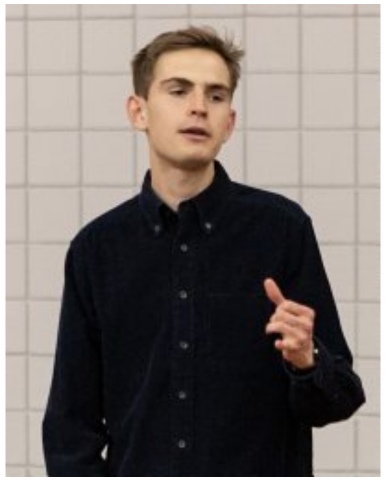
The basic idea: an ℓ_0-constrained core:

$$||\mathbf{\Lambda}||_0 \leq Q$$



(a) Tucker (b) CP (c) ALℓ_0CORE

Lead author:

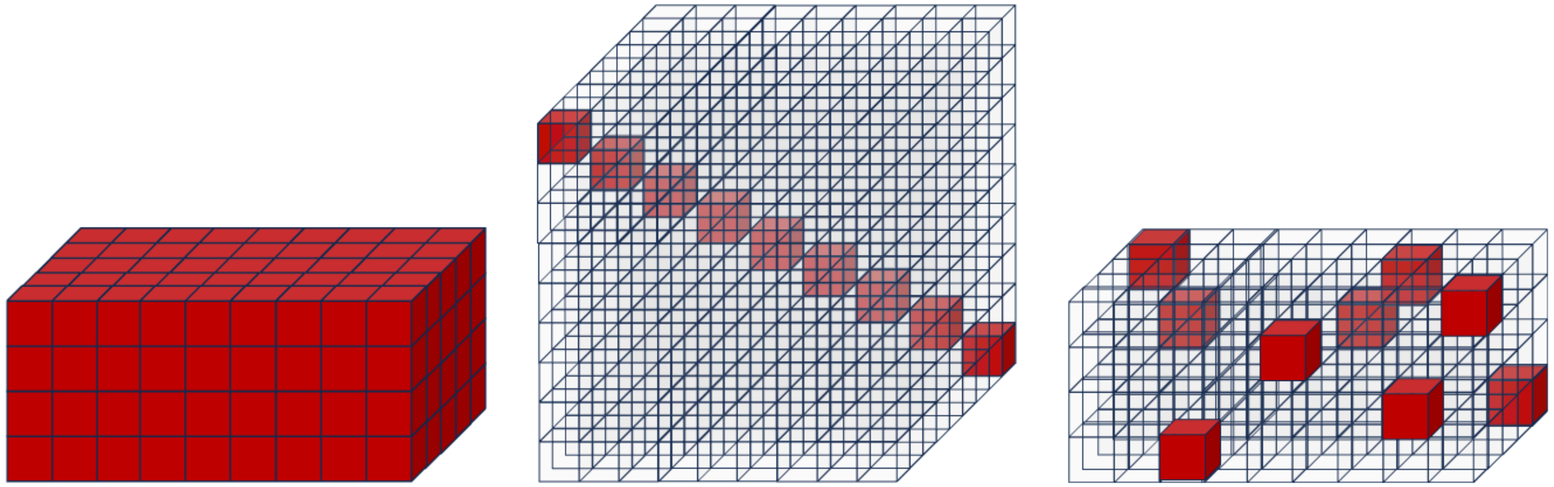


John Hood

ALℓ_0CORE [Hood & Schein, 2024]

The basic idea: an ℓ_0-constrained core:

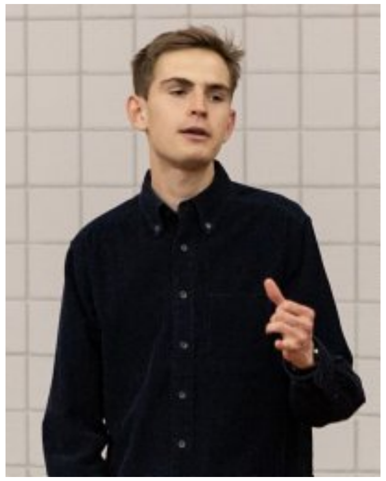
$$||\boldsymbol{\Lambda}||_0 \leq Q$$



(a) Tucker (b) CP (c) ALℓ_0CORE

- Q is a **budget** set by the user (for now)
- $\lambda_1, \ldots, \lambda_Q$ non-zero values
- $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_Q$ non-zero locations
  - $\kappa_q \equiv \big(c[q], d[q]\, a[q], r[q]\big)$

Lead author:

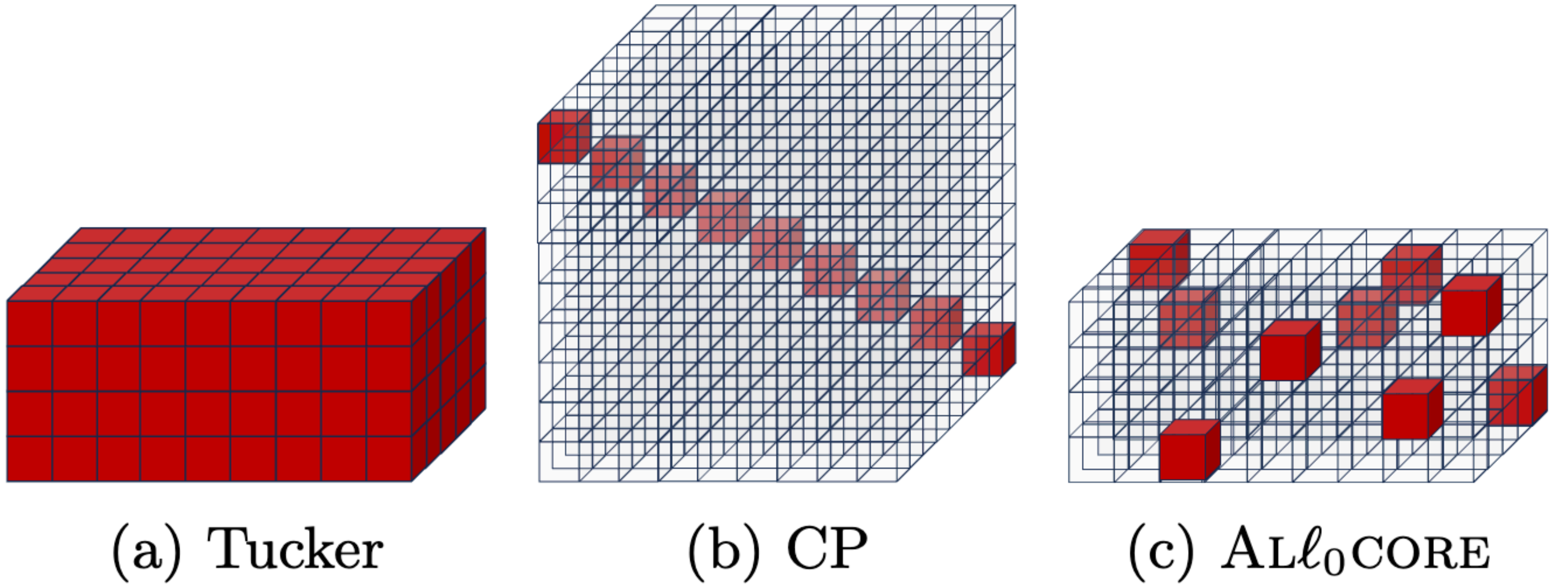


John Hood

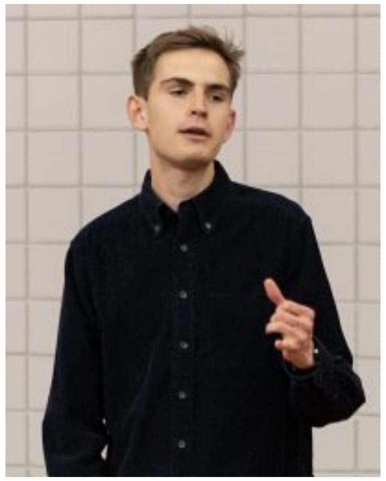
ALℓ_0CORE [Hood & Schein, 2024]

The basic idea: an ℓ_0-constrained core:

$$||\mathbf{\Lambda}||_0 \leq Q$$



Lead author:

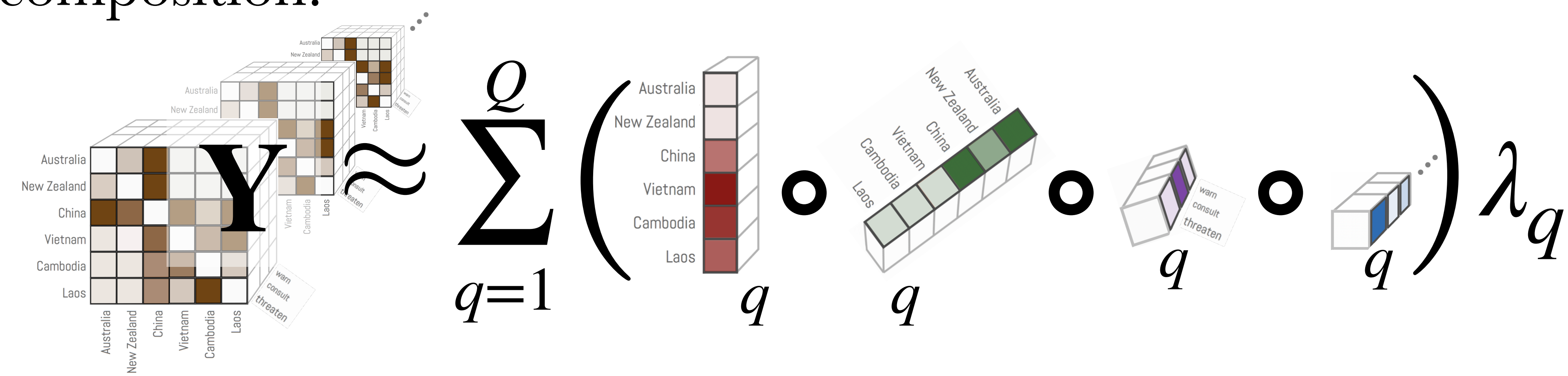


John Hood

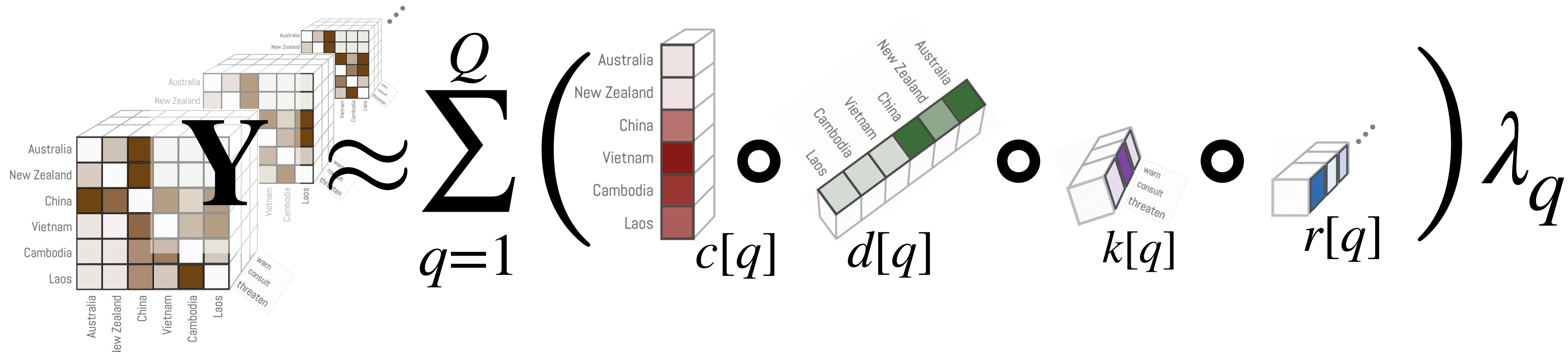
- Q is a **budget** set by the user (for now)
- $\lambda_1, \ldots, \lambda_Q$ non-zero values
- $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_Q$ non-zero locations
  - $\kappa_q \equiv \left(c[q], d[q]\, a[q], r[q]\right)$
- Inference involves **allocating non-zeros** across the core
- ALℓ_0CORE = "allocated ℓ_0-constrained core"

COMPARISON OF ALL THREE

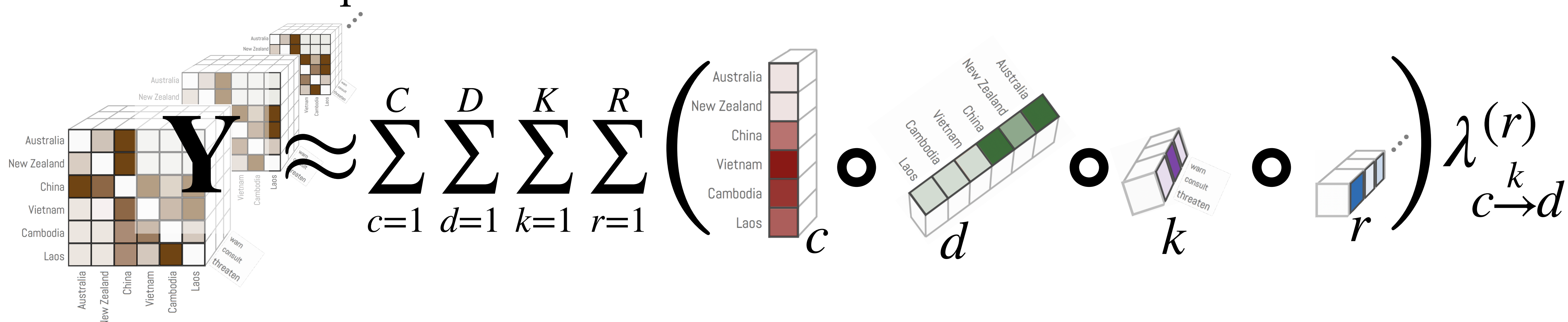
- CP decomposition:



- ALℓ_0CORE decomposition:



- Tucker decomposition



SAMPLING-BASED INFERENCE

- Prior for a $(K_1 \times \ldots \times K_M)$ core

$$\kappa_q \sim \text{Categorical}(\Pi)$$

SAMPLING-BASED INFERENCE

- Prior for a $(K_1 \times \ldots \times K_M)$ core

$$\kappa_q \sim \text{Categorical}(\Pi)$$

- For a rank-1 prior tensor $(\Pi \equiv \pi_1 \circ \ldots \circ \pi_M)$:

$$\kappa_{q,m} \sim \text{Categorical}(\boldsymbol{\pi}_m)$$

SAMPLING-BASED INFERENCE

- Prior for a $(K_1 \times \ldots \times K_M)$ core

$$\kappa_q \sim \mathrm{Categorical}(\Pi)$$

- For a rank-1 prior tensor $(\Pi \equiv \pi_1 \circ \ldots \circ \pi_M)$:

$$\kappa_{q,m} \sim \mathrm{Categorical}(\boldsymbol{\pi}_m)$$

- Naively sampling the multi-index is exponential $\boxed{\mathcal{O}(\prod_{m=1}^{M} K_m)}$

$$P(\kappa_q \mid -\,) = \frac{\cdots}{\sum_\kappa \cdots}$$

SAMPLING-BASED INFERENCE

- Prior for a $(K_1 \times \ldots \times K_M)$ core

$$\kappa_q \sim \text{Categorical}(\Pi)$$

- For a rank-1 prior tensor $(\Pi \equiv \pi_1 \circ \ldots \circ \pi_M)$:

$$\kappa_{q,m} \sim \text{Categorical}(\boldsymbol{\pi}_m)$$

- Naively sampling the multi-index is exponential $\boxed{\mathcal{O}(\prod_{m=1}^{M} K_m)}$

$$P(\kappa_q \mid -) = \frac{\cdots}{\sum_{\kappa} \cdots}$$

- Gibbs **sampling the sub-indices** is not $\boxed{\mathcal{O}(\sum_{m=1}^{M} K_m)}$

$$P(\kappa_{q,m} \mid \kappa_{q,\neg m} -) = \frac{\cdots}{\sum_{k_m} \cdots}$$

SAMPLING-BASED INFERENCE

- Prior for a $(K_1 \times \ldots \times K_M)$ core

$$\kappa_q \sim \text{Categorical}(\Pi)$$

- For a rank-1 prior tensor $(\Pi \equiv \pi_1 \circ \ldots \circ \pi_M)$:

$$\kappa_{q,m} \sim \text{Categorical}(\boldsymbol{\pi}_m)$$

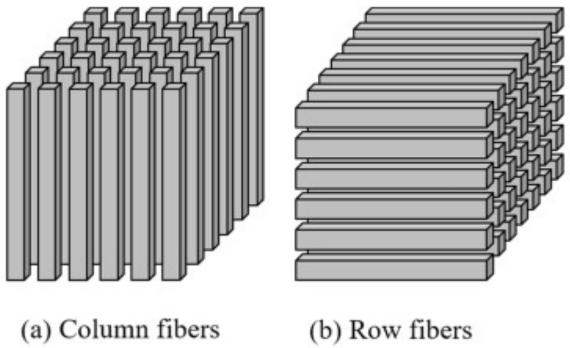
- Naively sampling the multi-index is exponential $\mathcal{O}(\prod_{m=1}^{M} K_m)$

$$P(\kappa_q \mid -) = \frac{\cdots}{\sum_{\kappa} \cdots}$$

- Gibbs **sampling the sub-indices** is not $\mathcal{O}(\sum_{m=1}^{M} K_m)$

$$P(\kappa_{q,m} \mid \kappa_{q,\neg m} -) = \frac{\cdots}{\sum_{k_m} \cdots}$$

- Each step re-allocates λ_q within a **fiber** of the core



SAMPLING-BASED INFERENCE

- Prior for a $(K_1 \times \ldots \times K_M)$ core

$$\kappa_q \sim \text{Categorical}(\Pi)$$

- For a rank-1 prior tensor $(\Pi \equiv \pi_1 \circ \ldots \circ \pi_M)$:

$$\kappa_{q,m} \sim \text{Categorical}(\boldsymbol{\pi}_m)$$

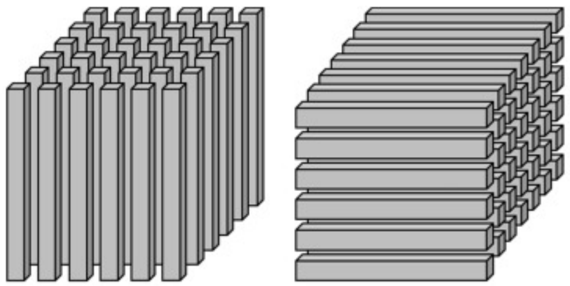
- Naively sampling the multi-index is exponential $\mathcal{O}(\prod_{m=1}^{M} K_m)$

$$P(\kappa_q \mid -) = \frac{\cdots}{\sum_{\kappa} \cdots}$$

- Gibbs **sampling the sub-indices** is not $\mathcal{O}(\sum_{m=1}^{M} K_m)$

$$P(\kappa_{q,m} \mid \kappa_{q,\neg m} -) = \frac{\cdots}{\sum_{k_m} \cdots}$$

- Each step re-allocates λ_q within a **fiber** of the core
- (We can do even better with specific priors, but this works surprisingly well.)

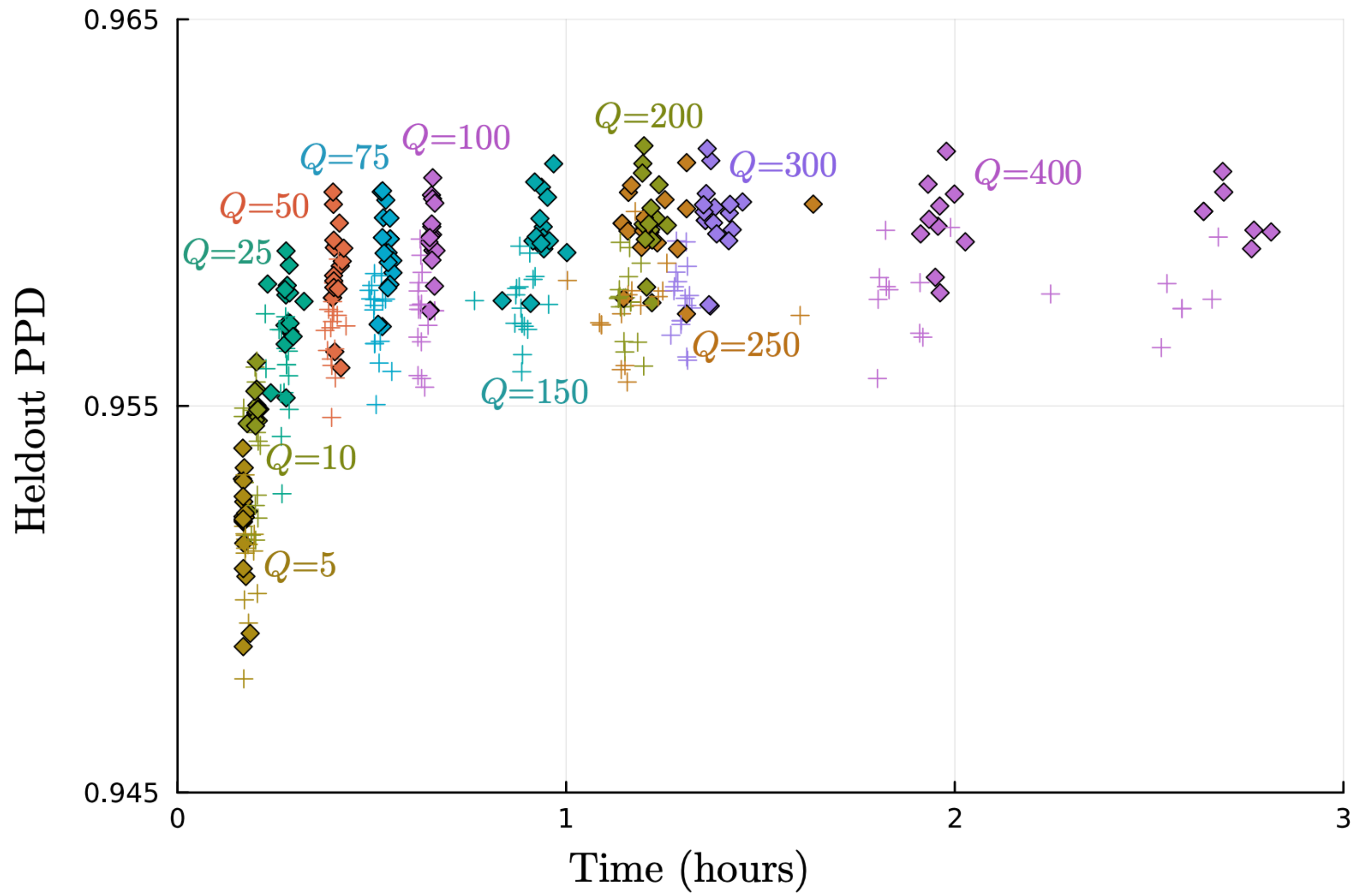


(a) Column fibers (b) Row fibers

Experiments

How dense does the core need to be?

- Consider two core shapes: small $(20 \times 20 \times 6 \times 3)$ and large $(50 \times 50 \times 6 \times 10)$
- Run $\text{A}\text{L}\ell_0\text{CORE}$ with $Q \in \{5, \ldots, 400\}$
- (Note: $Q = 400$ is 5% and 0.25% dense for the small and large cores, respectively)

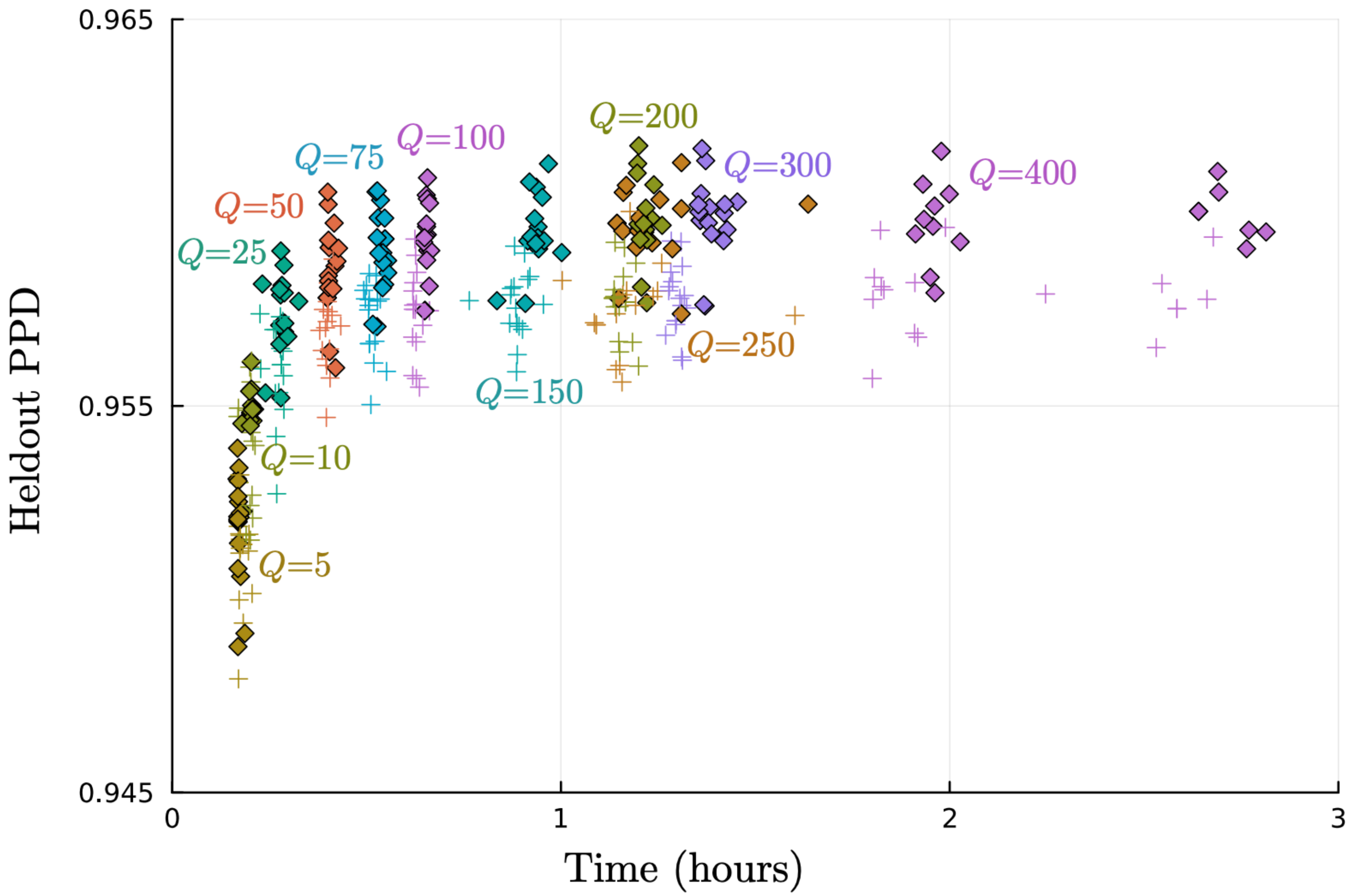


Performance plateaus very early → full Tucker is **wasting computation**.

Experiments

How dense does the core need to be?

- Consider two core shapes: small $(20 \times 20 \times 6 \times 3)$ and large $(50 \times 50 \times 6 \times 10)$
- Run $\text{AL}\ell_0\text{CORE}$ with $Q \in \{5, \ldots, 400\}$
- (Note: $Q = 400$ is 5% and 0.25% dense for the small and large cores, respectively)



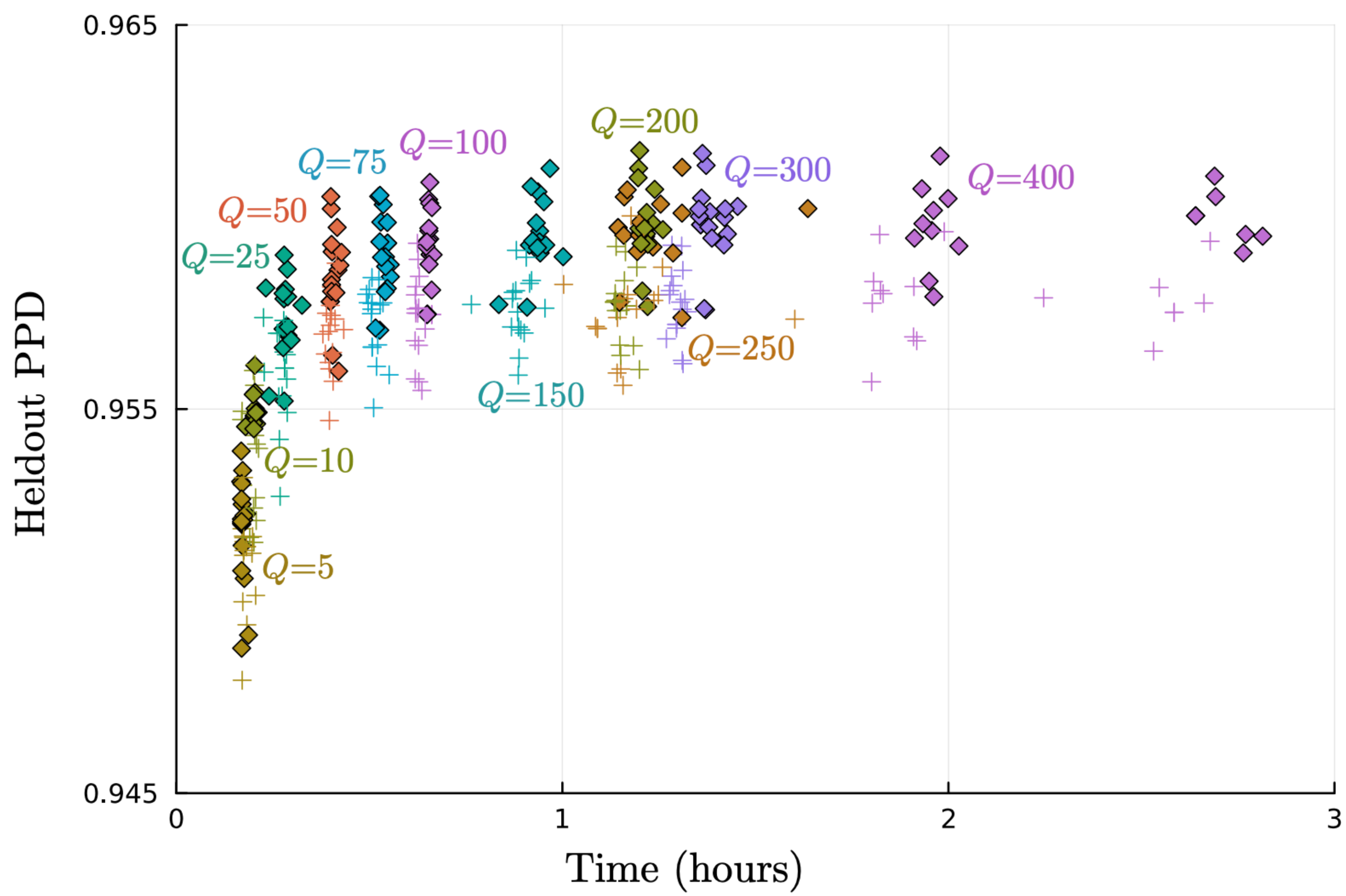
Performance plateaus very early → full Tucker is **wasting computation**.

- Gibbs sampling for full Tucker on the large core would take an estimated **30 days**

Experiments

How dense does the core need to be?

- Consider two core shapes: small $(20 \times 20 \times 6 \times 3)$ and large $(50 \times 50 \times 6 \times 10)$
- Run $\mathrm{AL}\ell_0\mathrm{CORE}$ with $Q \in \{5, \ldots, 400\}$
- (Note: $Q = 400$ is 5% and 0.25% dense for the small and large cores, respectively)



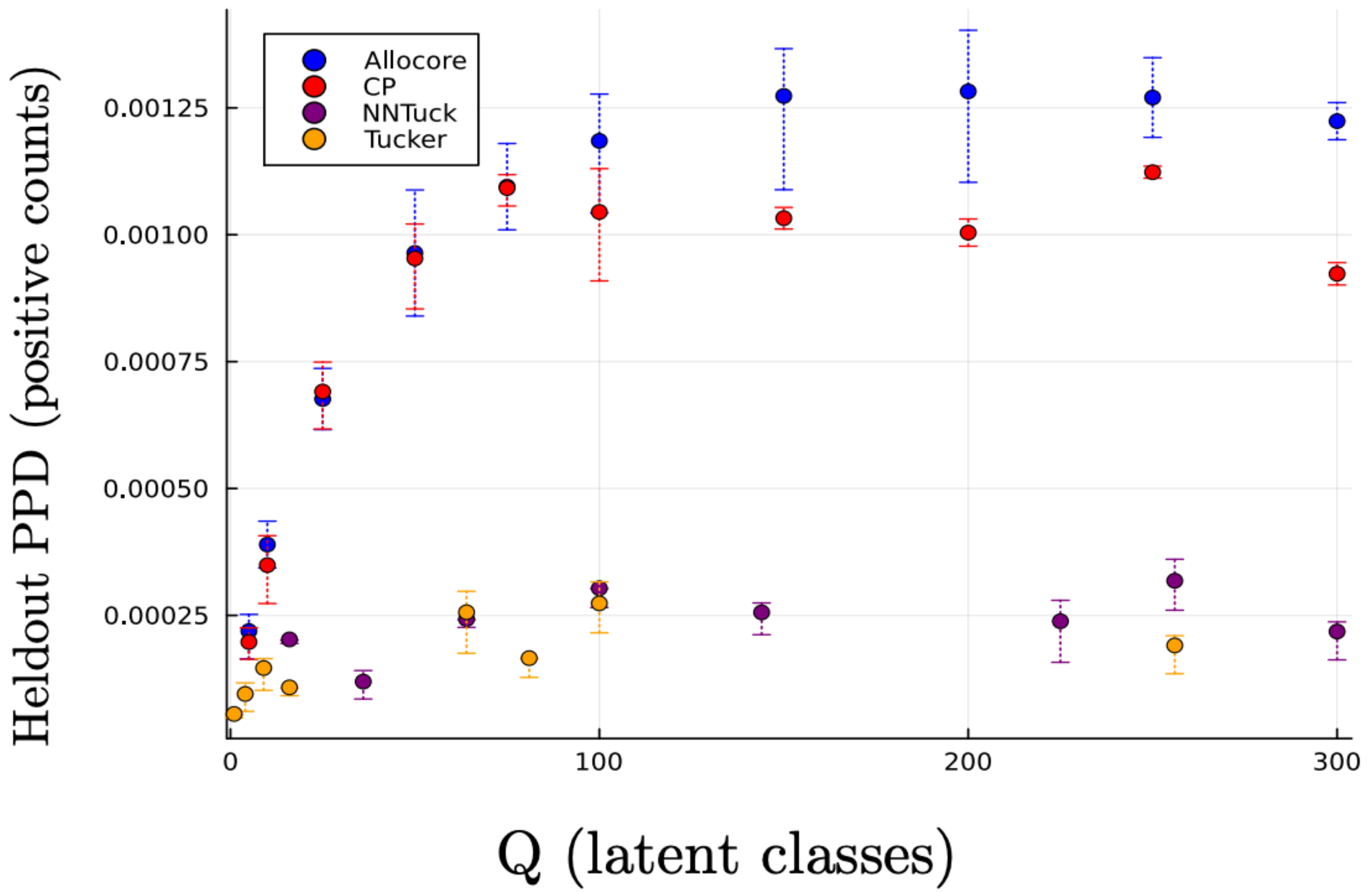
Performance plateaus very early → full Tucker is **wasting computation**.

- Gibbs sampling for full Tucker on the large core would take an estimated **30 days**
- EM with NNTUCK [Aguiar et al., 2023] would take an estimated **1.5 days**

EXPERIMENTS

How does it compare to CP and full Tucker with the **same number of parameters**?

- Set Q and use "**canonical**" $\textsc{Al}\ell_0\textsc{core}$ $(K_1 = \ldots K_M = Q)$
- Use CP decomposition with Q classes
- Use full Tucker ($\textsc{NNTuck}$) with a core size such that $|\boldsymbol{\Lambda}| \approx Q$



More parameter sharing than CP, and larger latent space than full Tucker.

POISSON ALℓ_0CORE

how active **country** *i* is
in (sender) **community** *c[q]*

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{i,c[q]}\, \psi^{(\leftarrow)}_{j,d[q]}\, \phi_{a,k[q]}\, \theta^{(t)}_{r[q]}\, \lambda_q\right)$$

POISSON ALℓ_0CORE

how active **country** *j* is
in (receiver) **community** *d[q]*

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{i,c[q]}\, \psi^{(\leftarrow)}_{j,d[q]}\, \phi_{a,k[q]}\, \theta^{(t)}_{r[q]}\, \lambda_q\right)$$

POISSON ALℓ_0CORE

how prevalent **action** a is
in **topic** $k[q]$

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{i,c[q]}\, \psi^{(\leftarrow)}_{j,d[q]}\, \phi_{a,k[q]}\, \theta^{(t)}_{r[q]}\, \lambda_q\right)$$

how active **regime** $r[q]$ is at **time** t

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{i,c[q]}\, \psi^{(\leftarrow)}_{j,d[q]}\, \phi_{a,k[q]}\, \theta^{(t)}_{r[q]}\, \lambda_q\right)$$

POISSON ALℓ_0CORE

allocated **non-zero value** q

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{q=1}^{Q} \psi^{(\rightarrow)}_{i,c[q]}\, \psi^{(\leftarrow)}_{j,d[q]}\, \phi_{a,k[q]}\, \theta^{(t)}_{r[q]}\, \lambda_q\right)$$

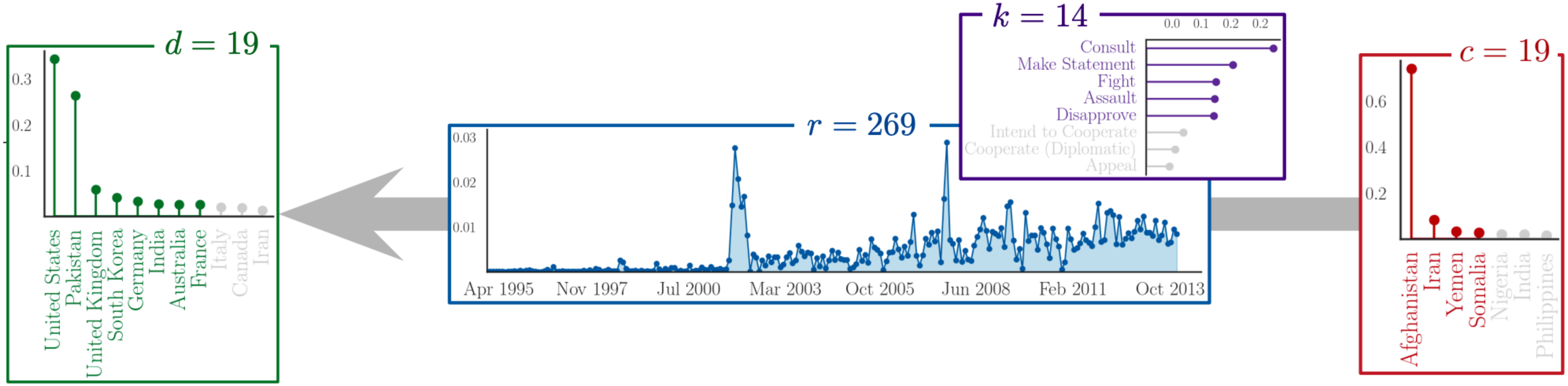
Example of inferred latent structure

Inferred non-zero core elements

rank	$\lambda^{(r)}_{c \xrightarrow{k} d}$	c	d	k	r
1	11.03	2	2	10	252
2	10.36	31	31	9	81
3	10.36	9	9	12	159
4	10.13	33	33	9	233
5	9.58	37	37	12	87
6	9.36	12	12	2	162
7	8.9	43	43	10	43
8	8.57	41	41	2	41
9	8.35	15	15	10	115
10	8.13	11	11	16	61
	...		...		
20	6.65	49	49	9	49
21	6.61	14	14	2	14
22	6.55	23	23	9	273
23	6.35	46	46	9	196
24	6.25	22	22	2	222
25	**6.24**	**40**	**40**	**19**	**40**
26	**6.23**	**19**	**19**	**14**	**269**
27	6.14	43	43	10	143
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	...		...		

- ICEWS tensor (1995-2013), monthly: (250 x 250 x 20 x 228)
- Core size: (50 x 50 x 20 x 300) → 15 million (possible) classes
- $Q = 400$

Example of inferred latent structure



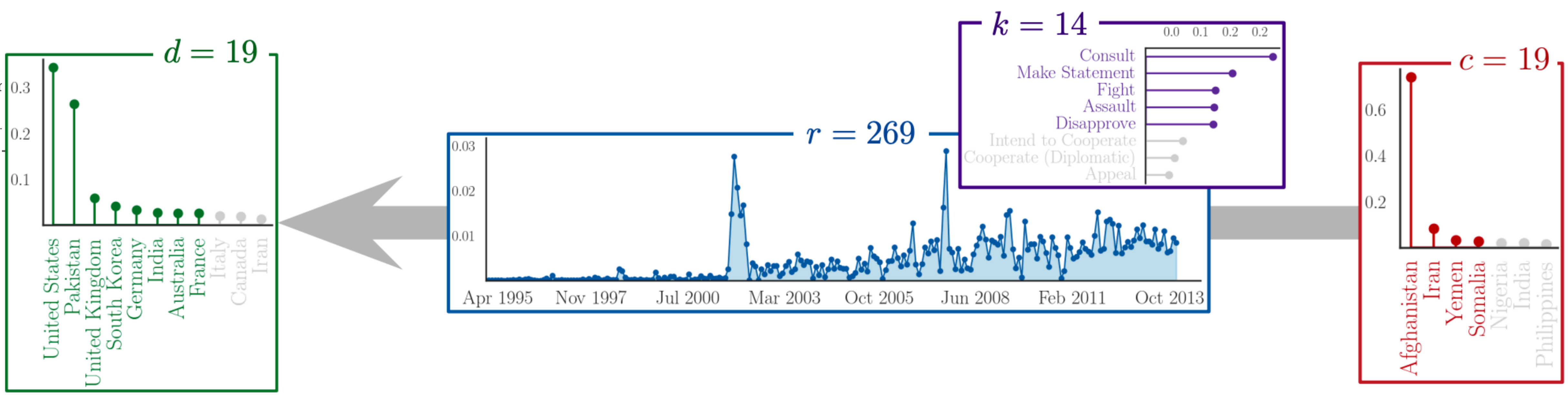
Inferred non-zero core elements

rank	$\lambda^{(r)}_{c \xrightarrow{k} d}$	c	d	k	r
1	11.03	2	2	10	252
2	10.36	31	31	9	81
3	10.36	9	9	12	159
4	10.13	33	33	9	233
5	9.58	37	37	12	87
6	9.36	12	12	2	162
7	8.9	43	43	10	43
8	8.57	41	41	2	41
9	8.35	15	15	10	115
10	8.13	11	11	16	61
	...		...		
20	6.65	49	49	9	49
21	6.61	14	14	2	14
22	6.55	23	23	9	273
23	6.35	46	46	9	196
24	6.25	22	22	2	222
25	**6.24**	**40**	**40**	**19**	**40**
26	**6.23**	**19**	**19**	**14**	**269**
27	6.14	43	43	10	143
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	...		...		

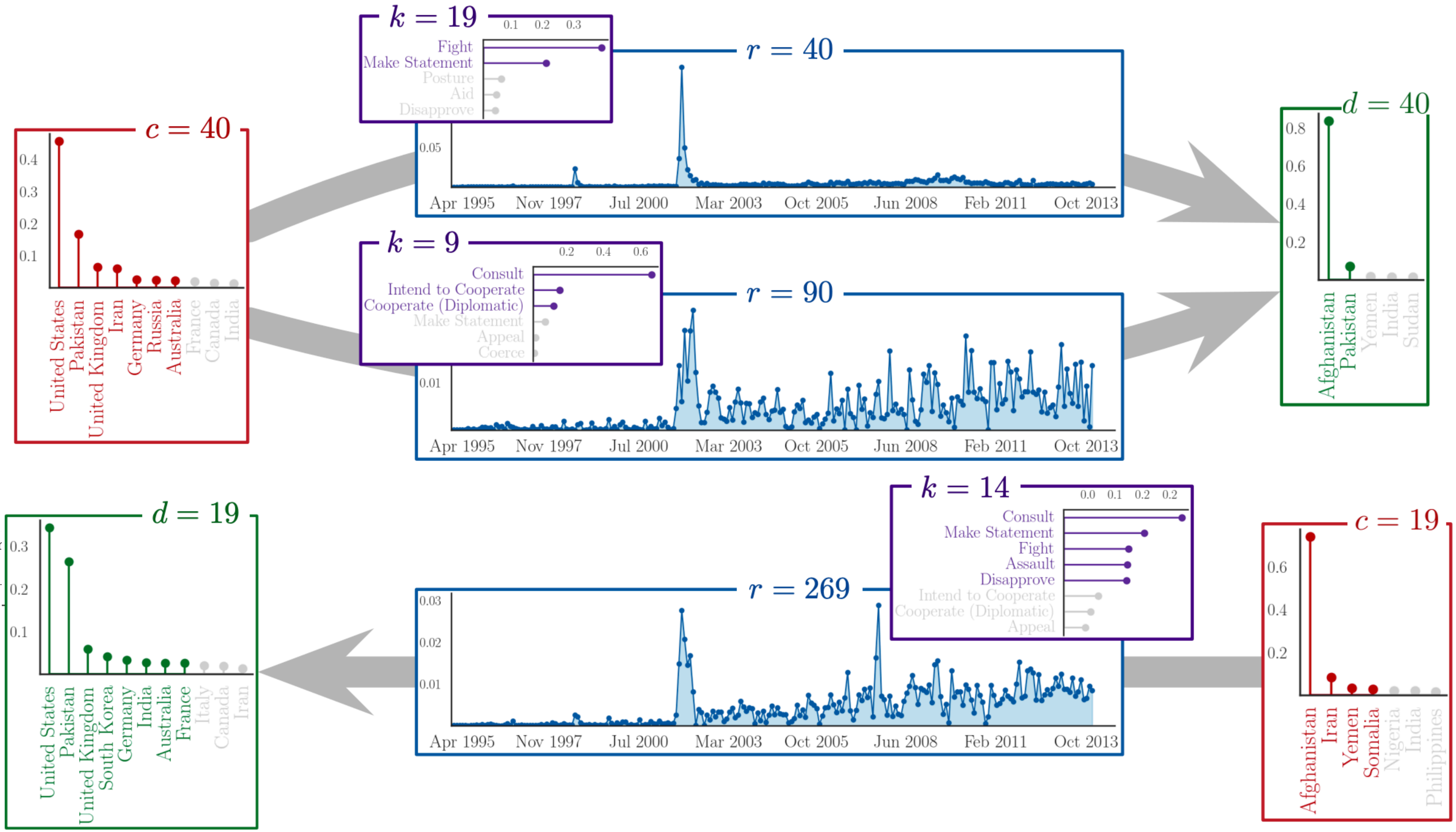
Example of inferred latent structure

Inferred nor

rank	$\lambda^{(r)}_{c \xrightarrow{k} d}$				
1	11.03				
2	10.36				
3	10.36				
4	10.13				
5	9.58				
6	9.36				
7	8.9				
8	8.57				
9	8.35				
10	8.13				
	…				
20	6.65				
21	6.61				
22	6.55				
23	6.35				
24	6.25				
25	**6.24**	**40**	**40**	**19**	**40**
26	**6.23**	**19**	**19**	**14**	**269**
27	6.14	43	43	10	143
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	…		…		

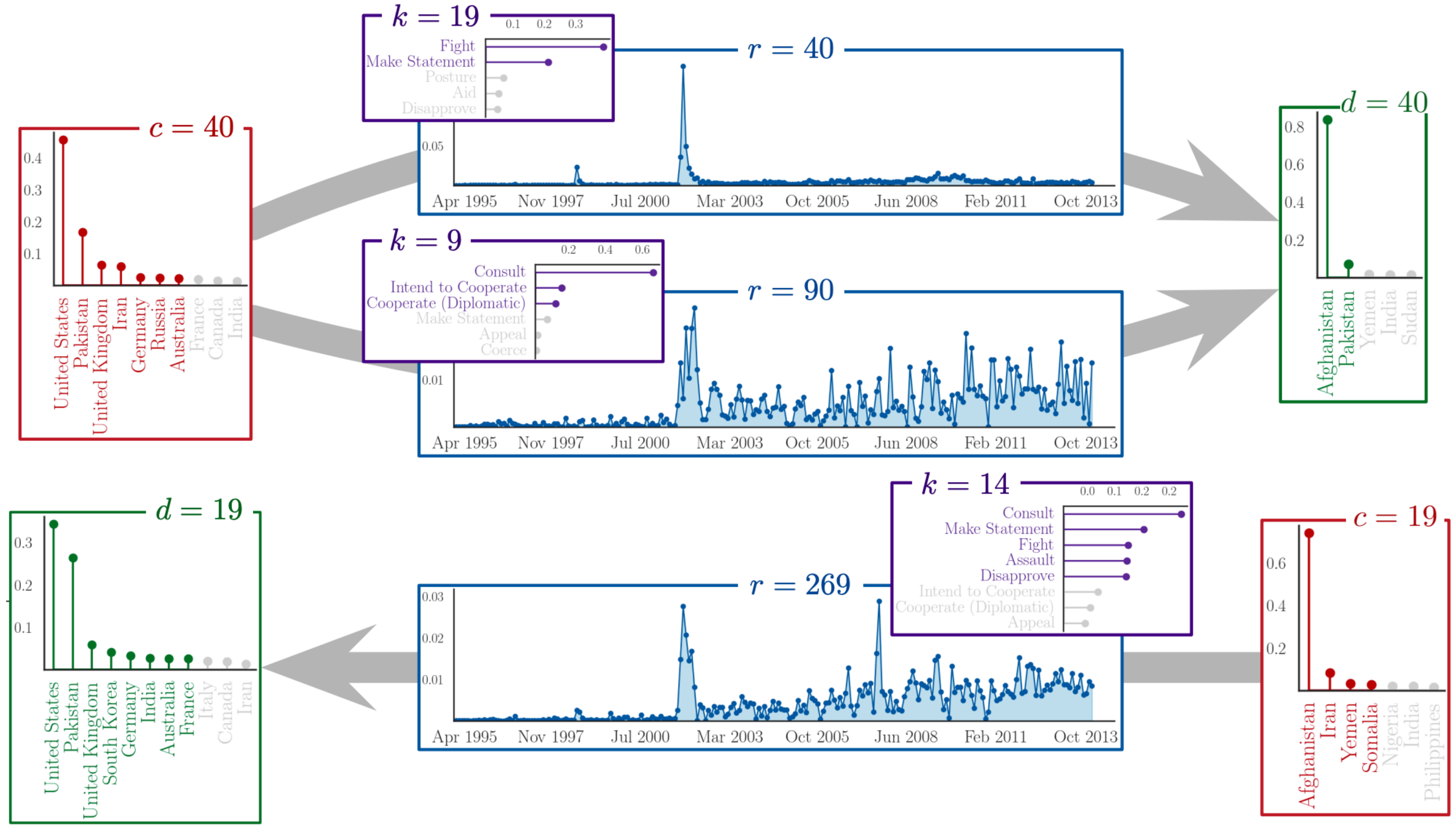


Example of inferred latent structure



Inferred nor

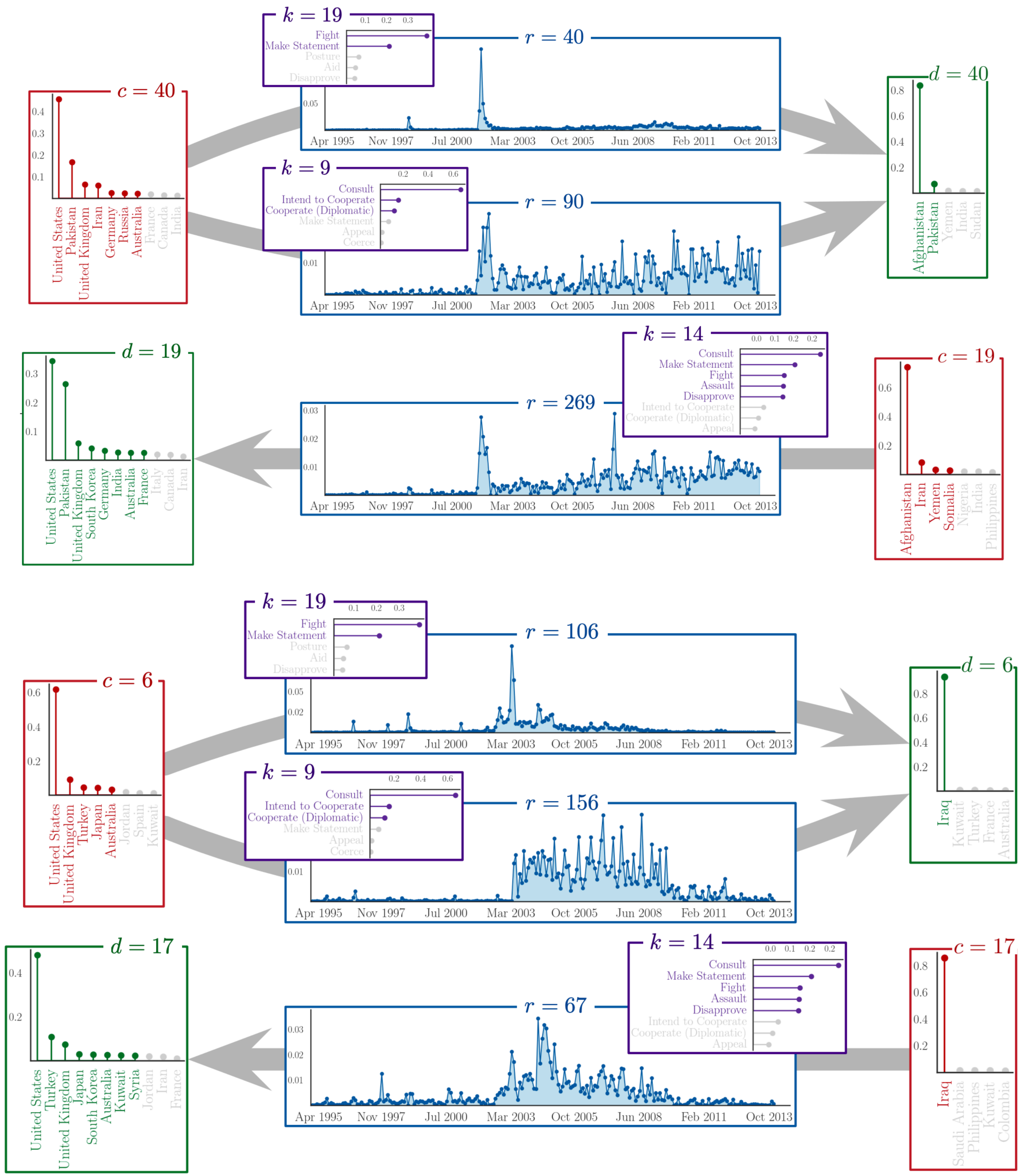
rank	$\lambda^{(r)}_{c \xrightarrow{k} d}$				
1	11.03				
2	10.36				
3	10.36				
4	10.13				
5	9.58				
6	9.36				
7	8.9				
8	8.57				
9	8.35				
10	8.13				
	...				
20	6.65				
21	6.61				
22	6.55				
23	6.35				
24	6.25				
25	**6.24**	**40**	**40**	**19**	**40**
26	**6.23**	**19**	**19**	**14**	**269**
27	6.14	43	43	10	143
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	...		...		



Inferred non-zero core elements

rank	$\lambda^{(r)}_{c \xrightarrow{k} d}$	c	d	k	r
1	11.03	2	2	10	252
2	10.36	31	31	9	81
3	10.36	9	9	12	159
4	10.13	33	33	9	233
5	9.58	37	37	12	87
6	9.36	12	12	2	162
7	8.9	43	43	10	43
8	8.57	41	41	2	41
9	8.35	15	15	10	115
10	8.13	11	11	16	61
	…		…		
20	6.65	49	49	9	49
21	6.61	14	14	2	14
22	6.55	23	23	9	273
23	6.35	46	46	9	196
24	6.25	22	22	2	222
25	**6.24**	**40**	**40**	**19**	**40**
26	**6.23**	**19**	**19**	**14**	**269**
27	6.14	43	43	10	143
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	…		…		

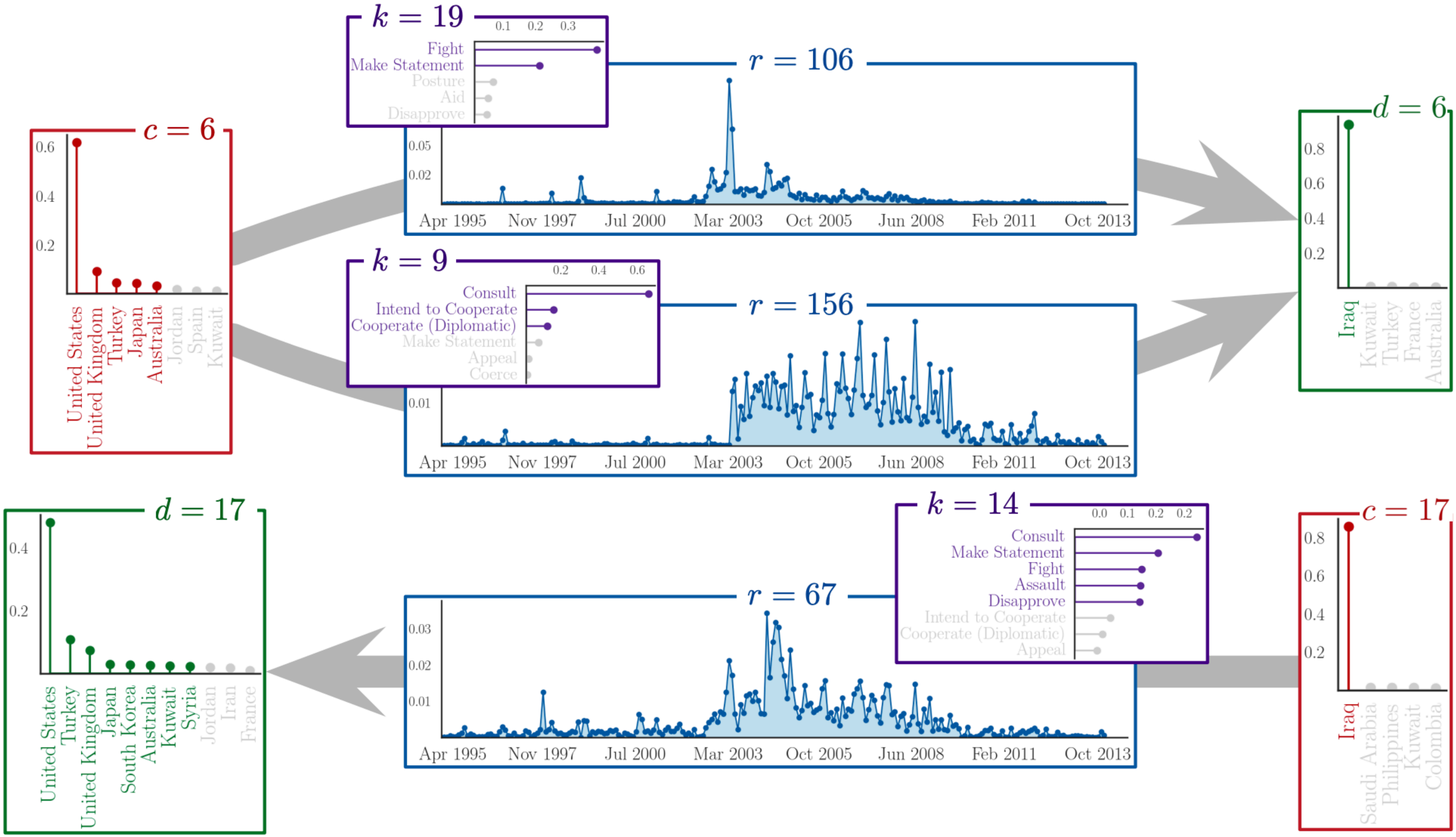
Example of inferred latent structure



Inferred non-zero core elements

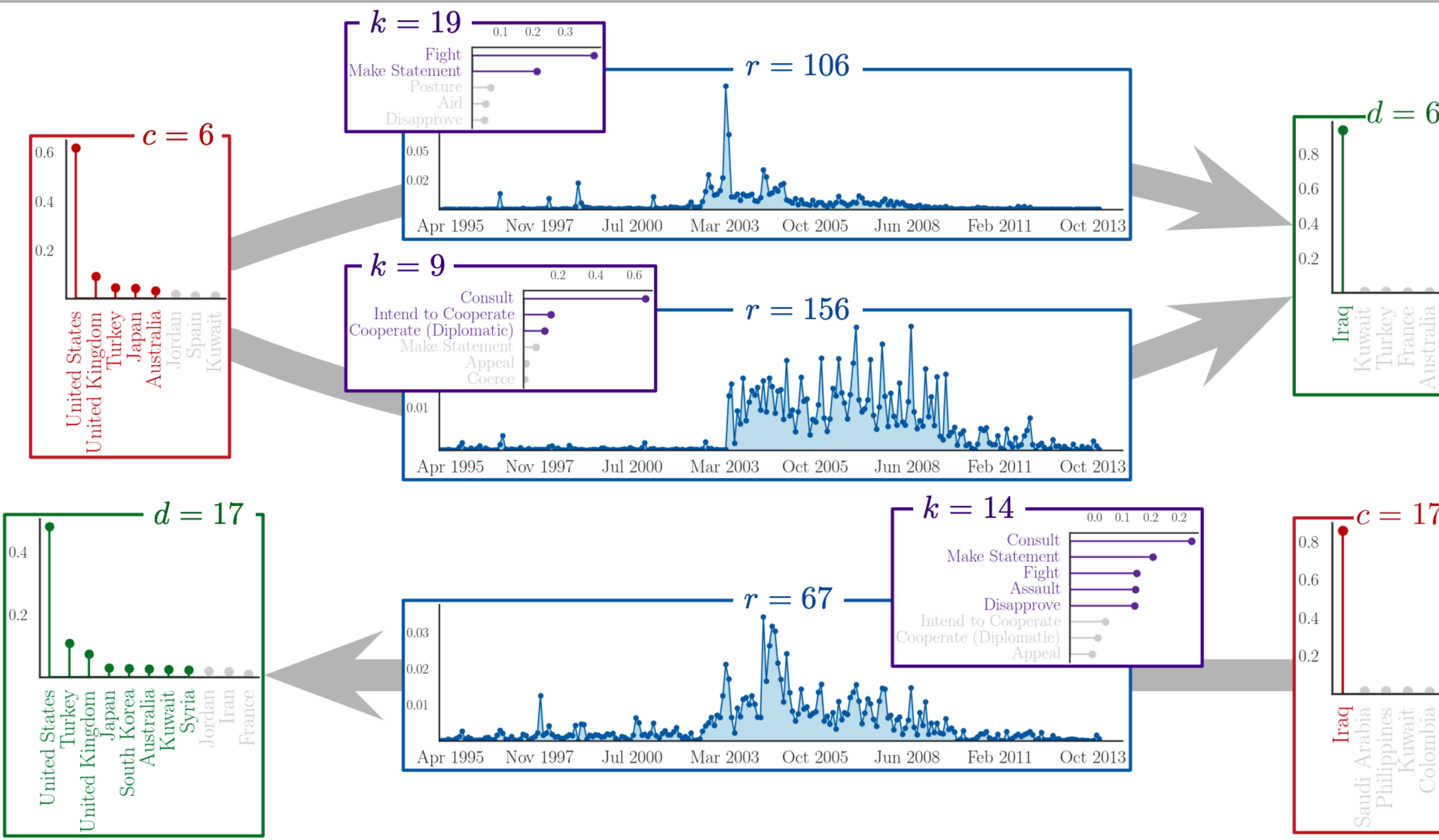
rank	$\lambda^{(r)}_{c \xrightarrow{k} d}$	c	d	k	r
1	11.03	2	2	10	252
2	10.36	31	31	9	81
3	10.36	9	9	12	159
4	10.13	33	33	9	233
5	9.58	37	37	12	87
6	9.36	12	12	2	162
7	8.9	43	43	10	43
8	8.57	41	41	2	41
9	8.35	15	15	10	115
10	8.13	11	11	16	61
	…		…		
20	6.65	49	49	9	49
21	6.61	14	14	2	14
22	6.55	23	23	9	273
23	6.35	46	46	9	196
24	6.25	22	22	2	222
25	**6.24**	**40**	**40**	**19**	**40**
26	**6.23**	**19**	**19**	**14**	**269**
27	6.14	43	43	10	143
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	…		…		

Example of inferred latent structure



Inferre rank					
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
20					
21					
22					
23					
24					
25					
26					
27					
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	...		...		

Example of inferred latent structure

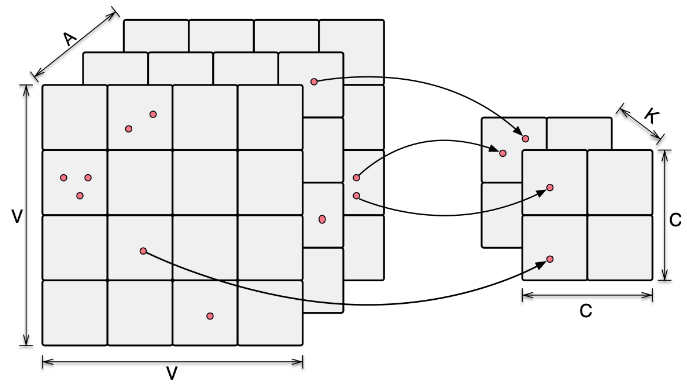


Inferre rank					
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
20					
21					
22					
23					
24					
25					
26					
27					
28	**6.14**	**6**	**6**	**19**	**106**
29	5.92	32	32	9	132
30	5.85	47	47	12	62
31	5.77	24	24	14	124
32	5.76	9	9	12	59
33	5.67	13	13	10	252
34	5.62	25	25	9	25
35	5.57	27	27	9	177
36	**5.53**	**17**	**17**	**14**	**67**
37	5.42	30	30	9	30
38	5.36	5	5	10	255
39	5.34	3	3	16	3
40	5.24	10	10	10	210
	…		…		

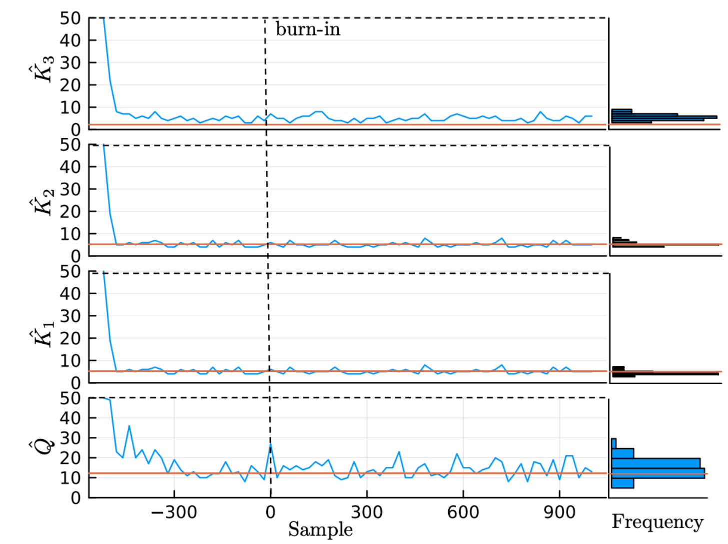
Same pattern found for **five** major wars in the observation period!

DISCUSSION / FUTURE

- **Sparse token allocation** as a principle (observed or latent)



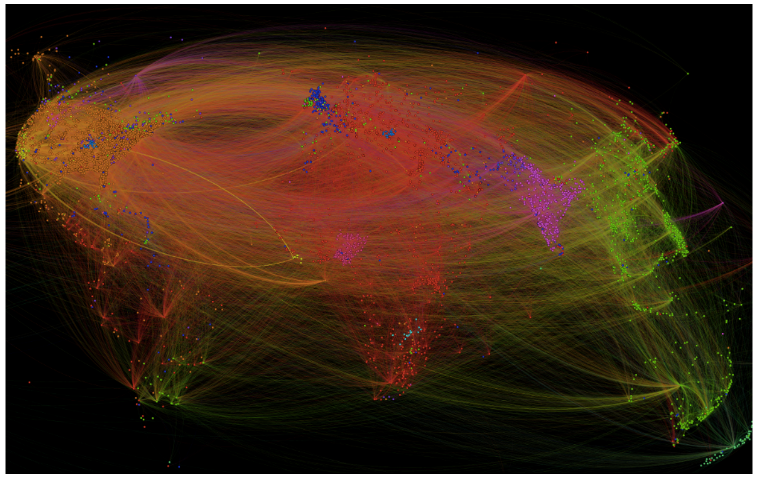
- Allocated **sparsity in factor matrices**; different motivation:
  - Interpretability (e.g., thresholding)
  - Identifiability (e.g., anchor features)



- Didn't talk about: **inferring** Q

- More on dyadic events:
  - Modeling everything by **publisher**
  - Modeling everything by **data set**

- One current extension / focus: **hypergraphs**
  - e.g., degree-5 networks → HUGE core tensors



BETTER INFERENCE

- **Dependent** prior for locations and values:

$$\mathbf{\Pi} \sim \text{Dirichlet}(1,\ldots,1)$$

$$\kappa_q \sim \text{Categorical}(\mathbf{\Pi}) \quad \text{for} \quad q = 1\ldots Q$$

$$\lambda_{\boldsymbol{\kappa}} \sim \begin{cases} \text{Gamma}(m_\kappa 1) & \text{if } m_\kappa > 0 \\ \delta_0 & \text{otherwise} \end{cases}$$

BETTER INFERENCE

- **Dependent** prior for locations and values:

$$m_{\boldsymbol{\kappa}} \triangleq \sum_{q=1}^{Q} \mathbf{1}(\boldsymbol{\kappa}_q = \boldsymbol{\kappa})$$

$$\boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1) \longrightarrow \boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1)$$

$$\kappa_q \sim \text{Categorical}(\boldsymbol{\Pi}) \;\; \text{for} \;\; q = 1\ldots Q \longrightarrow \mathbf{M} \sim \text{Multinomial}(Q, \boldsymbol{\Pi})$$

$$\lambda_{\boldsymbol{\kappa}} \sim \begin{cases} \text{Gamma}(m_\kappa 1) & \text{if } m_\kappa > 0 \\ \delta_0 & \text{otherwise} \end{cases}$$

BETTER INFERENCE

- **Dependent** prior for locations and values:

$$m_\kappa \triangleq \sum_{q=1}^{Q} \mathbf{1}(\boldsymbol{\kappa}_q = \boldsymbol{\kappa})$$

$$\lambda_{\boldsymbol{\kappa}} \triangleq \tilde{\lambda}_{\boldsymbol{\kappa}} \lambda_{\bullet}$$

$$\boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1) \longrightarrow \boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1)$$

$$\kappa_q \sim \text{Categorical}(\boldsymbol{\Pi}) \;\; \text{for} \;\; q = 1\ldots Q \longrightarrow \mathbf{M} \sim \text{Multinomial}(Q, \boldsymbol{\Pi})$$

Sum-proportion independence of gammas [Lukacs 1956]

"Delta Dirichlet" distribution [Lewy 1996]

$$\lambda_{\boldsymbol{\kappa}} \sim \begin{cases} \text{Gamma}(m_\kappa 1) & \text{if } m_\kappa > 0 \\ \delta_0 & \text{otherwise} \end{cases} \longrightarrow \begin{array}{l} \tilde{\boldsymbol{\Lambda}} \sim \text{DeltaDirichlet}(\mathbf{M}) \\ \lambda_\bullet \sim \text{Gamma}(Q,1) \end{array}$$

BETTER INFERENCE

- **Dependent** prior for locations and values:

$$m_\kappa \triangleq \sum_{q=1}^{Q} \mathbf{1}(\boldsymbol{\kappa}_q = \boldsymbol{\kappa})$$

$$\lambda_{\boldsymbol{\kappa}} \triangleq \tilde{\lambda}_{\boldsymbol{\kappa}} \lambda_{\bullet}$$

$$\boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1) \longrightarrow \boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1)$$

$$\kappa_q \sim \text{Categorical}(\boldsymbol{\Pi}) \;\; \text{for} \;\; q = 1\ldots Q \longrightarrow \mathbf{M} \sim \text{Multinomial}(Q, \boldsymbol{\Pi})$$

Sum-proportion independence of gammas [Lukacs 1956]

"Delta Dirichlet" distribution [Lewy 1996]

$$\lambda_{\boldsymbol{\kappa}} \sim \begin{cases} \text{Gamma}(m_\kappa\, 1) & \text{if } m_\kappa > 0 \\ \delta_0 & \text{otherwise} \end{cases} \longrightarrow \begin{array}{l} \tilde{\boldsymbol{\Lambda}} \sim \text{DeltaDirichlet}(\mathbf{M}) \\ \lambda_{\bullet} \sim \text{Gamma}(Q,1) \end{array}$$

$$y_{\mathbf{i},\boldsymbol{\kappa}} \sim \text{Pois}\left(\lambda_{\boldsymbol{\kappa}} \cdots\right)$$

- **Dependent** prior for locations and values:

$$m_{\kappa} \triangleq \sum_{q=1}^{Q} \mathbf{1}(\boldsymbol{\kappa}_q = \boldsymbol{\kappa})$$

$$\lambda_{\boldsymbol{\kappa}} \triangleq \tilde{\lambda}_{\boldsymbol{\kappa}} \lambda_{\bullet}$$

$$\boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1) \longrightarrow \boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1)$$

$$\kappa_q \sim \text{Categorical}(\boldsymbol{\Pi}) \text{ for } q = 1\ldots Q \longrightarrow \mathbf{M} \sim \text{Multinomial}(Q, \boldsymbol{\Pi})$$

Sum-proportion independence of gammas [Lukacs 1956]

"Delta Dirichlet" distribution [Lewy 1996]

$$\lambda_{\boldsymbol{\kappa}} \sim \begin{cases} \text{Gamma}(m_{\kappa}, 1) & \text{if } m_{\kappa} > 0 \\ \delta_0 & \text{otherwise} \end{cases} \longrightarrow \begin{array}{l} \tilde{\boldsymbol{\Lambda}} \sim \text{DeltaDirichlet}(\mathbf{M}) \\ \lambda_{\bullet} \sim \text{Gamma}(Q, 1) \end{array}$$

$$y_{\mathbf{i},\boldsymbol{\kappa}} \sim \text{Pois}\left(\lambda_{\boldsymbol{\kappa}} \cdots\right) \xrightarrow{\text{“Latent source” representation}} \begin{array}{l} y_{\bullet} \sim \text{Pois}\left(\lambda_{\bullet} \cdots\right) \\ \left(y_{\boldsymbol{\kappa}}\right)_{\boldsymbol{\kappa}} \sim \text{Multinomial}\left(y_{\bullet}, \tilde{\boldsymbol{\Lambda}}\right) \end{array}$$

BETTER INFERENCE

- **Dependent** prior for locations and values:

$$m_\kappa \triangleq \sum_{q=1}^{Q} \mathbf{1}(\boldsymbol{\kappa}_q = \boldsymbol{\kappa})$$

$$\lambda_{\boldsymbol{\kappa}} \triangleq \tilde{\lambda}_{\boldsymbol{\kappa}}\, \lambda_{\bullet}$$

$$\boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1) \longrightarrow \boldsymbol{\Pi} \sim \text{Dirichlet}(1,\ldots,1)$$

$$\kappa_q \sim \text{Categorical}(\boldsymbol{\Pi}) \text{ for } q = 1\ldots Q \longrightarrow \mathbf{M} \sim \text{Multinomial}(Q, \boldsymbol{\Pi})$$

Sum-proportion independence of gammas [Lukacs 1956]

"Delta Dirichlet" distribution [Lewy 1996]

$$\lambda_{\boldsymbol{\kappa}} \sim \begin{cases} \text{Gamma}(m_\kappa, 1) & \text{if } m_\kappa > 0 \\ \delta_0 & \text{otherwise} \end{cases} \longrightarrow \begin{array}{l} \tilde{\boldsymbol{\Lambda}} \sim \text{DeltaDirichlet}(\mathbf{M}) \\ \lambda_\bullet \sim \text{Gamma}(Q,1) \end{array}$$

"Latent source" representation

$$y_{\mathbf{i},\boldsymbol{\kappa}} \sim \text{Pois}\left(\lambda_{\boldsymbol{\kappa}} \cdots\right) \longrightarrow \begin{array}{l} y_\bullet \sim \text{Pois}\left(\lambda_\bullet \cdots\right) \\ \left(y_{\boldsymbol{\kappa}}\right)_{\boldsymbol{\kappa}} \sim \text{Multinomial}\left(y_\bullet, \tilde{\boldsymbol{\Lambda}}\right) \end{array}$$

$$\boxed{\begin{array}{ll} \mathbf{M} \sim \text{DirMult}(Q,1,\ldots,1) & \lambda_\bullet \sim \text{Gamma}(Q,1) \\ \left(y_{\boldsymbol{\kappa}}\right)_{\boldsymbol{\kappa}} \sim \text{DeltaDirMult}\left(y_\bullet, \mathbf{M}\right) & y_\bullet \sim \text{Pois}\left(\lambda_\bullet \cdots\right) \end{array}}$$

BETTER INFERENCE

- Update to $\lambda_\bullet$ is conjugate
- New(?) form of "quasi"-conjugacy:

$$(\mathbf{M} \mid -) = \mathbf{1}(\mathbf{Y} > 0) + \mathbf{M}'$$

$$\mathbf{M}' \sim \text{DirMult}(Q - ||\mathbf{Y}||_0, 1 + \mathbf{Y})$$

- We can sample $\mathbf{M}'$ exactly using an urn scheme that is $\mathcal{O}(Q)$

$$\mathbf{M} \sim \text{DirMult}(Q, 1, \ldots, 1) \qquad \lambda_\bullet \sim \text{Gamma}(Q, 1)$$

$$\left(y_\kappa\right)_\kappa \sim \text{DeltaDirMult}\left(y_\bullet, \mathbf{M}\right) \qquad y_\bullet \sim \text{Pois}\left(\lambda_\bullet \cdots\right)$$

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

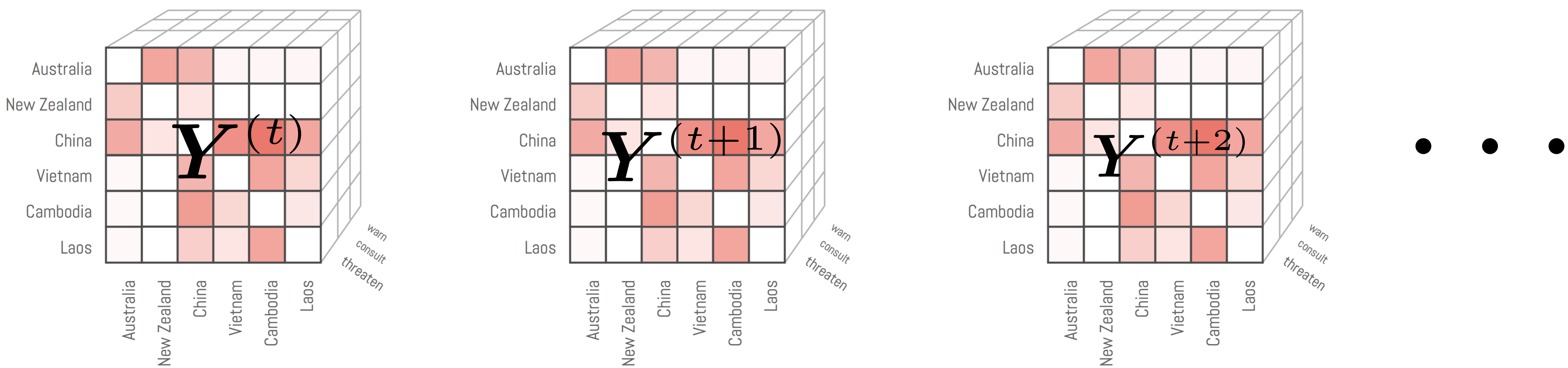
- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts “multilateral” structure [Schein et al., 2015]
  - **Tucker** extracts “communities” [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the “exponential blowup” [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling “escalation”** with a new matrix prior [Stoehr et al., 2023]

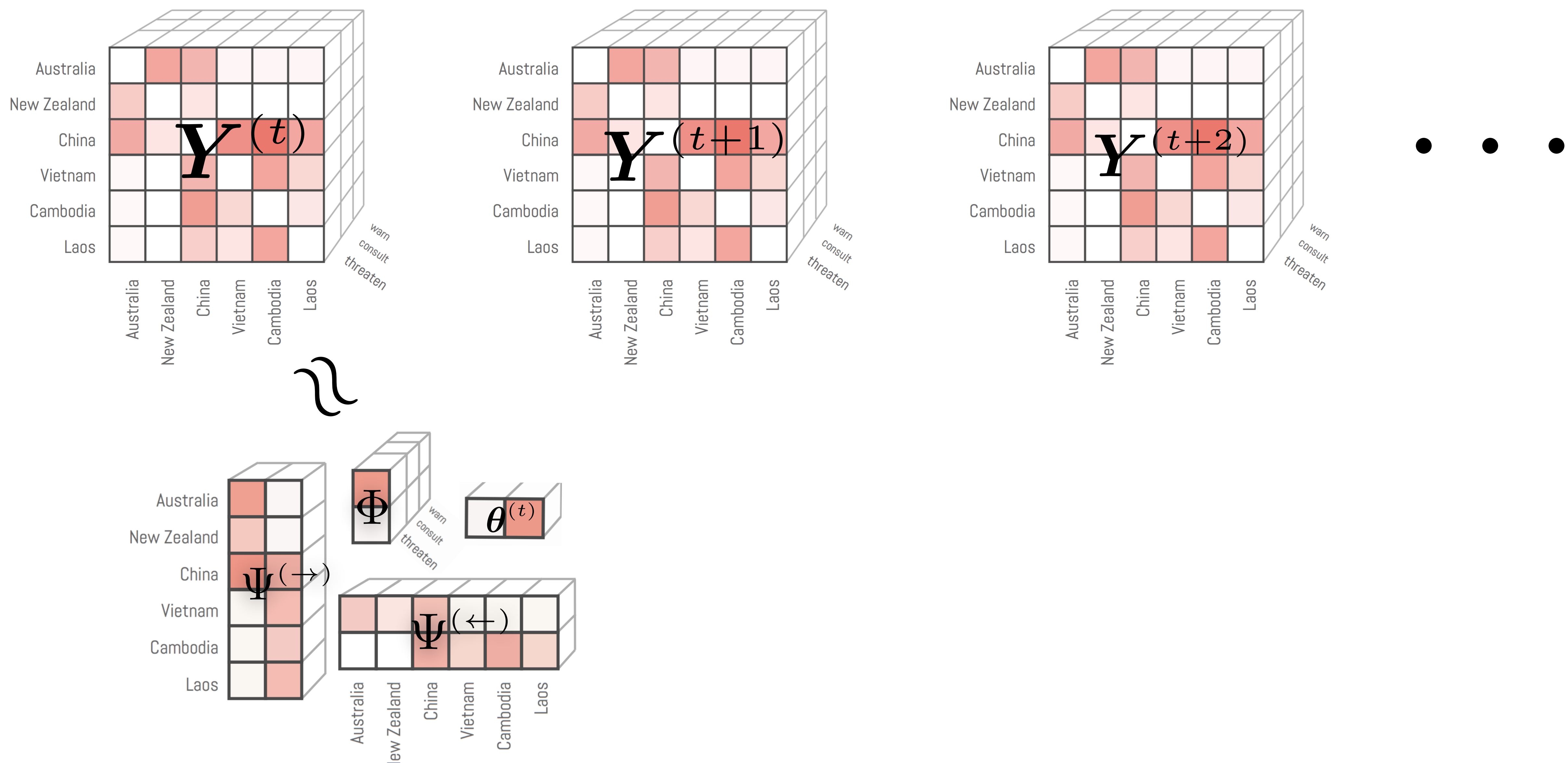
Building in **temporal structure**

dynamic tensor

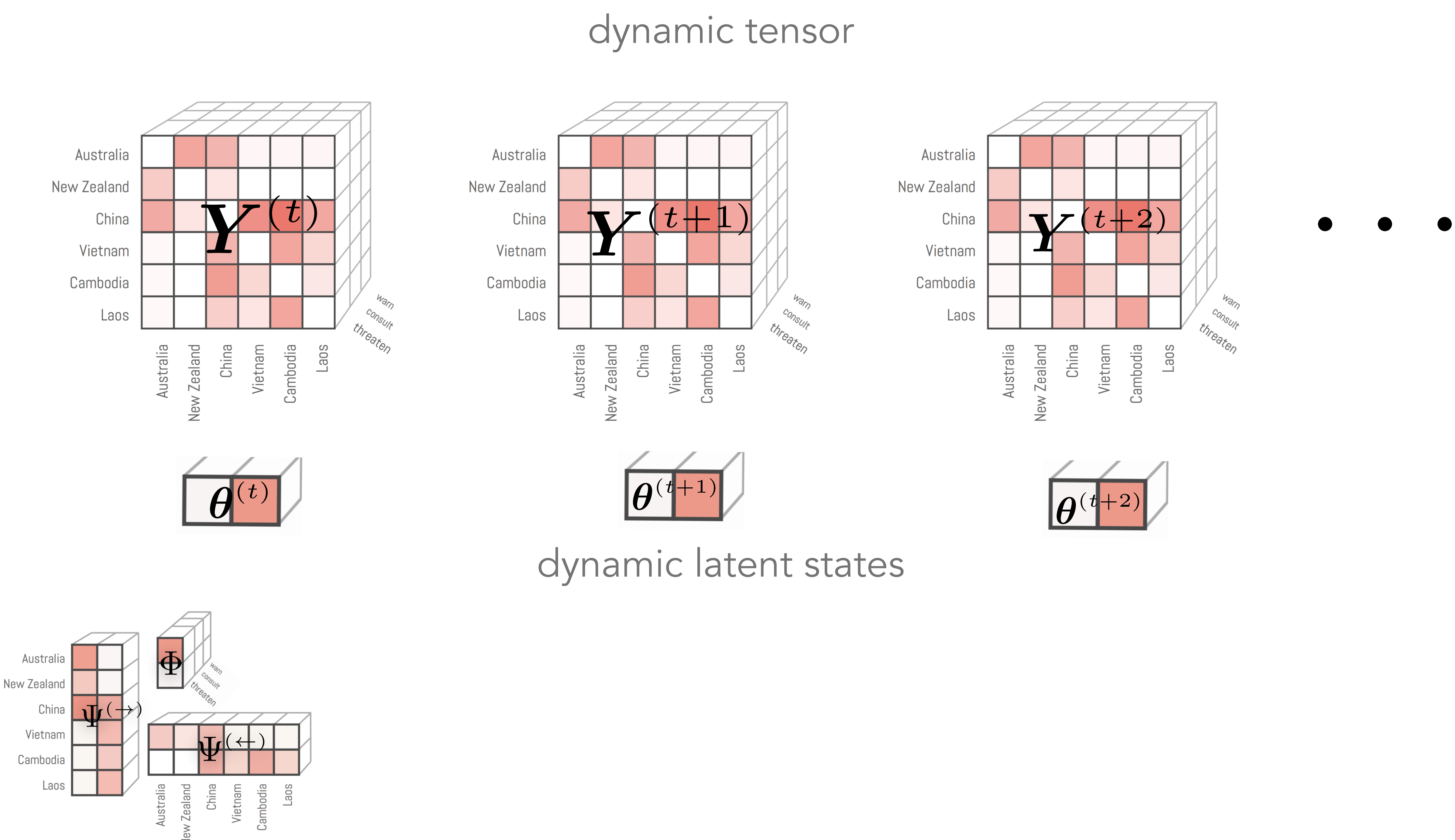


Building in **temporal structure**

dynamic tensor

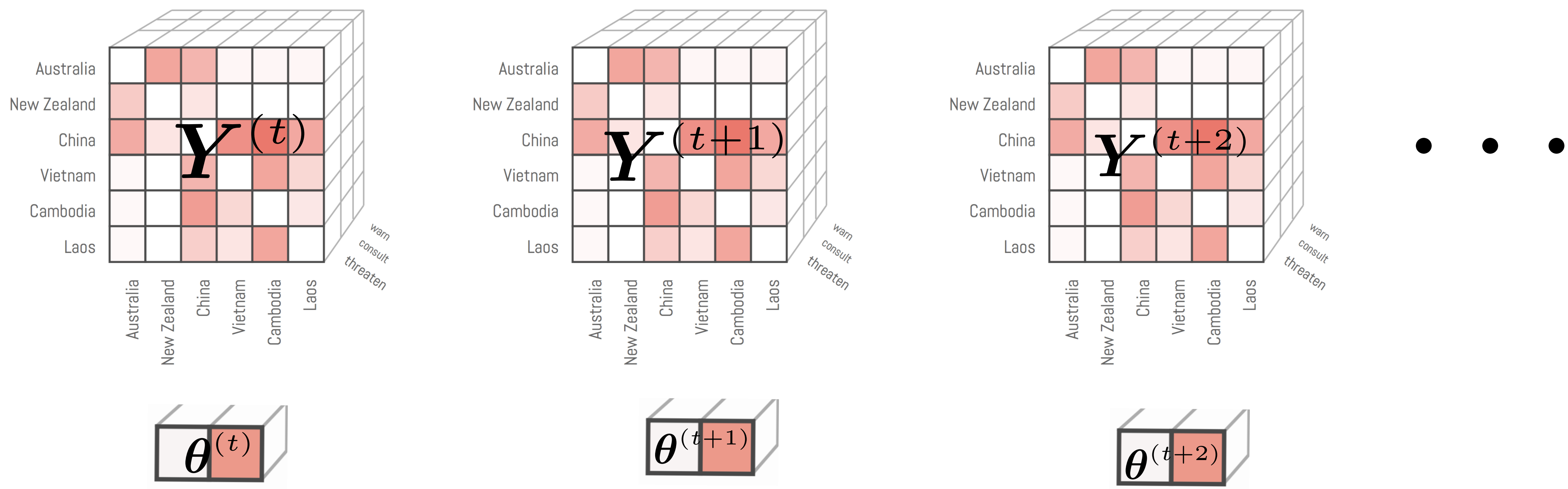


Building in **temporal structure**

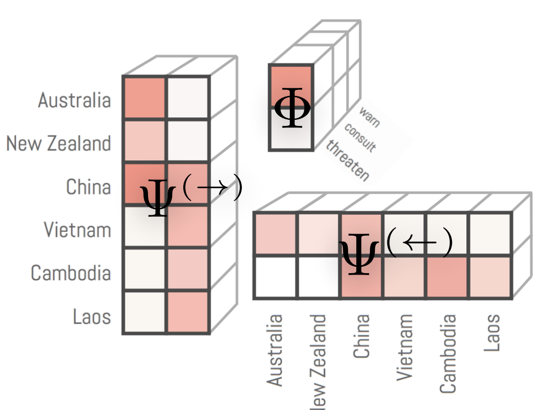


Building in **temporal structure**

dynamic tensor



dynamic latent states



Goal: Build in temporal structure.

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}^{(t-1)})$$

Building in **temporal structure**

Goal: Build in temporal structure.

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \,|\, \boldsymbol{\theta}^{(t-1)})$$

Challenge: Stay within *allocative Poisson factorization*

$$y_{\mathbf{i}} \sim \text{Pois}\Big(\sum \cdots\Big)$$

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

Building in **temporal structure**

Goal: Build in temporal structure.

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \,|\, \boldsymbol{\theta}^{(t-1)})$$

Challenge: Stay within *allocative Poisson factorization*

$$y_{\mathbf{i}} \sim \text{Pois}\Big(\sum \cdots\Big)$$

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

…but!

- only **non-negative priors**!
- (no Gaussians.)

Building in **temporal structure**

Goal: Build in temporal structure.

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \,|\, \boldsymbol{\theta}^{(t-1)})$$

Challenge: Stay within *allocative Poisson factorization*

$$y_{\mathbf{i}} \sim \text{Pois}\Big(\sum \cdots\Big)$$

Poisson LDS [Macke et al., 2011]

$$y_{\mathbf{i}} \sim \text{Pois}\Big(\exp\big(\sum \cdots\big)\Big)$$

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

…but!

- only **non-negative priors**!
- (no Gaussians.)

Gaussian priors**:**

toolkit *of approximate Bayesian inference*

Building in **temporal structure**

Goal: Build in temporal structure.

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \,|\, \boldsymbol{\theta}^{(t-1)})$$

Challenge: Stay within *allocative Poisson factorization*

$$y_{\mathbf{i}} \sim \text{Pois}\Big(\sum \cdots\Big)$$

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

…but!

- only **non-negative priors**!
- (no Gaussians.)

Non-negative priors**:**

***toolkit** of approximate Bayesian inference*

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \mathrm{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \mathrm{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \text{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

how active component *k'*
is at time step *t-1*

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \mathrm{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

the **probability** of transitioning
from component k' into k

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \text{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

$$\boxed{\sum_{k=1}^{K} \pi_{kk'} = 1}$$

the **probability** of transitioning from component *k′* into *k*

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \text{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\theta_k^{(t)} \sim \mathrm{Gam}\left(\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)},\ \tau_0\right)$$

concentration
hyperparameter

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\mathbb{E}\left[\theta_k^{(t)}\right] = \frac{\tau_0 \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)}}{\tau_0}$$

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\mathbb{E}\left[\theta_k^{(t)}\right] = \frac{\cancel{\tau_0} \sum_{k'=1}^{K} \pi_{kk'} \theta_{k'}^{(t-1)}}{\cancel{\tau_0}}$$

Poisson—gamma dynamical systems

[NeurIPS '16]

$$\mathbb{E}\left[\theta_k^{(t)}\right] = \sum_{k'=1}^{K} \pi_{kk'}\theta_{k'}^{(t-1)}$$

Poisson—gamma dynamical systems

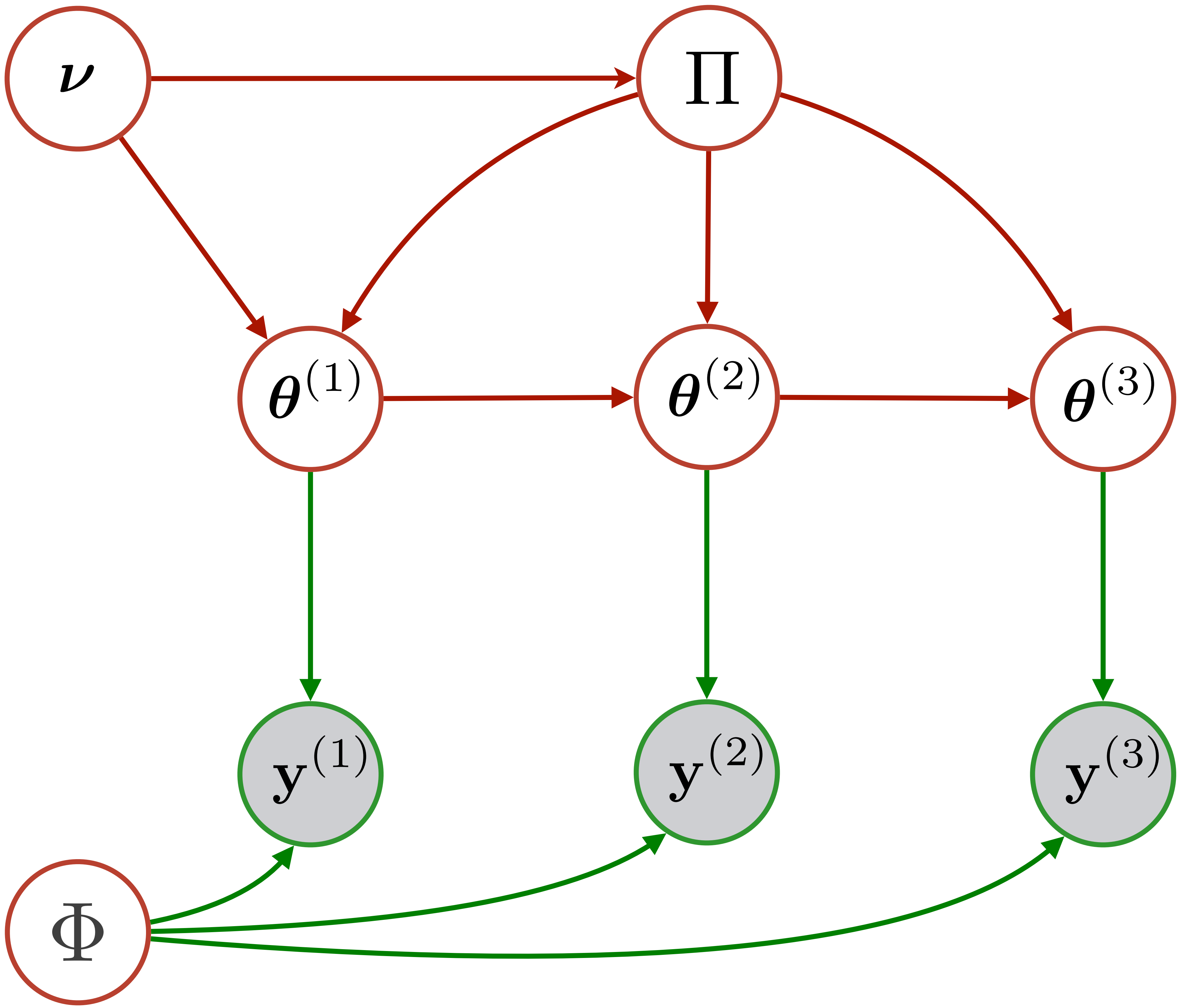
[NeurIPS '16]

$$\mathbb{E}\left[\boldsymbol{\theta}^{(t)}\right] = \Pi \boldsymbol{\theta}^{(t-1)}$$

$$\mathbb{E}\left[\boldsymbol{y}^{(t)}\right] = \Phi \boldsymbol{\theta}^{(t)}$$

(matches linear dynamical systems)

Technical challenge

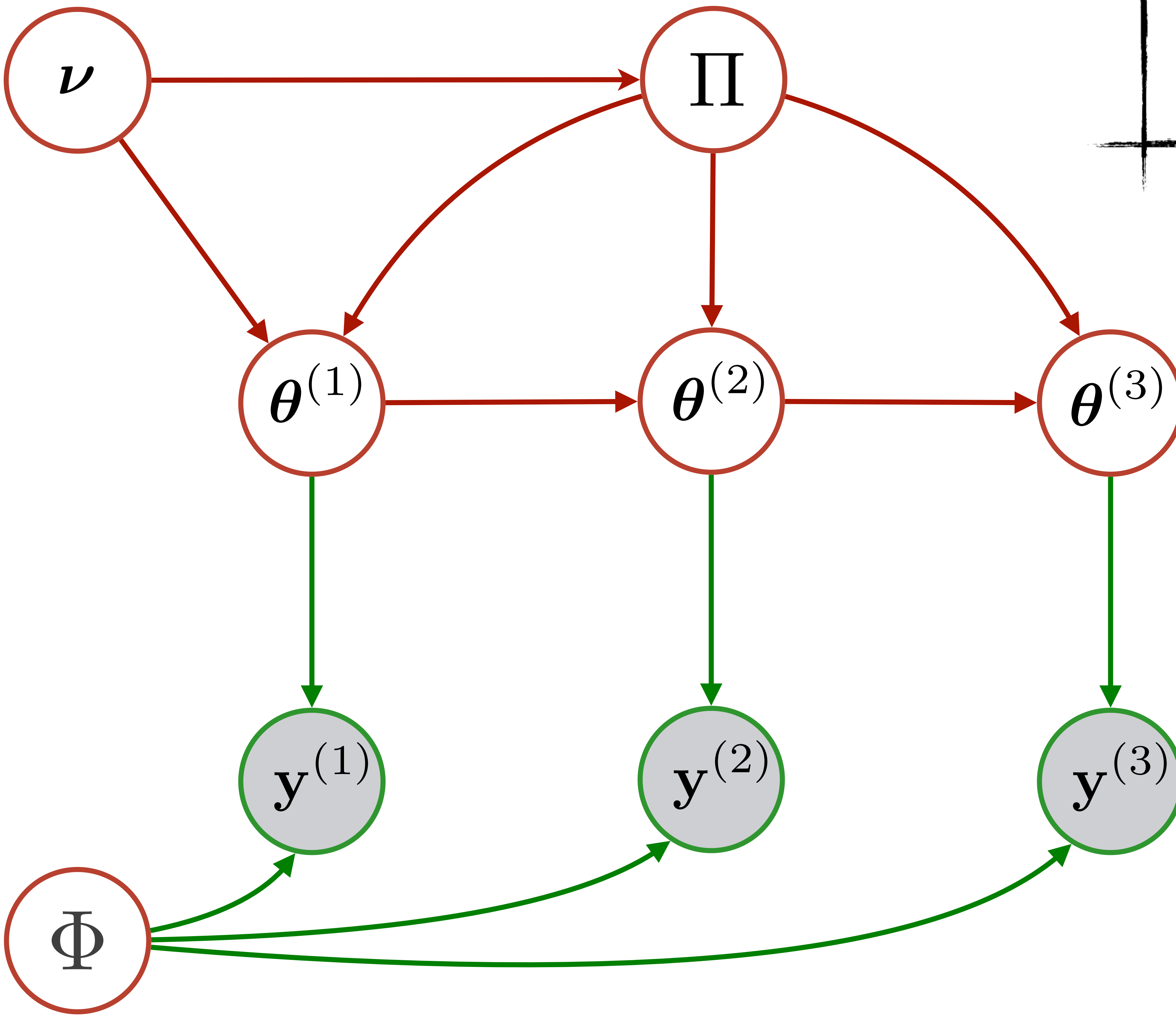


Technical challenge

Legend

Poisson/Multinomial

Gamma/Dirichlet

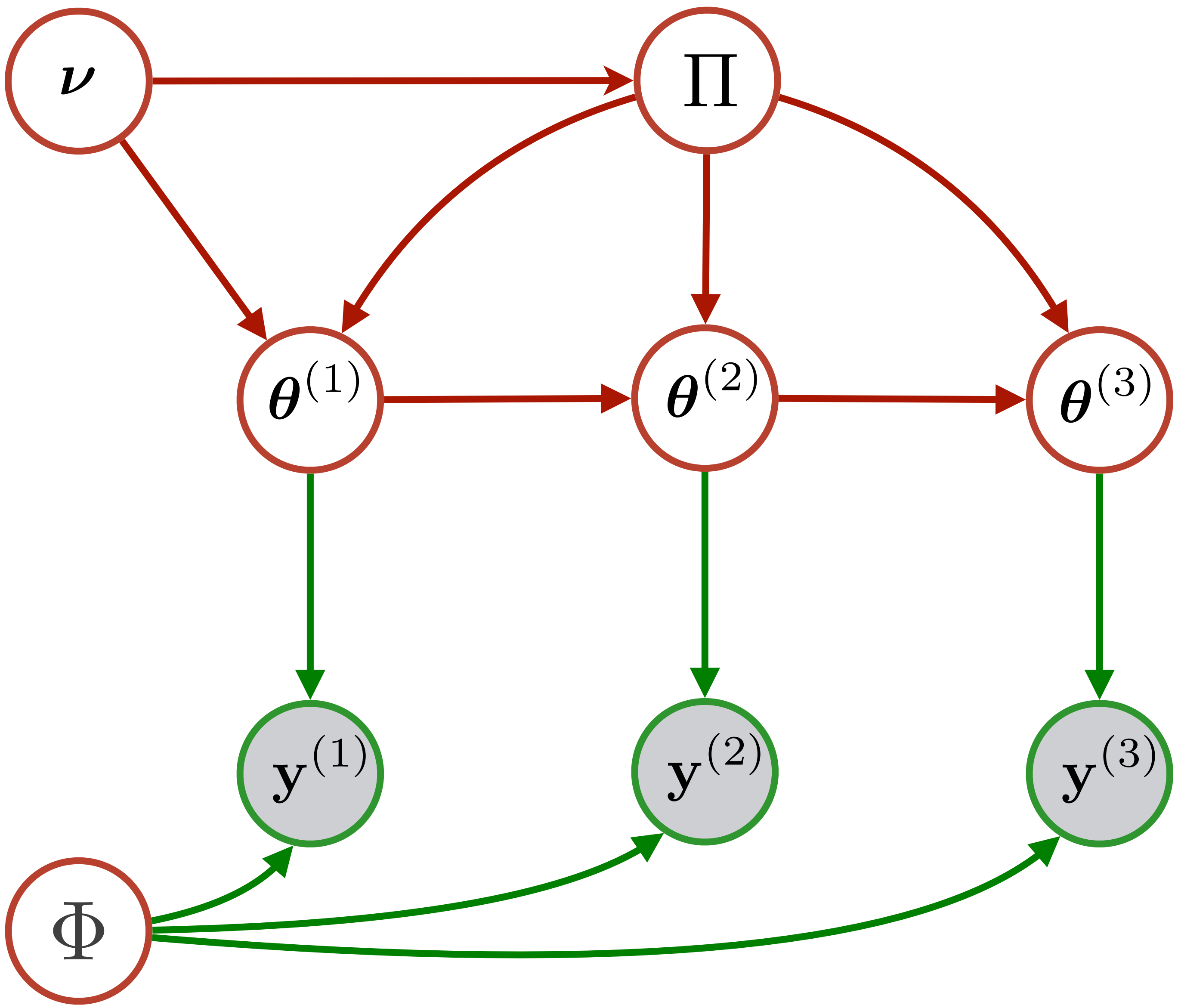


Technical challenge

Legend

Poisson/Multinomial

Gamma/Dirichlet

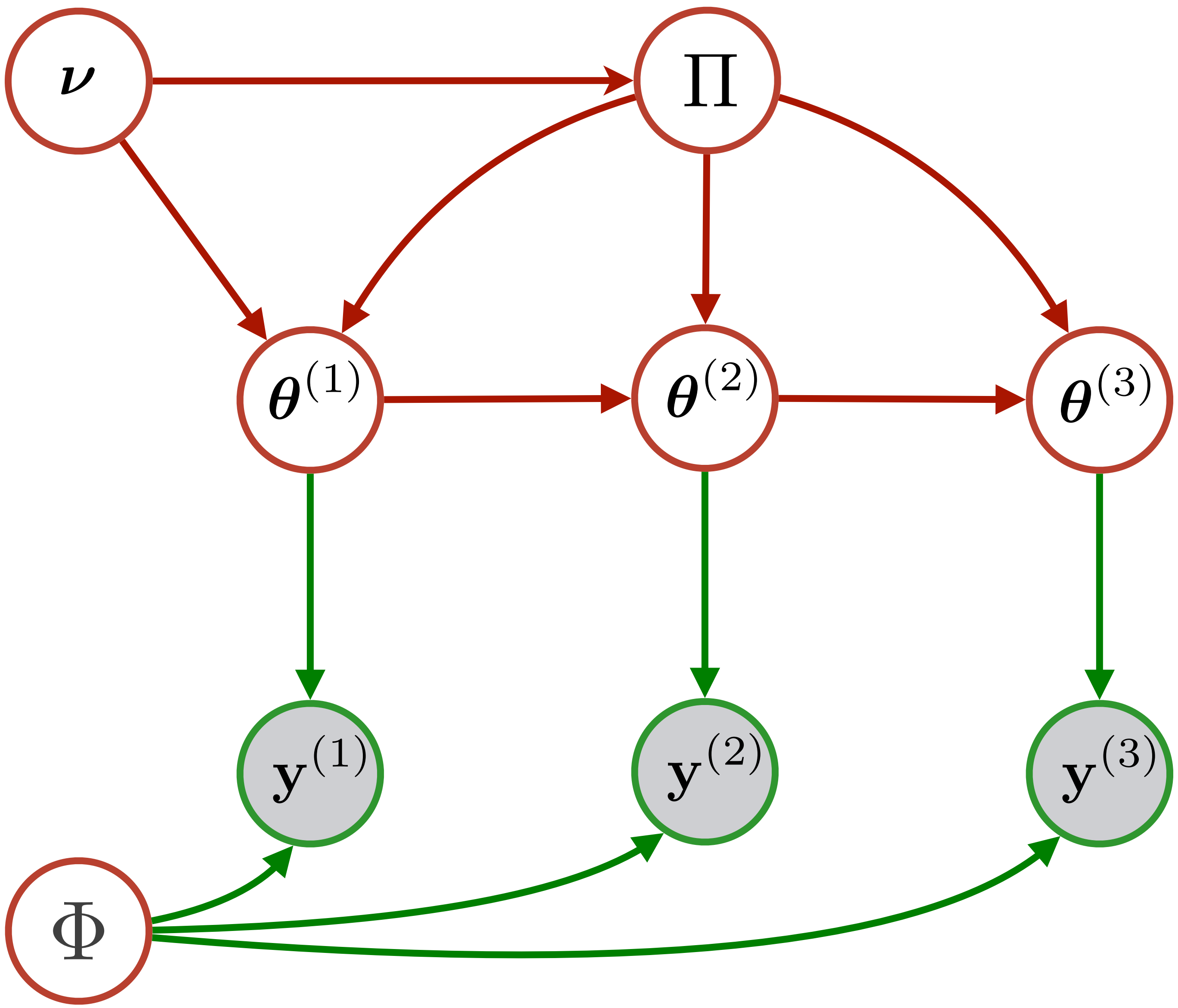


Technical challenge

Legend

Poisson/Multinomial

Gamma/Dirichlet



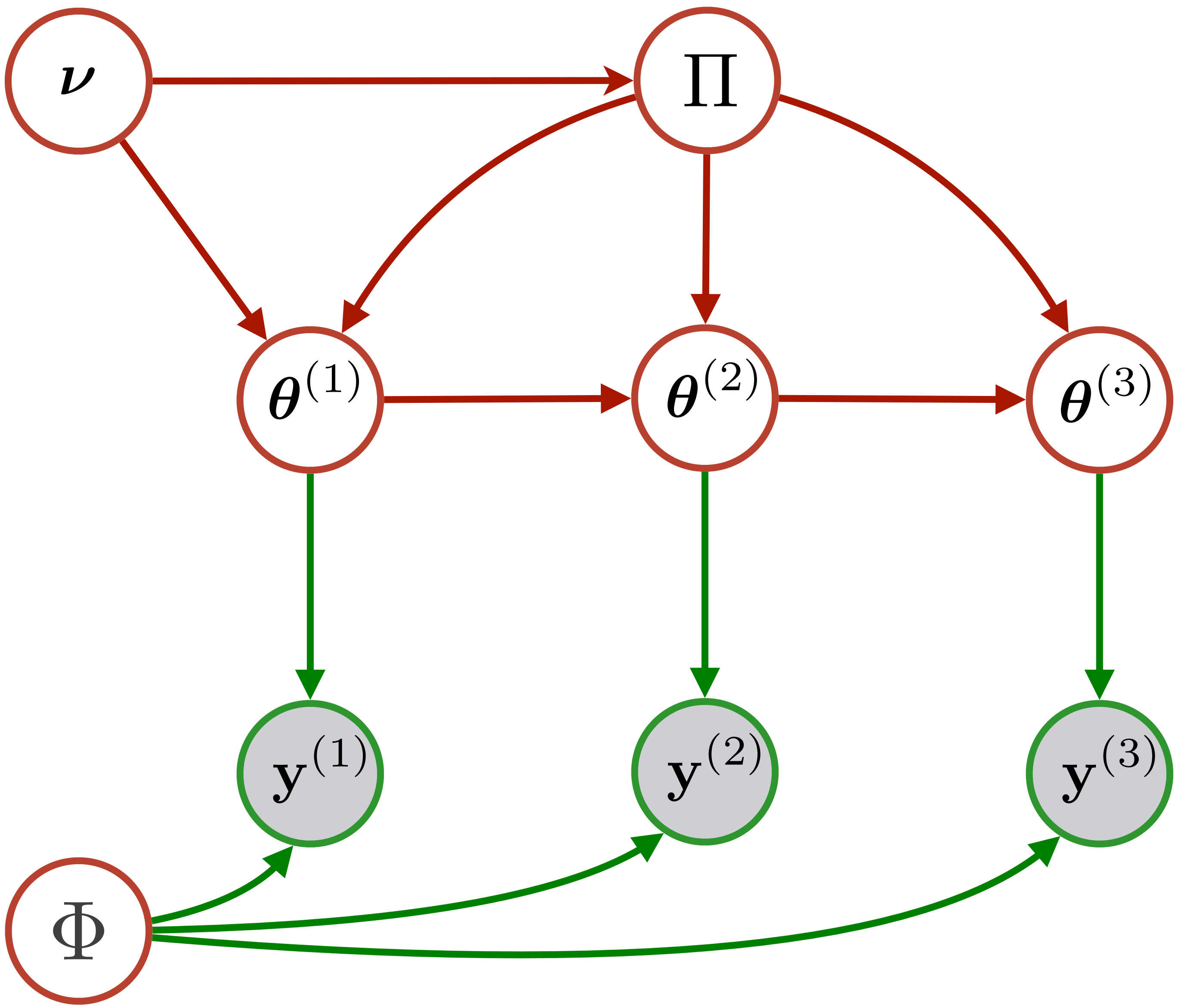
$$P(\Phi \mid Y)$$

Technical challenge

Legend

Poisson/Multinomial

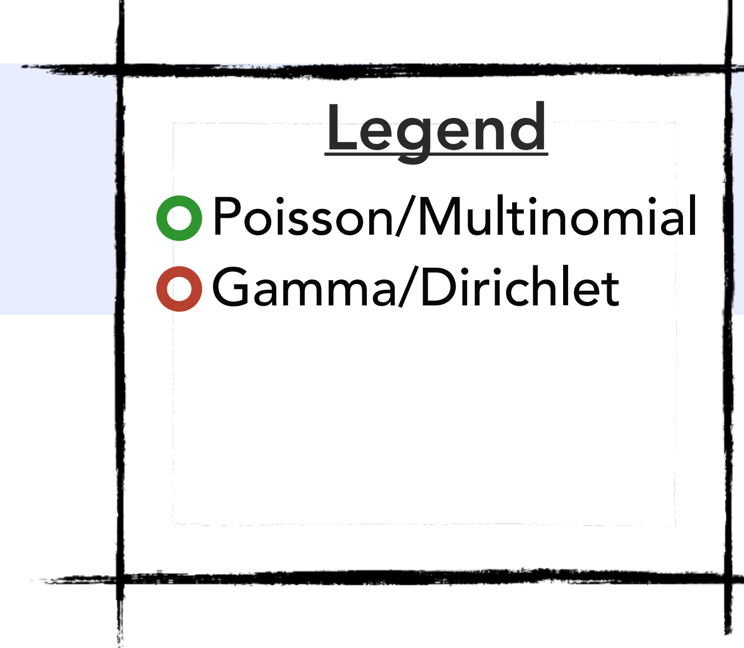
Gamma/Dirichlet

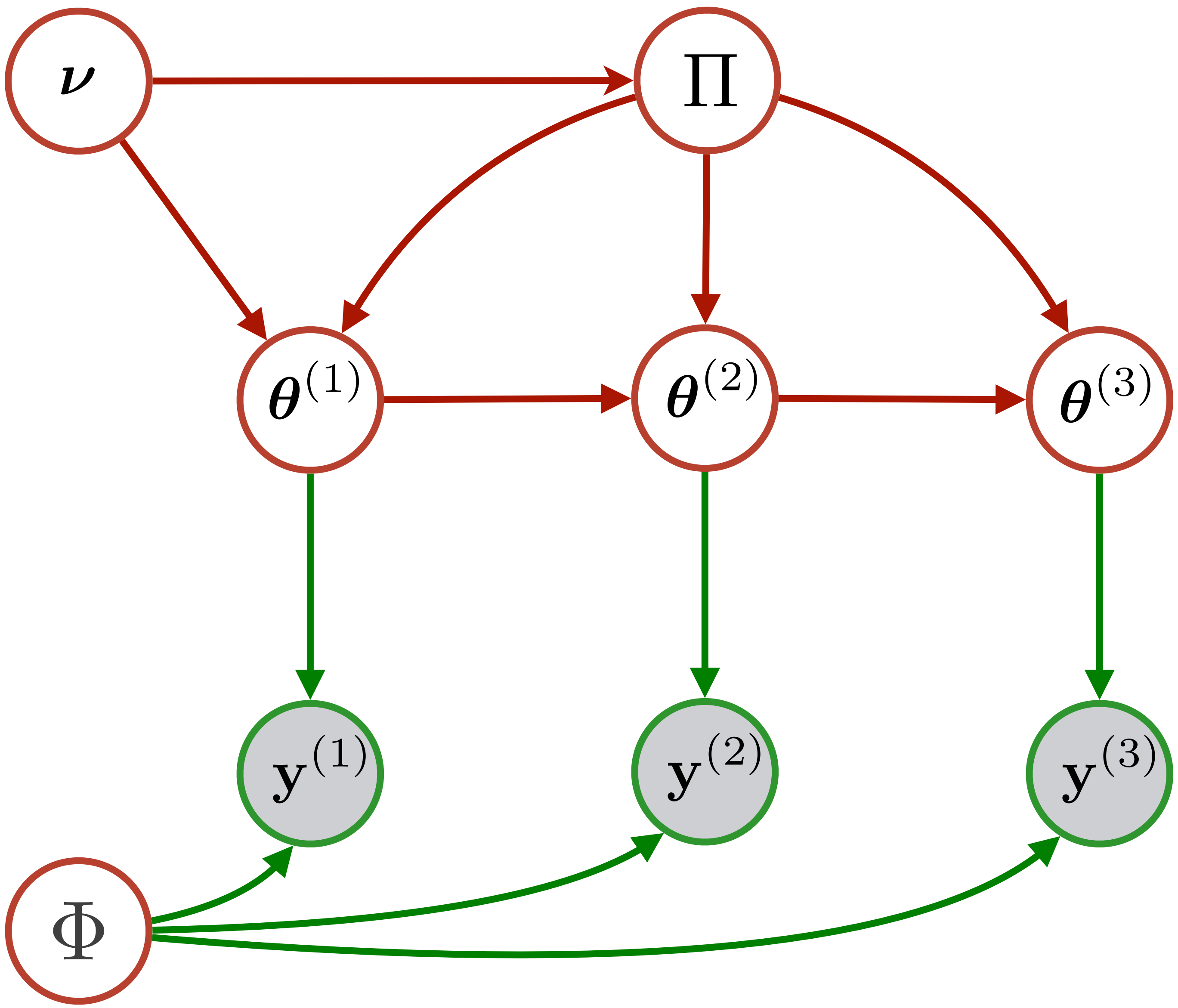


$$P(\Phi \mid Y) \checkmark$$

(conditional posterior has closed form)

Technical challenge

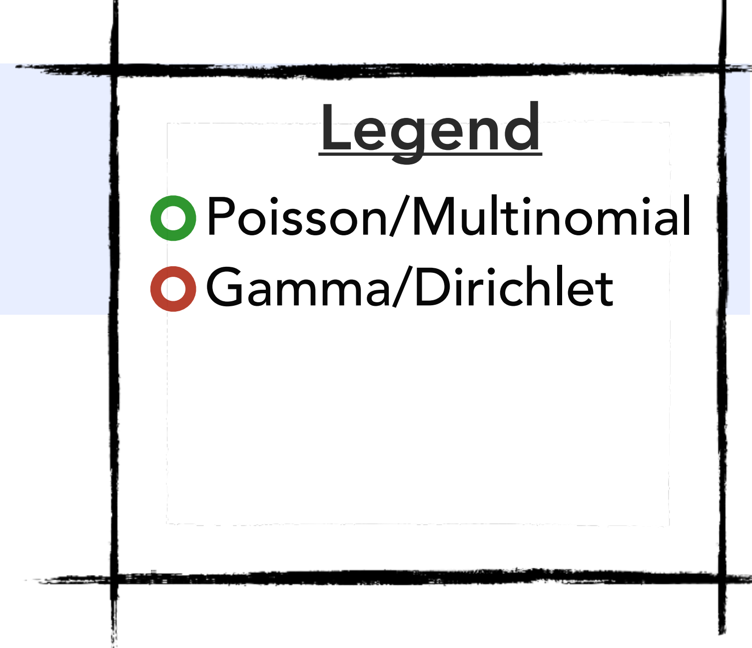


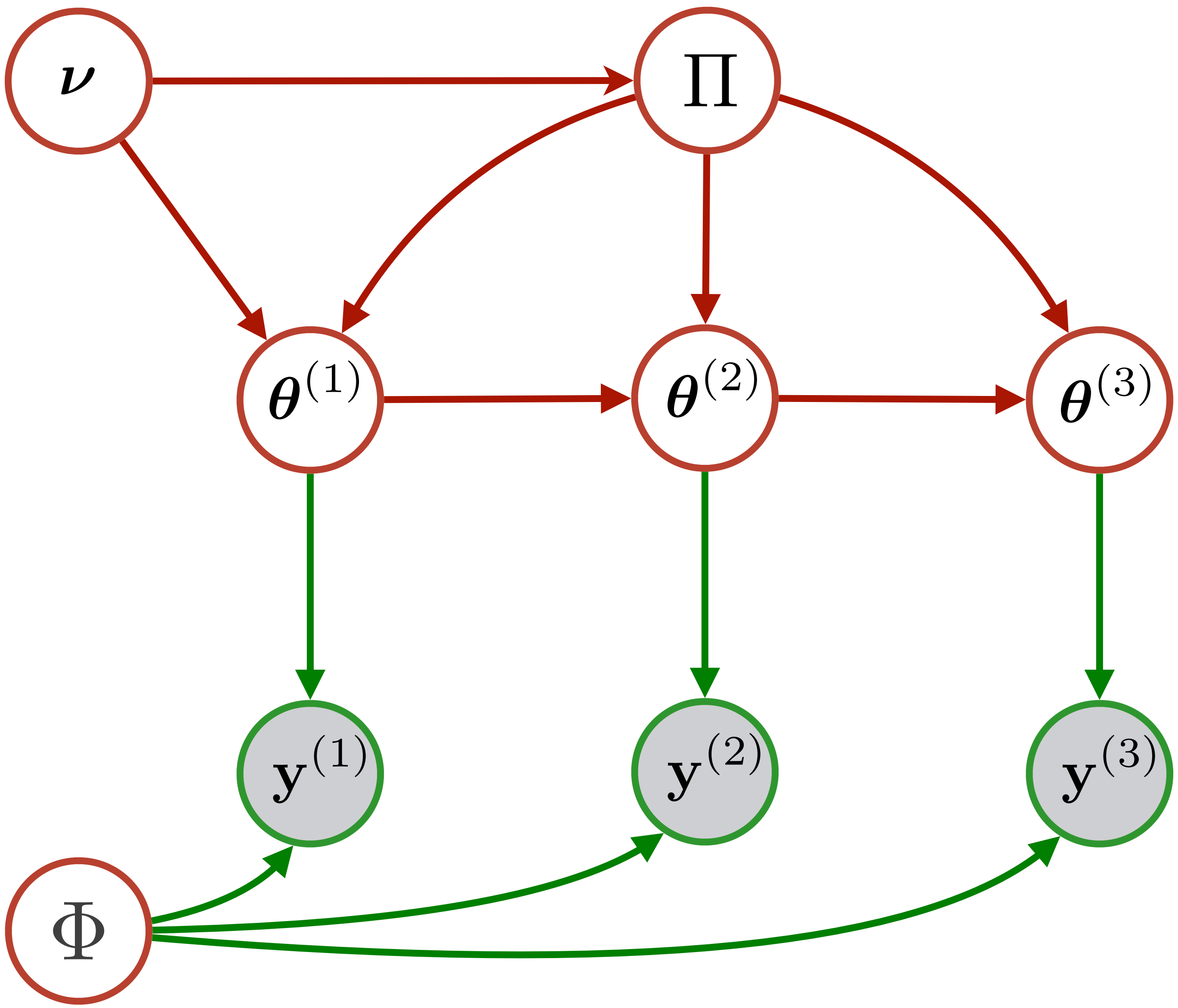


$$\Phi \sim P(\Phi \mid Y) \checkmark$$

(conditional posterior has closed form)

Technical challenge





$$\Pi \sim P(\Pi \mid Y, \Theta, \boldsymbol{\nu})$$ ✗

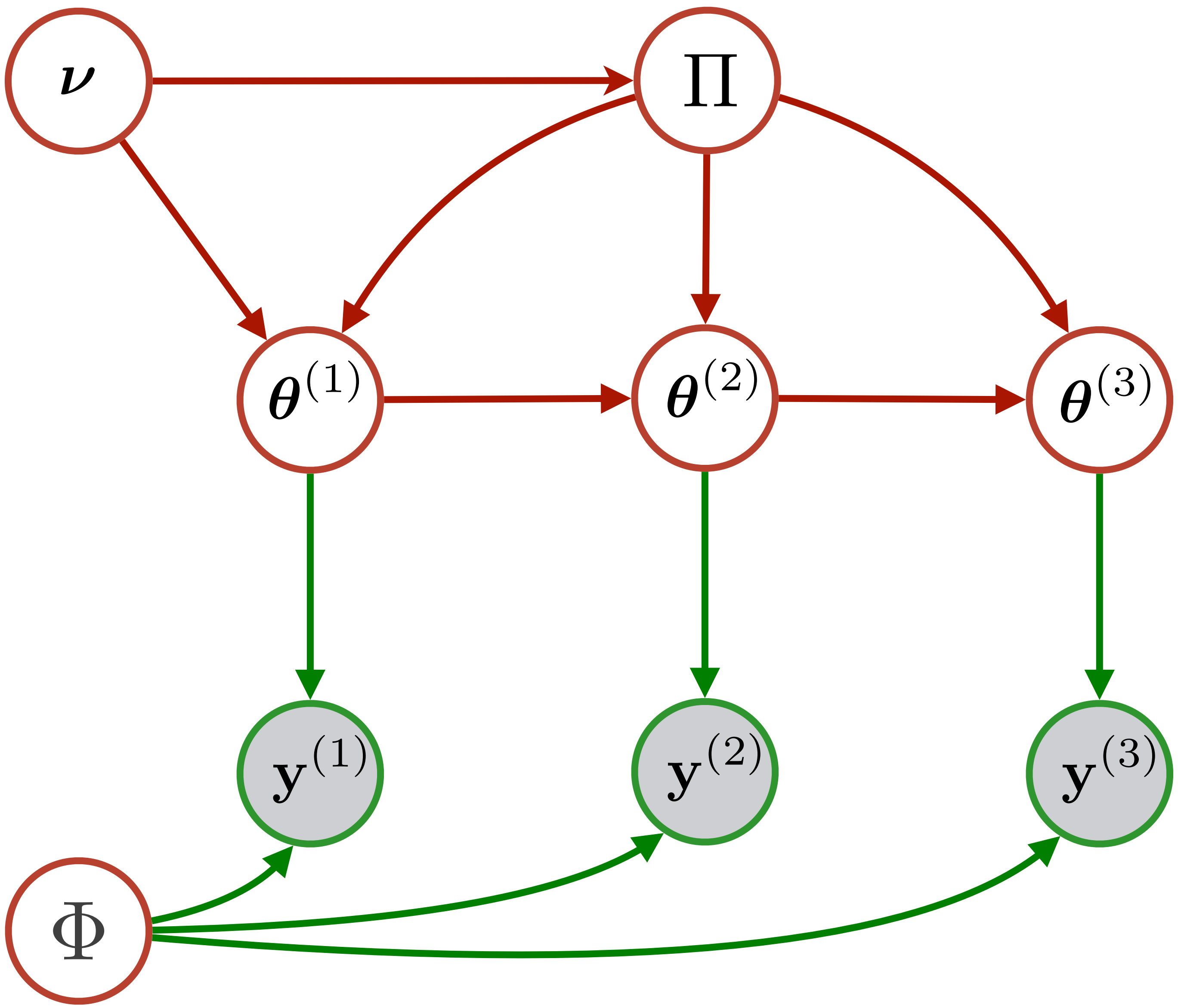
$$\Phi \sim P(\Phi \mid Y)$$ ✓

(conditional posterior has closed form)

Technical challenge

Legend

- Poisson/Multinomial
- Gamma/Dirichlet



$$\Pi \sim P(\Pi \,|\, Y, \Theta, \boldsymbol{\nu})$$ ✗

$$\Theta \sim P(\Theta \,|\, Y, \Pi, \boldsymbol{\nu})$$ ✗

$$\Phi \sim P(\Phi \,|\, Y)$$ ✓

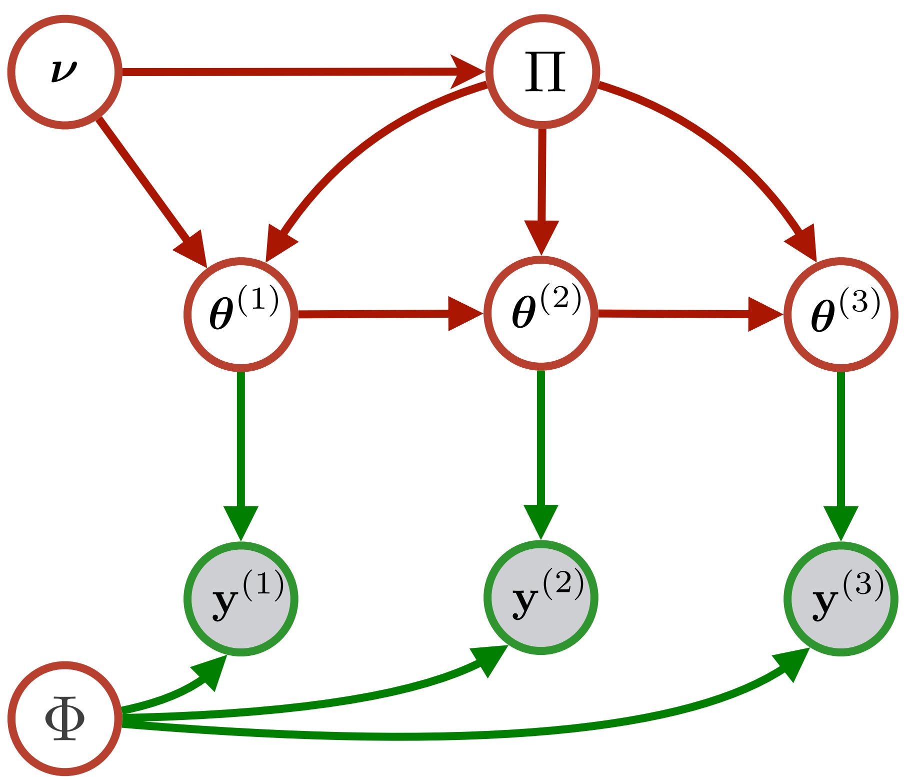
(conditional posterior has closed form)

Augment and Conquer

Legend

Poisson/Multinomial
Gamma/Dirichlet

Original model

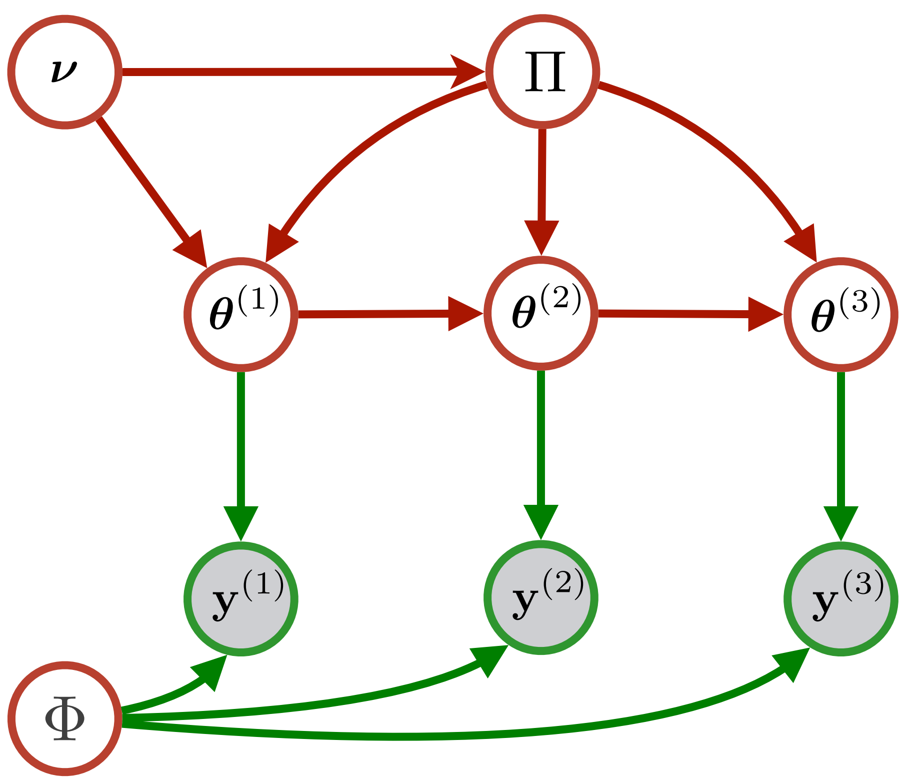


Augment and Conquer

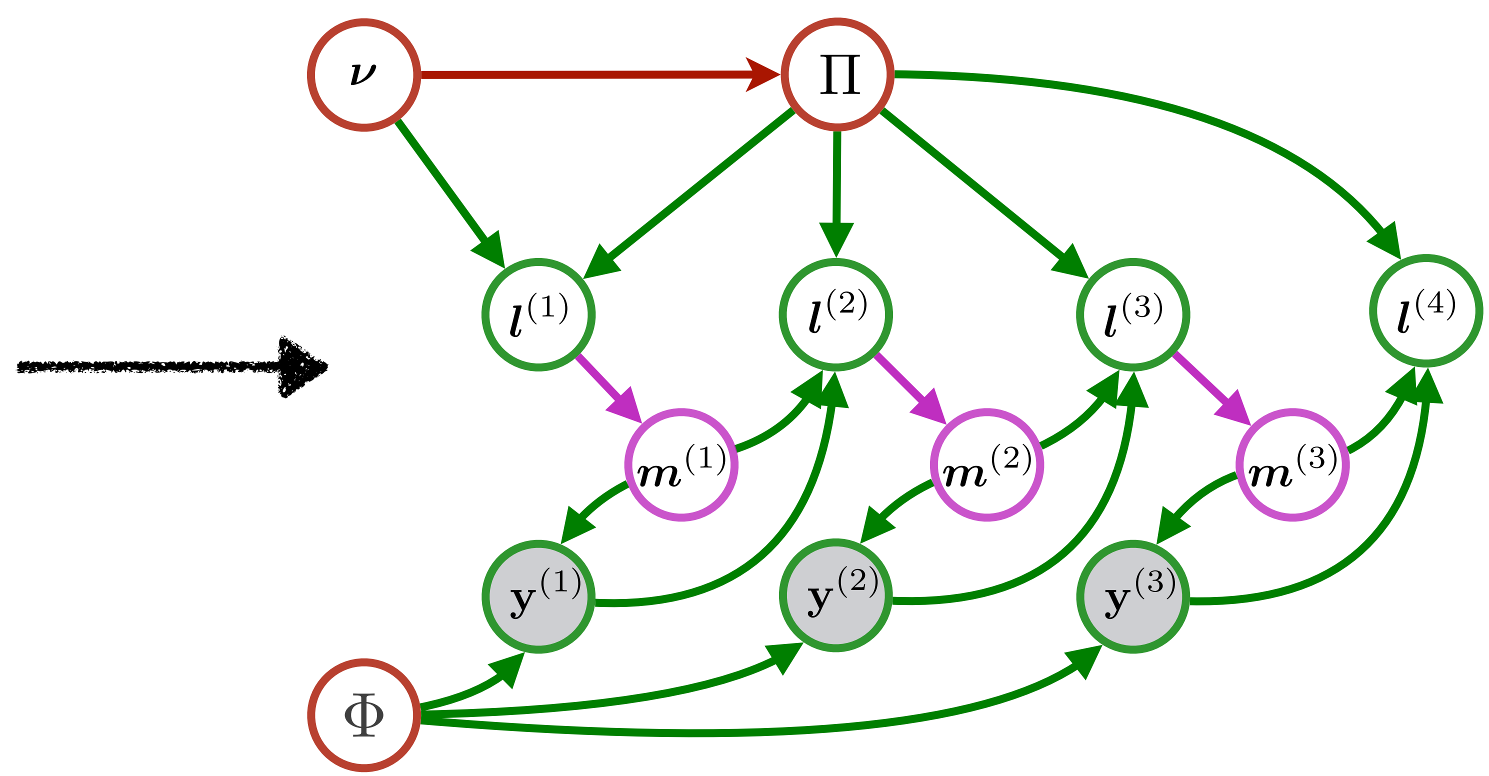
Legend

- Poisson/Multinomial
- Gamma/Dirichlet

Original model



Alternative model



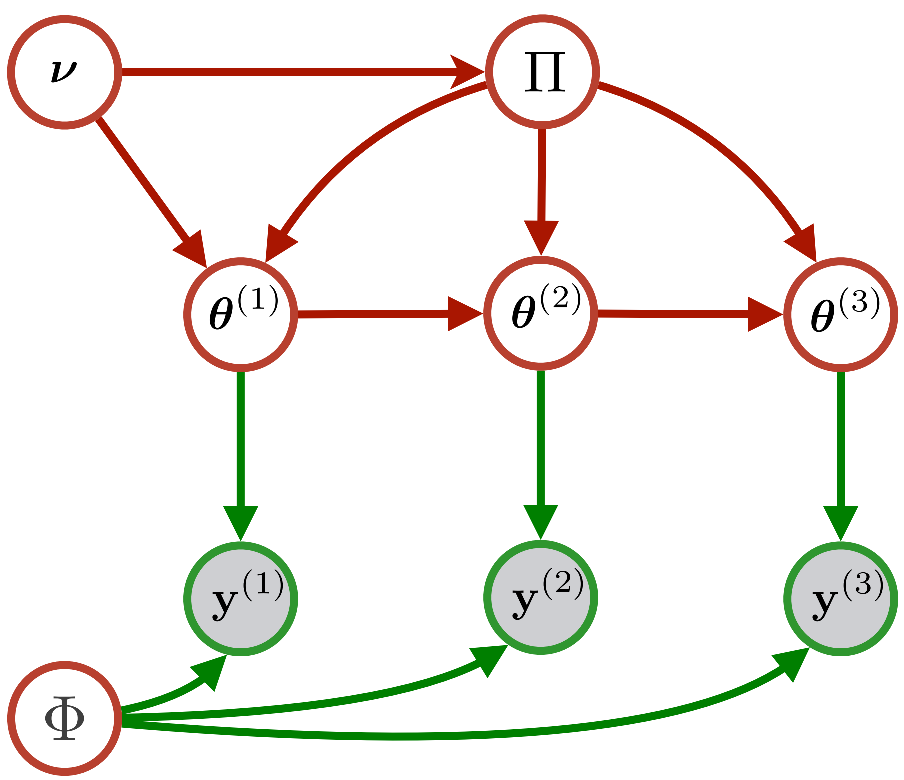
Augment and Conquer

Legend

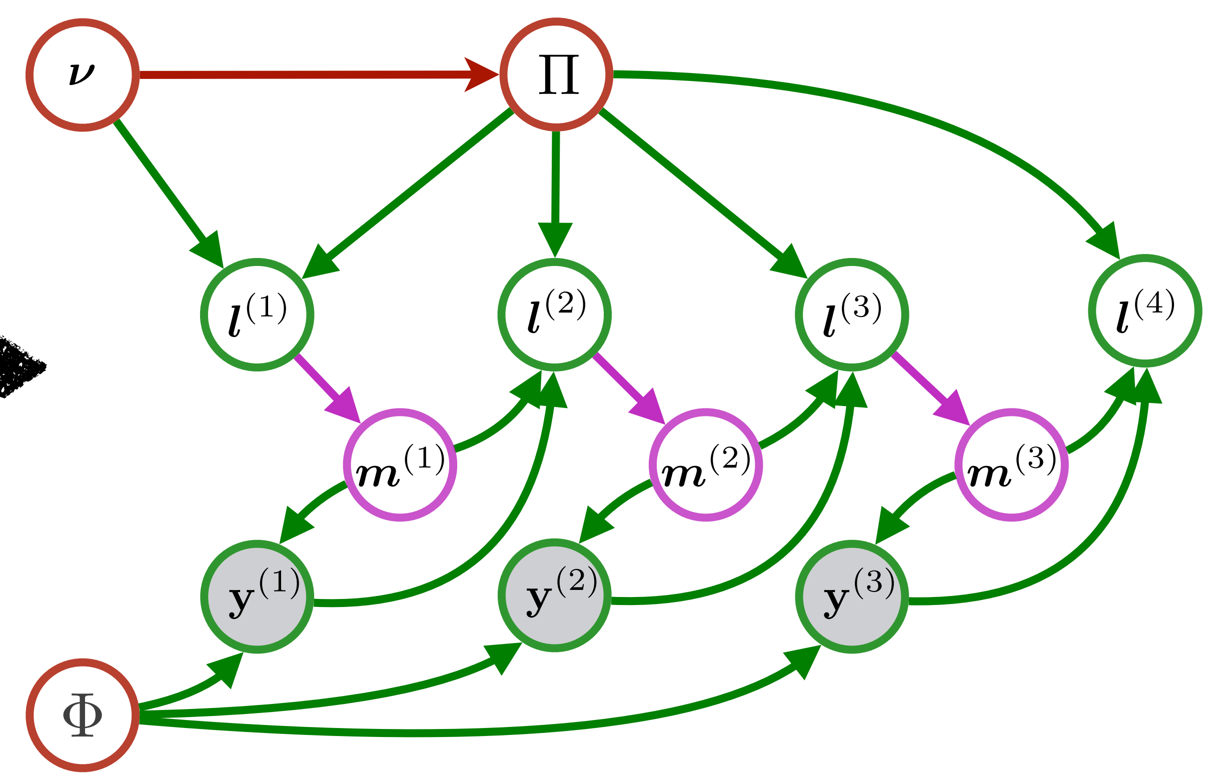
Poisson/Multinomial

Gamma/Dirichlet

Original model



Alternative model



$$\Pi \sim P\left(\Pi \,|\, \mathcal{A}, Y, \boldsymbol{\nu}\right) \checkmark$$

Poisson—gamma dynamical systems

[NeurIPS ’16]

Analogue to linear dynamical systems for high-dimensional **discrete time-series**

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

	T	$\hat{B}$	Task	Mean Relative Error (MRE) PGDS	GP-DPFA	LDS
GDELT	365	1.27	S	**2.335** ±0.19	2.951 ±0.32	3.493 ±0.53
			F	**2.173** ±0.41	2.207 ±0.42	2.397 ±0.29
ICEWS	365	1.10	S	**0.808** ±0.11	0.877 ±0.12	1.023 ±0.15
			F	**0.743** ±0.17	0.792 ±0.17	0.937 ±0.31
SOTU	225	1.45	S	**0.233** ±0.01	0.238 ±0.01	0.260 ±0.01
			F	**0.171** ±0.00	0.173 ±0.00	0.225 ±0.01
DBLP	14	1.64	S	0.417 ±0.03	0.422 ±0.05	**0.405** ±0.05
			F	**0.322** ±0.00	0.323 ±0.00	0.369 ±0.06
NIPS	17	0.33	S	0.415 ±0.07	**0.392** ±0.07	1.609 ±0.43
			F	0.343 ±0.01	**0.312** ±0.00	0.642 ±0.14

Poisson-randomized gamma dynamical systems

[NeurIPS '19]

Another analogue to linear dynamical systems
for high-dimensional discrete time-series

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

Poisson-randomized gamma dynamical systems

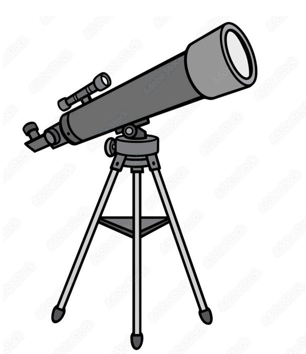
[NeurIPS '19]

Another analogue to linear dynamical systems
for high-dimensional discrete time-series

with APF, good things happen…

- **robust** to sparsity, overdispersion, burstiness
- model-fitting **scales with number of non-zeros**

Based on a **sparse discrete** representation…



…which lets it ``see'' things other models can't.

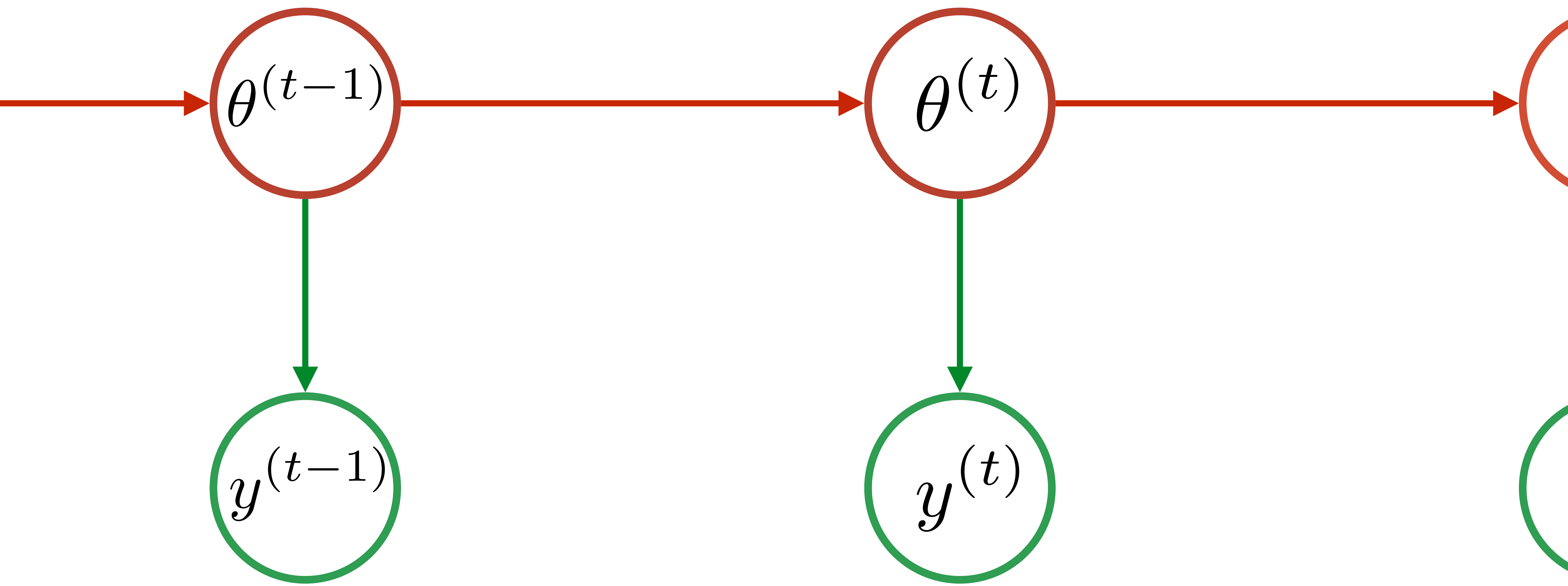
The gamma—Poisson—gamma chain

Legend

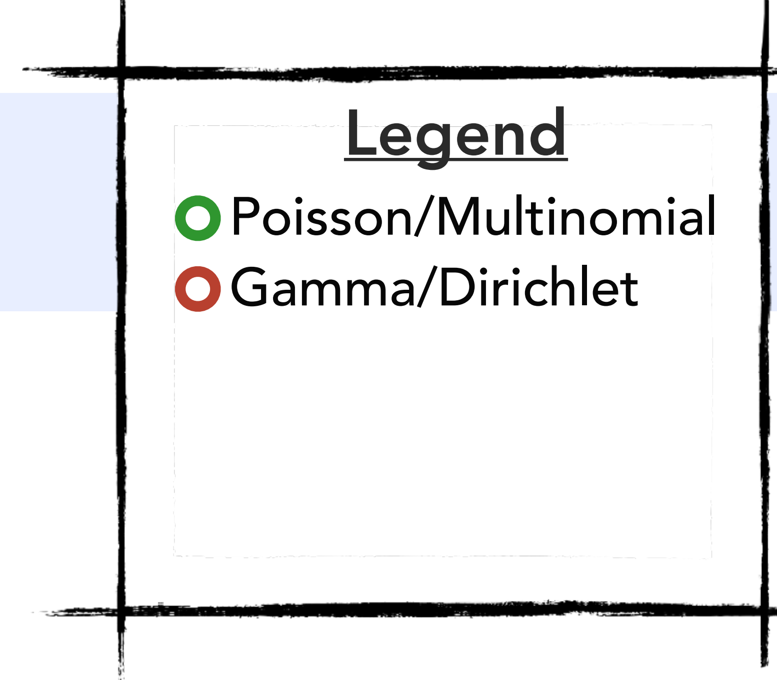
Poisson/Multinomial

Gamma/Dirichlet

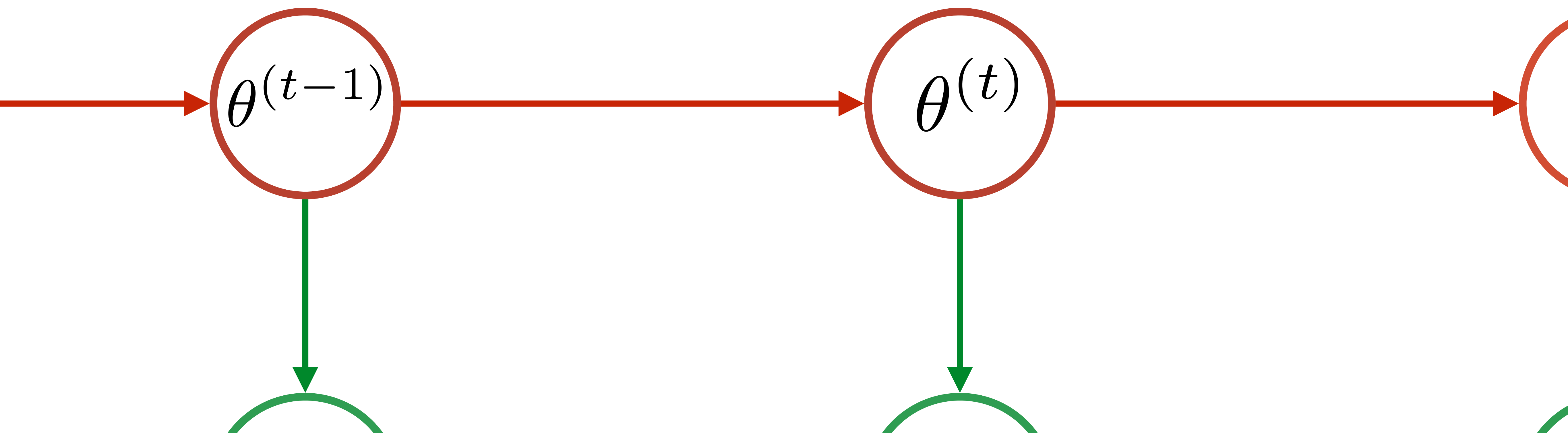
$$\theta^{(t)} \sim \mathrm{Gam}(\tau\,\theta^{(t-1)},\ \tau)$$



The gamma—Poisson—gamma chain



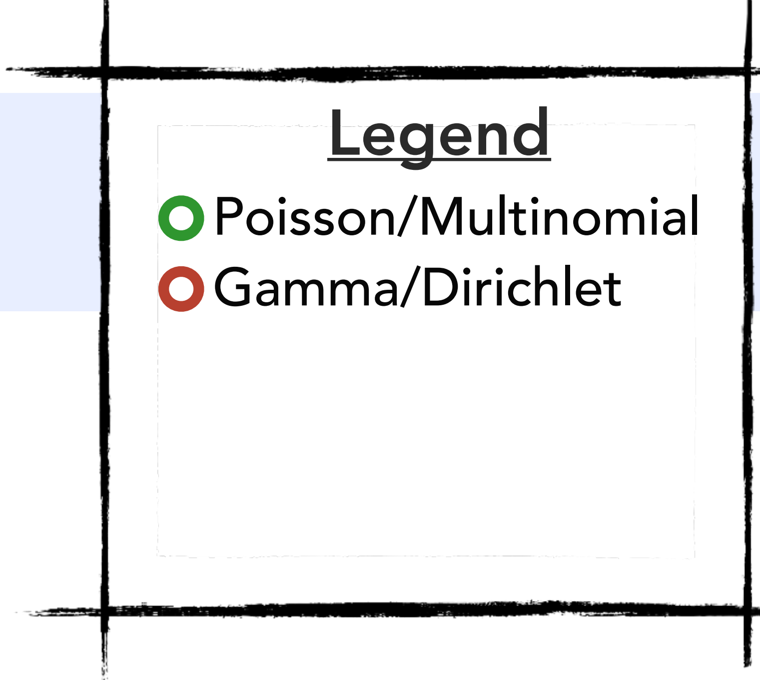
$$\theta^{(t)} \sim \mathrm{Gam}(\tau\, \theta^{(t-1)},\, \tau)$$

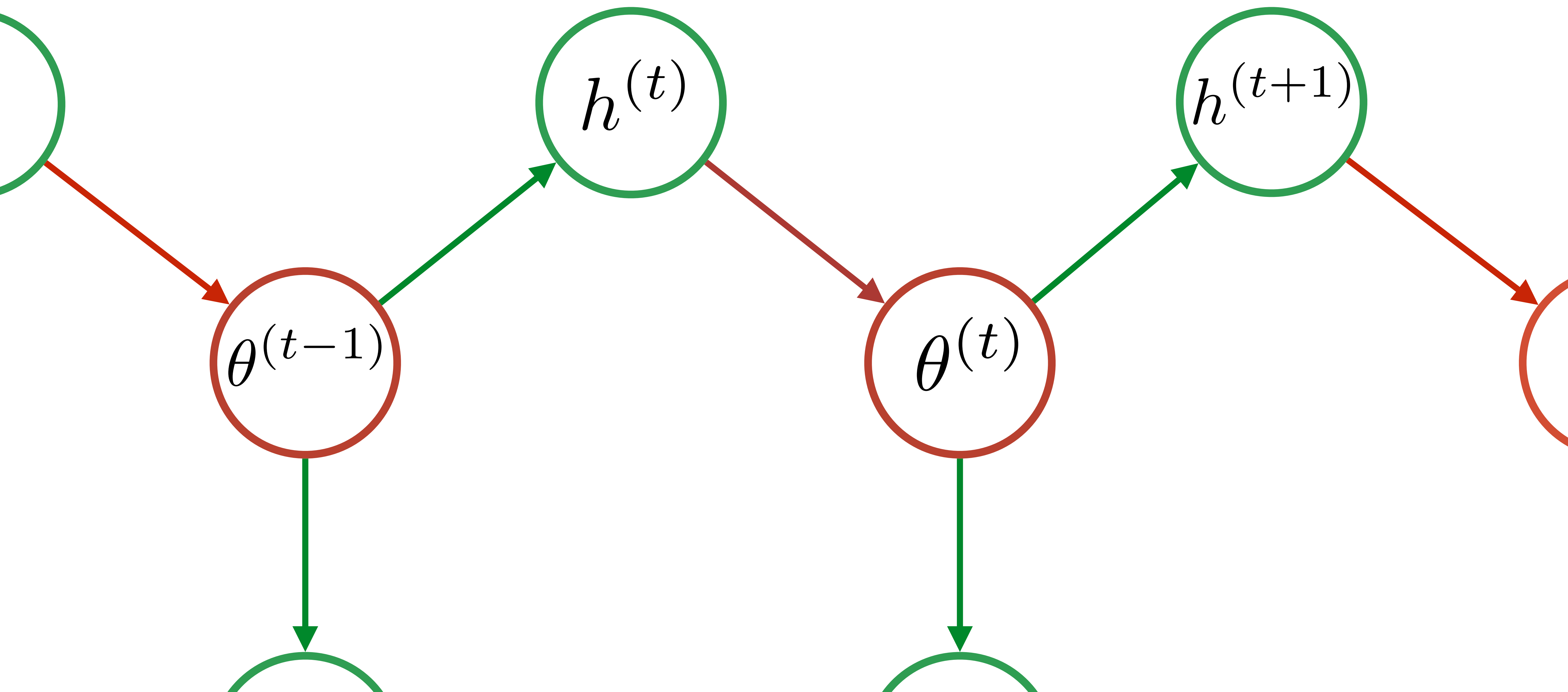


The gamma—Poisson—gamma chain

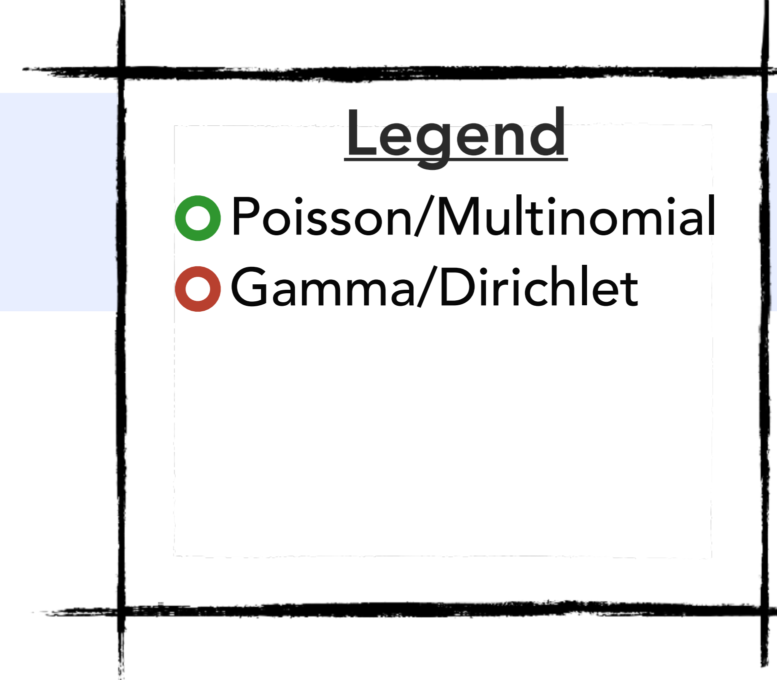
$$h^{(t)} \sim \mathrm{Pois}(\tau\theta^{(t-1)})$$ **discrete** latent state

$$\theta^{(t)} \sim \mathrm{Gam}(h^{(t)}, \tau)$$ **continuous** latent state





The gamma—Poisson—gamma chain

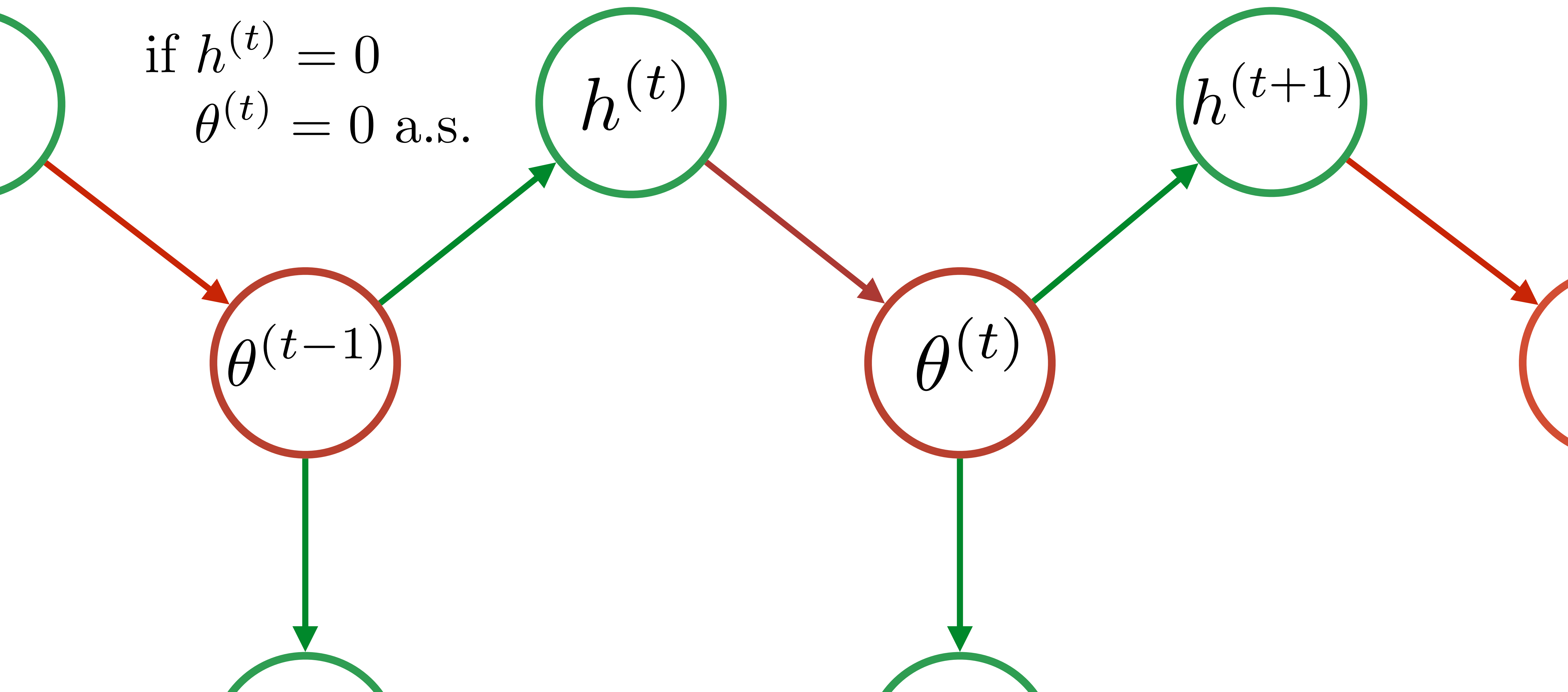


$$h^{(t)} \sim \text{Pois}(\tau\theta^{(t-1)})$$ **discrete** latent state

$$\theta^{(t)} \sim \text{Gam}(h^{(t)}, \tau)$$ **continuous** latent state

sparse representation

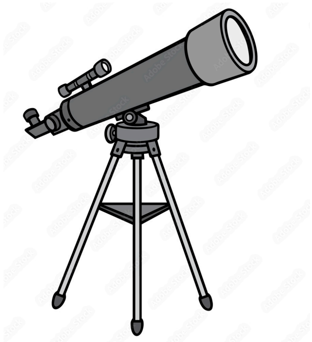
$$\text{if } h^{(t)} = 0$$
$$\theta^{(t)} = 0 \text{ a.s.}$$



Poisson-randomized gamma dynamical systems

[NeurIPS '19]

Sparse discrete representation…

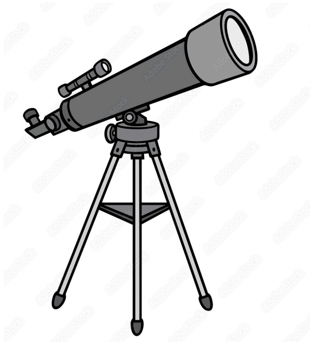


…lets it ``see'' things other models can't.

Poisson-randomized gamma dynamical systems

[NeurIPS '19]

Sparse discrete representation…



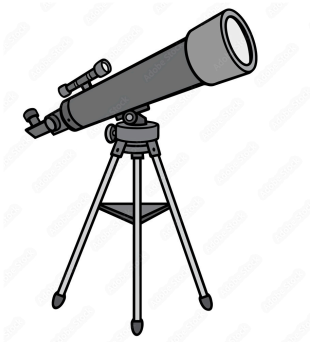
…lets it ``see'' things other models can't.

Qualitative study: PGDS [NeurIPS '16] **vs.** PRGDS [NeurIPS '19]

Poisson-randomized gamma dynamical systems

[NeurIPS '19]

Sparse discrete representation…



…lets it ``see'' things other models can't.

Qualitative study: PGDS [NeurIPS '16] **vs.** PRGDS [NeurIPS '19]

1) Fit them both to the same IR data.

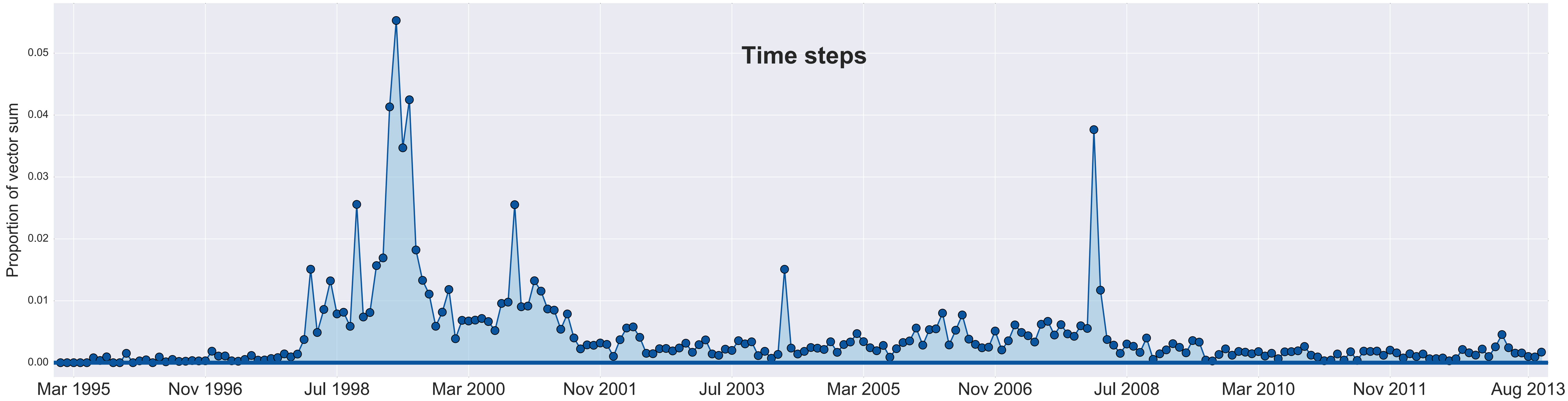
2) **Align** their latent components.

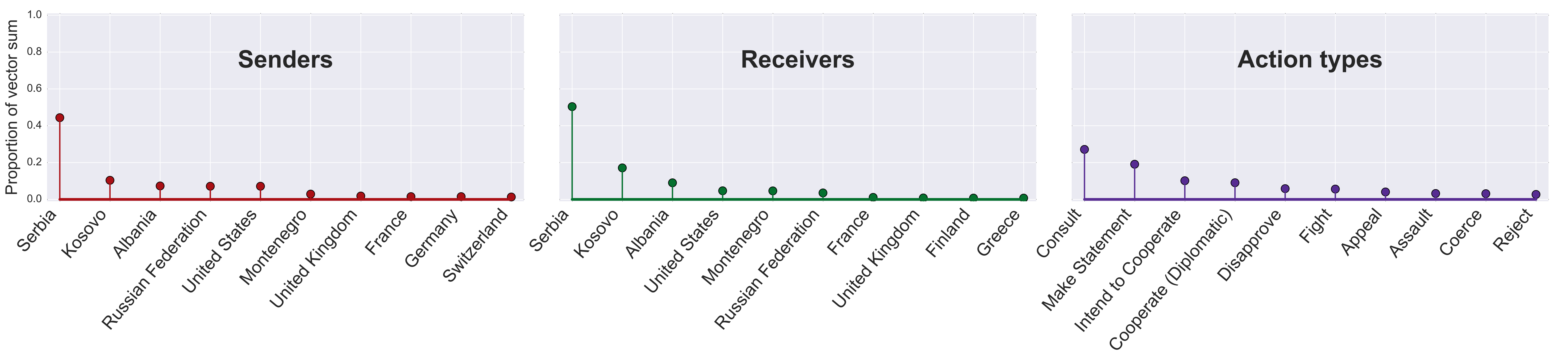
3) Check for systematic discrepancy.

Aligned component: **PGDS**

[NeurIPS '16]

Both models ``see'' this component…

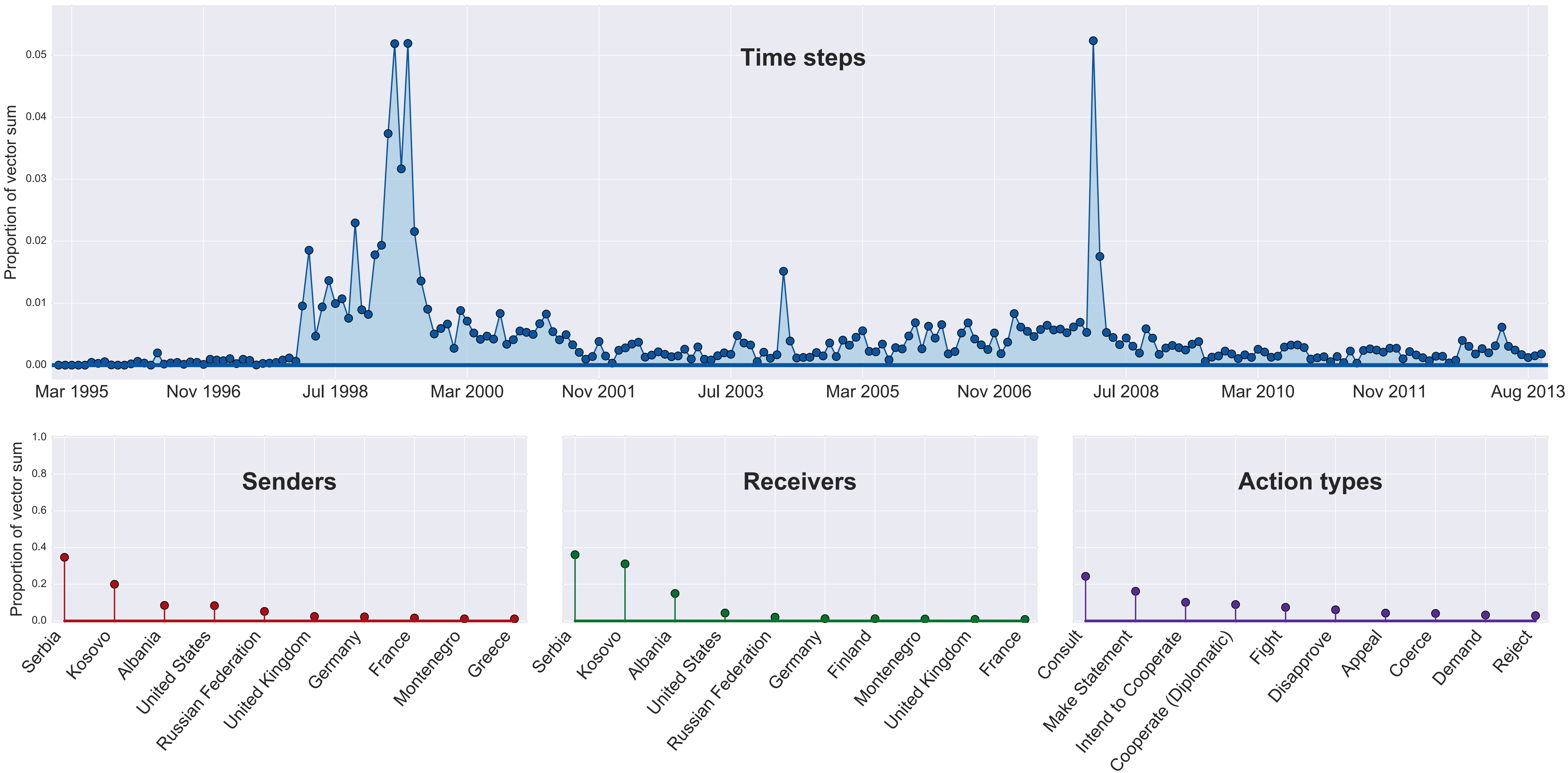




Aligned component: **PRGDS**

[NeurIPS '19]

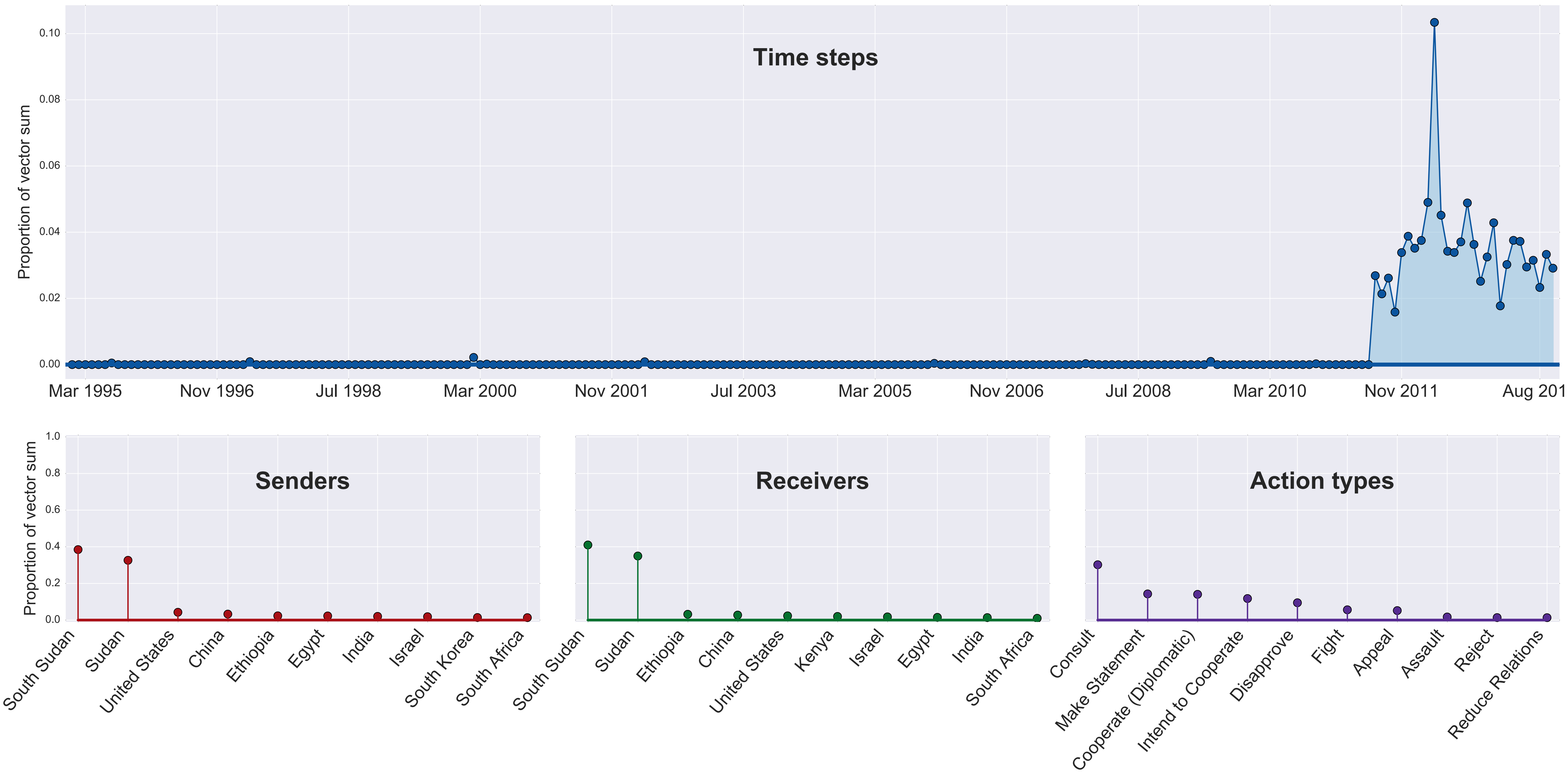
Both models "see" this component…



Discrepant component: **PRGDS**

[NeurIPS '19]

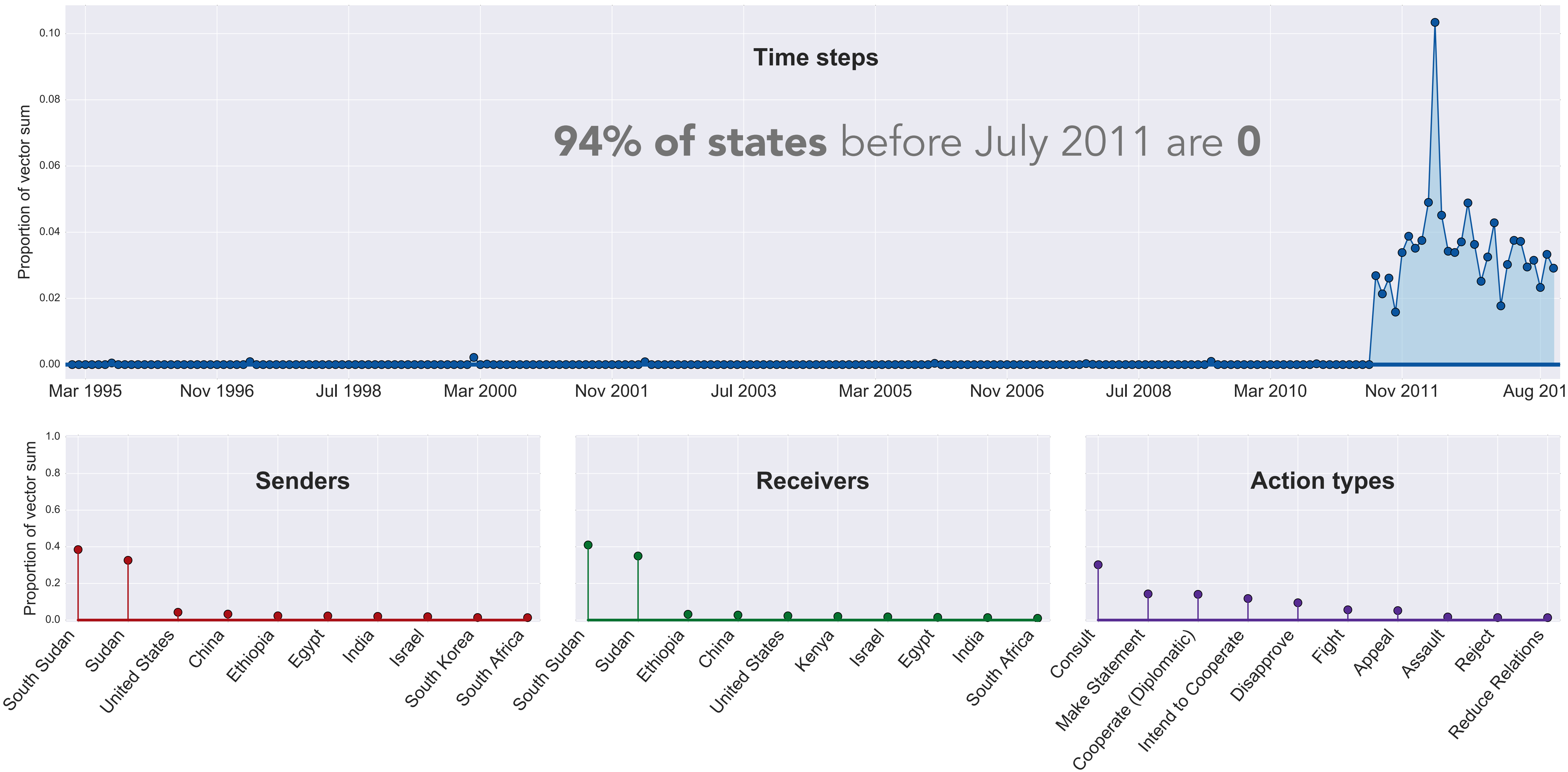
Only PRGDS ``sees'' this one.



Discrepant component: **PRGDS**

[NeurIPS ’19]

Only PRGDS ``sees” this one.

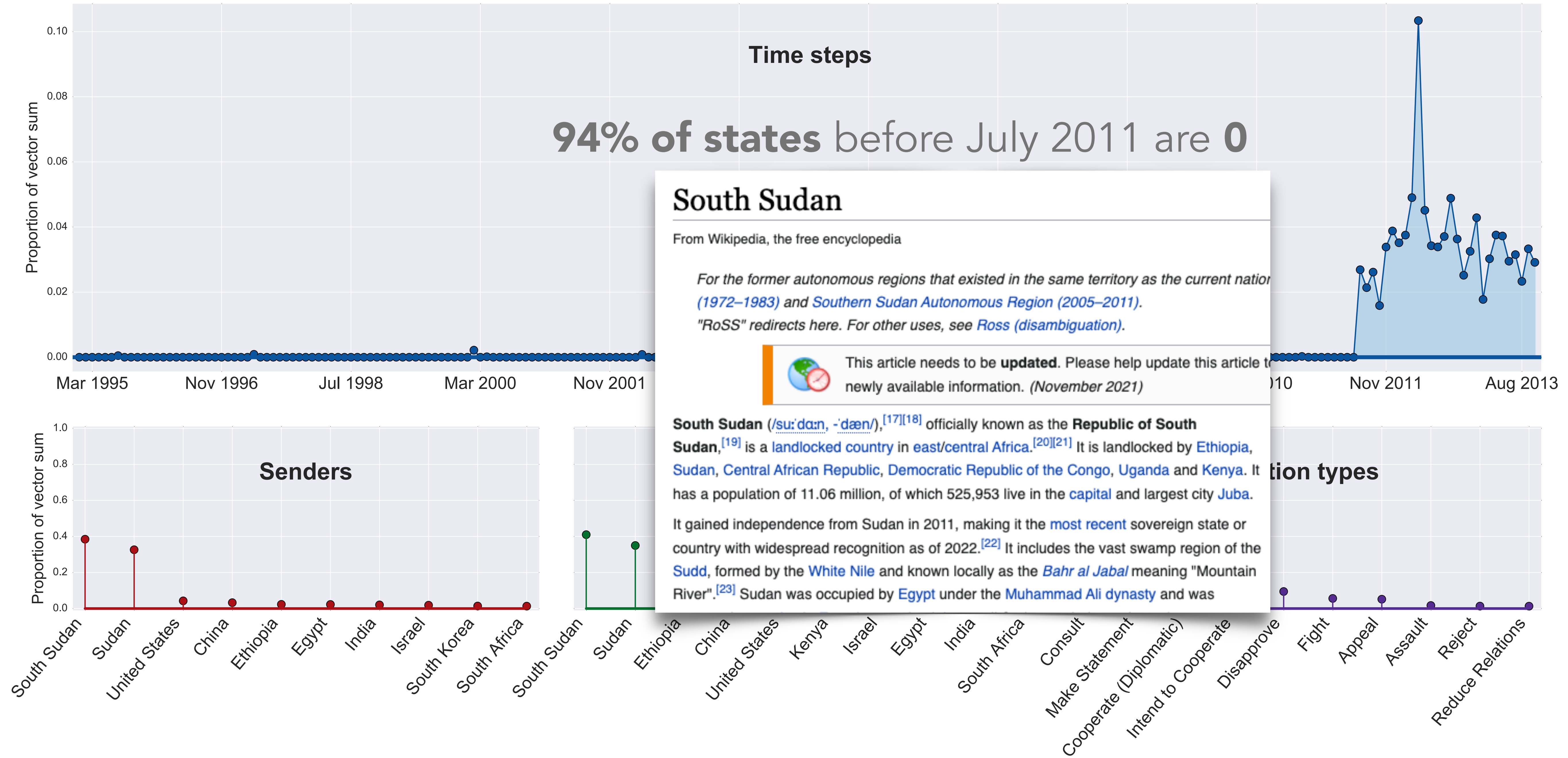


Discrepant component: **PRGDS**

[NeurIPS '19]

Only PRGDS ``sees'' this one.

94% of states before July 2011 are **0**



South Sudan

From Wikipedia, the free encyclopedia

For the former autonomous regions that existed in the same territory as the current natior (1972–1983) and Southern Sudan Autonomous Region (2005–2011). "RoSS" redirects here. For other uses, see Ross (disambiguation).

This article needs to be **updated**. Please help update this article t newly available information. *(November 2021)*

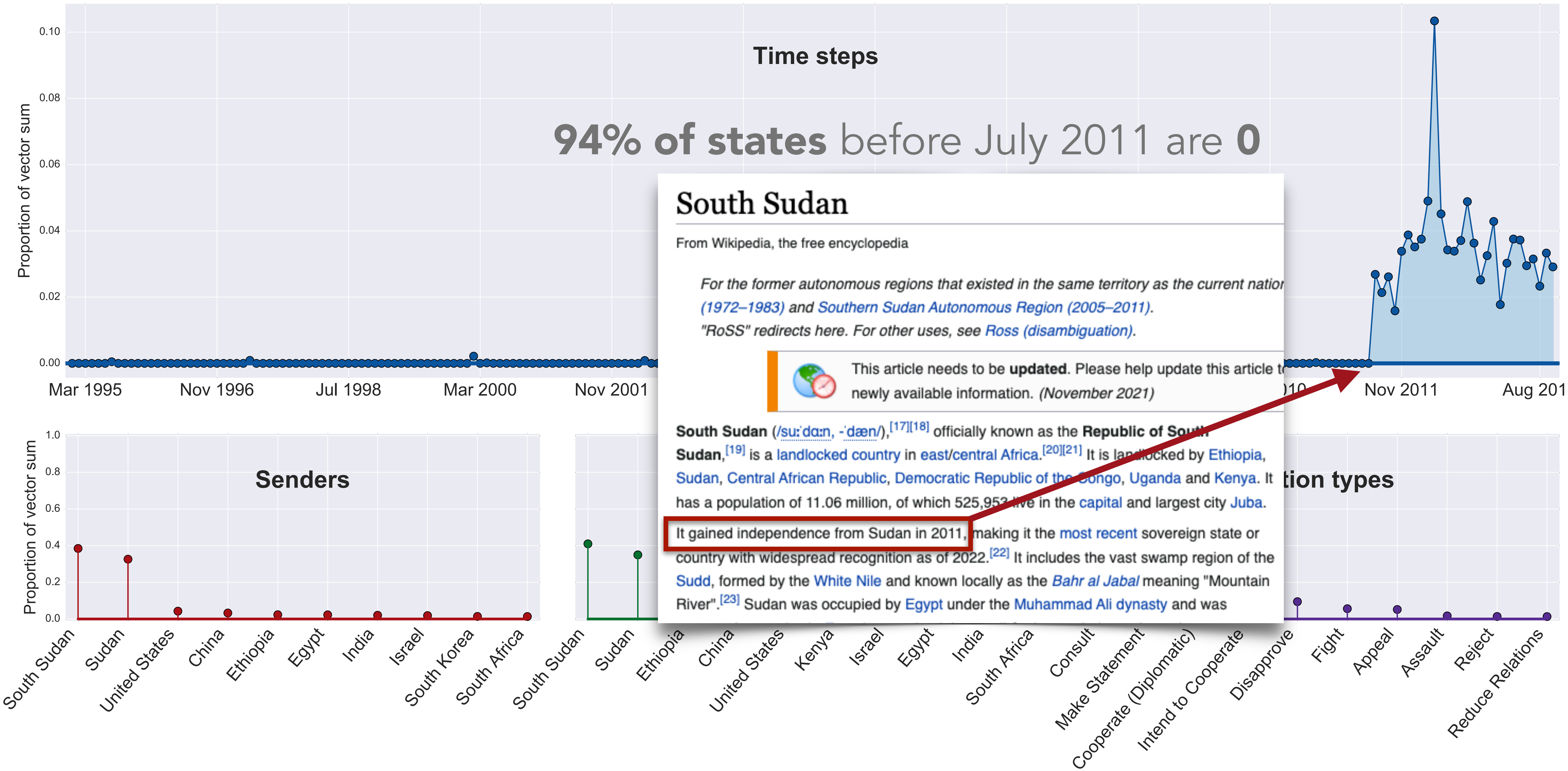
South Sudan (/suːˈdɑːn, -ˈdæn/),[17][18] officially known as the **Republic of South Sudan**,[19] is a landlocked country in east/central Africa.[20][21] It is landlocked by Ethiopia, Sudan, Central African Republic, Democratic Republic of the Congo, Uganda and Kenya. It has a population of 11.06 million, of which 525,953 live in the capital and largest city Juba.

It gained independence from Sudan in 2011, making it the most recent sovereign state or country with widespread recognition as of 2022.[22] It includes the vast swamp region of the Sudd, formed by the White Nile and known locally as the *Bahr al Jabal* meaning "Mountain River".[23] Sudan was occupied by Egypt under the Muhammad Ali dynasty and was

Discrepant component: **PRGDS**

[NeurIPS ’19]

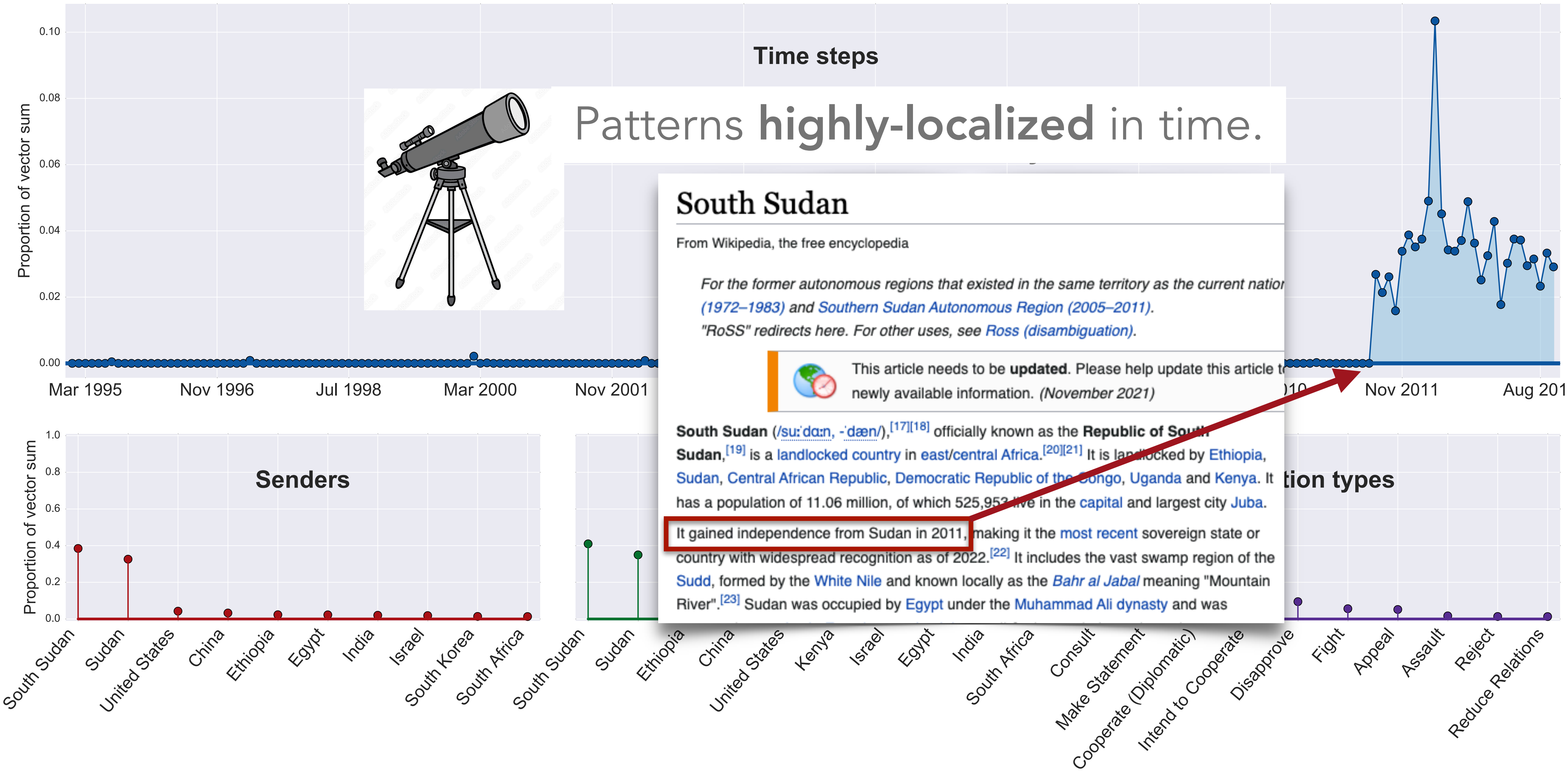
Only PRGDS ``sees” this one.



Discrepant component: **PRGDS**

[NeurIPS '19]

Only PRGDS ``sees'' this one.



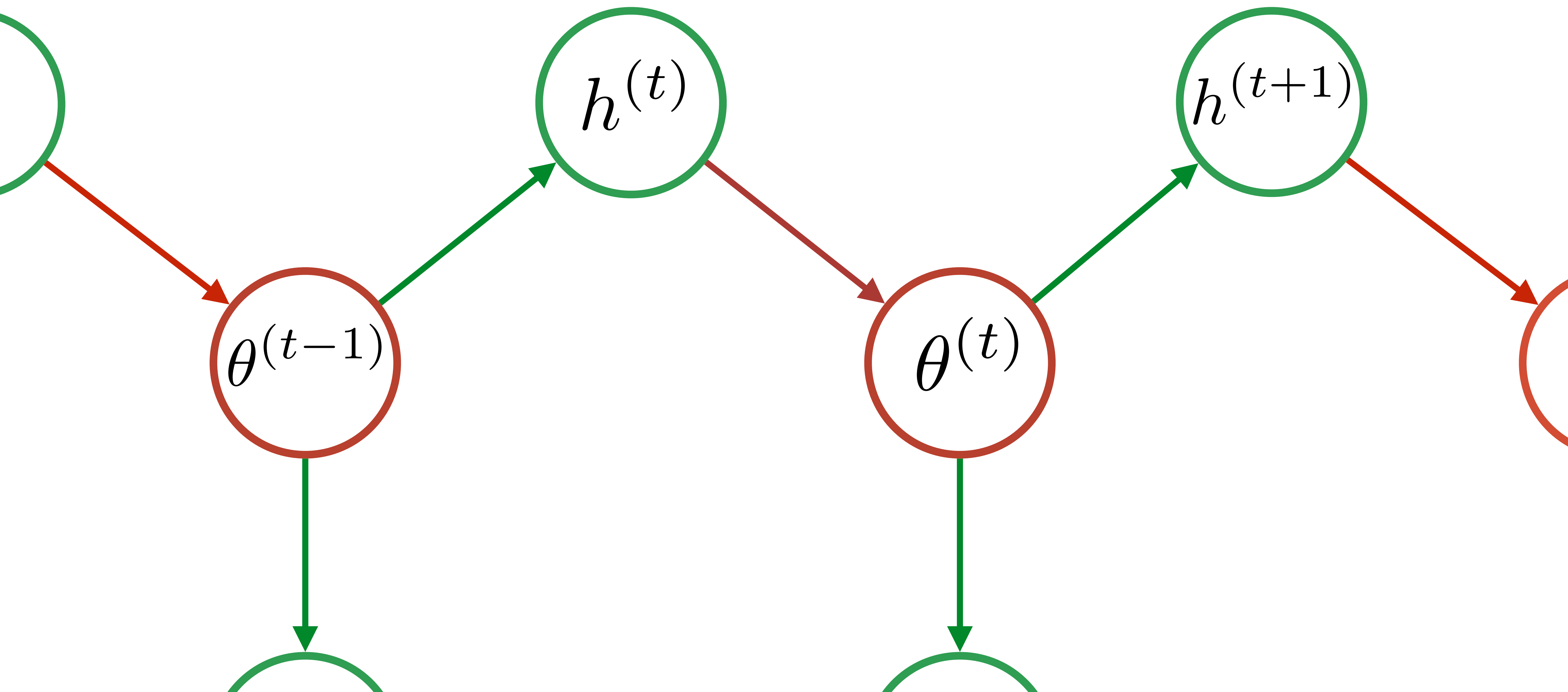
Approximate Bayesian inference

Legend

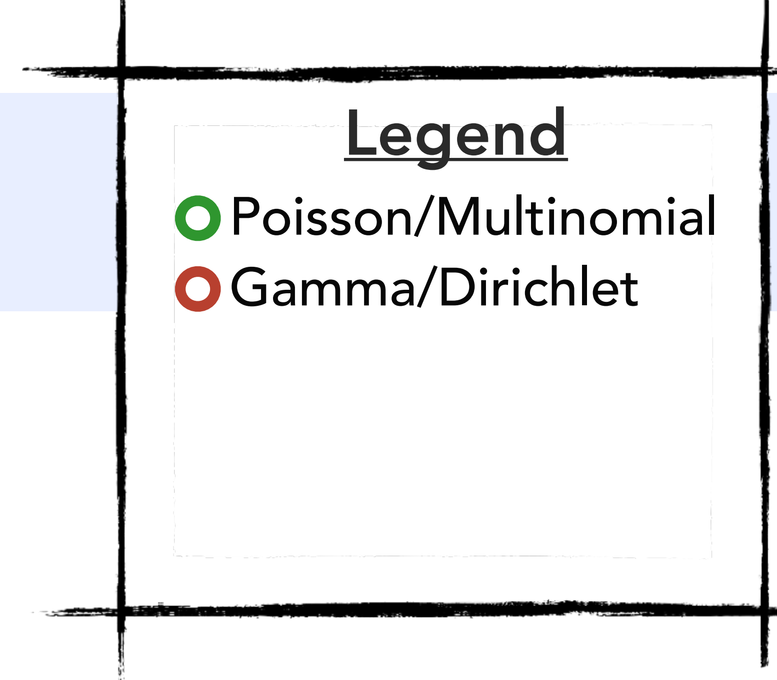
- Poisson/Multinomial
- Gamma/Dirichlet

Partially conditionally conjugate

$$\left(\theta^{(t)} \mid -\right) \sim P\left(\theta^{(t)} \mid h^{(t)}, h^{(t+1)}, -\right)$$



Approximate Bayesian inference

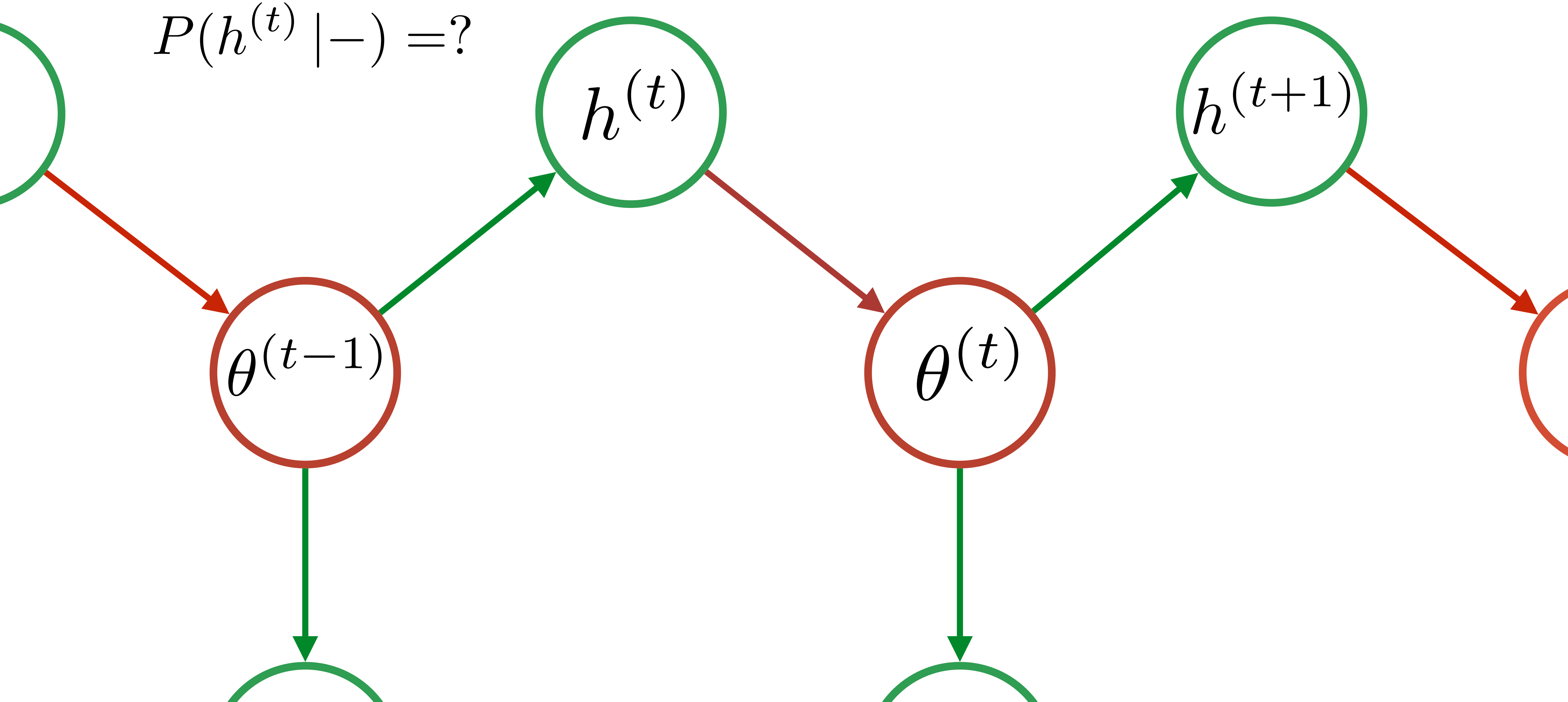


Partially conditionally conjugate

$$\left(\theta^{(t)} \mid -\right) \sim P\left(\theta^{(t)} \mid h^{(t)}, h^{(t+1)}, -\right) \checkmark$$

Partially not

$$P(h^{(t)} \mid -) = ?$$



Approximate Bayesian inference

Partially conditionally conjugate

$$\left(\theta^{(t)} \mid -\right) \sim P\left(\theta^{(t)} \mid h^{(t)}, h^{(t+1)}, -\right) \checkmark$$

Partially not

$$P(h^{(t)} \,|{-}) = ?$$

…but even without conjugacy…

closed-formedness!

Approximate Bayesian inference

Partially conditionally conjugate

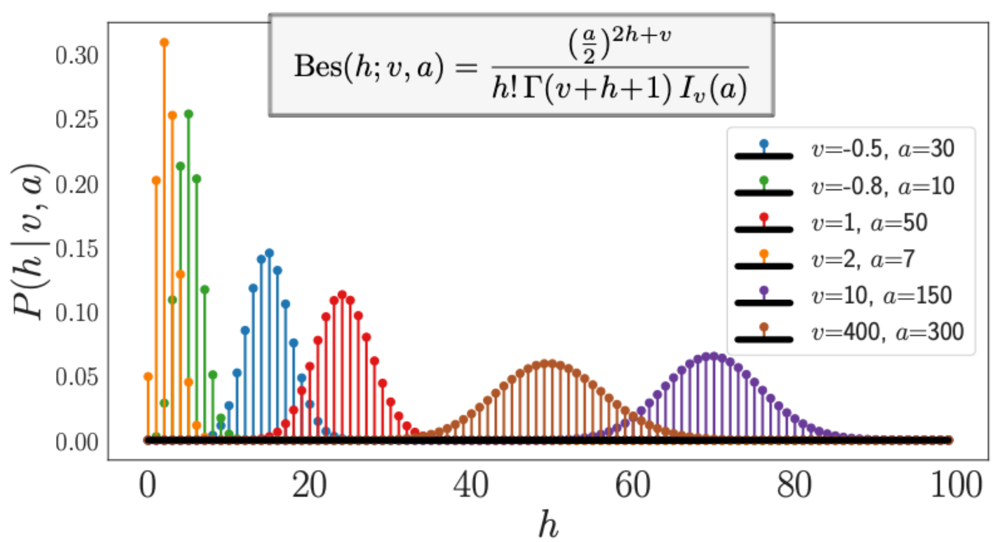
$$\left(\theta^{(t)} \mid -\right) \sim P\left(\theta^{(t)} \mid h^{(t)}, h^{(t+1)}, -\right) ✓$$

Partially not

$$P(h^{(t)} \mid -) = ✓$$ …but even without conjugacy…

closed-formedness!

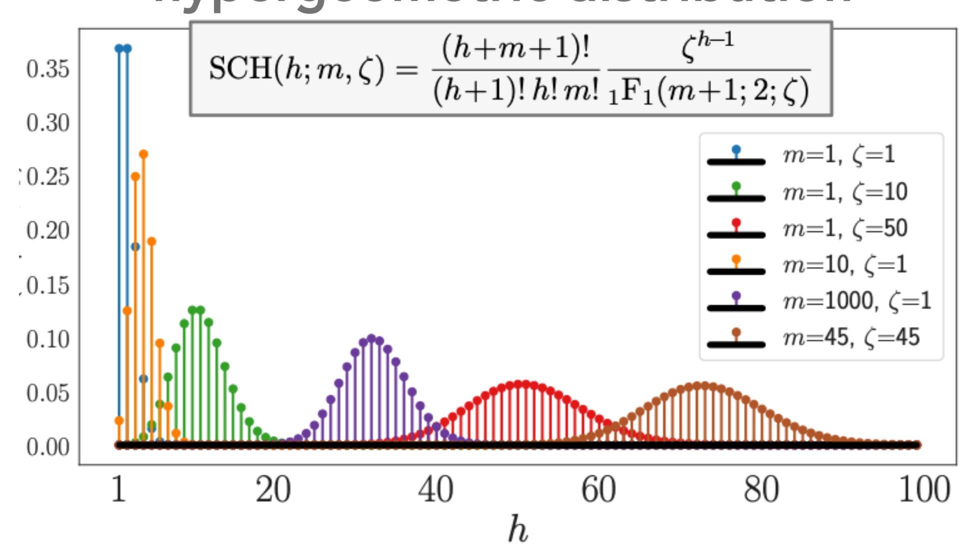
Bessel distribution



(Yuan and Kalbfleisch, 2000)

(Devroye 2002)

Size-biased confluent hypergeometric distribution



OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

OUTLINE

- **Motivation:** Dyadic event data (and its challenges)

- **Non-negative (Poisson) tensor decompositions**
  - **CP** extracts "multilateral" structure [Schein et al., 2015]
  - **Tucker** extracts "communities" [Schein et al., 2016]
  - Why (allocative) Poisson factorization
  - **ALℓ_0CORE** avoids the "exponential blowup" [Hood & Schein, 2024]

Appendix (*time permitting*)

- **State-space** modeling
  - **Gamma state-space** models [Schein et al., 2016b; Schein et al., 2019]
  - **Modeling "escalation"** with a new matrix prior [Stoehr et al., 2023]

Escalation in international relations

Escalation is a central **concept** in IR…

Table 2
WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	Type of Action
900717	IRQ	KUW	CHARGE
900725	IRQ	EGY	ASSURE
900727	IRQ	KUW	WARN
900731	IRQ	KUW	MOBILIZATION
900801	KUW	IRQ	REFUSE
900802	IRQ	KUW	MILITARY FORCE

…but **not observed** directly

AN ESCALATION LADDER
A Generalized (or Abstract) Scenario

AFTERMATHS

CIVILIAN CENTRAL WARS
44. Spasm or Insensate War
43. Some Other Kinds of Controlled General War
42. Civilian Devastation Attack
41. Augmented Disarming Attack
40. Countervalue Salvo
39. Slow-Motion Countercity War

(CITY TARGETING THRESHOLD)

MILITARY CENTRAL WARS
38. Unmodified Counterforce Attack
37. Counterforce-with-Avoidance Attack
36. Constrained Disarming Attack
35. Constrained Force-Reduction Salvo
34. Slow-Motion Counterforce War
33. Slow-Motion Counter-"Property" War
32. Formal Declaration of "General" War

(CENTRAL WAR THRESHOLD)

EXEMPLARY CENTRAL ATTACKS
31. Reciprocal Reprisals
30. Complete Evacuation (Approximately 95 per cent)
29. Exemplary Attacks on Population
28. Exemplary Attacks Against Property
27. Exemplary Attack on Military
26. Demonstration Attack on Zone of Interior

(CENTRAL SANCTUARY THRESHOLD)

BIZARRE CRISES
25. Evacuation (Approximately 70 per cent)
24. Unusual, Provocative, and Significant Countermeasures
23. Local Nuclear War—Military
22. Declaration of Limited Nuclear War
21. Local Nuclear War—Exemplary

(NO NUCLEAR USE THRESHOLD)

INTENSE CRISES
20. "Peaceful" World-Wide Embargo or Blockade
19. "Justifiable" Counterforce Attack
18. Spectacular Show or Demonstration of Force
17. Limited Evacuation (Approximately 20 per cent)
16. Nuclear "Ultimatums"
15. Barely Nuclear War
14. Declaration of Limited Conventional War
13. Large Compound Escalation
12. Large Conventional War (or Actions)
11. Super-Ready Status
10. Provocative Breaking Off of Diplomatic Relations

(NUCLEAR WAR IS UNTHINKABLE THRESHOLD)

TRADITIONAL CRISES
9. Dramatic Military Confrontations
8. Harassing Acts of Violence
7. "Legal" Harassment—Retortions
6. Significant Mobilization
5. Show of Force
4. Hardening of Positions—Confrontation of Wills

(DON'T ROCK THE BOAT THRESHOLD)

SUBCRISIS MANEUVERING
3. Solemn and Formal Declarations
2. Political, Economic, and Diplomatic Gestures
1. Ostensible Crisis

DISAGREEMENT—COLD WAR

Stoehr et al. [AISTATS 2023]

The Ordered Matrix Dirichlet for State-Space Models

Niklas Stoehr[ʕ] **Benjamin J. Radford**[ŋ] **Ryan Cotterell**[ʕ] **Aaron Schein**[ɘ]
[ʕ]ETH Zurich [ŋ]UNC Charlotte [ɘ]The University of Chicago
niklas.stoehr@inf.ethz.ch bradfor7@uncc.edu ryan.cotterell@inf.ethz.ch schein@uchicago.edu

State-space models to **measure (de-)escalation** in dyadic event data

Niklas Werner Stöhr

Lead author:

Address

ETH Zürich
Doctorate at D-INFK
Niklas Werner Stöhr
Professur für Informatik
OAT W 14
Andreasstrasse 5
8092 Zürich
Switzerland

Ordinal action types

action type *a*	action name	Goldstein value
0	provide aid	7.0
1	engage material cooperation	6.0
2	yield	5.0
3	express intent cooperate	4.0
4	engage diplomatic cooperation	3.5
5	appeal	3.0
6	consult	1.0
7	make public statement	0.0
9	investigate	-2.0
10	disapprove	-2.0
11	reject	-4.0
12	reduce relations	-4.0
13	demand	-5.0
14	threaten	-6.0
15	protest	-6.5
16	coerce	-7.0
17	exhibit force posture	-7.2
18	assault	-9.0
19	fight	-10.0
20	unconventional mass violence	-10.0

State-space model

$$\mathbb{E}\left[y_{i \xrightarrow{a} j}^{(t)}\right] = \sum_{k=1}^{K} \lambda_{i \xrightarrow{k} j}^{(t)} \phi_{ka}$$

State-space model

how much **dyad** $i \rightarrow j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

State-space model

how much **dyad** $i \rightarrow j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

State-space model

how much **dyad** $i \rightarrow j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i\overset{a}{\rightarrow}j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i\overset{k}{\rightarrow}j}\, \phi_{ka}$$

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i\overset{k}{\rightarrow}j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i\overset{k'}{\rightarrow}j}$$

State-space model

how much **dyad** $i \rightarrow j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

Φ is the $K \times A$ **emission matrix**

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**

State-space model

how much **dyad** $i \rightarrow j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

Φ is the $K \times A$ **emission matrix**
(row-stochastic) $\sum_{a=1}^{A} \phi_{ka} = 1$

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**
(row-stochastic) $\sum_{k'=1}^{K} \pi_{kk'} = 1$

State-space model

how much **dyad** $i \to j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

Φ is the $K \times A$ **emission matrix**

Idea: *Impose ordering on states:*
k^{th} *state is more conflictual than* $(k-1)^{th}$

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**

State-space model

how much **dyad** $i \to j$
is in **state** k at **time** t

how much **state** k
emits **action** a

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

Φ is the $K \times A$ **emission matrix**
(row-stochastic) $\sum_{a=1}^{A} \phi_{ka} = 1$

Idea: *Impose ordering on states:*
k^{th} *state is more conflictual than* $(k-1)^{th}$

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**
(row-stochastic) $\sum_{k'=1}^{K} \pi_{kk'} = 1$

Ordered state-space model

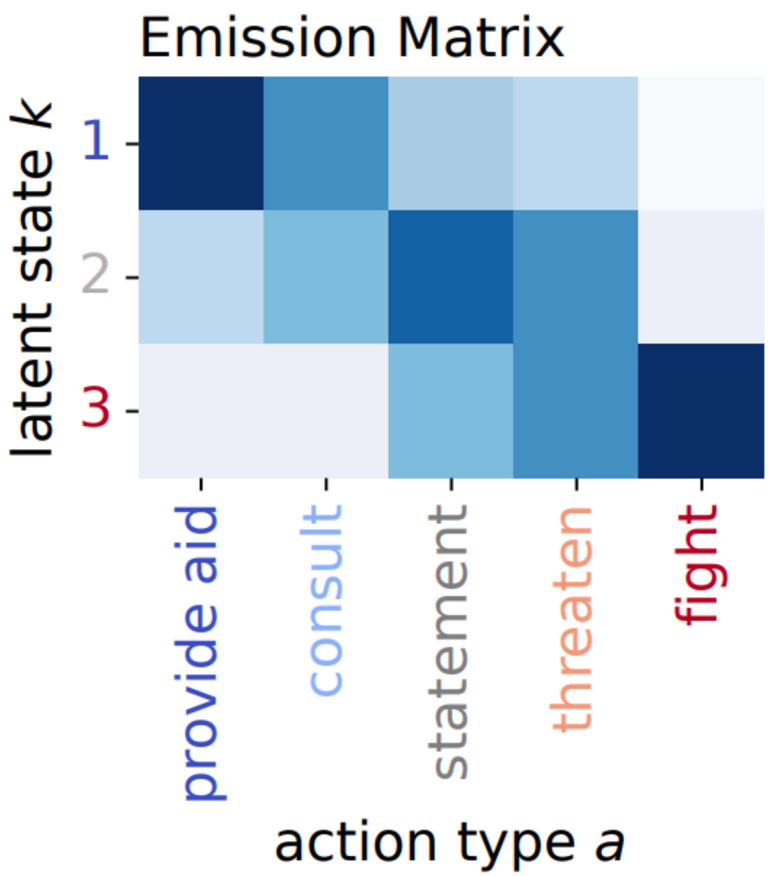
Idea: *Impose ordering on states:*
k^{th} *state is more conflictual than* $(k-1)^{th}$

Ordered state-space model

Idea: *Impose ordering on states:*
k^{th} *state is more conflictual than* $(k-1)^{th}$

Each state k defines a distribution over actions $(\phi_{k1}, \ldots, \phi_{kA})$

(k^{th} row of the emission matrix Φ)



Ordered state-space model

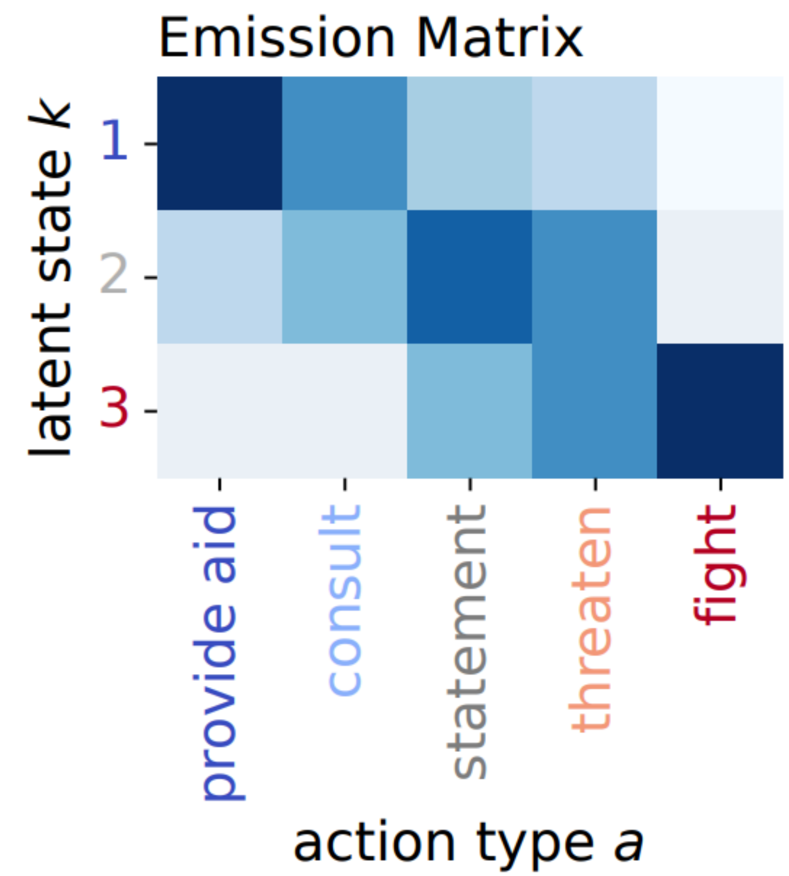
Idea: *Impose ordering on states:*
k^{th} *state is more conflictual than* $(k-1)^{th}$

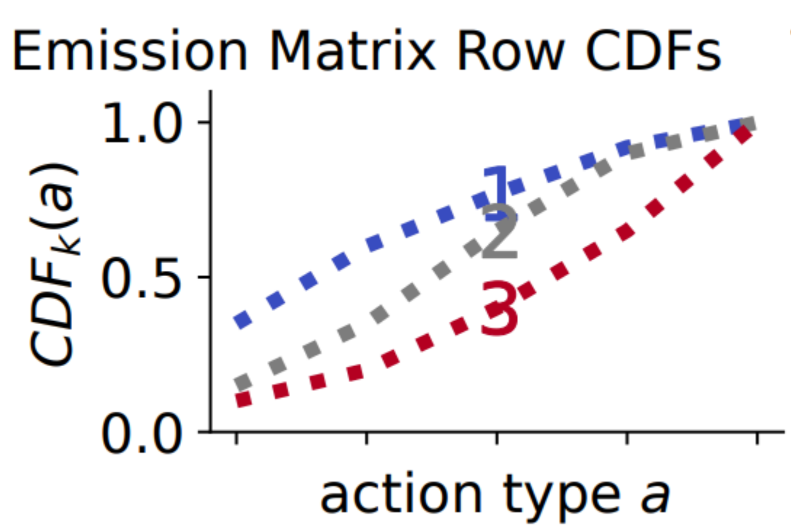
Each state k defines a distribution over actions $(\phi_{k1}, \ldots, \phi_{kA})$

(k^{th} row of the emission matrix Φ)

Operationalization:
k^{th} row *stochastically dominates* $(k-1)^{\text{th}}$

$$\sum_{a' \le a} \phi_{ka'} \ge \sum_{a' \le a} \phi_{k'a'} \quad \text{for any } a \text{ and } k < k'$$





Ordered state-space model

Idea: *Impose ordering on states:*
k^{th} *state is more conflictual than* $(k-1)^{th}$

Each state k defines a distribution over actions $(\phi_{k1}, \ldots, \phi_{kA})$

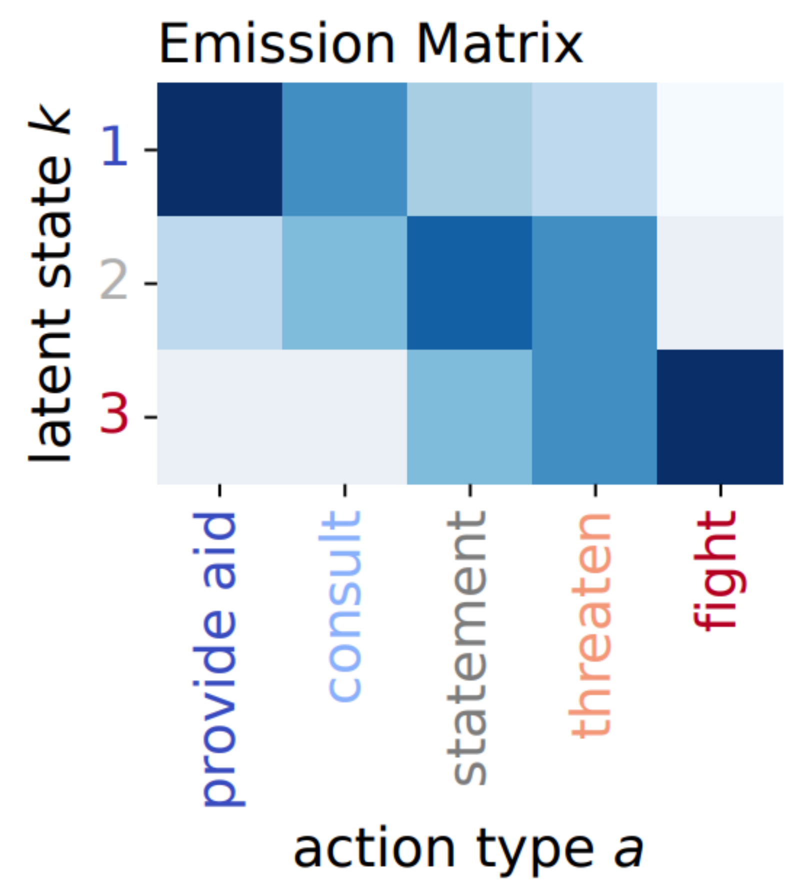
(k^{th} row of the emission matrix Φ)

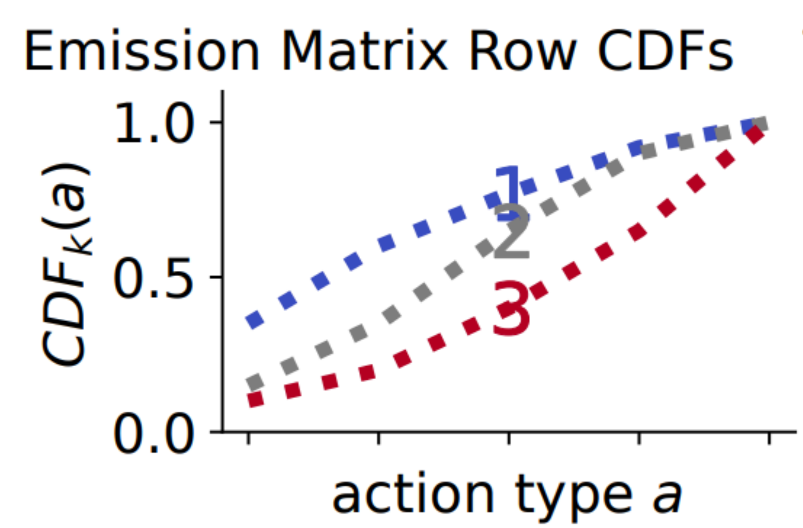
Operationalization:

k^{th} row *stochastically dominates* $(k-1)^{\text{th}}$

$$\sum_{a' \le a} \phi_{ka'} \ge \sum_{a' \le a} \phi_{k'a'} \quad \text{for any } a \text{ and } k < k'$$

``well-ordered stochastic matrix"
(probability mass shifts down and right)



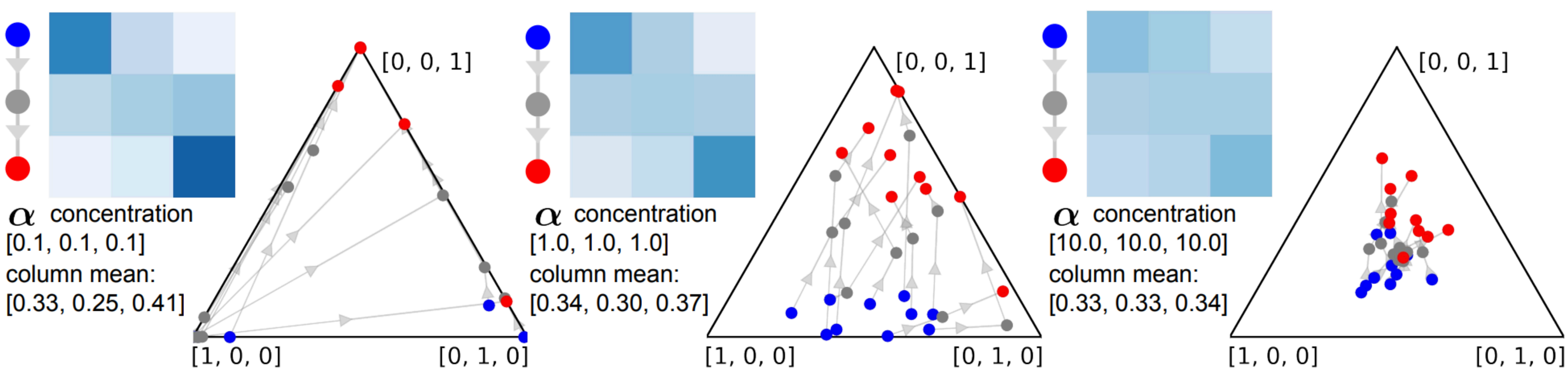


Ordered Matrix Dirichlet

A distribution over ``well-ordered" row-stochastic matrices

$$\Phi \sim \mathrm{OMD}(\alpha)$$

Defined implicitly using a **stick-breaking construction**



Ordered transitions

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**
(row-stochastic)

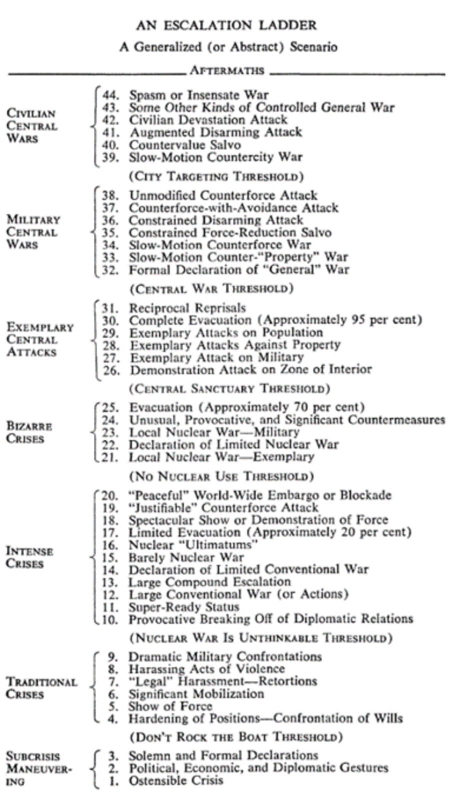
Ordered transitions

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**
(row-stochastic)

Ladder of escalation: *Allies do not become enemies overnight*



AN ESCALATION LADDER
A Generalized (or Abstract) Scenario

AFTERMATHS

CIVILIAN CENTRAL WARS
44. Spasm or Insensate War
43. Some Other Kinds of Controlled General War
42. Civilian Devastation Attack
41. Augmented Disarming Attack
40. Countervalue Salvo
39. Slow-Motion Countercity War

(CITY TARGETING THRESHOLD)

MILITARY CENTRAL WARS
38. Unmodified Counterforce Attack
37. Counterforce-with-Avoidance Attack
36. Constrained Disarming Attack
35. Constrained Force-Reduction Salvo
34. Slow-Motion Counterforce War
33. Slow-Motion Counter-"Property" War
32. Formal Declaration of "General" War

(CENTRAL WAR THRESHOLD)

EXEMPLARY CENTRAL ATTACKS
31. Reciprocal Reprisals
30. Complete Evacuation (Approximately 95 per cent)
29. Exemplary Attacks on Population
28. Exemplary Attacks Against Property
27. Exemplary Attack on Military
26. Demonstration Attack on Zone of Interior

(CENTRAL SANCTUARY THRESHOLD)

BIZARRE CRISES
25. Evacuation (Approximately 70 per cent)
24. Unusual, Provocative, and Significant Countermeasures
23. Local Nuclear War—Military
22. Declaration of Limited Nuclear War
21. Local Nuclear War—Exemplary

(NO NUCLEAR USE THRESHOLD)

INTENSE CRISES
20. "Peaceful" World-Wide Embargo or Blockade
19. "Justifiable" Counterforce Attack
18. Spectacular Show or Demonstration of Force
17. Limited Evacuation (Approximately 20 per cent)
16. Nuclear "Ultimatums"
15. Barely Nuclear War
14. Declaration of Limited Conventional War
13. Large Compound Escalation
12. Large Conventional War (or Actions)
11. Super-Ready Status
10. Provocative Breaking Off of Diplomatic Relations

(NUCLEAR WAR IS UNTHINKABLE THRESHOLD)

TRADITIONAL CRISES
9. Dramatic Military Confrontations
8. Harassing Acts of Violence
7. "Legal" Harassment—Retortions
6. Significant Mobilization
5. Show of Force
4. Hardening of Positions—Confrontation of Wills

(DON'T ROCK THE BOAT THRESHOLD)

SUBCRISIS MANEUVERING
3. Solemn and Formal Declarations
2. Political, Economic, and Diplomatic Gestures
1. Ostensible Crisis

DISAGREEMENT—COLD WAR

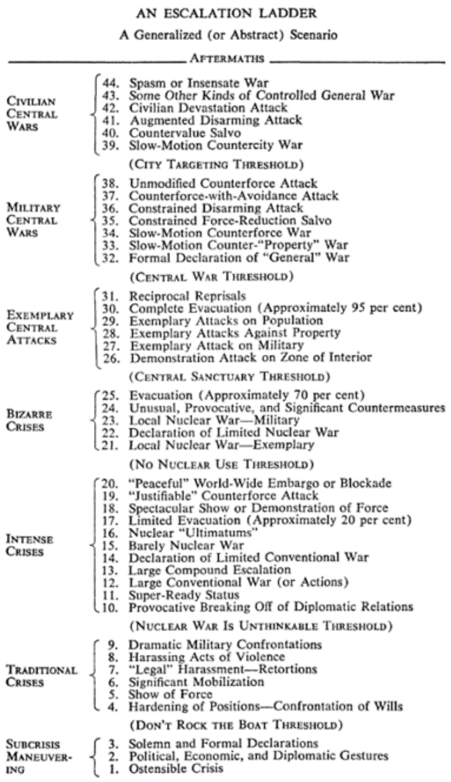
Ordered transitions

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**
(row-stochastic)

Ladder of escalation: *Allies do not become enemies overnight*



AN ESCALATION LADDER
A Generalized (or Abstract) Scenario

AFTERMATHS

CIVILIAN CENTRAL WARS
44. Spasm or Insensate War
43. Some Other Kinds of Controlled General War
42. Civilian Devastation Attack
41. Augmented Disarming Attack
40. Countervalue Salvo
39. Slow-Motion Countercity War

(CITY TARGETING THRESHOLD)

MILITARY CENTRAL WARS
38. Unmodified Counterforce Attack
37. Counterforce-with-Avoidance Attack
36. Constrained Disarming Attack
35. Constrained Force-Reduction Salvo
34. Slow-Motion Counterforce War
33. Slow-Motion Counter-"Property" War
32. Formal Declaration of "General" War

(CENTRAL WAR THRESHOLD)

EXEMPLARY CENTRAL ATTACKS
31. Reciprocal Reprisals
30. Complete Evacuation (Approximately 95 per cent)
29. Exemplary Attacks on Population
28. Exemplary Attacks Against Property
27. Exemplary Attack on Military
26. Demonstration Attack on Zone of Interior

(CENTRAL SANCTUARY THRESHOLD)

BIZARRE CRISES
25. Evacuation (Approximately 70 per cent)
24. Unusual, Provocative, and Significant Countermeasures
23. Local Nuclear War—Military
22. Declaration of Limited Nuclear War
21. Local Nuclear War—Exemplary

(NO NUCLEAR USE THRESHOLD)

INTENSE CRISES
20. "Peaceful" World-Wide Embargo or Blockade
19. "Justifiable" Counterforce Attack
18. Spectacular Show or Demonstration of Force
17. Limited Evacuation (Approximately 20 per cent)
16. Nuclear "Ultimatums"
15. Barely Nuclear War
14. Declaration of Limited Conventional War
13. Large Compound Escalation
12. Large Conventional War (or Actions)
11. Super-Ready Status
10. Provocative Breaking Off of Diplomatic Relations

(NUCLEAR WAR IS UNTHINKABLE THRESHOLD)

TRADITIONAL CRISES
9. Dramatic Military Confrontations
8. Harassing Acts of Violence
7. "Legal" Harassment—Retortions
6. Significant Mobilization
5. Show of Force
4. Hardening of Positions—Confrontation of Wills

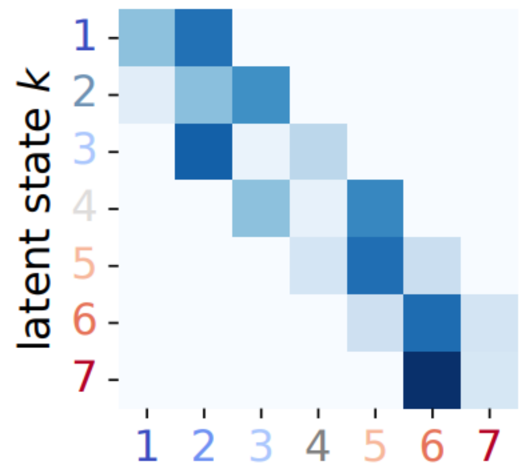
(DON'T ROCK THE BOAT THRESHOLD)

SUBCRISIS MANEUVERING
3. Solemn and Formal Declarations
2. Political, Economic, and Diplomatic Gestures
1. Ostensible Crisis

DISAGREEMENT—COLD WAR

Traditional way to operationalize this:

Banded transition matrix



Ordered transitions

how much **state** k'
transitions to **state** k

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Π is the $K \times K$ **transition matrix**
(row-stochastic)

Ladder of escalation: *Allies do not become enemies overnight*

AN ESCALATION LADDER

A Generalized (or Abstract) Scenario

AFTERMATHS

CIVILIAN CENTRAL WARS
- 44. Spasm or Insensate War
- 43. Some Other Kinds of Controlled General War
- 42. Civilian Devastation Attack
- 41. Augmented Disarming Attack
- 40. Countervalue Salvo
- 39. Slow-Motion Countercity War

(CITY TARGETING THRESHOLD)

MILITARY CENTRAL WARS
- 38. Unmodified Counterforce Attack
- 37. Counterforce-with-Avoidance Attack
- 36. Constrained Disarming Attack
- 35. Constrained Force-Reduction Salvo
- 34. Slow-Motion Counterforce War
- 33. Slow-Motion Counter-"Property" War
- 32. Formal Declaration of "General" War

(CENTRAL WAR THRESHOLD)

EXEMPLARY CENTRAL ATTACKS
- 31. Reciprocal Reprisals
- 30. Complete Evacuation (Approximately 95 per cent)
- 29. Exemplary Attacks on Population
- 28. Exemplary Attacks Against Property
- 27. Exemplary Attack on Military
- 26. Demonstration Attack on Zone of Interior

(CENTRAL SANCTUARY THRESHOLD)

BIZARRE CRISES
- 25. Evacuation (Approximately 70 per cent)
- 24. Unusual, Provocative, and Significant Countermeasures
- 23. Local Nuclear War—Military
- 22. Declaration of Limited Nuclear War
- 21. Local Nuclear War—Exemplary

(NO NUCLEAR USE THRESHOLD)

INTENSE CRISES
- 20. "Peaceful" World-Wide Embargo or Blockade
- 19. "Justifiable" Counterforce Attack
- 18. Spectacular Show or Demonstration of Force
- 17. Limited Evacuation (Approximately 20 per cent)
- 16. Nuclear "Ultimatums"
- 15. Barely Nuclear War
- 14. Declaration of Limited Conventional War
- 13. Large Compound Escalation
- 12. Large Conventional War (or Actions)
- 11. Super-Ready Status
- 10. Provocative Breaking Off of Diplomatic Relations

(NUCLEAR WAR IS UNTHINKABLE THRESHOLD)

TRADITIONAL CRISES
- 9. Dramatic Military Confrontations
- 8. Harassing Acts of Violence
- 7. "Legal" Harassment—Retortions
- 6. Significant Mobilization
- 5. Show of Force
- 4. Hardening of Positions—Confrontation of Wills

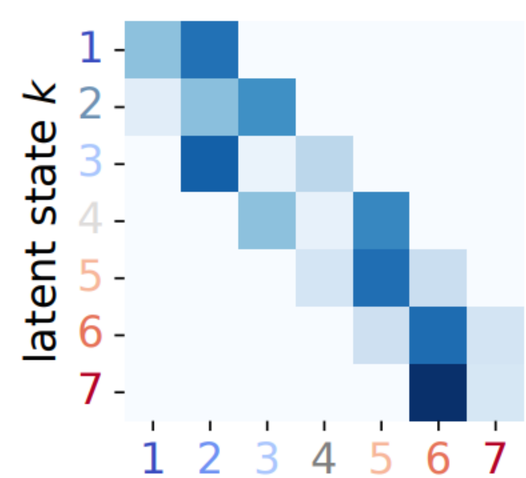
(DON'T ROCK THE BOAT THRESHOLD)

SUBCRISIS MANEUVERING
- 3. Solemn and Formal Declarations
- 2. Political, Economic, and Diplomatic Gestures
- 1. Ostensible Crisis

DISAGREEMENT—COLD WAR

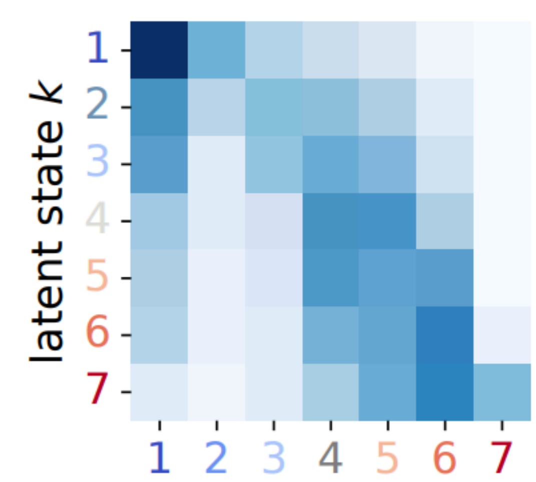
Traditional way to operationalize this:

Banded transition matrix

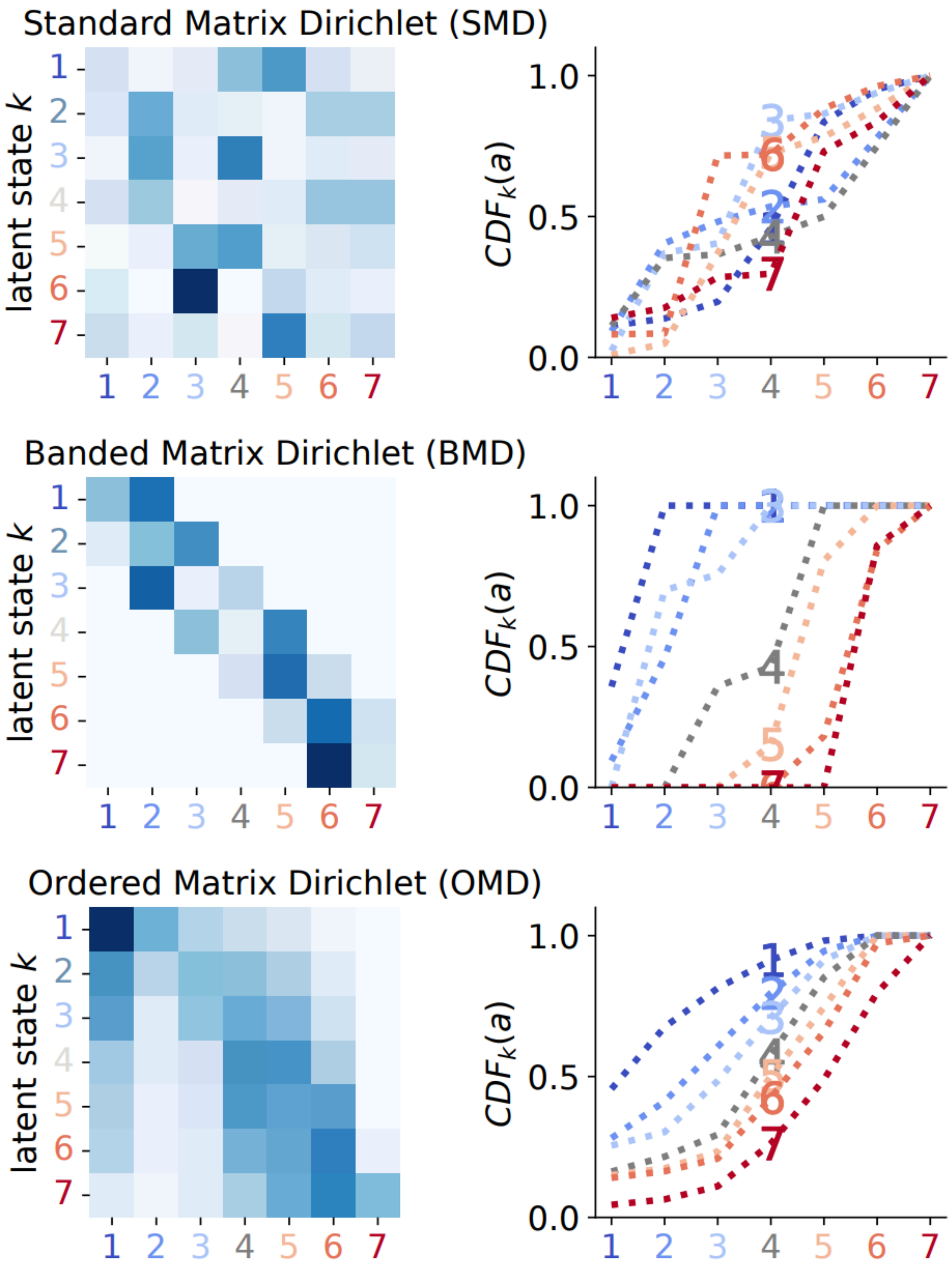


New way: $\Pi \sim \mathrm{OMD}(\boldsymbol{\alpha})$

OMD transition matrix



Ordered transitions



Putting it all together: **full model**

well-orderd **emission** matrix

$$\Phi \sim \text{OMD}\left(\boldsymbol{\alpha}\right)$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j} \phi_{ka}$$

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k} \, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \text{OMD}\,(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

One extra piece…

$$\lambda^{(t)}_{i \xrightarrow{k} j} \equiv \sum_{c=1}^{C} \sum_{d=1}^{C} \psi_{ic}\, \psi_{jd}\, \lambda^{(t)}_{c \xrightarrow{k} d}$$

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j} \, \phi_{ka}$$

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k} \, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

One extra piece…

$$\lambda^{(t)}_{i \xrightarrow{k} j} \equiv \sum_{c=1}^{C} \sum_{d=1}^{C} \psi_{ic} \, \psi_{jd} \, \lambda^{(t)}_{c \xrightarrow{k} d}$$

how much **countries** i and j
participate in **communities** c and d

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \lambda^{(t)}_{i \xrightarrow{k} j}\, \phi_{ka}$$

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[\lambda^{(t)}_{i \xrightarrow{k} j}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{i \xrightarrow{k'} j}$$

One extra piece…

$$\lambda^{(t)}_{i \xrightarrow{k} j} \equiv \sum_{c=1}^{C} \sum_{d=1}^{C} \psi_{ic}\, \psi_{jd}\, \lambda^{(t)}_{c \xrightarrow{k} d}$$

how much **countries** i and j participate in **communities** c and d

how much **dyad** $c \rightarrow d$ is in **state** k at **time** t

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \mathrm{OMD}\,(\boldsymbol{\alpha})$$

well-ordered **transition** matrix

$$\Pi \sim \mathrm{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i\overset{a}{\to}j}\right] = \sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{d=1}^{C} \psi_{ic}\,\psi_{jd}\,\lambda^{(t)}_{c\overset{k}{\to}d}\,\phi_{ka}$$

$$\mathbb{E}\left[\lambda^{(t)}_{c\overset{k}{\to}d}\right] = \sum_{k'=1}^{K} \pi_{k'k}\,\lambda^{(t-1)}_{c\overset{k'}{\to}d}$$

One extra piece…

$$\lambda^{(t)}_{i\overset{k}{\to}j} \equiv \sum_{c=1}^{C}\sum_{d=1}^{C} \psi_{ic}\,\psi_{jd}\,\lambda^{(t)}_{c\overset{k}{\to}d}$$

how much **countries** i and j
participate in **communities** c and d

how much **dyad** $c \to d$
is in **state** k at **time** t

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{d=1}^{C} \psi_{ic}\, \psi_{jd}\, \lambda^{(t)}_{c \xrightarrow{k} d}\, \phi_{ka}$$

Tucker decomposition

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[\lambda^{(t)}_{c \xrightarrow{k} d}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{c \xrightarrow{k'} d}$$

state-space model over the **core tensor**

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{d=1}^{C} \psi_{ic}\,\psi_{jd}\,\lambda^{(t)}_{c \xrightarrow{k} d}\,\phi_{ka}$$

Tucker decomposition

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{d=1}^{C} \psi_{ic}\,\psi_{jd}\,\lambda^{(t)}_{c \xrightarrow{k} d}\,\phi_{ka}\right)$$

Poisson Tucker decomposition [ICML '16]

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[\lambda^{(t)}_{c \xrightarrow{k} d}\right] = \sum_{k'=1}^{K} \pi_{k'k}\,\lambda^{(t-1)}_{c \xrightarrow{k'} d}$$

state-space model over the **core tensor**

$$\lambda^{(t)}_{c \xrightarrow{k} d} \sim \text{Gamma}\left(\tau_0 \sum_{k'=1}^{K} \pi_{k'k}\,\lambda^{(t-1)}_{c \xrightarrow{k'} d},\ \tau_0\right)$$

Poisson-Gamma dynamical system [NeurIPS '16]

Putting it all together: **full model**

well-ordered **emission** matrix

$$\Phi \sim \text{OMD}(\boldsymbol{\alpha})$$

$$\mathbb{E}\left[y^{(t)}_{i \xrightarrow{a} j}\right] = \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{d=1}^{C} \psi_{ic}\, \psi_{jd}\, \lambda^{(t)}_{c \xrightarrow{k} d}\, \phi_{ka}$$

Tucker decomposition

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{d=1}^{C} \psi_{ic}\, \psi_{jd}\, \lambda^{(t)}_{c \xrightarrow{k} d}\, \phi_{ka} \right)$$

Poisson Tucker decomposition
[ICML '16]

well-ordered **transition** matrix

$$\Pi \sim \text{OMD}(\boldsymbol{\alpha})$$

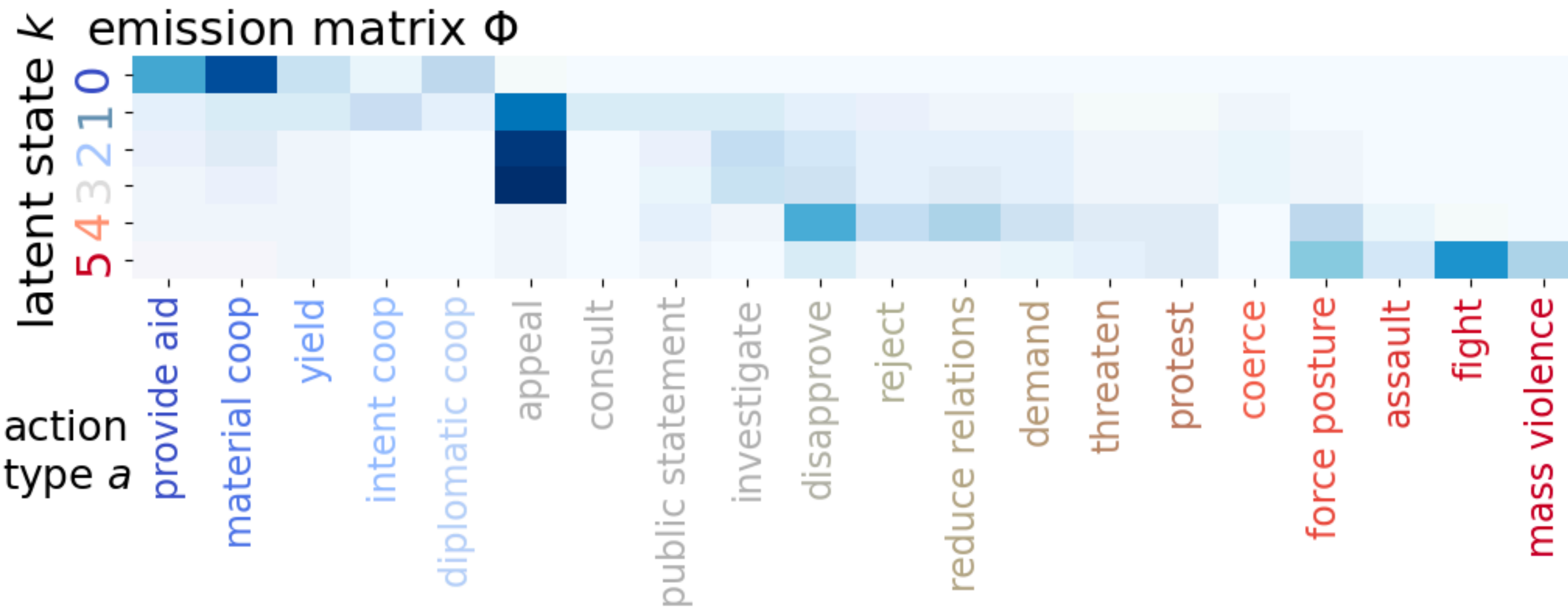
$$\mathbb{E}\left[\lambda^{(t)}_{c \xrightarrow{k} d}\right] = \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{c \xrightarrow{k'} d}$$

state-space model over
the **core tensor**

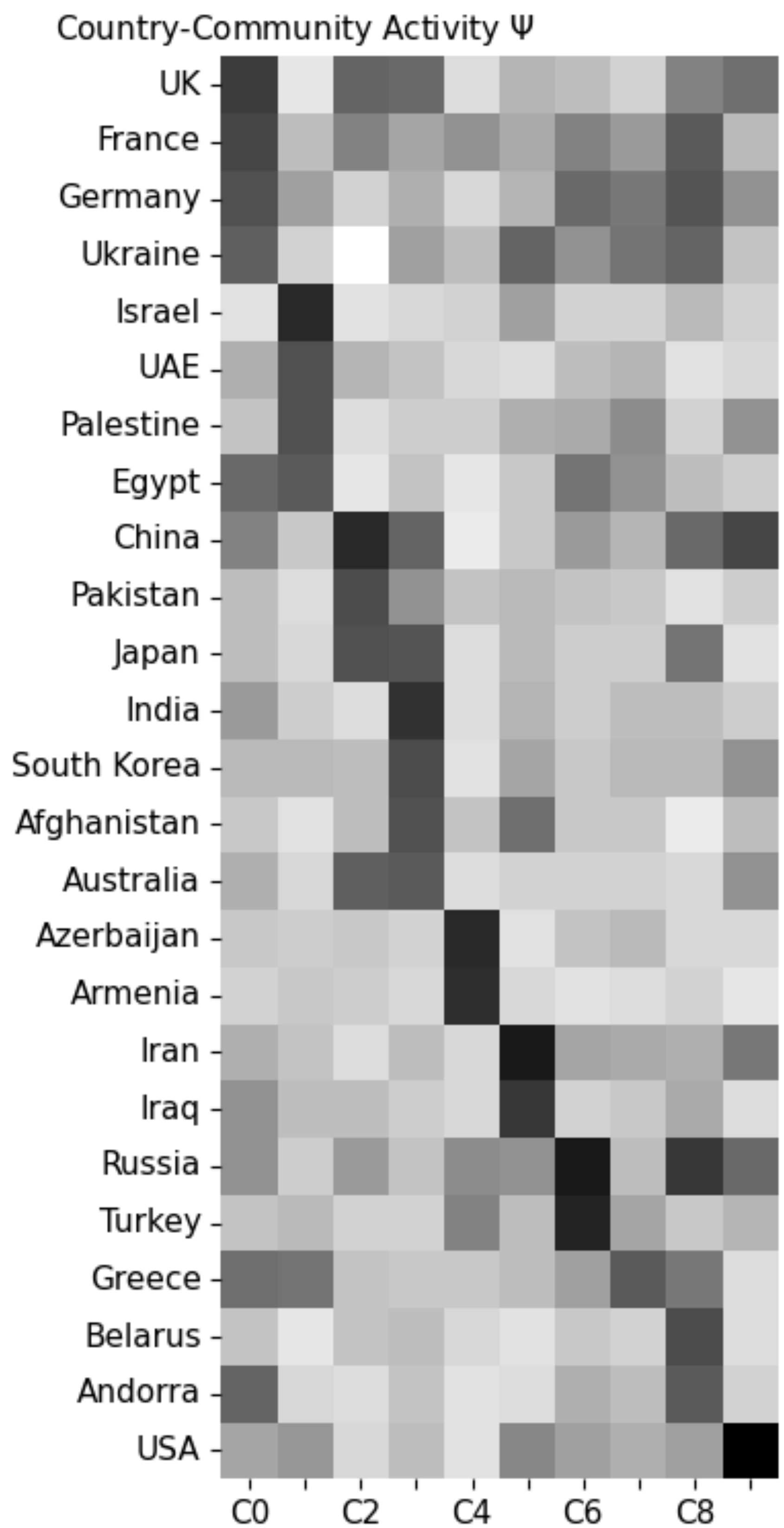
$$\lambda^{(t)}_{c \xrightarrow{k} d} \sim \text{Gamma}\left(\tau_0 \sum_{k'=1}^{K} \pi_{k'k}\, \lambda^{(t-1)}_{c \xrightarrow{k'} d},\ \tau_0 \right)$$

Poisson-Gamma dynamical system
[NeurIPS '16]

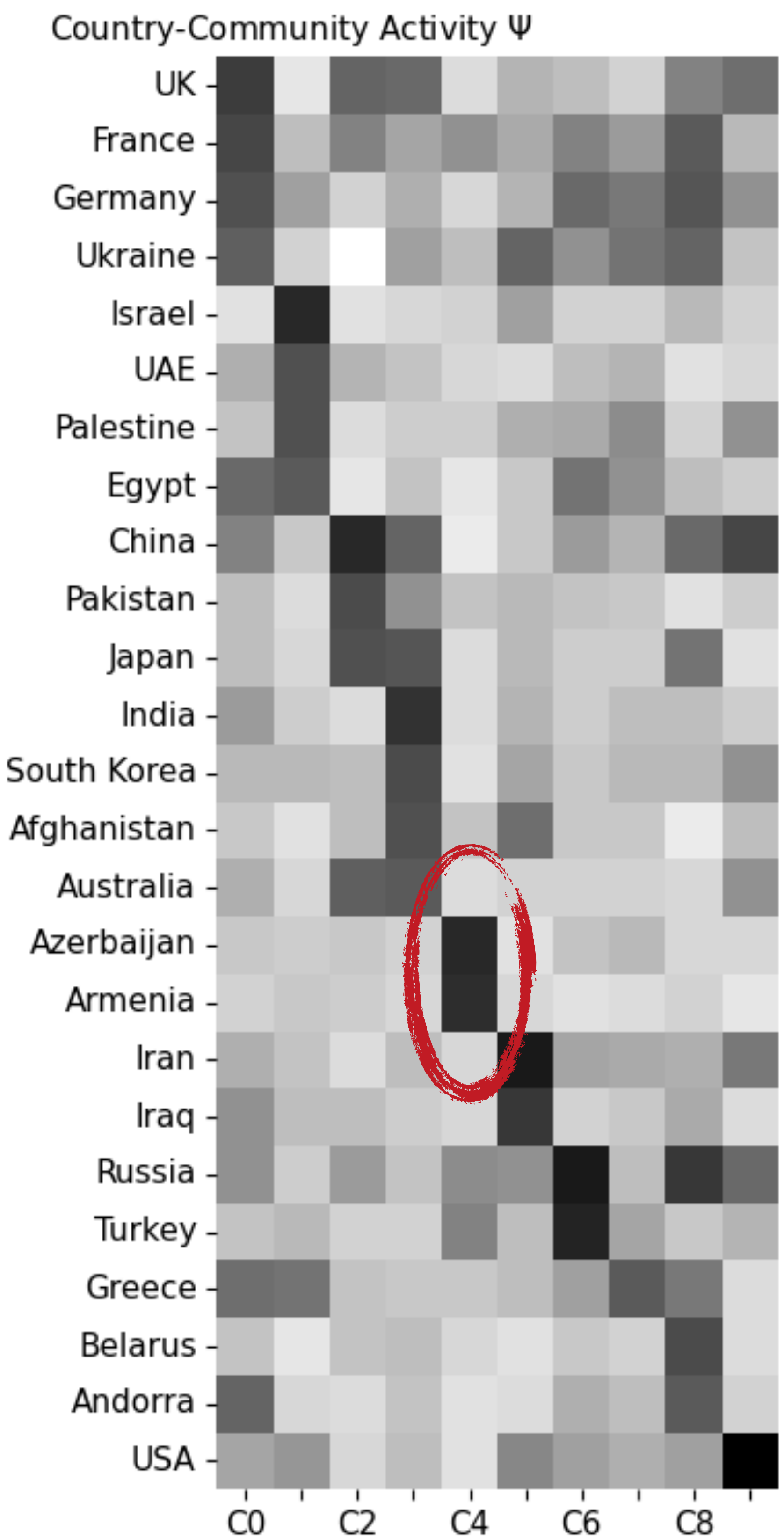
Results: ICEWS 2020 data

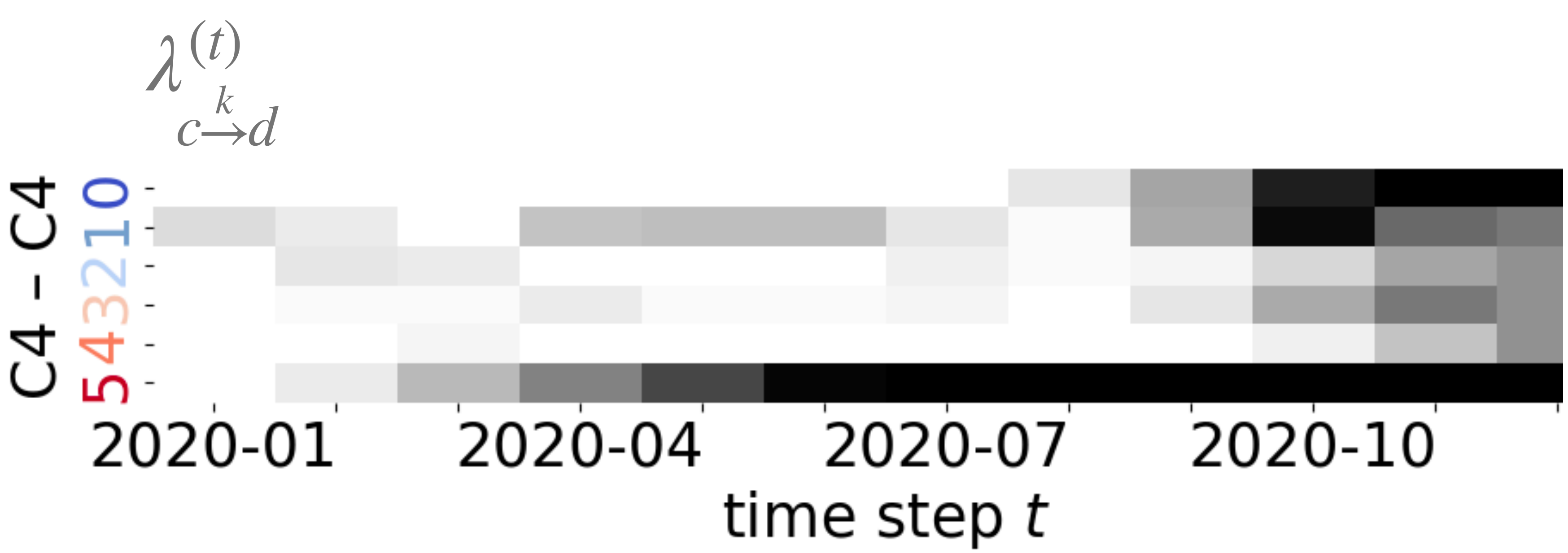


Results: ICEWS 2020 data

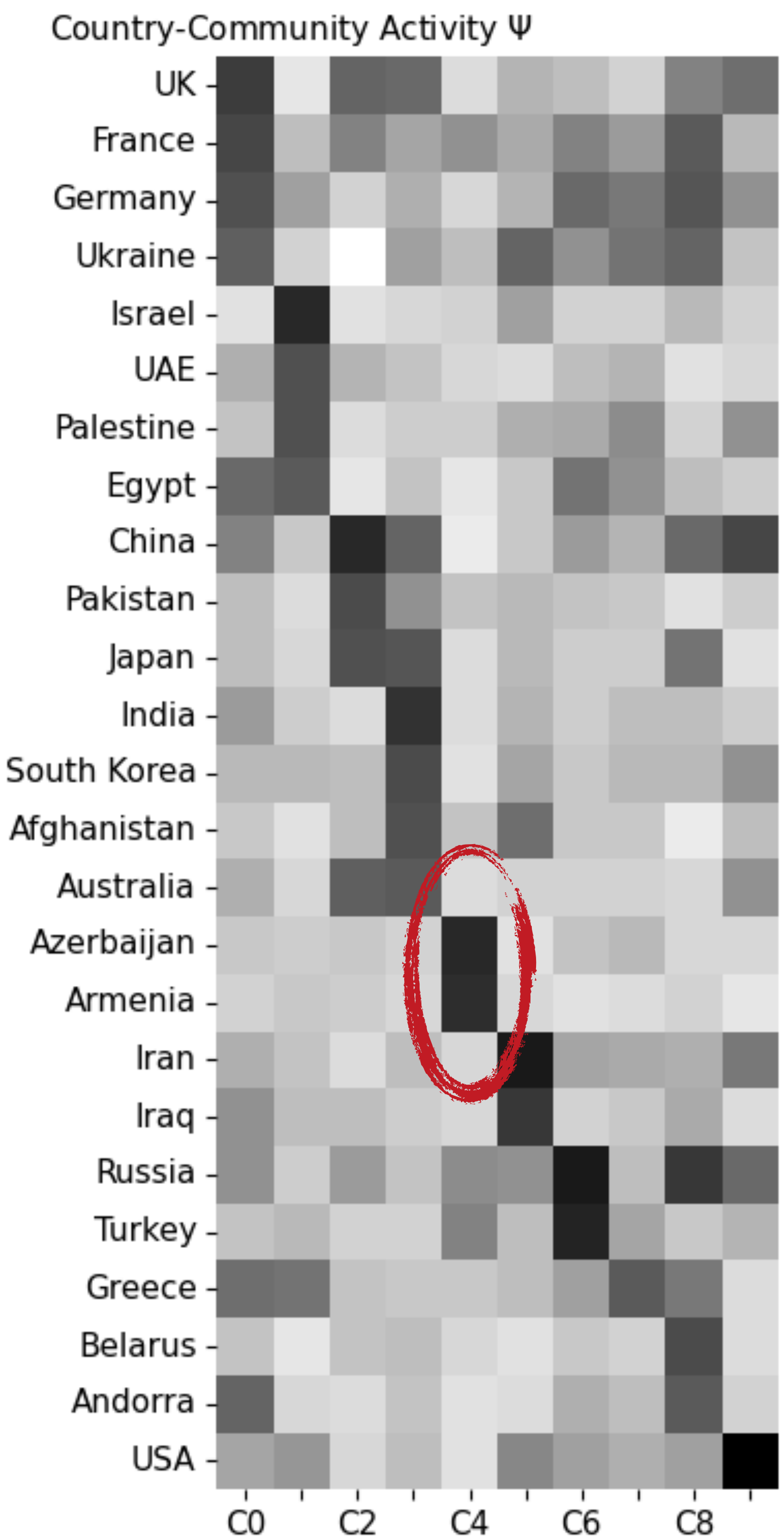


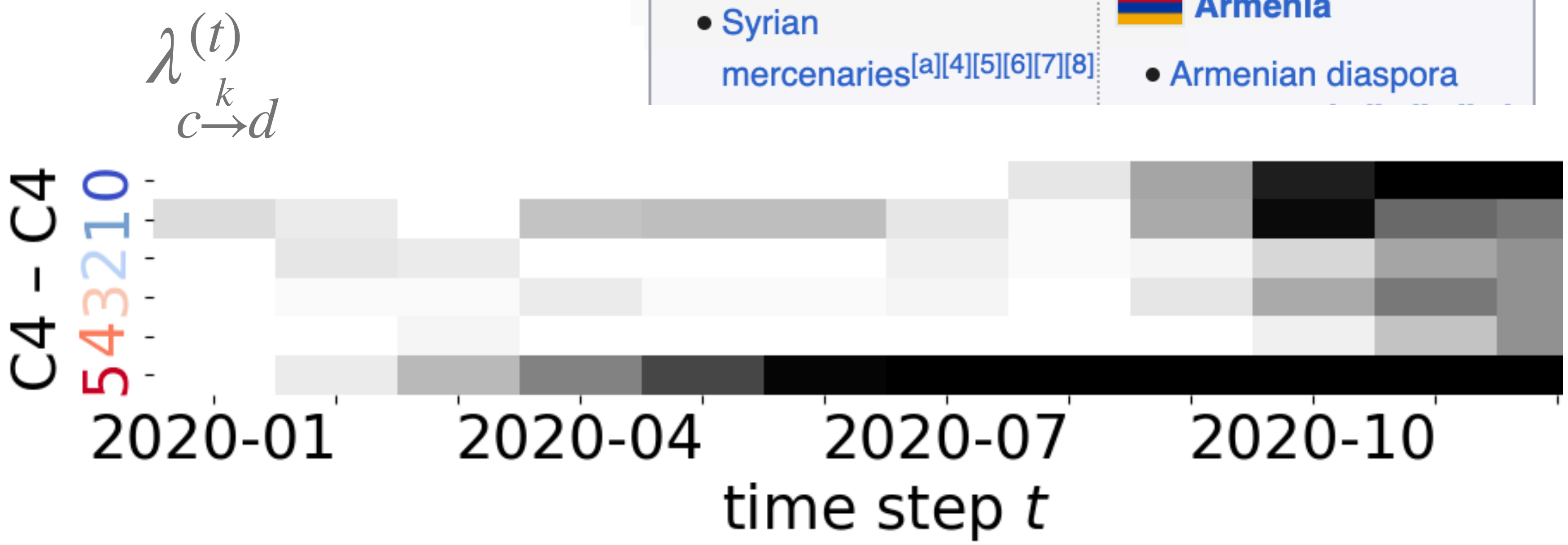
Results: ICEWS 2020 data



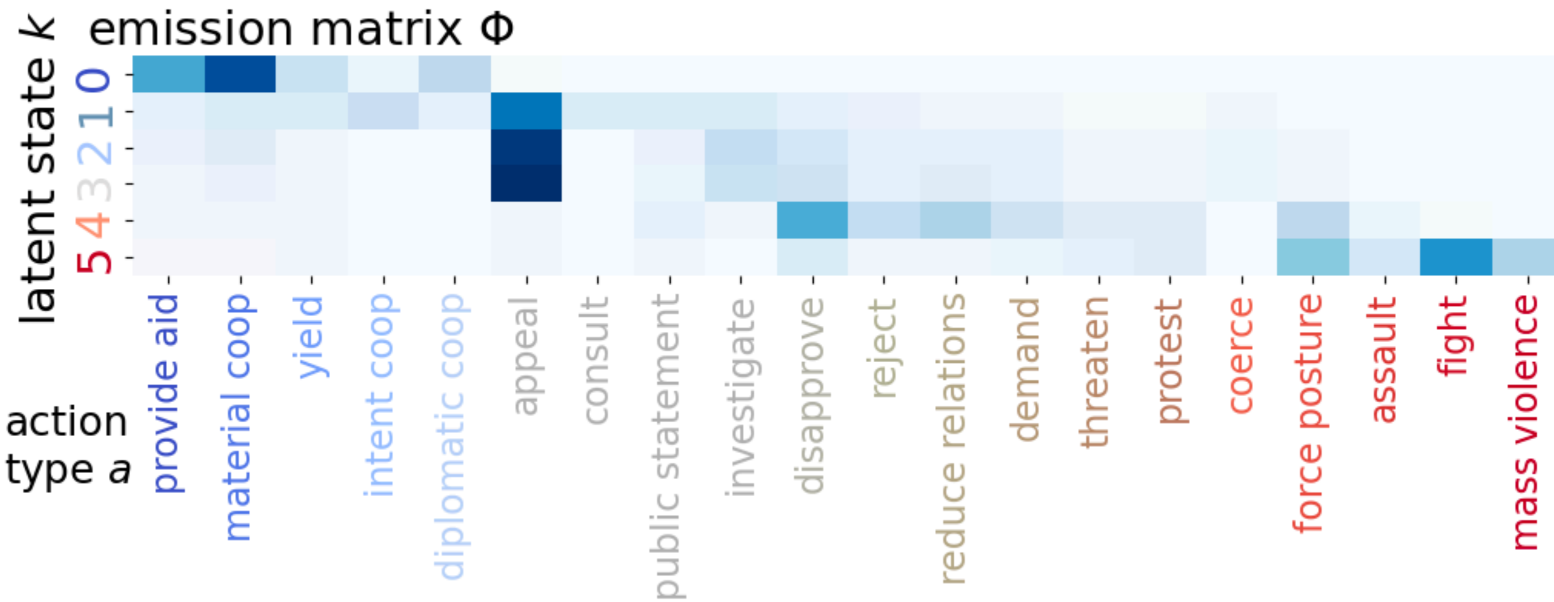


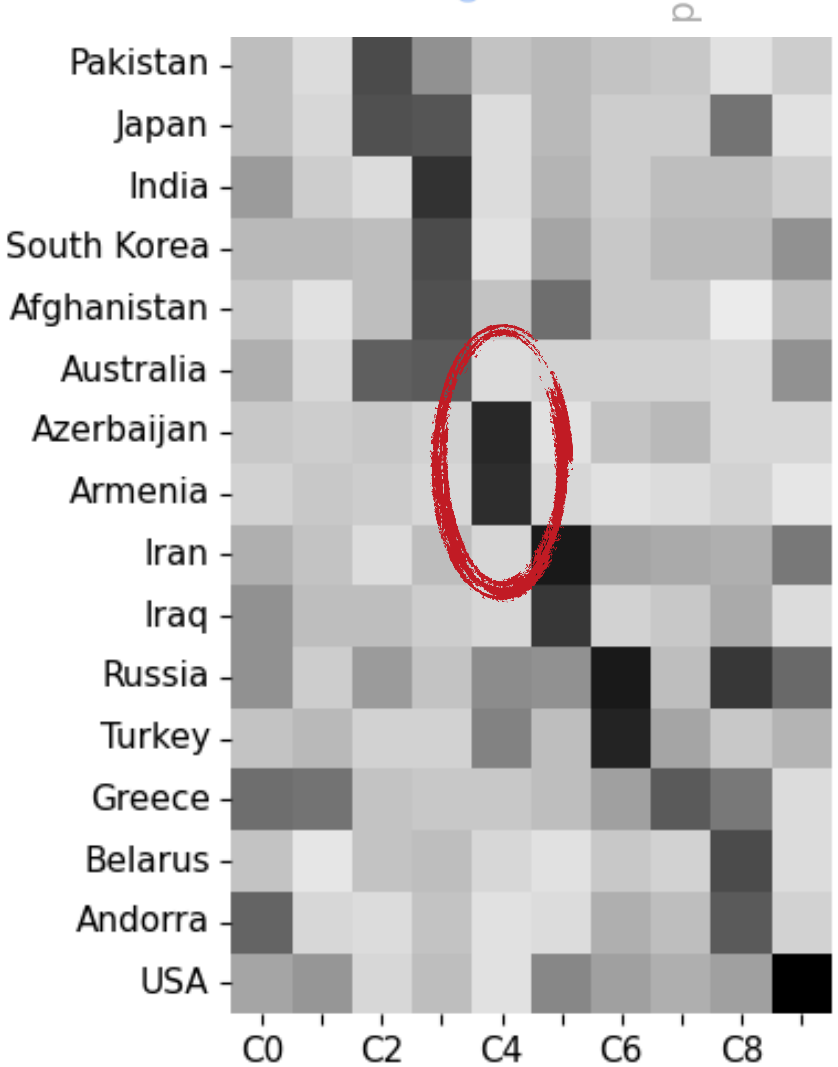
Results: ICEWS 2020 data

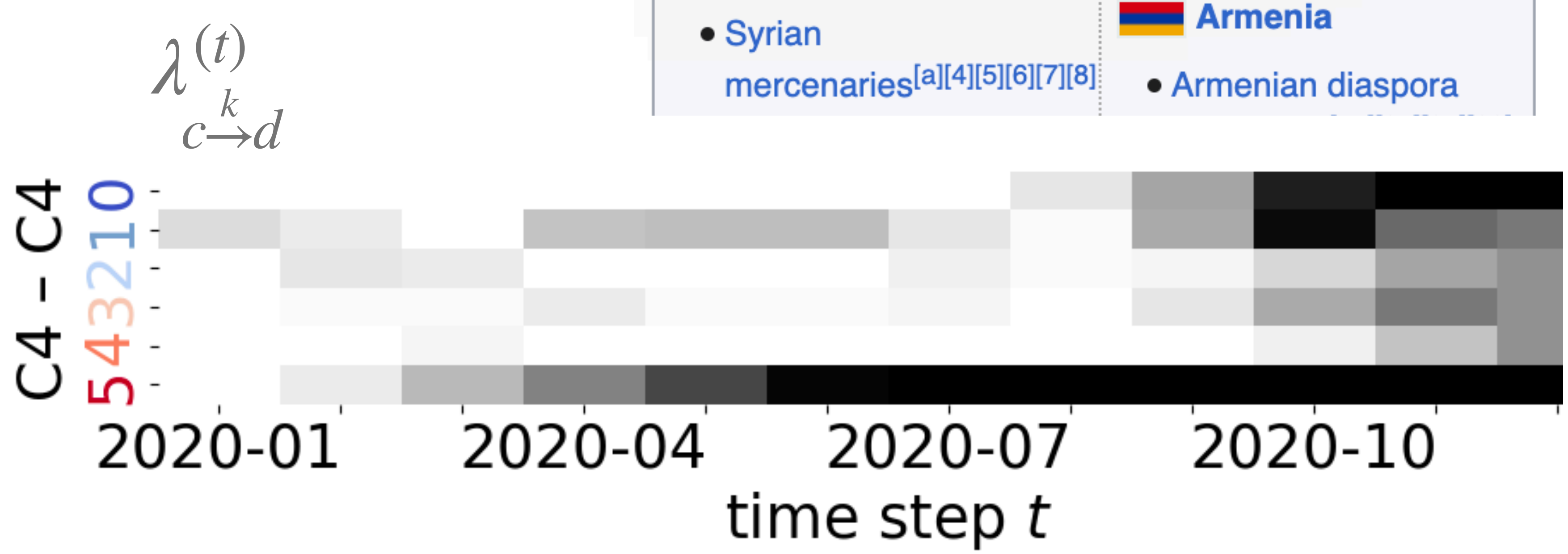




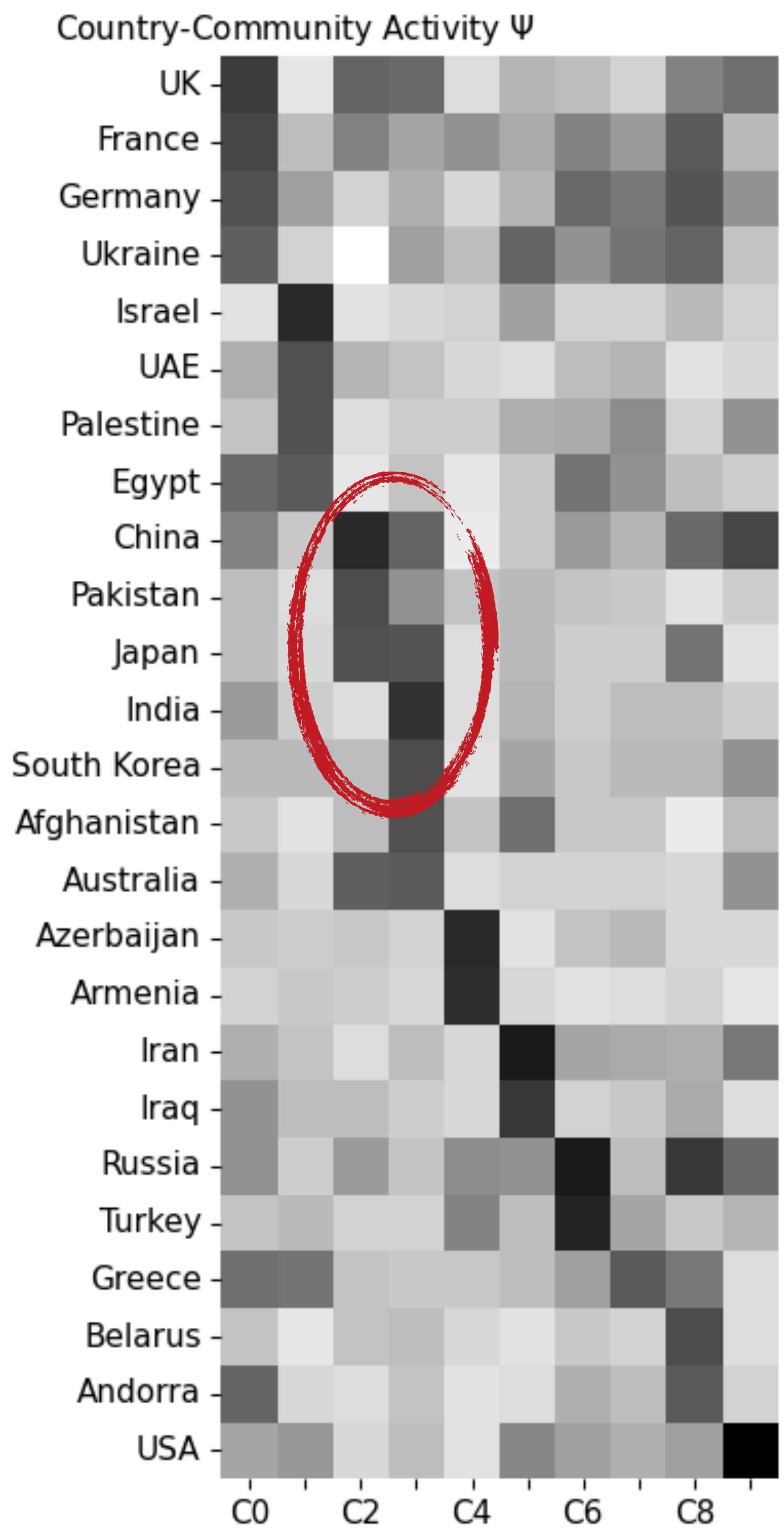
Results: ICEWS 2020 data

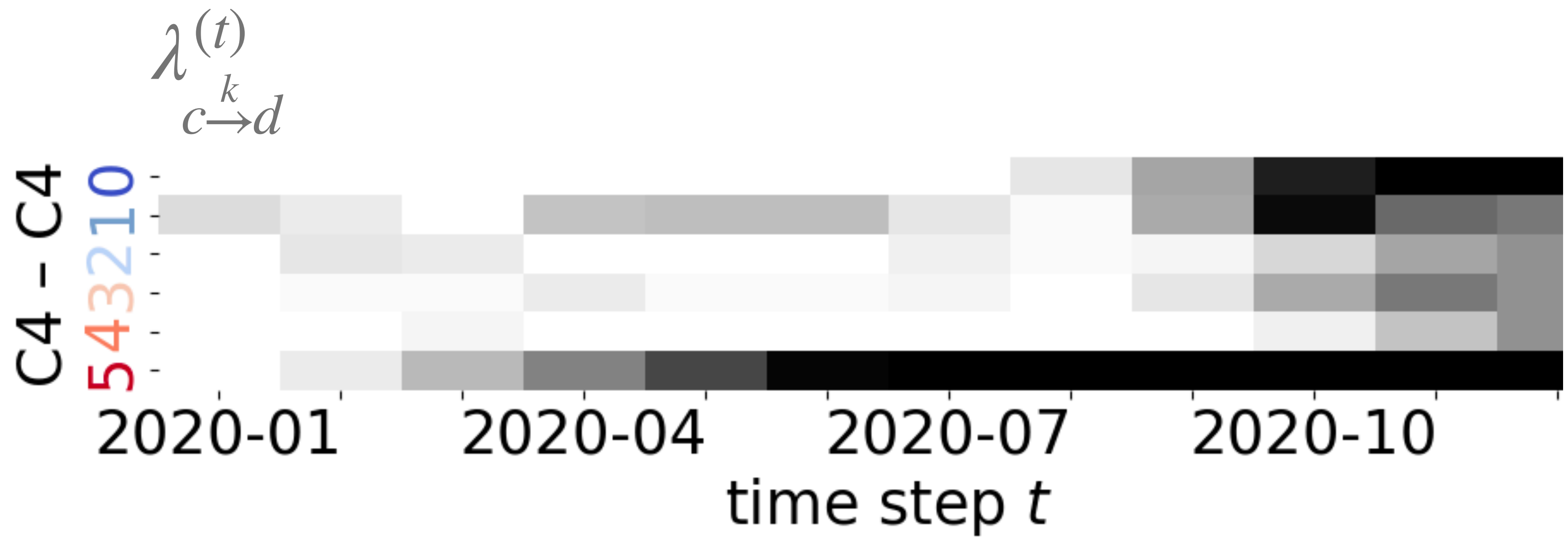




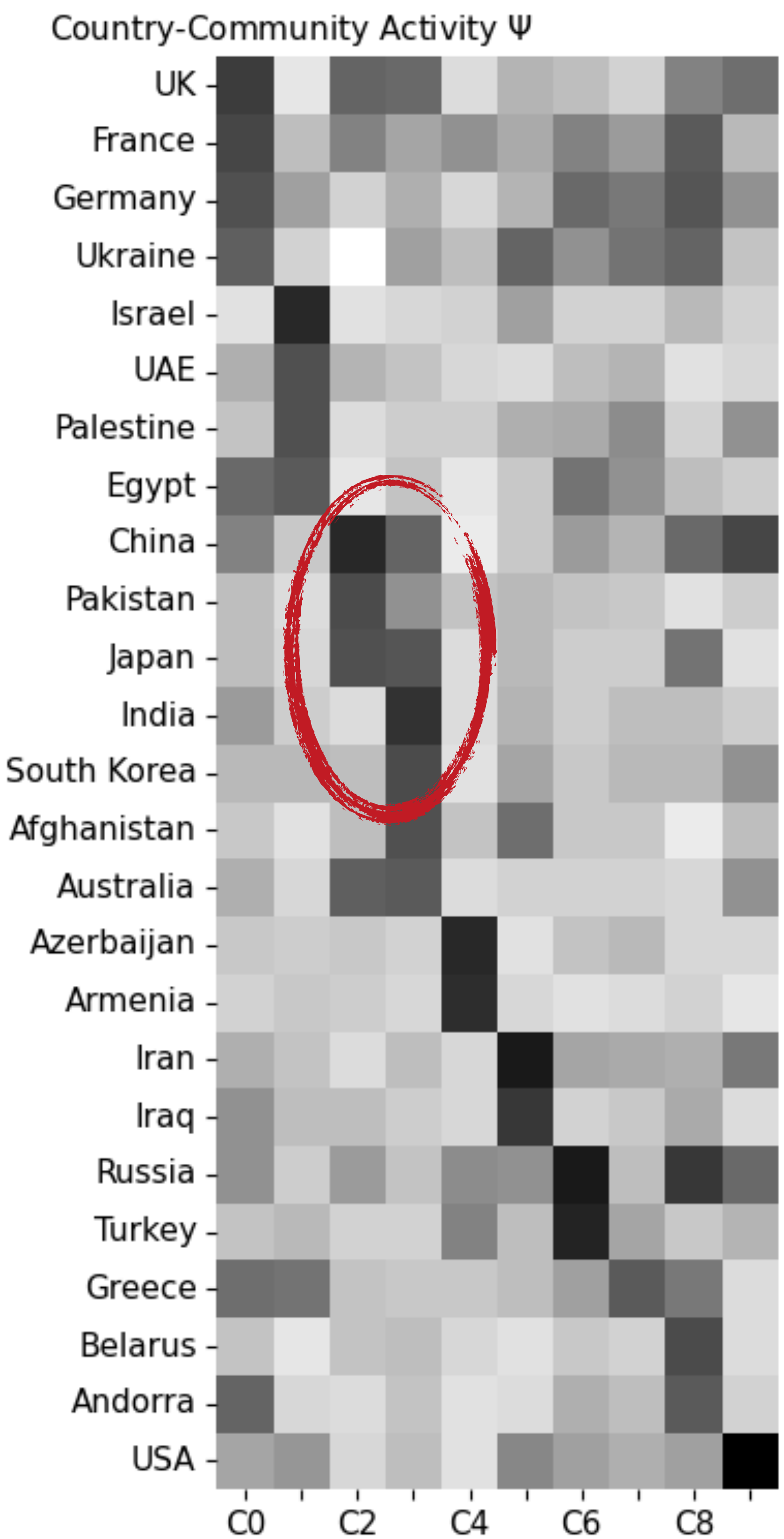


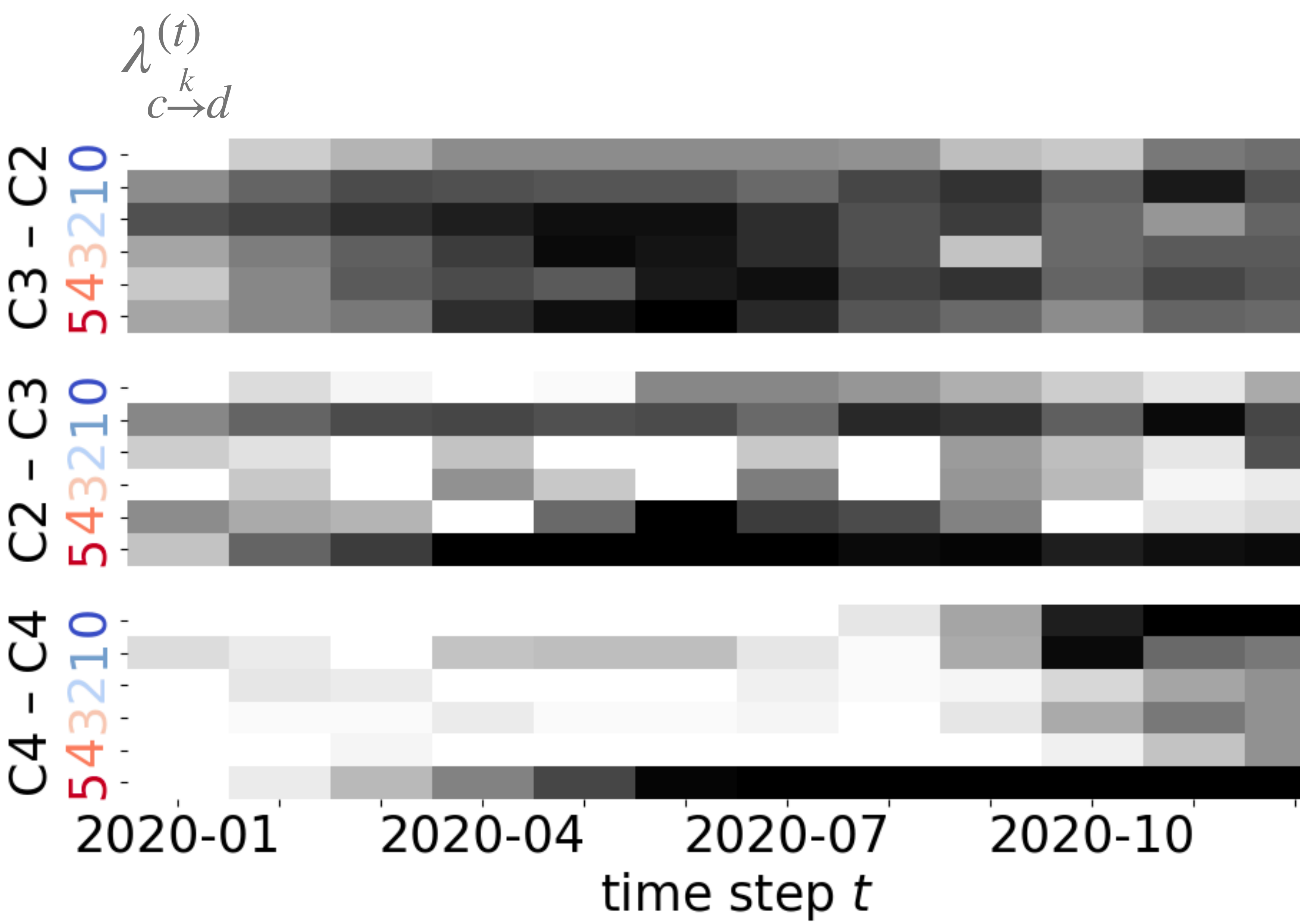
Results: ICEWS 2020 data



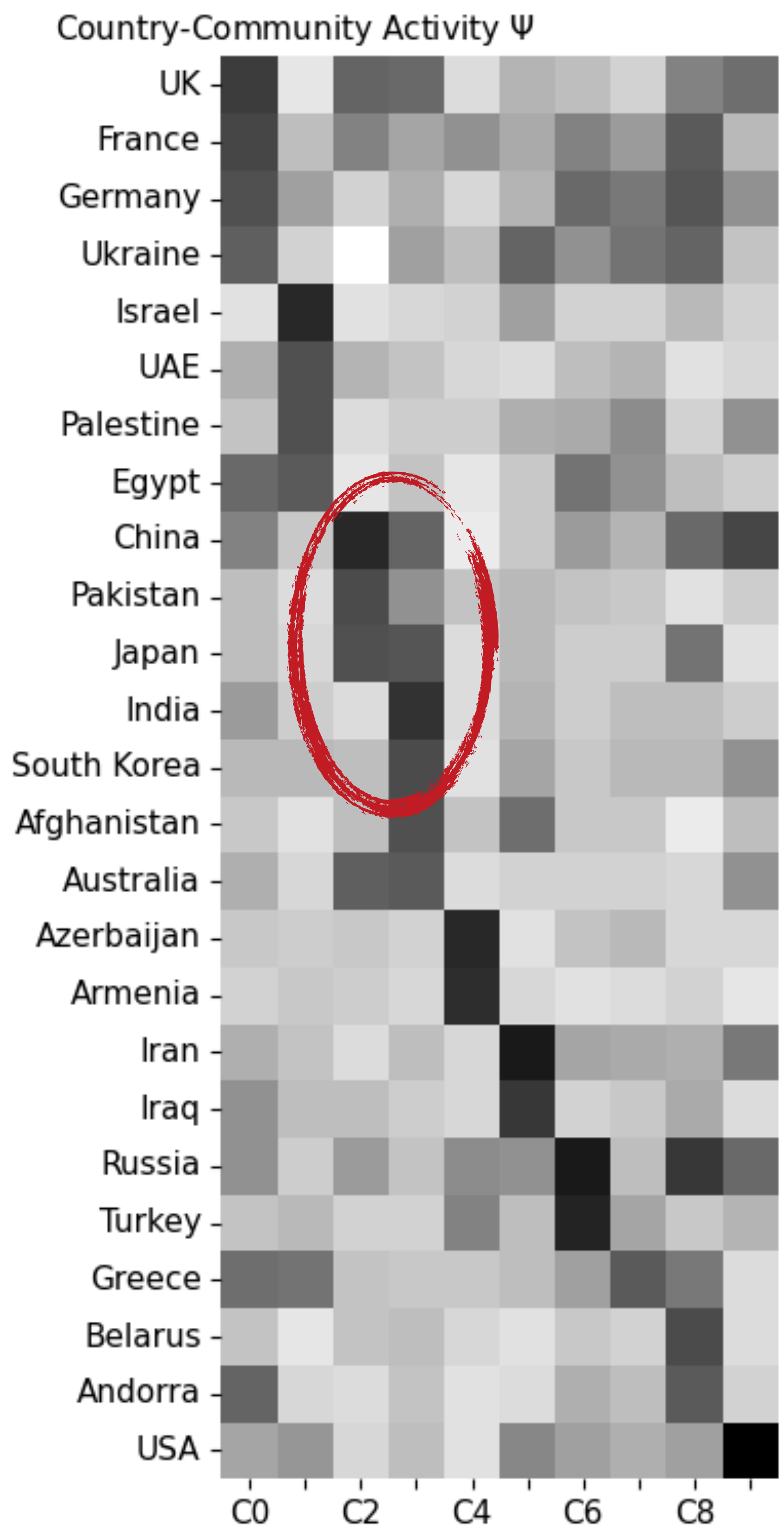


Results: ICEWS 2020 data





Results: ICEWS



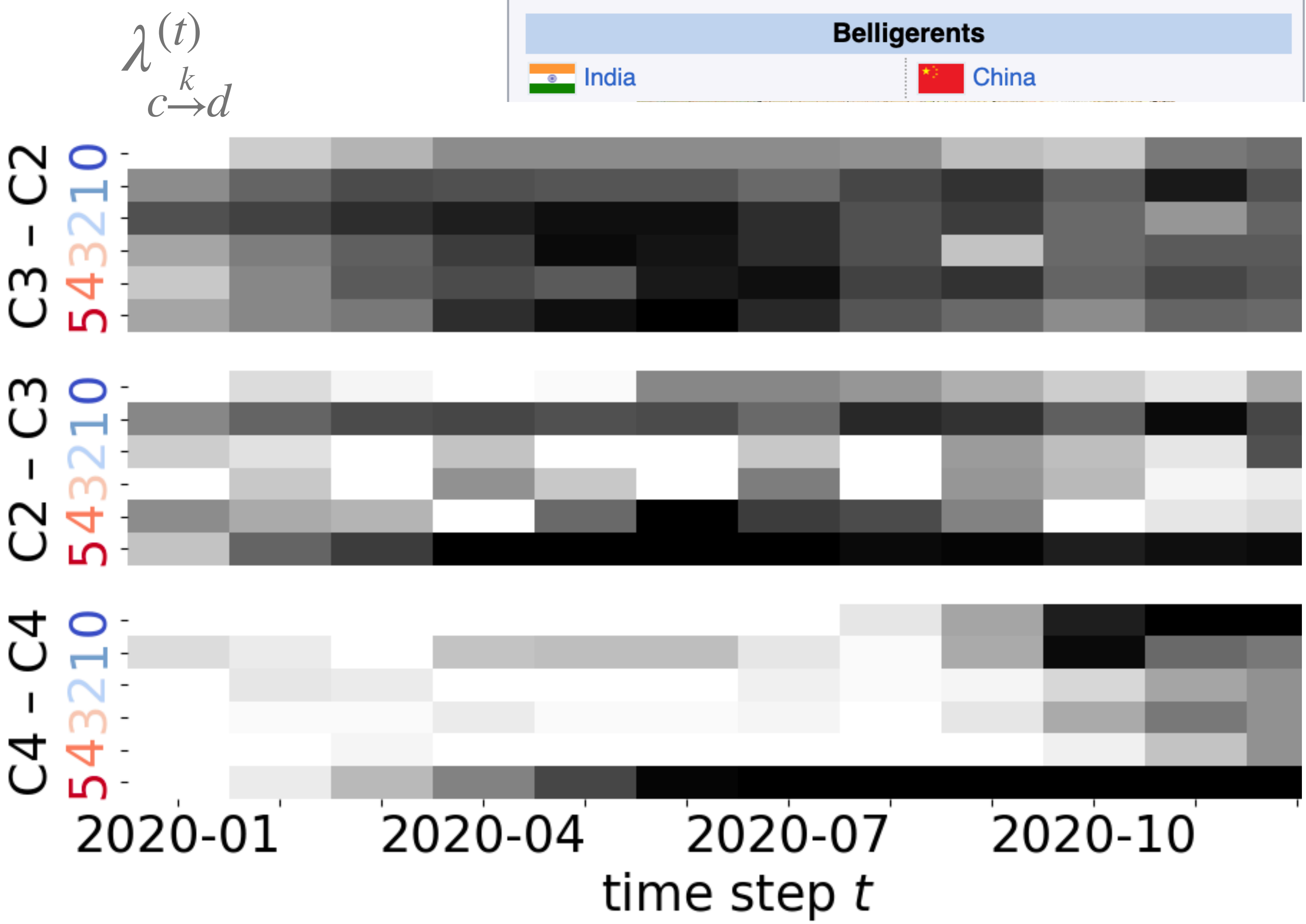
2020–2021 China–India skirmishes

Part of the Sino-Indian border dispute

Date	5 May 2020 – 20 January 2021 (8 months, 2 weeks and 1 day)
Location	Line of Actual Control (LAC), Sino-Indian border
Status	• Change in status quo of ground positions[1][2][3][4][5][6] • Disengagement and de-escalation in progress[7][8][9]

Belligerents

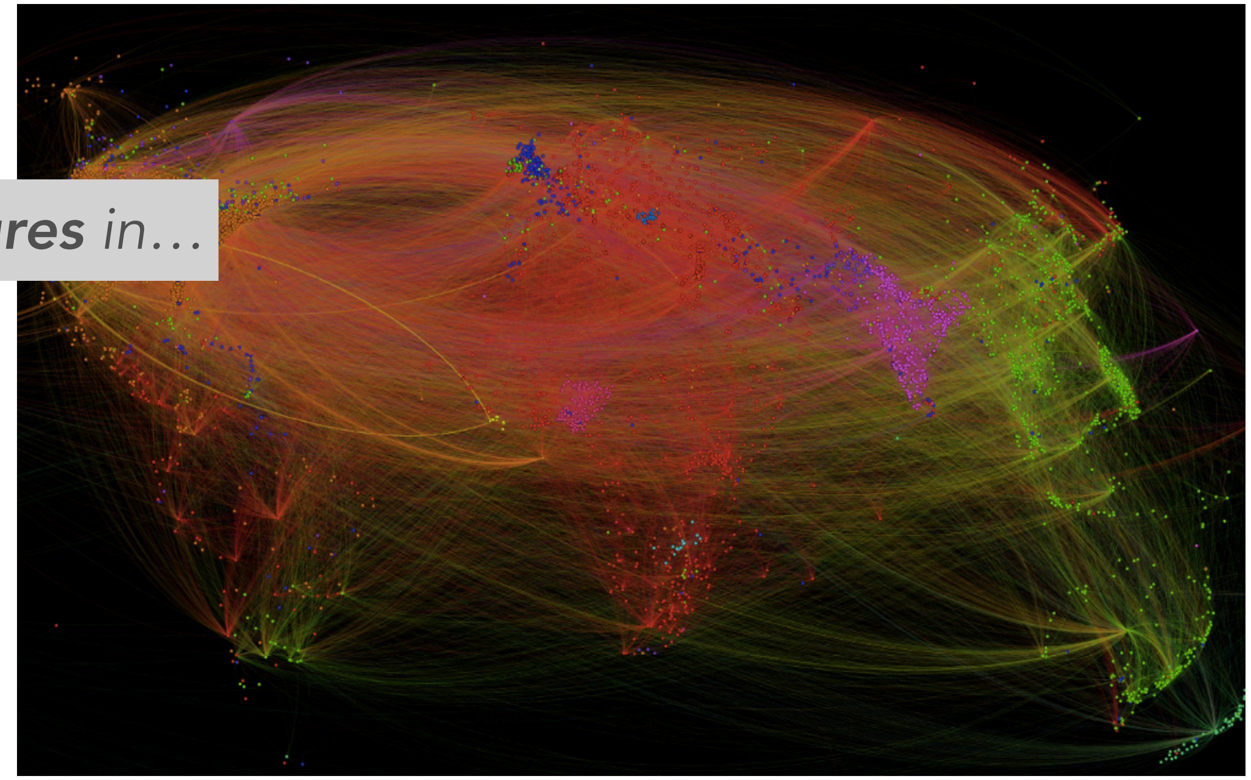
India	China



Summary and conclusion

Mission:

*Models to **measure complex dependence structures** in…*



…**high-dimensional discrete data**.

(sparse, bursty, overdispersed)

Requires:

- ***Blending** methodological frameworks:*
  - *Contingency table analysis*
  - *Tensor decomposition*
  - *Bayesian (ad)mixture / block modeling*
- ***New modeling motifs** for discrete and non-negative distributions*
- ***New techniques, theorems, tricks** for inference*

Data like this is **everywhere** e.g.,

- Genomics
- Neuroscience
- Demography

Broader data science goal:

*Develop the **toolkit** for **measuring scientifically-interesting complex structure** in ^^*