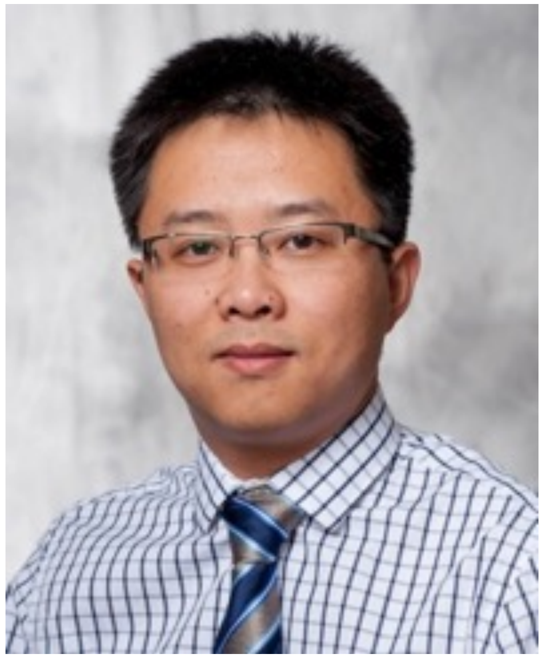
Poisson—Gamma Dynamical Systems

Aaron Schein
UMass Amherst

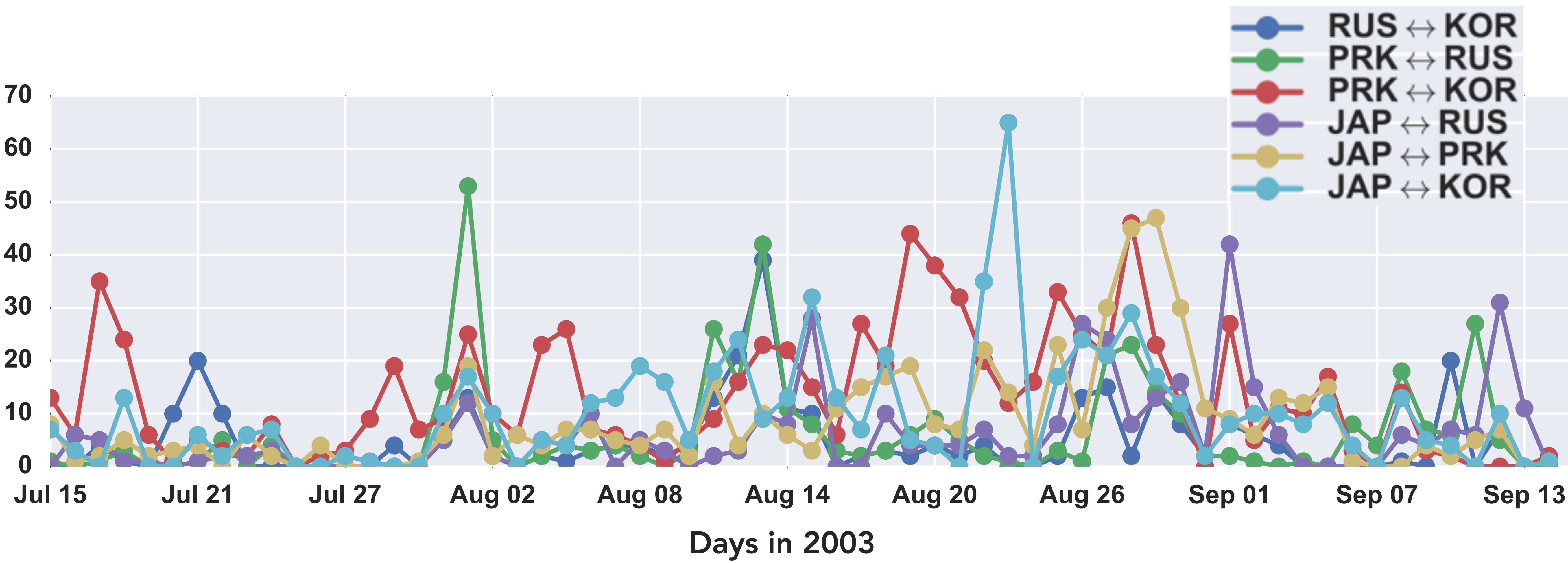
Joint work with:



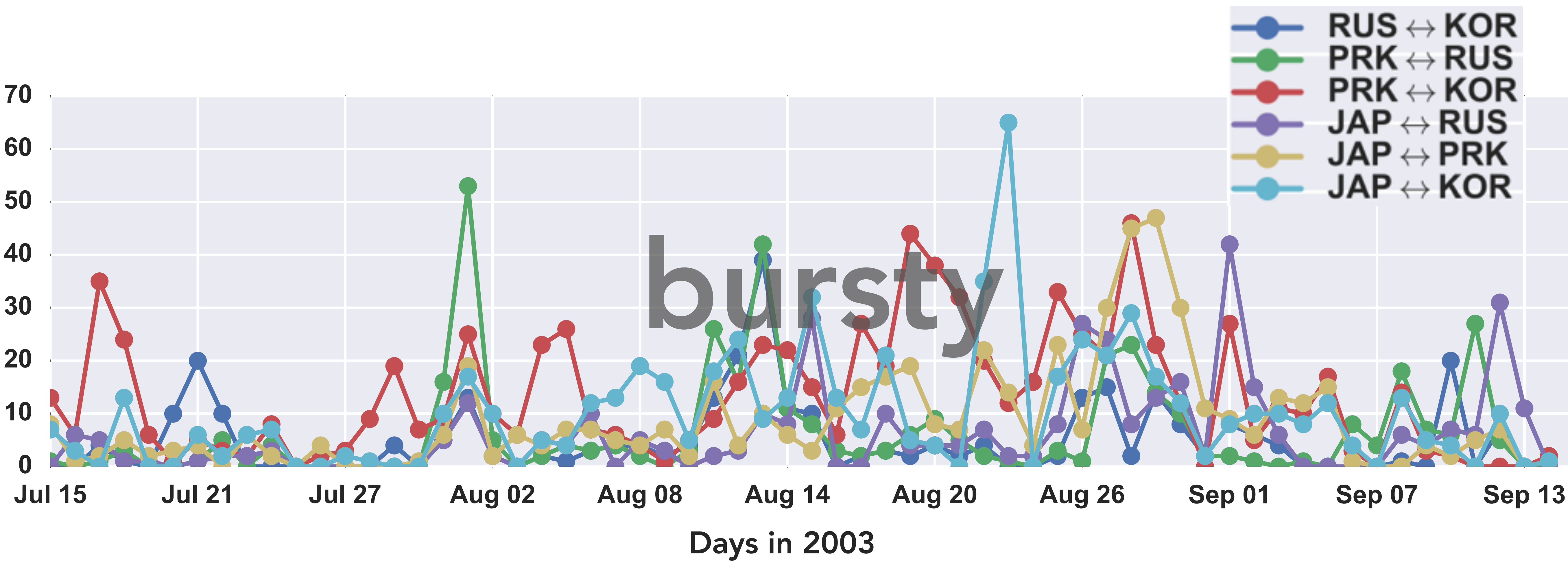
Mingyuan Zhou
Univ. Texas at Austin

Hanna Wallach
Microsoft Research

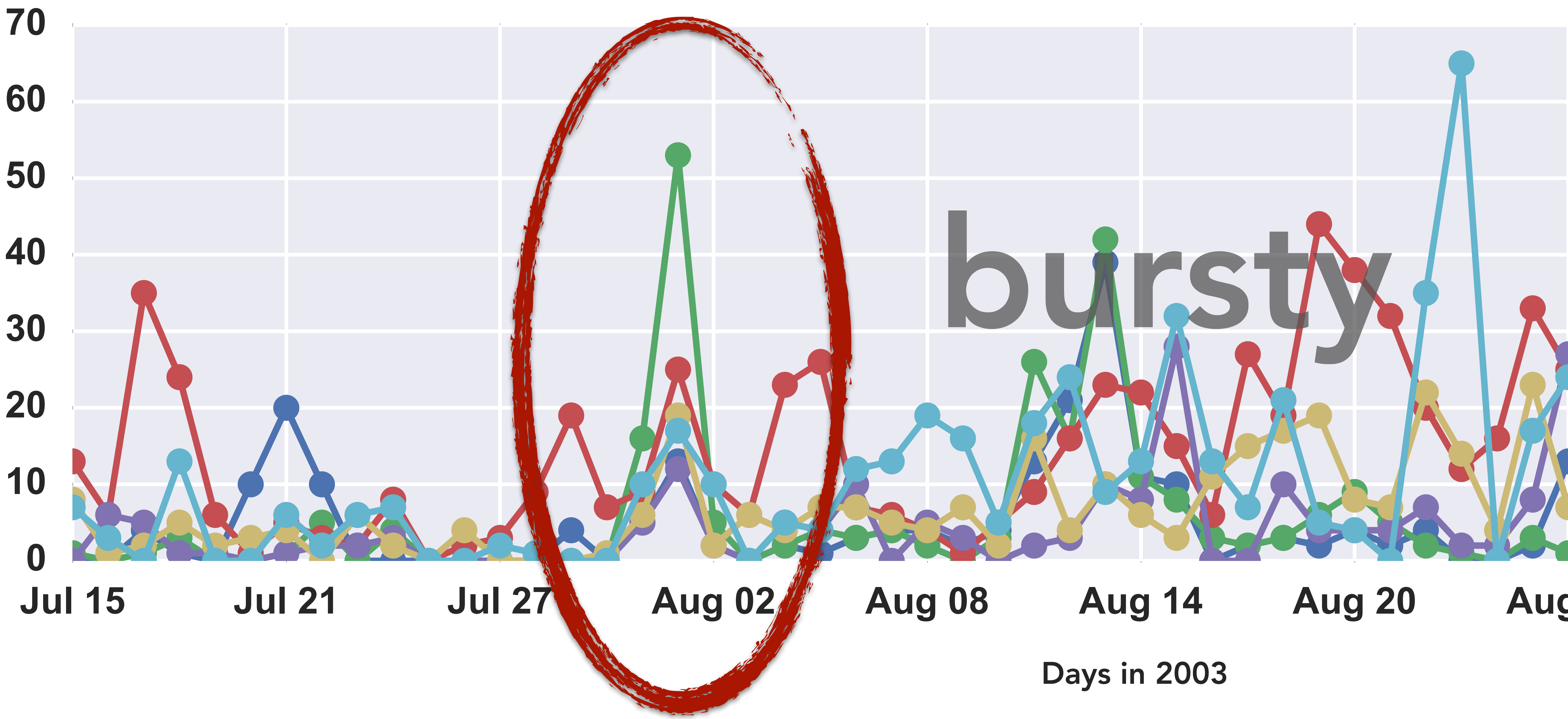
Sequential multivariate count data



Sequential multivariate count data

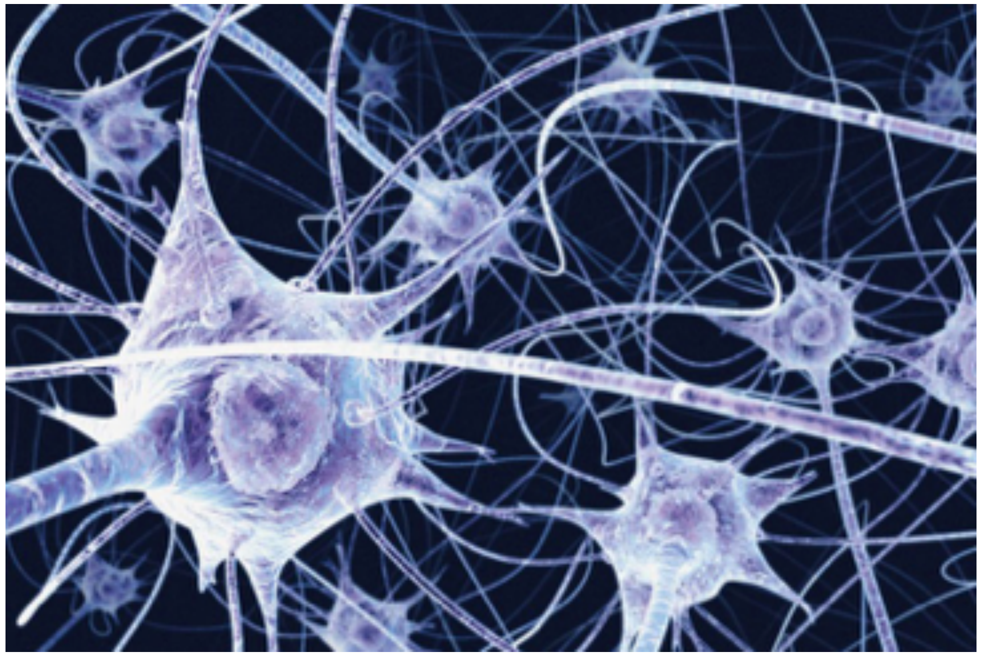


Sequential multivariate count data



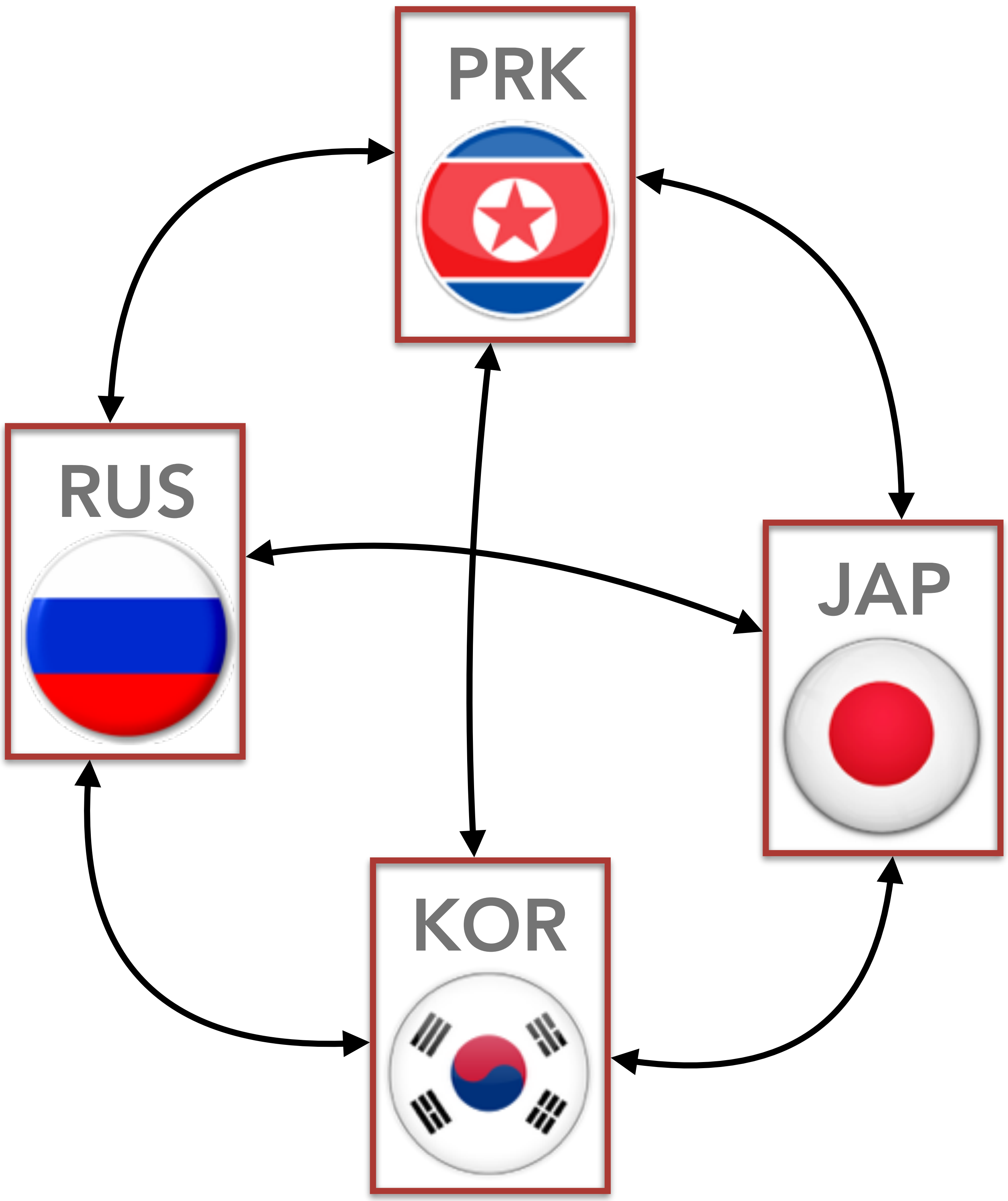
Sequential multivariate count data

It is everywhere.



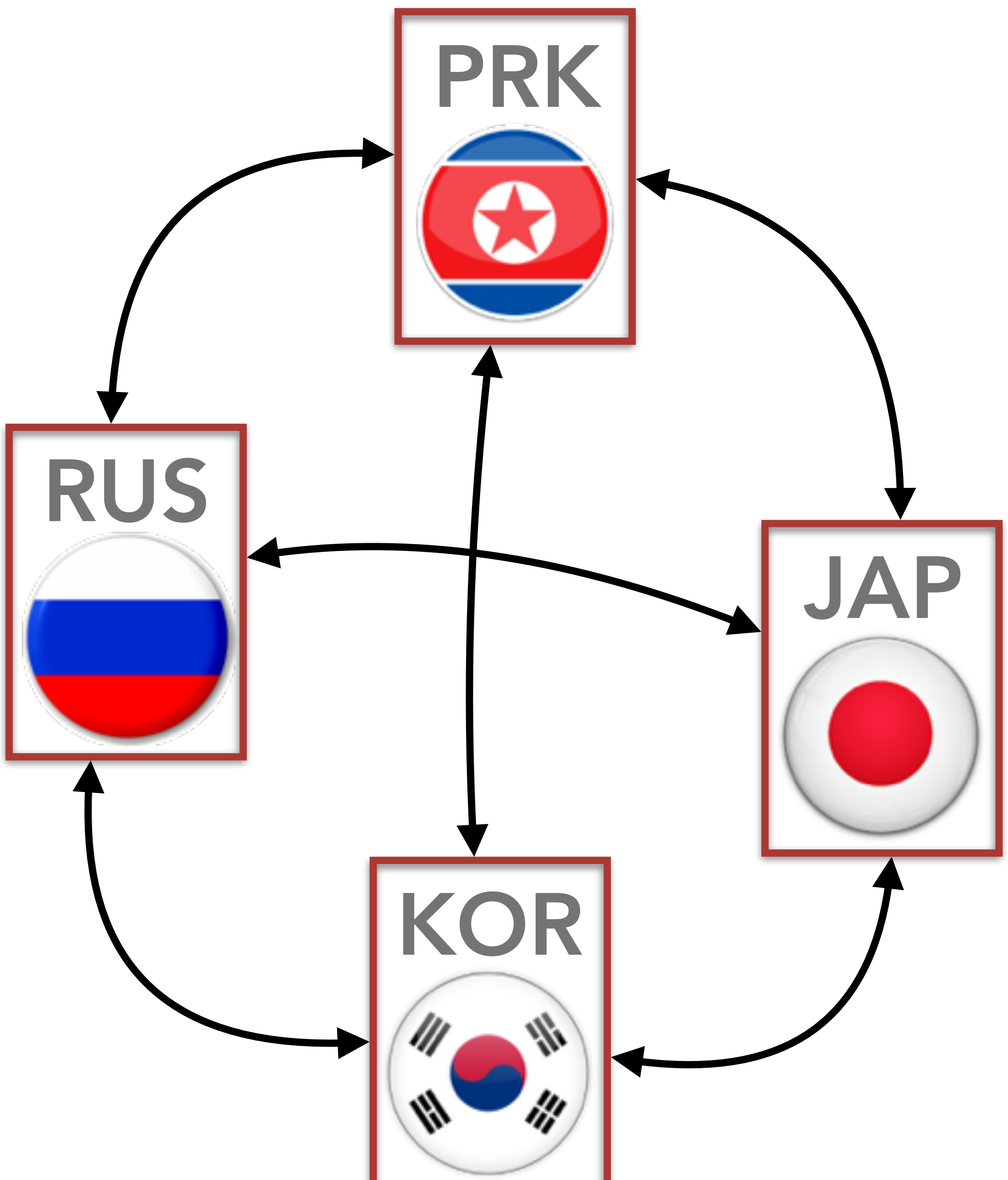
International relations

International relations



Sequentially observed count vectors

June 28, 2003

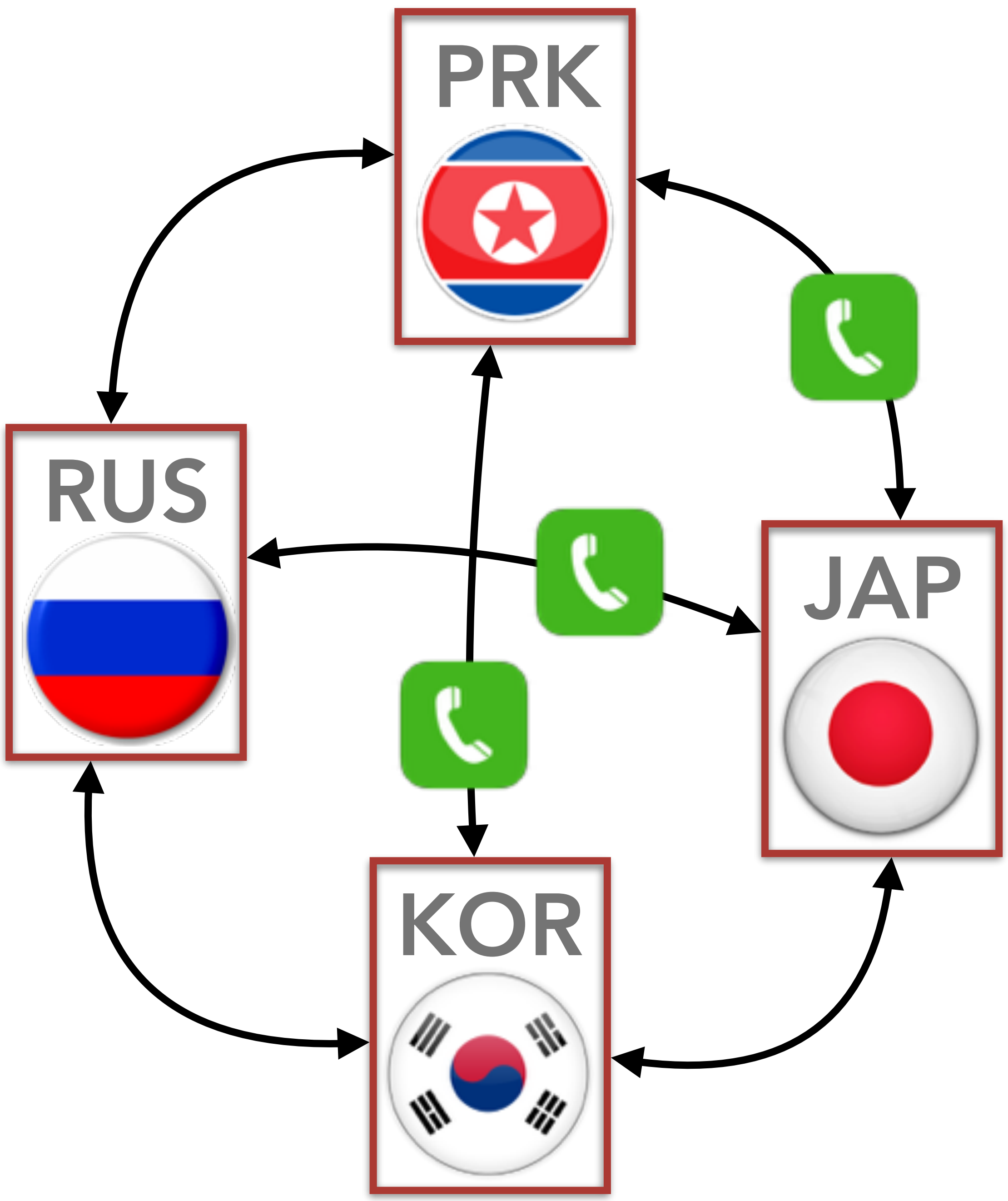


RUS-KOR
RUS-PRK
KOR-PRK
JAP-RUS
JAP-PRK
JAP-KOR

6/28

Sequentially observed count vectors

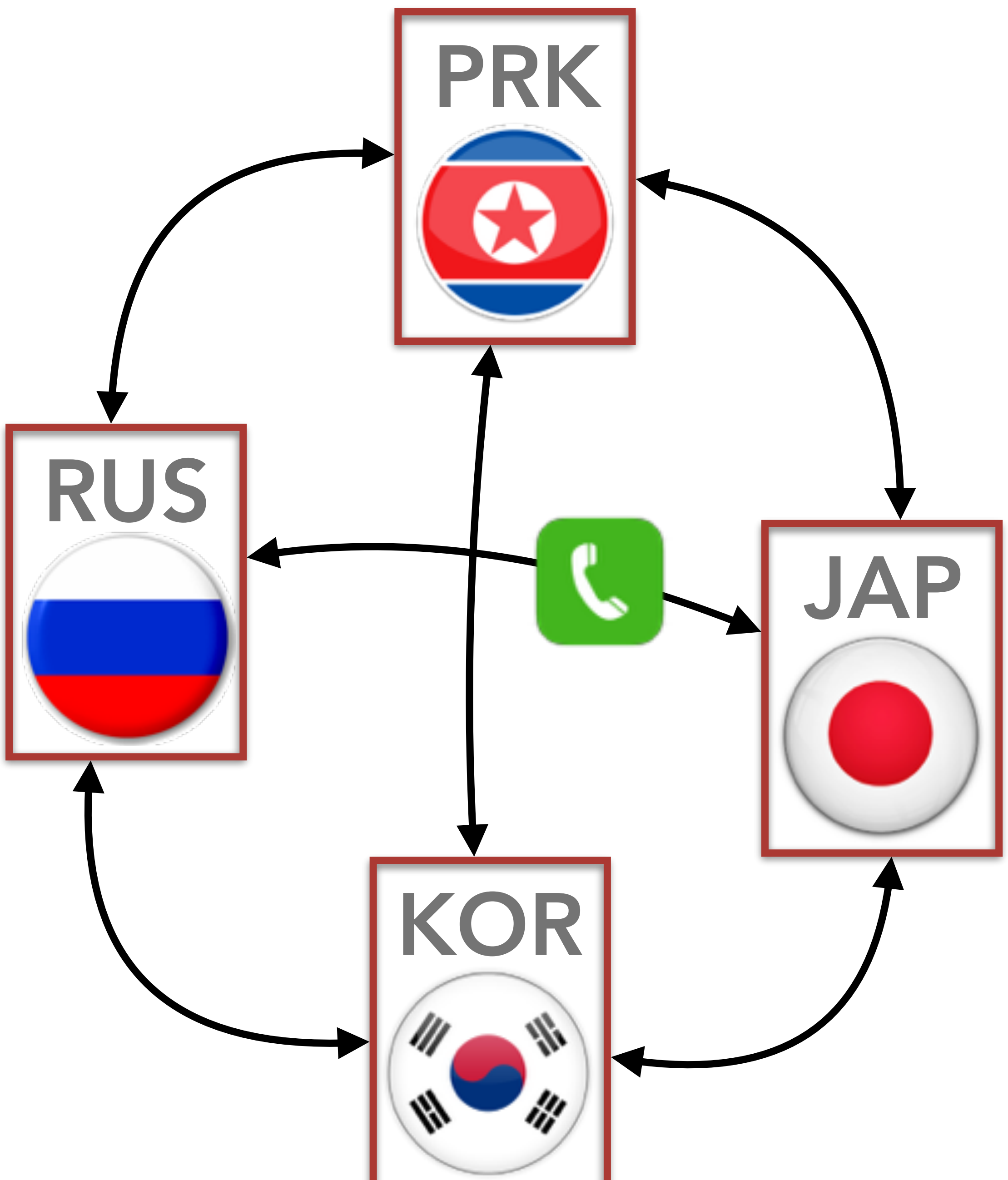
June 28, 2003



RUS-KOR	0
RUS-PRK	0
KOR-PRK	5
JAP-RUS	32
JAP-PRK	2
JAP-KOR	0

Sequentially observed count vectors

June 29, 2003



	6/28	6/29
RUS-KOR	0	0
RUS-PRK	0	0
KOR-PRK	5	0
JAP-RUS	32	40
JAP-PRK	2	0
JAP-KOR	0	0

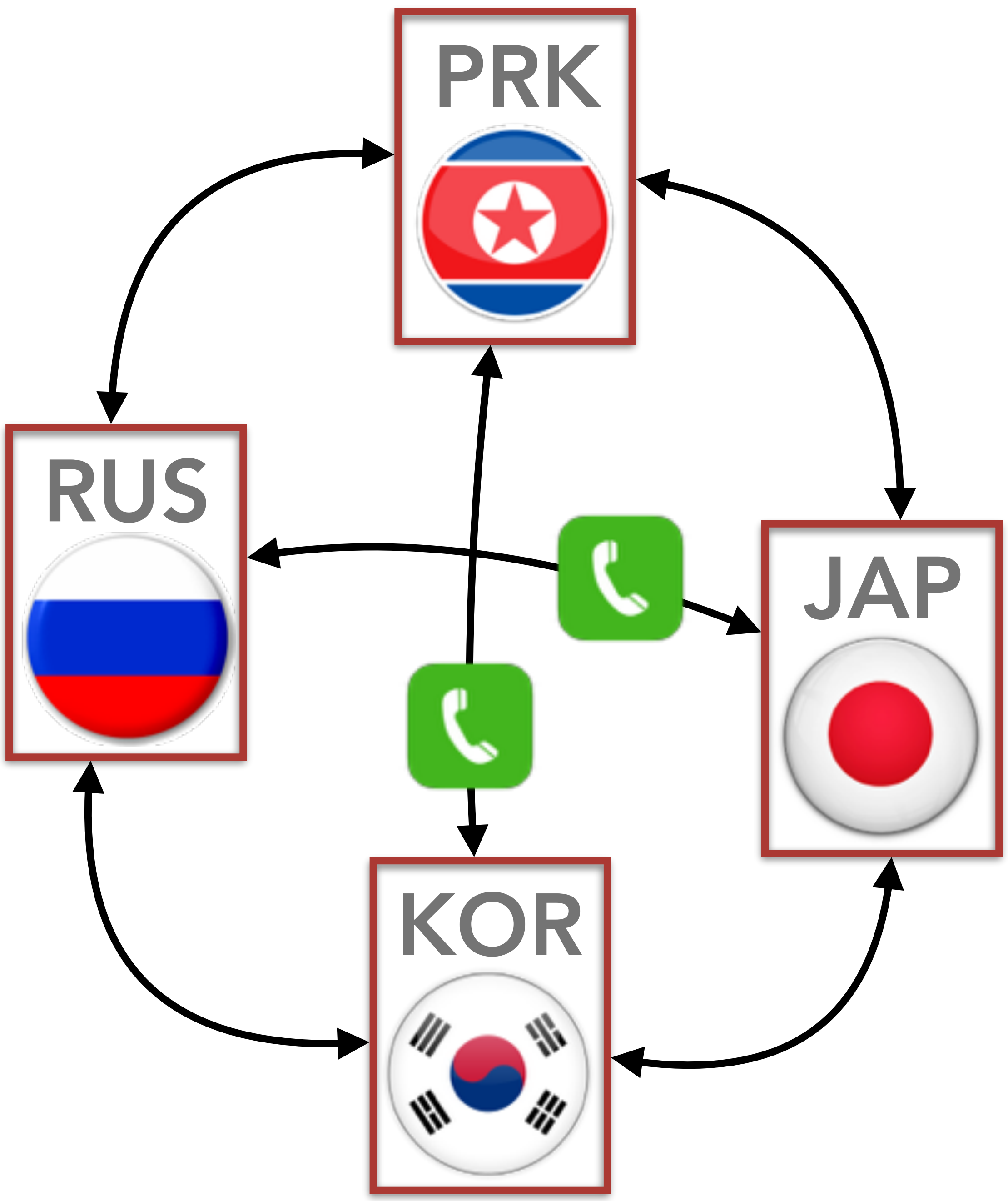
Sequentially observed count vectors

June 30, 2003

	6/28	6/29	6/30
RUS-KOR	0	0	5
RUS-PRK	0	0	6
KOR-PRK	5	0	10
JAP-RUS	32	40	7
JAP-PRK	2	0	0
JAP-KOR	0	0	2

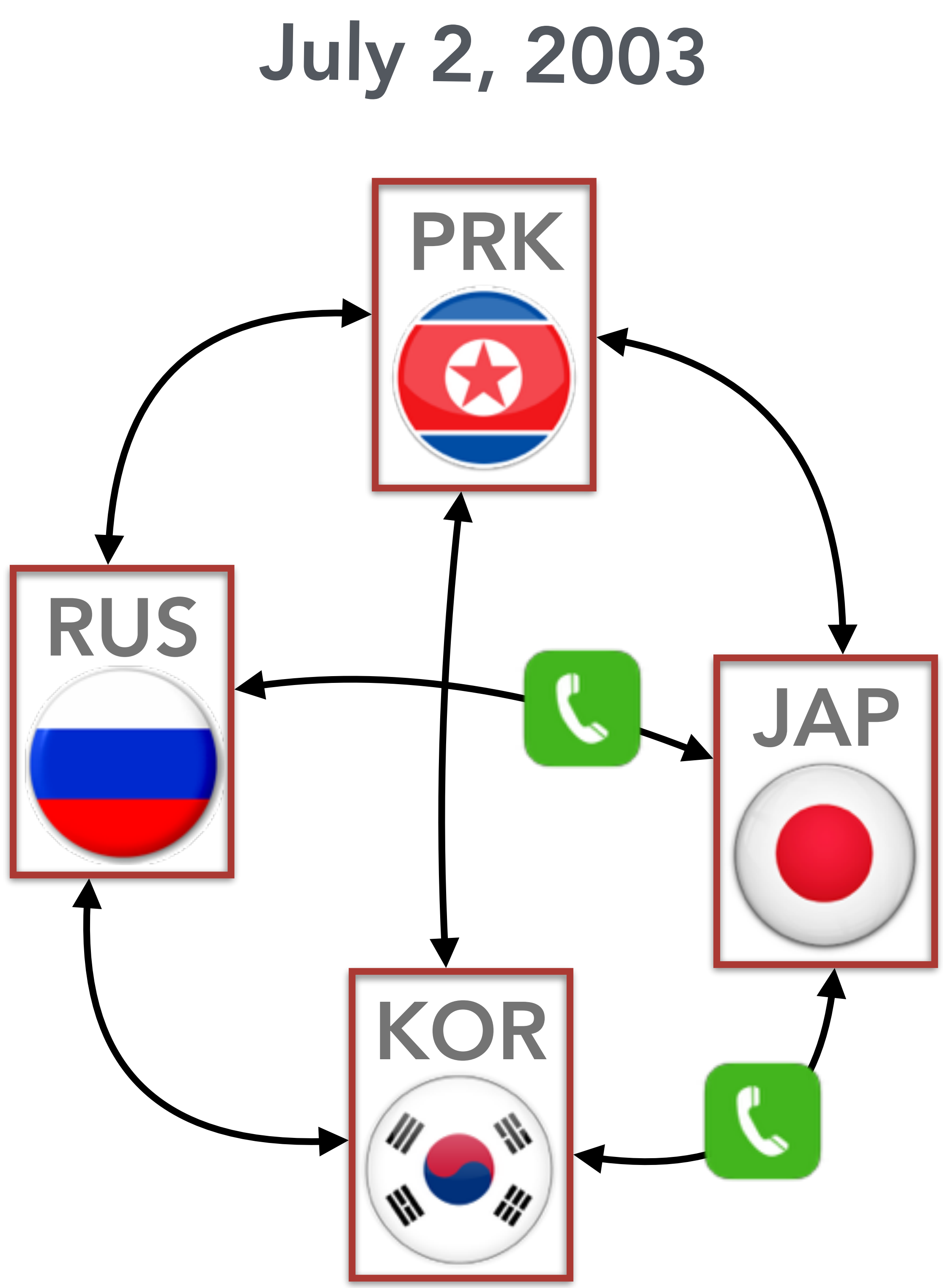
Sequentially observed count vectors

July 1, 2003



	6/28	6/29	6/30	7/1
RUS-KOR	0	0	5	0
RUS-PRK	0	0	6	0
KOR-PRK	5	0	10	13
JAP-RUS	32	40	7	3
JAP-PRK	2	0	0	0
JAP-KOR	0	0	2	0

Sequentially observed count vectors



	6/28	6/29	6/30	7/1	7/2
RUS-KOR	0	0	5	0	0
RUS-PRK	0	0	6	0	0
KOR-PRK	5	0	10	13	0
JAP-RUS	32	40	7	3	3
JAP-PRK	2	0	0	0	0
JAP-KOR	0	0	2	0	21

Sequentially observed count vectors

T

V	6/28	6/29	6/30	7/1	7/2
RUS-KOR	0	0	5	0	0
RUS-PRK	0	0	6	0	0
KOR-PRK	5	0	10	13	0
JAP-RUS	32	40	7	3	3
JAP-PRK	2	0	0	0	0
JAP-KOR	0	0	2	0	21

Sequentially observed count vectors

$V = 6{,}197$

$T = 365$

only 4% non-zero!

T

	6/28	6/29	6/30	7/1	7/2
RUS-KOR	0	0	5	0	0
RUS-PRK	0	0	6	0	0
KOR-PRK	5	0	10	13	0
JAP-RUS	32	40	7	3	3
JAP-PRK	2	0	0	0	0
JAP-KOR	0	0	2	0	21

V

sparse

Sequentially observed count vectors

$$\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(T)}$$

$$\mathbf{y}^{(t)} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 7 \\ 0 \\ 2 \\ 0 \end{bmatrix}}_{V}$$

Sequentially observed count vectors

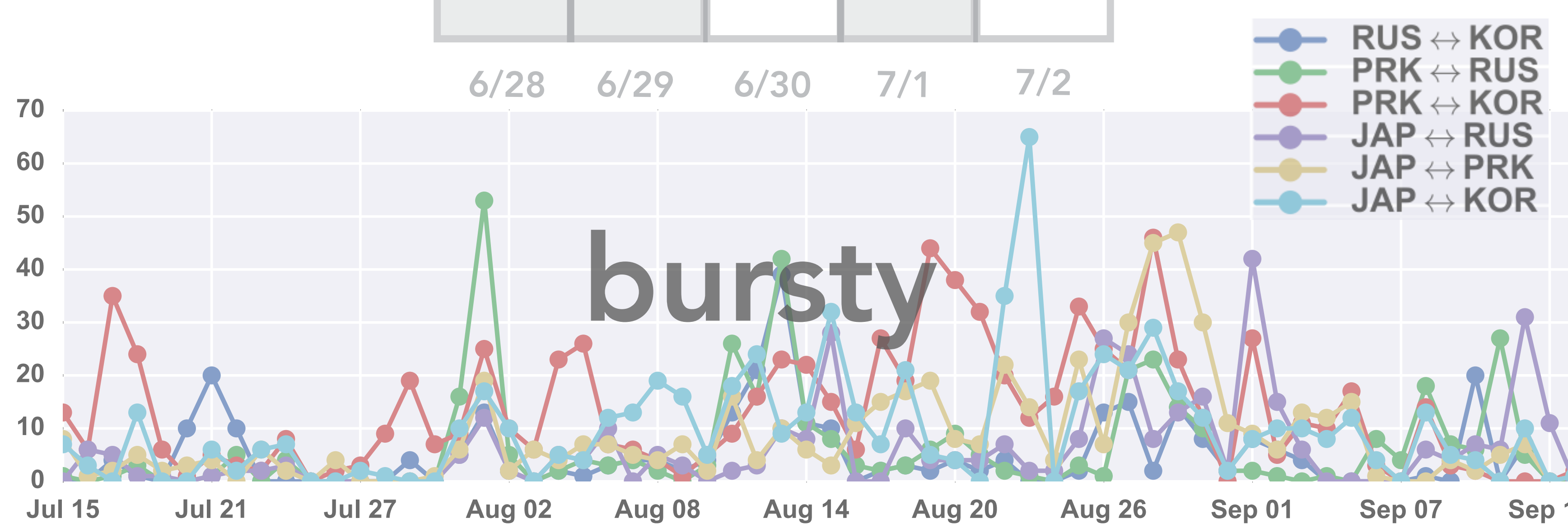
$$\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(T)}$$

$$\mathbf{y}^{(t)} = \begin{bmatrix} 0 \\ 1 \\ 7 \\ 0 \\ 2 \\ 0 \end{bmatrix} V$$

Especially for large *V*!

	6/28	6/29	6/30	7/1	7/2
RUS-KOR	0	0	5	0	0
RUS-PRK	0	0	6	0	0
KOR-PRK	5	0	10	13	0
JAP-RUS	32	40	7	3	3
JAP-PRK	2	0	0	0	0
JAP-KOR	0	0	2	0	21

sparse



Modeling sequential count data

Many reasons to want a probabilistic model

- Prediction (forecasting, smoothing)
- Explanation
- Exploration (interpretability!)

Modeling sequential count data

Many probabilistic models for sequential counts

- Autoregressive models
- Hidden Markov models
- Dynamic topic models
- Hawkes process models

Modeling sequential count data

Linear dynamical systems

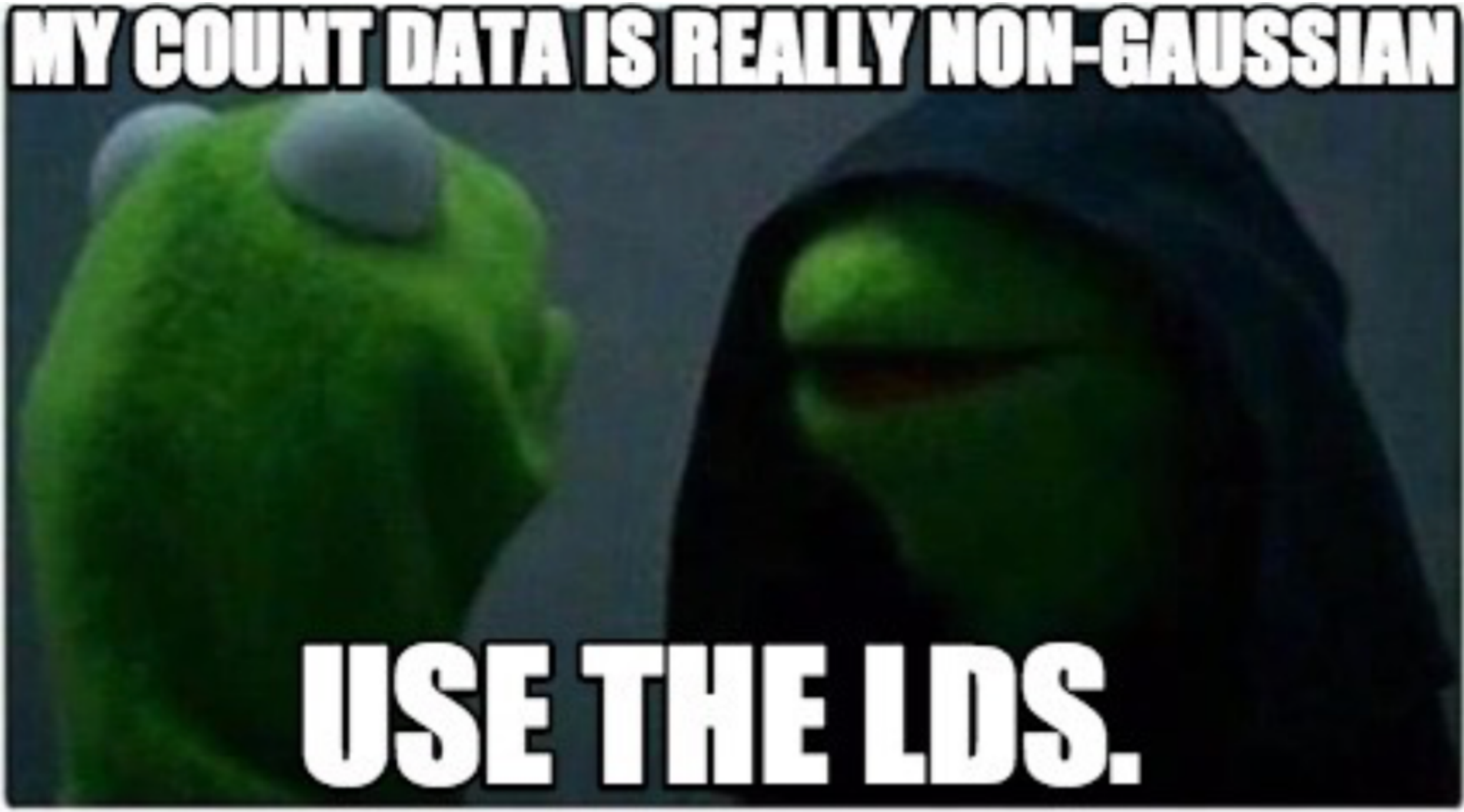
- Expressive
- Parsimonious
- Gaussian*

unnatural for count data

misspecified likelihood…

…or non-conjugate

Modeling sequential count data



The case for naturalness

Natural models

✓ likelihood matches the support of the data

✓ conjugate priors

- Statistically and computationally efficient
  - the right inductive bias constrains the hypothesis space
  - the right inductive bias supplements a lack of data
  - quantities of interest available in closed form
- Important for measurement

Poisson—gamma dynamical systems

Our contributions:

- Novel generative model
    - Matches the form of linear dynamical systems
    - Natural for count data
- Auxiliary variable-based MCMC inference 

Benefits of the model:

- Inference scales with the **number of non-zeros** (not size of data matrix, like the Gaussian LDS)
- Superior forecasting and smoothing performance
- Highly interpretable latent structure

Poisson–gamma dynamical systems

$$y_v^{(t)} \sim \text{Pois}\left(\quad \cdots \quad \right)$$

(**natural** likelihood for counts)

Poisson—gamma dynamical systems

$$y_v^{(t)} \sim \mathrm{Pois}\left(\sum_{k=1}^{K} \phi_{kv}\theta_k^{(t)}\right)$$

how active event type *v*
is in component *k*

Poisson—gamma dynamical systems

$$y_v^{(t)} \sim \text{Pois}\left(\sum_{k=1}^{K} \phi_{kv}\theta_k^{(t)}\right)$$

how active component *k*
is at time step *t*

Poisson—gamma dynamical systems

$$\mathbb{E}\left[y_v^{(t)}\right] = \sum_{k=1}^{K} \phi_{kv}\theta_k^{(t)}$$

Poisson—gamma dynamical systems

$$\theta_k^{(t)} \sim \text{Gam}\left(\qquad \cdots \qquad \right)$$

(**conjugate** prior to Poisson)

Poisson—gamma dynamical systems

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

how active component *k′*
is at time step *t-1*

Poisson—gamma dynamical systems

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

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

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

Poisson—gamma dynamical systems

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

concentration
hyperparameter

Poisson—gamma dynamical systems

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

Poisson—gamma dynamical systems

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

Poisson—gamma dynamical systems

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

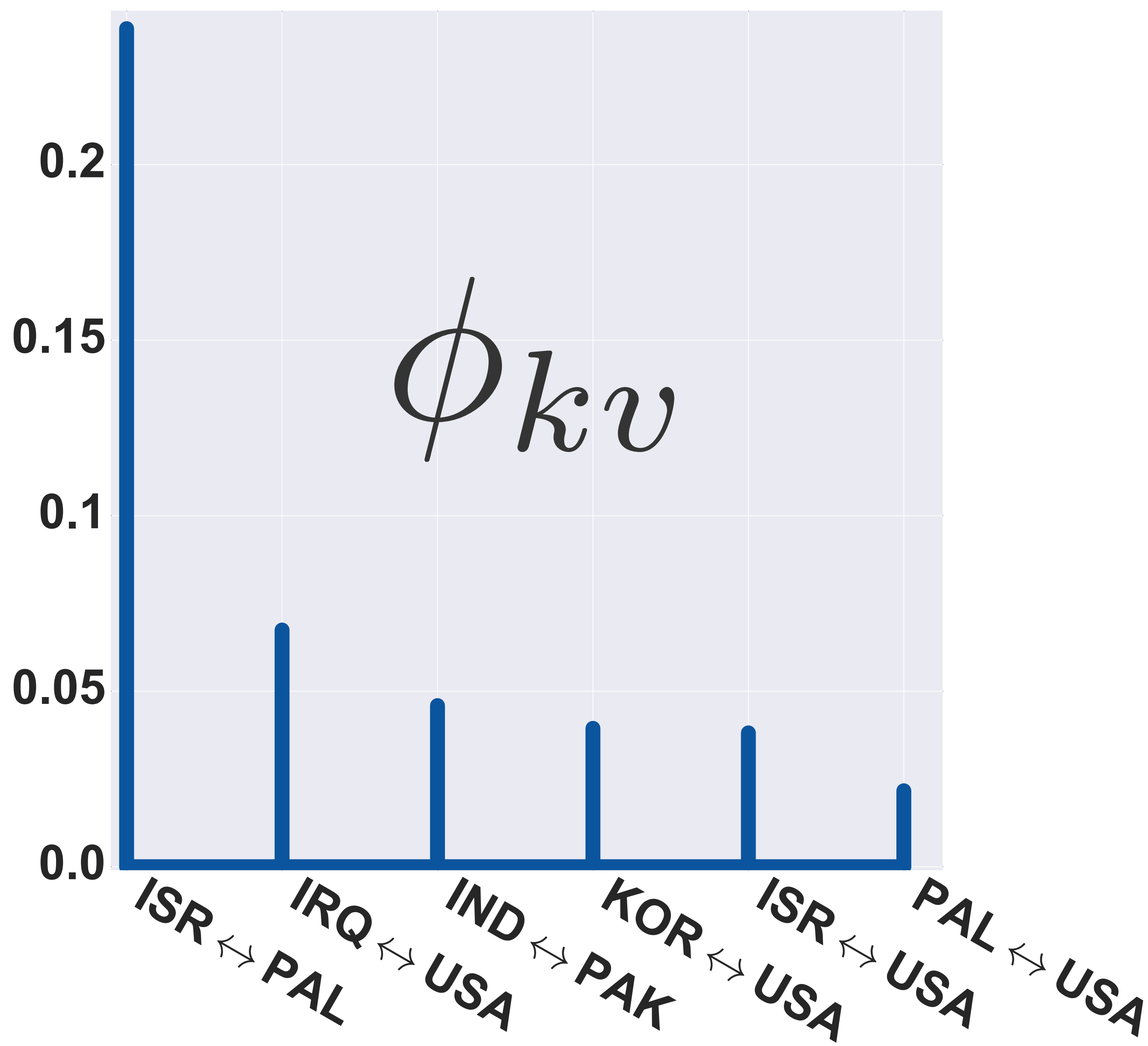
$$\mathbb{E}\left[\boldsymbol{y}^{(t)}\right] = \Phi\boldsymbol{\theta}^{(t)}$$

(matches linear dynamical systems)

Inferred latent structure

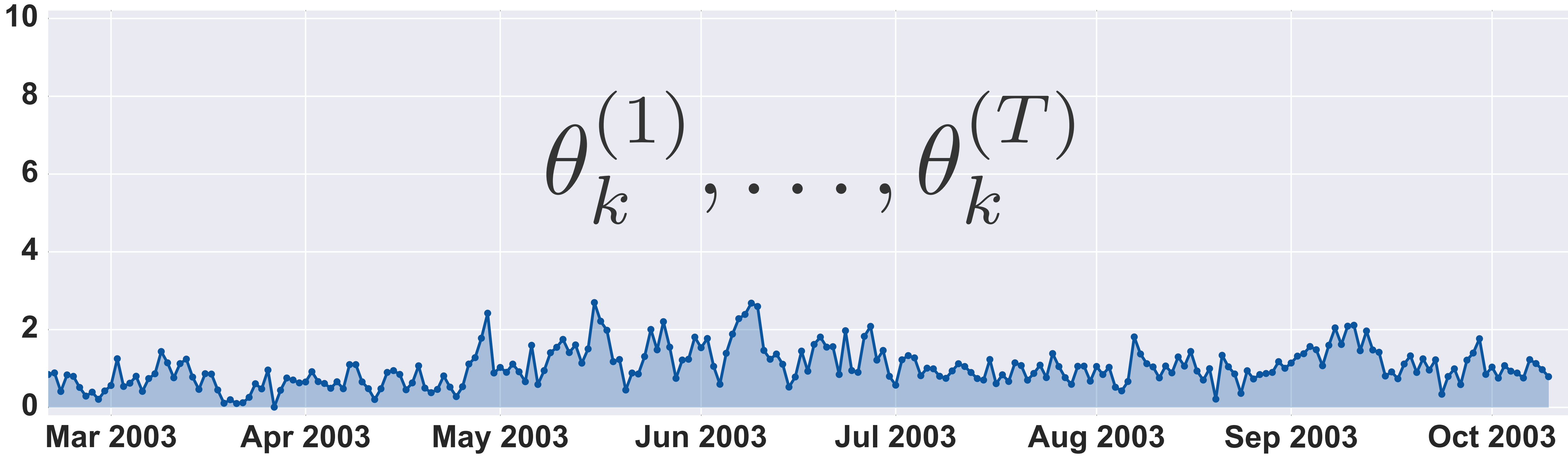
International relations data from 2003

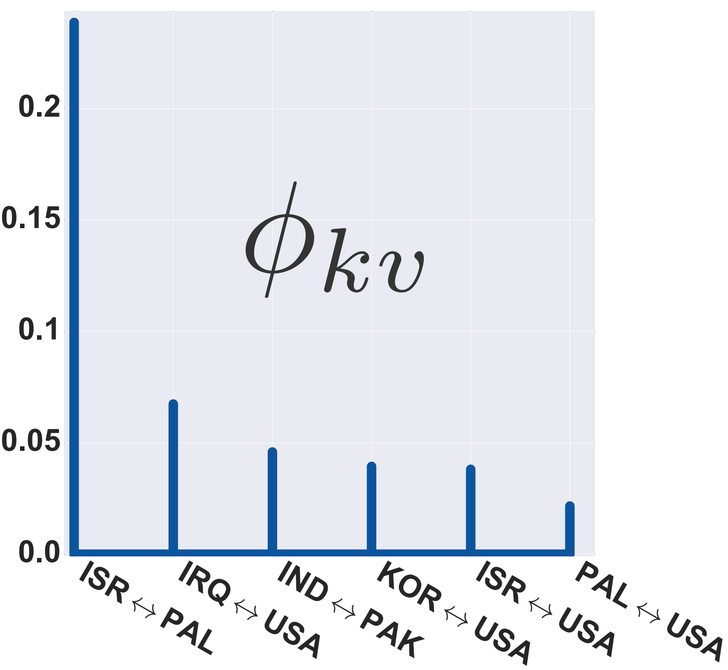
Inferred latent structure



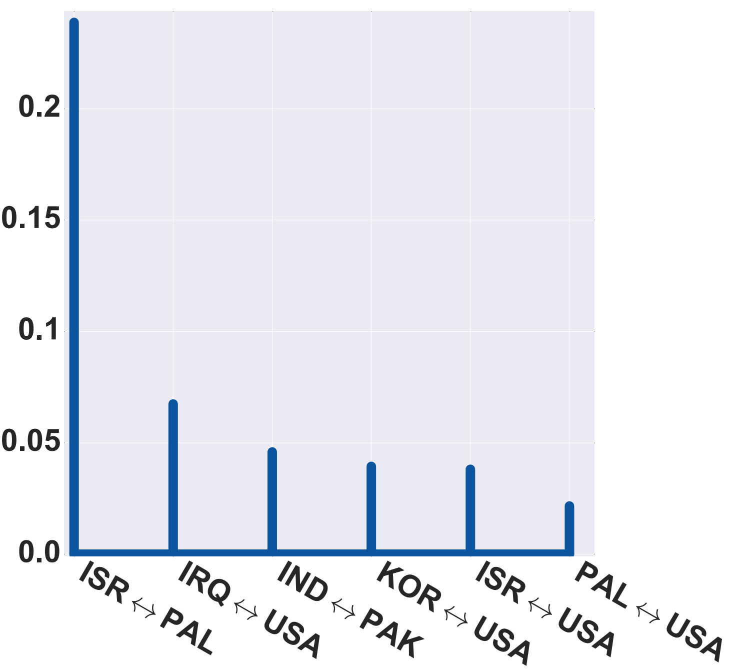
The top event types for component k=1

Inferred latent structure

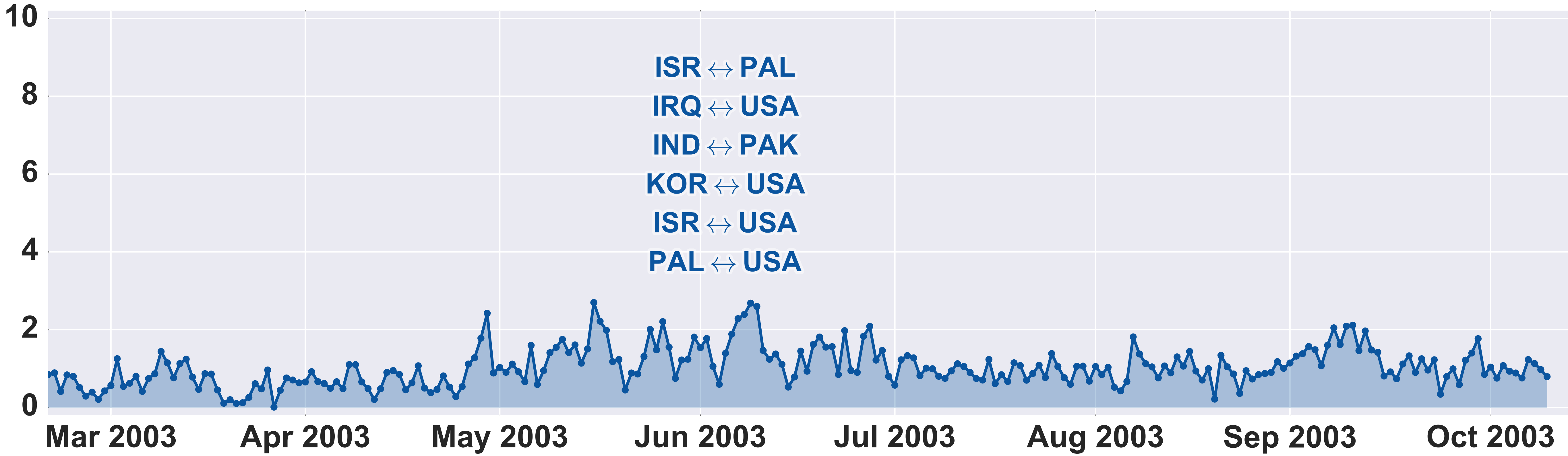


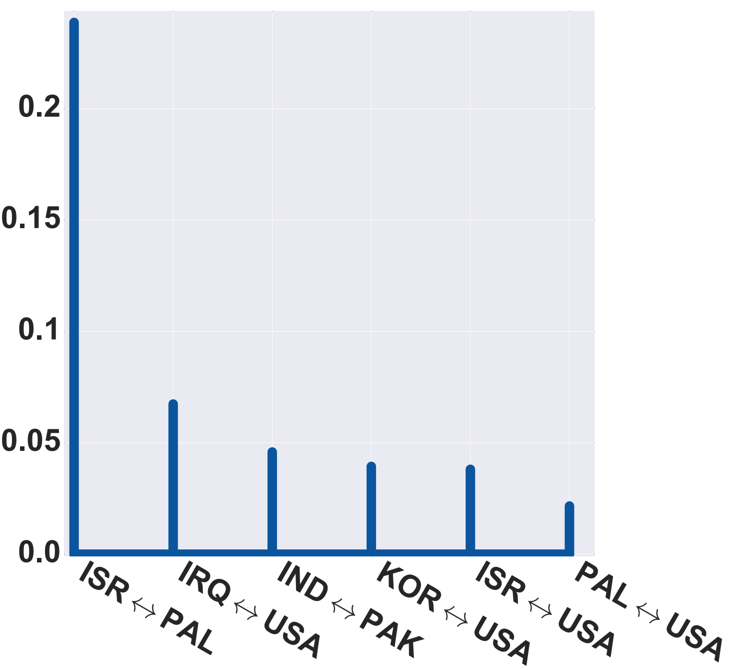


Inferred latent structure

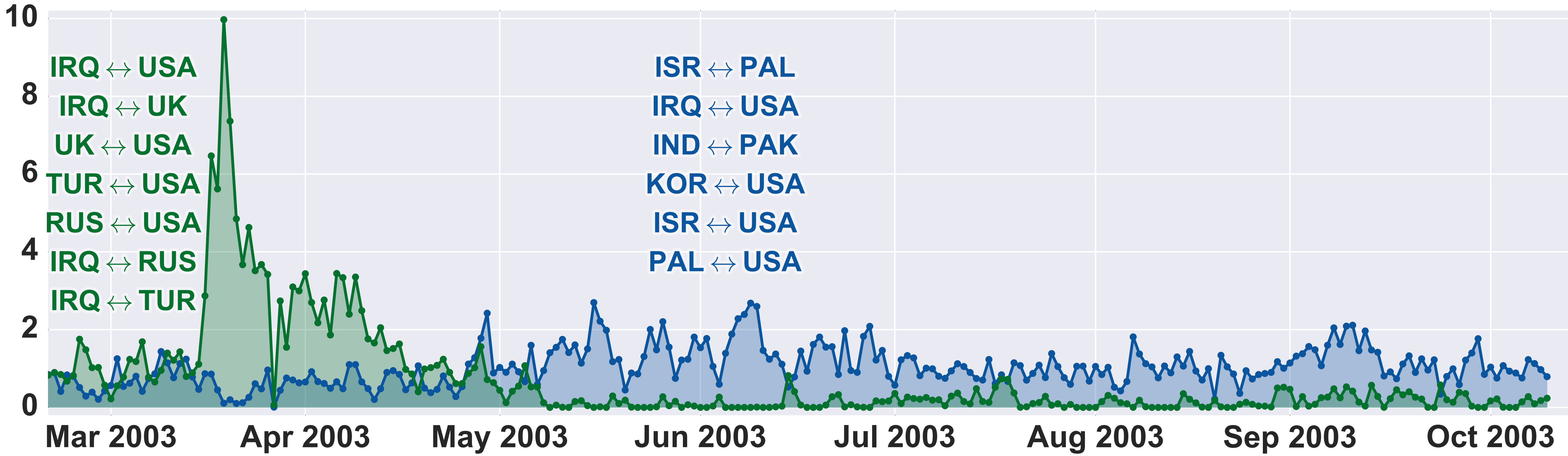


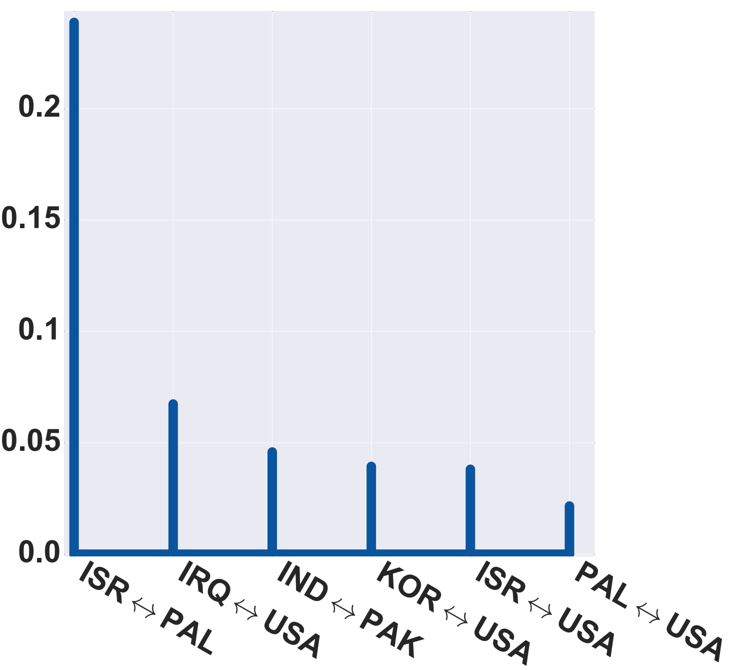
Inferred latent structure

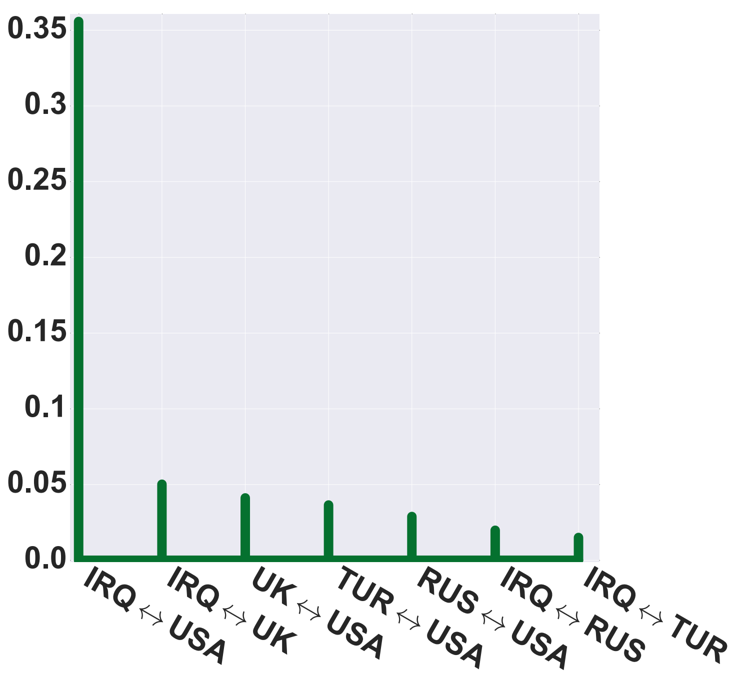




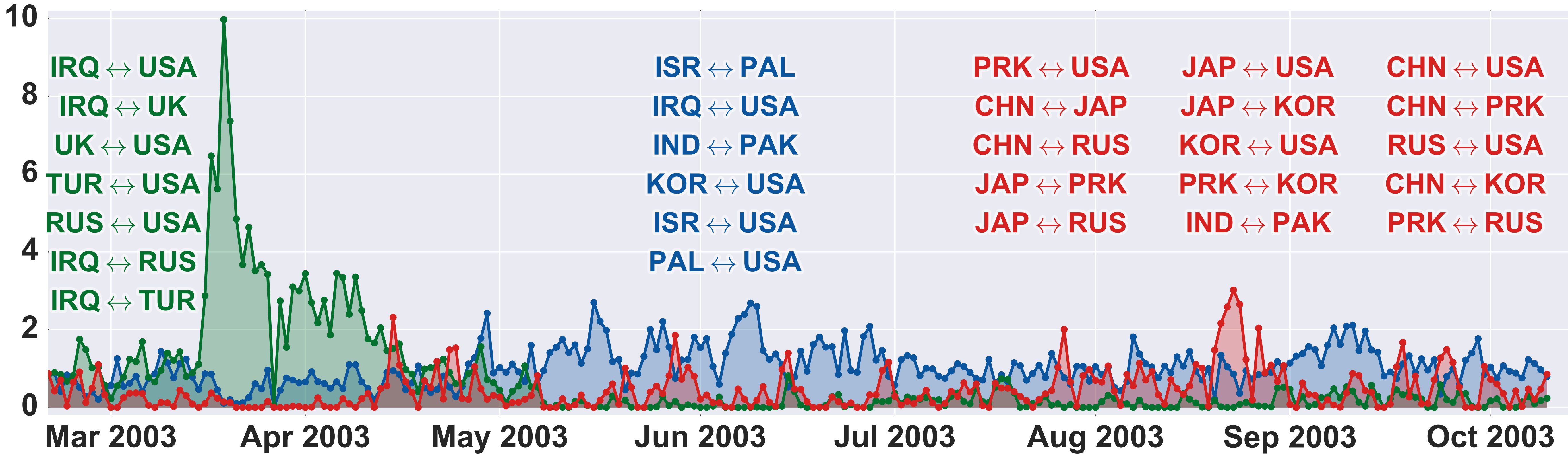
Inferred latent structure

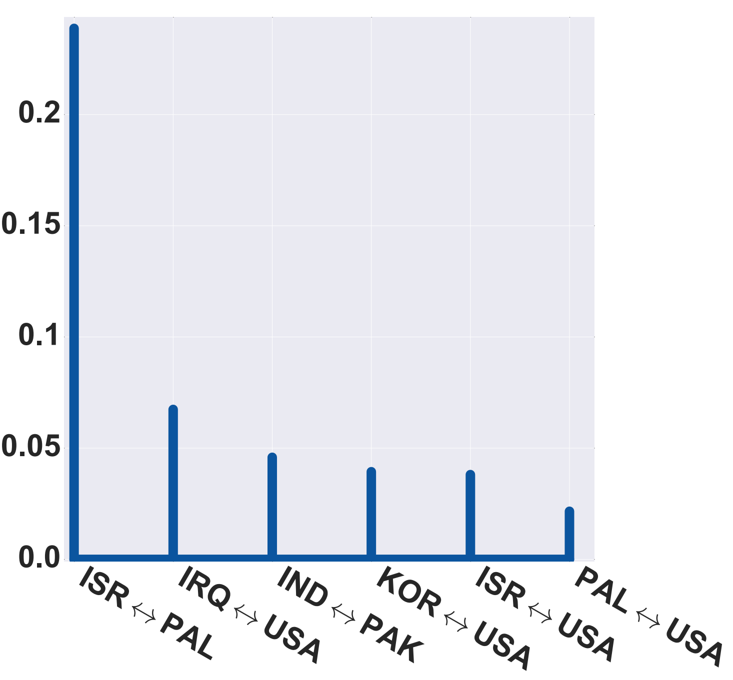


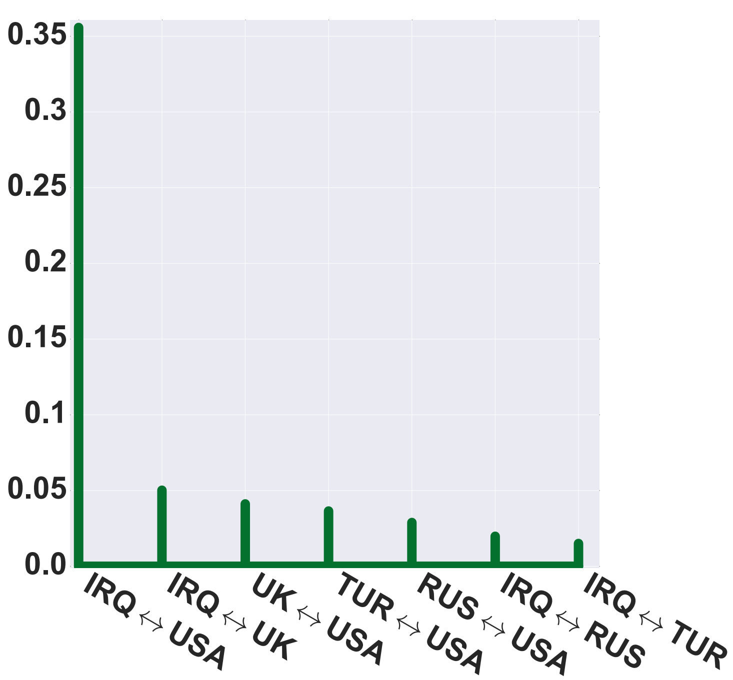


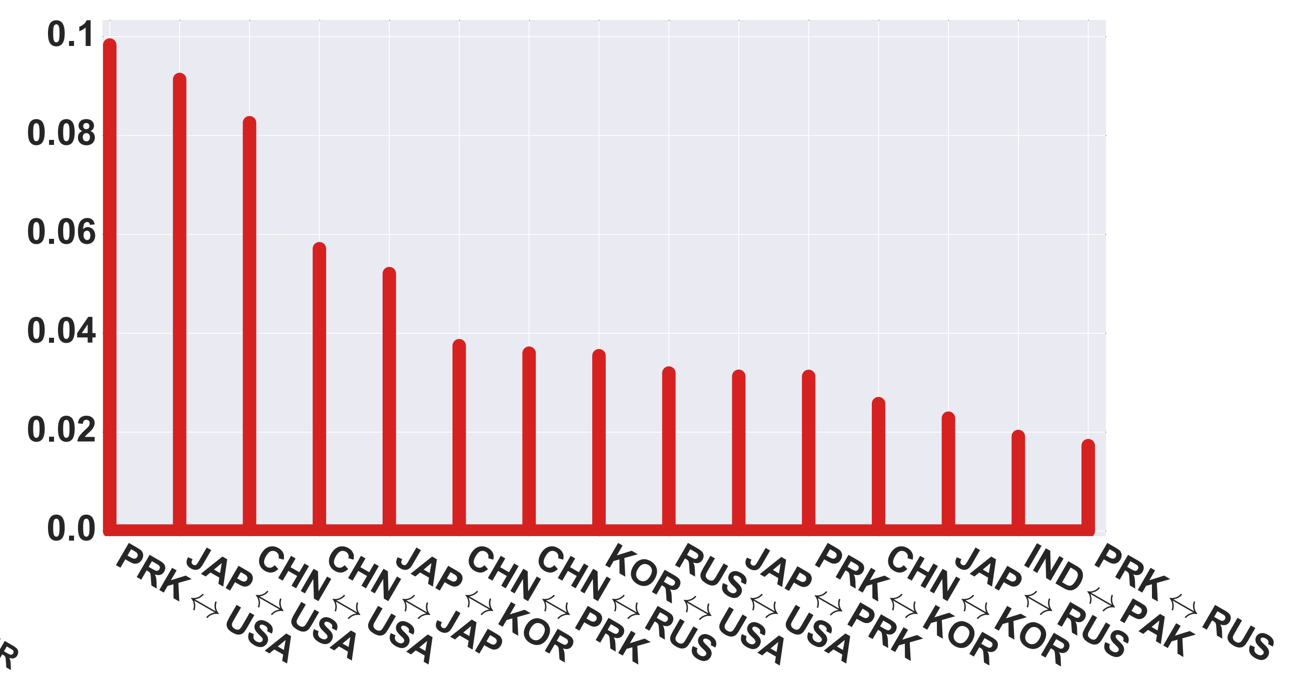


Inferred latent structure

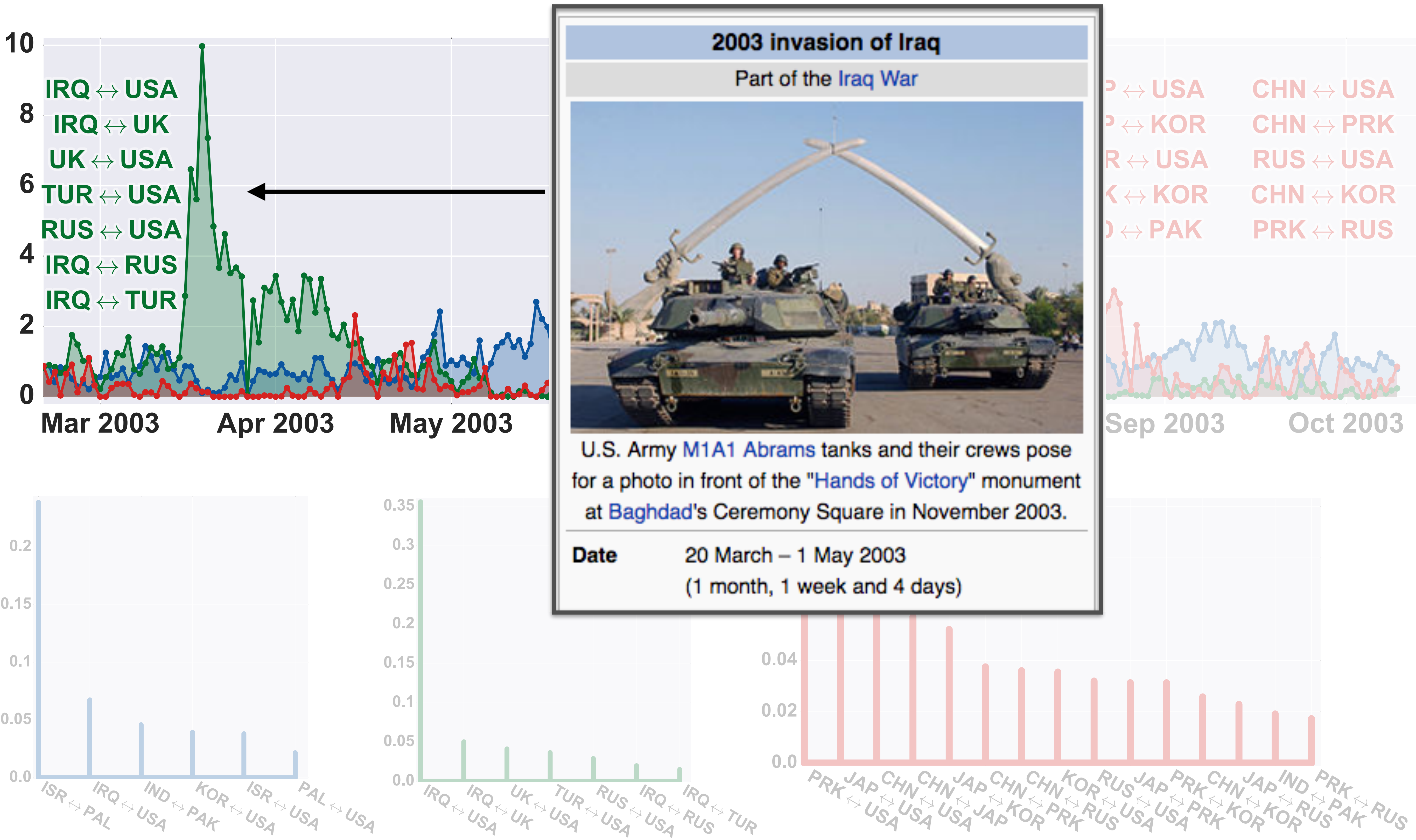




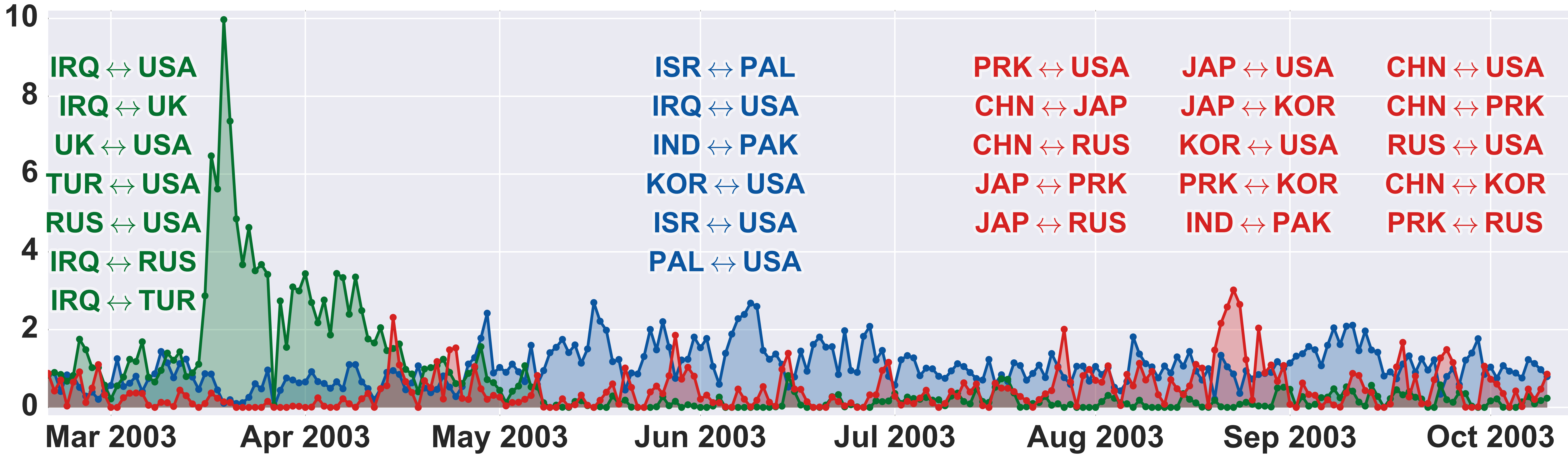


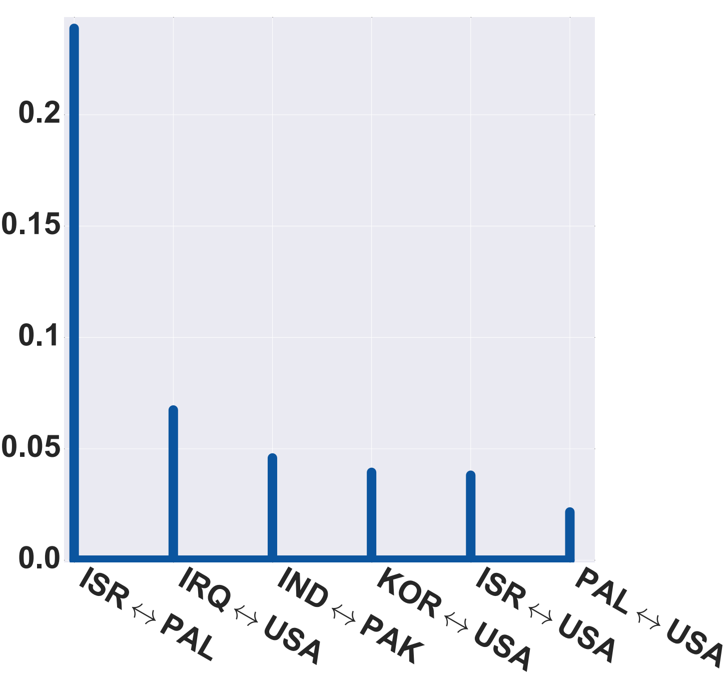


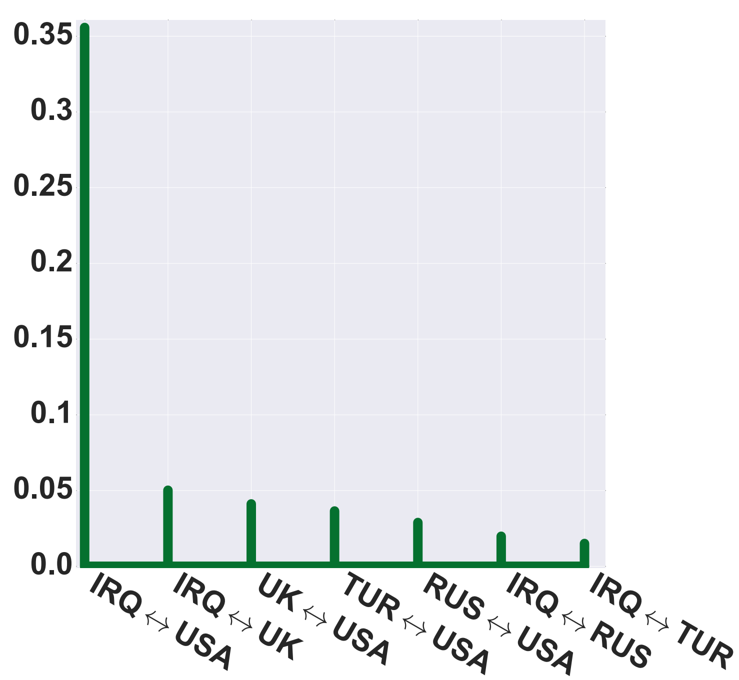
Inferred latent structure

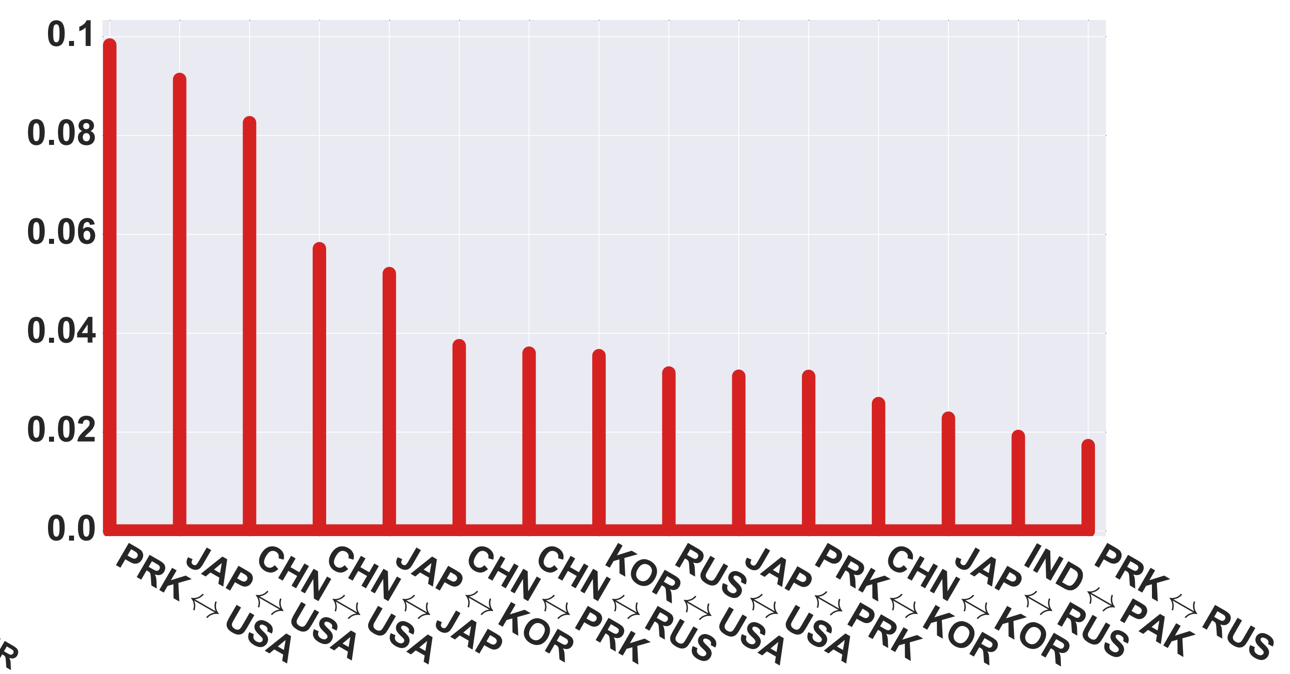


Inferred latent structure

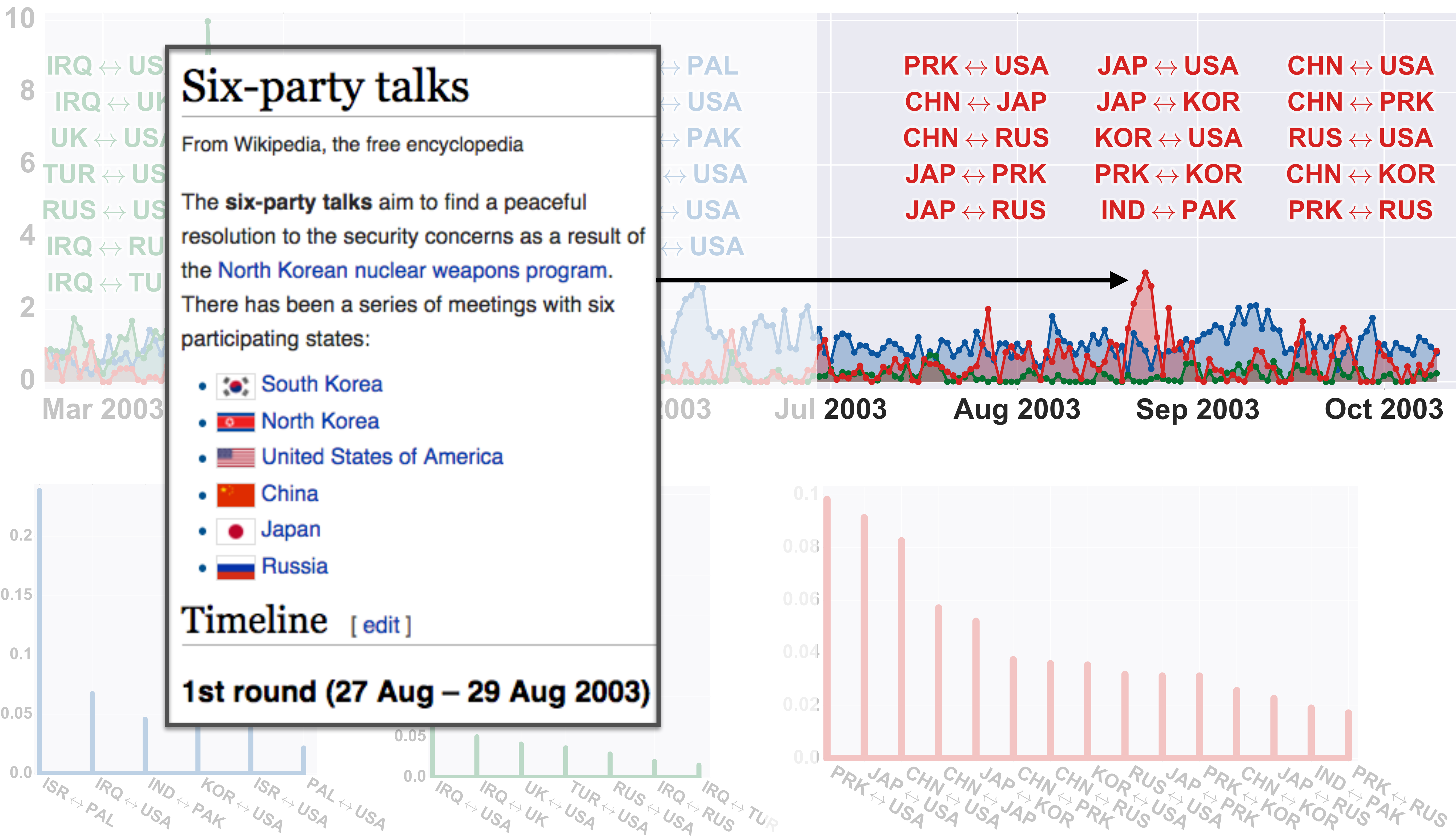




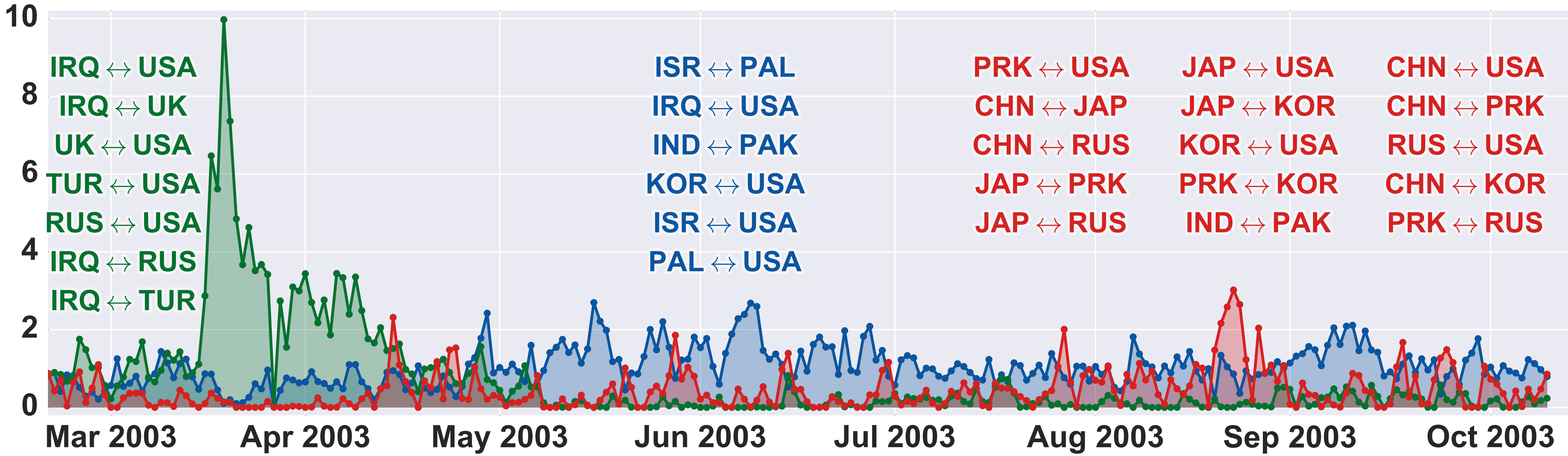


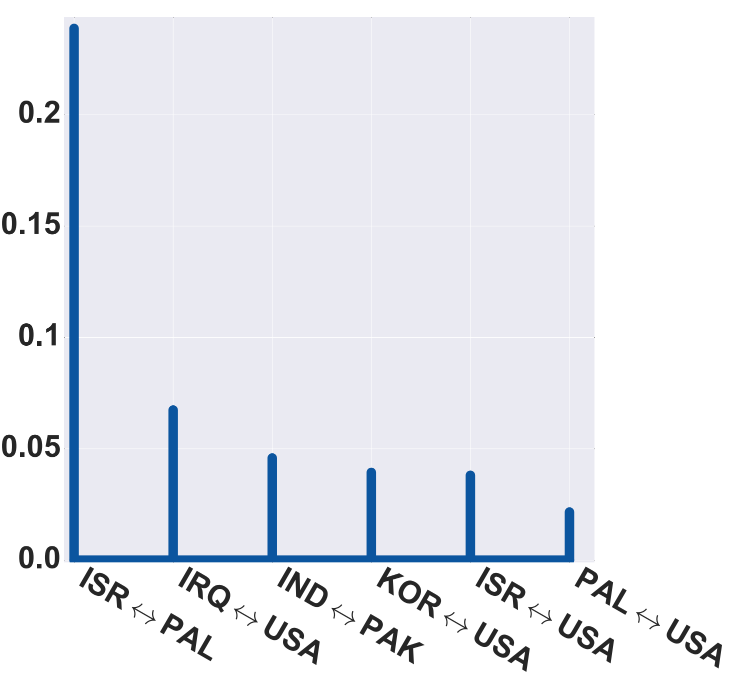


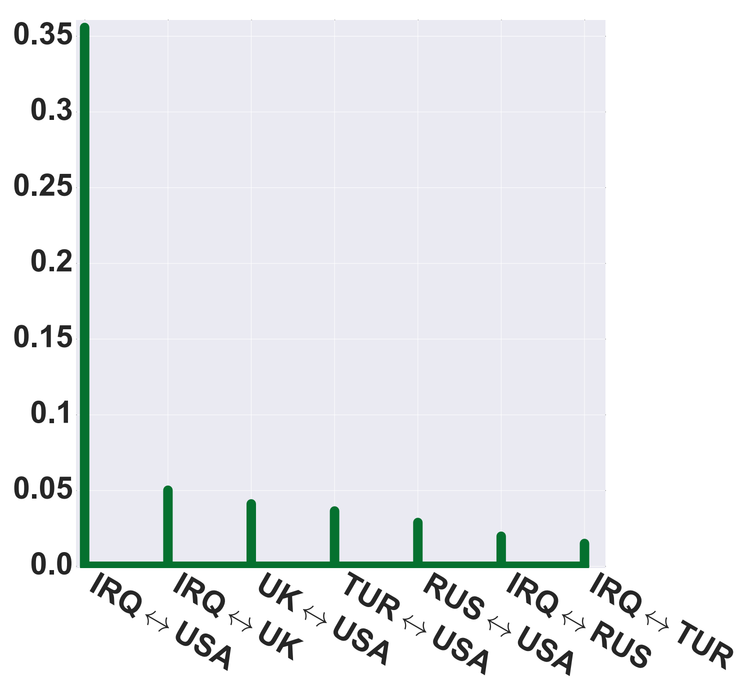
Inferred latent structure

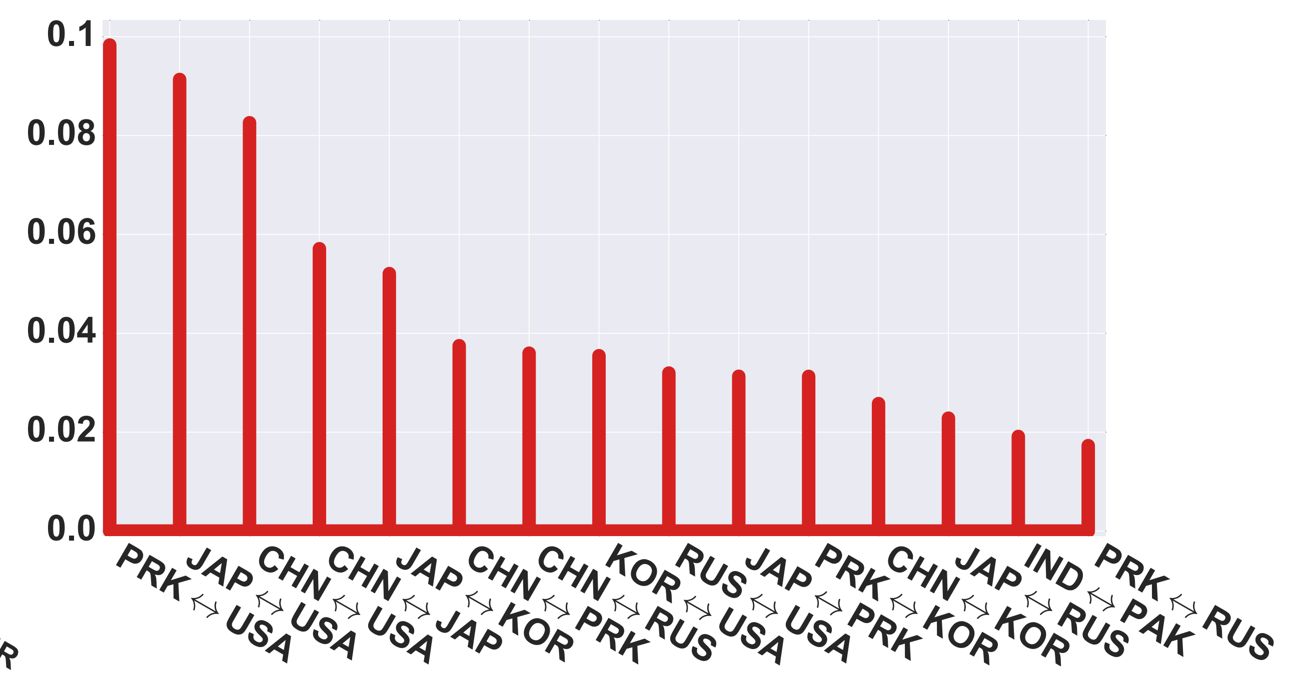


Inferred latent structure

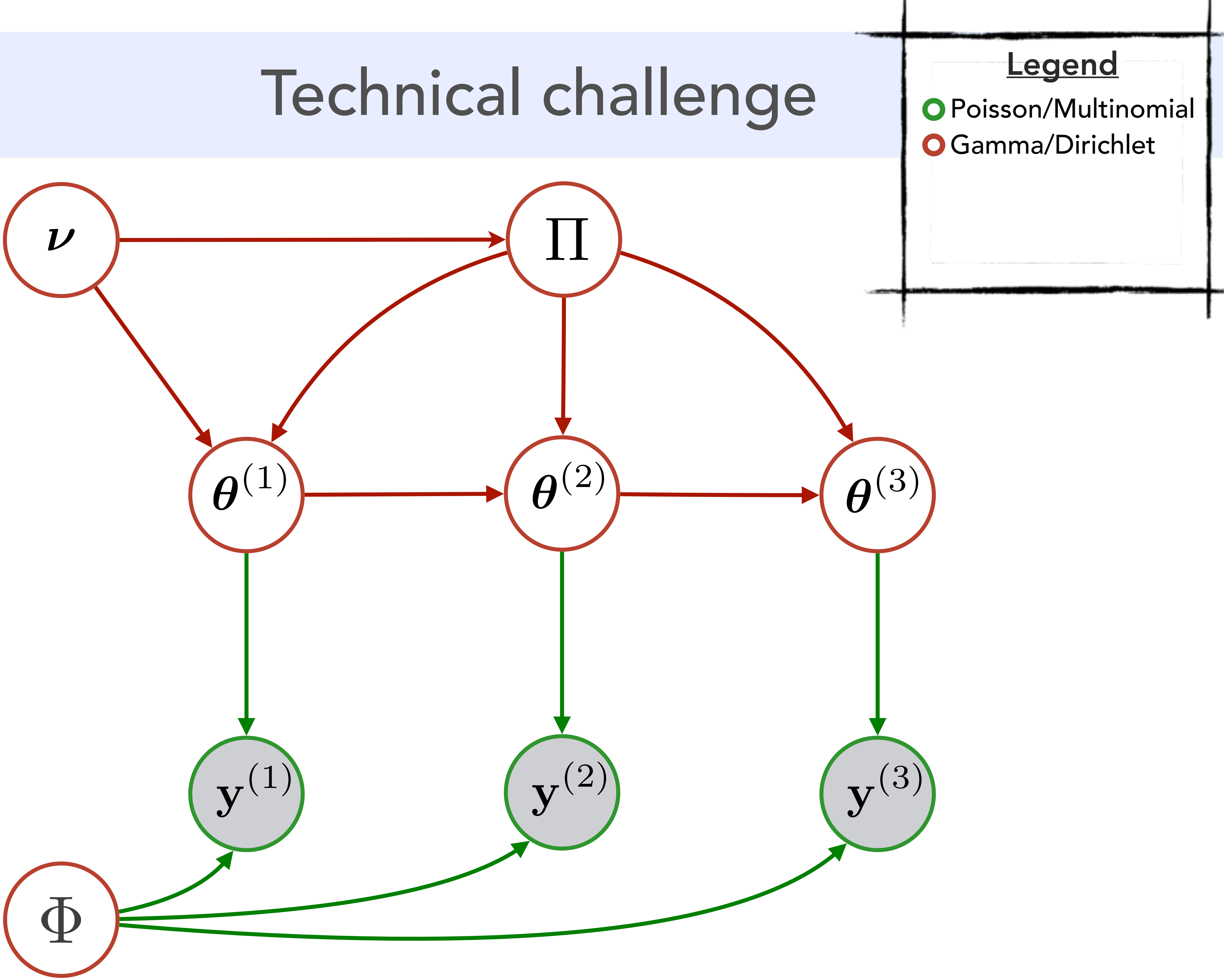








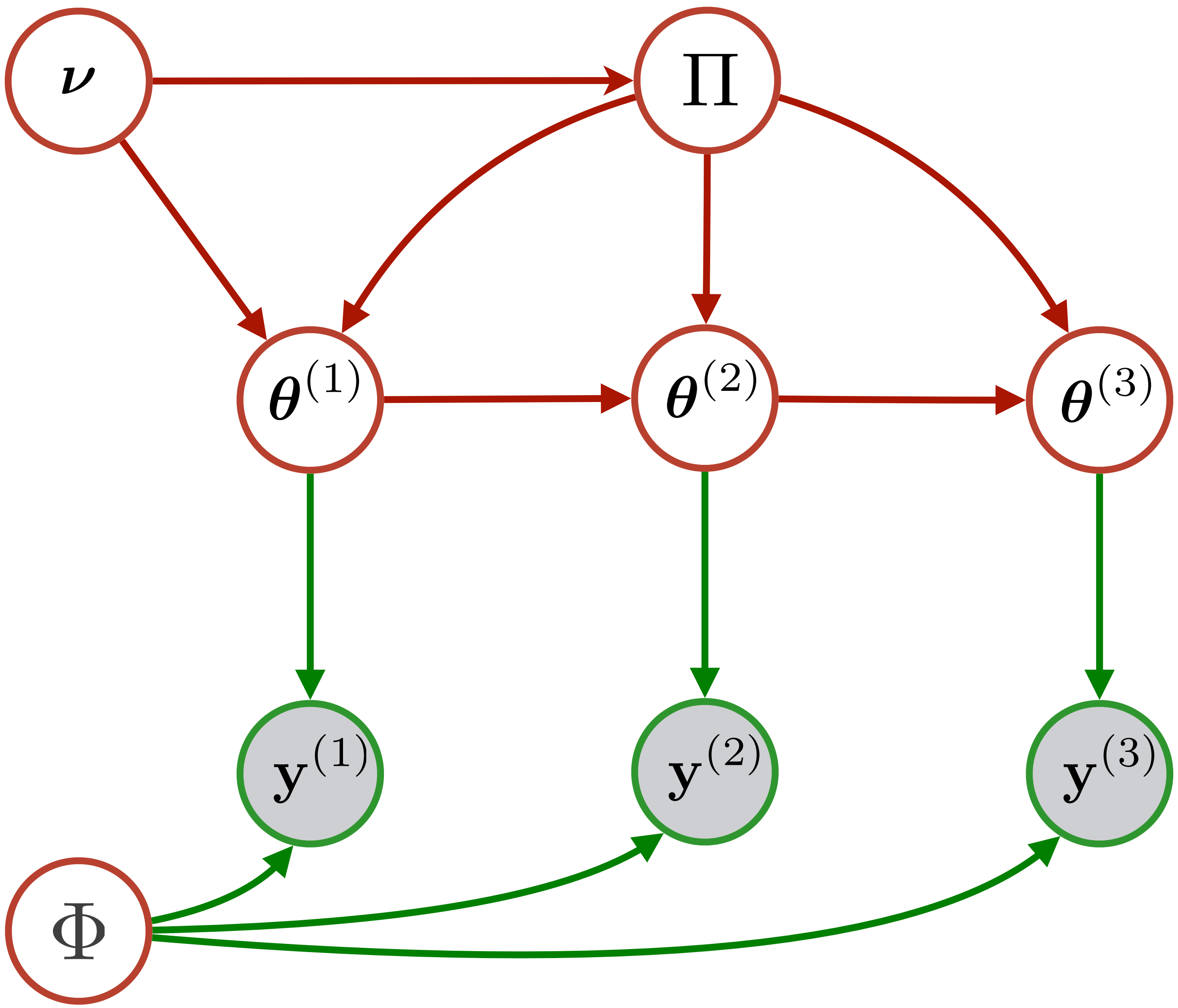
Technical challenge



Technical challenge

Legend

- Poisson/Multinomial
- Gamma/Dirichlet



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

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

$$\Phi \sim P(\Phi \mid Y)$$ ✓

(conditional posterior has closed form)

Augment and Conquer

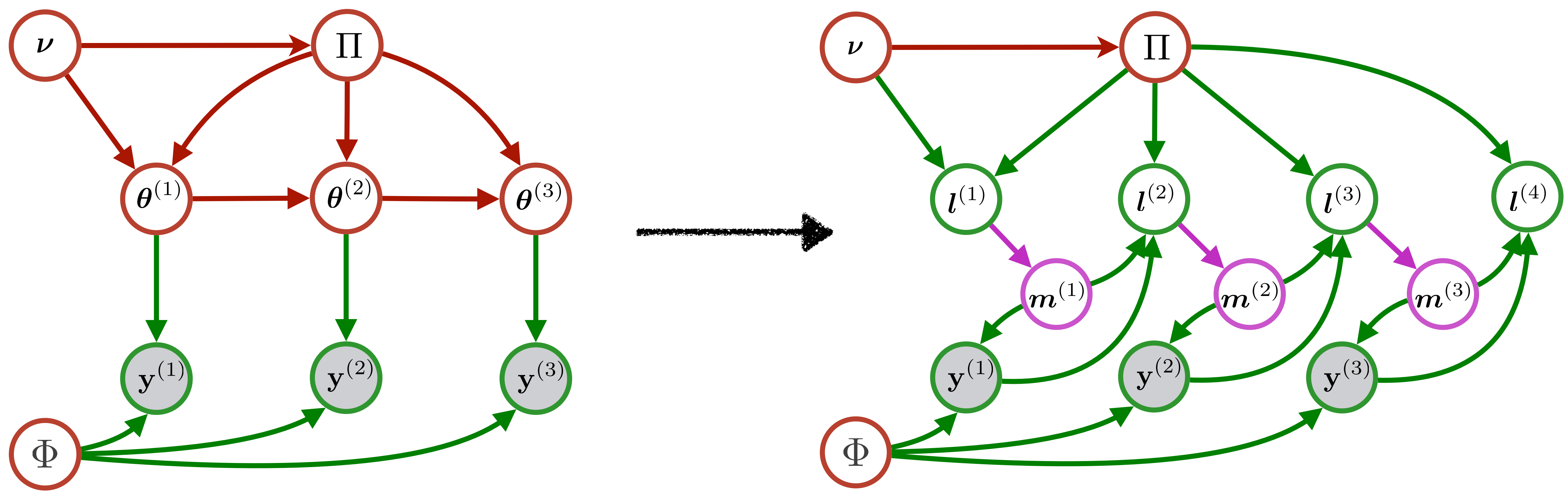
Legend

Poisson/Multinomial

Gamma/Dirichlet

Original model

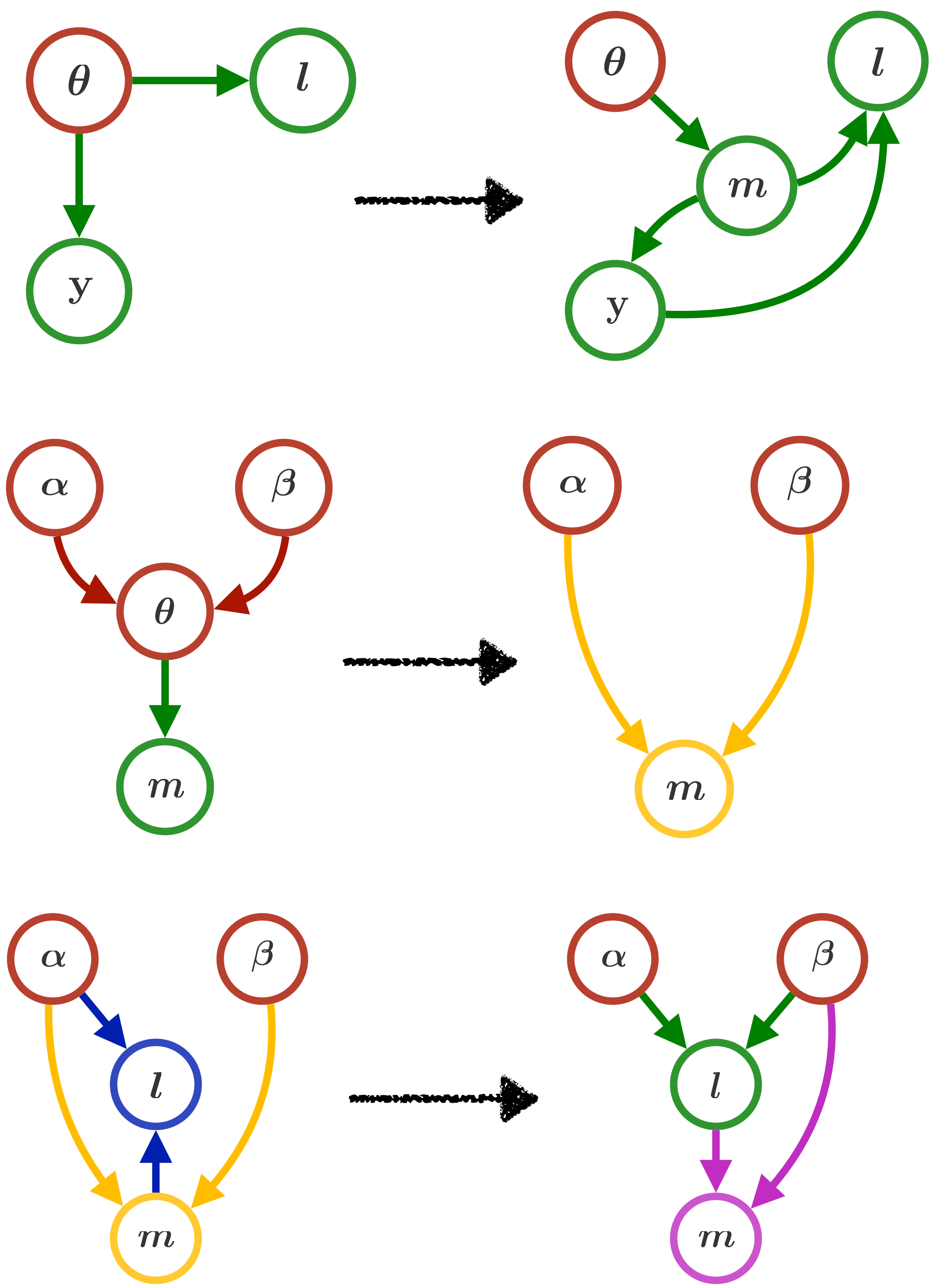
Alternative model



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

Three rules

Legend
Poisson/Multinomial
Gamma/Dirichlet



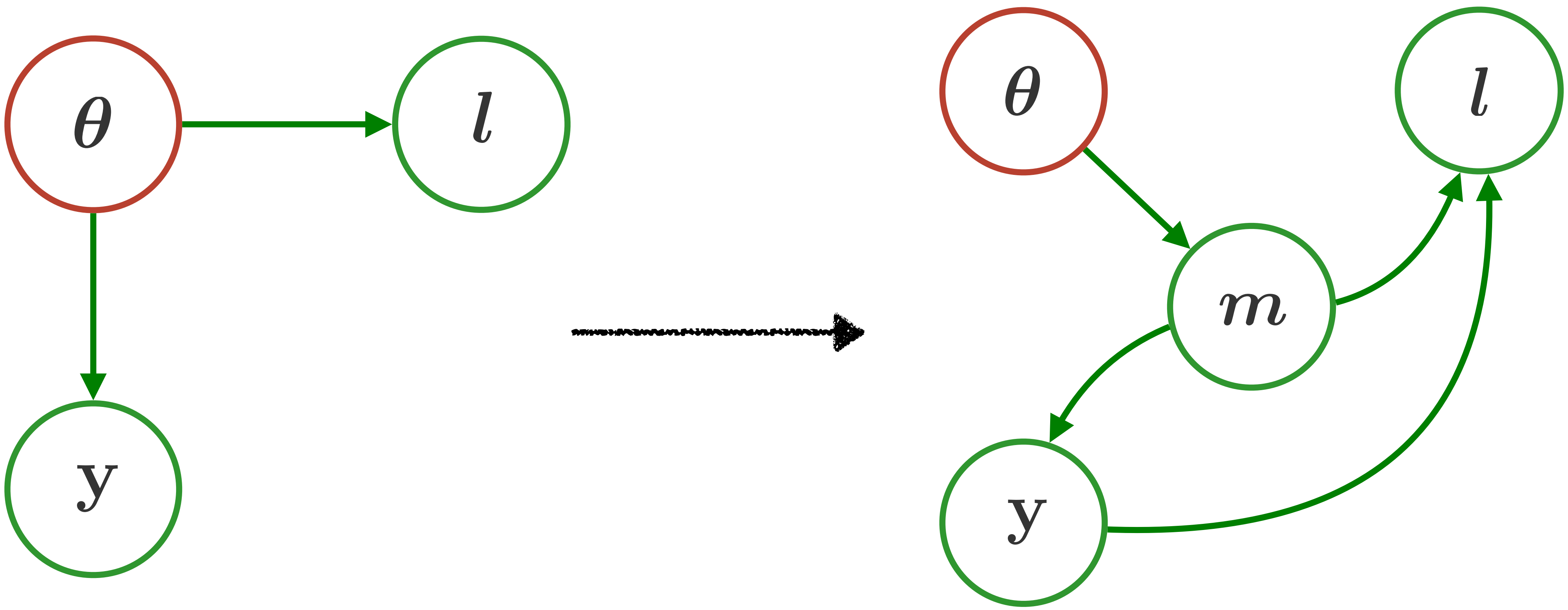
Three rules

Legend

- Poisson/Multinomial
- Gamma/Dirichlet

Relationship between
Poisson and Multinomial

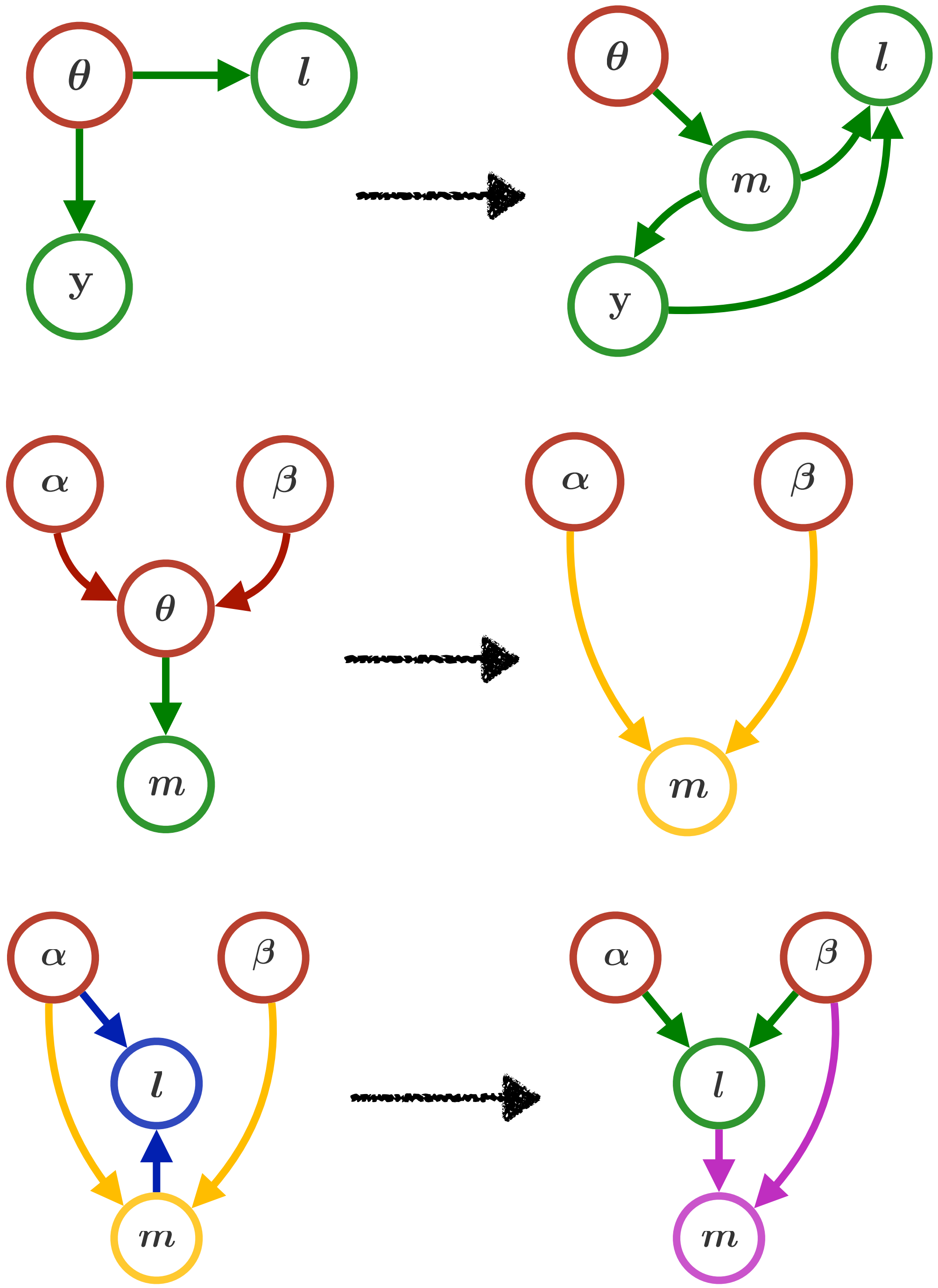
Steel (1953)



Three rules

Legend

- Poisson/Multinomial
- Gamma/Dirichlet



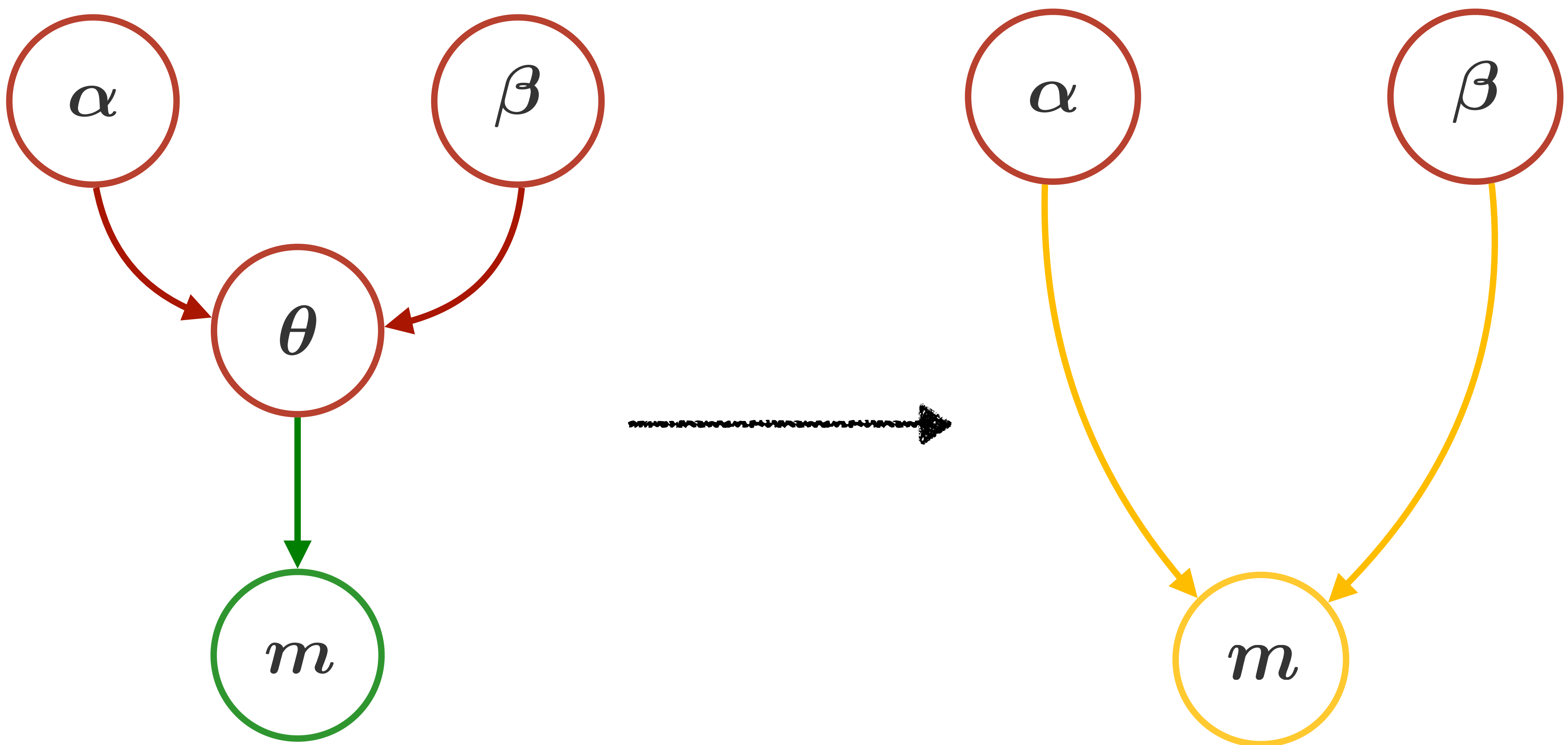
Poisson-multinomial relationship

Three rules

Legend
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial

Negative binomial as gamma-Poisson

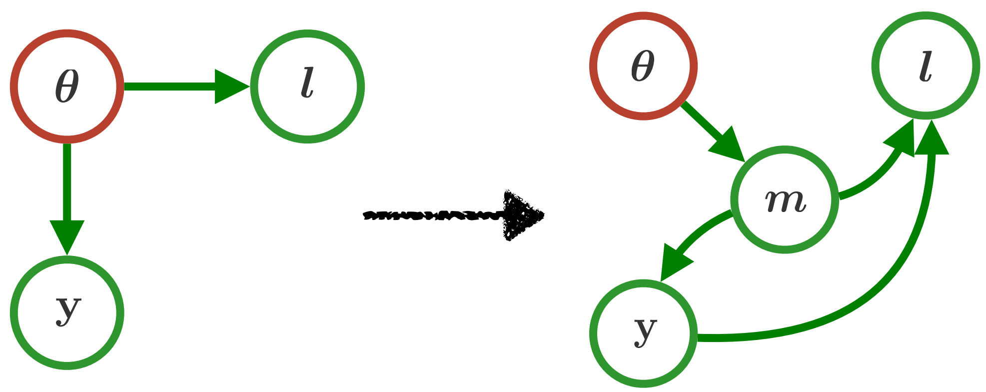
Greenwood & Yule (1920)



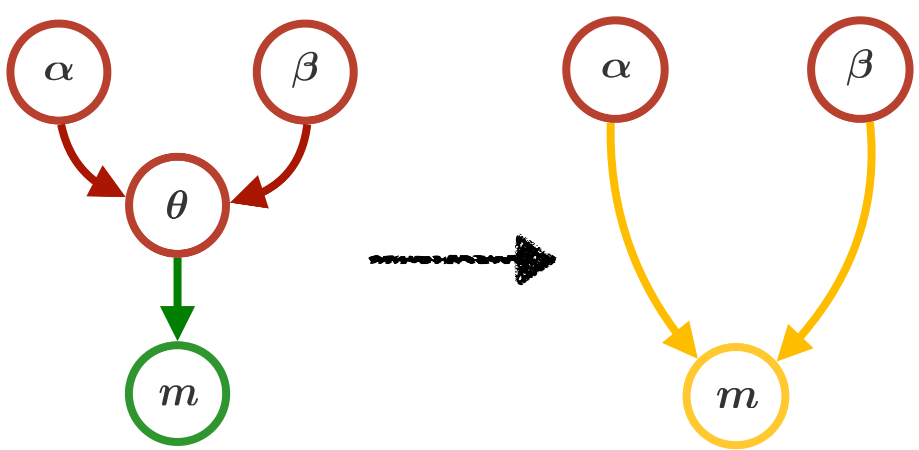
Three rules

Legend

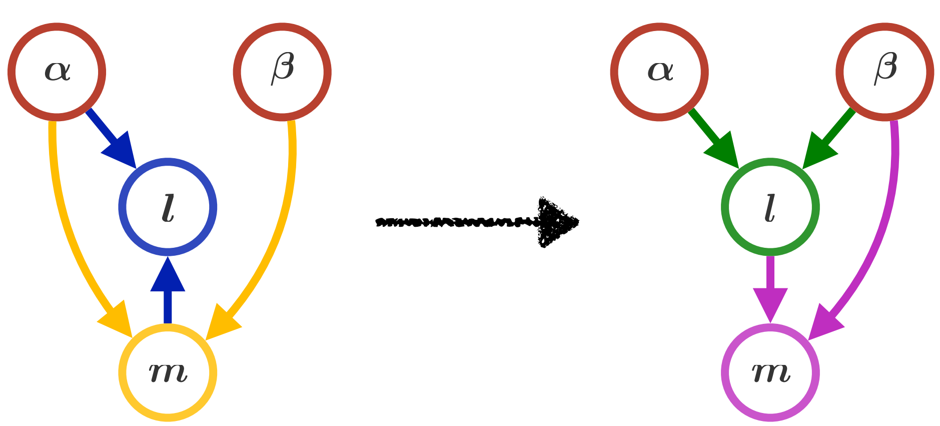
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial



Poisson-multinomial relationship



Negative binomial definition



Three rules

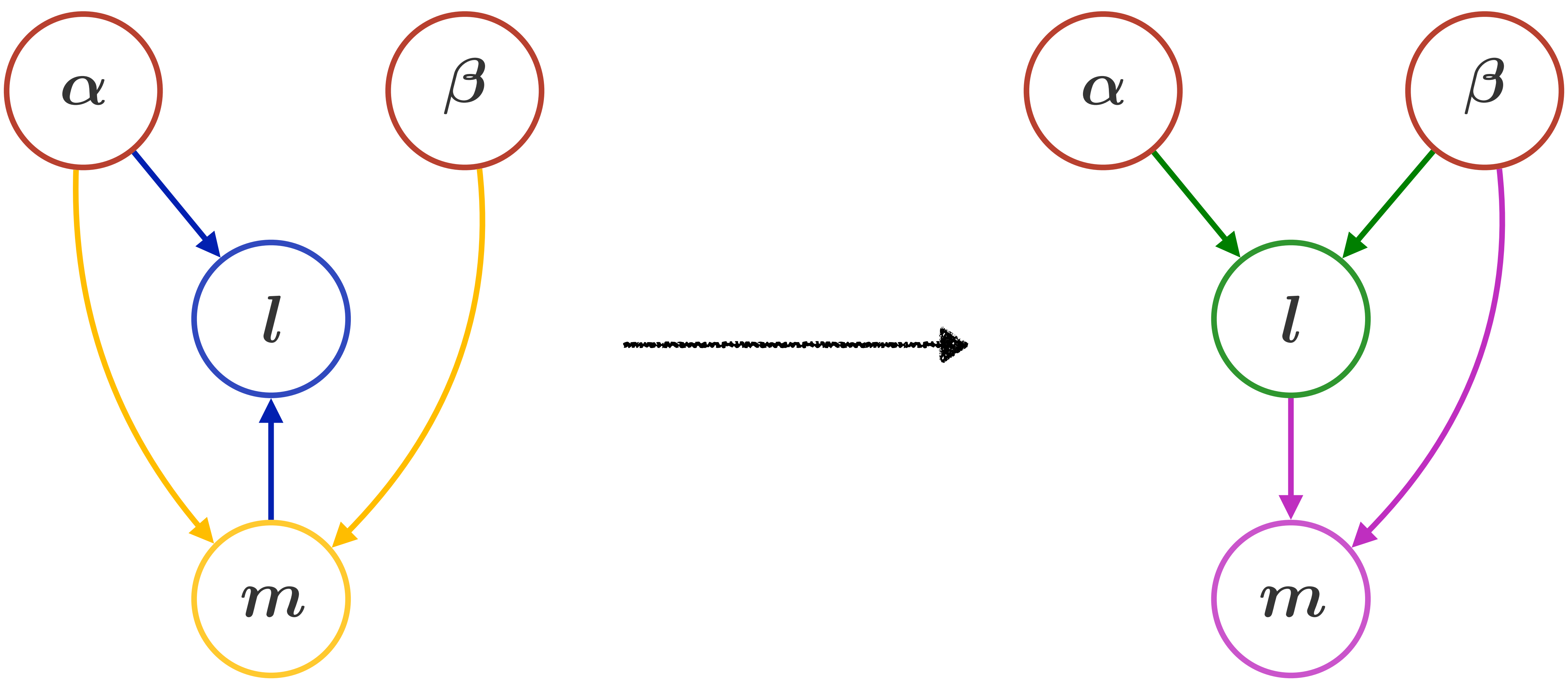
Legend

- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial

Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \mid \alpha, \beta)$$



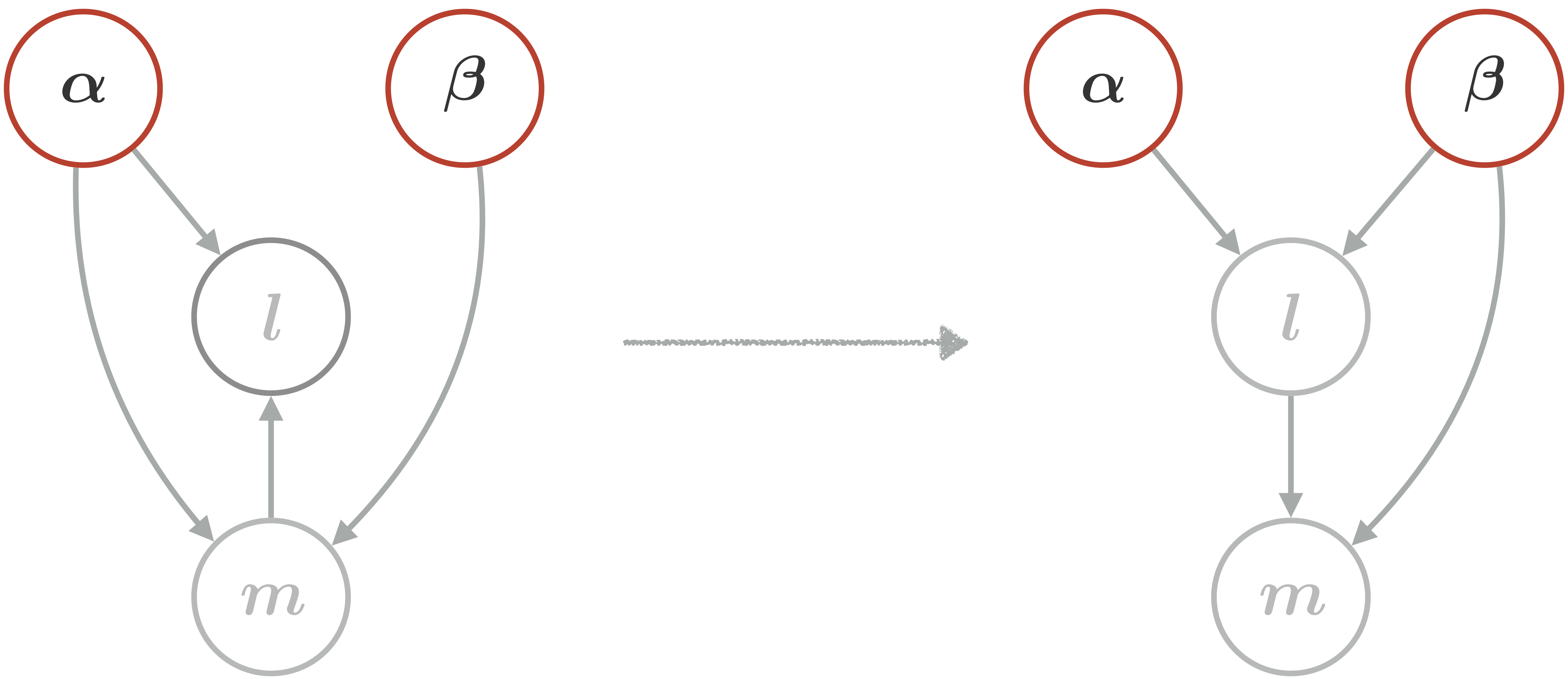
Three rules

Legend
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial

Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \,|\, \alpha, \beta)$$



Three rules

Legend
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial

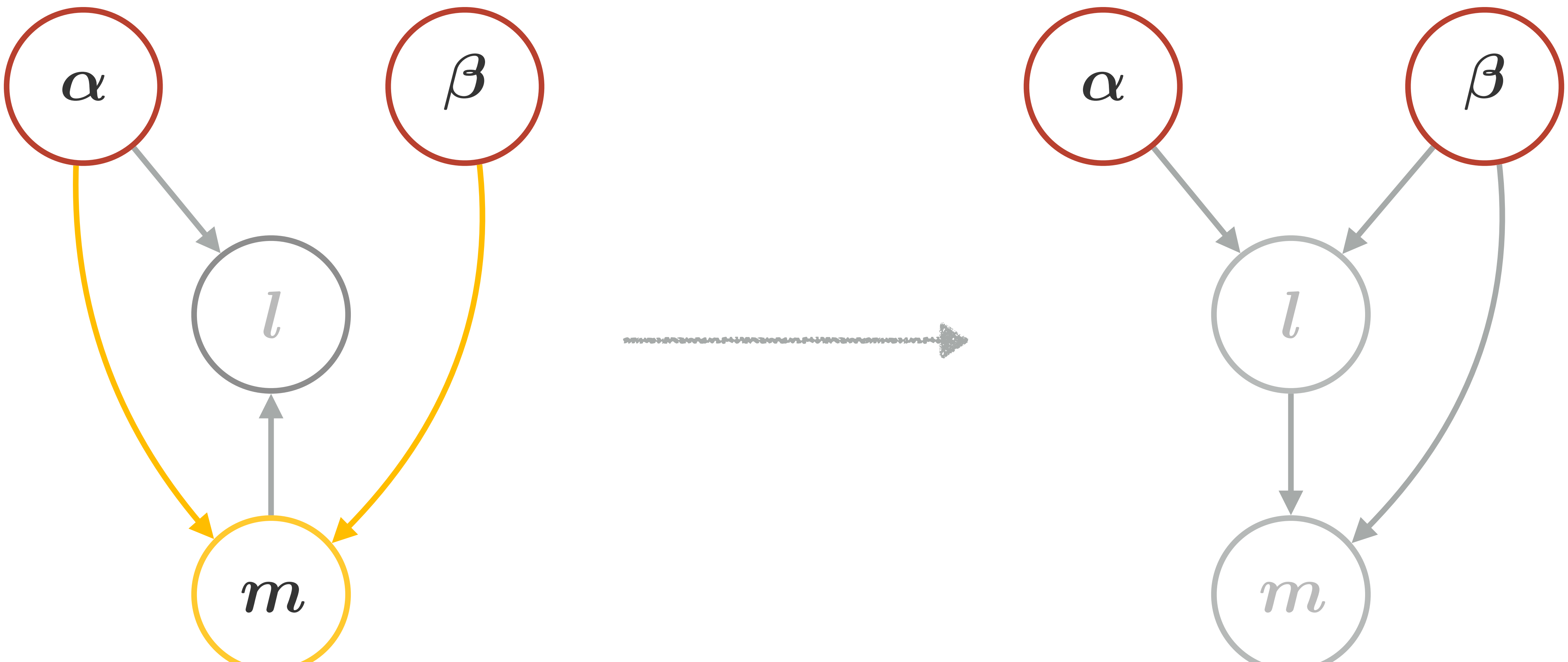
Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \mid \alpha, \beta)$$

//

$$P(m \mid \alpha, \beta)$$



Three rules

Legend

- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT

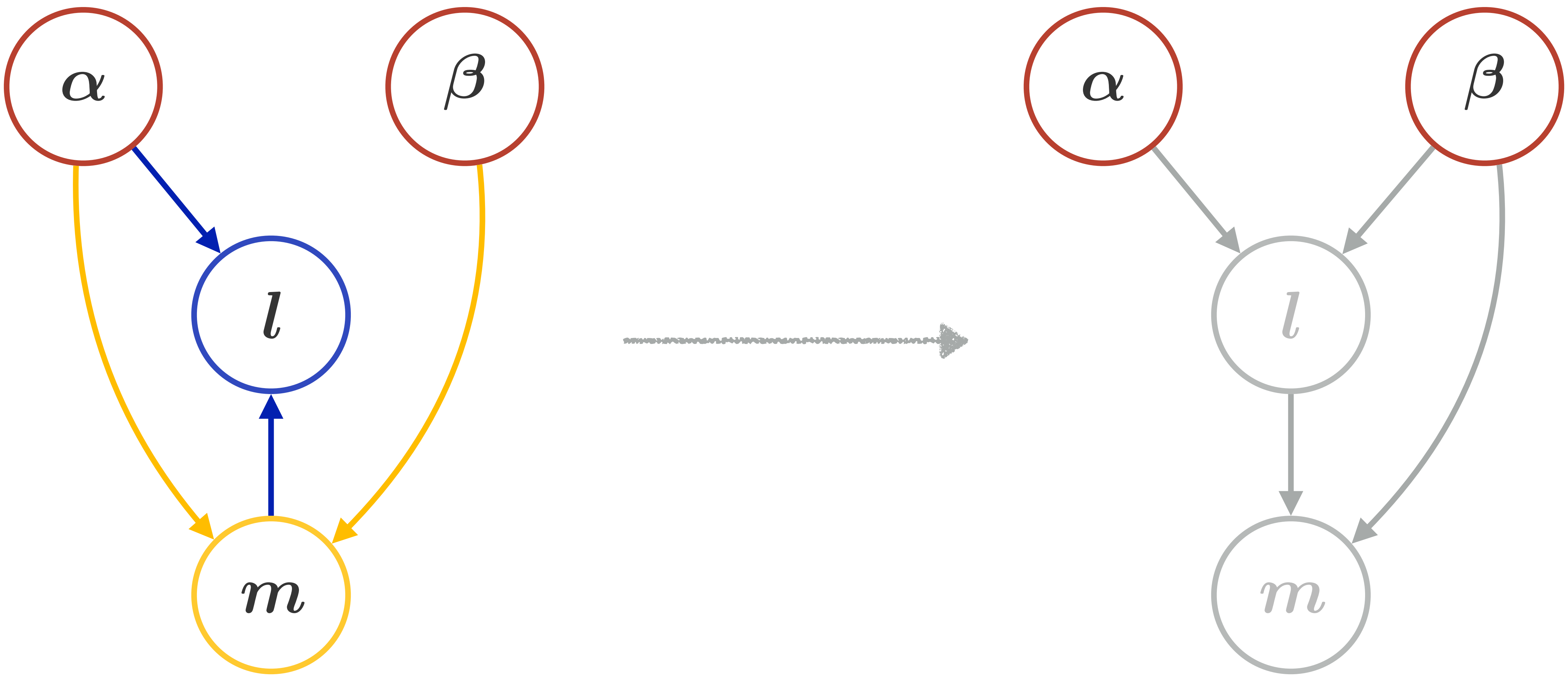
Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \mid \alpha, \beta)$$

//

$$P(l \mid m, \alpha)\, P(m \mid \alpha, \beta)$$



Three rules

Legend

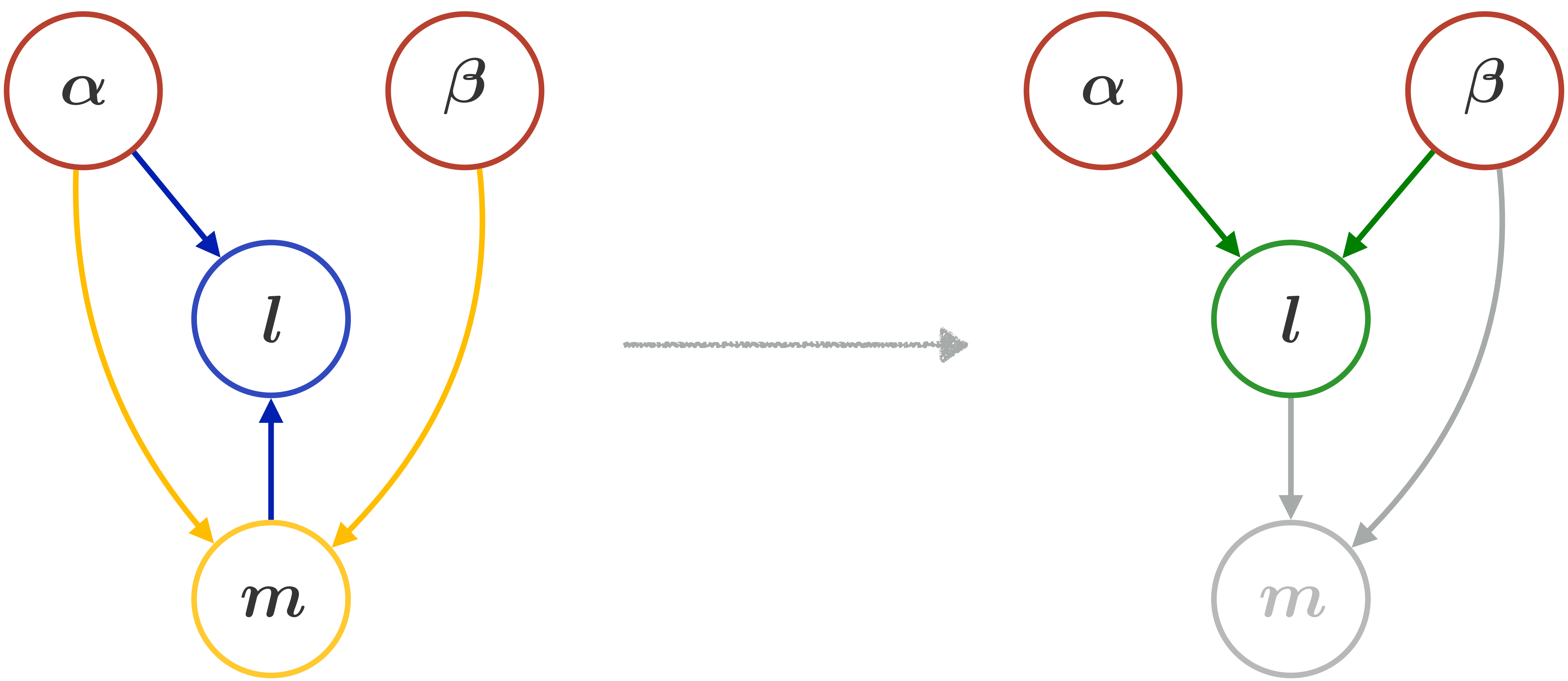
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT

Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \,|\, \alpha, \beta)$$

$$P(l \,|\, m, \alpha)\, P(m \,|\, \alpha, \beta) \qquad\qquad P(l \,|\, \alpha, \beta)$$



Three rules

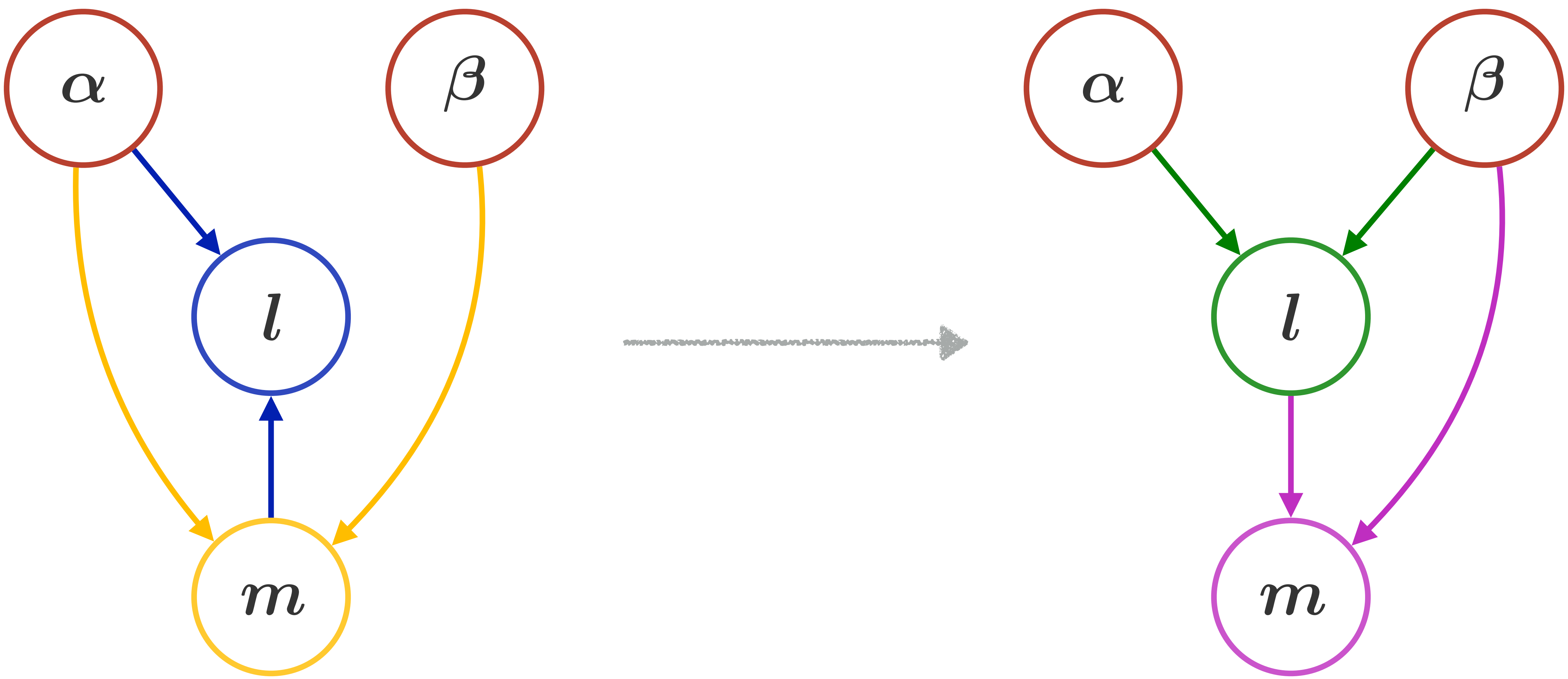
Legend
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog

Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \mid \alpha, \beta)$$

$$= P(l \mid m, \alpha)\, P(m \mid \alpha, \beta) \qquad = P(l \mid \alpha, \beta)\, P(m \mid l, \beta)$$



Three rules

Legend

- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog

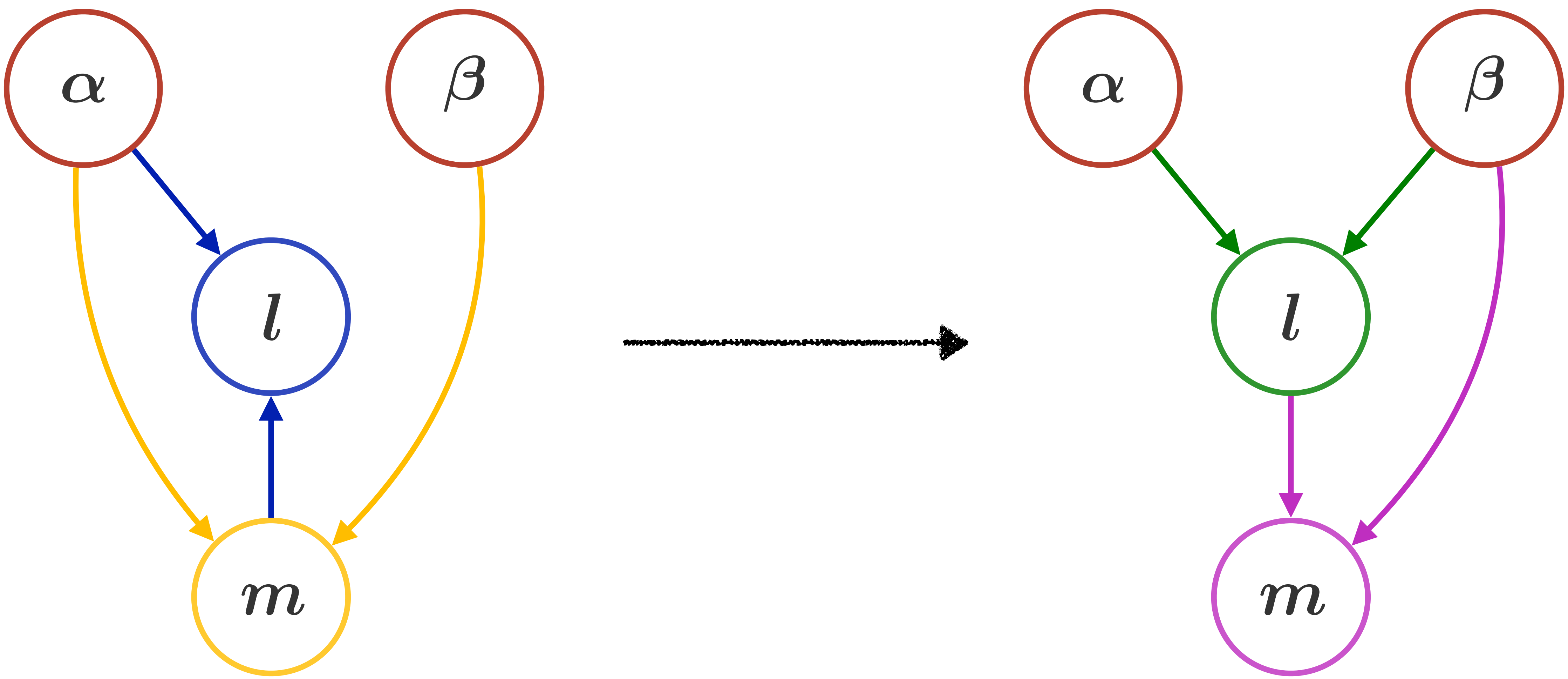
Magic bivariate count distribution

Zhou & Carin (2012)

$$P(l, m \mid \alpha, \beta)$$

$$= P(l \mid m, \alpha)\, P(m \mid \alpha, \beta)$$

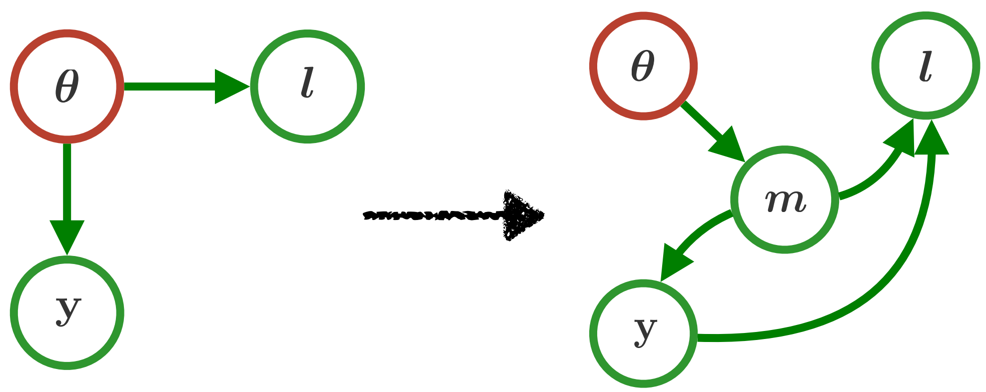
$$= P(l \mid \alpha, \beta)\, P(m \mid l, \beta)$$



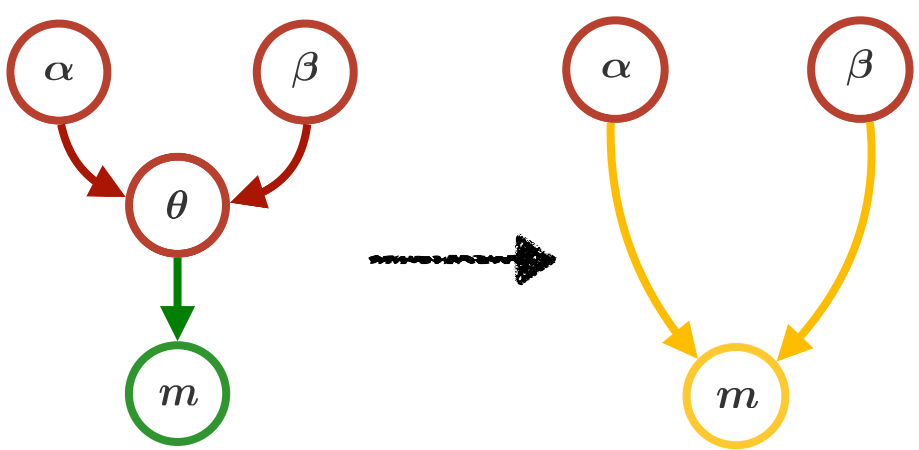
Three rules

Legend

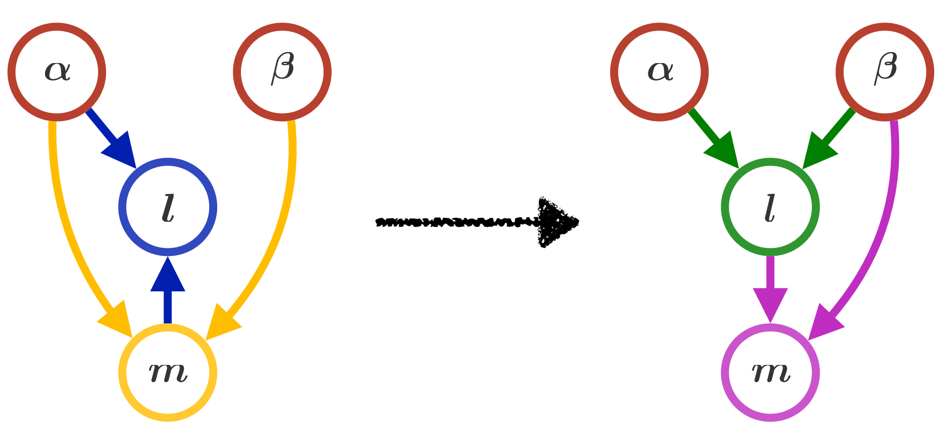
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog



Poisson-multinomial relationship



Negative binomial definition

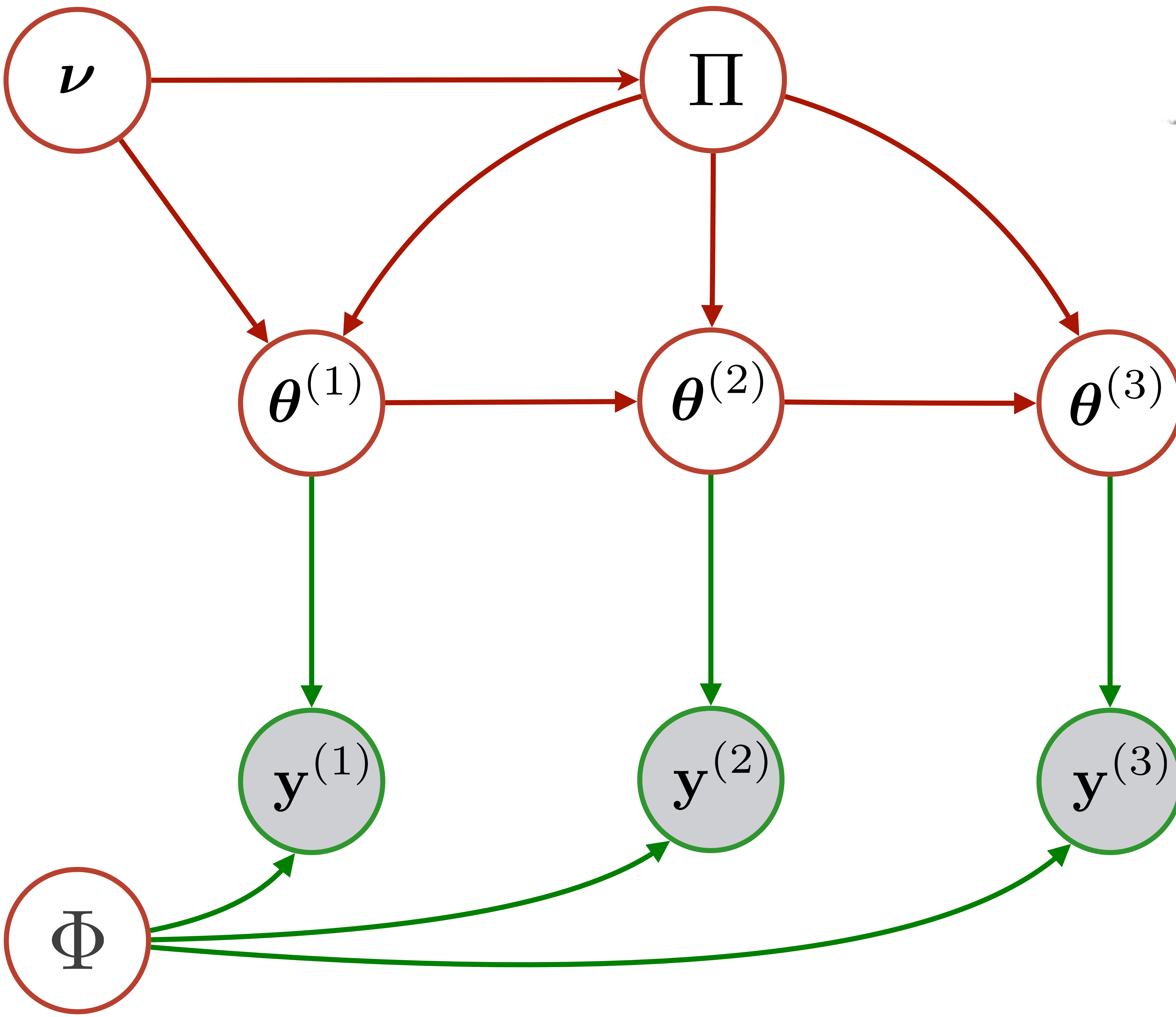


Magic bivariate count distribution

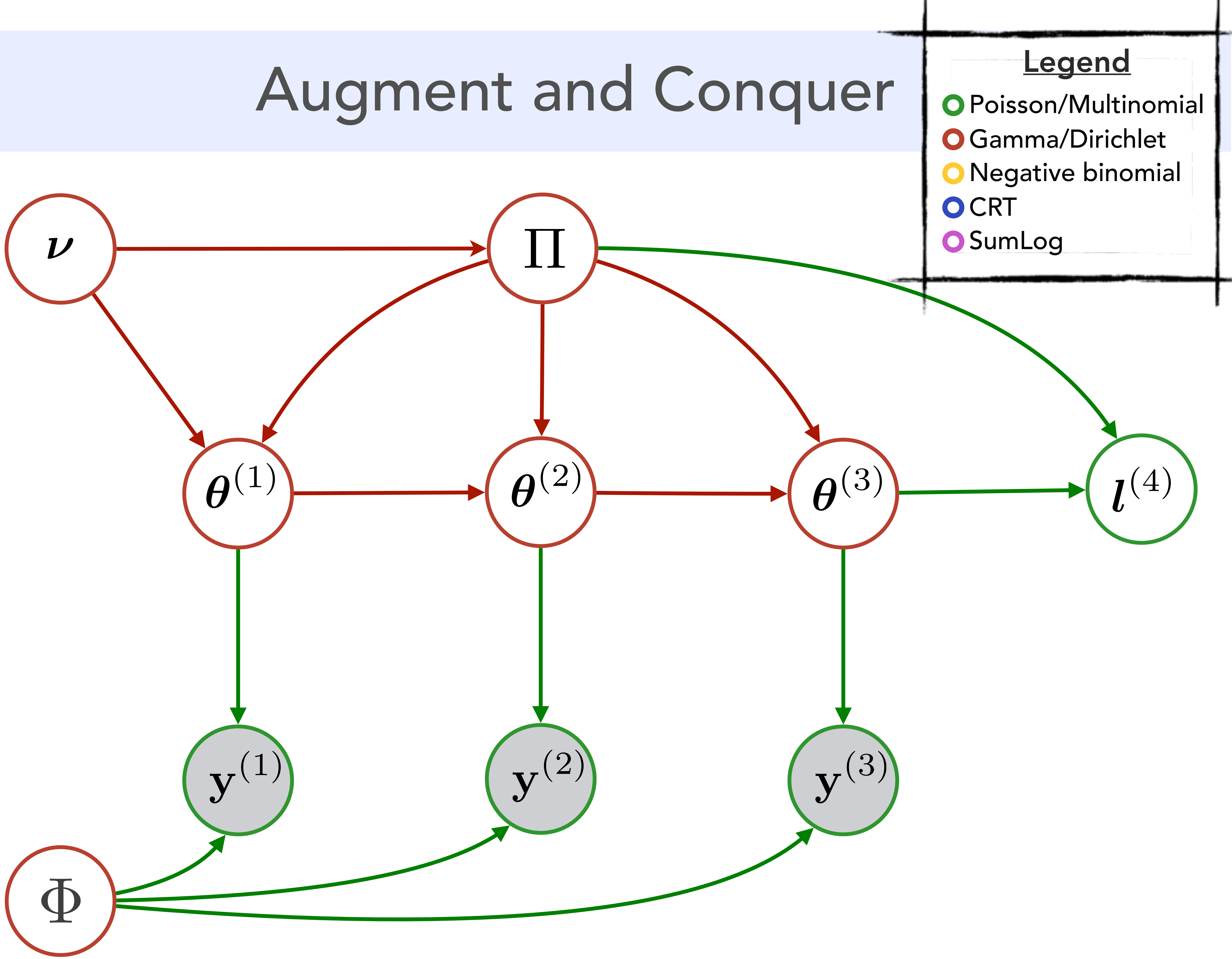
Augment and Conquer

Legend

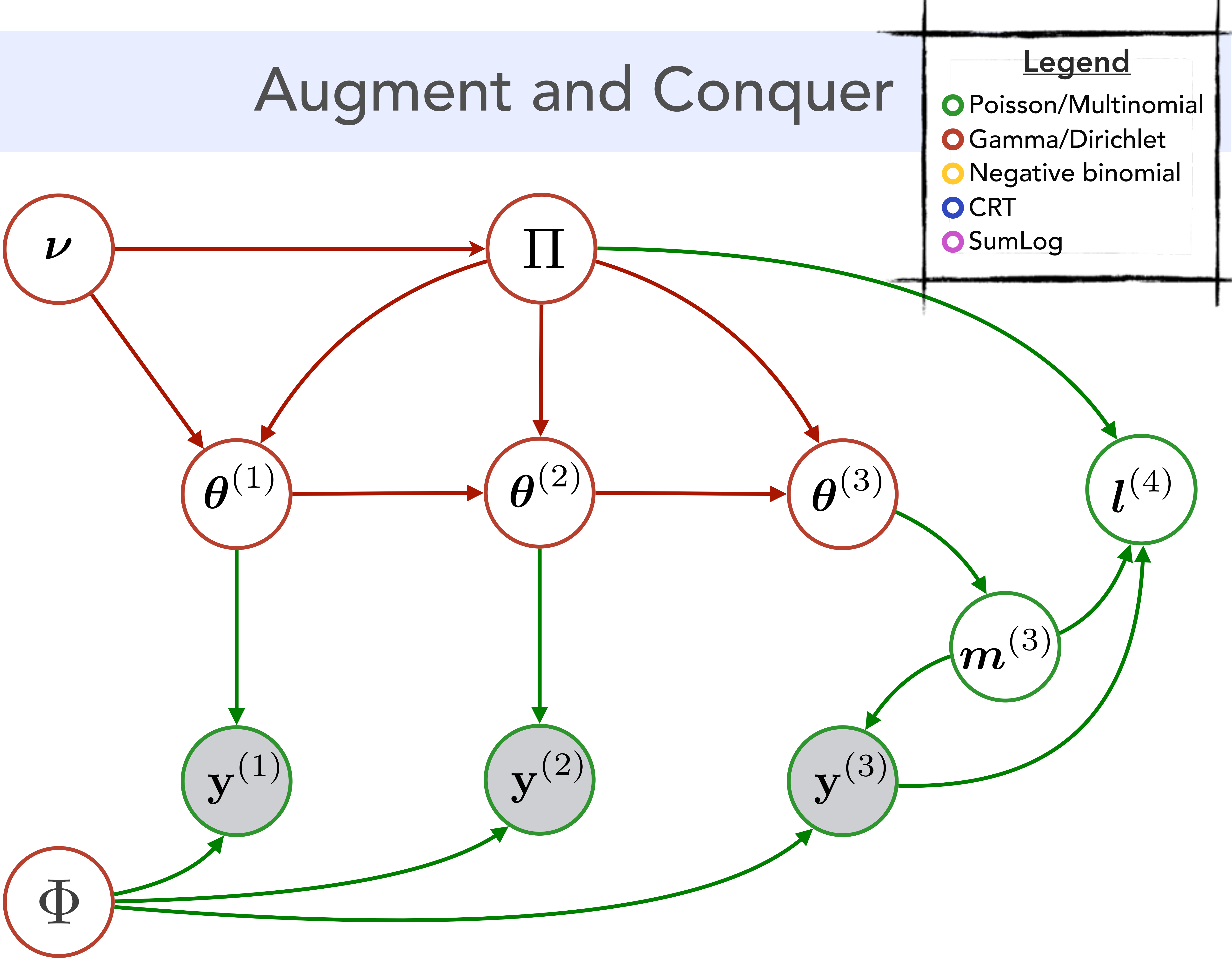
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog



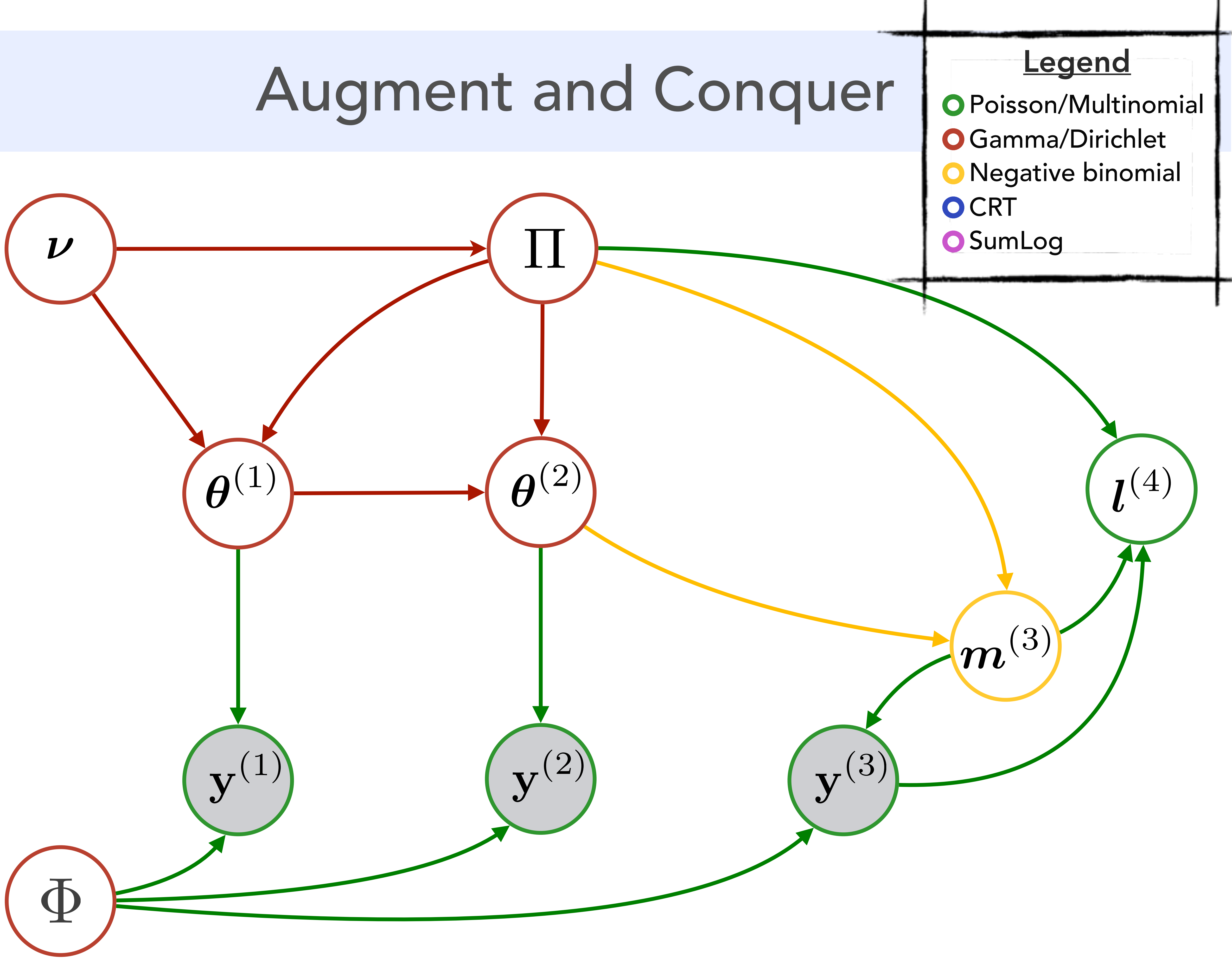
Augment and Conquer



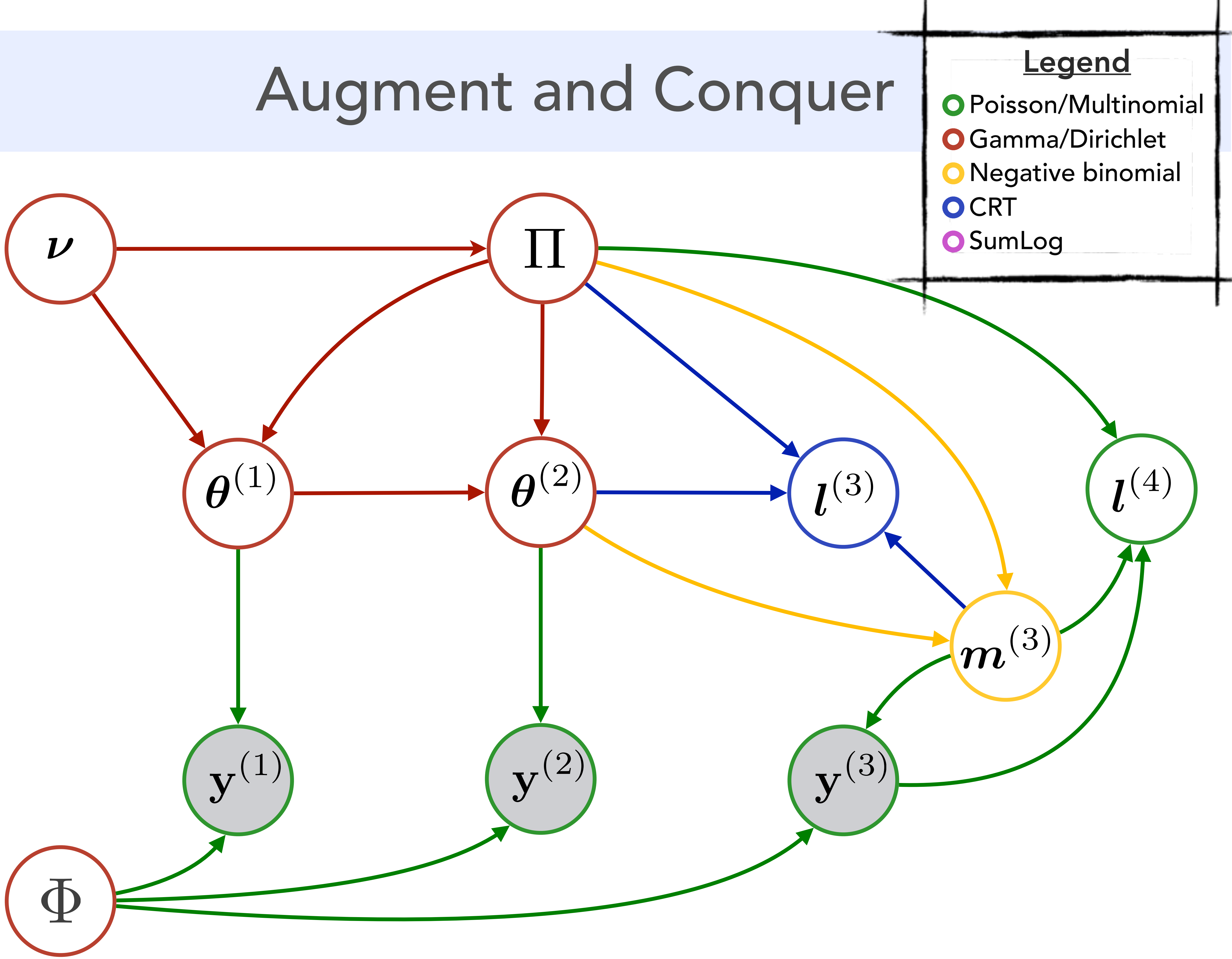
Augment and Conquer



Augment and Conquer



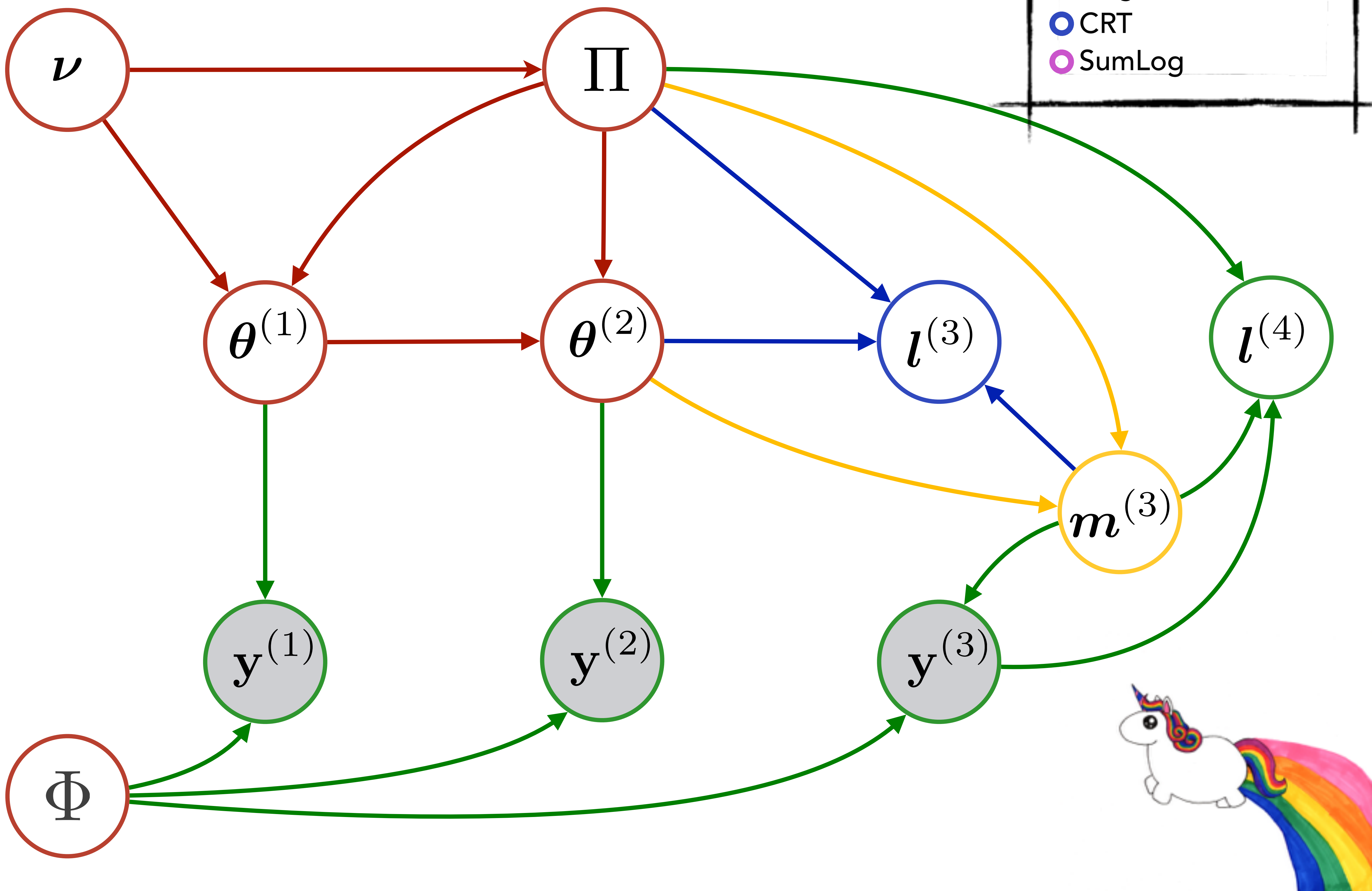
Augment and Conquer



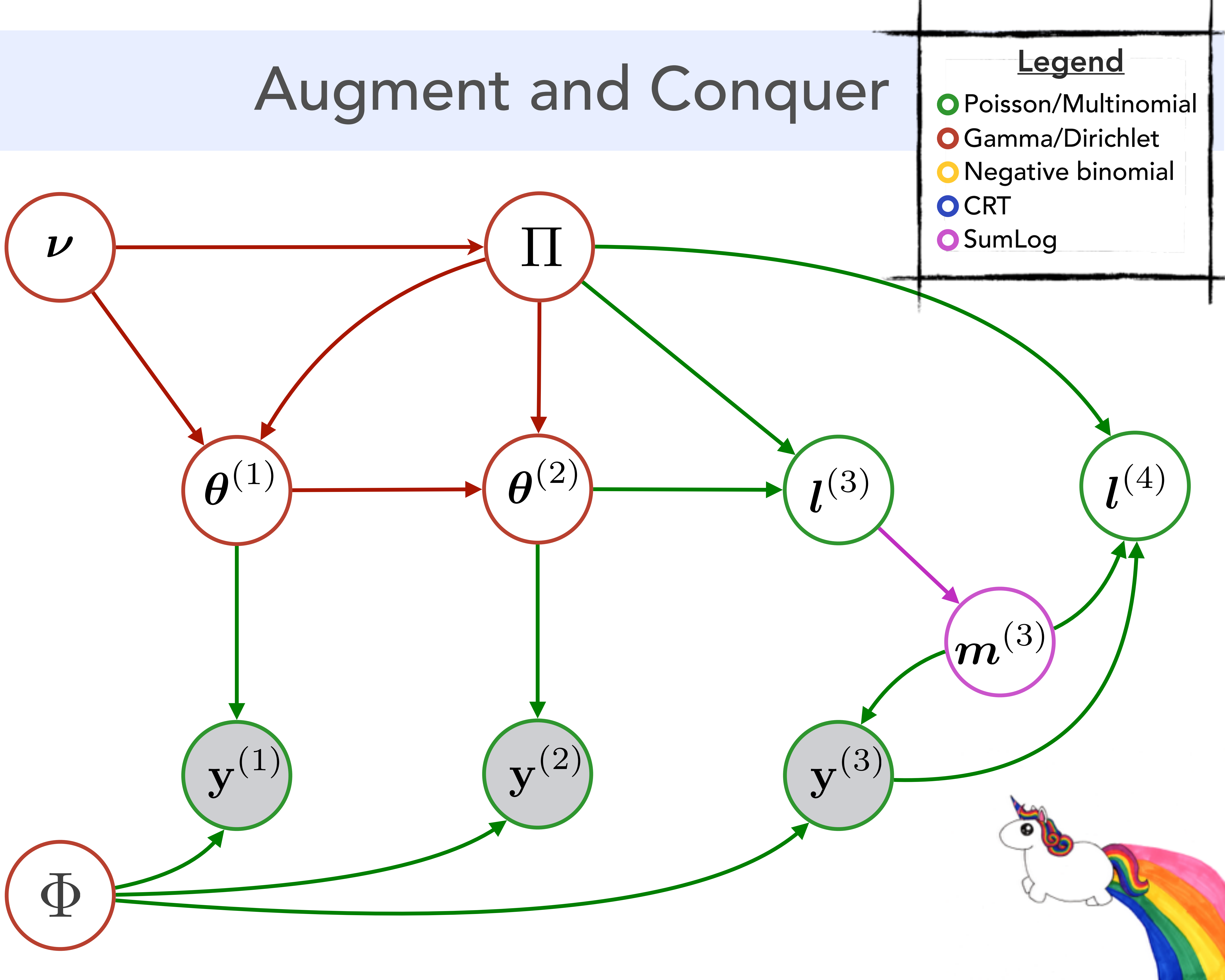
Augment and Conquer

Legend

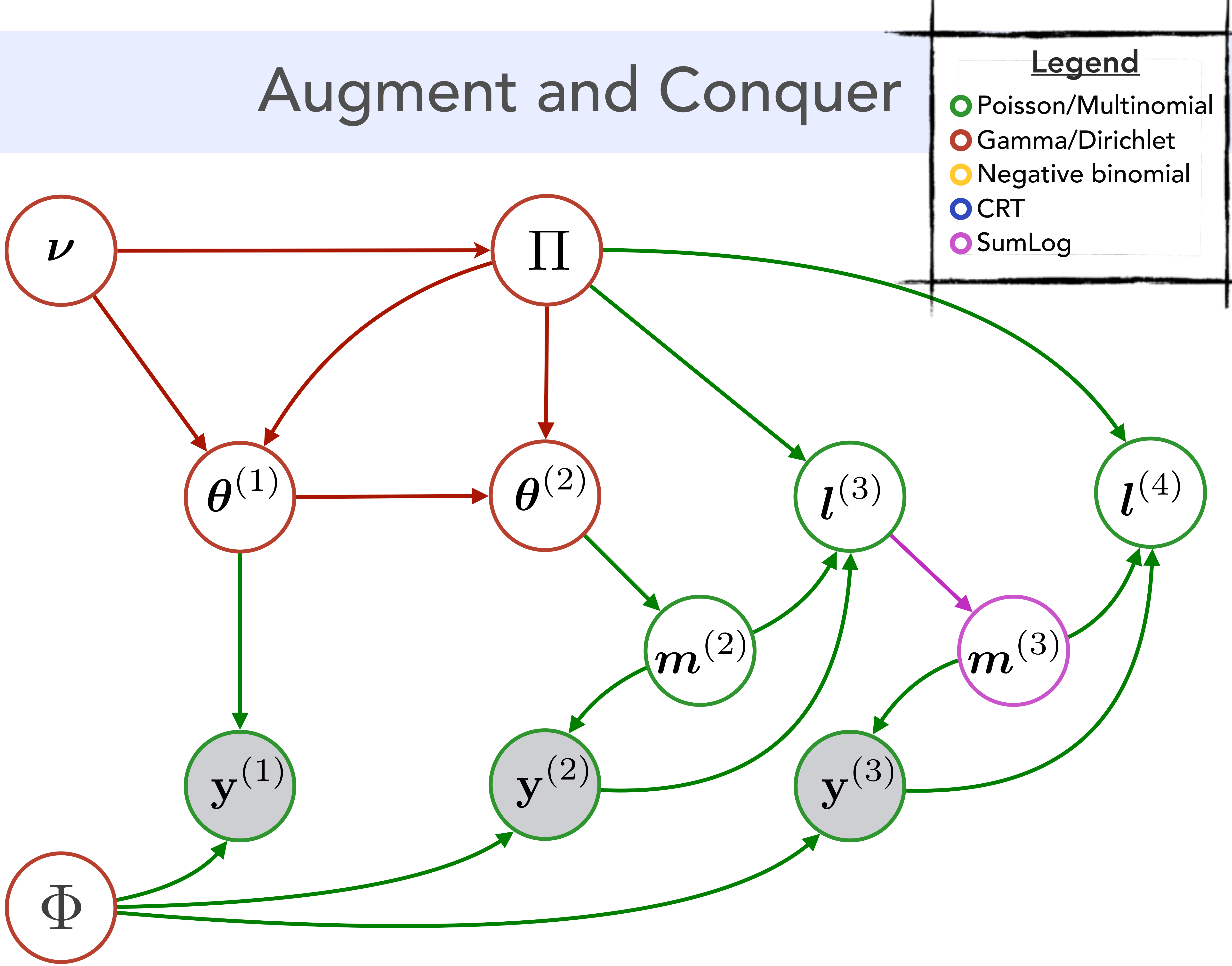
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog



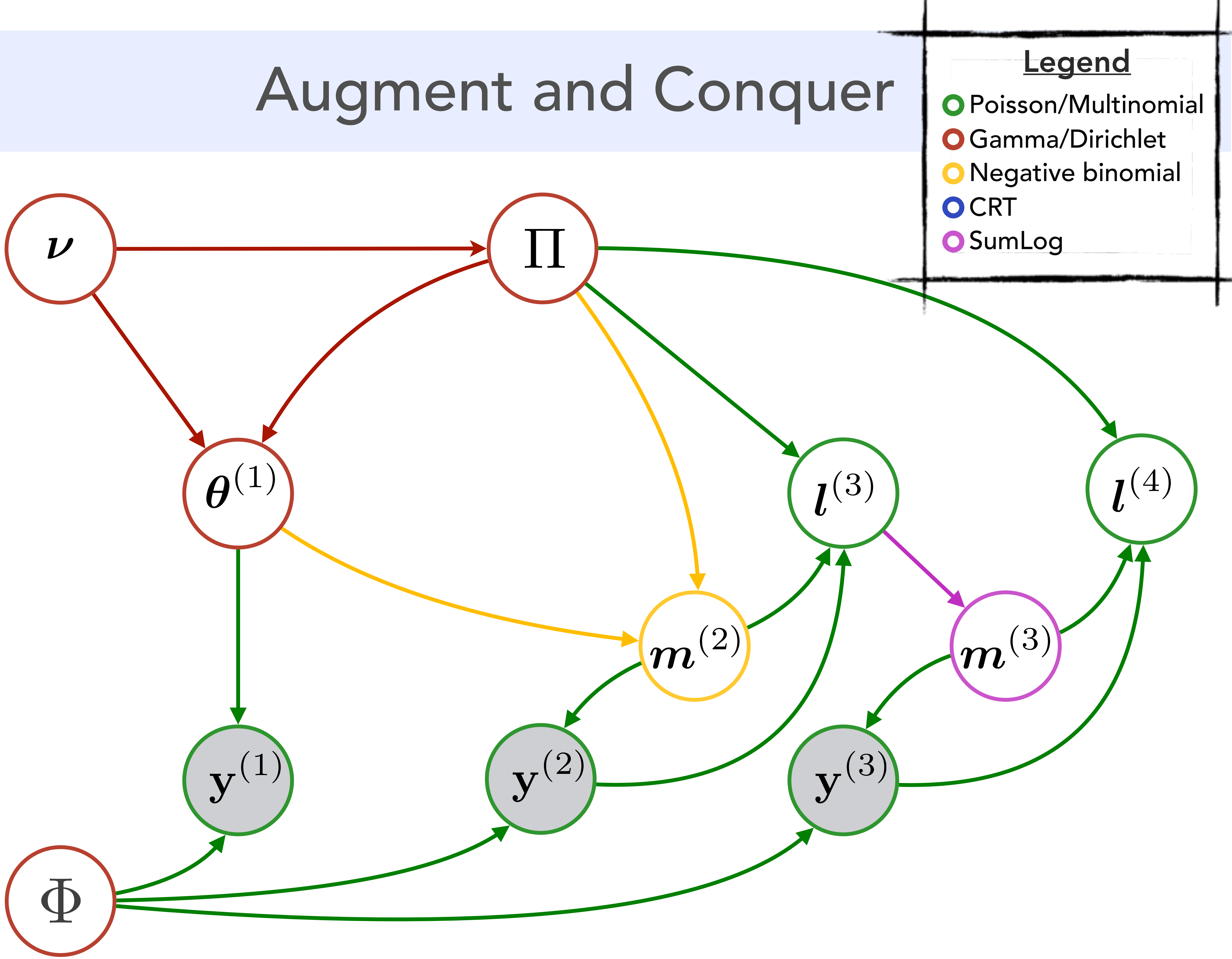
Augment and Conquer



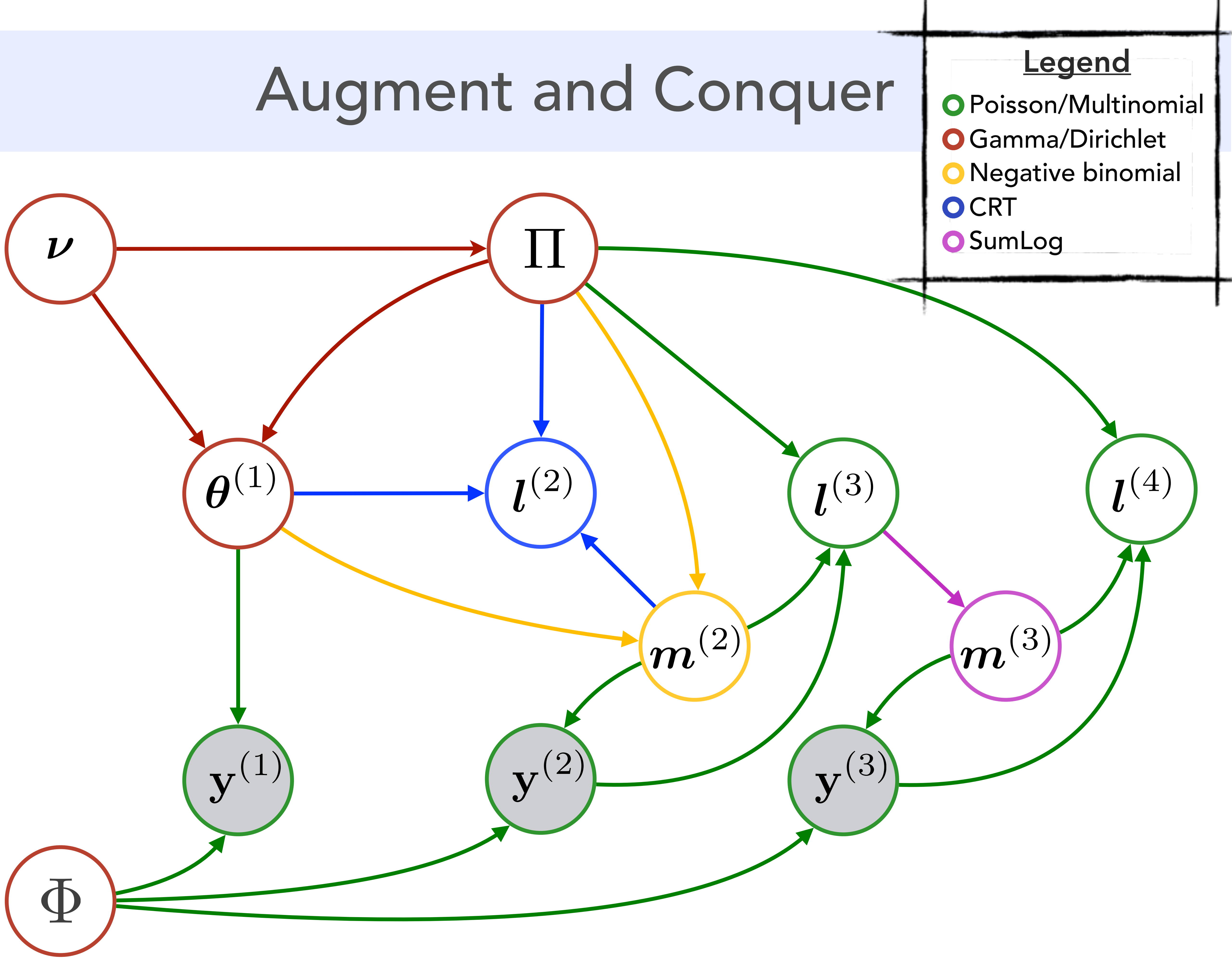
Augment and Conquer



Augment and Conquer



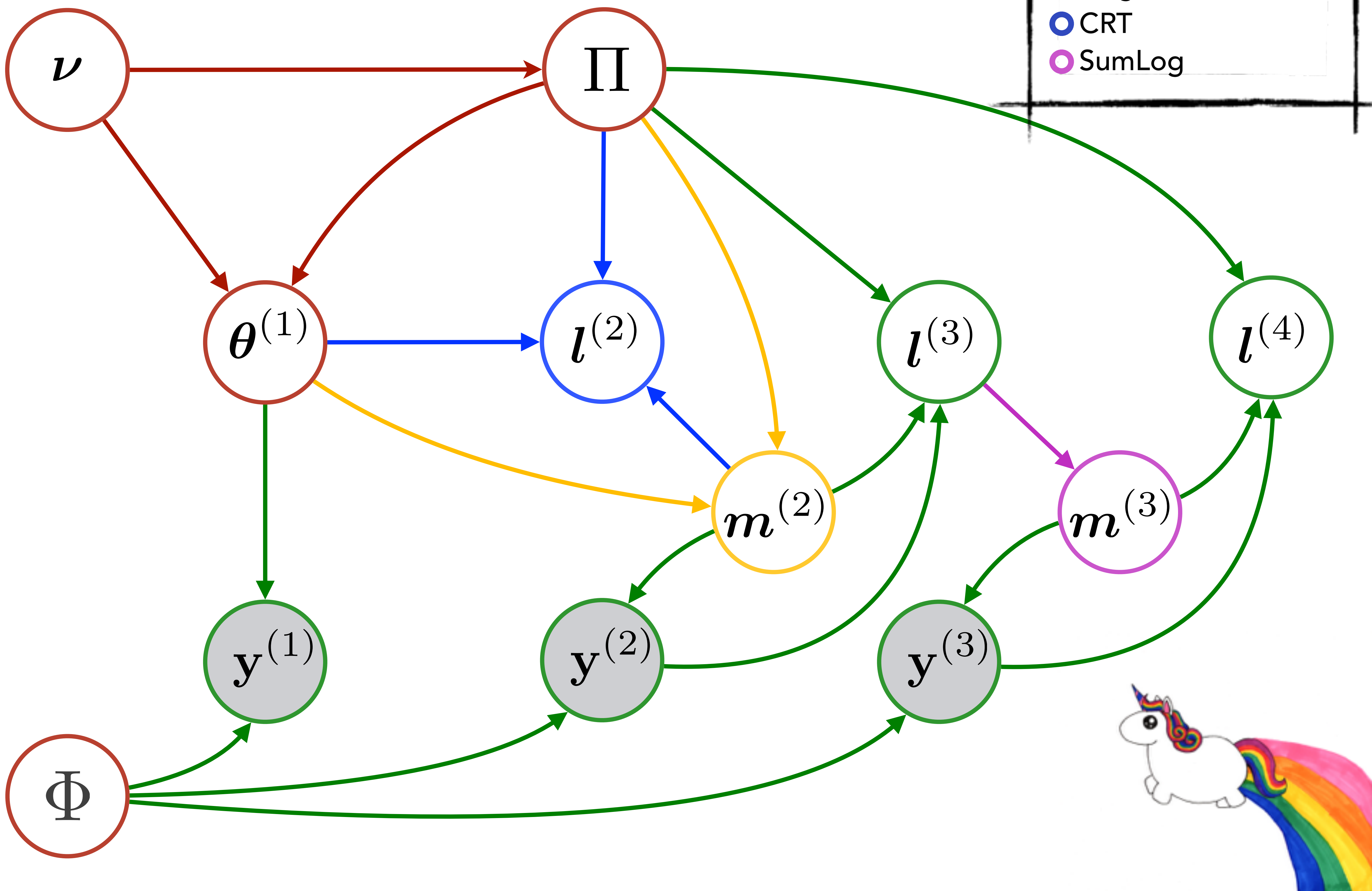
Augment and Conquer



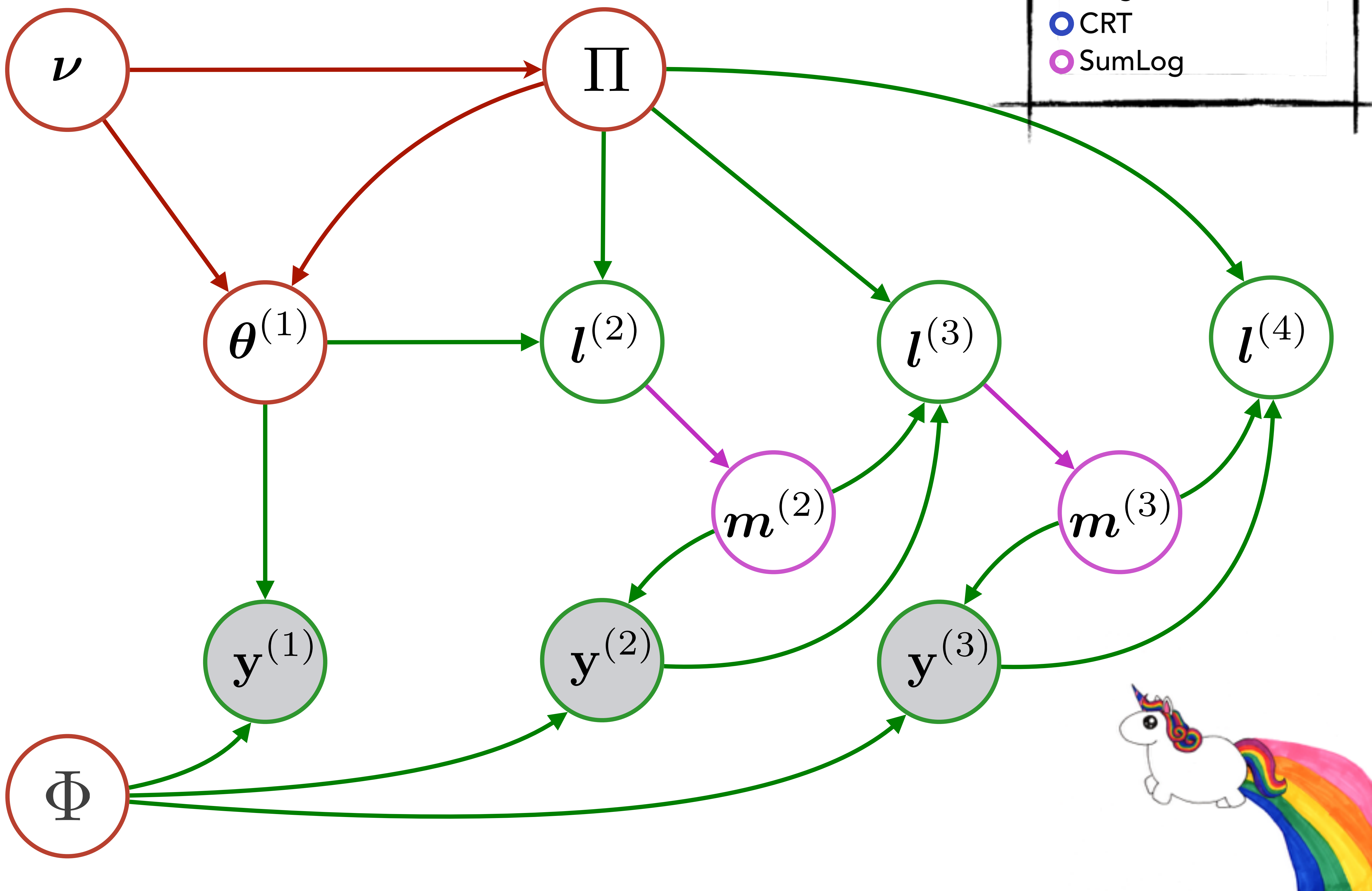
Augment and Conquer

Legend

- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog



Augment and Conquer



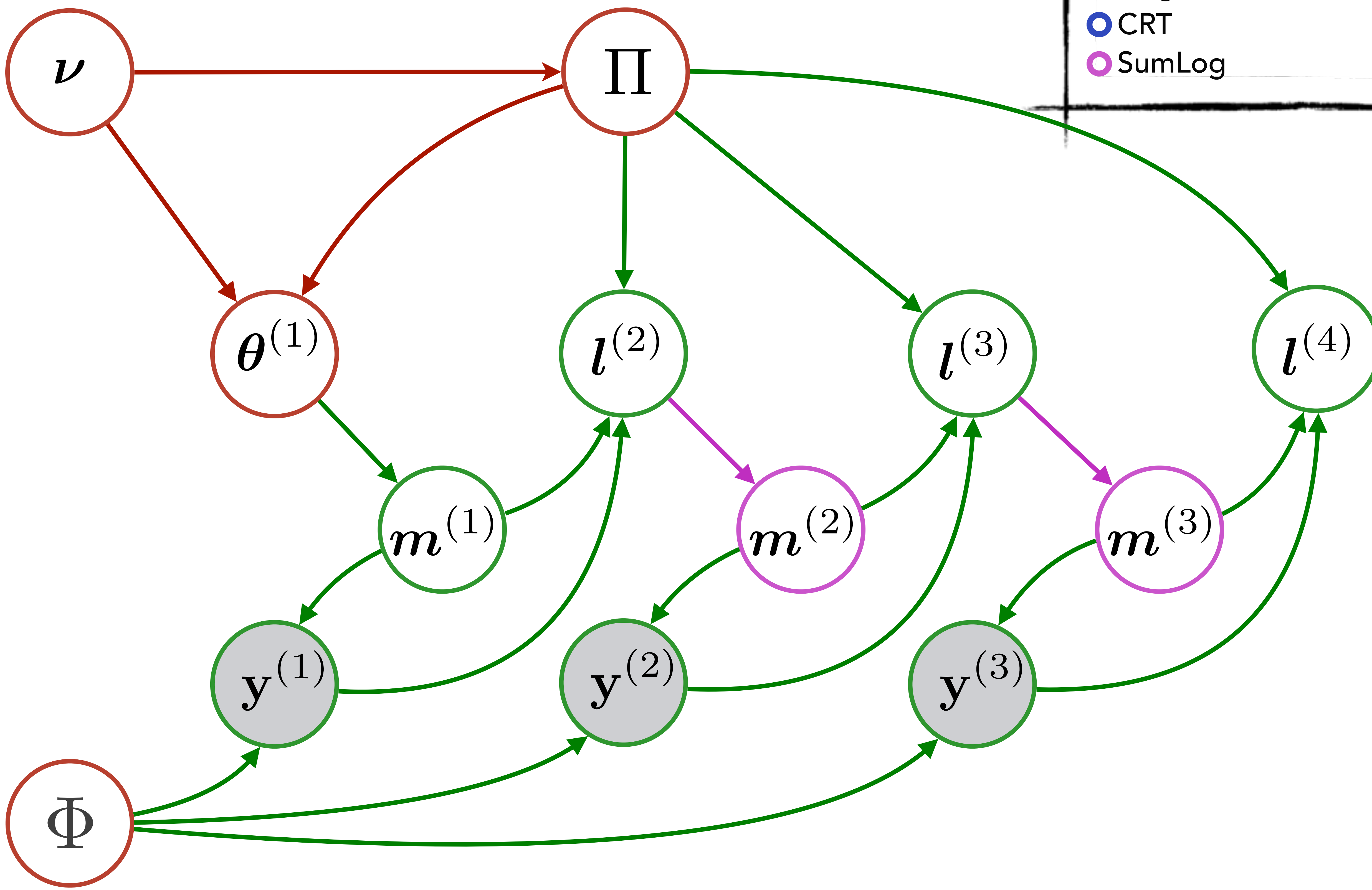
Legend

- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog

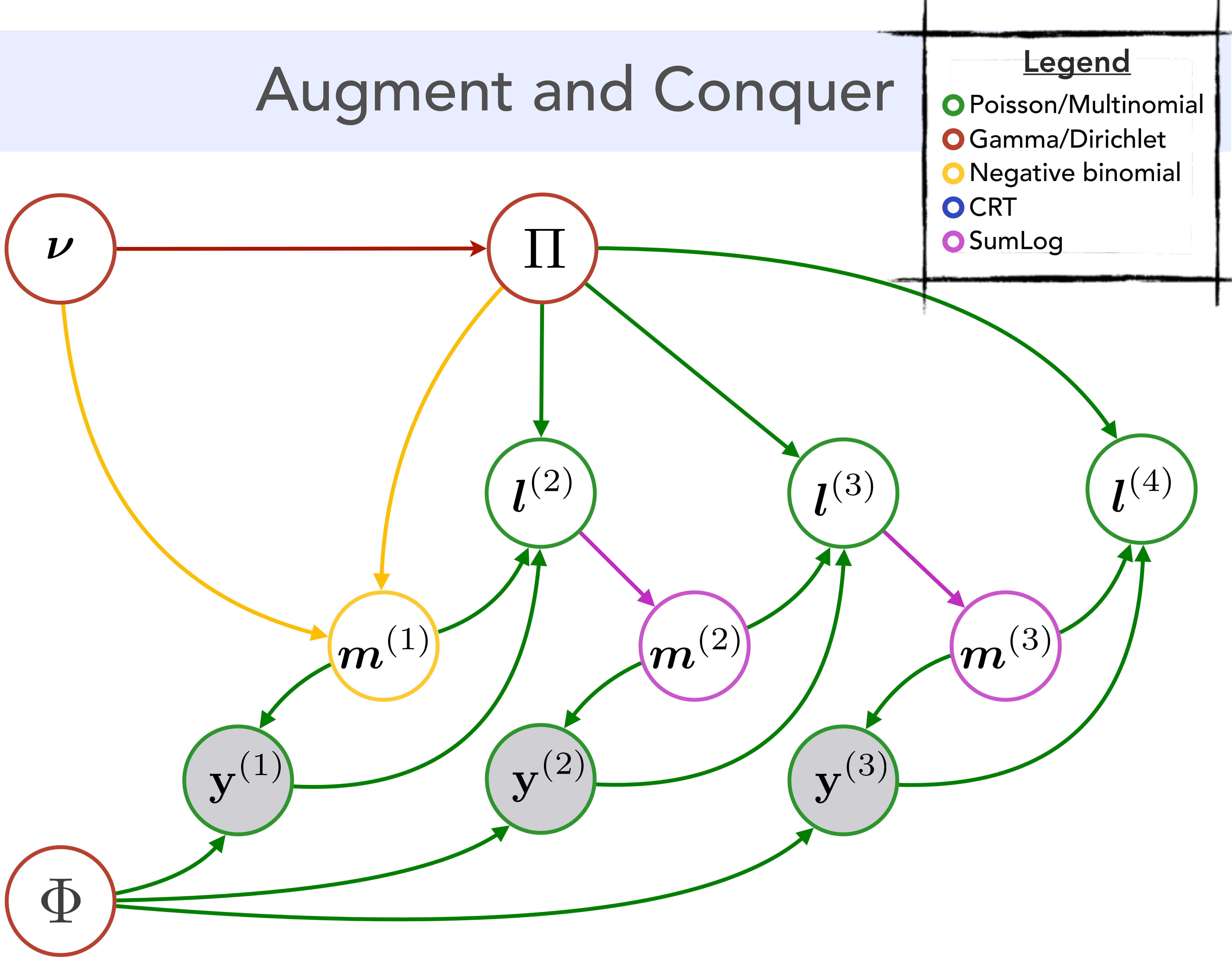
Augment and Conquer

Legend

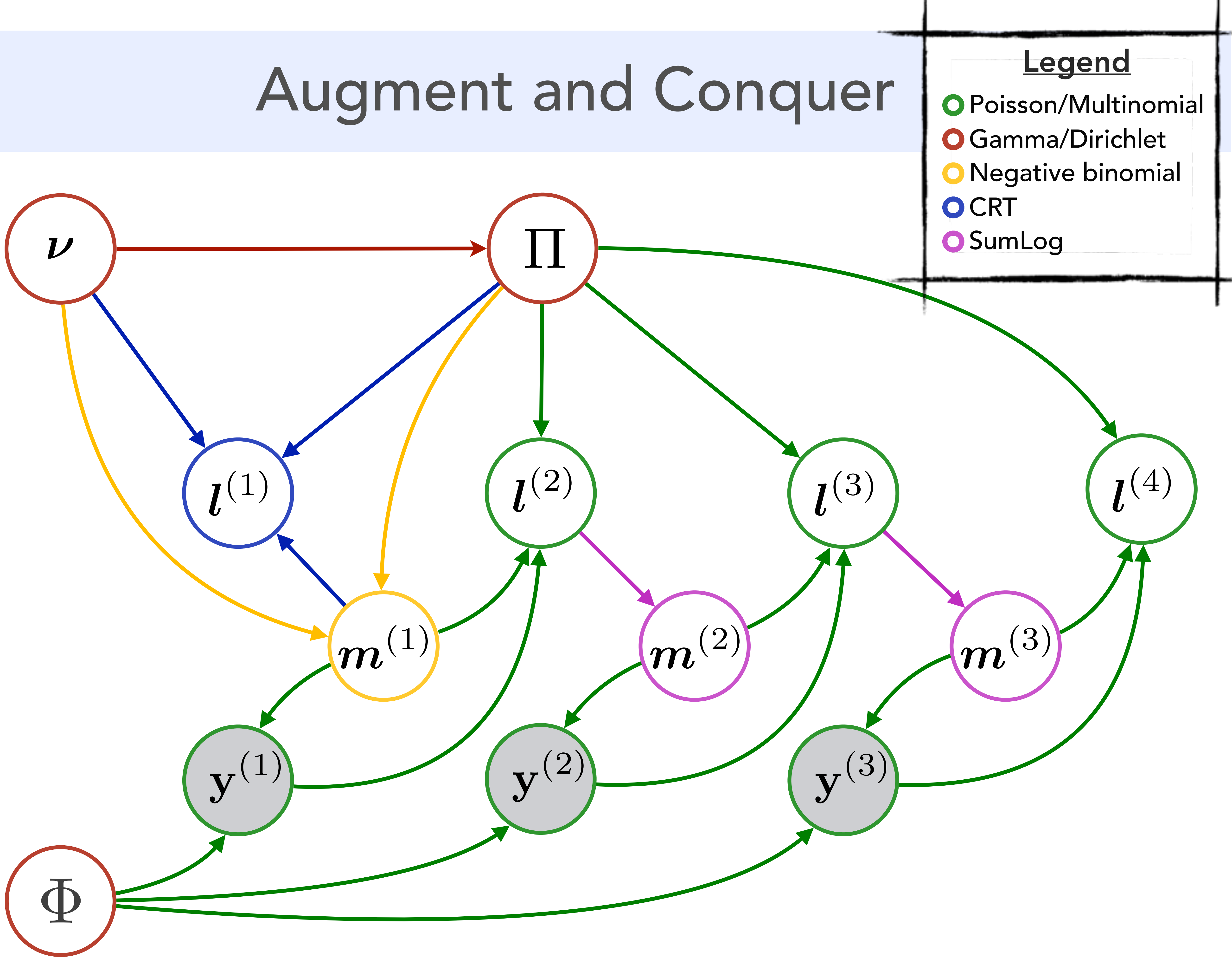
- Poisson/Multinomial
- Gamma/Dirichlet
- Negative binomial
- CRT
- SumLog



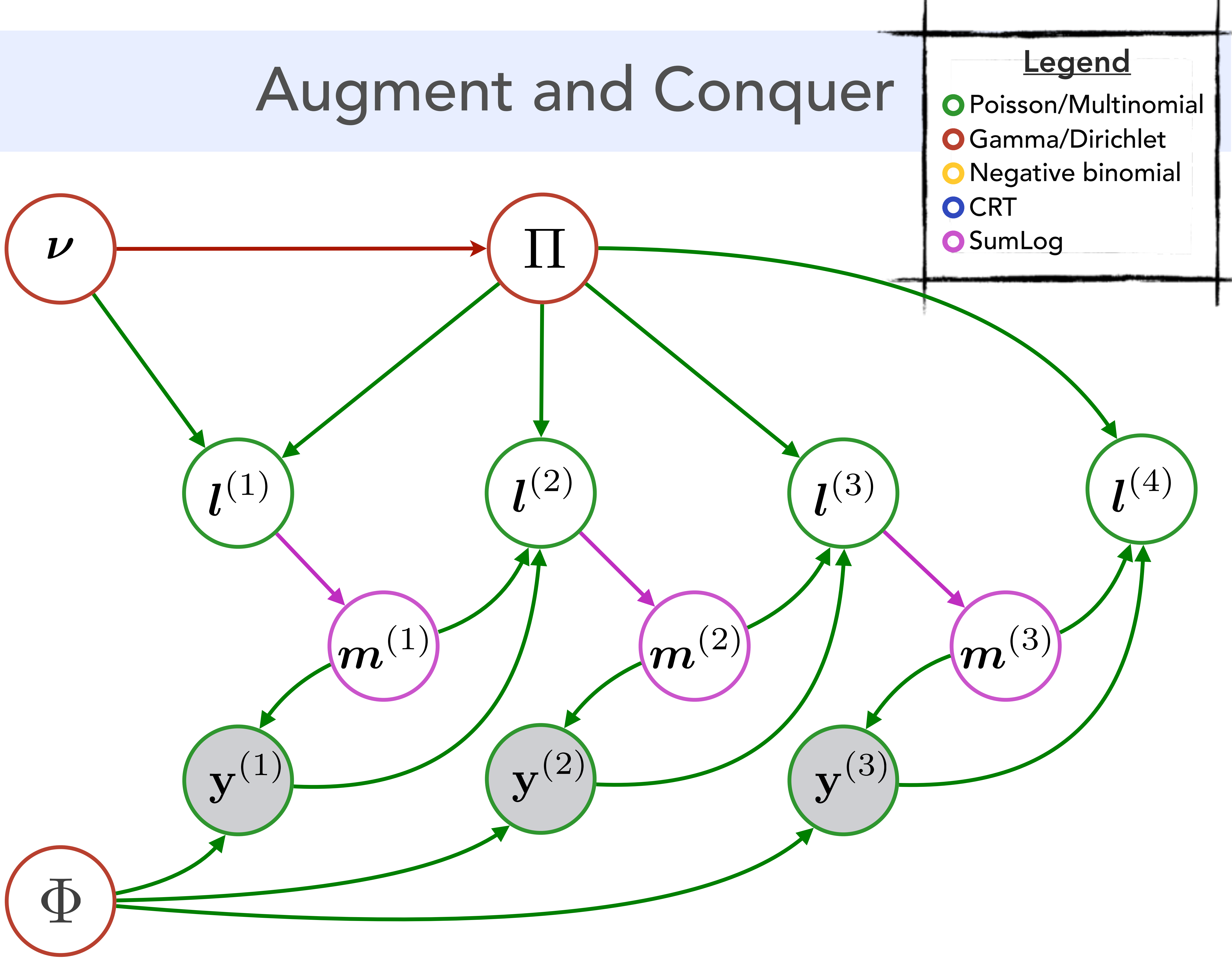
Augment and Conquer



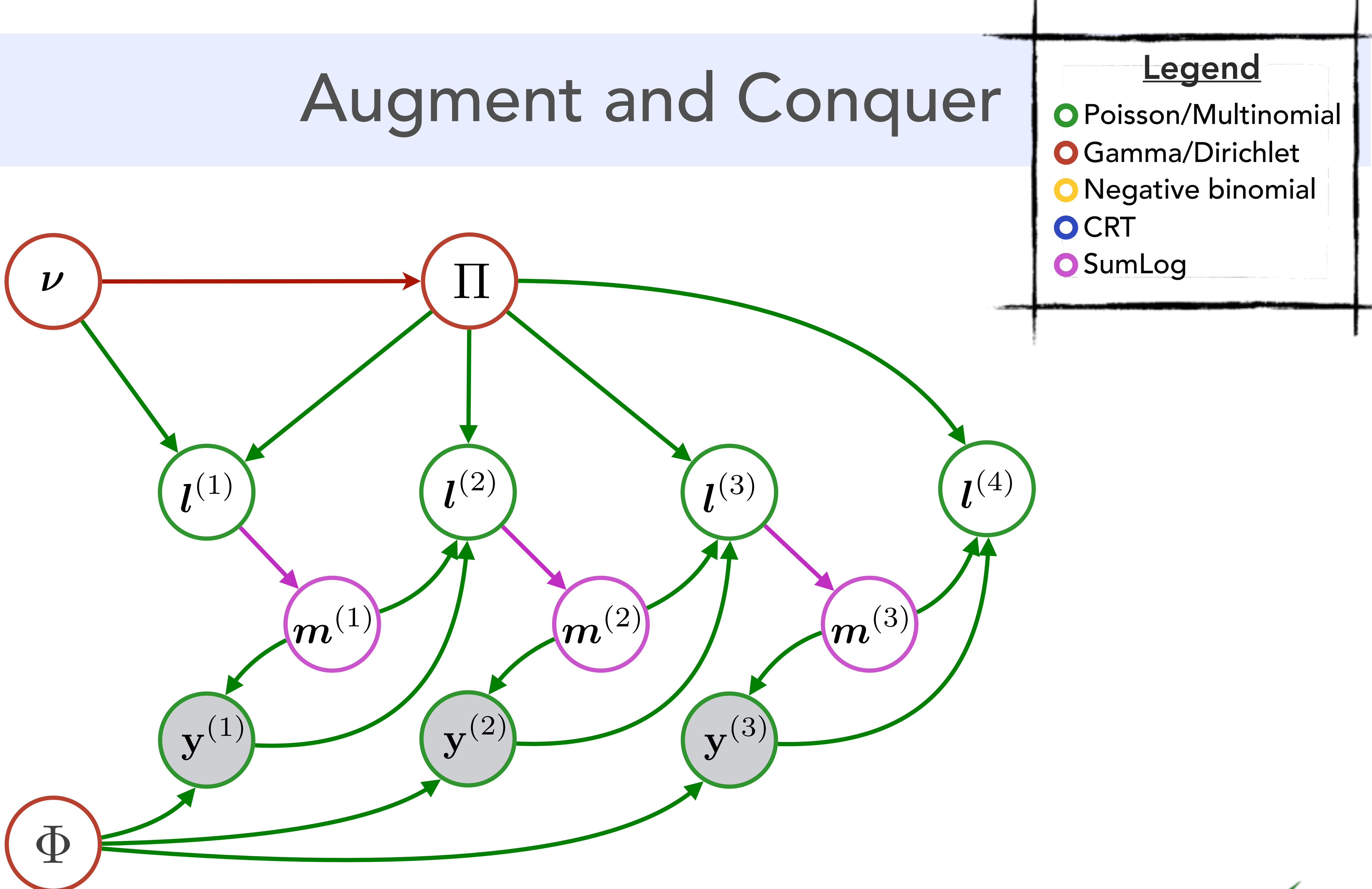
Augment and Conquer



Augment and Conquer

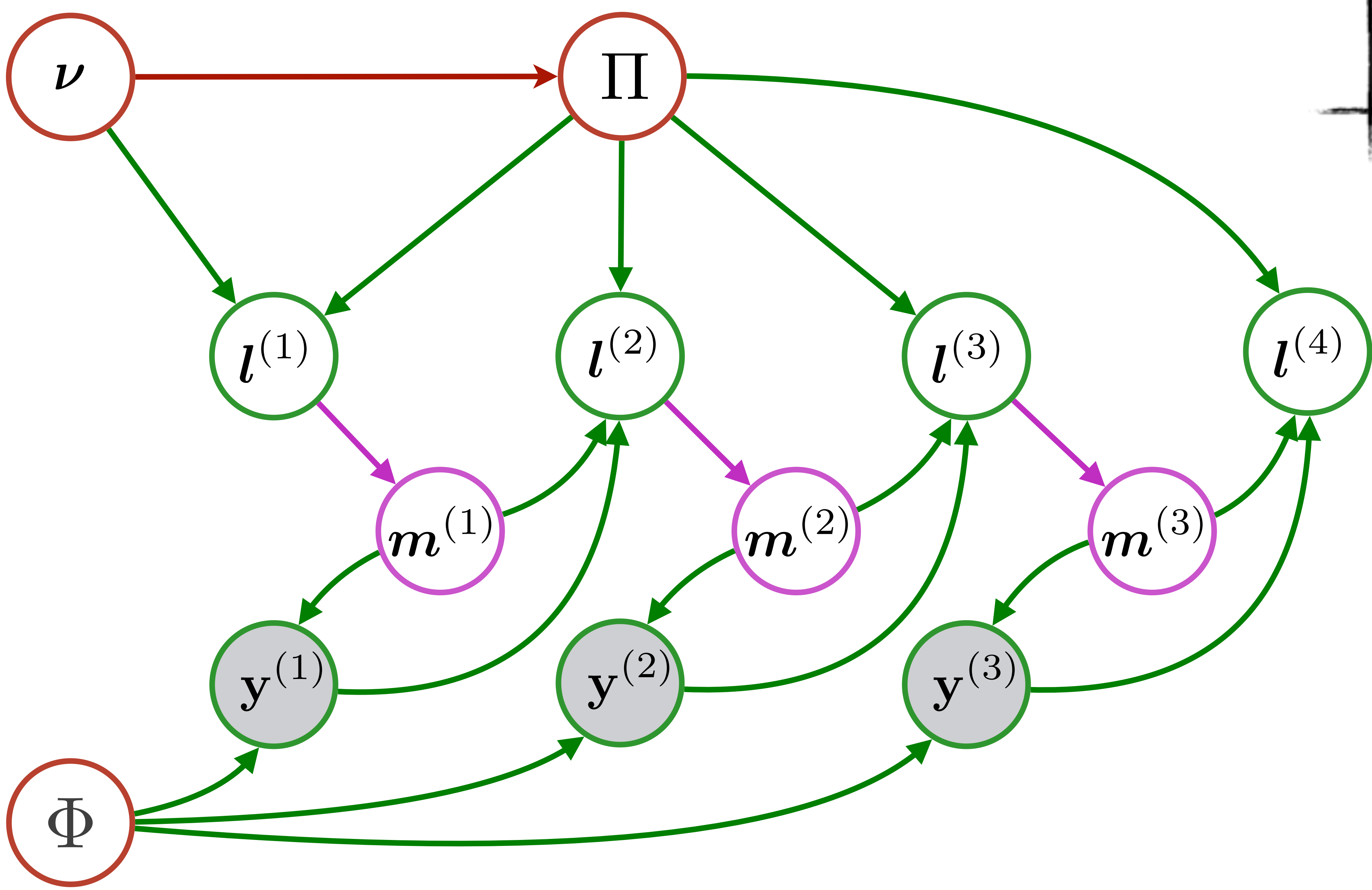


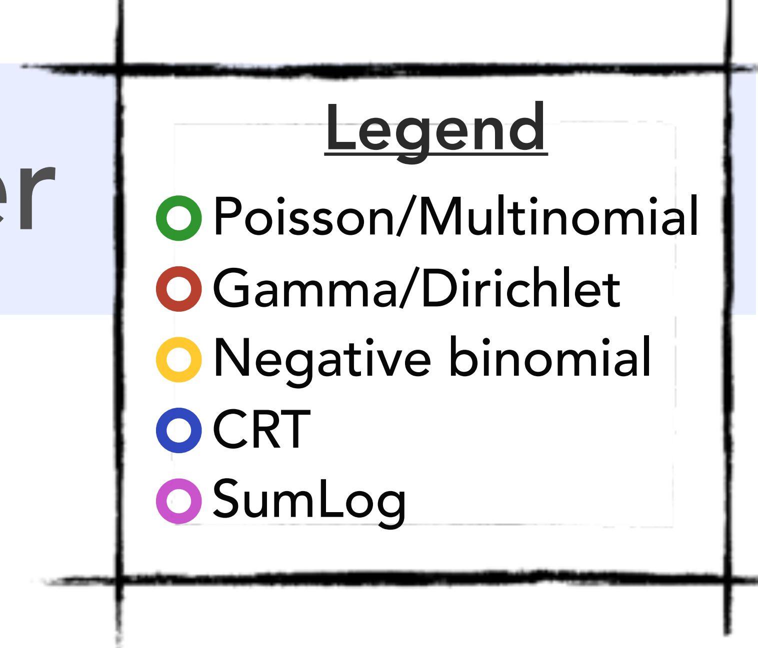
Augment and Conquer



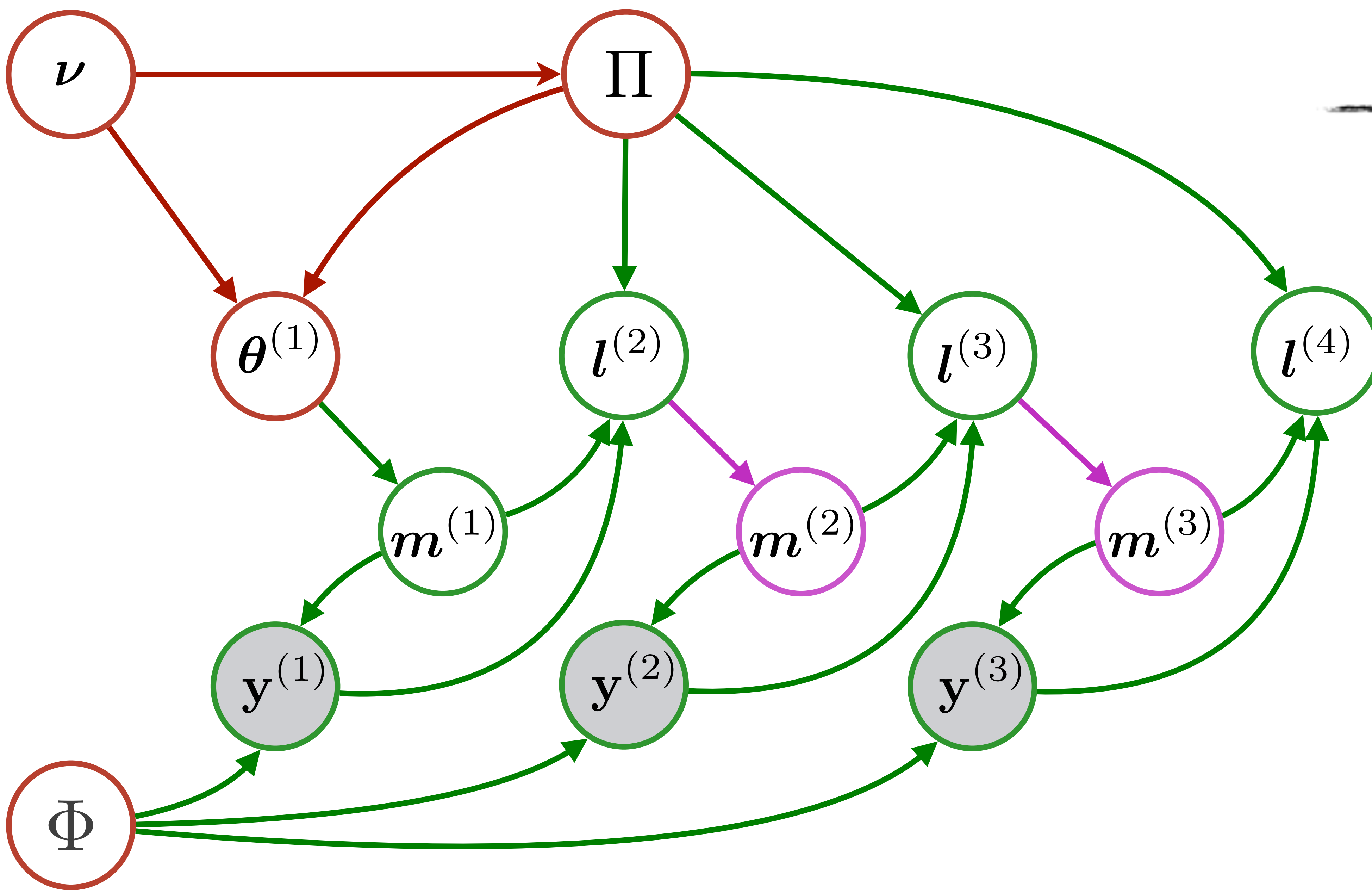
$$\Pi \sim P(\Pi \mid \mathcal{A}, Y, \boldsymbol{\nu}) \checkmark$$

Augment and Conquer



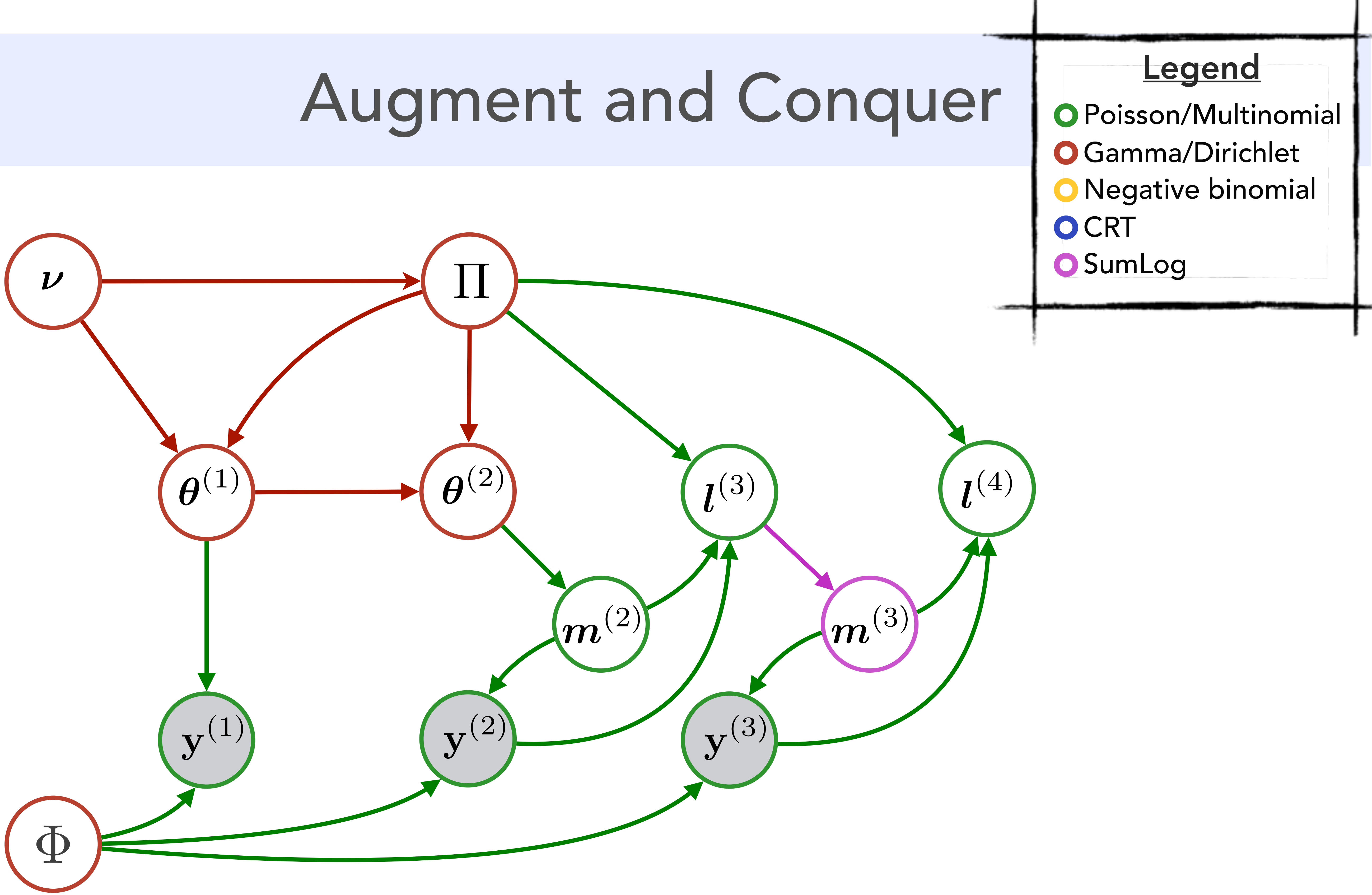


Augment and Conquer



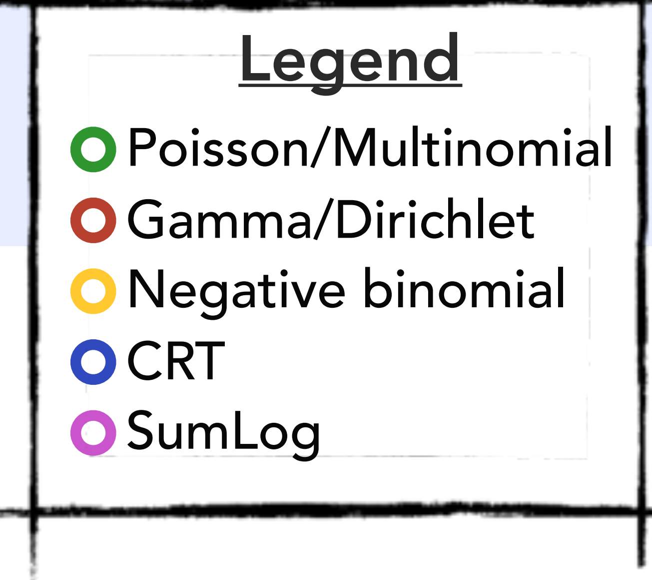
$$\boldsymbol{\theta}^{(1)} \sim P(\boldsymbol{\theta}^{(1)} \mid \mathcal{A}, Y, \Pi, \boldsymbol{\nu})$$ ✓

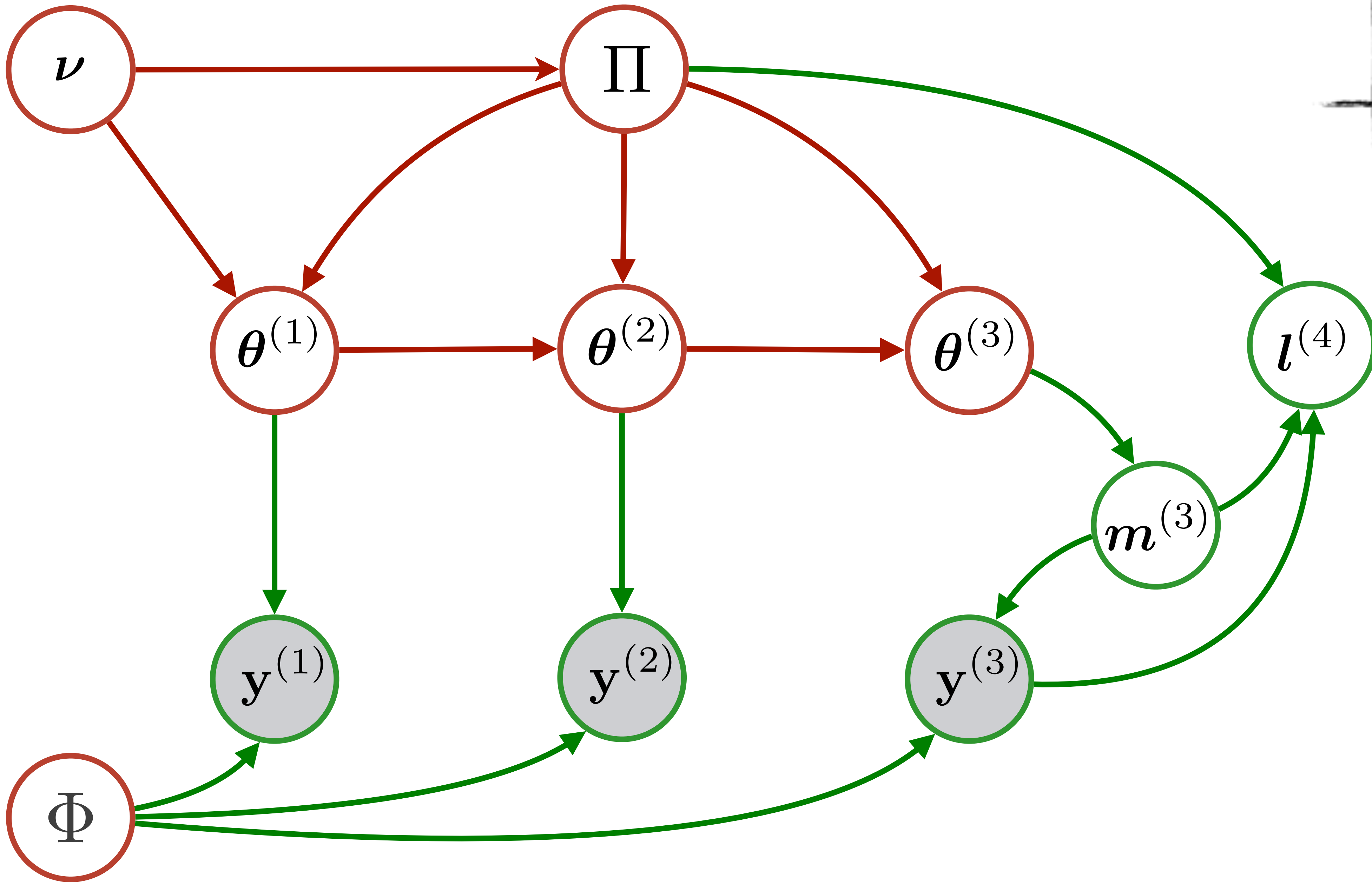
Augment and Conquer



$$\boldsymbol{\theta}^{(2)} \sim P(\boldsymbol{\theta}^{(2)} \mid \boldsymbol{\theta}^{(1)}, \mathcal{A}, Y, \Pi, \boldsymbol{\nu})$$ ✓

Augment and Conquer





$$\boldsymbol{\theta}^{(3)} \sim P(\boldsymbol{\theta}^{(3)} \mid \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(1)}, \mathcal{A}, Y, \Pi, \boldsymbol{\nu})$$ ✓

Solution

Backward filtering—forward sampling (BFFS)

Backward filtering

$$\mathcal{A}^{(t)} \sim P(\mathcal{A}^{(t)} \,|\, \mathcal{A}^{(t+1)}, Y, \Theta, \Pi, \boldsymbol{\nu})$$

Forward sampling

$$\boldsymbol{\theta}^{(t)} \sim P(\boldsymbol{\theta}^{(t)} \,|\, \boldsymbol{\theta}^{(t-1)}, \mathcal{A}, Y, \Pi, \boldsymbol{\nu})$$

Conclusion

Elegant inference in the **natural** model

that scales with the **number of non-zeros**

and relies on a **novel** auxiliary variable scheme

https://github.com/aschein/pgds