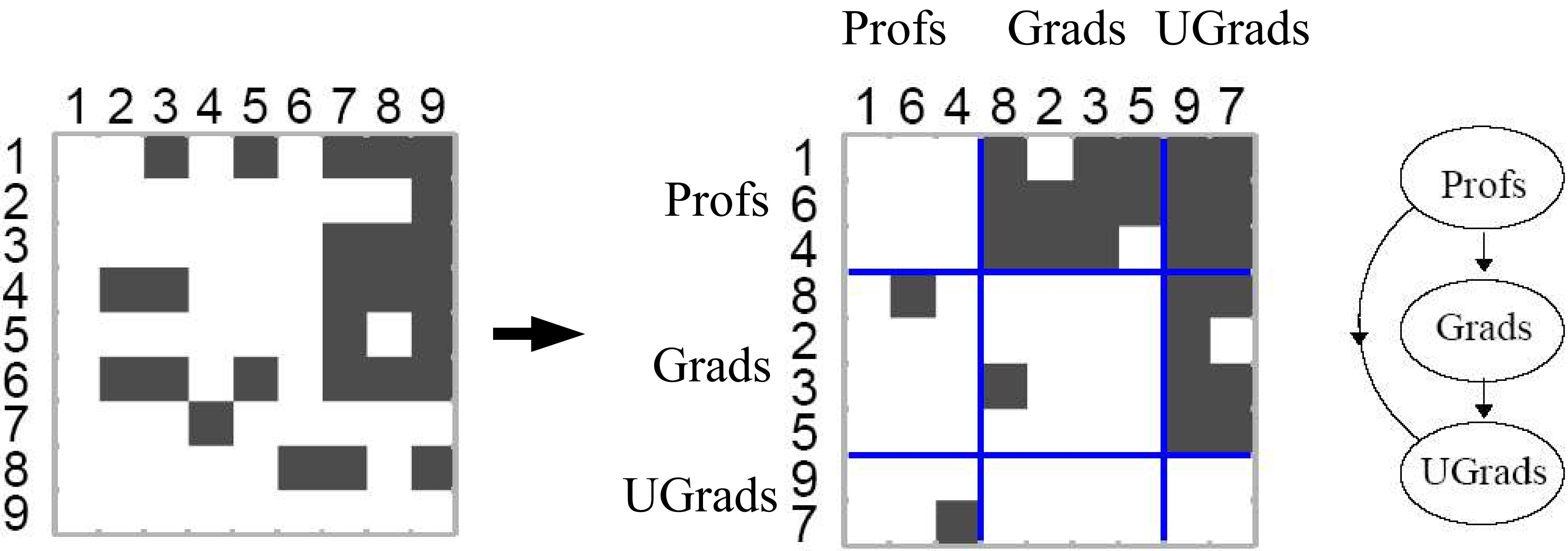
Learning systems of concepts with an Infinite Relational Model

Charles Kemp,[1] Josh Tenenbaum,[1] Tom Griffiths,[2]
Takeshi Yamada,[3] Naonori Ueda[3]

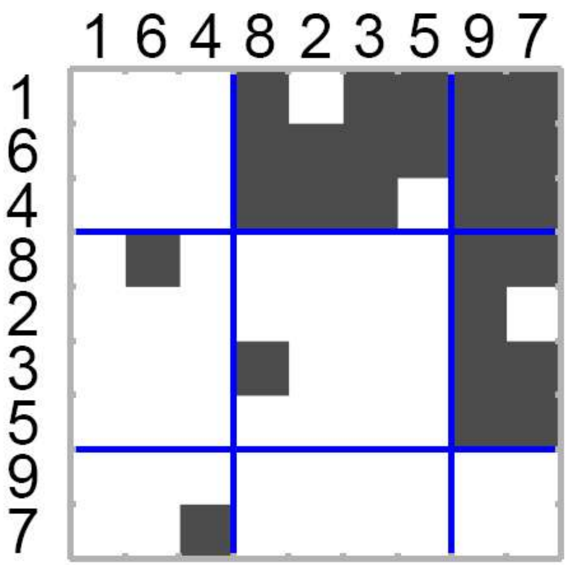
[1]MIT, [2]Brown, [3]NTT Corporation

Dominance relations at this conference



Predicate Invention

r(1,8). r(1,3). r(1,5) ...



a(1). a(6). a(4).

b(8). b(2). b(3). b(5).

c(9). c(7).

$r(X,Y) \leftarrow a(X),a(Y).$ (0.0)

$r(X,Y) \leftarrow a(X),b(Y).$ (0.9)

$r(X,Y) \leftarrow a(X),c(Y).$ (1.0)

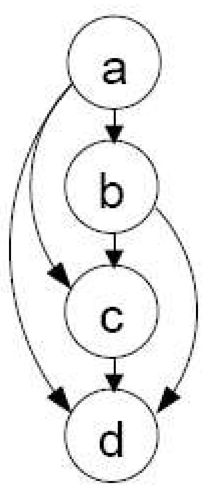
...

Outline

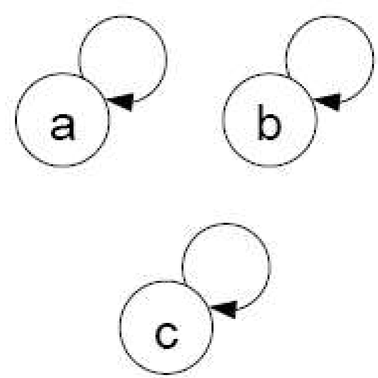
1) Discovering concepts with an Infinite Relational Model (IRM)

2) Discovering the *kind* of relational system that best explains a data set

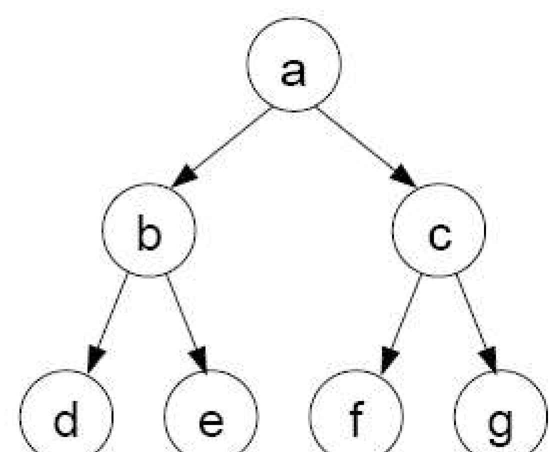
Dominance hierarchy



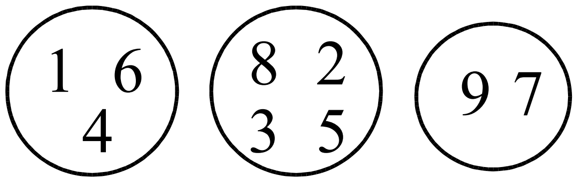
Cliques



Tree



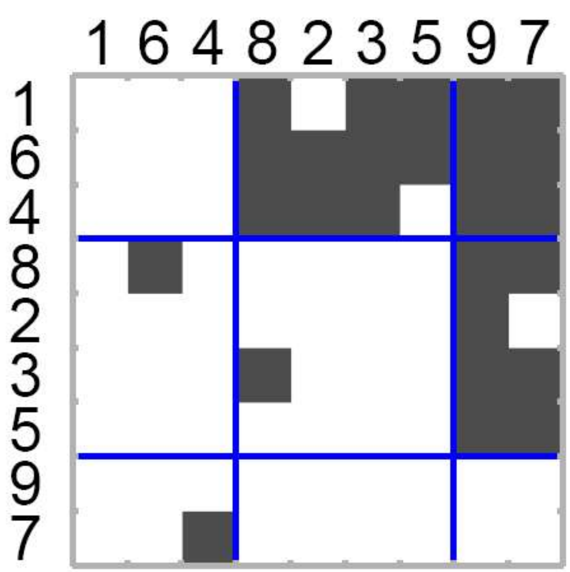
An Infinite Relational Model (IRM)



z $\quad z \mid \gamma \sim \mathrm{CRP}(\gamma)$

0.1	0.9	0.9
0.1	0.1	0.9
0.1	0.1	0.1

$\eta \quad \eta_{ab} \mid \alpha, \beta \sim \mathrm{Beta}(\alpha, \beta)$



R $\quad R_{ij} \mid z, \eta \sim \mathrm{Bernoulli}(\eta_{z_i z_j})$

- Goal: find *z* that maximizes $P(z|R)$

An Infinite Relational Model (IRM)

- Goal: find *z* that maximizes $P(\mathbf{z}|R) \propto P(R|\mathbf{z})P(\mathbf{z})$

$$P(R|\mathbf{z}) = \prod_{ab} \frac{B(m_{ab}+\alpha, \bar{m}_{ab}+\beta)}{B(\alpha,\beta)}$$

where m_{ab} is the number of 1-edges between classes *a* and *b*

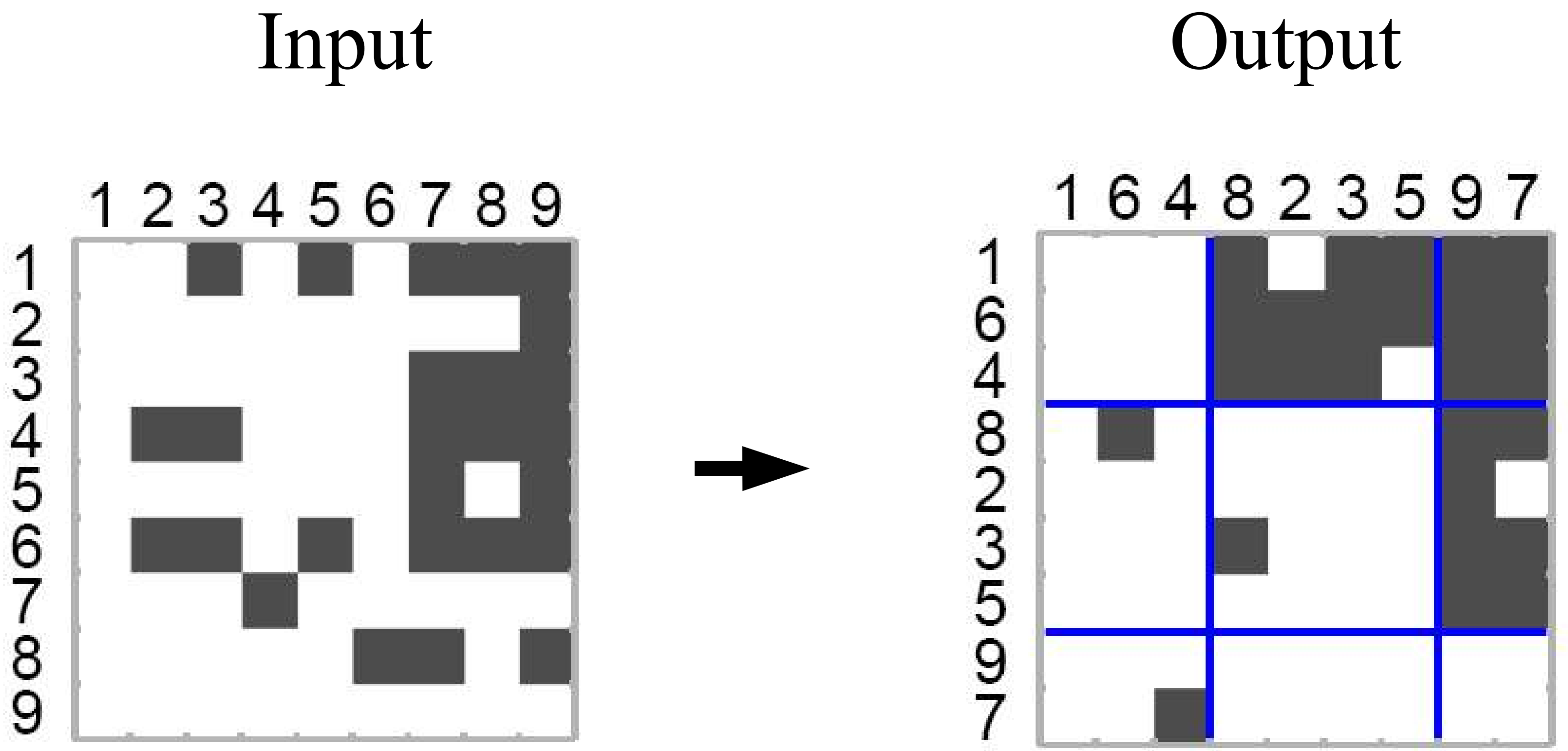
$\bar{m}_{ab}$ is the number of 0-edges between classes *a* and *b*

$B(\cdot,\cdot)$ is the Beta function

$$P(z_i = a | z_1, \ldots, z_{i-1}) = \begin{cases} \frac{n_a}{i-1+\gamma} & n_a > 0 \\ \frac{\gamma}{i-1+\gamma} & a \text{ is a new class} \end{cases}$$

where n_a is the number of entities in class *a*

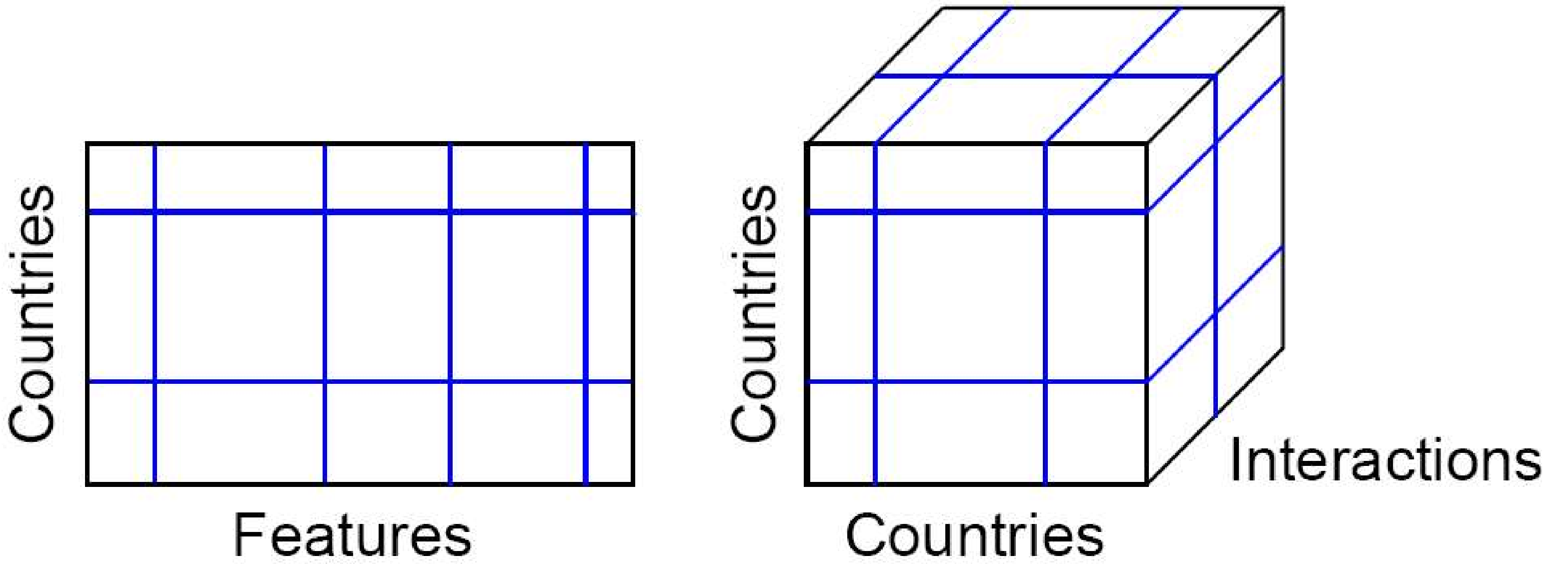
The IRM



Related Work

- Relational models
  - Sociology:
    - Wang and Wong (1987); Nowicki and Snijders (2001)
  - Machine learning:
    - Taskar, Segal and Koller (2001)
    - Wolfe and Jensen (2004)
    - Wang, Mohanty and McCallum (2005)
- Nonparametric Bayesian models
    - Ferguson (1973); Neal (1991)
- Nonparametric Bayesian relational models
    - Carbonetto, Kisynski, de Freitas and Poole (2005)
    - Xu, Tresp, Yu, Kriegel (2006)

Clustering arbitrary relational systems

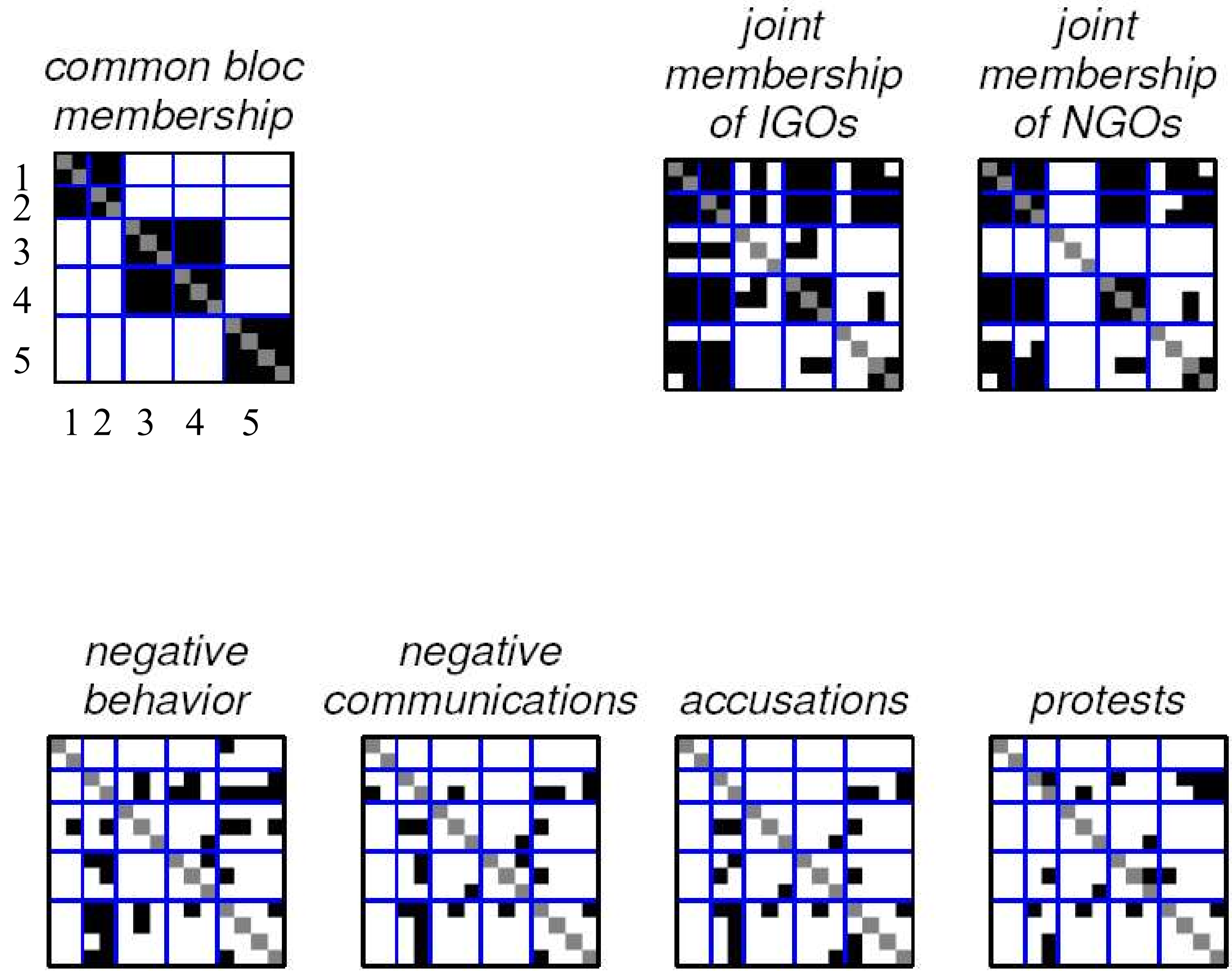


- 14 countries
- 54 binary relations representing interactions between countries (eg. exports to, protests against)
- 90 country features

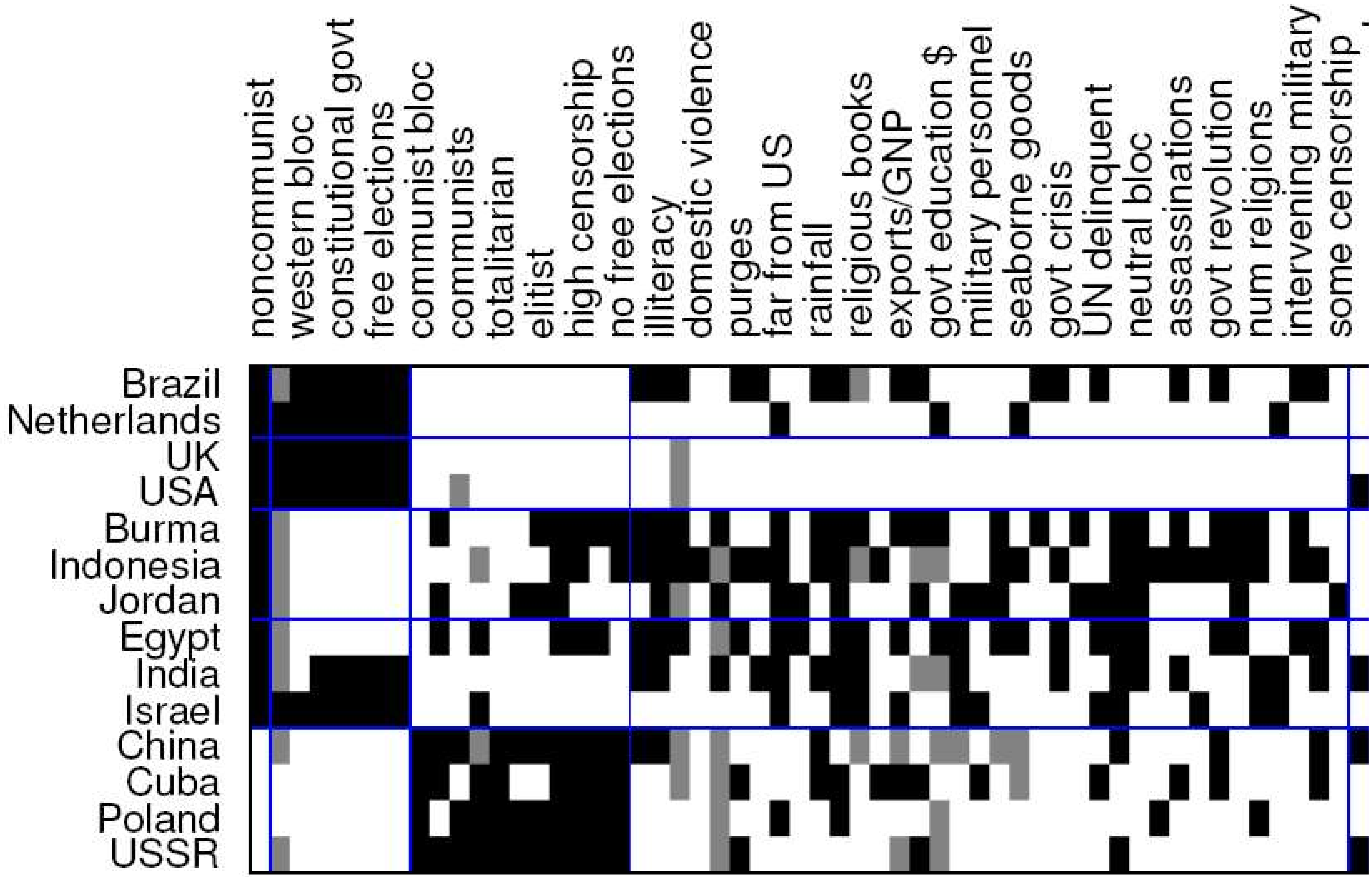
(Rummel, 1965)

Relation clusters (Rummel, 1965)

1. Brazil
   Netherlands
2. UK
   USA
3. Burma
   Indonesia
   Jordan
4. Egypt
   India
   Israel
5. China
   Cuba
   Poland
   USSR

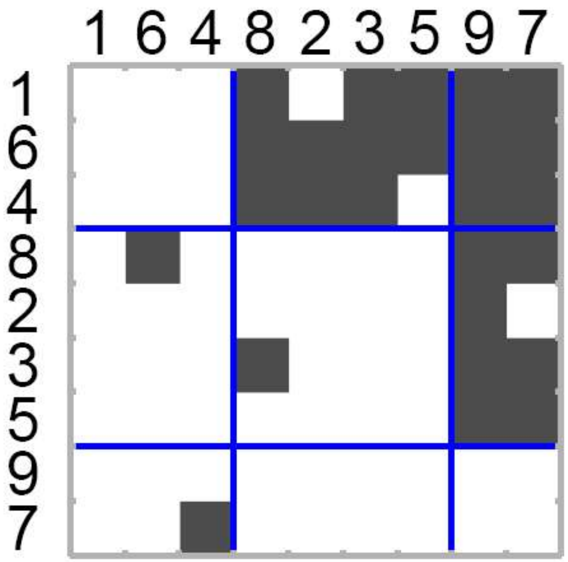


Feature clusters (Rummel, 1965)



Towards Richer Representations

r(1,8). r(1,3). r(1,5) ...



a(1). a(6). a(4).

b(8). b(2). b(3). b(5).

c(9). c(7).

r(X,Y) ← a(X),a(Y). (0.0)

r(X,Y) ← a(X),b(Y). (0.9)

r(X,Y) ← a(X),c(Y). (1.0)

...

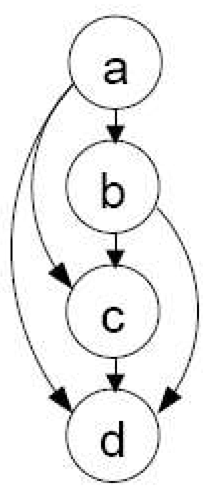
- The concepts discovered by the IRM can serve as primitives in complex logical theories
  - cf. Craven and Slattery (2001); Popescul and Ungar (2004)

Outline

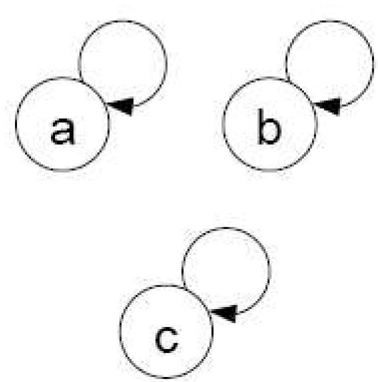
1) Discovering concepts with an Infinite Relational Model (IRM)

2) Discovering the *kind* of relational system that best explains a data set

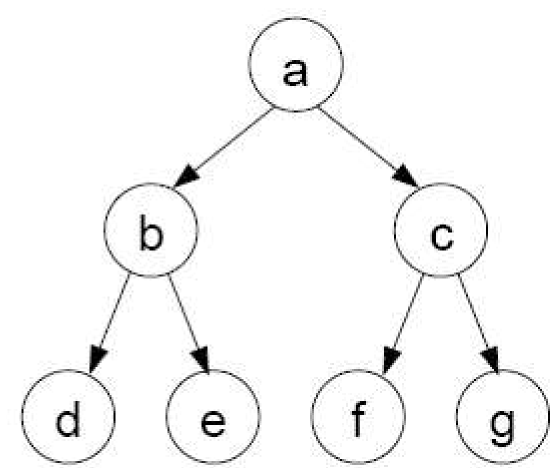
Dominance hierarchy



Cliques

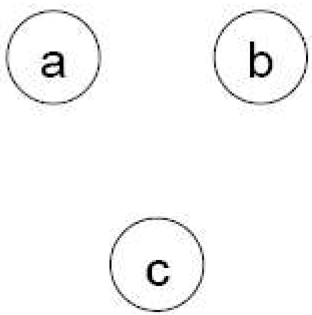


Tree

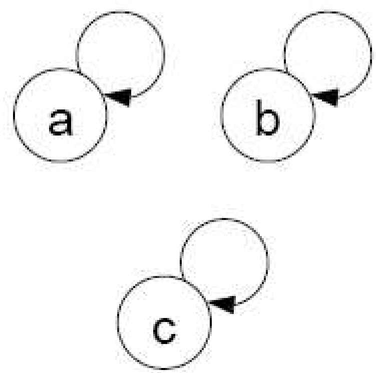


Structural forms

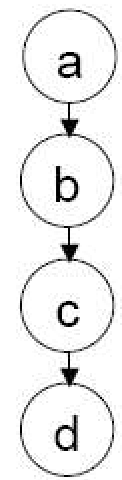
Partition



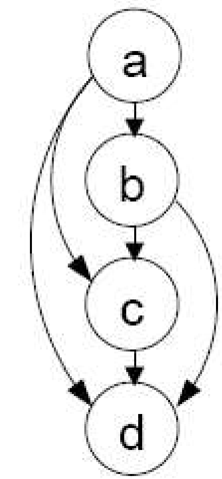
Cliques



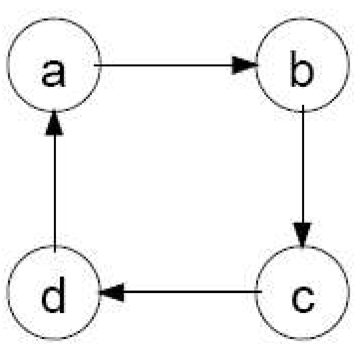
Chain



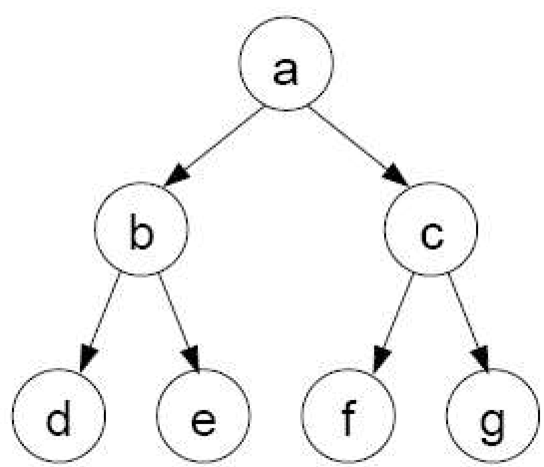
Dominance hierarchy



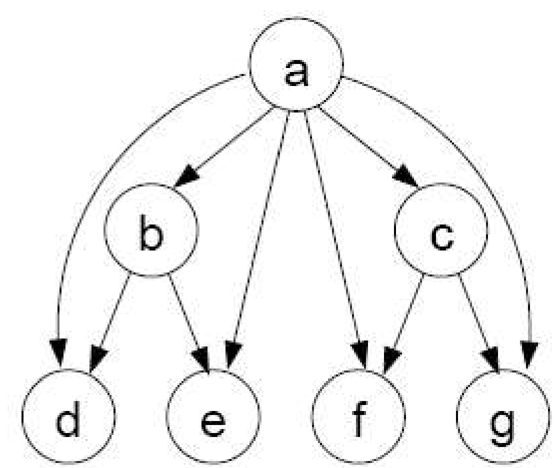
Ring

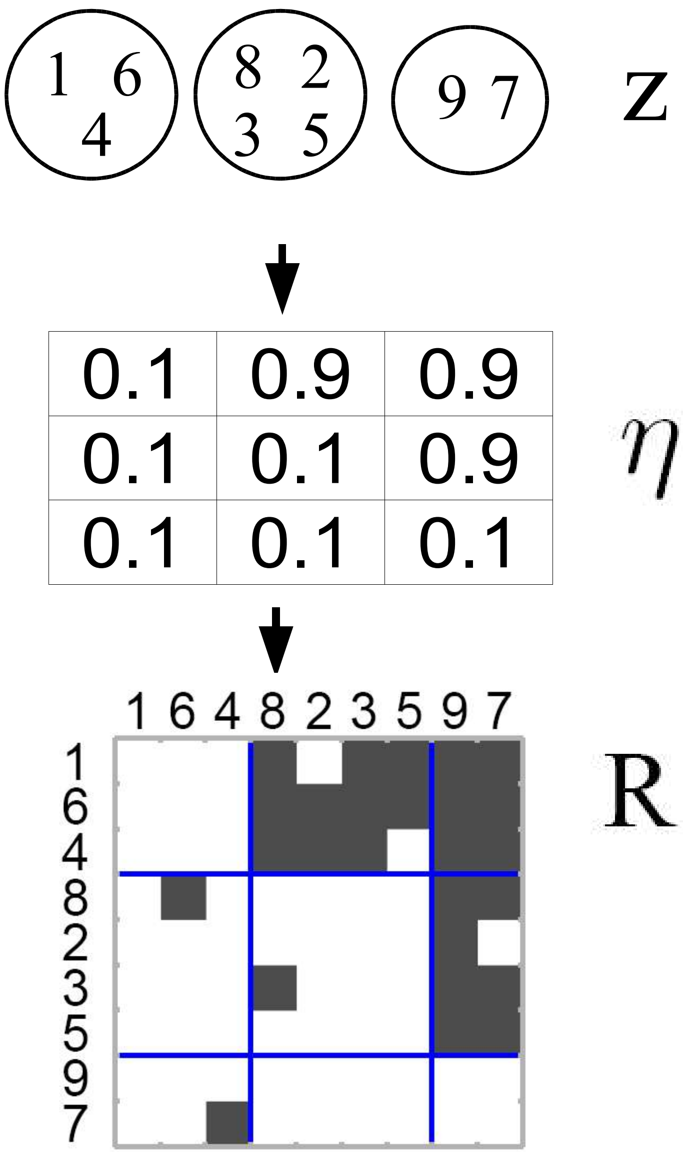


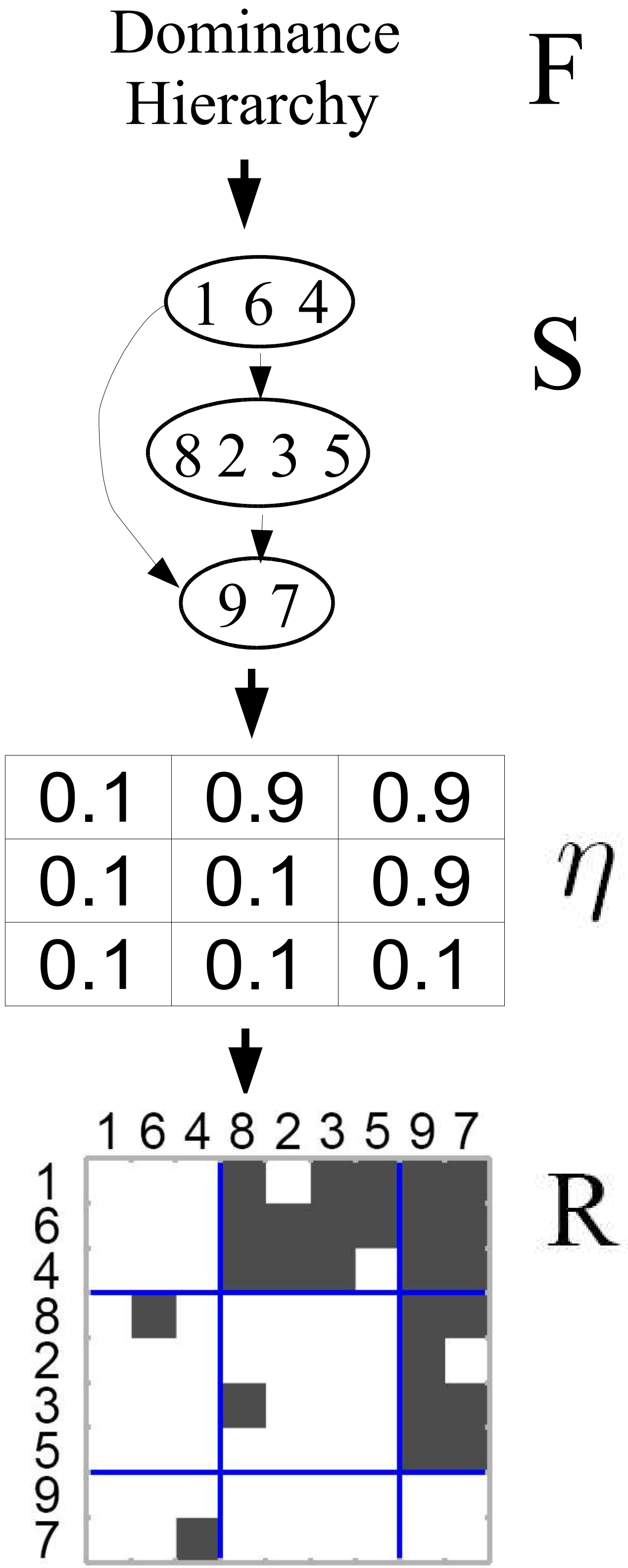
Tree



Dominance tree

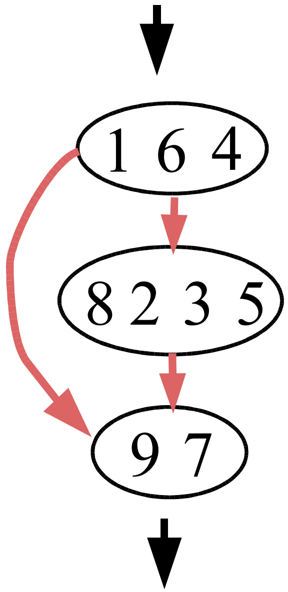



1 6 4
8 2 3 5
9 7
Z
0.1 0.9 0.9
0.1 0.1 0.9
0.1 0.1 0.1
η
1 6 4 8 2 3 5 9 7
1 6 4 8 2 3 5 9 7
R


Dominance Hierarchy
F
1 6 4
8 2 3 5
9 7
S
0.1 0.9 0.9
0.1 0.1 0.9
0.1 0.1 0.1
η
1 6 4 8 2 3 5 9 7
1 6 4 8 2 3 5 9 7
R

Dominance Hierarchy

F



S

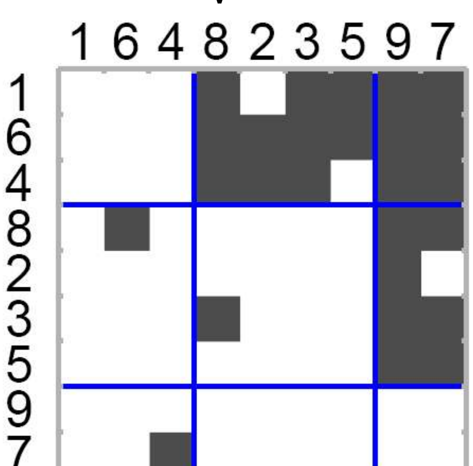
$$z \,|\, \gamma \quad \sim \mathrm{CRP}(\gamma)$$

$$P(S|z, F) \propto 1 \quad \text{if S consistent with z and F}$$

0.1	0.9	0.9
0.1	0.1	0.9
0.1	0.1	0.1

η

$$\eta_{ab} \,|\, G \sim \begin{cases} \mathsf{Beta}(\alpha_0, \beta_0), & \text{if } G_{ab} = 0 \\ \mathsf{Beta}(\alpha_1, \beta_1), & \text{if } G_{ab} = 1 \end{cases}$$



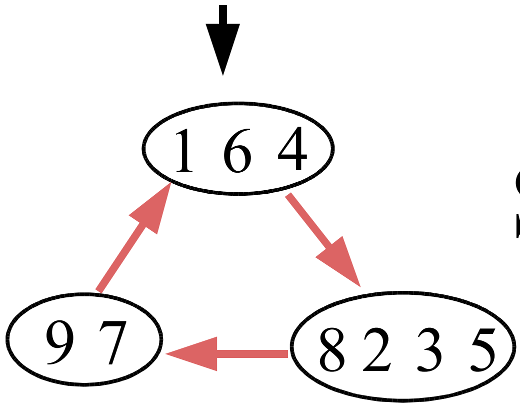
R

$$R_{ij} \,|\, z, \eta \sim \mathrm{Bernoulli}(\eta_{z_i z_j})$$

- Goal: find S that maximizes P(S|R,F)

Ring

F



S

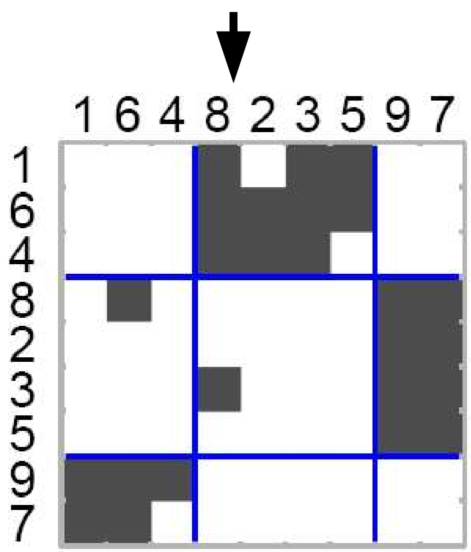
$$z \,|\, \gamma \quad \sim \mathrm{CRP}(\gamma)$$

$$P(S|z,F) \propto 1 \quad \text{if S consistent with z and F}$$

0.1	0.9	0.1
0.1	0.1	0.9
0.9	0.1	0.1

η

$$\eta_{ab} \,|\, G \sim \begin{cases} \mathrm{Beta}(\alpha_0, \beta_0), & \text{if } G_{ab} = 0 \\ \mathrm{Beta}(\alpha_1, \beta_1), & \text{if } G_{ab} = 1 \end{cases}$$

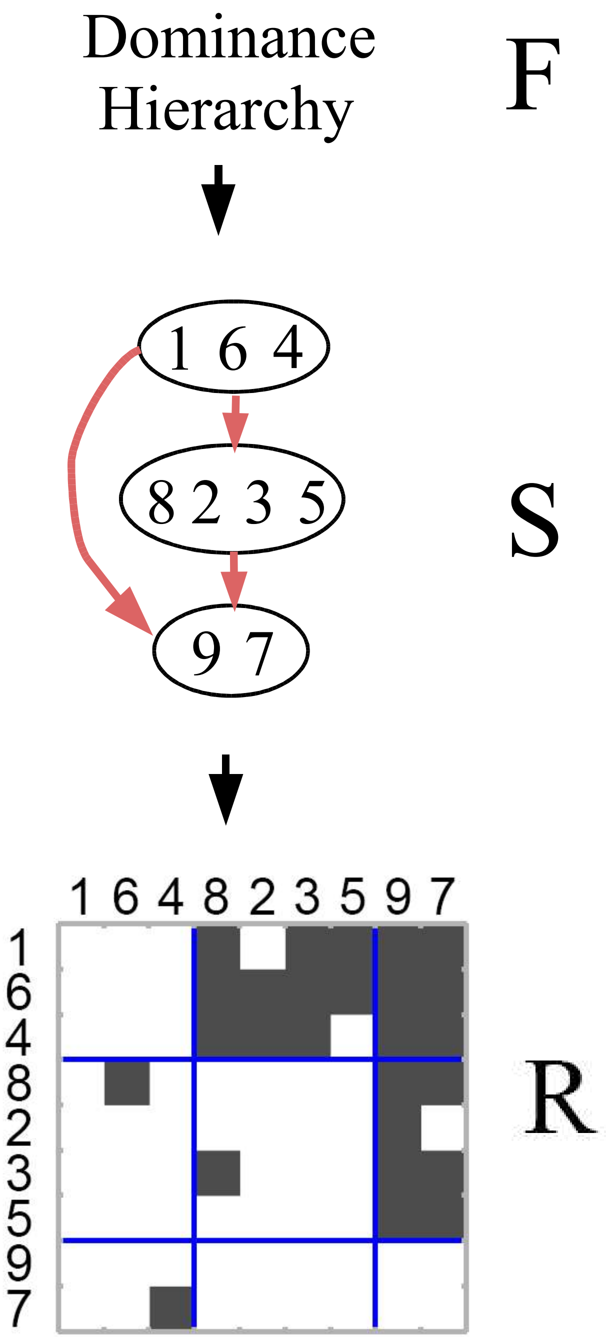


R

$$R_{ij} \,|\, z, \eta \ \sim \mathrm{Bernoulli}(\eta_{z_i z_j})$$

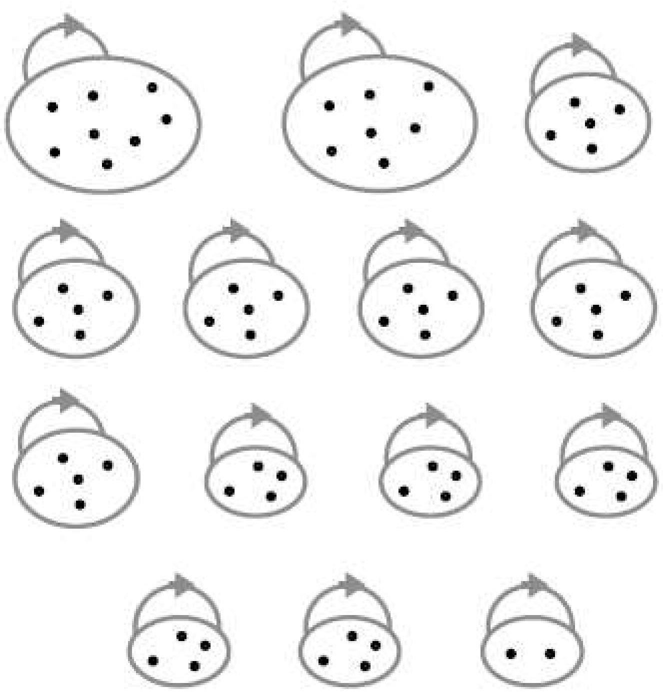
- Goal: find S that maximizes P(S|R,F)

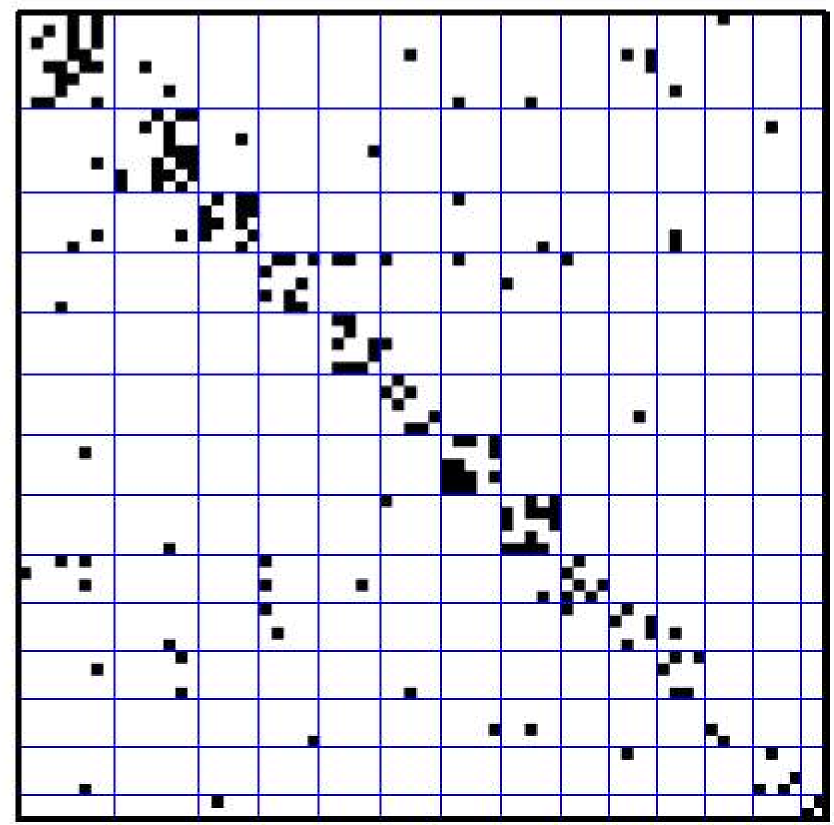
Learning structural forms



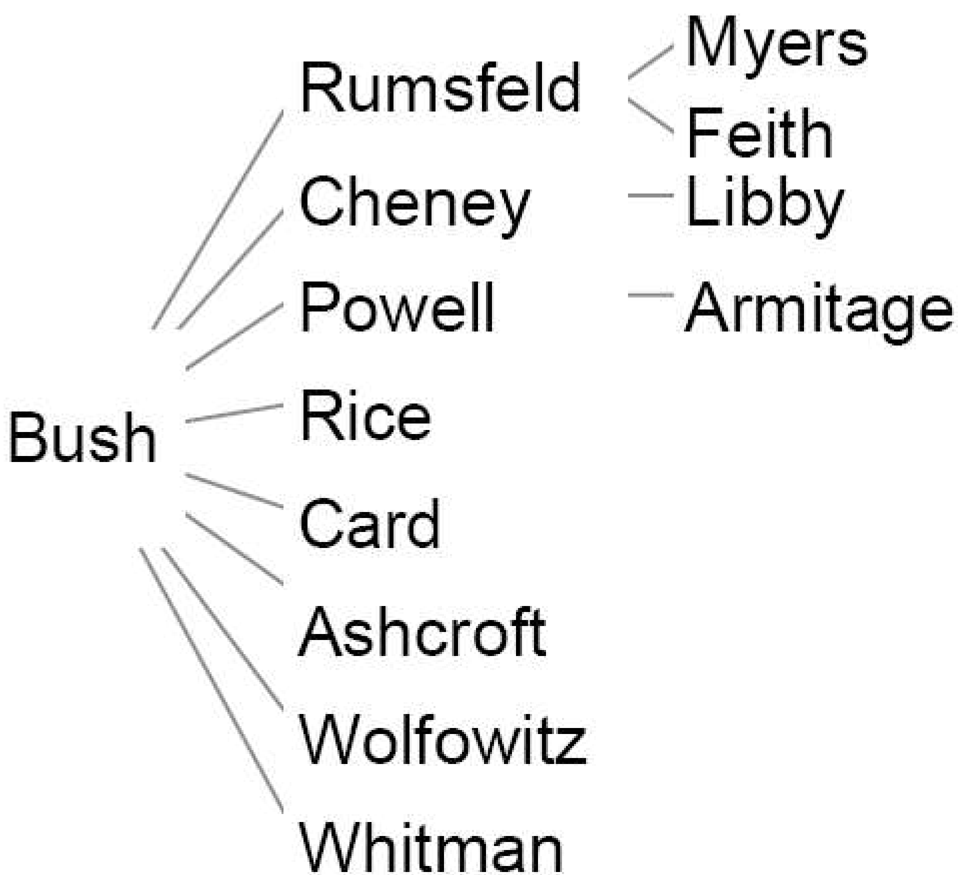
- We place a uniform prior over the set of forms and search for the S and F that maximize P(S,F|R)

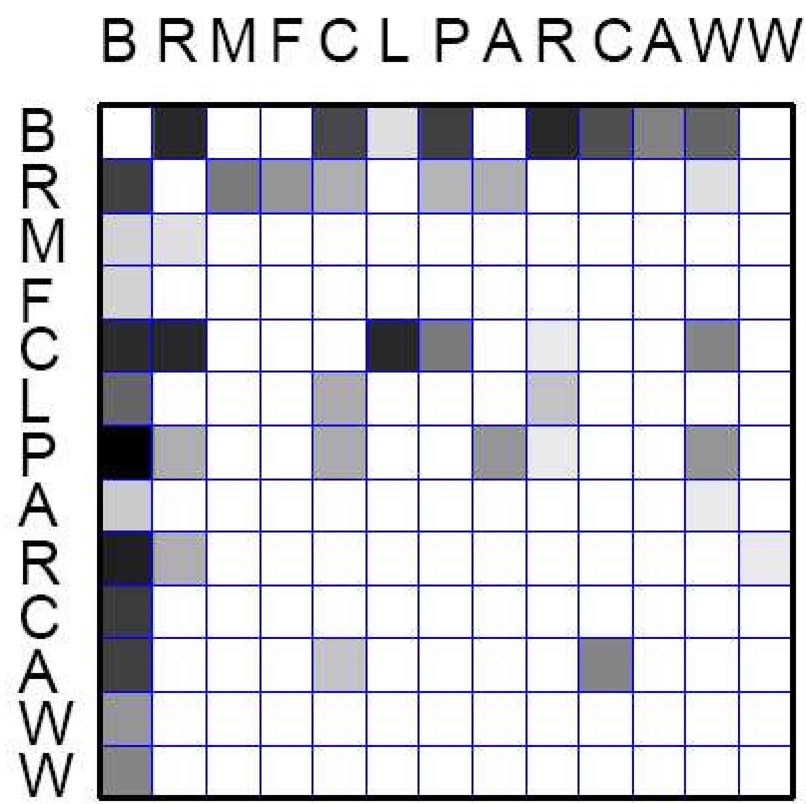
Friendship groups (MacRae, Gagnon)





Bush Cabinet





Conclusions

1) The IRM discovers concepts (unary predicates) and relationships between these concepts.

2) An extended version of the IRM can discover abstract structural properties of a relational system.