ICML 2016 NYC

Structure Learning of Partitioned Markov Networks

Song Liu[1], Taiji Suzuki[2], Masashi Sugiyama[3] and Kenji Fukumizu[1]

1. The Institute of Statistical Mathematics
2. Tokyo Institute of Technology, JST PRESTO
3. University of Tokyo

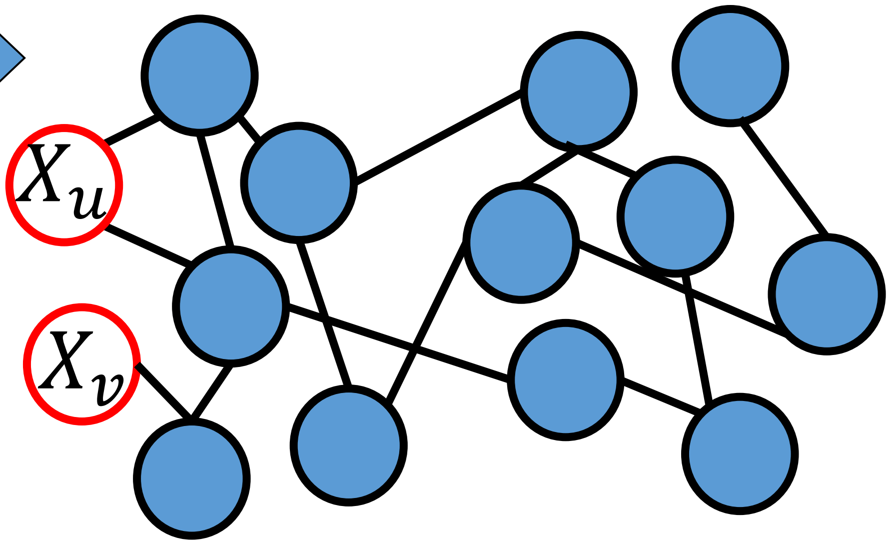
Outline

1. **Markov Network**
2. Partitioned Markov Network
3. Learning Partitioned Markov Network
4. Experiments
5. Conclusion

Markov Network (MN)

- Markov Network: *undirected* graphical model, encoding the *conditional independence* between random variables.

$$P(X_1, X_2, \dots, X_m) \rightarrow$$



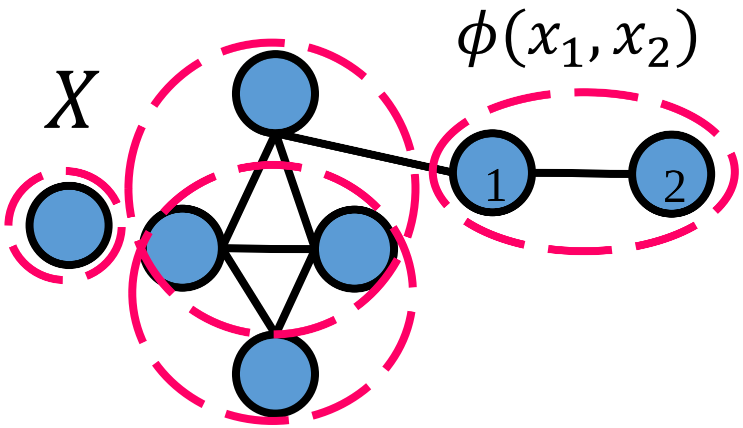
X_u and X_v are **not** connected

$$X_u \perp X_v \mid X_{\setminus \{X_u, X_v\}}$$

"x_u and x_v directly influences each other if connected."

Learning Sparse Structure of MN

- Hammersley-Clifford Theorem (Hammersley & Clifford,1971).



$$p(\boldsymbol{x}) \propto \prod_{c \in \boldsymbol{C}} \phi(\boldsymbol{x}_c)$$

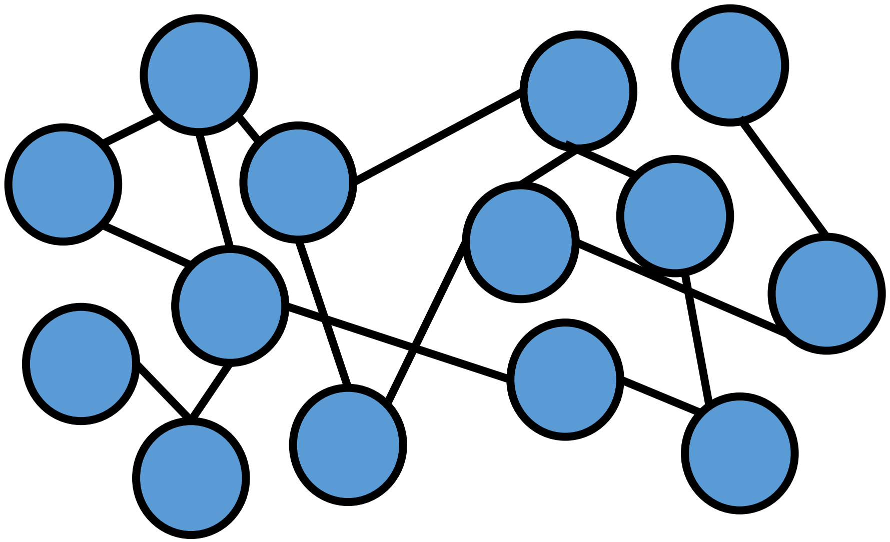
Cliques $\boldsymbol{c}$: fully connected sub-graph

- Learning a *sparse* factorization of density function.
- $\widehat{\boldsymbol{\Theta}} = \operatorname{argmax}_{\boldsymbol{\Theta}} \sum_{\boldsymbol{x}} \log p(\boldsymbol{x}; \boldsymbol{\Theta}) - \underline{\lambda \|\boldsymbol{\Theta}\|_1}, \boldsymbol{\Theta} \in \mathbb{R}^{m \times m}$.
  - (Friedman et al., 2008)

Sparsity inducing norm

Why not learn a full MN?

Not a good idea…



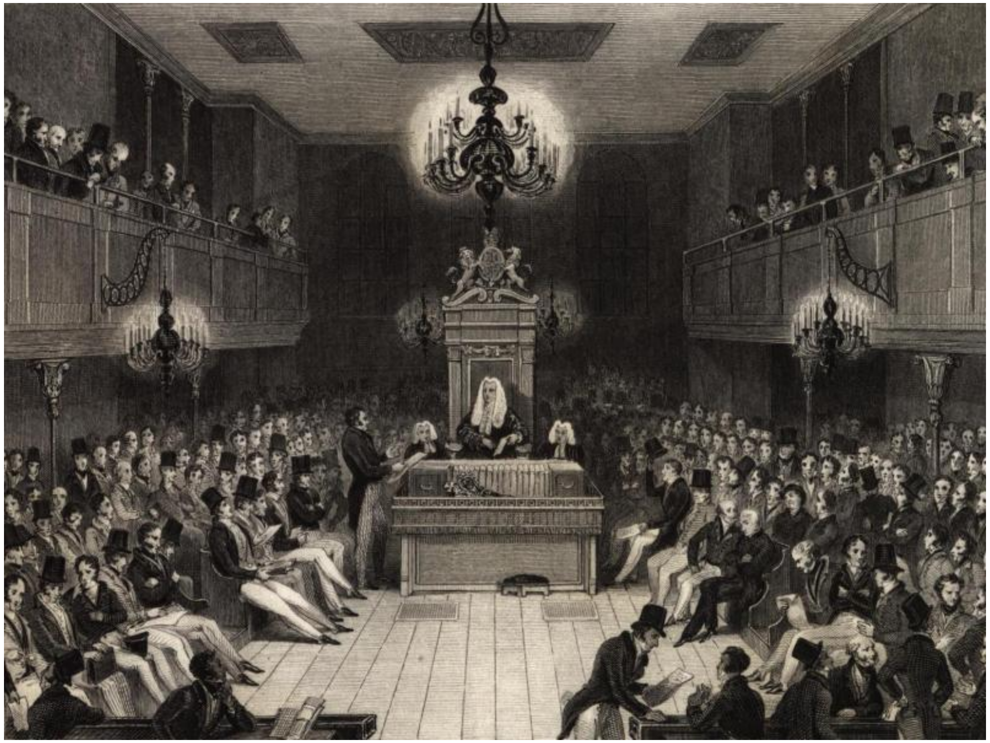
- Too few samples.
- Full model is beyond our understanding.
- **We are only interested in partial connections!**

Outline

1. Markov Network
2. **Partitioned Markov Network**
3. Learning Partitioned Markov Network
4. Experiments
5. Conclusion

Partitioned Networks

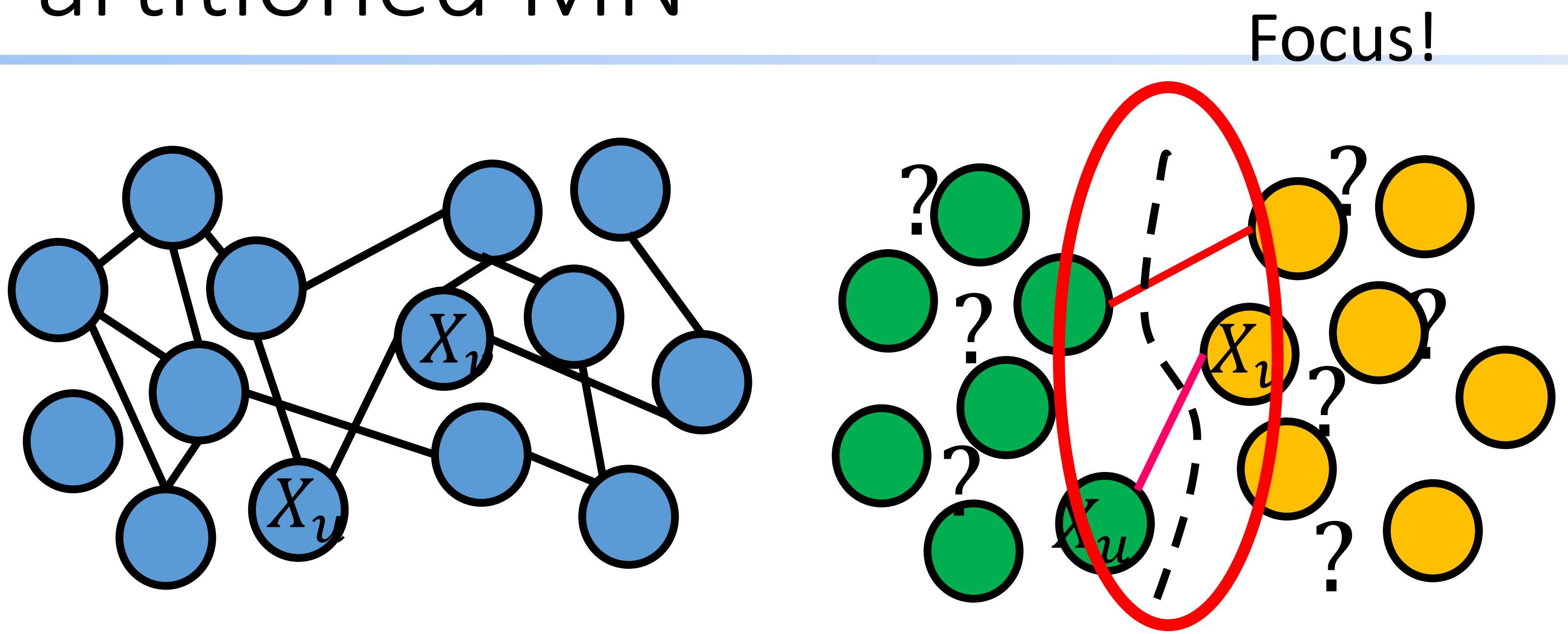
British Parliament



(WIKIPEDIA: House of Commons of the United Kingdom)

Some networks are naturally *partitioned*.

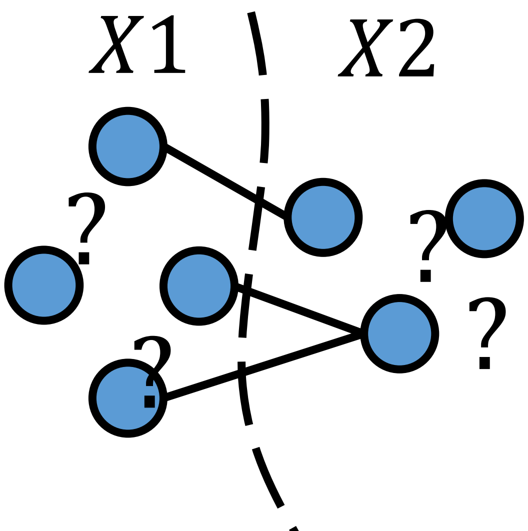
Partitioned MN



a **full** network a **partitioned** network

- Discover intergroup links of a partitioned network is key in many applications

Partitioned (Local) Markovian Property



$X = (X1, X2)$ is a partition of X.

<u>Partitioned Markovian Property</u>:

$\forall X_u \in X1, \forall X_v \in X2,$

If X_u and X_v are **not** connected,

✓ $X_u \perp X_v \mid X_{\backslash\{X_u, X_v\}}$

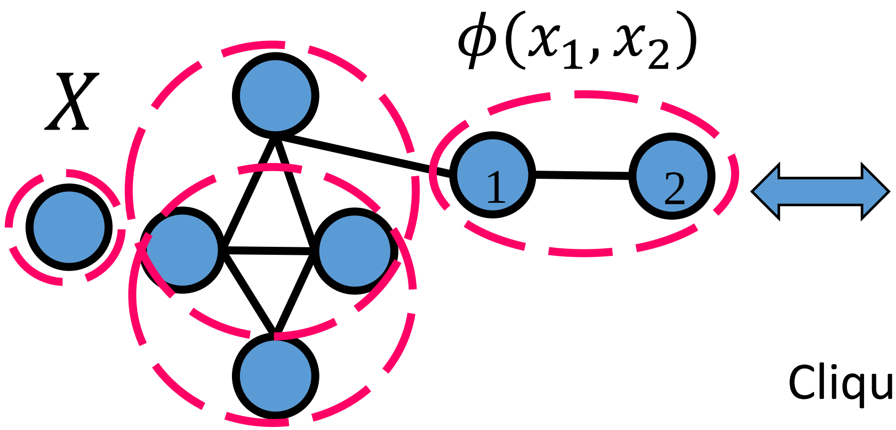
Partitioned MN does **not** guarantee:

$\forall X_u, X_v \in X,$ If X_u and X_v are not connected,

✗ $X_u \perp X_v \mid X_{\backslash\{X_u, X_v\}}$

Local Factorization of Partitioned MN

Recall, the full MN factorizes over the cliques (**Hammersley Clifford Theorem**).

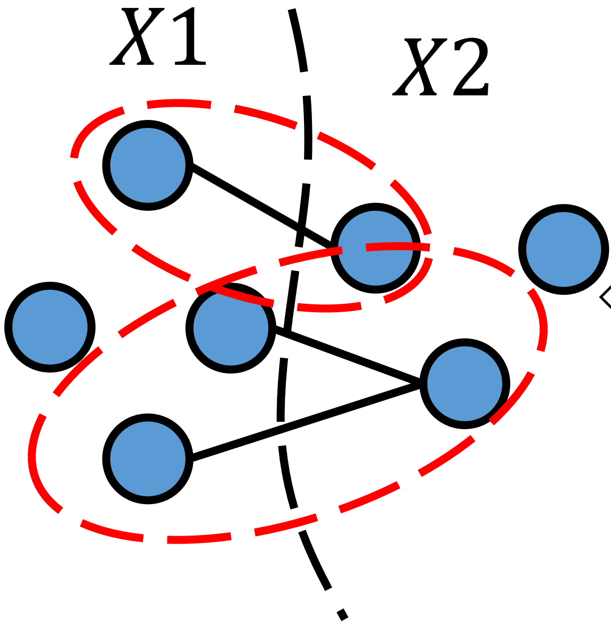


$$p(\boldsymbol{x}) \propto \prod_{c \in \boldsymbol{C}} \phi(\boldsymbol{x}_c)$$

Cliques $\boldsymbol{c}$: fully connected sub-graph

What can we say about the partitioned MN?

Factorization of Partitioned MN



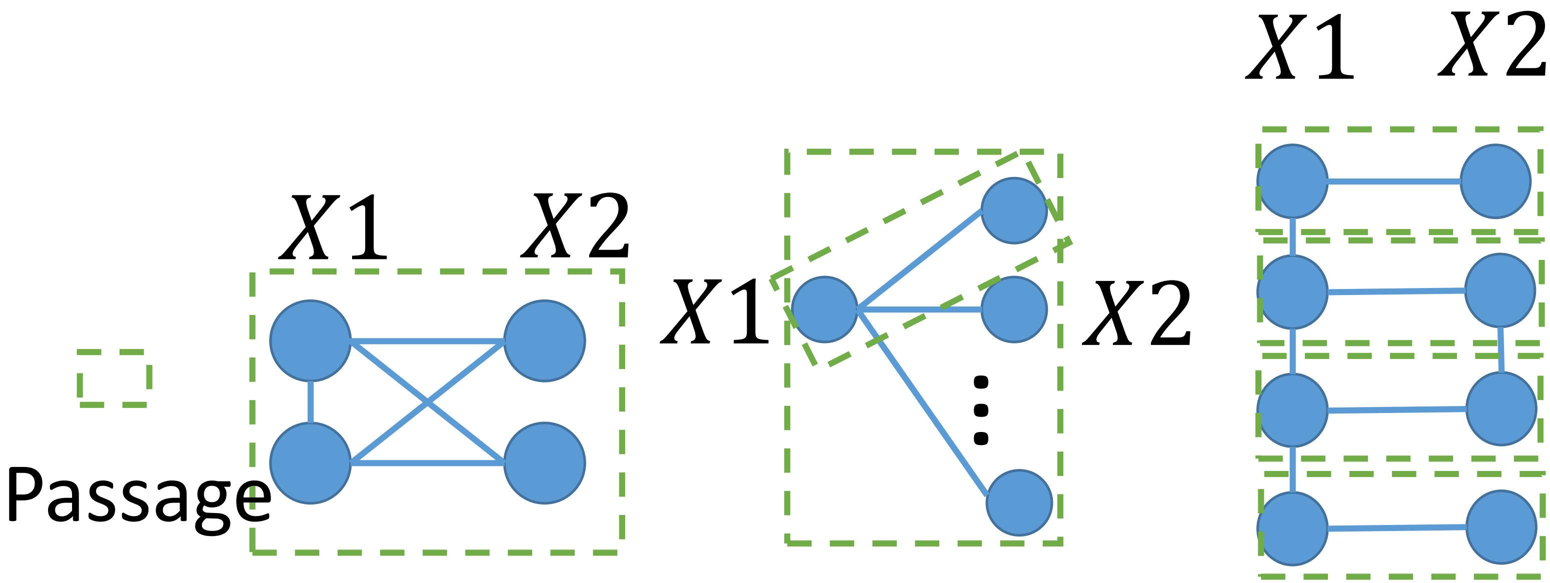
Partitioned Ratio

$$\boxed{\frac{p(x_1, x_2)}{p(x_1)p(x_2)}} \propto \boxed{\prod_{b \in \boldsymbol{B}} \phi(\boldsymbol{x}_b)}$$

$\boldsymbol{B}$: the set of **Passage** structures.

- An analog to Hamersley-Clifford Theorem
- Factorization is over a novel structure: **Passage**.

Passage



We define **passage** B a sub-graph of G, such that $X_B \cap X1 \neq \emptyset$, $X_B \cap X2 \neq \emptyset$, $\forall X_u \in X1 \cap X_B$ and $\forall X_v \in X2 \cap X_B$, (X_u, X_v) is in the edge set.

Outline

1. Markov Network
2. Partitioned Markov Network
3. **Learning Partitioned Markov Network**
4. Experiments
5. Conclusion

From Partitioned Ratio to Structure

- Knowing **factorization** of Partitioned Ratio = Knowing **passages structures** in Partitioned MN.

If $x_u \in X1, x_v \in X2$ do **not** appear in the same passage factor at least once, $X_u \perp X_v | \setminus X_u, X_v$ (See Proposition 2 in the paper).

- Pairwise assumption:

$$\frac{p(x_1,x_2)}{p(x_1)p(x_2)} \propto \prod_{b \in \boldsymbol{B}} \phi\left(\boldsymbol{x}_b\right) = \prod_{b \in \boldsymbol{B}} \prod_{X_u, X_v \in X_B} h_{u,v}(x_u, x_v)$$

Learning Pairwise Partitioned Ratio

- Given the partition $P(\boldsymbol{x}) = P(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{x} \in R^m$ and paired sample $\left\{\left(\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}\right)\right\}_{i=1}^{n} \sim P$, PR can be modelled as:

$$g(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{N(\boldsymbol{\theta})} \exp\left[\sum_{u \le v} \boldsymbol{\theta}_{u,v}^{\top} \boldsymbol{f}\left(\boldsymbol{x}_{u,v}\right)\right],$$

- $N(\boldsymbol{\theta})$ is a **normalization term**:
- $N(\boldsymbol{\theta}) := \int p(\boldsymbol{x}_1) \int p(\boldsymbol{x}_2) \exp\left[\sum_{u,v} \boldsymbol{\theta}_{u,v}^{\top} \boldsymbol{f}\left(\boldsymbol{x}_{u,v}\right)\right] d\boldsymbol{x}_2 d\boldsymbol{x}_1$

Learning Pairwise Partitioned Ratio

- $N(\boldsymbol{\theta})$ can be approximated by **U-statistics** (Hoeffding, 1963)
- $N(\boldsymbol{\theta}) \approx \widehat{N}(\boldsymbol{\theta}) = \frac{1}{\binom{2}{n}} \sum_{j \neq k} \exp\left[\sum_{u,v} \boldsymbol{\theta}_{u,v}^{\top} \boldsymbol{f}\left(\boldsymbol{x}_{u,v}^{[j,k]}\right)\right]$
  - where $\boldsymbol{x}^{[j,k]} = \left(\boldsymbol{x}_1^{(j)}, \boldsymbol{x}_2^{(k)}\right)$.

Maximum Likelihood Mutual Information: (Suzuki et al., 2008):

$$\hat{\theta} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\, P_{\mathrm{X}} \log g(\boldsymbol{x}; \boldsymbol{\theta}) \; - \lambda \underbrace{\sum_{u,v} ||\boldsymbol{\theta}_{u,v}||_2}$$

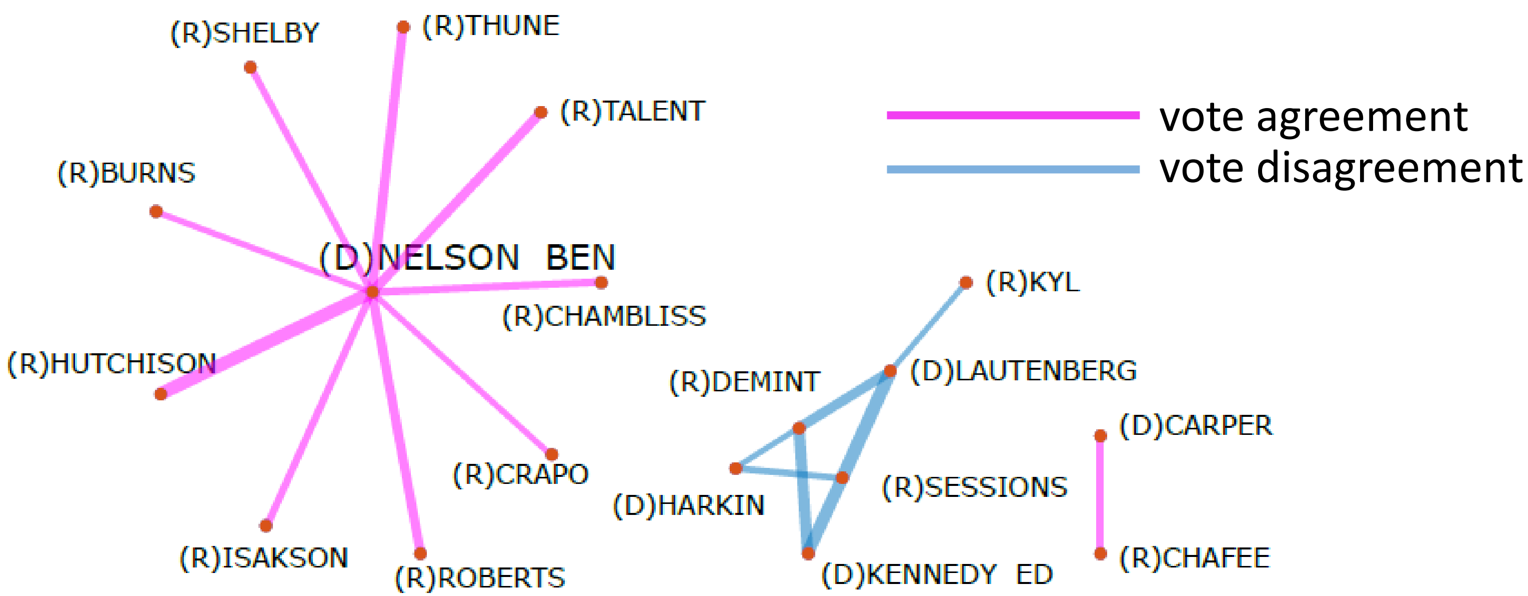
Included for sparsity

Outline

1. Markov Network
2. Partitioned Markov Network
3. Learning Partitioned Markov Network
4. **Experiments**
5. Conclusion

Experiments: Bipartisanship

- Learn the Bipartisanship in U.S. 109th senate.
- 45 Democrats, 55 Republicans.
- PMN learned over 645 recorded votes (Yea, Nay, not voting).
- Increasing regularization parameter until more than 15 edges.



Outline

1. Markov Network
2. Partitioned Markov Network
3. Learning Partitioned Markov Network
4. Experiments
5. **Conclusion**

Conclusion

- We learn interactions across two groups of random variables.
- Such interaction is expressed via **Partitioned Markov Network**.
- Learning sparse **factorization** of Partitioned Ratio leads to the discovery of **interactions** between two groups.