Title: Physics-inspired techniques for association rule mining

Date: Aug 09, 2016 09:30 AM

URL: http://pirsa.org/16080005

Abstract: Imagine you run a supermarket, and assume that for each customer â€œuâ€• you record what â€œuâ€• is buying. For instance, you may observe that u=1 typically buys bread and cheese and u=2 typically buys bread and salami. Studying your dataset you suspect that generally, customers who are likely to buy cheese are likely to buy bread as well. Rules of this kind are called association rules. Mining association rules is of significant practical importance in fields like market basket analysis and healthcare. In this talk I introduce a novel method for association rule mining which is inspired by ideas from classical statistical mechanics and quantum foundations.

Outline

- Topic models
- Their relation to statistical physics
- Association rule mining
- Matrix factorization techniques for ARM
- Numerical results
- Conclusions

Topic models

- Imagine you are the executive editor of the New York Times.
- Just digitalized all articles since 1851.
- Result: millions of text documents $d_1, d_2, \ldots, d_N$.
- Docs have **topics**. E.g., Politics, Sports, Finance, People, etc.
- For example, $d_{217} = 80\%$ Sports $+$ 20% Politics

- Inference of topic mixture is objective of topic modeling.

Topic models

- Most important topic models:
  - **probabilistic Latent Semantic Indexing (pLSI)**[1]; $\sim 4k$ citations
  - **Latent Dirichlet Allocation (LDA)**[2]; $\sim 15k$ citations
  - pLSI $\approx$ LDA[3]
- In pLSI, we regard each word w as iid random variable $\in \{vocabulary\}$. More precisely, we postulate

$$P[w_i = w|d_j] = \sum_{t=1}^{T} P[t|d_j]\ P[w|t]$$

t: abstract topic
$P[\bullet|d_j]$: multinomial dist
$P[\bullet|t]$: multinomial dist

[1] Thomas Hofmann. "Probabilistic latent semantic indexing.", 1999.
[2] D. Blei, A. Ng, and M. Jordan. "Latent dirichlet allocation.", 2003.
[3] M. Girolami, and A. Kaban. "On an equivalence between PLSI and LDA.", 2003.

Reformulation

- Define $\vec{F}_w, \vec{p}_d \in \mathbb{R}^T$ by

$$(\vec{F}_w)_t = P[w|t]$$
$$(\vec{p}_d)_t = P[t|d].$$

  Thus, $\sum_{w=1}^{W} \vec{F}_w = (1, ..., 1)^T$.

- Therefore, $(\vec{F}_1, ..., \vec{F}_W) \in \mathcal{M}$

$$\mathcal{M} = \left\{ (\vec{F}_1, ..., \vec{F}_W) \in [0,1]^{T \times W} \middle| \sum_{t=1}^{T} \vec{F}_t = (1, ..., 1)^T \right\}$$

- Consequently,

rep. of **words** is $(\vec{F}_1, ..., \vec{F}_W) \in \mathcal{M}$

rep. of **docs** is $\vec{p}_d \in \Delta_T$.

Statistical mechanics

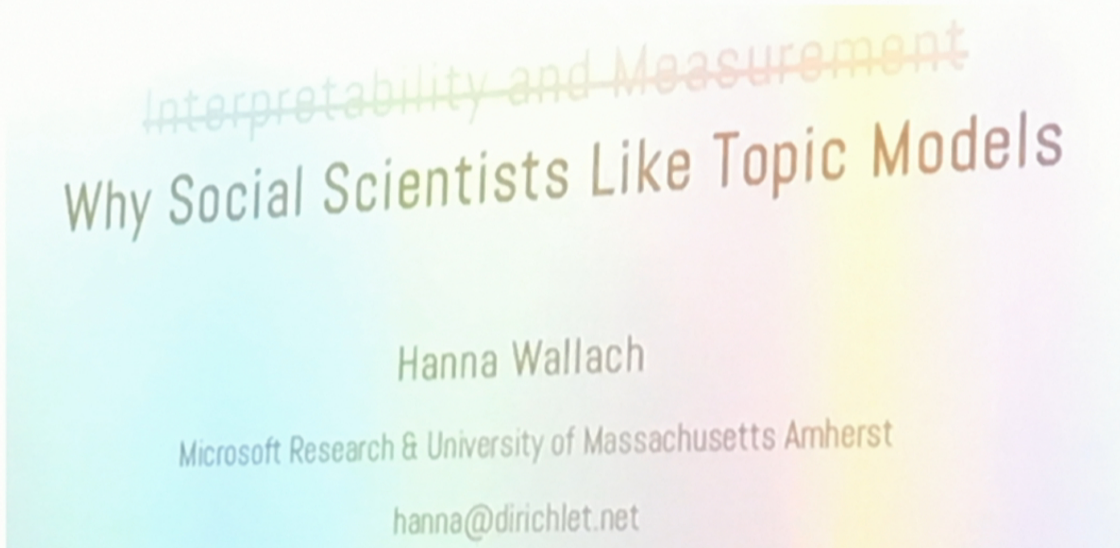
- **Traditionally:**
  - Let $\Gamma = \mathbb{R}^3 \times \mathbb{R}^3$ = phase space of particle in $\mathbb{R}^3$
  - States are p-distributions μ on Γ
  - Observables are functions $f : \Gamma \to \mathbb{R}$
  - $P\big[f \in [a,b]\big] = \mu(f^{-1}([a,b]))$
- **Finite formulation:**
  - Let $\Gamma = \{\omega_1, ..., \omega_T\}$
  - States are p-distributions $\vec{p}$ on Γ
  - Observables are functions $f : \Gamma \to \{1, ..., W\}$
  - $P\big[f = w\big] = P[f^{-1}(w)] = \sum_{t=1}^{T} (\vec{p})_t (\vec{E}_w)_t$ where

$$(\vec{E}_w)_t = \begin{cases} 1, & \text{if } \omega_t \in f^{-1}(w), \\ 0, & \text{otherwise.} \end{cases}$$

- $(\vec{E}_1, ..., \vec{E}_W) \in \mathcal{M}_0$ where
  $\mathcal{M}_0 := \left\{ (\vec{E}_1, ..., \vec{E}_W) \in \{0,1\}^{T \times W} \middle| \sum_{t=1}^{T} \vec{E}_t = (1, ..., 1)^T \right\}$
- $\mathcal{M} = \text{conv}(\mathcal{M}_0)$

Relation between topic modeling and statistical mechanics

- Hence, in pLSI, $\vec{p}_d$ is the *state* of doc d, and
- $(\vec{F}_1, \ldots, \vec{F}_W)$ is the representation of the ***measurement*** "What is the m-th word?" for arbitrary m.
- These **metaphors** are appreciated in social sciences...



Outline

- Topic models
- Their relation to statistical physics
- **Association rule mining**
- Matrix factorization techniques for ARM
- Numerical results
- Conclusion including Importance of study of foundations

Association rule mining

Example: market basket data

transactionID	bread	milk	cheese	beer	diaper	salami
1	1	1	1	0	0	0
2	1	0	1	1	1	0
3	1	1	0	1	0	1
4	1	0	0	1	1	1

- Note: $\underbrace{\text{cheese} \Rightarrow \text{bread}, \text{salami} \Rightarrow \text{bread}}_{\text{Association rules}}$
- Applications for instance in **market basket analysis** and **healthcare**.

From statistical mechanics to association rules

Statistical mechanics model for transaction data so that, e.g.,

$$P[\text{basket contains bread}|\text{customer } u] = \vec{p}_u{}^T \vec{E}_{\text{bread}}$$

Assume that

$$\text{supp}(\vec{E}_{\text{cheese}}) \subseteq \text{supp}(\vec{E}_{\text{bread}}).$$

Then, *Customers who are likely to buy cheese are likely to buy bread.* I.e., cheese $\Rightarrow$ bread.

Problem: These models are hard to infer.

Models that are easy to infer: Matrix factorizations.

Objective: Use previous ideas to infer ARs from matrix factorization models.

For this: we reformulate this def of ARs by noticing that...

Variational characterization

cheese $\Rightarrow$ bread

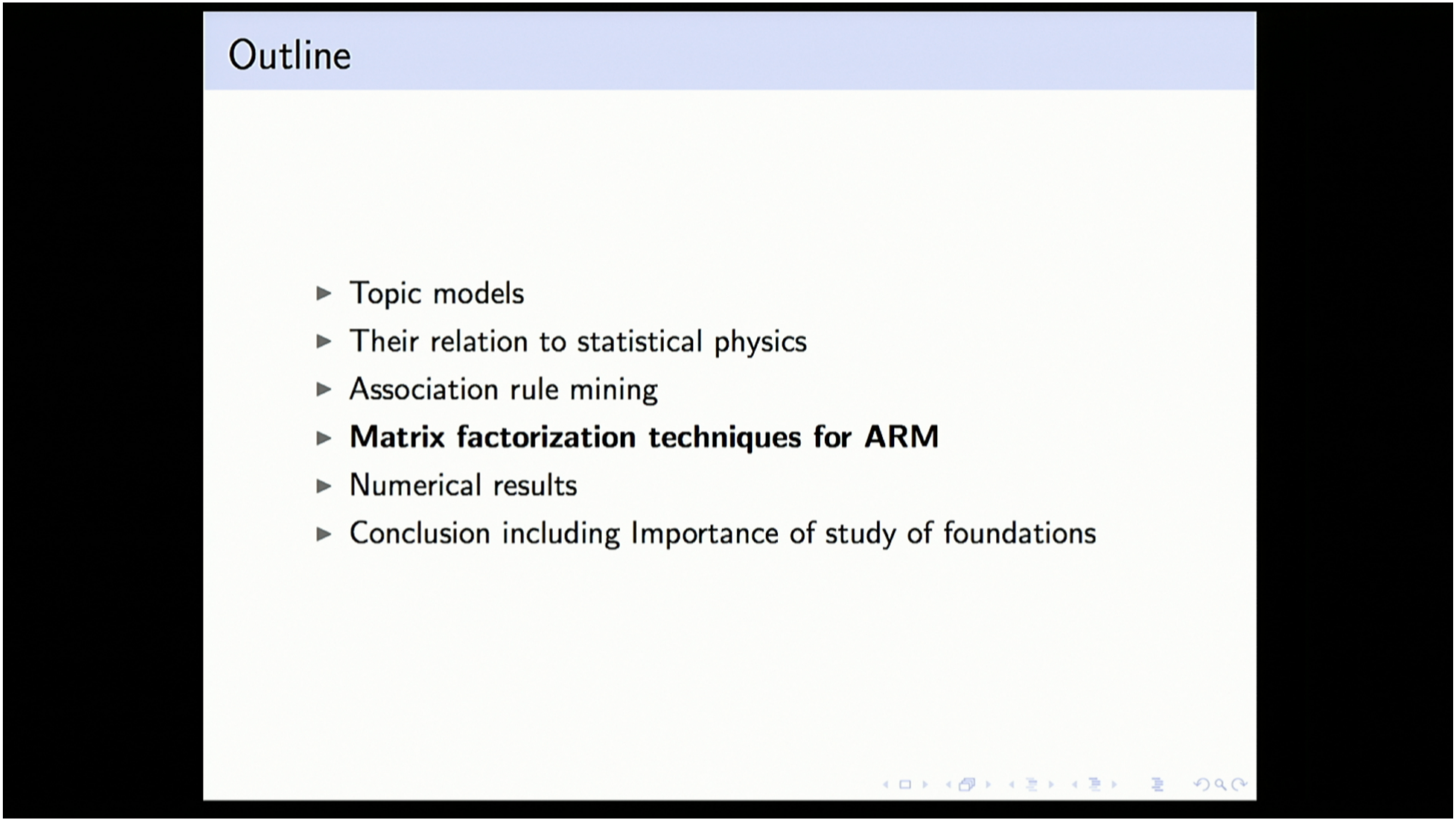
$:\Leftrightarrow \text{supp}(\vec{E}_{\text{cheese}}) \subseteq \text{supp}(\vec{E}_{\text{bread}})$

$\Leftrightarrow \eta_{\text{bread}}^{\text{cheese}} = 1$ where

$$\begin{aligned} &\eta_{\text{bread}}^{\text{cheese}} \\ &:= \min \quad \vec{p}^T \vec{E}_{\text{bread}} \ \text{ s.t. } \ \vec{p} \in \Delta_T, \ \vec{p}^T \vec{E}_{\text{cheese}} = 1 \\ &= \min \quad \vec{p}^T \vec{E}_{\text{bread}} \ \text{ s.t. } \ \vec{p} \in \underbrace{\{\text{every possible } \vec{p}\}}_{\text{STATE SPACE } \mathcal{S}}, \ \vec{p}^T \vec{E}_{\text{cheese}} = 1 \end{aligned}$$

- Hence, ARs expressed in terms of **optimization problems on state space**.
- Thus, same method applicable in topic modeling.[4]
- There exist datasets where traditional ARM methods fail to discover ARs where we physics succeeds in discovering ARs.

[4] E.g., to hierarchically classify docs; reverse roles of words and docs.


Outline
Topic models
Their relation to statistical physics
Association rule mining
Matrix factorization techniques for ARM
Numerical results
Conclusion including Importance of study of foundations

Matrix factorization techniques for AR mining

Matrix factorization models?

- Let R be market basket data matrix.
- Very common assumption: R is low-rank; set $d = \mathrm{rank}(R)$
- Thus, $\exists\ \vec{a}_u,\ \vec{b}_i \in \mathbb{R}^d$ s.t. $R_{ui} = \vec{a}_u^T \vec{b}_i$
- We think of $\vec{a}_u$ as **states** and of $\vec{b}_i$ as **measurements**.

Set $\eta^i_j :=$

$$\min \quad \vec{s}^T \vec{b}_j \ \text{ s.t. } \ \vec{s} \in \underbrace{\{\text{every possible } \vec{a}_u\}}_{\text{STATE SPACE } \mathcal{S}},\ \vec{s}^T \vec{b}_i = 1$$

Want to define: $i \Rightarrow j$ if $\eta^i_j = 1$

Q: $\mathcal{S} =??$

Matrix factorization techniques for AR mining

- **Q:** $\mathcal{S} = ??$
- $\mathcal{S} = \{\vec{a} \in \mathbb{R}^d | \vec{a}^T \vec{v} \in [0,1] \ \ \forall \vec{v} \in \mathcal{M}\}$
- $\mathcal{S}^{\text{in}} := \text{conv}\{\vec{a}_u\}_{u \in [U]} \subseteq \mathcal{S}$
- $\mathcal{S} \subseteq \{\vec{a} \in \mathbb{R}^d | \vec{a}^T \vec{b}_i \in [0,1] \ \ \forall i \in [I]\} =: \mathcal{S}^{\text{out}}$

Hence, $\eta_j^{\text{i,out}} \leq \eta_j^{\text{i}} \leq \eta_j^{\text{i,in}}$ where

$$\eta_j^{i,\text{in}} := \min \quad \vec{s}^T \vec{b}_j \ \text{ s.t. } \ \vec{s} \in \mathcal{S}^{\text{in}}, \ \vec{s}^T \vec{b}_i = 1$$

and

$$\eta_j^{i,\text{out}} := \min \quad \vec{s}^T \vec{b}_j \ \text{ s.t. } \ \vec{s} \in \mathcal{S}^{\text{out}}, \ \vec{s}^T \vec{b}_i = 1$$

In the remainder: use $\mathcal{S}^{\text{out}}$ because due to "mixedness" of $\vec{a}_u$'s, $\mathcal{S}^{\text{in}}$ is very small. **Hence,** $i \Rightarrow j$ if $\eta_j^{i,\text{out}} = 1$.
Neat: time complexity for $\mathcal{S}^{\text{out}}$ independent of U.

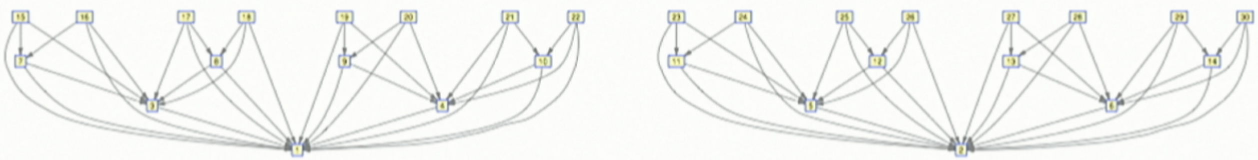
Outline

- Topic models
- Their relation to statistical physics
- Association rule mining
- Matrix factorization techniques for ARM
- **Numerical results**
- Conclusion including Importance of study of foundations

Synthetic data

Synthetic data = samples from statistical mechanics model with

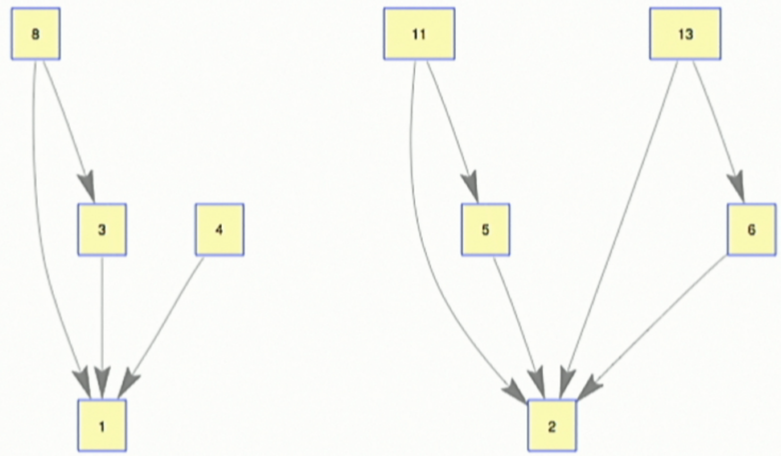
- highly mixed states, and
- with the measurements



- Baseline: **apriori algorithm**[5]

[5] R. Agrawal, and S. Ramakrishnan. 1994.

Synthetic data: apriori algorithm



Synthetic data: MF-AR

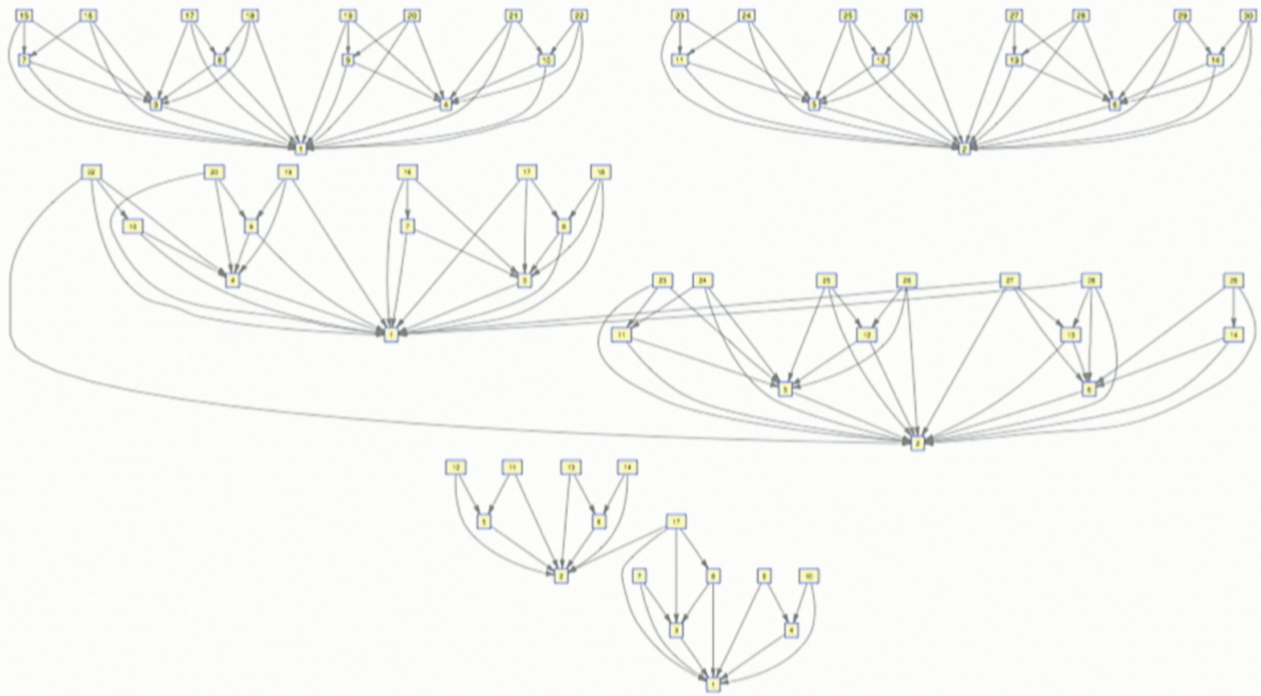


Figure: Synthetic data generated by a binary tree with **4 layers**.
Top: $d = 2^4$. *Middle*: $d = 14$. *Bottom*: $d = 12$.

Movie recommendation

Given:

	Jurassic Park	*Casablanca*	*Airplane!*	*Notting Hill*
Alice	⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆
Bob	⋆⋆⋆⋆	⋆⋆⋆⋆	?	⋆
Charlie	?	⋆⋆⋆	⋆⋆⋆	⋆⋆⋆
Donald	?	?	?	⋆⋆⋆⋆⋆

Dataset: MovieLens 1M
(1 million 5-star ratings, 6'040 users, 3'706 movies)

MovieLens association rules

Jurassic Park (1993) $\Rightarrow$

Star Wars IV (1977) Matrix (1999) Star Wars V (1980) Terminator 2 (1991) Men in Black (1997)	Stargate (1994) Independence Day (1996) Lost World: Jurassic Park (1997) Men in Black (1997) Total Recall (1990)
apriori (by support)	**MF-AR**

Summary

- Topic models $\approx$ Statistical Mechanics
- AR mining in statistical mechanics
- AR from MF
- Use notions of state and meas spaces.
  $\rightarrow$ Notions form quantum foundations.
- Quantum foundations can be highly practical!!
- Examples

Future work

- Dynamics![6]
- Thermodynamics?[7]

Thank you!

[6] D. Blei and J. Lafferty, 2006.
[7] Your paper, 2017