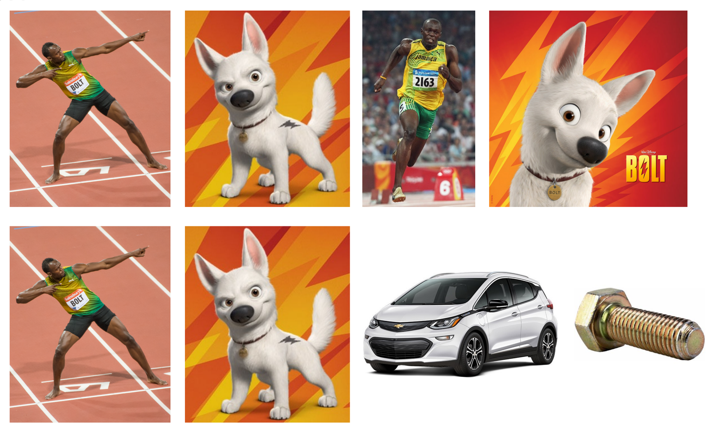
Efficient approximate sampling for projection DPPs

Guillaume Gautier[†,*], Rémi Bardenet[†], Michal Valko[*]

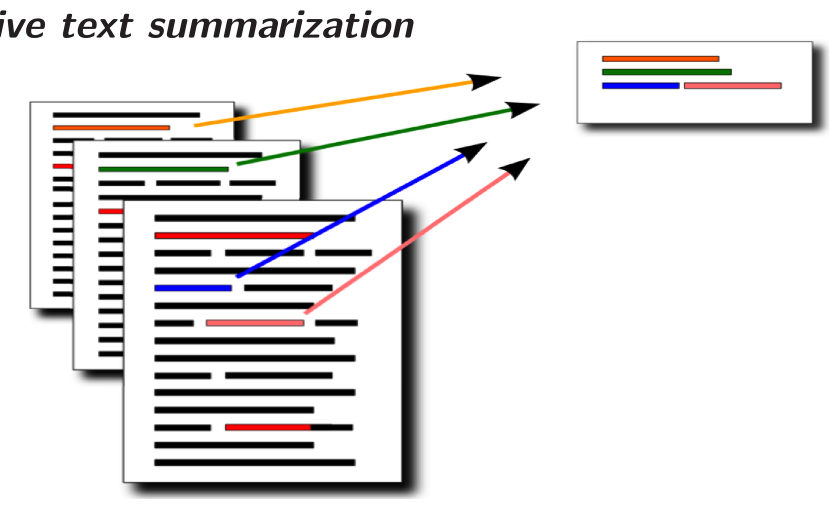
[†] Univ. Lille, CNRS, Centrale Lille, UMR 9189 — CRIStAL [*] INRIA Lille — Nord Europe, SequeL team

Motivation: Sample with diversity

Image search



Extractive text summarization



DPPs provide rigorous approach for sampling diverse subsets

Determinantal Point Processes

Definitions

- $\{1,\dots,N\}$ indices/labels of items
- $\mathbf{K}$ a $N\times N$ PSD matrix
- $\operatorname{DPP}(\mathbf{K})$ a measure on subsets of $\{1,\dots,N\}$
- $\mathcal{X}\sim\operatorname{DPP}(\mathbf{K})$ if $\forall S\subseteq\{1,\dots,N\}$,

$$\mathbb{P}[S\subseteq\mathcal{X}]=\det\mathbf{K}_S$$

- Existence when $\mathbf{0}_N\preceq\mathbf{K}\preceq\mathbf{I}_N$

Projection DPPs

- $\mathbf{K}$ is an orthogonal projection matrix

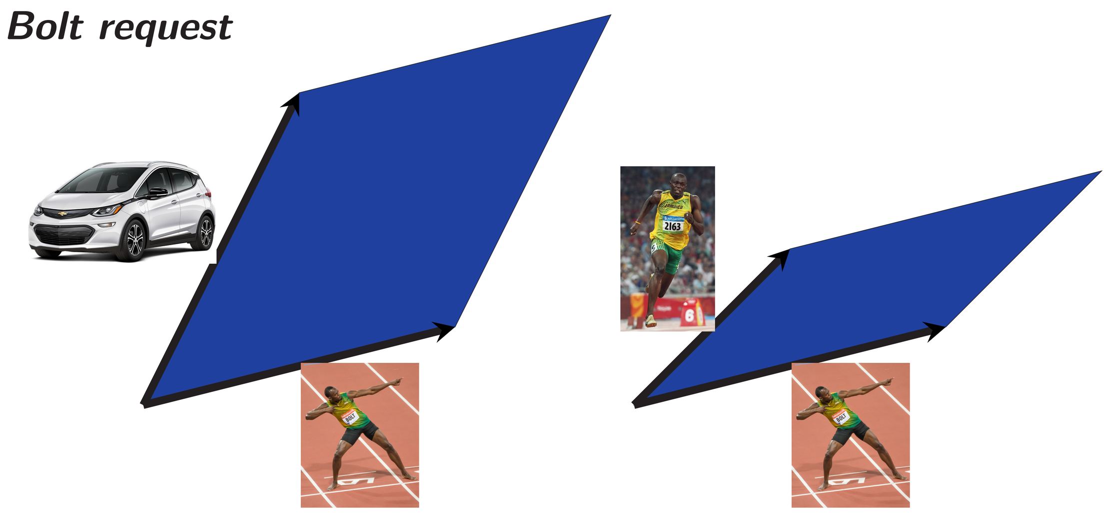
$$\mathbf{K}=\sum_{i=1}^{r}u^{(i)}{u^{(i)}}^{\intercal}=\mathbf{\Phi}^{\intercal}\mathbf{\Phi},\qquad\text{with }\langle u^{(i)},u^{(j)}\rangle=\delta_{ij}$$

- Interpretation

$$\mathbb{P}[S\subseteq\mathcal{X}]=\det\mathbf{K}_S=\operatorname{Vol}^2\{\phi_i\,;\,i\in S\}$$

- $|\mathcal{X}|\overset{a.s.}{=}\operatorname{Tr}\mathbf{K}=\operatorname{rank}\mathbf{K}=r$

Bolt request



Properties

- If $\mathcal{X}\sim\operatorname{DPP}(\mathbf{K})$, then $\forall i,j$

$$\begin{aligned}\mathbb{P}[\{i,j\}\subseteq\mathcal{X}]&=\begin{vmatrix}\mathbb{P}[i\in\mathcal{X}]&\mathbf{K}_{ij}\\\mathbf{K}_{ij}&\mathbb{P}[j\in\mathcal{X}]\end{vmatrix}\\&=\mathbb{P}[i\in\mathcal{X}]\,\mathbb{P}[j\in\mathcal{X}]-\mathbf{K}_{ij}^2\\&\leq\mathbb{P}[i\in\mathcal{X}]\,\mathbb{P}[j\in\mathcal{X}]\end{aligned}$$

- $|\mathbf{K}_{ij}|\approx$ similarity between i and j
  - The larger $|\mathbf{K}_{ij}|$ the smaller $\mathbb{P}[\{i,j\}\subseteq\mathcal{X}]$
  - Diversity/repulsion
  - $|\mathbf{K}_{ij}|$ yields departure from independence

Goal

Exact sampling is expensive

⇓

Provide efficient samplers for DPPs

Setup

Build an $r\times N$ feature matrix $\mathbf{A}=\left(\sqrt{q_1}\phi_1|\dots|\sqrt{q_N}\phi_N\right)$

Assumption 1. *$\mathbf{A}$ is full row rank i.e.* $\operatorname{rank}A=r$

Construct the projection kernel $\mathbf{K}=\mathbf{A}^{\intercal}[\mathbf{A}\mathbf{A}^{\intercal}]^{-1}\mathbf{A}$

Notations

- For $|B|=r$, $\mathbf{B}\triangleq\mathbf{A}_{:B}$

  Let $B=\{i_1,\dots,i_r\}$, then for $\mathcal{X}\sim\operatorname{DPP}(\mathbf{K})$

$$\mathbb{P}[\mathcal{X}=B]=\frac{|\det\mathbf{B}|^2}{\det\mathbf{A}\mathbf{A}^{\intercal}}\propto\operatorname{Vol}^2\mathbf{B}$$

- $\mathcal{B}\triangleq\{B\,;\,|B|=r,\det\mathbf{B}\neq0\}$
  - Indices of columns of $\mathbf{A}$ forming a basis of $\operatorname{Im}\mathbf{A}$
  - $\mathcal{B}$ is the support of $\operatorname{DPP}(\mathbf{K})$

State space represented as a tiling of a zonotope

The volume spanned by the feature vectors

$$\mathcal{Z}(\mathbf{A})\triangleq\mathbf{A}[0,1]^N$$

admits a natural tiling

$$\operatorname{Vol}\mathcal{Z}(\mathbf{A})=\sum_{B\in\mathcal{B}}\operatorname{Vol}\mathbf{B}=\sum_{B\in\mathcal{B}}|\det\mathbf{B}|$$

The optimal solution y^* of $P_x(\mathbf{A},c)$

$$\begin{aligned}\min_{y\in\mathbb{R}^N}\quad&c^{\intercal}y\\\text{s.t.}\quad&\mathbf{A}y=x\\&0\leq y\leq1\end{aligned}$$

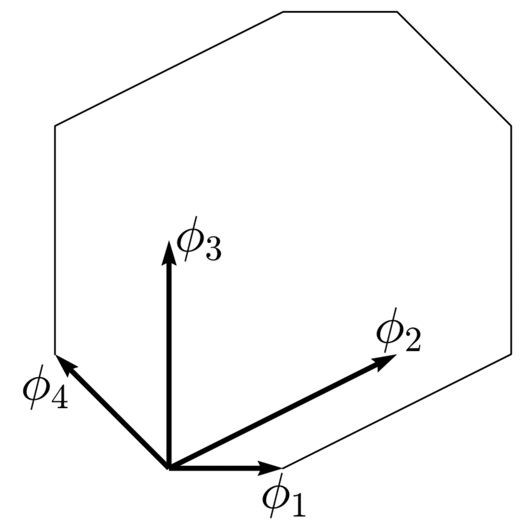
takes the form

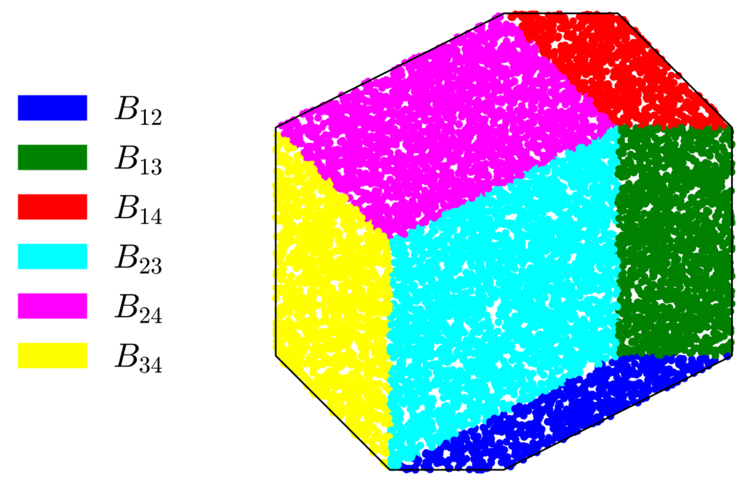
$$x=\mathbf{A}y^*=\mathbf{A}\xi(x)+\mathbf{B}_xu$$

for a unique $B_x\in\mathcal{B}$, where

- $u\in[0,1]^r$
- $\xi(x)\in\{0,1\}^N$ s.t. $\xi(x)_{|B_x}=0$

Any $x\in\mathcal{Z}(\mathbf{A})$ falls inside a uniquely defined parallelotope $\mathcal{Z}(\mathbf{B}_x)$ shifted by $\mathbf{A}\xi(x)$





Manipulating the optimality conditions:

- Each basis $B\in\mathcal{B}$ can be realized as a B_x for some x
- Any $x'\in\mathbf{A}\xi(x)+\mathcal{Z}(\mathbf{B}_x)$ will be assigned $B_{x'}=B_x$

$\mathcal{Z}(\mathbf{A})$ is tiled by all parallelotopes $\mathcal{Z}(\mathbf{B})$, $B\in\mathcal{B}$, with disjoint interiors

Random walk on tiles

Sample projection DPPs with hit-and-run + linear programming

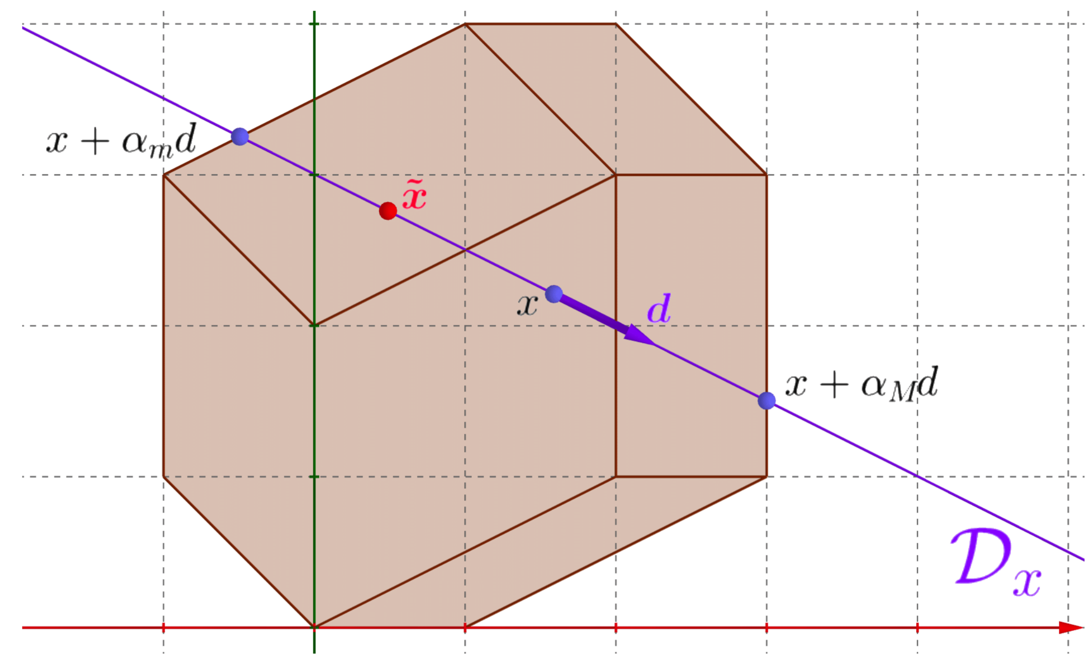
- Use hit-and-run to build an underlying continuous random walk (x_n) in $\mathcal{Z}(\mathbf{A})$ with limiting distribution

$$\pi(x)\,\mathrm{d}x=\sum_{B\in\mathcal{B}}C_{\mathbf{B}}\times\mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$

- Identify the tile in which x_n lies to get a discrete random walk (B_{x_n}) on $\mathcal{B}$ with limiting distribution

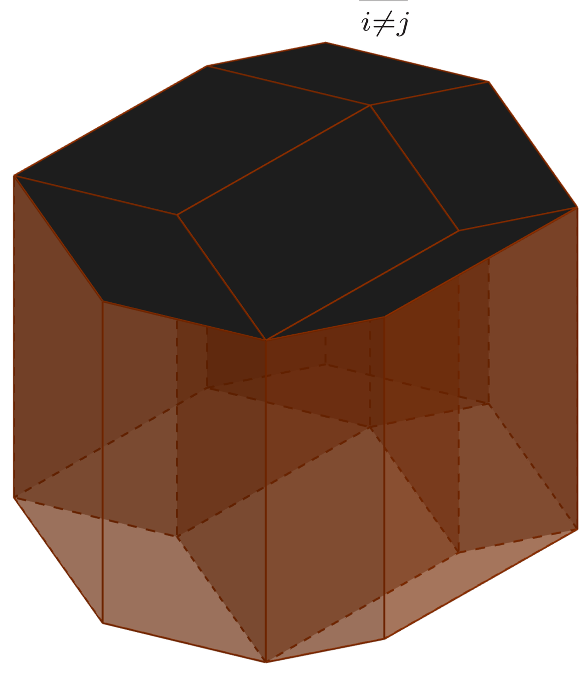
$$\mathbb{P}[B_x=B]\propto\int_{\mathbf{B}}\pi(x)\,\mathrm{d}x=C_{\mathbf{B}}\times\operatorname{Vol}\mathbf{B}$$

  - Solve $P_{x_n}(\mathbf{A},c)$
  - Extract the tile $B_{x_n}=\{i\,;\,y_i^*\in]0,1[\}$



Acceptance $=1$ leads to uniform limiting distribution

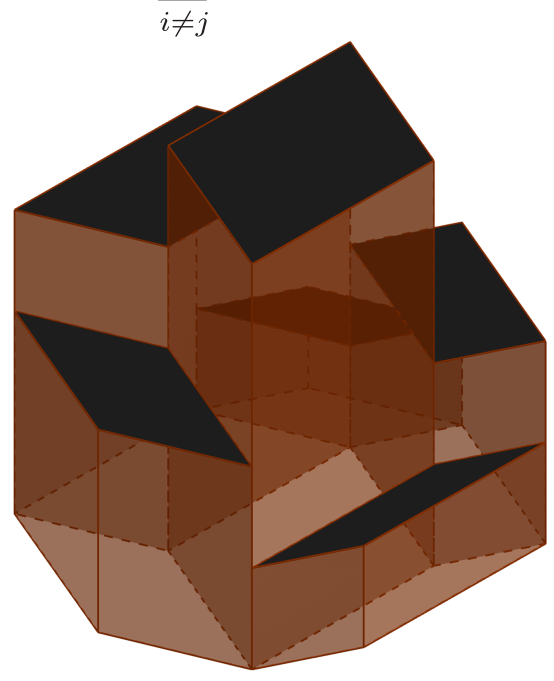
$$\pi(x)\,\mathrm{d}x=\mathbb{1}_{\mathcal{Z}(\mathbf{A})}(x)\,\mathrm{d}x=\sum_{i\neq j}1\times\mathbb{1}_{\mathbf{B}_{ij}}(x)\,\mathrm{d}x$$



$$\mathbb{P}[x\in\mathbf{B}_{ij}]\propto1\times\operatorname{Vol}\mathbf{B}_{ij}=\operatorname{Vol}^1\mathbf{B}_{ij}$$

Acceptance $=\frac{\operatorname{Vol}B(\tilde{x})}{\operatorname{Vol}B(x)}$ leads to volume limiting distribution

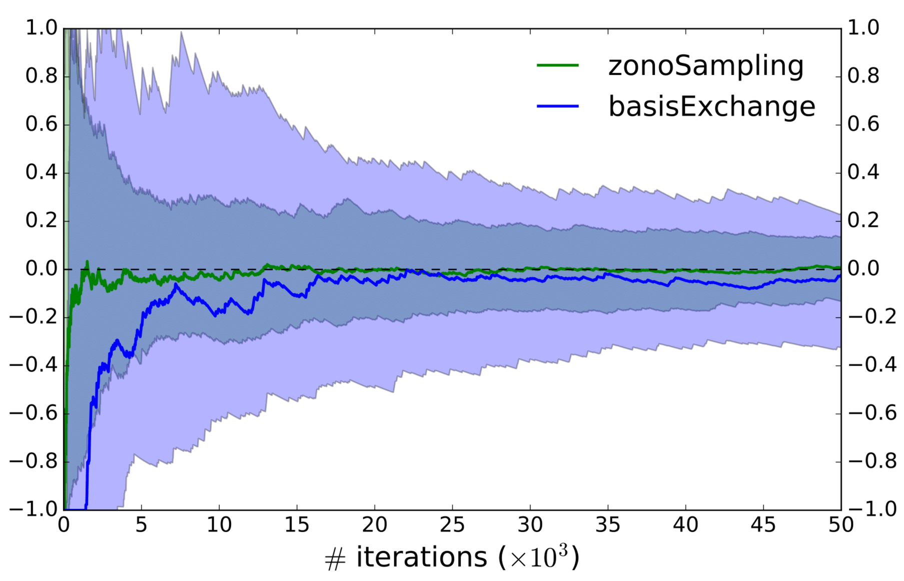
$$\pi(x)\,\mathrm{d}x=\sum_{i\neq j}\operatorname{Vol}\mathbf{B}_{ij}\times\mathbb{1}_{\mathbf{B}_{ij}}(x)\,\mathrm{d}x$$

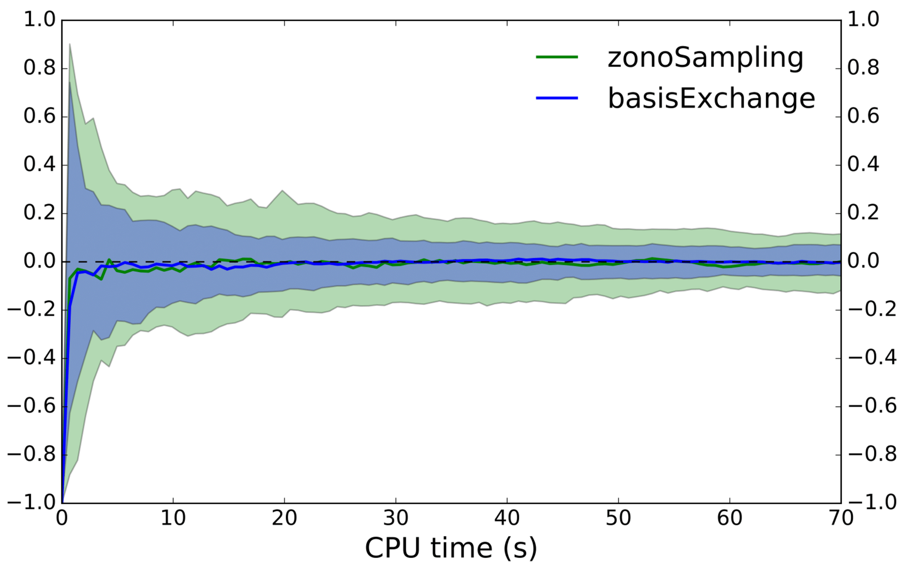


$$\mathbb{P}[x\in\mathbf{B}_{ij}]\propto\operatorname{Vol}\mathbf{B}_{ij}\operatorname{Vol}\mathbf{B}_{ij}=\operatorname{Vol}^2\mathbf{B}_{ij}$$

Some experiments

Relative error of the estimation of $\mathbb{P}[\{i_1,i_2,i_3\}\subseteq\mathcal{X}]$ for non-uniform spanning trees





Extractive text summarization

Chosen uniformly at random

- If you consider yourself to have even a passing familiarity with science, you likely find yourself in a state of disbelief as the president of the United States calls climate scientists "hoaxsters" and pushes conspiracy theories about vaccines.
- In fact, it's so wrong that it may have the opposite effect of what they're trying to achieve.
- Respondents who knew more about science generally, regardless of political leaning, were better able to identify the scientific consensus-in other words, the polarization disappeared.
- In fact, well-meaning attempts by scientists to inform the public might even backfire.
- Psychologists, aptly, dubbed this the "backfire effect."
- But if scientists are motivated to change minds-and many enrolled in science communication workshops do seem to have this goal-they will be sorely disappointed.
- That's not to say scientists should return to the bench and keep their mouths shut.
- Goldman also said scientists can do more than just educate the public: The Union of Concerned Scientists, for example, has created a science watchdogteam that keeps tabs on the activities of federal agencies.
- But I'm learning to better challenge scientists' assumptions about how communication works.
- It's very logical, and my hunch is that it comes naturally to scientists because most have largely spent their lives in school-whether as students, professors, or mentors-and the deficit model perfectly explains how a scientist learns science.
- So in the spirit of doing better, I'll not just write this article but also take the time to talk to scientists in person about how to communicate science strategically and to explain why it matters.

Chosen with our sampler

- If you are a scientist, this disregard for evidence probably drives you crazy.
- So what do you do about it?
- Across the country, science communication and advocacy groups report upticks in interest.
- In 2010, Dan Kahan, a Yale psychologist, essentially proved this theory wrong.
- If the deficit model were correct, Kahan reasoned, then people with increased scientific literacy, regardless of worldview, should agree with scientists that climate change poses a serious risk to humanity.
- Scientific literacy, it seemed, increased polarization.
- This lumps scientists in with the nebulous "left" and, as Daniel Engber pointed out here in Slate about the upcoming March for Science, rebrands scientific authority as just another form of elitism.
- Is it any surprise, then, that lectures from scientists built on the premise that they simply know more (even if it's true) fail to convince this audience?
- With that in mind, it may be more worthwhile to figure out how to talk about science with people they already know, through, say, local and community interactions, than it is to try to publish explainers on national news sites.
- Goldman also said scientists can do more than just educate the public: The Union of Concerned Scientists, for example, has created a science watchdogteam that keeps tabs on the activities of federal agencies.
- There's also a certain irony that, right here in this article, I'm lecturing scientists about what they might not know-in other words, I'm guilty of following the deficit model myself.

Conclusion and next steps

- New interpretable brigde between MCMC and optimization
- Random walk on tiles mixes empirically faster
- Extension to generic DPPs?
- Speed-up by dedicated LP solvers?
- Our sampler for Vol^1 inherits Lovasz & Vempala's mixing time from but not for Vol^2