

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

Jennifer Gillenwater
Advised by Ben Taskar and Emily Fox
Joint work with Alex Kulesza

OUTLINE

OUTLINE

Motivation & Background

OUTLINE

Motivation & Background

1. Dimensionality Reduction

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

2. MAP Estimation

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

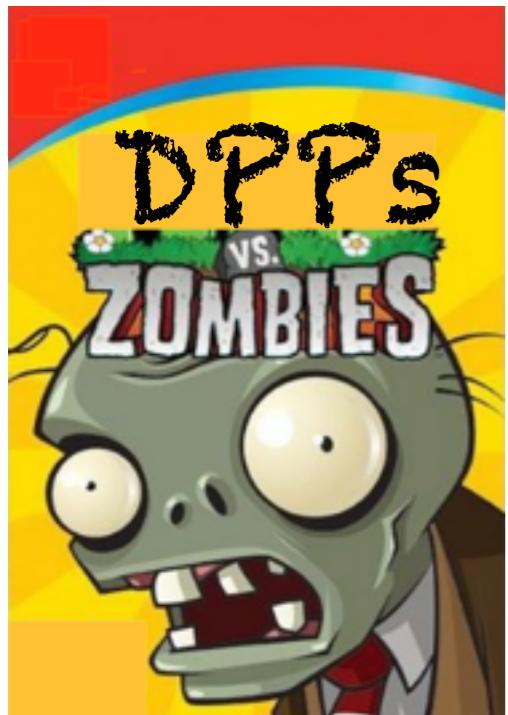
3. Likelihood Maximization

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

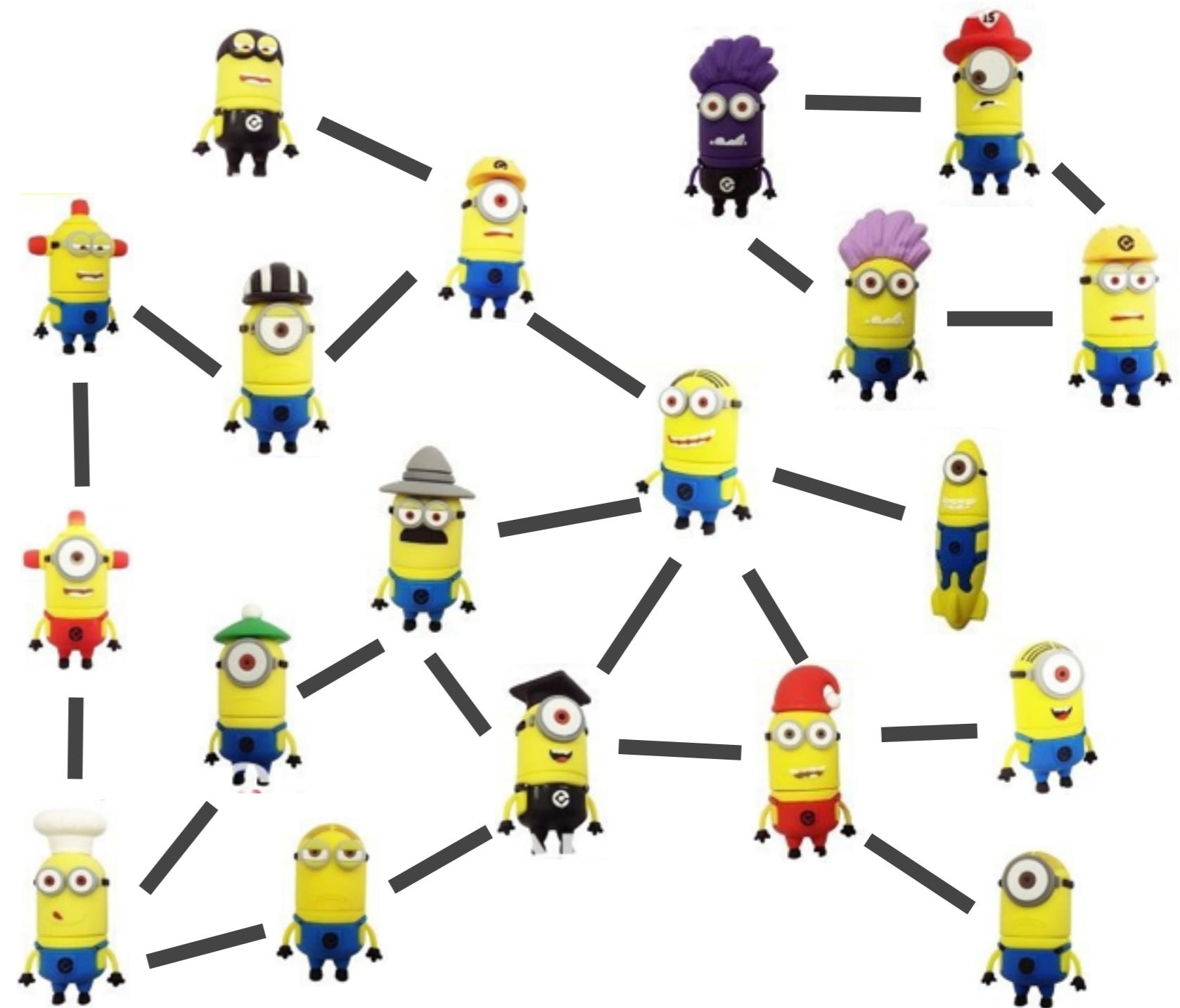
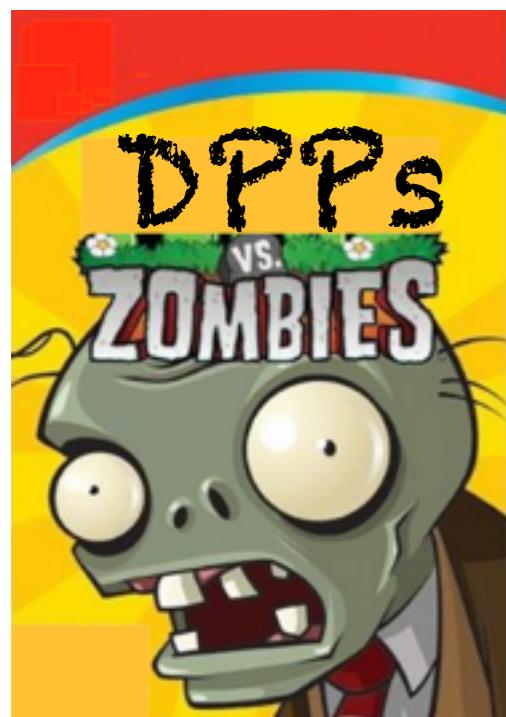
MOTIVATION & BACKGROUND

SOCIAL NETWORK MARKETING

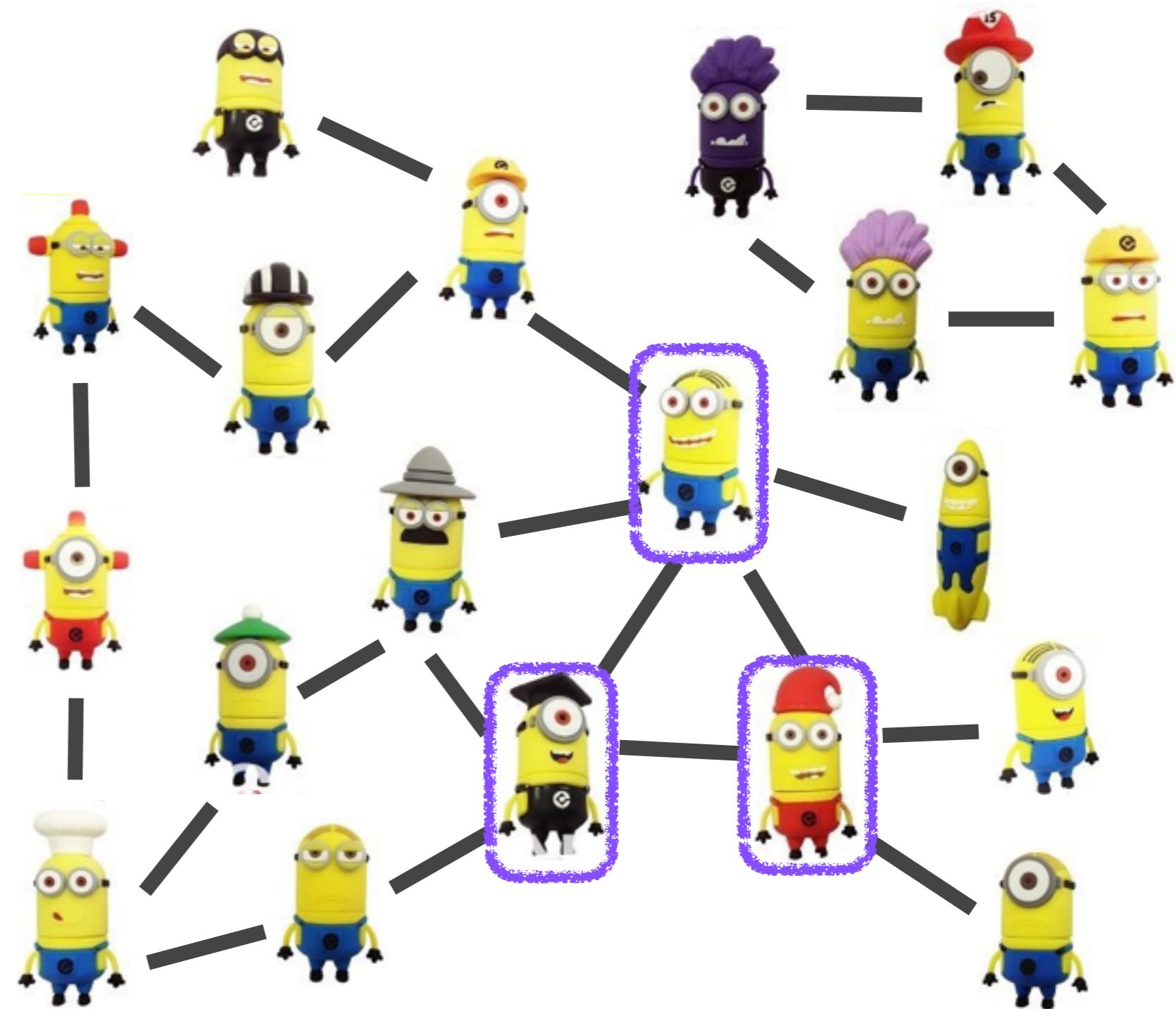
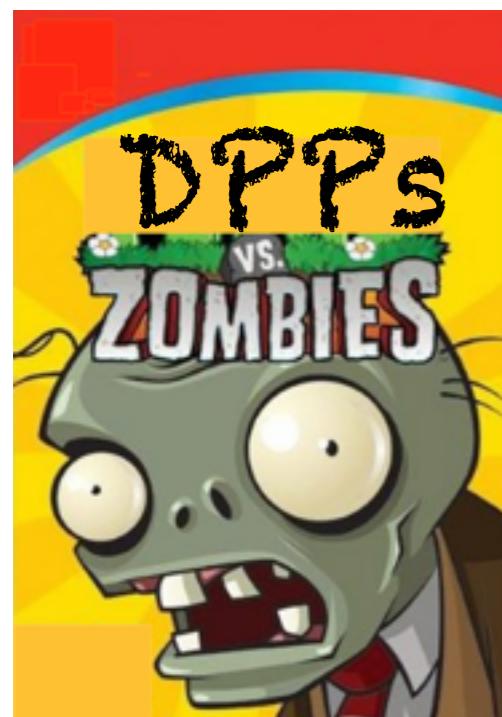
SOCIAL NETWORK MARKETING



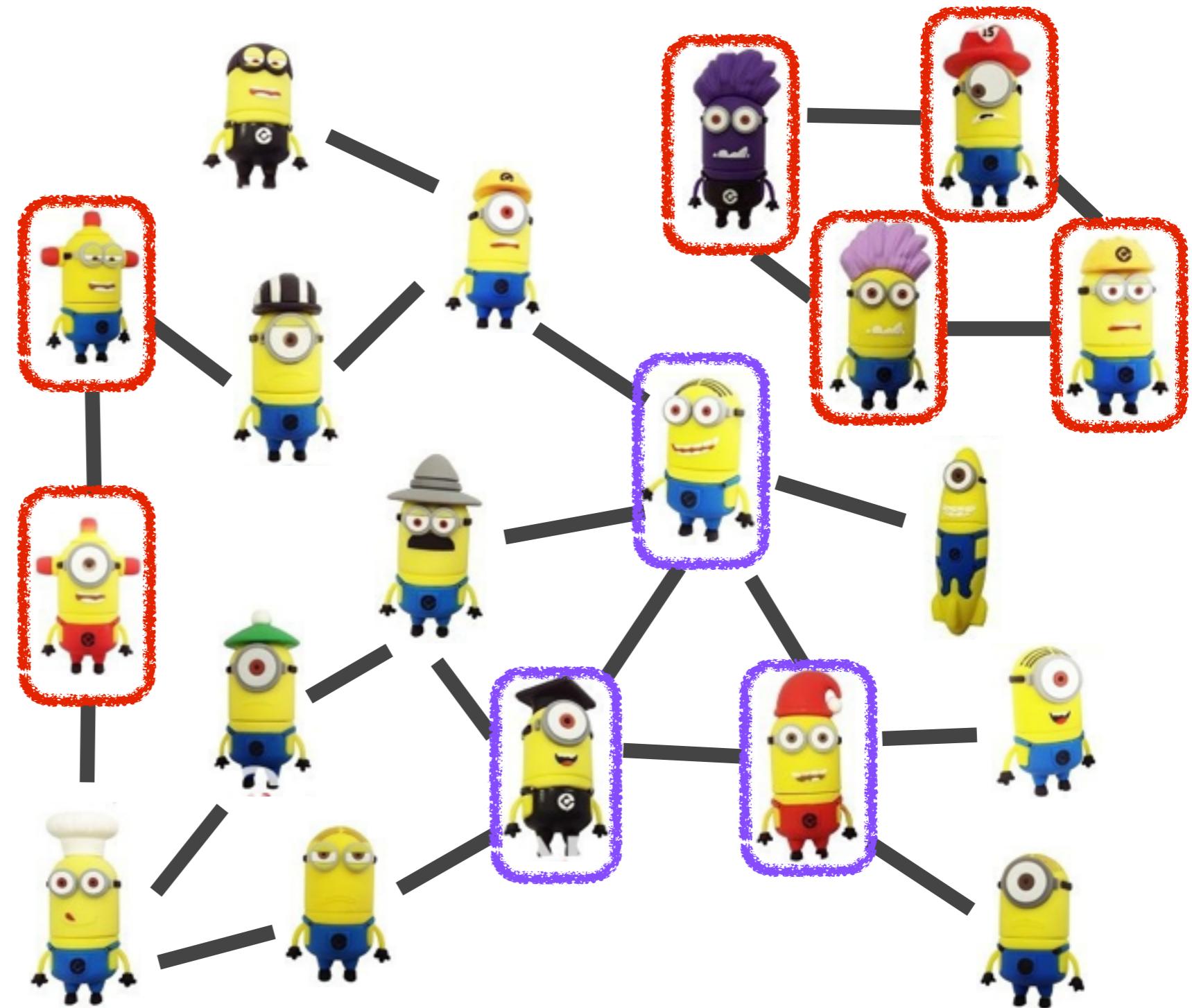
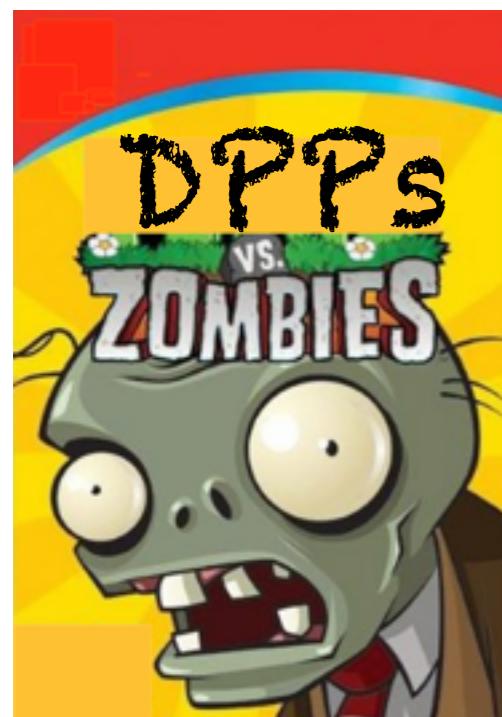
SOCIAL NETWORK MARKETING



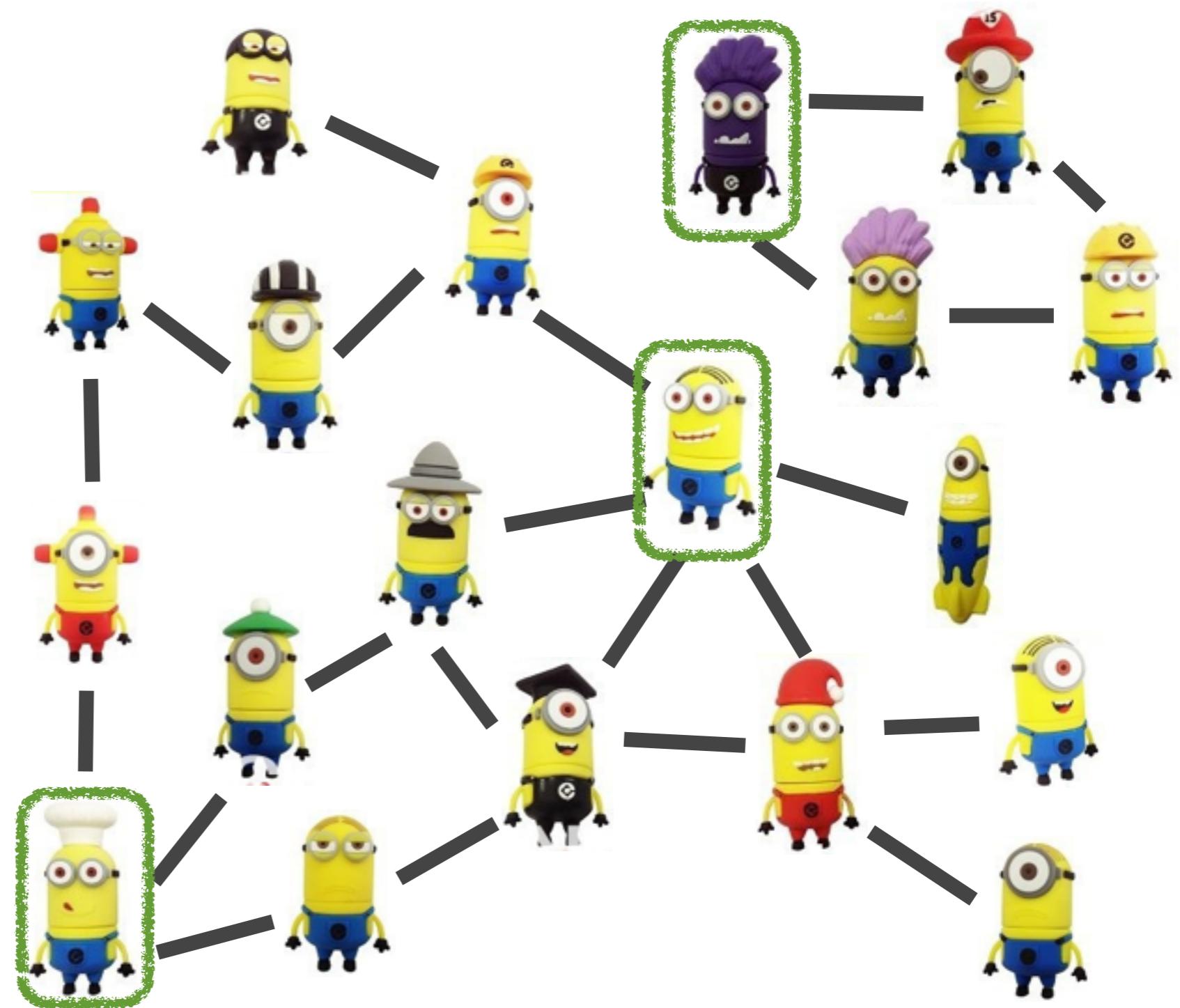
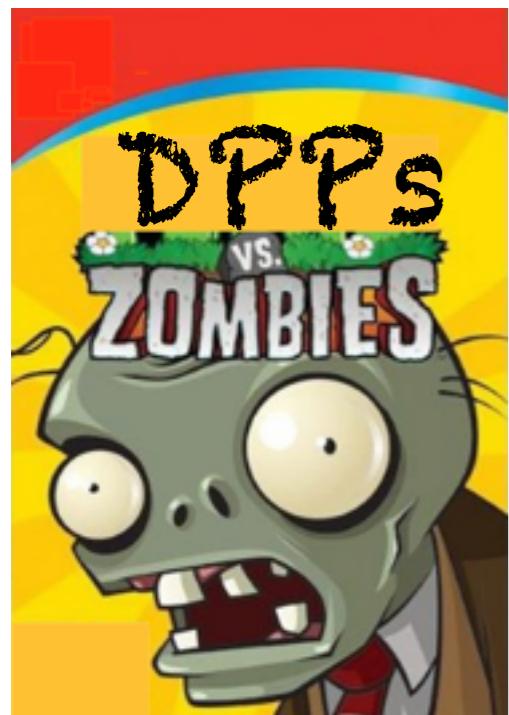
SOCIAL NETWORK MARKETING



SOCIAL NETWORK MARKETING



SOCIAL NETWORK MARKETING



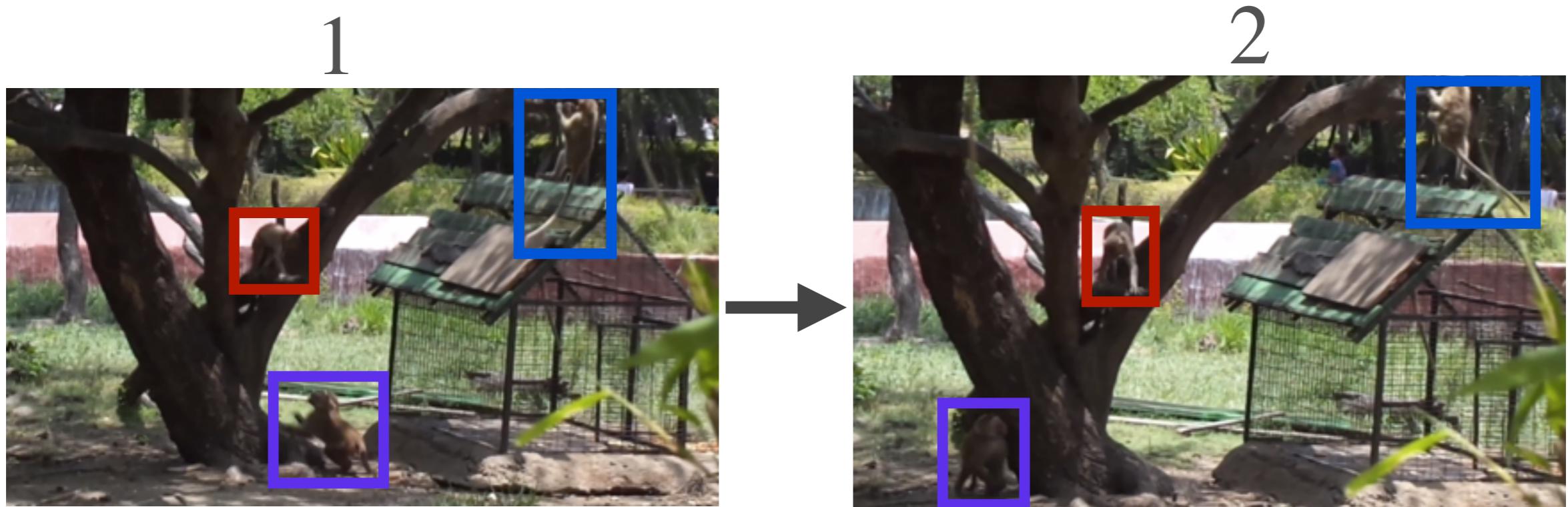
VIDEO TRACKING

VIDEO TRACKING

1



VIDEO TRACKING



VIDEOTRACKING

1



2



3

VIDEOTRACKING

1



2



3

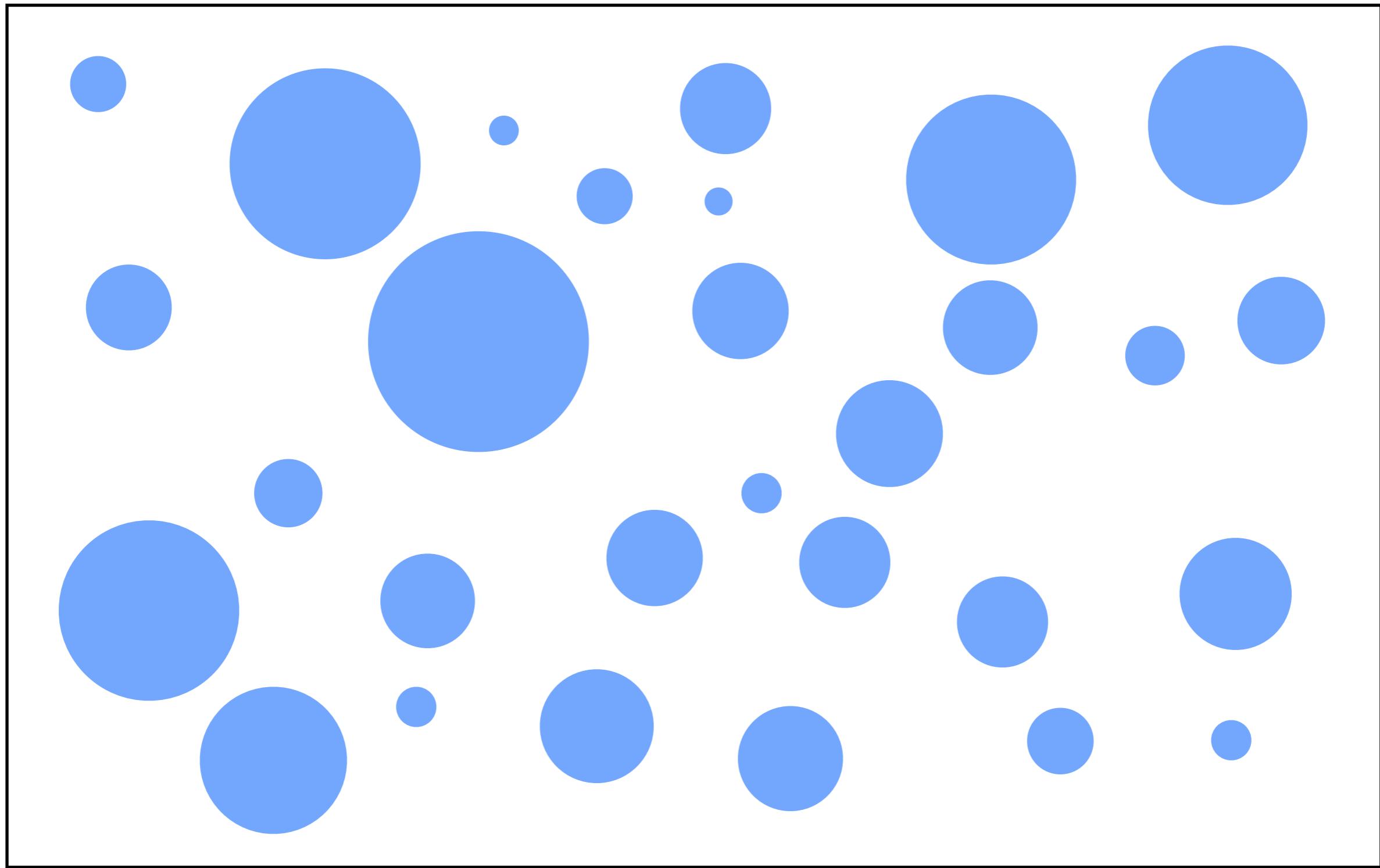


4

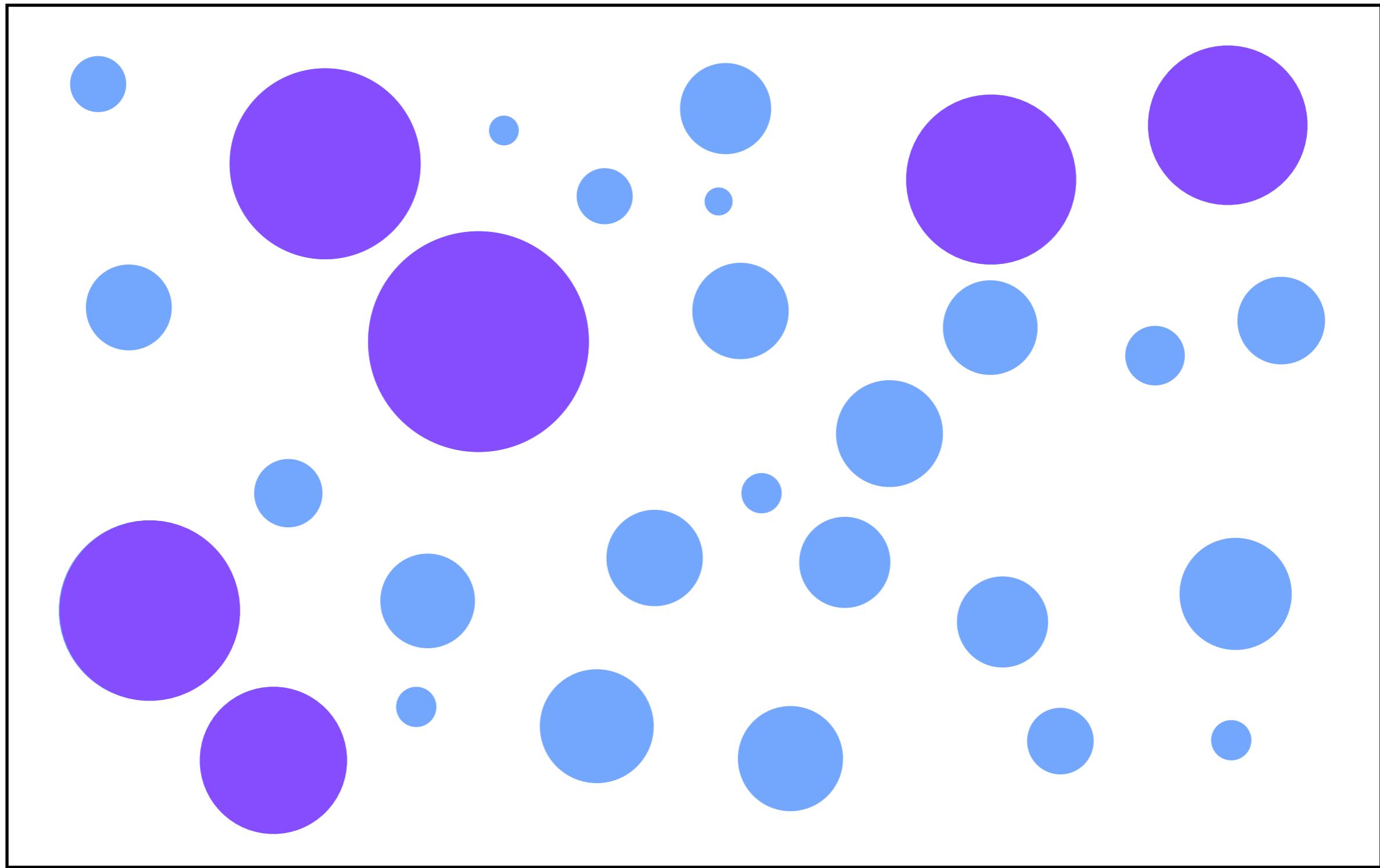


SUBSET SELECTION

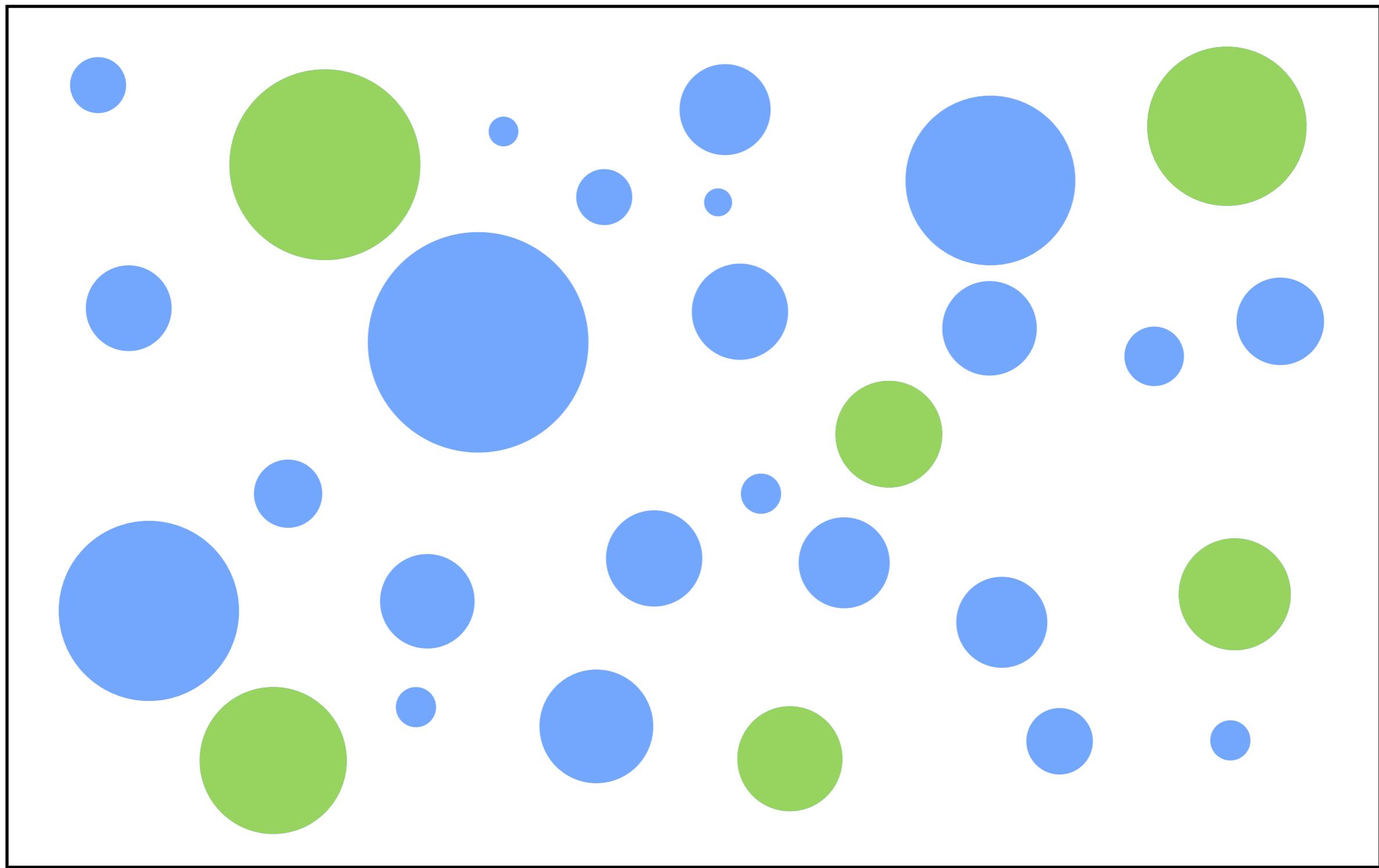
SUBSET SELECTION



SUBSET SELECTION



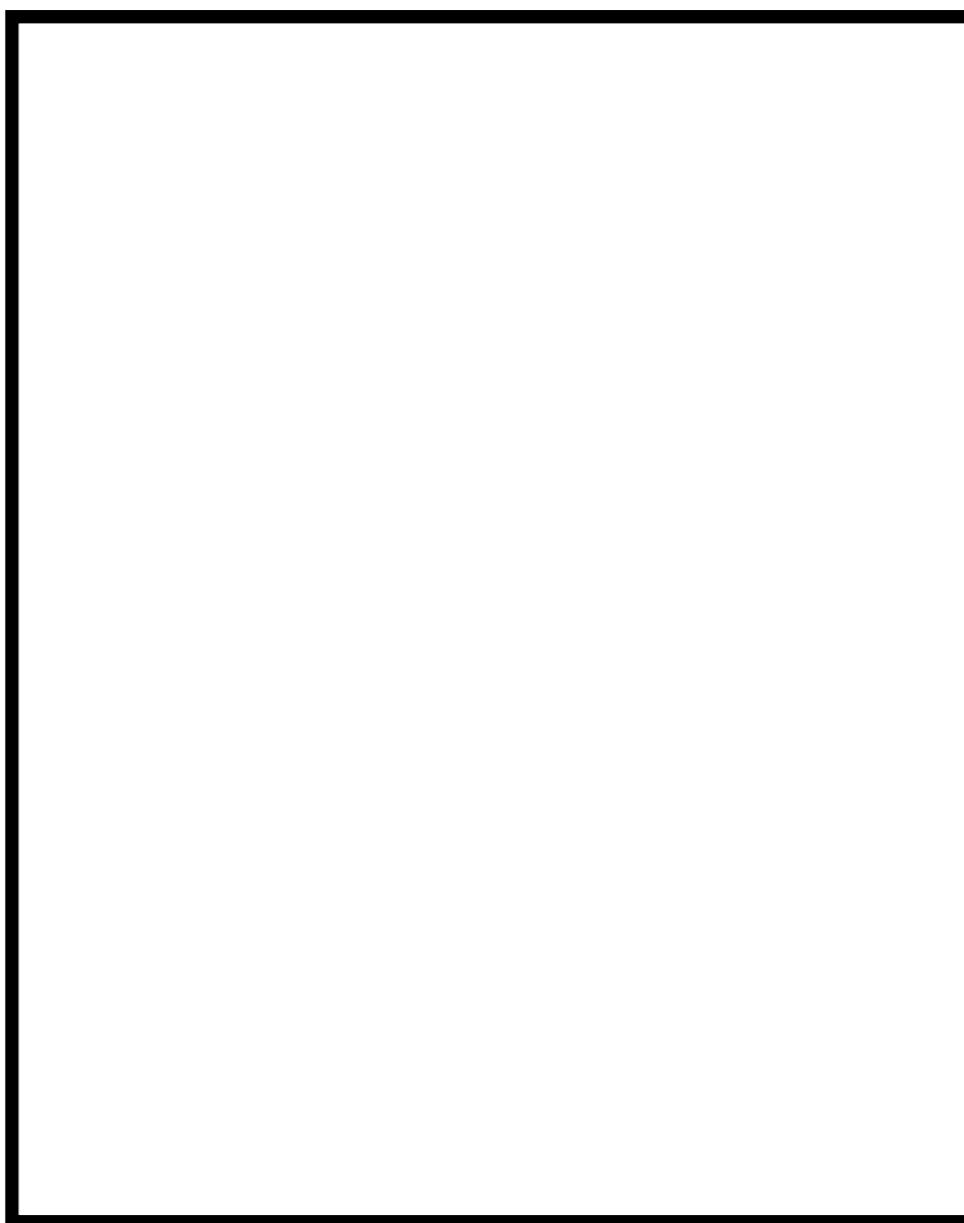
SUBSET SELECTION



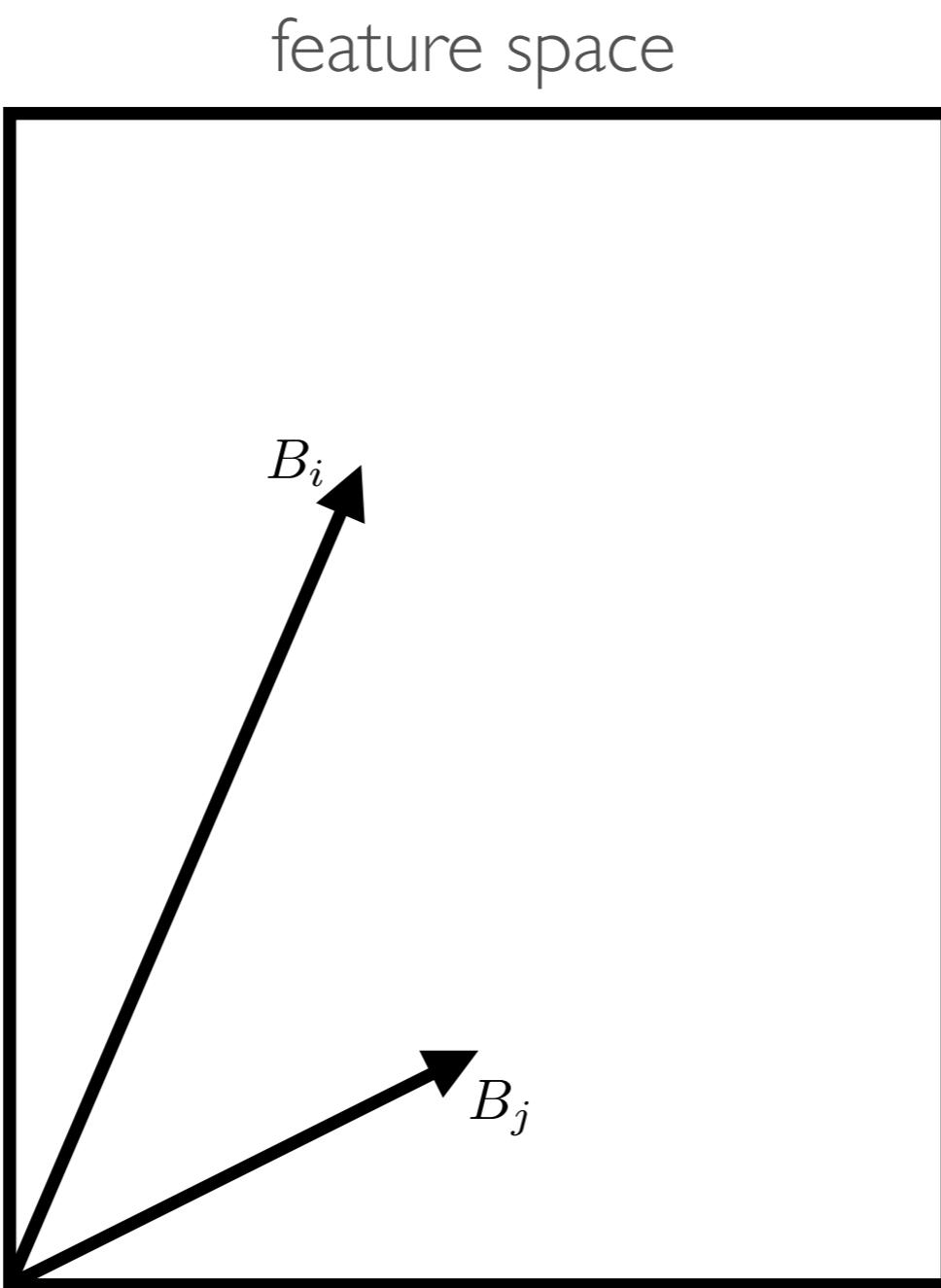
AREA AS SET-GOODNESS

AREA AS SET-GOODNESS

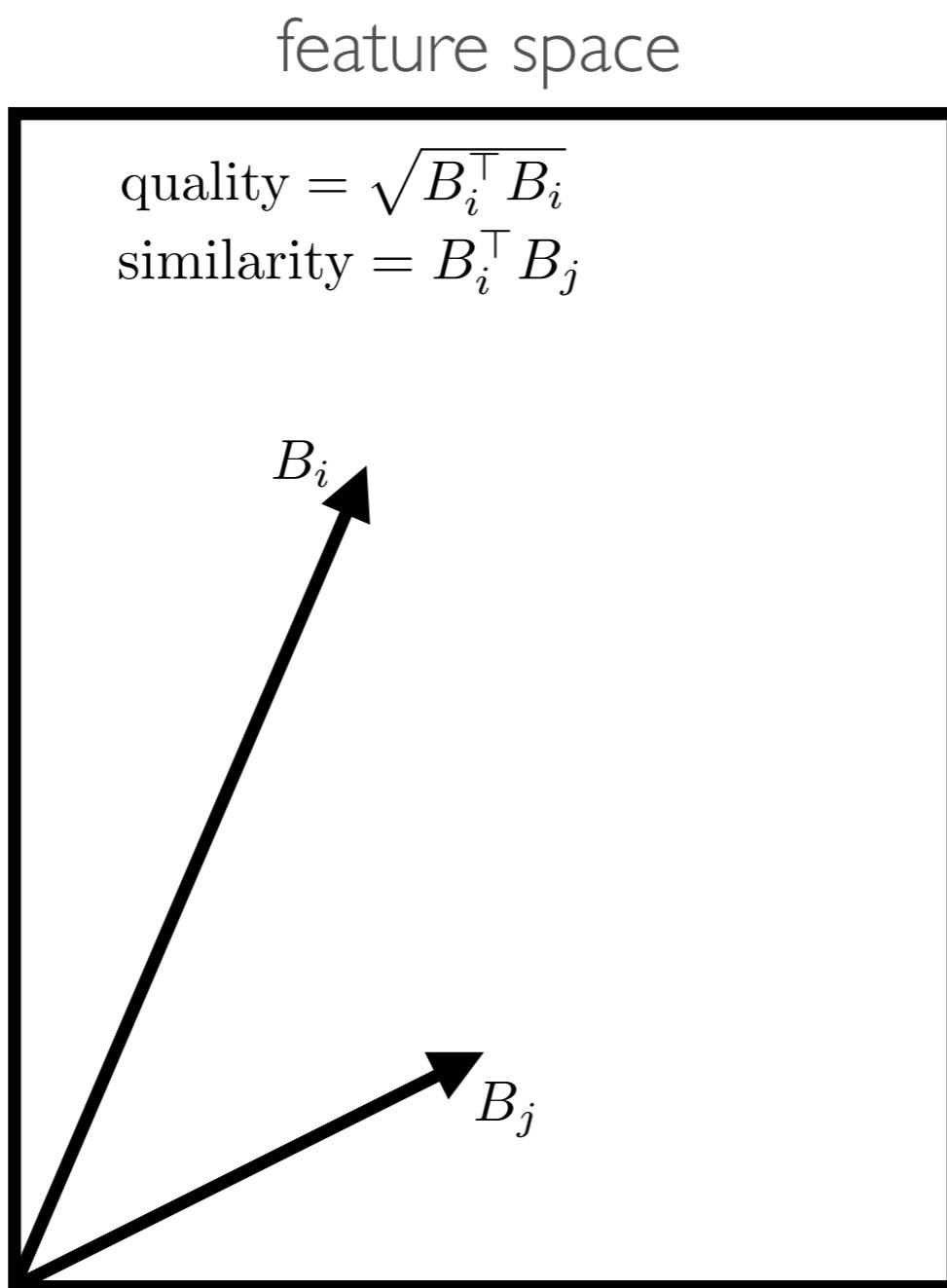
feature space



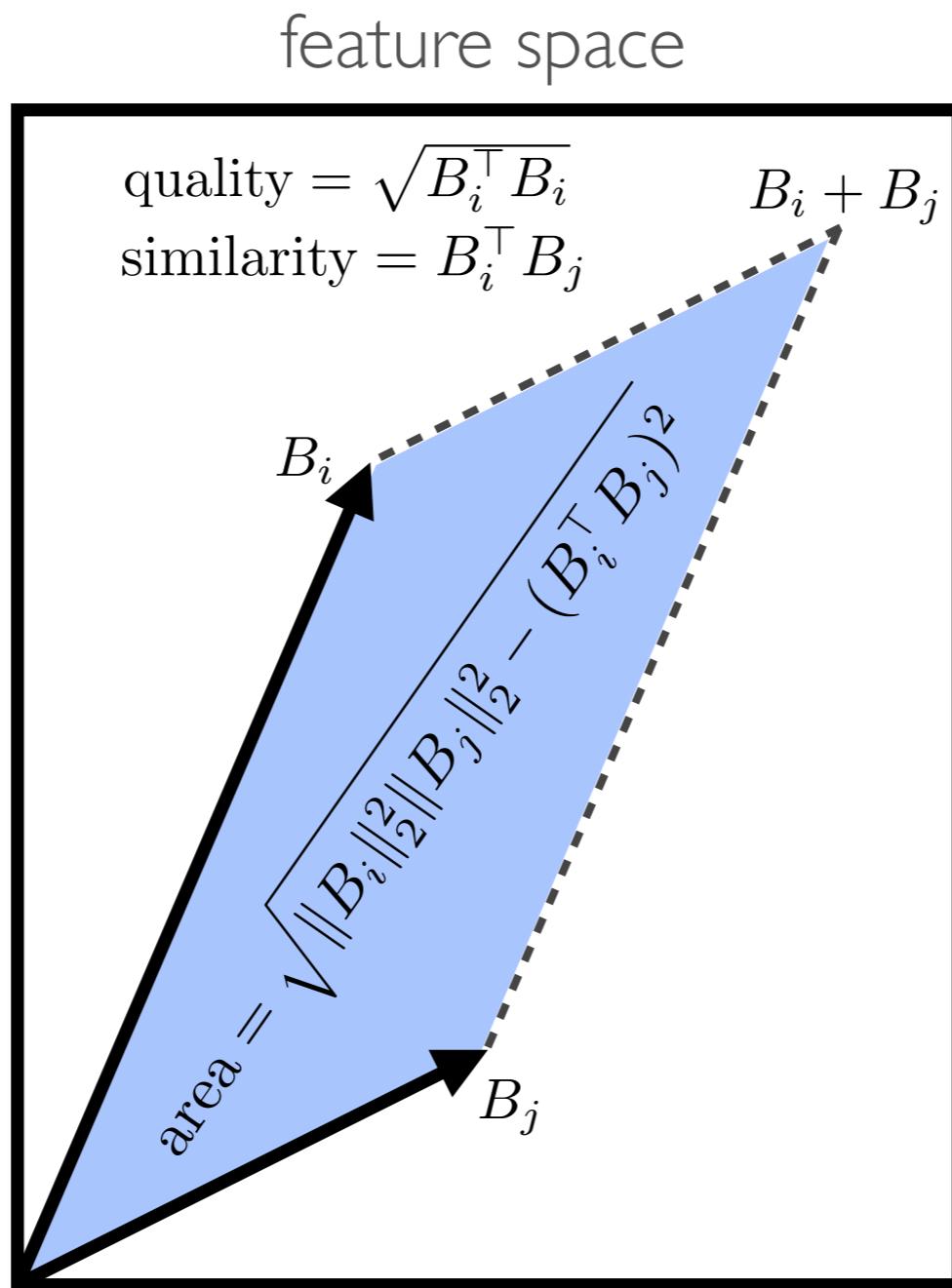
AREA AS SET-GOODNESS



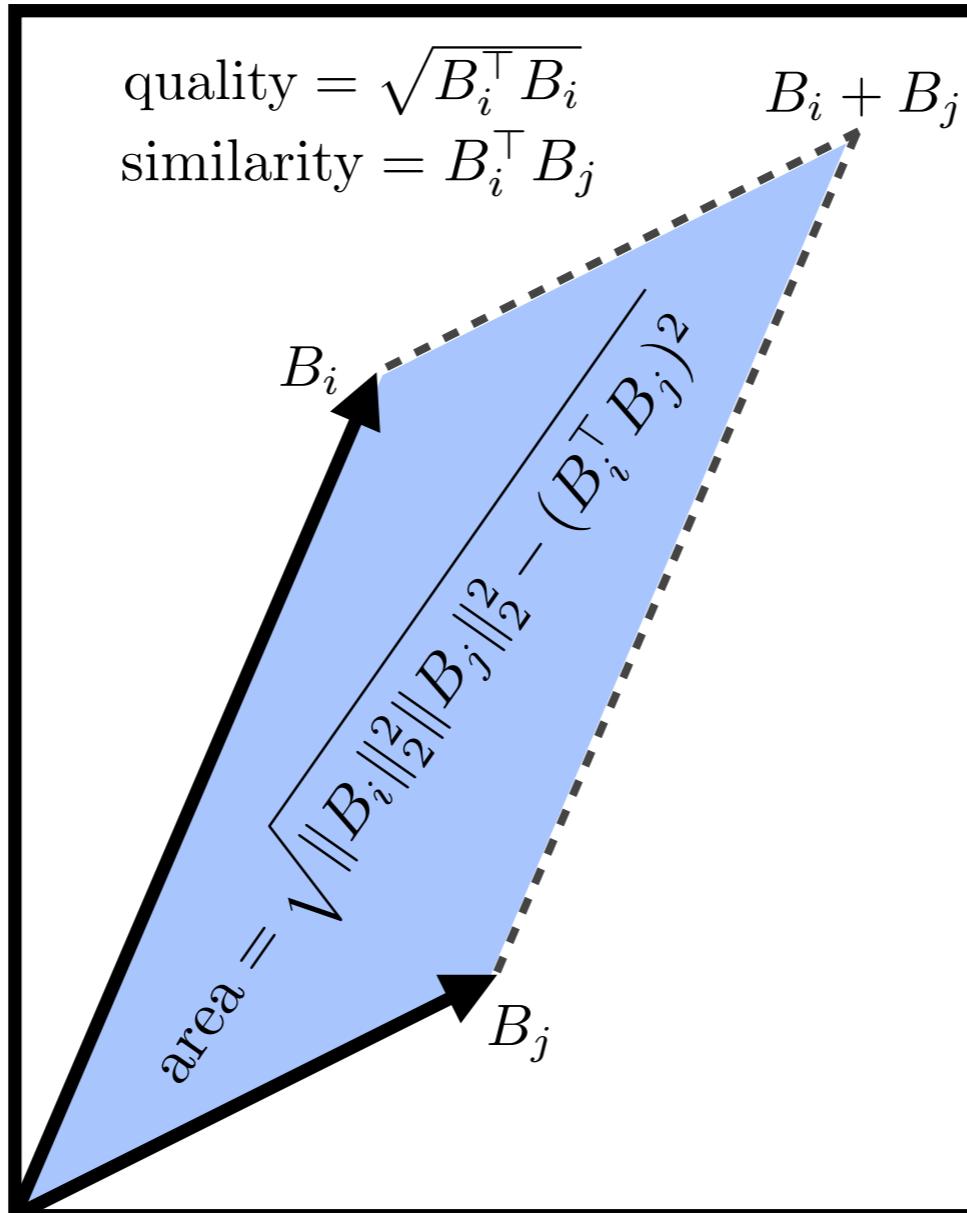
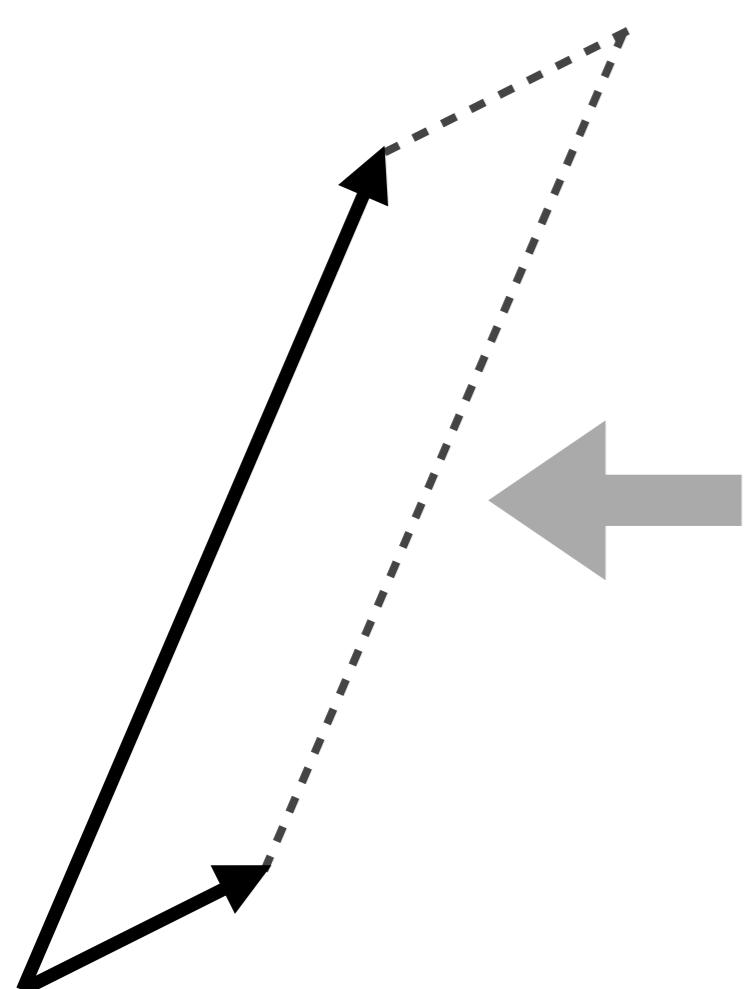
AREA AS SET-GOODNESS



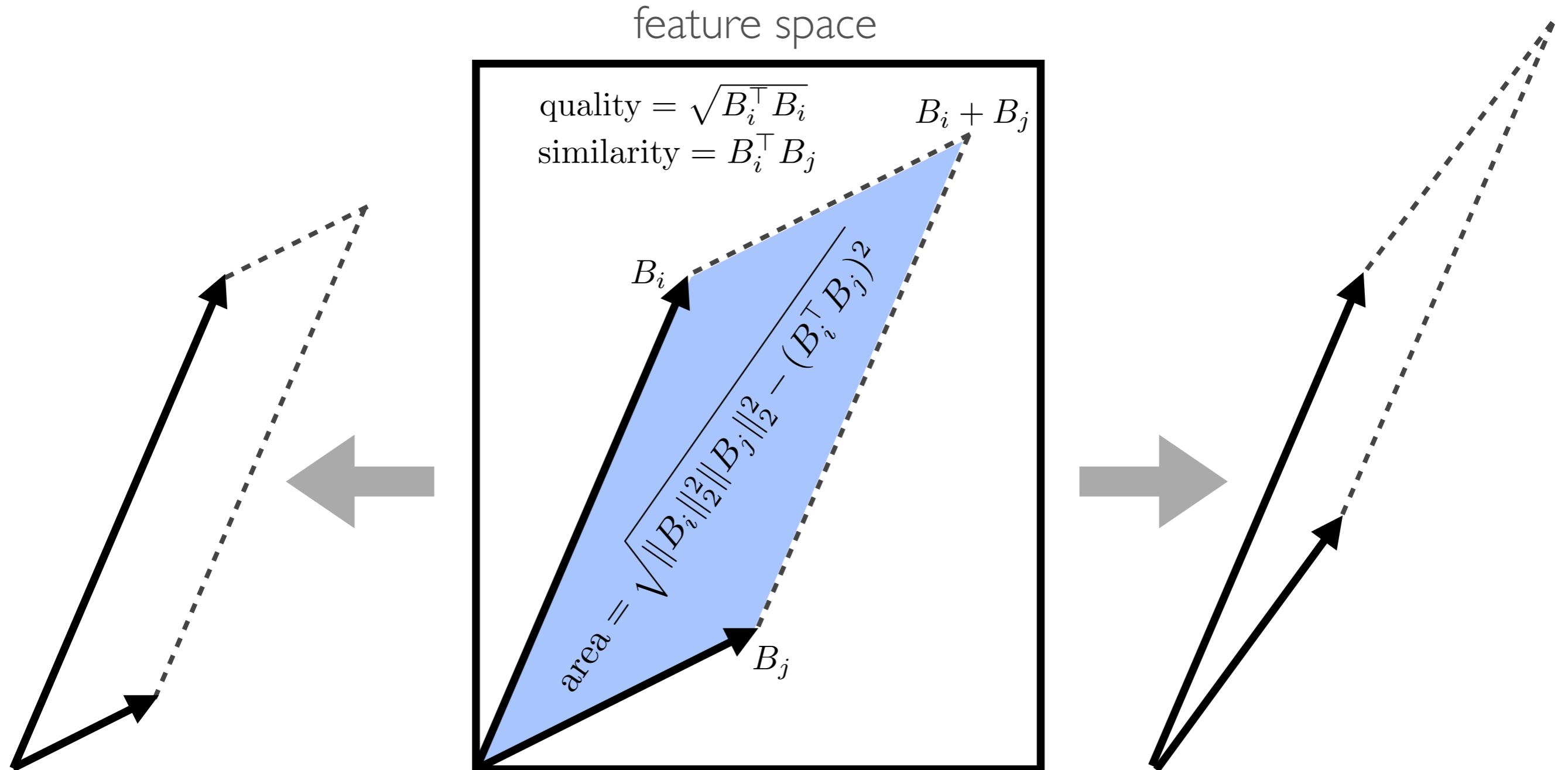
AREA AS SET-GOODNESS



AREA AS SET-GOODNESS



AREA AS SET-GOODNESS

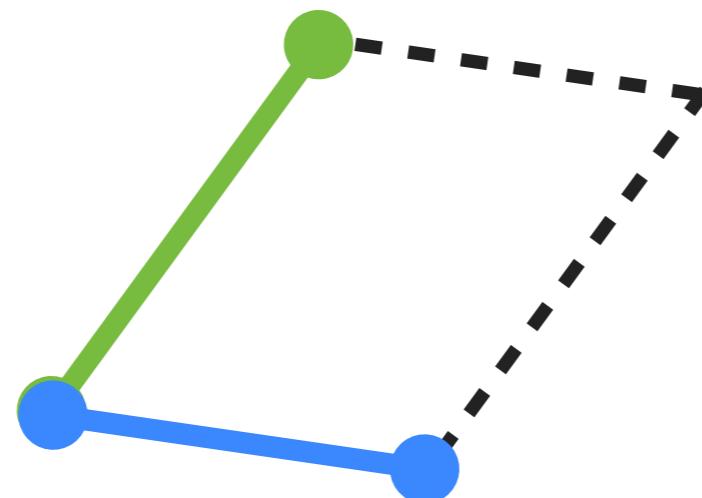


VOLUME AS SET-GOODNESS

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$

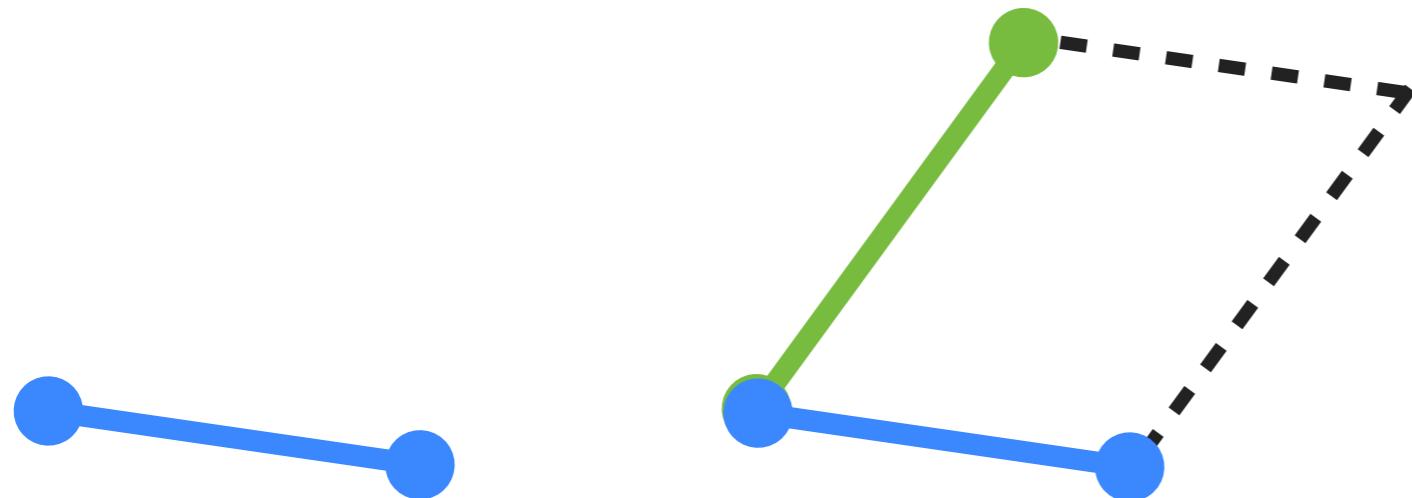
VOLUME AS SET-GOODNESS

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$



VOLUME AS SET-GOODNESS

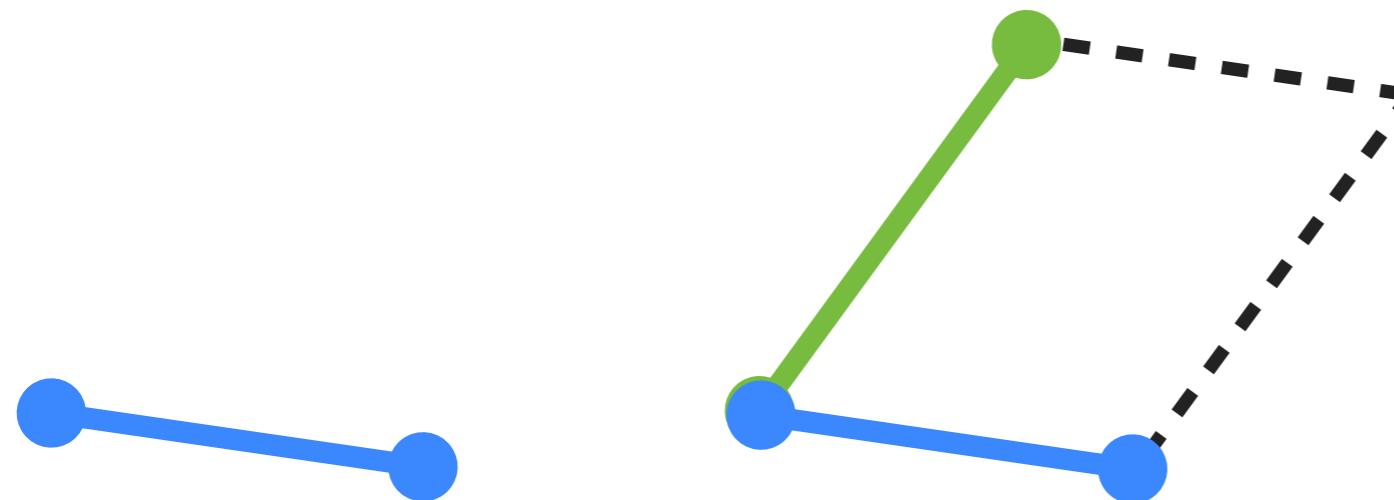
$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$



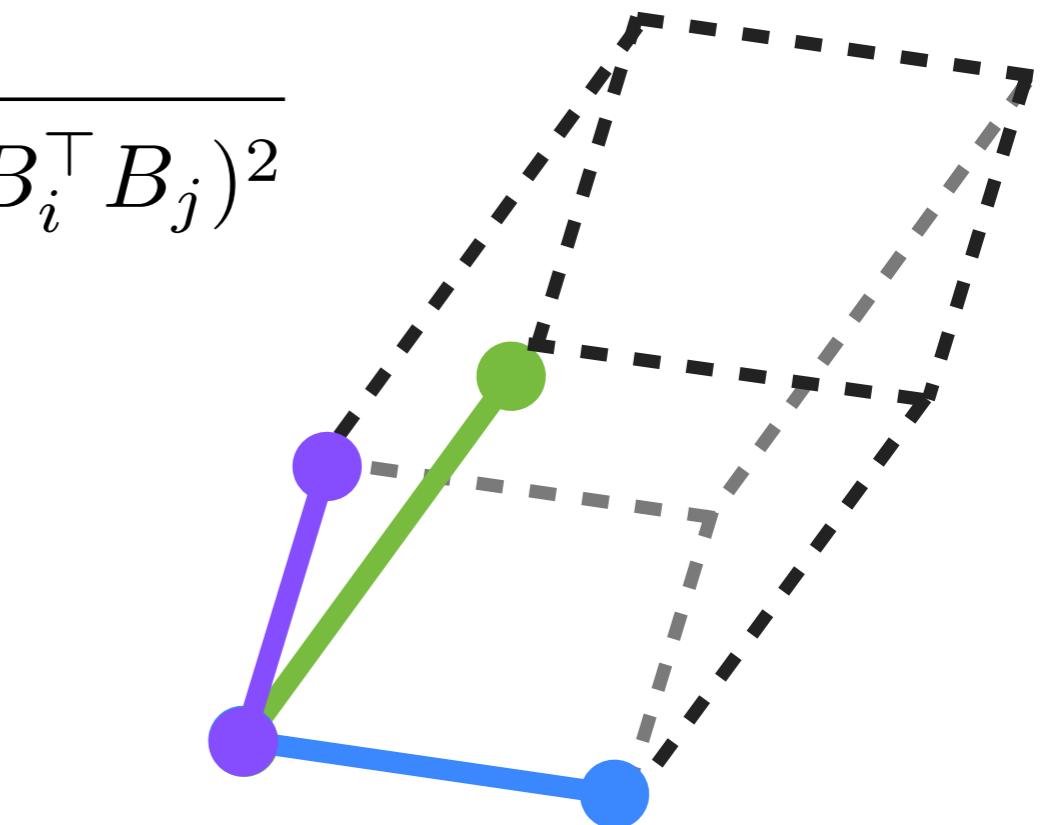
$$\text{length} = \|B_i\|_2$$

VOLUME AS SET-GOODNESS

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$



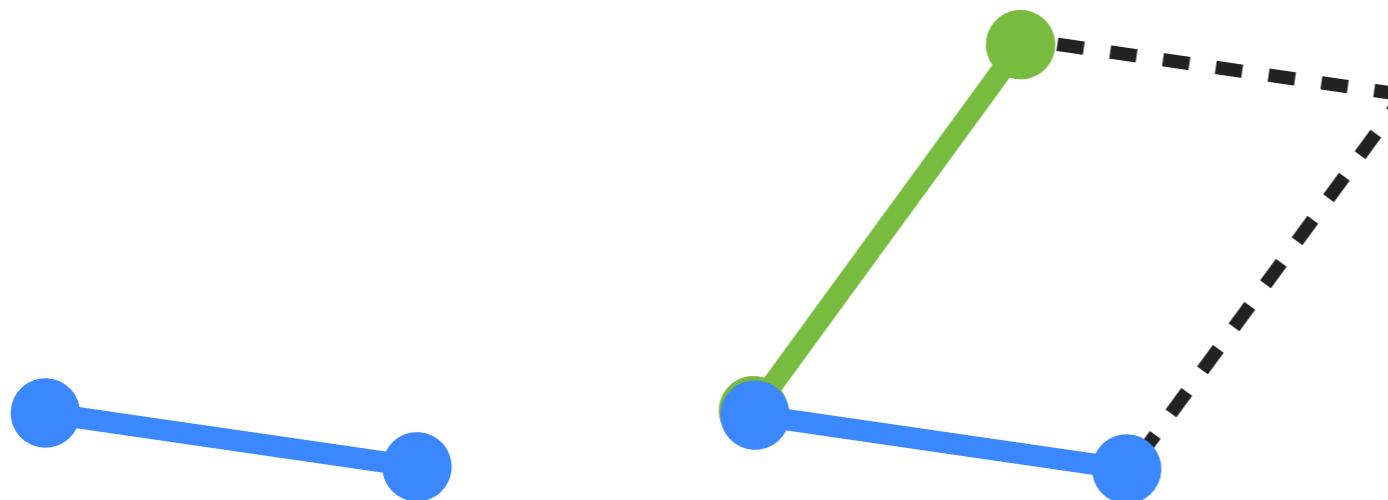
$$\text{length} = \|B_i\|_2$$



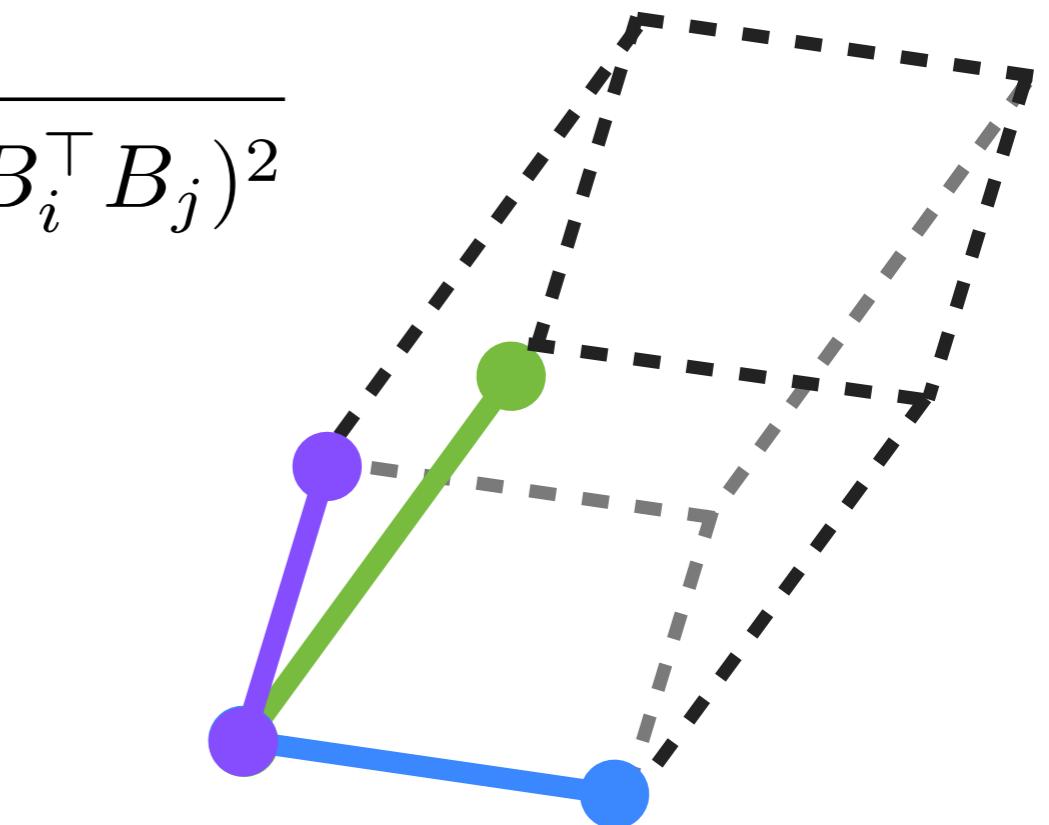
$$\text{volume} = \text{base} \times \text{height}$$

VOLUME AS SET-GOODNESS

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$



$$\text{length} = \|B_i\|_2$$



$$\text{volume} = \text{base} \times \text{height}$$

$$\begin{aligned}\text{vol}(B) &= \text{height} \times \text{base} \\ &= \|B_1\|_2 \text{vol}(\text{proj}_{\perp B_1}(B_{2:N}))\end{aligned}$$

AREA AS A DET

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$

AREA AS A DET

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$

$$= \det \begin{pmatrix} \|B_i\|_2^2 & B_i^\top B_j \\ B_i^\top B_j & \|B_j\|_2^2 \end{pmatrix}^{\frac{1}{2}}$$

AREA AS A DET

$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$

$$= \det \begin{pmatrix} \|B_i\|_2^2 & B_i^\top B_j \\ B_i^\top B_j & \|B_j\|_2^2 \end{pmatrix}^{\frac{1}{2}}$$

$$= \det \begin{pmatrix} \boxed{-B_i} & \boxed{-B_j} \\ \boxed{-B_i} & \boxed{-B_j} \end{pmatrix}^{2|1}$$

VOLUME AS A DET

$$\text{vol}(B_{\{i,j\}}) = \det \left(\begin{array}{c|c} \hline & B_i \\ \hline & B_j \\ \hline \end{array} \begin{array}{c|c} \hline & B_i \\ \hline & B_j \\ \hline \end{array} \right)^{2|1}$$

VOLUME AS A DET

$$\text{vol}(B_{\{i,j\}}) = \det \left(\begin{array}{c|c} \hline & B_i \\ \hline & B_j \\ \hline \end{array} \quad \begin{array}{c|c} \hline & B_i \\ \hline & B_j \\ \hline \end{array} \right)^{2|1}$$

$$\text{vol}(B) = \det \left(\begin{array}{c|c|c} \hline & B_1 & \\ \hline & \vdots & \\ \hline & B_N & \\ \hline \end{array} \quad \begin{array}{c|c|c} \hline & B_1 & \\ \hline & \ddots & \\ \hline & B_N & \\ \hline \end{array} \right)^{2|1}$$

$$\text{vol}(B)^2 = \det(B^\top B) = \det(L)$$

COMPLEX STATISTICS

COMPLEX STATISTICS



COMPLEX STATISTICS



COMPLEX STATISTICS

$$\sum$$



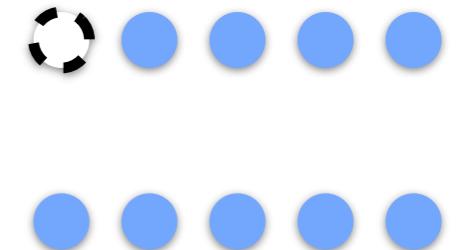
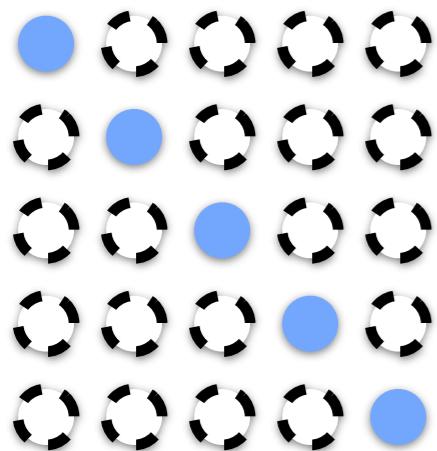
COMPLEX STATISTICS

$$\sum$$



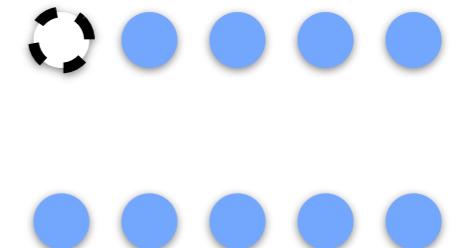
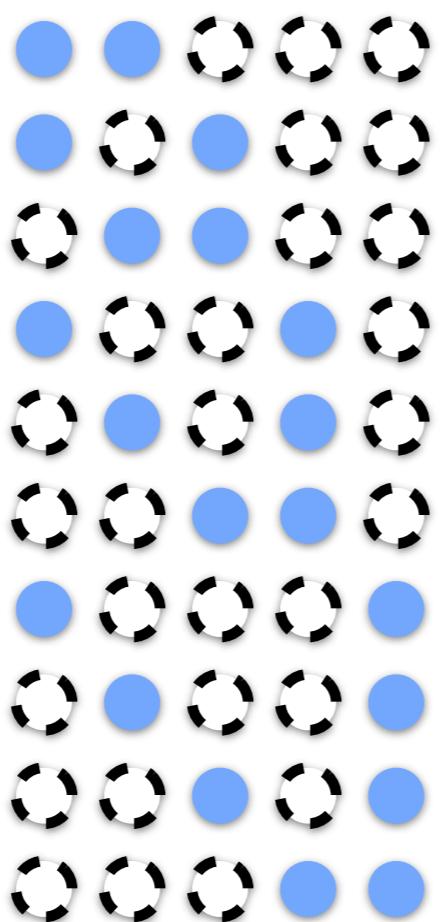
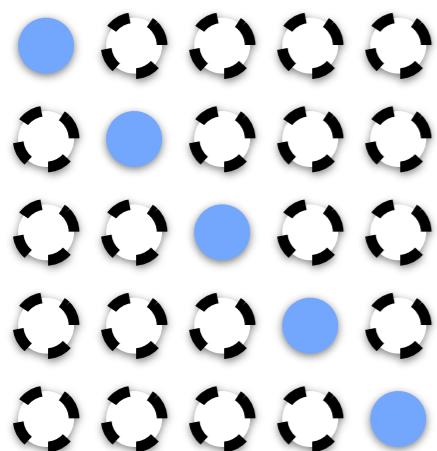
COMPLEX STATISTICS

$$\sum$$



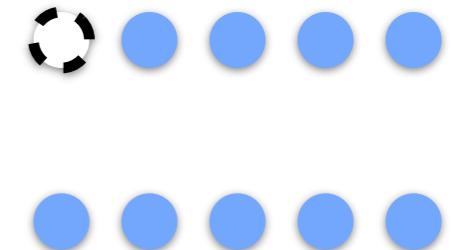
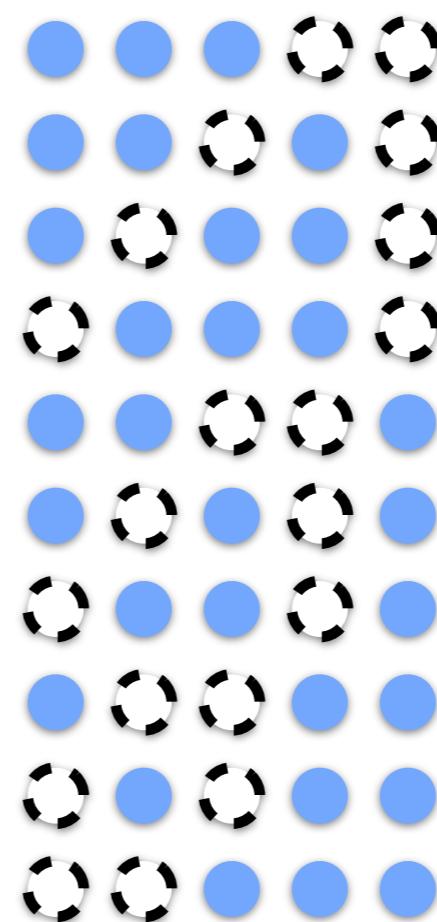
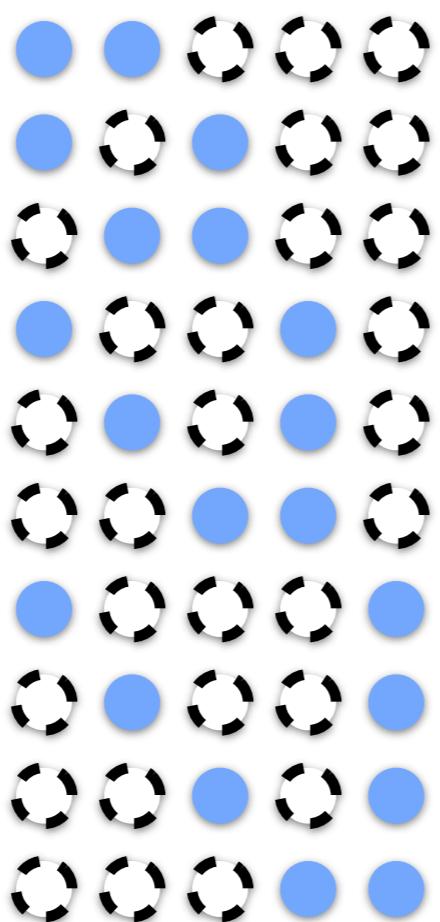
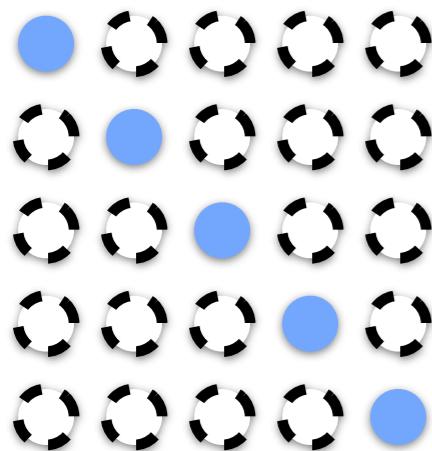
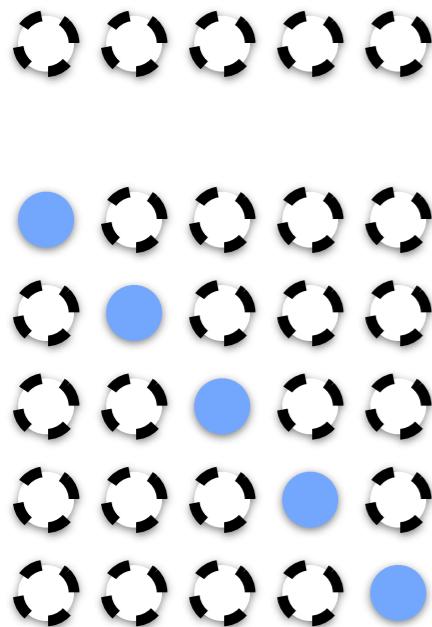
COMPLEX STATISTICS

$$\sum$$

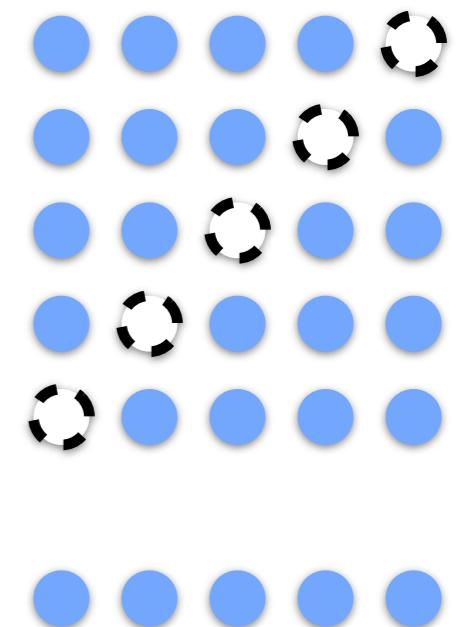
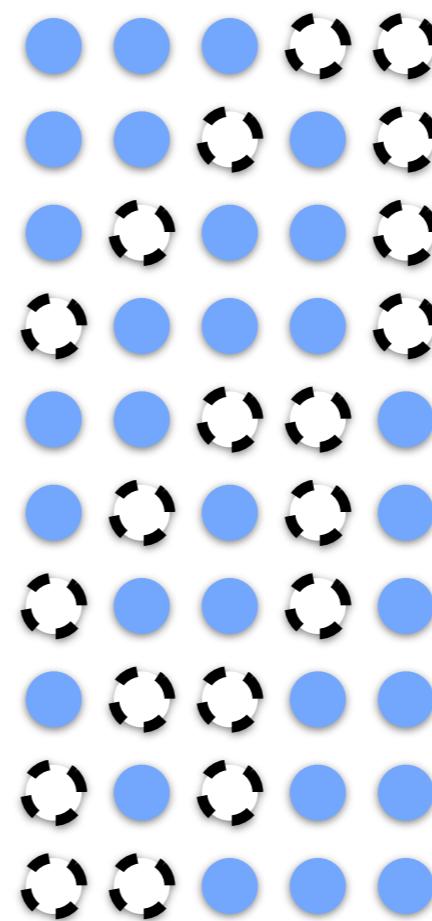
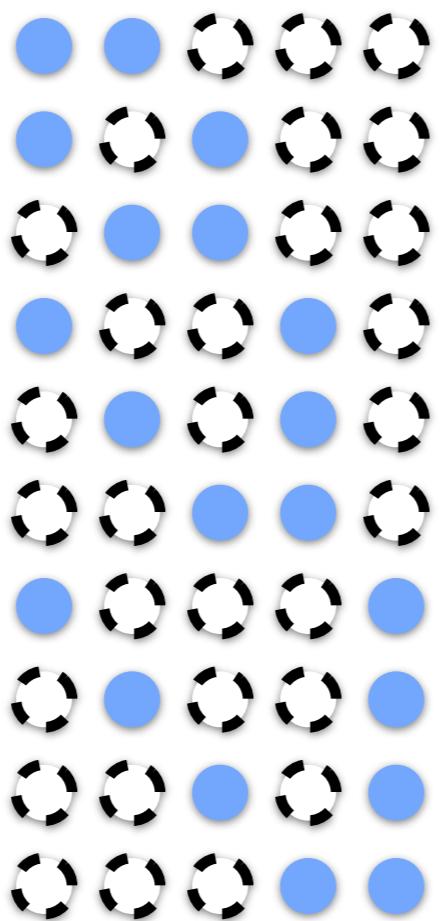
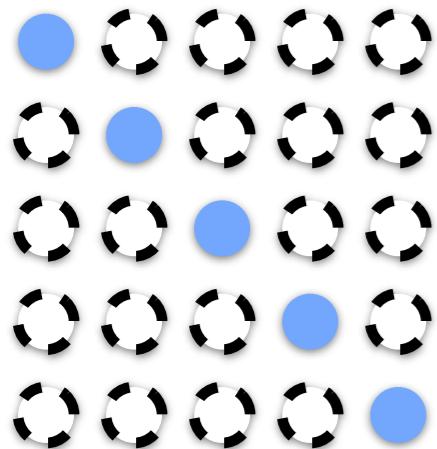
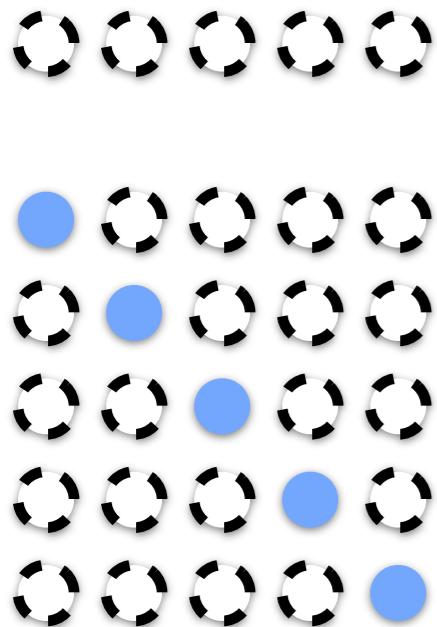


COMPLEX STATISTICS

$$\Sigma$$

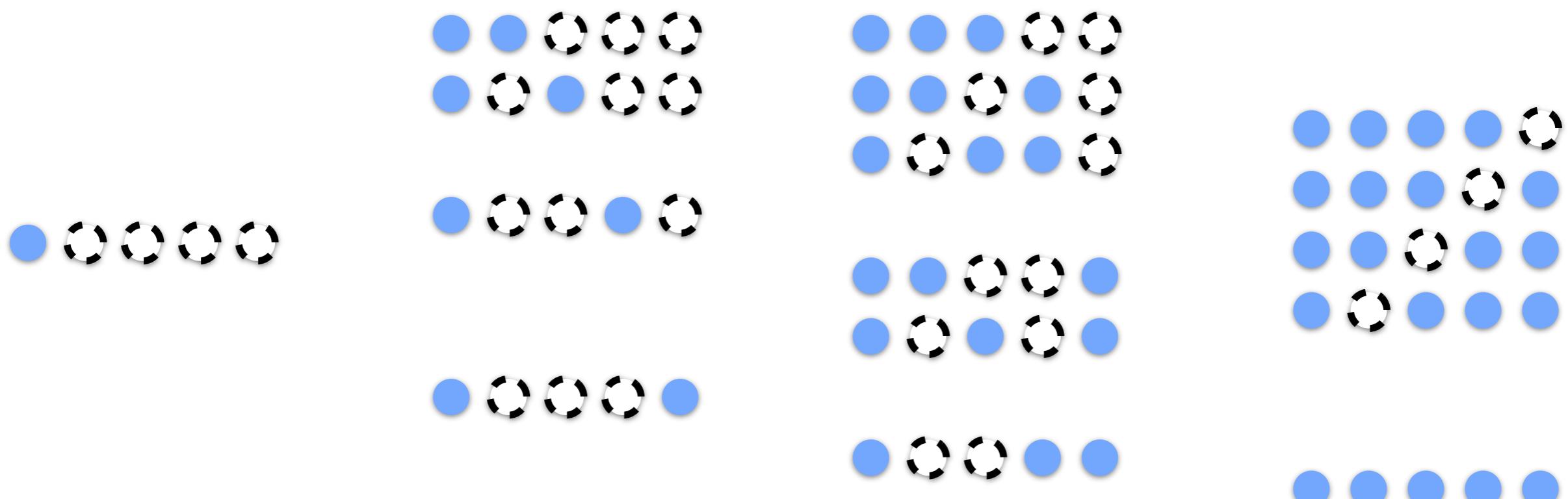


COMPLEX STATISTICS

 Σ 

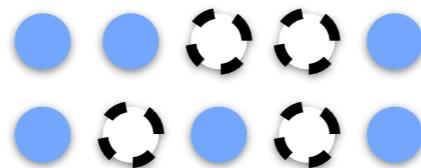
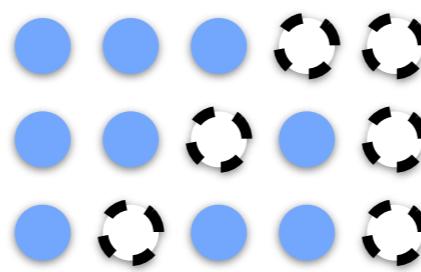
COMPLEX STATISTICS

Σ



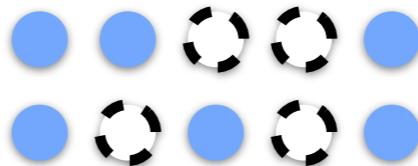
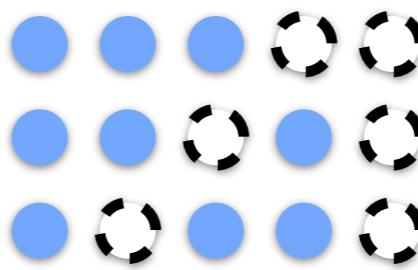
COMPLEX STATISTICS

$$\sum$$



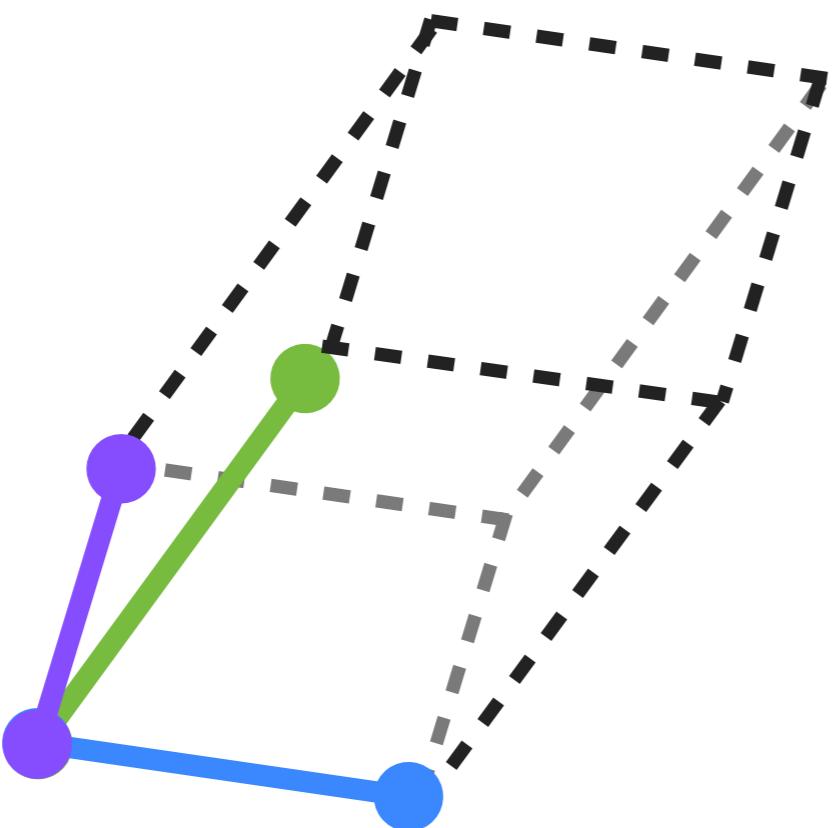
COMPLEX STATISTICS

N items $\Longrightarrow 2^N$ sets



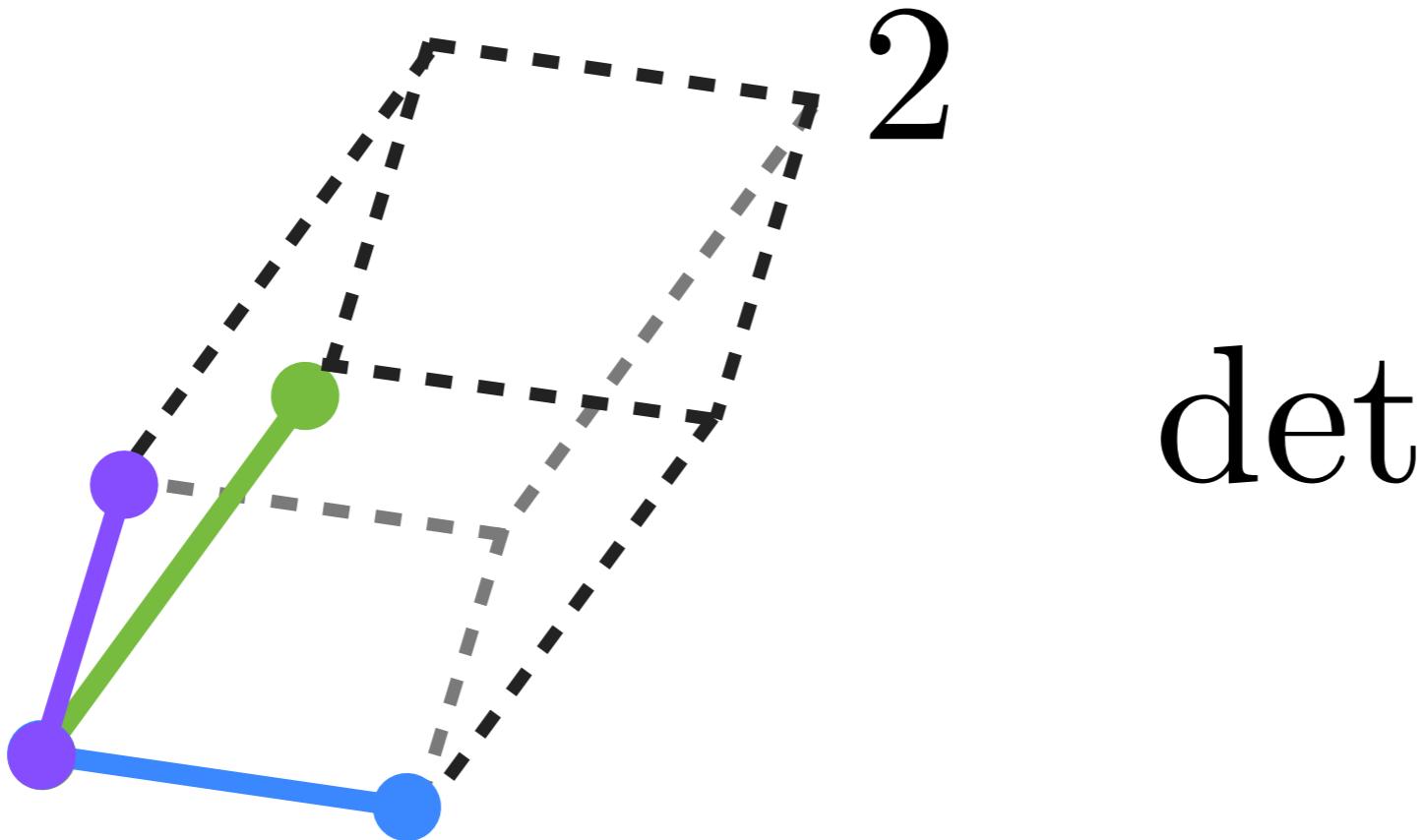
EFFICIENT COMPUTATION

EFFICIENT COMPUTATION

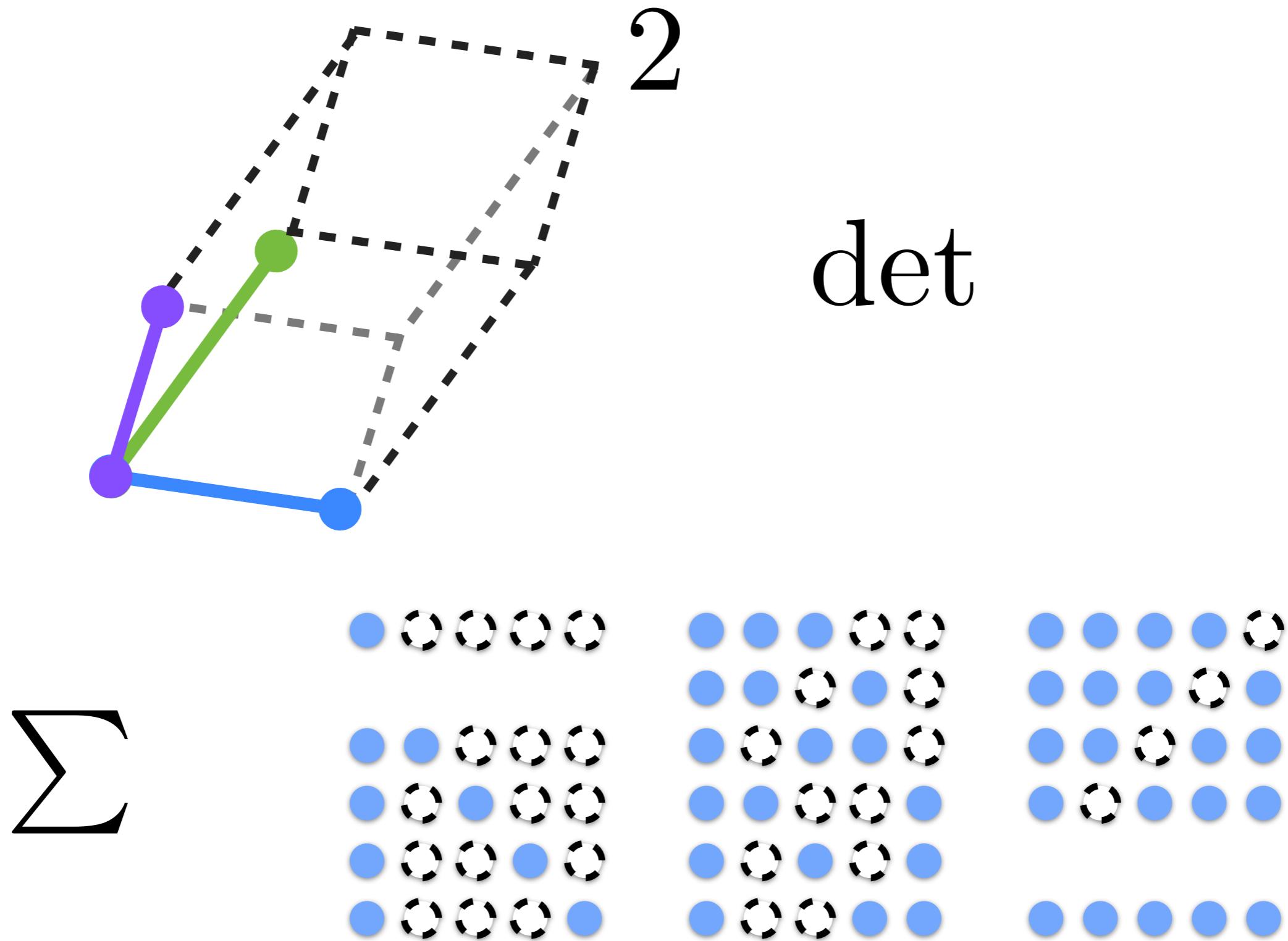


$$\det^{\frac{1}{2}}$$

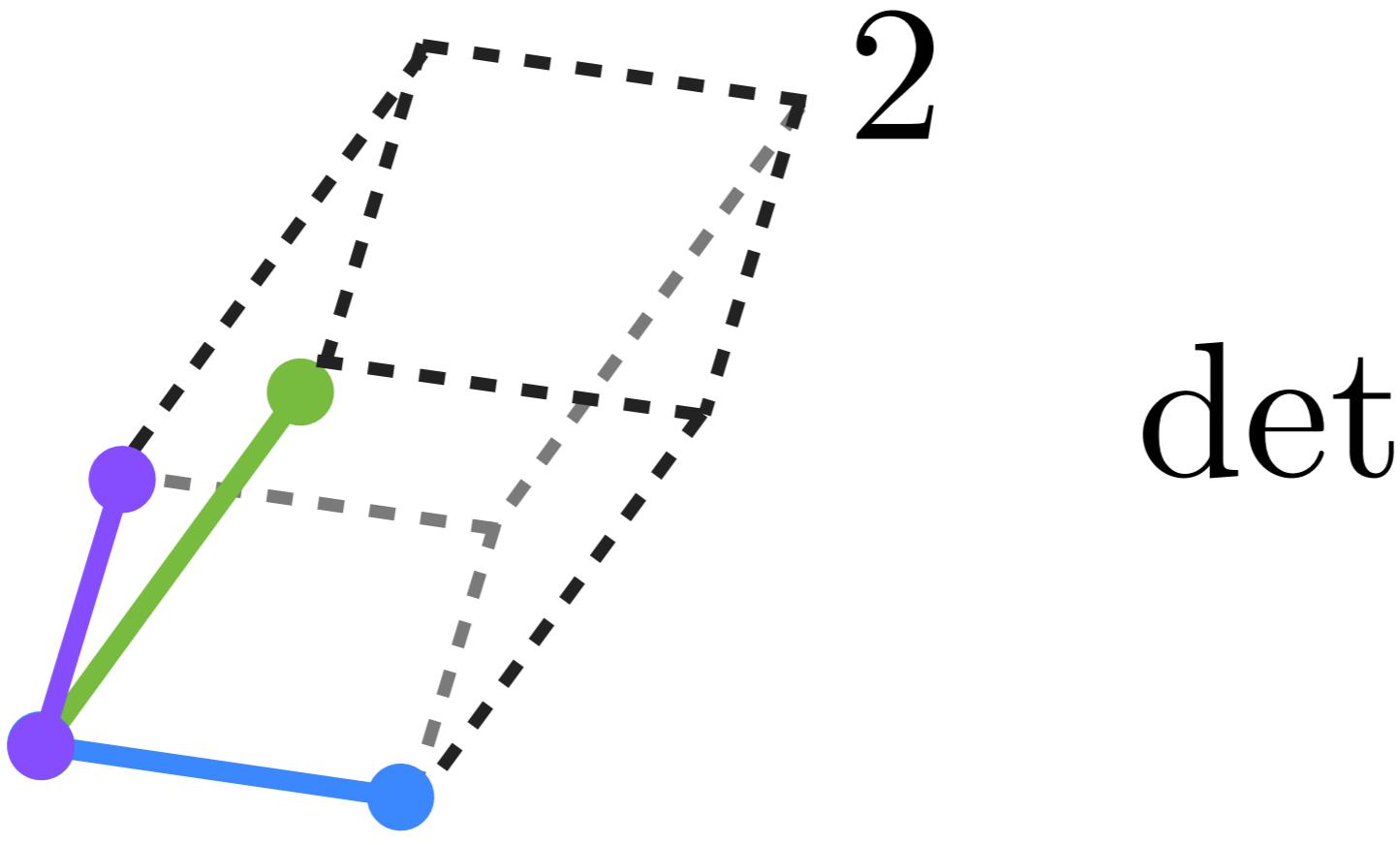
EFFICIENT COMPUTATION



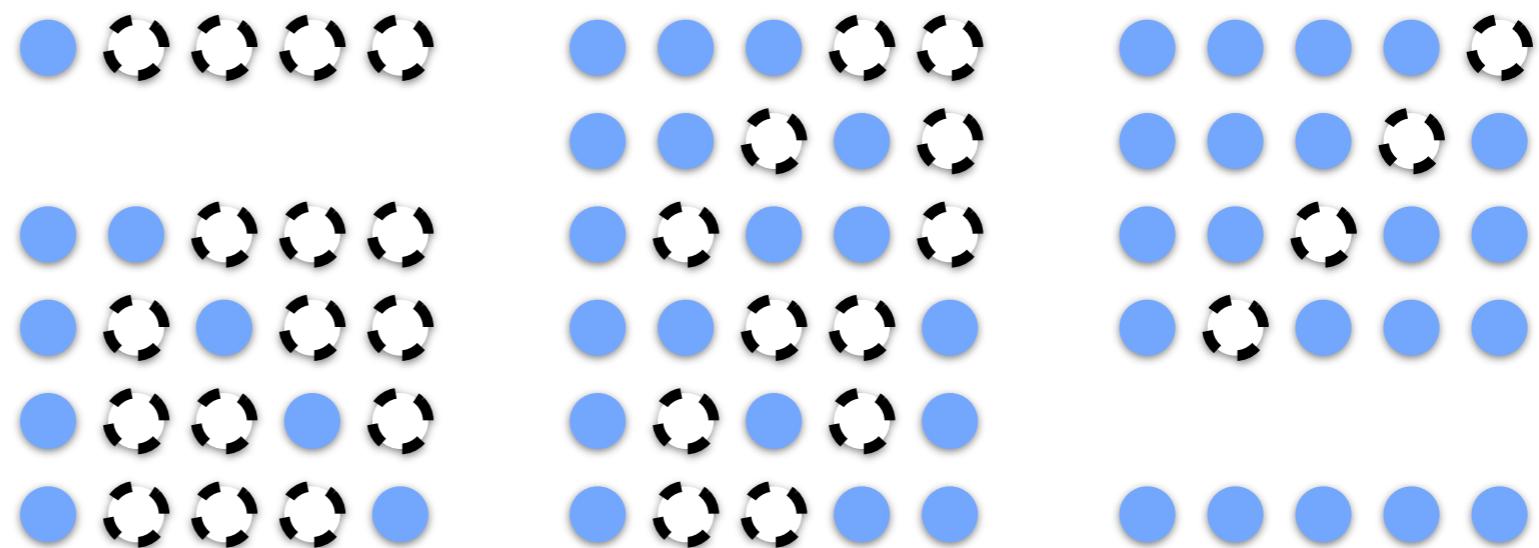
EFFICIENT COMPUTATION



EFFICIENT COMPUTATION



$O(N^3)$

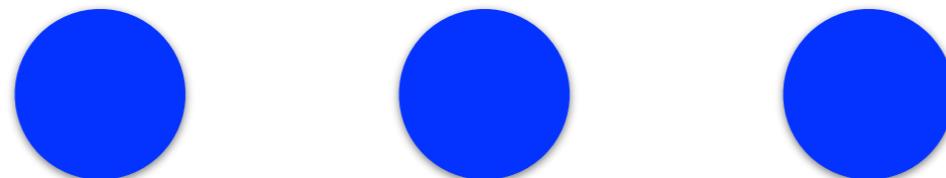
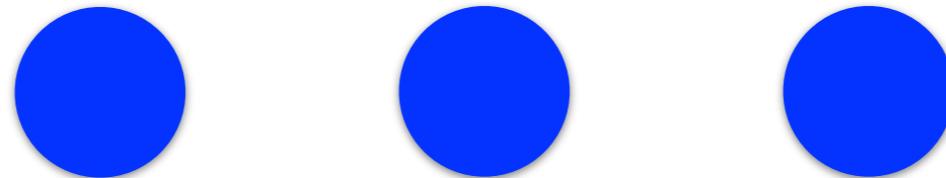
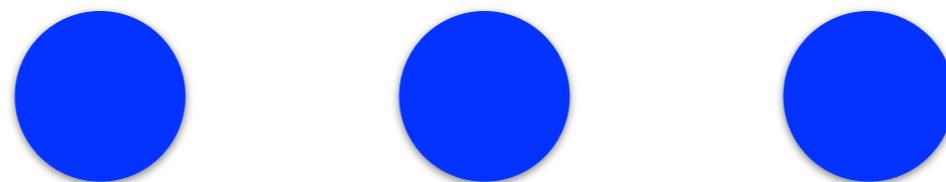


POINT PROCESSES

$$\mathcal{Y} = \{1, \dots, N\}$$

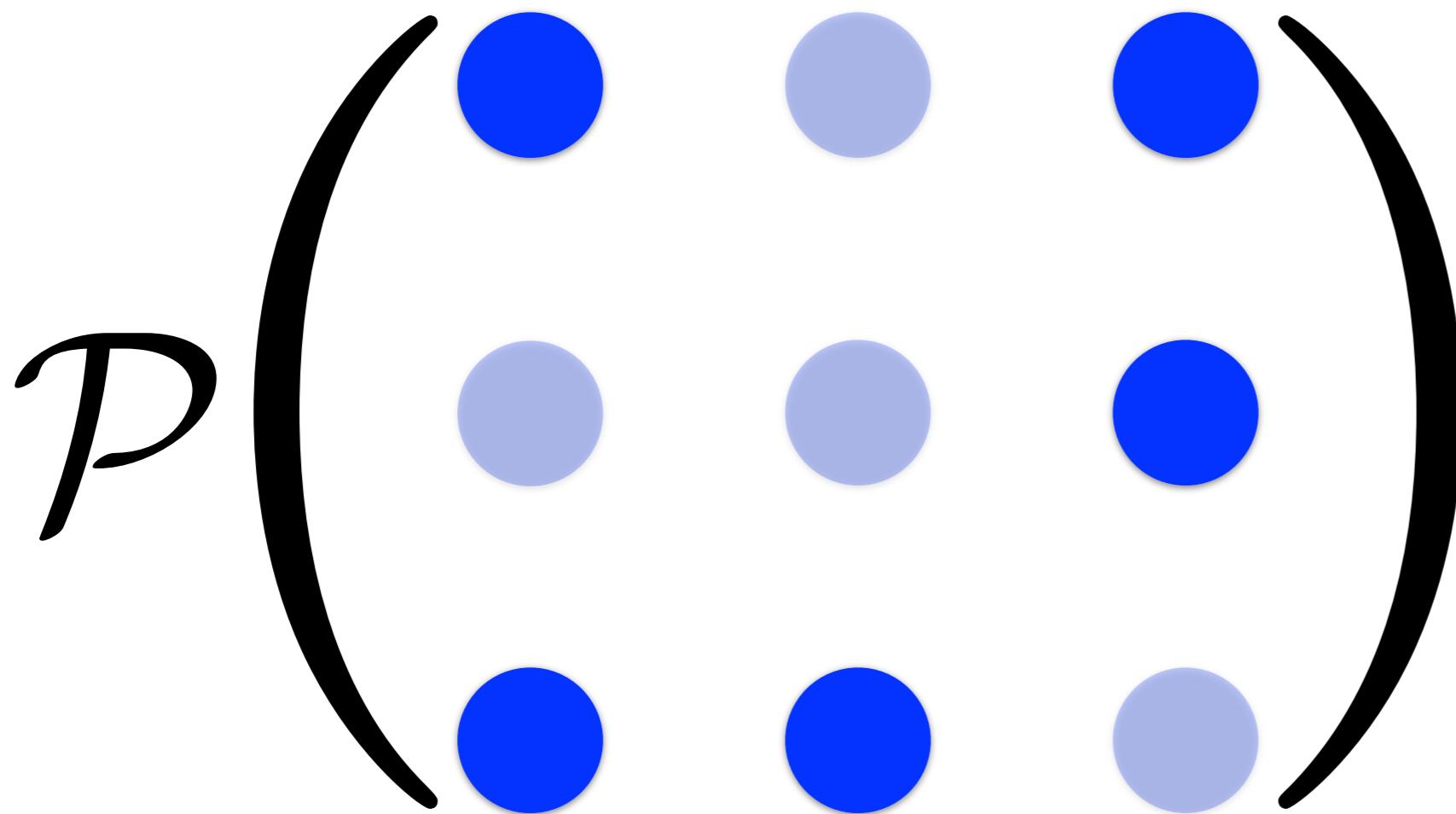
POINT PROCESSES

$$\mathcal{Y} = \{1, \dots, N\}$$



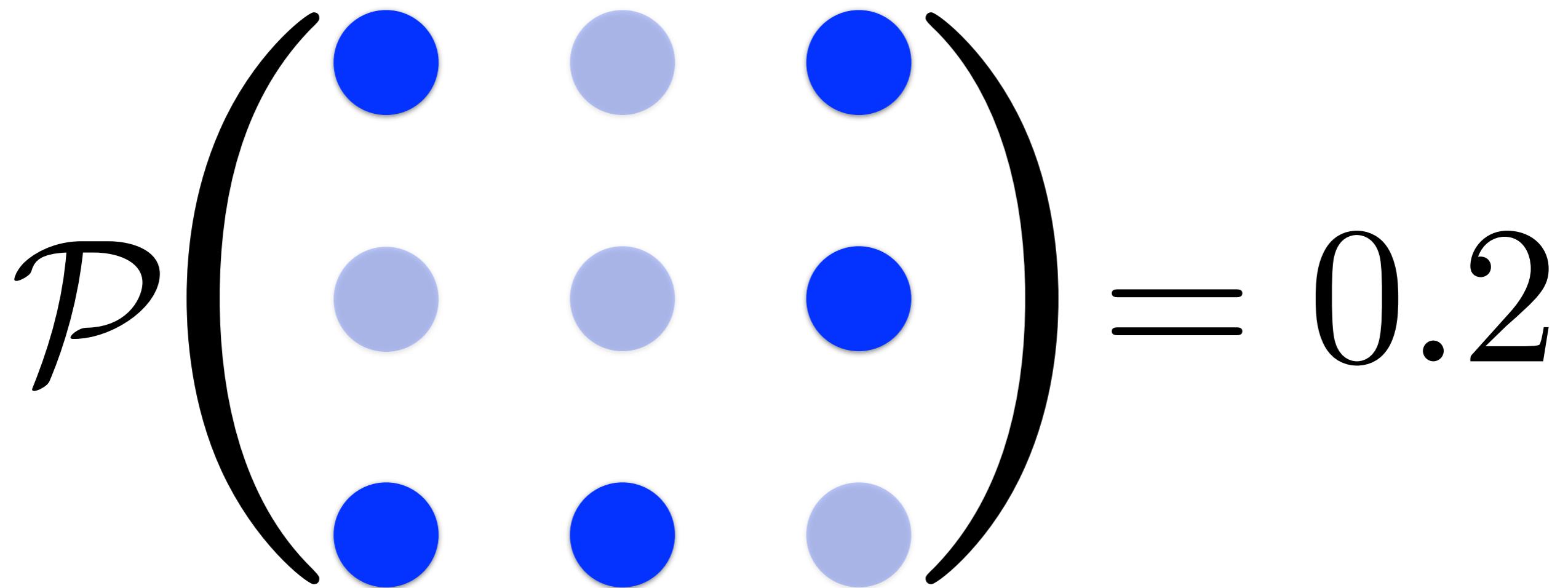
POINT PROCESSES

$$\mathcal{Y} = \{1, \dots, N\}$$



POINT PROCESSES

$$\mathcal{Y} = \{1, \dots, N\}$$



DETERMINANTAL

DETERMINANTAL

$$\mathcal{P}(\{2,3,5\}) \propto$$

DETERMINANTAL

$$\mathcal{P}(\{2,3,5\}) \propto$$

$$L_{11}\;\;\; L_{12}\;\;\; L_{13}\;\;\; L_{14}\;\;\; L_{15}$$

$$L_{21}\;\;\; L_{22}\;\;\; L_{23}\;\;\; L_{24}\;\;\; L_{25}$$

$$L_{31}\;\;\; L_{32}\;\;\; L_{33}\;\;\; L_{34}\;\;\; L_{35}$$

$$L_{41}\;\;\; L_{42}\;\;\; L_{43}\;\;\; L_{44}\;\;\; L_{45}$$

$$L_{51}\;\;\; L_{52}\;\;\; L_{53}\;\;\; L_{54}\;\;\; L_{55}$$

DETERMINANTAL

$$\mathcal{P}(\{2, 3, 5\}) \propto$$

L_{11}	L_{12}	L_{13}	L_{14}	L_{15}
L_{21}	L_{22}	L_{23}	L_{24}	L_{25}
L_{31}	L_{32}	L_{33}	L_{34}	L_{35}
L_{41}	L_{42}	L_{43}	L_{44}	L_{45}
L_{51}	L_{52}	L_{53}	L_{54}	L_{55}

DETERMINANTAL

$$L_{22} \quad L_{23} \quad L_{25}$$

$$\mathcal{P}(\{2,3,5\}) \propto \qquad \qquad L_{32} \quad L_{33} \quad L_{35}$$

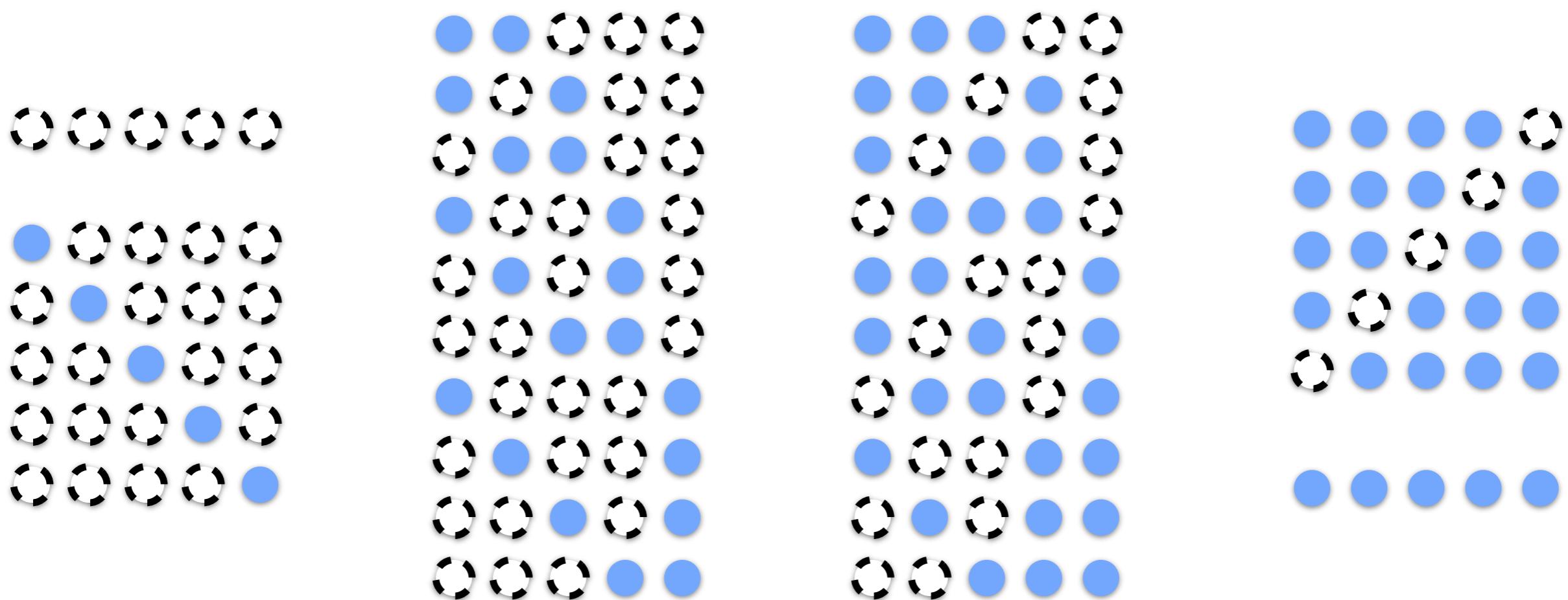
$$L_{52} \quad L_{53} \quad L_{55}$$

DETERMINANTAL

$$\mathcal{P}(\{2, 3, 5\}) \propto \det \begin{pmatrix} L_{22} & L_{23} & L_{25} \\ L_{32} & L_{33} & L_{35} \\ L_{52} & L_{53} & L_{55} \end{pmatrix}$$

DETERMINANTAL

$$\mathcal{P}(\{2, 3, 5\}) = \det \begin{pmatrix} L_{22} & L_{23} & L_{25} \\ L_{32} & L_{33} & L_{35} \\ L_{52} & L_{53} & L_{55} \end{pmatrix}$$



DETERMINANTAL

$$\mathcal{P}(\{2, 3, 5\}) = \det \begin{pmatrix} L_{22} & L_{23} & L_{25} \\ L_{32} & L_{33} & L_{35} \\ L_{52} & L_{53} & L_{55} \end{pmatrix} \frac{\det(L + I)}{\det(L + I)}$$

EFFICIENT INFERENCE

EFFICIENT INFERENCE

Normalizing: $\mathcal{P}_L(\mathbf{Y} = Y)$

EFFICIENT INFERENCE

Normalizing: $\mathcal{P}_L(\mathbf{Y} = Y)$

Marginalizing: $\mathcal{P}(Y \subseteq \mathbf{Y})$

EFFICIENT INFERENCE

Normalizing: $\mathcal{P}_L(\mathbf{Y} = Y)$

Marginalizing: $\mathcal{P}(Y \subseteq \mathbf{Y})$

Conditioning: $\mathcal{P}_L(\mathbf{Y} = B \mid A \subseteq \mathbf{Y})$

$\mathcal{P}_L(\mathbf{Y} = B \mid A \cap \mathbf{Y} = \emptyset)$

EFFICIENT INFERENCE

Normalizing: $\mathcal{P}_L(\mathbf{Y} = Y)$

Marginalizing: $\mathcal{P}(Y \subseteq \mathbf{Y})$

Conditioning: $\mathcal{P}_L(\mathbf{Y} = B \mid A \subseteq \mathbf{Y})$

$\mathcal{P}_L(\mathbf{Y} = B \mid A \cap \mathbf{Y} = \emptyset)$

Sampling: $\mathbf{Y} \sim \mathcal{P}_L$

EFFICIENT INFERENCE

Normalizing: $\mathcal{P}_L(\mathbf{Y} = Y)$

Marginalizing: $\mathcal{P}(Y \subseteq \mathbf{Y})$

Conditioning: $\mathcal{P}_L(\mathbf{Y} = B \mid A \subseteq \mathbf{Y})$

$\mathcal{P}_L(\mathbf{Y} = B \mid A \cap \mathbf{Y} = \emptyset)$

Sampling: $\mathbf{Y} \sim \mathcal{P}_L$

$O(N^3)$

1. DIMENSIONALITY REDUCTION

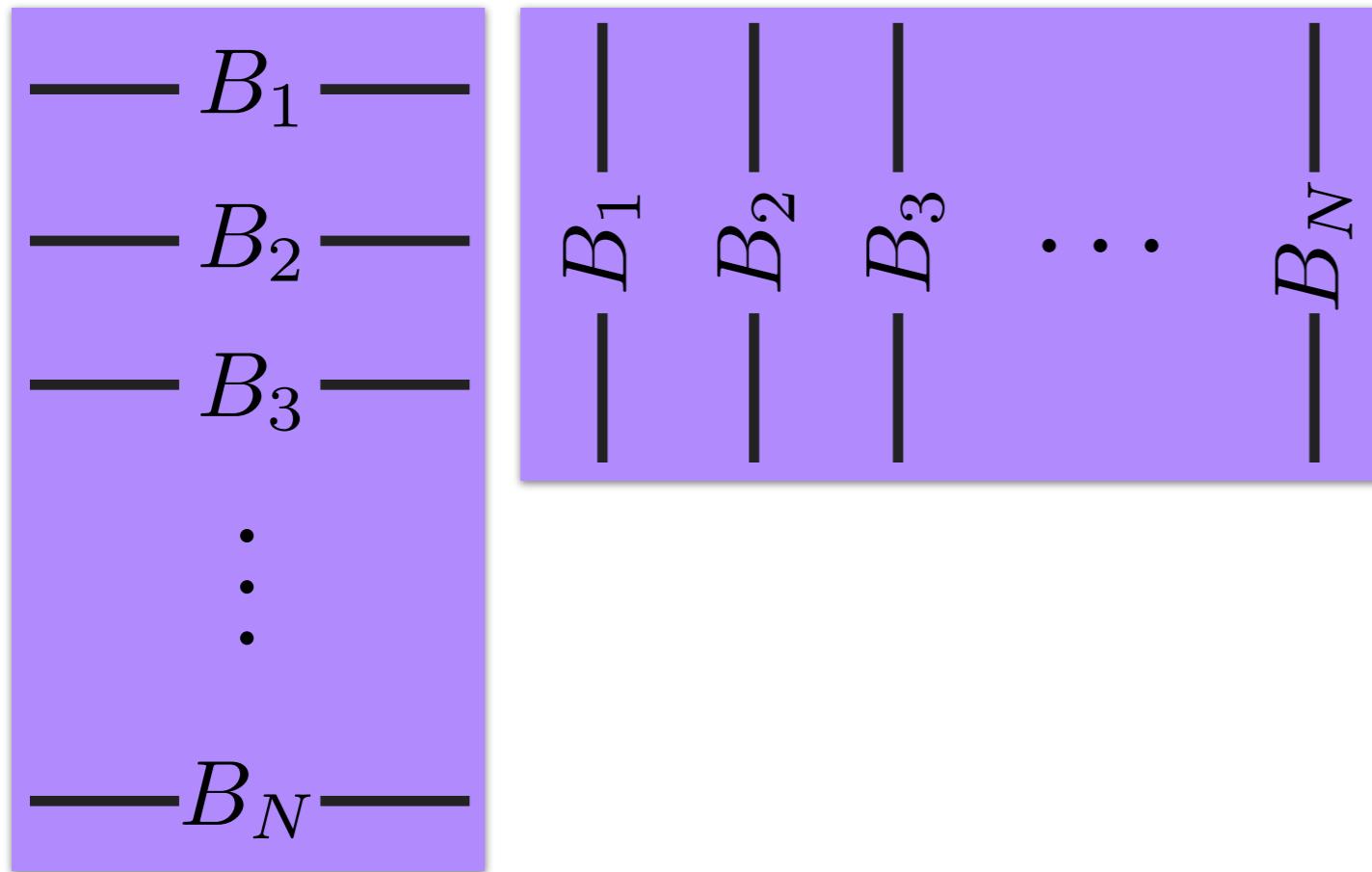
DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

L



DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

L

$$\begin{array}{c} \boxed{B_1} \\ \boxed{B_2} \\ \boxed{B_3} \\ \vdots \\ \boxed{B_N} \end{array} = \begin{array}{c} \boxed{B_1} \\ \boxed{B_2} \\ \boxed{B_3} \\ \dots \\ \boxed{B_N} \end{array} = N \times N$$

DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

C

$$\begin{array}{c} \boxed{\begin{array}{ccccccc} \hline & B_1 & & & & & \\ \hline & B_2 & & & & & \\ \hline & B_3 & & \cdots & & & \\ \hline & B_N & & & & & \end{array}} & = & \boxed{N \times N} \\ & & \end{array}$$
$$\begin{array}{c} \boxed{\begin{array}{c} \hline B_1 \hline \\ \hline B_2 \hline \\ \hline B_3 \hline \\ \vdots \\ \hline B_N \hline \end{array}} \end{array}$$

DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

C

$$\begin{array}{c} \text{---} B_1 \text{---} \\ \text{---} B_2 \text{---} \\ \text{---} B_3 \text{---} \\ \vdots \\ \text{---} B_N \text{---} \end{array} = \begin{array}{c} \text{---} B_1 \text{---} \\ \text{---} B_2 \text{---} \\ \text{---} B_3 \text{---} \\ \vdots \\ \text{---} B_N \text{---} \end{array}$$

DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

C

$$\begin{array}{c} \boxed{\begin{array}{ccccccc} \hline & B_1 & \hline & B_2 & \hline & B_3 & \hline & \cdots & \hline & B_N & \hline \end{array}} \\ \times \\ \boxed{\begin{array}{c} \hline B_1 \\ \hline B_2 \\ \hline B_3 \\ \vdots \\ \hline B_N \end{array}} \end{array} = \boxed{D \times D}$$

DUAL INFERENCE

DUAL INFERENCE

$$L = \textcolor{violet}{V} \Lambda V^\top$$

$$C = \hat{\textcolor{blue}{V}} \Lambda \hat{V}^\top$$

DUAL INFERENCE

$$L = V \Lambda V^\top \quad \longleftrightarrow \quad V = B^\top \hat{V} \Lambda^{-\frac{1}{2}} \quad C = \hat{V} \Lambda \hat{V}^\top$$

DUAL INFERENCE

$$L = V \Lambda V^\top \quad \longleftrightarrow \quad V = B^\top \hat{V} \Lambda^{-\frac{1}{2}} \quad C = \hat{V} \Lambda \hat{V}^\top$$

Normalizing $\sum_Y \det(L_Y) \quad O(\textcolor{blue}{D}^3)$

DUAL INFERENCE

$$L = V \Lambda V^\top \quad \begin{array}{c} \swarrow \\ V = B^\top \hat{V} \Lambda^{-\frac{1}{2}} \\ \searrow \end{array} \quad C = \hat{V} \Lambda \hat{V}^\top$$

Normalizing $\sum_Y \det(L_Y)$ $O(D^3)$

Marginalizing & Conditioning $O(D^3 + D^2 k^2)$

DUAL INFERENCE

$$L = V \Lambda V^\top \quad \xleftrightarrow{V = B^\top \hat{V} \Lambda^{-\frac{1}{2}}} \quad C = \hat{V} \Lambda \hat{V}^\top$$

Normalizing $\sum_Y \det(L_Y)$ $O(D^3)$

Marginalizing & Conditioning $O(D^3 + D^2 k^2)$

Sampling $\mathbf{Y} \sim \mathcal{P}_L$ $O(ND^2 k)$

EXPONENTIAL N

EXPONENTIAL N

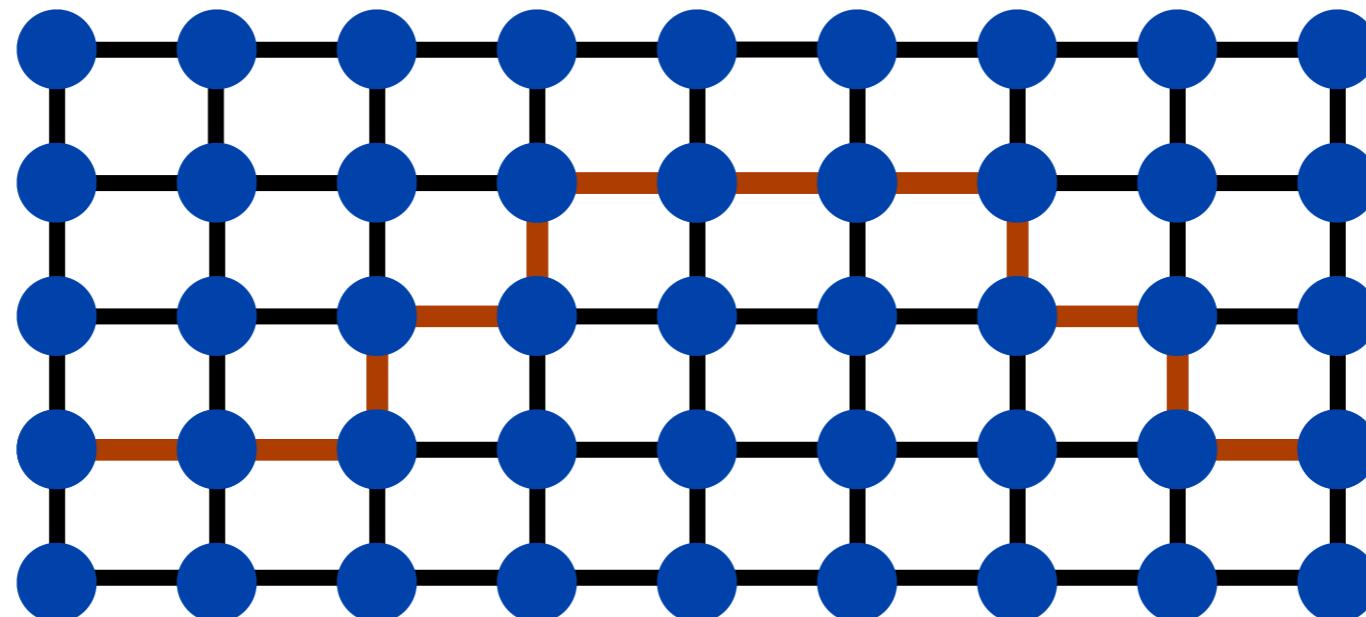
We want to select a diverse set of parses.

$$N = O(\{\text{sentence length}\}^{\{\text{sentence length}\}})$$

EXPONENTIAL N

We want to select a diverse set of parses.

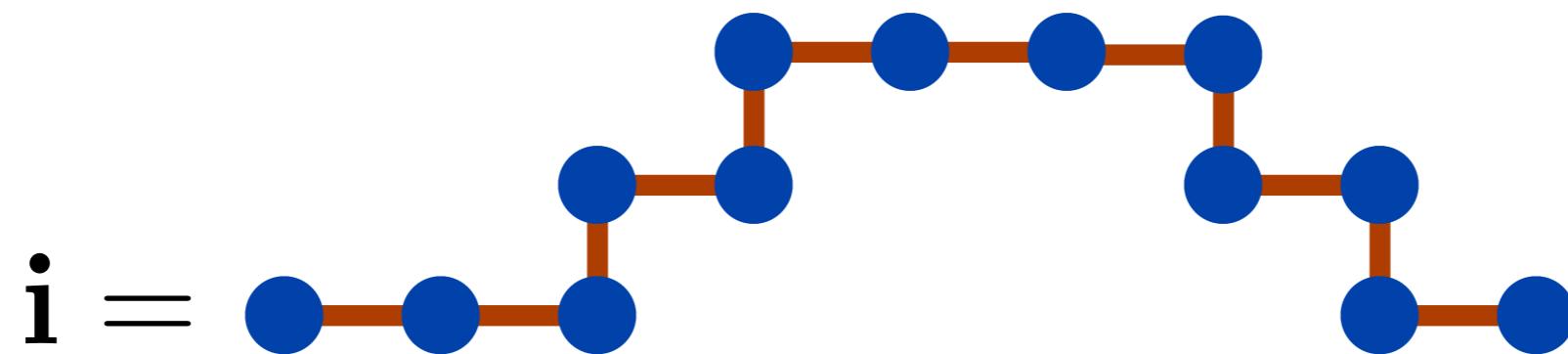
$$N = O(\{\text{sentence length}\}^{\{\text{sentence length}\}})$$



$$N = O(\{\text{node degree}\}^{\{\text{path length}\}})$$

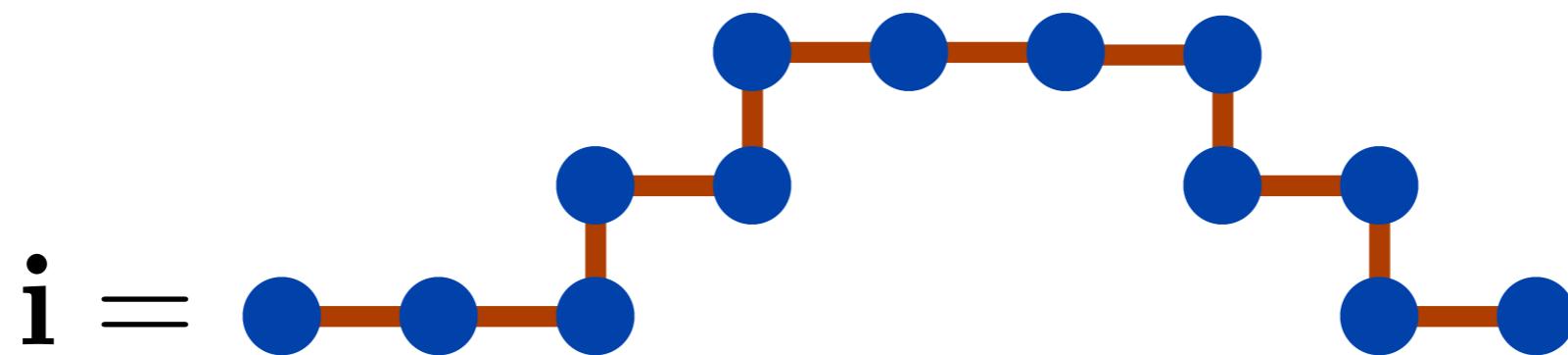
STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)



STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)



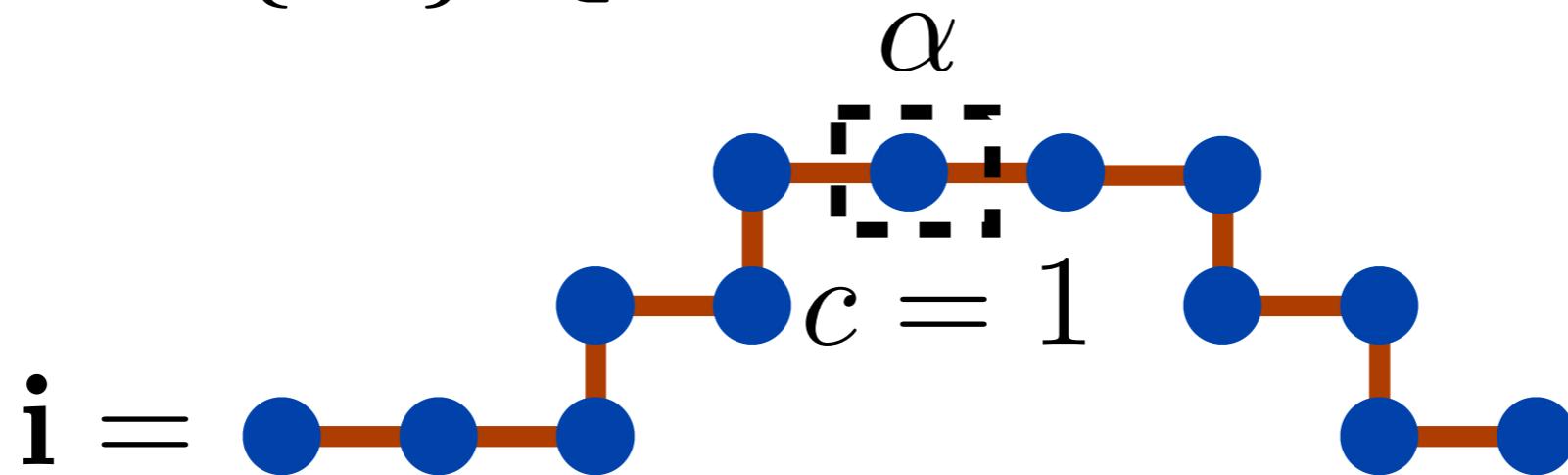
$$B_{\mathbf{i}} = q(\mathbf{i})\phi(\mathbf{i})$$

quality similarity

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$$\mathbf{i} = \{i_\alpha\}_{\alpha \in F}$$



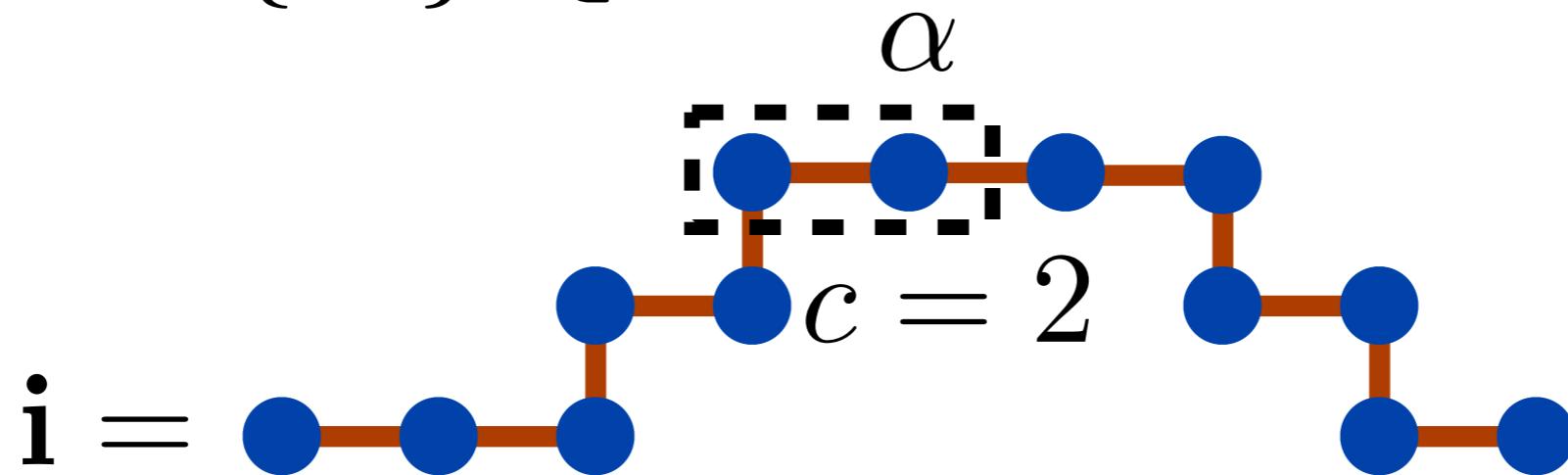
$$B_{\mathbf{i}} = q(\mathbf{i})\phi(\mathbf{i})$$

↑ ↑
quality similarity

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$$\mathbf{i} = \{i_\alpha\}_{\alpha \in F}$$



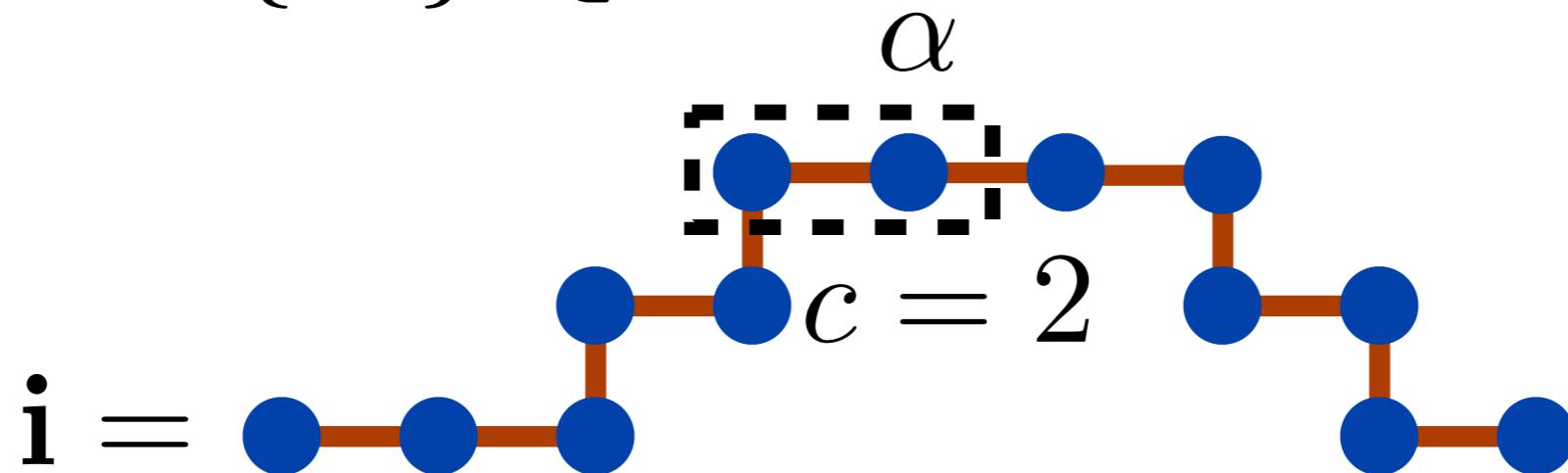
$$B_{\mathbf{i}} = q(\mathbf{i})\phi(\mathbf{i})$$

↑ ↑
quality similarity

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$$\mathbf{i} = \{i_\alpha\}_{\alpha \in F}$$

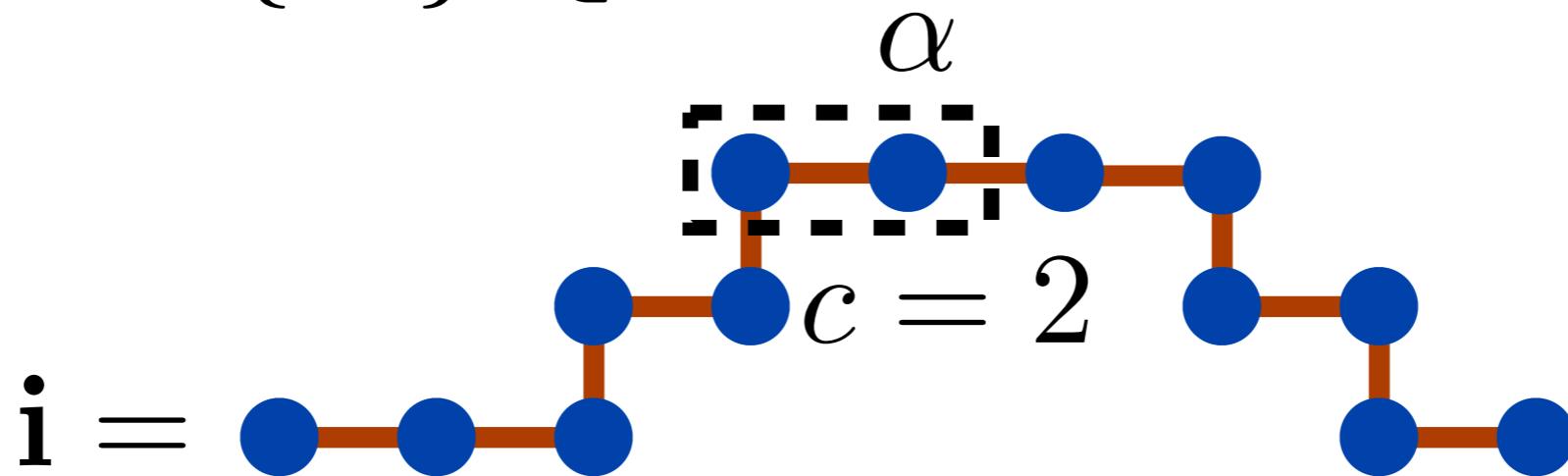


$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \phi(\mathbf{i})$$

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$$\mathbf{i} = \{i_\alpha\}_{\alpha \in F}$$

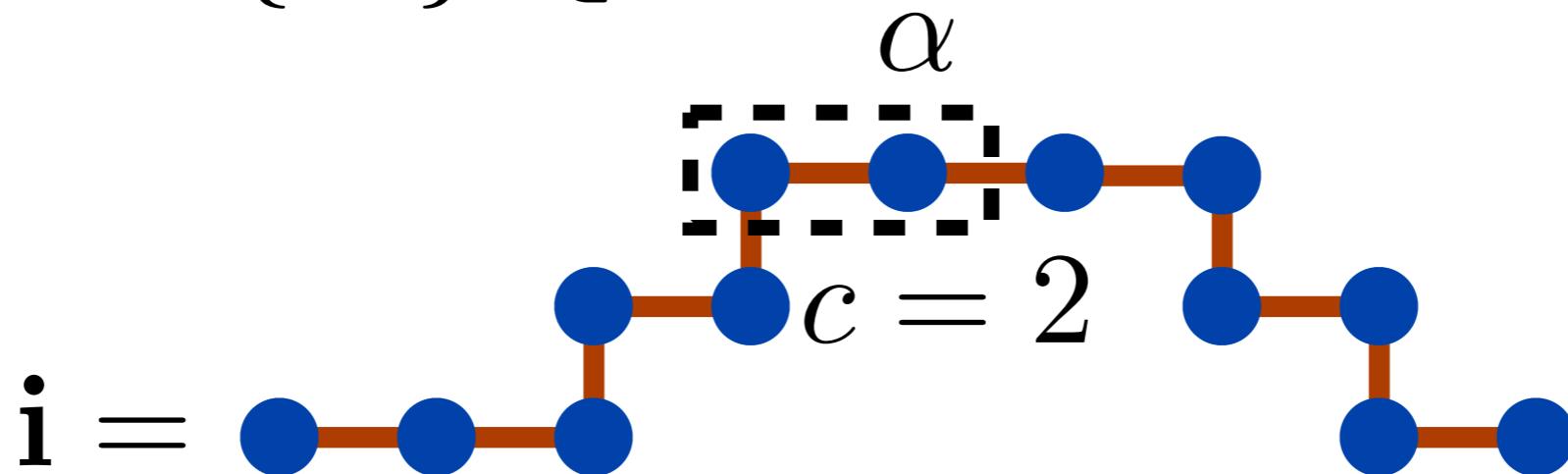


$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$$\mathbf{i} = \{i_\alpha\}_{\alpha \in F}$$

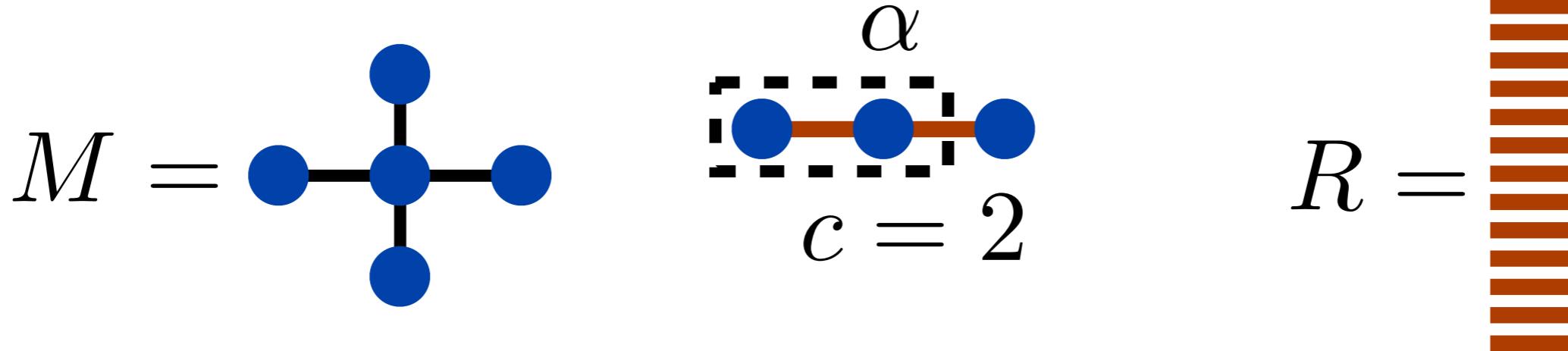


$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

$$\mathbf{Y} \sim \mathcal{P}_L \quad O(ND^2k)$$

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

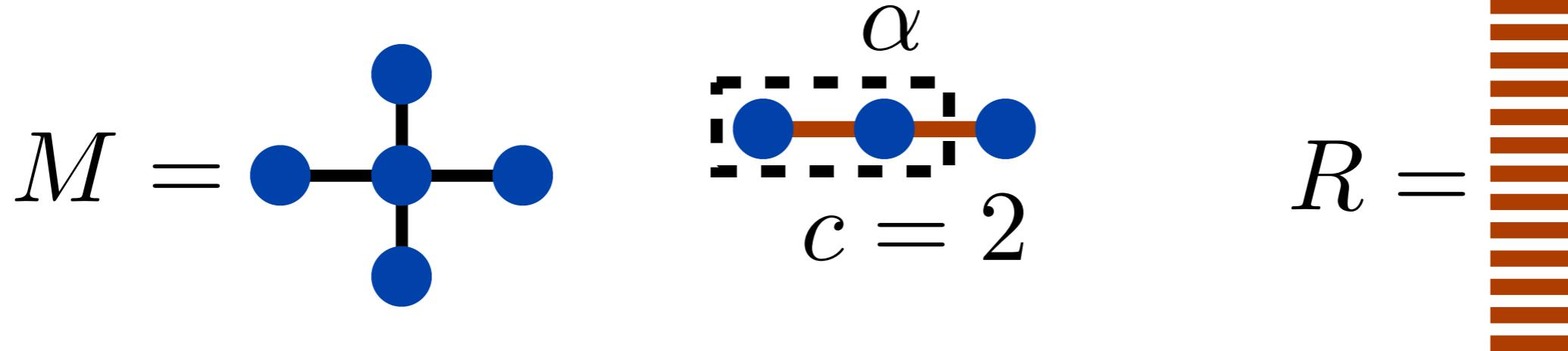


$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

$$\mathbf{Y} \sim \mathcal{P}_L \quad O(\textcolor{blue}{D}^2 \textcolor{green}{k}^3 + \textcolor{blue}{D} \textcolor{green}{k}^2 M^c R)$$

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)



$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

$$\mathbf{Y} \sim \mathcal{P}_L \quad O(\textcolor{blue}{D}^2 \textcolor{green}{k}^3 + \textcolor{blue}{D} \textcolor{green}{k}^2 M^c R)$$

$$M^c R = 4^2 * 12 = 192 \ll N = 4^{12} = 16,777,216$$

LARGE FEATURE SETS

LARGE FEATURE SETS

$N = \# \text{ of items}$

$D = \# \text{ of features}$

	Large	Exponential
Small	dual	dual + structure

LARGE FEATURE SETS

$N = \# \text{ of items}$

$D = \# \text{ of features}$	Small	Large
Large	?	?
Large	dual	dual + structure
Large	Exponential	

RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

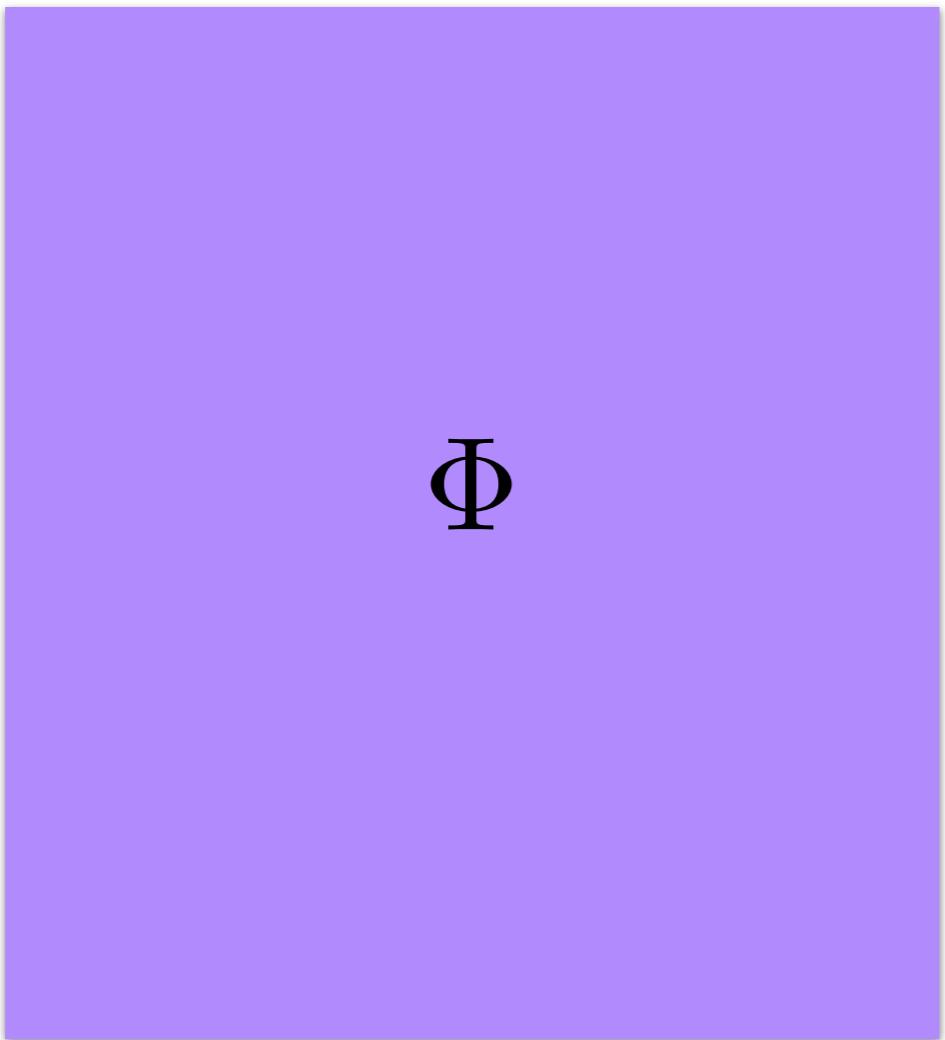
RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

D

N

Φ



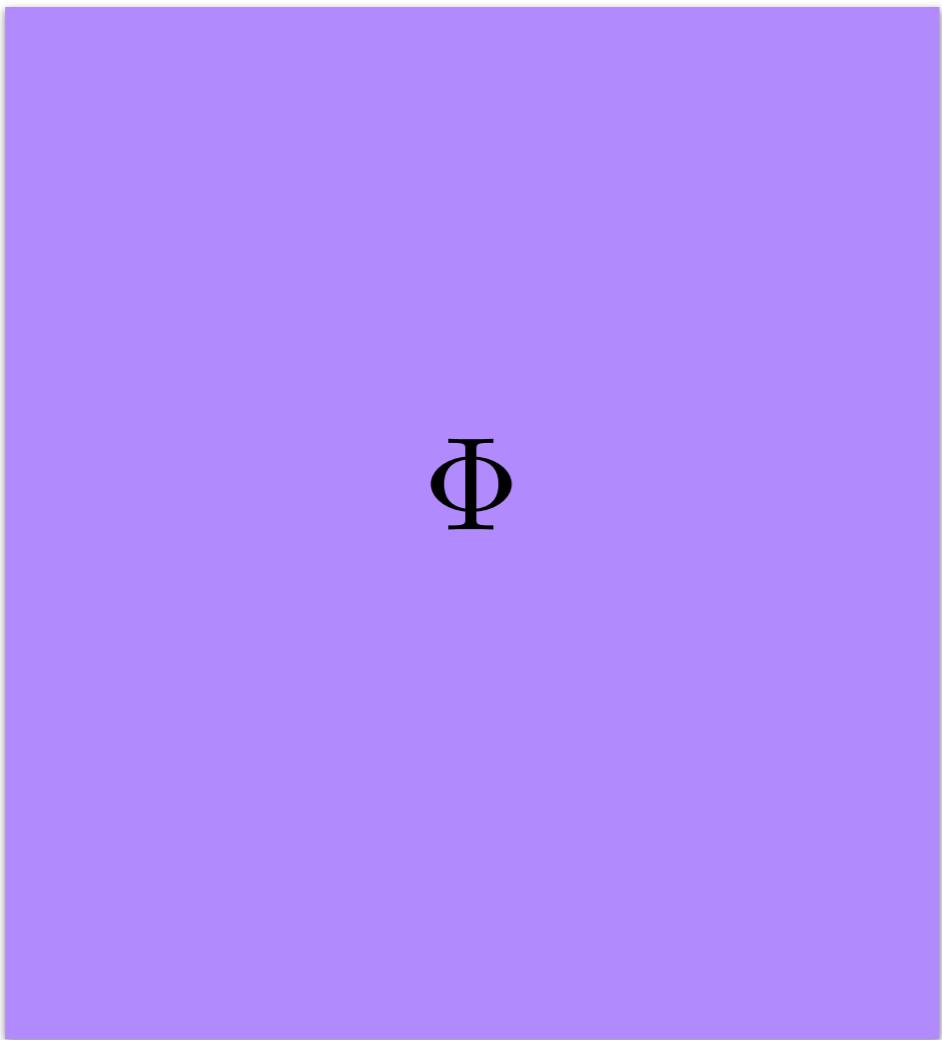
RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

D

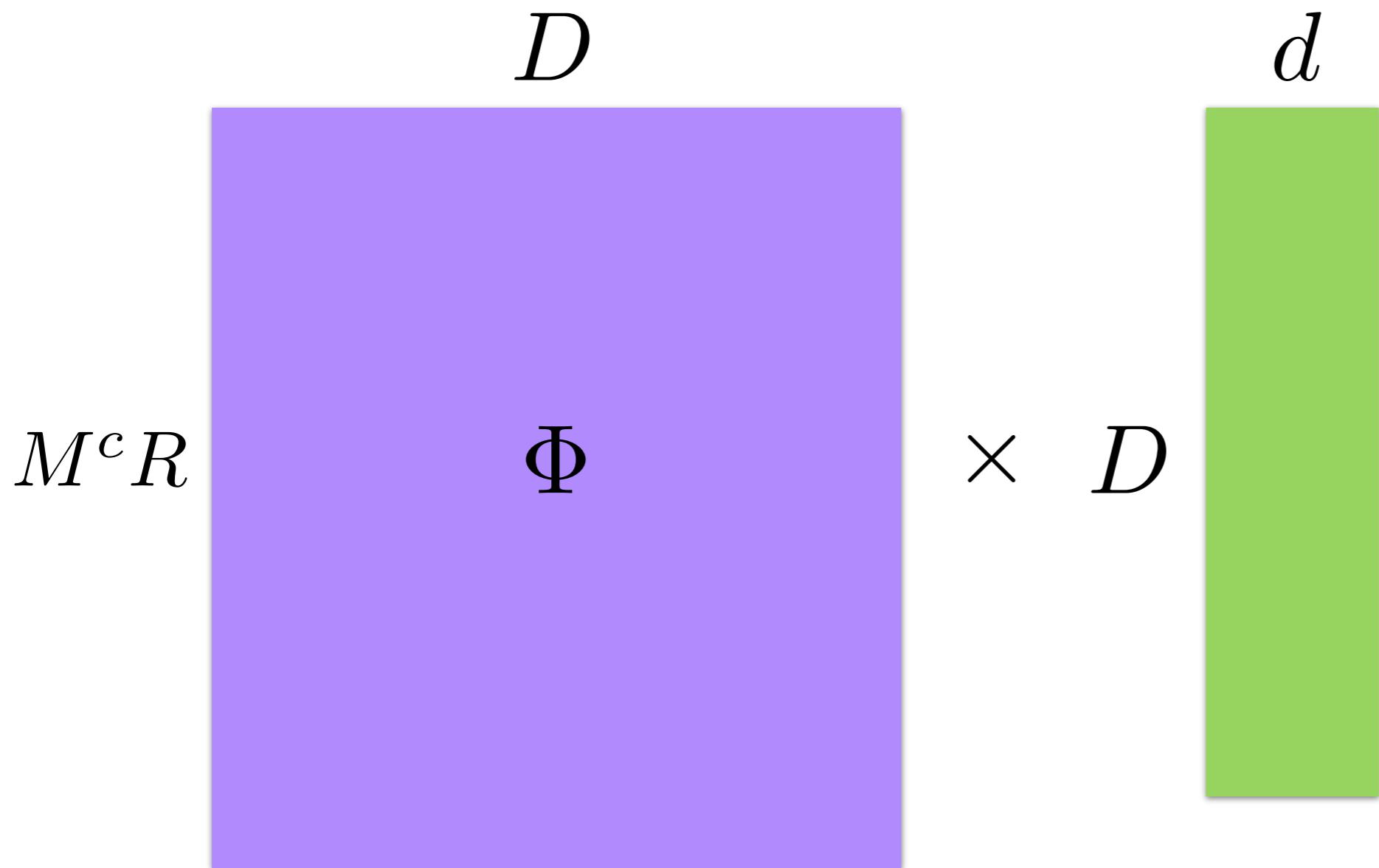
$M^c R$

Φ



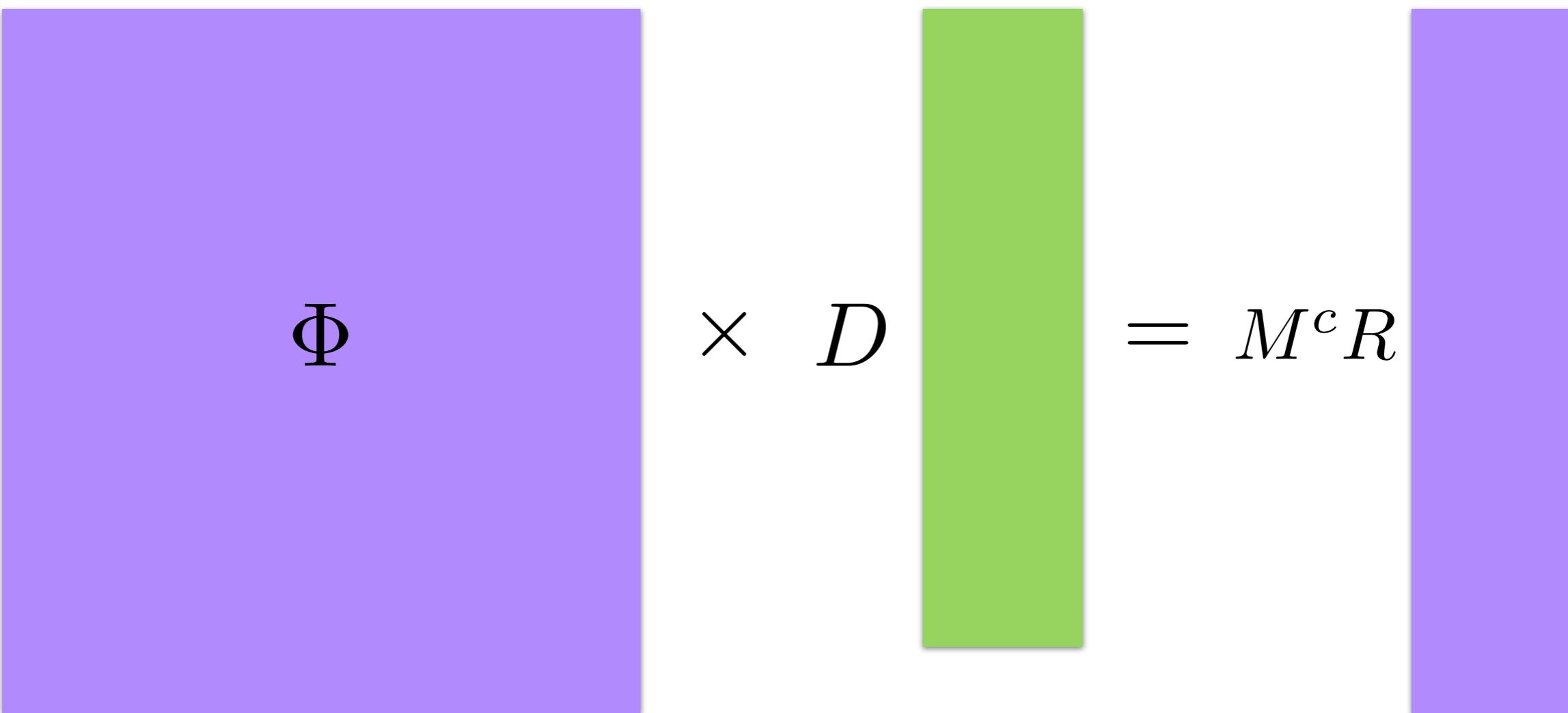
RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)



RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$M^c R \begin{matrix} D \\ \Phi \end{matrix} \times \begin{matrix} d \\ D \end{matrix} = M^c R \begin{matrix} d \\ \Phi \end{matrix}$$


RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

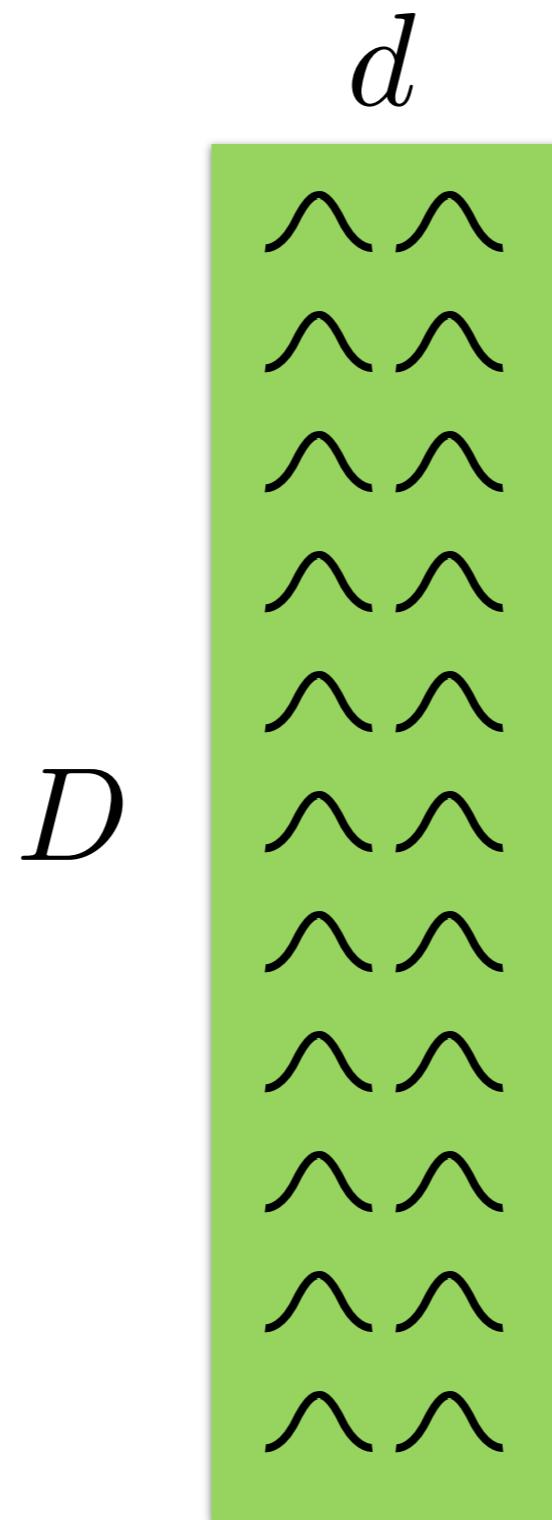
d



D

RANDOM PROJECTIONS

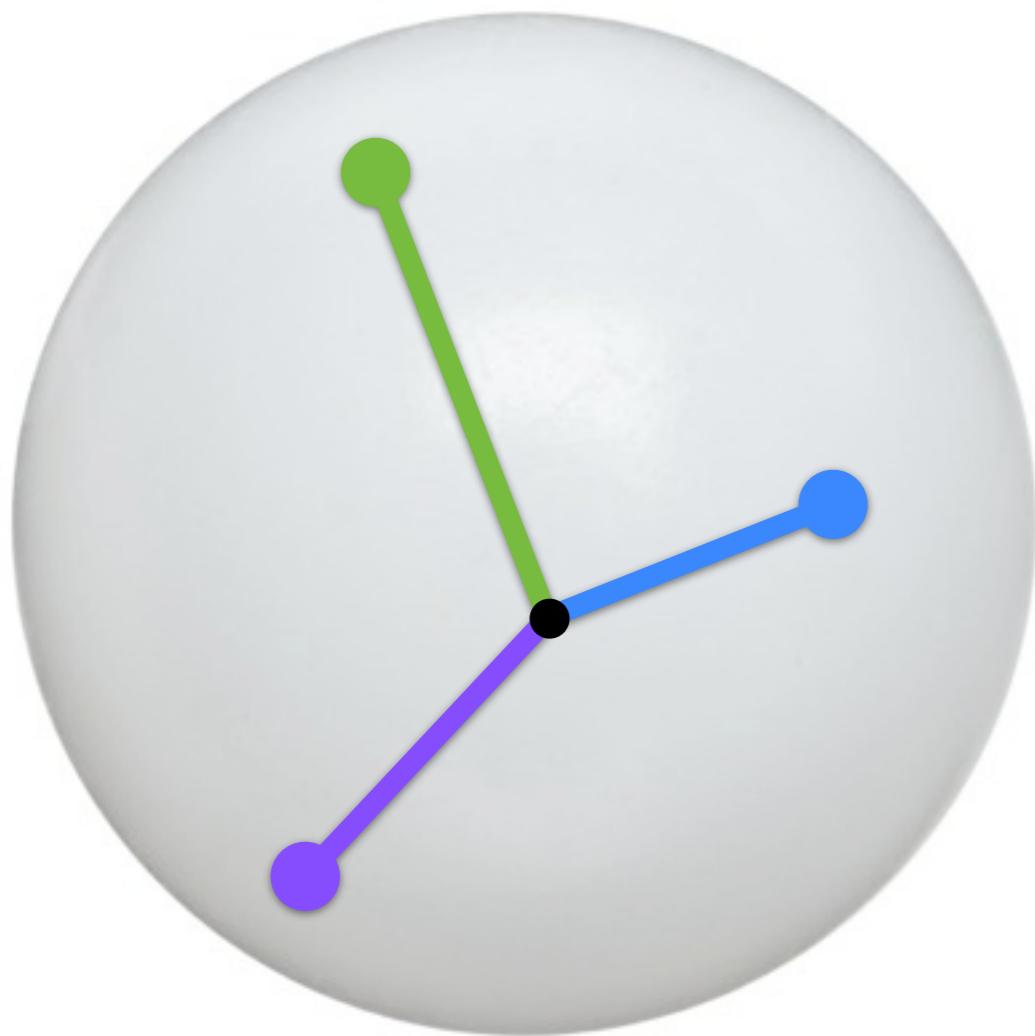
GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)



VOLUME PRESERVATION

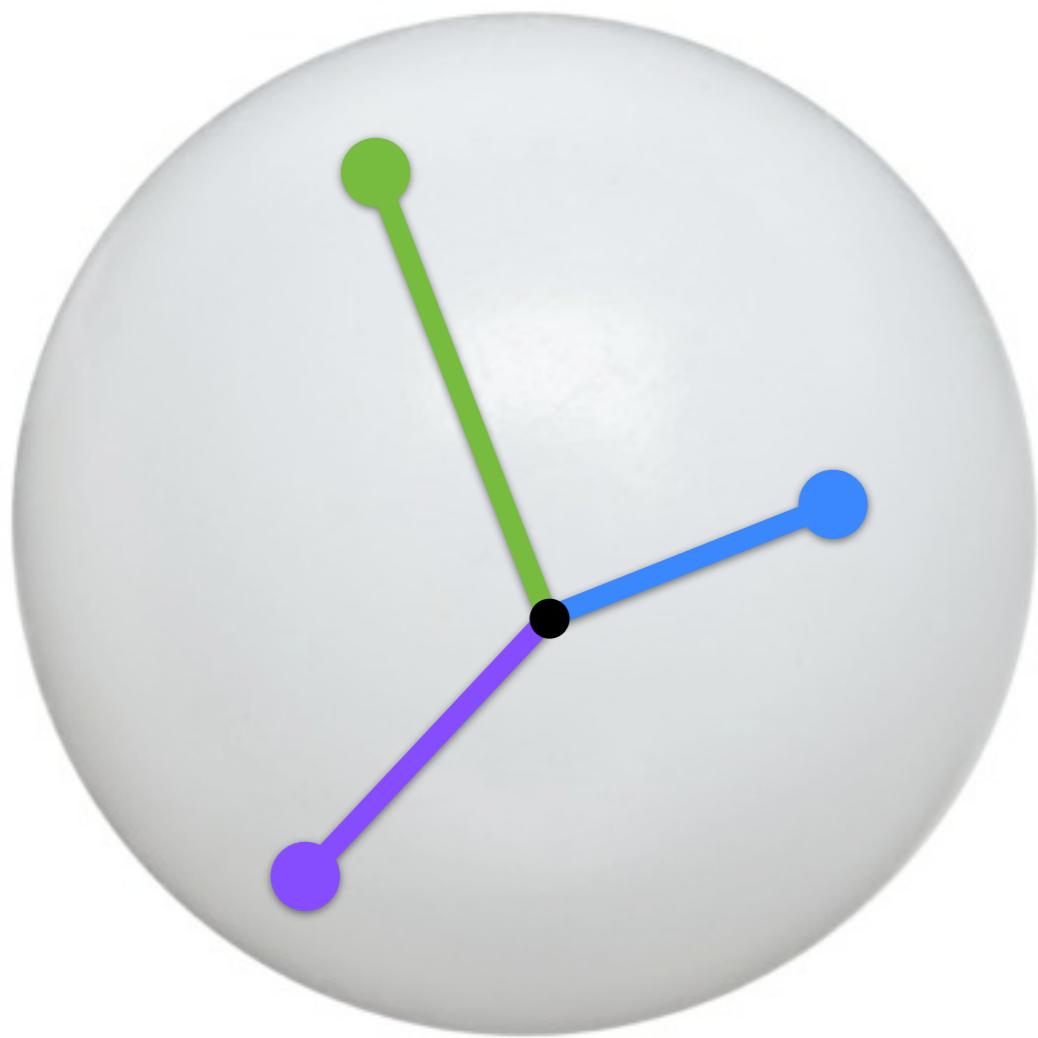
VOLUME PRESERVATION

JOHNSON AND LINDENSTRAUSS (1984)

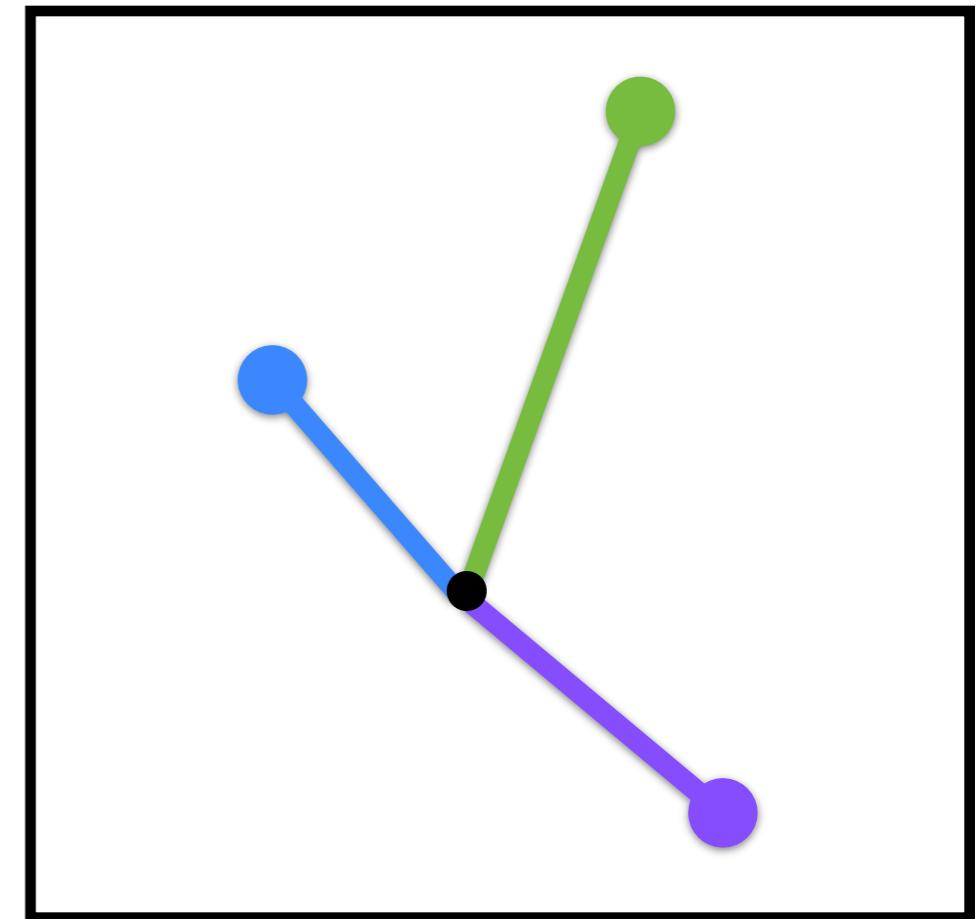


VOLUME PRESERVATION

JOHNSON AND LINDENSTRAUSS (1984)

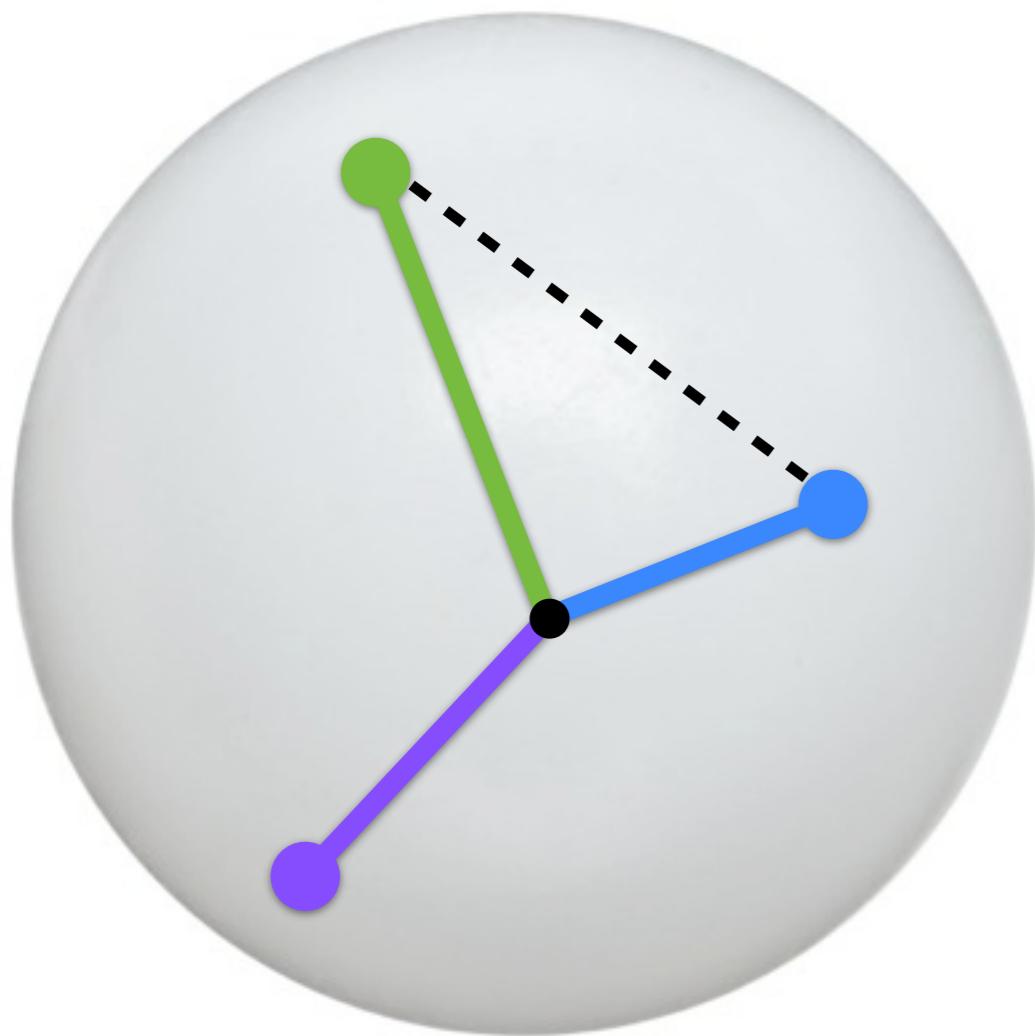


$\log N$

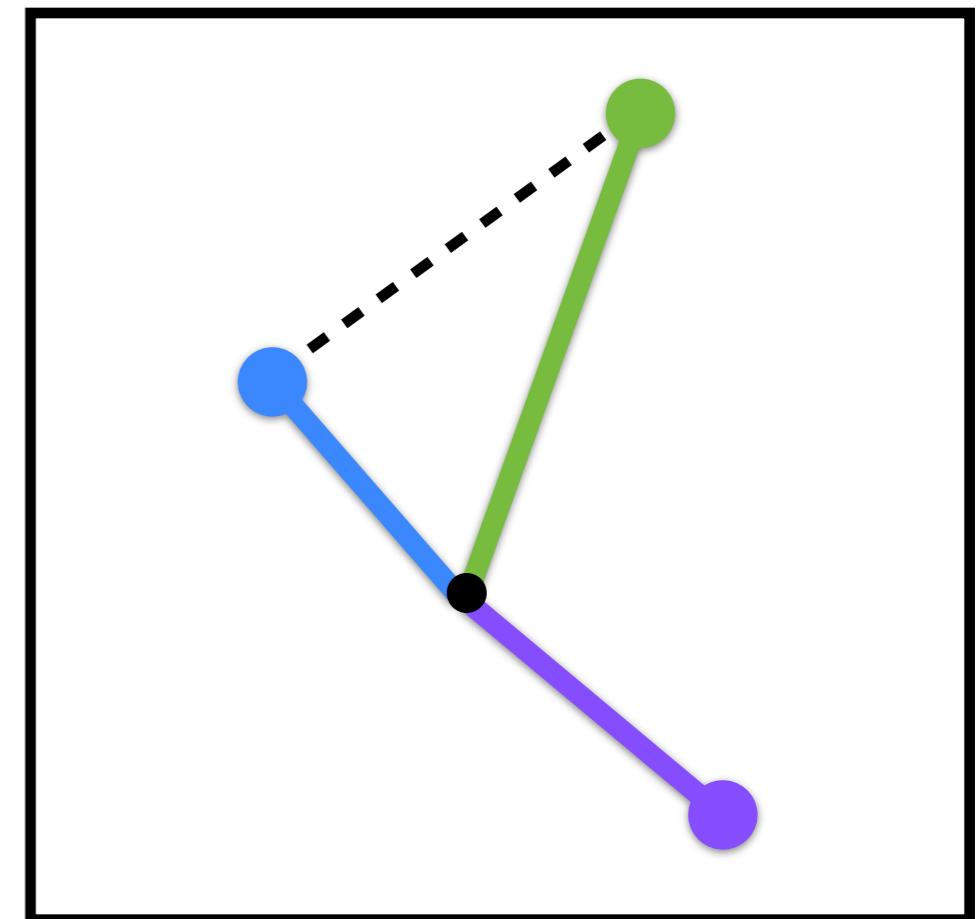


VOLUME PRESERVATION

JOHNSON AND LINDENSTRAUSS (1984)

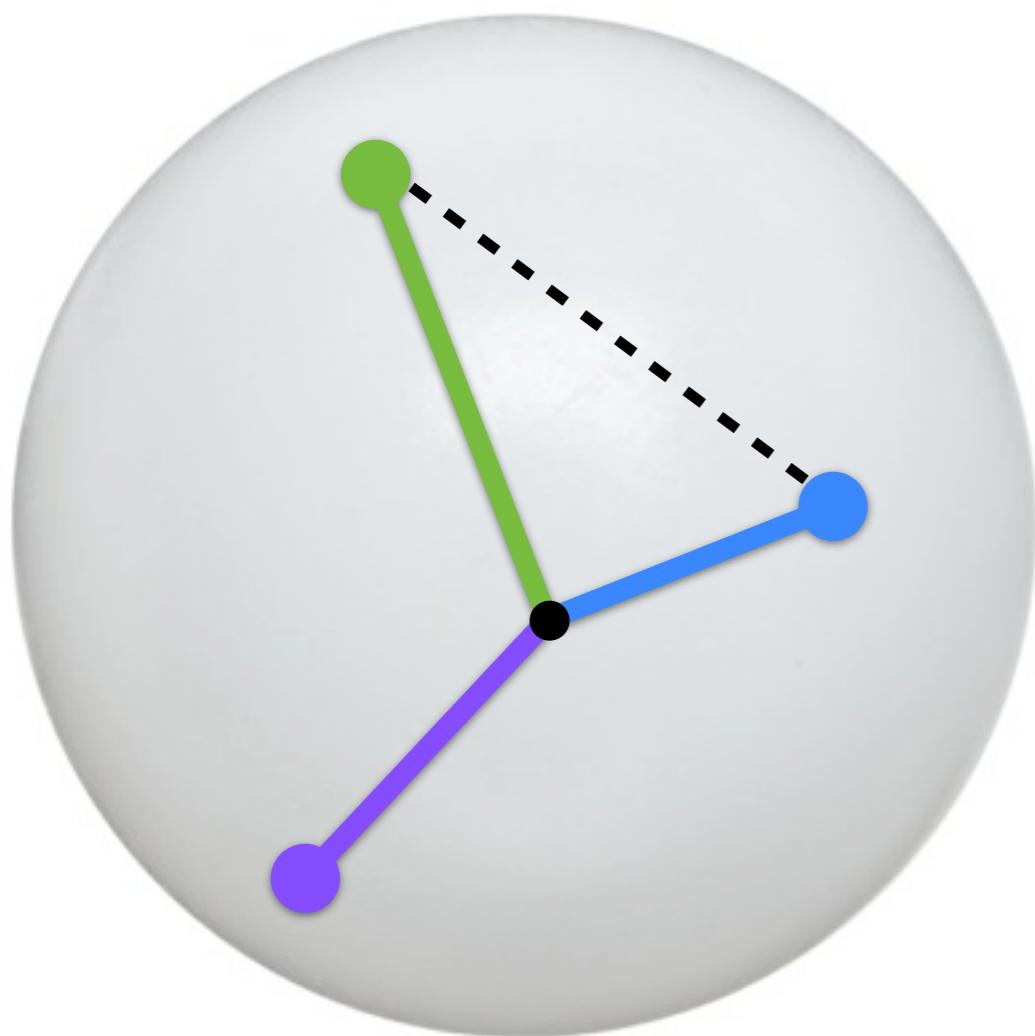


$\log N$

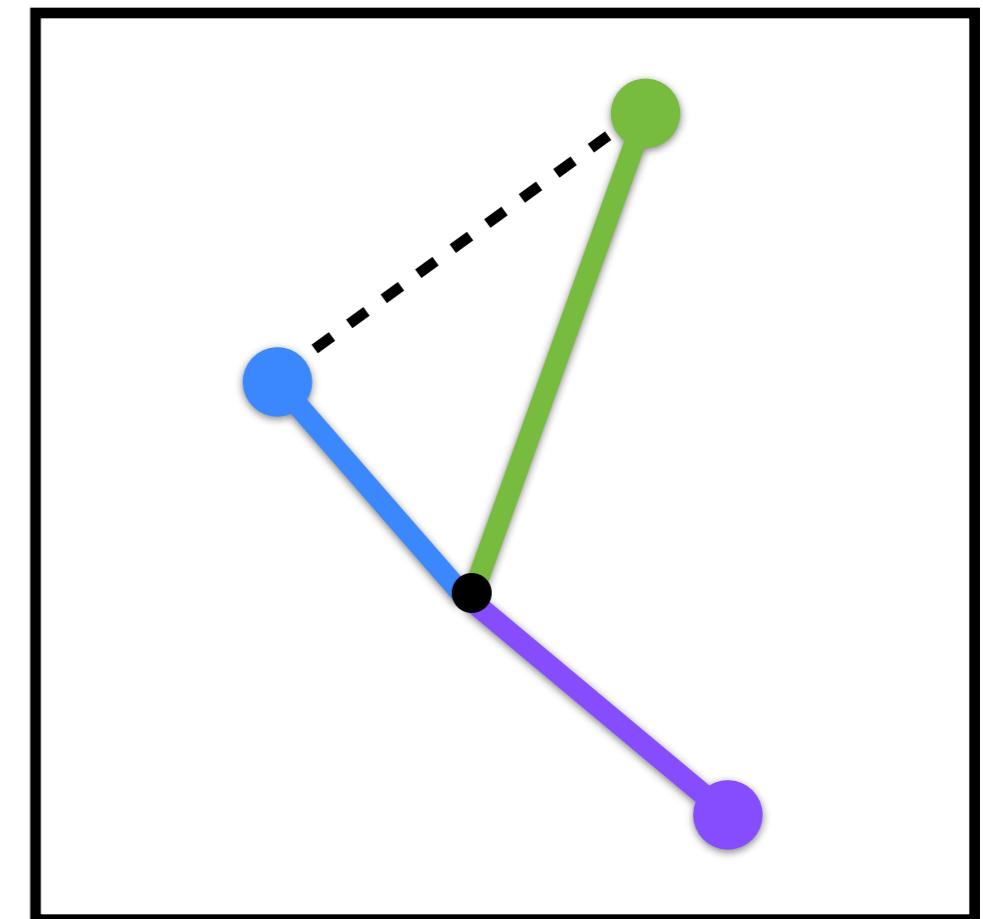


VOLUME PRESERVATION

MAGEN AND ZOUZIAS (2008)

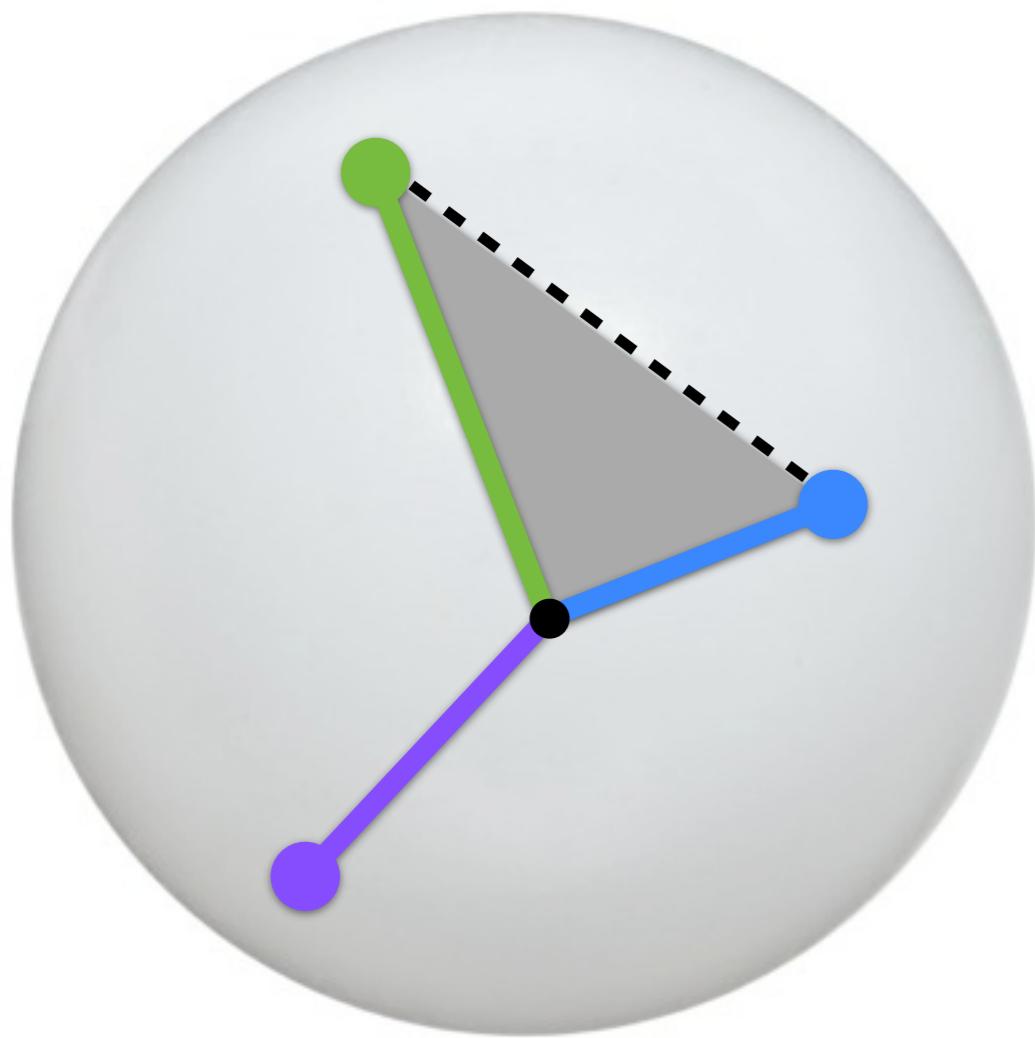


$\log N$

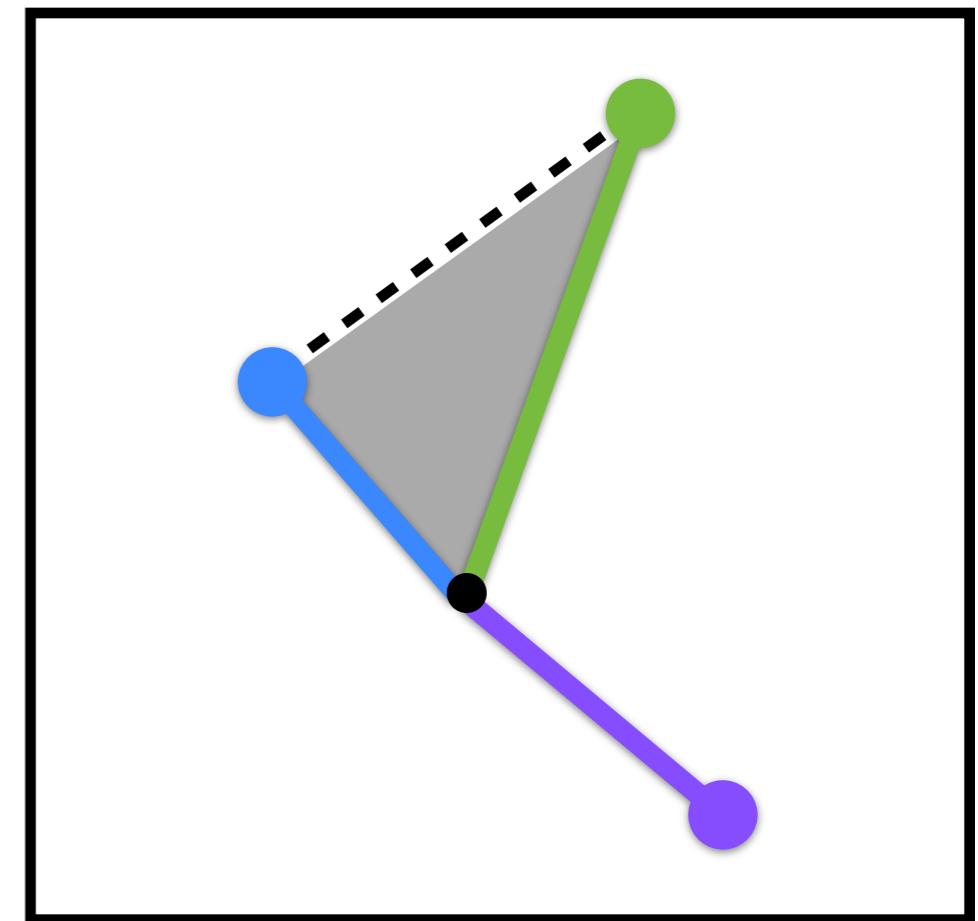


VOLUME PRESERVATION

MAGEN AND ZOUZIAS (2008)



$\log N$



DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$\text{vol}^2 = \det$$

DPP PRESERVATION

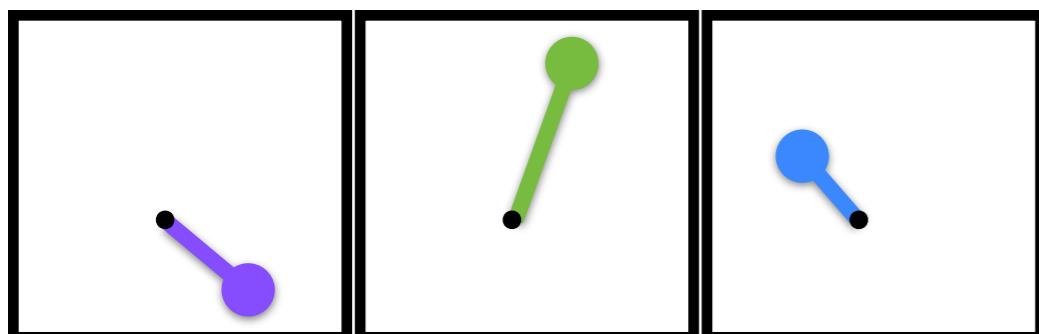
GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$\text{vol}^2 = \det$$

$$k = 1$$



\approx



DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$\text{vol}^2 = \det$$

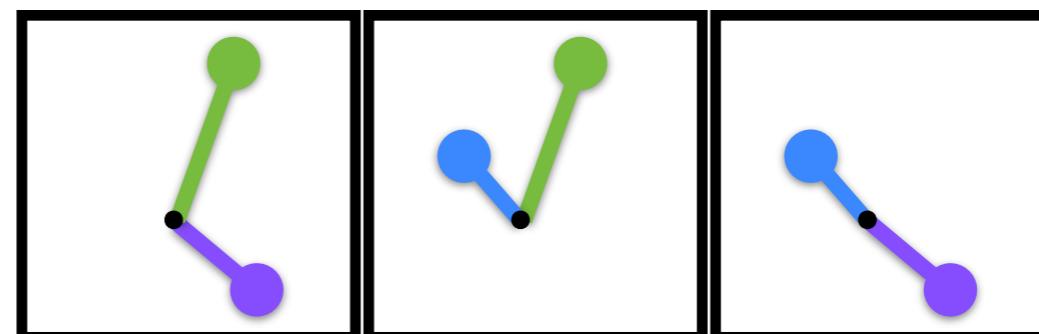
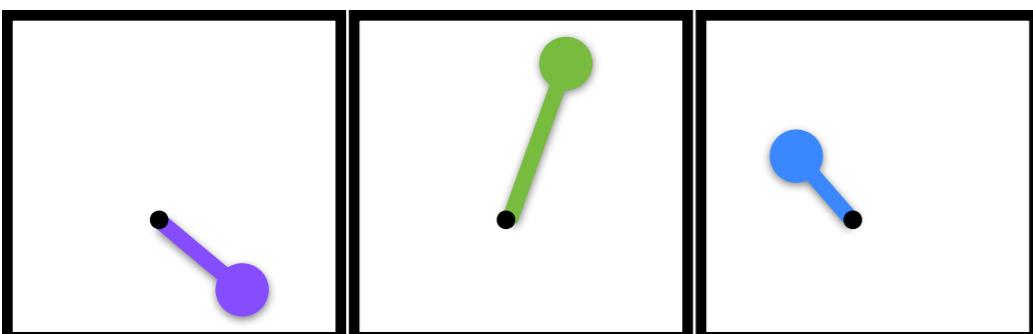
$$k = 1$$



$$k = 2$$



\approx



DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$\text{vol}^2 = \det$$

$k = 1$



$k = 2$



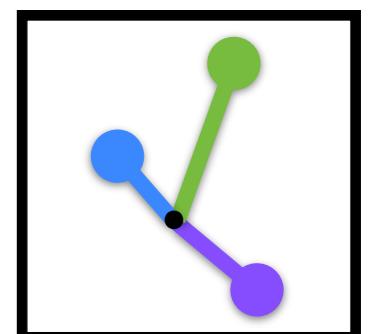
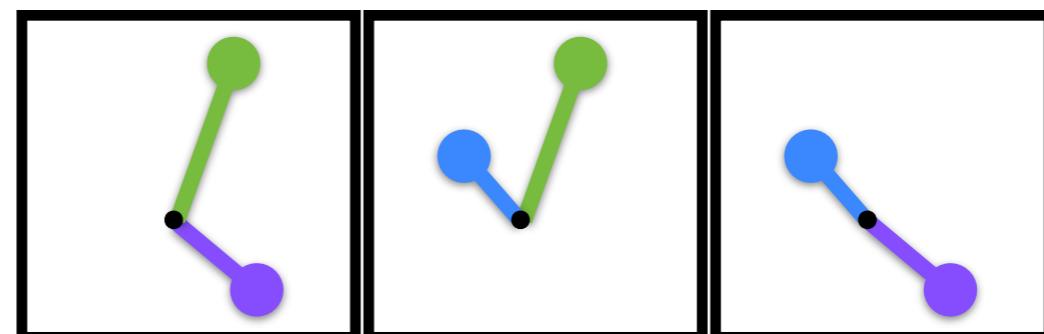
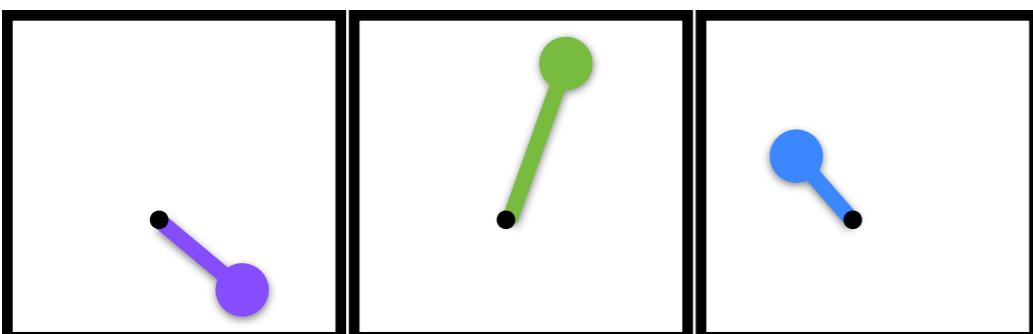
$k = 3$



\approx

\approx

\approx



DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$d = O \left(\max \left\{ \frac{k}{\epsilon}, \frac{\log(1/\delta) + \log(N)}{\epsilon^2} + k \right\} \right)$$

DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$d = O \left(\max \left\{ \frac{k}{\epsilon}, \frac{\log(1/\delta) + \log(N)}{\epsilon^2} + k \right\} \right)$$

subset size total # of items



DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$d = O \left(\max \left\{ \frac{k}{\epsilon}, \frac{\log(1/\delta) + \log(N)}{\epsilon^2} + k \right\} \right)$$

subset size total # of items
↓ ↓

$$\text{w.p. } 1 - \delta : \| \mathcal{P}^k - \tilde{\mathcal{P}}^k \|_1 \leq e^{6k\epsilon} - 1$$

DPP PRESERVATION

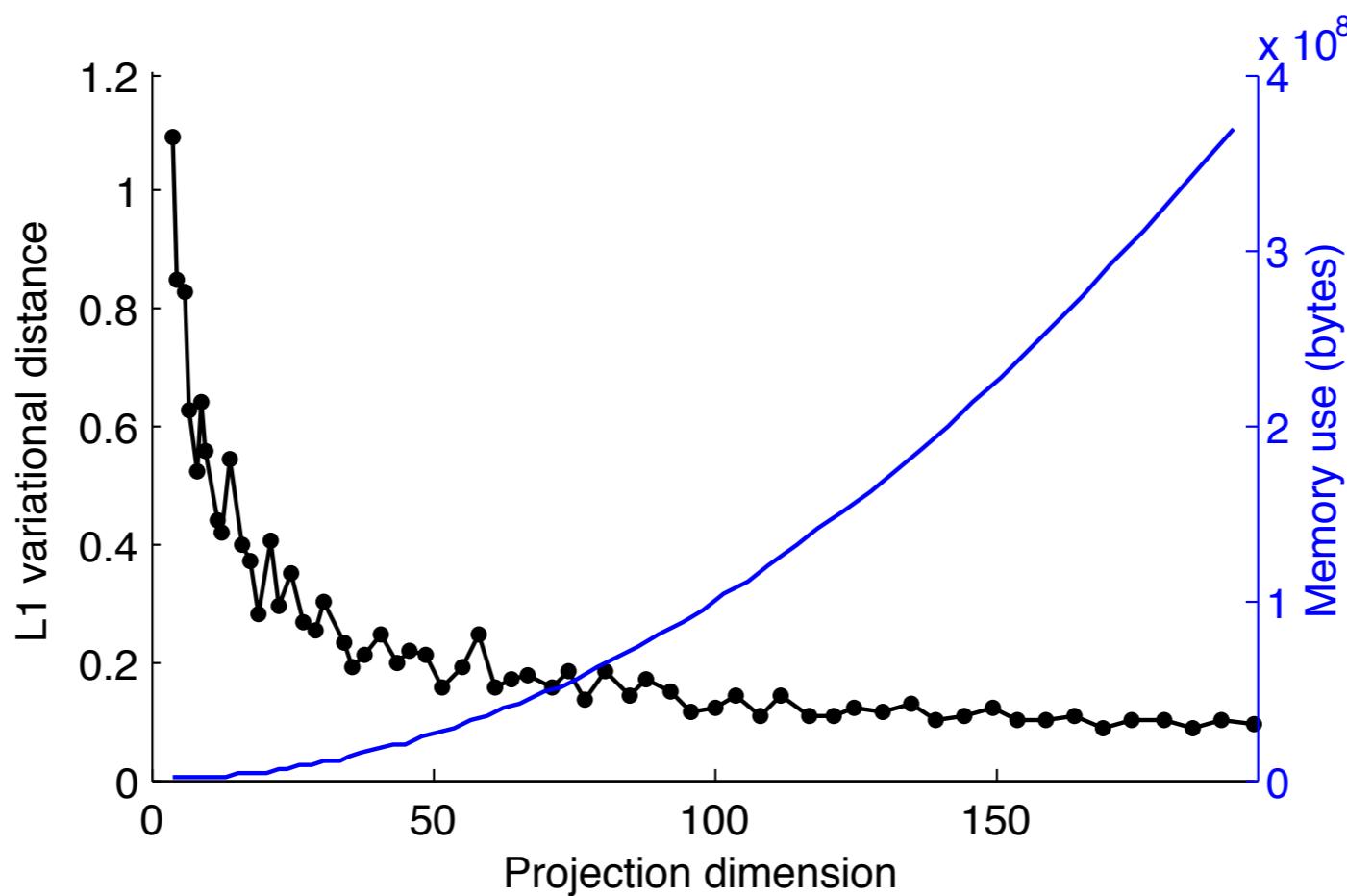
GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$d = O \left(\max \left\{ \frac{k}{\epsilon}, \frac{\log(1/\delta) + \log(N)}{\epsilon^2} + k \right\} \right)$$

subset size total # of items

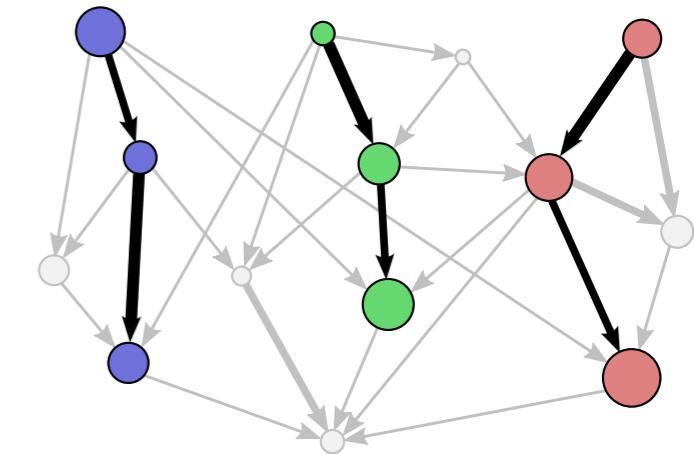
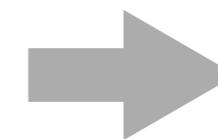
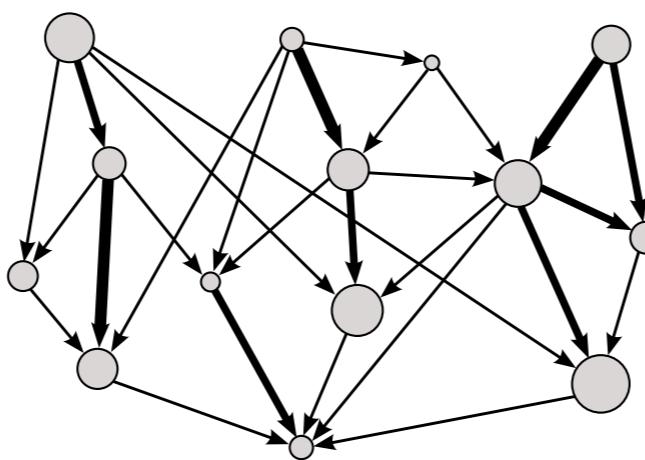
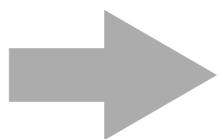
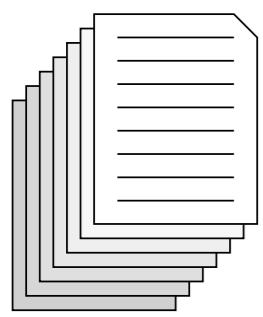
\downarrow \downarrow

$$\text{w.p. } 1 - \delta : \| \mathcal{P}^k - \tilde{\mathcal{P}}^k \|_1 \leq e^{6k\epsilon} - 1$$

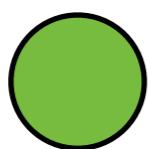
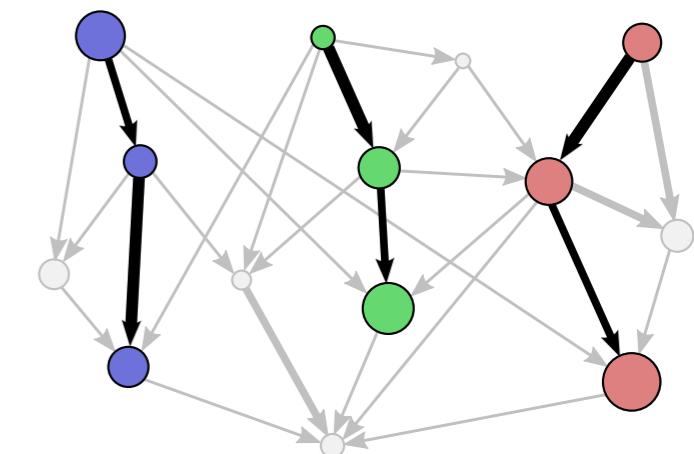
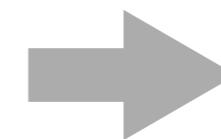
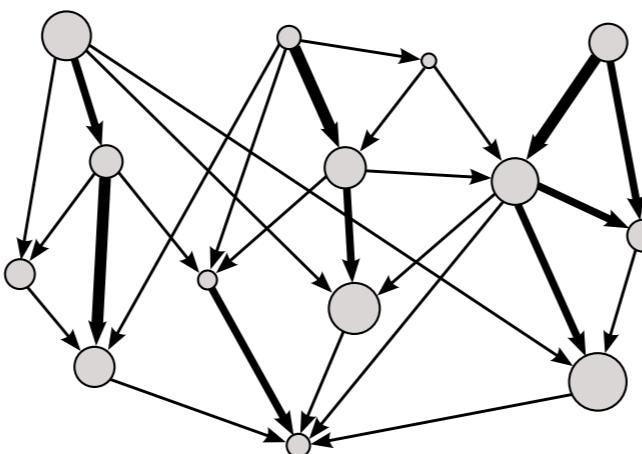
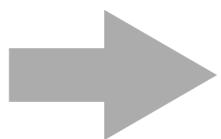
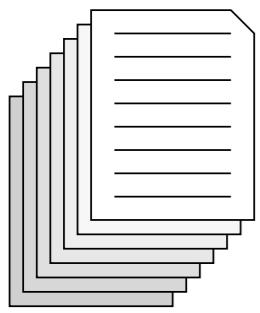


NEWS THREADING

NEWS THREADING

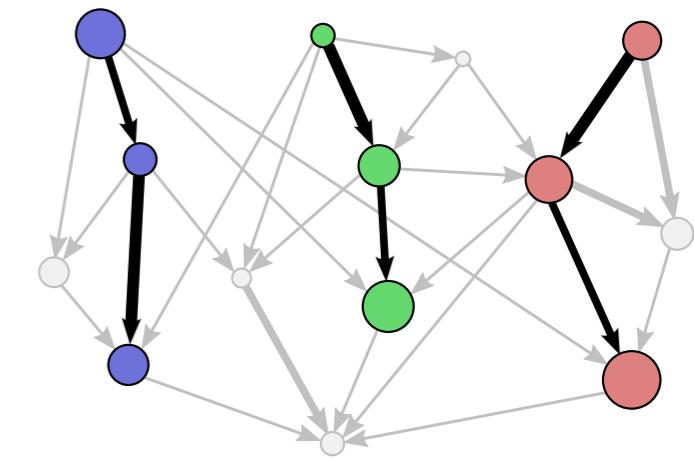
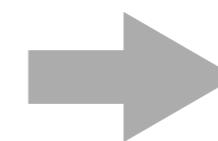
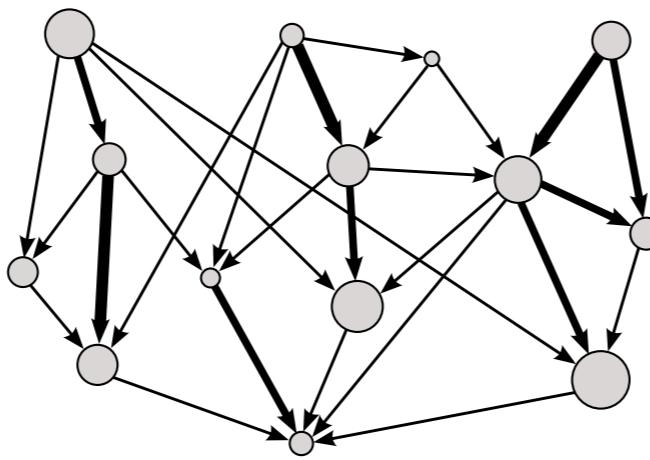
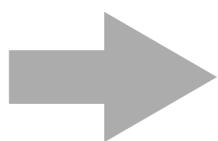
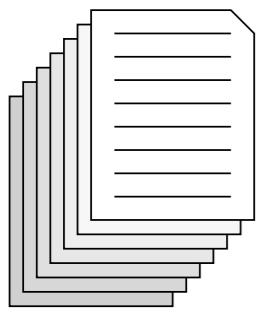


NEWS THREADING

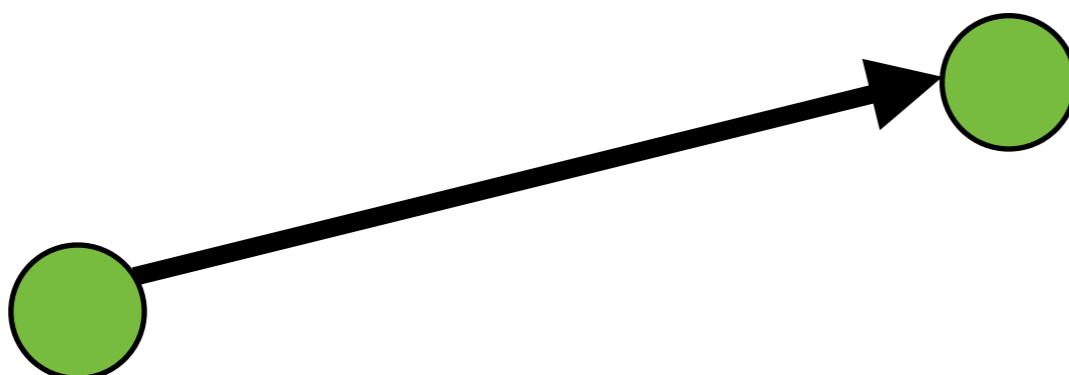


March 28: Health officials confirm
Ebola outbreak in Guinea's capital

NEWS THREADING

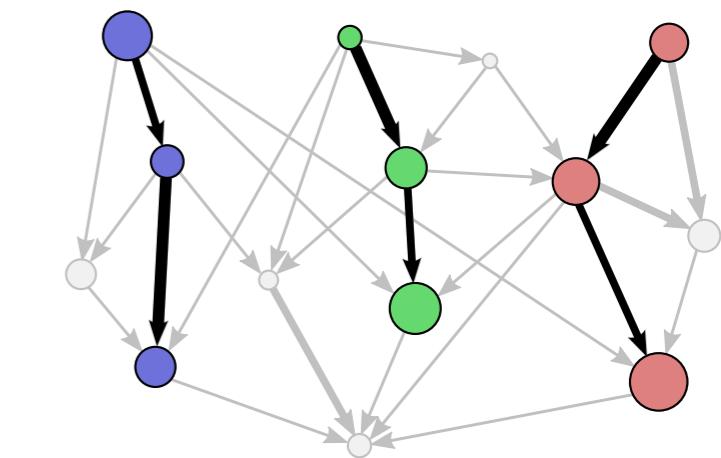
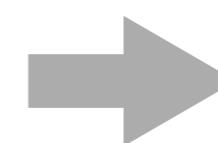
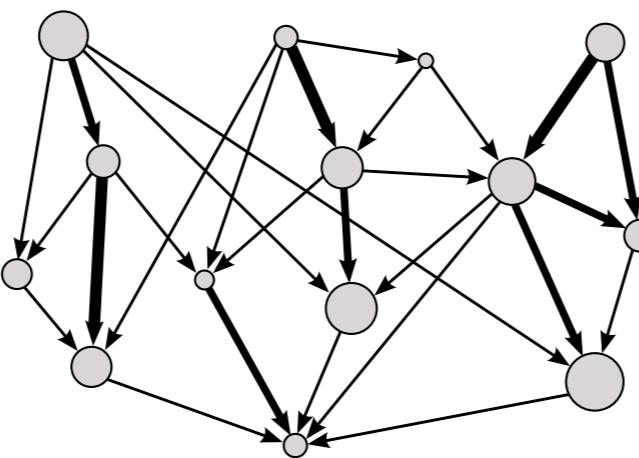
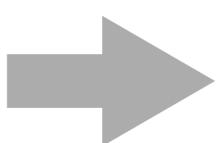
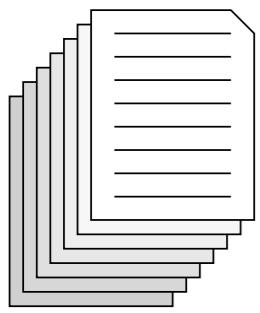


August 8: World Health Organization
declares Ebola epidemic an
international health emergency

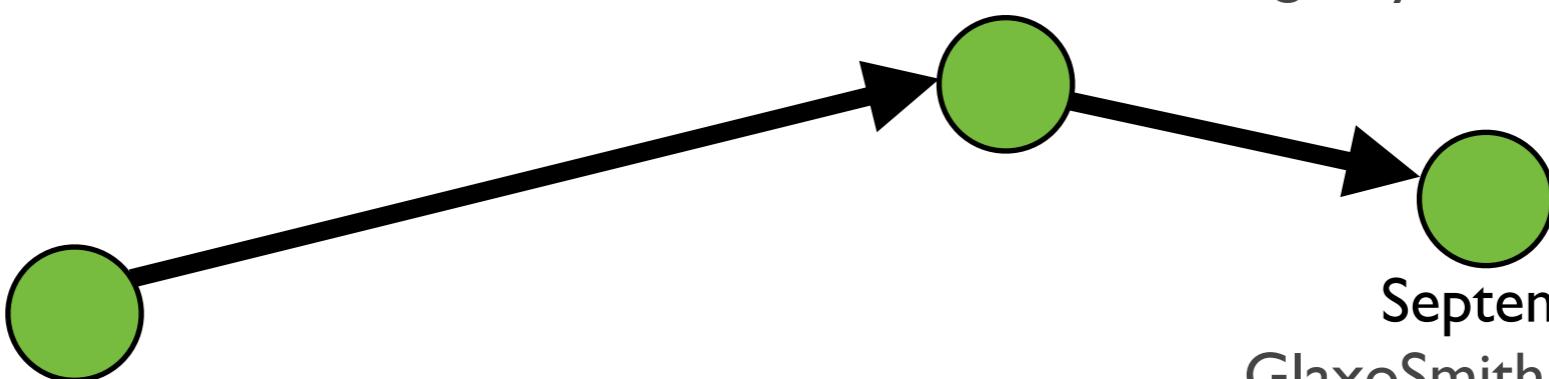


March 28: Health officials confirm
Ebola outbreak in Guinea's capital

NEWS THREADING



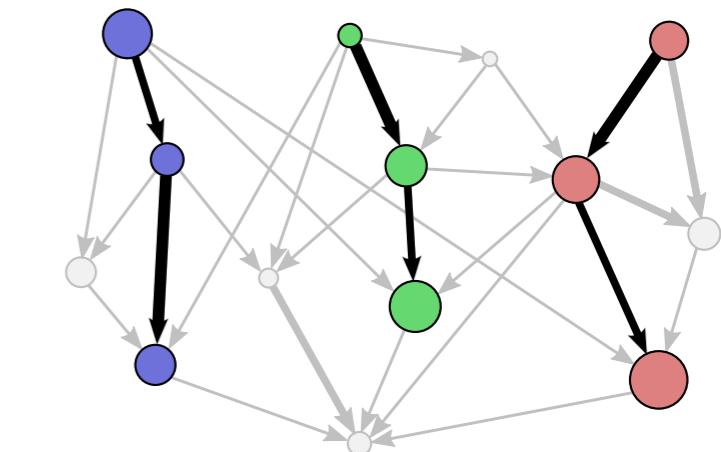
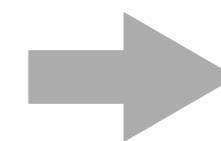
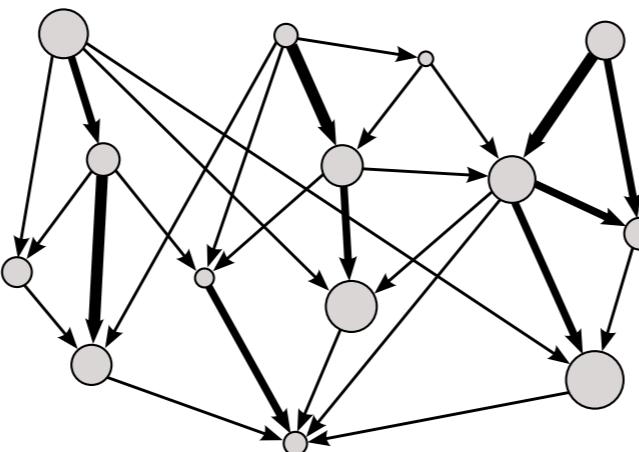
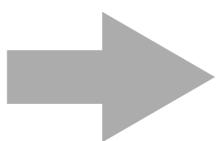
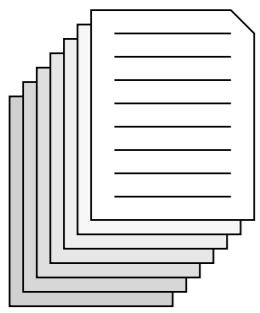
August 8: World Health Organization
declares Ebola epidemic an
international health emergency



March 28: Health officials confirm
Ebola outbreak in Guinea's capital

September 2:
GlaxoSmithKlein begins
Ebola vaccine drug trial

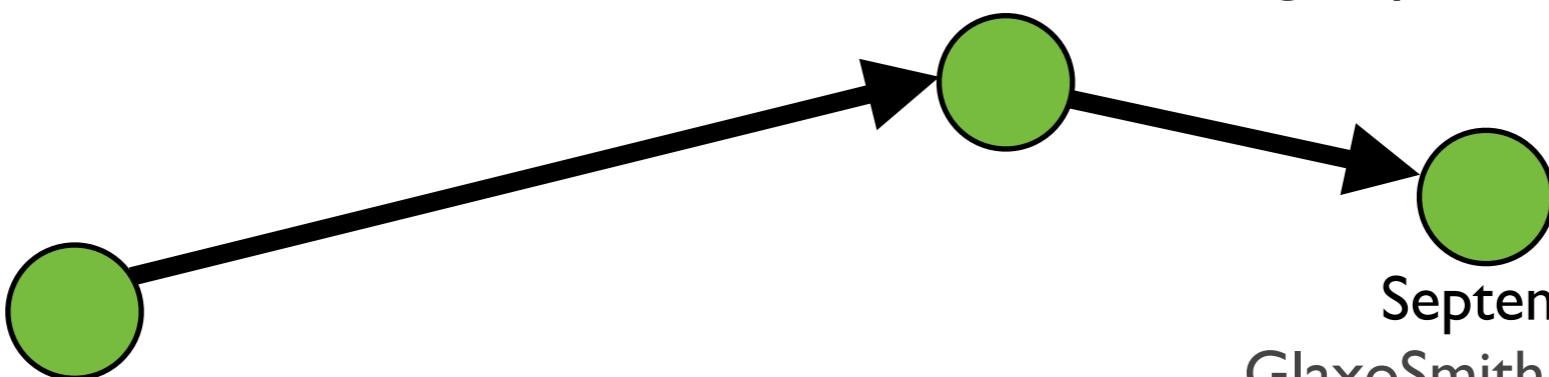
NEWS THREADING



$M \approx 35,000$

10^{360}

August 8: World Health Organization
declares Ebola epidemic an
international health emergency



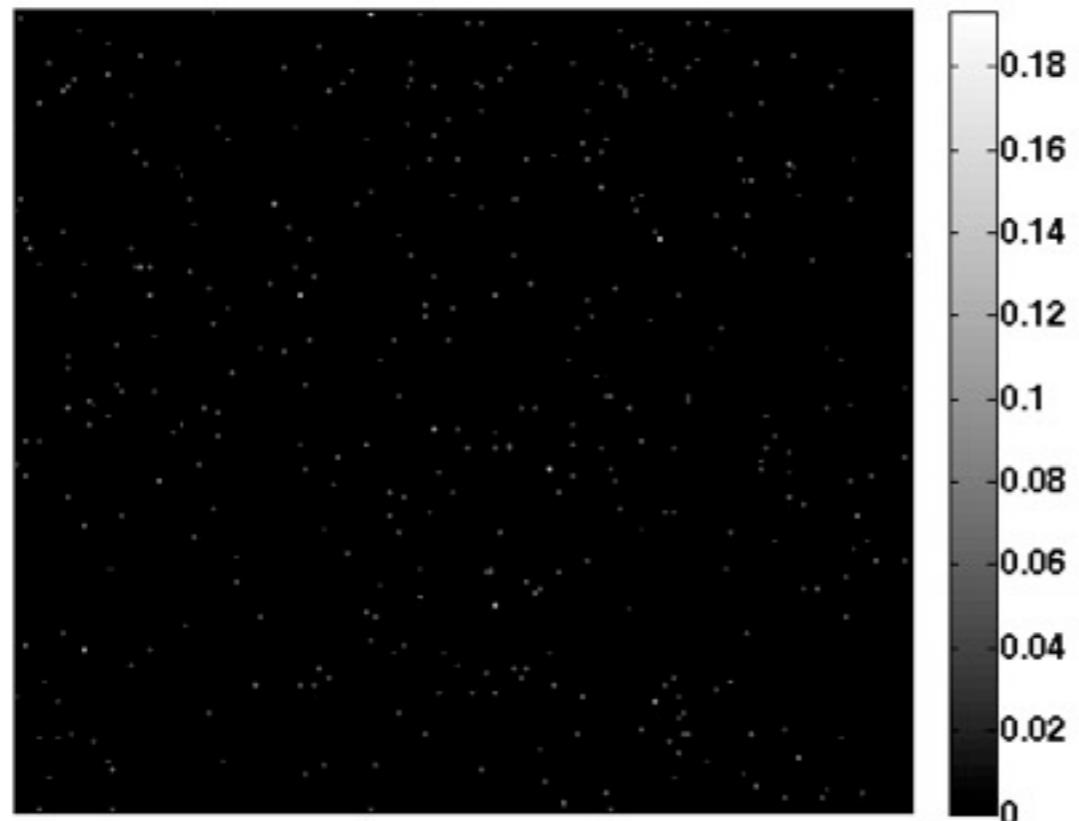
March 28: Health officials confirm
Ebola outbreak in Guinea's capital

September 2:
GlaxoSmithKlein begins
Ebola vaccine drug trial

PROJECTING NEWS FEATURES

PROJECTING NEWS FEATURES

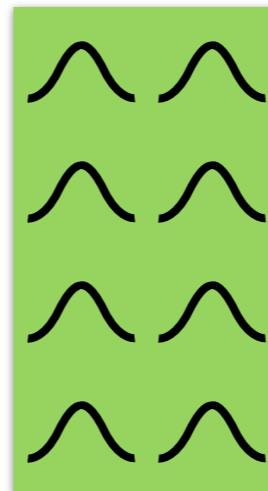
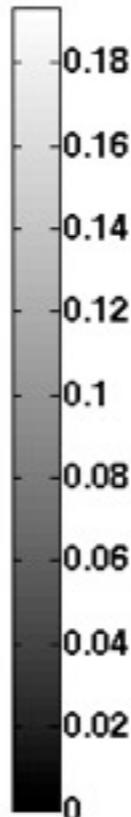
$\phi(\mathbf{i})$



$D = 36,356$

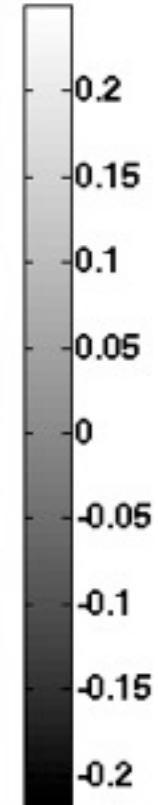
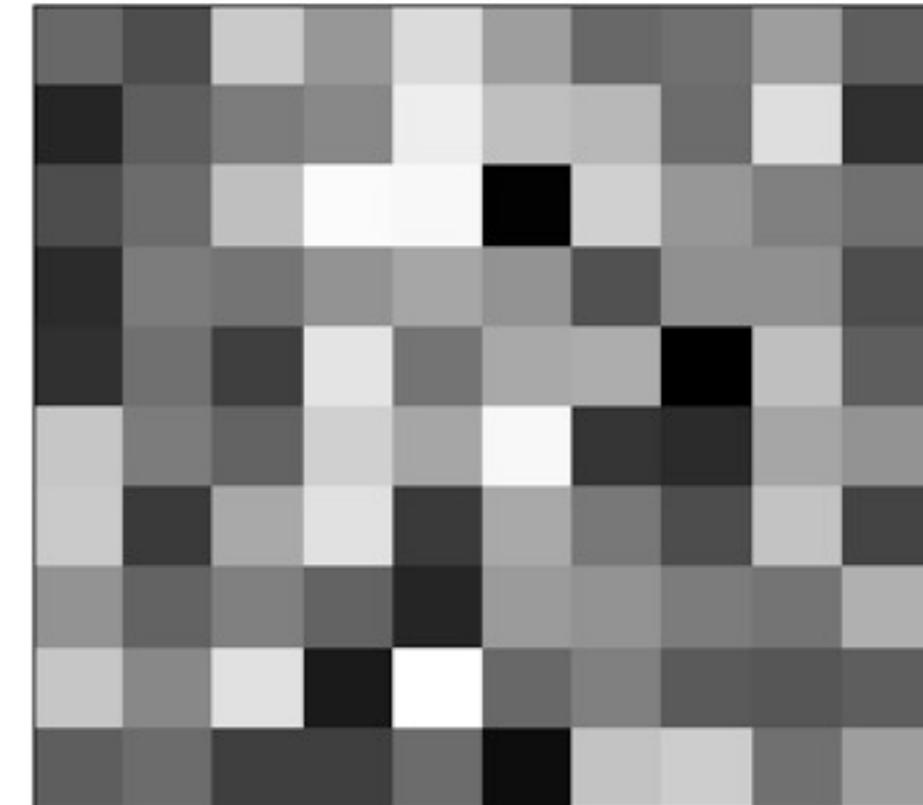
PROJECTING NEWS FEATURES

$\phi(\mathbf{i})$



G
→

$G\phi(\mathbf{i})$

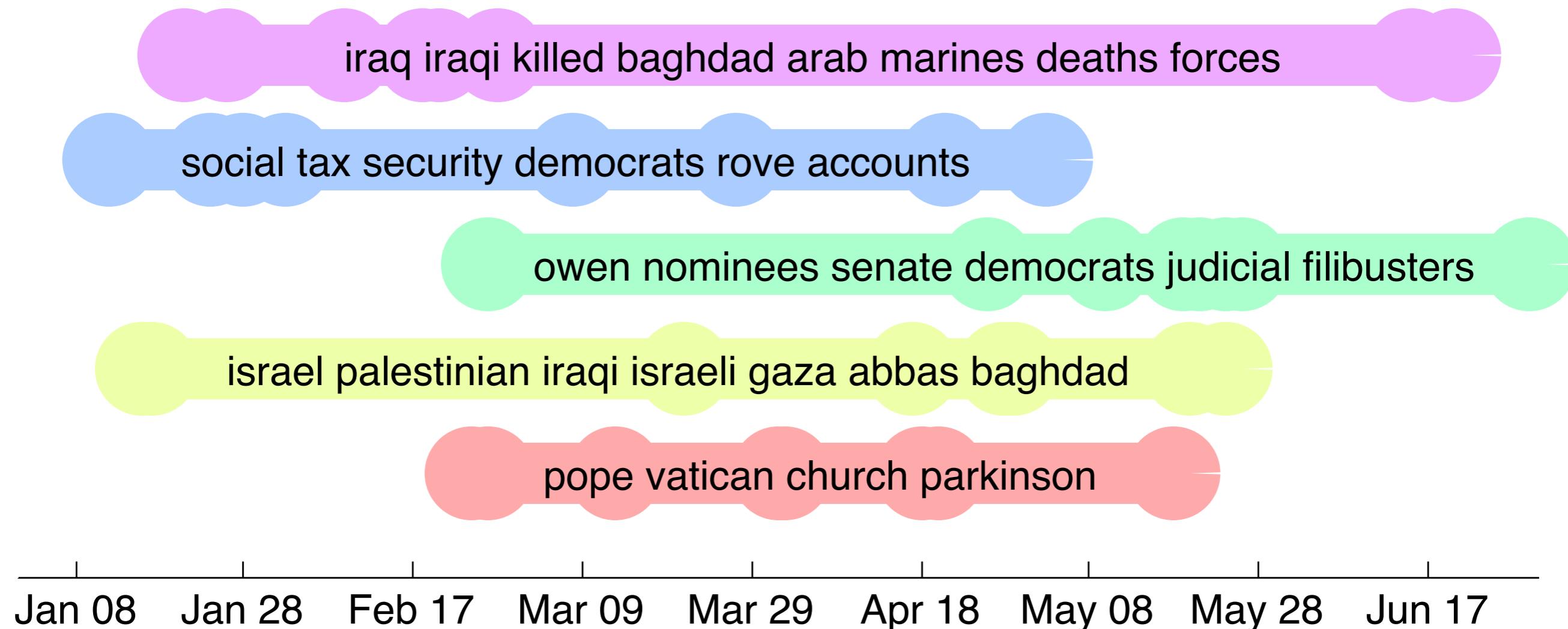


$D = 36,356$

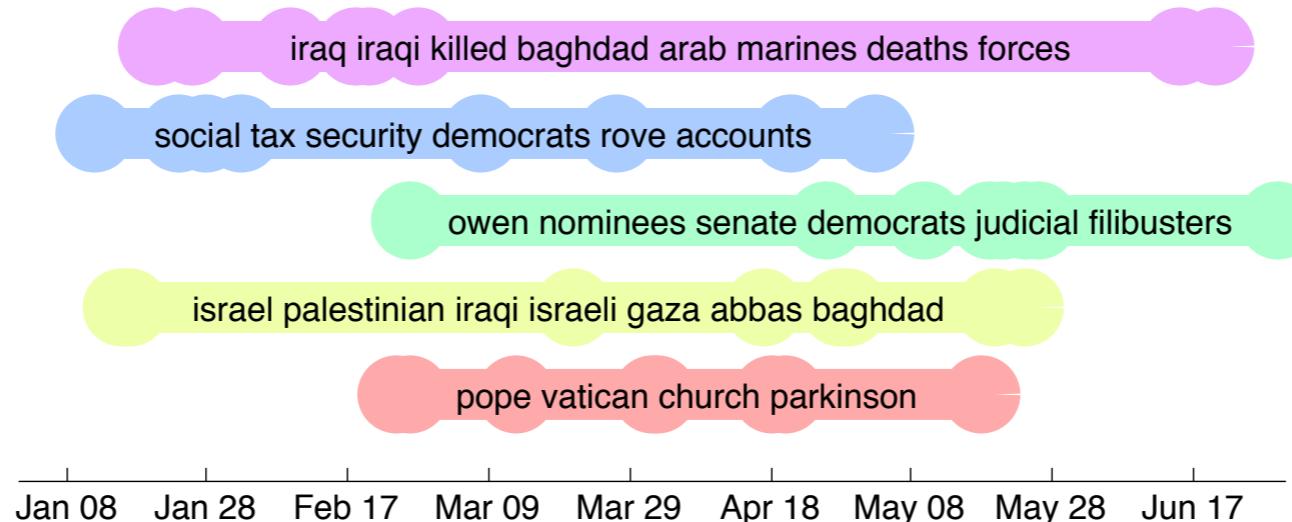
$d = 50$

DPP THREADS

DPP THREADS



DPP THREADS



Feb 24: Parkinson's Disease Increases Risks to Pope

Feb 26: Pope's Health Raises Questions About His Ability to Lead

Mar 13: Pope Returns Home After 18 Days at Hospital

Apr 01: Pope's Condition Worsens as World Prepares for End of Papacy

Apr 02: Pope, Though Gravely Ill, Utters Thanks for Prayers

Apr 18: Europeans Fast Falling Away from Church

Apr 20: In Developing World, Choice [of Pope] Met with Skepticism

May 18: Pope Sends Message with Choice of Name

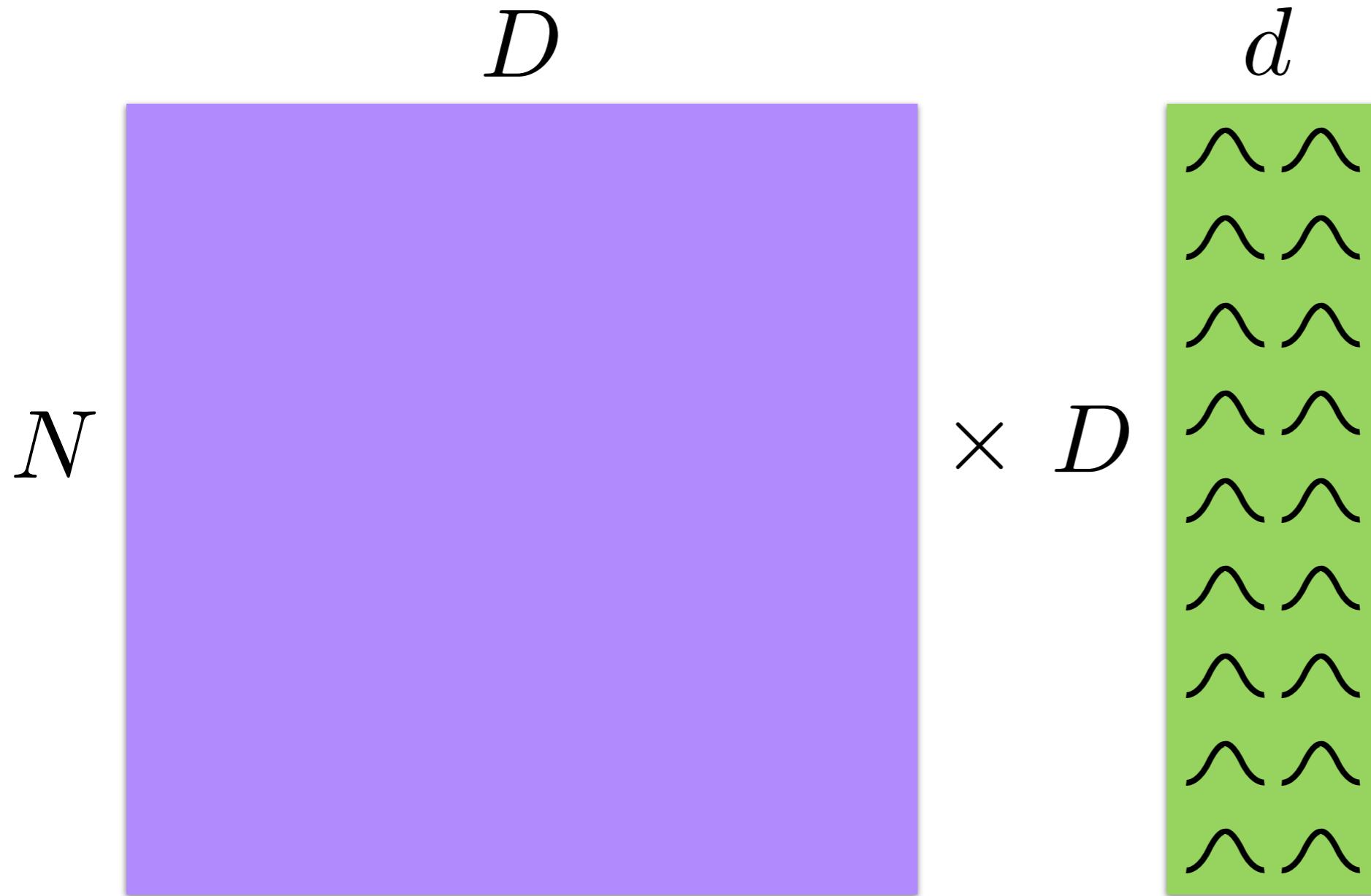
PROPOSED WORK

PROPOSED WORK

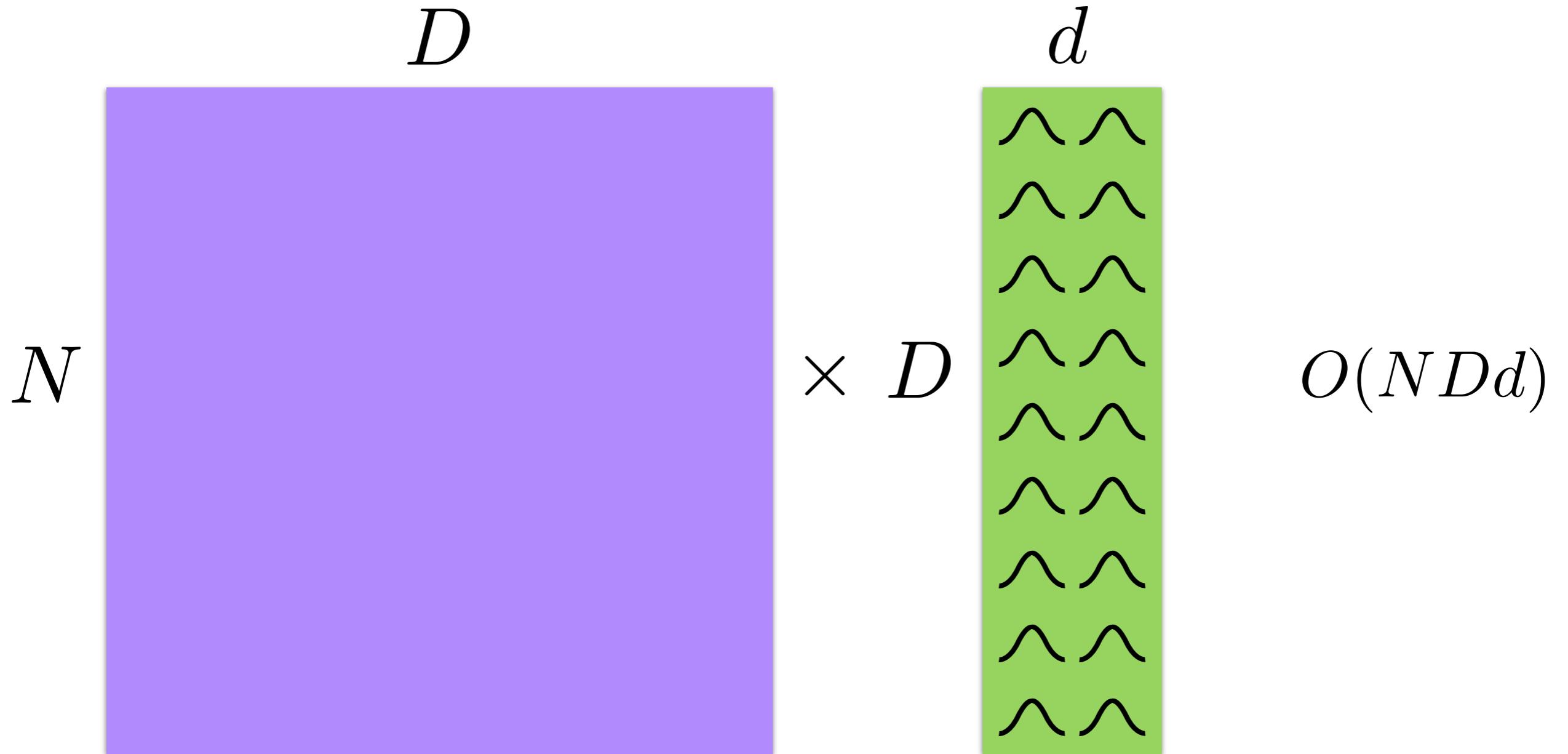
- Survey algorithms for large-scale eigendecomps

LARGE-SCALE EIGENDECOMP

LARGE-SCALE EIGENDECOMP

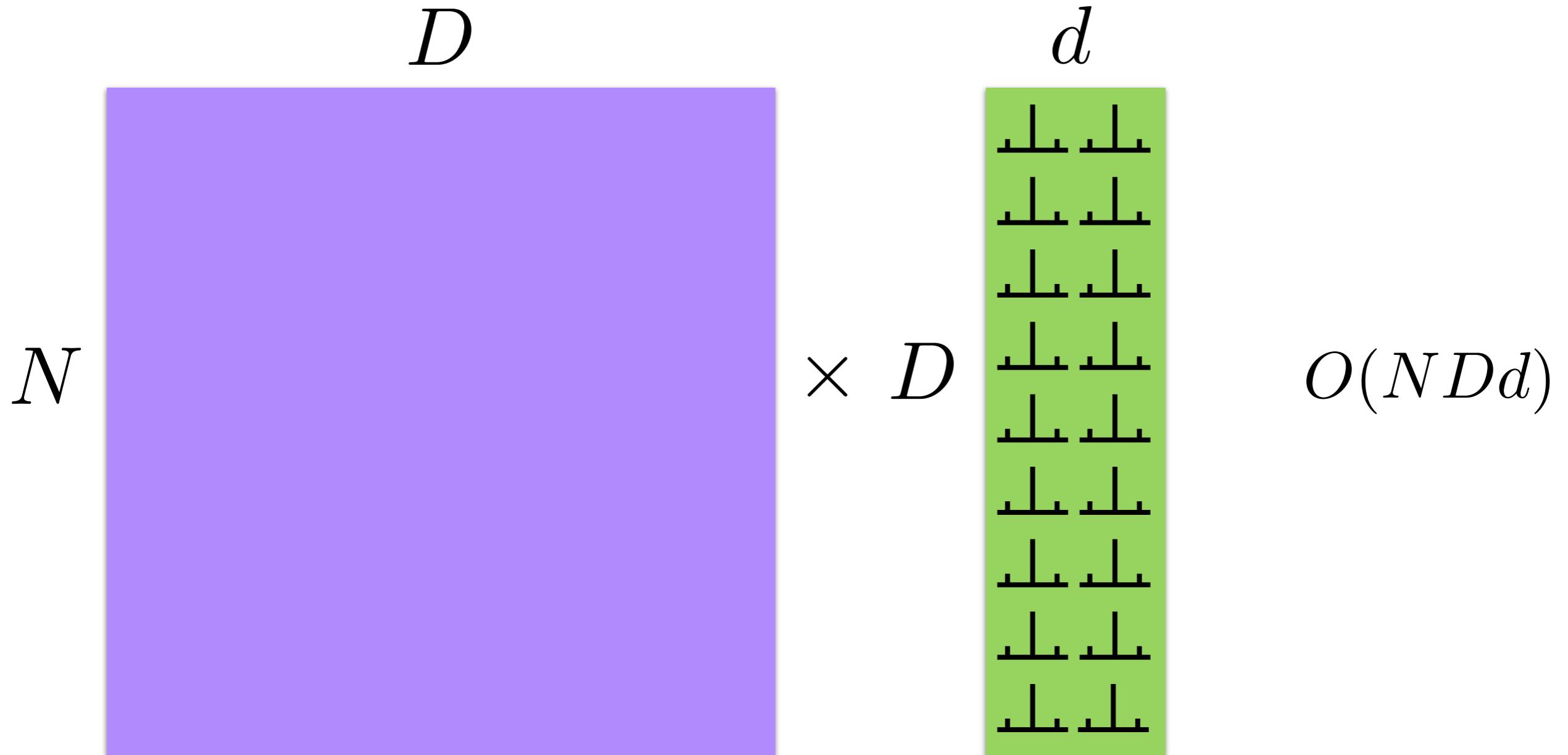


LARGE-SCALE EIGENDECOMP



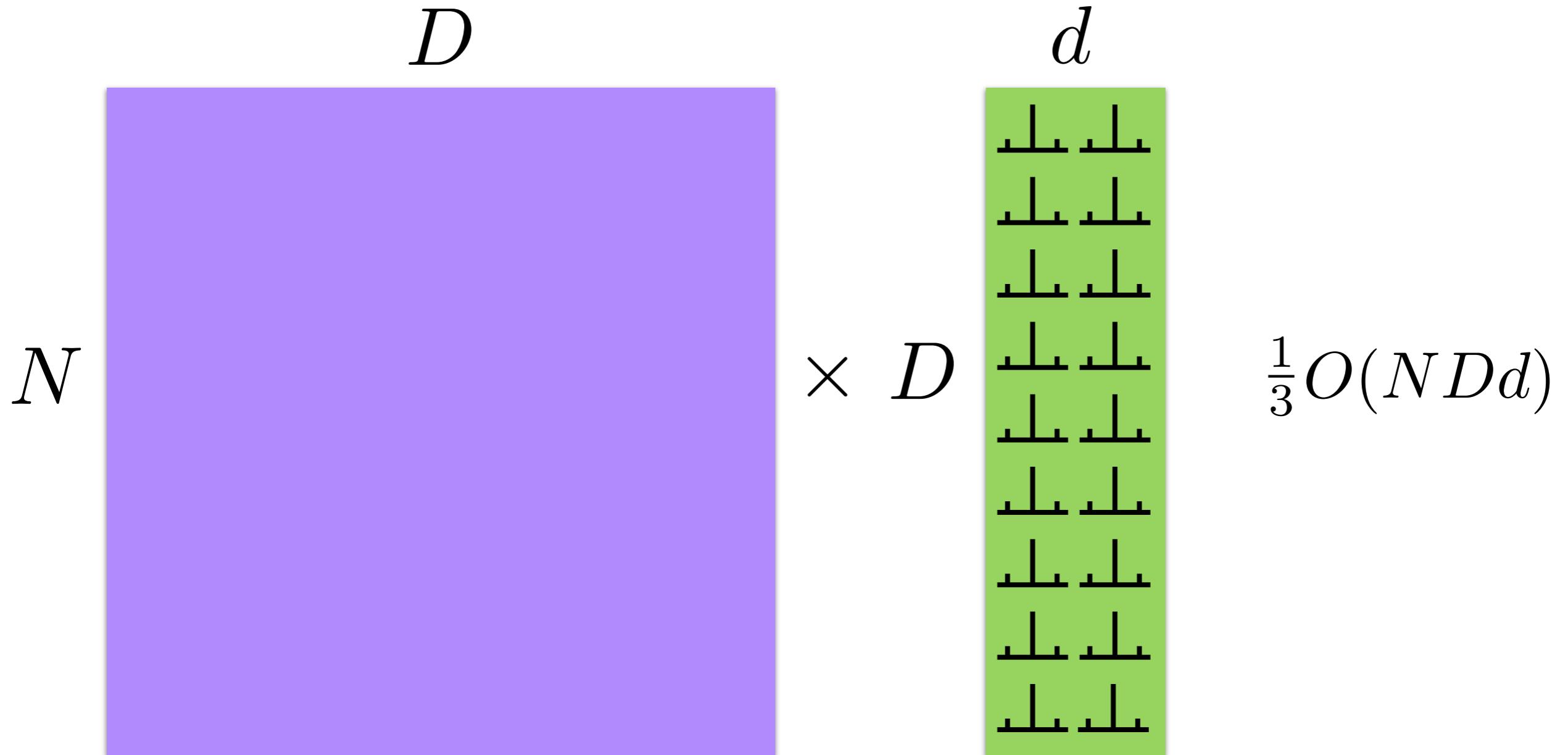
LARGE-SCALE EIGENDECOMP

ACHLIOPTAS (2002)



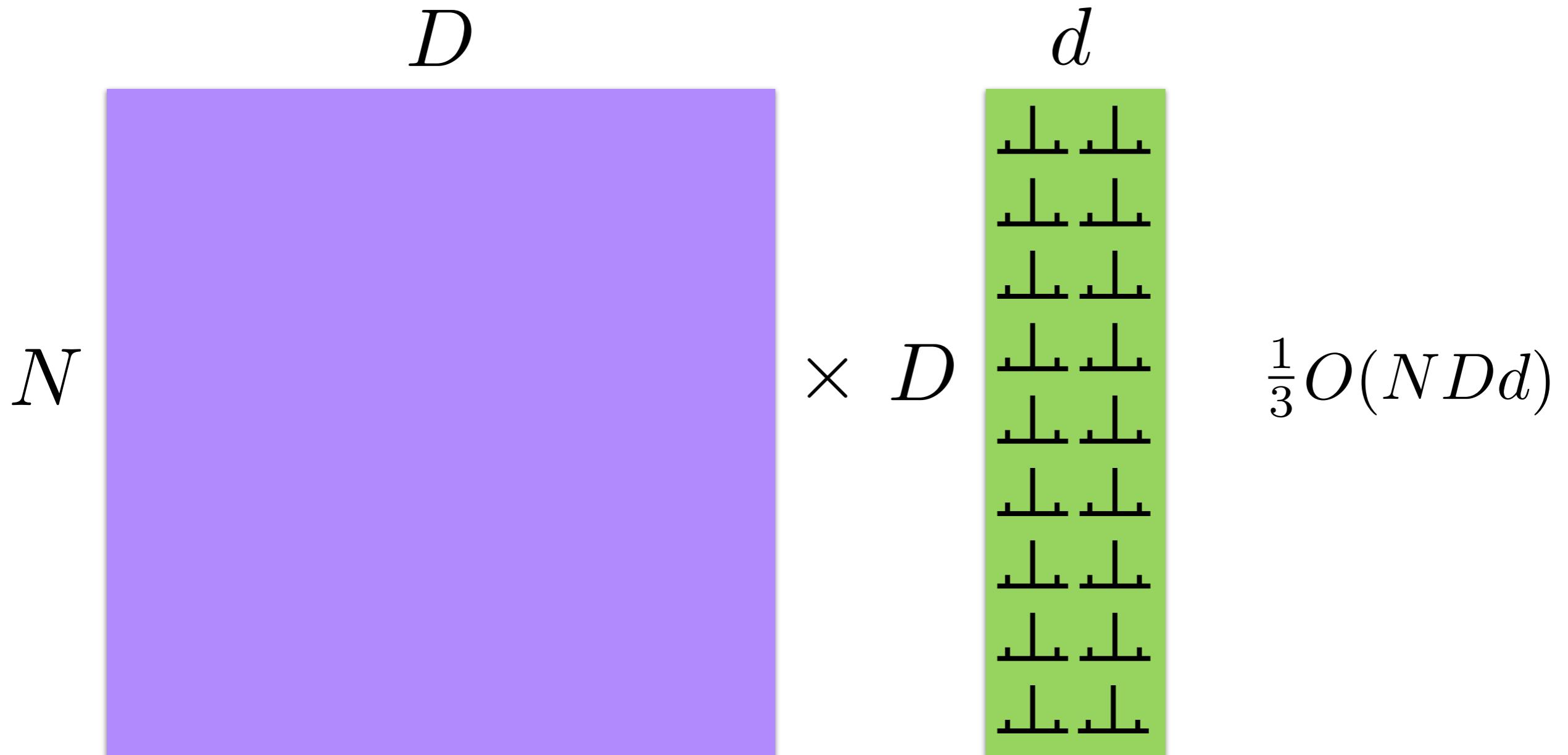
LARGE-SCALE EIGENDECOMP

ACHLIOPTAS (2002)



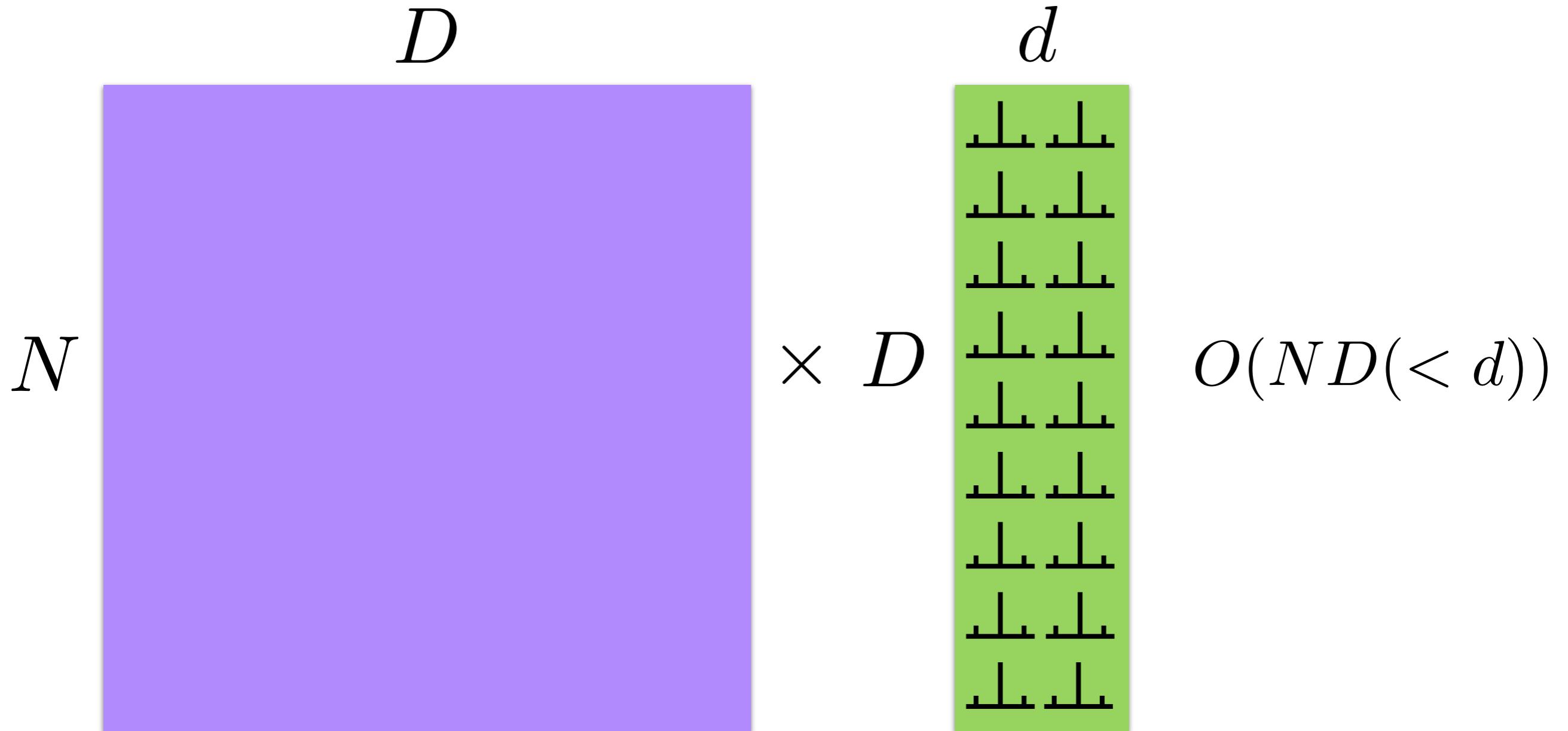
LARGE-SCALE EIGENDECOMP

AILON AND CHAZELLE (2009)



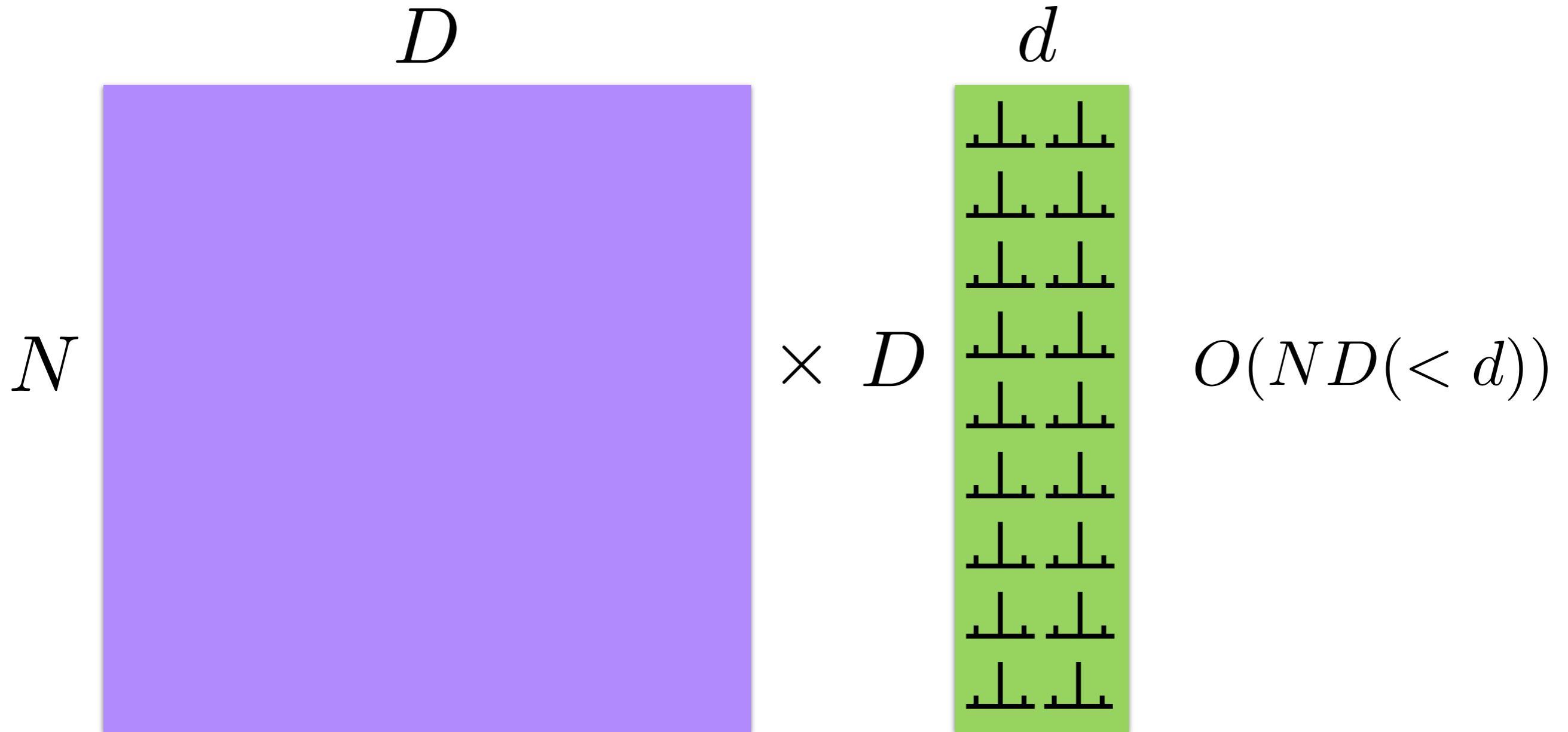
LARGE-SCALE EIGENDECOMP

AILON AND CHAZELLE (2009)



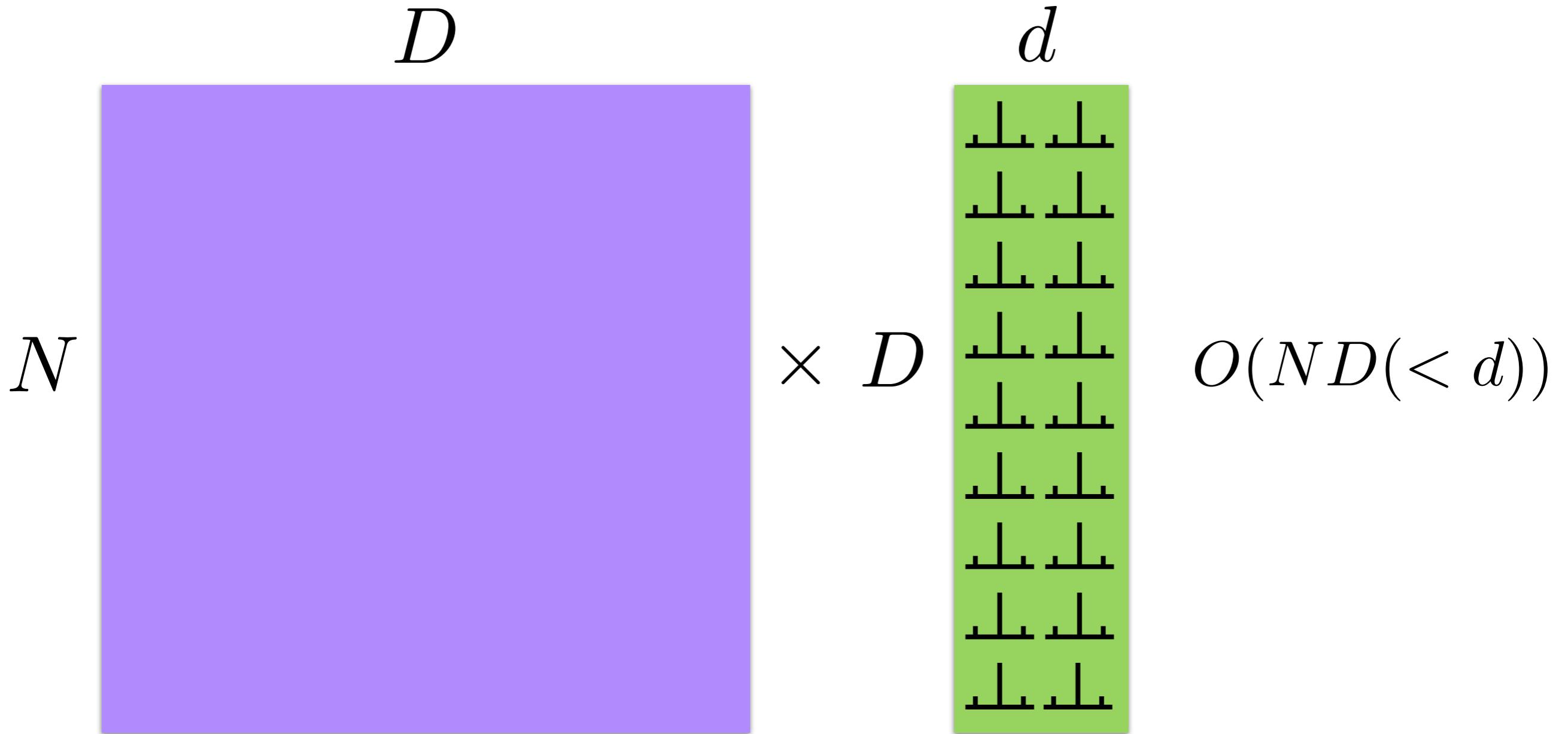
LARGE-SCALE EIGENDECOMP

MAGEN AND ZOUZIAS (2008)



LARGE-SCALE EIGENDECOMP

MAGEN AND ZOUZIAS (2008)



Take-away message for these other random projection methods:

- I) no volume-preservation guarantees, and 2) runtime will be dominated by DPP sampling anyway.

PROPOSED WORK

PROPOSED WORK

- Survey algorithms for large-scale eigendecomps

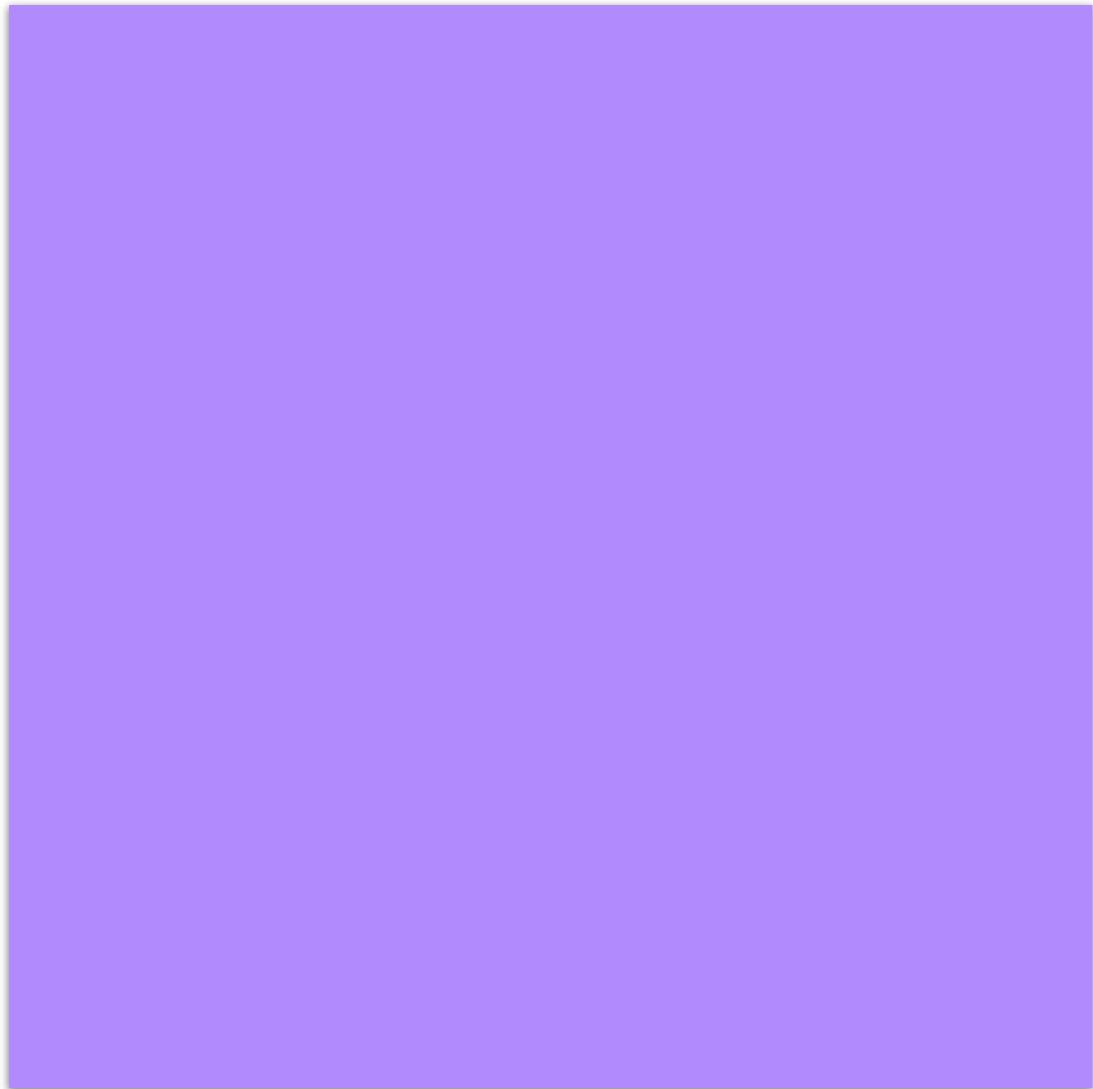
PROPOSED WORK

- Survey algorithms for large-scale eigendecomps
- Try to extend Nyström approximation from Affandi et al. (AIStats 2013) to structured DPPs

NYSTRÖM APPROX

D

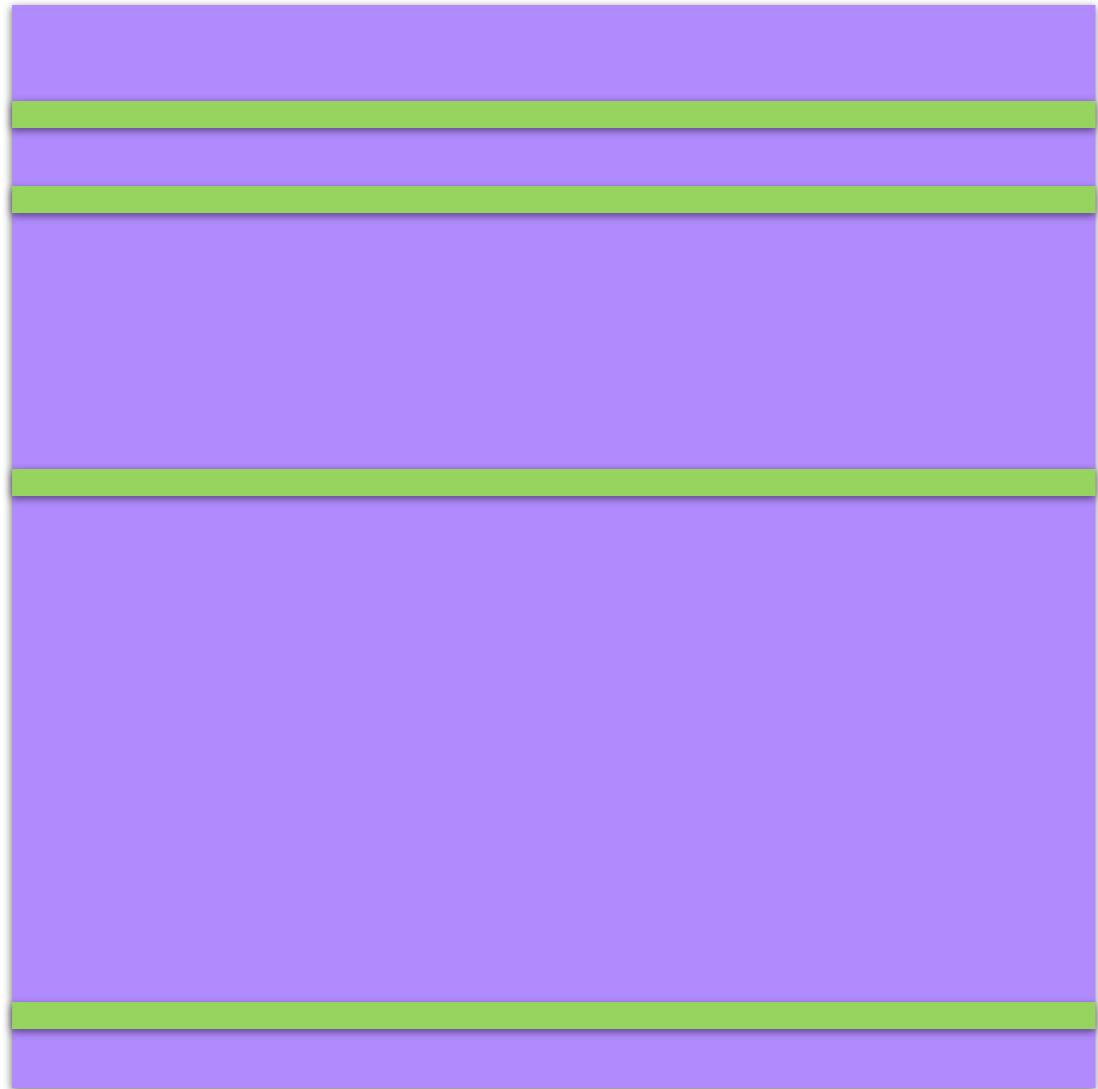
N



NYSTRÖM APPROX

D

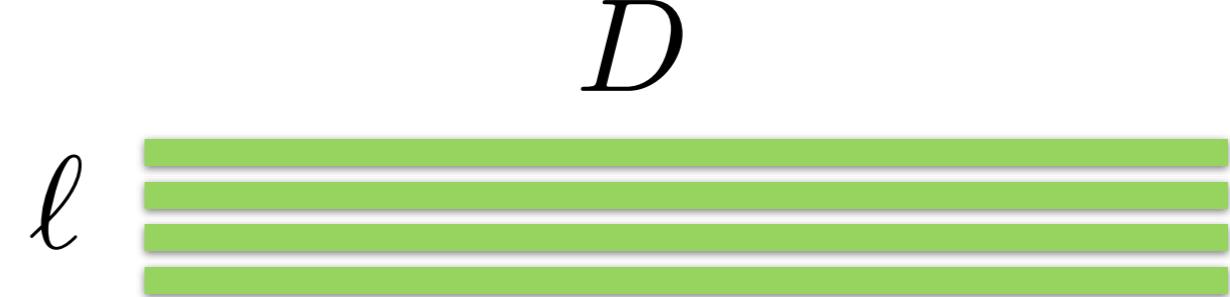
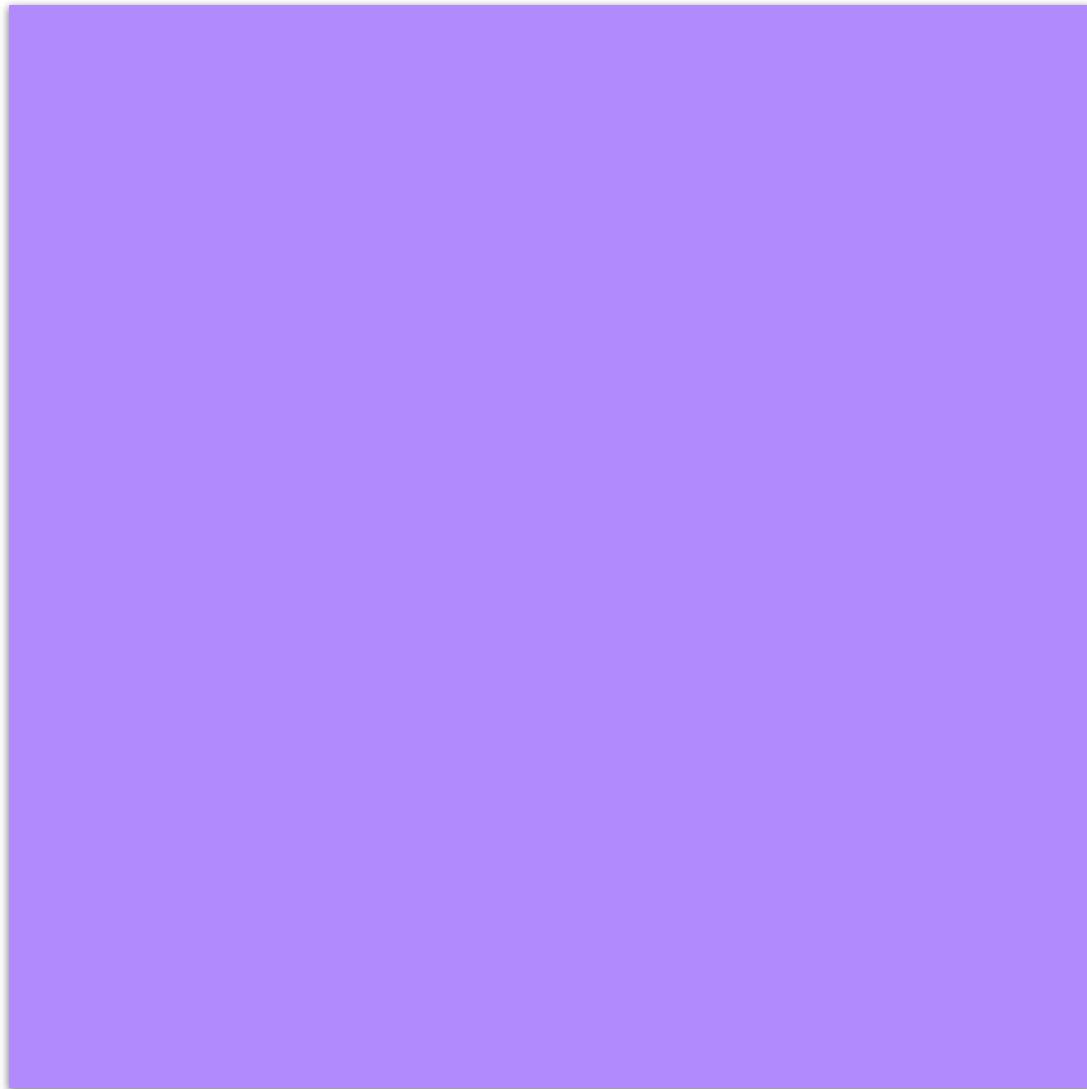
N



NYSTRÖM APPROX

D

N

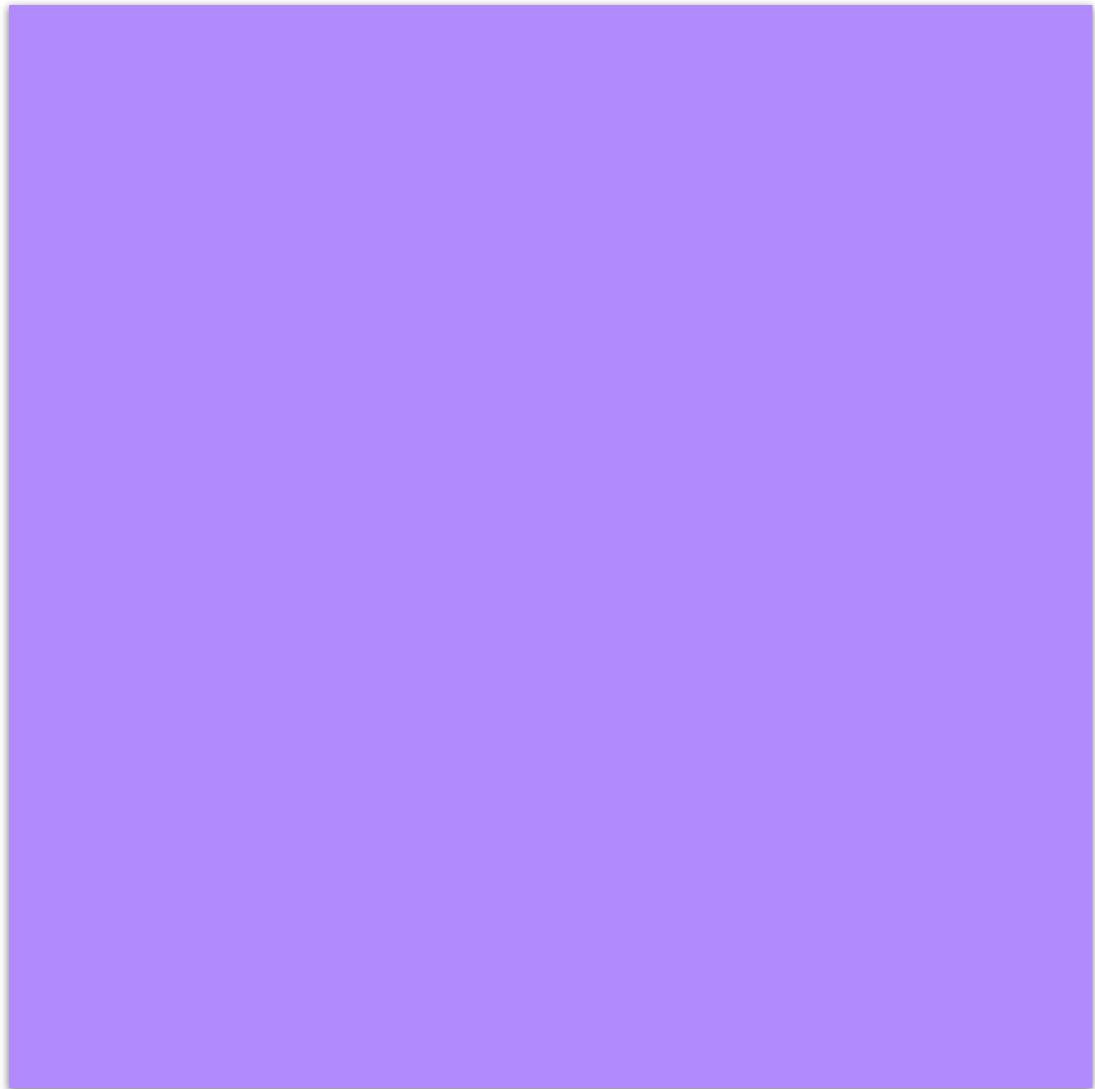


NYSTRÖM APPROX

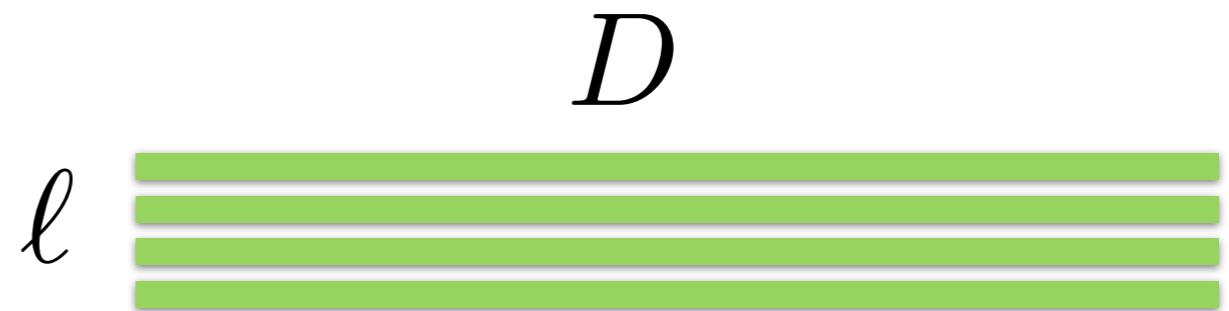
AFFANDI, KULESZA, FOX, AND TASKAR (AISTATS 2013)

D

N



$|\mathcal{P}(Y) - \tilde{\mathcal{P}}(Y)|$ bounded

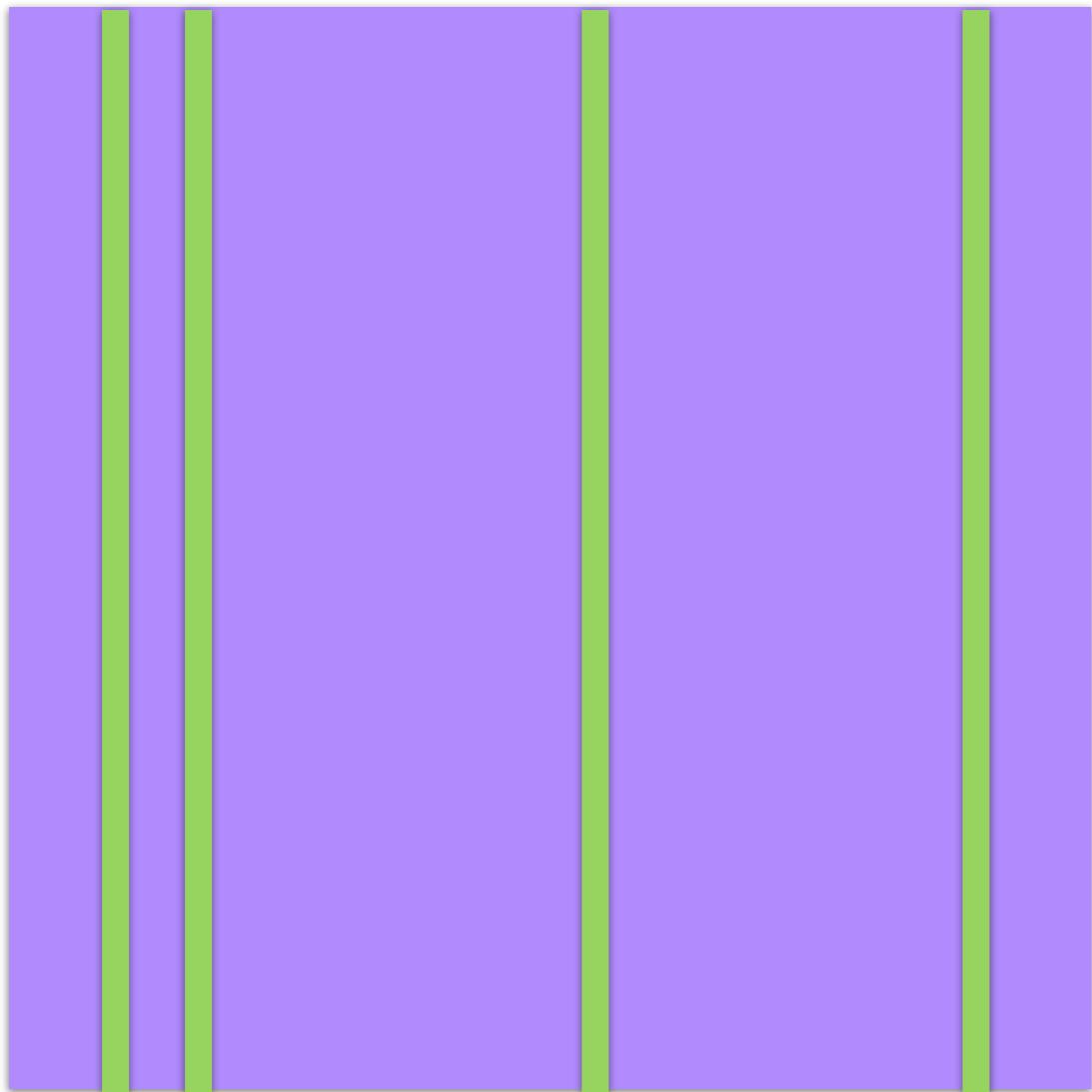


NYSTRÖM APPROX

AFFANDI, KULESZA, FOX, AND TASKAR (AISTATS 2013)

D

N



NYSTRÖM APPROX

AFFANDI, KULESZA, FOX, AND TASKAR (AISTATS 2013)

D

N

d

N



NYSTRÖM APPROX

d

N



NYSTRÖM APPROX

$$\tilde{C} = \begin{pmatrix} C_W & C_{W,\overline{W}} \\ C_{\overline{W},W} & C_{\overline{W},W} C_W^+ C_{W,\overline{W}} \end{pmatrix}$$

d



N

W

NYSTRÖM APPROX

d


$$\tilde{C} = \begin{pmatrix} C_W & C_{W,\overline{W}} \\ C_{\overline{W},W} & C_{\overline{W},W} C_W^+ C_{W,\overline{W}} \end{pmatrix}$$

N

W

$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

NYSTRÖM APPROX

d

 W

$$\tilde{C} = \begin{pmatrix} C_W & C_{W,\overline{W}} \\ C_{\overline{W},W} & C_{\overline{W},W} C_W^+ C_{W,\overline{W}} \end{pmatrix}$$

$$N \quad B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

$$\tilde{C} = \tilde{V} \tilde{\Lambda} \tilde{V}^\top$$

$$L \approx B^\top \tilde{V} \tilde{V}^\top B$$

PROPOSED WORK

PROPOSED WORK

- Survey algorithms for large-scale eigendecomps
- Try to extend Nyström approximation from Affandi et al. (AIStats 2013) to structured DPPs

PROPOSED WORK

- Survey algorithms for large-scale eigendecomps
- Try to extend Nyström approximation from Affandi et al. (AIStats 2013) to structured DPPs
- Try to establish mixing rates for MCMC sampling algorithms from Kang (NIPS 2013)

PROPOSED WORK

- Survey algorithms for large-scale eigendecomps
- Try to extend Nyström approximation from Affandi et al. (AIStats 2013) to structured DPPs
- Try to establish mixing rates for MCMC sampling algorithms from Kang (NIPS 2013)

$$O(k^3 \log(k/\epsilon))$$

PROPOSED WORK

- Survey algorithms for large-scale eigendecomps
- Try to extend Nyström approximation from Affandi et al. (AIStats 2013) to structured DPPs
- Try to establish mixing rates for MCMC sampling algorithms from Kang (NIPS 2013)

$$O(k^3 \log(\frac{1}{\epsilon}))$$

mixing time $\rightarrow \infty$ as eigenvalues of $L \rightarrow 0$

PROPOSED WORK

- Survey algorithms for large-scale eigendecomps
- Try to extend Nyström approximation from Affandi et al. (AIStats 2013) to structured DPPs
- Try to establish mixing rates for MCMC sampling algorithms from Kang (NIPS 2013)

$$O(k^3 \log(\frac{1}{\epsilon}))$$

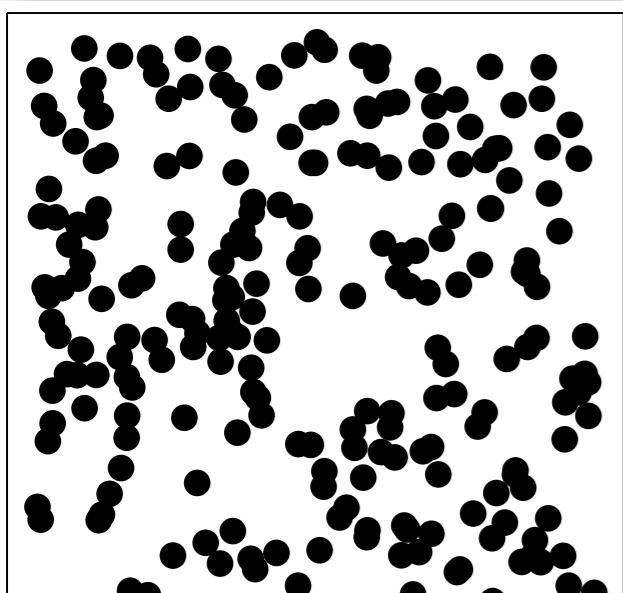
mixing time $\rightarrow \infty$ as eigenvalues of $L \rightarrow 0$

condition number bounded: $\frac{\lambda_{\max}}{\lambda_{\min}} \leq c$
 \implies mixing time bounded?

2. MAP ESTIMATION

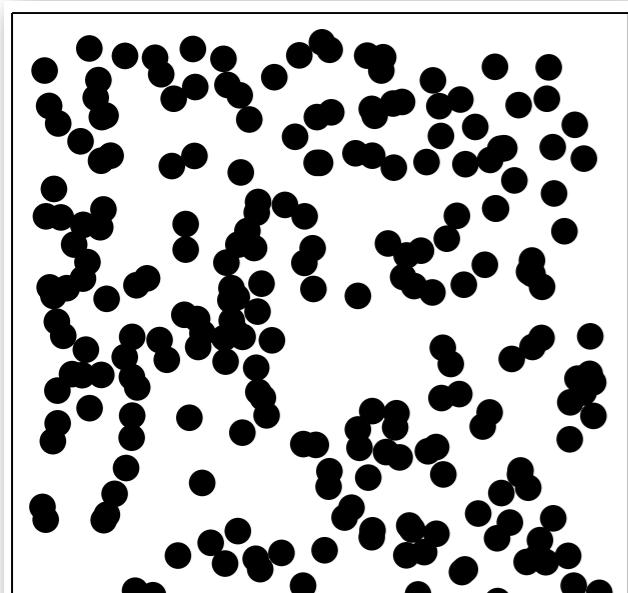
HARDNESS OF MAP

HARDNESS OF MAP

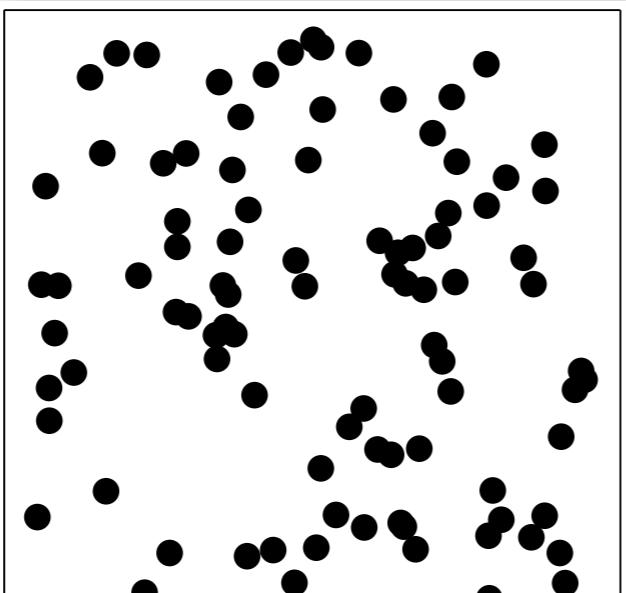


All points

HARDNESS OF MAP

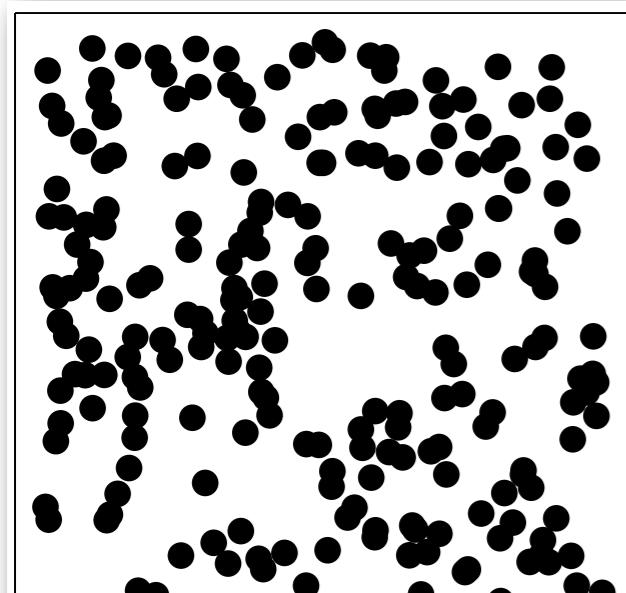


All points

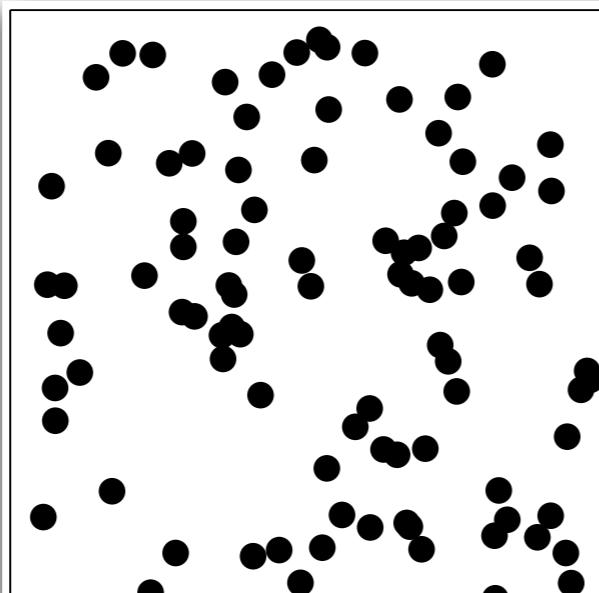


Independent sample

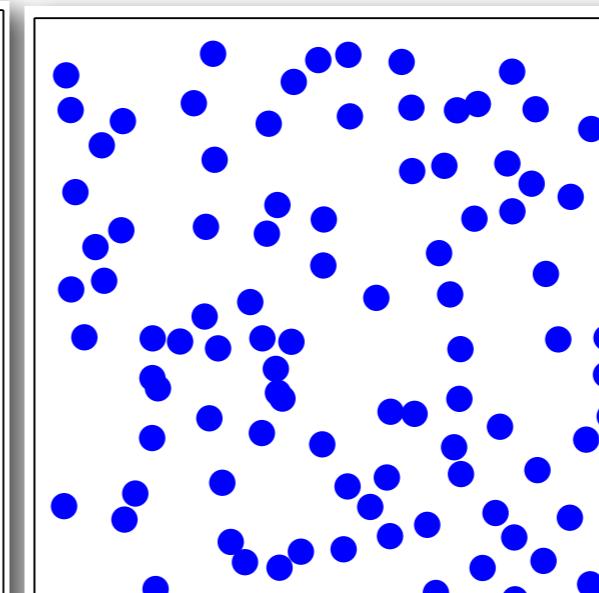
HARDNESS OF MAP



All points

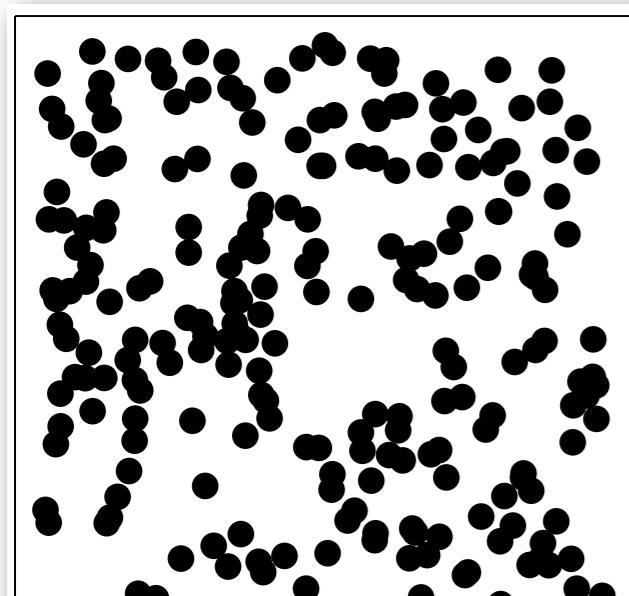


Independent sample

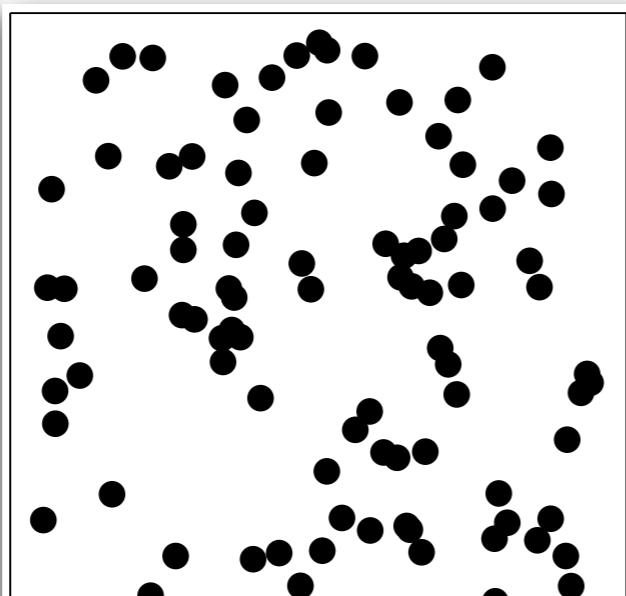


DPP sample

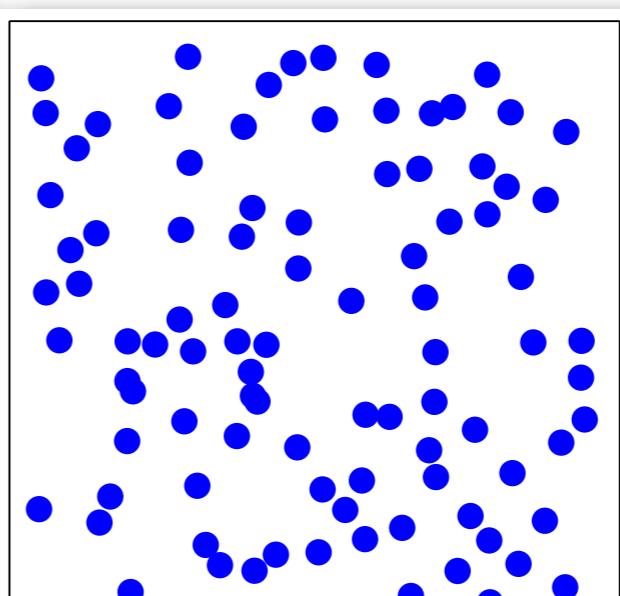
HARDNESS OF MAP



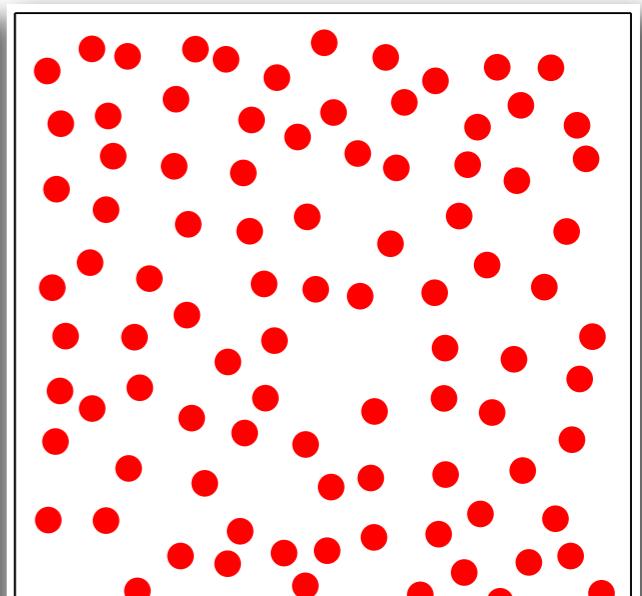
All points



Independent sample

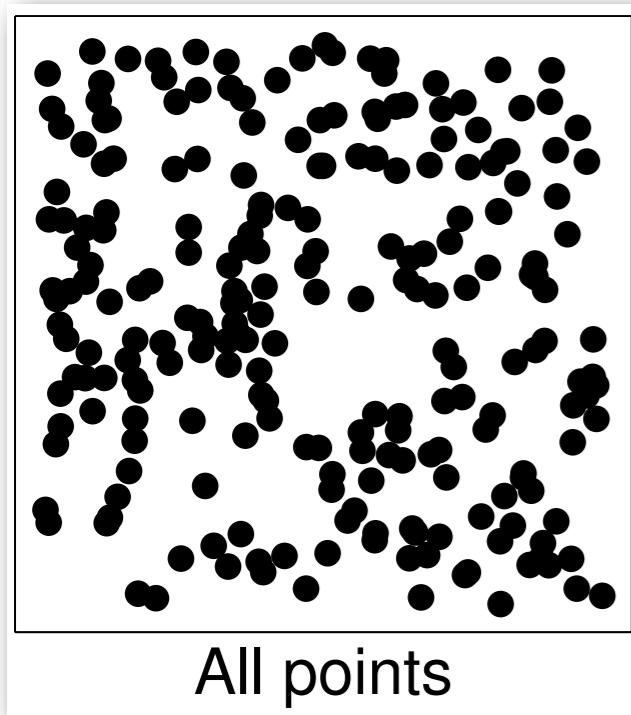


DPP sample

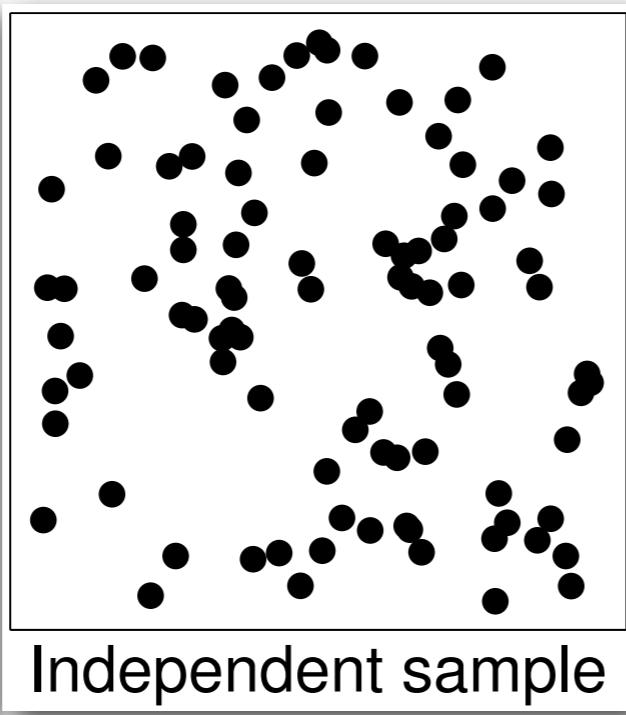


DPP (approx) MAP

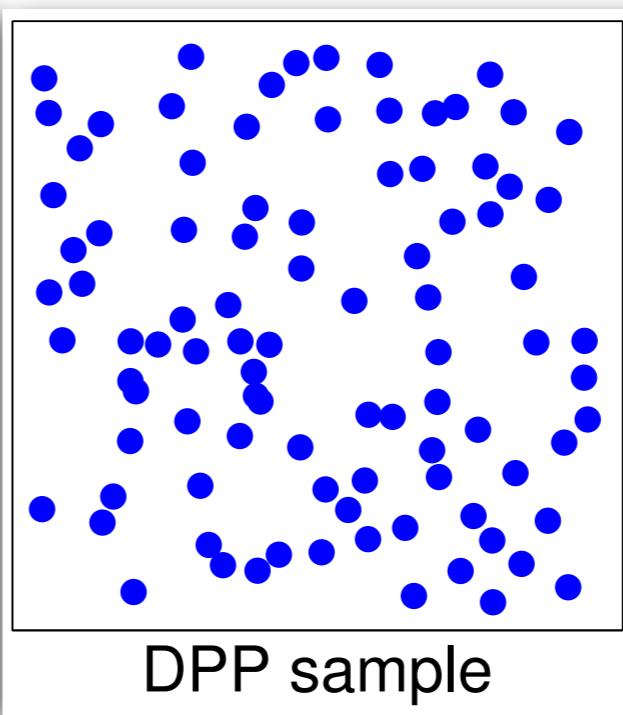
HARDNESS OF MAP



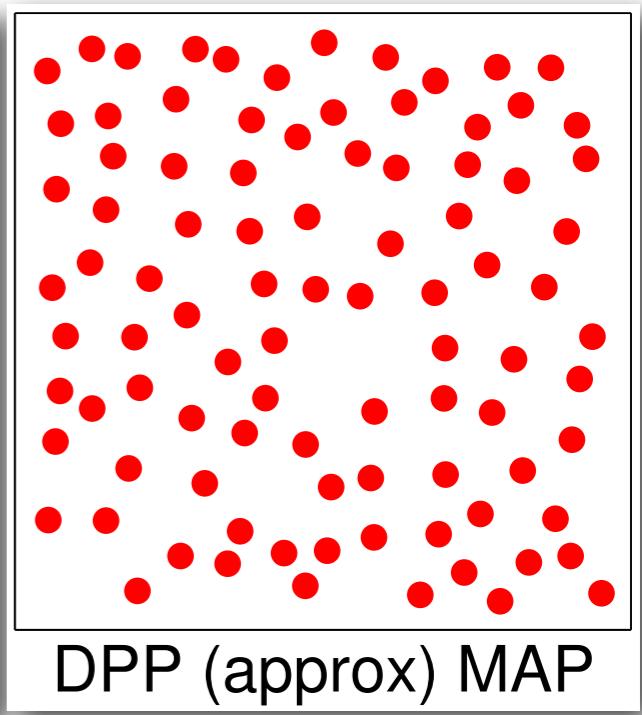
All points



Independent sample



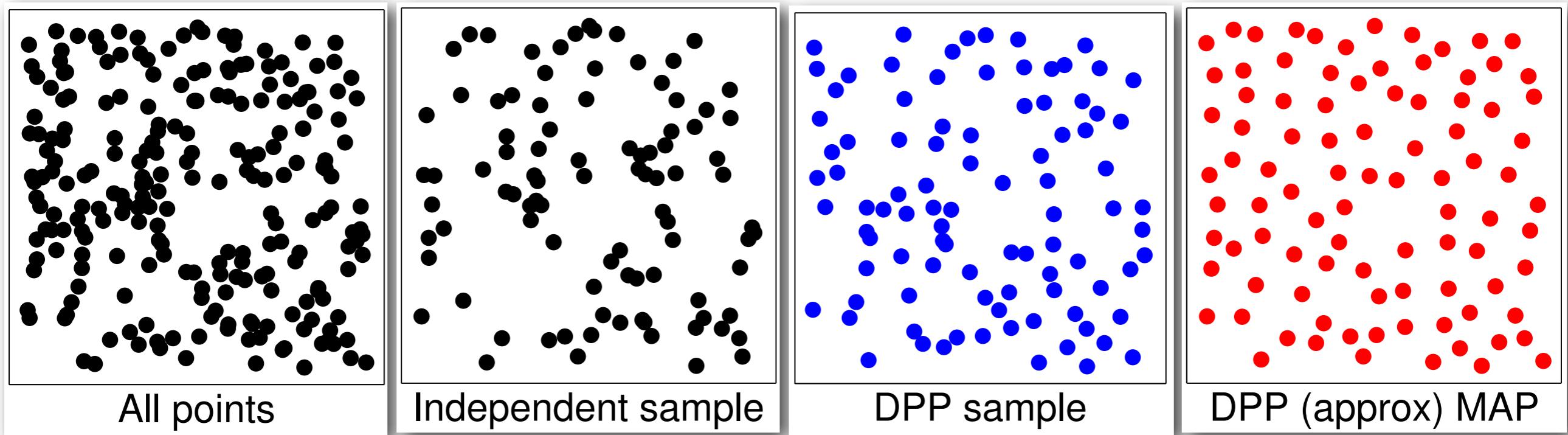
DPP sample



DPP (approx) MAP

$$Y = \arg \max_{Y': Y' \subseteq \mathcal{Y}} \det(L_{Y'})$$

HARDNESS OF MAP

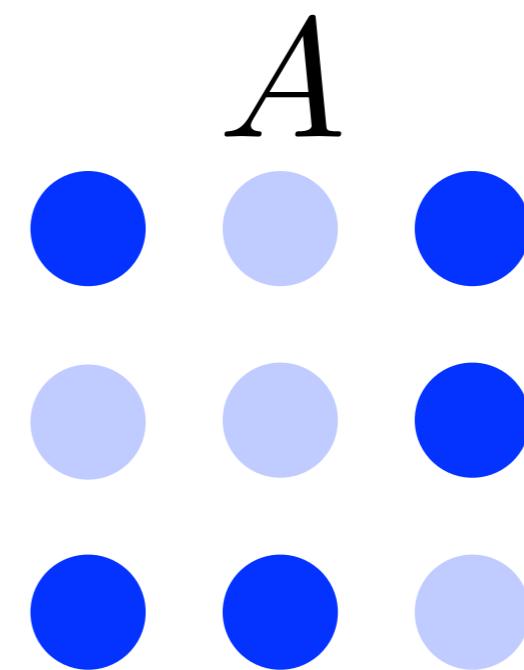


$$Y = \arg \max_{Y': Y' \subseteq \mathcal{Y}} \det(L_{Y'})$$

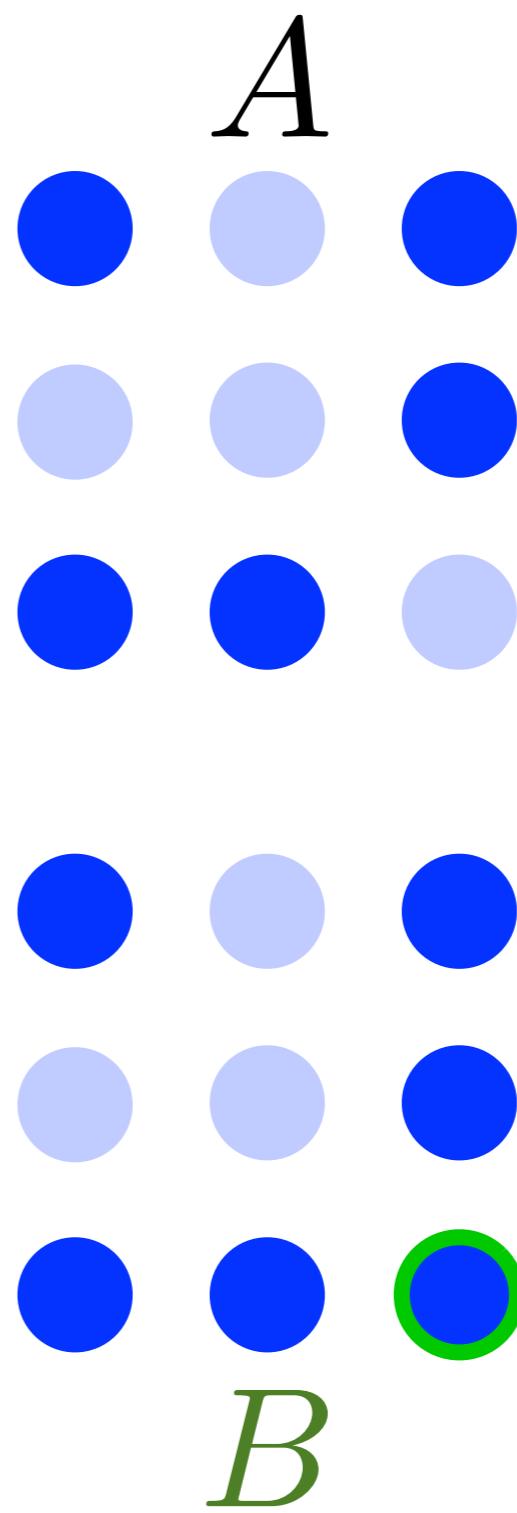
NP-hard, no PTAS: $\det(L_{\hat{Y}}) \geq \left(\frac{8}{9} + \epsilon\right) \det(L_{Y^*})$

SUBMODULARITY

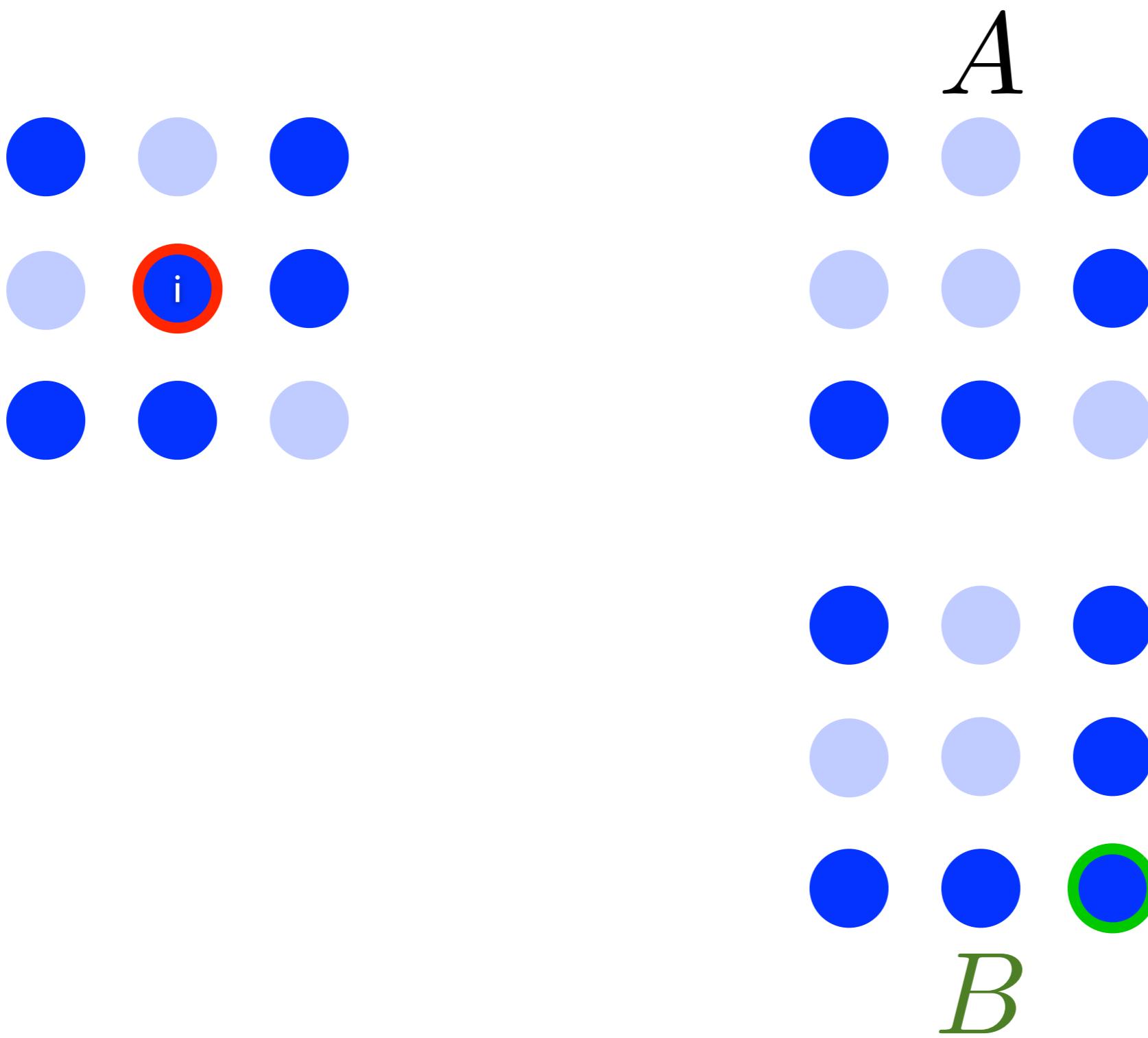
SUBMODULARITY



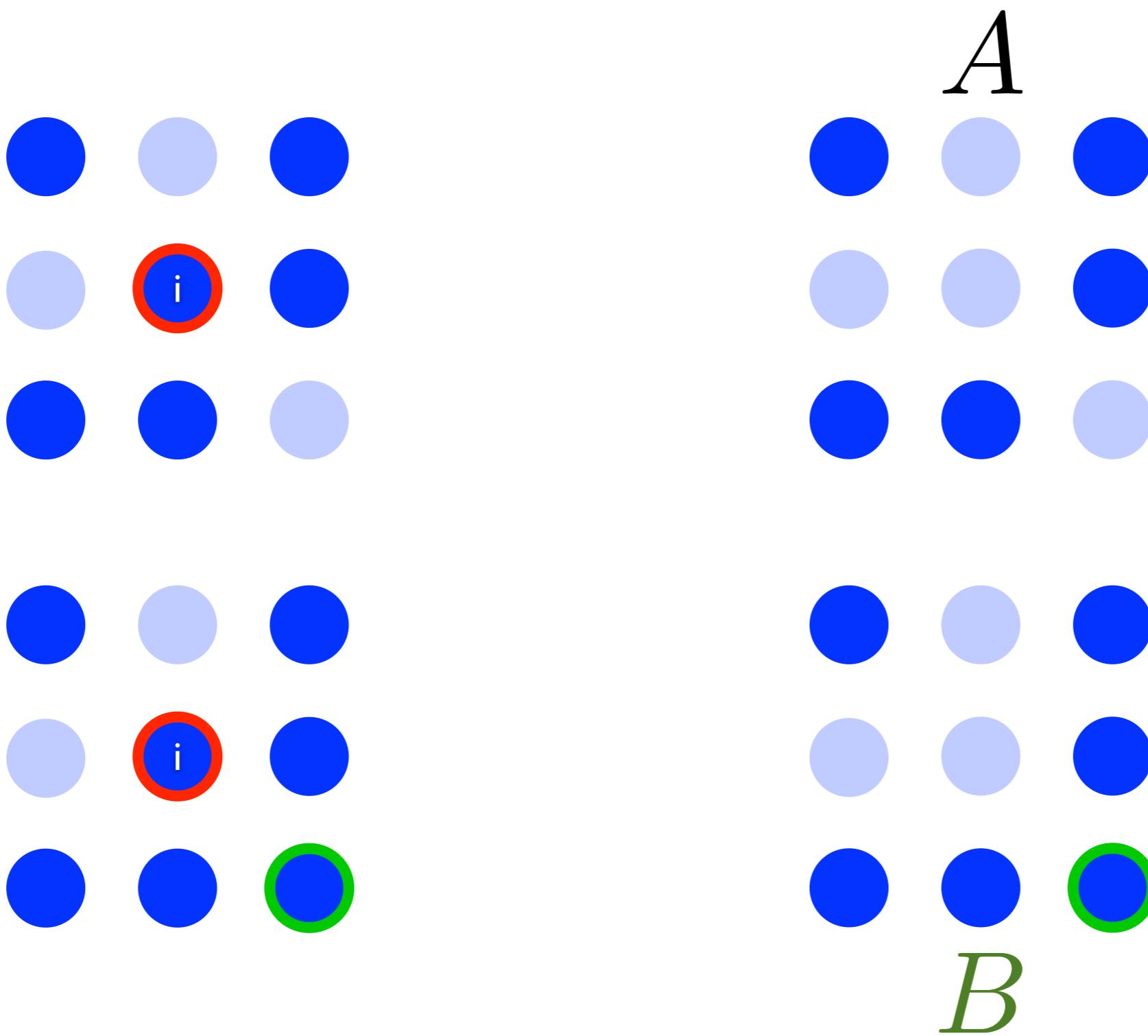
SUBMODULARITY



SUBMODULARITY



SUBMODULARITY



SUBMODULARITY

$$f\left(\begin{array}{c} \text{blue} \\ \text{light blue} \\ \text{light blue} \\ \text{red circle labeled } i \\ \text{blue} \end{array}\right) - f\left(\begin{array}{c} \text{blue} \\ \text{light blue} \\ \text{light blue} \\ \text{light blue} \\ \text{blue} \end{array}\right) \geq$$

A

$$f\left(\begin{array}{c} \text{blue} \\ \text{light blue} \\ \text{light blue} \\ \text{red circle labeled } i \\ \text{blue} \end{array}\right) - f\left(\begin{array}{c} \text{blue} \\ \text{light blue} \\ \text{light blue} \\ \text{light blue} \\ \text{green circle highlighted} \end{array}\right)$$

B

LOG-SUBMODULARITY

$$f\left(\begin{array}{ccc} \textcolor{blue}{\bullet} & \textcolor{lightblue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{lightblue}{\bullet} & \textcolor{red}{\text{---}}^{\textcolor{red}{i}} \textcolor{blue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{blue}{\bullet} & \textcolor{blue}{\bullet} & \textcolor{lightblue}{\bullet} \end{array}\right) - f\left(\begin{array}{ccc} \textcolor{blue}{\bullet} & \textcolor{lightblue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{lightblue}{\bullet} & \textcolor{lightblue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{blue}{\bullet} & \textcolor{blue}{\bullet} & \textcolor{lightblue}{\bullet} \end{array}\right) \geq$$

A

$$f\left(\begin{array}{ccc} \textcolor{blue}{\bullet} & \textcolor{lightblue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{lightblue}{\bullet} & \textcolor{red}{\text{---}}^{\textcolor{red}{i}} \textcolor{blue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{blue}{\bullet} & \textcolor{blue}{\bullet} & \textcolor{green}{\bullet} \end{array}\right) - f\left(\begin{array}{ccc} \textcolor{blue}{\bullet} & \textcolor{lightblue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{lightblue}{\bullet} & \textcolor{lightblue}{\bullet} & \textcolor{blue}{\bullet} \\ \textcolor{blue}{\bullet} & \textcolor{blue}{\bullet} & \textcolor{green}{\bullet} \end{array}\right)$$

B

LOG-SUBMODULARITY

$$\frac{f\left(\begin{array}{ccc} \text{blue} & \text{light blue} & \text{blue} \\ \text{light blue} & \text{red circle labeled } i & \text{blue} \\ \text{blue} & \text{blue} & \text{light blue} \end{array}\right)}{f\left(\begin{array}{ccc} \text{blue} & \text{light blue} & \text{blue} \\ \text{light blue} & \text{light blue} & \text{blue} \\ \text{blue} & \text{blue} & \text{light blue} \end{array}\right)} \geq \frac{f\left(\begin{array}{ccc} \text{blue} & \text{light blue} & \text{blue} \\ \text{light blue} & \text{red circle labeled } i & \text{blue} \\ \text{blue} & \text{blue} & \text{green circle} \end{array}\right)}{f\left(\begin{array}{ccc} \text{blue} & \text{light blue} & \text{blue} \\ \text{light blue} & \text{light blue} & \text{blue} \\ \text{blue} & \text{blue} & \text{green circle} \end{array}\right)}$$

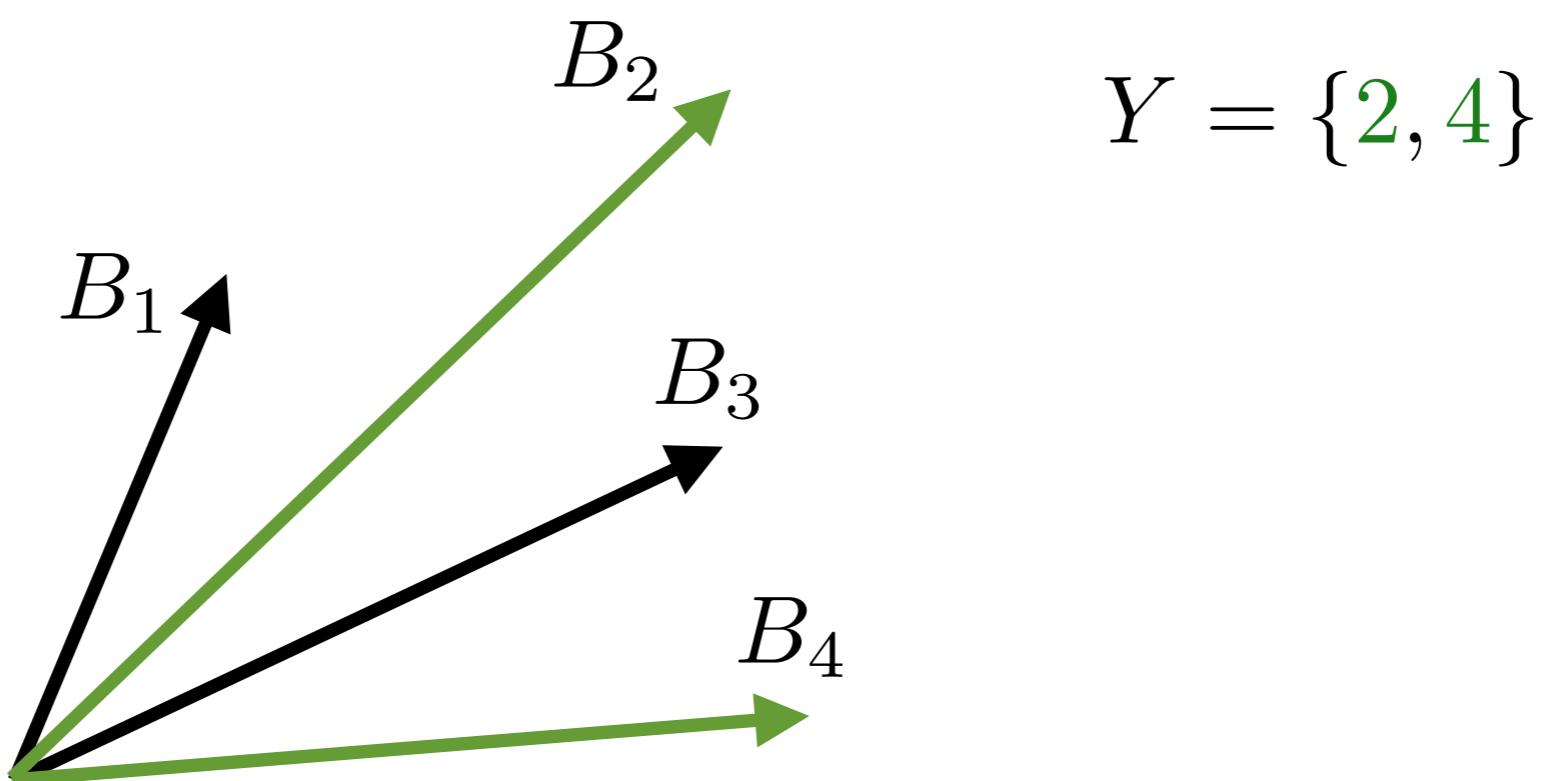
A

B

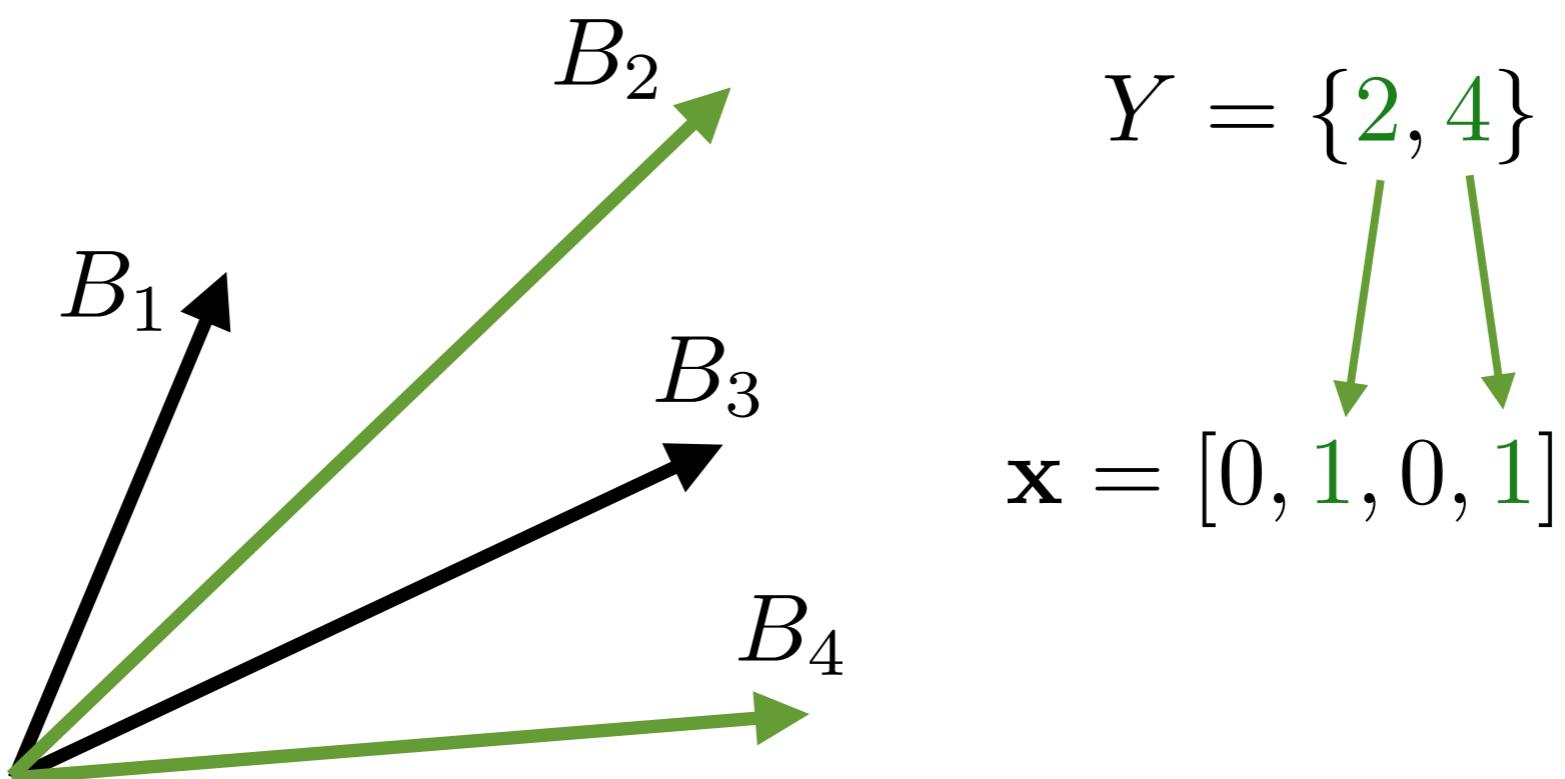
CHEKURI ET AL. 2011

CHEKURI ET AL. 2011

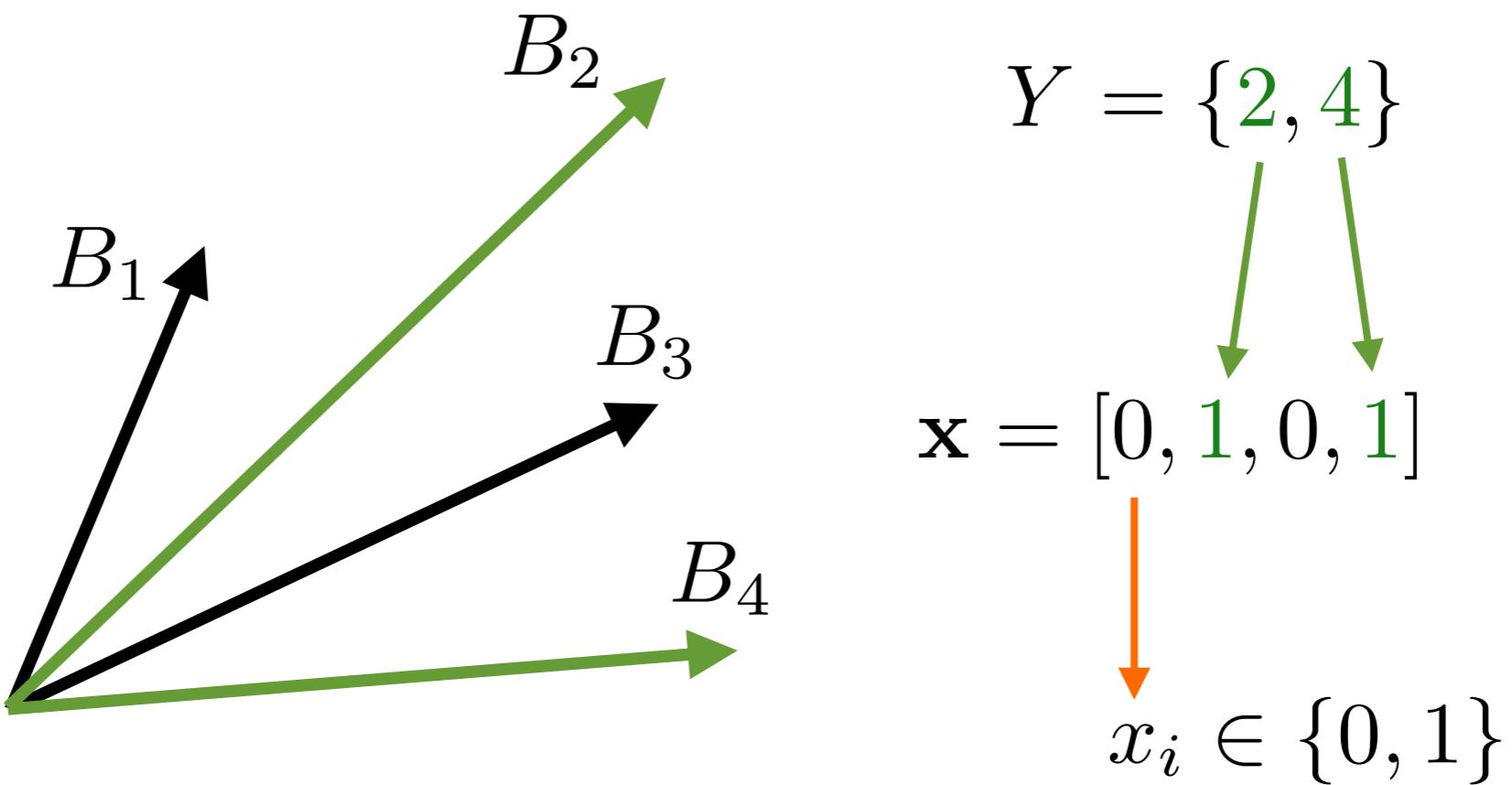
Step 1: Relax inclusion-exclusion



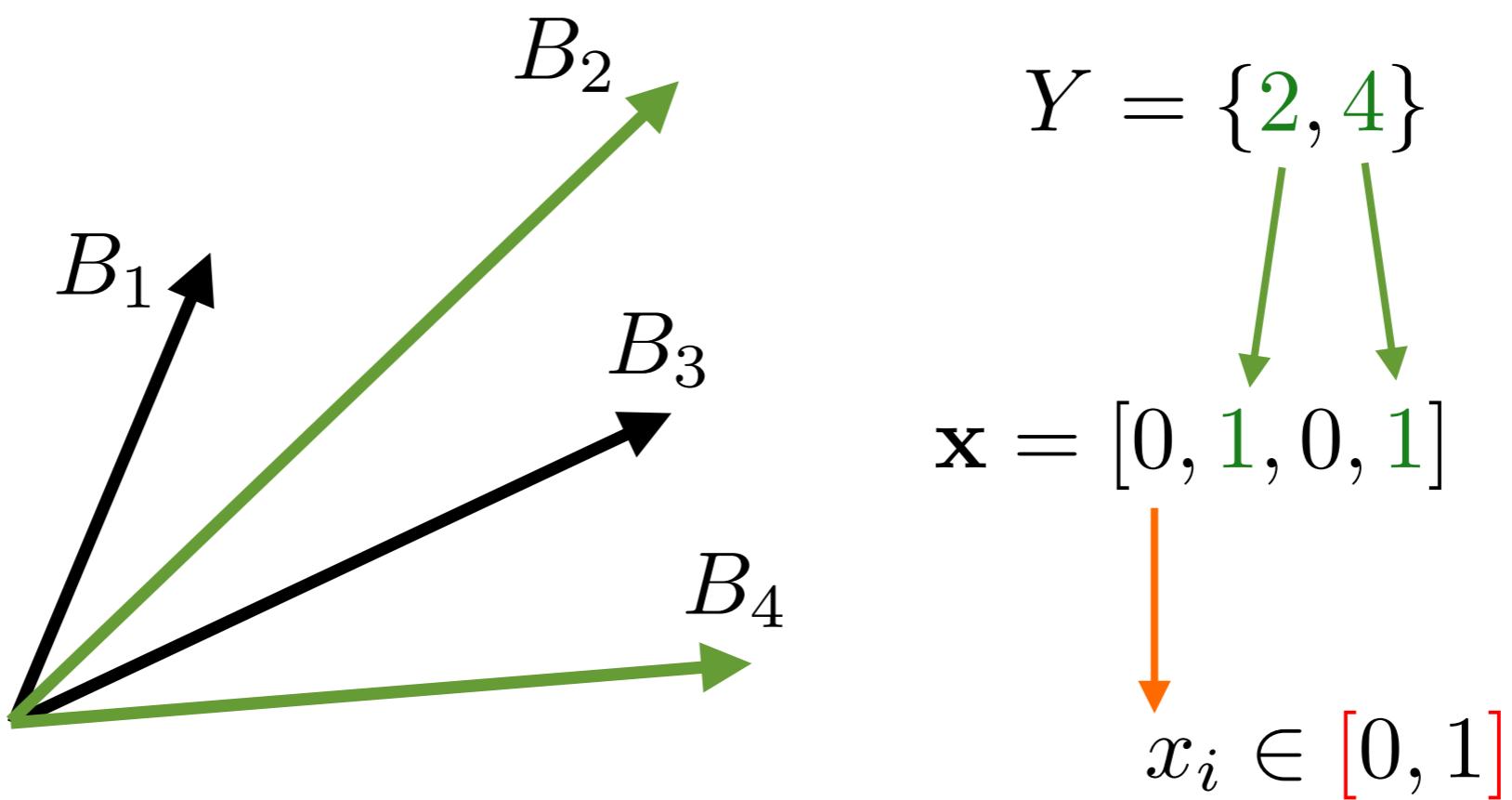
Step 1: Relax inclusion-exclusion



Step 1: Relax inclusion-exclusion



Step 1: Relax inclusion-exclusion



Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension}$$



log-submodular, like $\det(L_Y)$

Step 2: Extend objective

$$\begin{aligned} F(\mathbf{x}) &= E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension} \\ &= \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y) \end{aligned}$$

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension}$$

$$= \sum_Y \left[\prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \right] \log f(Y)$$

$$\downarrow \\ p_Y$$

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension}$$

$$= \boxed{\sum_Y} \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

\downarrow
 2^N subsets

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension}$$

$$= \boxed{\sum_Y} \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

\downarrow
 2^N subsets

Step 3: Optimize using gradient-based methods $\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}}$

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension}$$

$$= \boxed{\sum_Y} \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

\downarrow
 2^N subsets

Step 3: Optimize using gradient-based methods $\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}}$

Step 4: If unconstrained, solution will already be integer;
else, round solution: $x_i \in [0, 1] \rightarrow x_i \in \{0, 1\}$

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)] \quad \text{multilinear extension}$$

$$= \boxed{\sum_Y} \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

\downarrow
 2^N subsets \implies Monte Carlo required

Step 3: Optimize using gradient-based methods $\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}}$

Step 4: If unconstrained, solution will already be integer;
else, round solution: $x_i \in [0, 1] \rightarrow x_i \in \{0, 1\}$

SOFTMAX EXTENSION

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

Multilinear: $F(\mathbf{x}) = \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$

SOFTMAX EXTENSION

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

Multilinear: $F(\mathbf{x}) = \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$

Softmax: $\tilde{F}(\mathbf{x}) = \log \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) f(Y)$

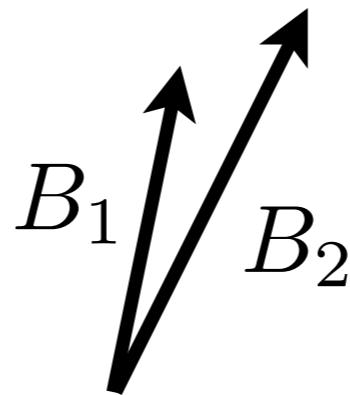
SOFTMAX EXTENSION

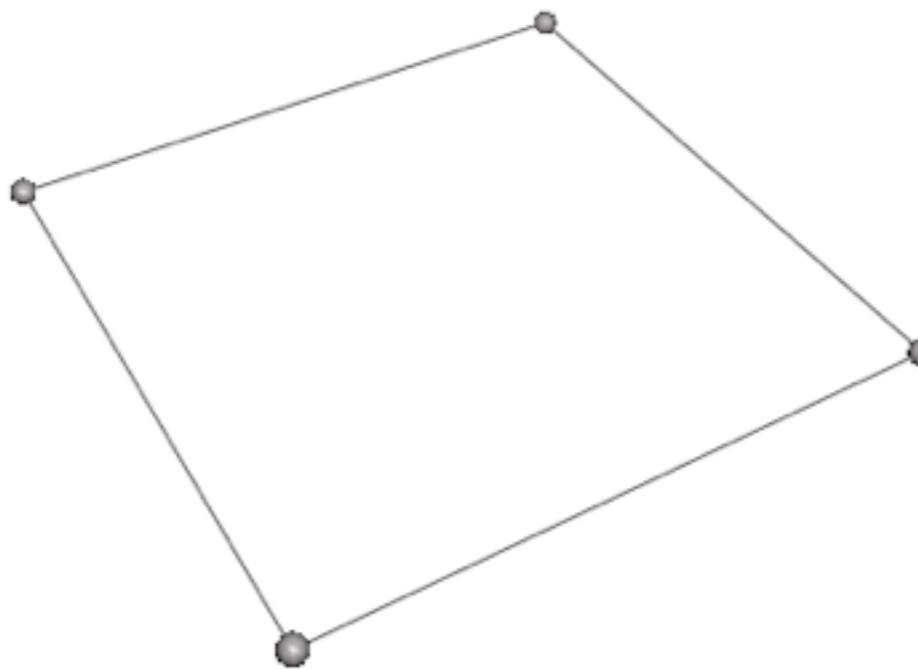
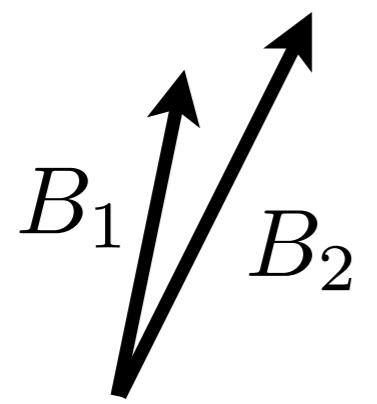
GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

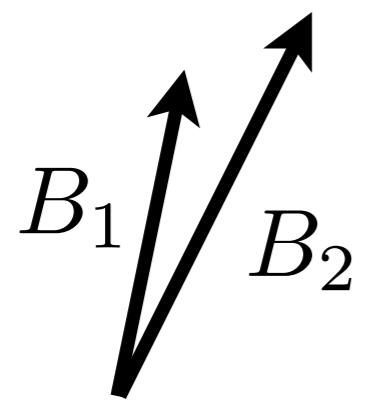
Multilinear: $F(\mathbf{x}) = \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$

Softmax: $\tilde{F}(\mathbf{x}) = \log \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) f(Y)$

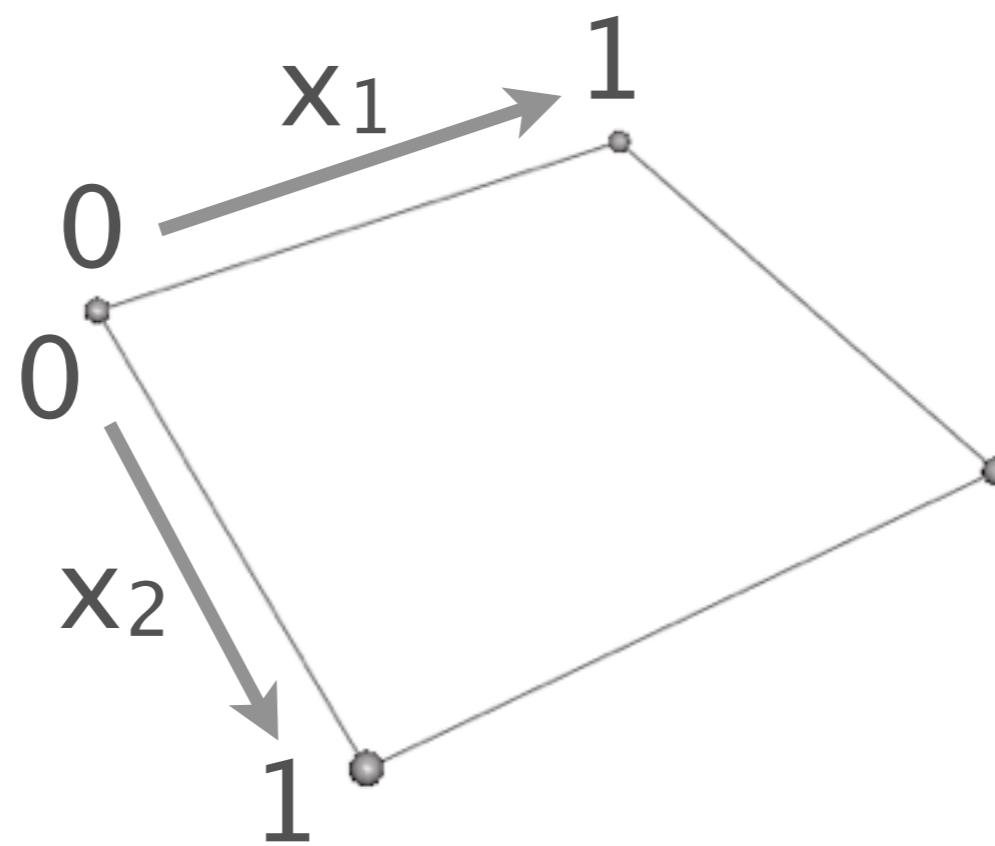
$$N = 2$$

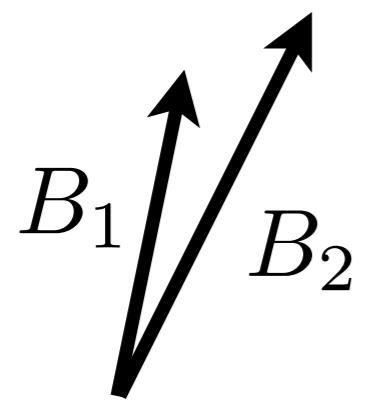




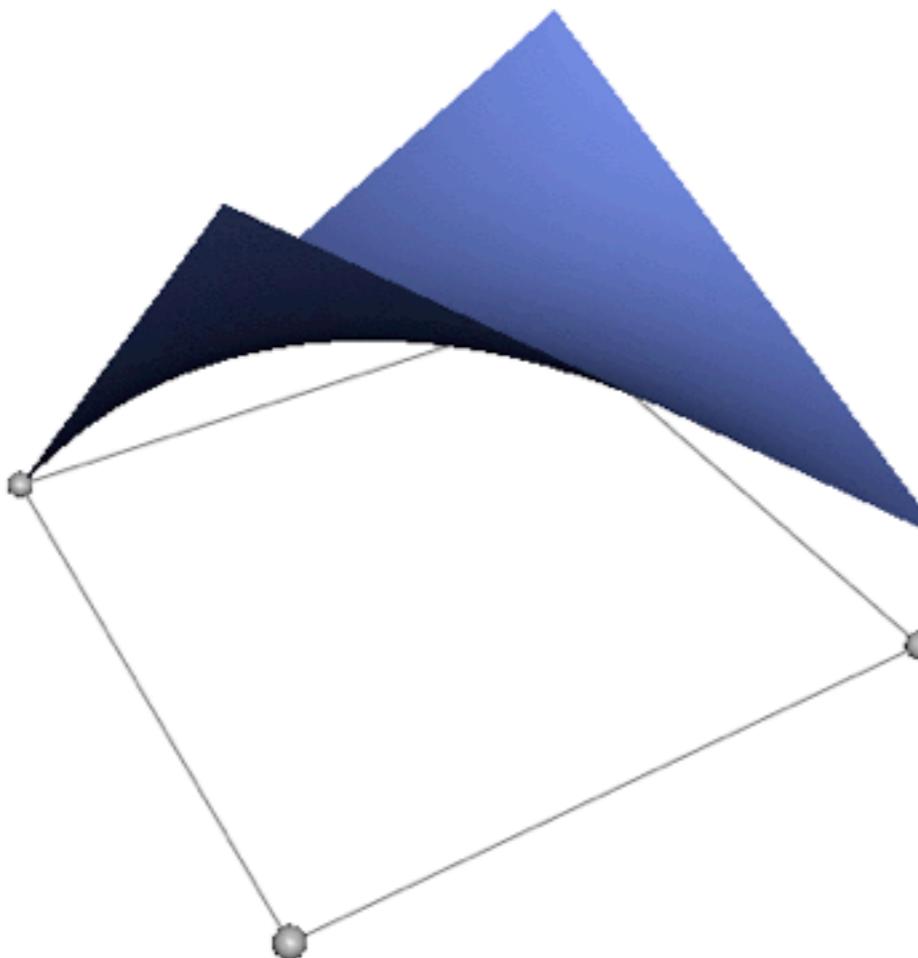


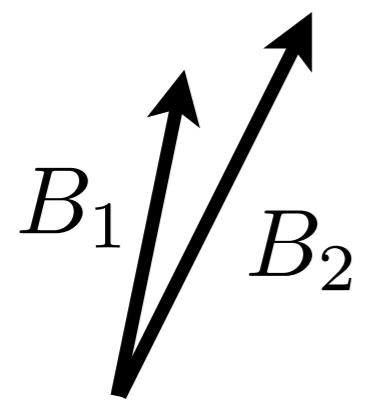
Relaxed domain



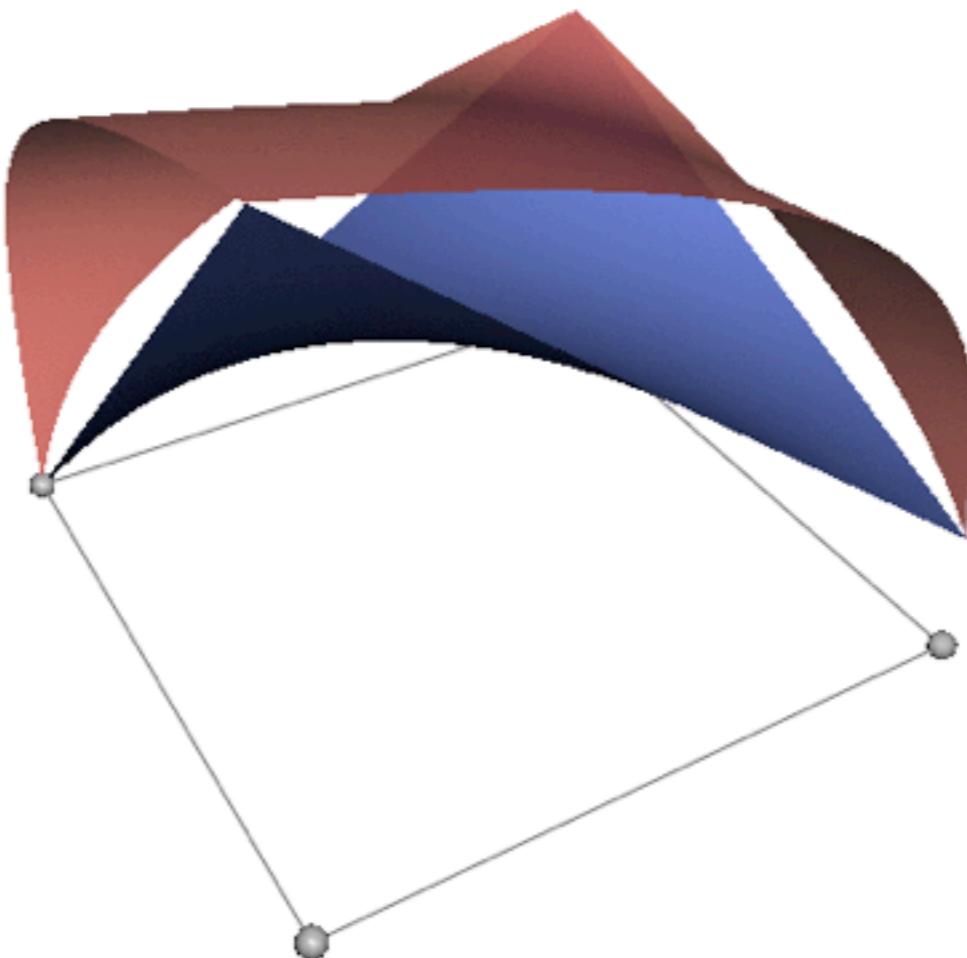


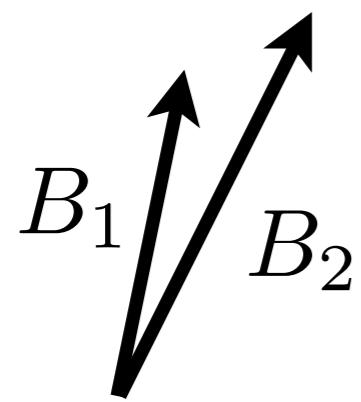
Multilinear extension



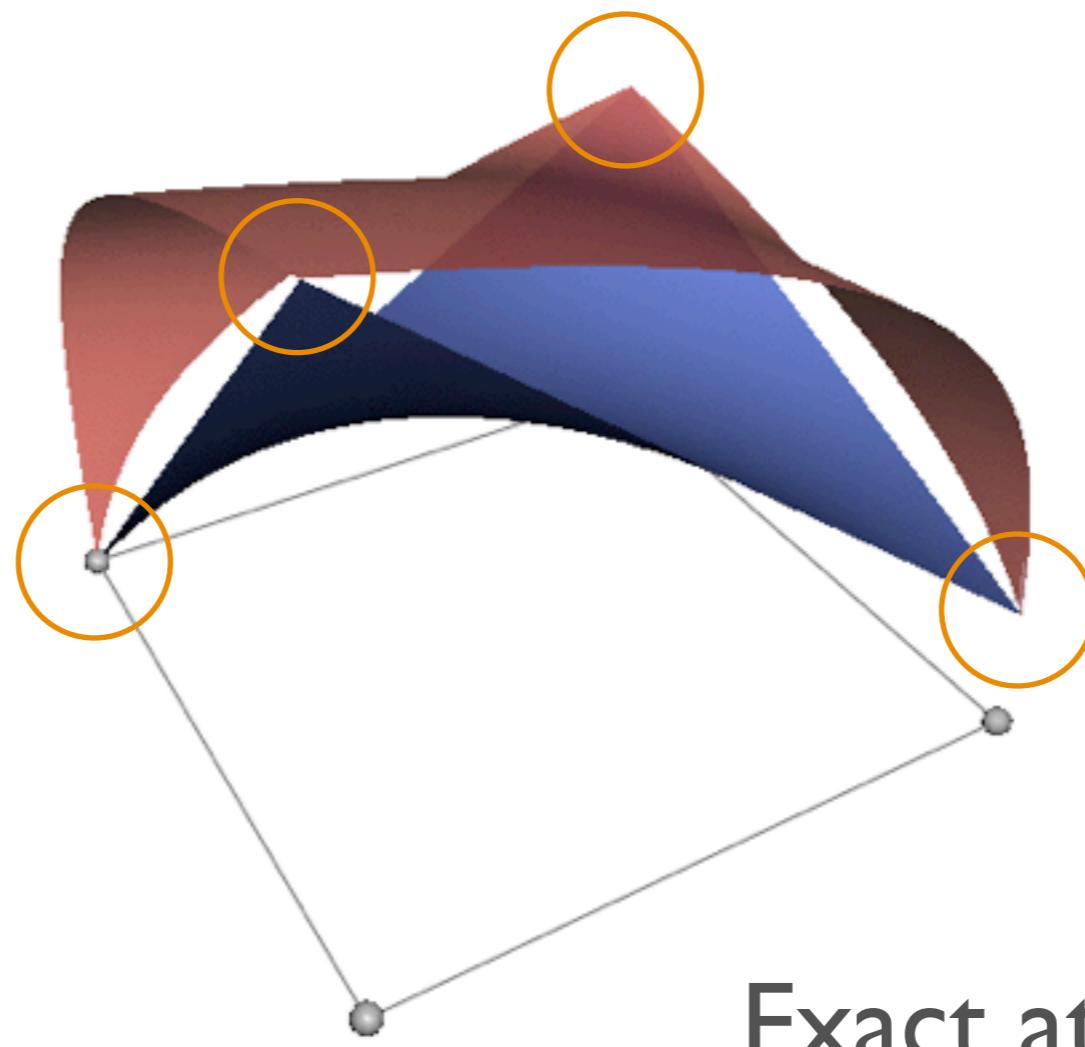


Softmax extension
Multilinear extension





Softmax extension
Multilinear extension



SOFTMAX EXTENSION

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

$$\tilde{F}(\mathbf{x}) = \log \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) f(Y)$$

SOFTMAX EXTENSION

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

$$\tilde{F}(\mathbf{x}) = \log \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) f(Y)$$

Theorem:

Efficiently computable for $f(Y) = \det(L_Y)$

$$O(N^3)$$

$$\tilde{F}(\mathbf{x}) = \log \det(\text{diag}(\mathbf{x})(L - I) + I)$$

EFFICIENCY PROOF

$$\exp(\tilde{F}(\boldsymbol{x})) = \sum_{Y: Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i: i \in Y} x_i \prod_{i: i \notin Y} (1 - x_i)$$

EFFICIENCY PROOF

$$\begin{aligned}\exp(\tilde{F}(\mathbf{x})) &= \sum_{Y: Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i:i \in Y} x_i \prod_{i:i \notin Y} (1 - x_i) \\ &= \prod_{i=1}^N (1 - x_i) \sum_{Y: Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i:i \in Y} \frac{x_i}{1 - x_i}\end{aligned}$$

EFFICIENCY PROOF

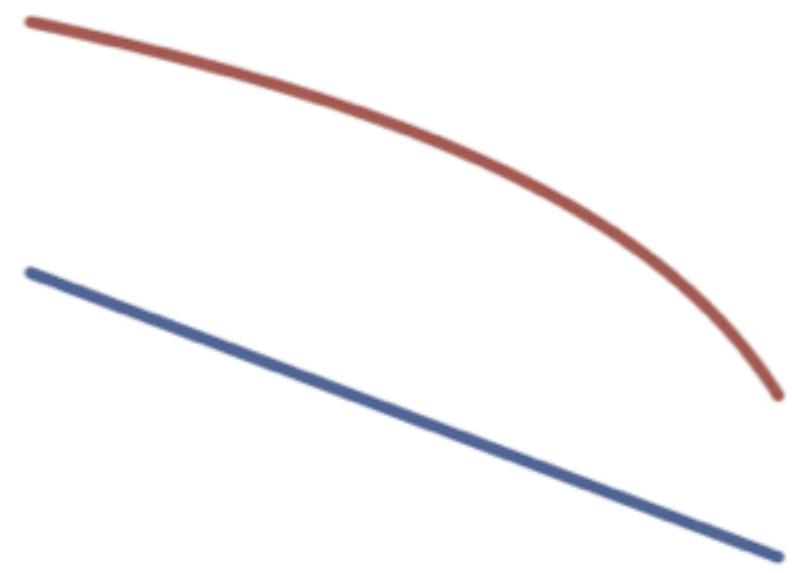
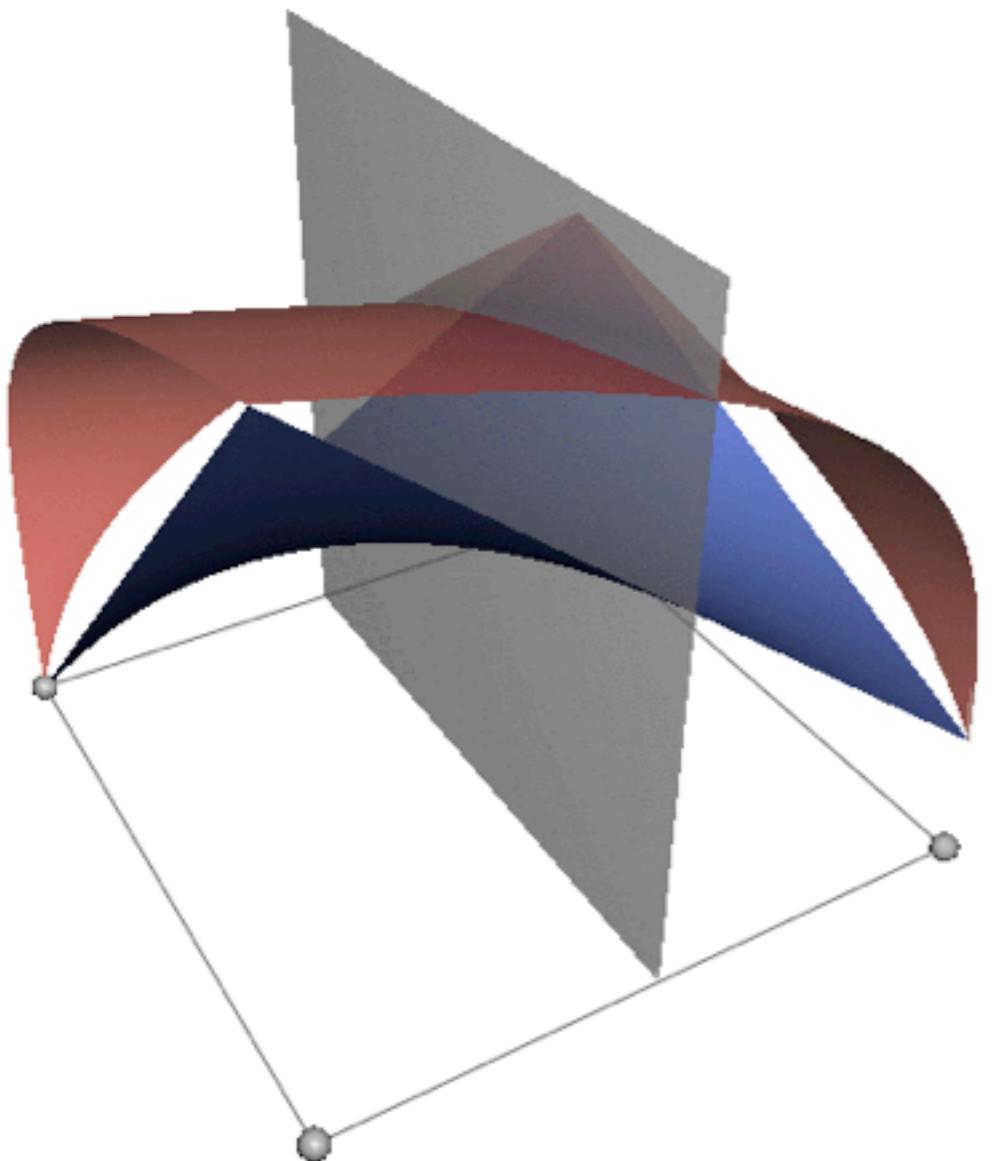
$$\begin{aligned}\exp(\tilde{F}(\boldsymbol{x})) &= \sum_{Y:Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i:i \in Y} x_i \prod_{i:i \notin Y} (1 - x_i) \\ &= \prod_{i=1}^N (1 - x_i) \sum_{Y:Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i:i \in Y} \frac{x_i}{1 - x_i} \\ &= \prod_{i=1}^N (1 - x_i) \sum_{Y:Y \subseteq \mathcal{Y}} \det \left([\text{diag}(\boldsymbol{x}) \text{diag}(1 - \boldsymbol{x})^{-1} L]_Y \right)\end{aligned}$$

EFFICIENCY PROOF

$$\begin{aligned}\exp(\tilde{F}(\boldsymbol{x})) &= \sum_{Y:Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i:i \in Y} x_i \prod_{i:i \notin Y} (1 - x_i) \\ &= \prod_{i=1}^N (1 - x_i) \sum_{Y:Y \subseteq \mathcal{Y}} \det(L_Y) \prod_{i:i \in Y} \frac{x_i}{1 - x_i} \\ &= \prod_{i=1}^N (1 - x_i) \sum_{Y:Y \subseteq \mathcal{Y}} \det([\text{diag}(\boldsymbol{x}) \text{diag}(1 - \boldsymbol{x})^{-1} L]_Y) \\ &= \prod_{i=1}^N (1 - x_i) \det(\text{diag}(\boldsymbol{x}) \text{diag}(1 - \boldsymbol{x})^{-1} L + I)\end{aligned}$$

Concave in all-positive/all-negative directions

Concave in all-positive/all-negative directions



CONCAVITY PROOF

For $u \geq 0$ and $0 < x + su < 1$: $\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) \leq 0$.

CONCAVITY PROOF

For $u \geq 0$ and $0 < x + su < 1$: $\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) \leq 0$.

$$\begin{aligned}\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) &= \text{tr}(-D^{-2} \text{diag}(u^2) \\ &\quad + 2M^{-1} D^{-3} \text{diag}(u^2) \\ &\quad - (M^{-1} D^{-2} \text{diag}(u))^2)\end{aligned}$$

CONCAVITY PROOF

For $u \geq 0$ and $0 < x + su < 1$: $\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) \leq 0$.

$$\begin{aligned}\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) &= \text{tr}(-D^{-2} \text{diag}(u^2) \\ &\quad + 2M^{-1} D^{-3} \text{diag}(u^2) \\ &\quad - (M^{-1} D^{-2} \text{diag}(u))^2) \\ &= - \text{tr}(SS^\top)\\ \text{where } S &= D^{-1} \text{diag}(u) \\ &\quad - M^{-1} D^{-2} \text{diag}(u)\end{aligned}$$

CONCAVITY PROOF

For $u \geq 0$ and $0 < x + su < 1$: $\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) \leq 0$.

$$\begin{aligned}\frac{\partial^2}{\partial s^2} \tilde{F}(x + su) &= \text{tr}(-D^{-2} \text{diag}(u^2) \\ &\quad + 2M^{-1} D^{-3} \text{diag}(u^2) \\ &\quad - (M^{-1} D^{-2} \text{diag}(u))^2) \\ &= - \text{tr}(SS^\top) \\ \text{where } S &= D^{-1} \text{diag}(u) \\ &\quad - M^{-1} D^{-2} \text{diag}(u) \\ &\leq 0\end{aligned}$$

APPROXIMATION GUARANTEE

Theorem: Concavity in all-positive directions

APPROXIMATION GUARANTEE

Theorem: Concavity in all-positive directions

+ Submodularity \longrightarrow

$$\text{LOCAL OPT of } \tilde{F} \geq \frac{1}{4} \max_{\mathbf{x}} \tilde{F}(\mathbf{x})$$

APPROXIMATION GUARANTEE

Theorem: Concavity in all-positive directions

+ Submodularity \longrightarrow

$$\text{LOCAL OPT of } \tilde{F} \geq \frac{1}{4} \max_{\mathbf{x}} \tilde{F}(\mathbf{x})$$

Theorem: In the unconstrained case, LOCAL OPT will be integer (no rounding necessary).

APPROXIMATION GUARANTEE

Theorem: Concavity in all-positive directions

+ Submodularity \longrightarrow

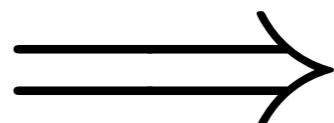
$$\text{LOCAL OPT of } \tilde{F} \geq \frac{1}{4} \max_{\mathbf{x}} \tilde{F}(\mathbf{x}) \geq \frac{1}{4} \max_Y \log \det(L_Y)$$

Theorem: In the unconstrained case, LOCAL OPT will be integer (no rounding necessary).

APPROXIMATION GUARANTEE

Theorem: Concavity in all-positive directions

+ Submodularity



$$\text{LOCAL OPT of } \tilde{F} \geq \frac{1}{4} \max_{\mathbf{x}} \tilde{F}(\mathbf{x}) \geq \frac{1}{4} \max_Y \log \det(L_Y)$$

Theorem: In the unconstrained case, LOCAL OPT will be integer (no rounding necessary).

Constrained: No guarantees, but in practice pipage
 $\max_{Y \in S}$ rounding and thresholding work well.

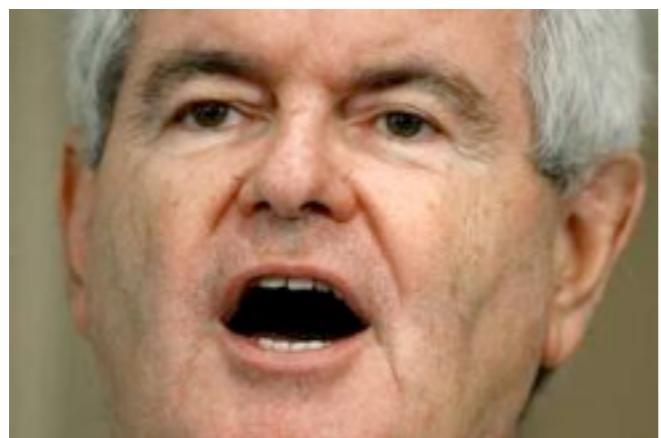
MATCHED SUMMARIZATION

MATCHED SUMMARIZATION

20 Republican primary debates

MATCHED SUMMARIZATION

20 Republican primary debates



MATCHED SUMMARIZATION

20 Republican primary debates



Average of 179 quotes per candidate

MATCHED SUMMARIZATION







‣ **R1 (taxes): No tax on interest, dividends, or capital gains.**



- **R1 (taxes): No tax on interest, dividends, or capital gains.**
- **R2 (law): We're not going to have Sharia law applied in U.S. courts.**



- **R1 (taxes): No tax on interest, dividends, or capital gains.**
- **R2 (law): We're not going to have Sharia law applied in U.S. courts.**
- **R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.**



- R1 (taxes): **No tax on interest, dividends, or capital gains.**
- R2 (law): **We're not going to have Sharia law applied in U.S. courts.**
- R3 (healthcare): **I will ... grant a waiver from Obamacare to all 50 states.**
- R4 (aid): **We're spending more on foreign aid than we ought to.**



- R1 (taxes): **No tax on interest, dividends, or capital gains.**
- R2 (law): **We're not going to have Sharia law applied in U.S. courts.**
- R3 (healthcare): **I will ... grant a waiver from Obamacare to all 50 states.**
- R4 (aid): **We're spending more on foreign aid than we ought to.**
- R5(healthcare): **If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.**



- R1 (taxes): **No tax on interest, dividends, or capital gains.**
- R2 (law): **We're not going to have Sharia law applied in U.S. courts.**
- R3 (healthcare): **I will ... grant a waiver from Obamacare to all 50 states.**
- R4 (aid): **We're spending more on foreign aid than we ought to.**
- R5(healthcare): **If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.**





- R1 (taxes): **No tax on interest, dividends, or capital gains.**
- R2 (law): **We're not going to have Sharia law applied in U.S. courts.**
- R3 (healthcare): **I will ... grant a waiver from Obamacare to all 50 states.**
- R4 (aid): **We're spending more on foreign aid than we ought to.**
- R5(healthcare): **If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.**



- S1 (taxes): **I don't believe in a zero capital gains tax rate.**



- R1 (taxes): No tax on interest, dividends, or capital gains.
- R2 (law): We're not going to have Sharia law applied in U.S. courts.
- R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.
- R4 (aid): We're spending more on foreign aid than we ought to.
- R5(healthcare): If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.



- S1 (taxes): I don't believe in a zero capital gains tax rate.
- S2 (taxes): Manufacture in America, you aren't going to pay any taxes.



- R1 (taxes): No tax on interest, dividends, or capital gains.
- R2 (law): We're not going to have Sharia law applied in U.S. courts.
- R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.
- R4 (aid): We're spending more on foreign aid than we ought to.
- R5(healthcare): If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.



- S1 (taxes): I don't believe in a zero capital gains tax rate.
- S2 (taxes): Manufacture in America, you aren't going to pay any taxes.
- S3 (aid): Zeroing out foreign aid ... that's absolutely the wrong course.



- R1 (taxes): No tax on interest, dividends, or capital gains.
- R2 (law): We're not going to have Sharia law applied in U.S. courts.
- R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.
- R4 (aid): We're spending more on foreign aid than we ought to.
- R5(healthcare): If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.



- S1 (taxes): I don't believe in a zero capital gains tax rate.
- S2 (taxes): Manufacture in America, you aren't going to pay any taxes.
- S3 (aid): Zeroing out foreign aid ... that's absolutely the wrong course.
- S4 (ethanol): I voted against ethanol subsidies my entire time in Congress.



- R1 (taxes): No tax on interest, dividends, or capital gains.
- R2 (law): We're not going to have Sharia law applied in U.S. courts.
- R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.
- R4 (aid): We're spending more on foreign aid than we ought to.
- R5 (healthcare): If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.



- S1 (taxes): I don't believe in a zero capital gains tax rate.
- S2 (taxes): Manufacture in America, you aren't going to pay any taxes.
- S3 (aid): Zeroing out foreign aid ... that's absolutely the wrong course.
- S4 (ethanol): I voted against ethanol subsidies my entire time in Congress.
- S5 (healthcare): Obamacare ... is going to blow a hole in the budget.

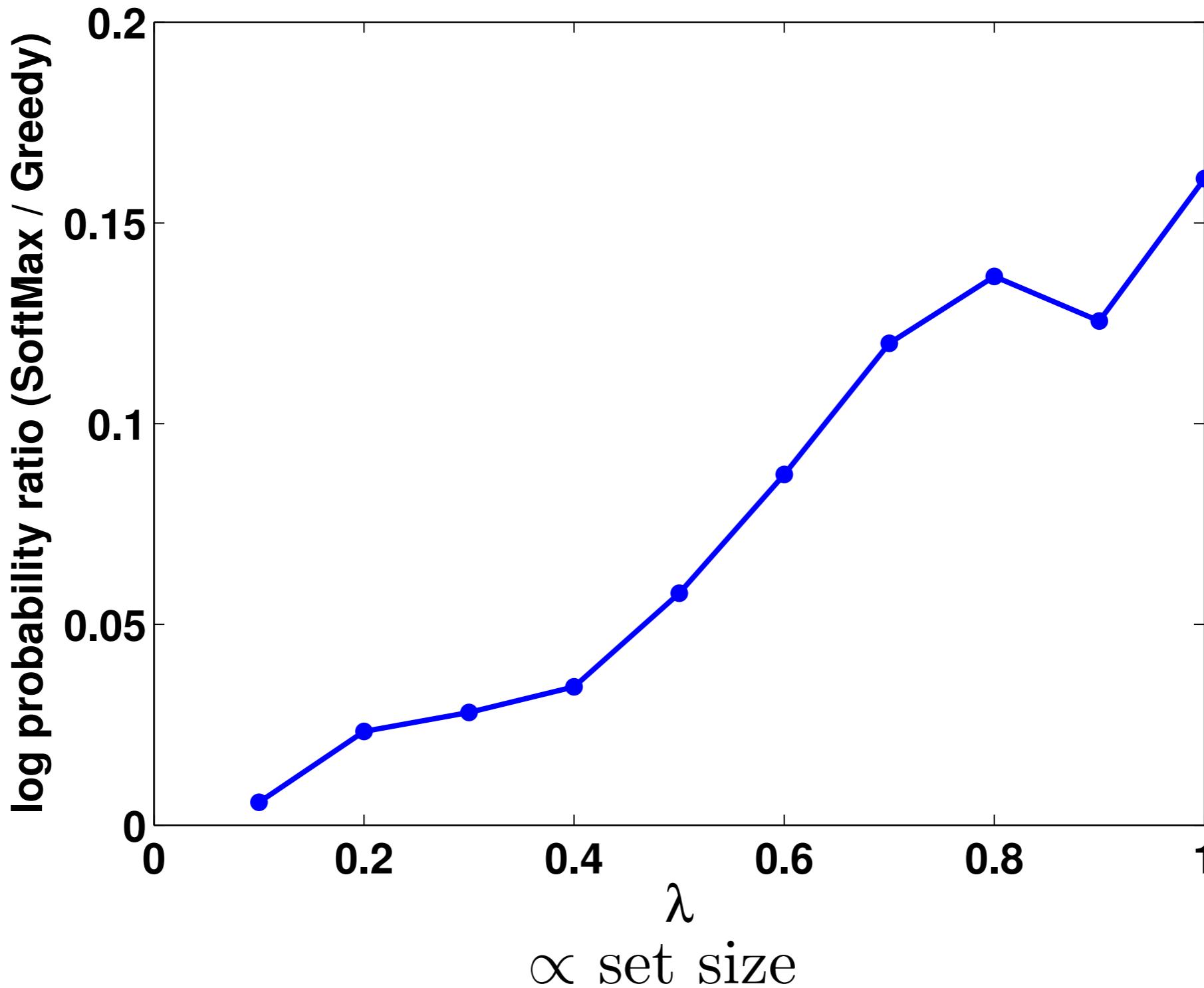
Matched summary

- **R1 (taxes): No tax on interest, dividends, or capital gains.**
- **S1 (taxes): I don't believe in a zero capital gains tax rate.**

- **R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.**
- **S5 (healthcare): Obamacare ... is going to blow a hole in the budget.**

- **R4 (aid): We're spending more on foreign aid than we ought to.**
- **S3 (aid): Zeroing out foreign aid ... that's absolutely the wrong course.**

PERFORMANCE



PROPOSED WORK

- Update experiments with faster algorithms and testing with random restarts
- k-DPP MAP estimation algorithm based on marginals, plus extension to structured k-DPPs

FUTURE WORK

- Update experiments with faster algorithms and testing with random restarts
- k-DPP MAP estimation algorithm based on marginals, plus extension to structured k-DPPs

3. LIKELIHOOD MAXIMIZATION

LIKELIHOOD

LIKELIHOOD

Y_1 

Y_2 

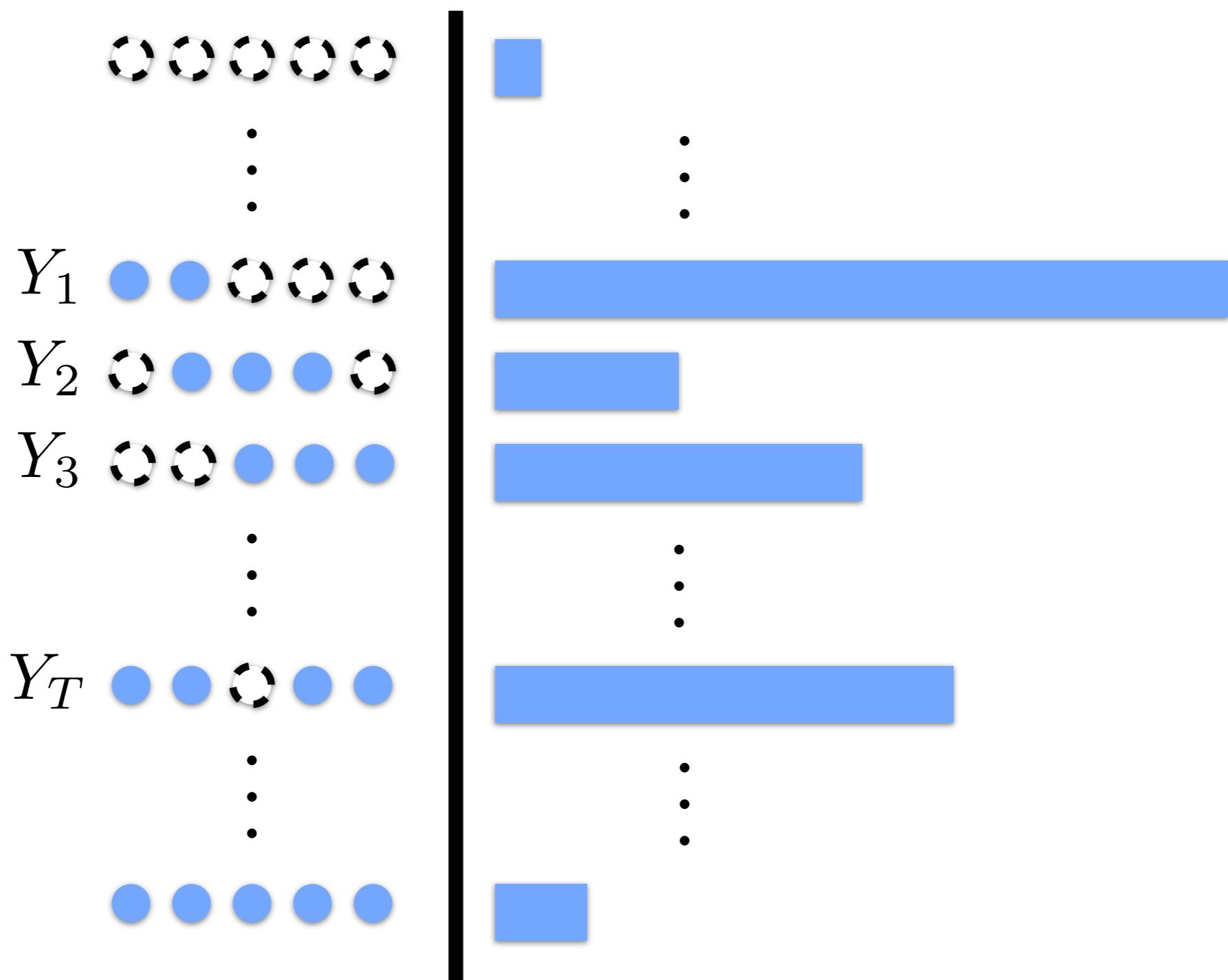
Y_3 

⋮
⋮

Y_T 

LIKELIHOOD

\mathcal{P} is a DPP



NON-CONCAVITY

NON-CONCAVITY

$$\mathcal{L}(L) = \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)]$$

NON-CONCAVITY

$$\begin{aligned}\mathcal{L}(L) &= \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)] \\ &\max_{L: L \succeq 0} \mathcal{L}(L)\end{aligned}$$

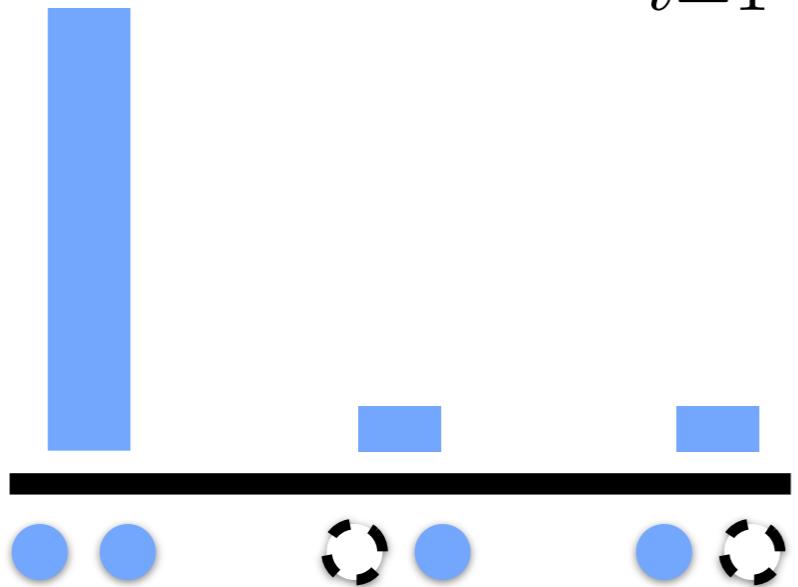
NON-CONCAVITY

$$\mathcal{L}(L) = \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)]$$
$$\max_{L:L \succeq 0} \mathcal{L}(L)$$



NON-CONCAVITY

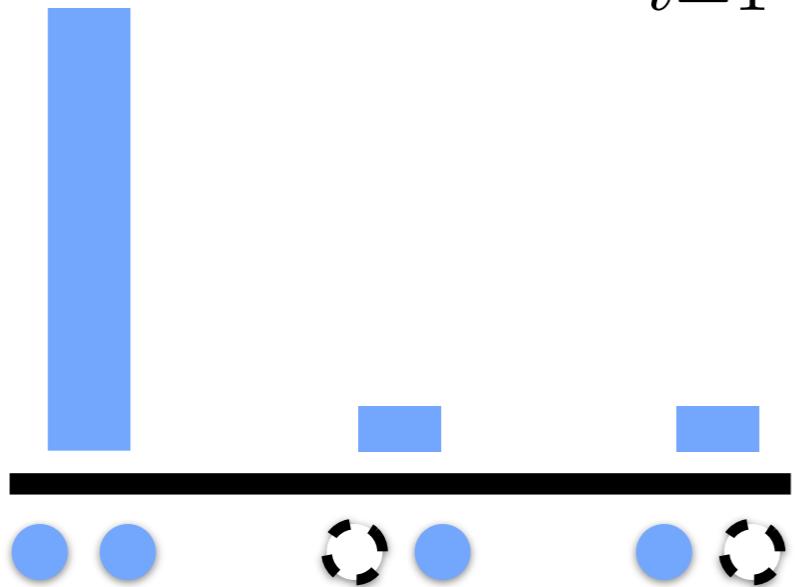
$$\mathcal{L}(L) = \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)]$$
$$\max_{L: L \succeq 0} \mathcal{L}(L)$$



$$L = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$

NON-CONCAVITY

$$\mathcal{L}(L) = \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)]$$
$$\max_{L: L \succeq 0} \mathcal{L}(L)$$

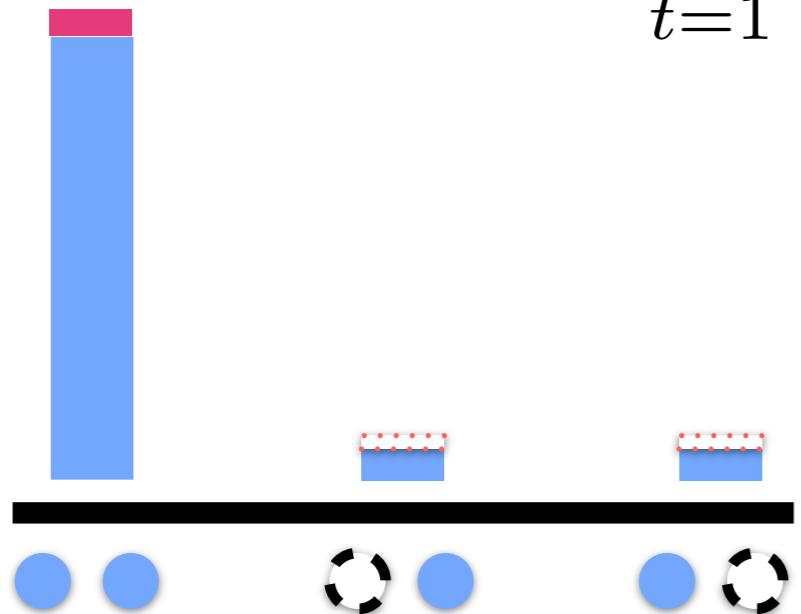


$$L = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$

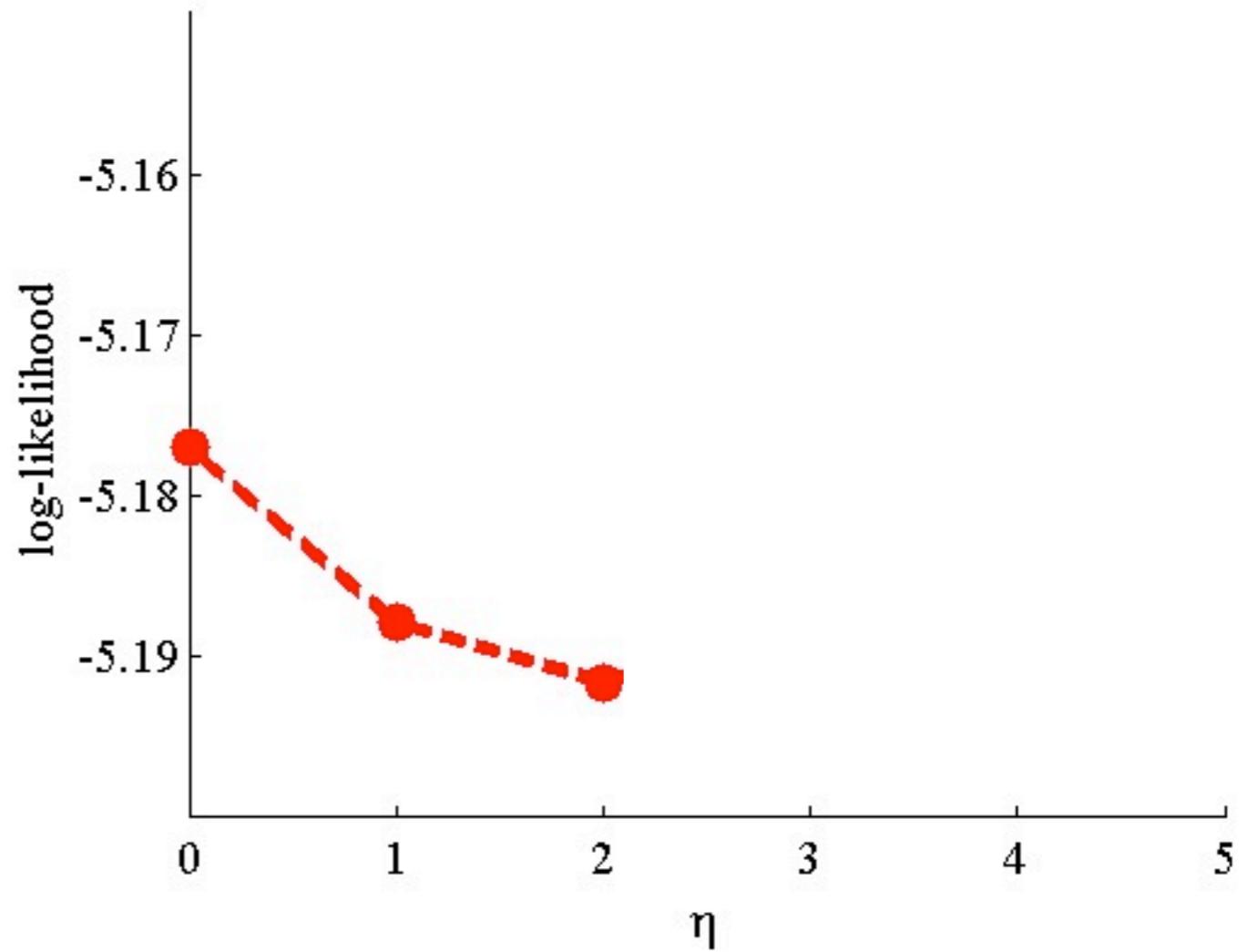
$$+ \eta \begin{bmatrix} 0.1 & 0.5 \\ 0.5 & 0.1 \end{bmatrix}$$

NON-CONCAVITY

$$\mathcal{L}(L) = \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)]$$
$$\max_{L: L \succeq 0} \mathcal{L}(L)$$

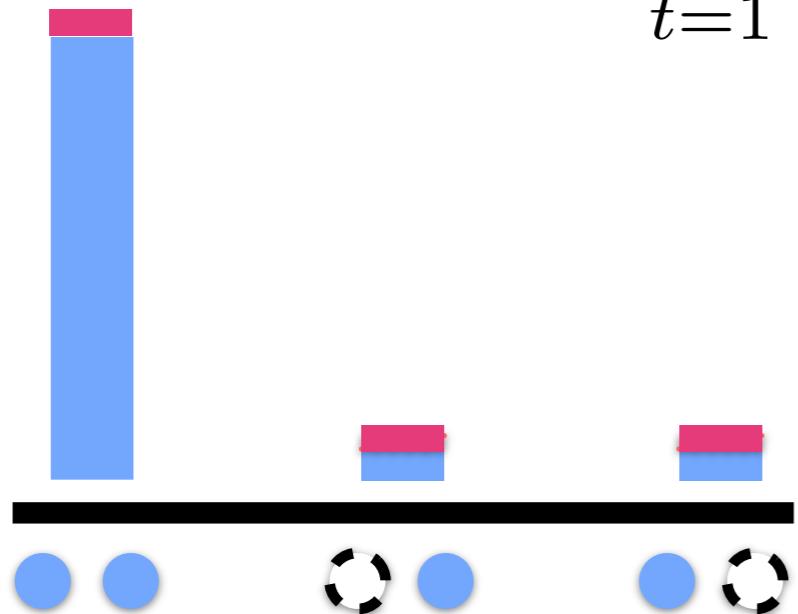


$$L = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$
$$+ \eta \begin{bmatrix} 0.1 & 0.5 \\ 0.5 & 0.1 \end{bmatrix}$$

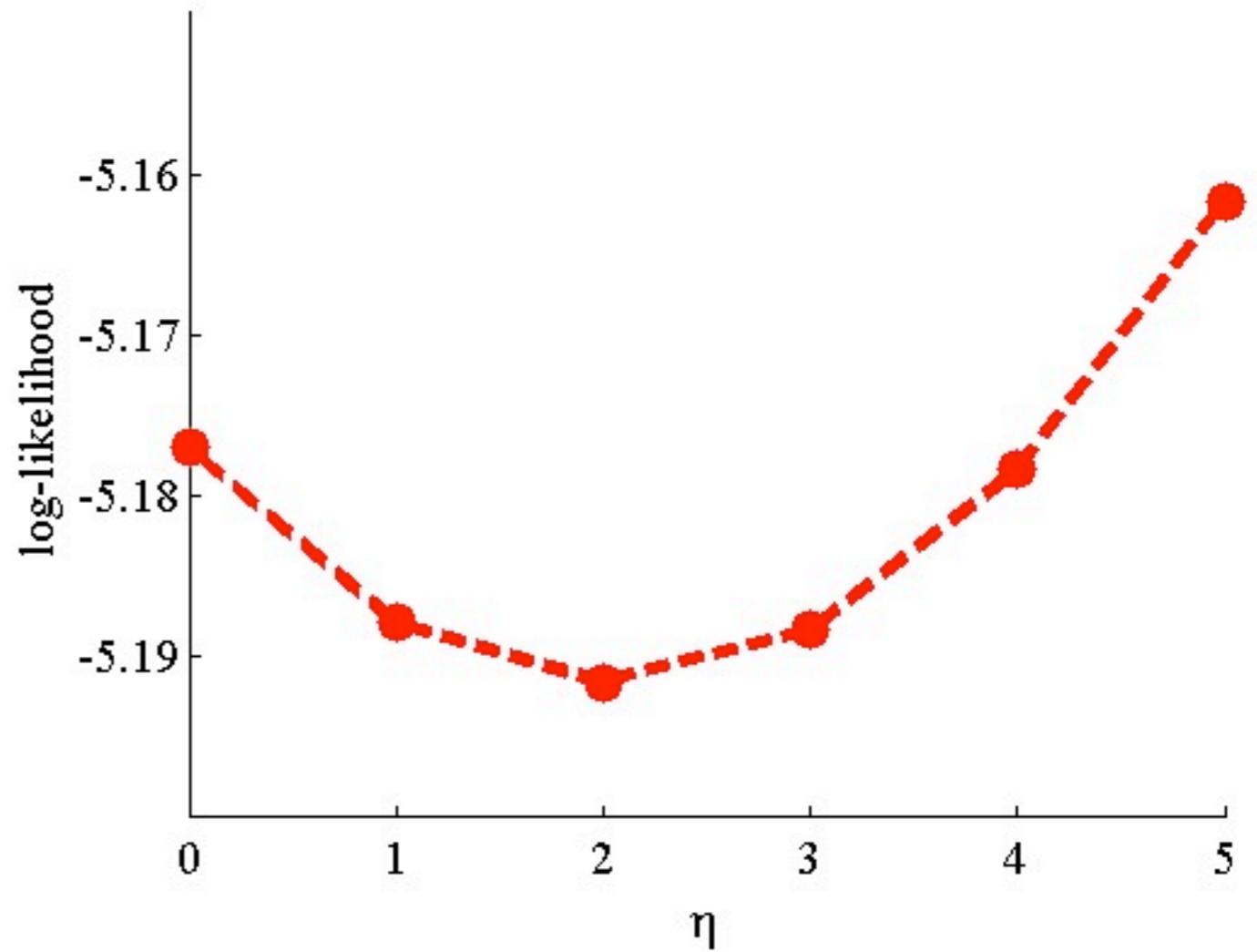


NON-CONCAVITY

$$\mathcal{L}(L) = \sum_{t=1}^T [\log \det(L_{Y_t}) - \log \det(L + I)]$$
$$\max_{L: L \succeq 0} \mathcal{L}(L)$$



$$L = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix} + \eta \begin{bmatrix} 0.1 & 0.5 \\ 0.5 & 0.1 \end{bmatrix}$$



K PARAMETERIZATION

K PARAMETERIZATION

Change of variables for developing EM algorithm:

$$L = V \text{diag}(\lambda) V^T$$

$$K = V \text{diag}\left(\frac{\lambda}{1+\lambda}\right) V^T$$

K PARAMETERIZATION

Change of variables for developing EM algorithm:

$$L = \textcolor{blue}{V} \text{diag}(\textcolor{brown}{\lambda}) \textcolor{blue}{V}^\top$$

$$K = \textcolor{blue}{V} \text{diag}\left(\frac{\lambda}{1+\lambda}\right) \textcolor{blue}{V}^\top$$

$$\mathcal{L}(K) = \sum_{t=1}^T \log |\det(K - I_{\bar{Y}_t})|$$

K PARAMETERIZATION

Change of variables for developing EM algorithm:

$$L = \textcolor{blue}{V} \text{diag}(\textcolor{brown}{\lambda}) \textcolor{blue}{V}^\top$$

$$K = \textcolor{blue}{V} \text{diag}\left(\frac{\lambda}{1+\lambda}\right) \textcolor{blue}{V}^\top$$

$$\mathcal{L}(K) = \sum_{t=1}^T \log |\det(K - I_{\bar{Y}_t})|$$

$$\max_{\substack{K: K \succeq 0, \\ I - K \succeq 0}} \mathcal{L}(K)$$

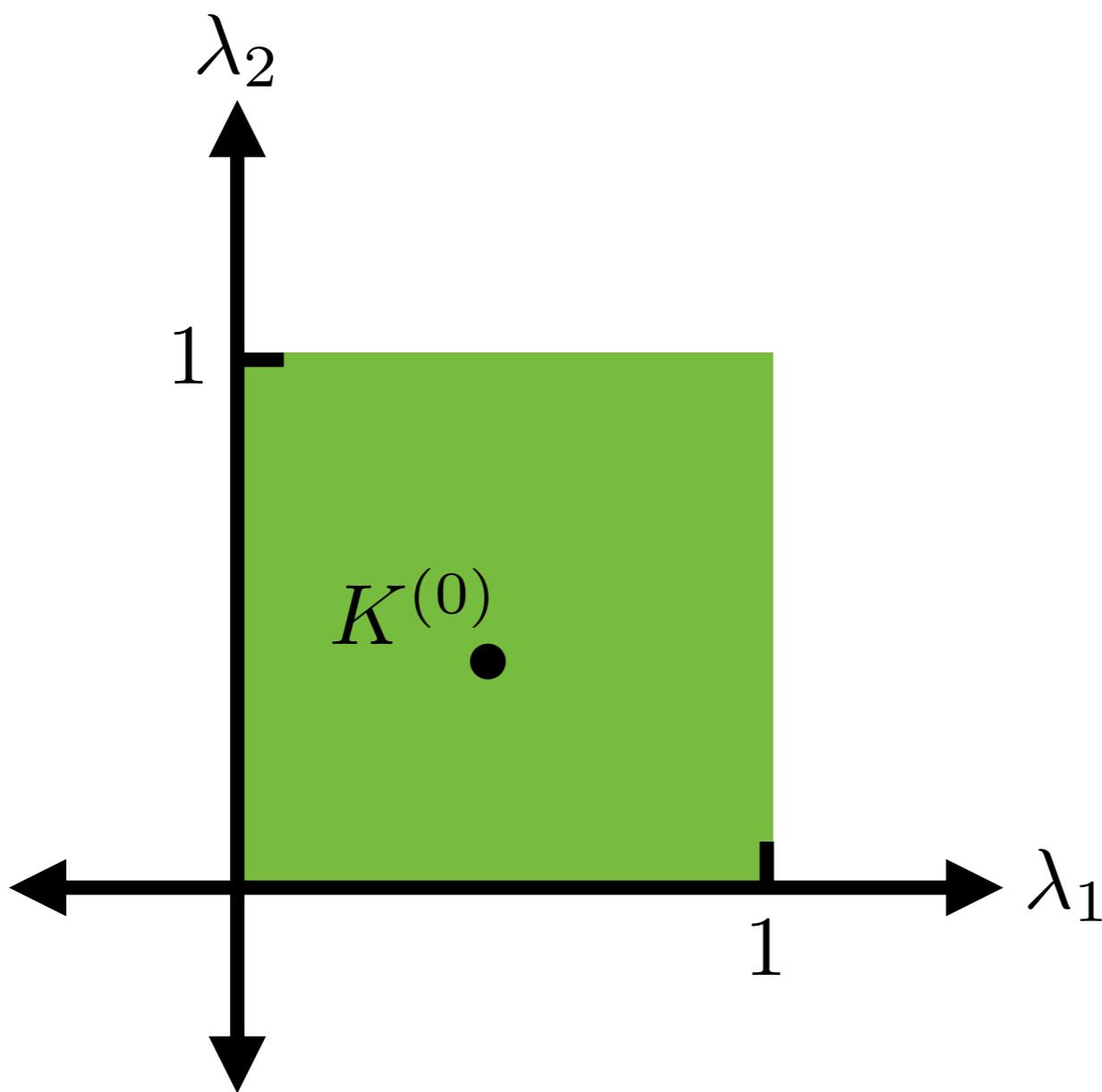
GRADIENT ASCENT

GRADIENT ASCENT

$$\frac{\partial \mathcal{L}(K)}{\partial K} = \sum_{t=1}^T (K - I_{\overline{Y}_t})^{-1}$$

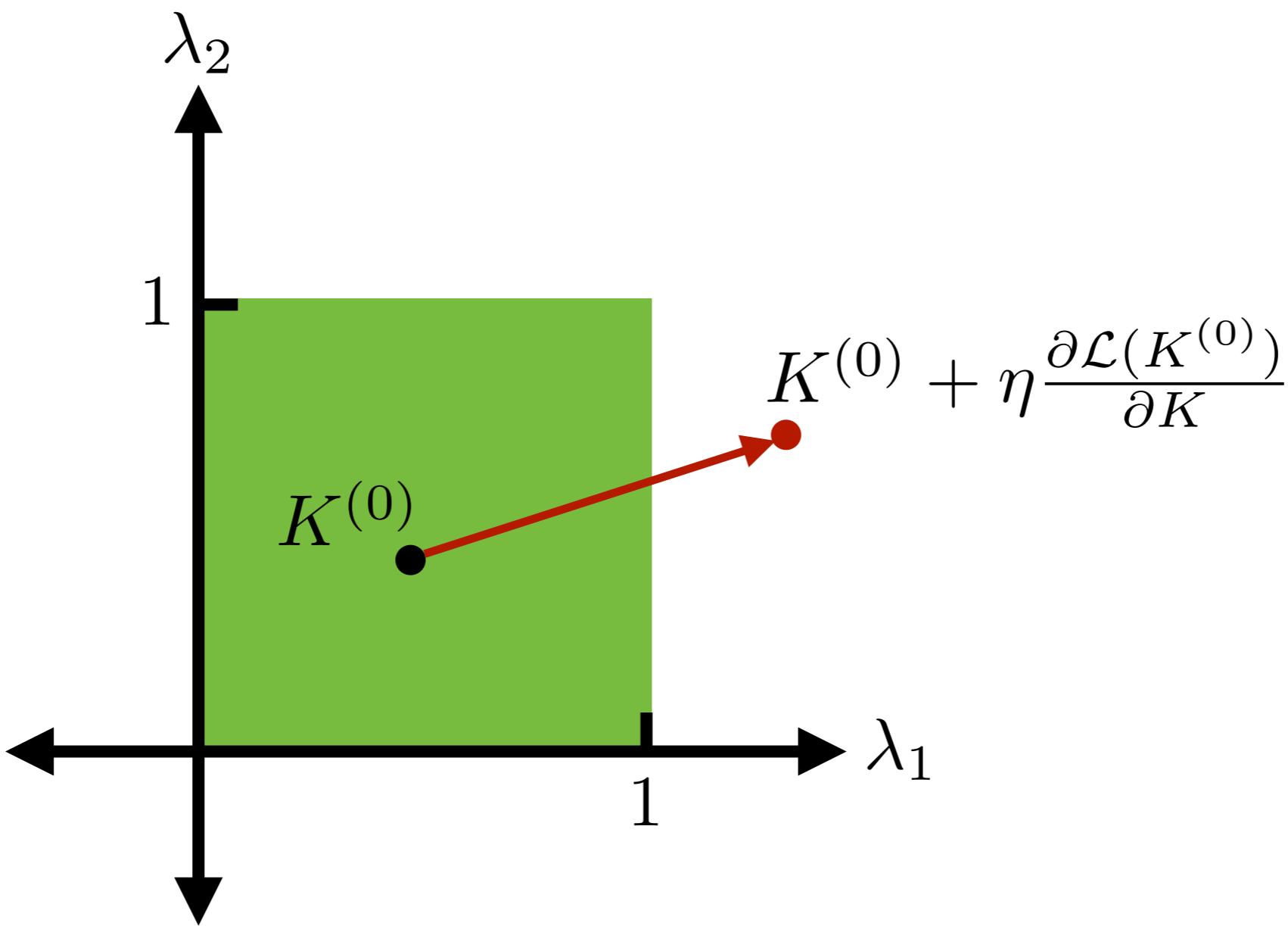
GRADIENT ASCENT

$$\frac{\partial \mathcal{L}(K)}{\partial K} = \sum_{t=1}^T (K - I_{\bar{Y}_t})^{-1}$$



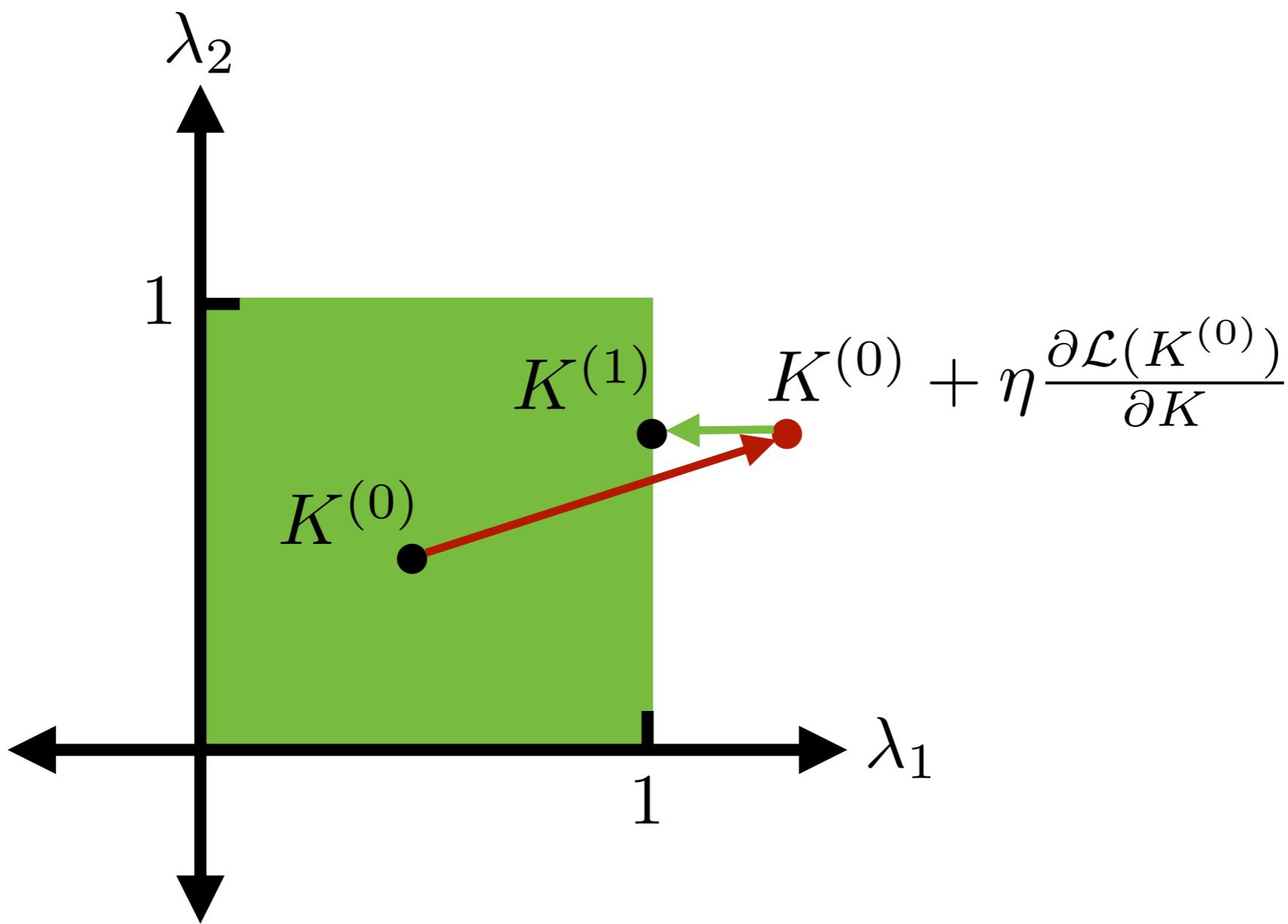
GRADIENT ASCENT

$$\frac{\partial \mathcal{L}(K)}{\partial K} = \sum_{t=1}^T (K - I_{\bar{Y}_t})^{-1}$$



GRADIENT ASCENT

$$\frac{\partial \mathcal{L}(K)}{\partial K} = \sum_{t=1}^T (K - I_{\bar{Y}_t})^{-1}$$



EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$K = V \Lambda V^\top$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$K = \mathcal{V} \Lambda \mathcal{V}^\top$$

$$K^J = \mathcal{V}^J \text{diag}(\mathbf{1}) (\mathcal{V}^J)^\top$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$K = \textcolor{blue}{V} \Lambda \textcolor{green}{V}^\top$$

$$K^{\textcolor{red}{J}} = \textcolor{blue}{V}^{\textcolor{red}{J}} \text{diag}(\mathbf{1}) (\textcolor{blue}{V}^{\textcolor{red}{J}})^\top$$

$$\mathcal{P}(Y) = \frac{\det(L_Y)}{\det(L + I)} = |\det(K - I_{\bar{Y}})|$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$K = \textcolor{blue}{V} \Lambda \textcolor{blue}{V}^\top$$

$$K^{\textcolor{red}{J}} = \textcolor{blue}{V}^{\textcolor{red}{J}} \text{diag}(\mathbf{1})(\textcolor{blue}{V}^{\textcolor{red}{J}})^\top$$

$$\begin{aligned} \mathcal{P}(Y) &= \frac{\det(L_Y)}{\det(L + I)} = |\det(K - I_{\bar{Y}})| \\ &= \sum_{\textcolor{red}{J}: \textcolor{red}{J} \subseteq \{1, \dots, N\}} \mathcal{P}_{K^{\textcolor{red}{J}}}(Y) \prod_{j:j \in \textcolor{red}{J}} \lambda_j \prod_{j:j \notin \textcolor{red}{J}} (1 - \lambda_j) \end{aligned}$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$K = V \Lambda V^\top$$

$$K^{\textcolor{red}{J}} = V^{\textcolor{red}{J}} \text{diag}(\mathbf{1})(V^{\textcolor{red}{J}})^\top$$

$$\begin{aligned}\mathcal{P}(Y) &= \frac{\det(L_Y)}{\det(L + I)} = |\det(K - I_{\bar{Y}})| \\ &= \sum_{\textcolor{red}{J}: \textcolor{red}{J} \subseteq \{1, \dots, N\}} \mathcal{P}_{K^{\textcolor{red}{J}}}(Y) \prod_{j:j \in \textcolor{red}{J}} \lambda_j \prod_{j:j \notin \textcolor{red}{J}} (1 - \lambda_j) \\ &= \sum_{\textcolor{red}{J}: \textcolor{red}{J} \subseteq \{1, \dots, N\}} \Pr_K(Y \mid \textcolor{red}{J}) \Pr_K(\textcolor{red}{J})\end{aligned}$$

Hidden variable $\textcolor{red}{J}$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\mathcal{L}(K) = \mathcal{L}(V, \Lambda) = \sum_{t=1}^T \log p_K(Y_t) = \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

eigenvectors eigenvalues

$$\mathcal{L}(K) = \mathcal{L}(V, \Lambda) = \sum_{t=1}^T \log p_K(Y_t) = \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right)$$


EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

eigenvectors eigenvalues

$$\mathcal{L}(K) = \mathcal{L}(V, \Lambda) = \sum_{t=1}^T \log p_K(Y_t) = \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right)$$

choice of eigenvectors

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

eigenvectors eigenvalues

$$\mathcal{L}(K) = \mathcal{L}(V, \Lambda) = \sum_{t=1}^T \log p_K(Y_t) = \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right)$$

choice of eigenvectors

$$= \sum_{t=1}^T \log \left(\sum_J q(J \mid Y_t) \frac{p_K(J, Y_t)}{q(J \mid Y_t)} \right)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\begin{aligned} \text{eigenvectors } & \text{ eigenvalues} & & \text{choice of eigenvectors} \\ \mathcal{L}(K) = \mathcal{L}(V, \Lambda) &= \sum_{t=1}^T \log p_K(Y_t) &= \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right) \\ &= \sum_{t=1}^T \log \left(\sum_J q(J | Y_t) \frac{p_K(J, Y_t)}{q(J | Y_t)} \right) \\ &\geq \sum_{t=1}^T \sum_J q(J | Y_t) \log \left(\frac{p_K(J, Y_t)}{q(J | Y_t)} \right) \equiv F(q, V, \Lambda) \end{aligned}$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

eigenvectors eigenvalues

$$\mathcal{L}(K) = \mathcal{L}(V, \Lambda) = \sum_{t=1}^T \log p_K(Y_t) = \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right)$$

$$= \sum_{t=1}^T \log \left(\sum_J q(J | Y_t) \frac{p_K(J, Y_t)}{q(J | Y_t)} \right)$$

$$\geq \sum_{t=1}^T \sum_J q(J | Y_t) \log \left(\frac{p_K(J, Y_t)}{q(J | Y_t)} \right) \equiv F(q, V, \Lambda)$$

$$\mathcal{P}(Y) = \sum_{\mathbf{J}: \mathbf{J} \subseteq \{1, \dots, N\}} \mathcal{P}_{K^{\mathbf{J}}}(Y) \prod_{j:j \in \mathbf{J}} \lambda_j \prod_{j:j \notin \mathbf{J}} (1 - \lambda_j)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

eigenvectors eigenvalues

$$\begin{aligned}
 \mathcal{L}(K) = \mathcal{L}(V, \Lambda) &= \sum_{t=1}^T \log p_K(Y_t) = \sum_{t=1}^T \log \left(\sum_J p_K(J, Y_t) \right) \\
 &= \sum_{t=1}^T \log \left(\sum_J q(J | Y_t) \frac{p_K(J, Y_t)}{q(J | Y_t)} \right) \\
 &\geq \sum_{t=1}^T \sum_J q(J | Y_t) \log \left(\frac{p_K(J, Y_t)}{q(J | Y_t)} \right) \equiv F(q, V, \Lambda)
 \end{aligned}$$

$$\mathcal{P}(Y) = \sum_{\mathbf{J}: \mathbf{J} \subseteq \{1, \dots, N\}} \mathcal{P}_{K^{\mathbf{J}}}(Y) \prod_{j:j \in \mathbf{J}} \lambda_j \prod_{j:j \notin \mathbf{J}} (1 - \lambda_j)$$

$$\log(p_K(\mathbf{J}, Y)) = \log \mathcal{P}_{K^{\mathbf{J}}}(Y) + \sum_{j:j \in \mathbf{J}} \log \lambda_j + \sum_{j:j \notin \mathbf{J}} \log(1 - \lambda_j)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$F(q, V, \Lambda) = \sum_{t=1}^T -\mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t)) + \mathcal{L}(V, \Lambda)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$F(q, V, \Lambda) = \sum_{t=1}^T -\mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t)) + \mathcal{L}(V, \Lambda)$$

E-step: $\min_q \sum_{t=1}^T \mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t))$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\begin{aligned} F(q, V, \Lambda) &= \sum_{t=1}^T -\mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t)) + \mathcal{L}(V, \Lambda) \\ &= \sum_{t=1}^T \mathbb{E}_q[\log p_K(J, Y_t)] + H(q) \end{aligned}$$

E-step: $\min_q \sum_{t=1}^T \mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t))$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\begin{aligned} F(q, V, \Lambda) &= \sum_{t=1}^T -\mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t)) + \mathcal{L}(V, \Lambda) \\ &= \sum_{t=1}^T \mathbb{E}_q[\log p_K(J, Y_t)] + H(q) \end{aligned}$$

E-step: $\min_q \sum_{t=1}^T \mathbf{KL}(q(J \mid Y_t) \parallel p_K(J \mid Y_t))$

M-step: $\max_{V, \Lambda} \sum_{t=1}^T \mathbb{E}_q[\log p_K(J, Y_t)] \text{ s.t. } \mathbf{0} \leq \boldsymbol{\lambda} \leq \mathbf{1}, V^\top V = I$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\text{E-step: } \min_q \sum_{t=1}^T \mathbf{KL}(q(J | Y_t) \| p_K(J | Y_t))$$

$$q(J | Y_t) = p_K(J | Y_t)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\text{E-step: } \min_q \sum_{t=1}^T \mathbf{KL}(q(J | Y_t) \| p_K(J | Y_t))$$

$$q(J | Y_t) = p_K(J | Y_t)$$

$P(Y_t | J)$ is a DPP $\implies P(J | Y_t)$ is a DPP

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\text{E-step: } \min_q \sum_{t=1}^T \mathbf{KL}(q(J | Y_t) \| p_K(J | Y_t))$$

$$q(J | Y_t) = p_K(J | Y_t)$$

$P(Y_t | J)$ is a DPP $\implies P(J | Y_t)$ is a DPP

$$\text{M-step: } \lambda_j = \frac{1}{T} \sum_{t=1}^T \sum_{J:j \in J} q(J | Y_t)$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\text{E-step: } \min_q \sum_{t=1}^T \mathbf{KL}(q(J | Y_t) \| p_K(J | Y_t))$$

$$q(J | Y_t) = p_K(J | Y_t)$$

$P(Y_t | J)$ is a DPP $\implies P(J | Y_t)$ is a DPP

$$\text{M-step: } \lambda_j = \frac{1}{T} \sum_{t=1}^T \sum_{J:j \in J} q(J | Y_t)$$

$$O(T N k^2), \text{ for } k = \max_t |Y_t|$$

EXPECTATION-MAXIMIZATION

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

$$\text{E-step: } \min_q \sum_{t=1}^T \mathbf{KL}(q(J | Y_t) \| p_K(J | Y_t))$$

$$q(J | Y_t) = p_K(J | Y_t)$$

$P(Y_t | J)$ is a DPP $\implies P(J | Y_t)$ is a DPP

$$\text{M-step: } \lambda_j = \frac{1}{T} \sum_{t=1}^T \sum_{J:j \in J} q(J | Y_t)$$

$$O(T N k^2), \text{ for } k = \max_t |Y_t|$$

V update is more complicated, but still efficient

PRODUCT RECOMMENDATION

PRODUCT RECOMMENDATION

~ 30,000 Amazon baby registries

PRODUCT RECOMMENDATION

~ 30,000 Amazon baby registries

13 categories

PRODUCT RECOMMENDATION

~ 30,000 Amazon baby registries

13 categories



furniture

PRODUCT RECOMMENDATION

~ 30,000 Amazon baby registries

13 categories



furniture



carseats

PRODUCT RECOMMENDATION

~ 30,000 Amazon baby registries

13 categories



furniture



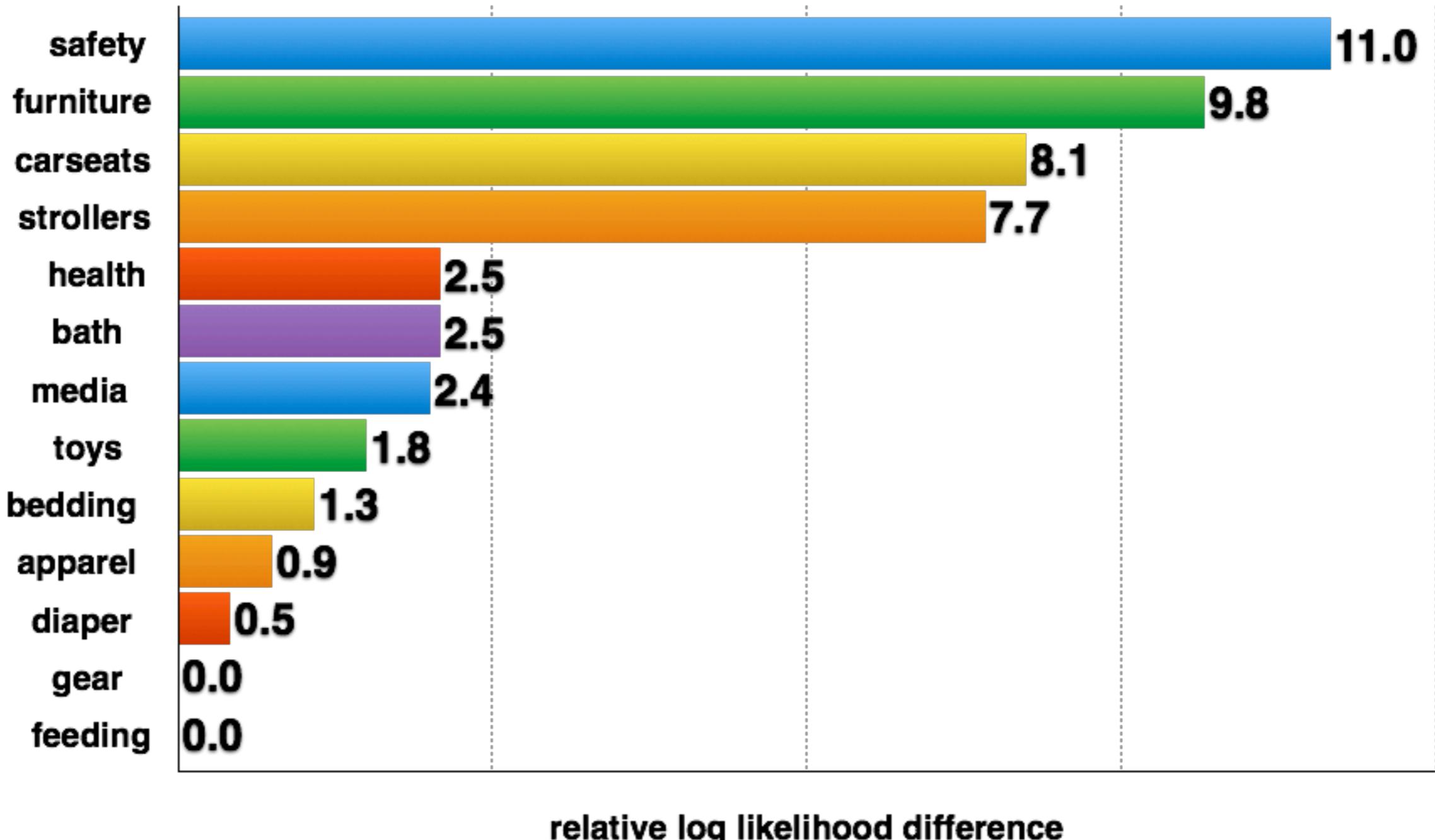
carseats



toys

EM VS PROJECTED GRADIENT

EM VS PROJECTED GRADIENT



“SAFETY” SELECTION

“SAFETY” SELECTION

Graco Sweet Slumber
Sound Machine



Boppy Noggin Nest
Head Support



Cloud b Twilight
Night Light



Braun ThermoScan
Lens Filters



Aquatopia Bath
Thermometer Alarm



Britax EZ-Cling
Sun Shades



TL Care Organic
Cotton Mittens



Regalo Easy Step
Walk Thru Gate



VTech Comm.
Audio Monitor



Infant Optics
Video Monitor

“SAFETY” SELECTION

Graco Sweet Slumber Sound Machine



Motorola Video Monitor



Cloud b Twilight Night Light



Summer Infant Video Monitor



Aquatopia Bath Thermometer Alarm



Britax EZ-Cling Sun Shades



TL Care Organic Cotton Mittens



Regalo Easy Step Walk Thru Gate



VTech Comm. Audio Monitor



Infant Optics Video Monitor

PROPOSED WORK

PROPOSED WORK

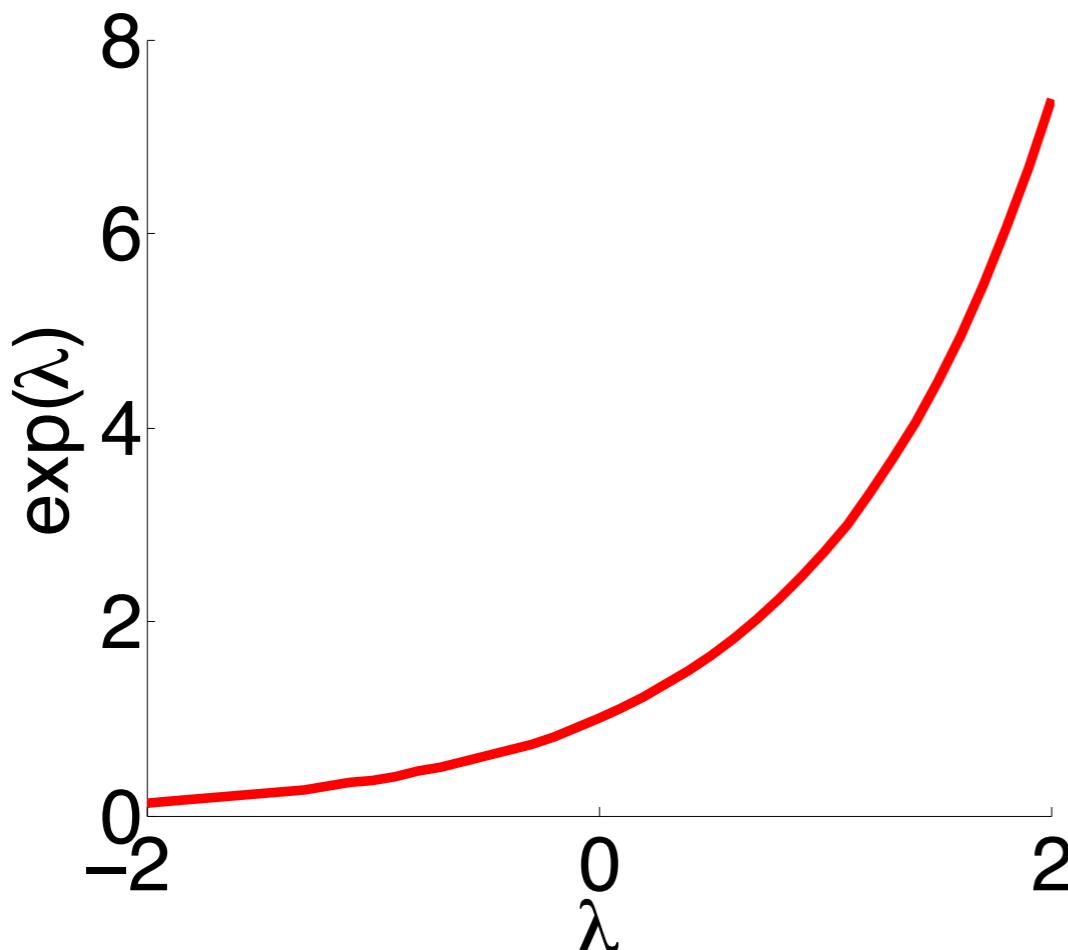
Applying matrix *exponentiated gradient*,
proposed by Tsuda et al. (JMLR 2005)

$$K^{(t+1)} \leftarrow \exp \left(\log(K^{(t)}) + \eta \nabla \mathcal{L}(K^{(t)}) \right)$$

PROPOSED WORK

Applying matrix exponentiated gradient,
proposed by Tsuda et al. (JMLR 2005)

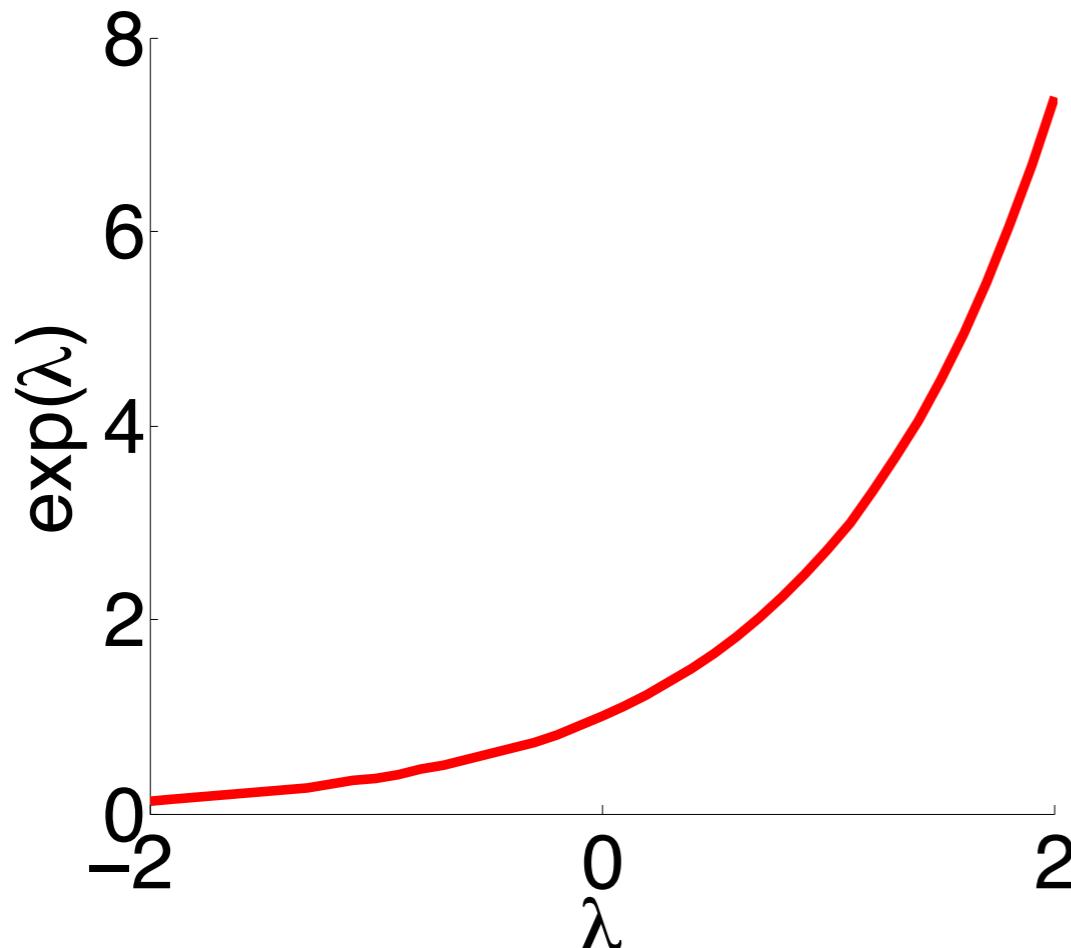
$$K^{(t+1)} \leftarrow \exp \left(\log(K^{(t)}) + \eta \nabla \mathcal{L}(K^{(t)}) \right)$$



PROPOSED WORK

Applying matrix exponentiated gradient,
proposed by Tsuda et al. (JMLR 2005)

$$K^{(t+1)} \leftarrow \exp \left(\log(K^{(t)}) + \eta \nabla \mathcal{L}(K^{(t)}) \right)$$



Relative log-likelihood
 $\frac{|EM-EG|}{|EG|} : 0.57\%$

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

1. Dimensionality Reduction

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

1. Dimensionality Reduction

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

2. MAP Estimation

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

1. Dimensionality Reduction

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

2. MAP Estimation

GILLENWATER, KULESZA, AND TASKAR (NIPS 2012)

3. Likelihood Maximization

GILLENWATER, KULESZA, FOX, AND TASKAR (NIPS 2014)

APPROXIMATE INFERENCE FOR DETERMINANTAL POINT PROCESSES

Thanks to the committee:
Michael Kearns
Sasha Rakhlin
Ali Jadbabaie
Jeff Bilmes