DETERMINANTAL POINT PROCESSES
FOR
NATURAL LANGUAGE PROCESSING

Jennifer Gillenwater
Joint work with Alex Kulesza and Ben Taskar

OUTLINE

OUTLINE

Motivation & background on DPPs

OUTLINE

Motivation & background on DPPs

Large-scale settings

OUTLINE

Motivation & background on DPPs

Large-scale settings

Structured summarization

OUTLINE

Motivation & background on DPPs

Large-scale settings

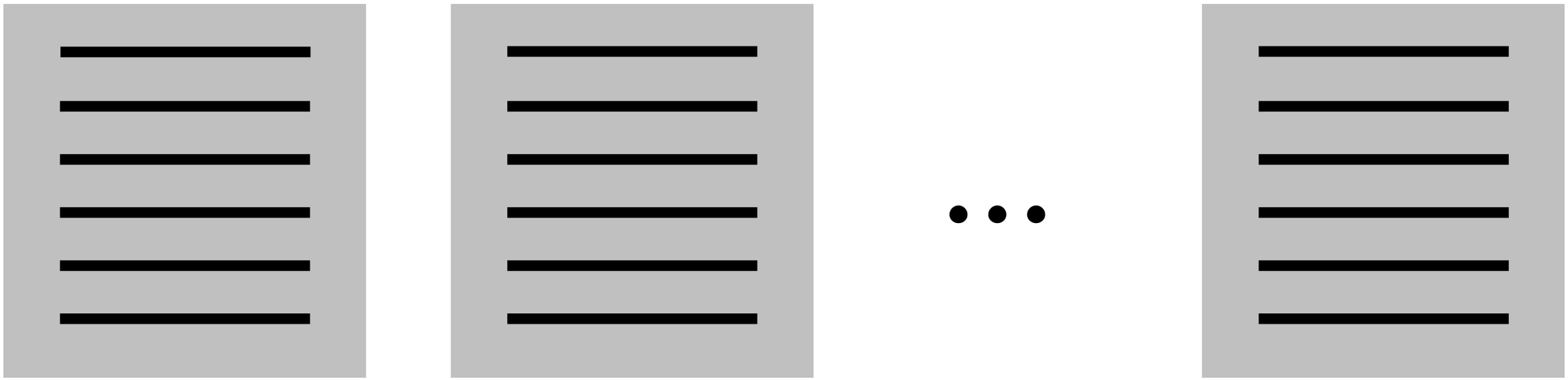
Structured summarization

Other potential NLP applications

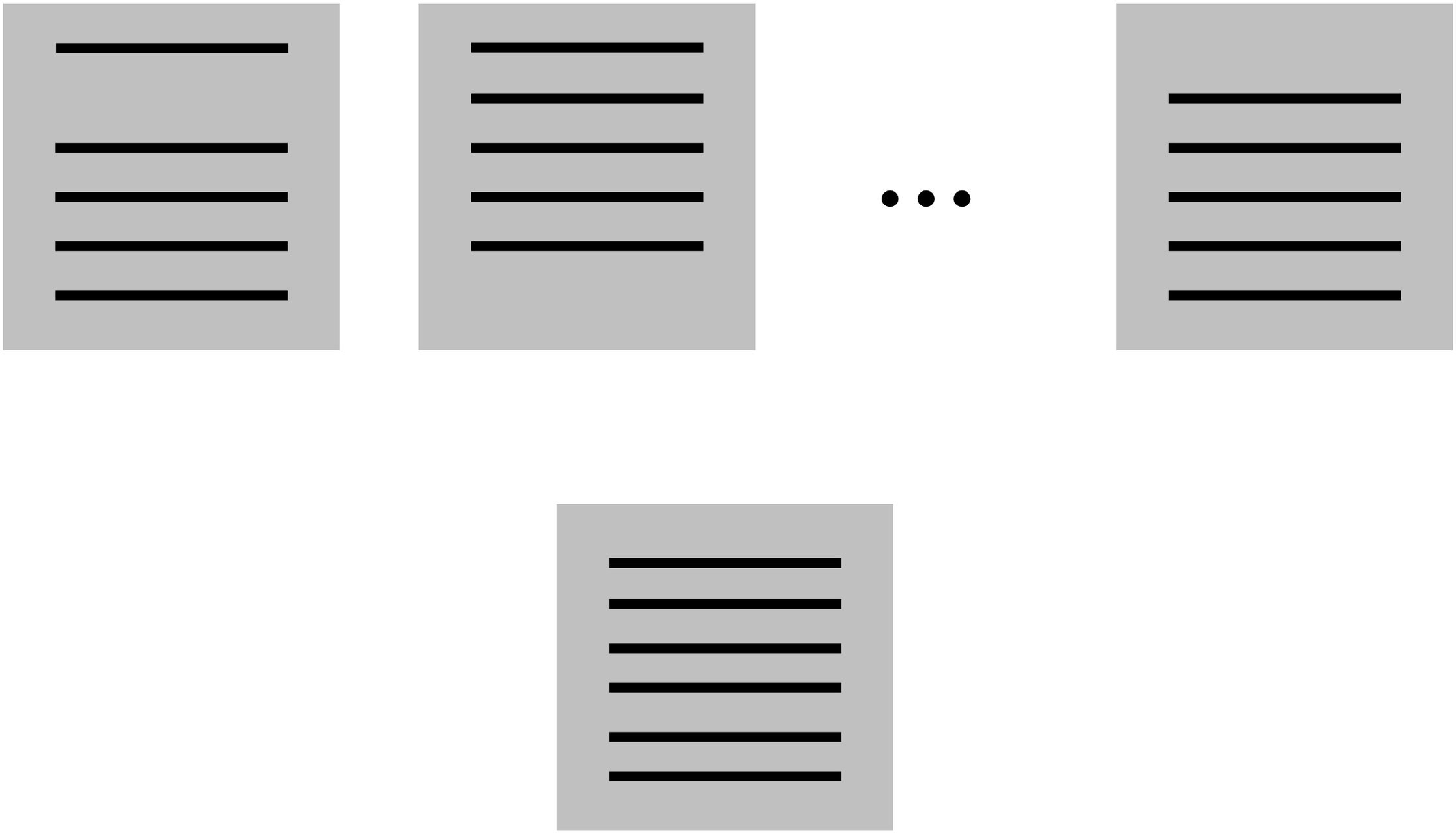
MOTIVATION & BACKGROUND

SUMMARIZATION

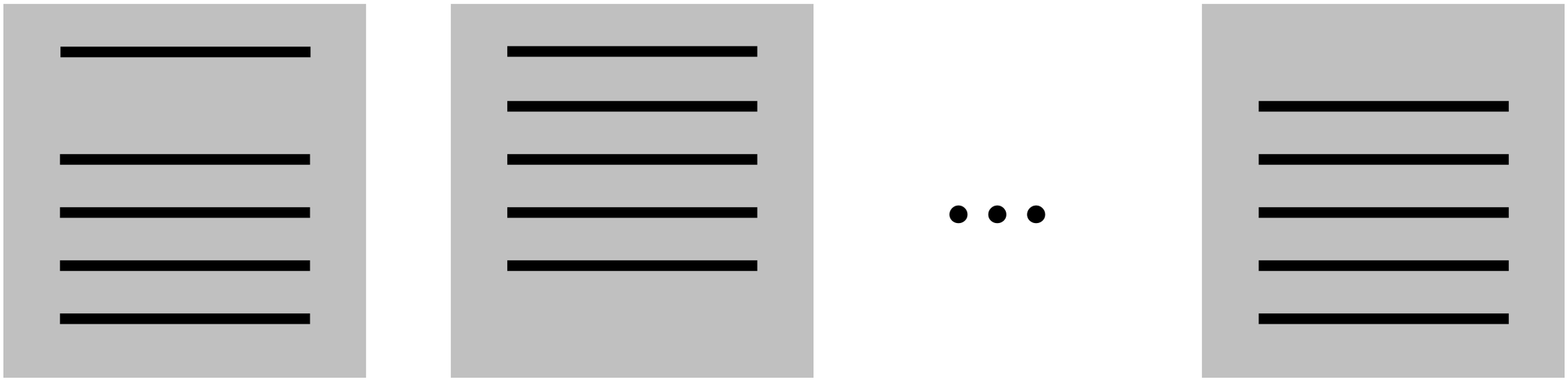
SUMMARIZATION



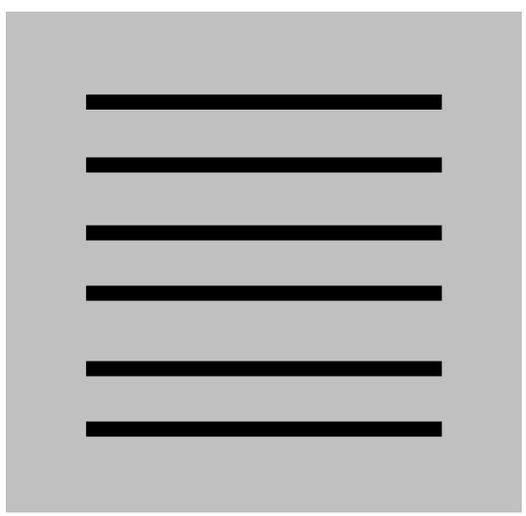
SUMMARIZATION



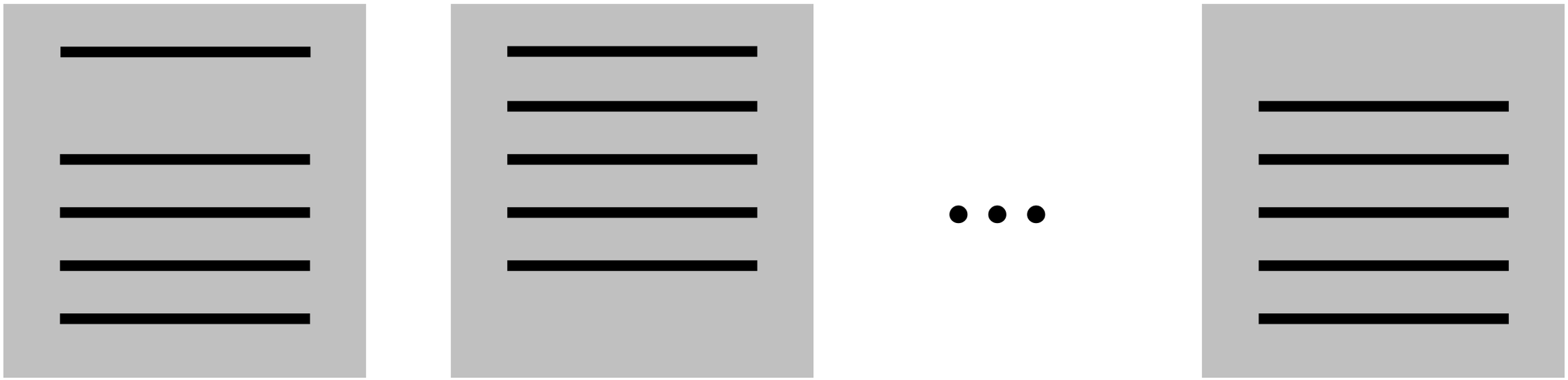
SUMMARIZATION



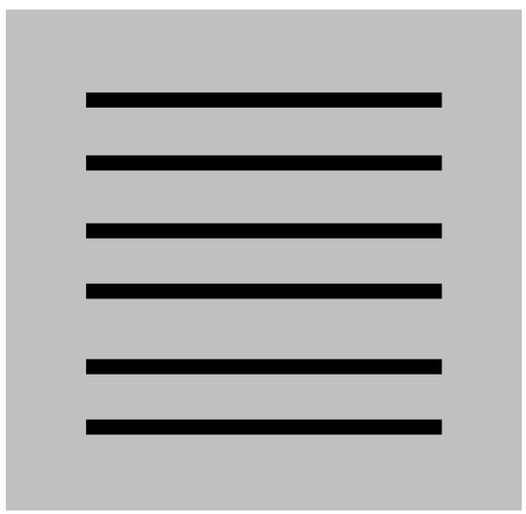
Quality:
relevance to
the topic



SUMMARIZATION



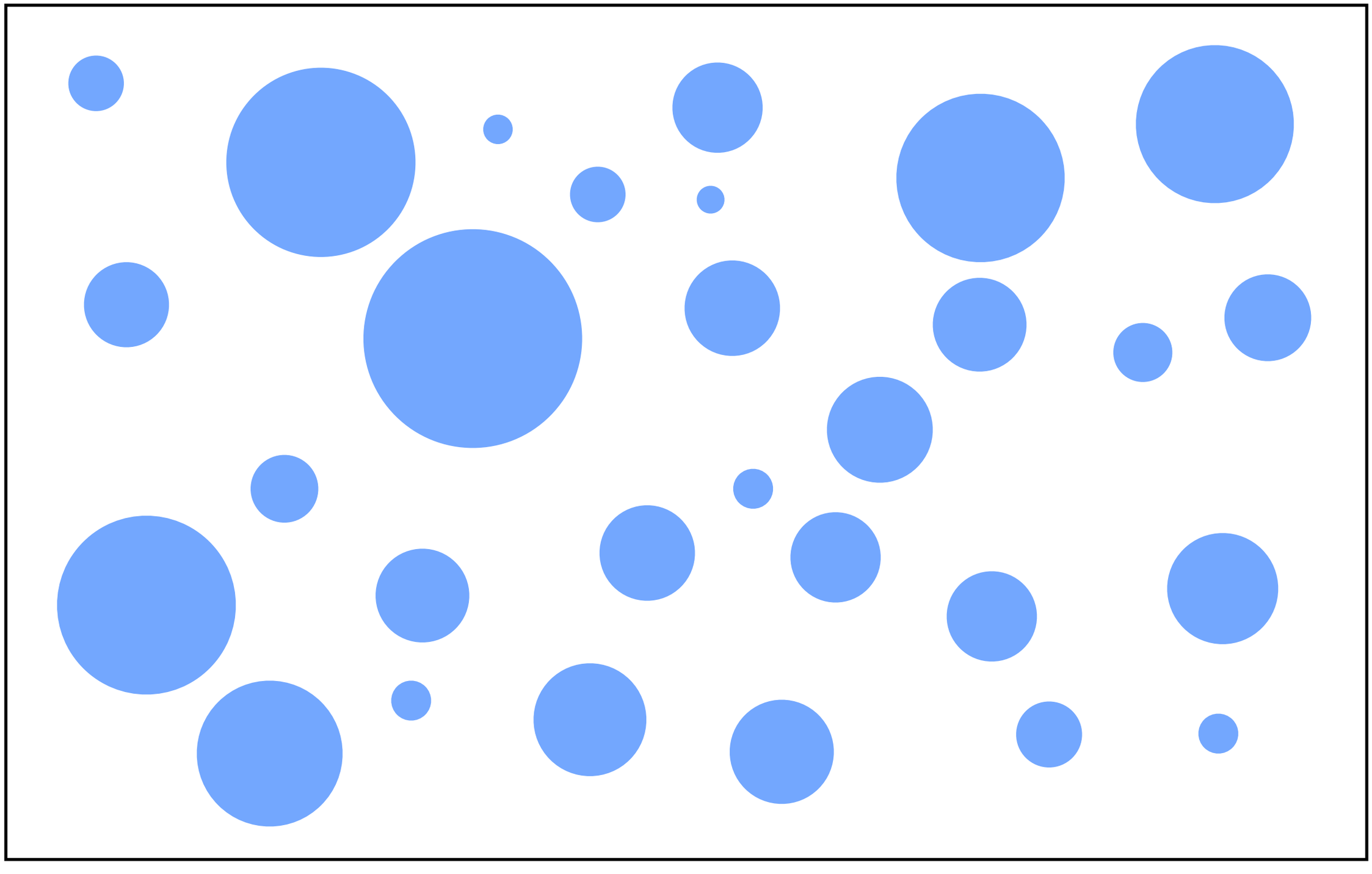
Quality:
relevance to
the topic



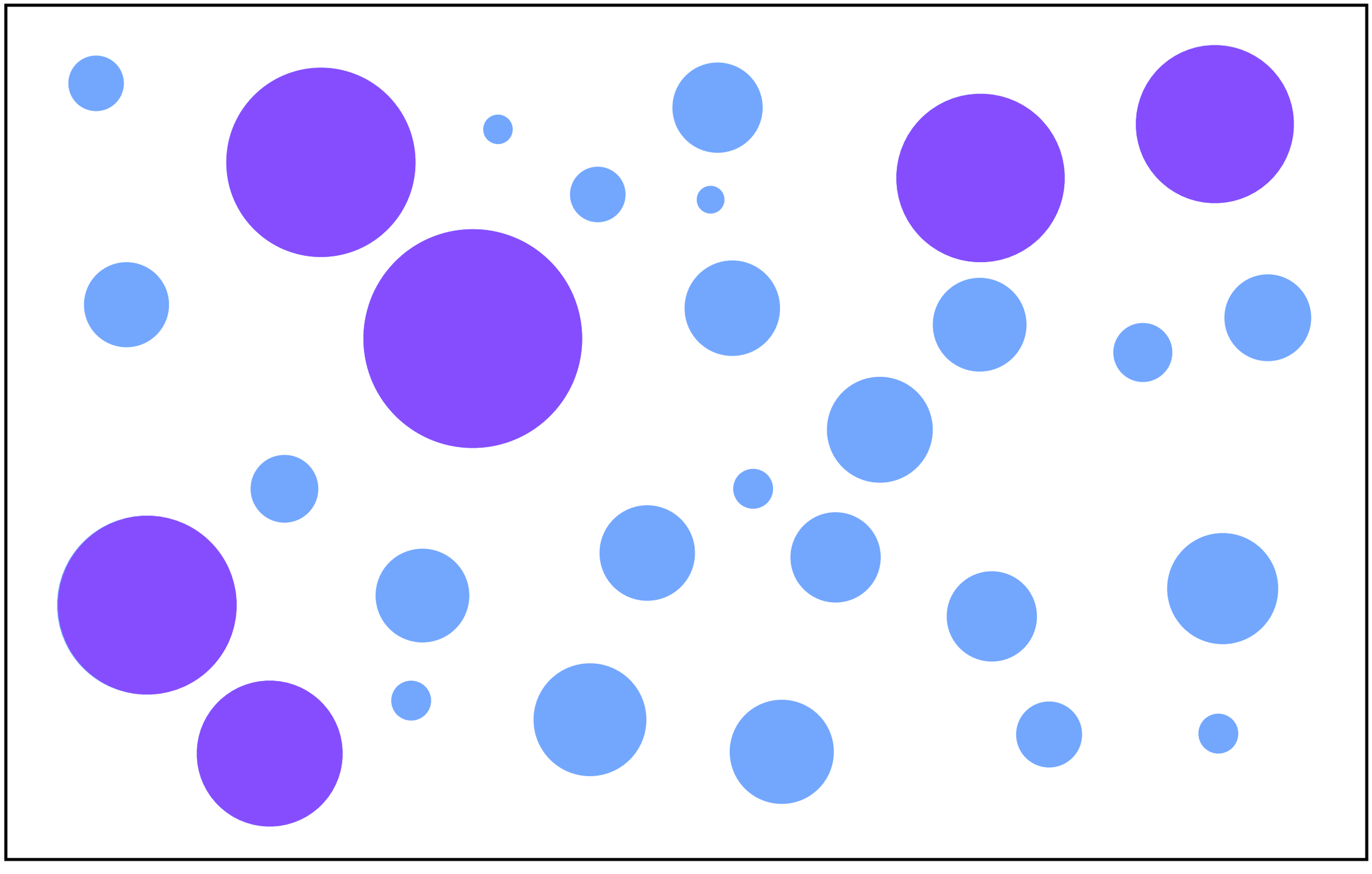
Diversity:
coverage of
core ideas

SUBSET SELECTION

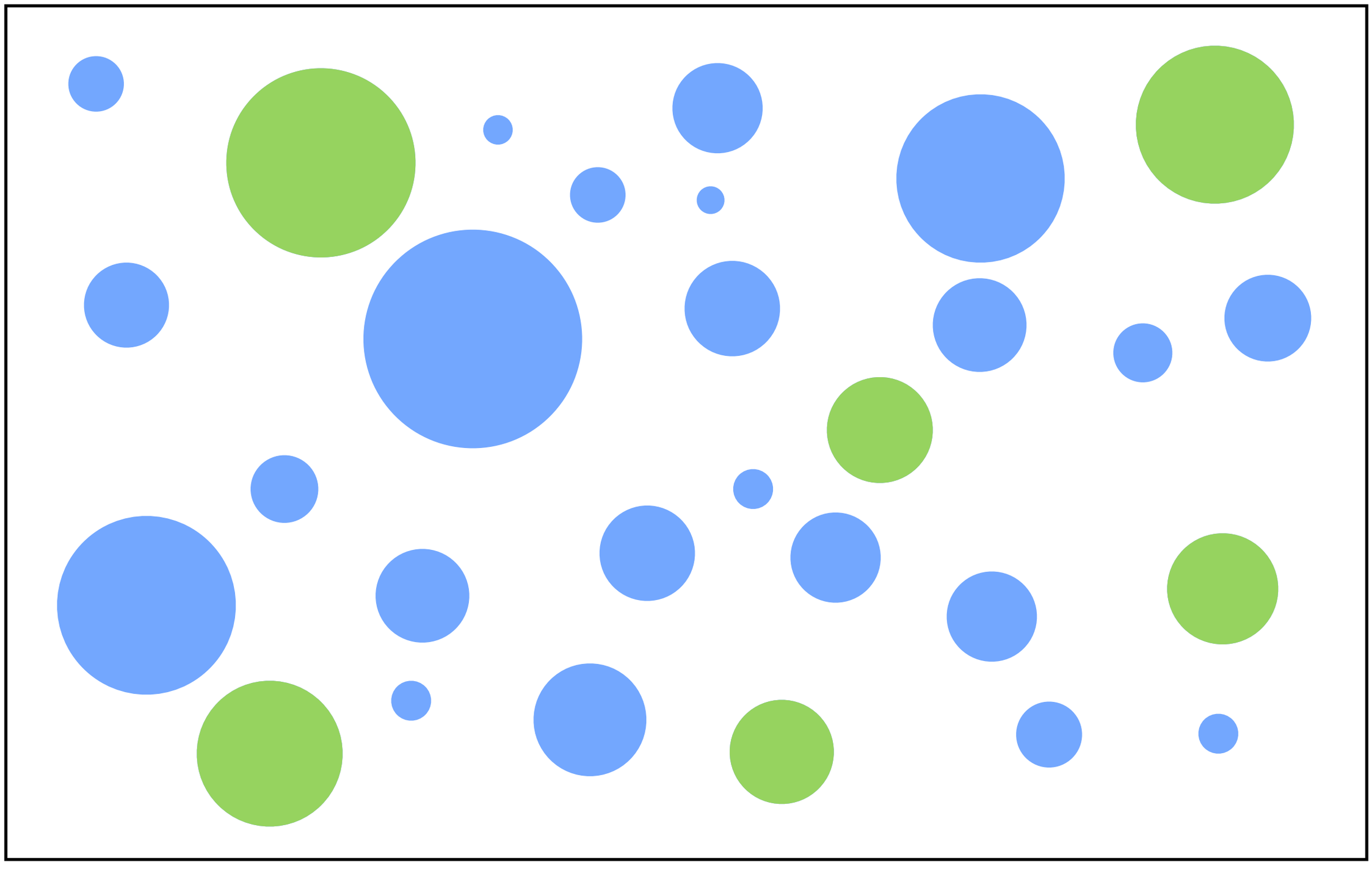
SUBSET SELECTION



SUBSET SELECTION

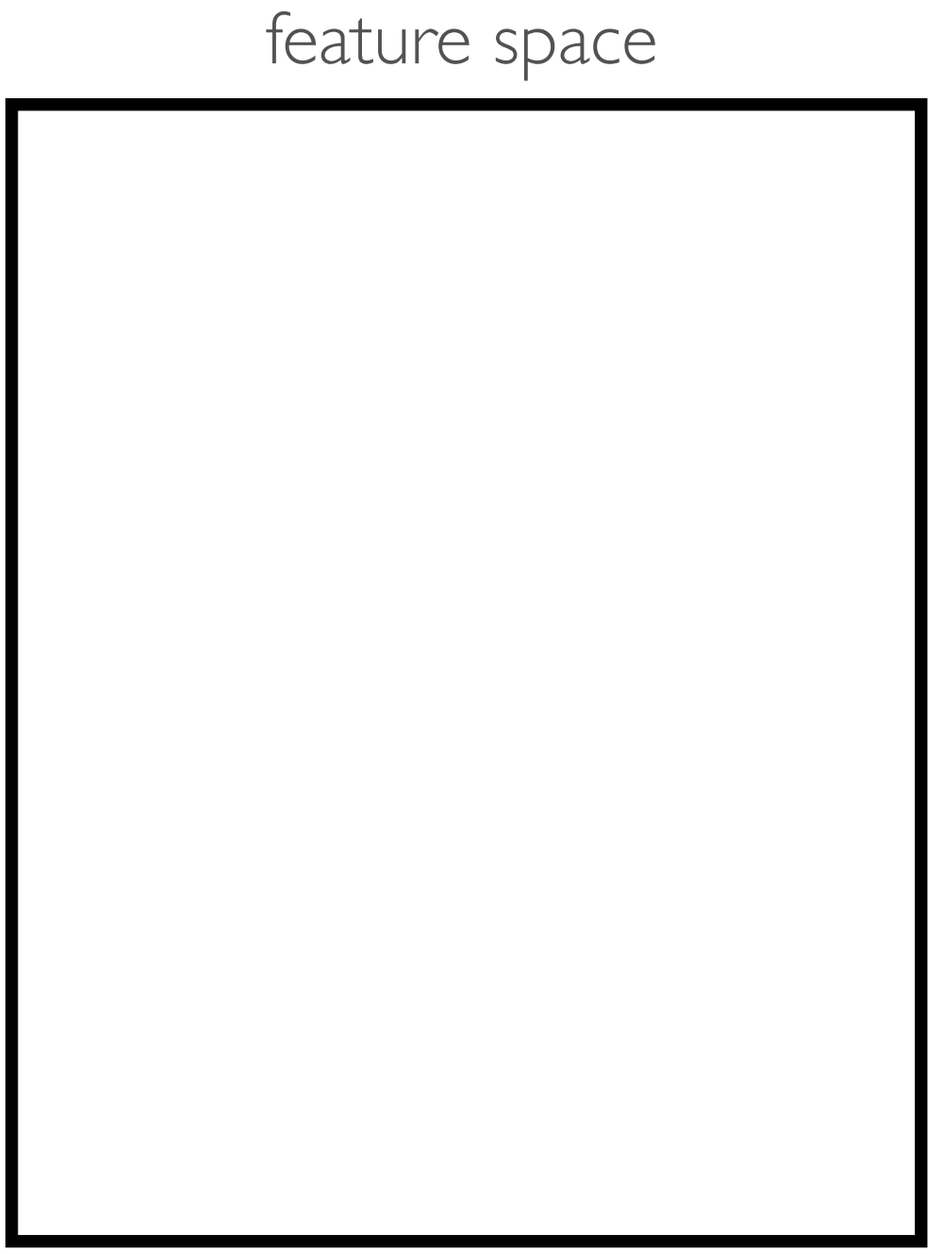


SUBSET SELECTION

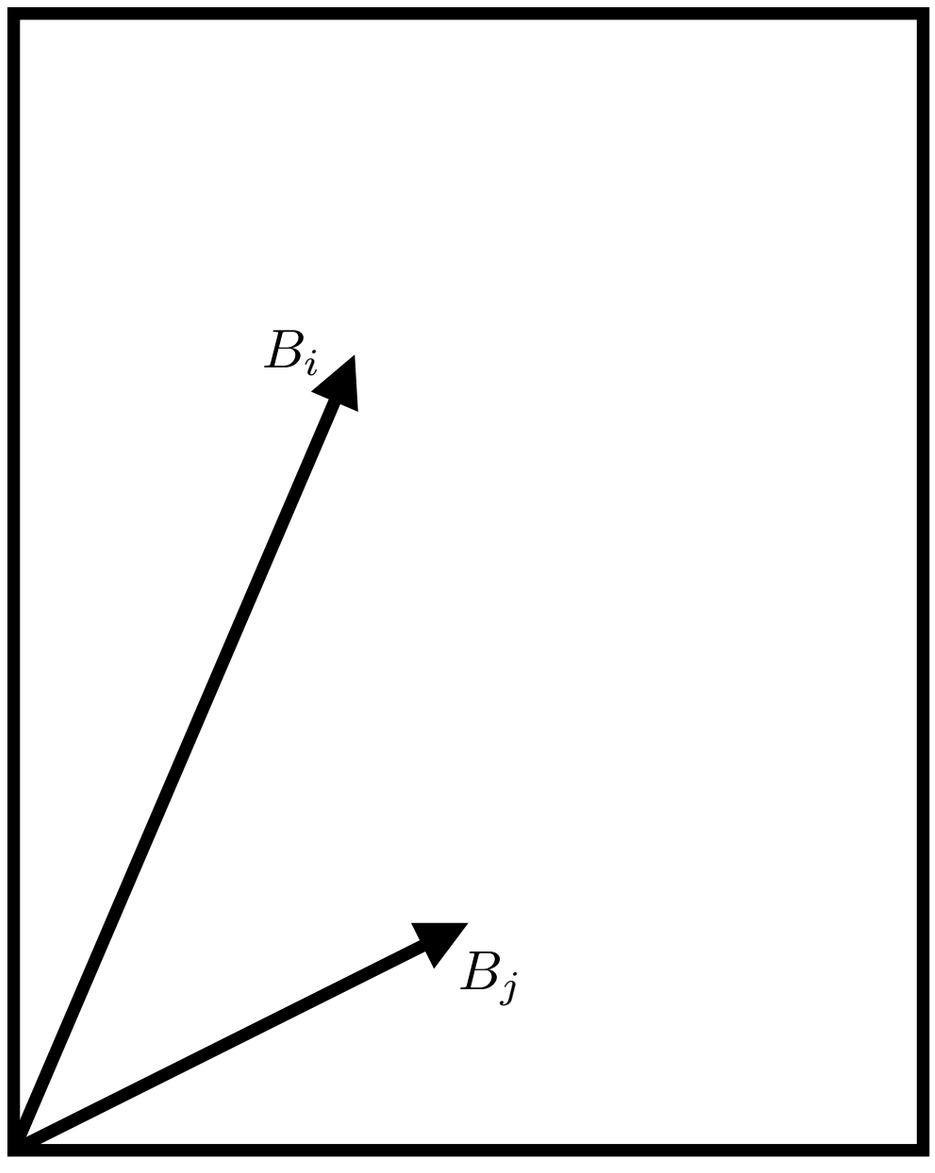


AREA AS SET-GOODNESS

AREA AS SET-GOODNESS

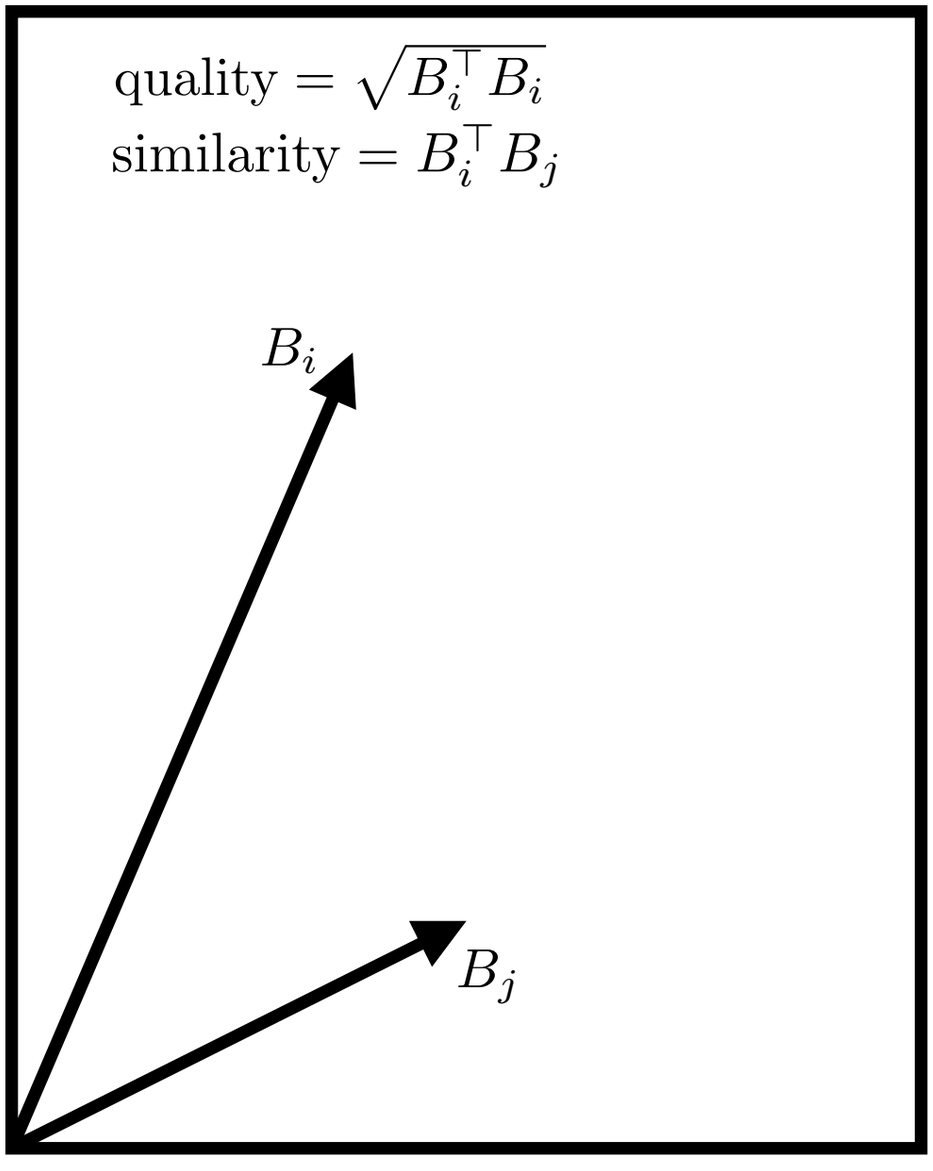


AREA AS SET-GOODNESS

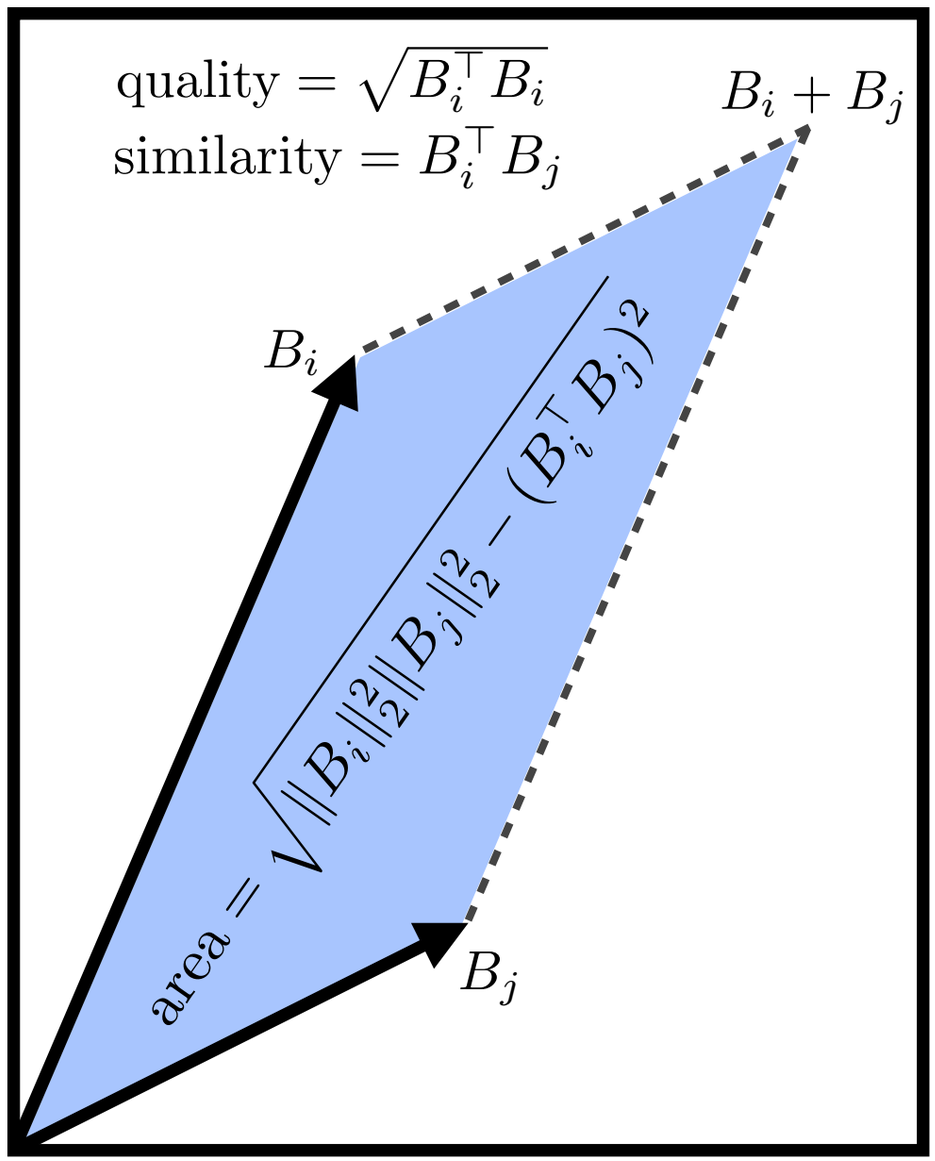


AREA AS SET-GOODNESS

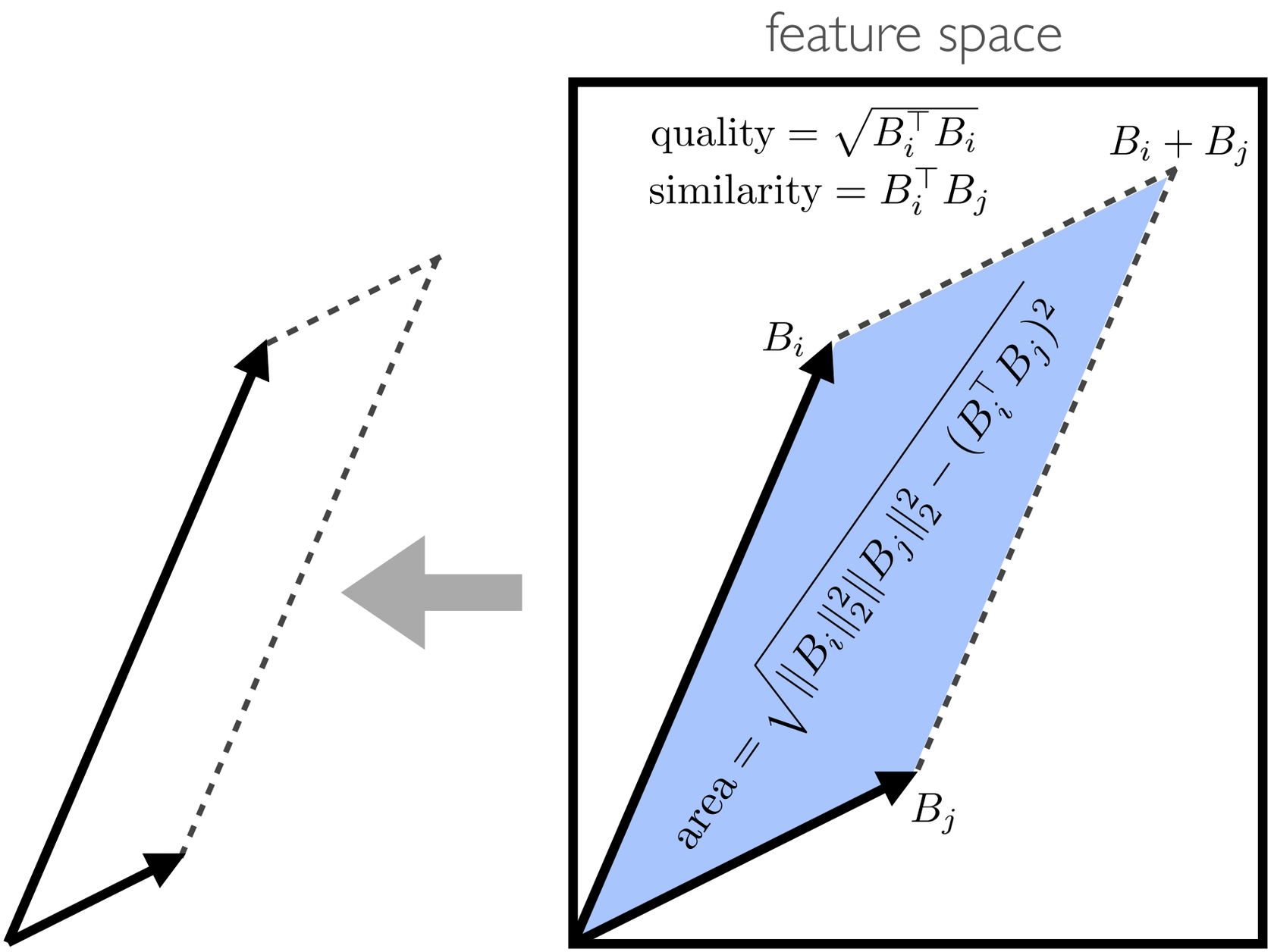
feature space



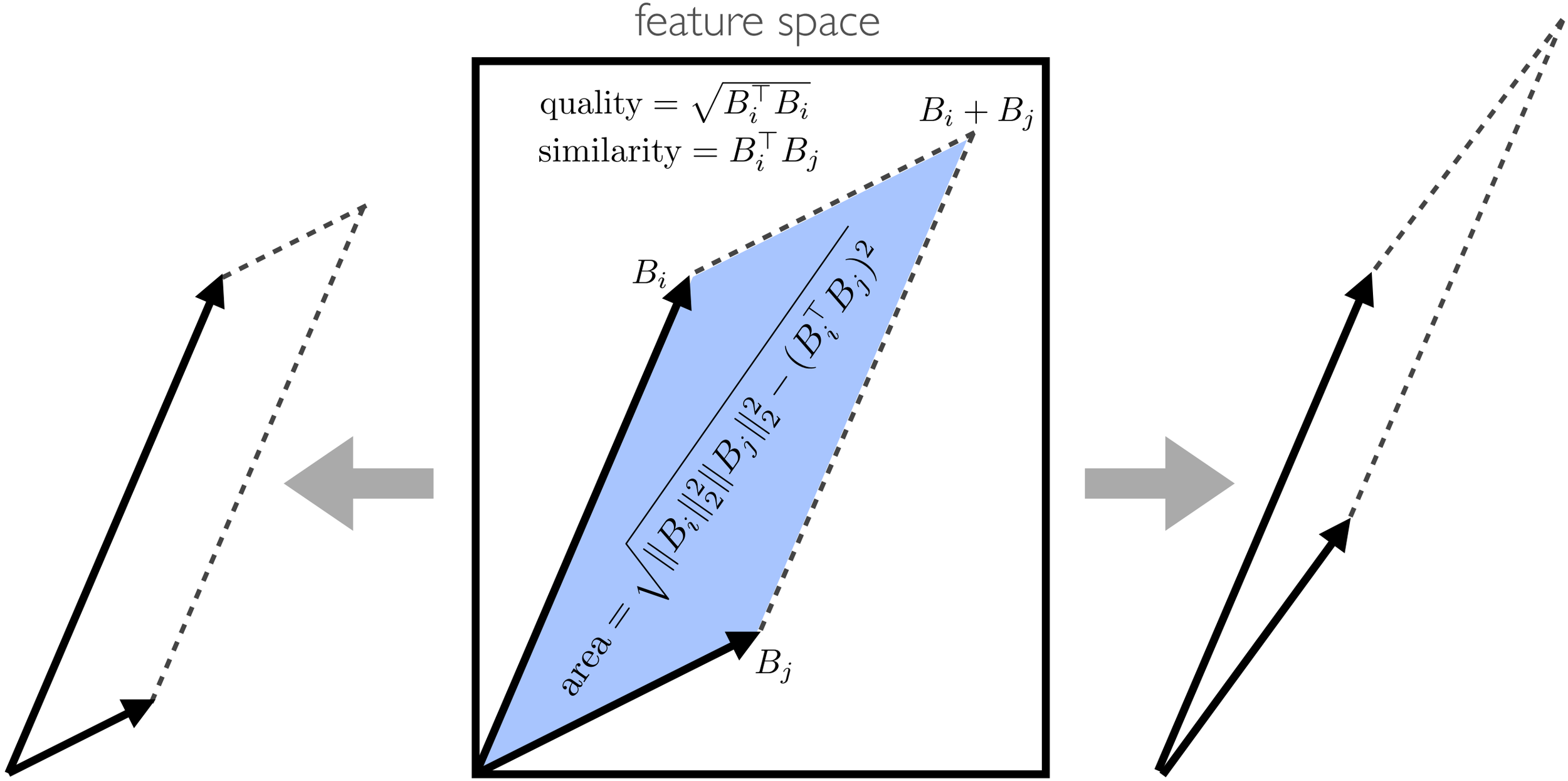
AREA AS SET-GOODNESS



AREA AS SET-GOODNESS



AREA AS SET-GOODNESS

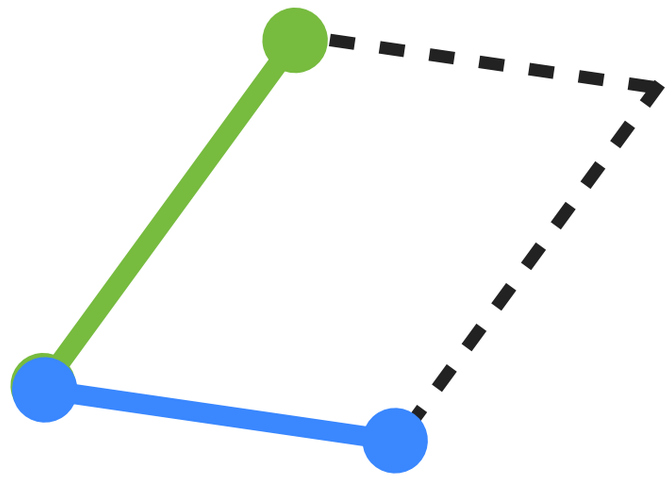


VOLUME AS SET-GOODNESS

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

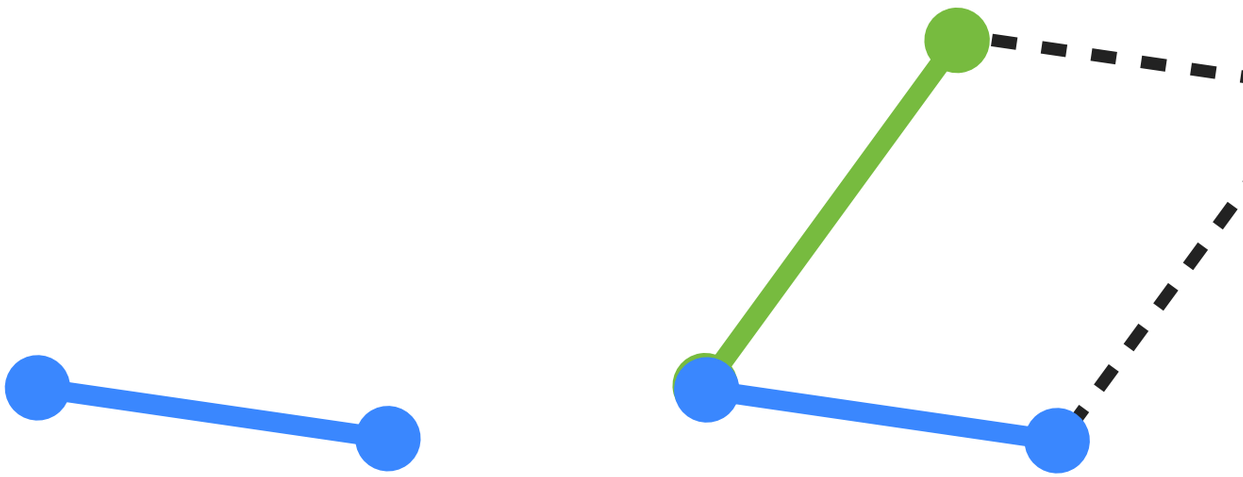
VOLUME AS SET-GOODNESS

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



VOLUME AS SET-GOODNESS

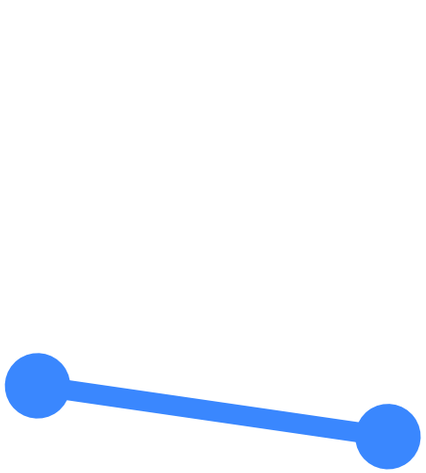
$$\text{area} = \sqrt{\|B_i\|_2^2 \|B_j\|_2^2 - (B_i^\top B_j)^2}$$



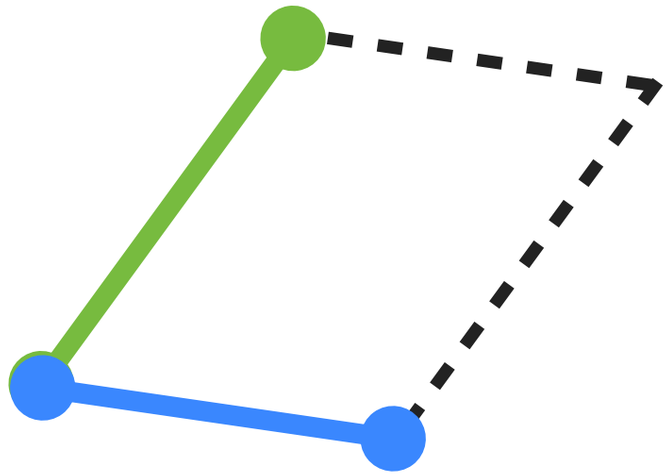
$$\text{length} = \|B_i\|_2$$

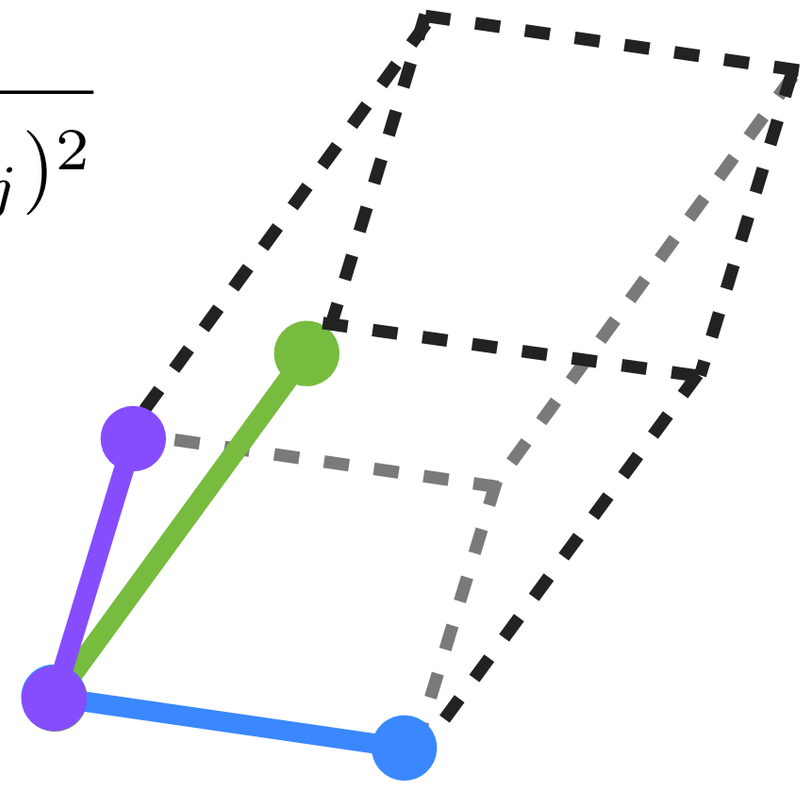
VOLUME AS SET-GOODNESS

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



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

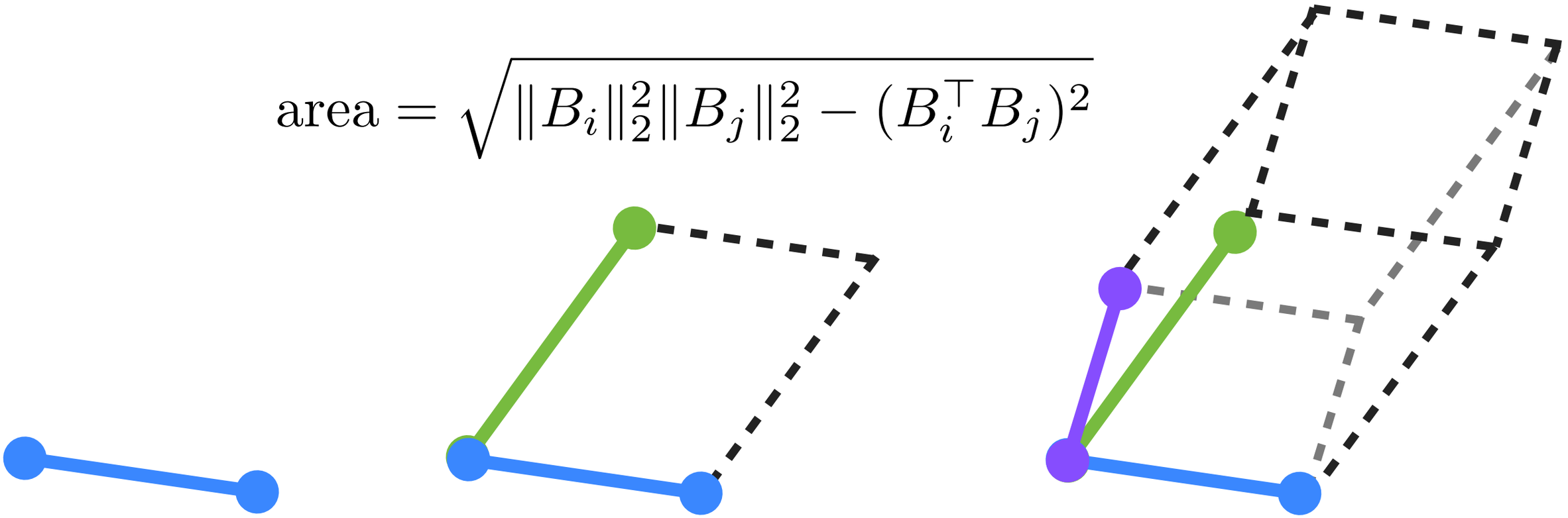




$$\text{volume} = \text{base} \times \text{height}$$

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

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

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

$$= \det \left(\begin{array}{cc} \multicolumn{2}{c}{\multirow{2}{*}{$\begin{matrix} \text{—} & B_i & \text{—} \\ \text{—} & B_j & \text{—} \end{matrix}$}} & \\ & & \begin{matrix} \vert & \vert \\ B_i & B_j \\ \vert & \vert \end{matrix} \end{array} \right)^{\frac{1}{2}}$$

VOLUME AS A DET

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

VOLUME AS A DET

$$\mathrm{vol}(B_{\{i,j\}}) = \det\left(\begin{array}{cc} \multicolumn{2}{c}{B_i} \\ \multicolumn{2}{c}{B_j} \end{array} \right)^{\frac{1}{2}}$$

$$\mathrm{vol}(B) = \det\left(\begin{array}{c} B_1 \\ \vdots \\ B_N \end{array} \right)^{\frac{1}{2}}$$

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

COMPLEX STATISTICS

COMPLEX STATISTICS

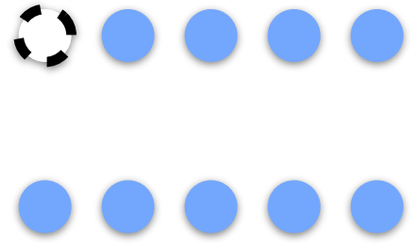
COMPLEX STATISTICS

COMPLEX STATISTICS

$$\sum$$

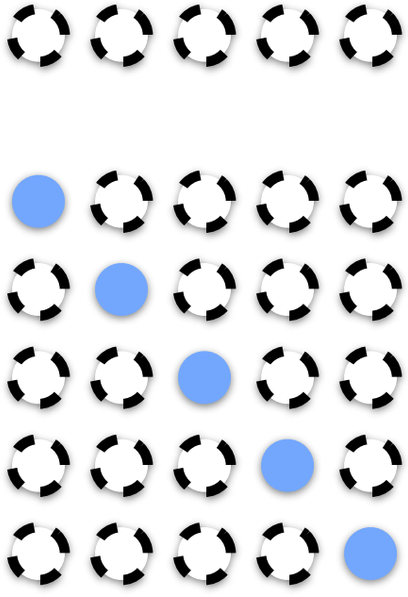
COMPLEX STATISTICS

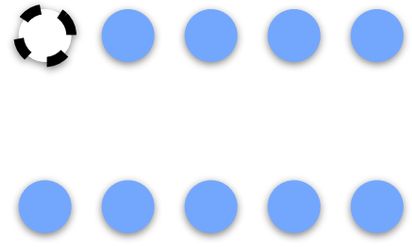
$$\sum$$



COMPLEX STATISTICS

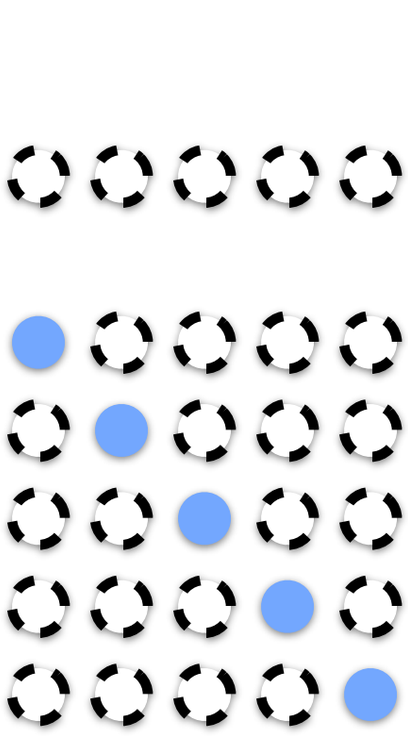
$$\sum$$

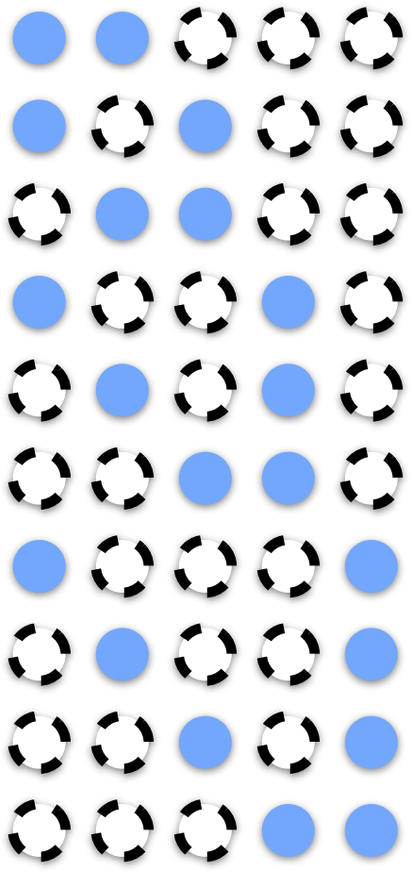


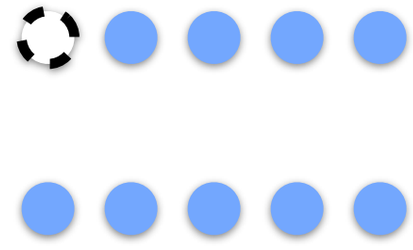


COMPLEX STATISTICS

$$\sum$$

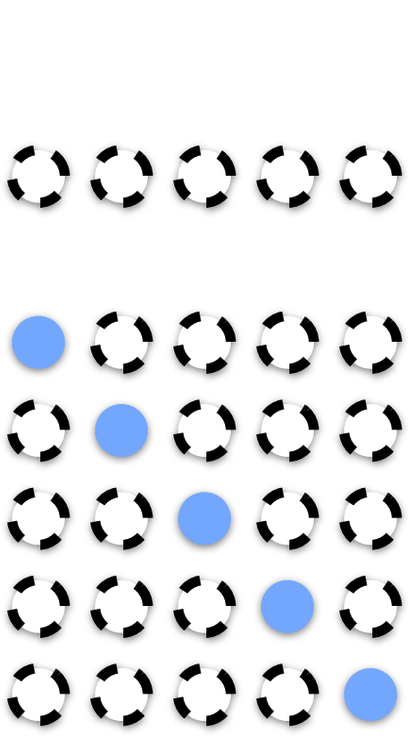


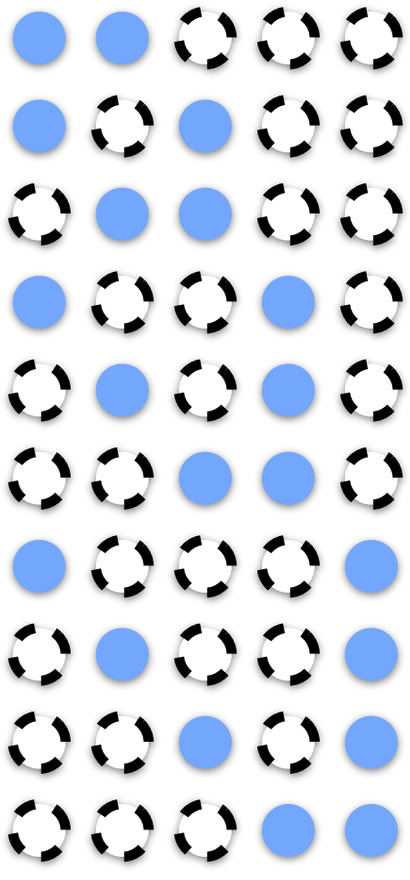


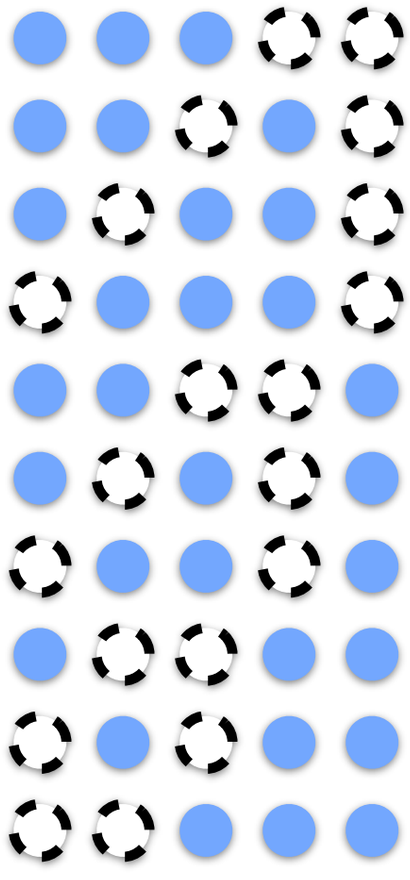


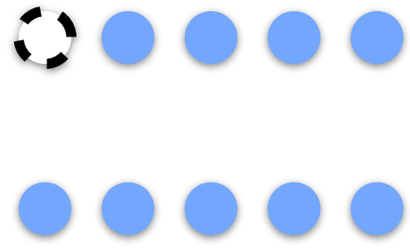
COMPLEX STATISTICS

$$\sum$$



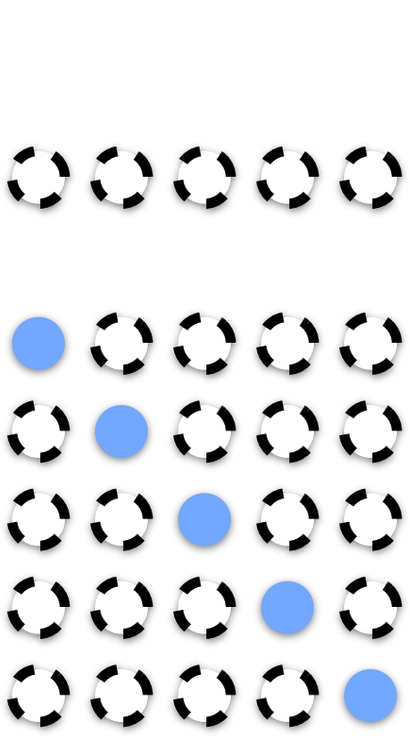


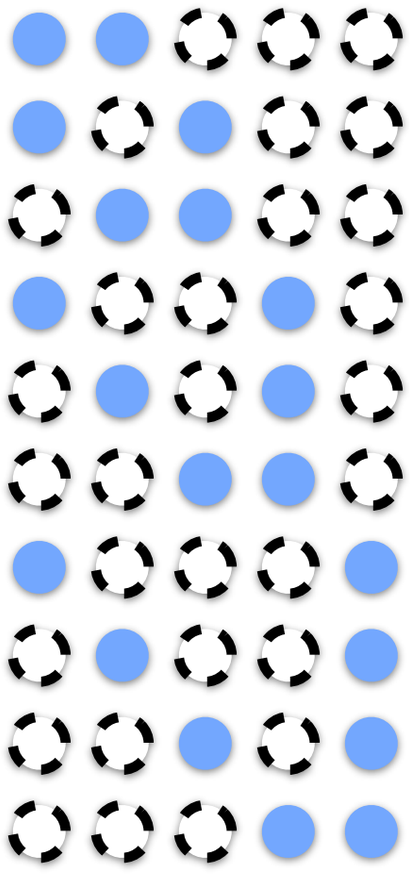


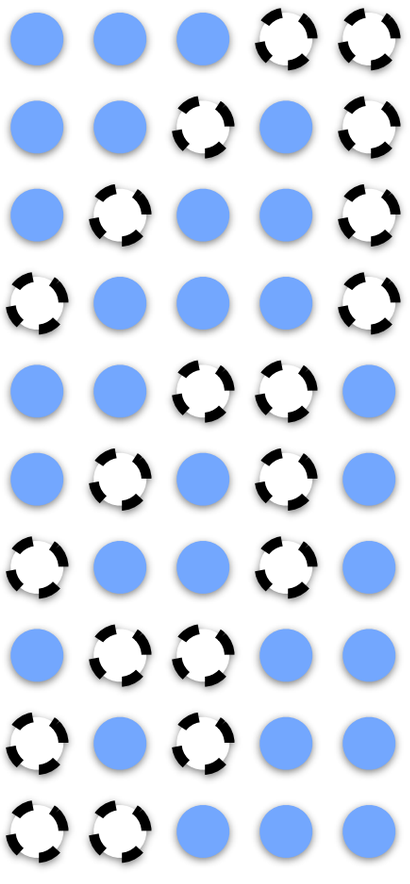


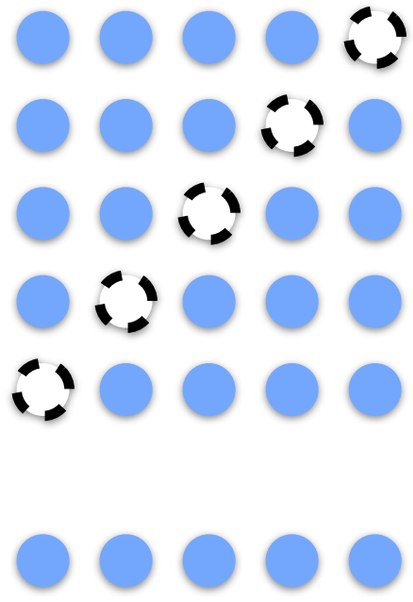
COMPLEX STATISTICS

$$\sum$$



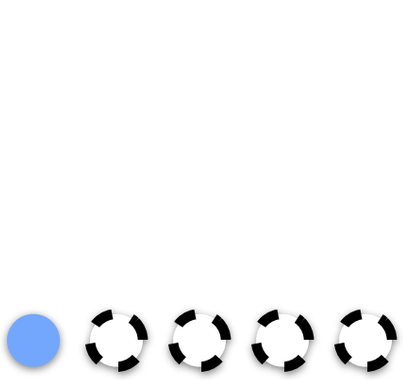


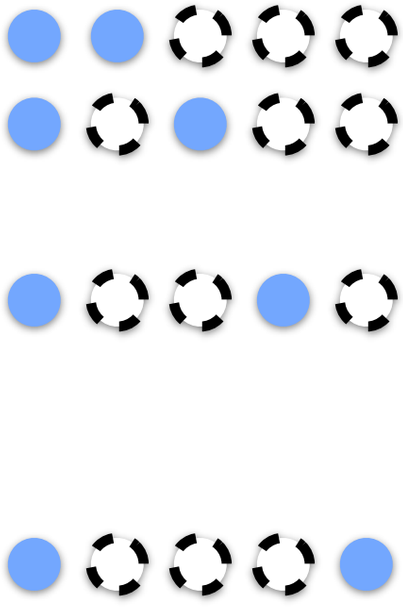


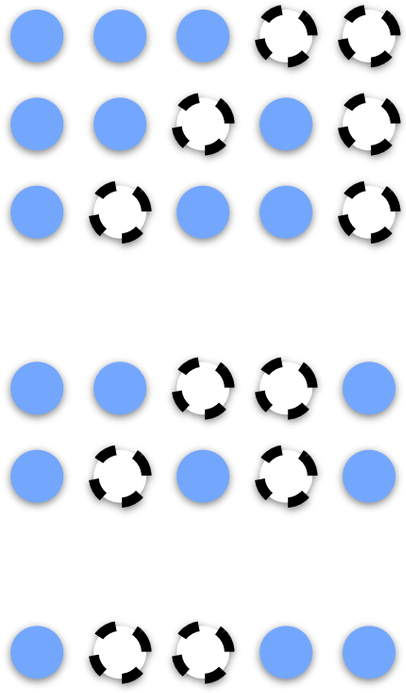


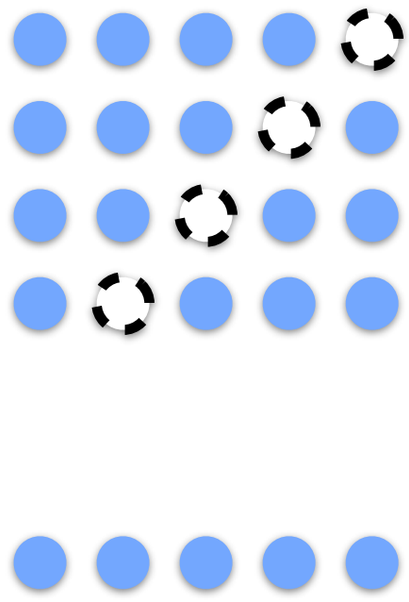
COMPLEX STATISTICS

$$\sum$$



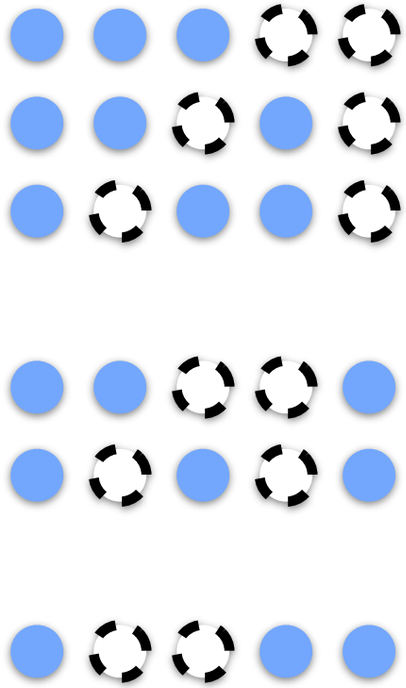






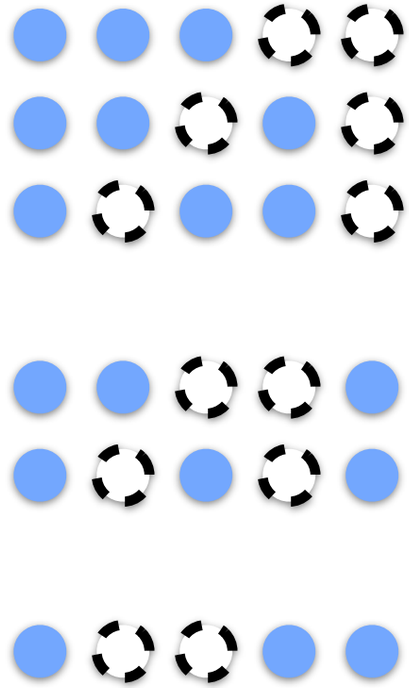
COMPLEX STATISTICS

$$\sum$$



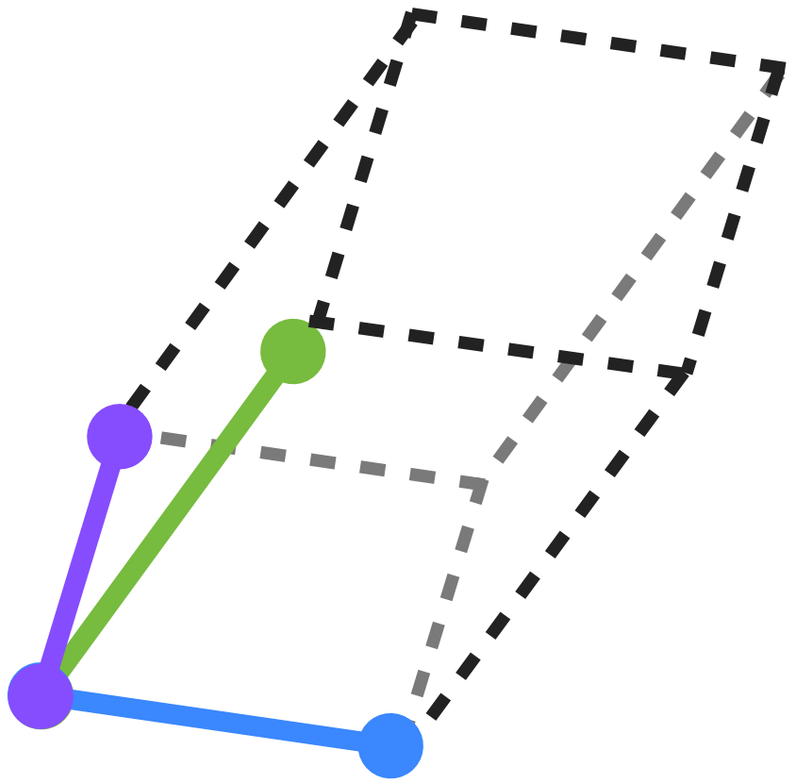
COMPLEX STATISTICS

$$N \text{ items} \implies 2^N \text{ sets}$$



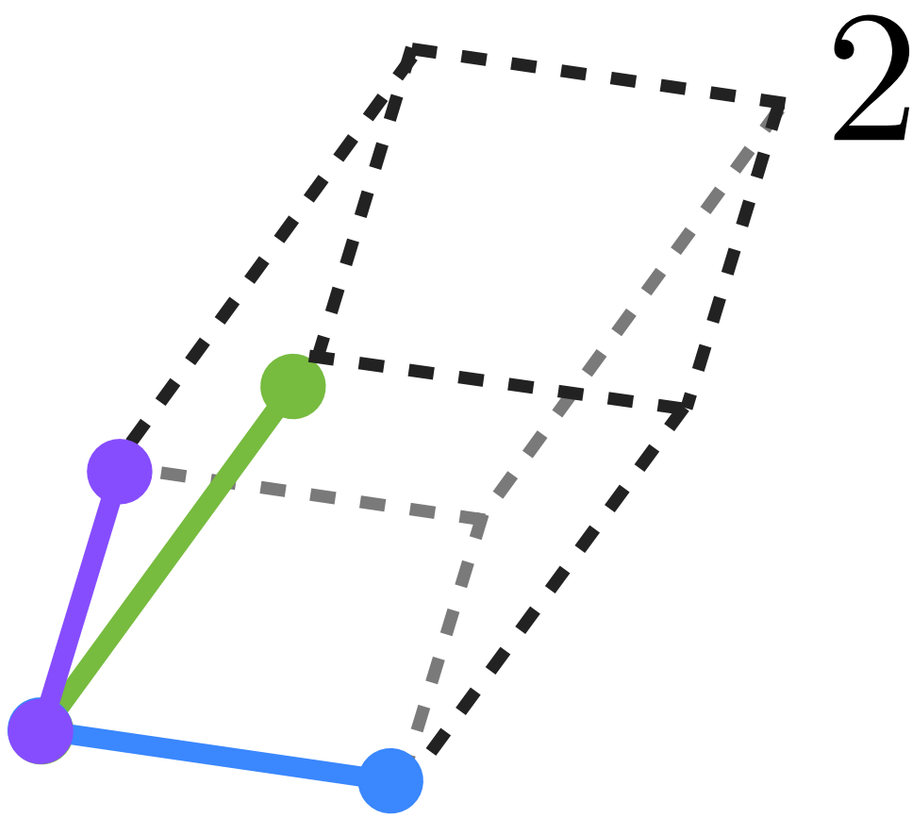
EFFICIENT COMPUTATION

EFFICIENT COMPUTATION



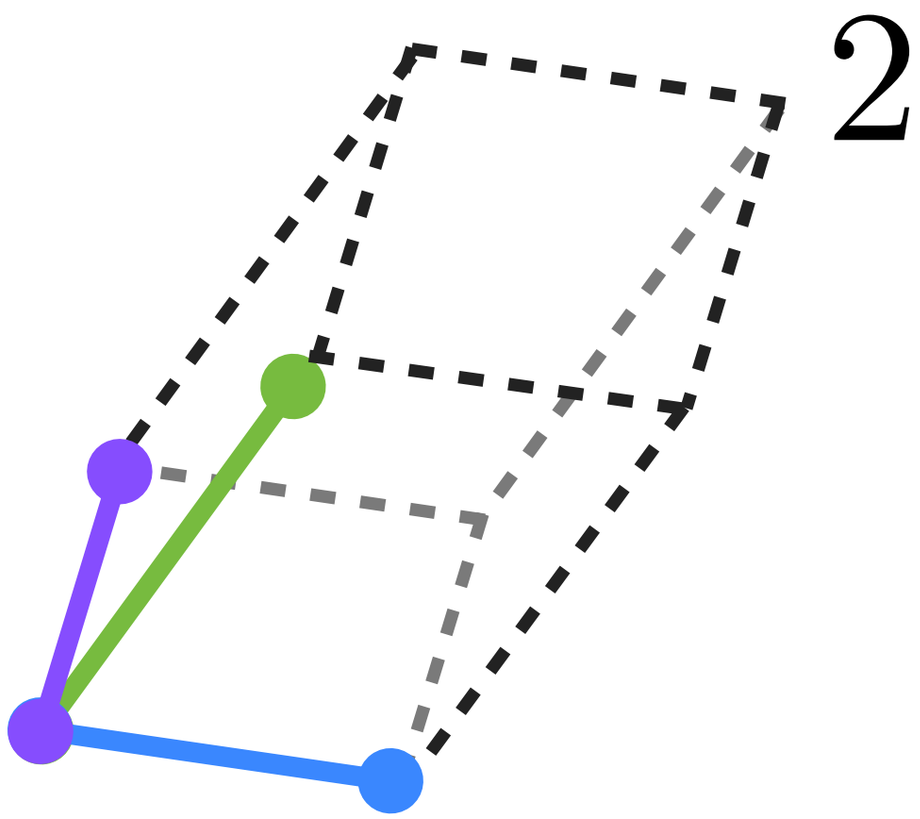
$$\det{}^{\frac{1}{2}}$$

EFFICIENT COMPUTATION



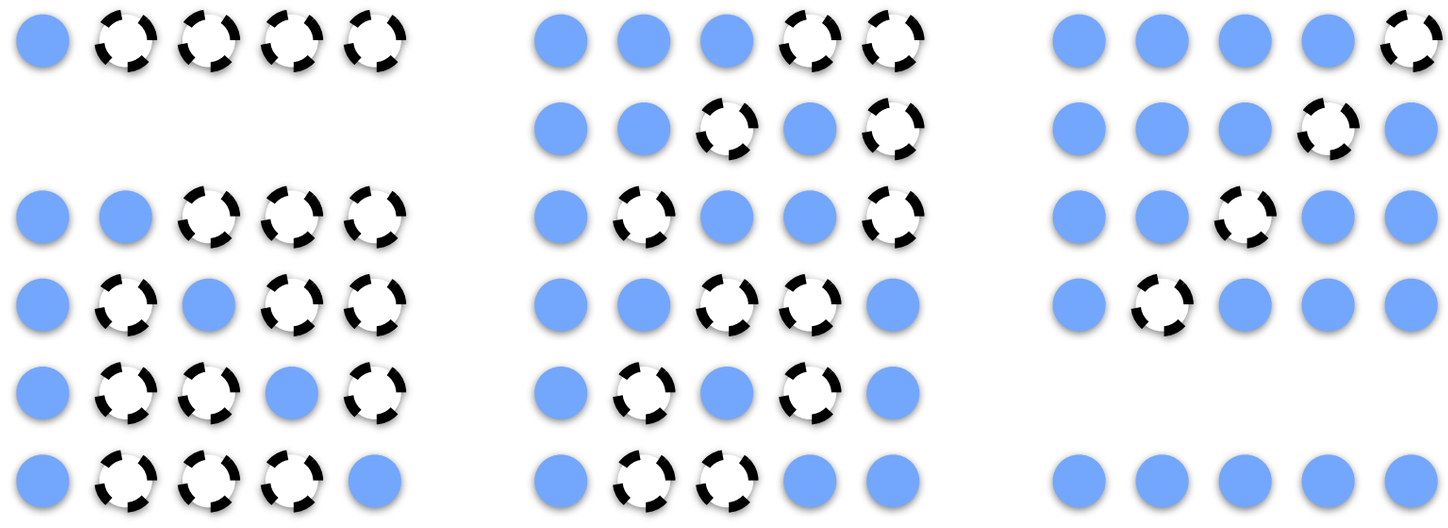
$$\det$$

EFFICIENT COMPUTATION

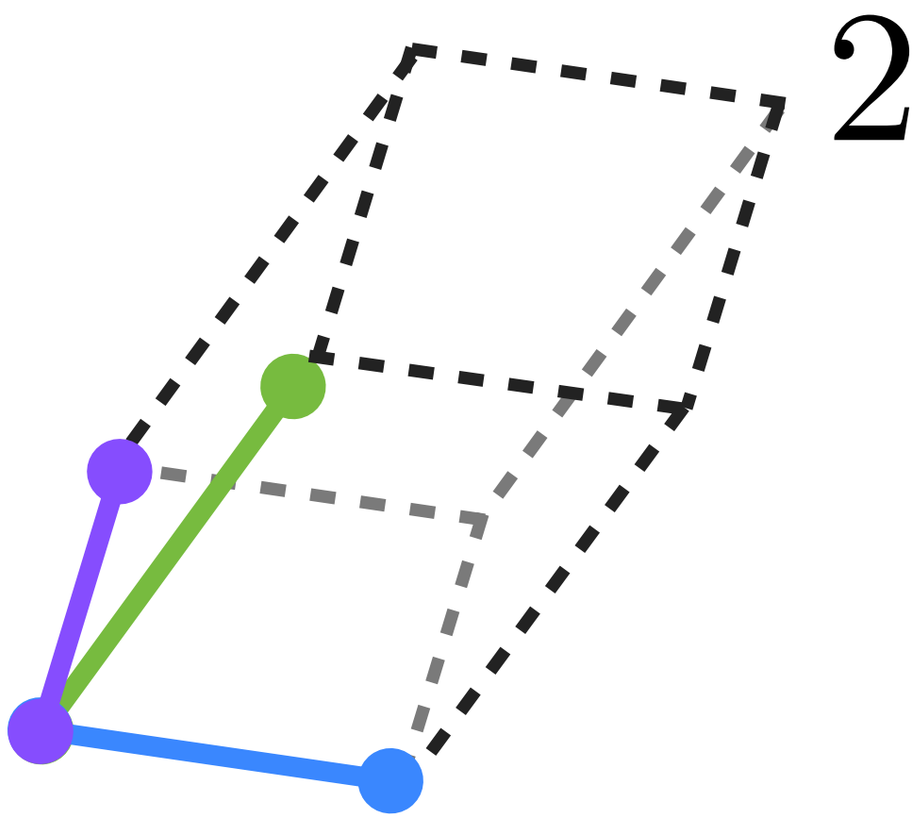


$$\det$$

$$\sum$$

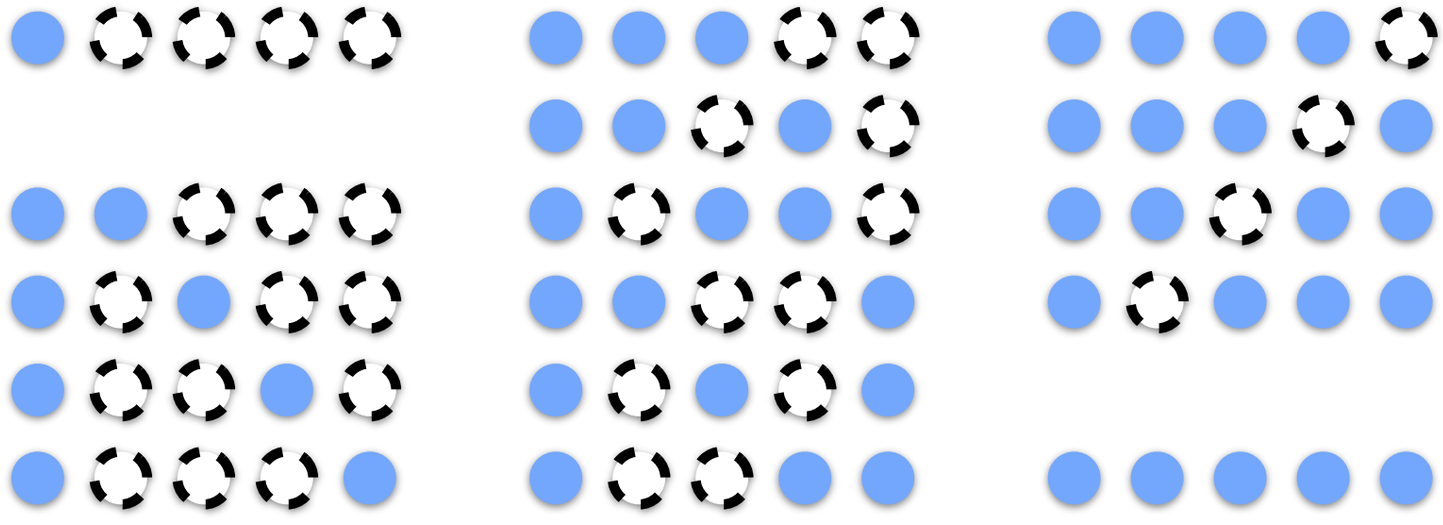


EFFICIENT COMPUTATION



$$\det$$

$$O(N^3)$$

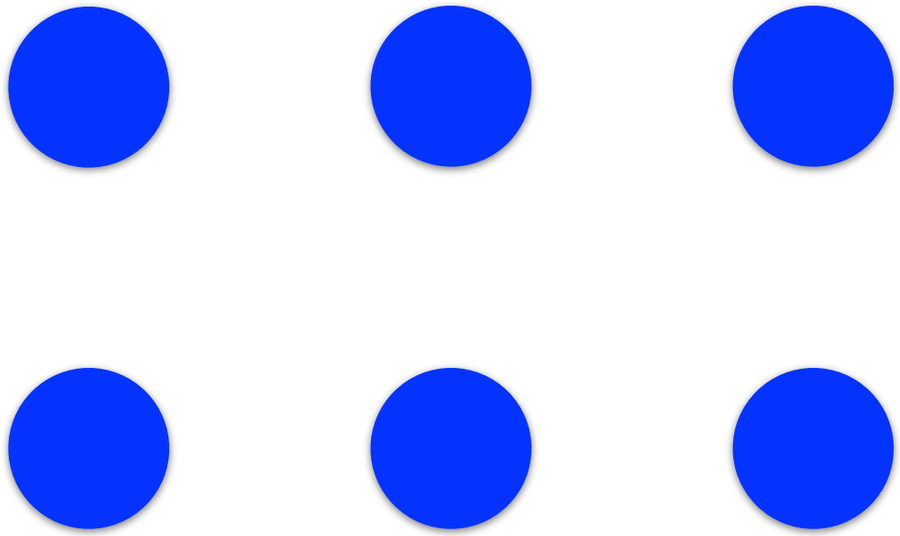


POINT PROCESSES

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

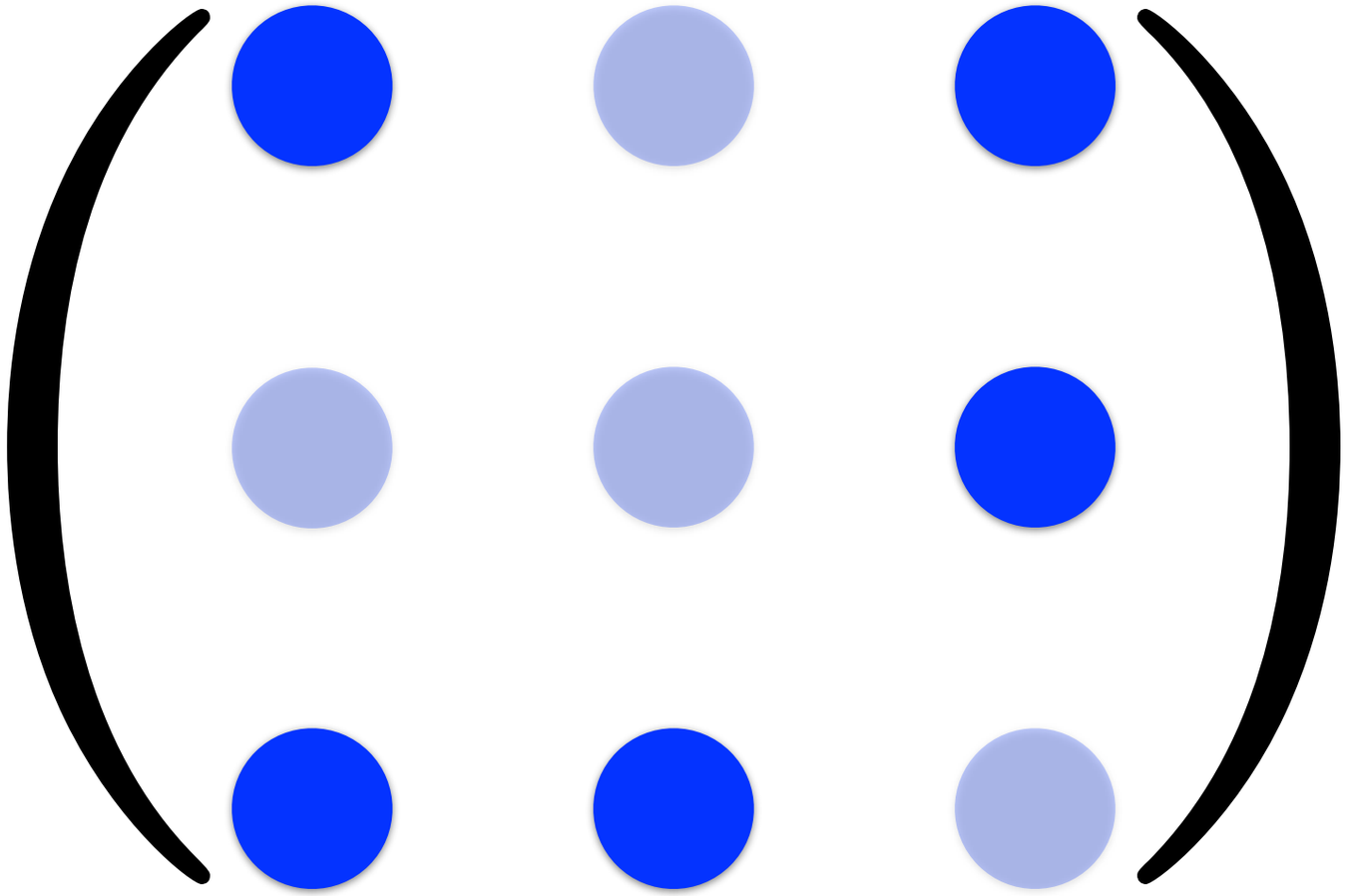
POINT PROCESSES

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



POINT PROCESSES

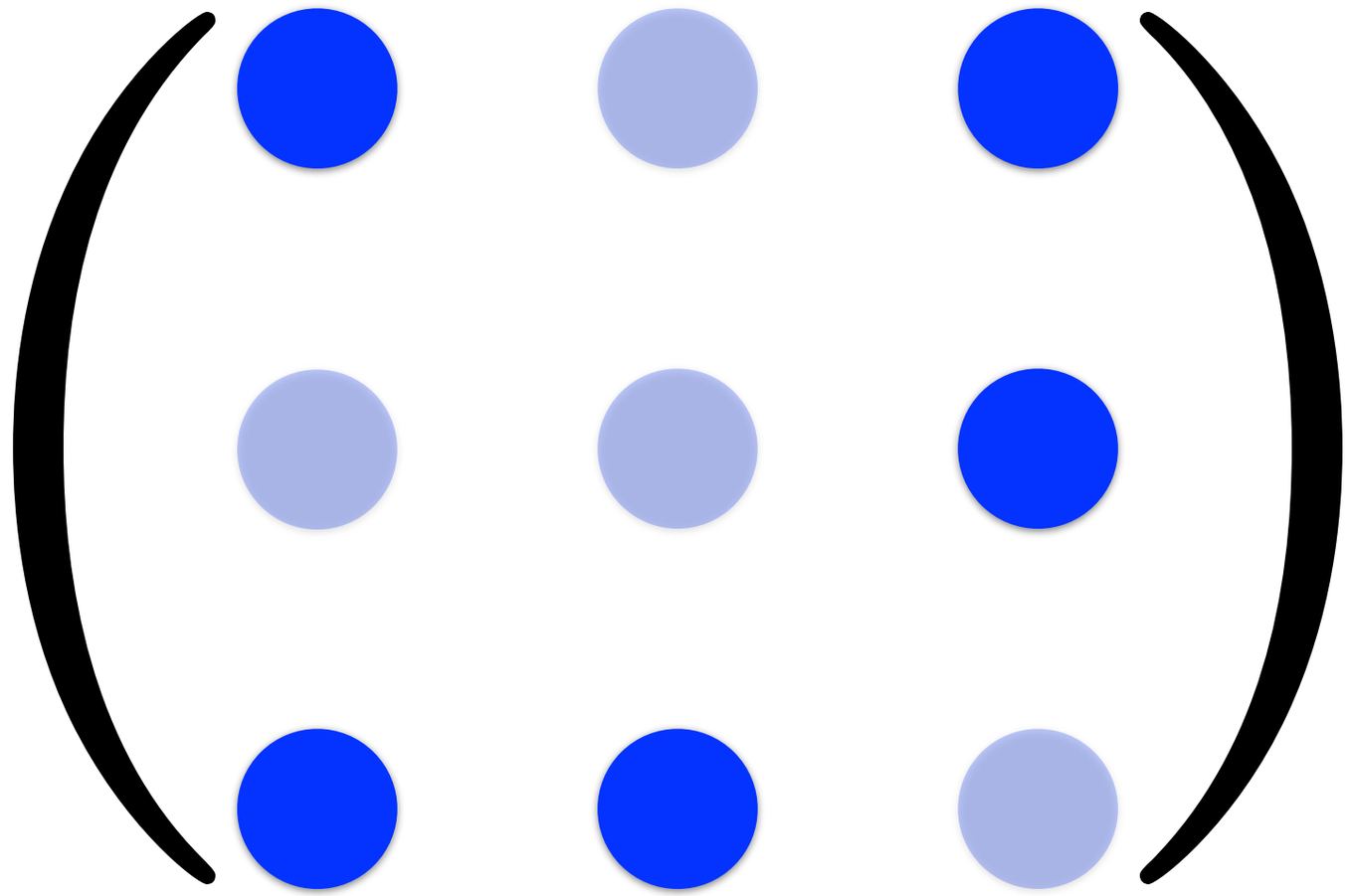
$$\mathcal{Y} = \{1, \ldots, N\}$$

$\mathcal{P}$ 

POINT PROCESSES

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

$$\mathcal{P}\left(\right) = 0.2$$



DETERMINANTAL

DETERMINANTAL

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

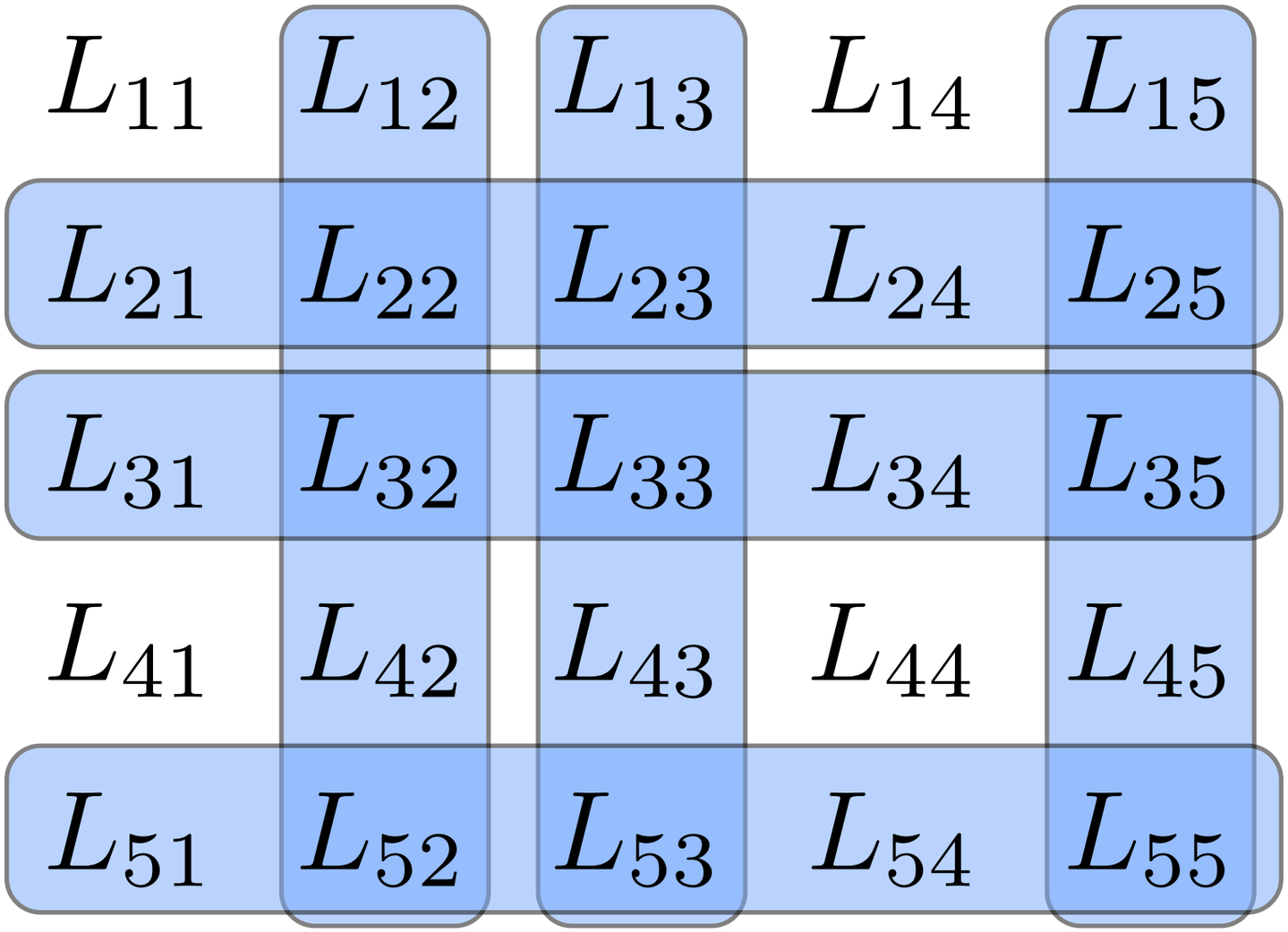
DETERMINANTAL

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

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

DETERMINANTAL

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



DETERMINANTAL

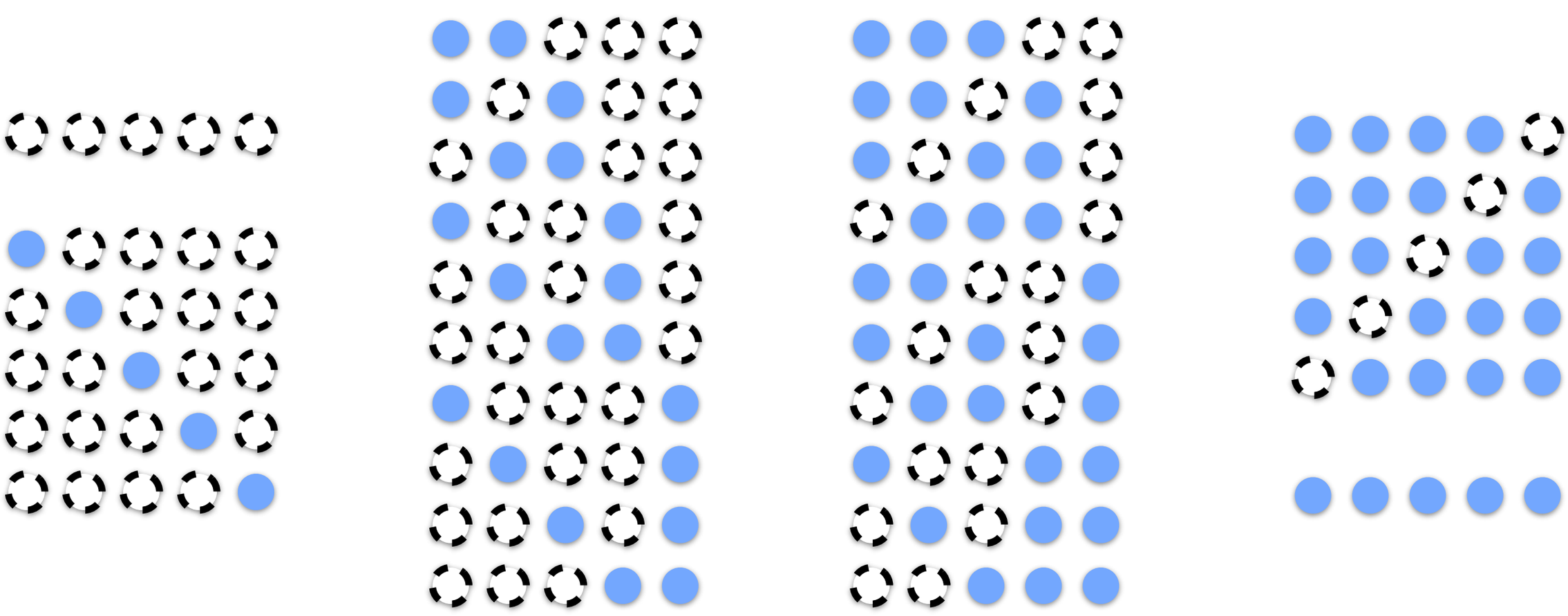
$$\mathcal{P}(\{2,3,5\}) \propto \begin{matrix} L_{22} & L_{23} & L_{25} \\ L_{32} & L_{33} & L_{35} \\ L_{52} & L_{53} & L_{55} \end{matrix}$$

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

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

$$\det(L+I)$$

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

$O(N^3)$

LARGE-SCALE SETTINGS

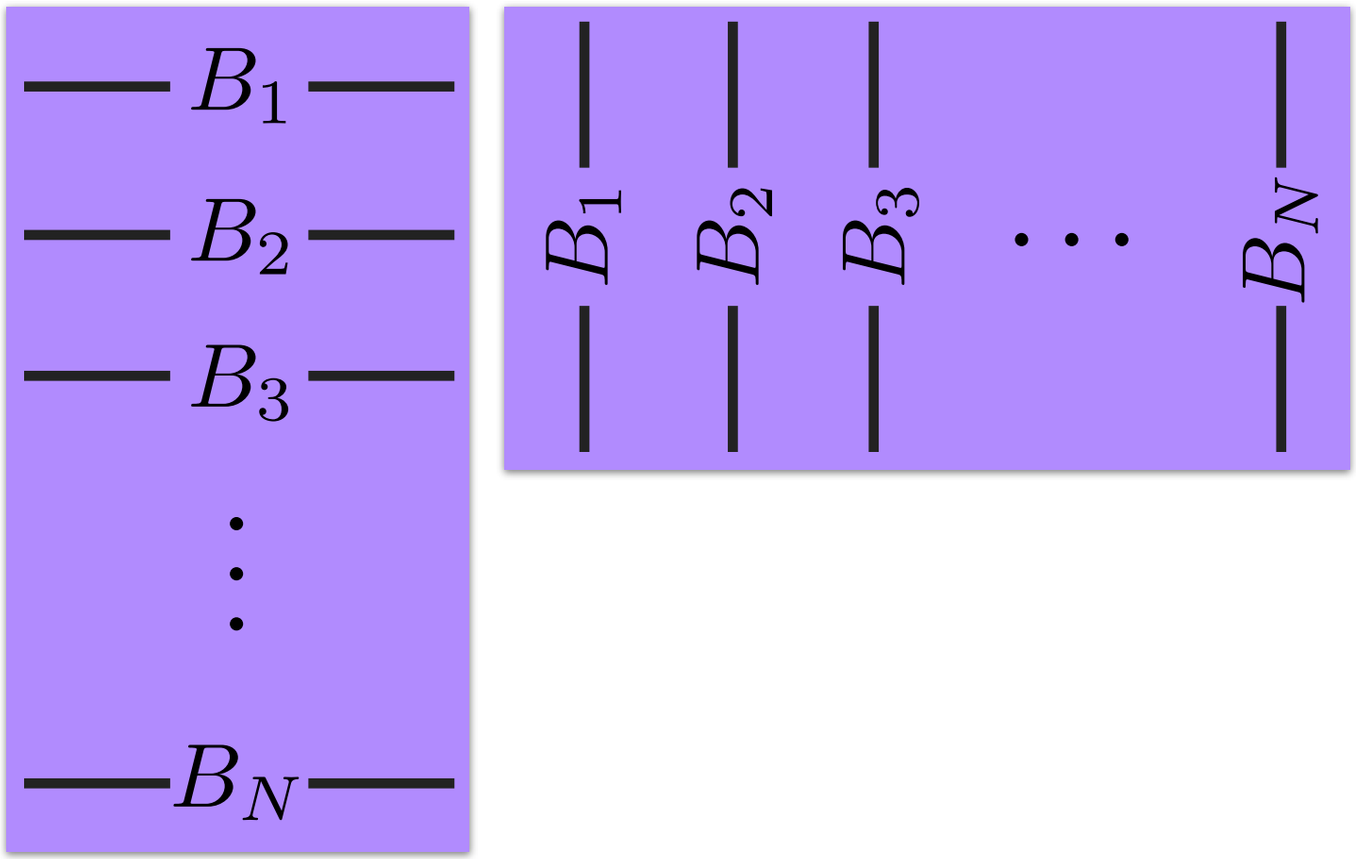
DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

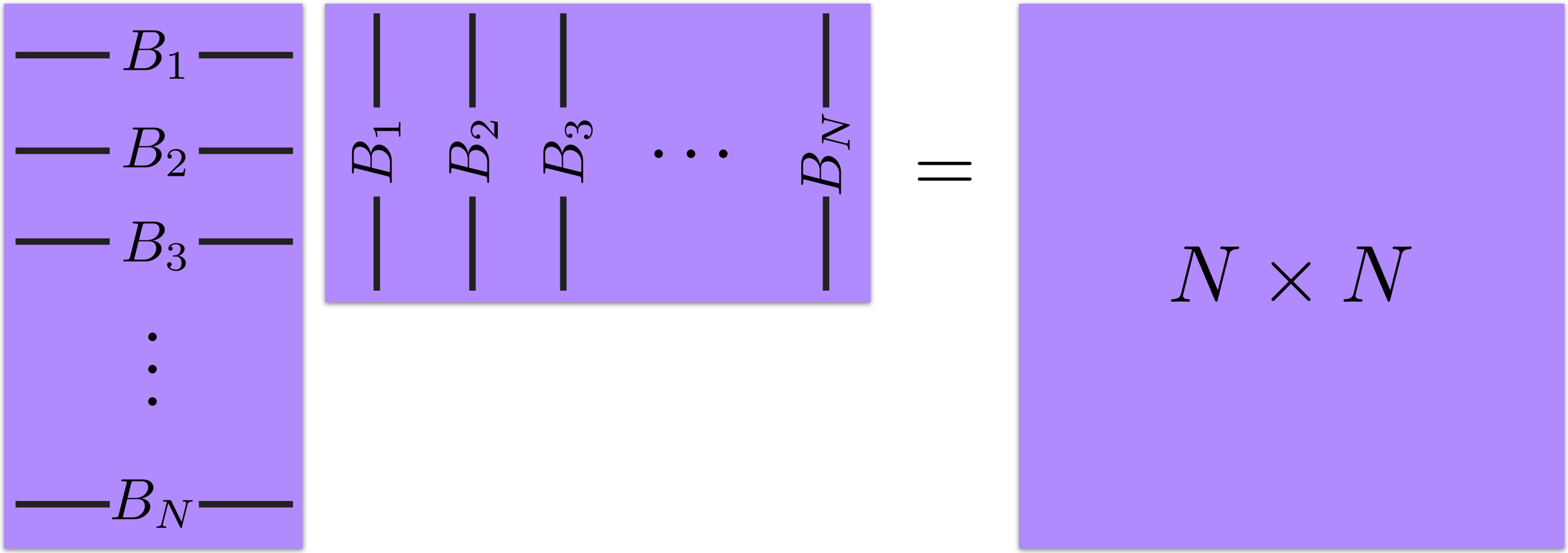
$$L$$



DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

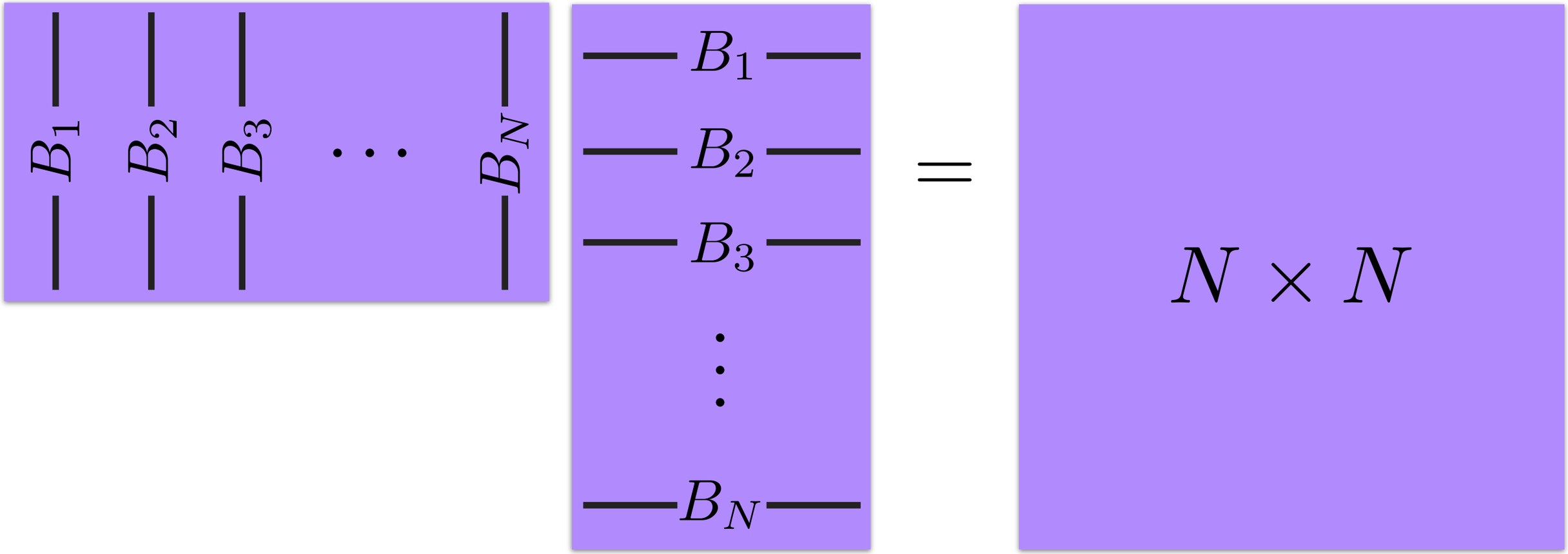
$$L$$



DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

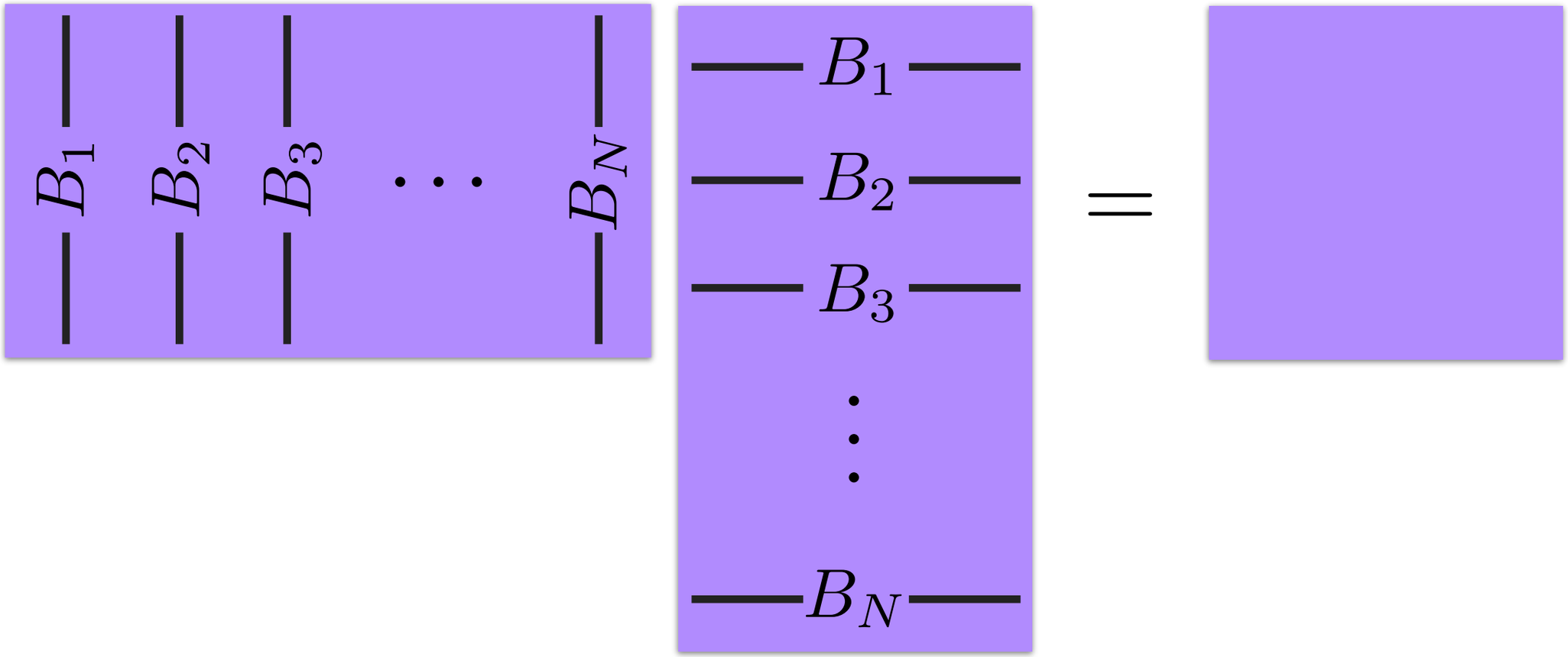
$$C$$



DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

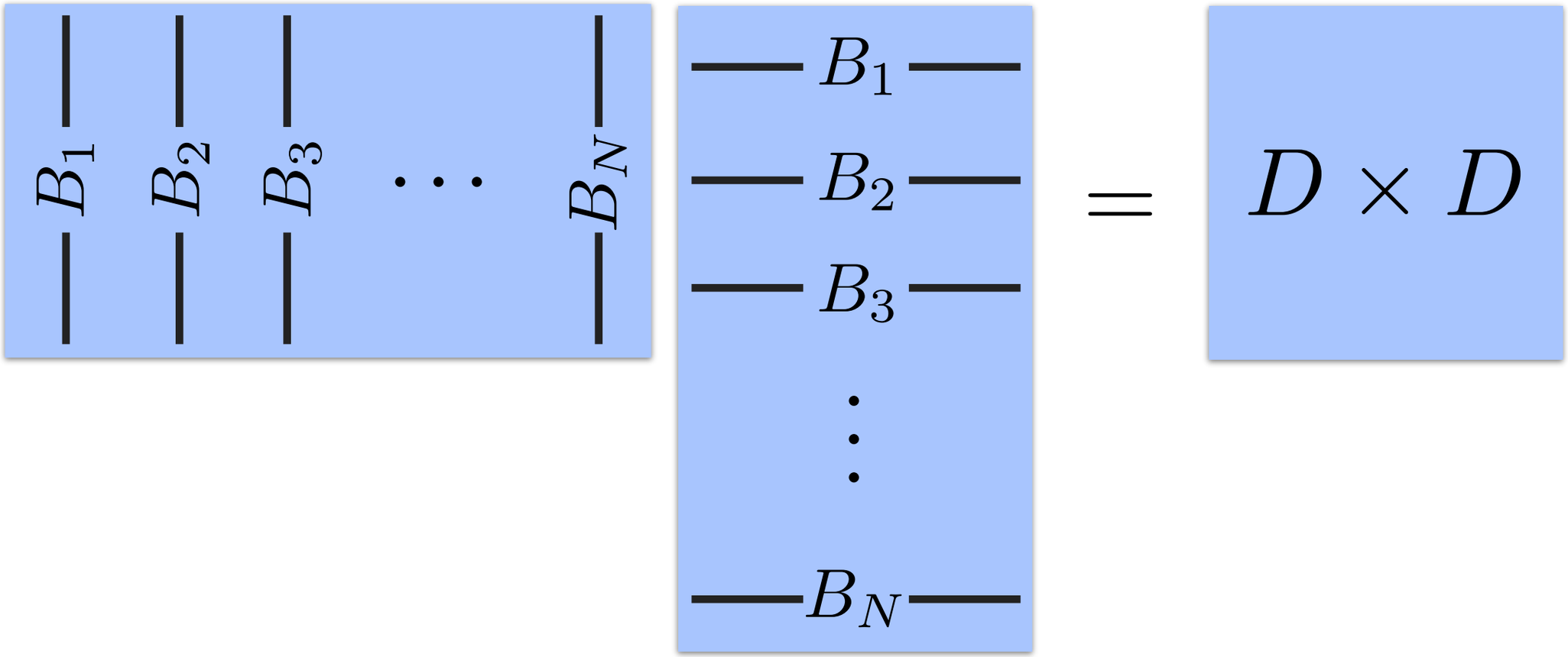
$$C$$



DUAL KERNEL

KULESZA AND TASKAR (NIPS 2010)

$$C$$



DUAL INFERENCE

DUAL INFERENCE

$$L = V \Lambda V^\top$$

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

DUAL INFERENCE

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

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

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	$\sum_Y \det(L_Y)$	$O(D^3)$
Marginalizing & Conditioning		$O(D^3 + D^2 k^2)$

DUAL INFERENCE

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

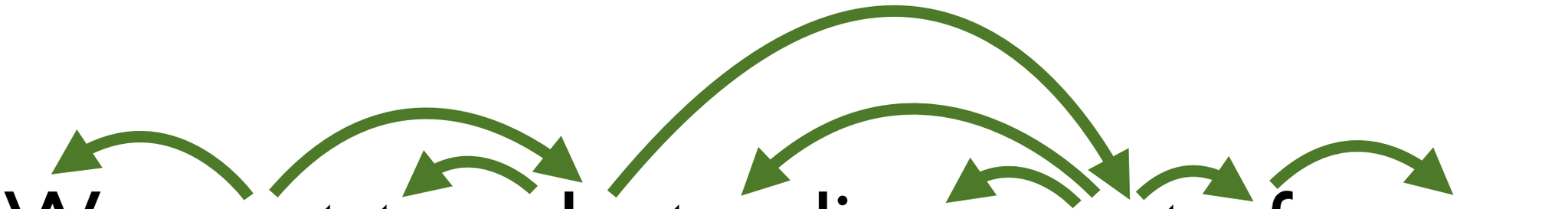
Normalizing $\sum_Y \det(L_Y)$	$O(D^3)$
Marginalizing & Conditioning	$O(D^3 + D^2 k^2)$
Sampling $\mathbf{Y} \sim \mathcal{P}_L$	$O(N D^2 k)$

EXPONENTIAL N

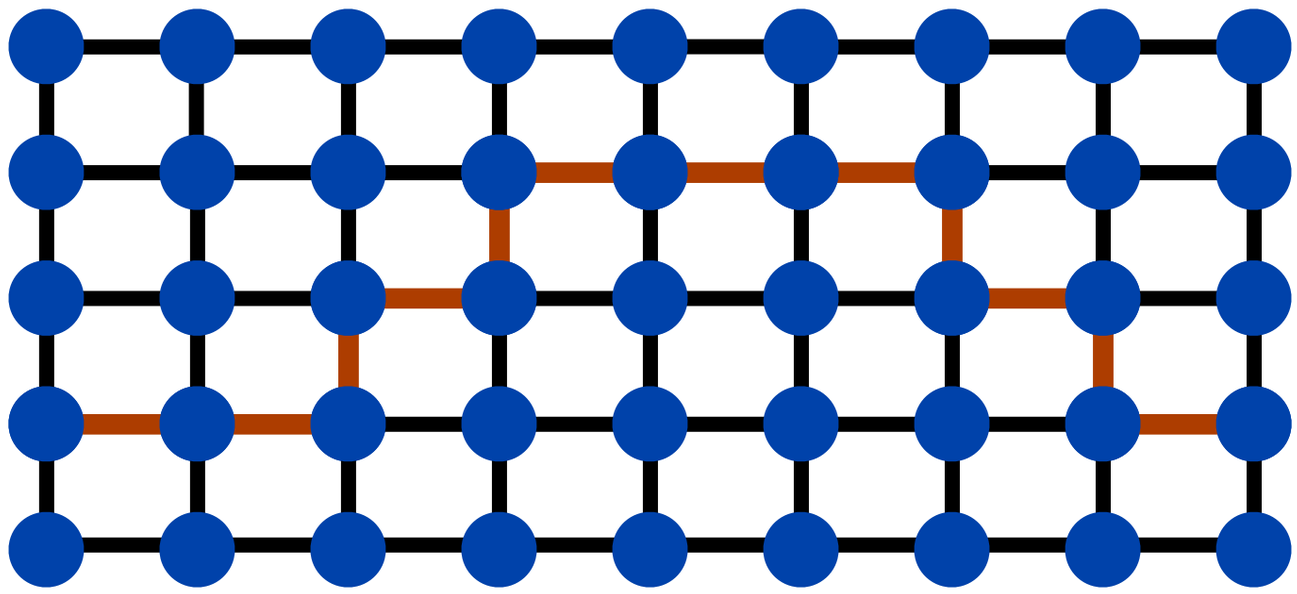
EXPONENTIAL N

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

EXPONENTIAL N



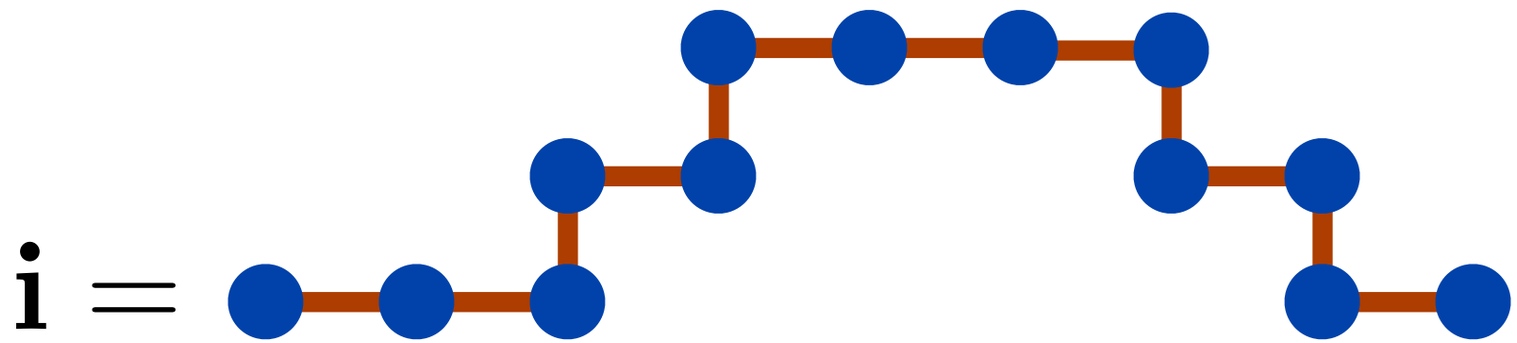
$$N = O(\{\text{sentence length}\}^{\{\text{sentence length}\}})$$



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

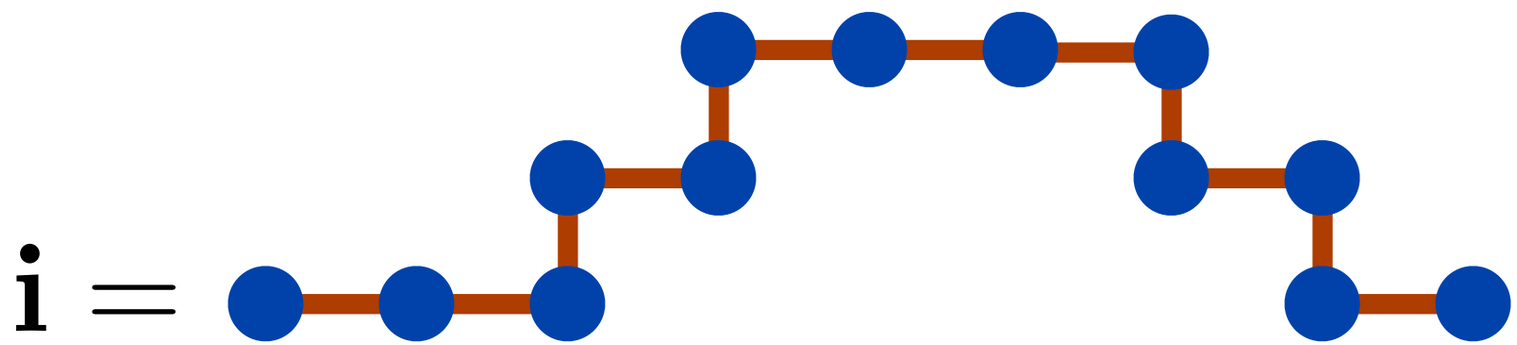
STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)



STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)



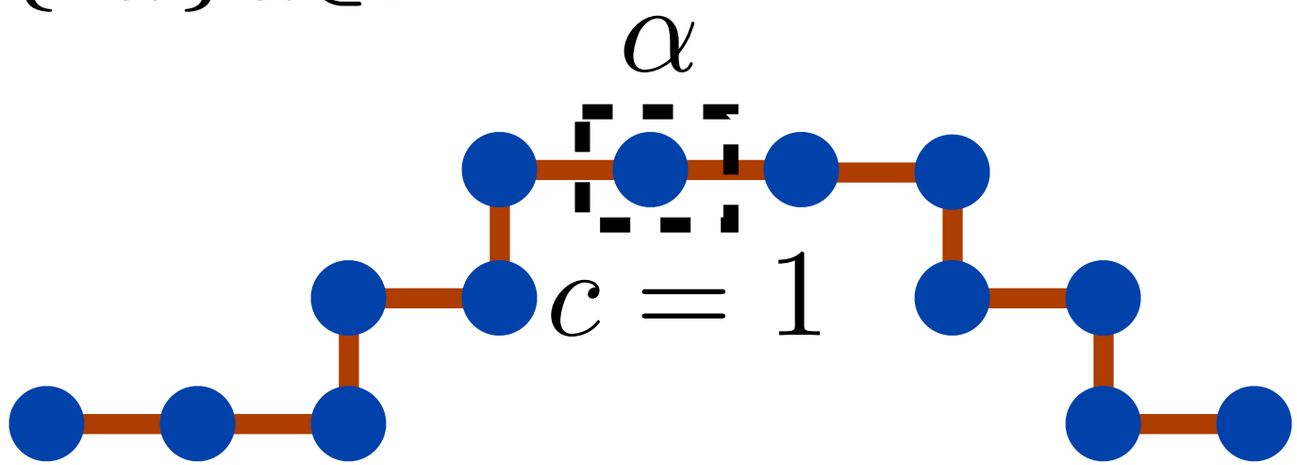
$$B_{\mathbf{i}} = q(\mathbf{i})\phi(\mathbf{i})$$

quality similarity

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

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



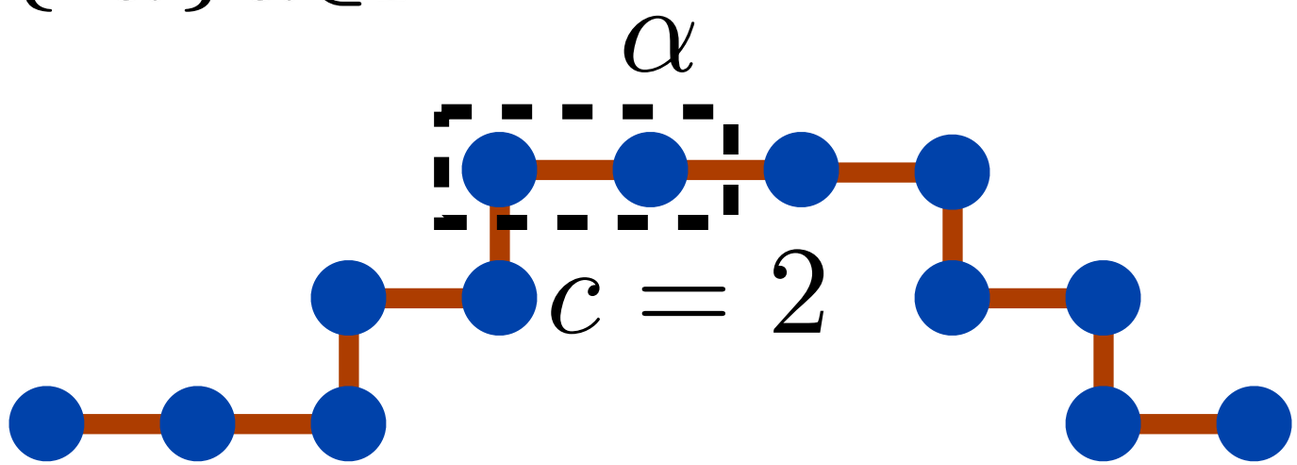
$$B_{\mathbf{i}} = q(\mathbf{i})\phi(\mathbf{i})$$

quality similarity

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

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



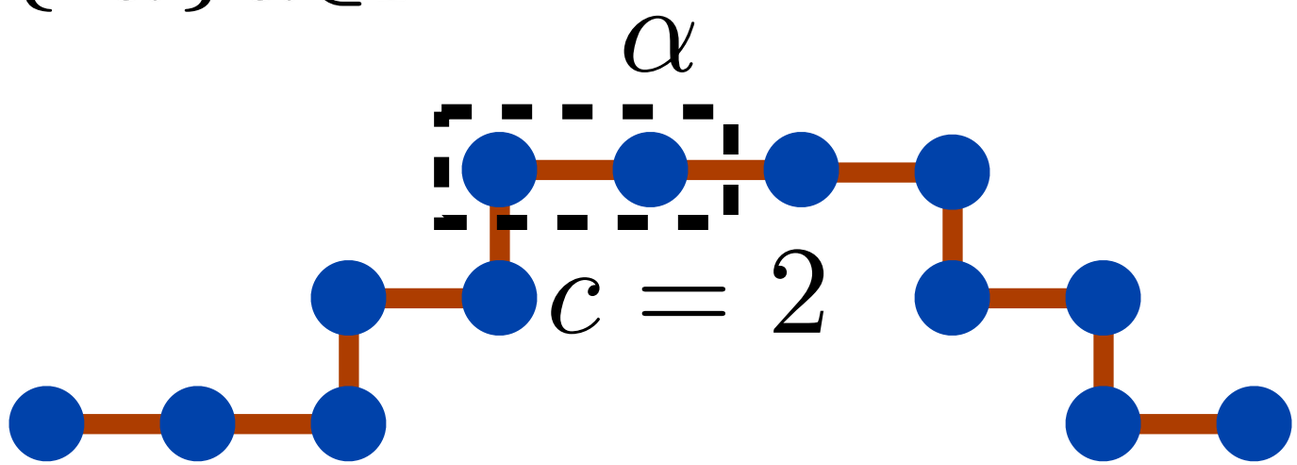
$$B_{\mathbf{i}} = q(\mathbf{i})\phi(\mathbf{i})$$

quality similarity

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

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



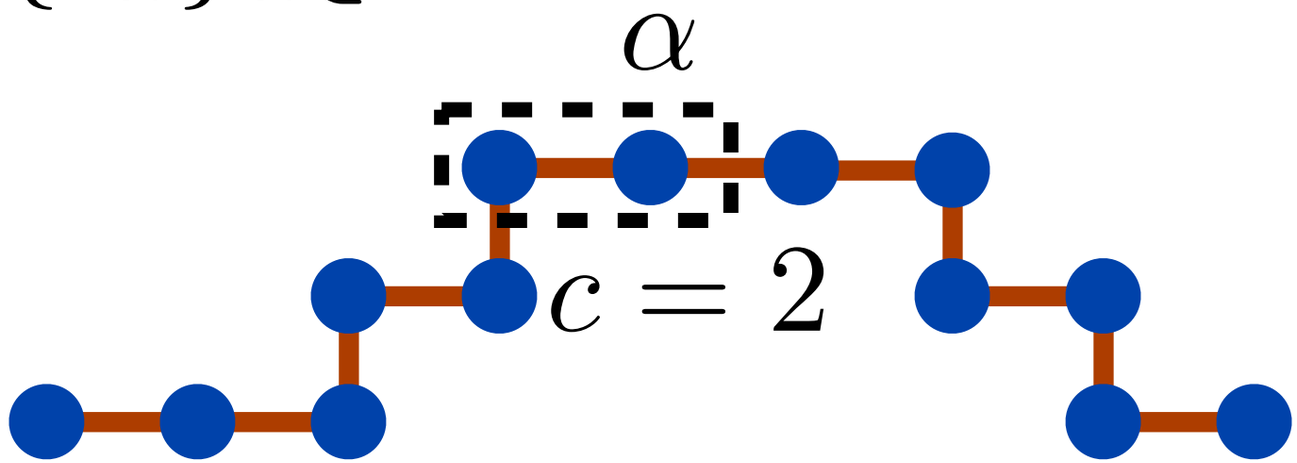
$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \phi(\mathbf{i})$$

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

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

$$\mathbf{i} =$$

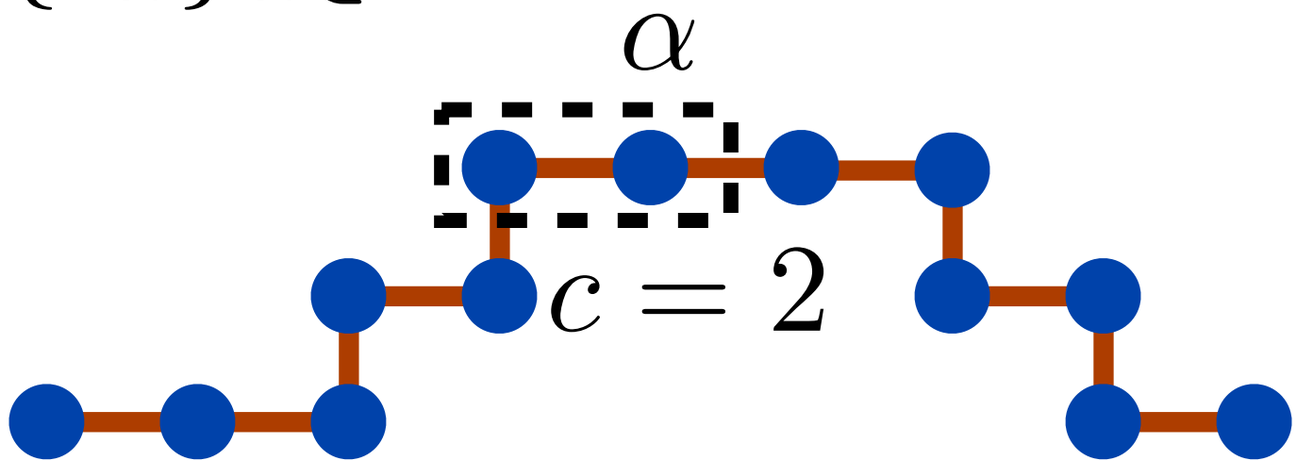


$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

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

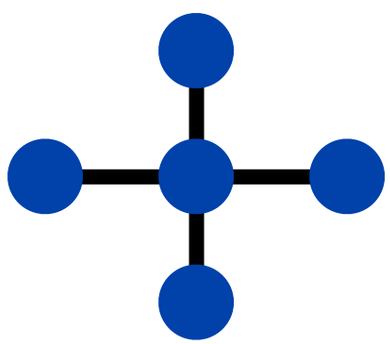
$\mathbf{i} =$ 

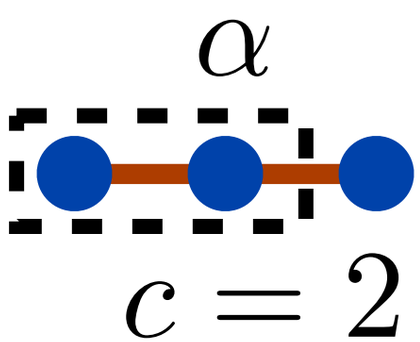
$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

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

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$M =$ 



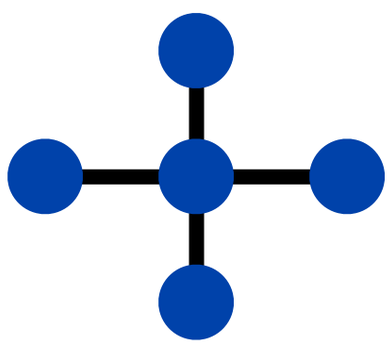
$R =$ 

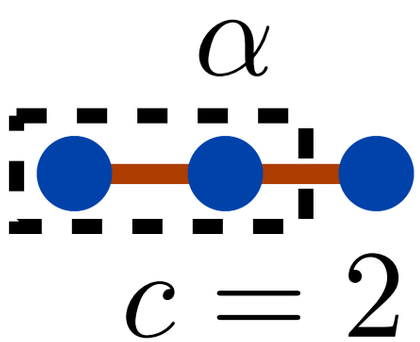
$$B_{\mathbf{i}} = \left[\prod_{\alpha \in F} q(i_\alpha) \right] \left[\sum_{\alpha \in F} \phi(i_\alpha) \right]$$

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

STRUCTURE FACTORIZATION

KULESZA AND TASKAR (NIPS 2010)

$M =$ 



$R =$ 

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

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

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

LARGE FEATURE SETS?

LARGE FEATURE SETS?

$N = \#$ of items

$D = \#$ of features

	Large	**Exponential**
Small	dual	dual + structure

LARGE FEATURE SETS?

$N = \#$ of items

$D = \#$ of features

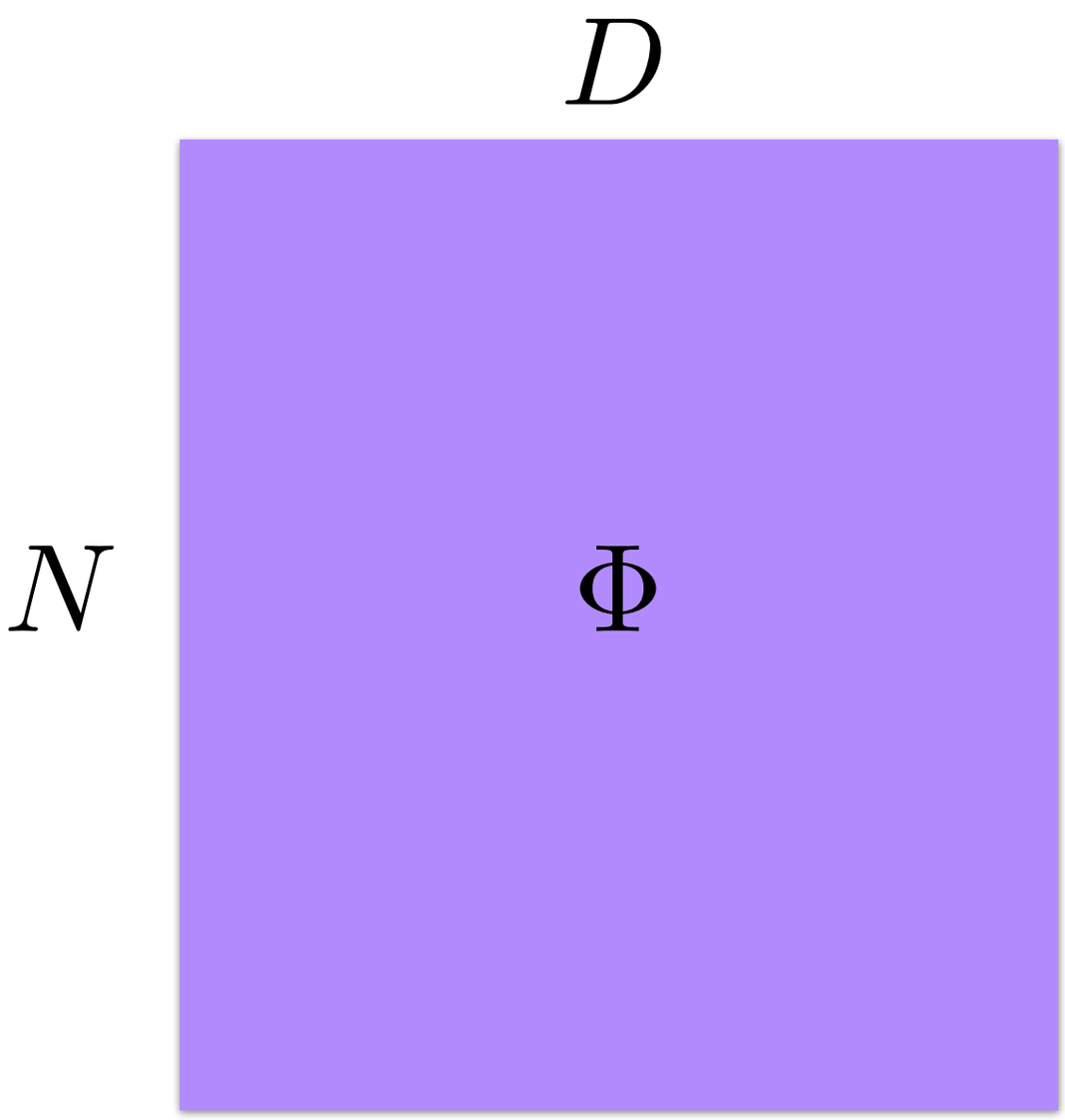
	Large	**Exponential**
Small	dual	dual + structure
Large	?	?

RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)



RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

D

$M^c R$

Φ

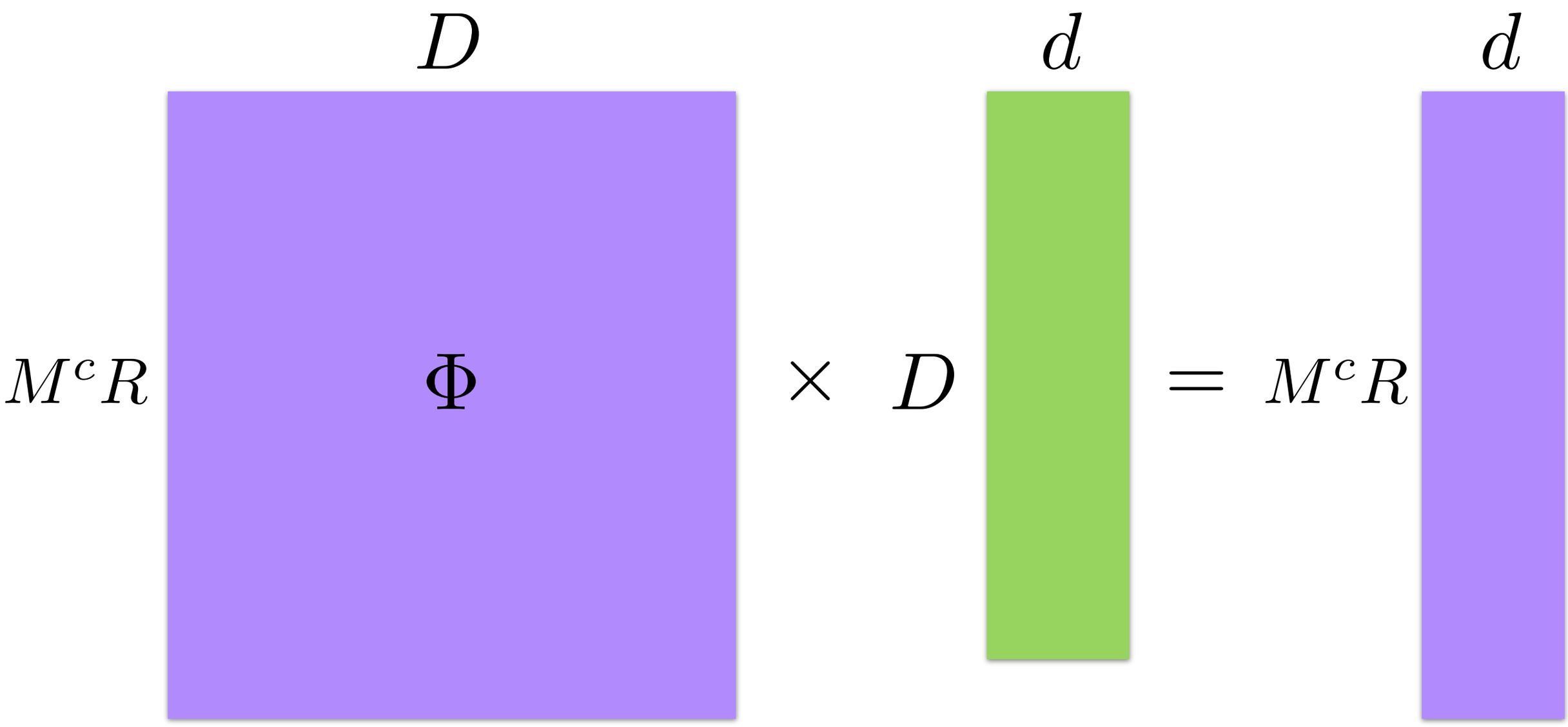
RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

$$M^c R \quad \underset{D}{\Phi} \quad \times \quad \underset{d}{D}$$

RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

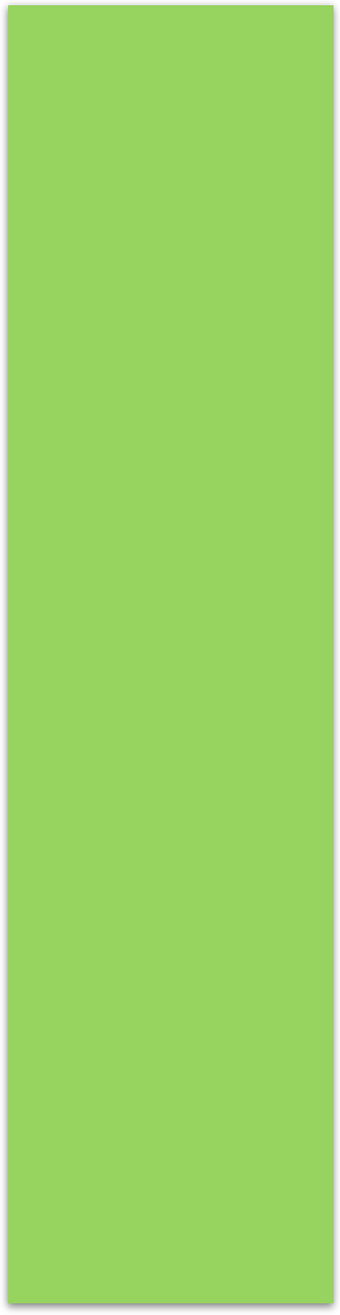


RANDOM PROJECTIONS

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

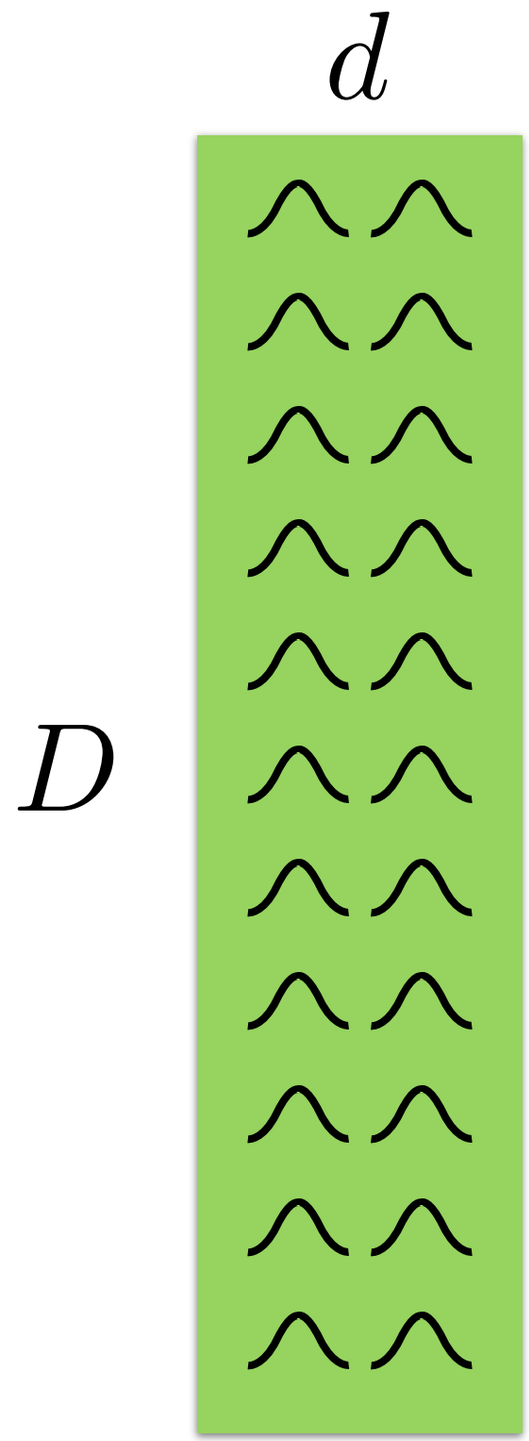
d

D



RANDOM PROJECTIONS

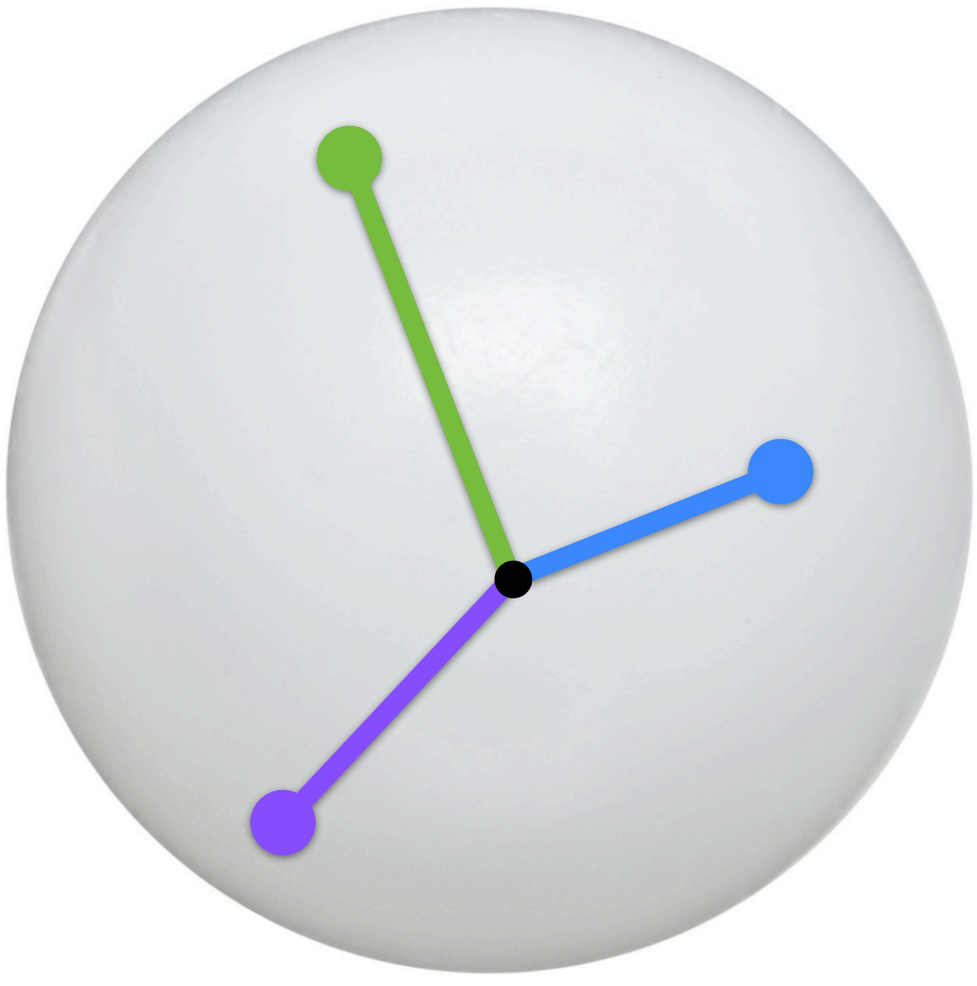
GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)



VOLUME PRESERVATION

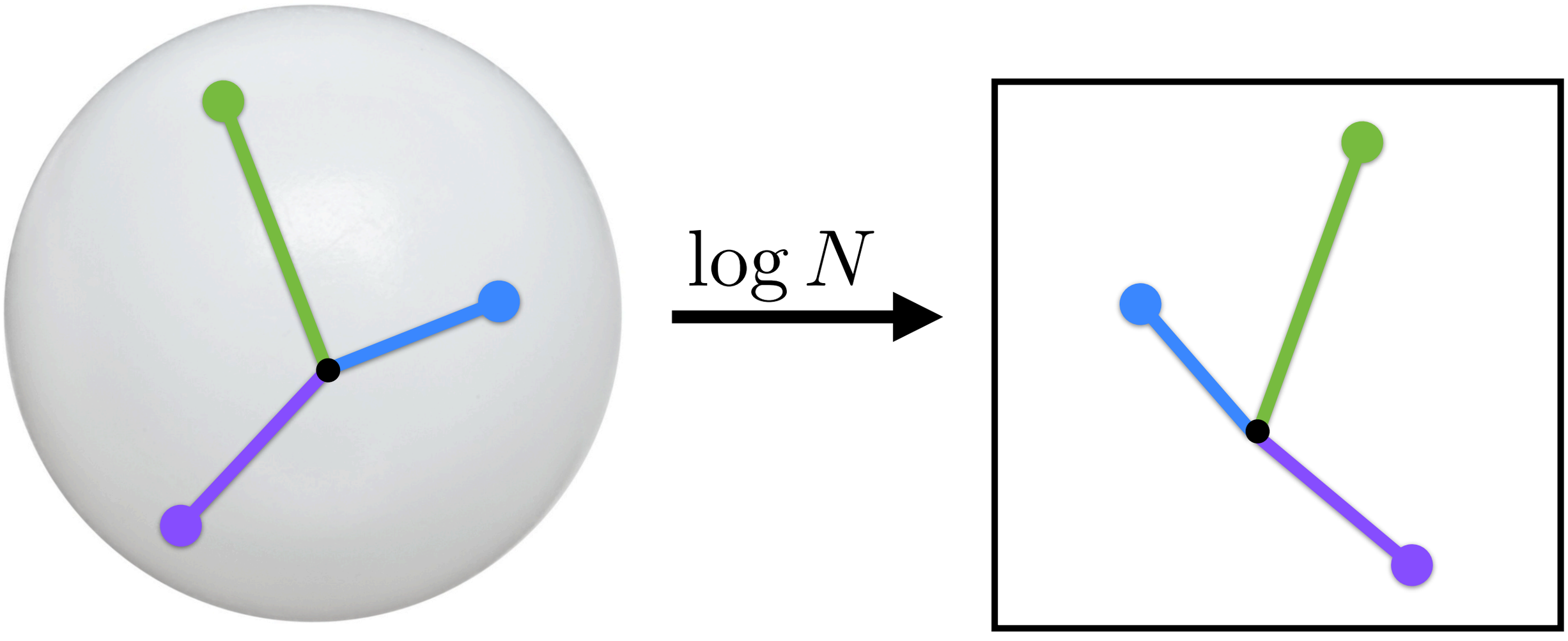
VOLUME PRESERVATION

JOHNSON AND LINDENSTRAUSS (1984)



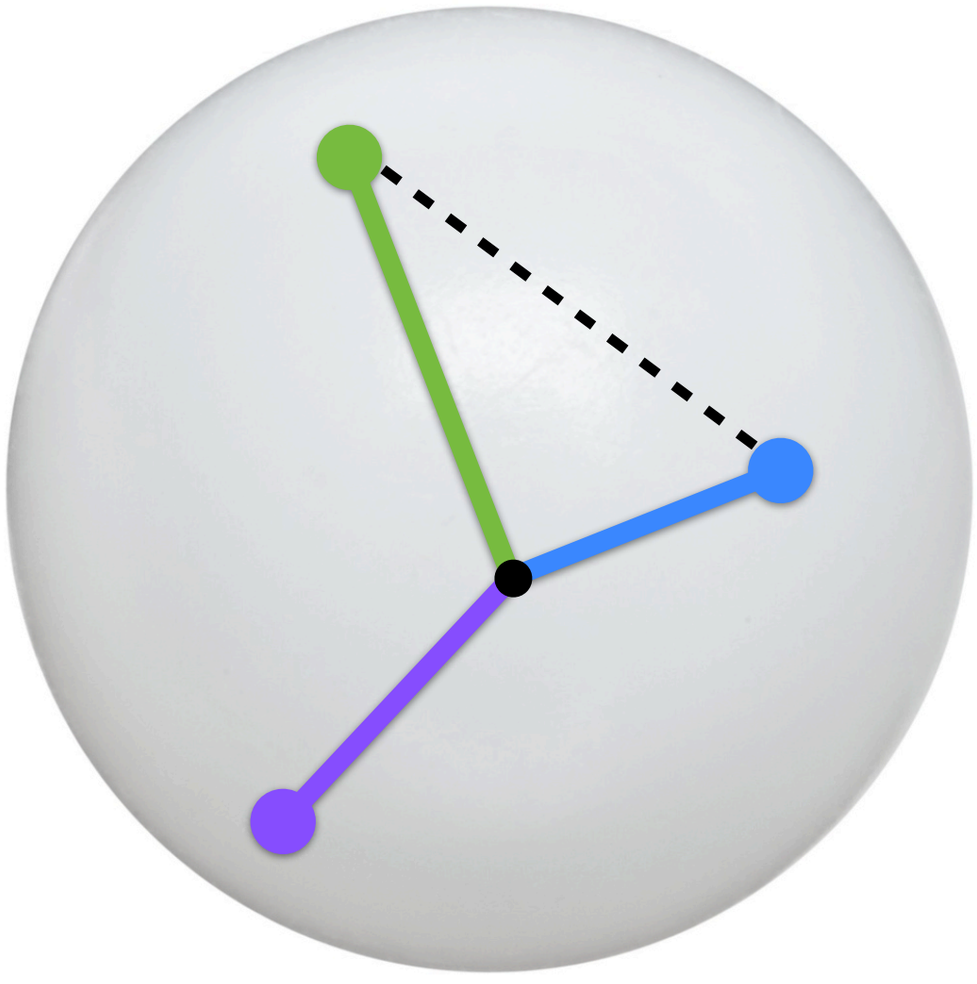
VOLUME PRESERVATION

JOHNSON AND LINDENSTRAUSS (1984)

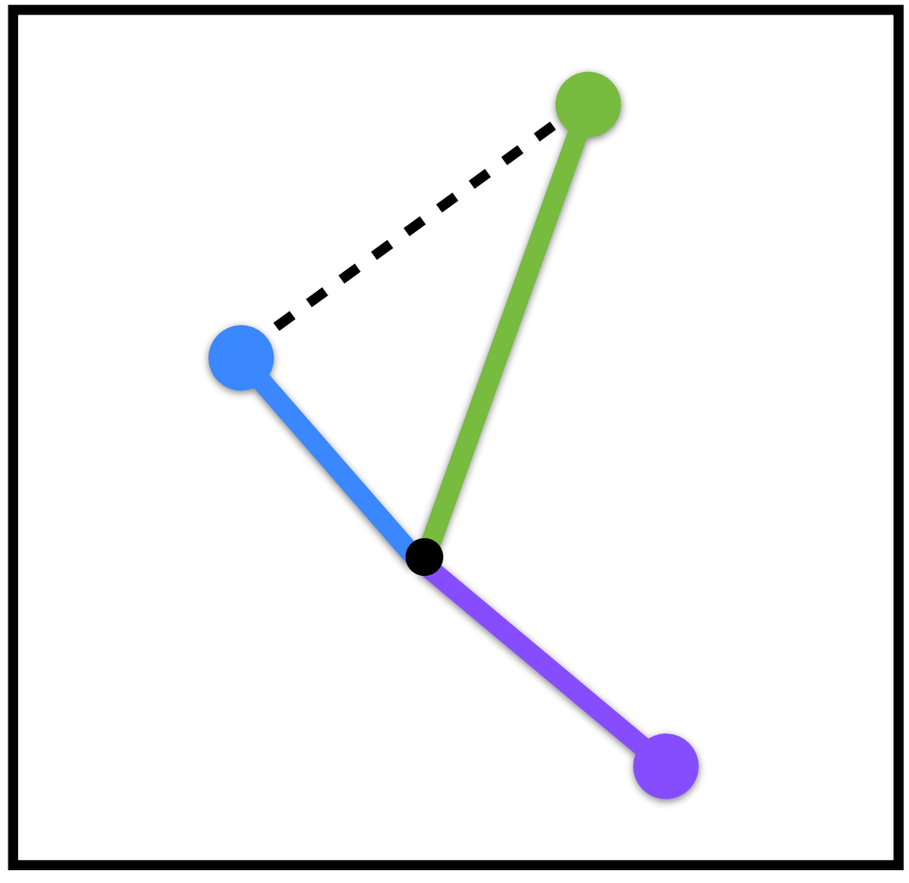


VOLUME PRESERVATION

JOHNSON AND LINDENSTRAUSS (1984)

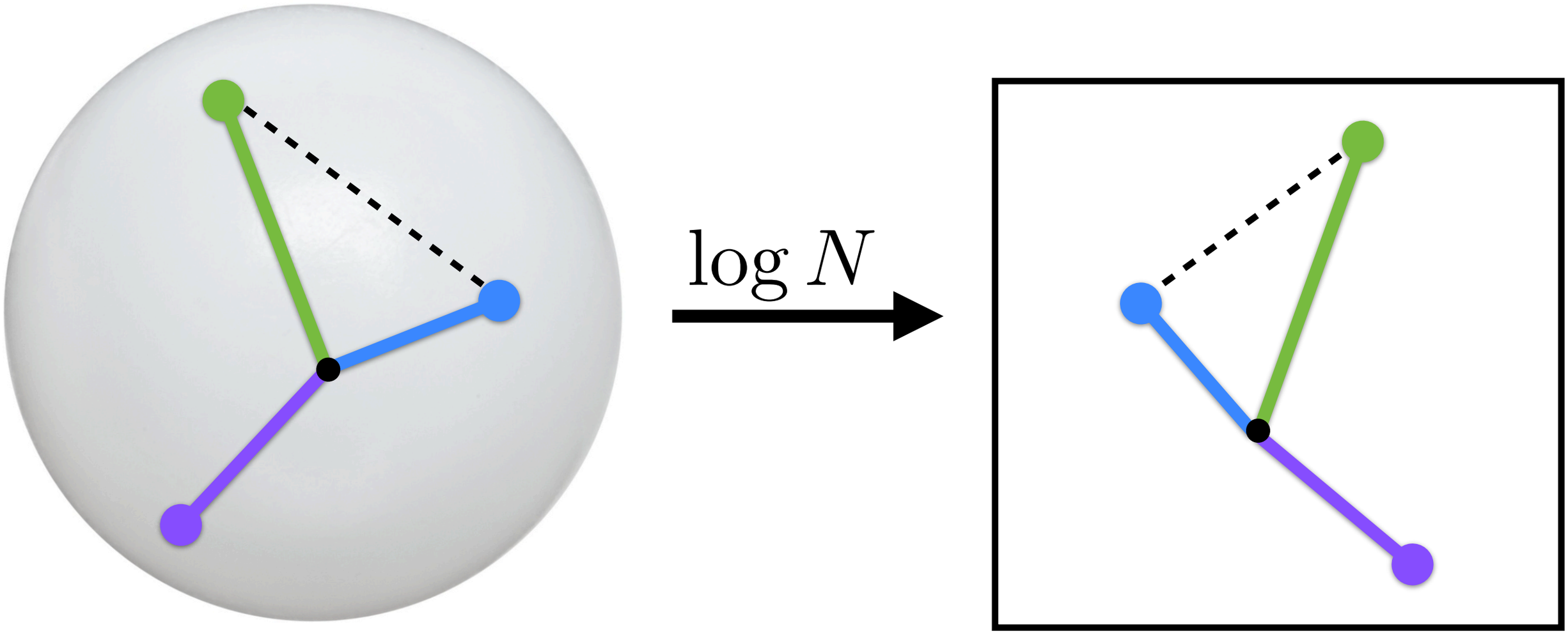


$\log N$



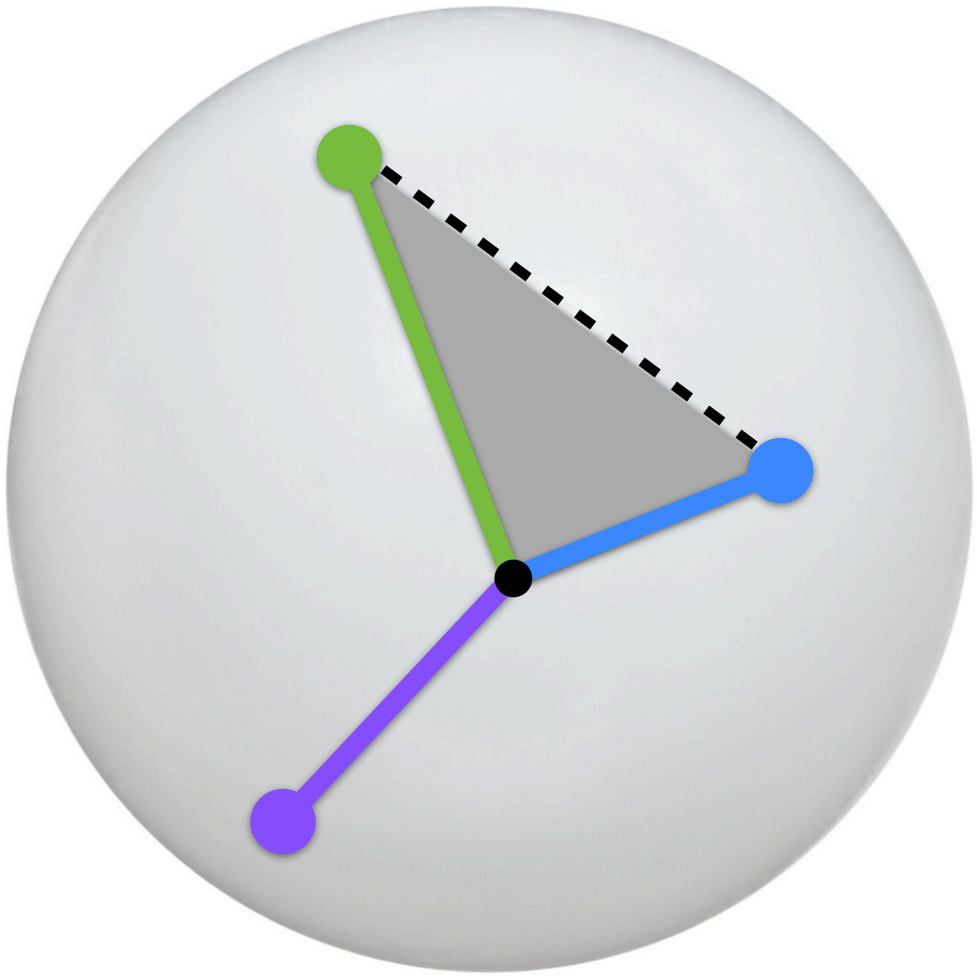
VOLUME PRESERVATION

MAGEN AND ZOUZIAS (2008)

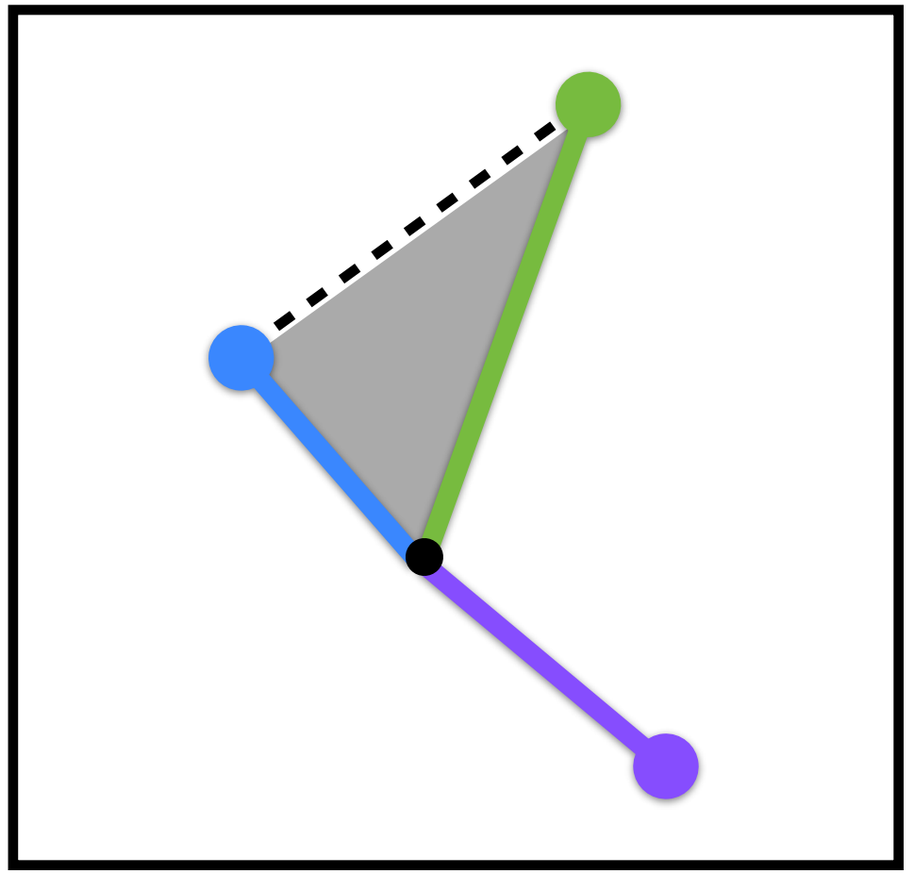


VOLUME PRESERVATION

MAGEN AND ZOUZIAS (2008)



$\log N$



DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

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

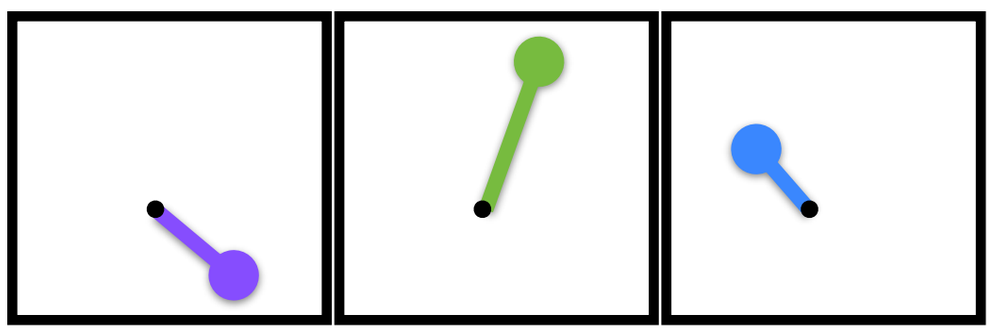
DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

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

$k = 1$

$\approx$

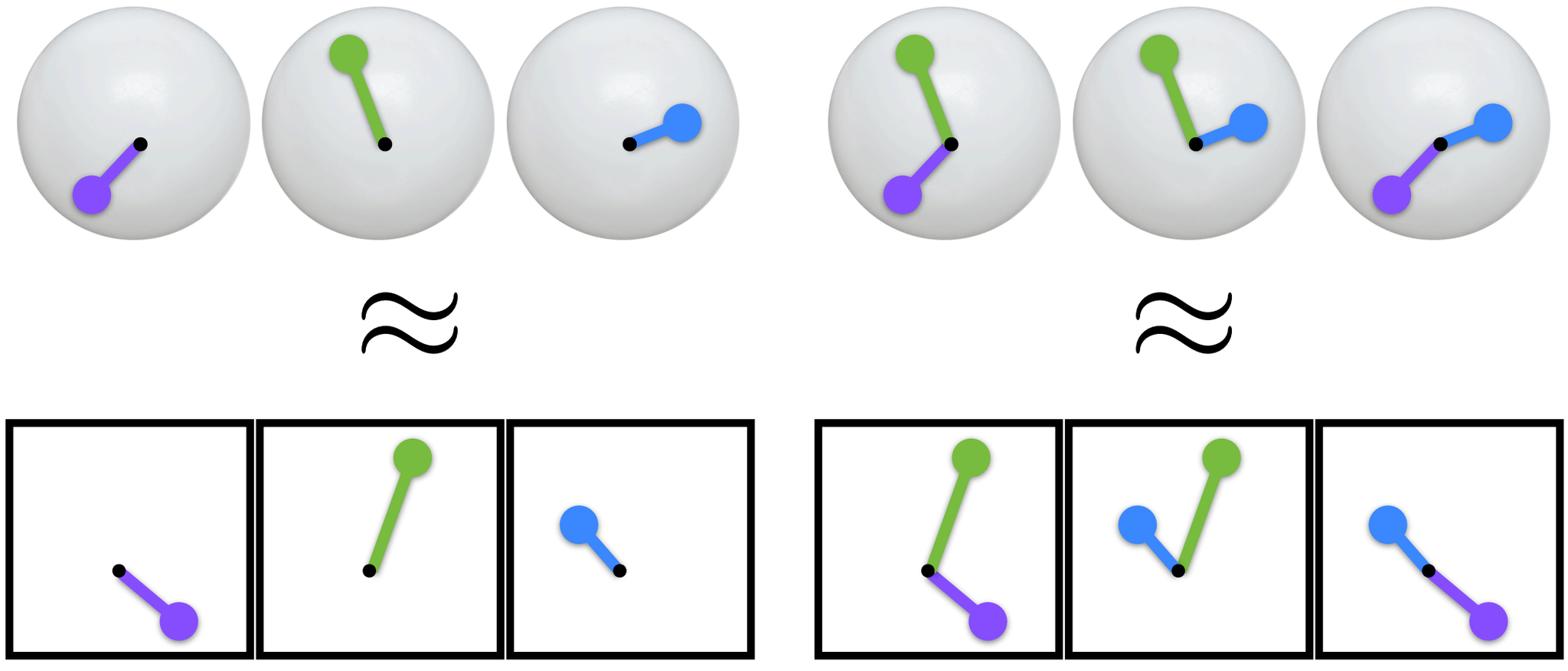


DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

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

$k = 1$ $\qquad$ $k = 2$

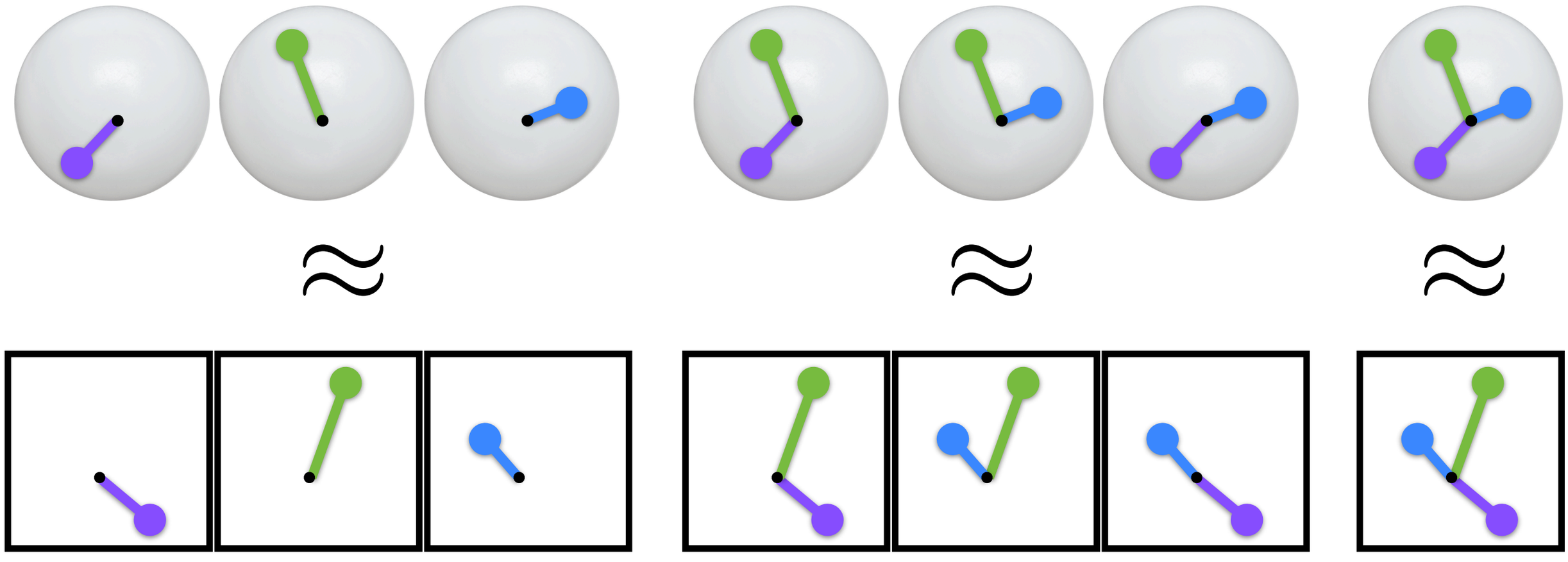


DPP PRESERVATION

GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

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

$k = 1$ $k = 2$ $k = 3$



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

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

DPP PRESERVATION

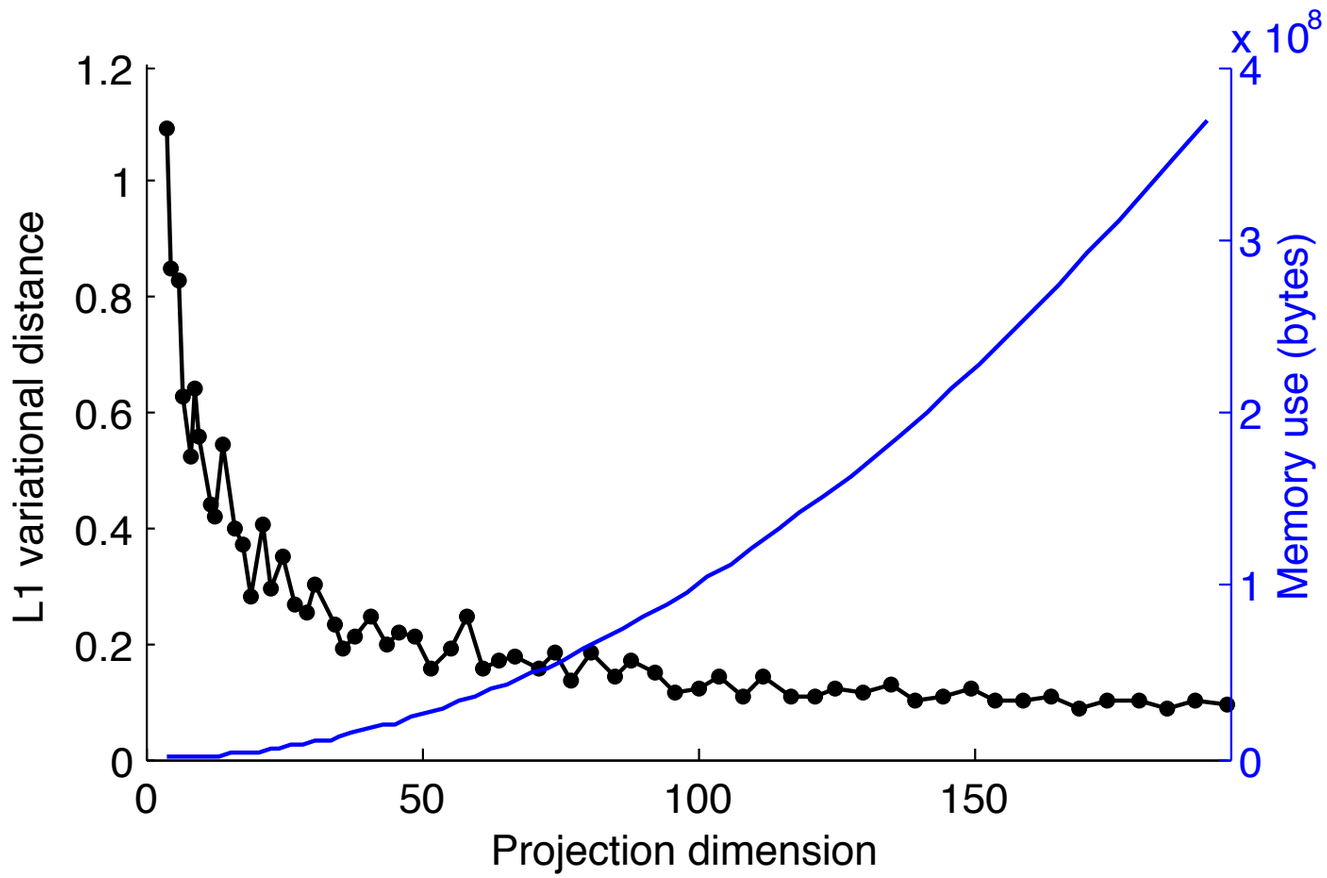
GILLENWATER, KULESZA, AND TASKAR (EMNLP 2012)

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

subset size

total # of items

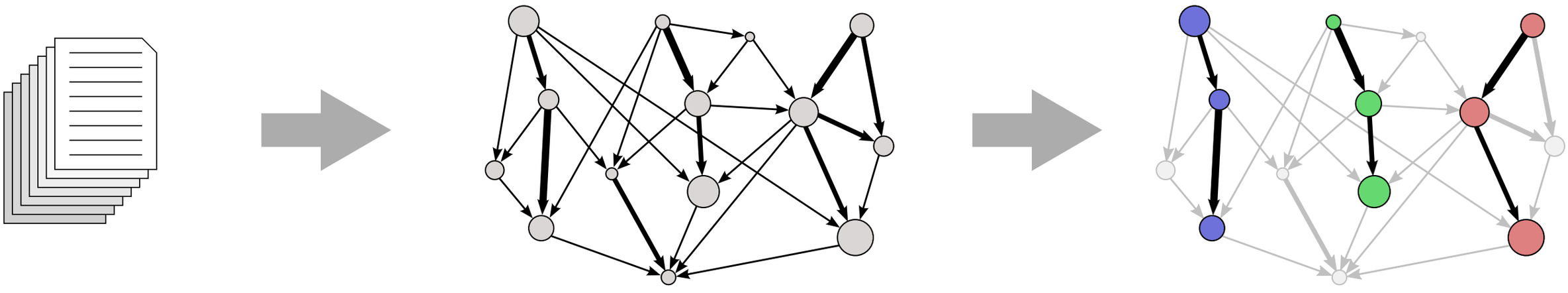
$$\text{w.p. } 1-\delta: \quad \|\mathcal{P}^k - \tilde{\mathcal{P}}^k\|_1 \leq e^{6k\epsilon} - 1$$



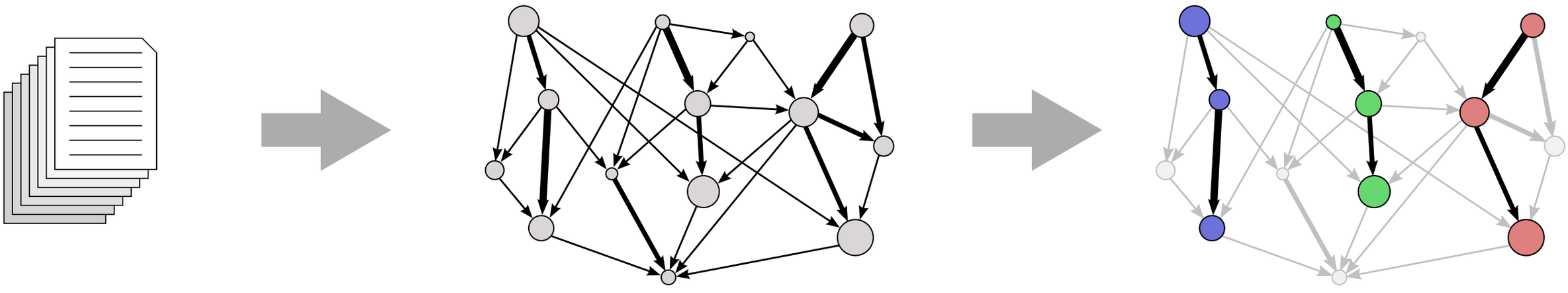
STRUCTURED SUMMARIZATION

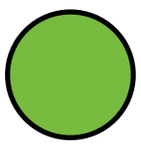
NEWS THREADING

NEWS THREADING



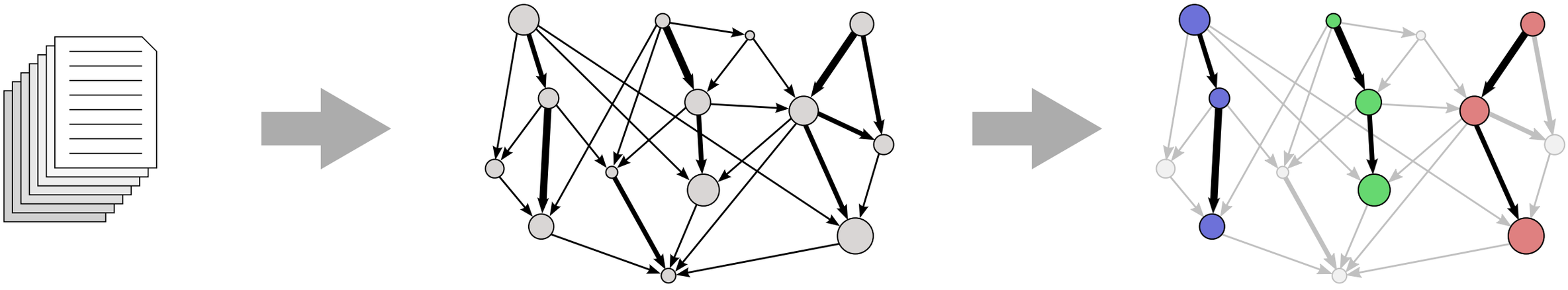
NEWS THREADING

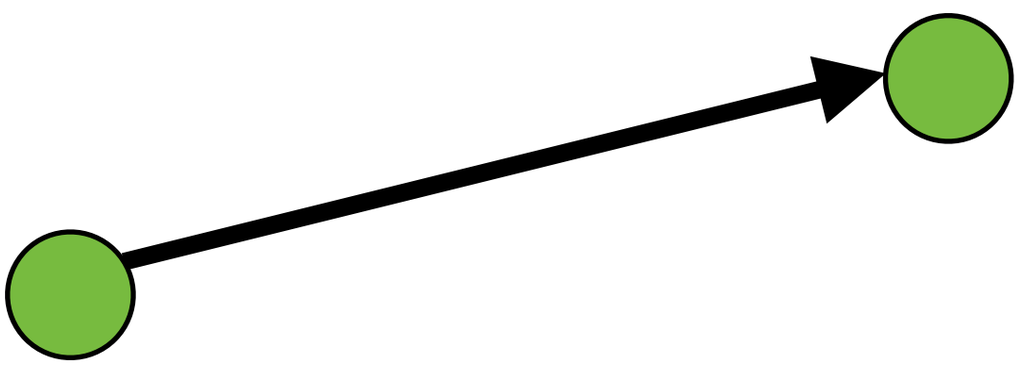




March 28: Health officials confirm Ebola outbreak in Guinea's capital

NEWS THREADING

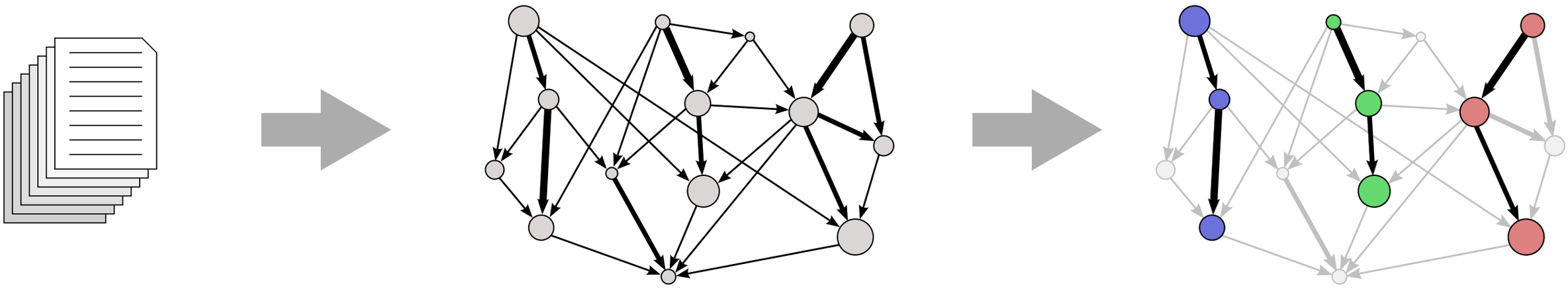


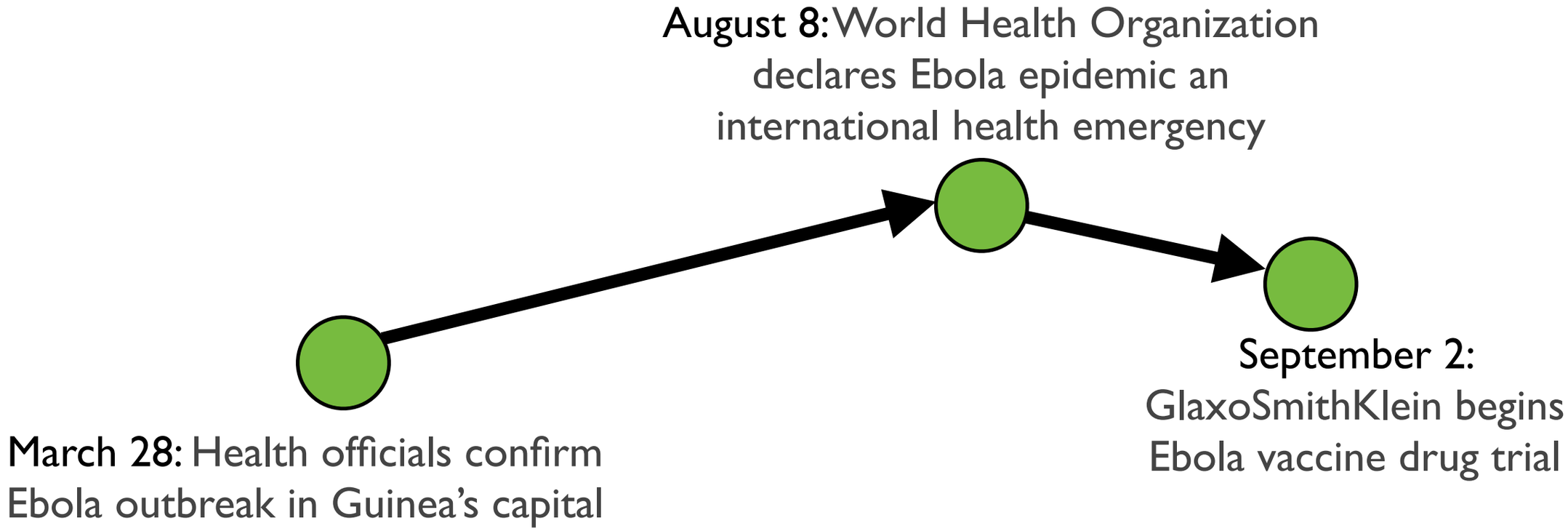


August 8: World Health Organization declares Ebola epidemic an international health emergency

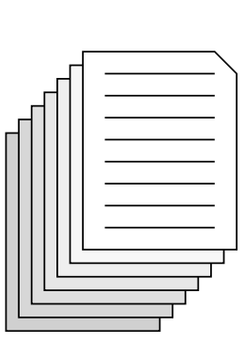
March 28: Health officials confirm Ebola outbreak in Guinea's capital

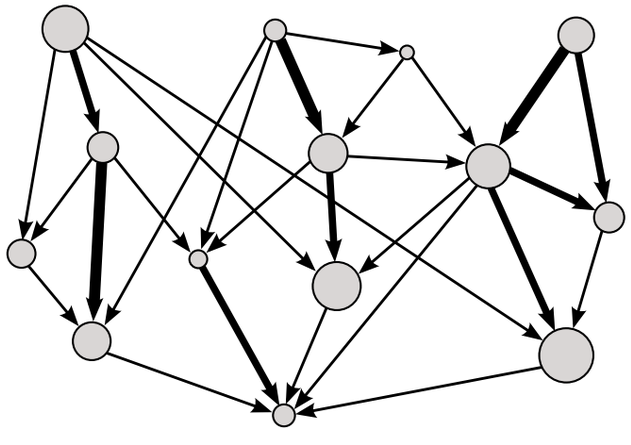
NEWS THREADING

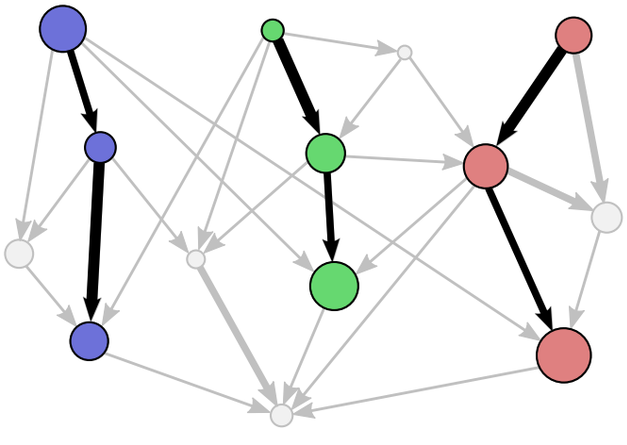




NEWS THREADING

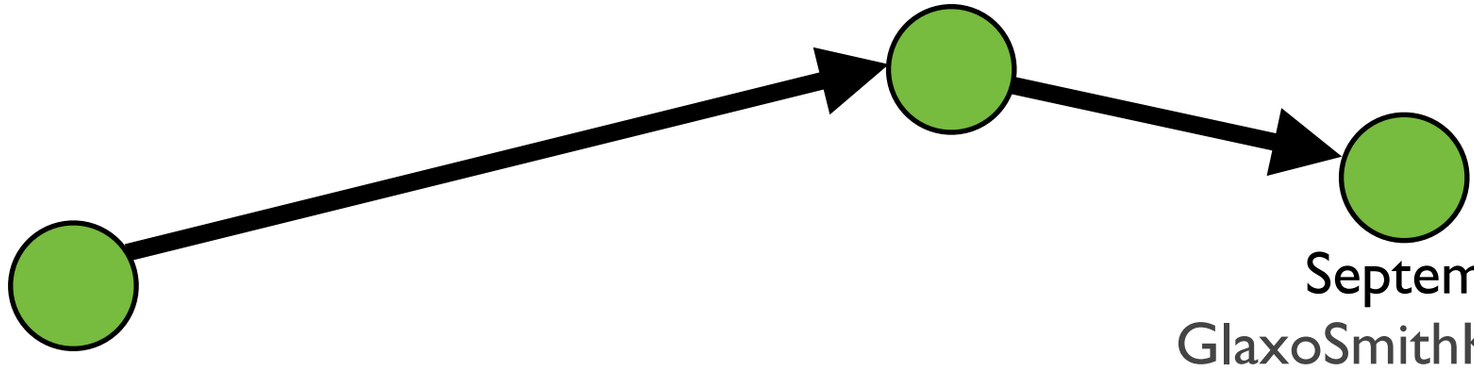






$M \approx 35{,}000$

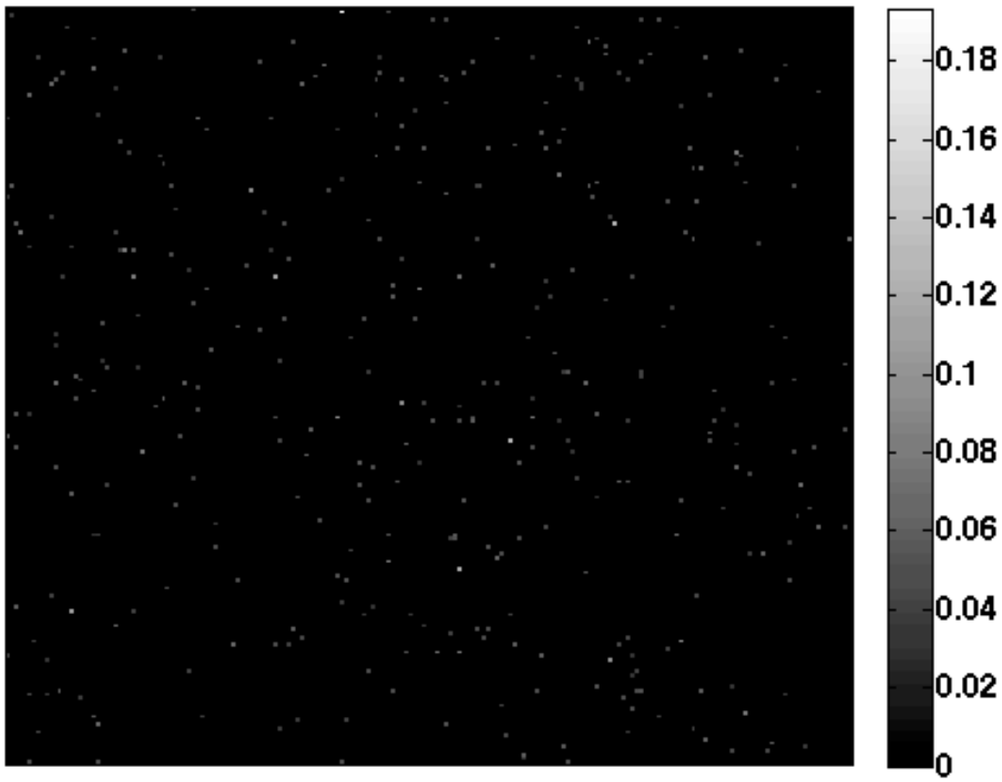
10^{360}



PROJECTING NEWS FEATURES

PROJECTING NEWS FEATURES

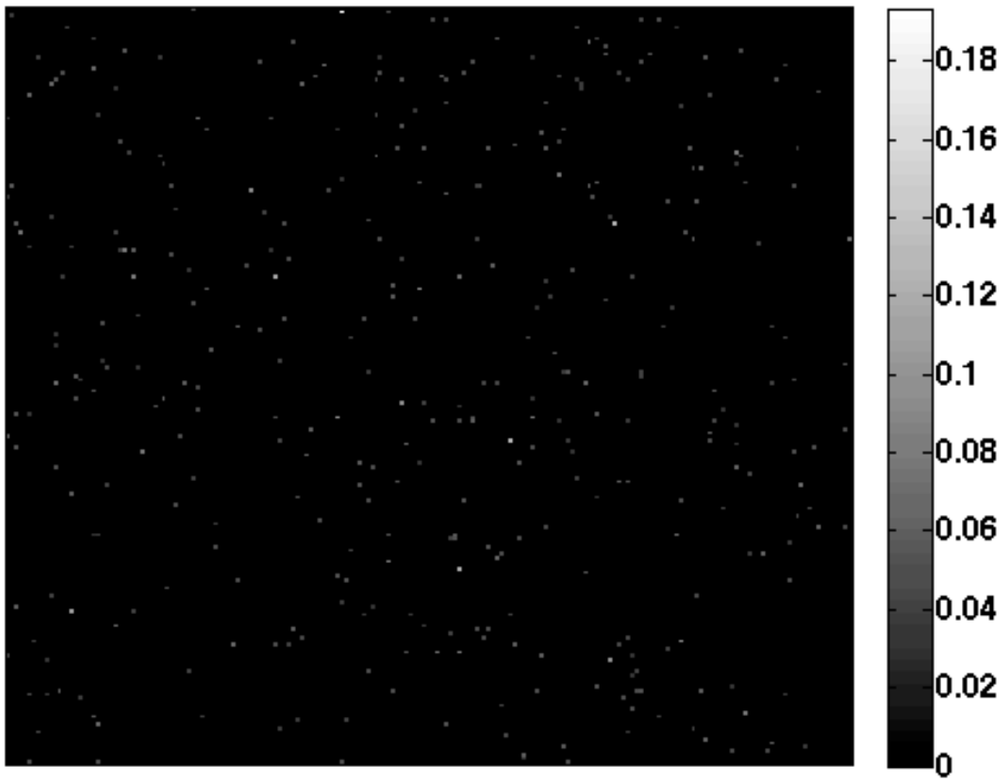
$$\phi(\mathbf{i})$$



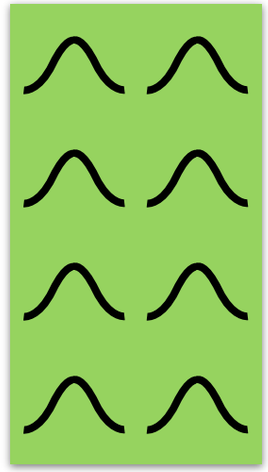
$$D = 36{,}356$$

PROJECTING NEWS FEATURES

$\phi(\mathbf{i})$

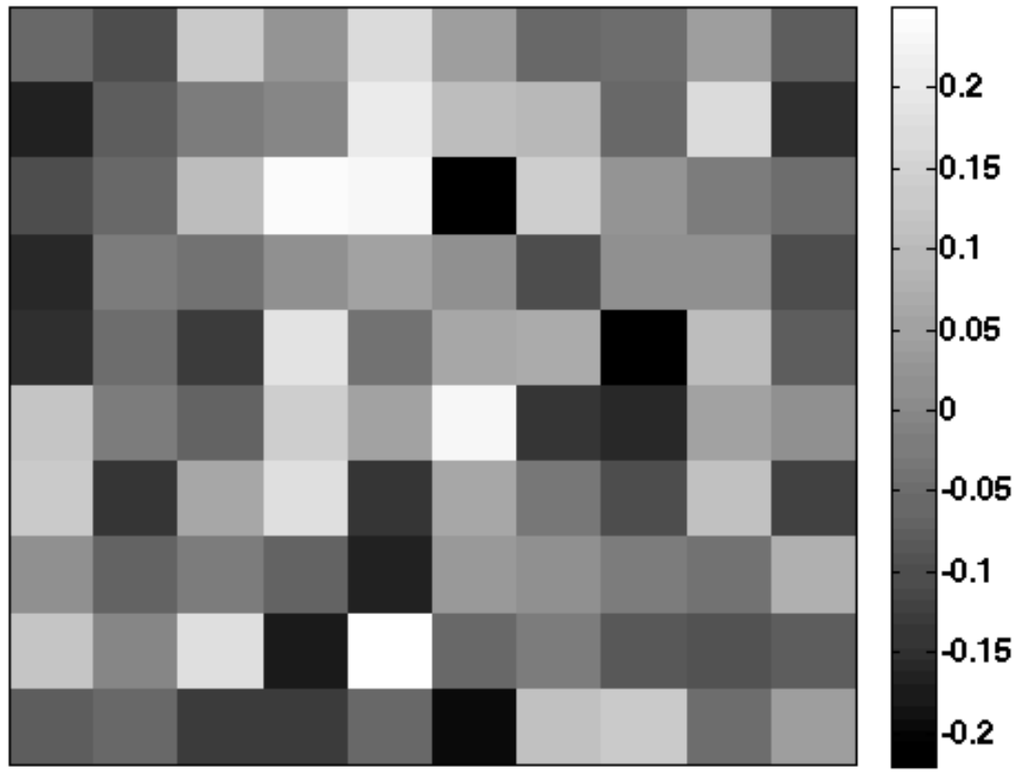


$D = 36{,}356$

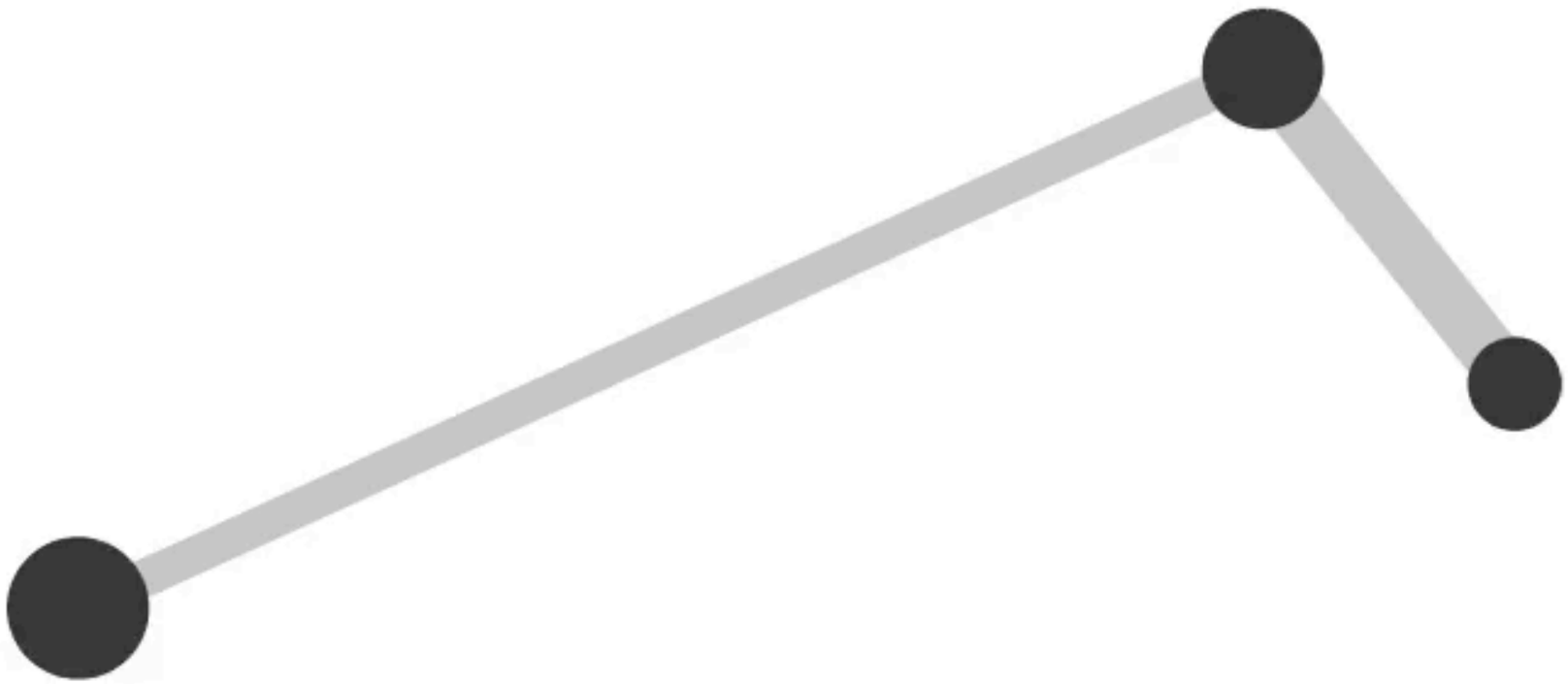


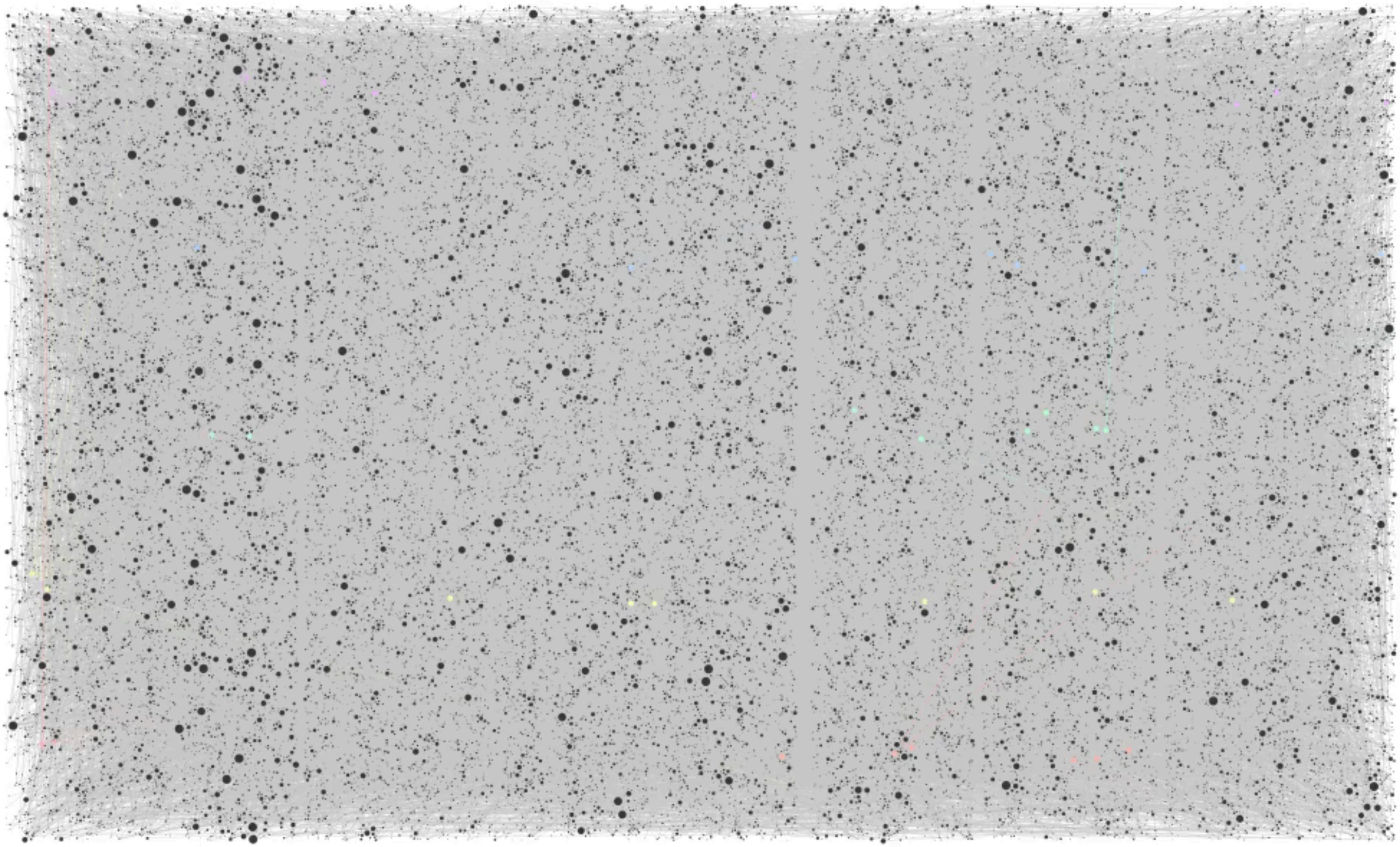
G

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



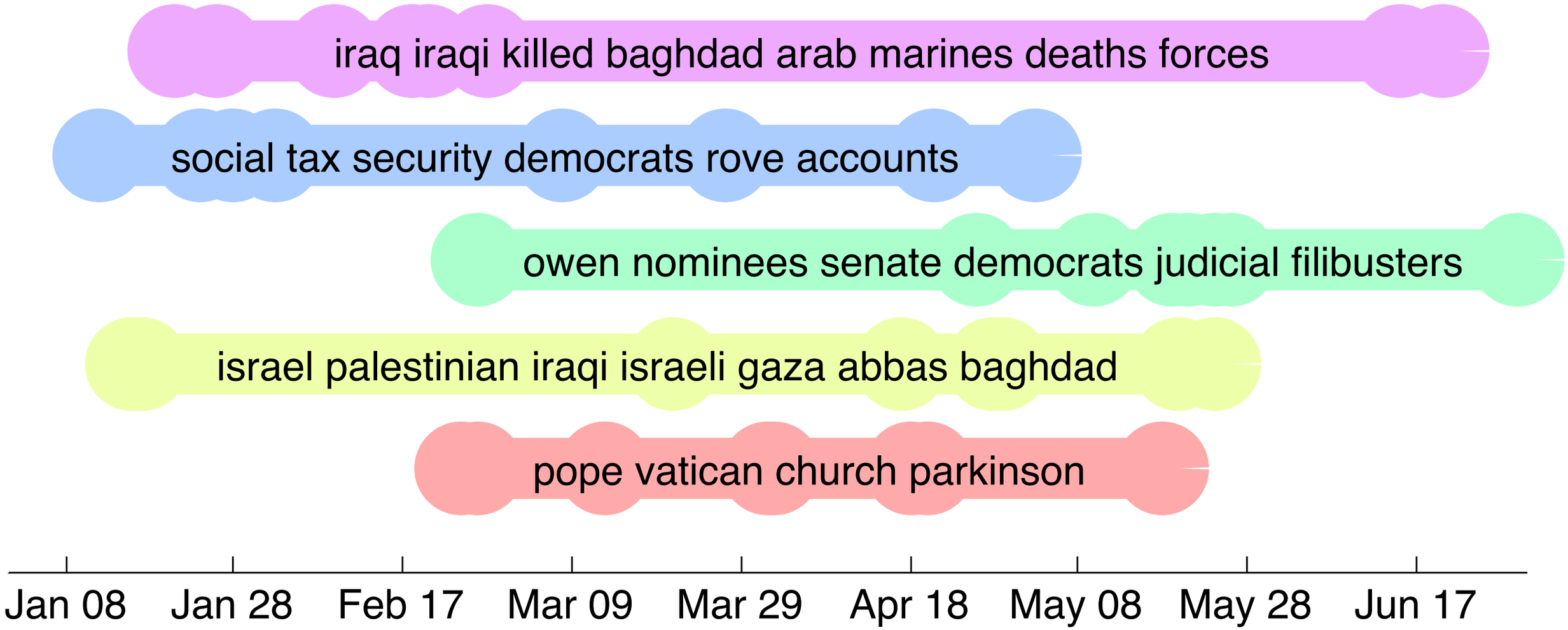
$d = 50$



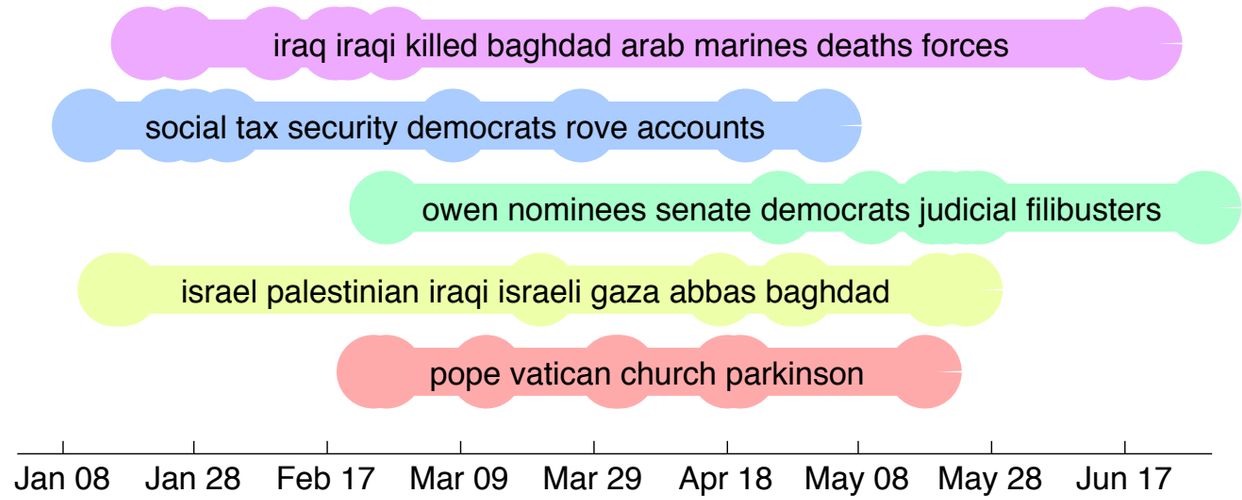


DPP THREADS

DPP THREADS


iraq iraqi killed baghdad arab marines deaths forces
social tax security democrats rove accounts
owen nominees senate democrats judicial filibusters
israel palestinian iraqi israeli gaza abbas baghdad
pope vatican church parkinson
Jan 08
Jan 28
Feb 17
Mar 09
Mar 29
Apr 18
May 08
May 28
Jun 17

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

QUANTITATIVE RESULTS

System											
ROUGE-1F											
R-SU4F											
Coherence											

QUANTITATIVE RESULTS

System	k-means
ROUGE-1F	16.5
R-SU4F	3.76
Coherence	2.73

QUANTITATIVE RESULTS

System	k-means	DTM
ROUGE-1F	16.5	14.7
R-SU4F	3.76	3.44
Coherence	2.73	3.2

QUANTITATIVE RESULTS

System	k-means	DTM	DPP
ROUGE-1F	16.5	14.7	**17.2**
R-SU4F	3.76	3.44	**3.98**
Coherence	2.73	3.2	**3.3**

QUANTITATIVE RESULTS

System	k-means	DTM	DPP
ROUGE-1F	16.5	14.7	**17.2**
R-SU4F	3.76	3.44	**3.98**
Coherence	2.73	3.2	**3.3**
Runtime (s)	626	19,434	**252**

OTHER POTENTIAL NLP APPLICATIONS

POTENTIAL APP: RE-RANKING

POTENTIAL APP: RE-RANKING

- Parser: simple model with local features defines basic scores for all possible parse trees

POTENTIAL APP: RE-RANKING

- Parser: simple model with local features defines basic scores for all possible parse trees
- Re-ranker: more complex model with non-local features provides more refined scores

POTENTIAL APP: RE-RANKING

- Parser: simple model with local features defines basic scores for all possible parse trees
- Re-ranker: more complex model with non-local features provides more refined scores
- Typical pipeline: find the k highest-scoring parses under the simple model, then score these k with the more complex model and output the best

POTENTIAL APP: RE-RANKING

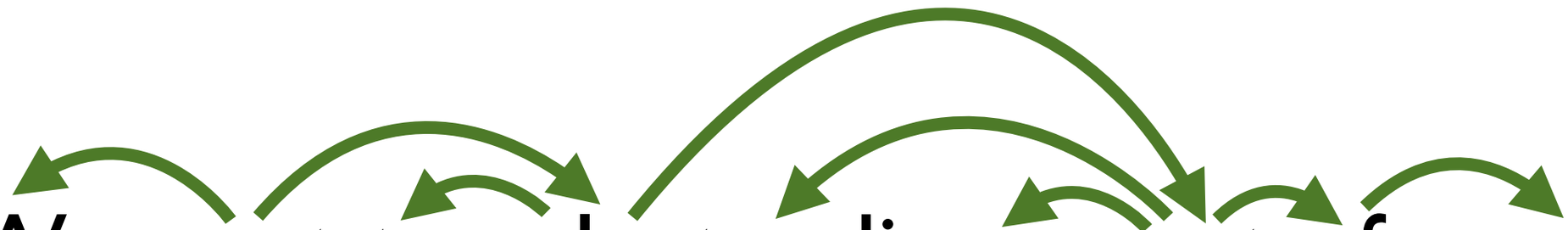
- Parser: simple model with local features defines basic scores for all possible parse trees
- Re-ranker: more complex model with non-local features provides more refined scores
- Typical pipeline: find the k highest-scoring parses under the simple model, then score these k with the more complex model and output the best
- Issue: the k may be largely redundant, so re-ranker does not get to consider significantly different parses

IDEA: USE DPPS FOR SELECTING RE-RANKER INPUT

IDEA: USE DPPS FOR SELECTING RE-RANKER INPUT

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

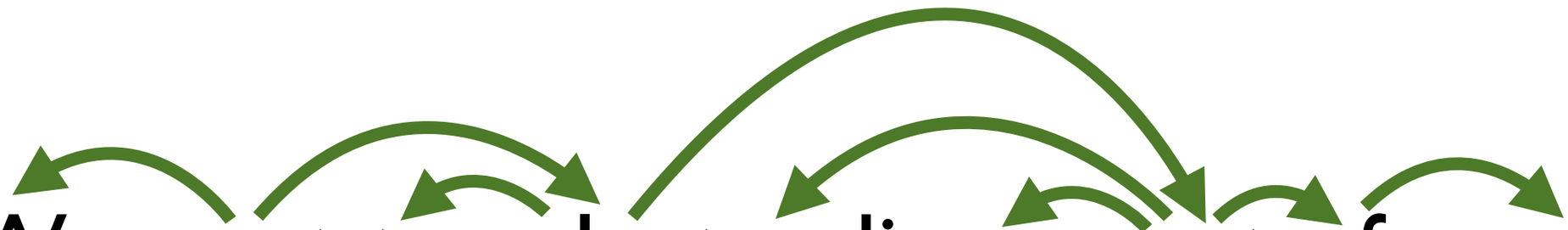
IDEA: USE DPPS FOR SELECTING RE-RANKER INPUT



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

Quality:
standard
parser scores

IDEA: USE DPPS FOR SELECTING RE-RANKER INPUT



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

Quality:
standard
parser scores

Diversity:
edge lengths,
POS pairs, etc.

POTENTIAL APP: WORD SENSE INDUCTION

POTENTIAL APP: WORD SENSE INDUCTION

- Goal: identify all possible senses of ambiguous words (e.g. river bank vs bank deposit)

POTENTIAL APP: WORD SENSE INDUCTION

- Goal: identify all possible senses of ambiguous words (e.g. river bank vs bank deposit)
- Typical approach: unsupervised clustering, cluster centers represent word senses

POTENTIAL APP: WORD SENSE INDUCTION

- Goal: identify all possible senses of ambiguous words (e.g. river bank vs bank deposit)
- Typical approach: unsupervised clustering, cluster centers represent word senses
- Why DPPs fit: can re-express finding cluster centers as the problem of finding a high-quality, diverse set

POTENTIAL APP: WORD SENSE INDUCTION

- Goal: identify all possible senses of ambiguous words (e.g. river bank vs bank deposit)
- Typical approach: unsupervised clustering, cluster centers represent word senses
- Why DPPs fit: can re-express finding cluster centers as the problem of finding a high-quality, diverse set

Quality:
centrality (density of points nearby)

POTENTIAL APP: WORD SENSE INDUCTION

- Goal: identify all possible senses of ambiguous words (e.g. river bank vs bank deposit)
- Typical approach: unsupervised clustering, cluster centers represent word senses
- Why DPPs fit: can re-express finding cluster centers as the problem of finding a high-quality, diverse set

Quality: centrality (density of points nearby)

Diversity: same as standard WSI features

QUESTIONS?