

# A Brief Overview of Sublinear Matrix Approximation Techniques

UMassAmherst

Manning College of Information  
& Computer Sciences

## Matrices in the Wild

# Distance matrices

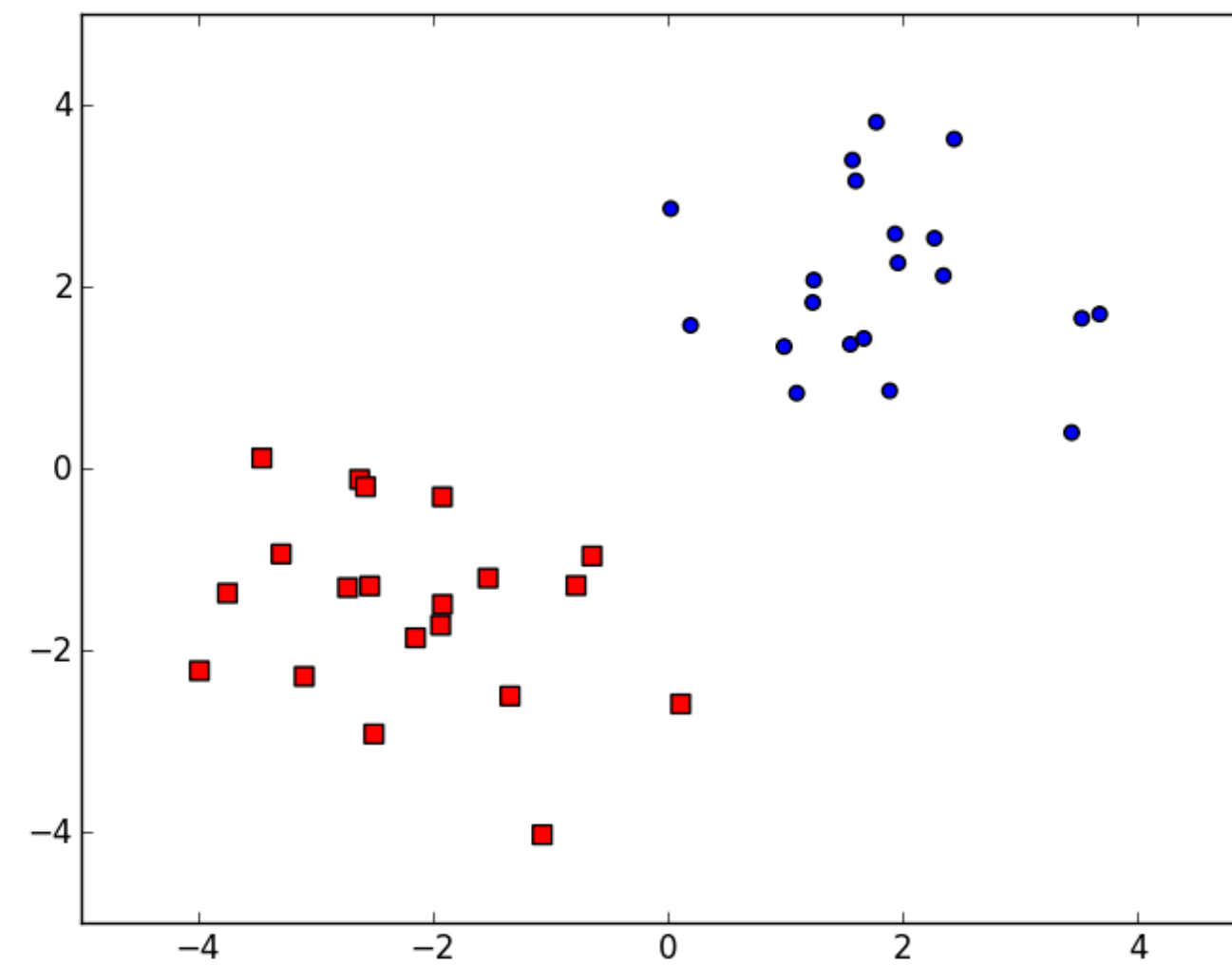


Image source: <http://blog.sairahul.com/2014/01/linear-separability.html>

## Matrices in the Wild

# Similarity matrices

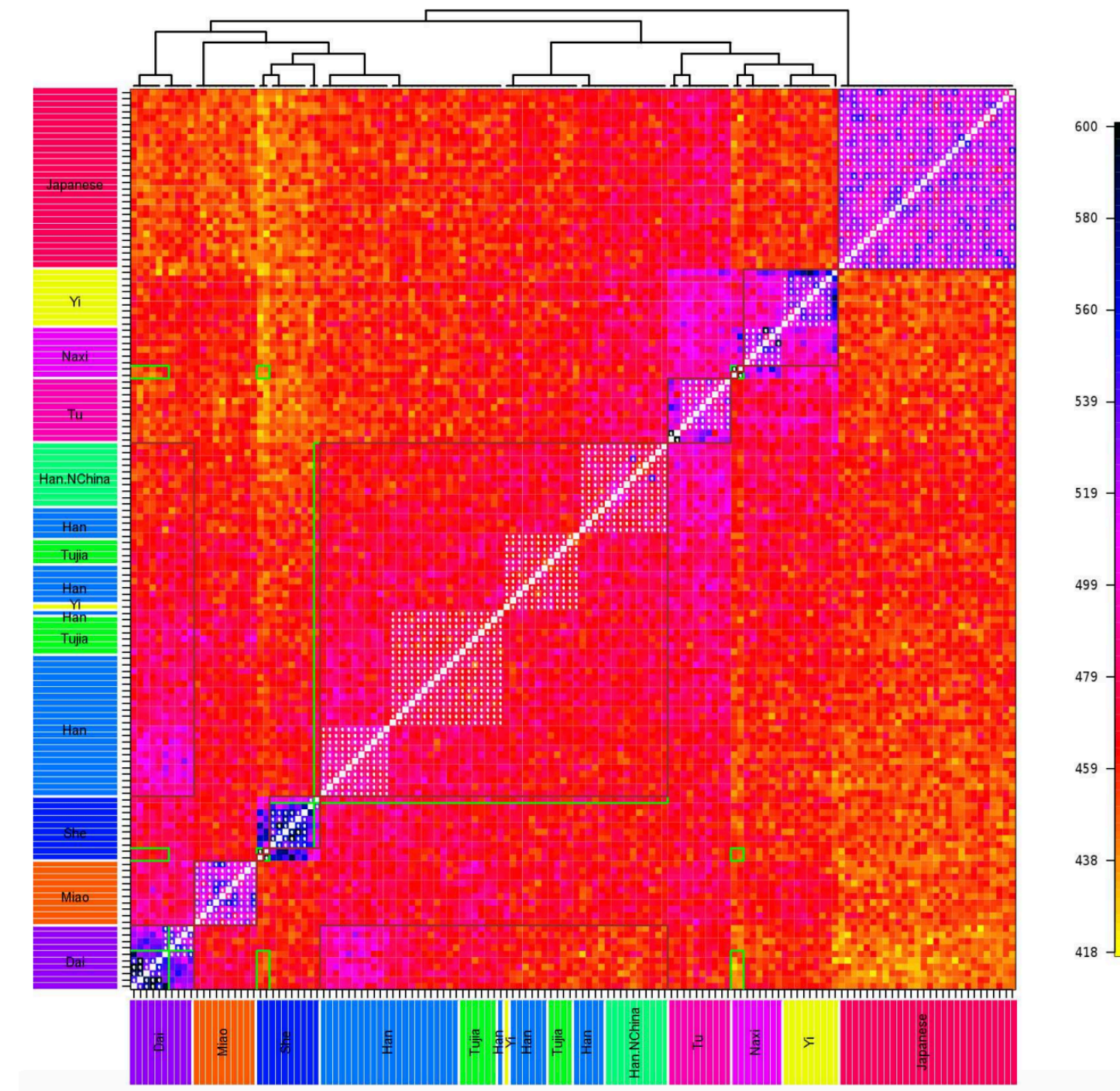


Image source: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.727.7925&rep=rep1&type=pdf>

## Matrices in the Wild

# Kernel matrices

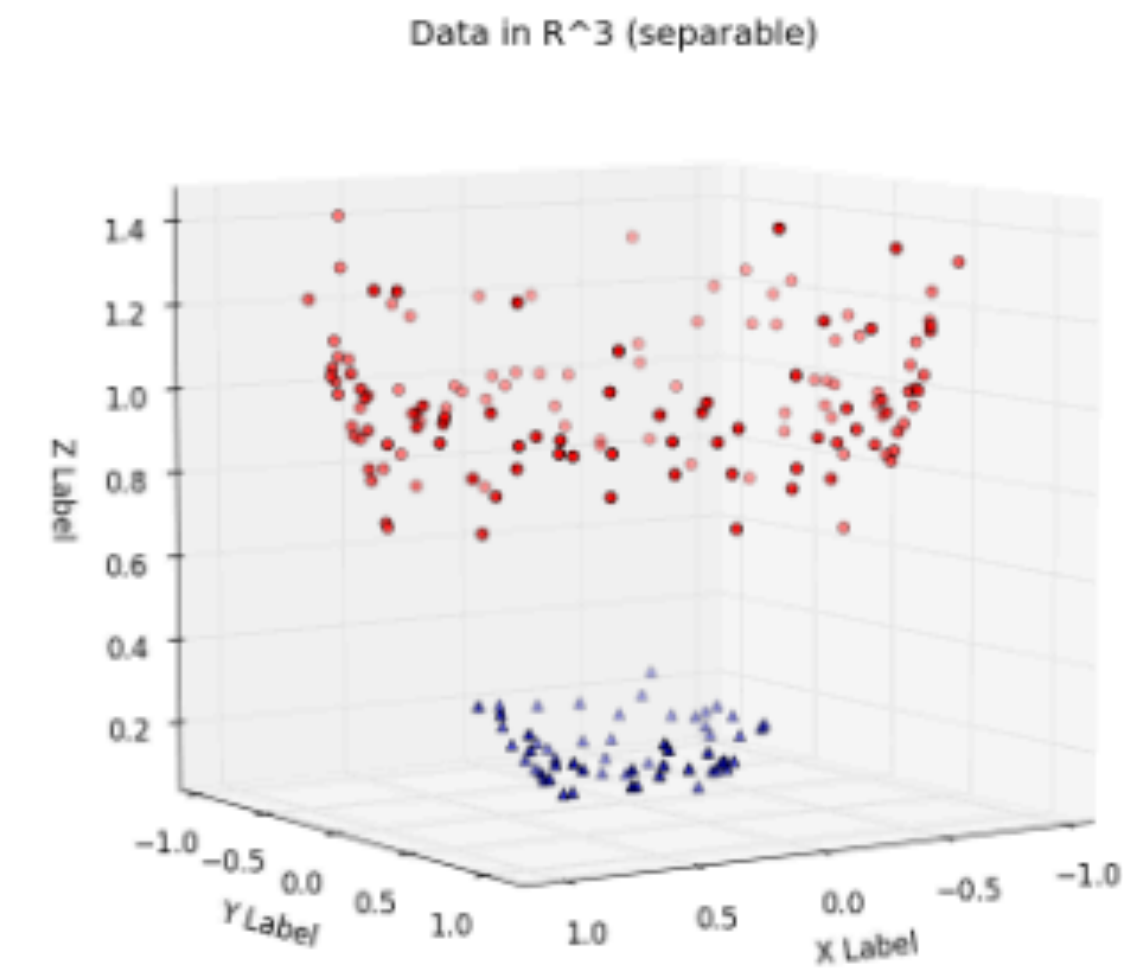
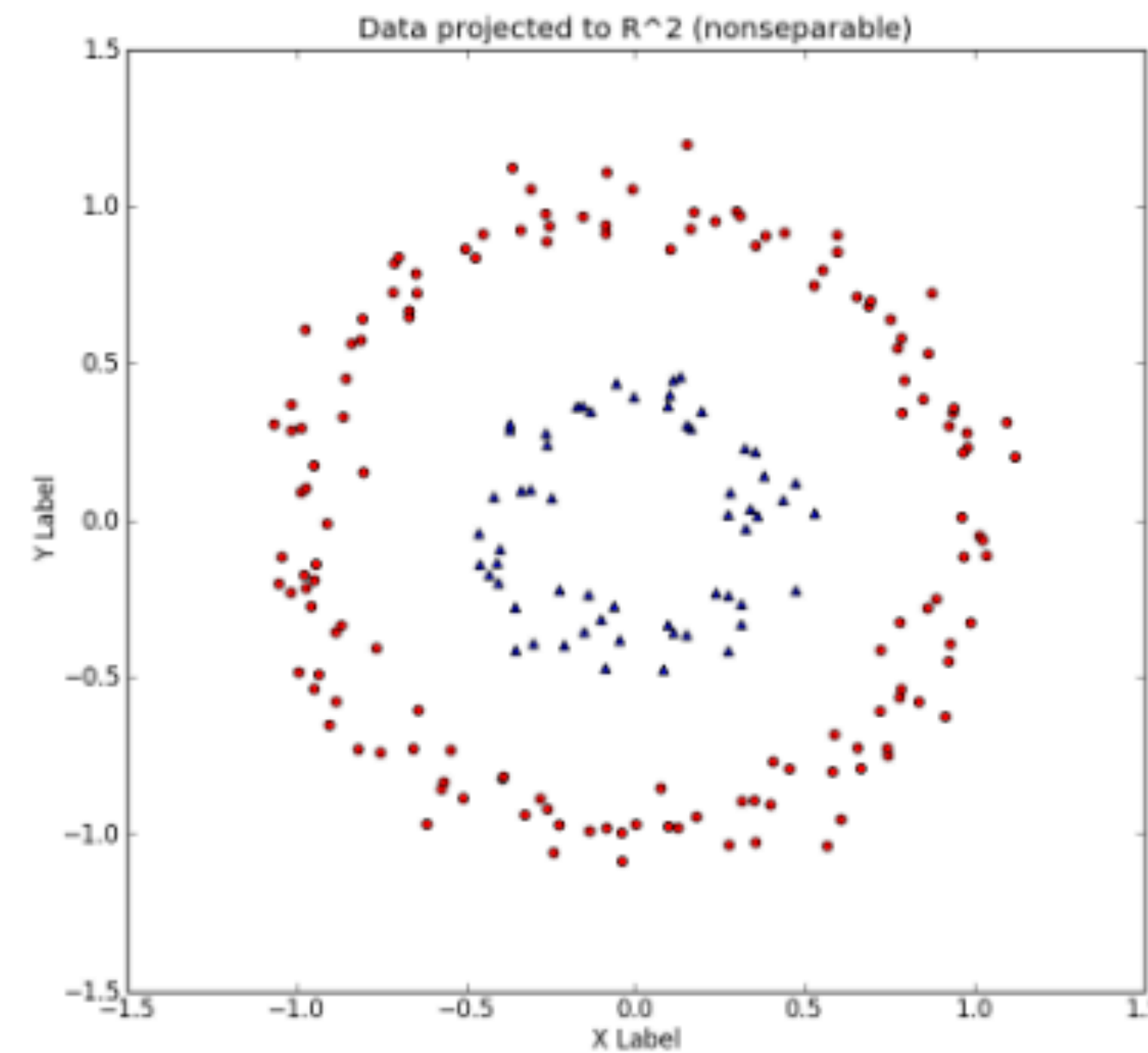


Image source: [http://www.eric-kim.net/eric-kim-net/posts/1/kernel\\_trick.html](http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html)

Define implicit mapping function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  such that  $\exists K(x, y) = \langle \phi(x), \phi(y) \rangle$  is easy to compute



## Computational Cost

$\Omega(n^2)$  comparisons among datapoints

Expensive comparisons between documents results in computational bottleneck

## Computational Cost

$\Omega(n^2)$  comparisons among datapoints

Expensive comparisons between documents results in computational bottleneck

Can we approximate these in sublinear time without losing much information

## Sublinear Methods

Define implicit mapping function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  such that  $\exists K(x, y) = \langle \phi(x), \phi(y) \rangle$  is easy to compute

### Random features approximations of kernel matrices [1]

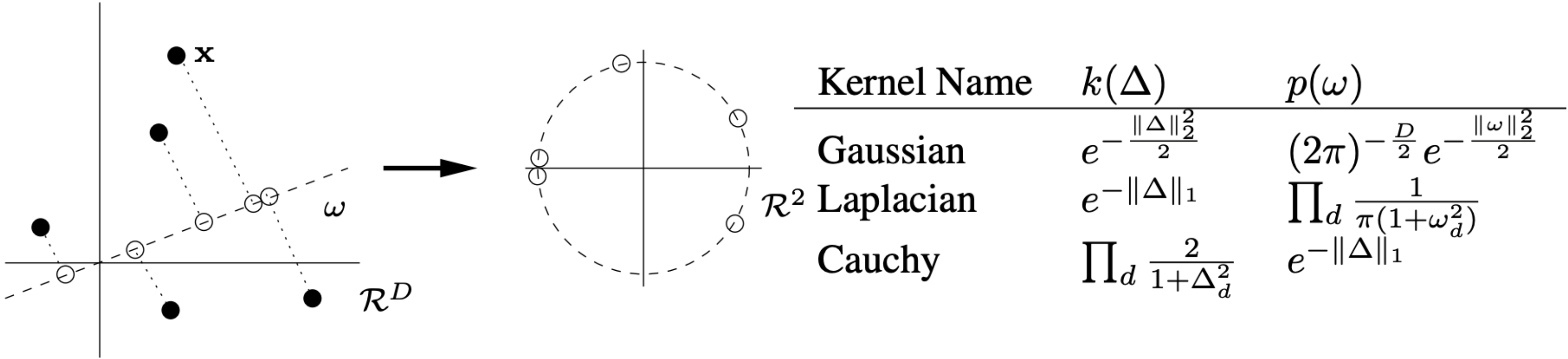
Works on shift invariant kernel's:  $K(x, y) = K(x - y)$

[1] Rahimi, A. and Recht, B., 2007. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20

# Sublinear Methods

Define implicit mapping function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  such that  $\exists K(x, y) = \langle \phi(x), \phi(y) \rangle$  is easy to compute

## Random features approximations of kernel matrices [1]



Transparent dots are  $z(x)$  and  $K(x, y) \approx \langle z(x), z(y) \rangle$

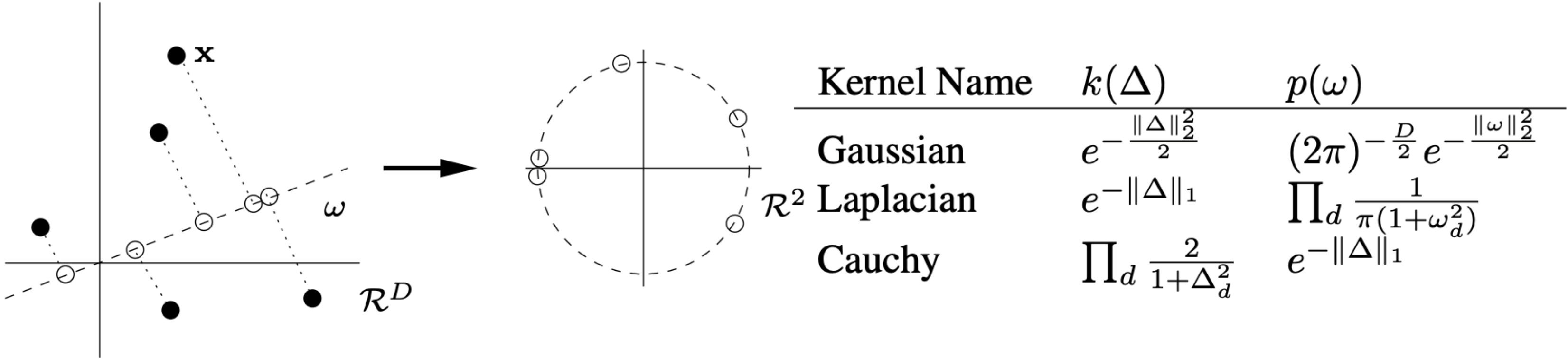
[1] Rahimi, A. and Recht, B., 2007. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20



# Sublinear Methods

Define implicit mapping function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  such that  $\exists K(x, y) = \langle \phi(x), \phi(y) \rangle$  is easy to compute

## Random features approximations of kernel matrices [1]



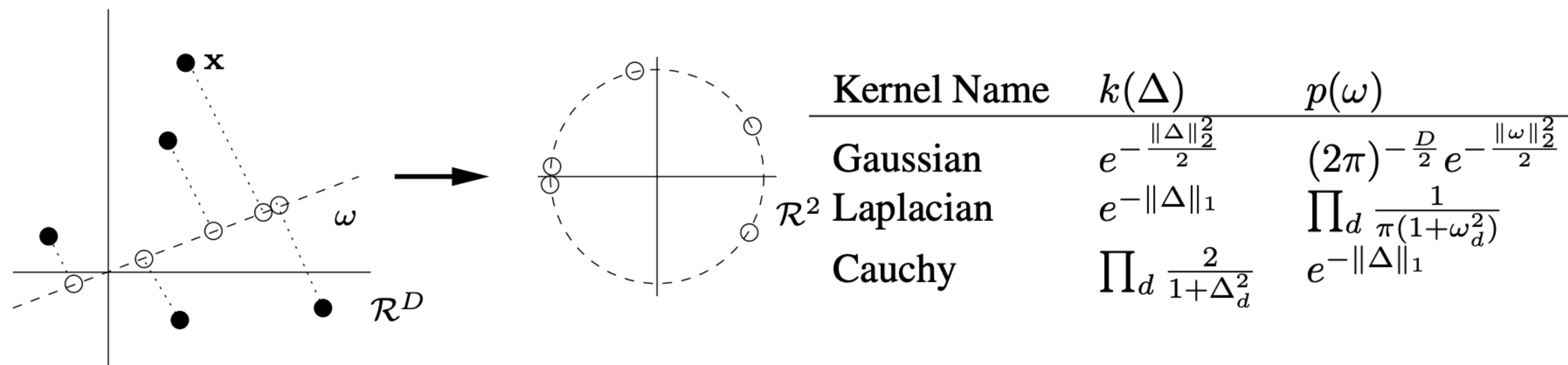
We can store  $Z \in \mathbb{R}^{N \times R}$  where  $R \ll N$

[1] Rahimi, A. and Recht, B., 2007. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20

## Sublinear Methods

Define implicit mapping function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  such that  $\exists K(x, y) = \langle \phi(x), \phi(y) \rangle$  is easy to compute

### Random features approximations of kernel matrices [1]



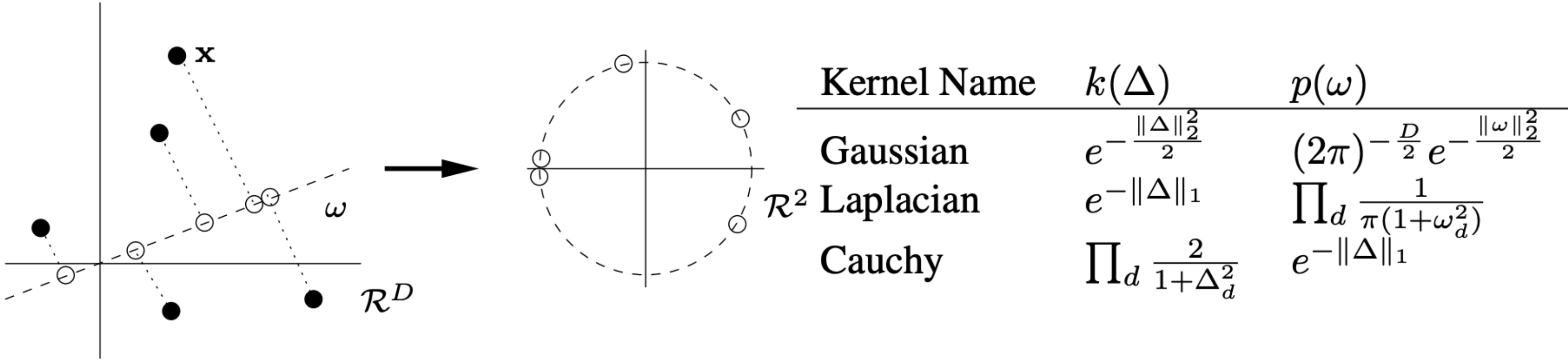
Prediction in say linear regression then becomes:  $\beta = (Z^T Z)^{-1} Z^T Y$

[1] Rahimi, A. and Recht, B., 2007. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20

# Sublinear Methods

Define implicit mapping function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  such that  $\exists K(x, y) = \langle \phi(x), \phi(y) \rangle$  is easy to compute

## Random features approximations of kernel matrices [1]



Matrix inversion is  $O(NR^2)$  instead of  $O(N^3)$

[1] Rahimi, A. and Recht, B., 2007. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20

## Sublinear Methods

### Nyström approximation [1]

For any matrix  $K \in \mathbb{R}^{n \times n}$  compute approximation as  $KS(S^T KS)^+ S^T K$

$S \in \mathbb{R}^{n \times s}$  samples columns of  $K$  at random

[1] Nyström, E.J., 1930. Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben. *Acta Mathematica*, 54, pp.185-204.



## Sublinear Methods

### Nyström approximation [1]

For any matrix  $K \in \mathbb{R}^{n \times n}$  compute approximation as  $KS(S^T KS)^+ S^T K$

$S \in \mathbb{R}^{n \times s}$  samples columns of  $K$  at random

Uniform sampling may fail in natural datasets where relative importance of data is not uniform

[1] Nyström, E.J., 1930. Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben. *Acta Mathematica*, 54, pp.185-204.

## Sublinear Methods

### Nyström approximation

For any matrix  $K \in \mathbb{R}^{n \times n}$  compute approximation as  $KS(S^T KS)^+ S^T K$

$S \in \mathbb{R}^{n \times s}$  samples columns of  $K$  at random

We can then use alternate sampling methods like leverage scores [1,2] to choose  $S$

- [1] Cohen, M.B., Musco, C. and Pachocki, J., 2016. Online row sampling. *arXiv preprint arXiv:1604.05448*.
- [2] Musco, C. and Musco, C., 2017. Recursive sampling for the nystrom method. *Advances in Neural Information Processing Systems*, 30.

## Sublinear Methods

### Pseudoskeleton approximation [1]

For any matrix  $K \in \mathbb{R}^{n \times n}$  compute approximation as  $KS_2(S_2^T KS_1)^+ S_1^T K$

$S_1, S_2 \in \mathbb{R}^{n \times s}$  samples columns of  $K$  at random

[1] Goreinov, S.A., Tyrtyshnikov, E.E. and Zamarashkin, N.L., 1997. A theory of pseudoskeleton approximations. *Linear algebra and its applications*, 261(1-3), pp.1-21.

## Sublinear Methods

CUR [1]

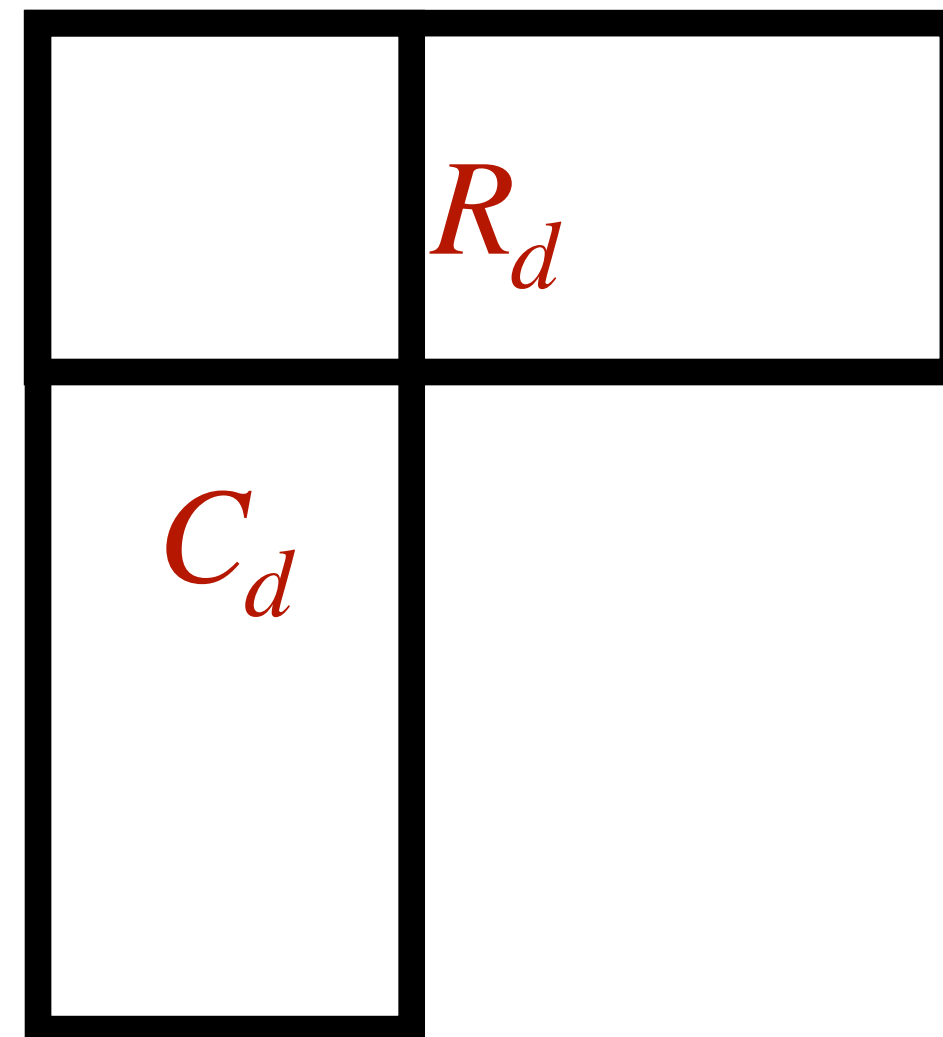
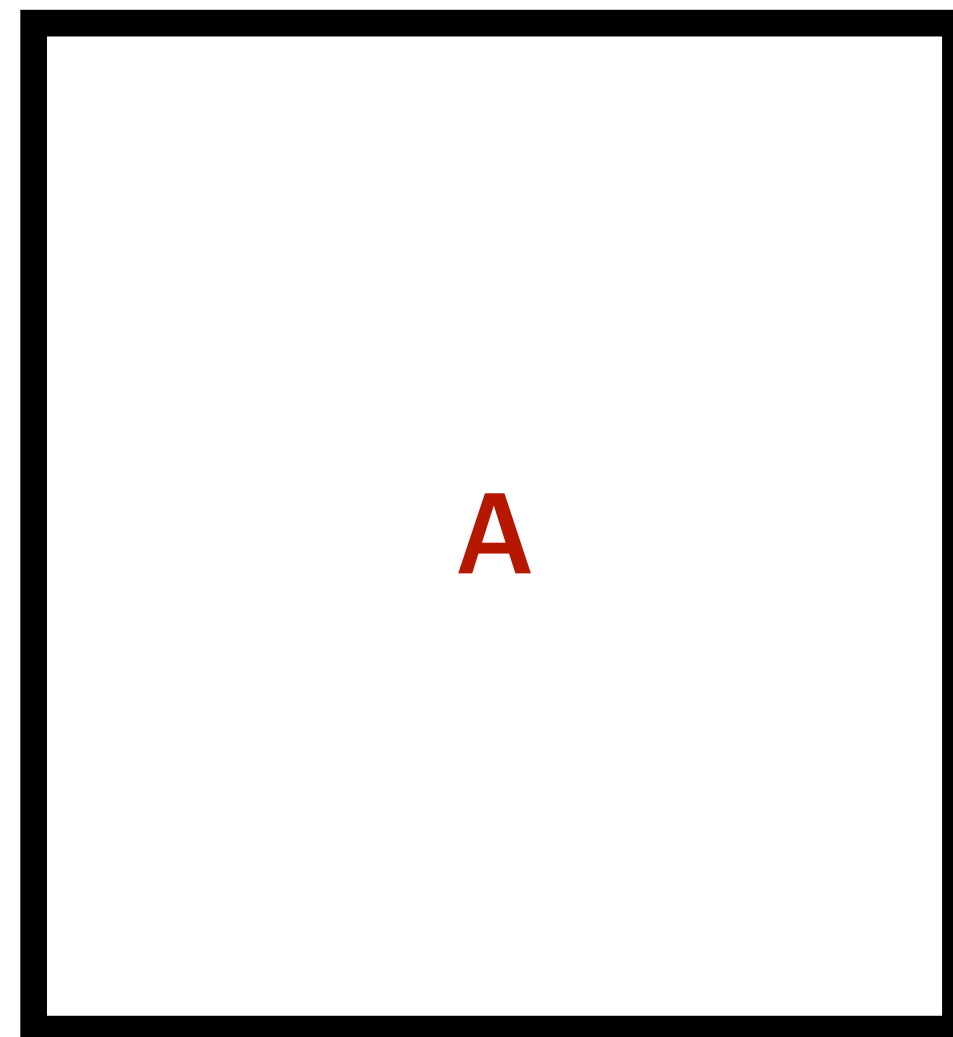


[1] Mahoney, M.W. and Drineas, P., 2009. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3), pp.697-702.



## Sublinear Methods

CUR [1]

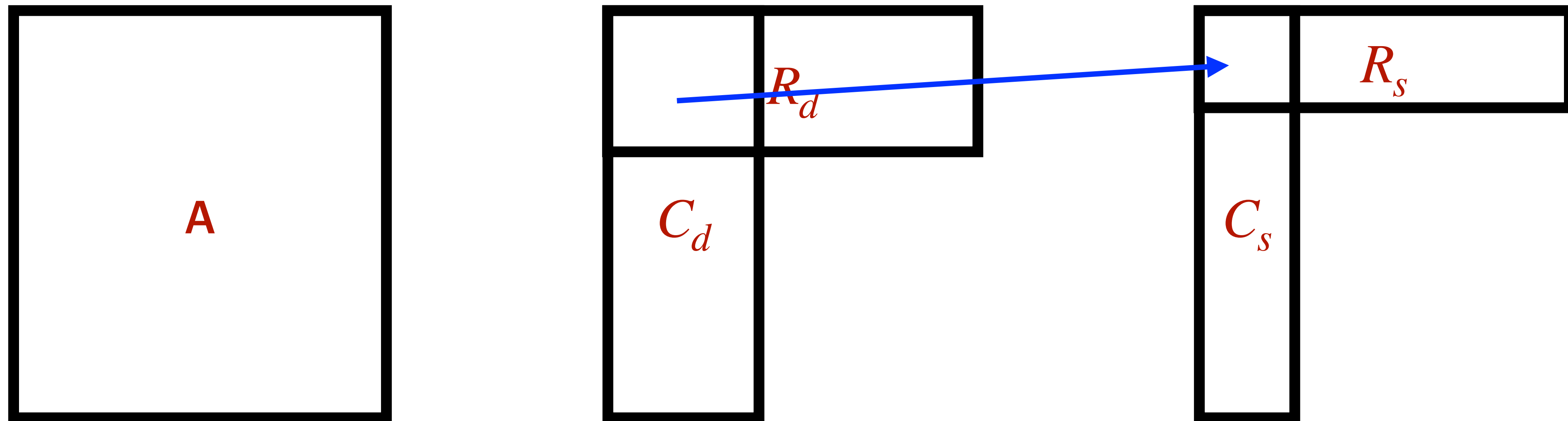


[1] Mahoney, M.W. and Drineas, P., 2009. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3), pp.697-702.

## Sublinear Methods

### CUR [1]

Construct small  $U$

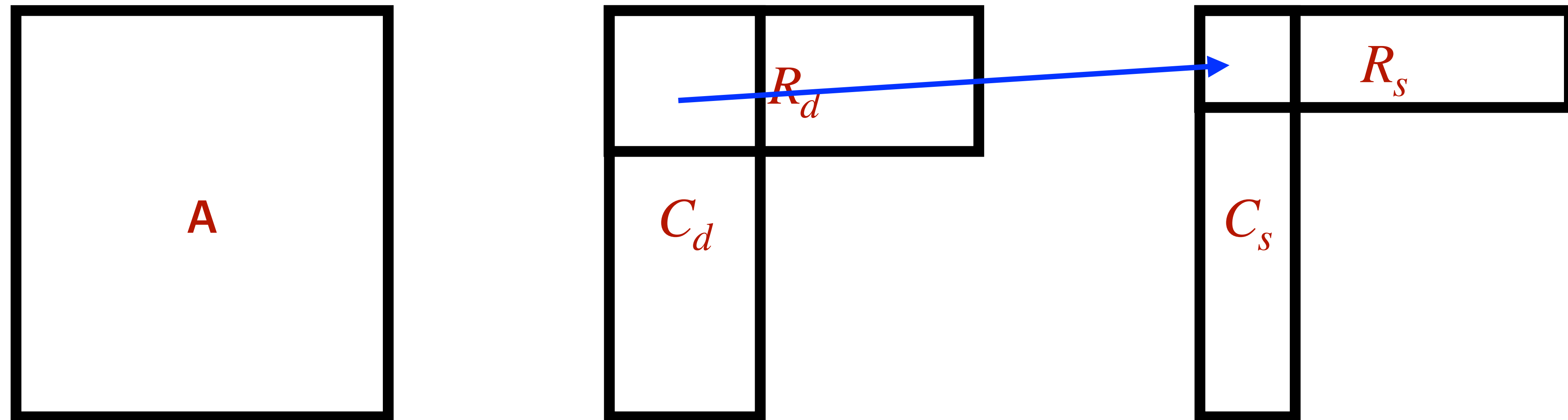


[1] Mahoney, M.W. and Drineas, P., 2009. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3), pp.697-702.

## Sublinear Methods

### CUR [1]

Construct small  $U$



For multiple copies keep one column/row but multiply it by  $\sqrt{d}$

# Sublinear Time Approximation of Text Similarity Matrices

Archan Ray, Nicholas Monath, Andrew McCallum,  
Cameron Musco

UMassAmherst

Manning College of Information  
& Computer Sciences



## A Rapid and Specific Assay for the Detection of MERS-CoV

Pei Huang<sup>1,2</sup>, Hualei Wang<sup>2,3,4\*</sup>, Zengguo Cao<sup>2,3</sup>, Hongli Jin<sup>2,3</sup>, Hang Beibei Yu<sup>5</sup>, Feihu Yan<sup>6</sup>, Xingxing Hu<sup>1,2</sup>, Fangfang Wu<sup>6</sup>, Cuicui Jiao<sup>7</sup>, Shangnan Xu<sup>1,2</sup>, Yongkun Zhao<sup>8,9</sup>, Na Feng<sup>6,8</sup>, Jianzhong Wang<sup>1</sup>, W Tiecheng Wang<sup>2,4</sup>, Yuwei Gao<sup>1,4</sup>, Songtao Yang<sup>4,4</sup> and Xianzhu Xia<sup>2,3</sup>

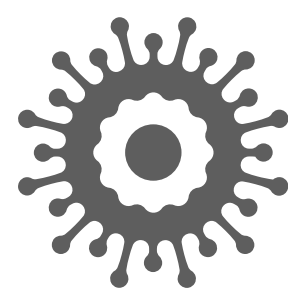
<sup>1</sup>Animal Science and Technology College, Jilin Agricultural University, Changchun, China, <sup>2</sup>Institute of Zoonosis Prevention and Control, Institute of Military Veterinary, Academy of Military Medical Science, Beijing, China, <sup>3</sup>College of Veterinary Medicine, Jilin University, Changchun, China, <sup>4</sup>Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, China, <sup>5</sup>State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, <sup>6</sup>Guangzhou Eighth People's Hospital of Guangzhou Medical University, Guangzhou, China, <sup>7</sup>Department of Clinical Laboratory, College of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China

### OPEN ACCESS

**Edited by:** Dirk Dittmer, University of North Carolina at Chapel Hill, United States  
**Reviewed by:** Timothy Sheahan, University of North Carolina at Chapel Hill, United States; Yibo Li, University of Pennsylvania, United States  
**\*Correspondence:** Hualei Wang, wanghz@journals.frontiersin.org

**Specialty section:** This article was submitted to Microbiology and Immunology, a specialty of the journal Frontiers in Microbiology.

Middle East respiratory syndrome coronavirus (MERS-CoV) is a novel human coronavirus that can cause severe respiratory illness. In this study, we established a rapid and specific reverse transcription loop isotherm amplification (RT-LAMP) assay. The result was visible by the naked eye. The detection limit of MERS-CoV RNA was 10<sup>3</sup> copies per reaction. The assay showed high specificity. Compared to the World Health Organization (WHO) reference method, the RT-LAMP assay requires less expensive equipment and is more suitable for point-of-care testing.



Amazingly, it is effective against SARS and MERS.

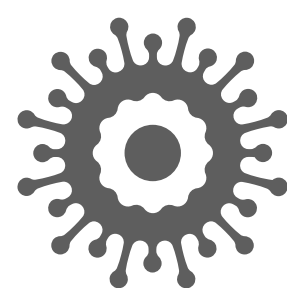


Contents lists available at ScienceDirect  
**Vaccine**  
journal homepage: www.elsevier.com/locate/vaccine

## DNA vaccine encoding Middle East respiratory syndrome coronavirus S1 protein induces protective immune responses in mice

Hang Chi<sup>a</sup>, Xuexing Zheng<sup>a,b</sup>, Xiwen Wang<sup>a</sup>, Chong Wang<sup>a</sup>, Hualei Wang<sup>a,c</sup>, Weiwei Gai<sup>a</sup>, Stanley Perlman<sup>d</sup>, Songtao Yang<sup>a,c,e</sup>, Jincun Zhao<sup>e,f</sup>, Xianzhu Xia<sup>a,c,g</sup>

<sup>a</sup>Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Institute of Military Veterinary, Academy of Military Medical Science, Changchun, China, <sup>b</sup>School of Public Health, Jilin University, Changchun, China, <sup>c</sup>Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, China, <sup>d</sup>Department of Microbiology, University of North Carolina at Chapel Hill, United States, <sup>e</sup>State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, <sup>f</sup>Guangzhou Eighth People's Hospital of Guangzhou Medical University, Guangzhou, China, <sup>g</sup>Department of Clinical Laboratory, College of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China



The Middle East respiratory syndrome coronavirus (MERS-CoV) is an emerging pathogen...

**ARTICLE IN PRESS**  
Article history:  
Received 10 June 2016  
Received in revised form 20 February 2017  
Accepted 28 February 2017  
Available online 14 March 2017

**Keywords:**  
MERS-CoV  
DNA vaccine  
Spike protein

### 1. Introduction

Middle East respiratory syndrome (MERS)-coronavirus (MERS-CoV), an emerging zoonotic virus, is the causative agent of MERS. MERS-CoV was first identified in Saudi Arabia in 2012 and MERS cases have been reported in 27 countries since then [1,2]. As of February 10, 2017, 1905 laboratory-confirmed cases, including 677 deaths related to MERS-CoV, had been reported to WHO (~36% mortality). Several family clusters and nosocomial clusters cases have been reported, revealing the human-to-human transmissibility of MERS-CoV, and raising the concern of a MERS-CoV global pandemic [3–5]. Currently, no licensed therapeutic or vaccine is available, which highlights the need for efficient vaccines against MERS-CoV.

To date, several vaccine candidates have been developed, such as viral vector-based recombinants [6–11], subunit vaccines [12–19], DNA vaccines [20], DNA prime/protein-boost vaccines [21] and a reverse genetics-constructed recombinant coronavirus vaccine [22]. Among them, DNA vaccines present a range of unique advantages such as proper antigen protein folding, rapid design and production, cost-effectiveness, and stability at non-refrigerated temperatures for convenient storage and shipping [23]. Furthermore, it has been reported that DNA vaccines can induce both humoral and cellular immune responses against MERS-CoV and SARS-CoV infection [20,24,25].

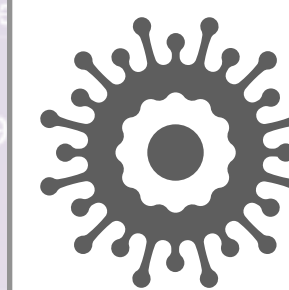
MERS-CoV is the first lineage of *Betacoronavirus* known to infect humans [26]. The genome of MERS-CoV encodes four structural proteins – spike (S), envelope (E), membrane (M) and nucleocapsid (N) [27]. The S protein, a class I fusion protein forming protruding

## Unexpected outbreaks of arbovirus infections: lessons learned from the Pacific and tropical America

Didier Musso, MD, Prof Alfonso J Rodriguez-Morales, MD, José Eduardo Levi, PhD

Van-Mai Cao-Lormeau

Published: June 19, 2018



Pandemic arboviruses have emerged as a major global health problem in the past four decades.

Check for updates

### Summary

Pandemic arboviruses have emerged as a major global health

decades. Predicting where and when the next outbreak will occur is a challenge, but history tells us that such swan events (epidemics that are predicted to have an extreme effect) will continue to occur as globalisation expands. We briefly review the history of arboviral epidemics that have occurred in the past 50 years in the American and Pacific regions, to provide a perspective on the future, and to highlight the need for improved surveillance, including laboratory-based

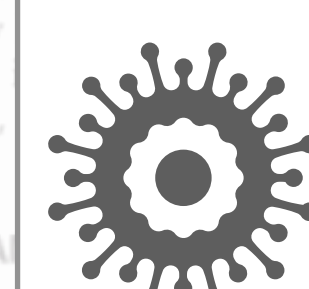
TICK-BORNE DISEASES

0025-7125/02 \$15.00 + .00

## COLORADO TICK FEVER

Richard Klasco, MD

Colorado tick fever (CTF), also known as *mountain fever* and *mountain tick fever*, is a well-described viral tick-borne disease common to the Rocky Mountain region of North America. The disease is characterized by a biphasic illness with a prodromal phase followed by a febrile phase. The disease is caused by the Colorado tick fever virus (CTFV), a member of the *Coltivirus* genus, family *Reoviridae*. The disease is transmitted from the bite of an infected wood tick.

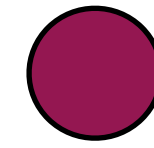
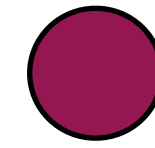
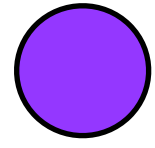
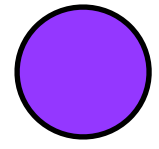


The arboviral infection, CTF, is transmitted from the bite of an infected wood tick.

CAUSE AND EFFECT

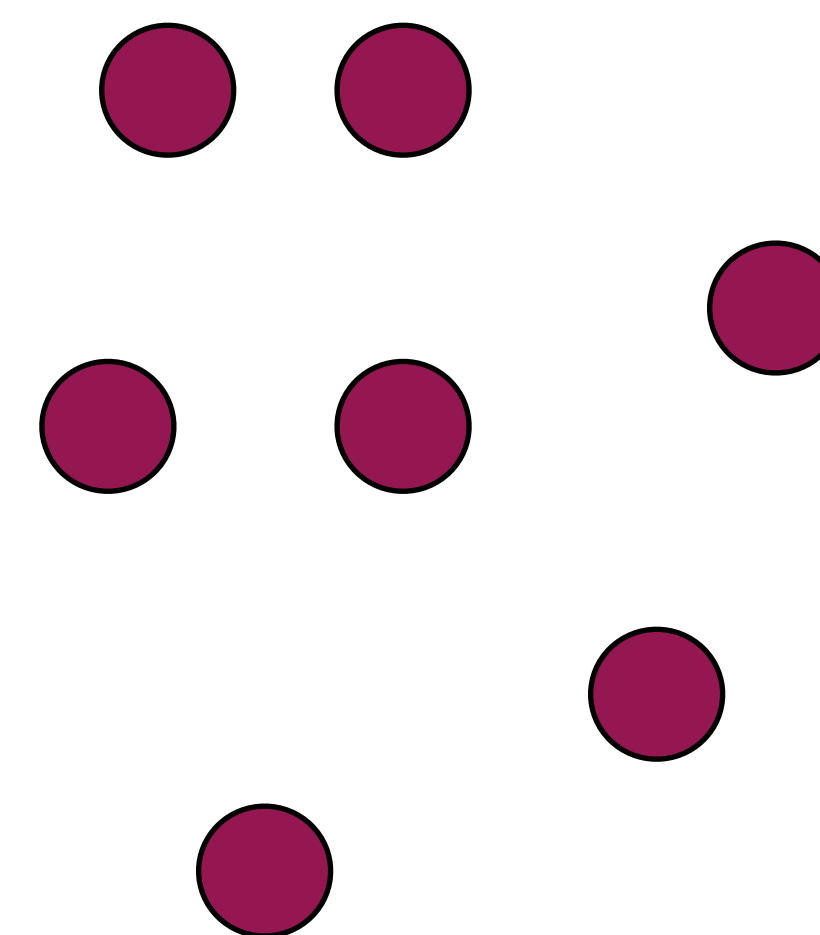
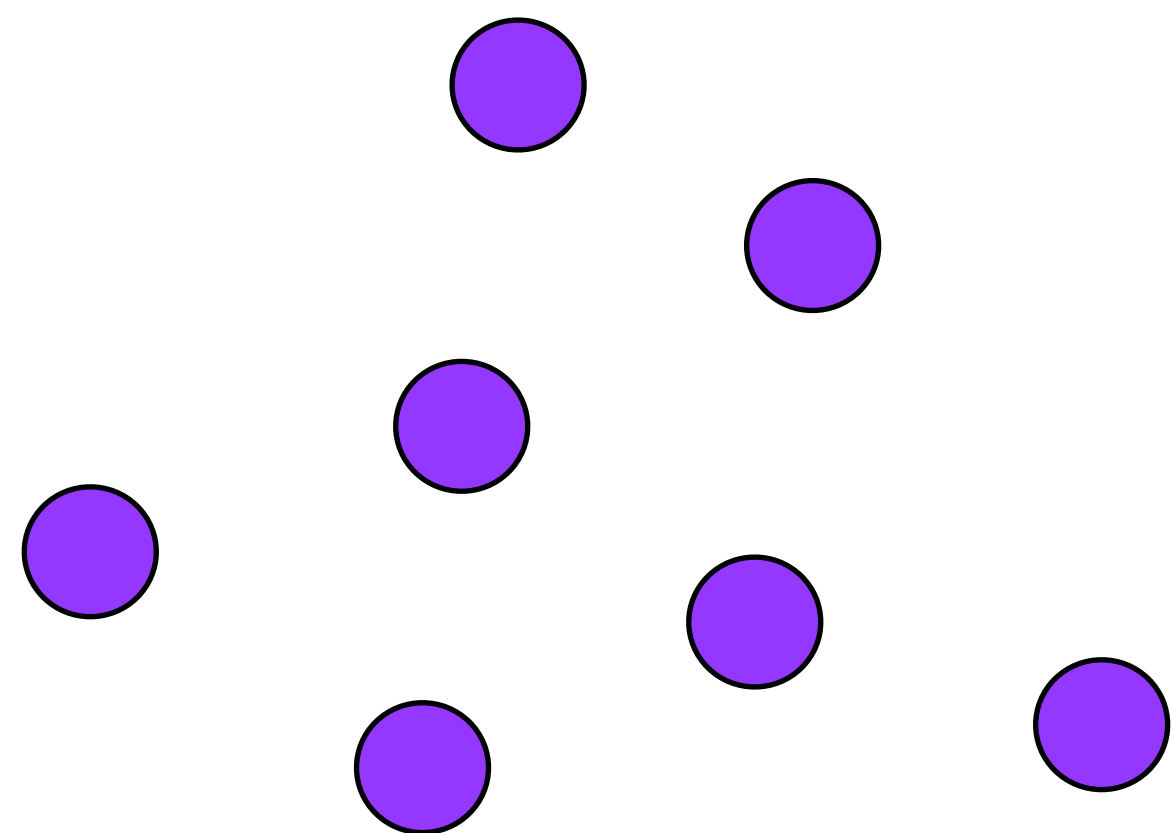
Colorado tick fever virus (CTFV), an orbivirus, is the causative agent of CTF. Formerly classified as an orbivirus, the sixth report of the International Committee on Taxonomy of Viruses identified CTFV as a member of the genus *Coltivirus* (group A), family *Reoviridae* (virus code, 60.0.4.0.001; virus accession number, 60040001).<sup>24</sup> At least 22 strains of CTFV are known,<sup>5, 24</sup> many of which cause disease in humans.<sup>8</sup> Of these, the Florio strain is the best characterized.<sup>8</sup> Eyach, a group A *Coltivirus* closely related to CTFV, has been detected in European Ixodidae ticks and has been implicated in human disease in Czechoslovakia.<sup>12, 24</sup>

In 2000, the CTFV genome was sequenced and was found to consist of 12 dsRNA segments that encode several important proteins.<sup>3</sup> These include VP1, the viral RNA dependent RNA polymerase; methyltransferases; RGD-binding proteins, extracellular proteins that mediate cell-



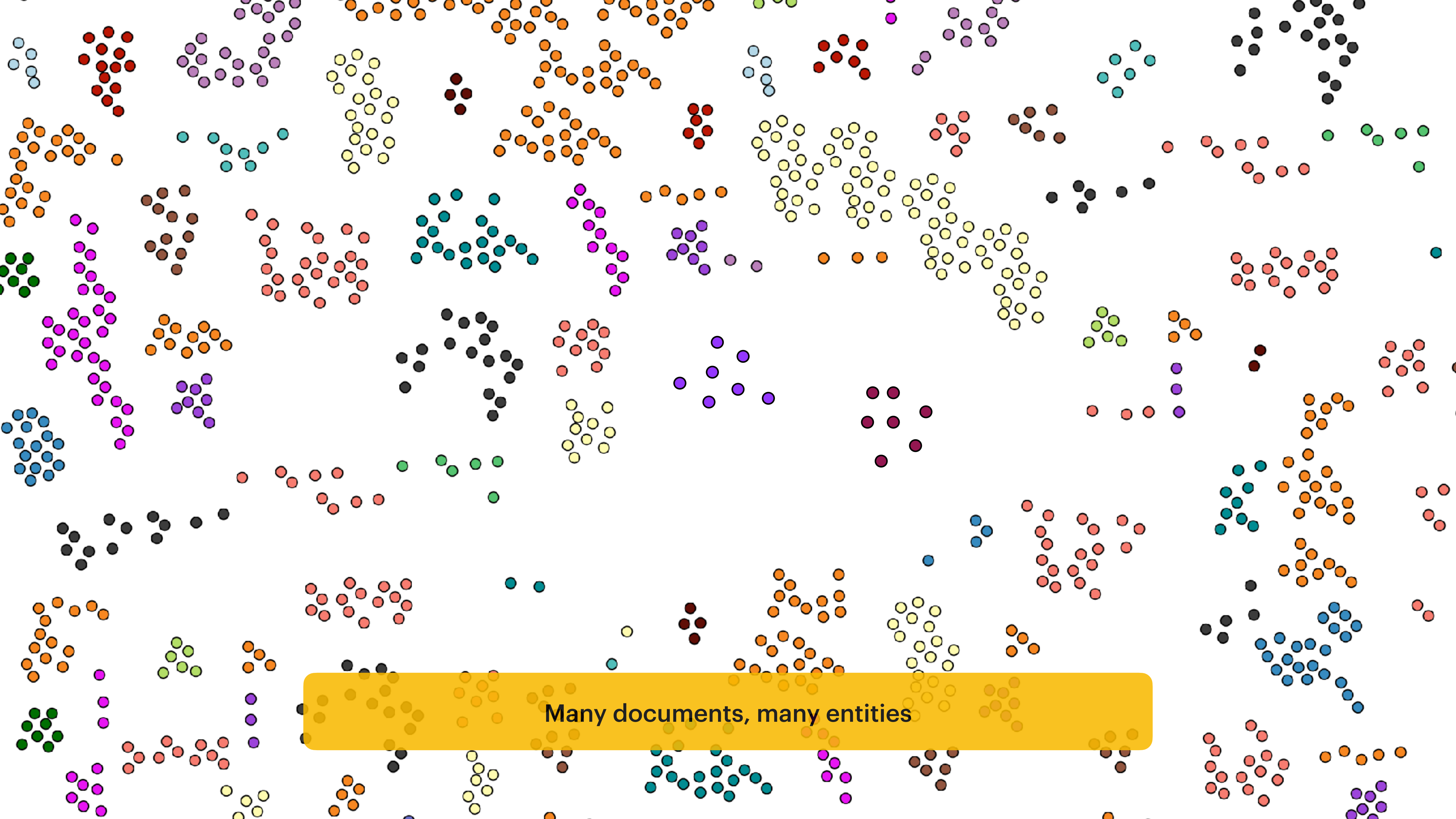
$\mathbb{R}^L$

$\mathbb{R}^L$



Numerous documents mentioning the same entities





Many documents, many entities





$O(n^2)$  similarities

Many documents, many entities

0.0089	0.3262	..	..	..	0.1820	0.2993
0.3262	0.6635	..	..	..	0.4246	0.1990
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
0.1820	0.4246	..	..	..	0.3865	0.2350
0.2993	0.1990	..	..	..	0.2350	0.8414

*O(n<sup>2</sup>) similarities*

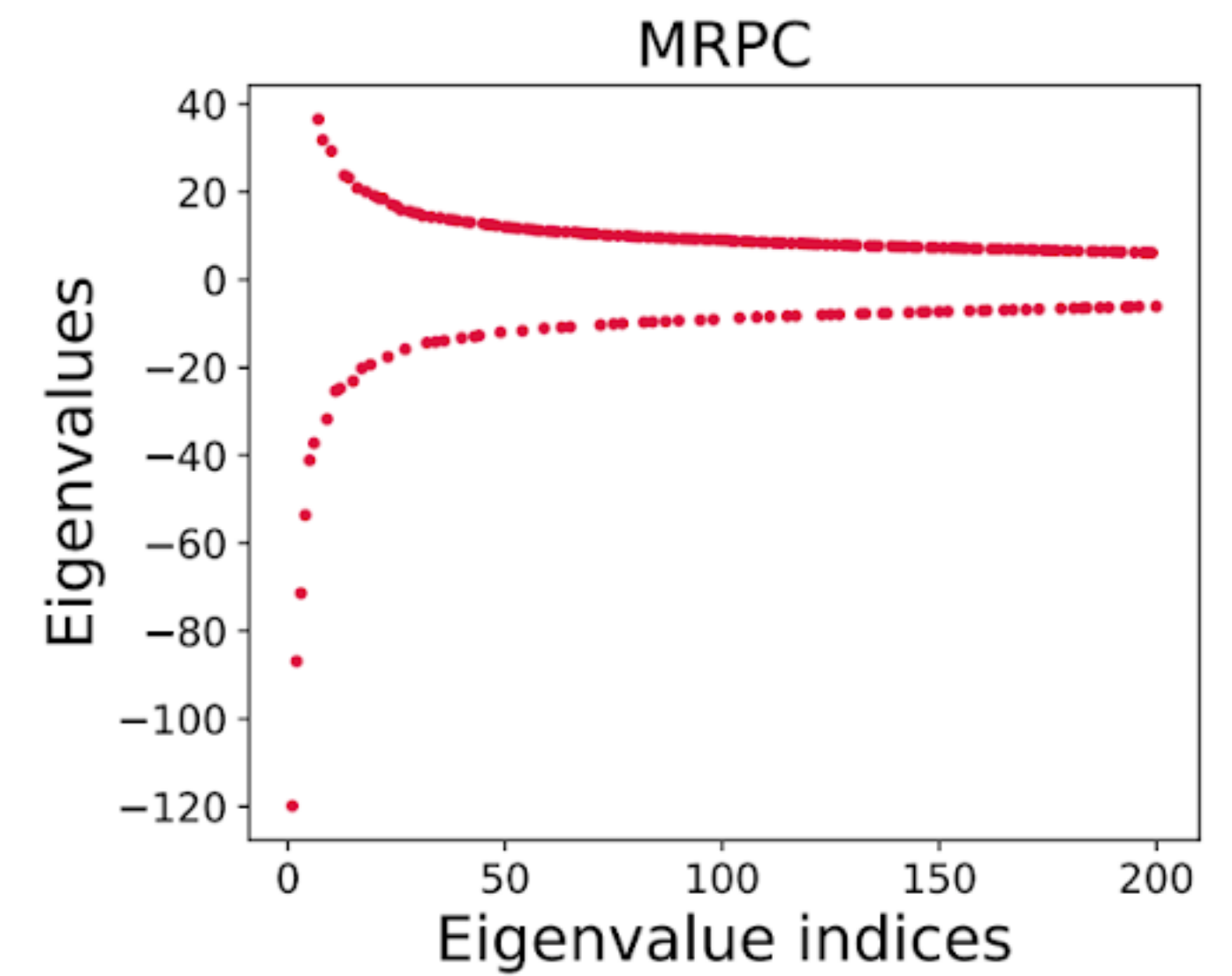
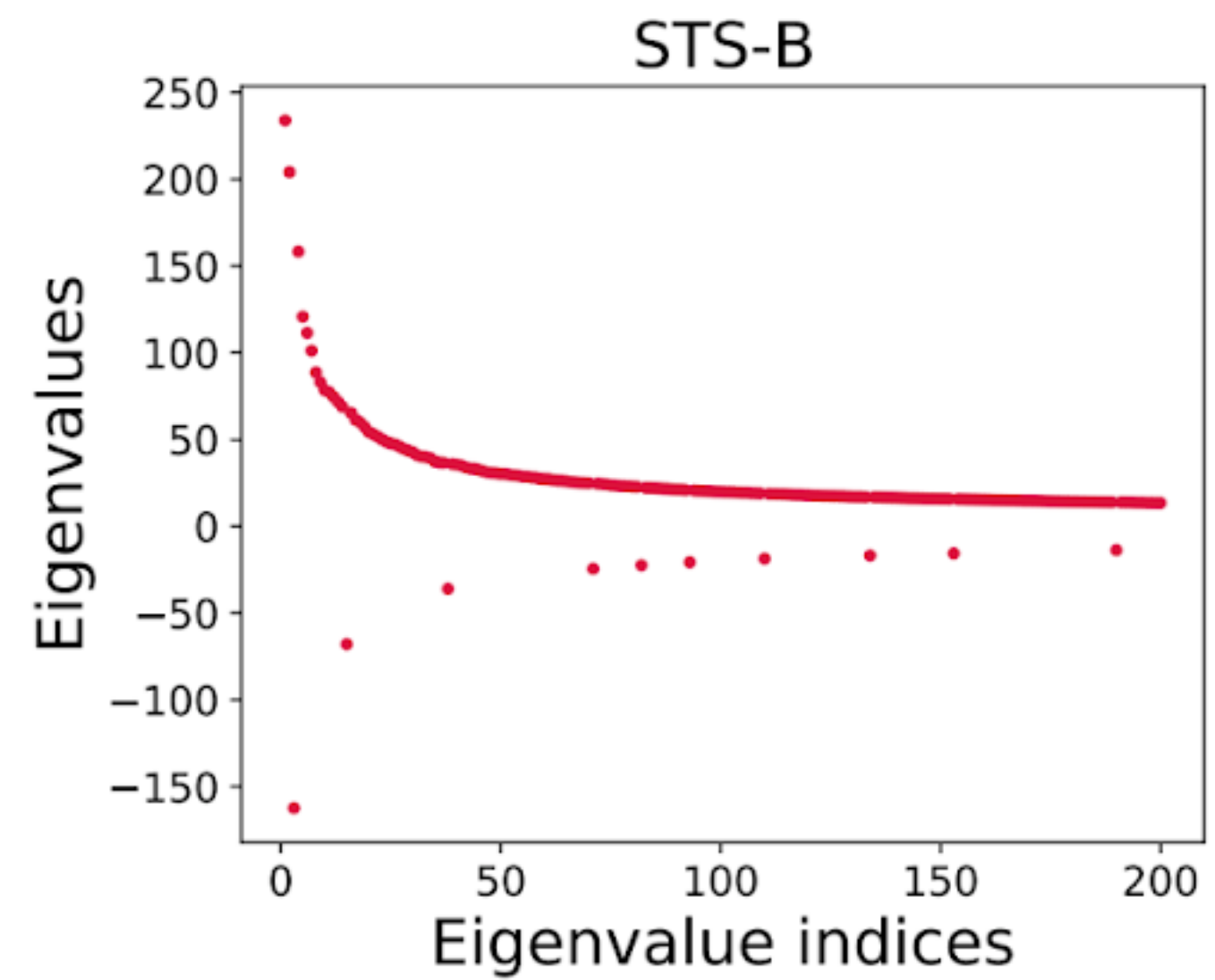
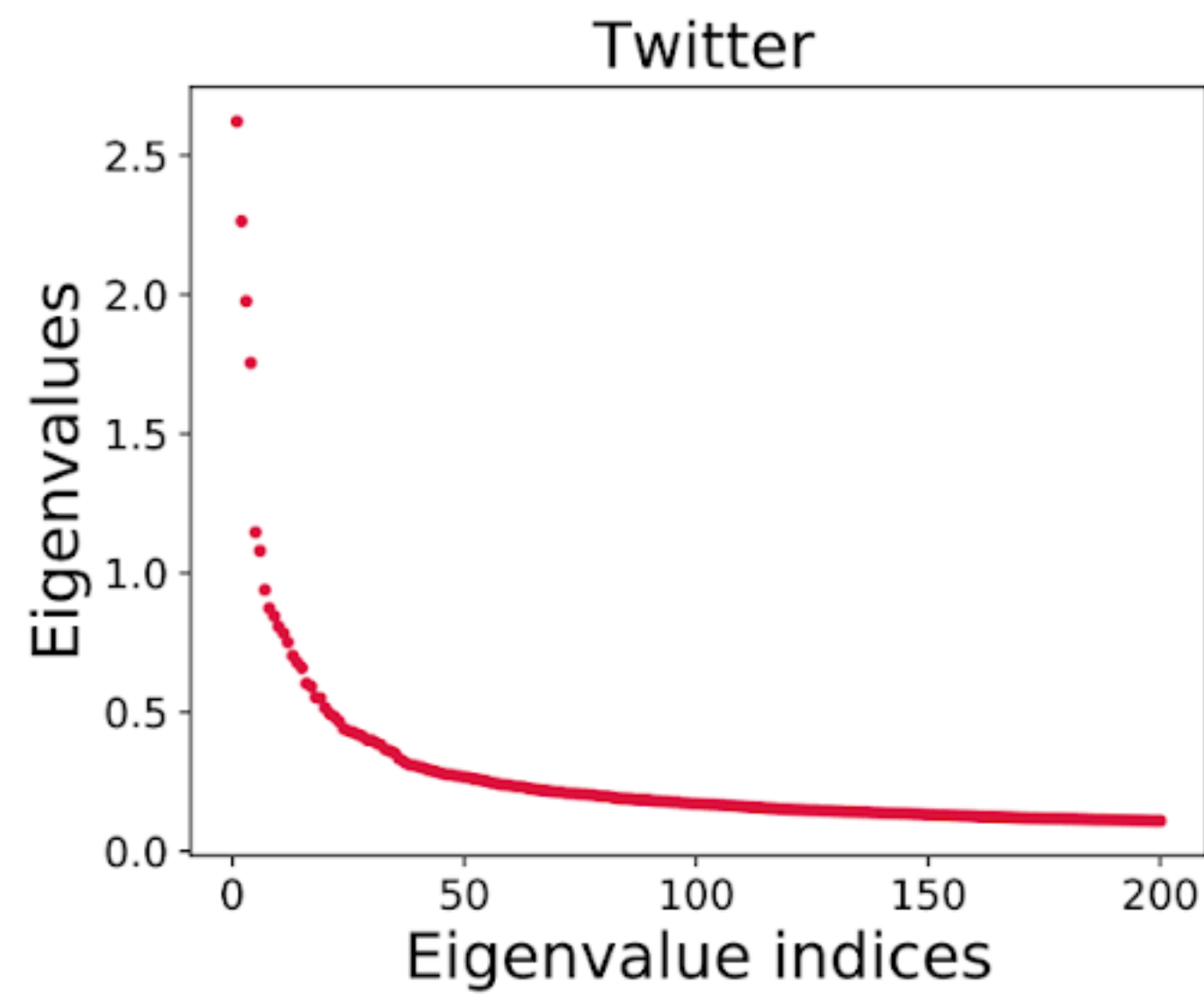
0.0091	0.3020	..	..	..	0.1653	0.2900
0.3020	We'll do okay with <i>approximate</i> similarities					0.1983
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
0.1653	0.4337	..	..	..	0.3769	0.2375
0.2900	0.1983	..	..	..	0.2375	0.8342

0.0091	0.3020	..	..	..	0.1653	0.2900
0.3020	We'll do okay with <i>approximate</i> similarities					0.1983
..	..	..	..	..	..	..
..	We want these methods to be <i>sublinear</i> [ $o(n)$ ]					..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
0.1653	0.4337	..	..	..	0.3769	0.2375
0.2900	0.1983	..	..	..	0.2375	0.8342



0.0091	0.3020	..	..	..	0.1653	0.2900
0.3020	We'll do okay with <i>approximate</i> similarities					0.1983
..	..	..	..	..	..	..
..	We want these methods to be <i>sublinear</i> [ $o(n)$ ]					..
..	..	..	..	..	..	..
..	<i>In NLP</i> most similarity matrices are <i>indefinite</i>					..
..	..	..	..	..	..	..
0.1653	0.4337	..	..	..	0.3769	0.2375
0.2900	0.1983	..	..	..	0.2375	0.8342

***In NLP*** most similarity matrices are ***indefinite***



0.0091	0.3020	..	..	..	0.1653	0.2900
0.3020	We'll do okay with <i>approximate</i> similarities					0.1983
..	..	..	..	..	..	..
..	We want these methods to be <i>sublinear</i> [ $o(n)$ ]					..
..	..	..	..	..	..	..
..	<i>In NLP</i> most similarity matrices are <i>indefinite</i>					..
..	..	..	..	..	..	..
0.1653	Many methods exist for approximating PSD matrices					0.2375
0.2900	0.1983	..	..	..	0.2375	0.8342

0.0091	0.3020	..	..	..	0.1653	0.2900
0.3020	We'll do okay with <i>approximate</i> similarities					0.1983
..	..	..	..	..	..	..
..	We want these methods to be <i>sublinear</i> [ $o(n)$ ]					..
..	..	..	..	..	..	..
..	<i>In NLP</i> most similarity matrices are <i>indefinite</i>					..
..	..	..	..	..	..	..
0.1653	Not many methods for indefinite matrices exist					0.2375
0.2900	0.1983	..	..	..	0.2375	0.8342

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Columns of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Columns of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

Inverse of the principal sub matrix of  $K$

$$\begin{bmatrix} -6.6019 & 3.2457 \\ 3.24573 & -0.0886 \end{bmatrix}$$

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Columns of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

Inverse of the principal sub matrix of  $K$

$$\begin{bmatrix} -6.6019 & 3.2457 \\ 3.24573 & -0.0886 \end{bmatrix}$$

Rows of  $K$

$$\begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}^T$$

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$



Nyström Method:  $\tilde{K} = KS(S^T KS)^+ S^T K$

$$\tilde{K} = \begin{bmatrix} 0.0091 & 0.3020 & \dots & \dots & \dots & 0.1653 & 0.2900 \\ 0.3020 & 0.6571 & \dots & \dots & \dots & 0.4337 & 0.1983 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1653 & 0.4337 & \dots & \dots & \dots & 0.3769 & 0.2375 \\ 0.2900 & 0.1983 & \dots & \dots & \dots & 0.2375 & 0.8342 \end{bmatrix}$$

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Nyström Method:  $\tilde{K} = KS(S^T KS)^+ S^T K$

$$\tilde{K} = \begin{bmatrix} 0.0091 & 0.3020 & \dots & \dots & \dots & 0.1653 & 0.2900 \\ 0.3020 & 0.6571 & \dots & \dots & \dots & 0.4337 & 0.1983 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1653 & 0.4337 & \dots & \dots & \dots & 0.3769 & 0.2375 \\ 0.2900 & 0.1983 & \dots & \dots & \dots & 0.2375 & 0.8342 \end{bmatrix}$$

Benefits: Runs in **sublinear** time with respect to  $K$ .

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Nyström Method:  $\tilde{K} = KS(S^T KS)^+ S^T K$

$$\tilde{K} = \begin{bmatrix} 0.0091 & 0.3020 & \dots & \dots & \dots & 0.1653 & 0.2900 \\ 0.3020 & 0.6571 & \dots & \dots & \dots & 0.4337 & 0.1983 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1653 & 0.4337 & \dots & \dots & \dots & 0.3769 & 0.2375 \\ 0.2900 & 0.1983 & \dots & \dots & \dots & 0.2375 & 0.8342 \end{bmatrix}$$

Benefits: Runs in **sublinear** time with respect to  $K$ .

But becomes **unstable** for **indefinite matrices** unless they are near PSD

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

But becomes *unstable* for *indefinite matrices* unless they are near PSD

Columns of  $K$

Rows of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

Inverse of the principal sub matrix of  $K$

$$\begin{bmatrix} -6.6019 & 3.2457 \\ 3.24573 & -0.0886 \end{bmatrix}$$

$$\begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}^T$$

$(S^T K S)^+$

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Columns of  $K$

Rows of  $K$

Inverse of the principal  
sub matrix of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

$$\begin{bmatrix} -6.6019 & 3.2457 \\ 3.24573 & -0.0886 \end{bmatrix}$$

$$\begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}^T$$

If  $S^T K S$  is ill conditioned,  $(S^T K S)^+$  will fail

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Smaller eigenvalues of  $S^T K S$  can get blown up leading to large error

Columns of  $K$

Rows of  $K$

Inverse of the principal  
sub matrix of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

$$\begin{bmatrix} -6.6019 & 3.2457 \\ 3.24573 & -0.0886 \end{bmatrix}$$

$$\begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}^T$$

If  $S^T K S$  is ill conditioned,  $(S^T K S)^+$  will fail

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Smaller eigenvalues of  $S^T K S$  can get blown up leading to large error

Columns of  $K$

Rows of  $K$

$$\tilde{K} = \begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}$$

Inverse of the principal sub matrix of  $K$

$$\begin{bmatrix} -6.6019 & 3.2457 \\ 3.24573 & -0.0886 \end{bmatrix}$$

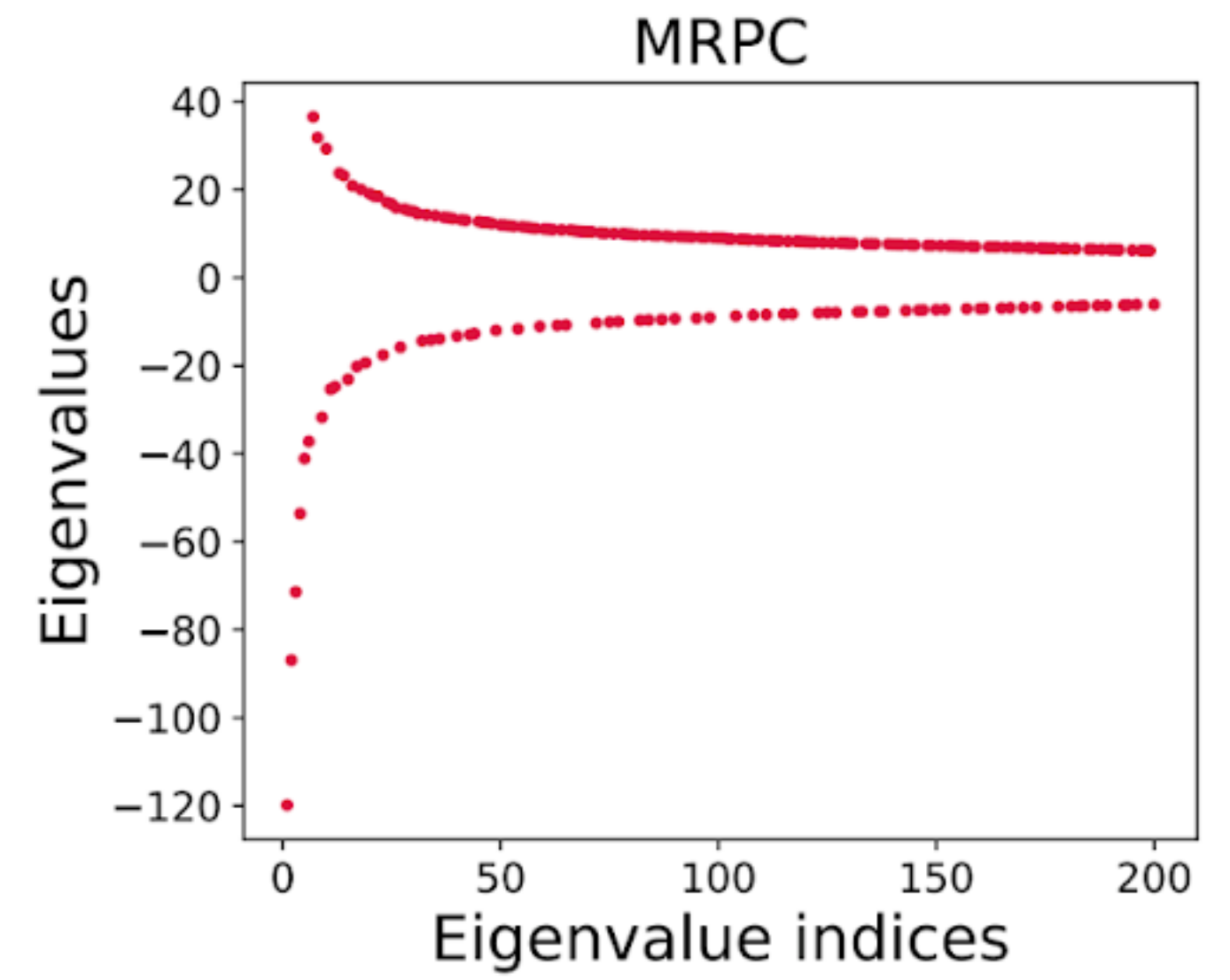
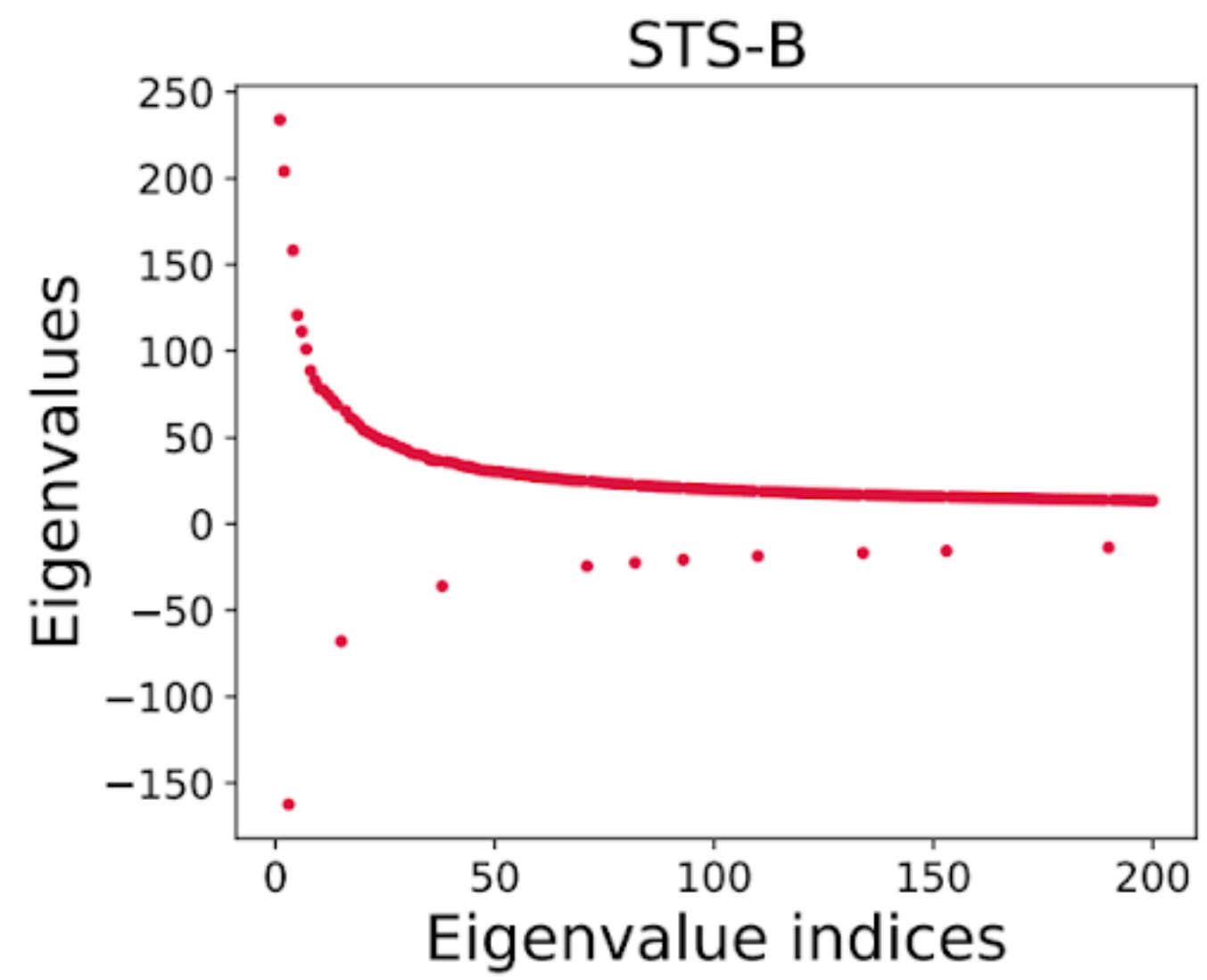
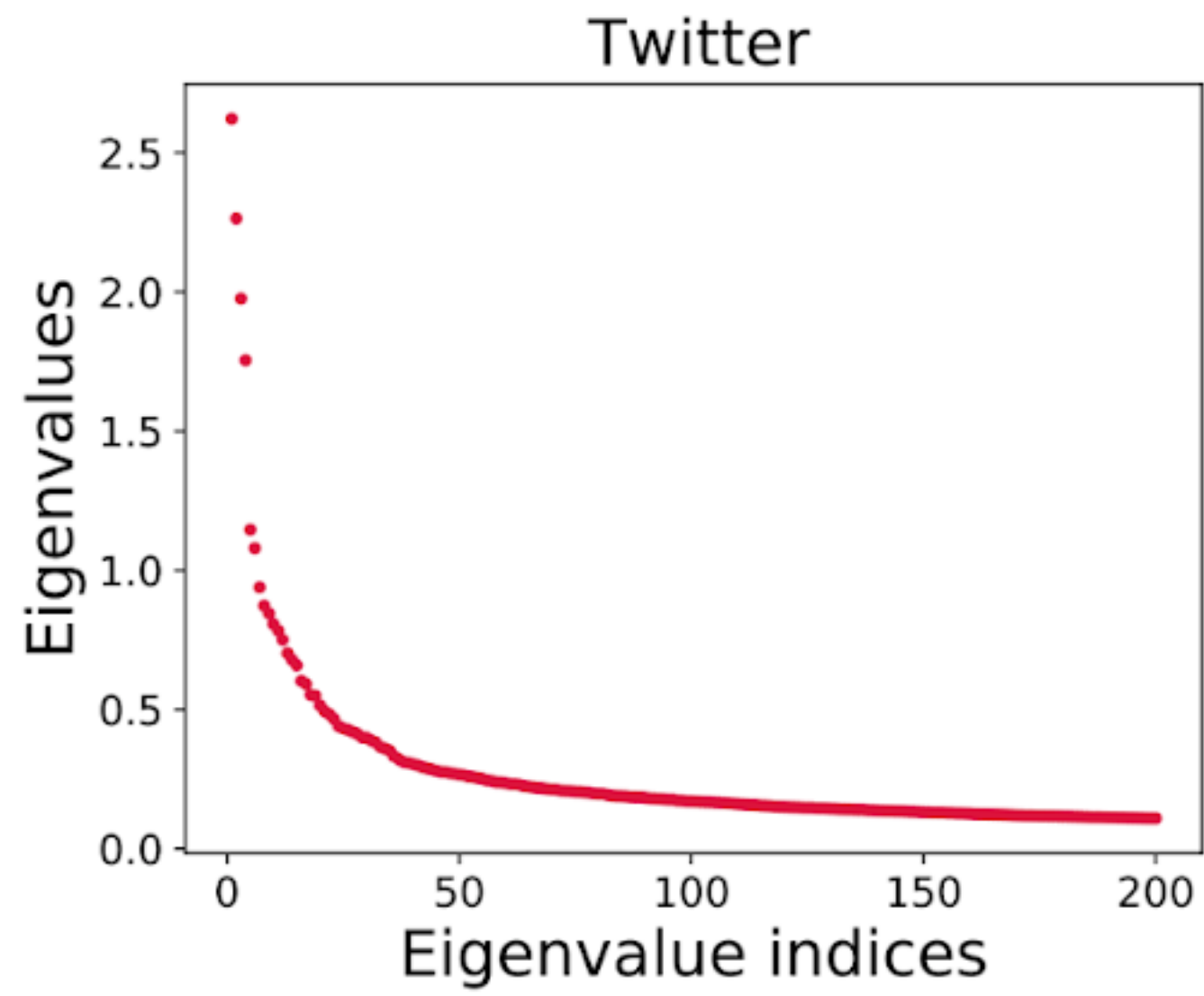
$$\begin{bmatrix} 0.0089 & 0.3262 \\ 0.3262 & 0.6635 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.1820 & 0.4246 \\ 0.2993 & 0.1990 \end{bmatrix}^T$$

If  $S^T K S$  is ill conditioned,  $(S^T K S)^+$  will fail

Endemic to indefinite matrices

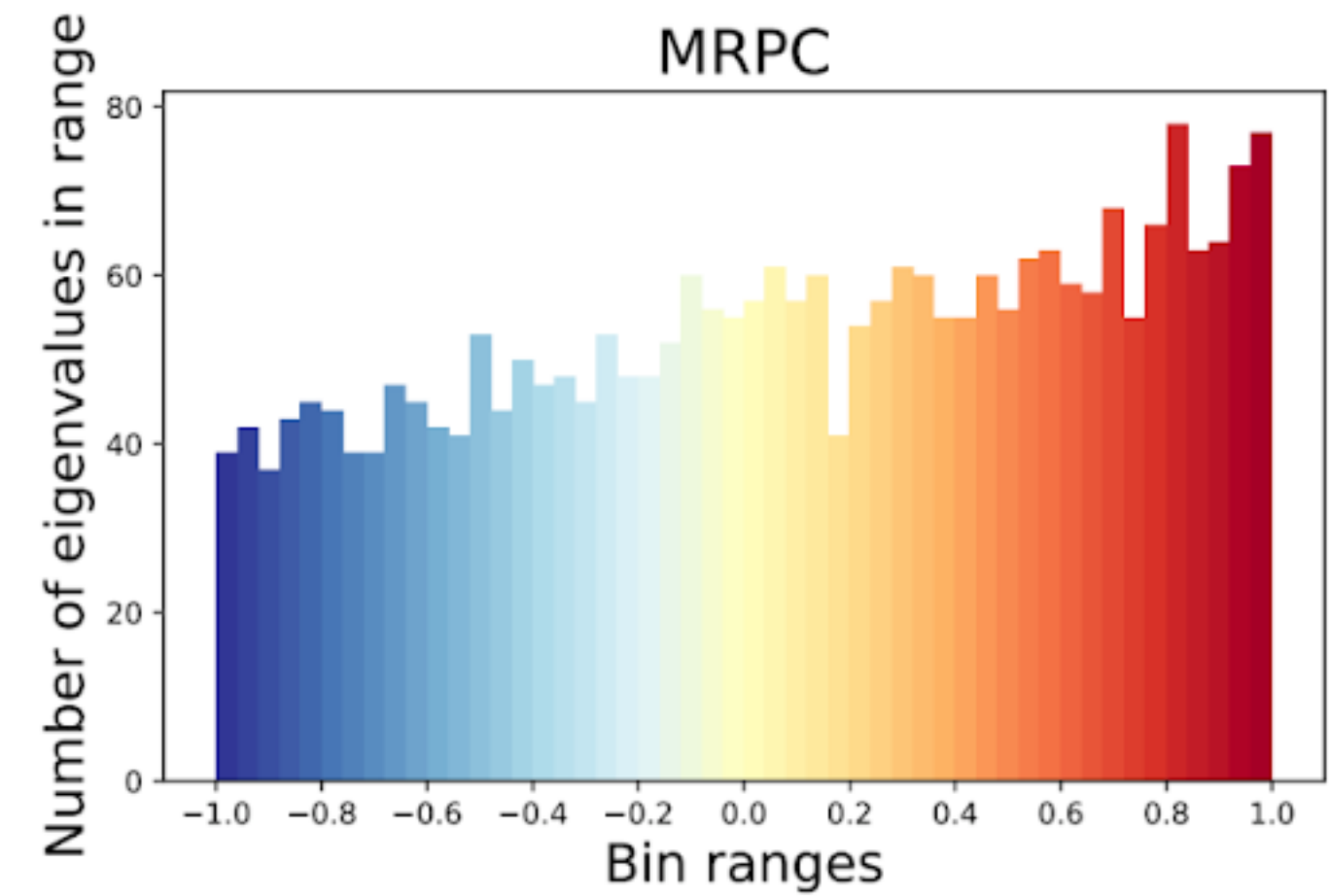
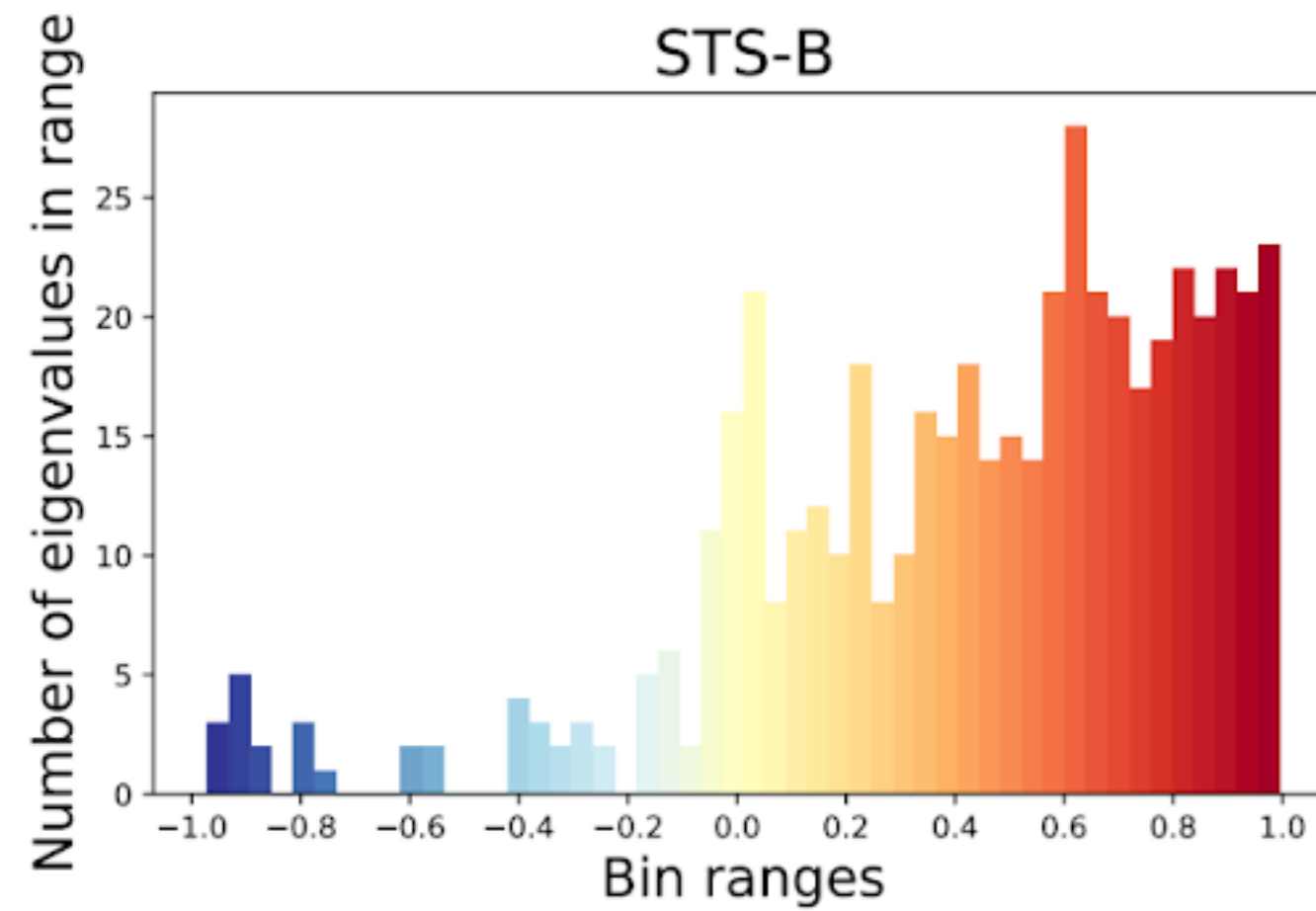
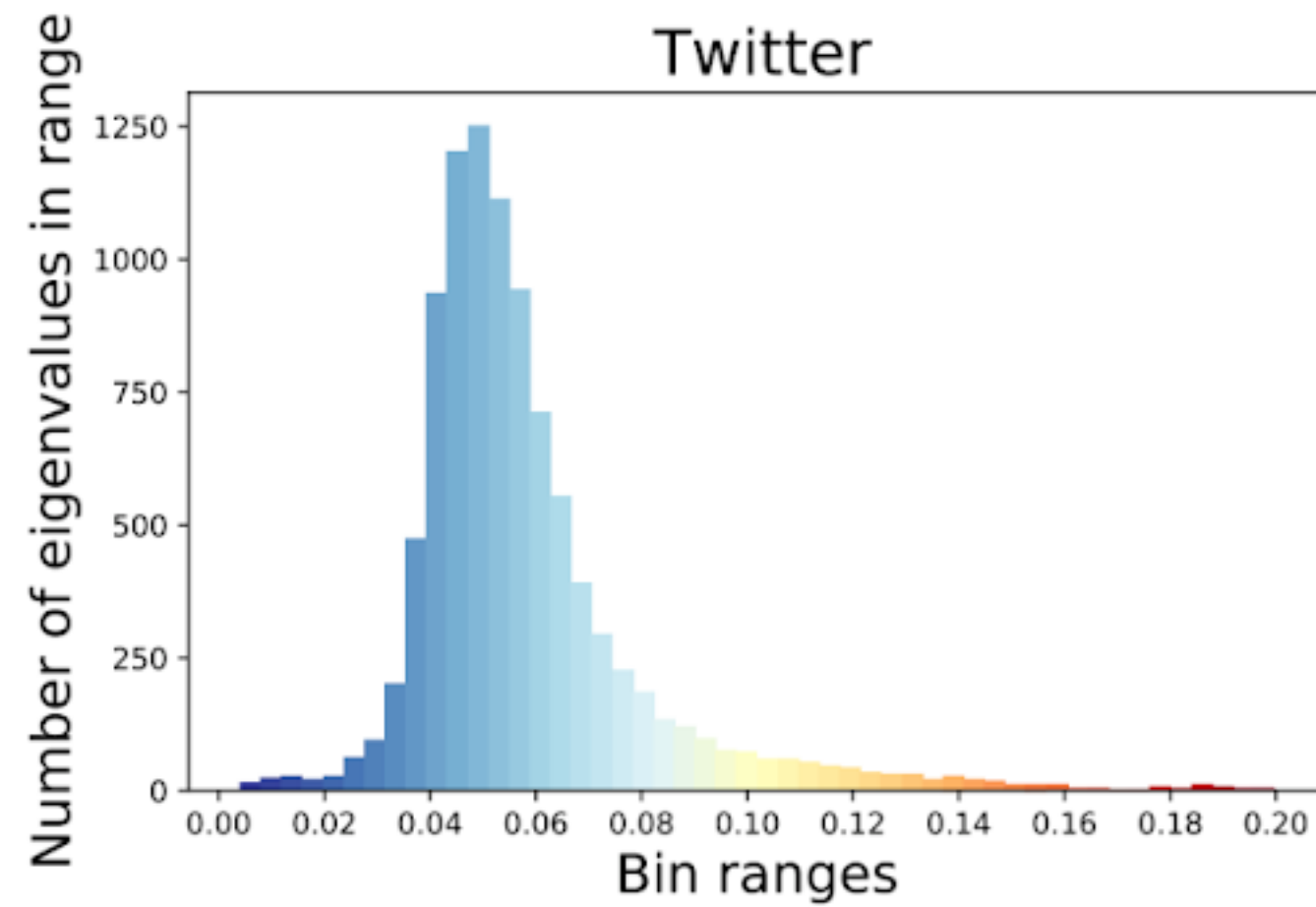
$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

Endemic to indefinite matrices





Endemic to indefinite matrices



Independently sample  $S^T K S$ , plot histogram of the eigenvalues

# Sublinear methods for approximating indefinite similarity matrices

**SMS-Nyström**

**Other Approaches**

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^TKS)^+S^TK$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

Sample  $S_1$  and  $S_2$  independently at random

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

Sample  $S_1$  and  $S_2$  independently at random

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

# Sublinear methods for approximating indefinite similarity matrices

$O(n^3)$  computations for  $\lambda_{\min}(K)$

If  $|\lambda_{\min}(K)|$  is large, **bad approximation**

**SMS-Nyström**

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$



# Sublinear methods for approximating indefinite similarity matrices

$O(n^3)$  computations for  $\lambda_{\min}(K)$

If  $|\lambda_{\min}(K)|$  is large, **bad approximation**

Instead compute  $\lambda_{\min}(S_2^T K S_2)$ , where  $|S_2| = 2|S|$

**SMS-Nyström**

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

# Sublinear methods for approximating indefinite similarity matrices

$O(n^3)$  computations for  $\lambda_{\min}(K)$

If  $|\lambda_{\min}(K)|$  is large, **bad approximation**

Instead compute  $\lambda_{\min}(S_2^T K S_2)$ , where  $|S_2| = 2|S|$

Shift  $S^T K S$  using  $\lambda_{\min}(S_2^T K S_2)$

**SMS-Nyström**

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

# Sublinear methods for approximating indefinite similarity matrices

$O(n^3)$  computations for  $\lambda_{\min}(K)$

If  $|\lambda_{\min}(K)|$  is large, **bad approximation**

Instead compute  $\lambda_{\min}(S_2^T K S_2)$ , where  $|S_2| = 2|S|$

Shift  $S^T K S$  using  $\lambda_{\min}(S_2^T K S_2)$

$S^T K S$  is guaranteed to be PSD

**SMS-Nyström**

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

# Sublinear methods for approximating indefinite similarity matrices

$O(n^3)$  computations for  $\lambda_{\min}(K)$

If  $|\lambda_{\min}(K)|$  is large, **bad approximation**

Instead compute  $\lambda_{\min}(S_2^T K S_2)$ , where  $|S_2| = 2|S|$

Shift  $S^T K S$  using  $\lambda_{\min}(S_2^T K S_2)$

$S^T K S$  is guaranteed to be PSD

Shift is large, so no eigenvalues near 0

**SMS-Nyström**

$K =$

0.0089	0.3262	..	..	..	0.1820	0.2993
0.3262	0.6635	..	..	..	0.4246	0.1990
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
0.1820	0.4246	..	..	..	0.3865	0.2350
0.2993	0.1990	..	..	..	0.2350	0.8414

# Sublinear methods for approximating indefinite similarity matrices

Overall computation overheads

$(|S_2| - |S|)^2$  similarity computations

$O(|S_2|^3)$  computation for  $\lambda_{\min}(S_2^T K S_2)$

**SMS-Nyström**

$$K = \begin{bmatrix} 0.0089 & 0.3262 & \dots & \dots & \dots & 0.1820 & 0.2993 \\ 0.3262 & 0.6635 & \dots & \dots & \dots & 0.4246 & 0.1990 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.1820 & 0.4246 & \dots & \dots & \dots & 0.3865 & 0.2350 \\ 0.2993 & 0.1990 & \dots & \dots & \dots & 0.2350 & 0.8414 \end{bmatrix}$$

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

Very similar to Nyström Approximation, rows and columns sampled independently

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

$|S_2| > |S_1|$  works amazingly! In our experiments  $|S_2| = 2|S_1|$  works reliably

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

$(S^T KS)^+$  is replaced with  $(KSS^T K)^+ S^T KS$



# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

Benefits: No parameters to tune!

# Sublinear methods for approximating indefinite similarity matrices

## SMS-Nyström

$$K = KS(S^T KS)^+ S^T K$$

$$\bar{K} = K - \lambda_{\min}(K)I_n$$

If  $\lambda_{\min}(K)$  is **small**, we can apply Nyström approximation

## Other Approaches

Skeleton approximation:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

SiCUR:

$$\tilde{K} = KS_2(S_2^T KS_1)^+ S_1^T K$$

$$|S_2| = 2|S_1|$$

StaCUR:

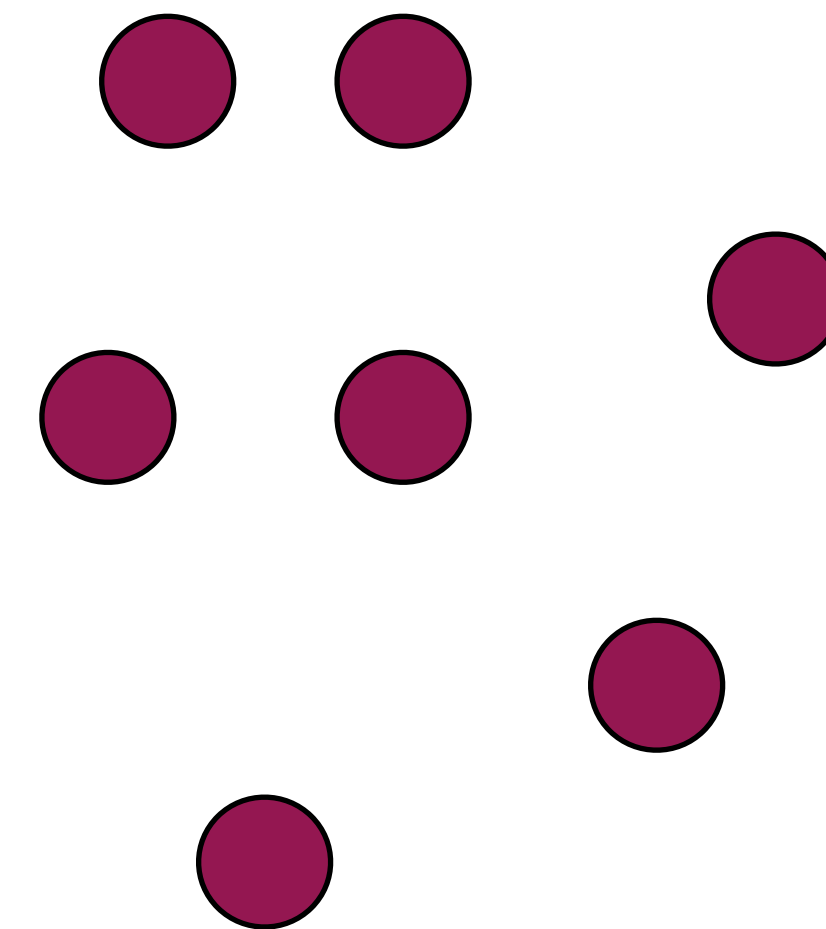
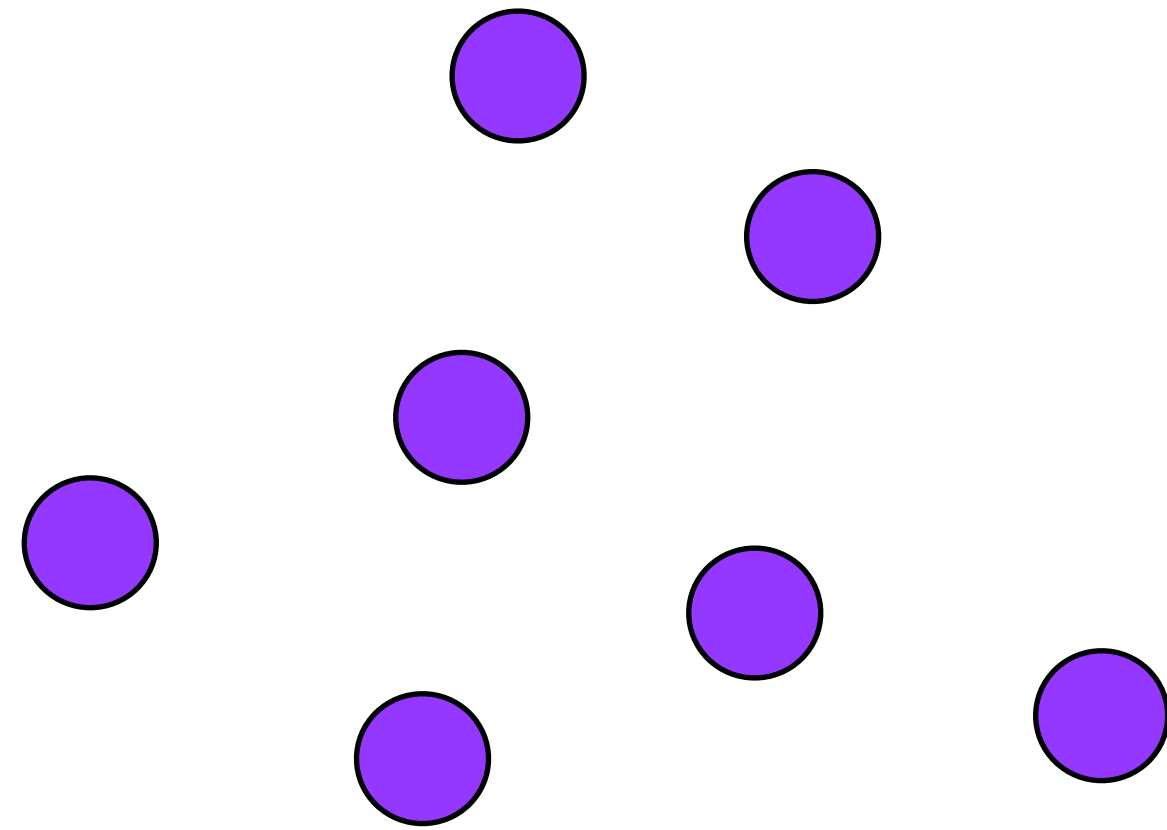
$$\tilde{K} = \frac{n}{s} KS(KSS^T K)^+ S^T KSS^T K$$

Benefits: No two stage sampling!



# Empirical evaluation

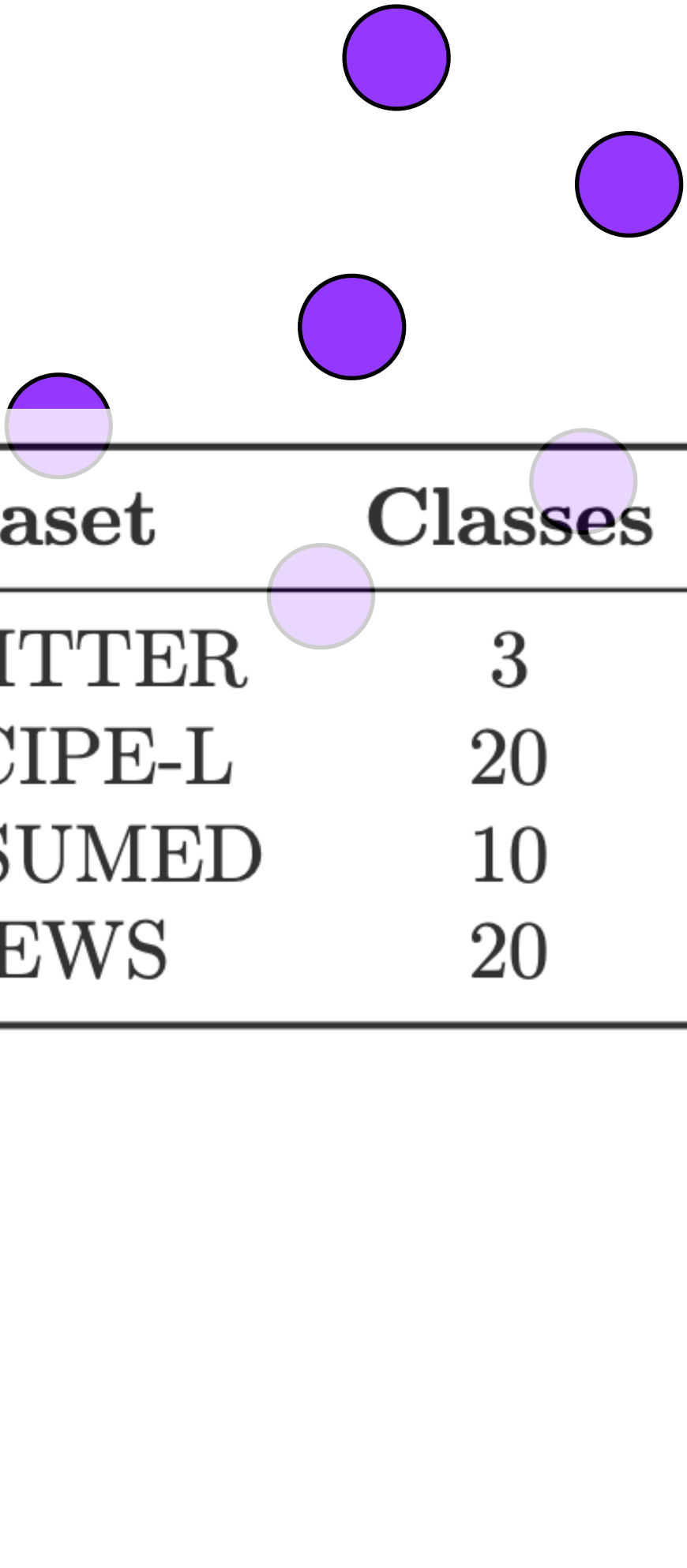
$\mathbb{R}^L$



Consider the task of document classification

# Empirical evaluation

$\mathbb{R}^L$



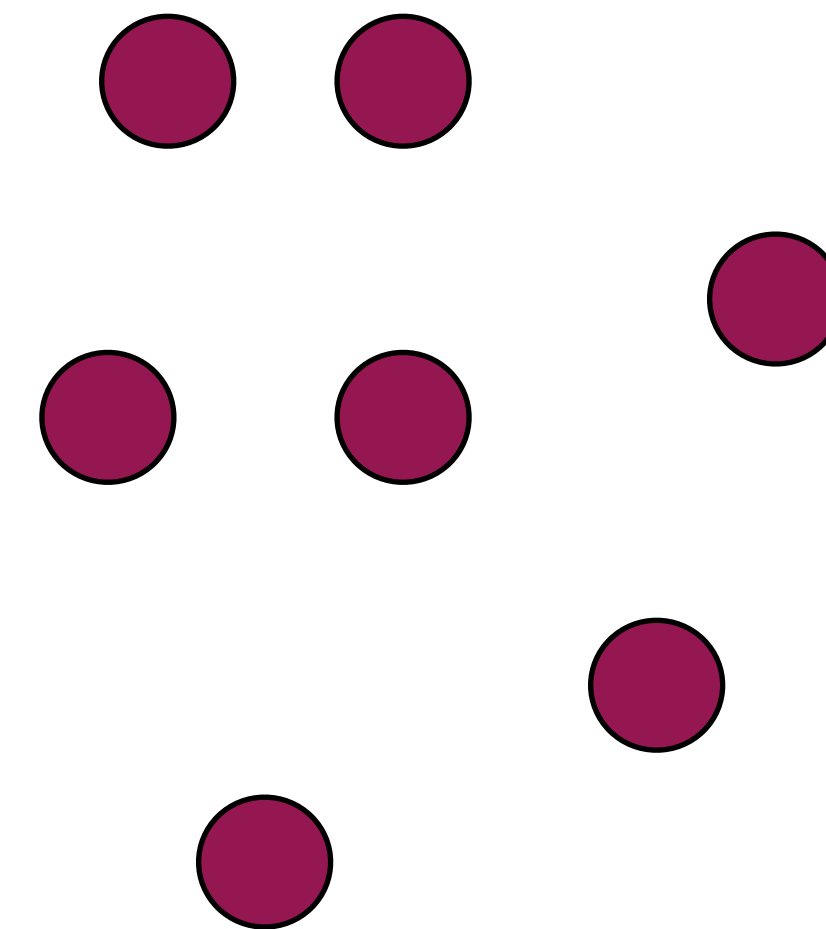
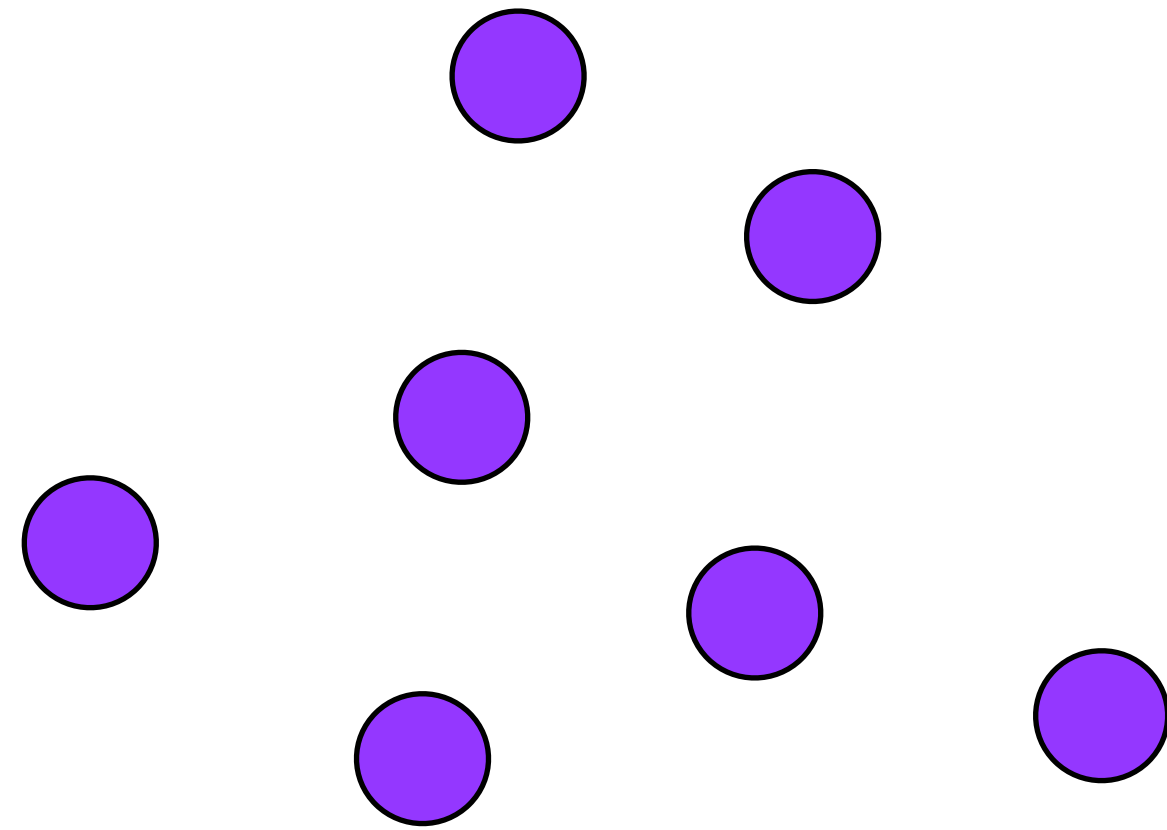
Dataset	Classes	Train	Test	BOW Dim	Length	Application
TWITTER	3	2176	932	6344	9.9	Tweets categorized by sentiment
RECIPE-L	20	27841	11933	3590	18.5	Recipe procedures labeled by origin
OHSUMED	10	3999	5153	31789	59.2	medical Abstracts (class subsampled)
20NEWS	20	11293	7528	29671	72	Canonical User-written posts dataset

Consider the task of document classification

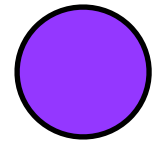


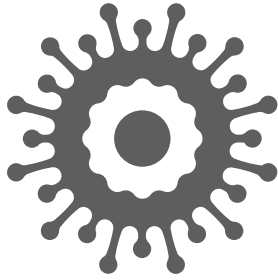
# Empirical evaluation

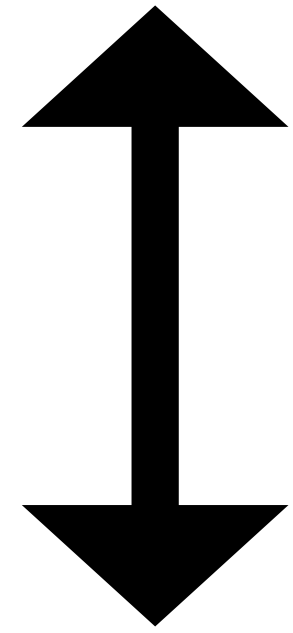
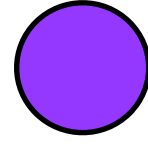
$\mathbb{R}^L$

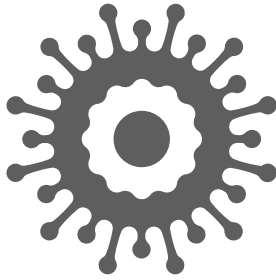


We can check which documents belongs to same class by just checking how well they align with each other

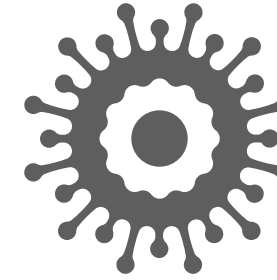


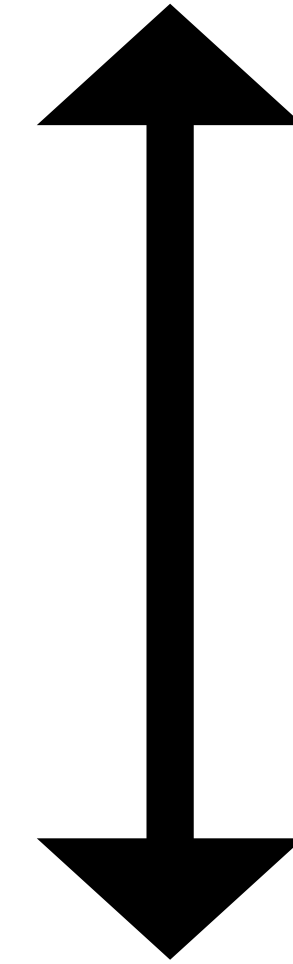
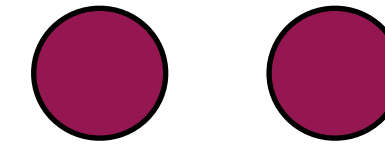
 Amazingly, it is effective against SARS and **MERS**.

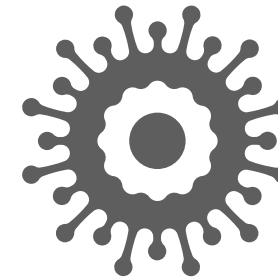


 The Middle East respiratory syndrome coronavirus (MERS-CoV) is an emerging pathogen...

*RL*

 **Pandemic arboviruses** have emerged as a major global health problem in the past four decades.



 The arboviral infection, **CTF**, is transmitted from the bite of an infected wood tick.

$\mathbb{R}^d$

MERS  
↑  
MERS-COV

Pandemic  
↑  
Infection

Arboviruses  
↑  
Arboviral

Word embedding space (here: word2vec)



Compute cumulative distance between all words for each pair of documents

$\mathbb{R}^d$

MERS  
↑  
MERS-COV

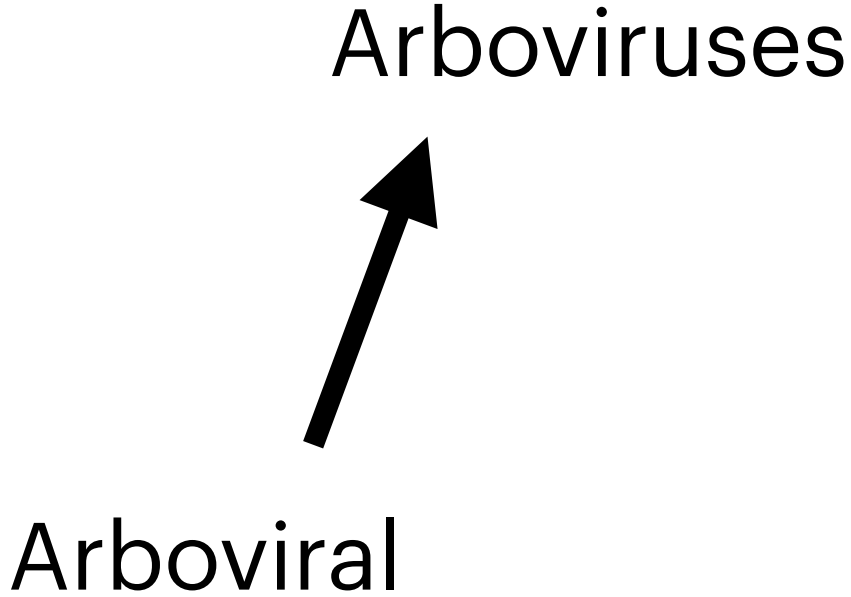
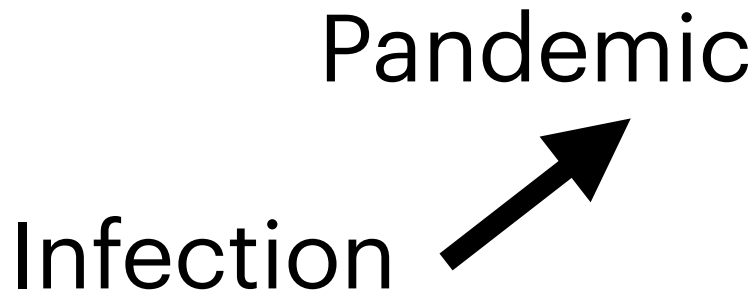
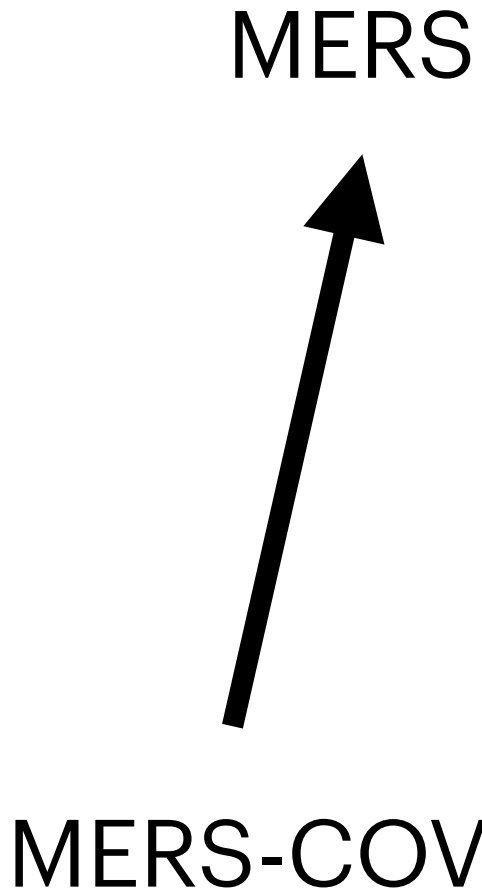
Pandemic  
↑  
Infection

Arboviruses  
↑  
Arboviral

Word embedding space (here: word2vec)

Total less distance travelled implies nearness in  $\mathbb{R}^L$

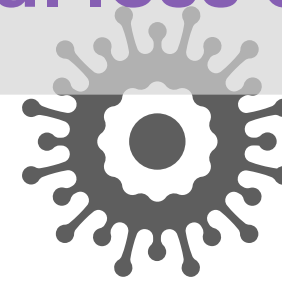
$\mathbb{R}^d$

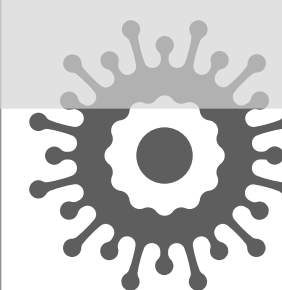


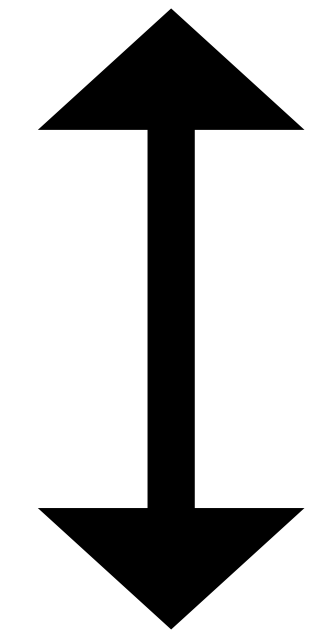
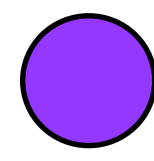
Word embedding space (here: word2vec)

$\mathbb{R}^L$

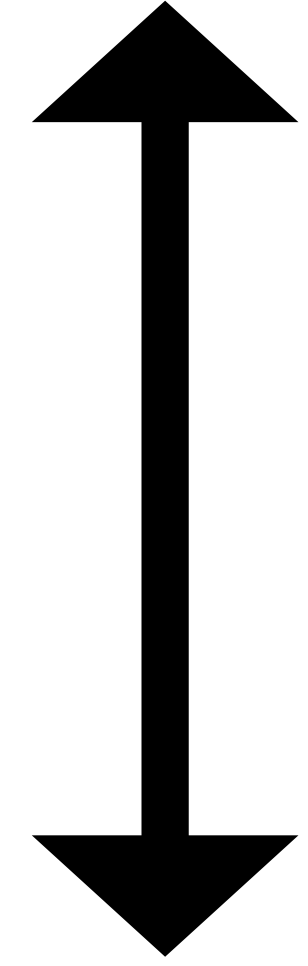
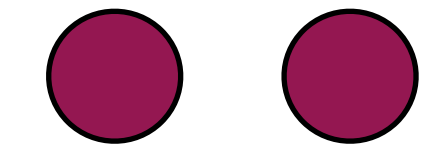
Total less distance travelled implies nearness in  $\mathbb{R}^L$

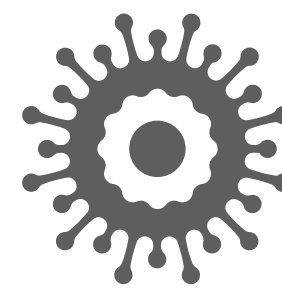
 Amazingly, it is effective against SARS and **MERS**.

 **Pandemic arboviruses** have emerged as a major global health problem in the past four decades.



 The Middle East respiratory syndrome coronavirus (MERS-CoV) is an emerging pathogen...



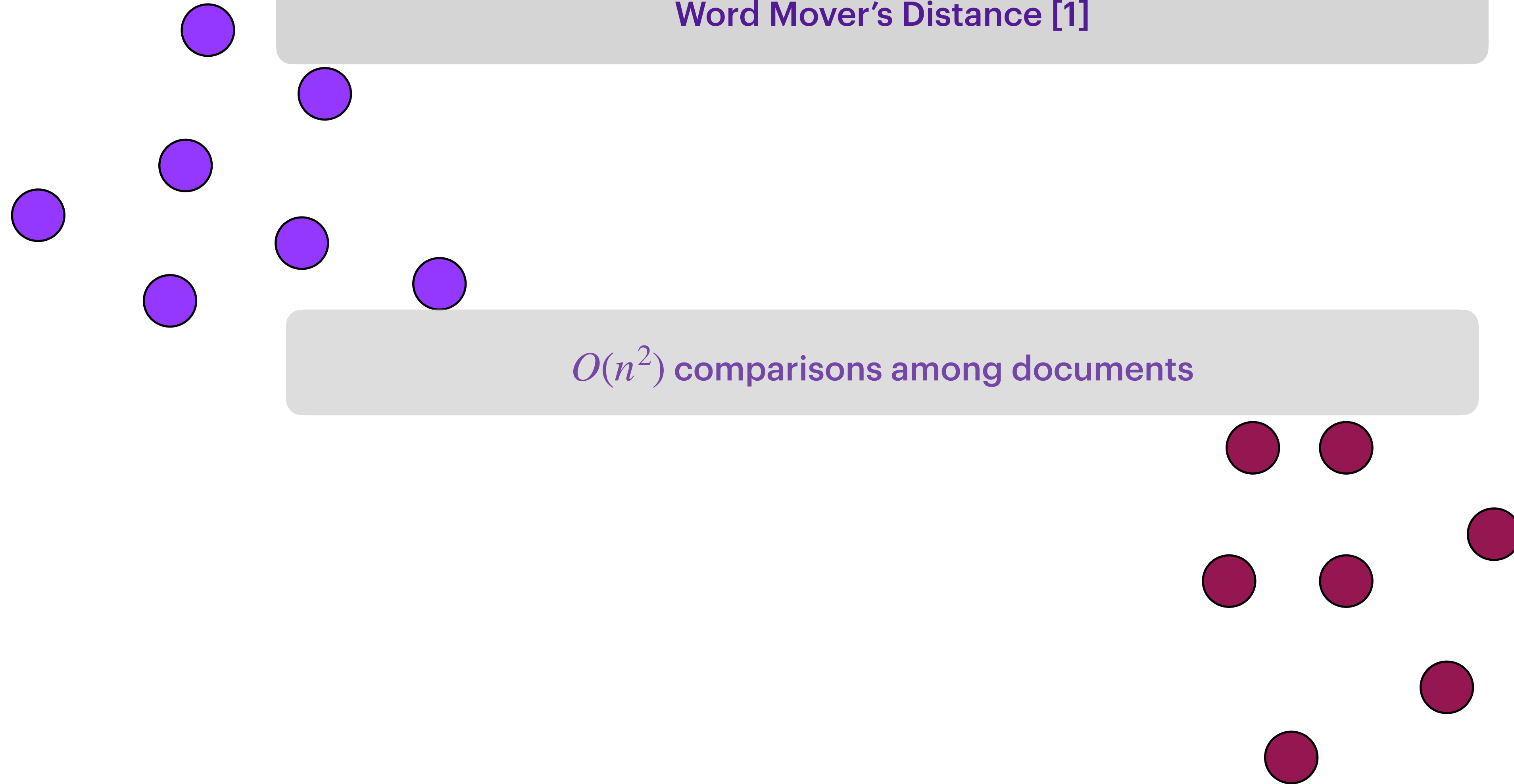
 The arboviral infection, **CTF**, is transmitted from the bite of an infected wood tick.

Word Mover's Distance [1]

[1] Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In *International conference on machine learning* (pp. 957-966). PMLR.

## Word Mover's Distance [1]

$O(n^2)$  comparisons among documents



[1] Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In *International conference on machine learning* (pp. 957-966). PMLR.

## Word Mover's Distance [1]

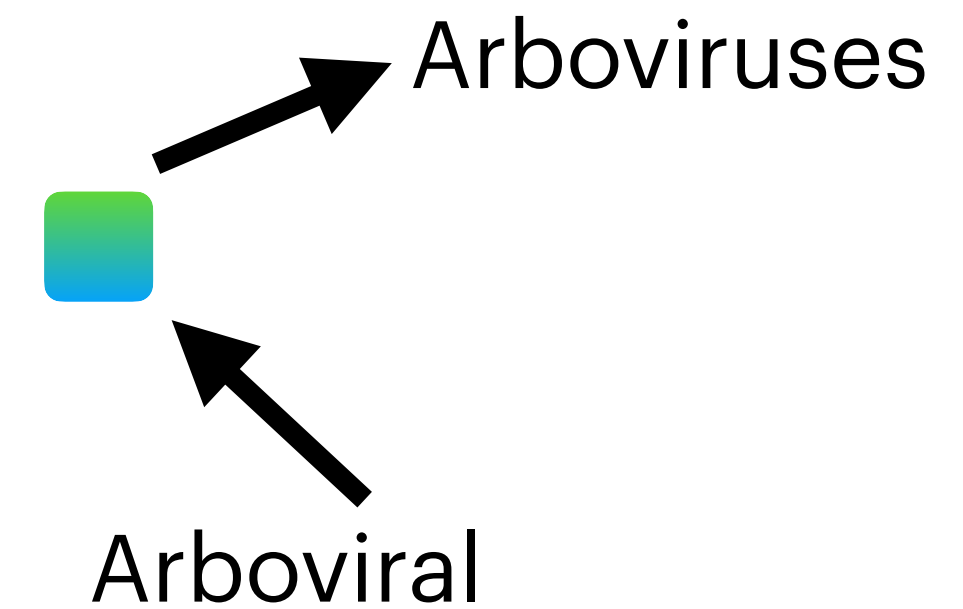
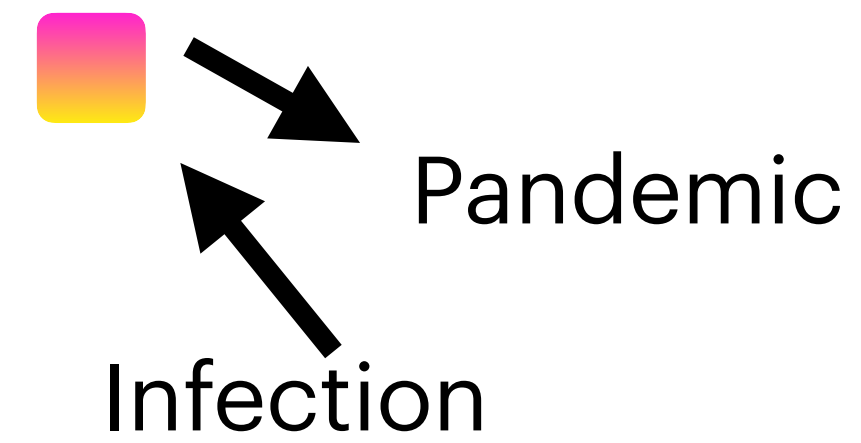
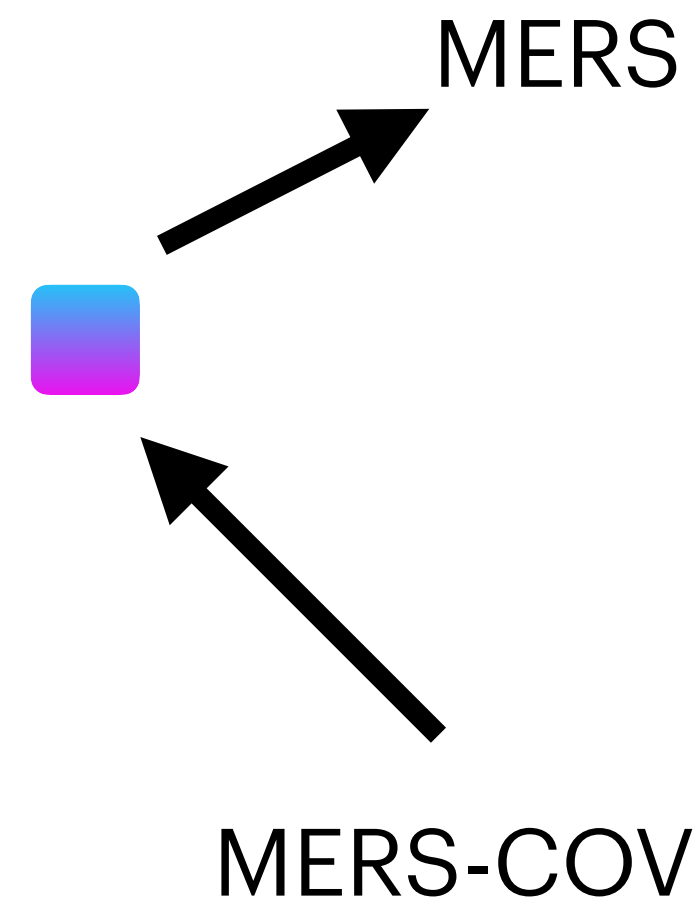
$O(n^2)$  comparisons among documents

$O(L^3 \log L)$  comparisons between documents

[1] Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In *International conference on machine learning* (pp. 957-966). PMLR.

# Word Mover's Embedding [1]

$\mathbb{R}^d$

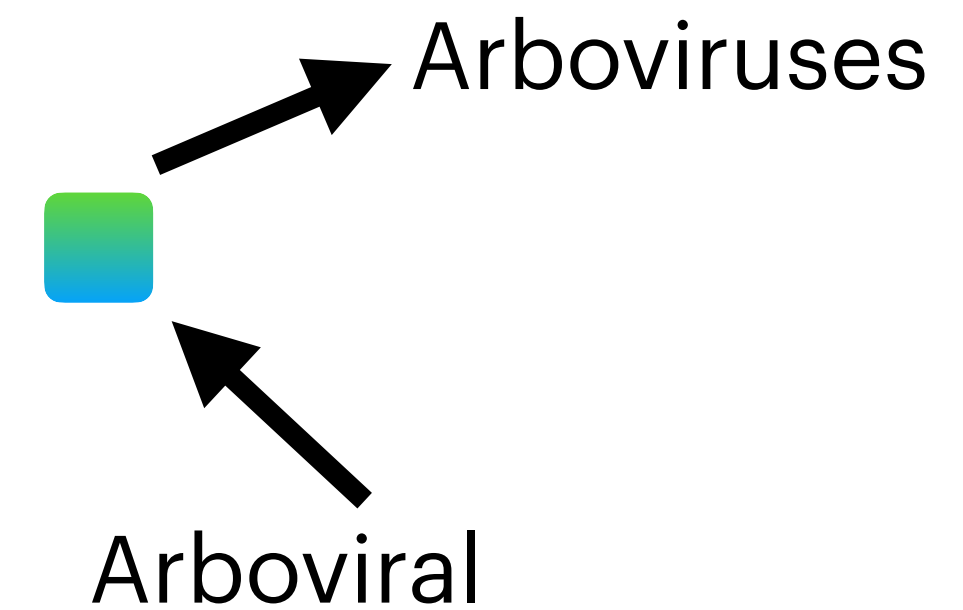
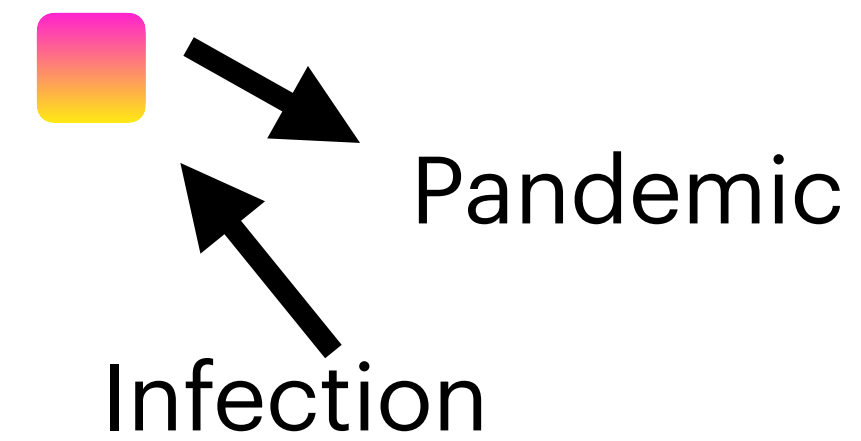
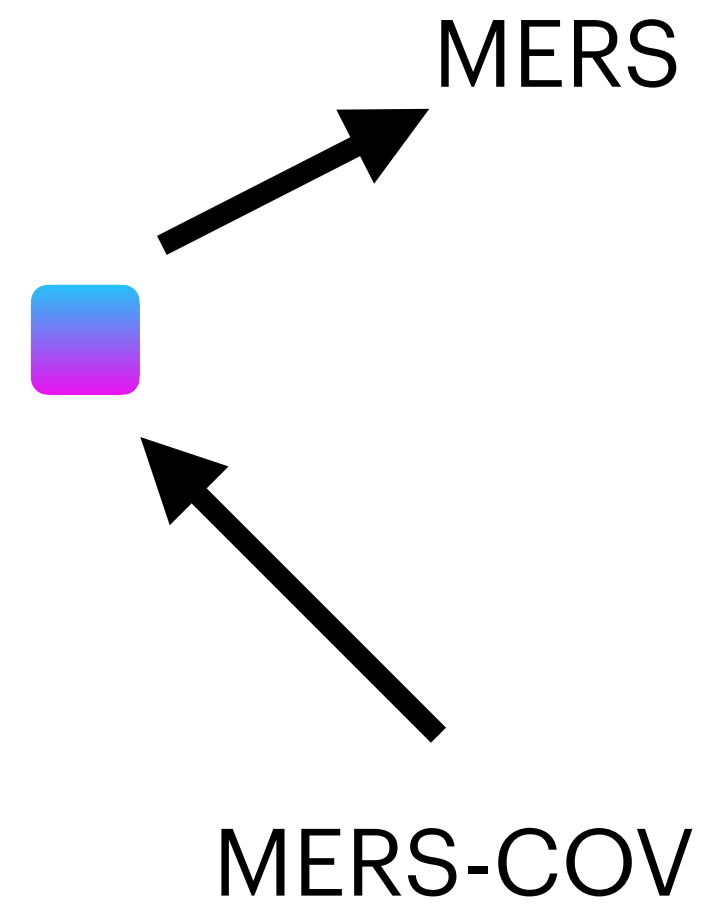


are words from random documents

[1] Wu, L., Yen, I.E., Xu, K., Xu, F., Balakrishnan, A., Chen, P.Y., Ravikumar, P. and Witbrock, M.J., 2018. Word mover's embedding: From word2vec to document embedding. *arXiv preprint arXiv:1811.01713*.

# Word Mover's Embedding [1]

$\mathbb{R}^d$

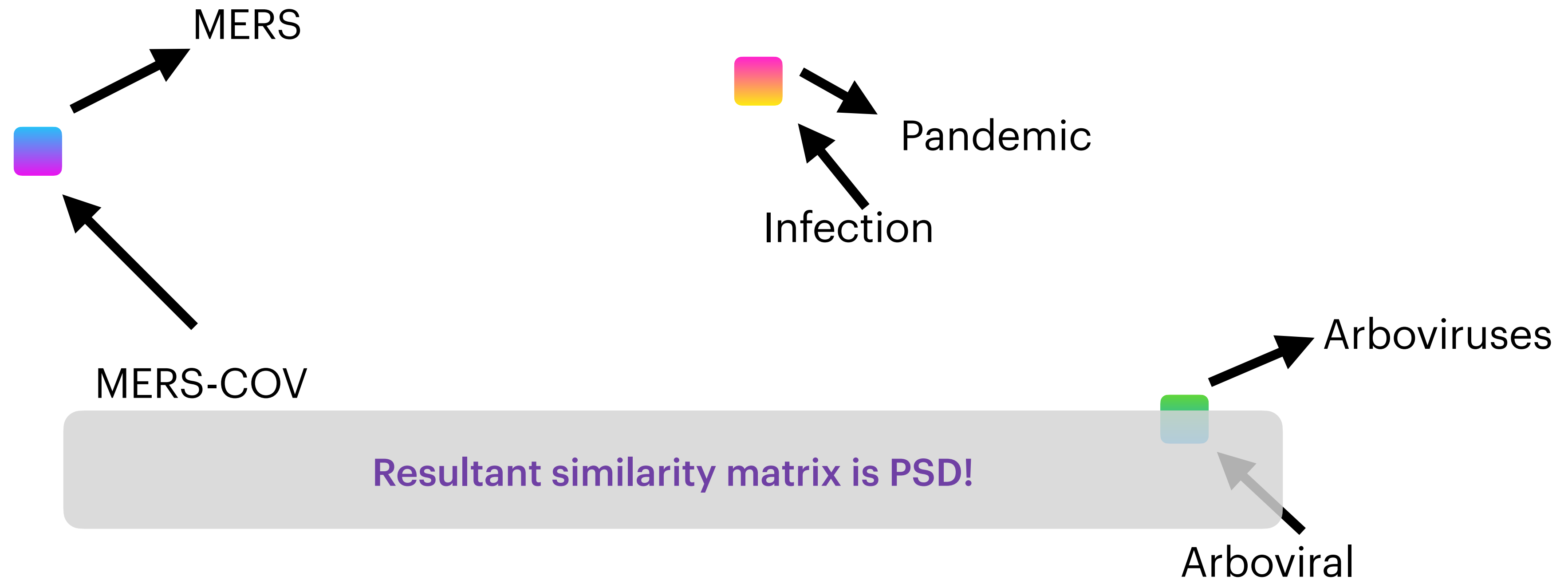


Construct Word Mover's Kernel (WMK) using infinite dimensional feature map to random documents from a given distribution

[1] Wu, L., Yen, I.E., Xu, K., Xu, F., Balakrishnan, A., Chen, P.Y., Ravikumar, P. and Witbrock, M.J., 2018. Word mover's embedding: From word2vec to document embedding. *arXiv preprint arXiv:1811.01713*.

# Word Mover's Embedding [1]

$\mathbb{R}^d$



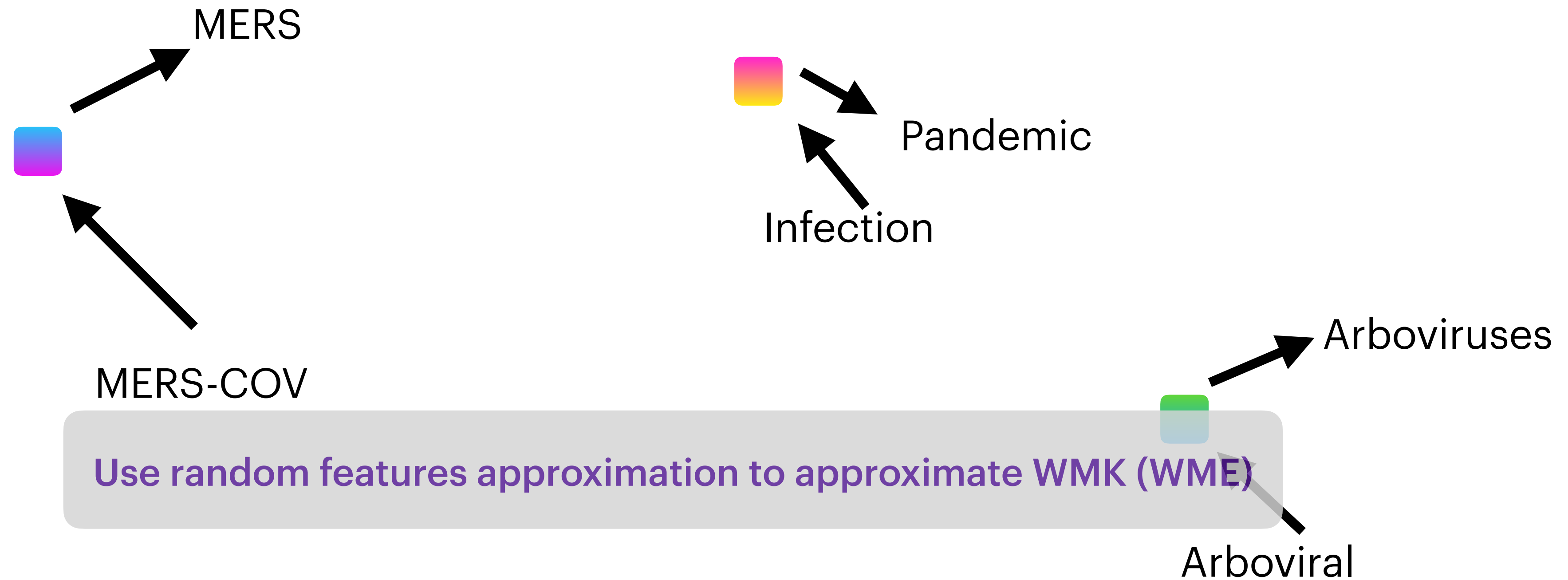
Construct Word Mover's Kernel (WMK) using infinite dimensional feature map to random documents from a given distribution

[1] Wu, L., Yen, I.E., Xu, K., Xu, F., Balakrishnan, A., Chen, P.Y., Ravikumar, P. and Witbrock, M.J., 2018. Word mover's embedding: From word2vec to document embedding. *arXiv preprint arXiv:1811.01713*.



# Word Mover's Embedding [1]

$\mathbb{R}^d$

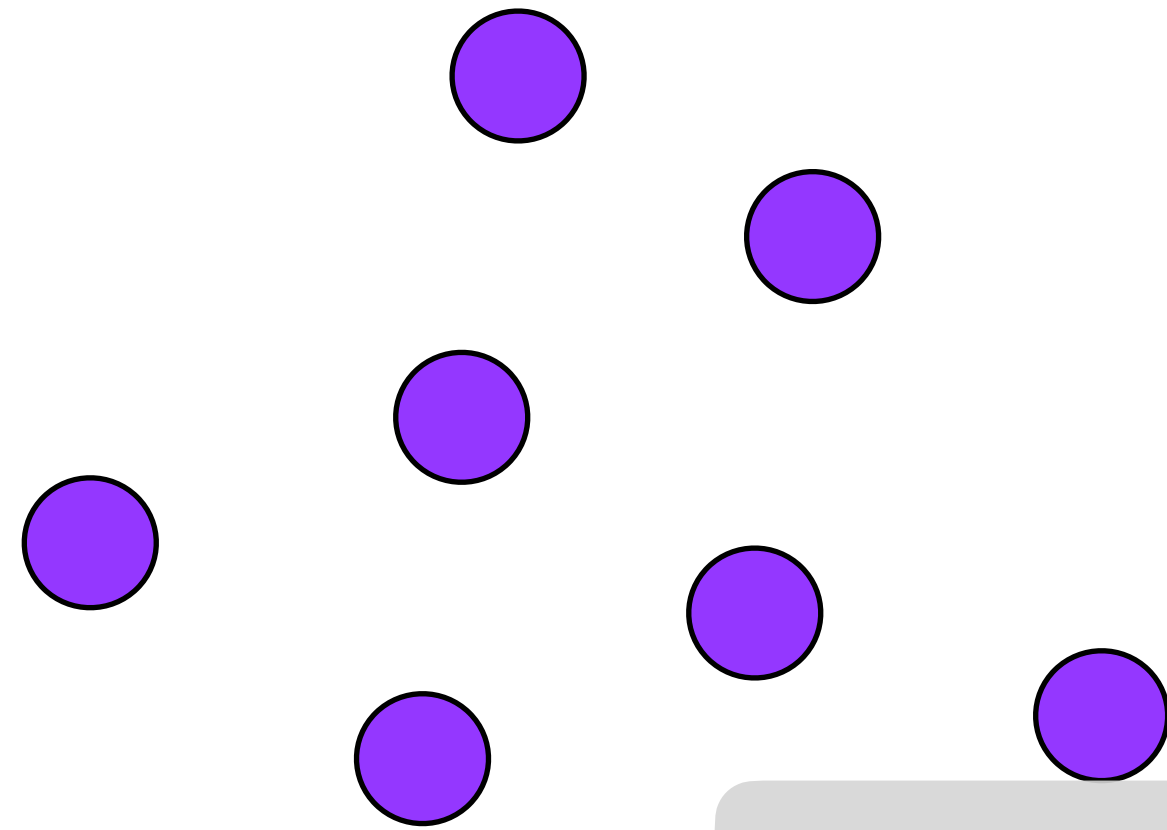


Construct Word Mover's Kernel (WMK) using infinite dimensional feature map to random documents from a given distribution

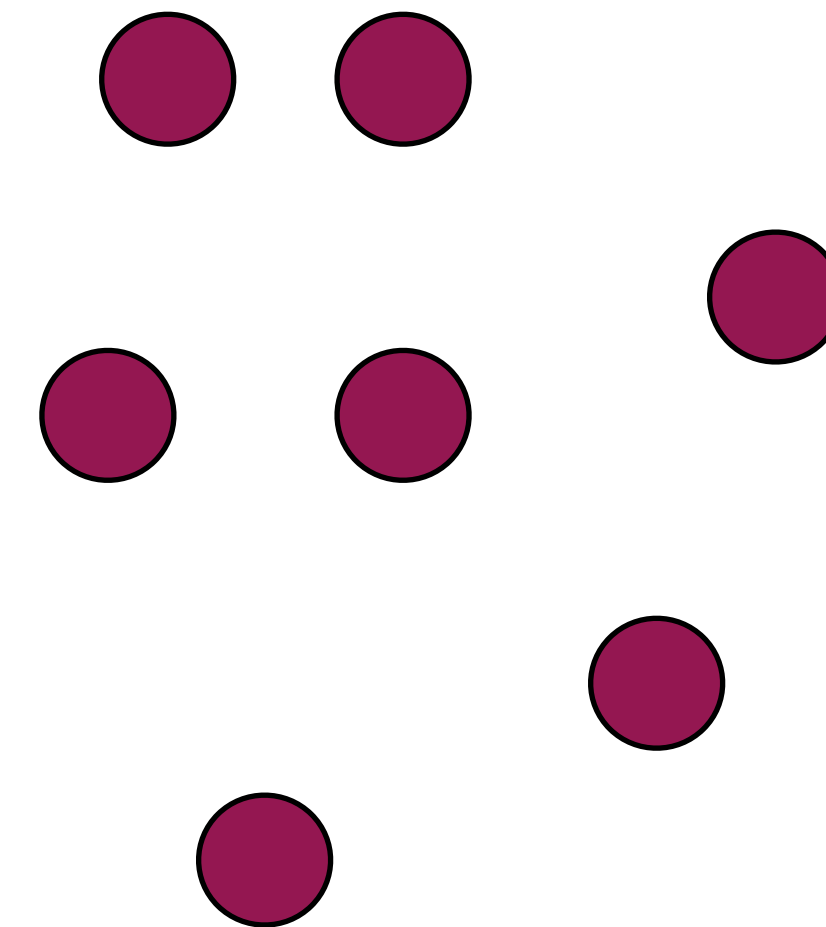
[1] Wu, L., Yen, I.E., Xu, K., Xu, F., Balakrishnan, A., Chen, P.Y., Ravikumar, P. and Witbrock, M.J., 2018. Word mover's embedding: From word2vec to document embedding. *arXiv preprint arXiv:1811.01713*.

# Our Approach

$\mathbb{R}^L$

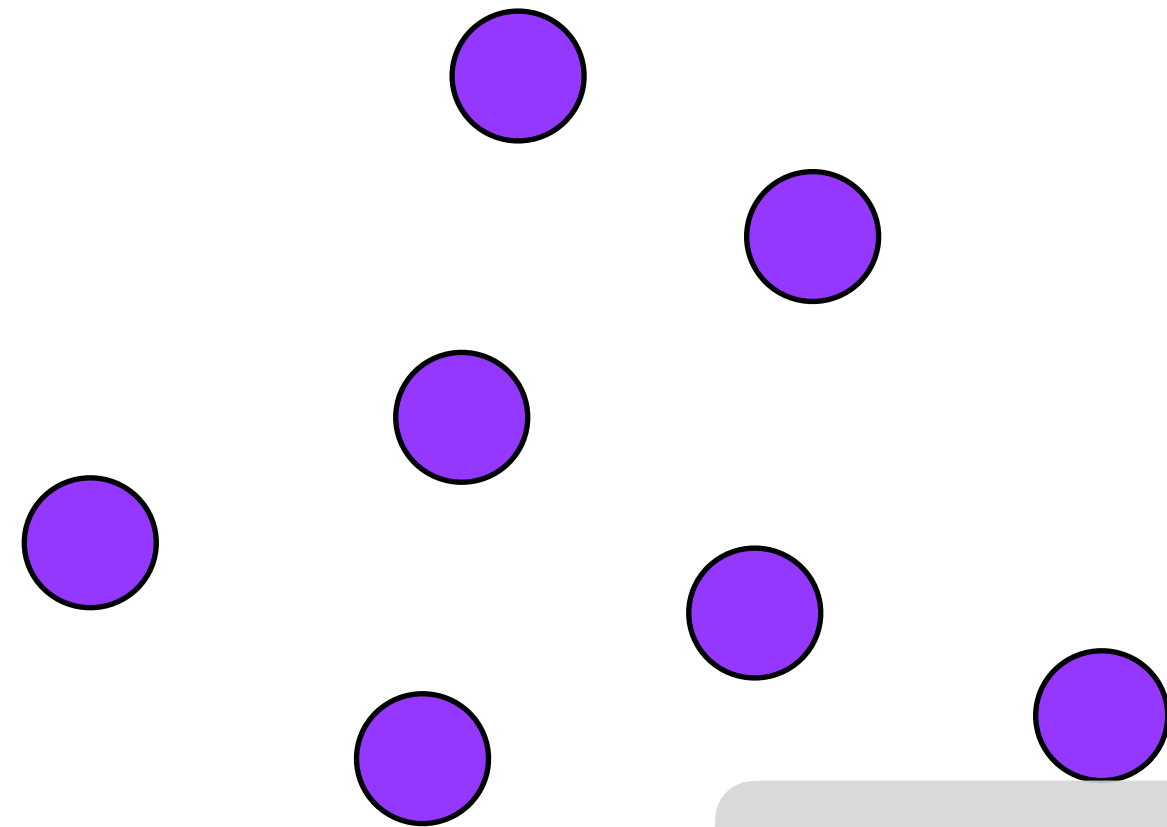


Similarity between documents  $x, y$ :  $\exp(-\gamma \text{WMD}(x, y))$

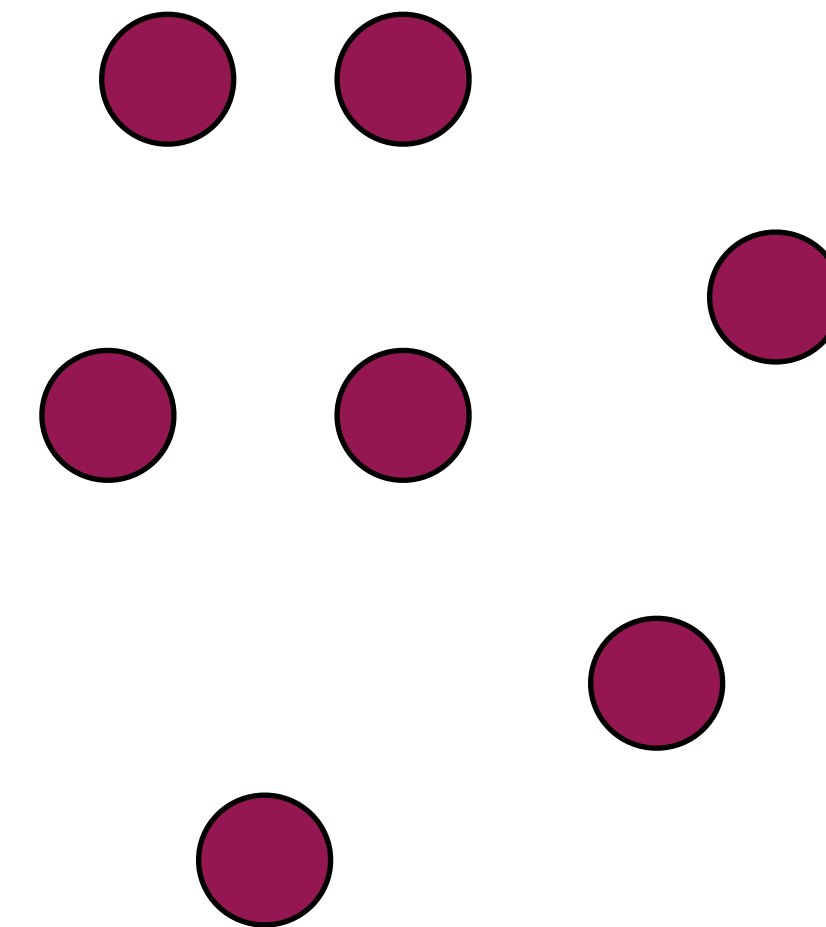


# Our Approach

$\mathbb{R}^L$

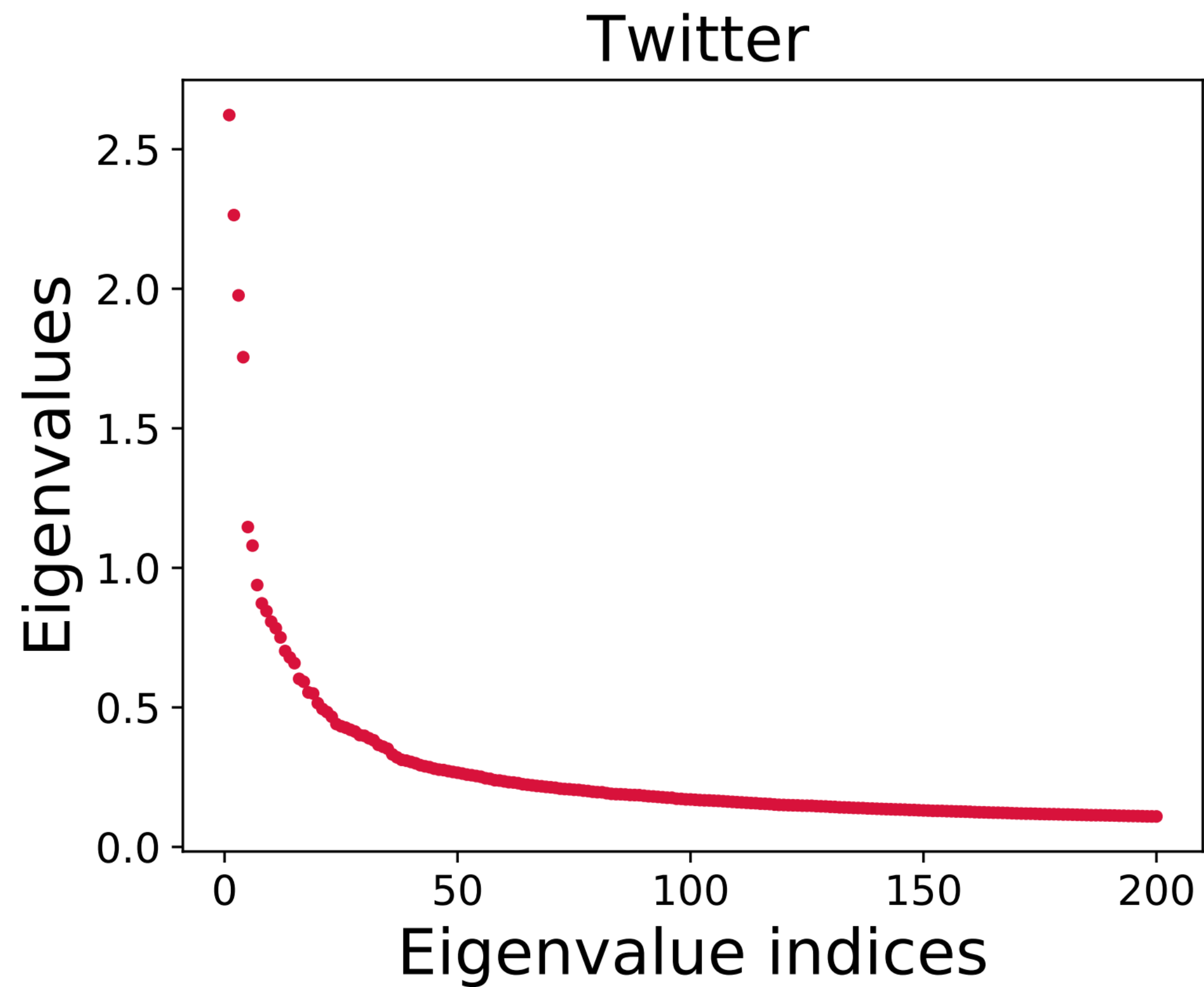


Similarity between documents  $x, y$ :  $\exp(-\gamma \text{WMD}(x, y))$



Similarity matrix not PSD but near-PSD

# Our Approach



Similarity matrix not PSD but near-PSD

# Our Approach

$\mathbb{R}^L$



Similarity between documents  $x, y$ :  $\exp(-\gamma \text{WMD}(x, y))$



But  $O(n^2)$  WMD computations is prohibitively expensive!

Similarity matrix not PSD but near-PSD



## Sentence Similarity Task

Score	English	Spanish
5/4	<i>The two sentences are completely equivalent, as they mean the same thing.</i>	
	The bird is bathing in the sink. Birdie is washing itself in the water basin.	El pájaro se esta bañando en el lavabo. El pájaro se está lavando en el aguamanil.
4	<i>The two sentences are mostly equivalent, but some unimportant details differ.</i>	
	In May 2010, the troops attempted to invade Kabul. The US army invaded Kabul on May 7th last year, 2010.	
3	<i>The two sentences are roughly equivalent, but some important information differs/missing.</i>	
	John said he is considered a witness but not a suspect. "He is not a suspect anymore." John said.	John dijo que él es considerado como testigo, y no como sospechoso. "Él ya no es un sospechoso," John dijo.
2	<i>The two sentences are not equivalent, but share some details.</i>	
	They flew out of the nest in groups. They flew into the nest together.	Ellos volaron del nido en grupos. Volaron hacia el nido juntos.
1	<i>The two sentences are not equivalent, but are on the same topic.</i>	
	The woman is playing the violin. The young lady enjoys listening to the guitar.	La mujer está tocando el violín. La joven disfruta escuchar la guitarra.
0	<i>The two sentences are completely dissimilar.</i>	
	John went horse back riding at dawn with a whole group of friends. Sunrise at dawn is a magnificent view to take in if you wake up early enough for it.	Al amanecer, Juan se fue a montar a caballo con un grupo de amigos. La salida del sol al amanecer es una magnífica vista que puede presenciar si usted se despierta lo suficientemente temprano para verla.



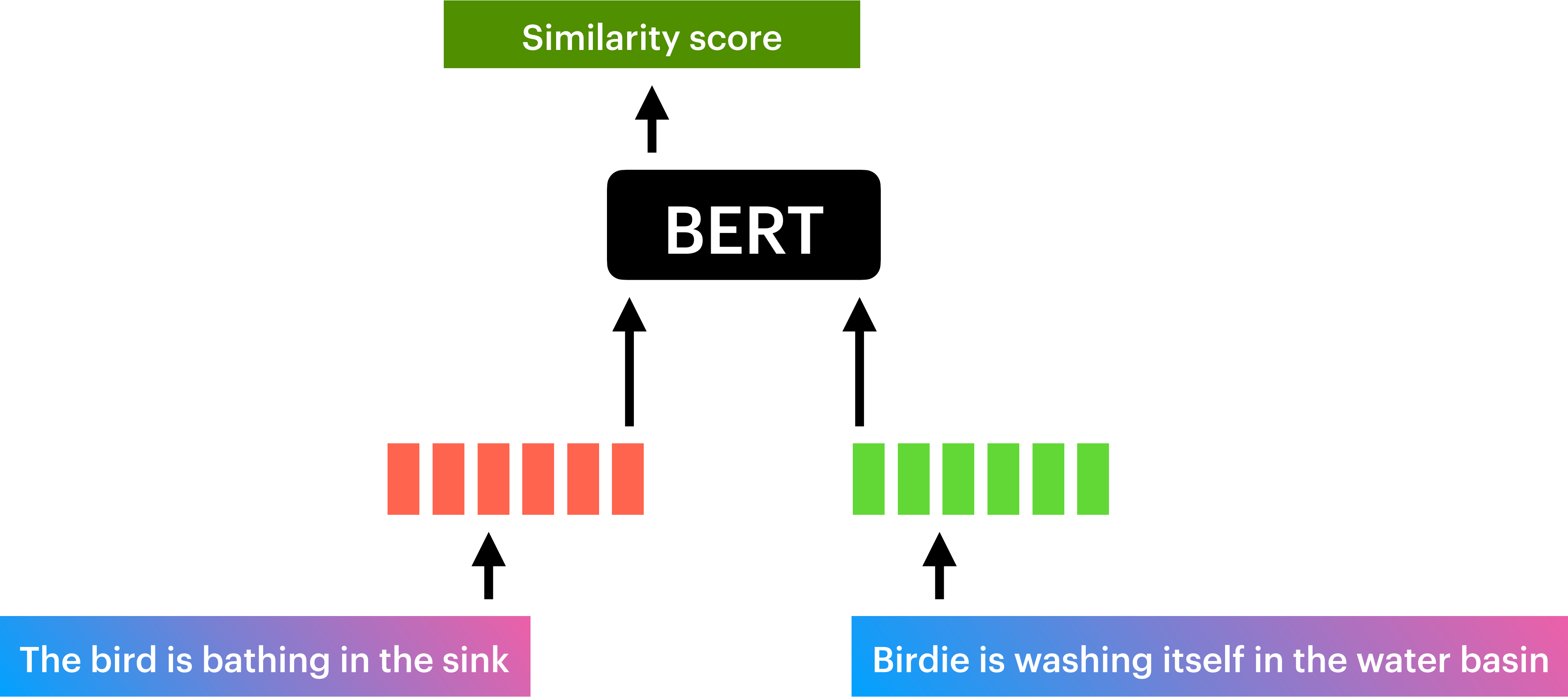
## Sentence Similarity Task

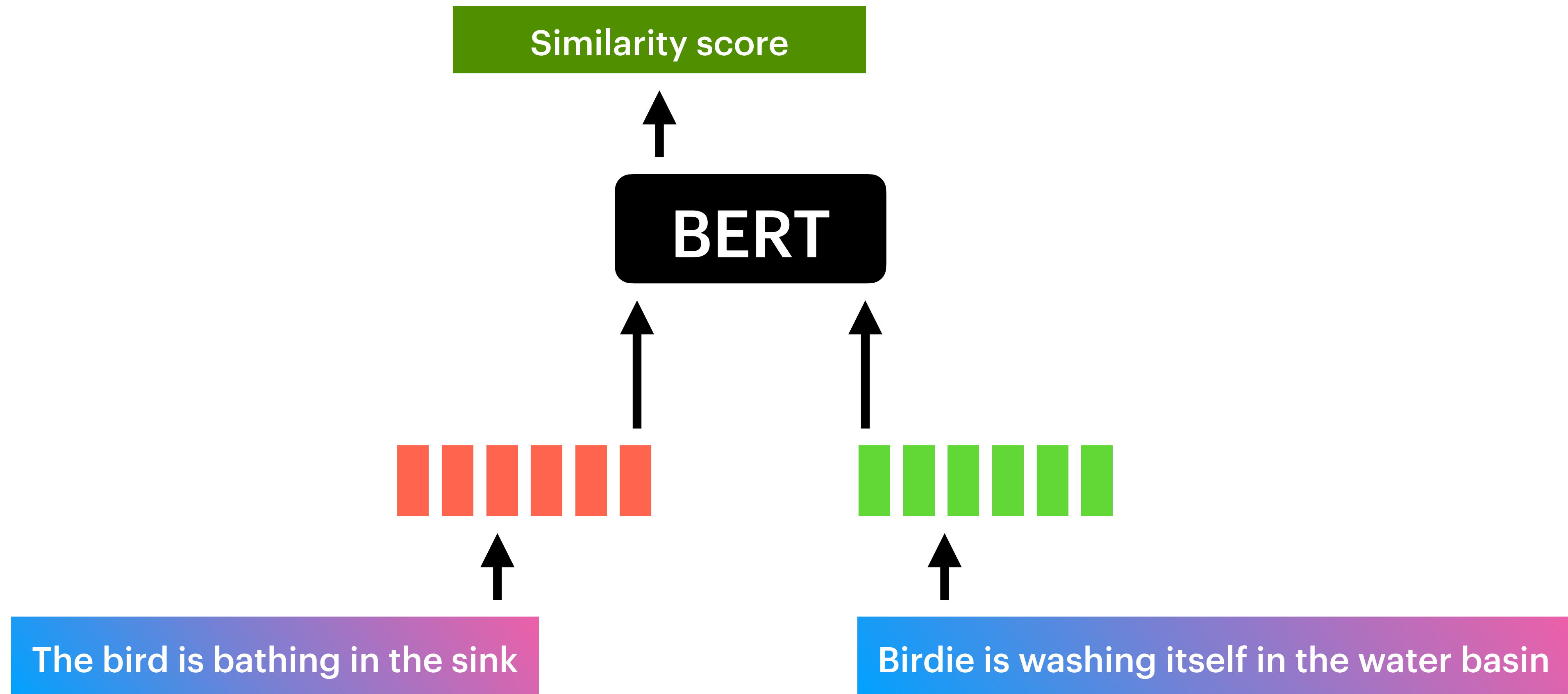
Score	English	Spanish
5/4	<i>The two sentences are completely equivalent, as they mean the same thing.</i>	
	The bird is bathing in the sink. Birdie is washing itself in the water basin.	El pájaro se esta bañando en el lavabo. El pájaro se está lavando en el aguamanil.
4	<i>The two sentences are mostly equivalent, but some unimportant details differ.</i>	
	In May 2010, the troops attempted to invade Kabul.	

Dataset	Score range	Train	Test	Application
STS-B	1-5	5749	3000	Semantic similarity of sentence pairs based on human annotations
MRPC	0-1	3668	816	Semantic equivalence of sentence pairs
RTE	0-1	2490	554	Text entailment of news and Wikipedia articles

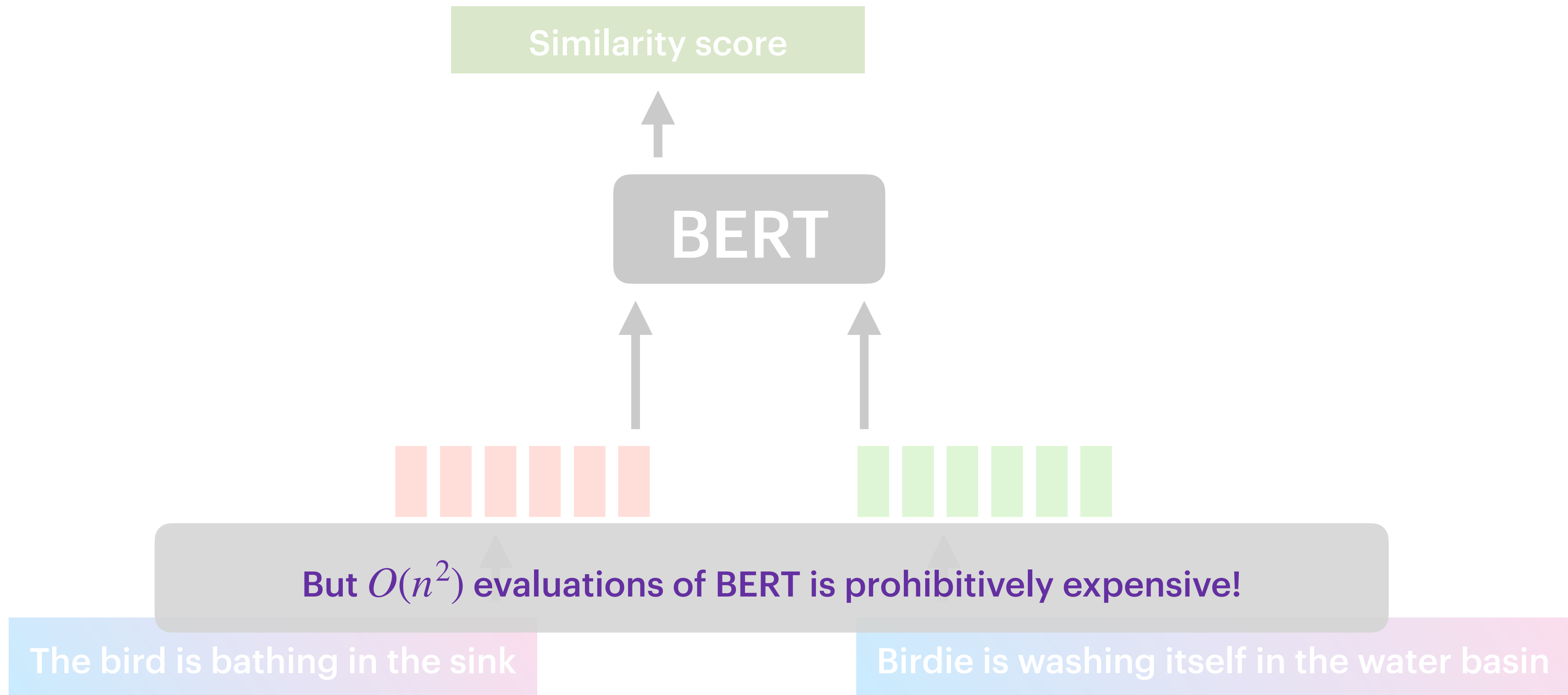
	They flew out of the nest in groups. They flew into the nest together.	Ellos volaron del nido en grupos. Volaron hacia el nido juntos.
1	<i>The two sentences are not equivalent, but are on the same topic.</i>	
	The woman is playing the violin. The young lady enjoys listening to the guitar.	La mujer está tocando el violín. La joven disfruta escuchar la guitarra.
0	<i>The two sentences are completely dissimilar.</i>	
	John went horse back riding at dawn with a whole group of friends. Sunrise at dawn is a magnificent view to take in if you wake up early enough for it.	Al amanecer, Juan se fue a montar a caballo con un grupo de amigos. La salida del sol al amanecer es una magnífica vista que puede presenciar si usted se despierta lo suficientemente temprano para verla.





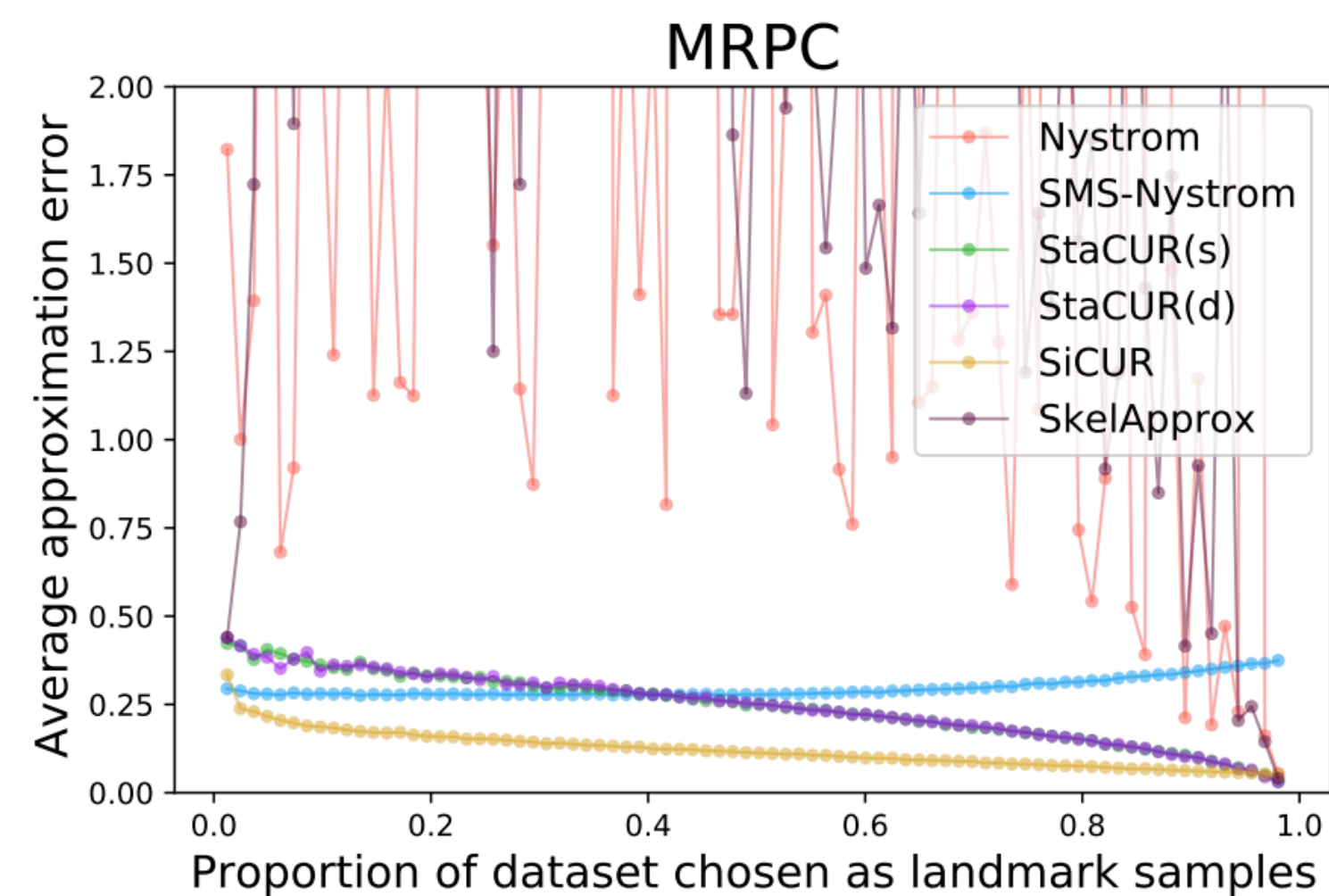
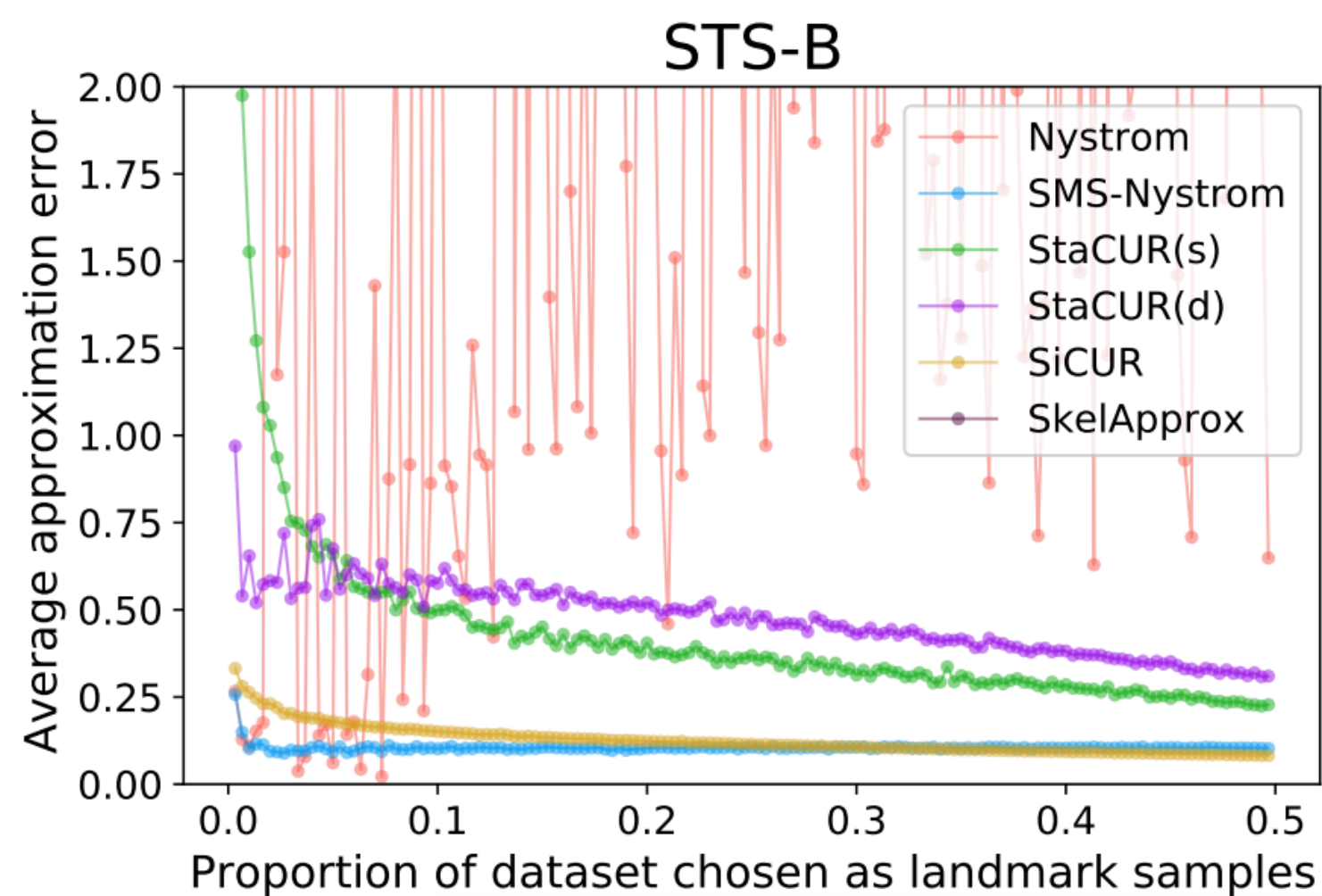
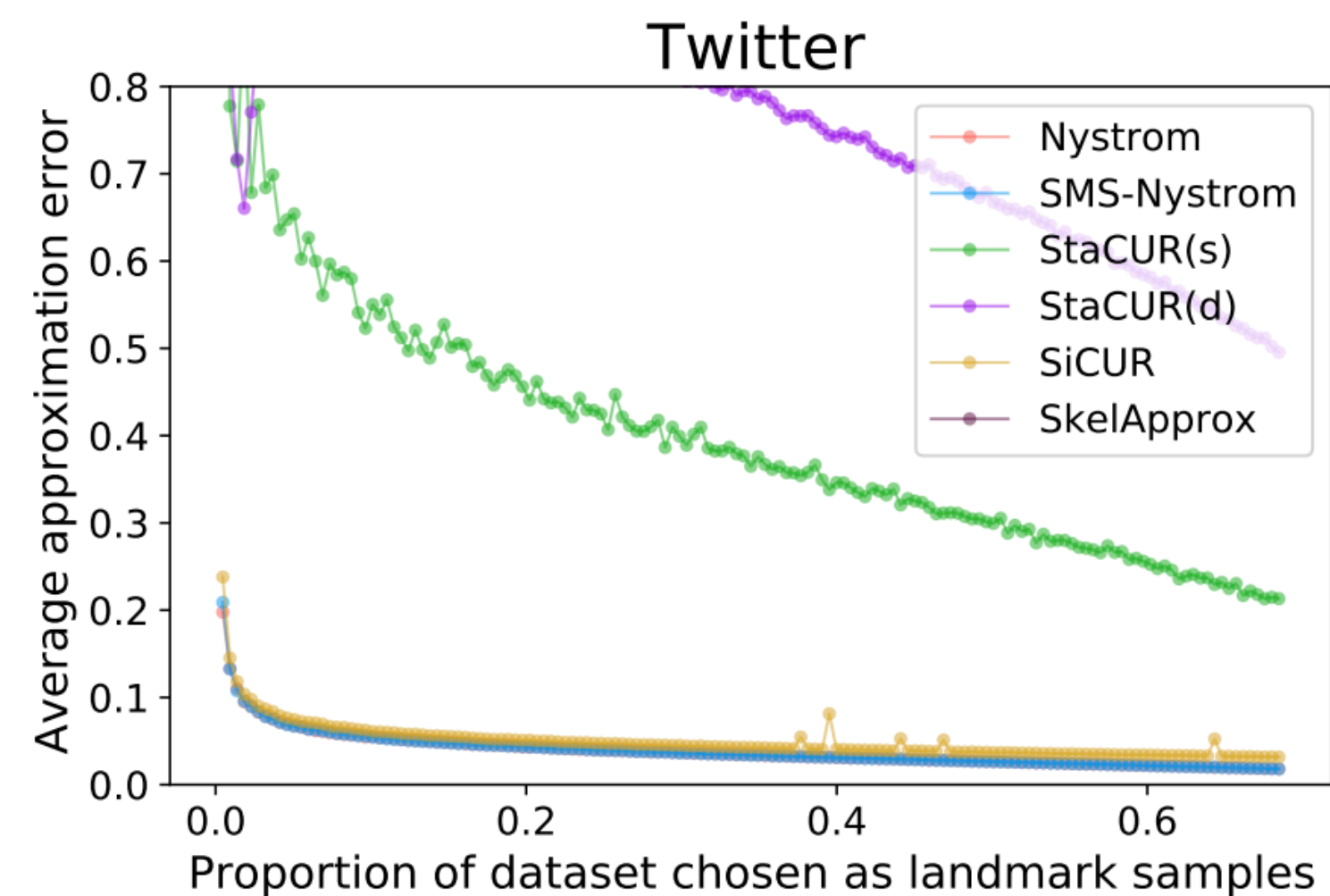
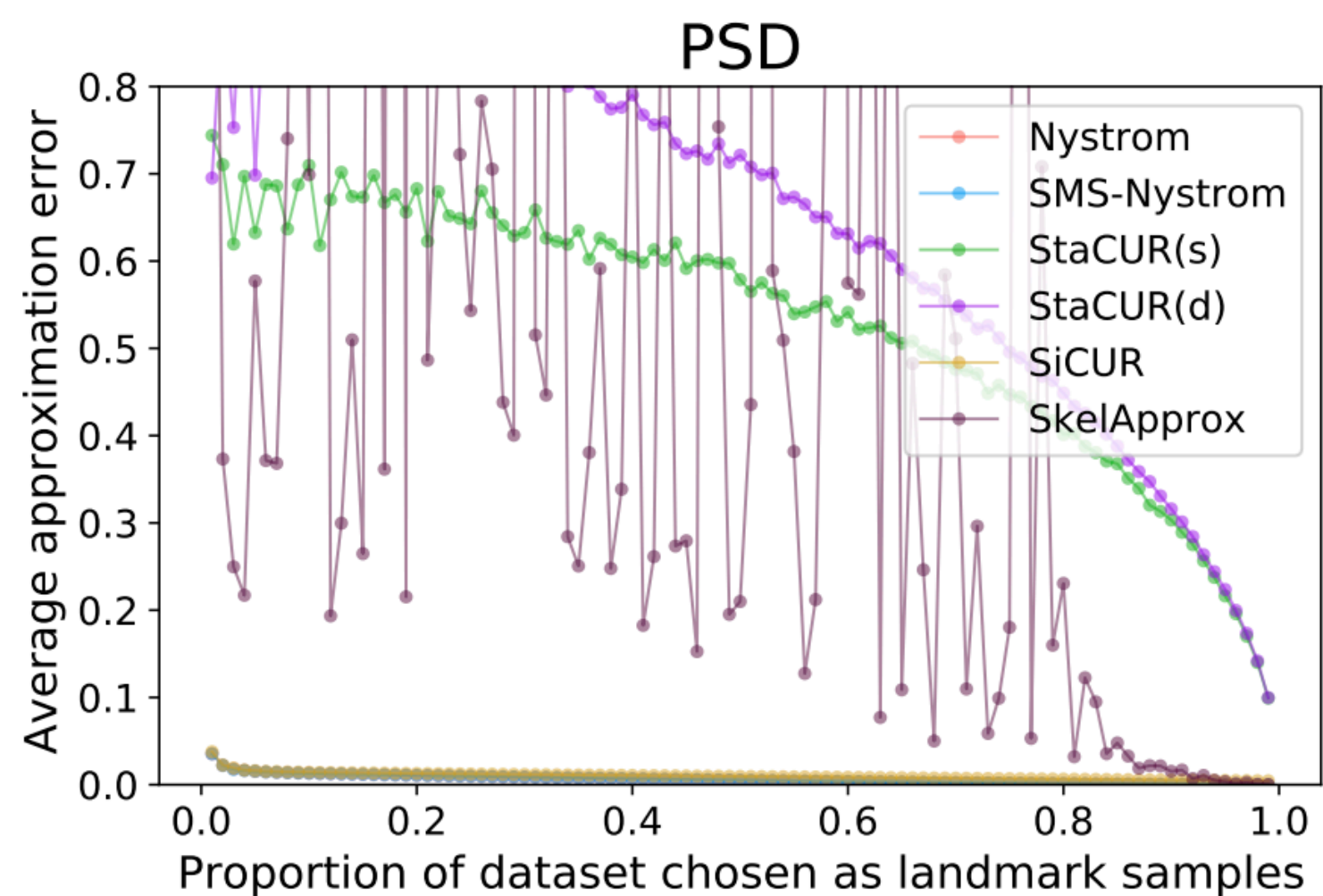


$O(n^2)$  similarities for all pairs



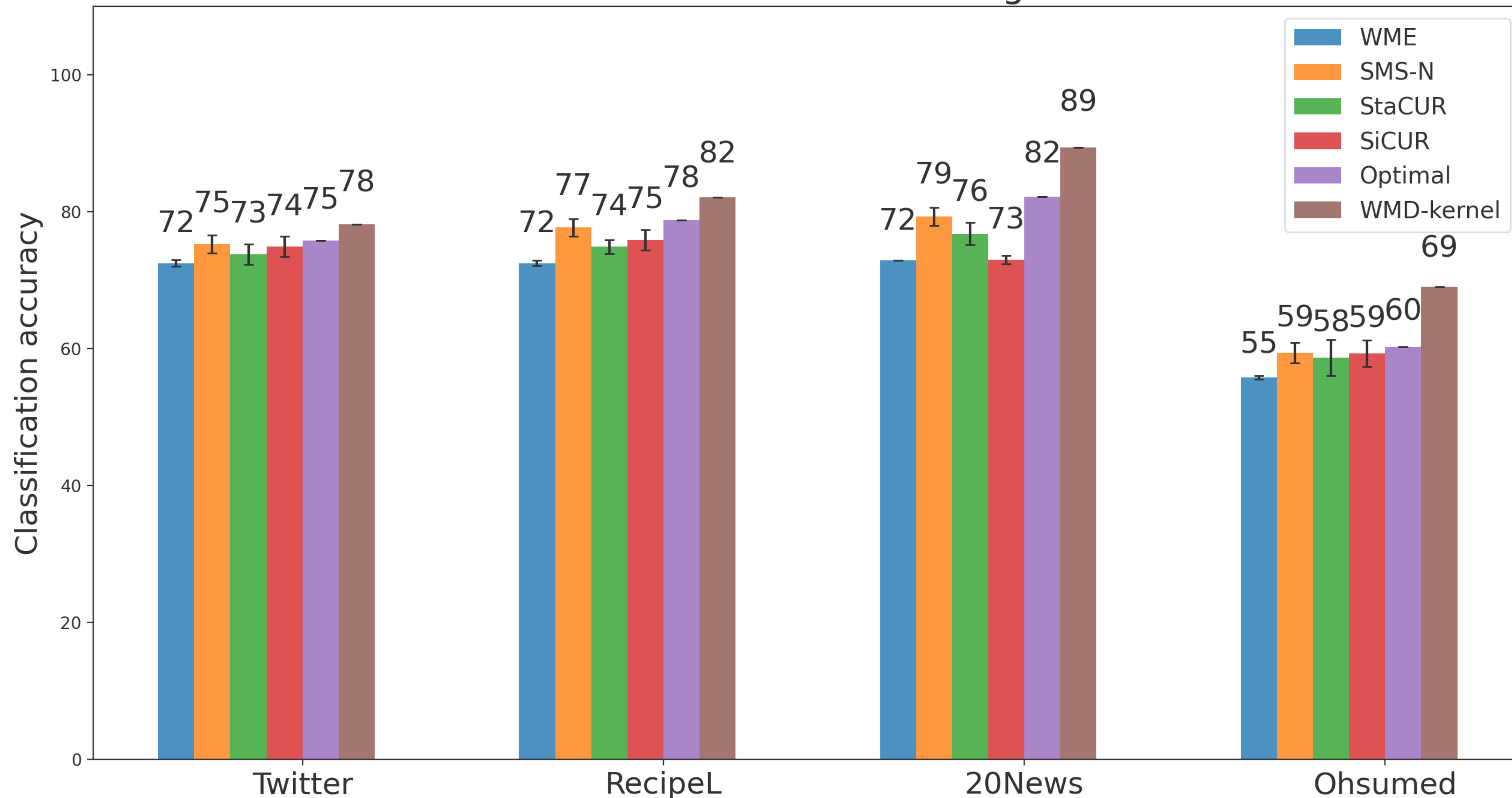
$O(n^2)$  similarities for all pairs

# Comparing Approximation on Given Datasets



# Document Classification Task

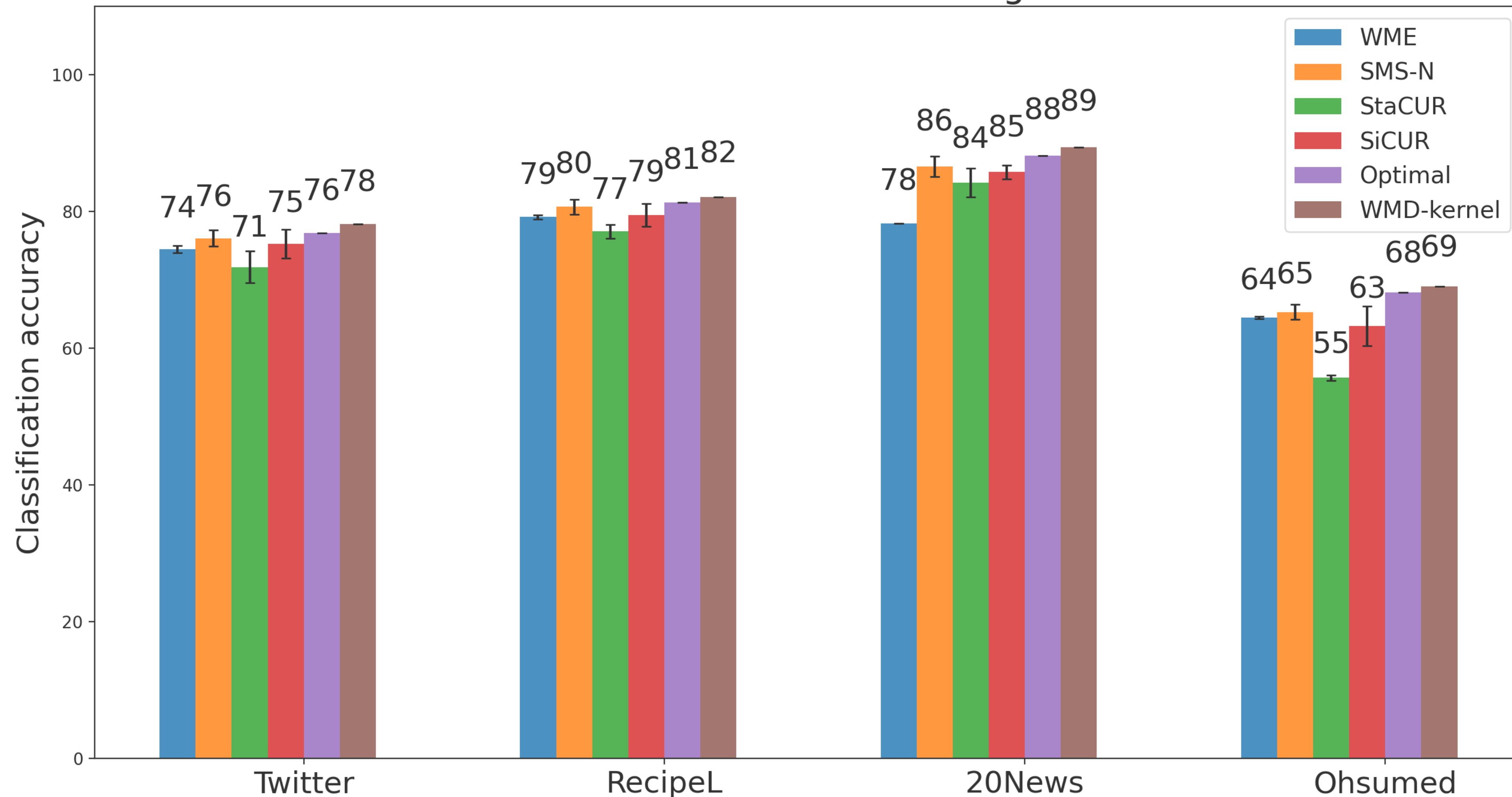
Document classification using WMD



Small rank accuracy scores

# Document Classification Task

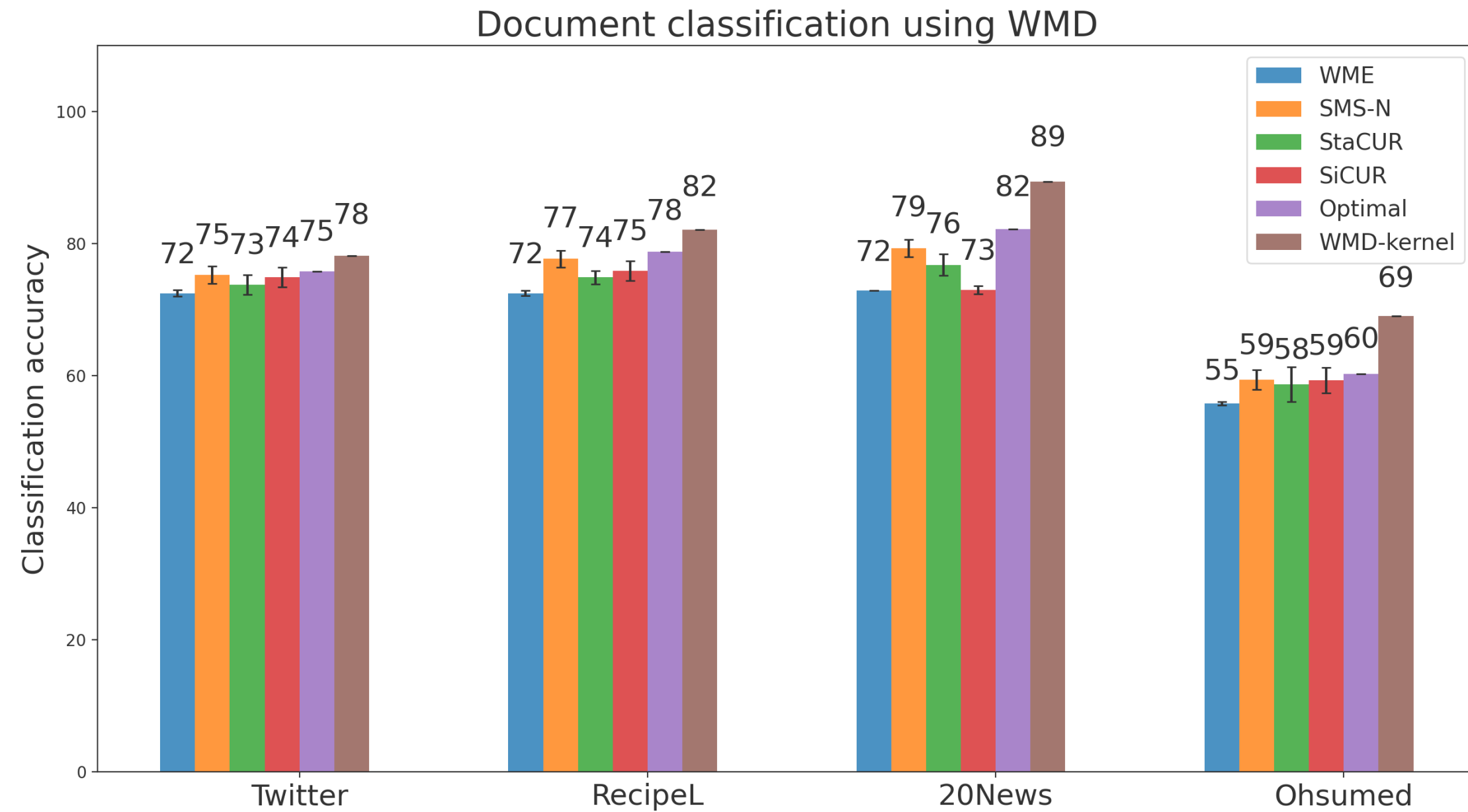
Document classification using WMD



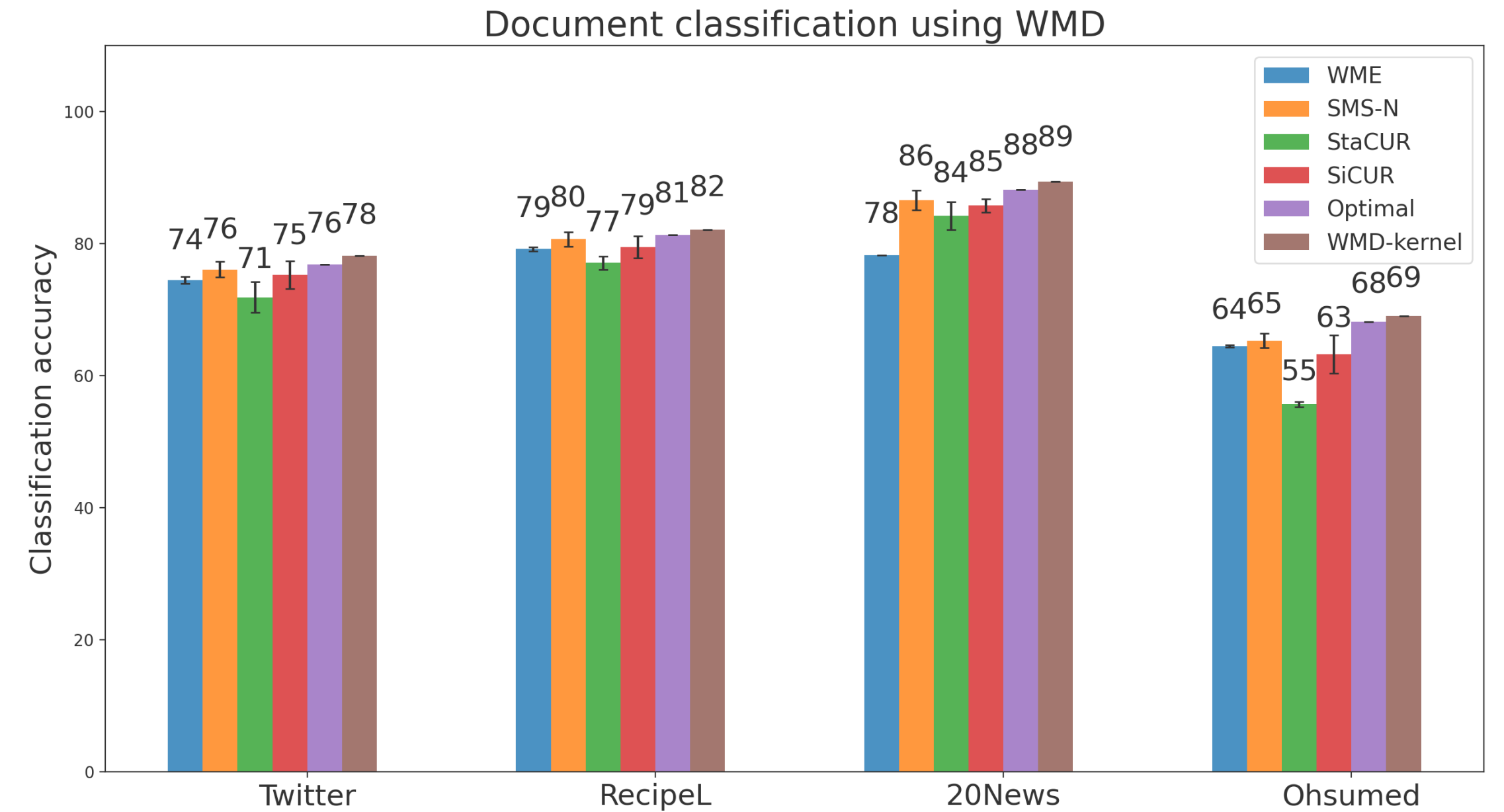
Large rank accuracy scores



# Document Classification Task



Small rank accuracy scores



Large rank accuracy scores

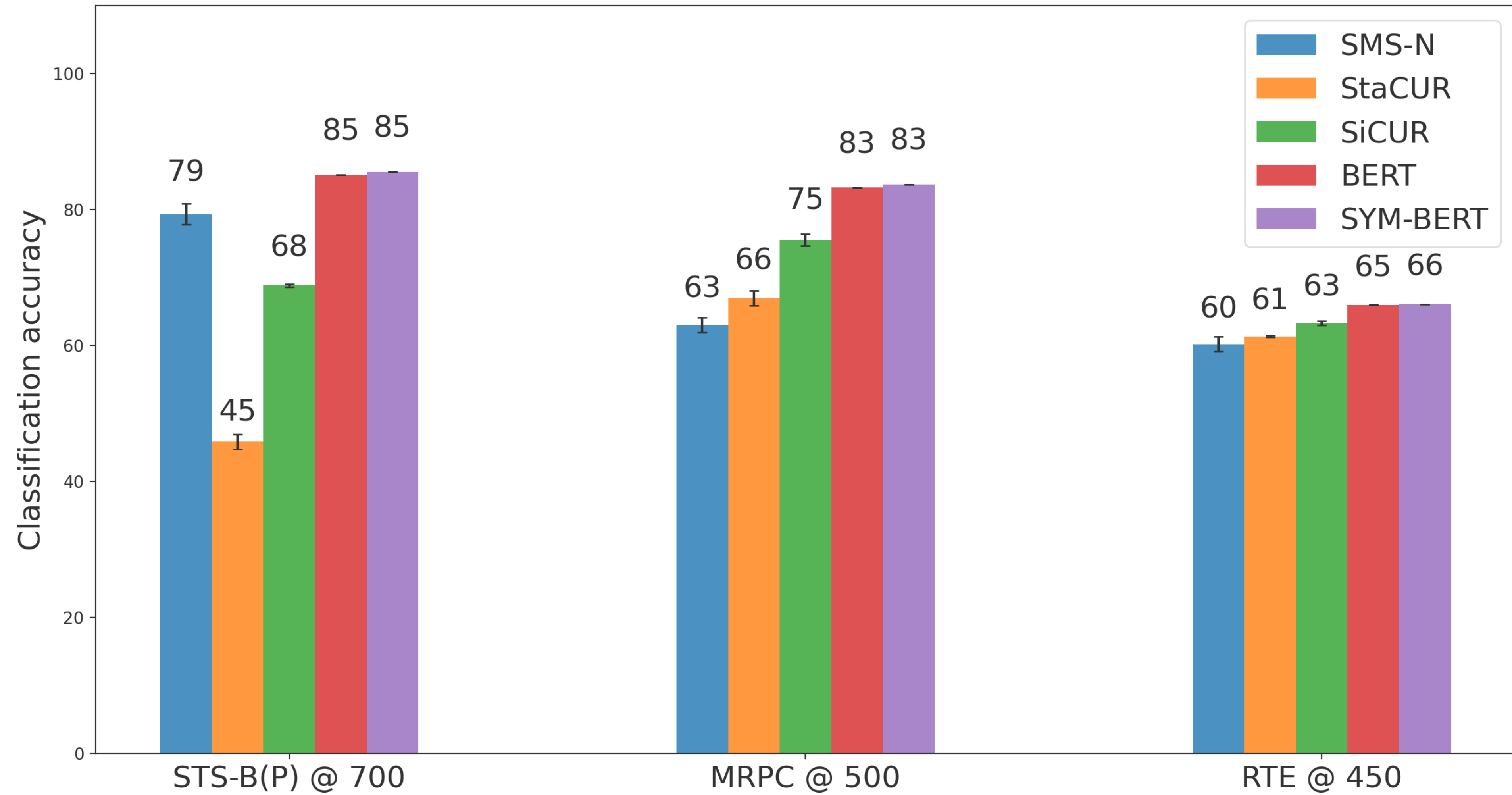
Method	Twitter	Recipe_L	Ohsumed	20News
WME(SR)	13.02	5639.06	85.43	2712.12
SMS-N(SR)	86.23	13979.01	629.54	9422.05
WME(LR)	102.06	29238.48	2787.00	13021.13
SMS-N(LR)	1014.06	223902.32	21246.65	130342.28

Time in seconds



# Approximation of GLUE tasks

Performance on GLUE tasks

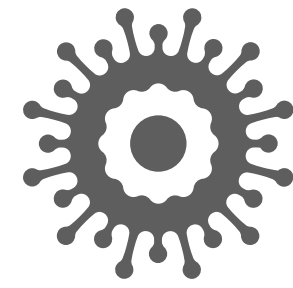


# Cross-Document Entity & Coreference

## A Rapid and Specific Assay for the Detection of MERS-CoV

Pei Huang<sup>1,2</sup>, Hualai Wang<sup>2,3,4\*</sup>, Zengguo Cao<sup>2,3</sup>, Hongli Jin<sup>2,3</sup>, Hang Beibei Yu<sup>5</sup>, Feihu Yan<sup>6</sup>, Xingxing Hu<sup>1,2</sup>, Fangfang Wu<sup>6</sup>, Cuicui Jiao<sup>6</sup>, Shangnan Xu<sup>1,2</sup>, Yongkun Zhao<sup>6,7</sup>, Na Feng<sup>6,8</sup>, Jianzhong Wang<sup>1</sup>, W Tiecheng Wang<sup>2,4</sup>, Yuwei Gao<sup>1,4</sup>, Songtao Yang<sup>4,4</sup> and Xianzhu Xia<sup>2,3</sup>

<sup>1</sup>Animal Science and Technology College, Jilin Agricultural University, Changchun, China, <sup>2</sup>Institute for Zoonosis Prevention and Control, Institute of Military Veterinary, Academy of Military Medical Sciences, <sup>3</sup>College of Veterinary Medicine, Jilin University, Changchun, China, <sup>4</sup>Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, China, <sup>5</sup>State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, <sup>6</sup>Guangzhou Eighth People's Hospital of Guangzhou Medical University, Guangzhou, China, <sup>7</sup>Department of Clinical Laboratory, College of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China



Amazingly, it is effective against SARS and MERS.

OPEN ACCESS

**Edited by:** Dirk Dittmar, University of North Carolina at Chapel Hill, United States  
**Reviewed by:** Timothy Sheahan, University of North Carolina at Chapel Hill, United States; Yibo Li, University of Pennsylvania, United States  
**\*Correspondence:** Hualai Wang, wangh20@ornl.com

**Specialty section:** This article was submitted to

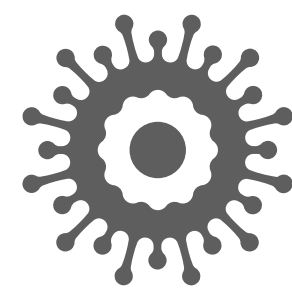
Middle East respiratory syndrome coronavirus (MERS-CoV) is a novel human coronavirus that can cause severe illness. In this study, we established a reverse transcription loop isotherm amplification (RT-LAMP) assay. The result was visible by the naked eye. The detection limit was 2 × 10<sup>3</sup> copies of MERS-CoV RNA. The assay showed high specificity. Compared to the World Health Organization (WHO) reference method, the RT-LAMP assay requires less expensive equipment.

Contents lists available at ScienceDirect  
**Vaccine**  
journal homepage: www.elsevier.com/locate/vaccine

## DNA vaccine encoding Middle East respiratory syndrome coronavirus S1 protein induces protective immune responses in mice

Hang Chi<sup>a</sup>, Xuexing Zheng<sup>a,b</sup>, Xiwen Wang<sup>a</sup>, Chong Wang<sup>a</sup>, Hualai Wang<sup>a,c</sup>, Weiwei Gai<sup>a</sup>, Stanley Perlman<sup>d</sup>, Songtao Yang<sup>a,c,e</sup>, Jincun Zhao<sup>a,c,e</sup>, Xianzhu Xia<sup>a,c,e</sup>

<sup>a</sup>Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Institute of Military Veterinary, Academy of Military Medical Science, Changchun, China; <sup>b</sup>School of Public Health, Jilin University, Changchun, China; <sup>c</sup>Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, China; <sup>d</sup>Department of Microbiology, University of North Carolina at Chapel Hill, United States; <sup>e</sup>State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China



The Middle East respiratory syndrome coronavirus (MERS-CoV) is an emerging pathogen...

ARTICLE IN PRESS

Article history:  
Received 10 June 2016  
Received in revised form 20 February 2017  
Accepted 28 February 2017  
Available online 14 March 2017

**Keywords:**  
MERS-CoV  
DNA vaccine  
Spike protein

1. Introduction

Middle East respiratory syndrome (MERS)-coronavirus (MERS-CoV), an emerging zoonotic virus, is the causative agent of MERS. MERS-CoV was first identified in Saudi Arabia in 2012 and MERS cases have been reported in 27 countries since then [1,2]. As of February 10, 2017, 1905 laboratory-confirmed cases, including 677 deaths related to MERS-CoV, had been reported to WHO (~36% mortality). Several family clusters and nosocomial clusters cases have been reported, revealing the human-to-human transmissibility of MERS-CoV, and raising the concern of a MERS-CoV global pandemic [3–5]. Currently, no licensed therapeutic or vaccine is available, which highlights the need for efficient vaccines against MERS-CoV.

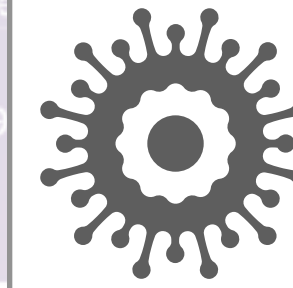
To date, several vaccine candidates have been developed, such as viral vector-based recombinants [6–11], subunit vaccines [12–19], DNA vaccines [20], DNA prime/protein-boost vaccines [21] and a reverse genetics-constructed recombinant coronavirus vaccine [22]. Among them, DNA vaccines present a range of unique advantages such as proper antigen protein folding, rapid design and production, cost-effectiveness, and stability at non-refrigerated temperatures for convenient storage and shipping [23]. Furthermore, it has been reported that DNA vaccines can induce both humoral and cellular immune responses against MERS-CoV and SARS-CoV infection [20,24,25]. MERS-CoV is the first lineage of *Betacoronavirus* known to infect humans [26]. The genome of MERS-CoV encodes four structural proteins – spike (S), envelope (E), membrane (M) and nucleocapsid (N) [27]. The S protein, a class I fusion protein forming protruding

## Unexpected outbreaks of arbovirus infections: lessons from Central and South America

Didier Musso, MD, Prof Alfonso J Rodriguez-Morales, MD, José Eduardo Levi, PhD

Van-Mai Cao-Lormeau

Published: June 19, 2018



Pandemic arboviruses have emerged as a major global health problem in the past four decades.

Check for updates

## Summary

Pandemic arboviruses have emerged as a major global health

decades. Predicting where and when the next outbreak will occur is a challenge, but history tells us that such swan events (epidemics that are predicted to have an extreme effect) will continue to occur as globalisation expand. We briefly review the history of arboviral epidemics that have occurred in the past 50 years in the American and Pacific regions, to illustrate the need for surveillance, and to highlight the need for improved diagnostic tools, including laboratory-based

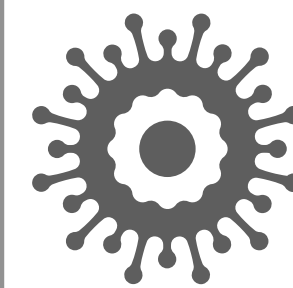
TICK-BORNE DISEASES

0025-7125/02 \$15.00 + .00

## COLORADO TICK FEVER

Richard Klasco, MD

Colorado tick fever (CTF), also known as *mountain fever* and *mountain tick fever*, is a well-described viral tick-borne disease common to the Rocky Mountain region of North America. The disease is characterized by a biphasic illness with a prodromal phase followed by a febrile phase. The disease is caused by the Colorado tick fever virus (CTFV), a member of the *Coltivirus* genus, family *Reoviridae*. The disease is transmitted from the bite of an infected wood tick.



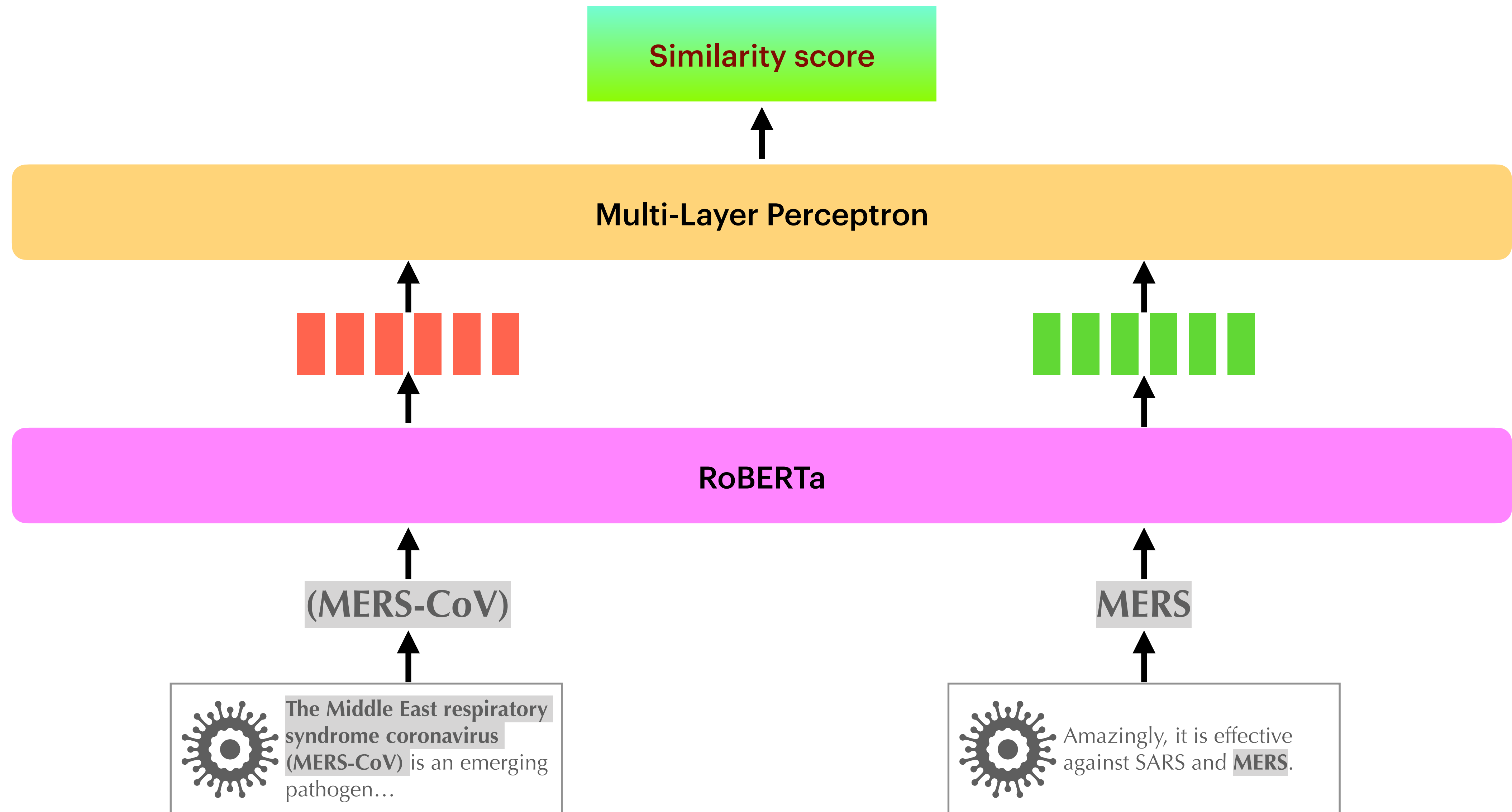
The arboviral infection, CTF, is transmitted from the bite of an infected wood tick.

CAUSE AND EFFECT

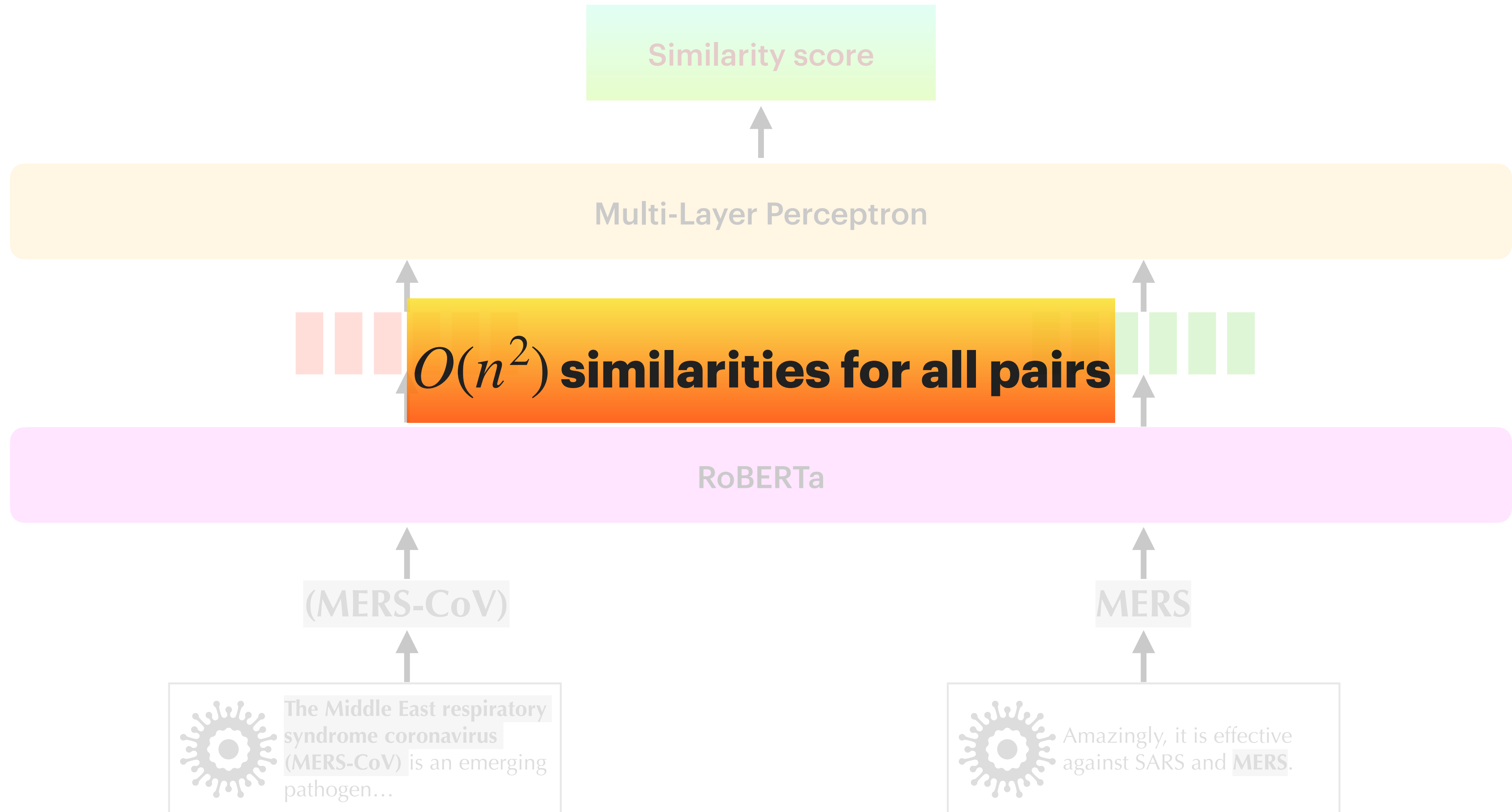
Colorado tick fever (CTF) is a biphasic illness caused by the Colorado tick fever virus (CTFV), a member of the *Coltivirus* genus, family *Reoviridae*. The disease is transmitted from the bite of an infected wood tick. CTFV is the causative agent of CTF. Formerly classified as an orbivirus, the sixth report of the International Committee on Taxonomy of Viruses identified CTFV as a member of the genus *Coltivirus* (group A), family *Reoviridae* (virus code, 60.0.4.0.001; virus accession number, 60040001).<sup>24</sup> At least 22 strains of CTFV are known,<sup>5,24</sup> many of which cause disease in humans.<sup>8</sup> Of these, the Florio strain is the best characterized.<sup>8</sup> Eyach, a group A *Coltivirus* closely related to CTFV, has been detected in European Ixodidae ticks and has been implicated in human disease in Czechoslovakia.<sup>12,24</sup>

In 2000, the CTFV genome was sequenced and was found to consist of 12 dsRNA segments that encode several important proteins.<sup>3</sup> These include VP1, the viral RNA dependent RNA polymerase; methyltransferases; RGD-binding proteins, extracellular proteins that mediate cell-

# Cross-Document Entity & Coreference



# Cross-Document Entity & Coreference

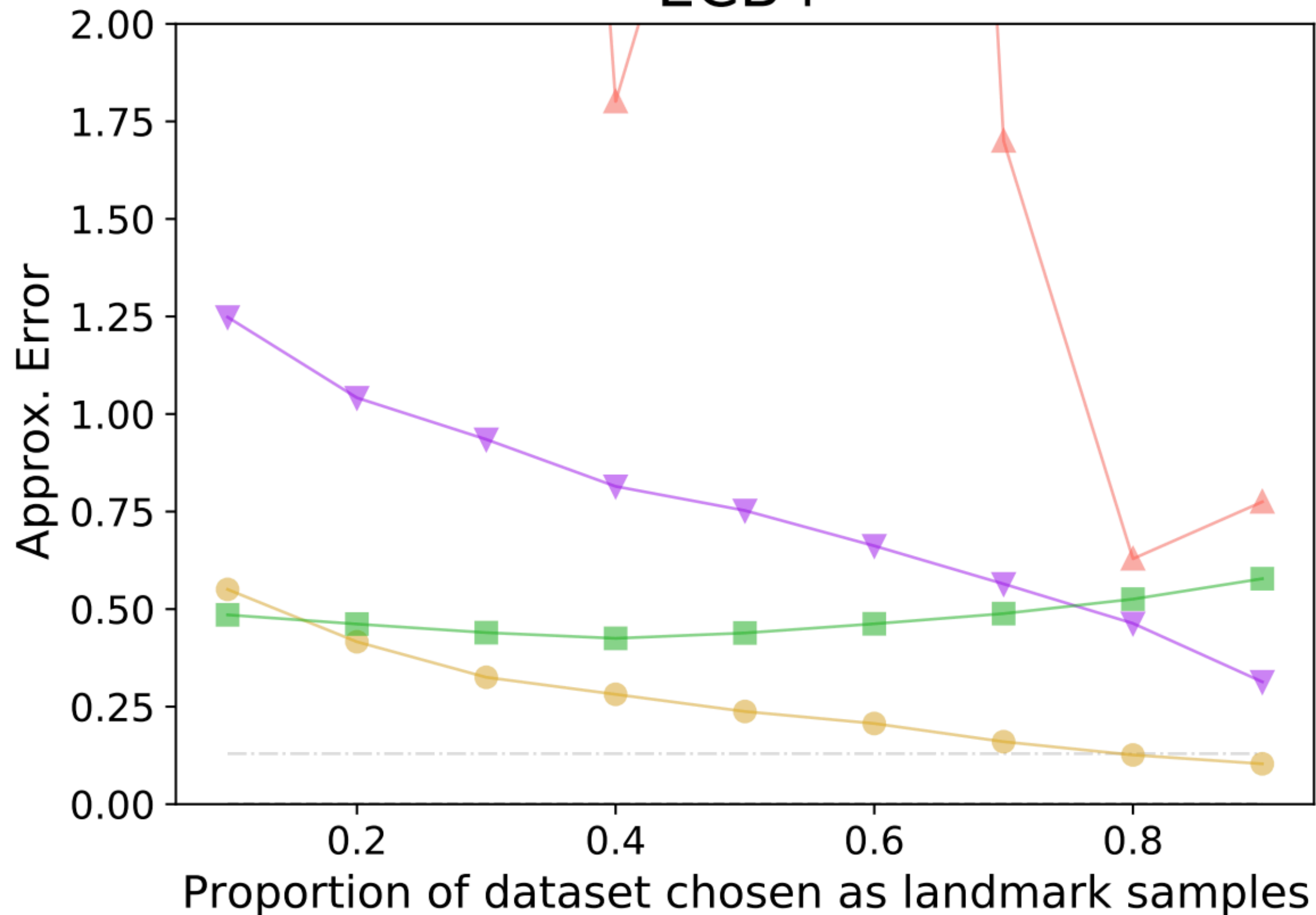


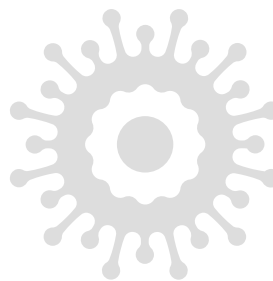


# Cross-Document Entity & Coreference

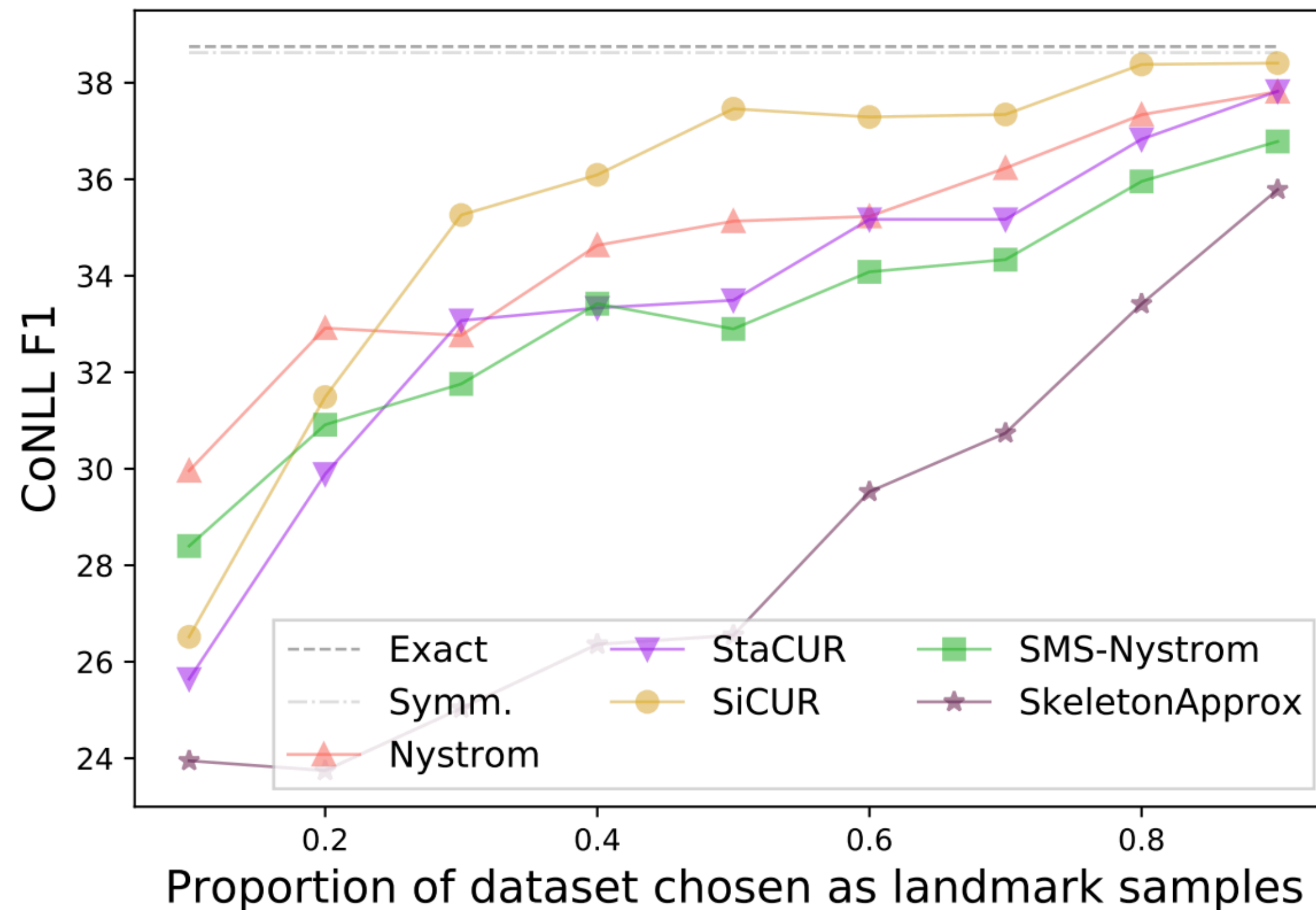
Similarity score

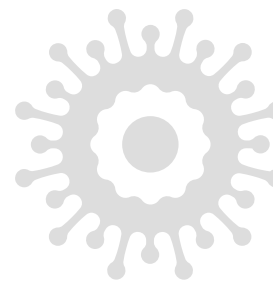
ECB+



 The Middle East respiratory syndrome coronavirus (MERS-CoV) is an emerging pathogen...

ECB+



 Amazingly, it is effective against SARS and MERS.

# Conclusions

We show that indefinite matrices arising in NLP can be approximated using sublinear algorithms

Simple variant of Nyström and variants of CUR display strong performance in variety of tasks

***Thank you! Questions?***