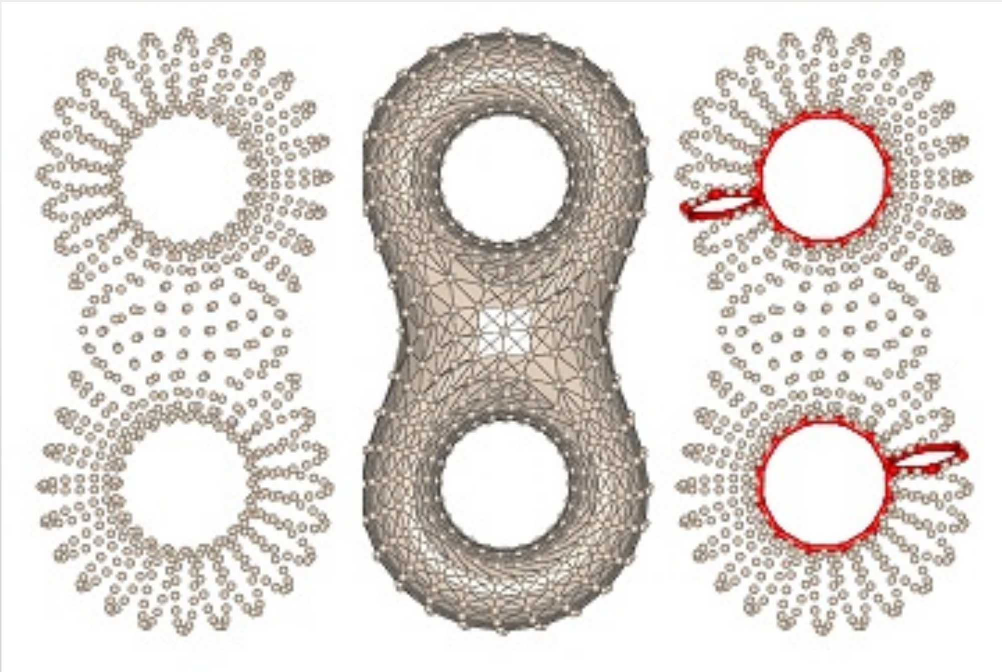
# Algebraic Topology, Quantum Algorithms and BIG DATA



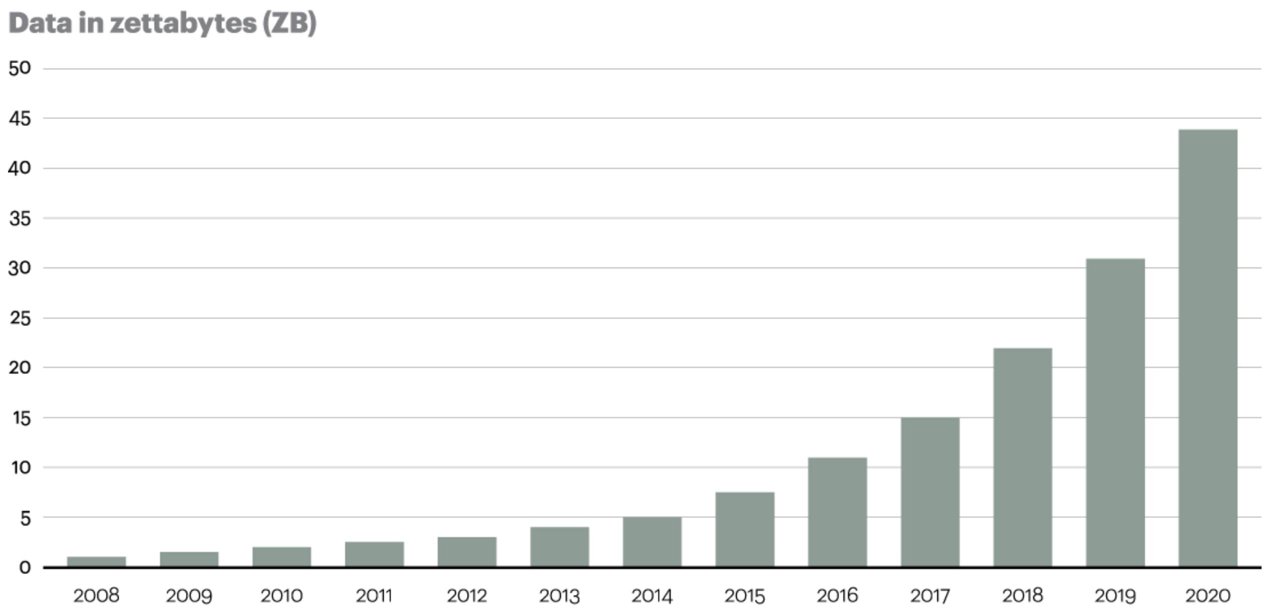
Quantum Physics and Geometry" (2017)

Paolo Zanardi, USC Los Angeles

# A new Buzzword: BIG DATA

Figure 1

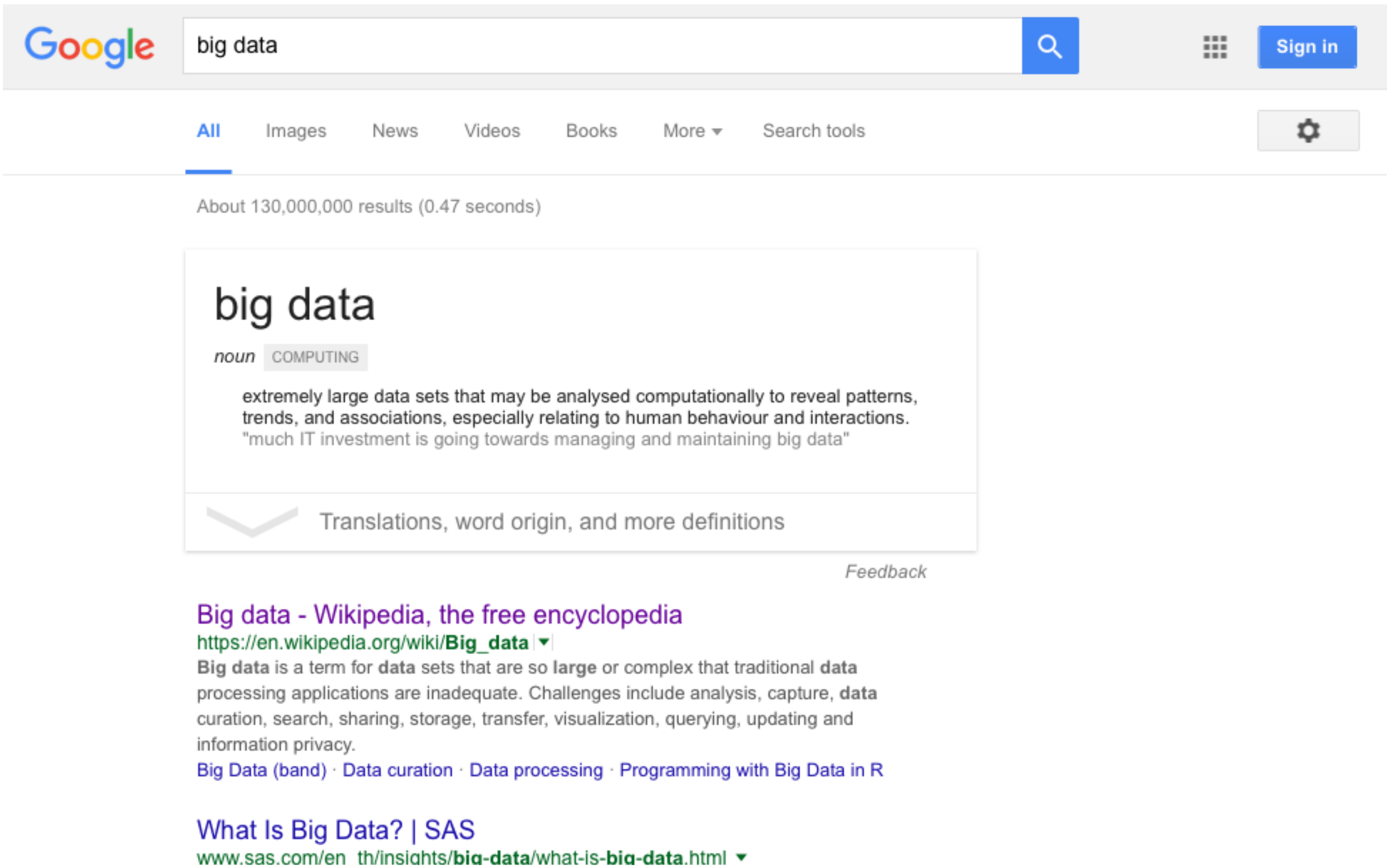
**Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020**

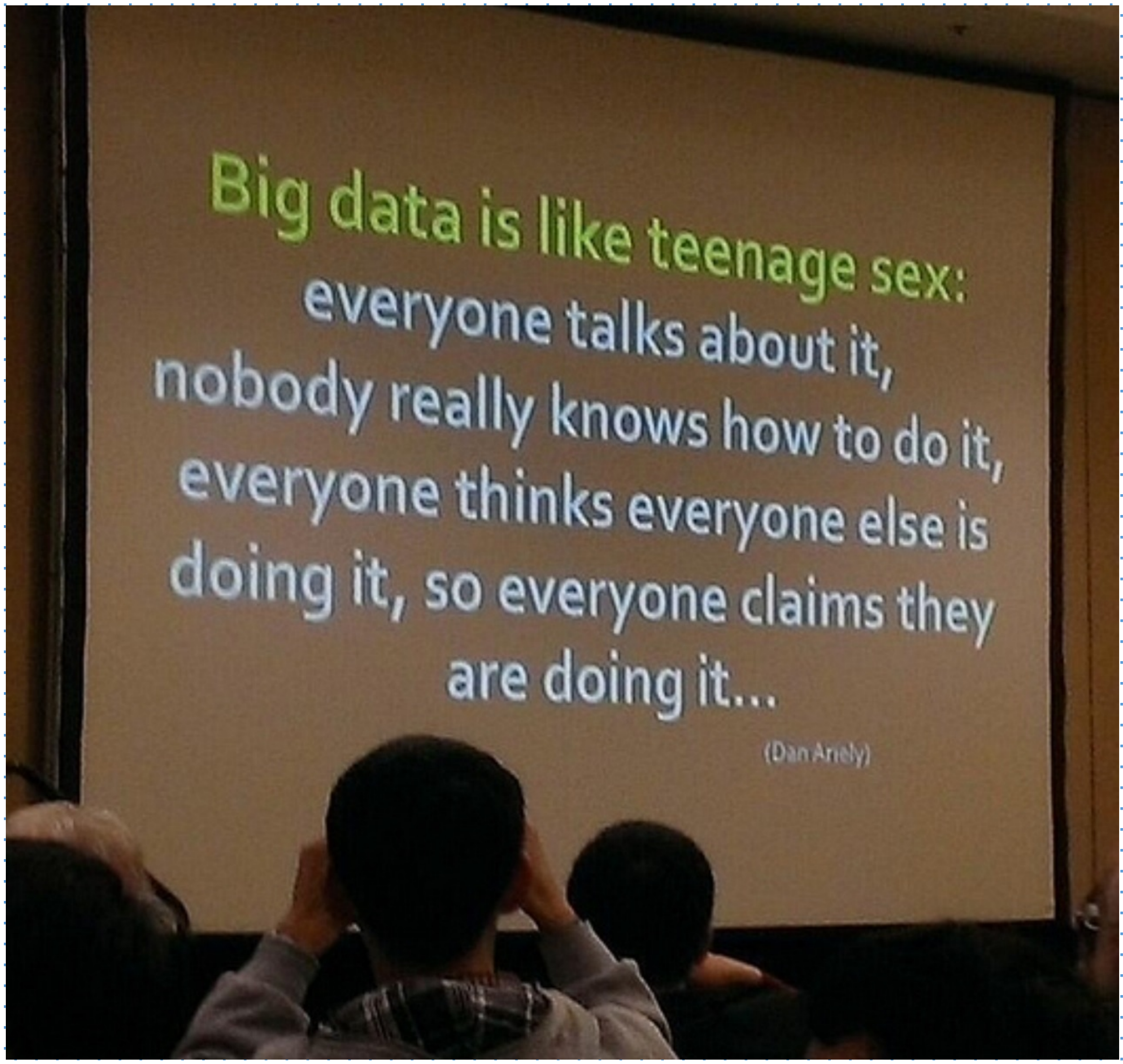



Source: Oracle, 2012

**ZB**=1,000,000,000,000,000,000,000 =10^21 Bytes

# What’s BIG DATA?




Big data is like teenage sex:
everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is
doing it, so everyone claims they
are doing it...
(Dan Ariely)

# And We(=Quantum Information Folks) make no difference…..

PRL 108, 230506 (2012) PHYSICAL REVIEW LETTERS week ending 8 JUNE 2012

## Adiabatic Quantum Algorithm for Search Engine Ranking

Silvano Garnerone,[1,2,5] Paolo Zanardi,[2,5] and Daniel A. Lidar[2,3,4,5]
[1]*Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[2]*Department of Physics & Astronomy, University of Southern California, Los Angeles, California 90089, USA*
[3]*Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089, USA*
[4]*Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA*
[5]*Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
(Received 25 October 2011; published 4 June 2012)

We propose an adiabatic quantum algorithm for generating a quantum pure state encoding of the PageRank vector, the most widely used tool in ranking the relative importance of internet pages. We present extensive numerical simulations which provide evidence that this algorithm can prepare the quantum PageRank state in a time which, on average, scales polylogarithmically in the number of web pages. We argue that the main topological feature of the underlying web graph allowing for such a scaling is the out-degree distribution. The top-ranked $\log(n)$ entries of the quantum PageRank state can then be estimated with a polynomial quantum speed-up. Moreover, the quantum PageRank state can be used in "*q*-sampling" protocols for testing properties of distributions, which require exponentially fewer measurements than all classical schemes designed for the same task. This can be used to decide whether to run a classical update of the PageRank.

PRL 113, 130503 (2014) PHYSICAL REVIEW LETTERS week ending 26 SEPTEMBER 2014

## Quantum Support Vector Machine for Big Data Classification

Patrick Rebentrost,[1,*] Masoud Mohseni,[2] and Seth Lloyd[1,3,†]
[1]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Google Research, Venice, California 90291, USA*
[3]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
(Received 12 February 2014; published 25 September 2014)

Supervised machine learning is the classification of new data based on already classified training examples. In this work, we show that the support vector machine, an optimized binary classifier, can be implemented on a quantum computer, with complexity logarithmic in the size of the vectors and the number of training examples. In cases where classical sampling algorithms require polynomial time, an exponential speedup is obtained. At the core of this quantum big data algorithm is a nonsparse matrix exponentiation technique for efficiently performing a matrix inversion of the training data inner-product (kernel) matrix.

v:1307.0411v2 [quant-ph] 4 Nov 2013

## Quantum algorithms for supervised and unsupervised machine learning

*Seth Lloyd[1,3], Masoud Mohseni[2], Patrick Rebentrost[1]*
1. Massachusetts Institute of Technology, Research Laboratory for Electronics
2. Google Research
3. To whom correspondence should be addressed: slloyd@mit.edu

*Abstract:* Machine-learning tasks frequently involve problems of manipulating and classifying large numbers of vectors in high-dimensional spaces. Classical algorithms for solving such problems typically take time polynomial in the number of vectors and the dimension of the space. Quantum computers are good at manipulating high-dimensional vectors in large tensor product spaces. This paper provides supervised and unsupervised quantum machine learning algorithms for cluster assignment and cluster finding. Quantum machine learning can take time logarithmic in both the number of vectors and their dimension, an exponential speed-up over classical algorithms.

nature physics LETTERS
PUBLISHED ONLINE: 27 JULY 2014 | DOI: 10.1038/NPHYS3029

## Quantum principal component analysis

Seth Lloyd[1,2*], Masoud Mohseni[3] and Patrick Rebentrost[2]

**The usual way to reveal properties of an unknown quantum state, given many copies of a system in that state, is to perform measurements of different observables and to analyse the results statistically[1,2]. For non-sparse but low-rank quantum states, revealing eigenvectors and corresponding eigenvalues in classical form scales super-linearly with the system dimension[3–6]. Here we show that multiple copies of a quantum system with density matrix $\rho$ can be used to construct the unitary transformation $e^{-i\rho t}$. As a result, one can perform quantum principal component analysis of an unknown low-rank density matrix, revealing in quantum form the eigenvectors corresponding to the large eigenvalues in time exponentially faster than any existing algorithm. We discuss applications to data analysis, process tomography and state discrimination.**

Quantum tomography is the process of discovering features of

significant advantages for quantum tomography. Moreover, it allows us to perform quantum PCA (qPCA) of an unknown low-rank density matrix to construct the eigenvectors corresponding to the large eigenvalues of the state (the principal components) in time $O(\log d)$, an exponential speed-up over existing algorithms. We also show how qPCA can provide new methods of state discrimination and cluster assignment.

Suppose that one is presented with $n$ copies of $\rho$. A simple trick allows one to apply the unitary transformation $e^{-i\rho t}$ to any density matrix $\sigma$ up to $n$th order in $t$. Note that

$$\mathrm{tr}_P\, e^{-iS\Delta t}\rho\otimes\sigma\, e^{iS\Delta t} = (\cos^2\Delta t)\sigma + (\sin^2\Delta t)\rho - i\sin\Delta t\cos\Delta t[\rho,\sigma]$$
$$= \sigma - i\Delta t[\rho,\sigma] + O(\Delta t^2) \quad (1)$$

Here $\mathrm{tr}_P$ is the partial trace over the first variable and $S$ is the

## Today's Talk is about yet another Big Quantum Data approach:

nature COMMUNICATIONS

ARTICLE

Received 17 Sep 2014 | Accepted 9 Nov 2015 | Published 25 Jan 2016 DOI: 10.1038/ncomms10138 OPEN

Quantum algorithms for topological and geometric analysis of data

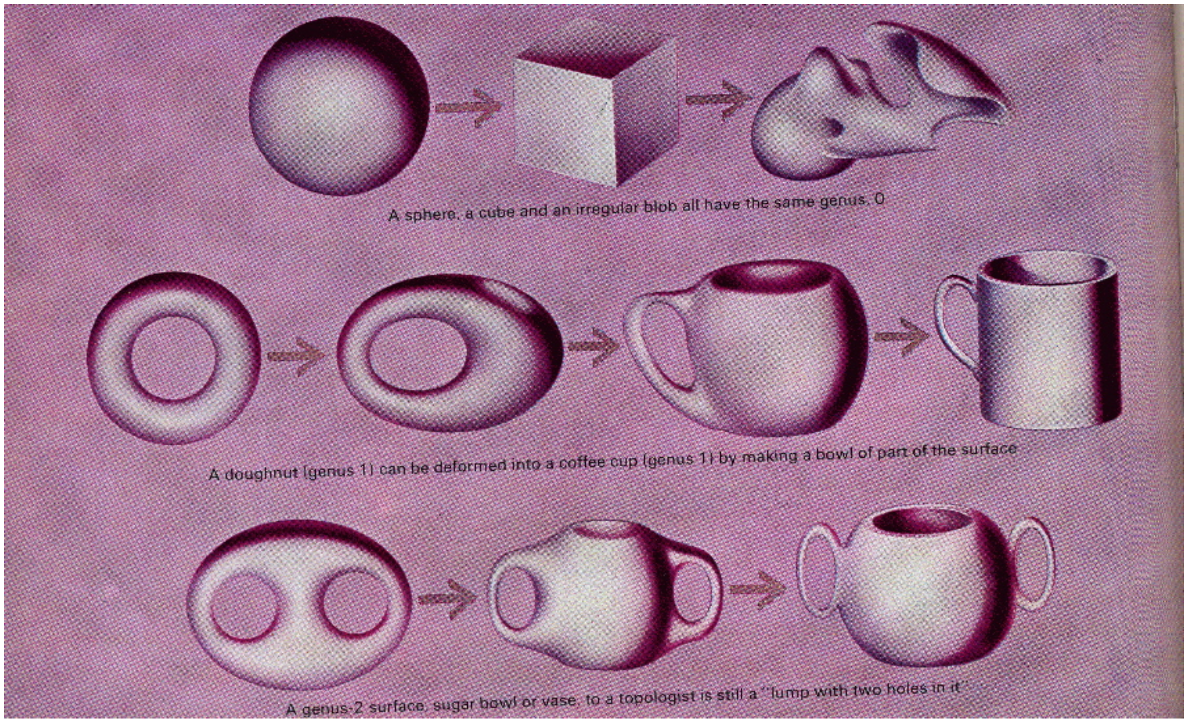
Seth Lloyd[1], Silvano Garnerone[2] & Paolo Zanardi[3]

*The Bad*

*The Good*

*The Ugly*

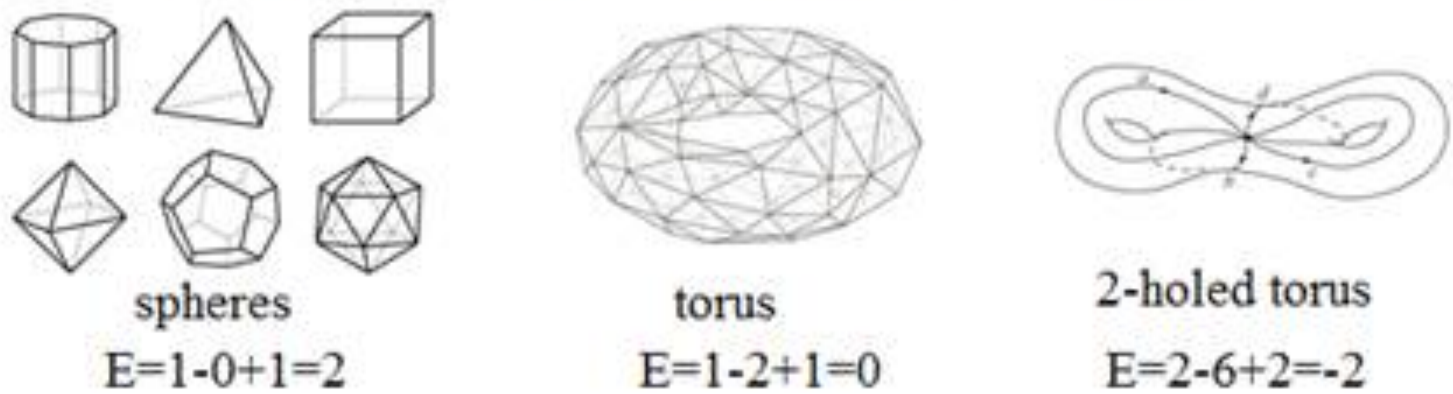
# Extracting Small Patterns out Big Data: **TOPOLOGY**




Classification of vast sets of complex objects in terms of simple topological invariants

# Topological Invariants: Euler's Characteristics

$$E = (\#\ \text{pieces}) - (\#\ \text{holes}) + (\#\ \text{voids})$$

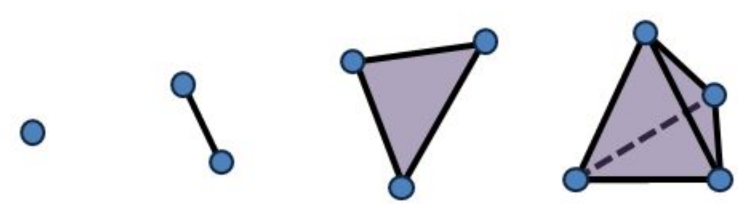



$E=2(1-g)$, $g$= genus

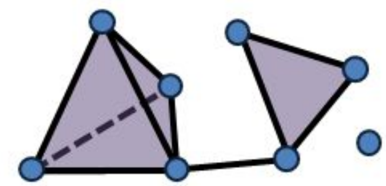
Let's refine this concept for triangulable spaces (homeomorphic to polyhedra): **Simplicial Complexes**

## Background

A **simplicial complex** is built from points, edges, triangular faces, etc.

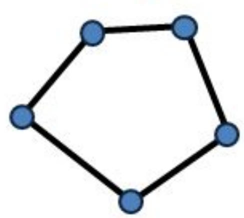


0-simplex 1-simplex 2-simplex 3-simplex (solid)

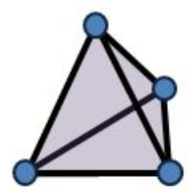


example of a simplicial complex

**Homology** counts components, holds, voids, etc.



hole



void (contains faces but empty interior)

Homology of a simplicial complex is computable via linear algebra.

# Boundary Map & Chain Complexes

$C_n$=*n-th* *Chain Group= Formal linear combinations of simplices of the complex*

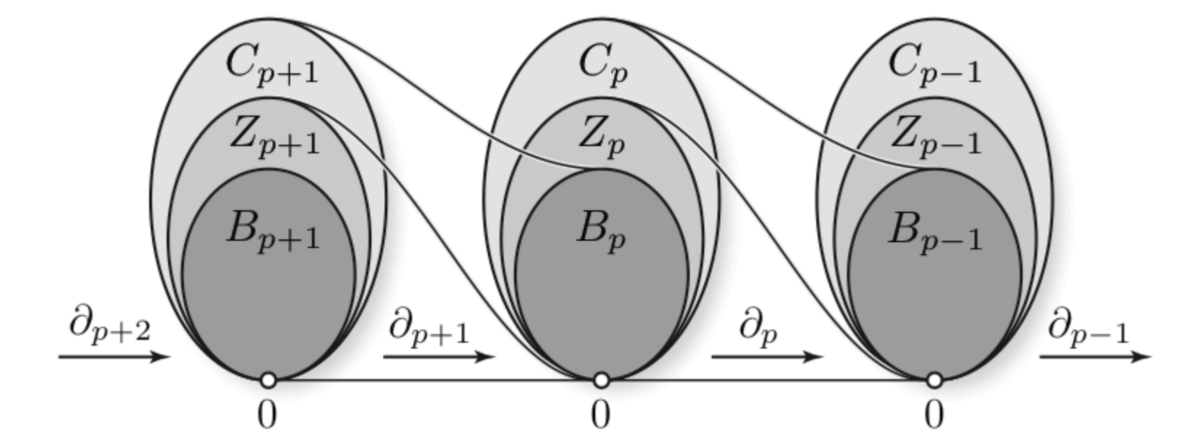
$$\partial_n \sigma_n(\Delta^n) = \sum_{k=0}^{n} (-1)^k [p_0, \cdots, p_{k-1}, p_{k+1}, \cdots p_n]$$

Boundary Map: sends a simplex to a combination of its faces

$$\cdots \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

$$\partial_n \circ \partial_{n+1} = 0_{n+1,n-1},$$

Nilpotency=boundary of boundary is 0



# Homology Groups

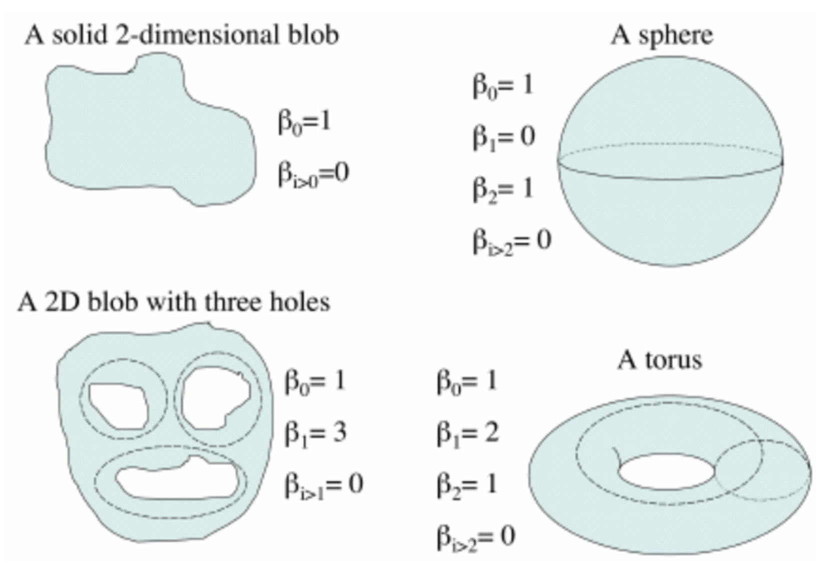
$$H_n(X) := \ker(\partial_n)/\mathrm{im}(\partial_{n+1}) = Z_n(X)/B_n(X),$$

$\beta_k$=# of generators of $H_k$= Betti Number

$\beta_0$=#of connected components,

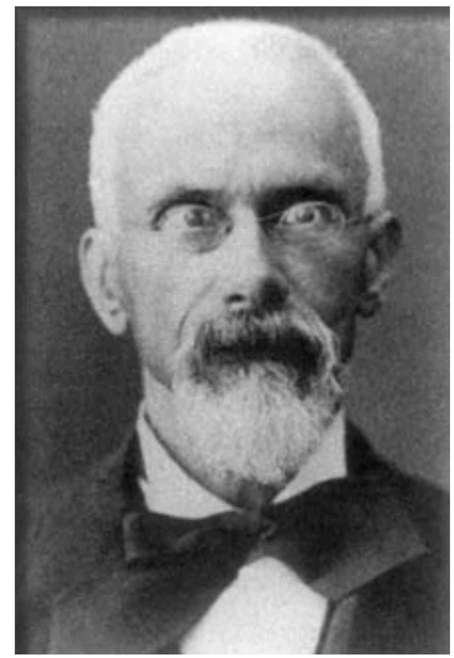
$\beta_1$=#of holes,

$\beta_2$=# voids, ……




$$\chi = \sum_{k=} (-1)^k \beta_k$$

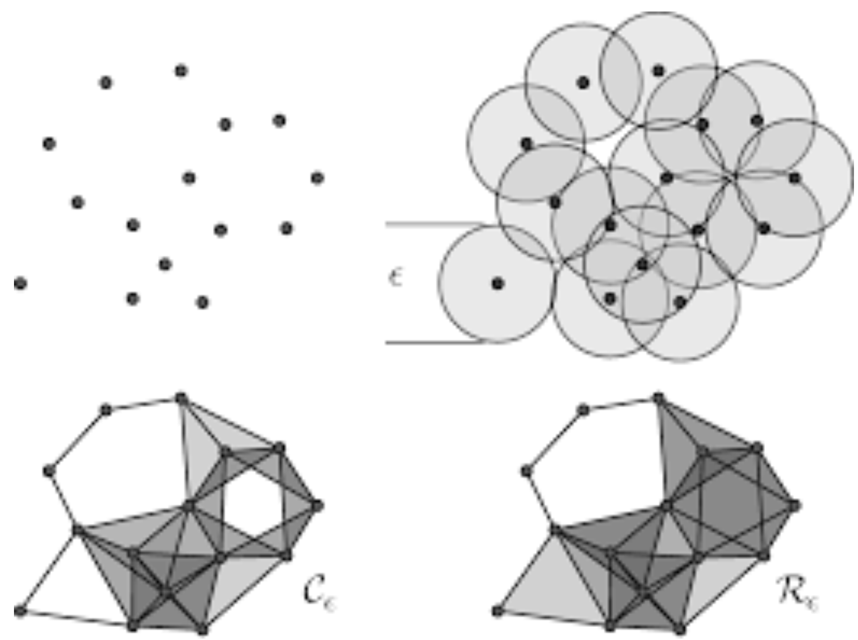
Euler's Characteristic



E. Betti, 1823-1892

# Complexes from Point Cloud Data (PCD)

Data can be represented by "clouds" of points in a high dimensional space: **how do we do topology with that?!?**



Čech complex

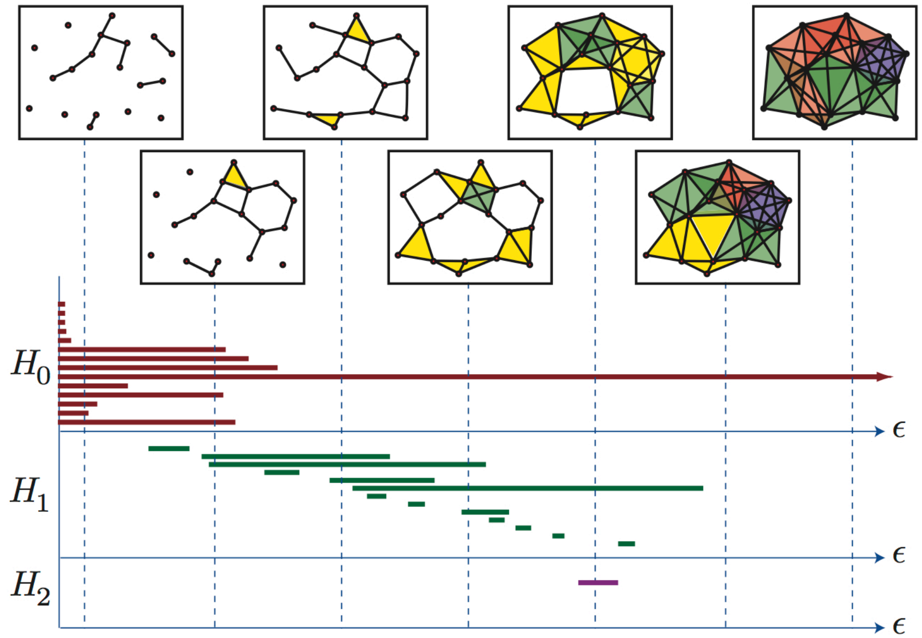
Vietoris-Rips complex

For each scale of $\boldsymbol{\varepsilon}$ one builds a simplicial complex $S_\varepsilon$ out of the PCD increasing $\boldsymbol{\varepsilon}$ makes $S_\varepsilon$ growing

Varying $\boldsymbol{\varepsilon}$ over a range of scales one obtains a family of nested simplicial complexes aka a Filtration

# Complex Filtrations and BarCodes

Tracking how Betti numbers change as function of the scale $\varepsilon$ reveals how topological features come into existence and go away as the data is analyzed at different $\varepsilon$




A topological feature that persists over many length scales can be identified with a ‘true’ feature of the structure: **Persistent Homology**

# Quantum Algorithm for Persistent Homology: The Sketch


nature COMMUNICATIONS

ARTICLE

Received 17 Sep 2014 | Accepted 9 Nov 2015 | Published 25 Jan 2016 DOI: 10.1038/ncomms10138 OPEN

Quantum algorithms for topological and geometric analysis of data

Seth Lloyd[1], Silvano Garnerone[2] & Paolo Zanardi[3]

**0)** Store (or compute) distances between data points in a Q-RAM

**1)** Fix $\boldsymbol{\varepsilon}$, construct a quantum state encoding simplicial complex at the scale $\boldsymbol{\varepsilon}$ (Grover Search Algorithm)

**2)** Find the kernel of the Laplacian to get the Betti Numbers (Quantum Phase estimation Algorithm)

**3)** Iterate over the $\boldsymbol{\varepsilon}$ and look for persistent features across scales

**How About computational complexity?!?**

# Computational Complexity: Classical vs Quantum

**Table 1 | Computational cost comparison.**

| Procedural steps | Classical cost | Quantum cost |
|---|---|---|
| Input pairwise distances, $n$ points | $O(n^2)$ bits | $O(n^2)$ bits |
| Construct simplicial complex | $O(2^n)$ ops | $O(n^2)$ ops on $O(n)$ qubits |
| Diagonalize Laplacian/find Betti numbers | $O(2^{2n} \log(1/\delta))$ ops | $O(n^5/\delta)$ quantum ops |

$\delta$ is the multiplicative accuracy to which the Betti numbers and the eigenvalues of the combinatorial Laplacian are determined. Note the trade-off between the exponential quantum speed-up and accuracy: the quantum algorithms obtain an exponential speed-up over classical algorithms but provide an accuracy that scales polynomially in $1/\delta$ rather than exponentially. This feature arises from the nature of the quantum phase estimation/matrix inversion algorithms, which obtain their exponential speed-up by estimating eigenvectors and eigenvalues using a 'pointer-variable' measurement interaction[38–40]. By contrast, classical algorithms need only keep $O(\log(1/\delta))$ bits of precision, but must perform $O(2^{2n})$ steps to diagonalize $2^n \times 2^n$ sparse matrices.

NATURE COMMUNICATIONS | 6:10138 | DOI: 10.1038/ncomms10138 | www.nature.com/naturecommunications 3

**Our Quantum algorithm provides and exponential speed-up over the classical one!**

*QUANTUM SUPREMACY......* $|\Psi\rangle$

nature COMMUNICATIONS

ARTICLE

Received 17 Sep 2014 | Accepted 9 Nov 2015 | Published 25 Jan 2016 DOI: 10.1038/ncomms10138 OPEN

Quantum algorithms for topological and geometric analysis of data

Seth Lloyd[1], Silvano Garnerone[2] & Paolo Zanardi[3]

# The Guts of the Quantum Algorithm I

Let $s_k$ a k-simplex we map it onto a quantum state $|s_k\rangle = |j_1, j_2, \ldots, j_n\rangle$ where $j_p=1$ iff p is in $s_k$

$\mathcal{H}_k^\epsilon$ | Space generated by the k-simplex states in the $\boldsymbol{\varepsilon}$-Complex, $|S_k^\epsilon|$ --dimensional

## Quantum Pipeline 1: Encoding the $\boldsymbol{\varepsilon}$-Complex

Grover's Search Algorithm:

$$|\psi\rangle_k^\epsilon = \frac{1}{\sqrt{|S_k^\epsilon|}} \sum_{s_k \in S_k^\epsilon} |s_k\rangle,$$

Takes time $O\left(n^2 (\zeta_k^\epsilon)^{-1/2}\right)$ where $\zeta_k^\epsilon$ is fraction of simplices actually present in the $\boldsymbol{\varepsilon}$-Complex; Classical time $O(2^n)$

# The Guts of the Quantum Algorithm II


nature COMMUNICATIONS

ARTICLE

Received 17 Sep 2014 | Accepted 9 Nov 2015 | Published 25 Jan 2016 DOI: 10.1038/ncomms10138 OPEN

Quantum algorithms for topological and geometric analysis of data

Seth Lloyd[1], Silvano Garnerone[2] & Paolo Zanardi[3]

**Combinatorial Hodge Theory**: Betti numbers are the dimensions of the kernels of the $\boldsymbol{\varepsilon}$-complex Laplacian operators (*0*-eigenvectors=Harmonic forms ≅ to Homology classes)

$$\Delta_k = \tilde{\partial}_k^\dagger \tilde{\partial}_k + \tilde{\partial}_{k+1} \tilde{\partial}_{k+1}^\dagger$$

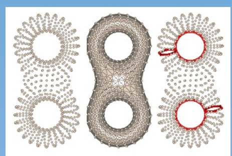
If $B_k^\epsilon = \begin{pmatrix} 0 & \tilde{\partial}_k \\ \tilde{\partial}_k^\dagger & 0 \end{pmatrix}$. then $B^\epsilon = B_1^\epsilon \oplus B_2^\epsilon \oplus \ldots \oplus B_n^\epsilon$. is the $\boldsymbol{\varepsilon}$-complex Dirac operator

$$B^{\epsilon 2} = \Delta_0 \oplus \Delta_1 \oplus \ldots \oplus \Delta_n$$

## Quantum Pipeline 2

Run the **Quantum Phase Algorithm** for $B^\epsilon$ over the uniform mixture of all simplices
Determines the dimensions of $\mathrm{Ker}\,\Delta_k$. i.e., the Betti's numbers

Classically: $O\left(\binom{n}{k}^2\right) \sim O(2^{2n})$ Quantumly (*n*-sparsity→) $O(n^5)$



# Summary & Conclusions

$|\Psi\rangle$

We live in the **BIG DATA** age AND in the **Quantum Information** Age

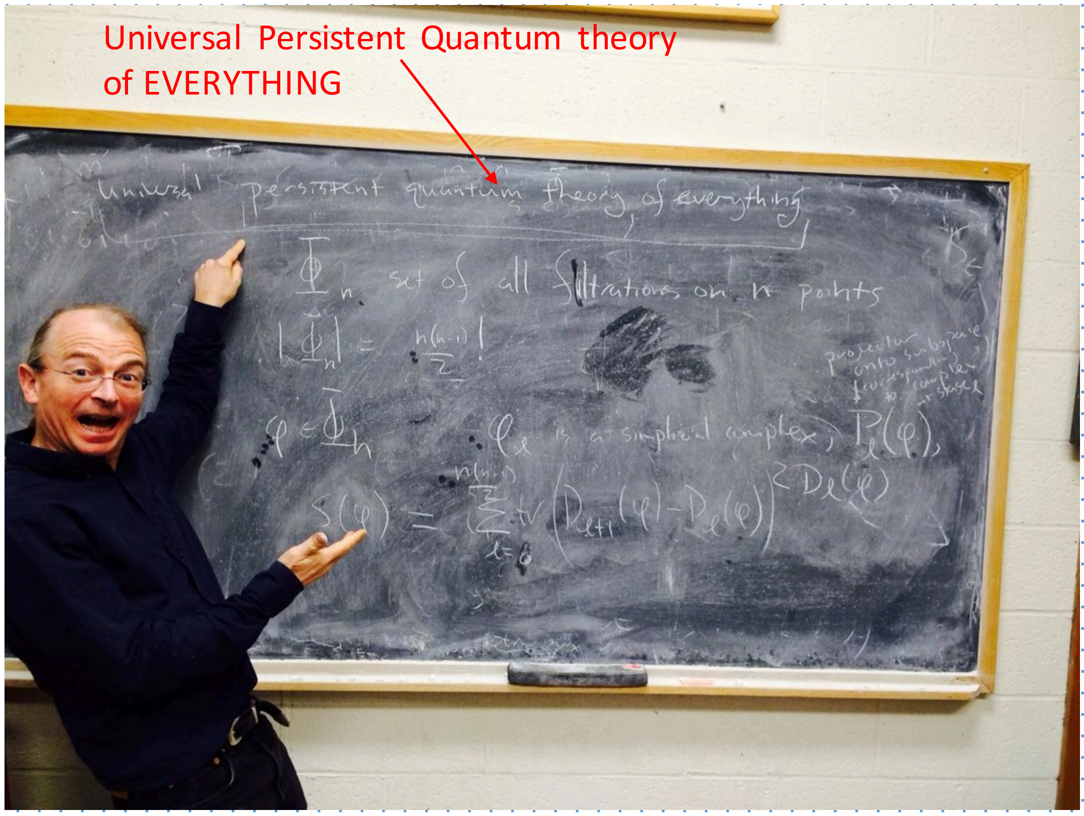
**Big Quantum Data** algorithms with exponential speedups e.g., **Q-machine learning**

We described a **topological data analysis** quantum algorithm for **persistent homology**

Quantum Information Processing in kicking in BIG time in the Big DATA scene

**we are excited!**

## Well, someone @ MIT a little over-excited, perhaps….




Seth Lloyd

# THANKS FOR THE ATTENTION!!