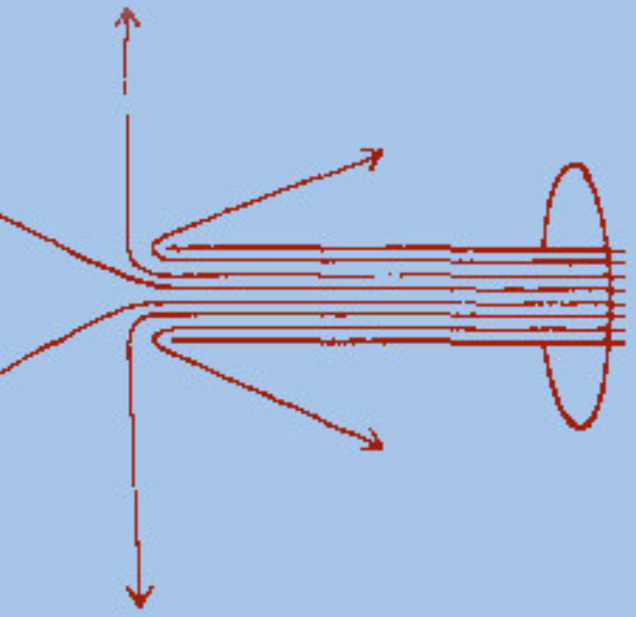
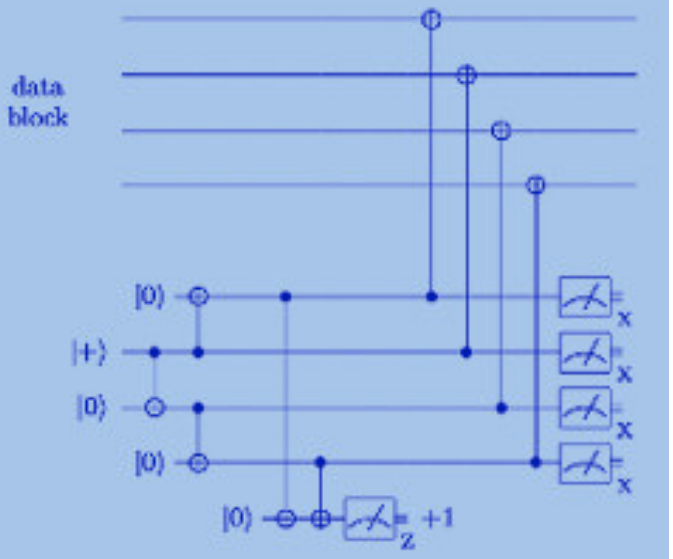
Paul Ginsparg (Cornell University) at:

From Monopoles to Fault-Tolerant Quantum Computation:





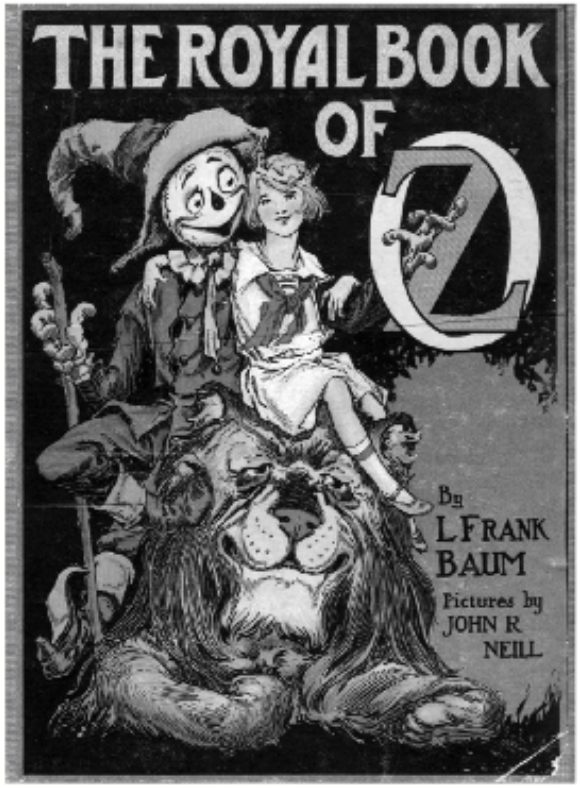
Conference in honor of John Preskill's 60th birthday

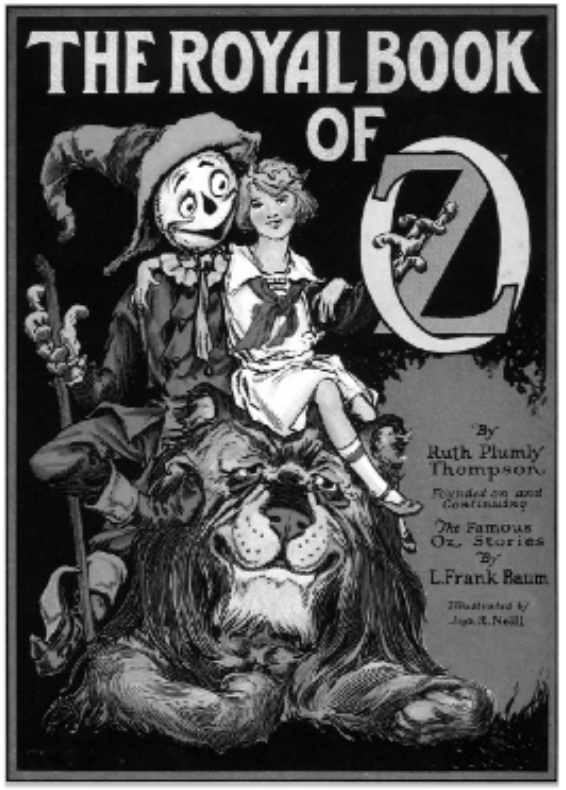
March 14-16, 2013

José Nilo G. Binongo

Chance vol 16 (2003)

Who Wrote the 15th Book of Oz?





Who is the author of *The Royal Book of Oz*?

The Royal Historians of Oz

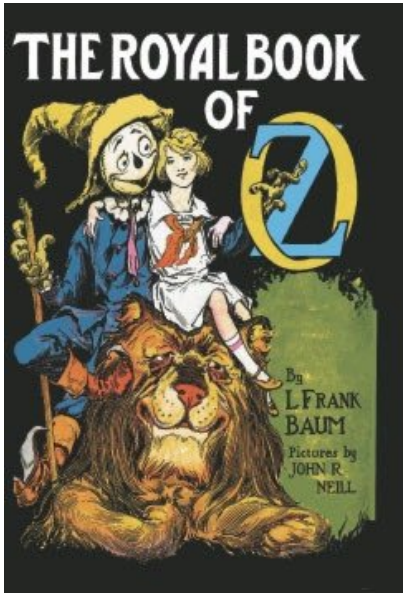
Lyman Frank Baum (1856–1919), according to Martin Gardner, was "America's greatest writer of children's fantasy." "His *Wonderful Wizard of Oz* it became a children's best seller by Christmas of 1900. Such an instant success led biographer Russell MacFall to write a chapter about 1900, calling it Baum's *annus mirabilis* (Baum and MacFall 1961).

sisting of 14 books. He consequently earned the title, "The Royal Historian of Oz." Nye describes the scenario in the early 1900s:

> *The Wizard* was apparently written with no intention of supplying a sequel; it is a complete unit, with nothing in it to anticipate a successor, much less thirteen of them. … [Baum did try] to end the series in 1910 with *The Emerald City of Oz*, but he was driven back to Oz by the demands of his readers...
>
> Finally, promising that "as long as you care to read them I shall try to write them," he resigned himself to at least one Oz story each year. (Gardner and Nye, p. 5)

Even in frail health, Baum tried hard to keep that promise. Carpenter and Shirley (1992, p. 117) write the details:

> In 1918, Frank agreed to have his gallbladder removed. By this time he had written two extra Oz



http://www.slate.com/articles/podcasts/lexicon_valley/2012/06/lexicon_valley_resolving_authorship_controversies_in_the_federalist_papers_and_the_wizard_of_oz.html

http://www.mhpbooks.com/mapping-the-oz-genome/
Mapping the Oz genome

http://www.ssc.wisc.edu/~zzeng/soc357/OZ.pdf

Who Wrote the 15th Book of Oz?

An Application of Multivariate Analysis to Authorship Attribution

J. Binongo, Chance vol 16 (2003)

L. Frank Baum wrote 14 books starting in 1900, 'til death in 1919 (published: '00, '04, '07, '08–'10, '13–'20). **1918**: gallbladder removed, had written two extra: The Magic of Oz and Glinda of Oz for reserve, then from bed finished:

#12. The Tin Woodsman of Oz (1918). Other two published posthumously:

#13. The Magic of Oz (1919)

#14. Glinda of Oz (1920, edited by his son)

19 more appeared, one per year from '21-'39, by 1939 (the movie!) there were **33** by Baum and children's author Ruth Thompson. Burning question:

#15. The royal book of Oz (1921): Baum's last or Thompson's first?

“Stopwords”

the (6.7%)	with (0.7%)	up (0.3%)	into (0.2%)	just (0.2%)
and (3.7%)	but (0.7%)	no (0.3%)	now (0.2%)	very (0.2%)
to (2.6%)	for (0.7%)	out (0.3%)	down (0.2%)	where (0.2%)
a/an (2.3%)	at (0.6%)	what (0.3%)	over (0.2%)	before (0.2%)
of (2.1%)	this/these (0.5%)	then (0.3%)	back (0.2%)	upon (0.1%)
in (1.3%)	so (0.5%)	if (0.3%)	or (0.2%)	about (0.1%)
that/those (1.0%)	all (0.5%)	there (0.3%)	well (0.2%)	after (0.1%)
it (1.0%)	on (0.5%)	by (0.3%)	which (0.2%)	more (0.1%)
not (0.9%)	from (0.4%)	who (0.3%)	how (0.2%)	why (0.1%)
as (0.7%)	one/ones (0.3%)	when (0.2%)	here (0.2%)	some (0.1%)

Figure 3. Fifty stylometric variables.

Known Works

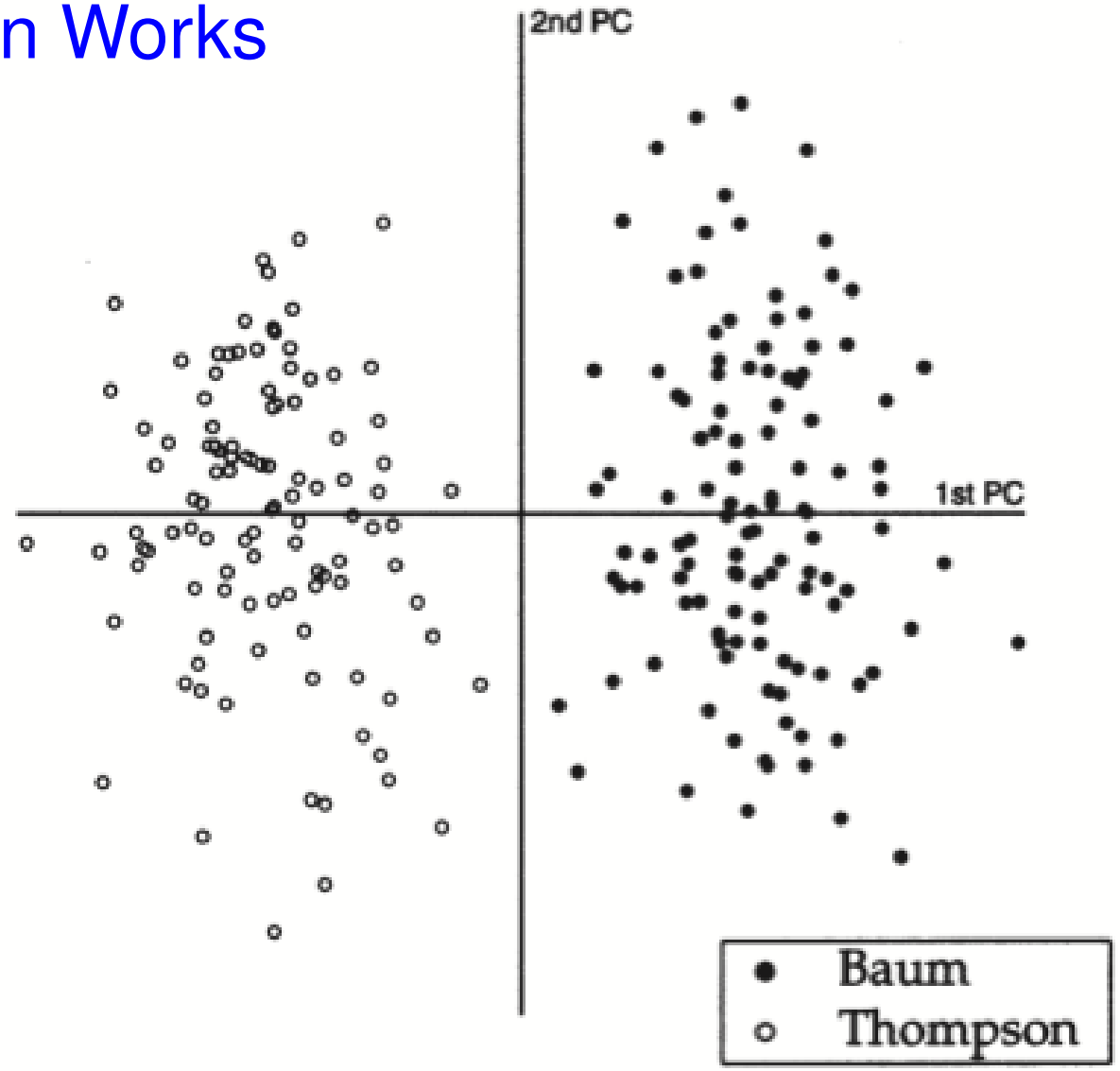
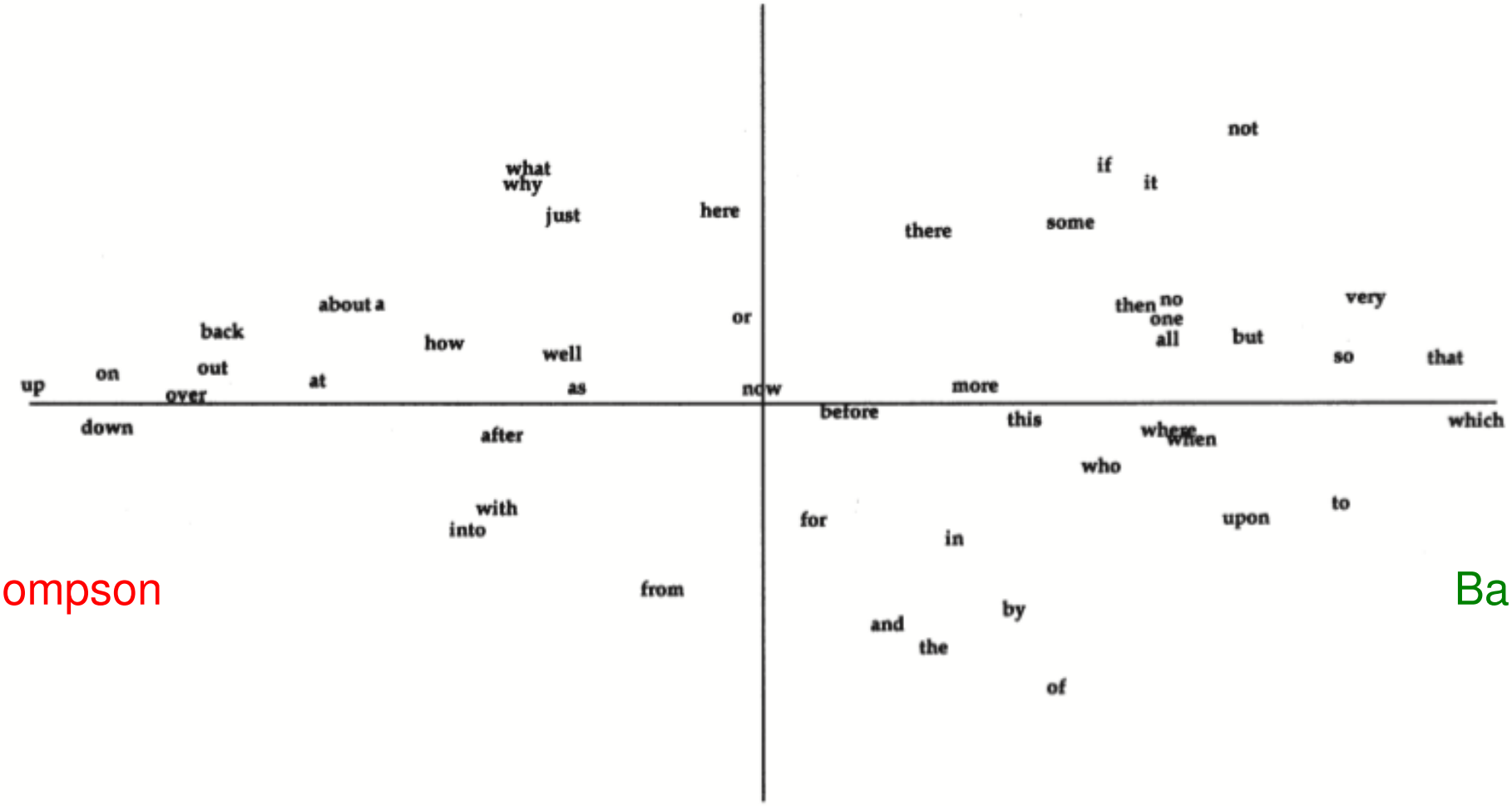


Figure 5. Baum vs. Thompson.

Stopwords



Thompson

Baum

Figure 6. Component loadings.

Known Works

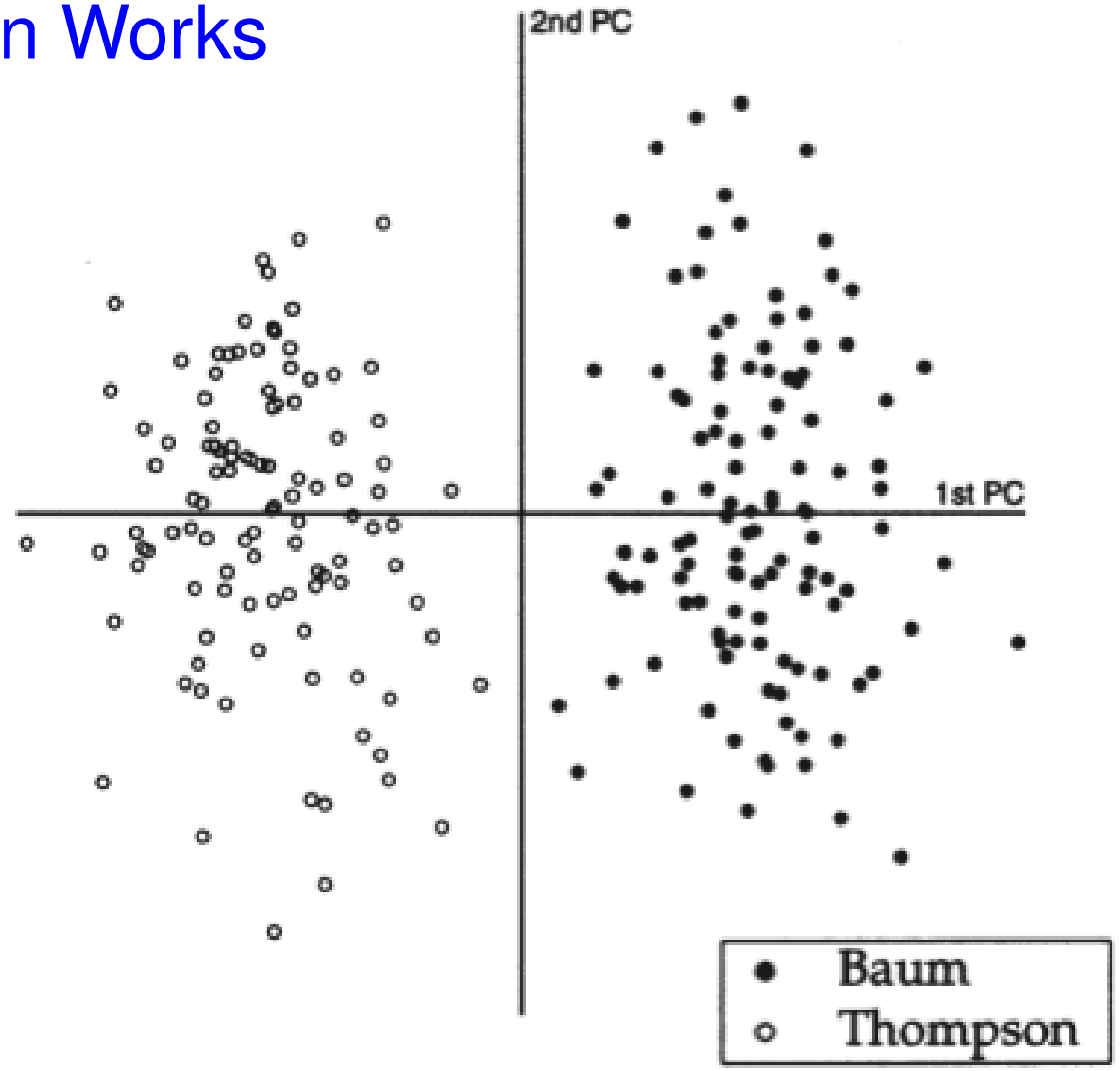
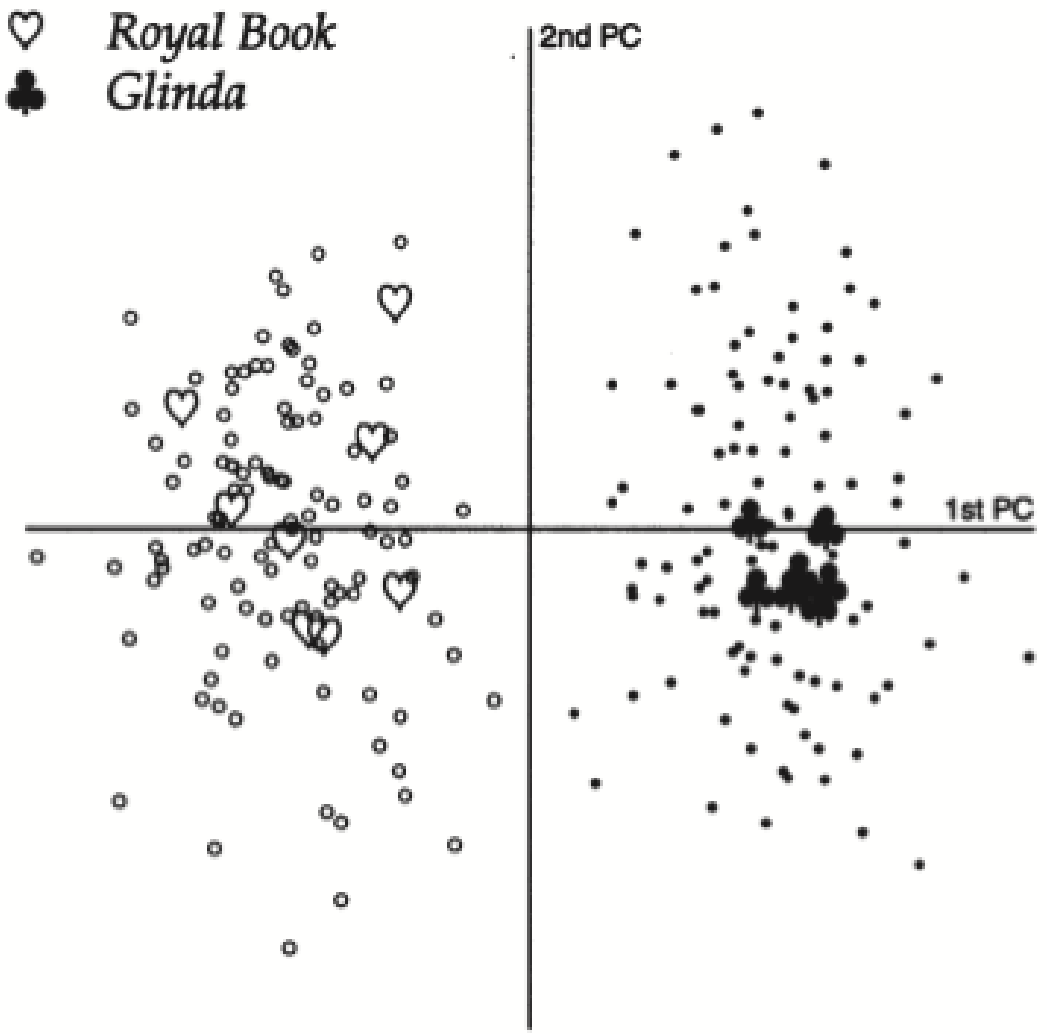


Figure 5. Baum vs. Thompson.

Redux

Thompson



Baum

Figure 10. *The Royal Book of Oz* (1921).

Other Applications

- **J.F. Burrows (1992), "Computers and the study of literature", in C. Butler (Ed.), Computers and Written Texts, pp. 167–204 (Oxford, Blackwell):**
  18th century differences between male and female writers gone by 20th century

  J.F. Burrows (1992), http://llc.oxfordjournals.org/content/7/2/91.full.pdf
  "Not Unless You Ask Nicely: The Interpretative Nexus between Analysis and Information", Literary and Linguistic Computing, 7(2): 91–109
  e.g., Jane Austen vs. Henry James

- **"The Federalist Revisited: New Directions in Authorship Attribution", D.I. Holmes and R.S. Forsyth (1995), Literary and Linguistic Computing, Vol 10, 111**
  http://llc.oxfordjournals.org/content/10/2/111.full.pdf
  Twelve of 85 Federalist papers ("Publius", 1788; 49–58,62,63) disputed between Hamilton and Madison (→Madison, e.g., used 'upon' 24x less often)

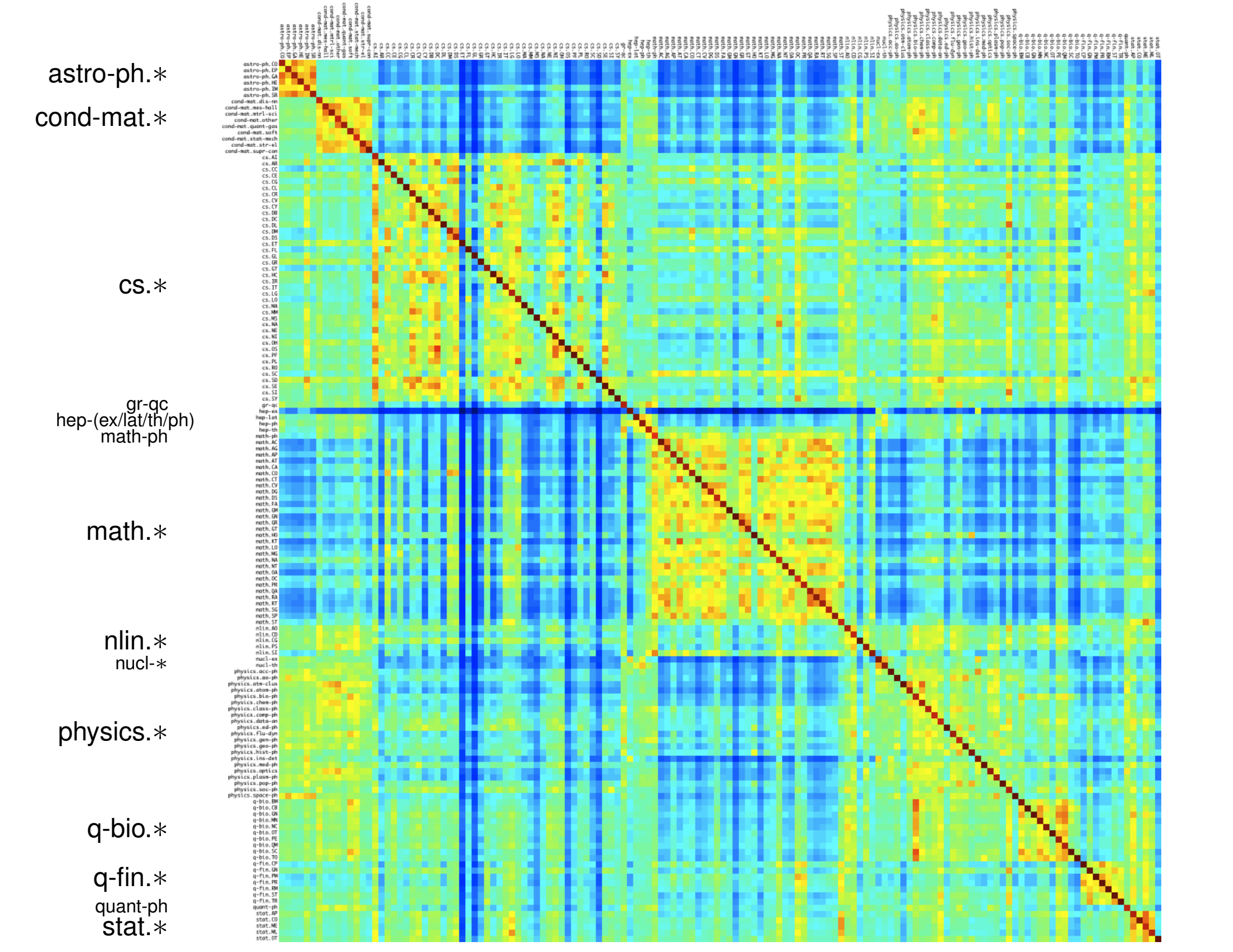
Other Applications, cont'd

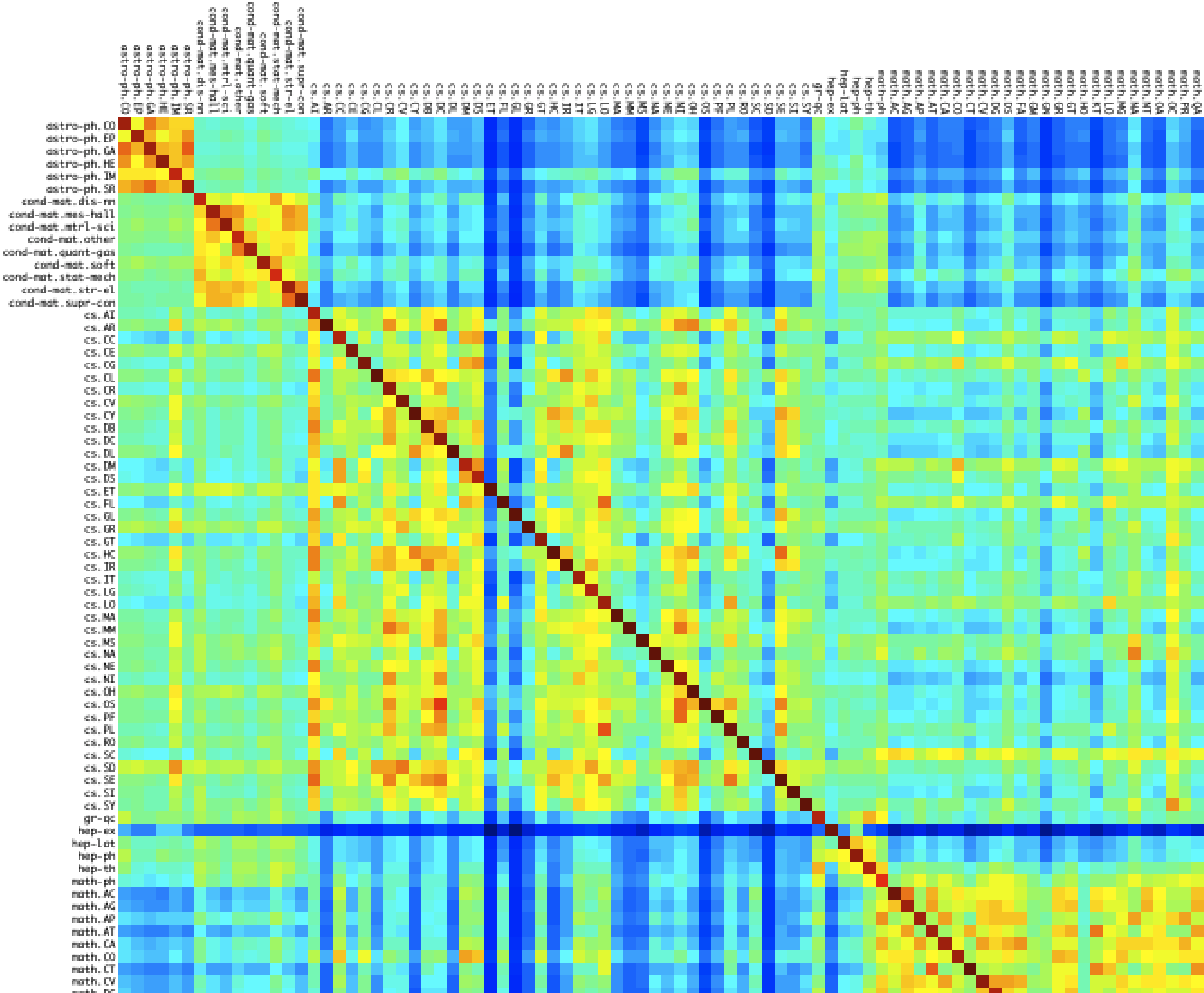
- **J.Binongo, M.Smith (1999), http://llc.oxfordjournals.org/content/14/4/445.full.pdf**
  "The Application of Principal Component Analysis to Stylometry",
  First two acts of Pericles (Shakespeare vs George Wilkins).

  J.Binongo, M.Smith (1999), J. Applied Statistics, Vol. 26, No. 7, 1999, 781- 787
  "A bridge between statistics and literature: the graphs of Oscar Wilde's literary genres"
  http://www.tandfonline.com/doi/pdf/10.1080/02664769922025
  Distinguish between Oscar Wilde plays and essays.

  J.Binongo, M.Smith (2000), The Mathematics Teacher , 93: 338-344, 2000.
  "Project Jacobean: A Mathematical Exploration of a Literary Era"
  http://www.nctm.org/publications/article.aspx?id=18016
  Jacobean dramatist Thomas Middleton likely author of (anonymous) play
  "The Revenger's Tragedy" , not William Tourneur.

- **Other languages, e.g.:**
  "Principal component analysis for authorship attribution", A.Jamak, A.Savatić, M.Can,
  Business Systems Research,Vol. 3 (Sep 2012), http://hrcak.srce.hr/file/128421
  Texts of three 20th century Bosnian authors


astro-ph.∗
cond-mat.∗
cs.∗
gr-qc
hep-(ex/lat/th/ph)
math-ph
math.∗
nlin.∗
nucl-∗
physics.∗
q-bio.∗
q-fin.∗
quant-ph
stat.∗


astro-ph.CO
astro-ph.EP
astro-ph.GA
astro-ph.HE
astro-ph.IM
astro-ph.SR
cond-mat.dis-nn
cond-mat.mes-hall
cond-mat.mtrl-sci
cond-mat.other
cond-mat.quant-gas
cond-mat.soft
cond-mat.stat-mech
cond-mat.str-el
cond-mat.supr-con
cs.AI
cs.AR
cs.CC
cs.CE
cs.CG
cs.CL
cs.CR
cs.CV
cs.CY
cs.DB
cs.DC
cs.DL
cs.DM
cs.DS
cs.ET
cs.FL
cs.GL
cs.GR
cs.GT
cs.HC
cs.IR
cs.IT
cs.LG
cs.LO
cs.MA
cs.MM
cs.MS
cs.NA
cs.NE
cs.NI
cs.OH
cs.OS
cs.PF
cs.PL
cs.RO
cs.SC
cs.SD
cs.SE
cs.SI
cs.SY
gr-qc
hep-ex
hep-lat
hep-ph
hep-th
math-ph
math.AC
math.AG
math.AP
math.AT
math.CA
math.CO
math.CT
math.CV
math.DG
math.DS
math.FA
math.GM
math.GN
math.GR
math.GT
math.HO
math.KT
math.LO
math.MG
math.MP
math.NT
math.OA
math.OC
math.PR
math.QA

Hard Work

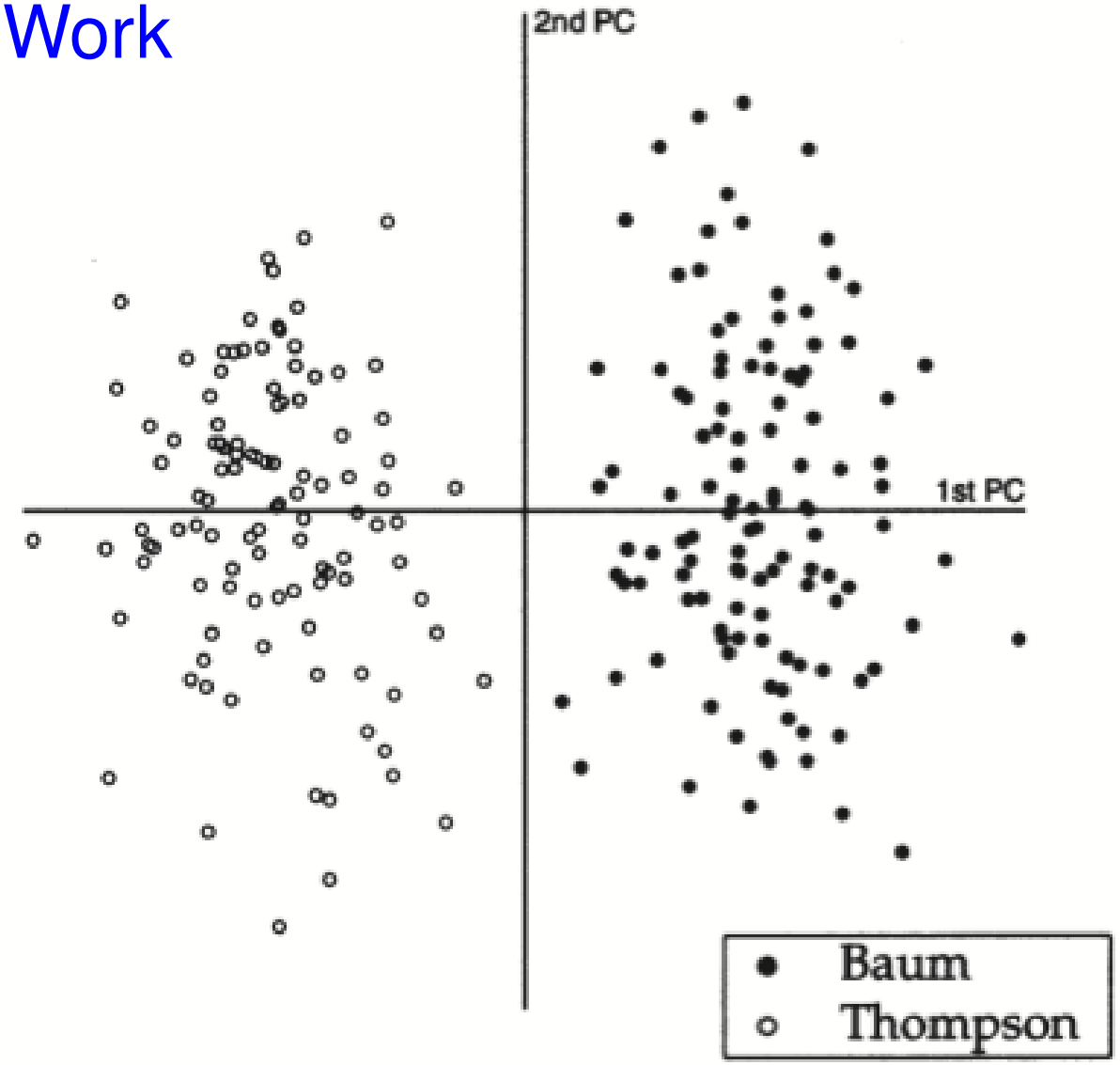
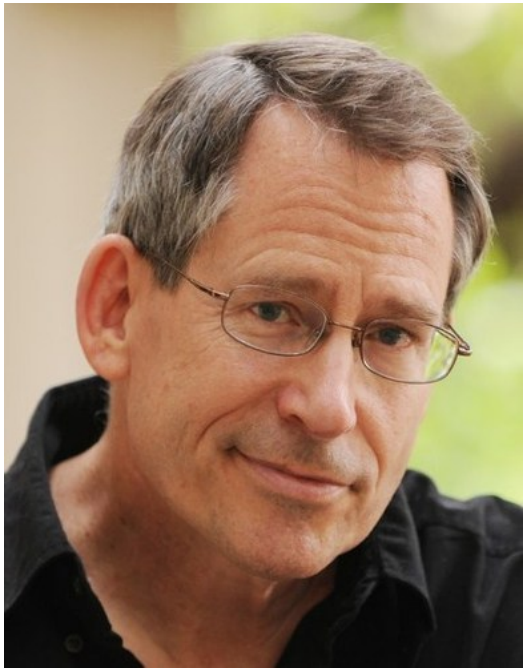


Figure 5. Baum vs. Thompson.

Raw Data I



J. Preskill:

hep-ph/9206216 11831

hep-th/9209058 7336

quant-ph/9705031 13486

quant-ph/9705032 9946

quant-ph/9712048 22217

quant-ph/9904022 5427

quant-ph/0010098 5824

1203.5813 7975

1207.6131 6173

Raw Data II

A. Kitaev:

quant-ph/9511026 10744

quant-ph/9707021 12656

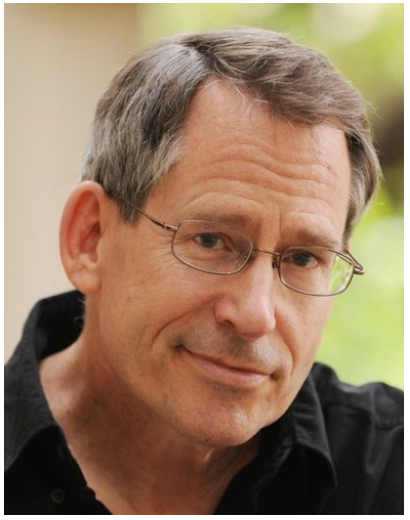
cond-mat/0010440 4749

cond-mat/0506438 33687

math/0606562 18240

cond-mat/0609441 2845

0901.2686 7298



J. Preskill:

hep-ph/9206216 11831

hep-th/9209058 7336

quant-ph/9705031 13486

quant-ph/9705032 9946

quant-ph/9712048 22217

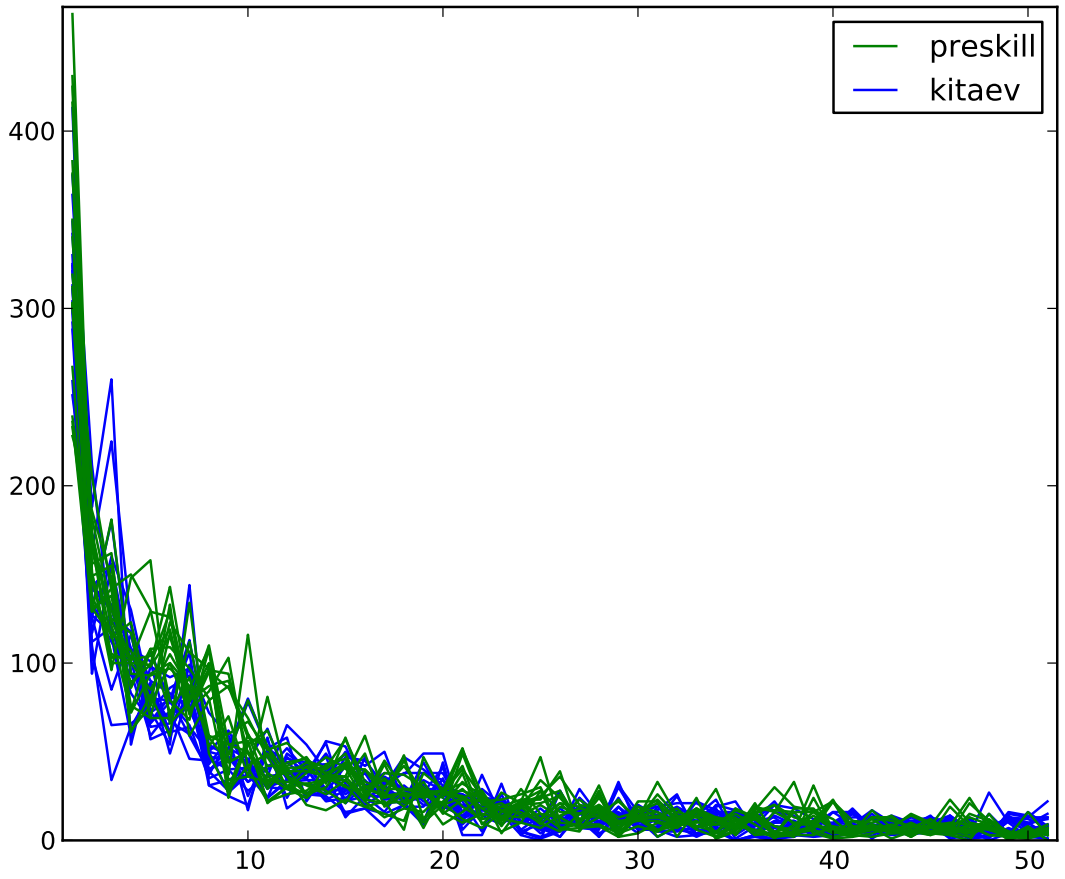
quant-ph/9904022 5427

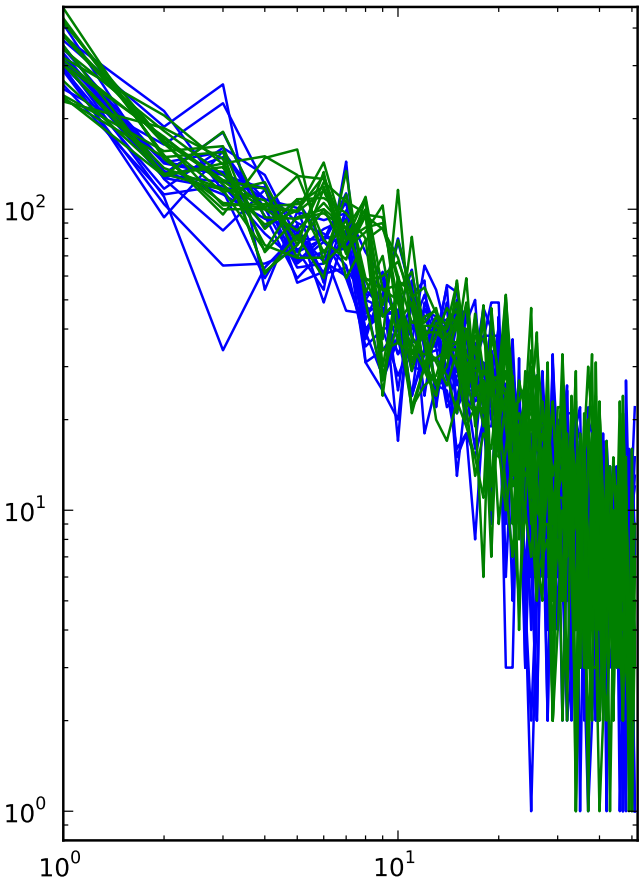
quant-ph/0010098 5824

1203.5813 7975

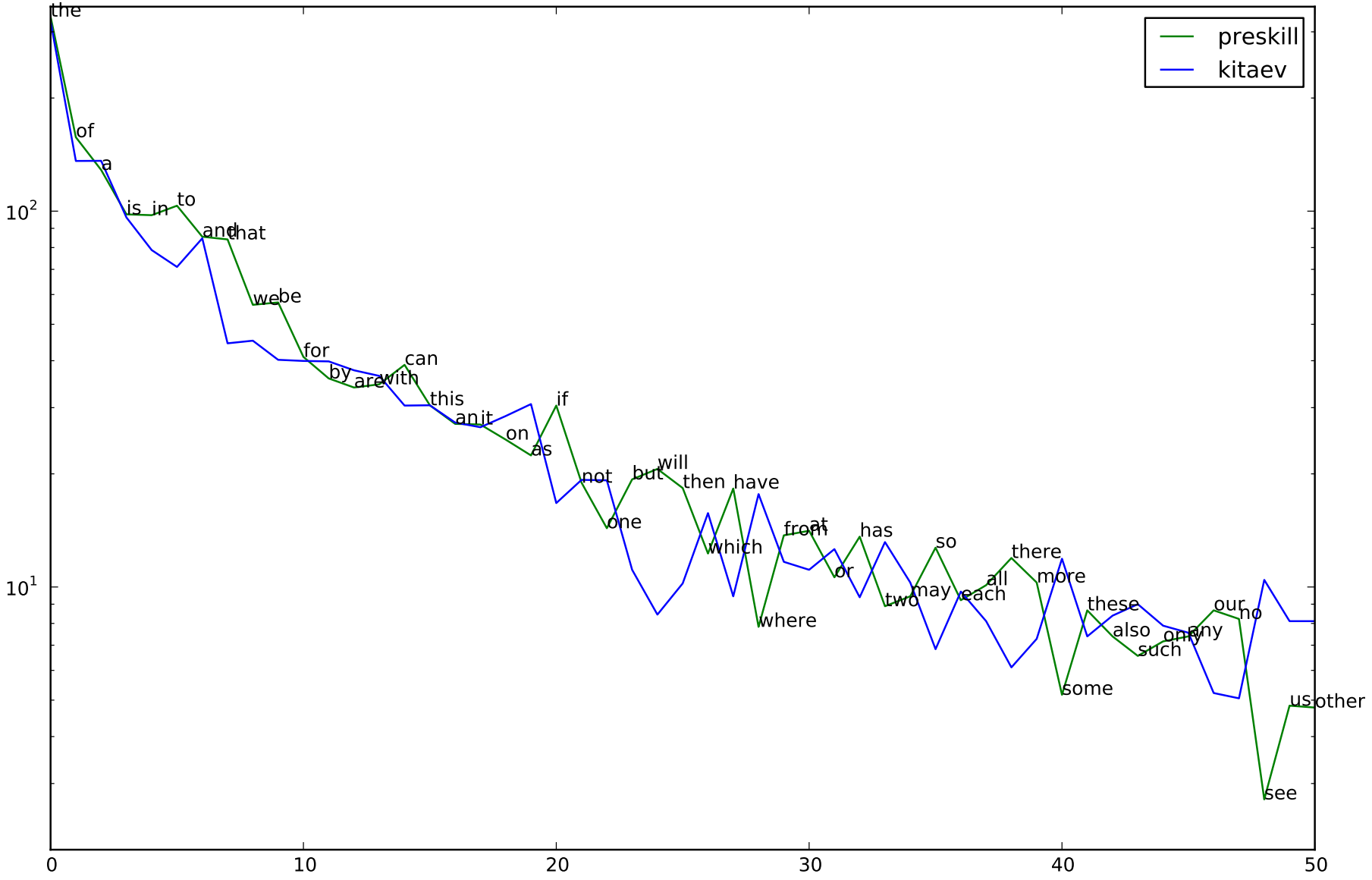
1207.6131 6173

5000 Word Blocks





Averages (10% stopword depletion)



Singular Value Decomposition

$$M = U\Sigma V^T$$

(generalizes $M = O\Lambda O^T$ to non-symmetric, non-square matrices)

- **weather data**
- **document – word (LSA, 1988)**
- **stock data**
- **genomic data**
- **apple itunes genius**
- **microarray data**
- **netflix challenge (500k $\times$ 17k)**
- . . .

a.k.a. Schmidt decomposition

$$M = U\Sigma V^\dagger$$

(generalizes $M = U\Lambda U^\dagger$)

Familiar to physicists as the Schmidt decomposition

$$|\psi\rangle = M_{ij}|\phi_A^i\rangle \otimes |\phi_B^j\rangle = \sum_i \sigma_i |\psi_A^i\rangle \otimes |\psi_B^i\rangle$$

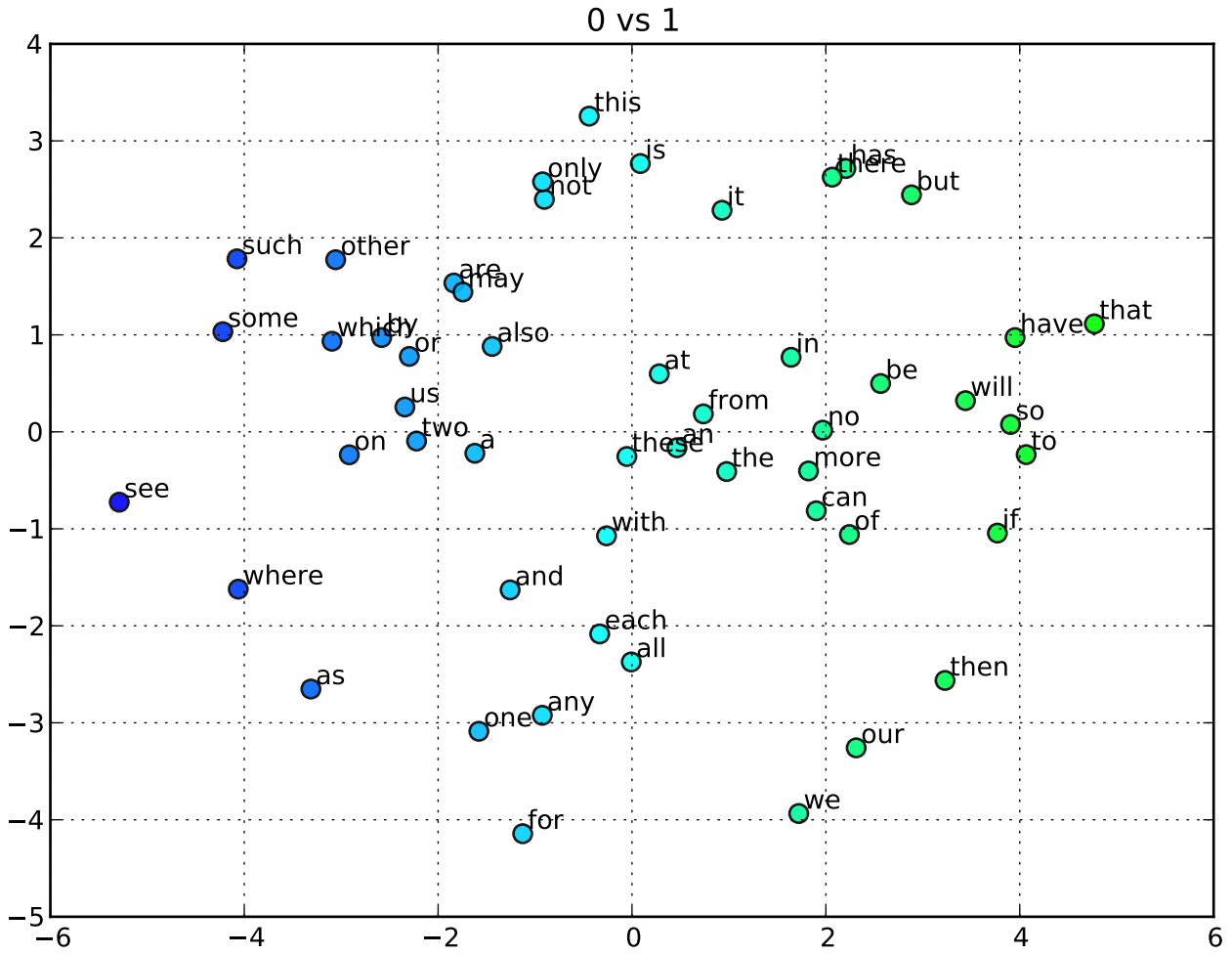
where orthonormal bases: $\langle\psi_A^i|\psi_A^j\rangle = \langle\psi_B^i|\psi_B^j\rangle = \delta_{ij}$
(components correspond to columns of U and V).

With $\sigma_i = \exp(-\xi_i/2)$, entanglement spectrum "energy levels" ξ_i give more info than entanglement entropy $S = \sum_i \xi_i \exp(-\xi_i)$ (a single number, thermodynamic entropy at $T = 1$) , and probe topological order of ground state (Li/Haldane, arXiv:0805.0332)

PCA / dimensional reduction

- $M = \mathbf{U}\Sigma V^T$ **(** M **= 36**$\times$**51 "document"–word matrix)**
- **Each of the 51 words is represented by a point in a 36-dimensional space, with vector components equal to the number of times it occurs in each of those 36 blocks of text**
- **Project points to 2d plane where they have maximal variance, i.e., maximal summed spread from the origin (since their euclidean length is the standard deviation)**
- **That plane in the 36-dim space is specified by the two left eigenvectors with largest eigenvalues**
- **Plot the 2d projections**

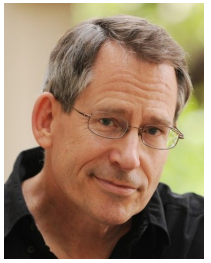
Kitaev/Preskill?

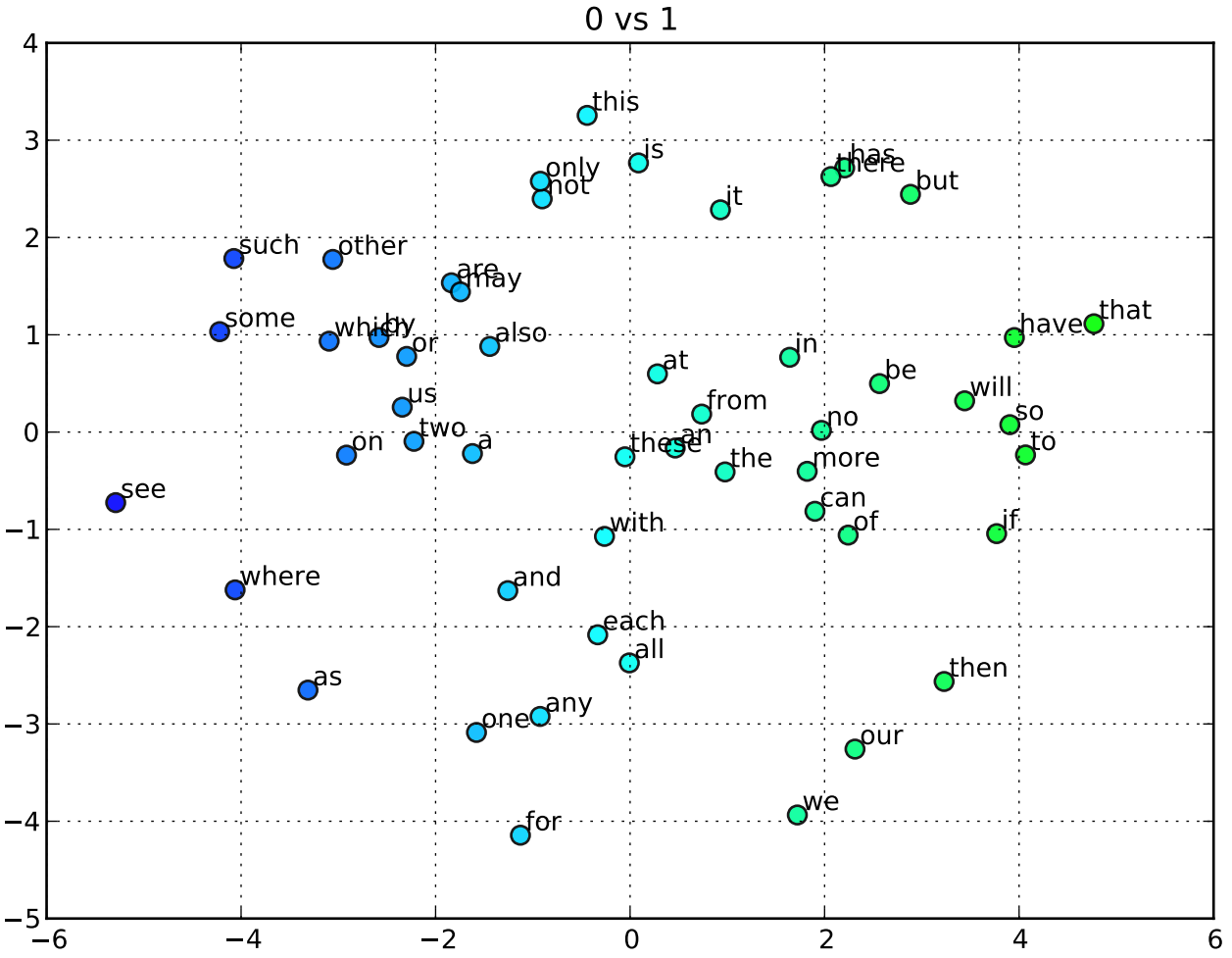


“Component Loadings”

Kitaev

Preskill





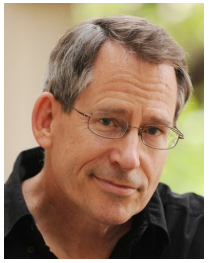
“Component Loadings”

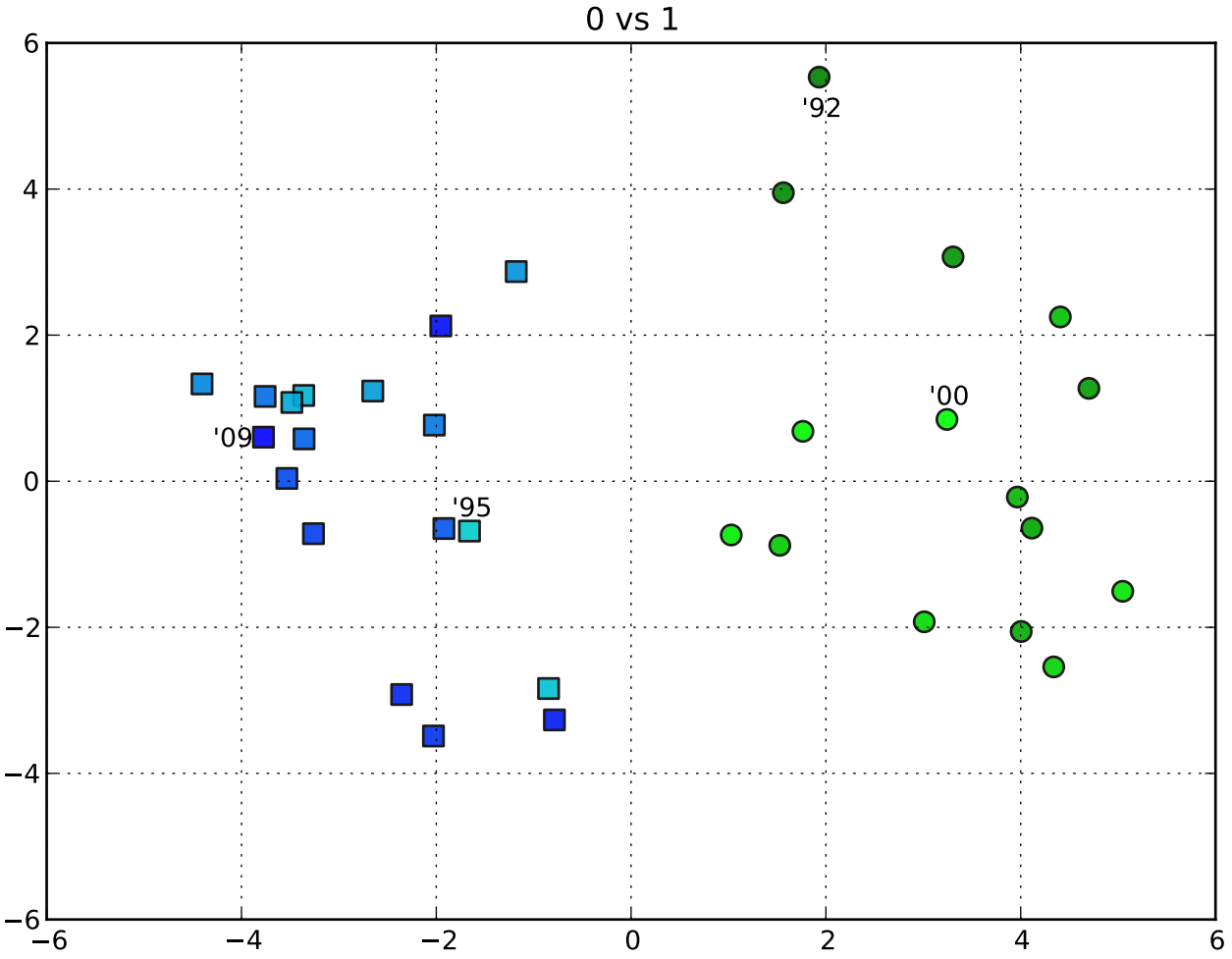
PCA / dimensional reduction

- $M = U\Sigma\mathbf{V^T}$ (M = 36$\times$51 "document"–word matrix)
- Each of the thirty-six 5000 word blocks is represented by a point in a 51-dimensional space, with vector components equal to the number of times each of the 51 words occurs
- Project points to 2d plane where they have maximal variance, i.e., maximal summed spread from the origin (since their euclidean length is the standard deviation)
- That plane in the 51-dim space is specified by the two right eigenvectors with largest eigenvalues
- Plot the 2d projections

Kitaev

Preskill





So Who Wrote hep-th/0510092?

Topological entanglement entropy

Alexei Kitaev, John Preskill

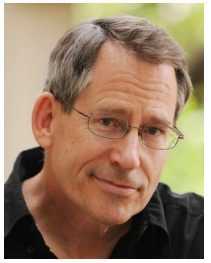
(Submitted on 11 Oct 2005 (v1), last revised 23 Jan 2006 (this version, v2))

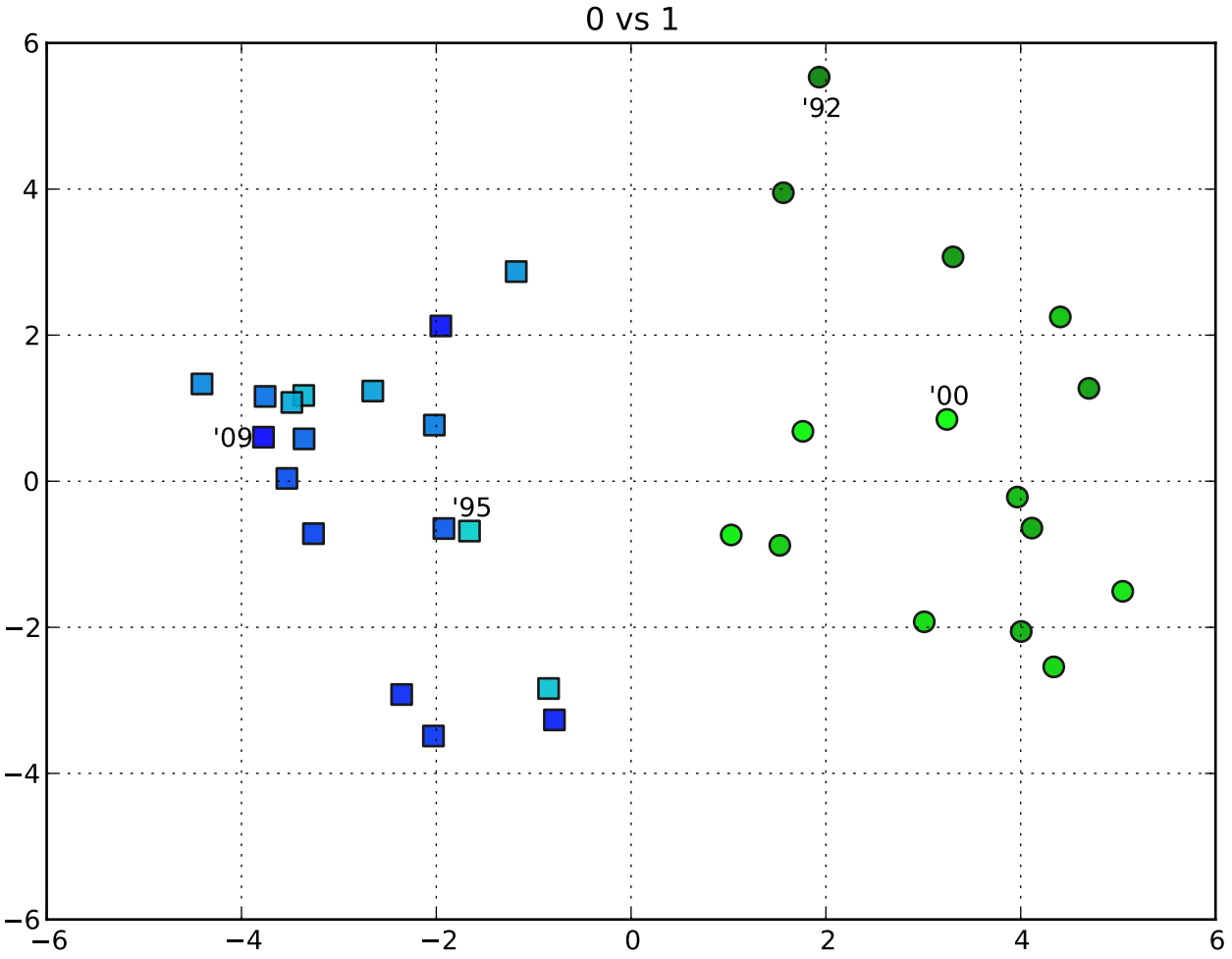
We formulate a universal characterization of the many-particle quantum entanglement in the ground state of a topologically ordered two-dimensional medium with a mass gap. We consider a disk in the plane, with a smooth boundary of length L, large compared to the correlation length. In the ground state, by tracing out all degrees of freedom in the exterior of the disk, we obtain a marginal density operator \rho for the degrees of freedom in the interior. The von Neumann entropy S(\rho) of this density operator, a measure of the entanglement of the interior and exterior variables, has the form S(\rho)= \alpha L -\gamma + ..., where the ellipsis represents terms that vanish in the limit L\to\infty. The coefficient \alpha, arising from short wavelength modes localized near the boundary, is nonuniversal and ultraviolet divergent, but -\gamma is a universal additive constant characterizing a global feature of the entanglement in the ground state. Using topological quantum field theory methods, we derive a formula for \gamma in terms of properties of the superselection sectors of the medium.

Comments:	4 pages, 3 eps figures. v2: reference added
Subjects:	**High Energy Physics - Theory (hep-th)**; Strongly Correlated Electrons (cond-mat.str-el); Quantum Physics (quant-ph)
Journal reference:	Phys.Rev.Lett. 96 (2006) 110404
DOI:	10.1103/PhysRevLett.96.110404
Report number:	CALT-68-2578
Cite as:	**arXiv:hep-th/0510092**

Kitaev

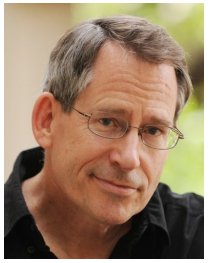
Preskill

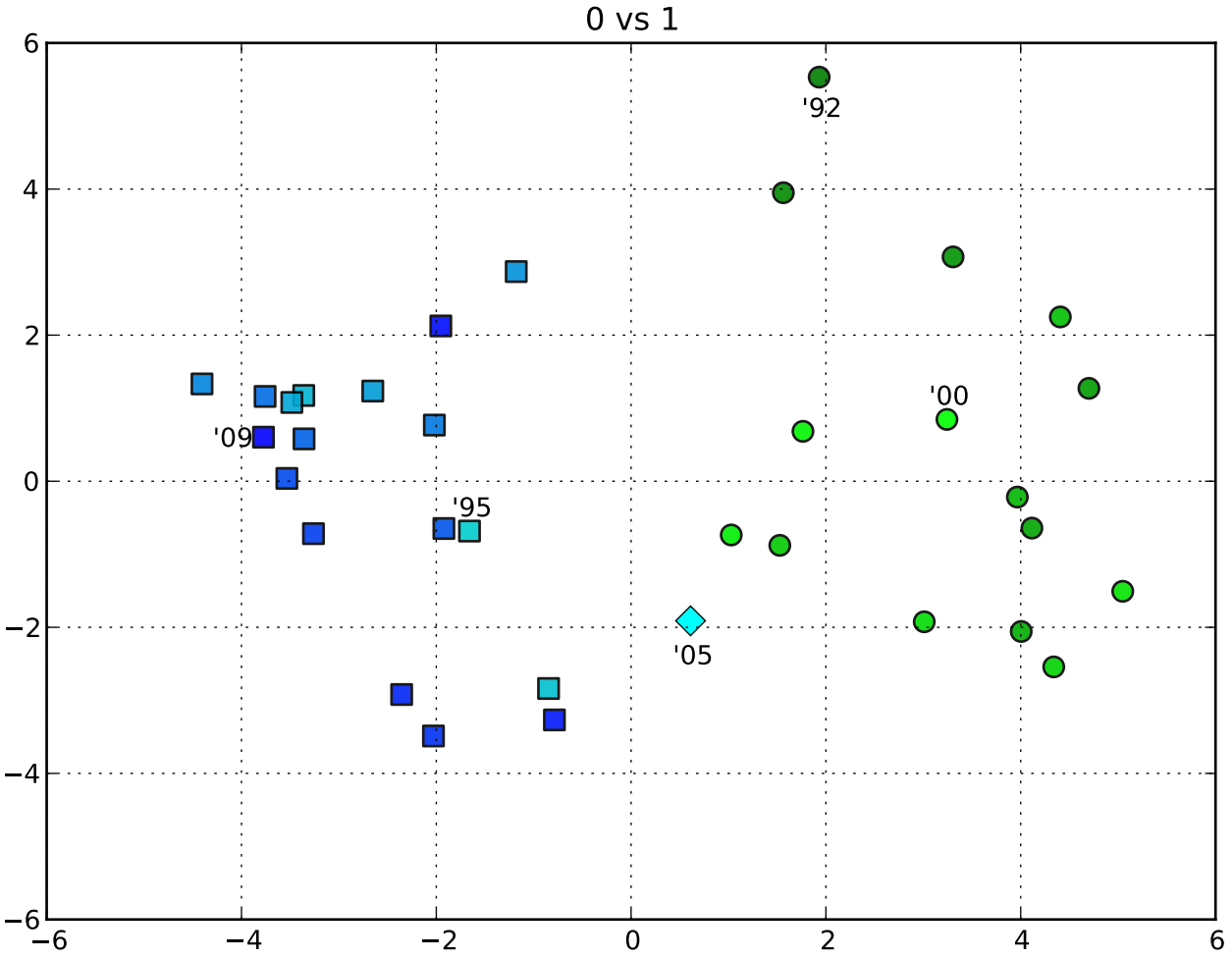




Kitaev

Preskill





Raw Data II

A. Kitaev:

quant-ph/9511026 10744

quant-ph/9707021 12656

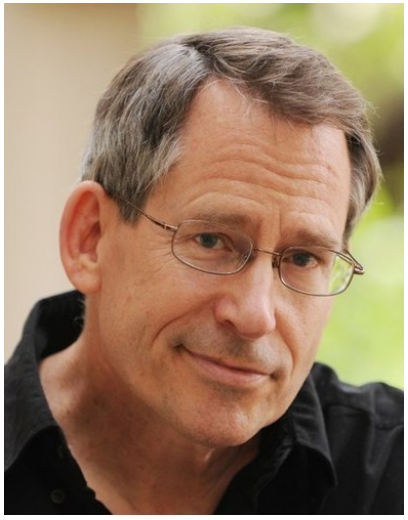
cond-mat/0010440 4749

cond-mat/0506438 33687

math/0606562 18240

cond-mat/0609441 2845

0901.2686 7298



J. Preskill:

hep-ph/9206216 11831

hep-th/9209058 7336

quant-ph/9705031 13486

quant-ph/9705032 9946

quant-ph/9712048 22217

quant-ph/9904022 5427

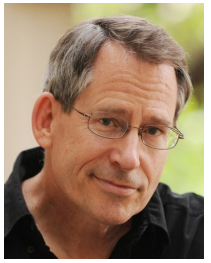
quant-ph/0010098 5824

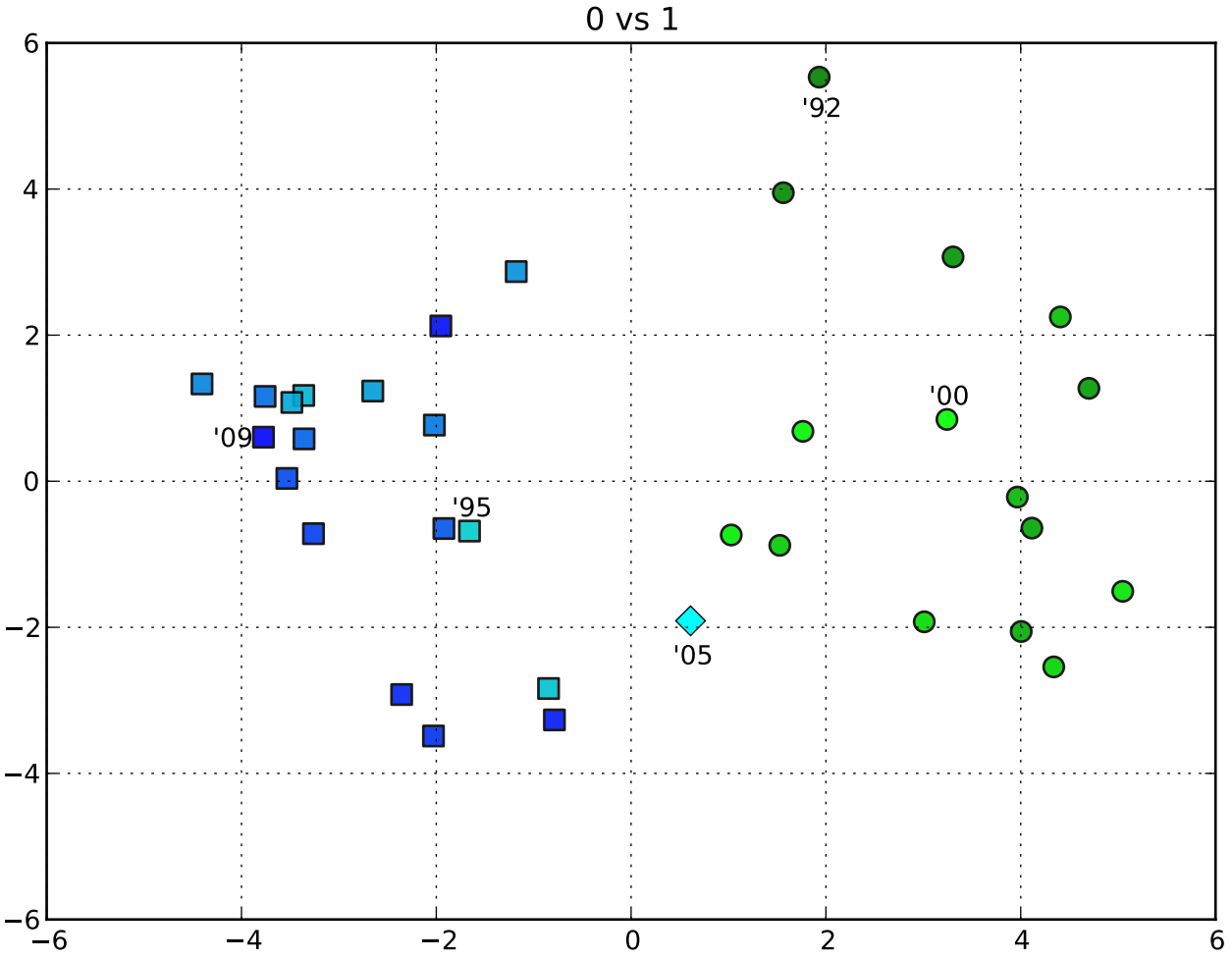
1203.5813 7975

1207.6131 6173

Kitaev

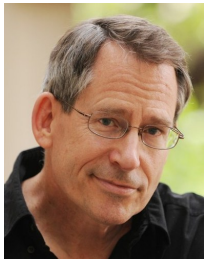
Preskill

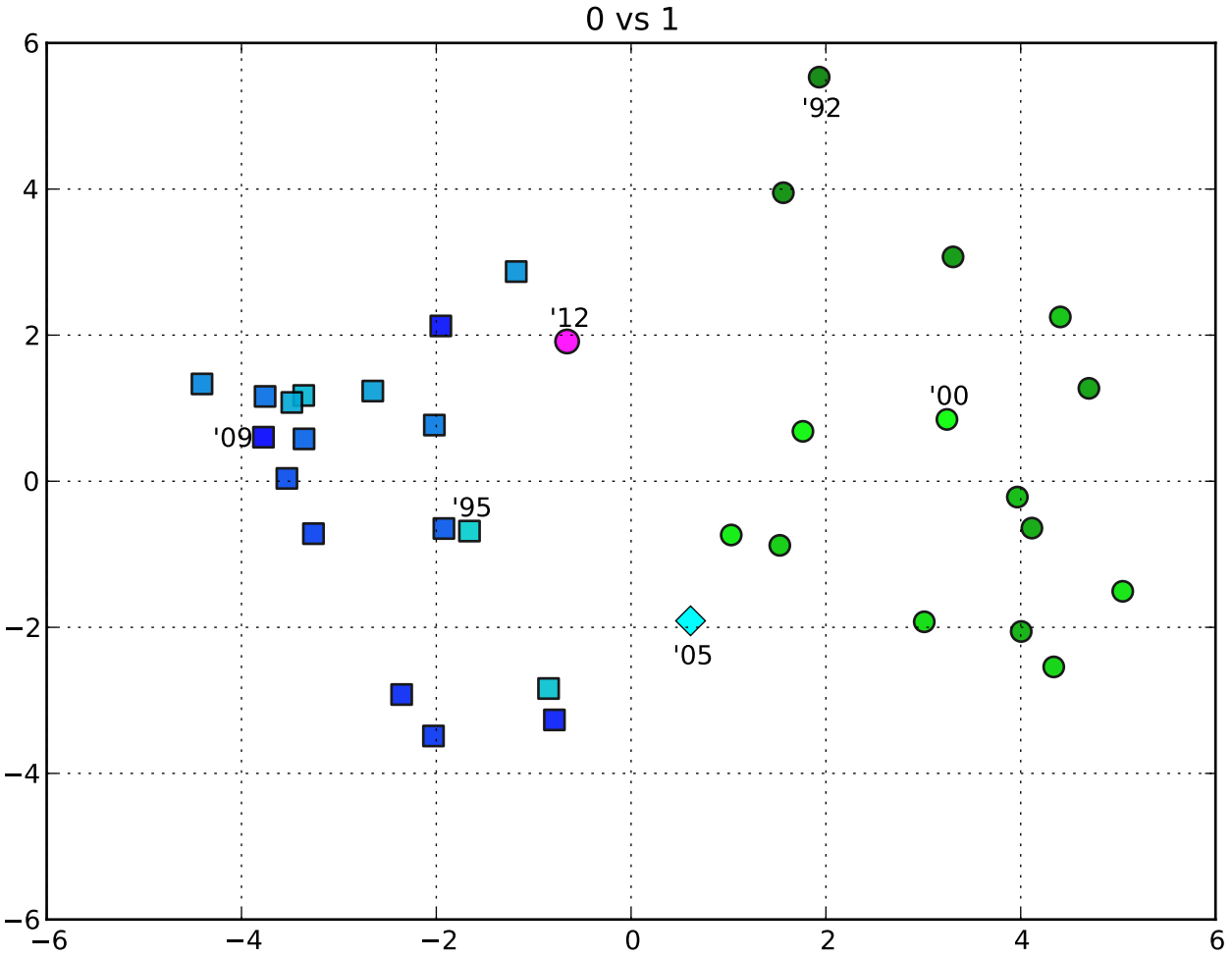




Kitaev

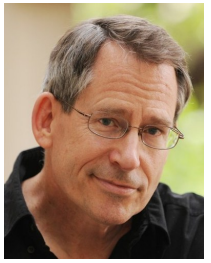
Preskill

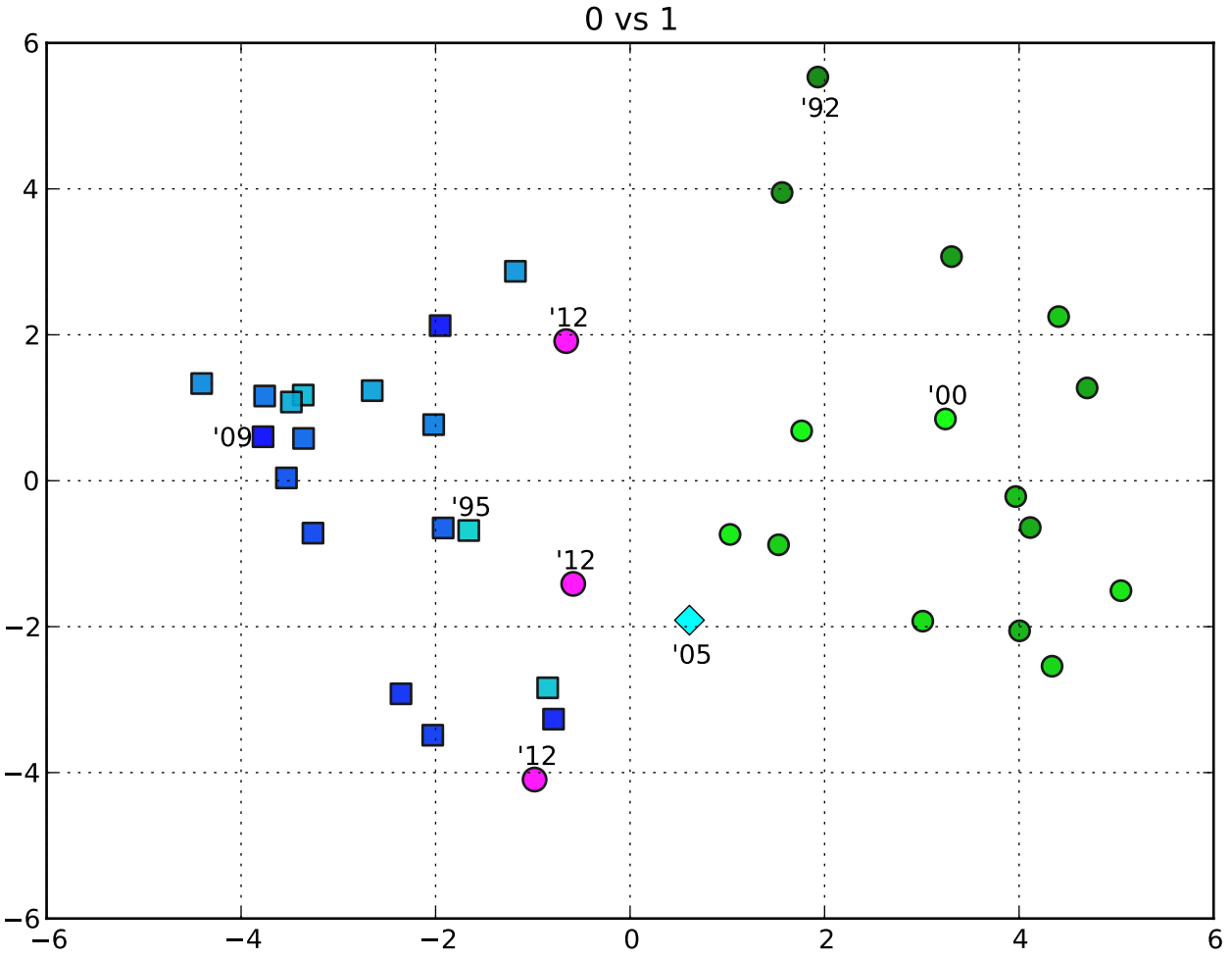




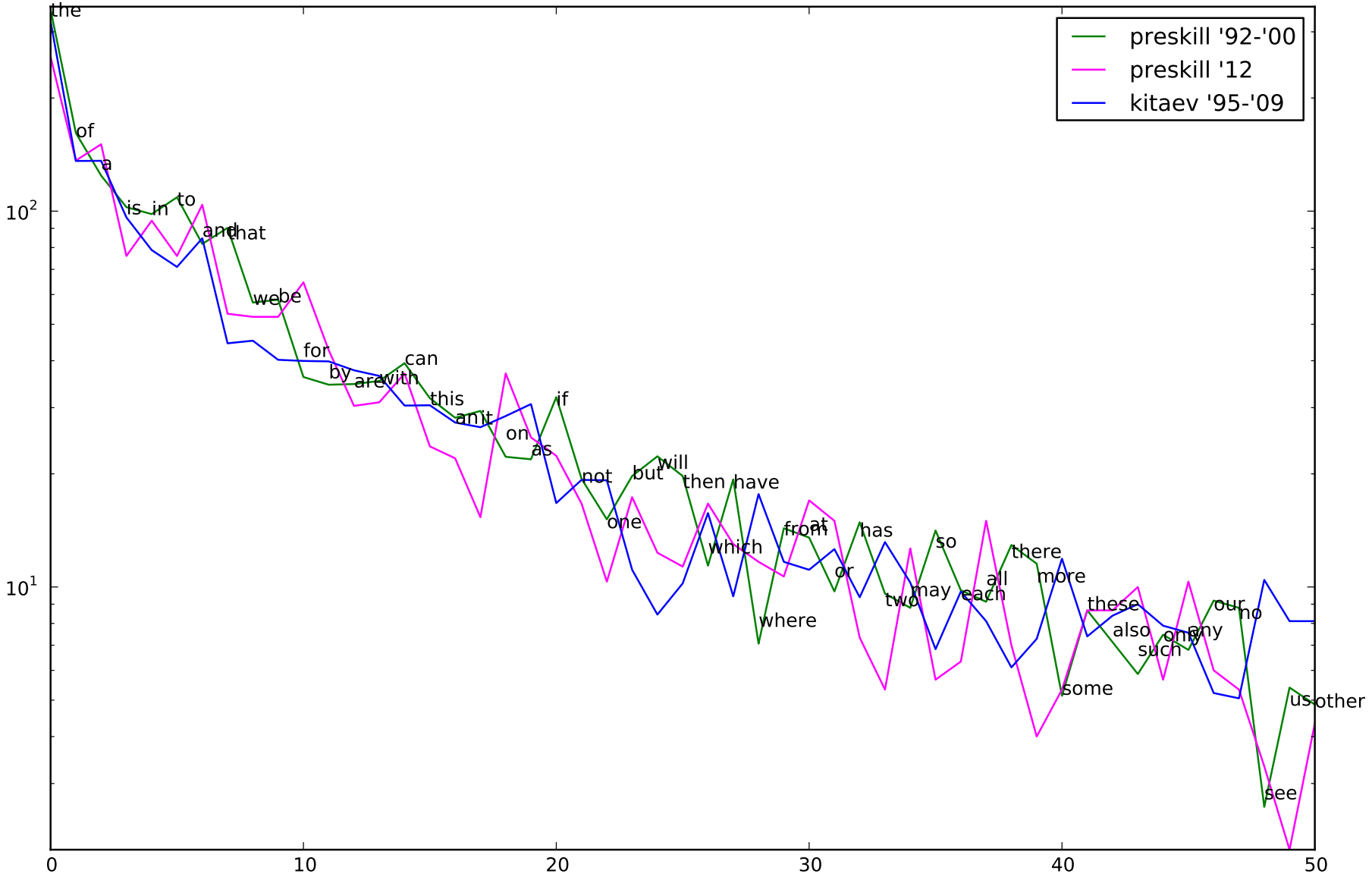
Kitaev

Preskill





Where have all the stopwords gone?



It all started here . . .

(1)

A. Kitaev — Anyons + Fault tolerance — 9 April 97

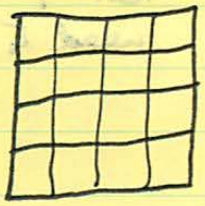
Classical Fault tolerance
– Not needed! Why?

Magnetic disk: $H = -J\sum \sigma_i^z \sigma_{i+1}^z$ (spins align) -- A "repetition code"

$X_i = X_{i+1}$

Rep. code has no quantum analog

⇓ closest thing is "toric code"



Torus -- qubits on edges of lattice

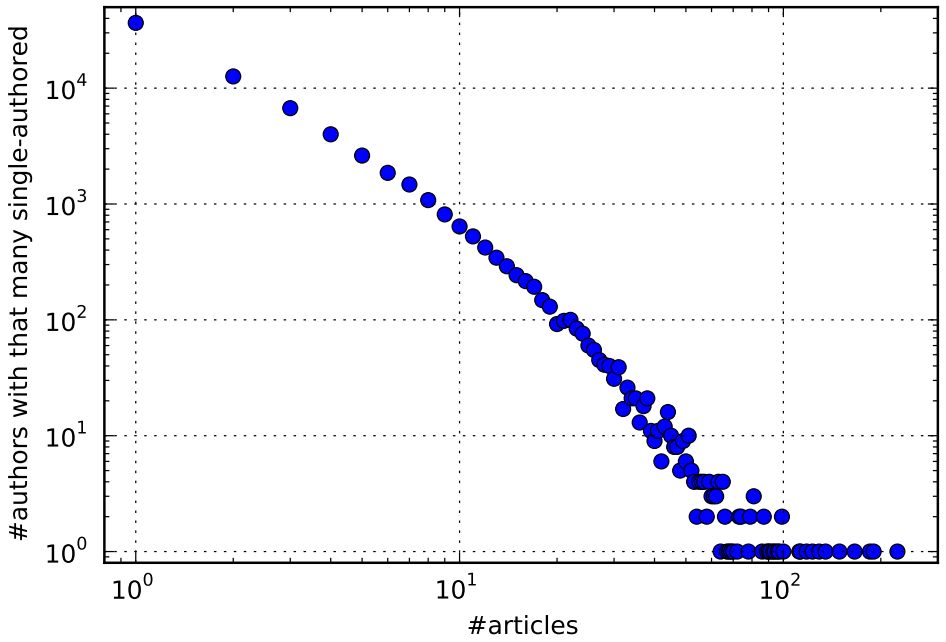
Stabilizer generators:

[r] $A_r = \prod_{j \in \text{plaquette}} \sigma_j^x$

[ℓ] $B_\ell = \prod_{j \in \text{star}} \sigma_j^z$

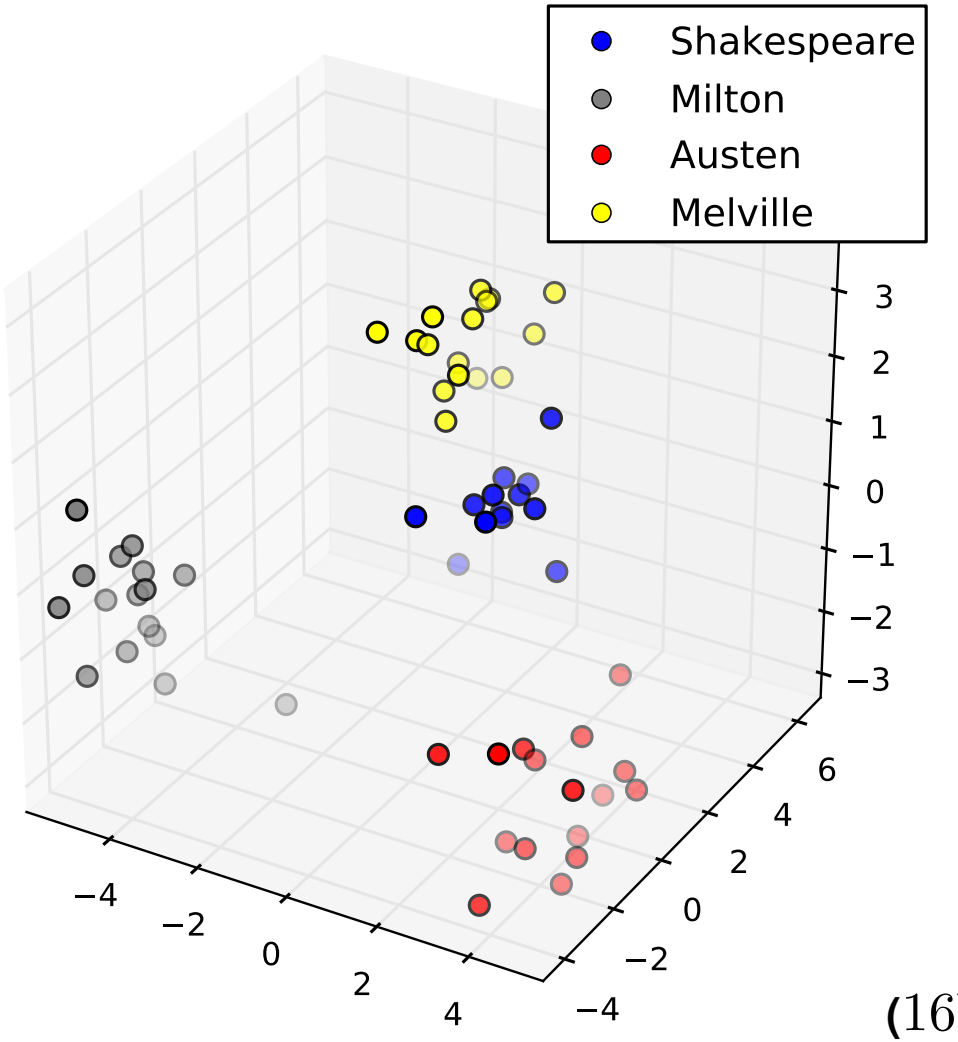
All mutually commuting

Big Data: Solo Authors



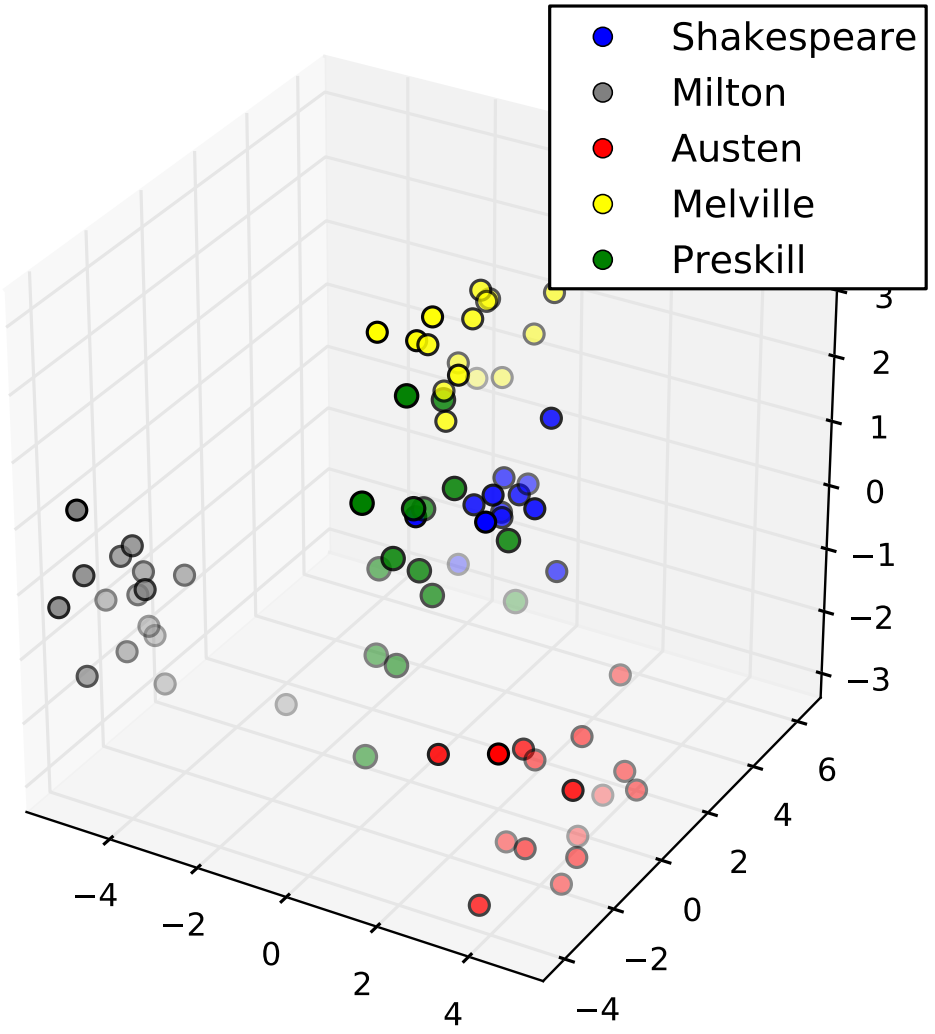
roughly 25% single authored, total of 225199 (36596 authors with a single such article, 12621 with two, . . . 640 authors with exactly 10 single authored articles apiece, out to the max author with 225 single-authored submissions).

nltk.org: J. Preskill vs J. Milton?

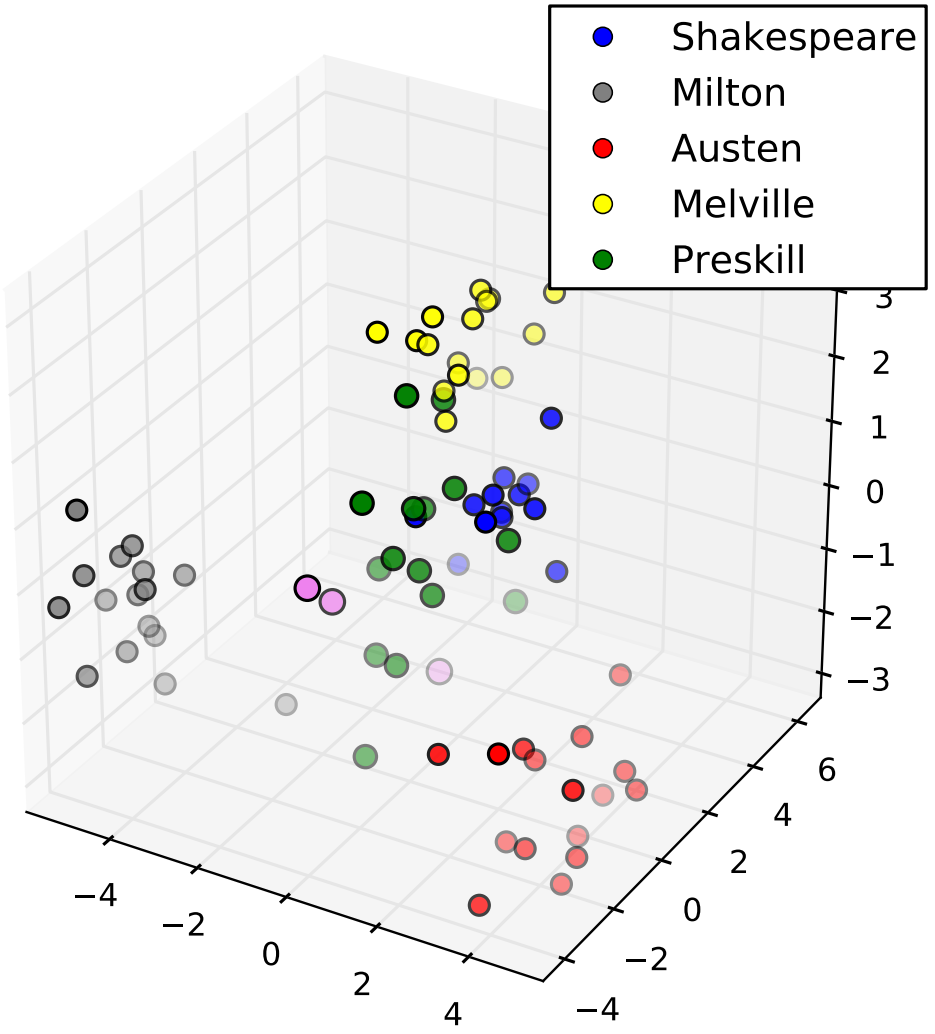


$(16^{\text{th}}, 17^{\text{th}}, 18^{\text{th}}, 19^{\text{th}})$

J. Preskill: You are no W. Shakespeare

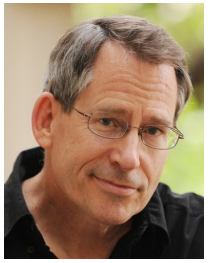


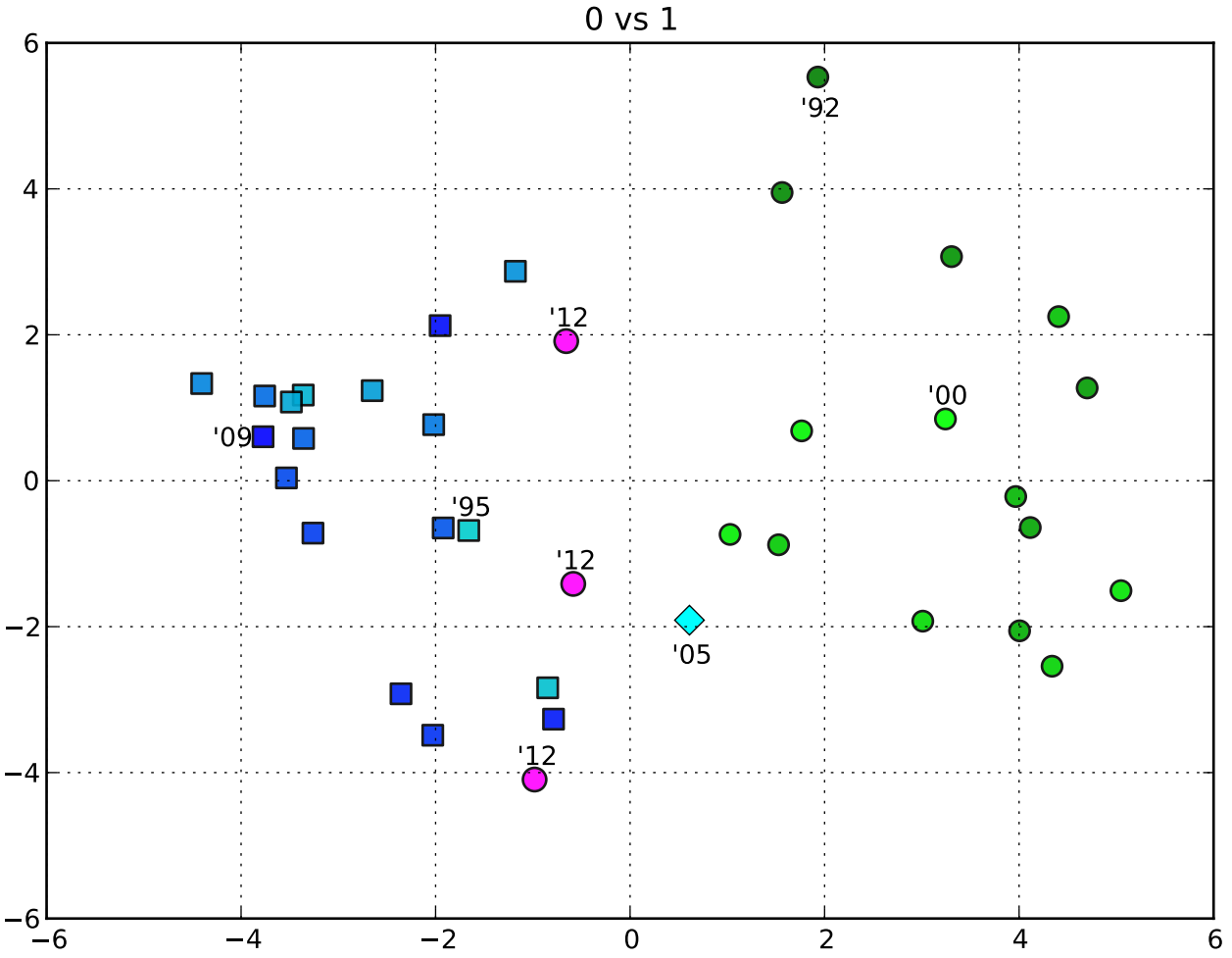
J. Preskill: You are no H. Melville



Kitaev

Preskill





“And Many Happy Returns”