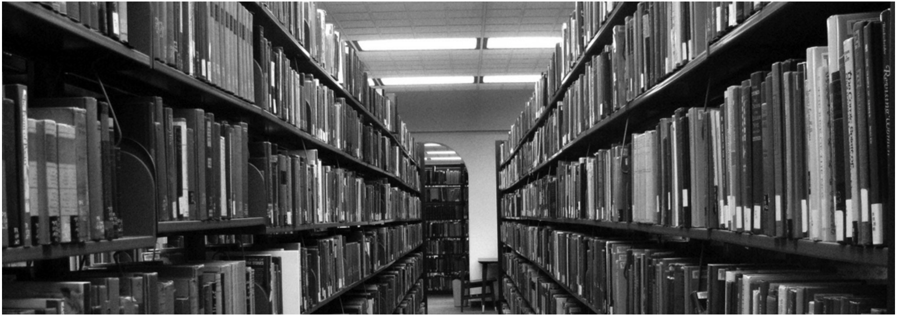
Probabilistic Topic Models and User Behavior

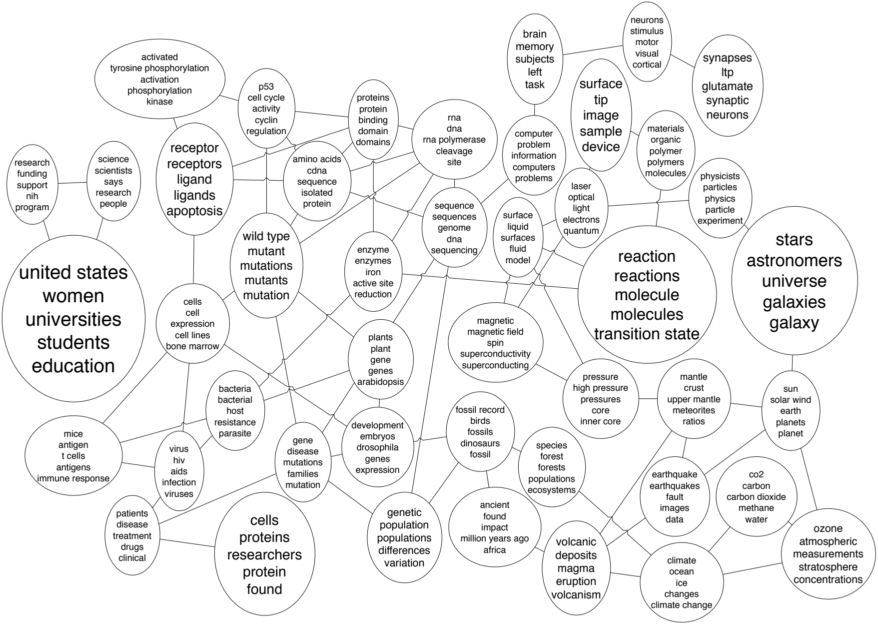
David M. Blei
Columbia University

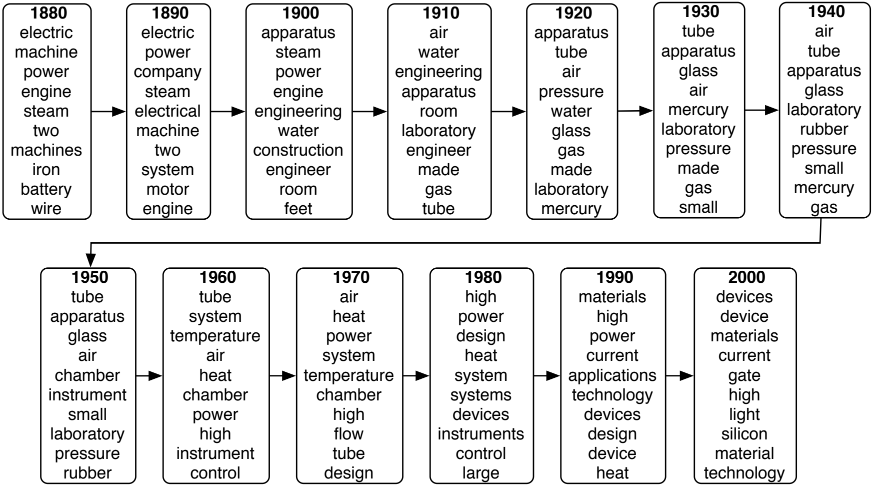
- **ORGANIZE**
- **VISUALIZE**
- **SUMMARIZE**
- **SEARCH**
- **PREDICT**
- **UNDERSTAND**

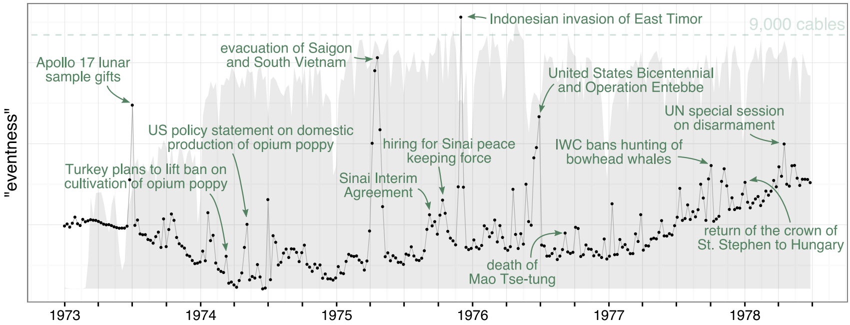


TOPIC MODELING

1. **Discover** the thematic structure
2. **Annotate** the documents
3. **Use** the annotations to visualize, organize, summarize, ...

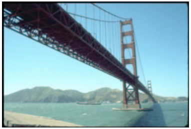

research funding support nih program
science scientists says research people
united states women universities students education
activated tyrosine phosphorylation activation phosphorylation kinase
receptor receptors ligand ligands apoptosis
p53 cell cycle activity cyclin regulation
proteins protein binding domain domains
amino acids cdna sequence isolated protein
rna dna rna polymerase cleavage site
wild type mutant mutations mutants mutation
enzyme enzymes iron active site reduction
sequence sequences genome dna sequencing
cells cell expression cell lines bone marrow
plants plant gene genes arabidopsis
bacteria bacterial host resistance parasite
mice antigen t cells antigens immune response
virus hiv aids infection viruses
gene disease mutations families mutation
development embryos drosophila genes expression
patients disease treatment drugs clinical
cells proteins researchers protein found
genetic population populations differences variation
brain memory subjects left task
neurons stimulus motor visual cortical
synapses ltp glutamate synaptic neurons
surface tip image sample device
computer problem information computers problems
materials organic polymer polymers molecules
laser optical light electrons quantum
physicists particles physics particle experiment
surface liquid surfaces fluid model
reaction reactions molecule molecules transition state
stars astronomers universe galaxies galaxy
magnetic magnetic field spin superconductivity superconducting
pressure high pressure pressures core inner core
mantle crust upper mantle meteorites ratios
sun solar wind earth planets planet
fossil record birds fossils dinosaurs fossil
species forest forests populations ecosystems
ancient found impact million years ago africa
earthquake earthquakes fault images data
co2 carbon carbon dioxide methane water
ozone atmospheric measurements stratosphere concentrations
volcanic deposits magma eruption volcanism
climate ocean ice changes climate change


1880
electric
machine
power
engine
steam
two
machines
iron
battery
wire
1890
electric
power
company
steam
electrical
machine
two
system
motor
engine
1900
apparatus
steam
power
engine
engineering
water
construction
engineer
room
feet
1910
air
water
engineering
apparatus
room
laboratory
engineer
made
gas
tube
1920
apparatus
tube
air
pressure
water
glass
gas
made
laboratory
mercury
1930
tube
apparatus
glass
air
mercury
laboratory
pressure
made
gas
small
1940
air
tube
apparatus
glass
laboratory
rubber
pressure
small
mercury
gas
1950
tube
apparatus
glass
air
chamber
instrument
small
laboratory
pressure
rubber
1960
tube
system
temperature
air
heat
chamber
power
high
instrument
control
1970
air
heat
power
system
temperature
chamber
high
flow
tube
design
1980
high
power
design
heat
system
systems
devices
instruments
control
large
1990
materials
high
power
current
applications
technology
devices
design
device
heat
2000
devices
device
materials
current
gate
high
light
silicon
material
technology


"eventness"
9,000 cables
Indonesian invasion of East Timor
evacuation of Saigon
and South Vietnam
Apollo 17 lunar
sample gifts
United States Bicentennial
and Operation Entebbe
UN special session
on disarmament
US policy statement on domestic
production of opium poppy
hiring for Sinai peace
keeping force
IWC bans hunting of
bowhead whales
Turkey plans to lift ban on
cultivation of opium poppy
Sinai Interim
Agreement
return of the crown of
St. Stephen to Hungary
death of
Mao Tse-tung
1973
1974
1975
1976
1977
1978

SKY WATER TREE
MOUNTAIN PEOPLE

SCOTLAND WATER
FLOWER HILLS TREE

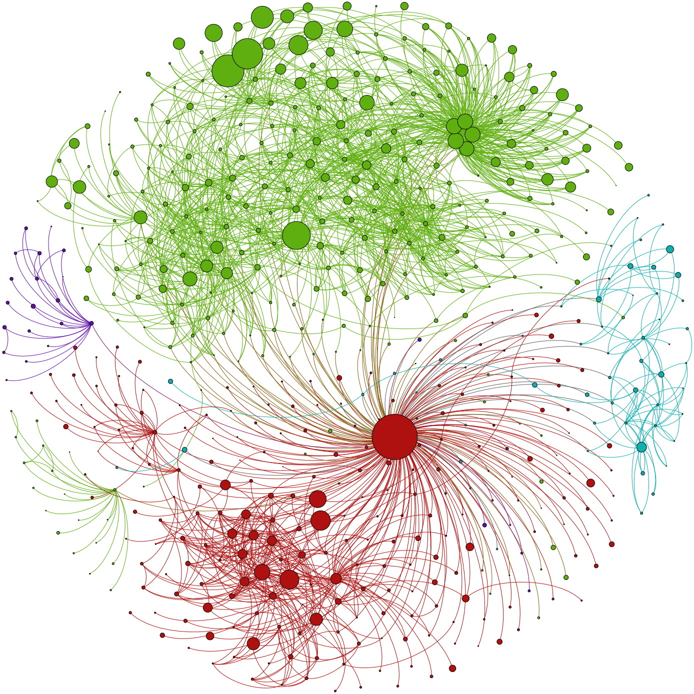


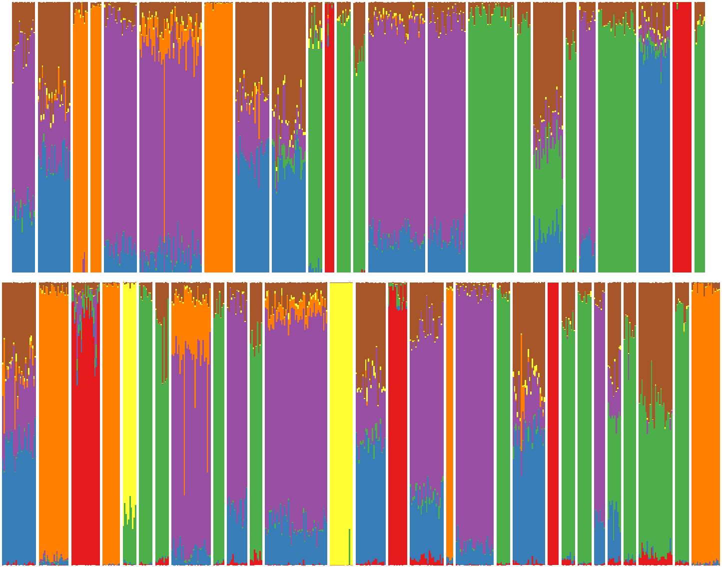
SKY WATER BUILDING
PEOPLE WATER

FISH WATER OCEAN
TREE CORAL

PEOPLE MARKET PATTERN
TEXTILE DISPLAY

BIRDS NEST TREE
BRANCH LEAVES





Charles Darwin's library

The NYC subway

- **People read documents.**
- These might be people for whom we want to form predictions.
- And, their behavior is an additional signal about the meaning of the documents and the organization of the collection.

This talk

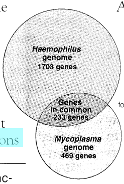
1. Introduction to topic modeling
2. Recommendation and exploration with collaborative topic models
3. The bigger picture: Using probability models to solve problems with data

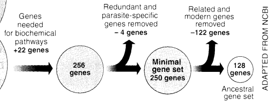
Introduction to Topic Modeling

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

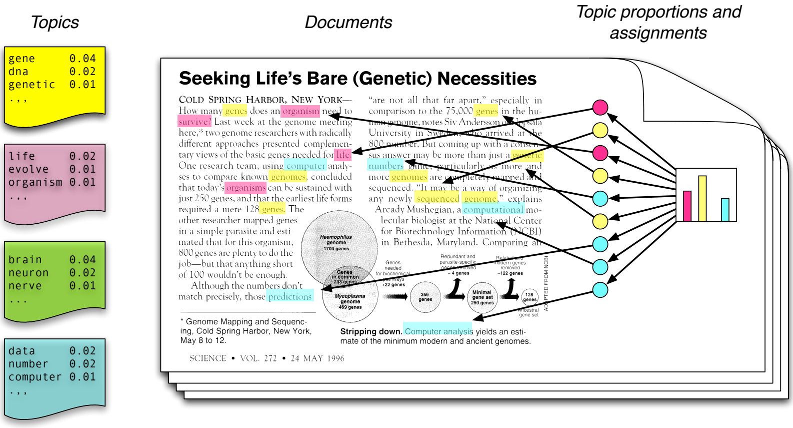




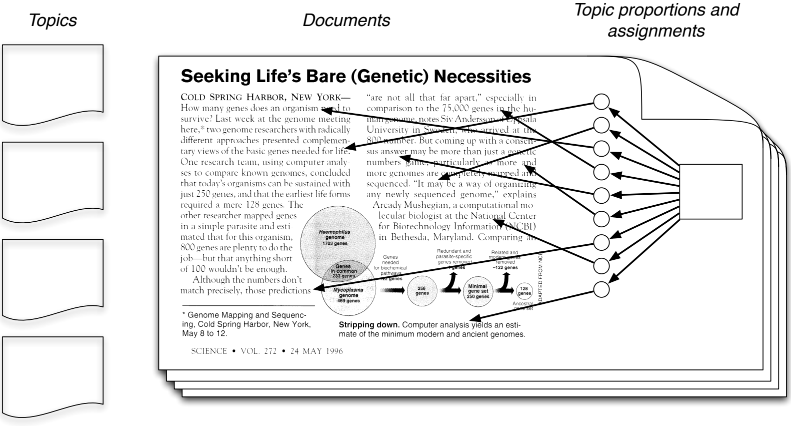
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

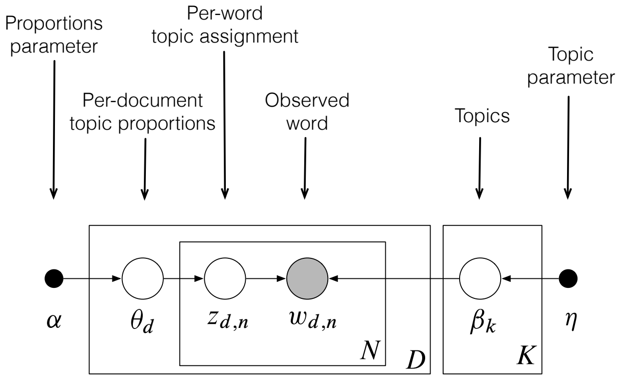
Documents exhibit multiple topics.



Latent Dirichlet Allocation

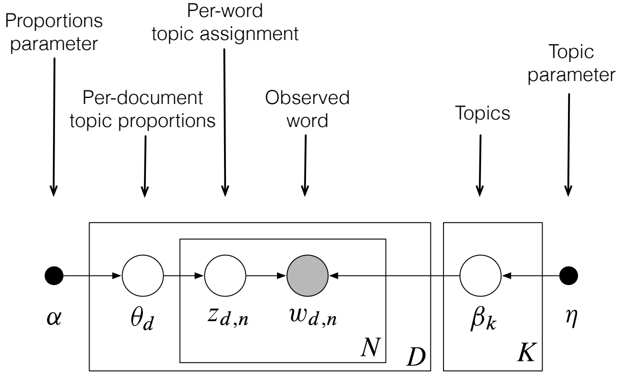


Latent Dirichlet Allocation



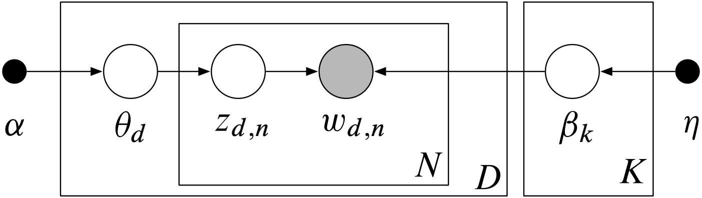
LDA as a graphical model

- Nodes are random variables; edges indicate dependence.
- Shaded nodes are observed; unshaded nodes are hidden.
- Plates indicate replicated variables.

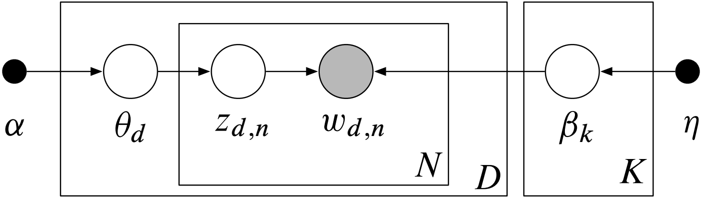


LDA as a graphical model

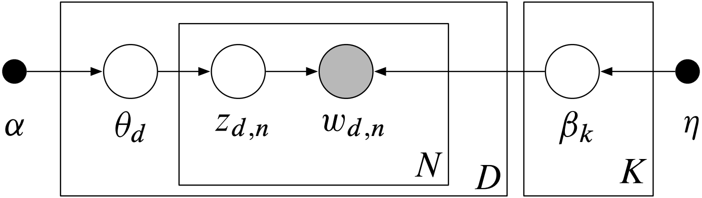
- Defines a factorization of the joint probability distribution
- Encodes independence assumptions about the variables
- Connects to algorithms for computing with data



- The joint defines a posterior, $p(\theta, z, \beta \mid w)$.
- From a collection of documents, infer
  - Per-word topic assignment $z_{d,n}$
  - Per-document topic proportions θ_d
  - Per-corpus topic distributions β_k
- Then use posterior expectations to perform the task at hand: information retrieval, document similarity, exploration, and others.



- Mean field variational methods (Blei et al., 2001, 2003)
- Expectation propagation (Minka and Lafferty, 2002)
- Collapsed Gibbs sampling (Griffiths and Steyvers, 2002)
- Distributed sampling (Newman et al., 2008; Ahmed et al., 2012)
- Collapsed variational inference (Teh et al., 2006)
- Stochastic inference (Hoffman et al., 2010, 2013; Mimno et al., 2012)
- Factorization inference (Arora et al., 2012; Anandkumar et al., 2012)
- Amortized inference (Srivastava and Sutton, 2016)



- LDA in R [https://cran.r-project.org/web/packages/lda/]
- GenSim [https://radimrehurek.com/gensim]
- Mallet [http://mallet.cs.umass.edu]
- Vowpal Wabbit [http://hunch.net/∼vw/]
- Apache Spark [http://spark.apache.org/]
- SciKit Learn [http://scikit-learn.org/]

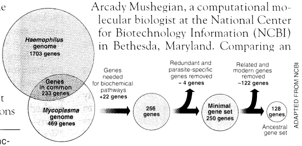
- **Data**: The OCR'ed collection of *Science* from 1990–2000
  - 17K documents
  - 11M words
  - 20K unique terms (stop words and rare words removed)

- **Model**: 100-topic LDA model using variational inference.

Seeking Life's Bare (Genetic) Necessities

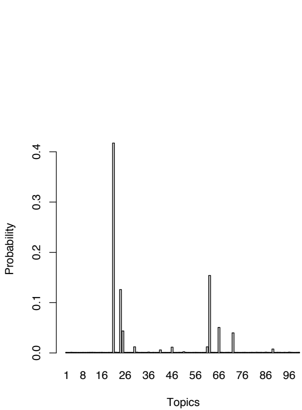
COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an



Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.



human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

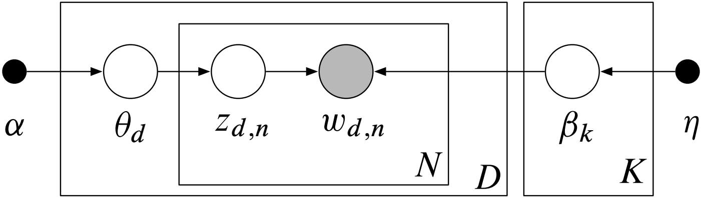
1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says	Children	Garden

6	7	8	9	10
Bush	Building	Won	Yankees	Government
Campaign	Street	Team	Game	War
Clinton	Square	Second	Mets	Military
Republican	Housing	Race	Season	Officials
House	House	Round	Run	Iraq
Party	Buildings	Cup	League	Forces
Democratic	Development	Open	Baseball	Iraqi
Political	Space	Game	Team	Army
Democrats	Percent	Play	Games	Troops
Senator	Real	Win	Hit	Soldiers

11	12	13	14	15
Children	Stock	Church	Art	Police
School	Percent	War	Museum	Yesterday
Women	Companies	Women	Show	Man
Family	Fund	Life	Gallery	Officer
Parents	Market	Black	Works	Officers
Child	Bank	Political	Artists	Case
Life	Investors	Catholic	Street	Found
Says	Funds	Government	Artist	Charged
Help	Financial	Jewish	Paintings	Street
Mother	Business	Pope	Exhibition	Shot

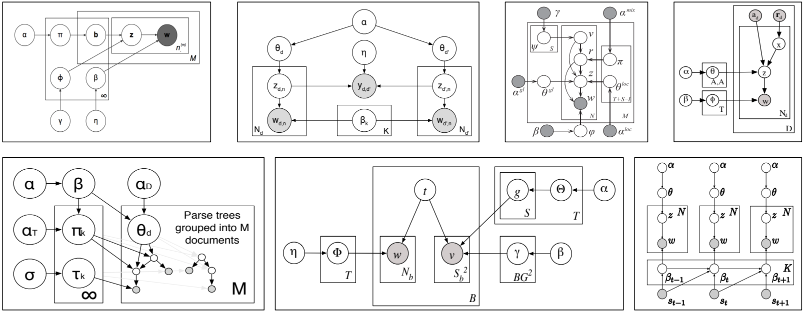
How does LDA "work"?

- LDA trades off two goals.
  1. In each **document**, allocate its words to **few topics**.
  2. In each **topic**, assign high probability to **few terms**.
- These goals are at odds.
  - Putting a document in a single topic makes #2 hard:
    All of its words must have probability under that topic.
  - Putting very few words in each topic makes #1 hard:
    To cover a document's words, it must assign many topics to it.
- Trading off these goals finds groups of tightly co-occurring words.

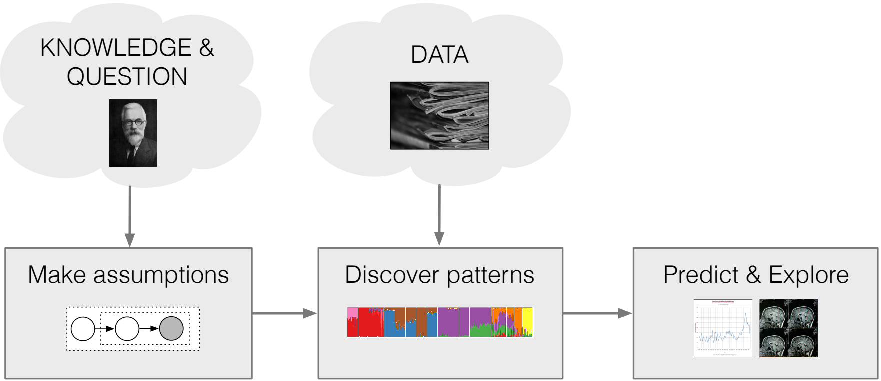


- Summary: LDA discovers themes through posterior inference.

- Other perspectives
  - Latent semantic analysis [Deerwester et al., 1990; Hofmann, 1999]
  - A mixed-membership model [Erosheva, 2004]
  - PCA and matrix factorization [Jakulin and Buntine, 2002]
  - Was independently invented for genetics [Pritchard et al., 2000]



- LDA has become a building block that enables many applications.
- Algorithmic improvements let us fit models to massive data. (See VW, Gensim, Mallet, others.)
- Organizing and finding patterns in text is important in the sciences, humanities, industry, and culture.



- Case study in **text analysis with probability models**
- Topic modeling research
  - develops new models.
  - develops new inference algorithms.
  - develops new applications, visualizations, tools.

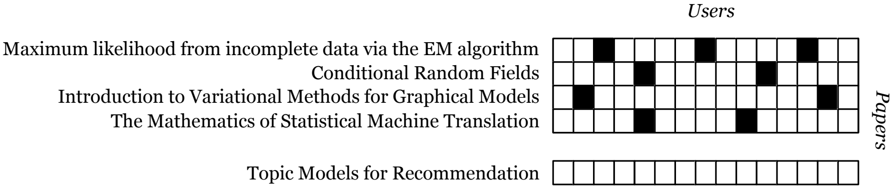
Collaborative Topic Models

with Prem Gopalan, Laurent Charlin, and Chong Wang

Charles Darwin’s library

Reading on the New York subway

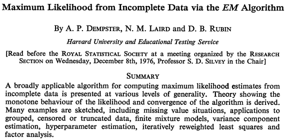
- **People read documents.**
- *Collaborative topic models* connect content to consumption



- Example: Scientists share their research libraries.

- Collaborative topic models can
  - Helps readers discover documents, old and new.
  - Describe readers in terms of topical preferences
  - Identify documents that are impactful, interdisciplinary

- Consider EM (Dempster et al., 1977). We infer topics from its text:



Maximum Likelihood from Incomplete Data via the *EM* Algorithm

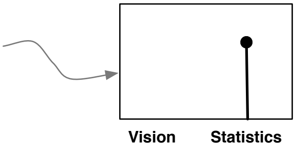
By A. P. Dempster, N. M. Laird and D. B. Rubin

Harvard University and Educational Testing Service

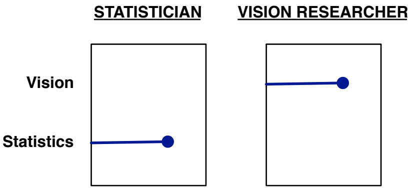
[Read before the Royal Statistical Society at a meeting organized by the Research Section on Wednesday, December 8th, 1976, Professor S. D. Silvey in the Chair]

Summary

A broadly applicable algorithm for computing maximum likelihood estimates from incomplete data is presented at various levels of generality. Theory showing the monotone behaviour of the likelihood and convergence of the algorithm is derived. Many examples are sketched, including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimation, iteratively reweighted least squares and factor analysis.

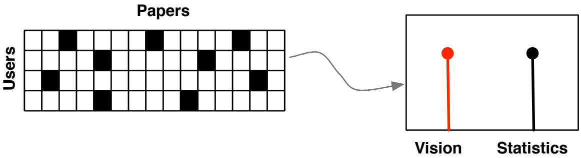


- Suppose there are two types of scientists

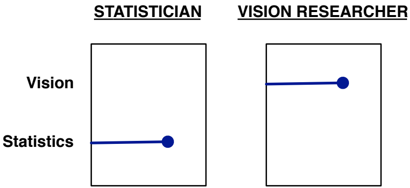


- We first recommend the EM paper to **statisticians**.

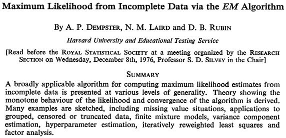
- With user data, we can adjust the topics to account for who liked it:



- Consider again the scientists



- We now recommend the EM paper to **vision researchers**.



Maximum Likelihood from Incomplete Data via the *EM* Algorithm

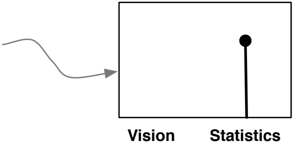
By A. P. Dempster, N. M. Laird and D. B. Rubin

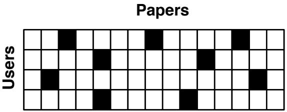
Harvard University and Educational Testing Service

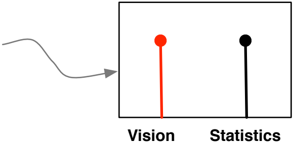
[Read before the Royal Statistical Society at a meeting organized by the Research Section on Wednesday, December 8th, 1976, Professor S. D. Silvey in the Chair]

Summary

A broadly applicable algorithm for computing maximum likelihood estimates from incomplete data is presented at various levels of generality. Theory showing the monotone behaviour of the likelihood and convergence of the algorithm is derived. Many examples are sketched, including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimation, iteratively reweighted least squares and factor analysis.

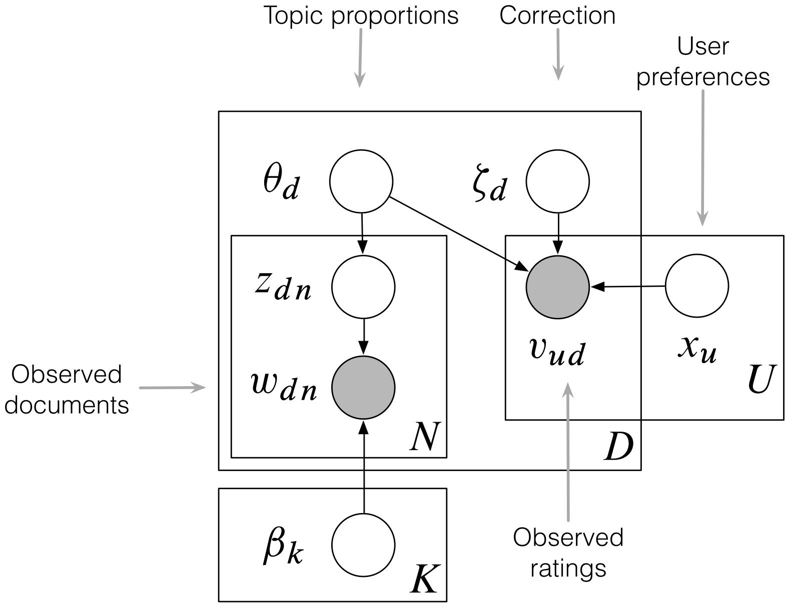


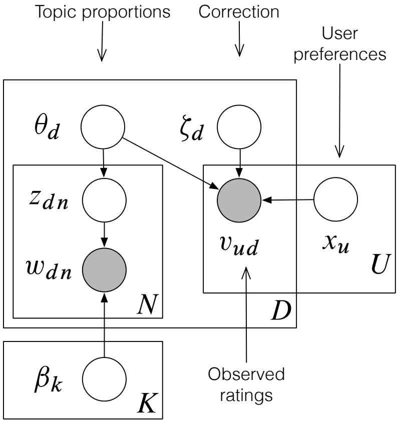




Without text, we cannot initially recommend to anyone.

Without user data, we cannot recommend to vision researchers.


Topic proportions
Correction
User preferences
Observed documents
Observed ratings
θ_d
ζ_d
z_{dn}
w_{dn}
v_{ud}
x_u
β_k
N
D
U
K



$$
\begin{aligned}
\theta_{dk} &\sim \mathrm{Gamma}(\cdot,\cdot) \\
\zeta_{dk} &\sim \mathrm{Gamma}(\cdot,\cdot) \\
x_{uk} &\sim \mathrm{Gamma}(\cdot,\cdot) \\
v_{ud} &\sim \mathrm{Poisson}((\theta_d + \zeta_d)^\top x_u)
\end{aligned}
$$

- Blends factorization-based and content-based recommendation
- Describes user preferences with interpretable topics
- Builds on Poisson factorization

  [Canney 2004; Dunson and Herring 2005; Gopalan et al. 2014]

- Big data set from Mendeley.com
- The data:
  - 261K documents
  - 80K users
  - 10K vocabulary terms
  - 25M observed words
  - 5.1M entries (sparsity is 0.02%)

Maximum Likelihood from Incomplete Data via the *EM* Algorithm

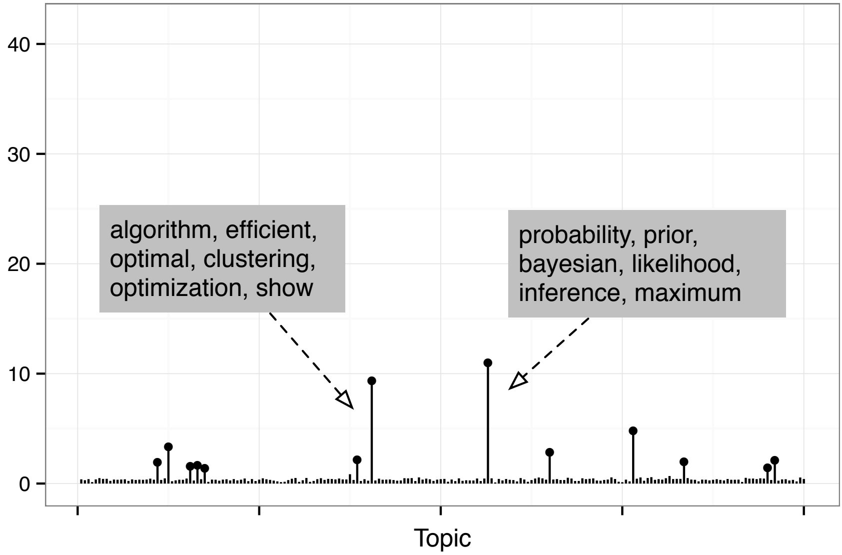
By A. P. DEMPSTER, N. M. LAIRD and D. B. RUBIN

Harvard University and Educational Testing Service

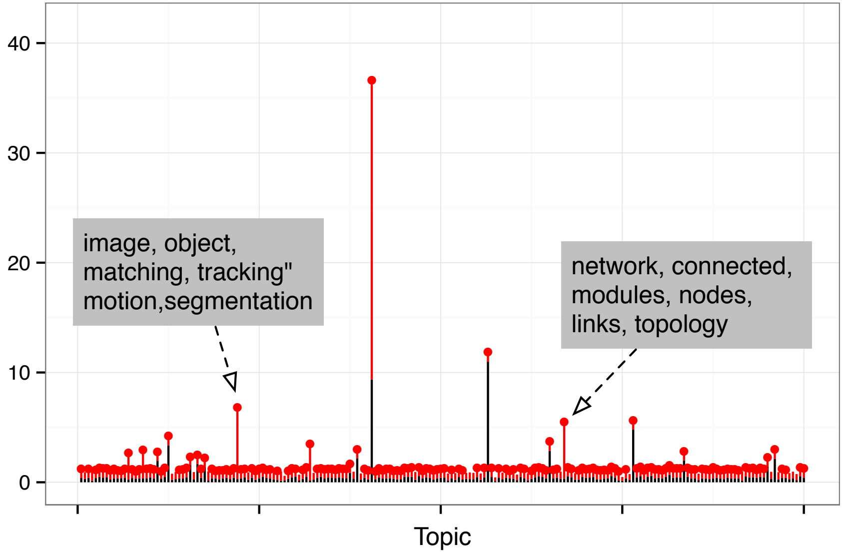
[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, December 8th, 1976, Professor S. D. SILVEY in the Chair]

SUMMARY

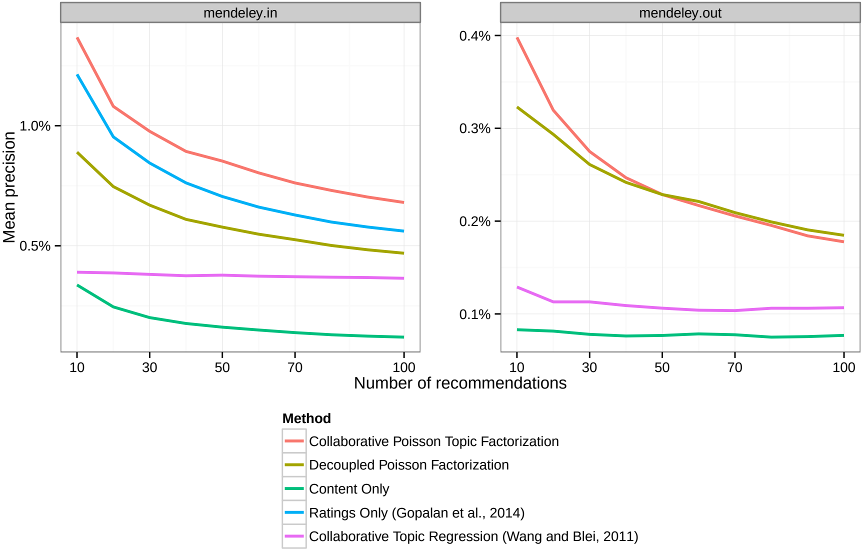
A broadly applicable algorithm for computing maximum likelihood estimates from incomplete data is presented at various levels of generality. Theory showing the monotone behaviour of the likelihood and convergence of the algorithm is derived. Many examples are sketched, including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimation, iteratively reweighted least squares and factor analysis.



40
30
20
10
0
algorithm, efficient, optimal, clustering, optimization, show
probability, prior, bayesian, likelihood, inference, maximum
Topic

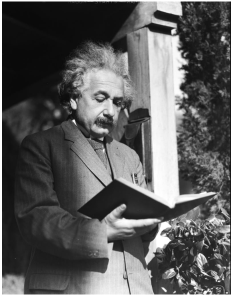

40
30
20
10
0
image, object,
matching, tracking"
motion,segmentation
network, connected,
modules, nodes,
links, topology
Topic

Mendeley

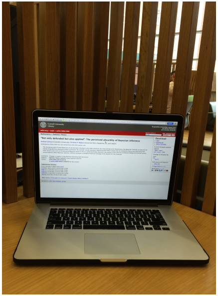


mendeley.in
mendeley.out
Mean precision
1.0%
0.5%
0.4%
0.3%
0.2%
0.1%
10
30
50
70
100
Number of recommendations
Method
Collaborative Poisson Topic Factorization
Decoupled Poisson Factorization
Content Only
Ratings Only (Gopalan et al., 2014)
Collaborative Topic Regression (Wang and Blei, 2011)

Darwin's library



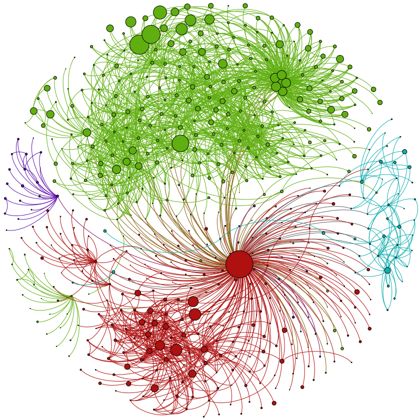
Einstein reading



Another scientist reading

- The readers also **tell us about the articles**.
- We can look at posterior estimates to find
  - Interdisciplinary articles
  - Influential articles within a field
  - Outside influences on a field

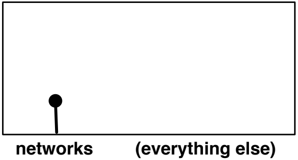
“Network Analysis”



network; connected; modules; nodes; links; topology; connectivity; graph; robustness; connections; modular; world; degree; properties

Assortative mixing in networks

M. E. J. Newman
Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120 and
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501



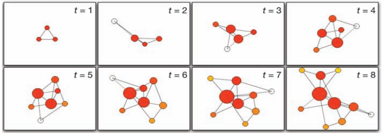
About networks

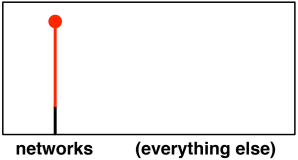
- Assortative mixing in networks

  (Newman, 2002)
- Mixing patterns in networks

  (Newman, 2002)
- Catastrophic cascade of failures in interdependent networks

  (Buldyrev et al., 2010)





About networks; for readers of networks

- Emergence of scaling in random networks
  (Barabassi and Albert, 1999)
- Statistical mechanics of complex networks
  (Albert and Barabassi, 2002)
- Complex networks: Structure and dynamics
  (Boccaletti et al., 2006)

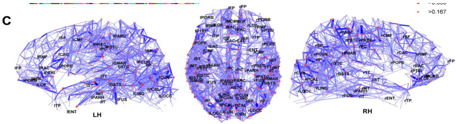
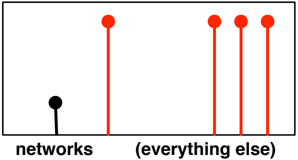


Figure 3. High-Resolution Connection Matrix, Network Layout and Connectivity Backbone (Participant A, scan 2)



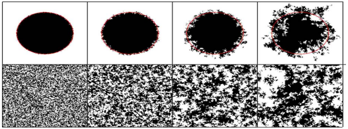
About networks; for readers of other fields

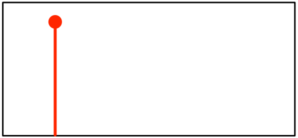
- Mapping the Structural Core of Human Cerebral Cortex

  (Hagmann et al., 2008)
- Network thinking in ecology and evolution

  (Proulx et al., 2005)
- Linked: The New Science of Networks

  (Barabasi, 2002)





***Not* about networks; for readers of networks**

- Power-law distributions in empirical data
  (Clauset et al., 2009)
- Statistical physics of social dynamics
  (Castellano et al., 2009)
- The origin of bursts and heavy tails in human dynamics
  (Barabasi, 2005)

“Statistical Modeling”

About this field; read by users in this field

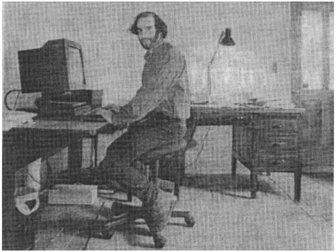
- A Bayesian analysis of some nonparametric problems
- Bayesian measures of model complexity and fit
- Monte Carlo Methods in Bayesian Computation

About this field; read by users in other fields

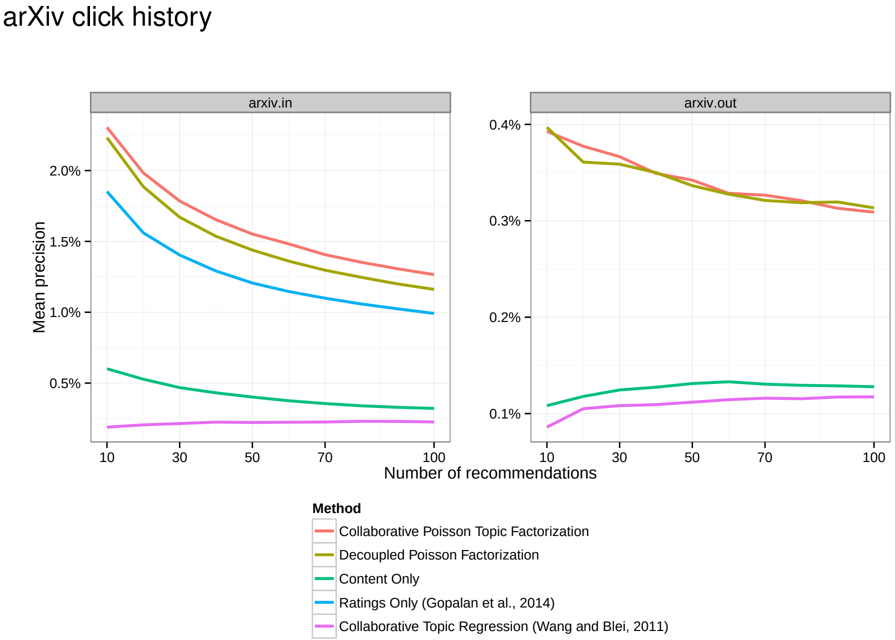
- A tutorial on HMMs and selected applications in speech recognition
- An Introduction to Bayesian Networks and Influence Diagrams
- Maximum likelihood from incomplete data via the EM algorithm

About other fields; read by users in this field

- Second Thoughts on the Bootstrap
- A guide to Eclipse and the R plug-in StatET
- Using Multivariate Statistics



- A decade of clicks on arXiv.org (2003–2013)
- The data:
  - 826K documents
  - 120K users
  - 14K vocabulary terms
  - 54M observed words
  - 43.6M entries (sparsity is 0.04%)


arXiv click history
arxiv.in
arxiv.out
Mean precision
2.0%
1.5%
1.0%
0.5%
0.4%
0.3%
0.2%
0.1%
10
30
50
70
100
Number of recommendations
Method
Collaborative Poisson Topic Factorization
Decoupled Poisson Factorization
Content Only
Ratings Only (Gopalan et al., 2014)
Collaborative Topic Regression (Wang and Blei, 2011)

Stat.ML: Machine Learning

In stat.ML; for stat.ML readers

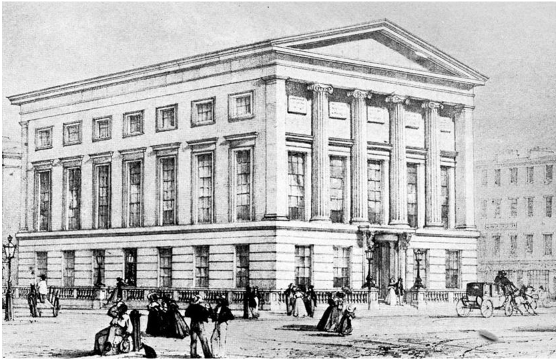
- Noisy matrix decomposition via convex relaxation
- Robust computation of linear models, or how to find a needle in a haystack
- High-dimensional regression with noisy and missing data

In stat.ML; for other readers

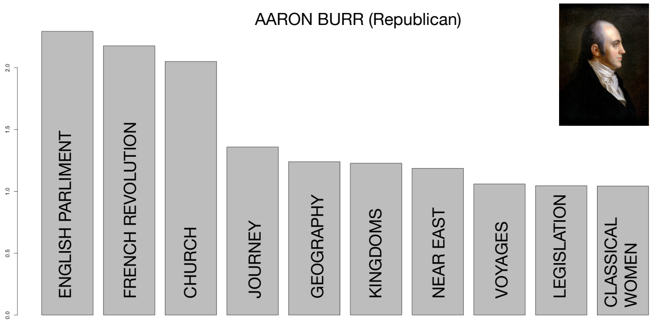
- Co-evolution of selection and influence in social networks
- Hierarchical structure and the prediction of missing links in networks
- Learning continuous-time social network dynamics

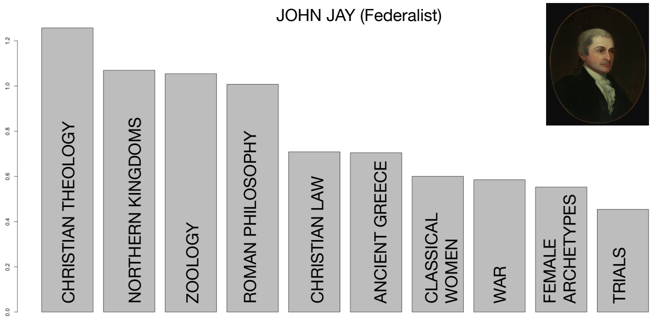
In other categories; for stat.ML readers

- Finding structure with randomness
- Representation learning: A review and new perspectives
- Computational and statistical tradeoffs via convex relaxation



- The New York Society library is the first library in New York City (1754)
- Mark Hoffman and Peter Bearman (sociology) are using collaborative topic models to explore the usage patterns of important figures in U.S. History
- The data
  - 1789 – 1806
  - 847 users (people like Aaron Burr, John Jay, etc.)
  - 2,327 items (items like The Prince)
  - 33M words; vocabulary of 8,000

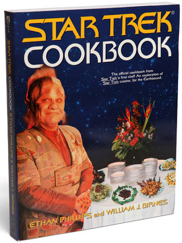

AARON BURR (Republican)
0.0
0.5
1.0
1.5
2.0
ENGLISH PARLIMENT
FRENCH REVOLUTION
CHURCH
JOURNEY
GEOGRAPHY
KINGDOMS
NEAR EAST
VOYAGES
LEGISLATION
CLASSICAL WOMEN


JOHN JAY (Federalist)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
CHRISTIAN THEOLOGY
NORTHERN KINGDOMS
ZOOLOGY
ROMAN PHILOSOPHY
CHRISTIAN LAW
ANCIENT GREECE
CLASSICAL WOMEN
WAR
FEMALE ARCHETYPES
TRIALS

Collaborative topic models

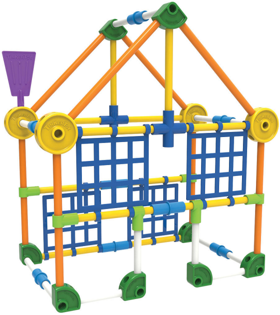
- Connect text to usage, content to consumption
- Blend content-based and user-based recommendation
- Opens new windows into how people read

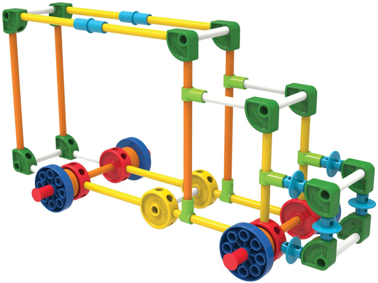
Discussion: Modern Probabilistic Modeling



How to use traditional machine learning and statistics to solve modern problems

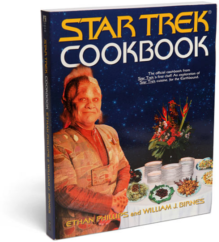
Probabilistic machine learning: tailored models for the problem at hand.



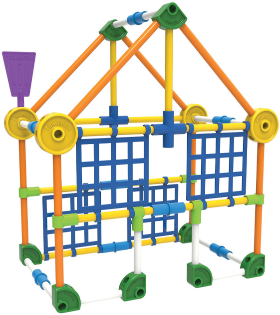


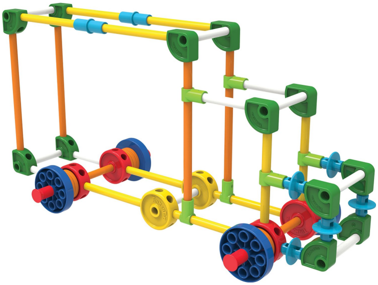
Probabilistic machine learning: tailored models for the problem at hand.

- Compose and connect reusable parts
- Driven by disciplinary knowledge and its questions
- Focus on discovering and using structure in unstructured data
- Exploratory, observational, causal analyses

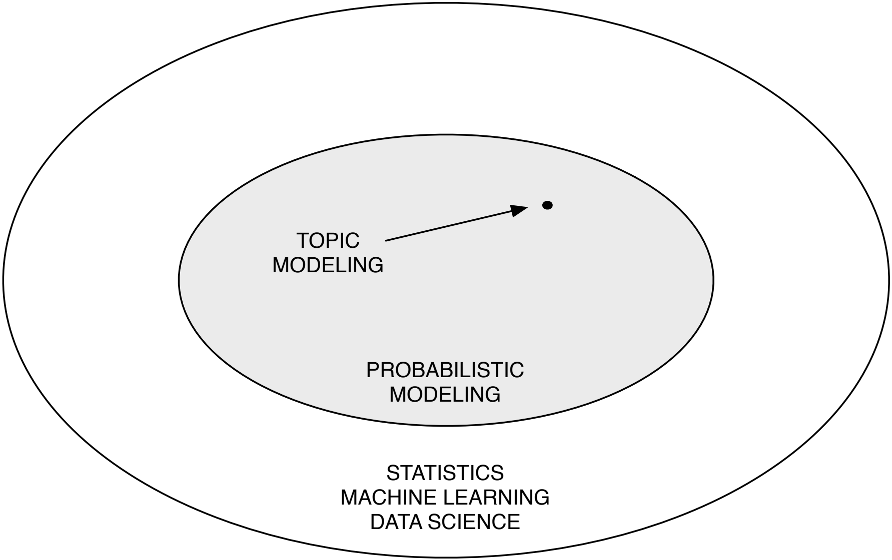


Many software packages available; typically fast and scalable

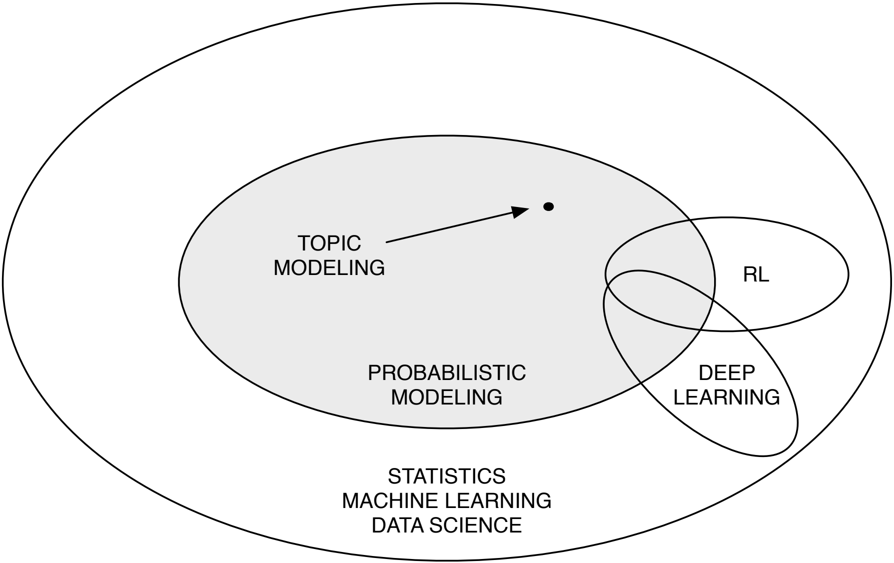




More challenging to implement; may not be fast or scalable

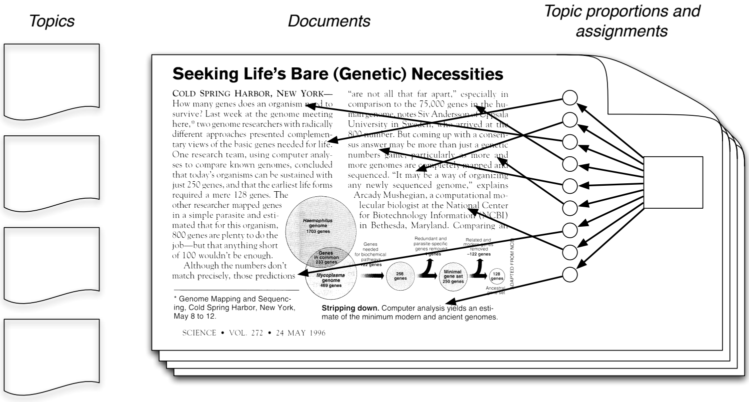


A big picture



A big picture (not to scale)

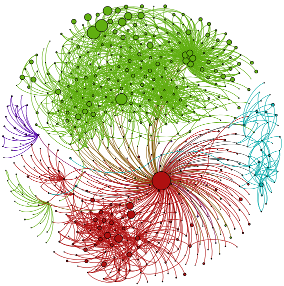
I. Assume our data come from a model with hidden patterns at work

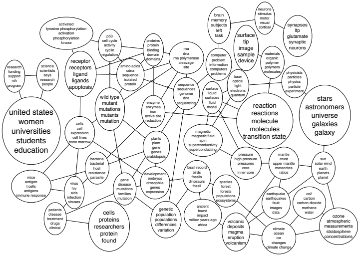


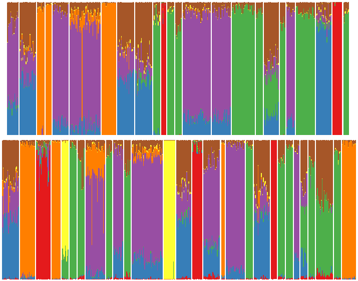
II. Discover those patterns from data

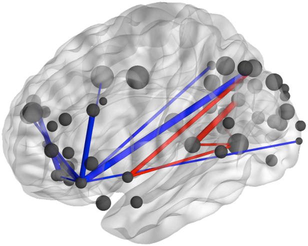
$$\boxed{\nu^* = \arg\max_{\nu} \; \mathbb{E}_q \left[\log p(x, z, \beta \,|\, \alpha)\right] + \mathbb{H}\left[q(z, \beta \,|\, \nu)\right]}$$

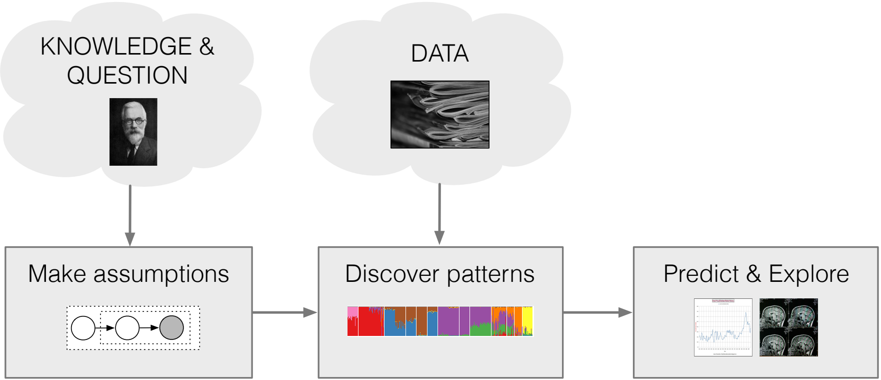
III. Use the discovered patterns to predict about and explore the data





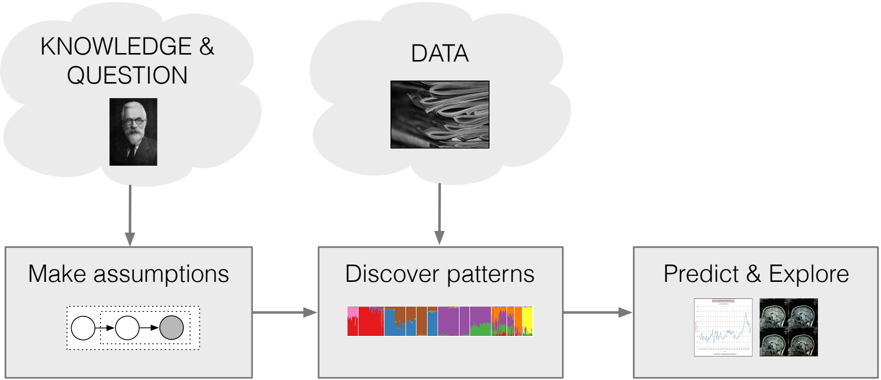






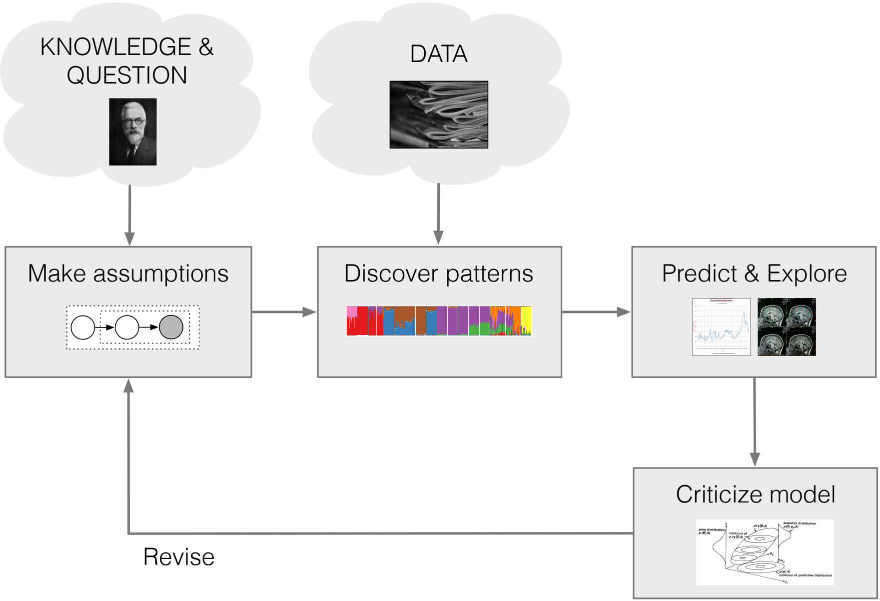
Our perspective:

- Customized data analysis is important to many fields.
- This pipeline separates assumptions, computation, application.
- It facilitates solving data science problems.



What we need:

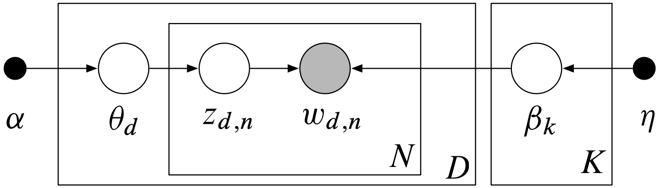
- **Flexible** and **expressive** components for building models
- **Scalable** and **generic** inference algorithms
- **Easy to use** software to stretch probabilistic modeling into new areas

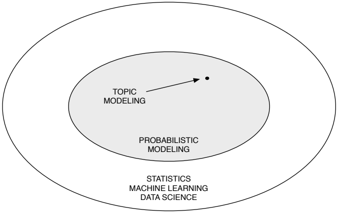

KNOWLEDGE &
QUESTION
DATA
Make assumptions
Discover patterns
Predict & Explore
Criticize model
Revise

Edward: Probabilistic modeling, inference, and criticism

`github.com/blei-lab/edward`

(lead by Dustin Tran)


α
θ_d
$z_{d,n}$
$w_{d,n}$
N
D
β_k
K
η

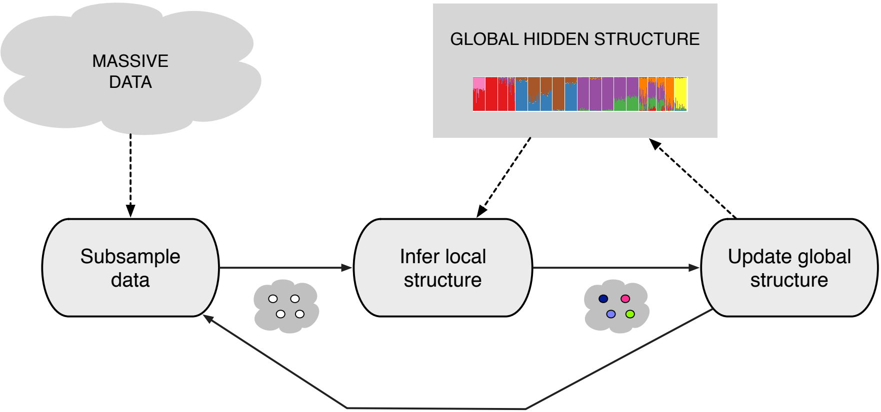

TOPIC
MODELING
PROBABILISTIC
MODELING
STATISTICS
MACHINE LEARNING
DATA SCIENCE

We should seek out unfamiliar summaries of observational material, and establish their useful properties... And still more novelty can come from finding, and evading, still deeper lying constraints.

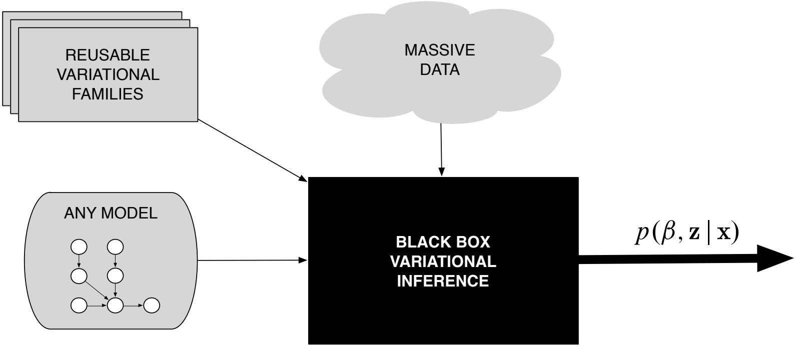
(John Tukey, *The Future of Data Analysis*, 1962)

A few slides about inference

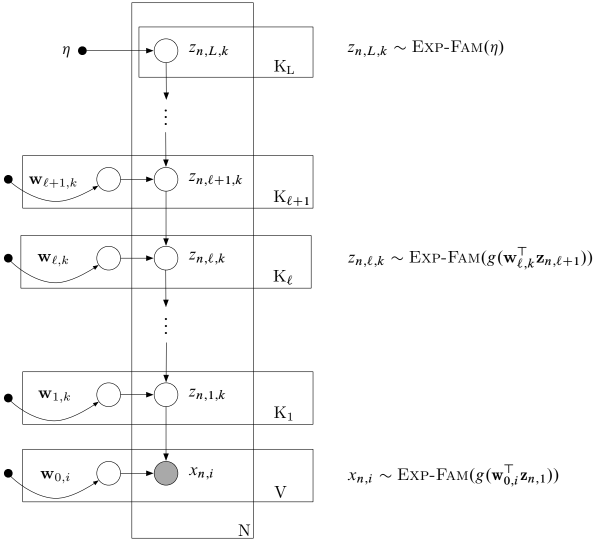
Stochastic variational inference [Hoffman et al., 2013]



Black box variational inference [Ranganath et al., 2014]

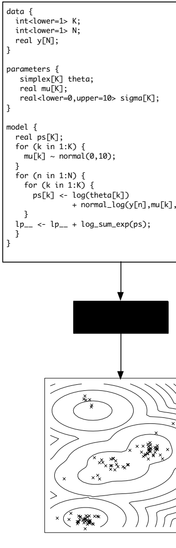


- Easily use variational inference with *any model*
- No exponential family requirements
- No mathematical work beyond specifying the model



Deep Exponential Families

[Ranganath et al., 2015]

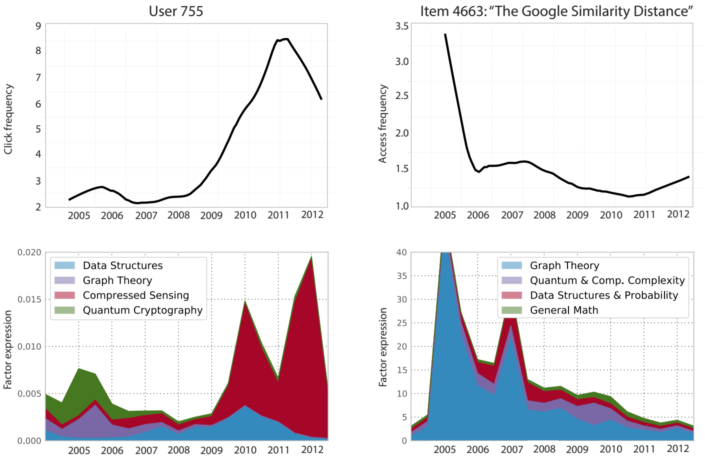


Probabilistic Progra

[Kucukelbir et al., 2

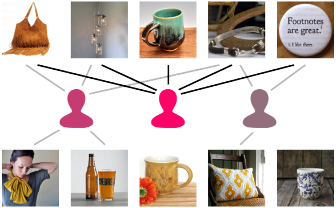
Some recent work about recommendation

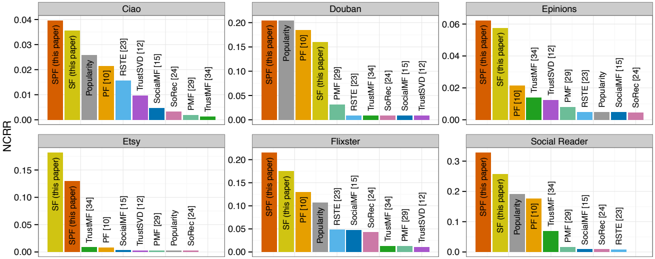
Time series recommendation



(with Laurent Charlin, James McInerney, Rajesh Ranganath)

Social recommendation





(with Allison Chaney)