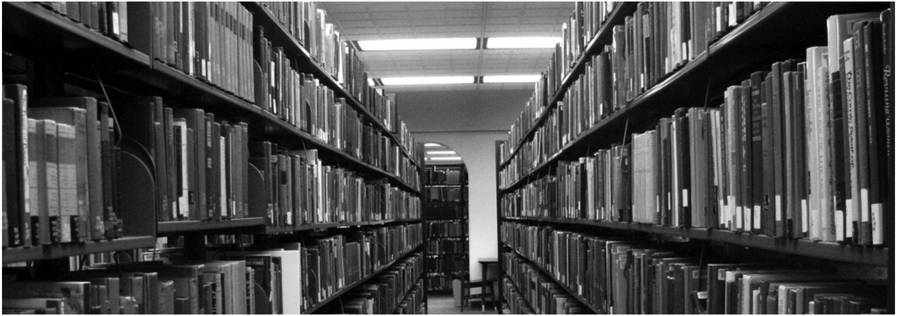
Topic Models for Understanding History

David M. Blei
Columbia University

joint work with Allison Chaney, Hanna Wallach, and Matt Connelly

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



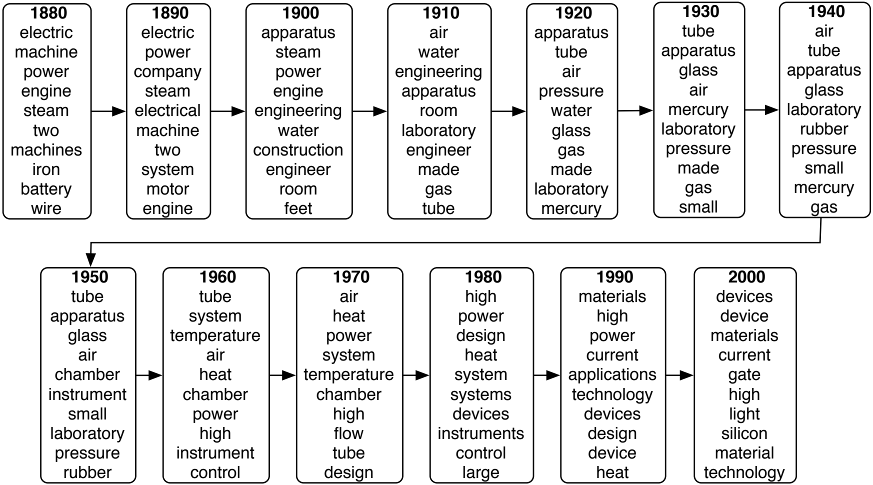
TOPIC MODELING

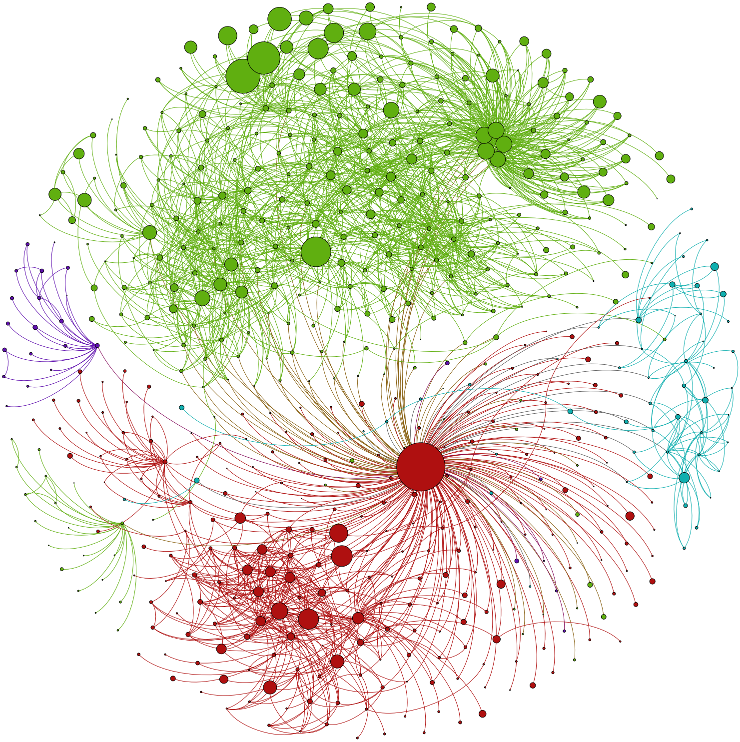
1. **Discover** the thematic structure
2. **Annotate** the documents
3. **Use** the annotations to visualize, organize, summarize, ...

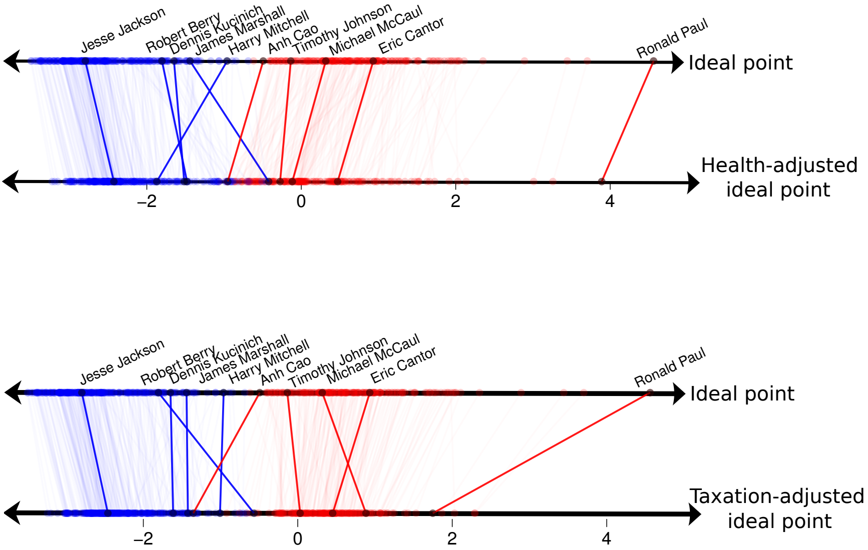
1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says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

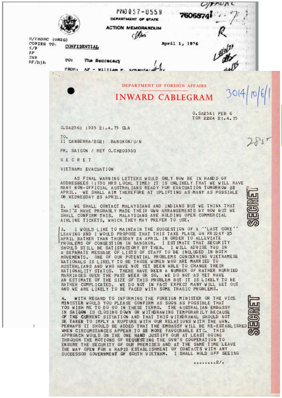
6	7	8	9	10
Bush	Building	Won	Yankees	Government
Campaign	Street	Team	Game	War
Clinton	Square	Second	Mets	Military
Republican	Housing	Race	Season	Officials
House	House	Round	Run	Iraq
Party	Buildings	Cup	League	Forces
Democratic	Development	Open	Baseball	Iraqi
Political	Space	Game	Team	Army
Democrats	Percent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


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




Jesse Jackson
Robert Berry
Dennis Kucinich
James Marshall
Harry Mitchell
Anh Cao
Timothy Johnson
Michael McCaul
Eric Cantor
Ronald Paul
Ideal point
Health-adjusted ideal point
−2
0
2
4
Jesse Jackson
Robert Berry
Dennis Kucinich
James Marshall
Harry Mitchell
Anh Cao
Timothy Johnson
Michael McCaul
Eric Cantor
Ronald Paul
Ideal point
Taxation-adjusted ideal point
−2
0
2
4



P760057-0559

DEPARTMENT OF STATE

ACTION MEMORANDUM

7606874

O/FADRC (ORIG)
COPIES TO: S/P, RF, INR, RF/hjb

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. Schaufele, Jr.

DEPARTMENT OF FOREIGN AFFAIRS

INWARD CABLEGRAM

3014/10/6/1

O.SA2561 PEB 6
TOR 2224 21.4.75

O.SA2561 1935 21.4.75 CLA

TO.
II CANBERRA/2021 BANGKOK/U/N

FM. SAIGON / REF O.CH203930

S E C R E T

VIETNAM: EVACUATION

AS FINAL WARNING LETTERS WOULD ONLY NOW BE IN HANDS OF ADDRESSEES (1730 HRS LOCAL TIME) IT IS UNLIKELY THAT WE WILL HAVE MANY NON-OFFICIAL AUSTRALIANS READY FOR EVACUATION TOMORROW 22 APRIL. WE SHALL AIM THEREFORE AT UPLIFTING AS MANY AS POSSIBLE ON WEDNESDAY 23 APRIL.

2. WE SHALL CONTACT MALAYSIANS AND INDIANS BUT WE THINK THAT THAI'S HAVE PROBABLY MADE THEIR OWN ARRANGEMENTS BY NOW BUT WE SHALL CONFIRM THIS. MALAYSIANS ARE HOLDING OPEN COMMERCIAL AIRLINE TICKETS, WHICH THEY MAY PREFER TO USE.

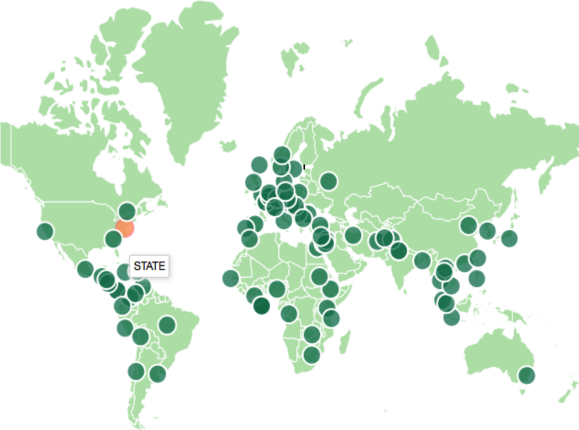
3. I WOULD LIKE TO MAINTAIN THE SUGGESTION OF A ''LAST CORE'' LEAVING AND I WOULD PROPOSE THAT THIS TAKE PLACE ON FRIDAY 25 APRIL RATHER THAN THURSDAY 24 APRIL IN ORDER TO ALLEVIATE PROBLEMS OF CONGESTION IN BANGKOK. I ESTIMATE THAT SECURITY SHOULD STILL BE SATISFACTORY BY THEN. I WILL ADVISE YOU IN A SEPARATE MESSAGE OF LISTS OF STAFF TO BE INCLUDED IN BOTH MOVEMENTS. ONE OF OUR POTENTIAL PROBLEMS CONCERNING VIETNAMESE NATIONALS IS LIKELY TO BE THOSE WOMEN WHO ARE MARRIED TO AUSTRALIANS AND WHO HAVE NOT YET BEEN ABLE TO CHANGE THEIR NATIONALITY STATUS. THERE HAVE BEEN A NUMBER OF RATHER HURRIED MARRIAGES OVER THE PAST WEEK OR SO. WE DO NOT AS YET HAVE AN ESTIMATE OF THE SIZE OF THIS PROBLEM BUT IT IS LIKELY TO BE RATHER COMPLICATED. WE DO NOT IN FACT EXPECT MANY WILL GET OUT AND WE ARE LIKELY TO BE FACED WITH SOME TRAGIC PROBLEMS.

4. WITH REGARD TO INFORMING THE FOREIGN MINISTER OR THE VICE MINISTER WOULD YOU PLEASE CONFIRM AS SOON AS POSSIBLE THAT YOU WISH ME TO DO SO IN THESE TERMS. THE AUSTRALIAN EMBASSY IN SAIGON IS CLOSING DOWN OR WITHDRAWING TEMPORARILY BECAUSE OF THE CURRENT SITUATION AND THAT THIS WITHDRAWAL SHOULD NOT BE TAKEN TO IMPLY A RUPTURE WITH OUR RELATIONS WITH THE GVN. PERHAPS IT SHOULD BE ADDED THAT THE EMBASSY WILL BE RE-ESTABLISHED WHEN CIRCUMSTANCES APPEAR TO BE MORE FAVOURABLE ETC. THIS APPROACH WOULD ON THE ONE HAND JUSTIFY OUR AT LEAST GOING THROUGH THE MOTIONS OF REQUESTING THE GVN'S COOPERATION TO ENSURE THE SECURITY OF OUR PREMISES AND AT THE SAME TIME LEAVE THE WAY OPEN FOR A RAPID ESTABLISHMENT OF CONTACTS WITH ANY SUCCESSOR GOVERNMENT OF SOUTH VIETNAM. I SHALL HOLD OFF SEEING

........2/.

SECRET

SECRET



- Historians want to identify important events from primary sources.
- Example: Embassies send cables to each other during the 1970s
- Goal: Use topic models to discover **events** in this data set

This talk

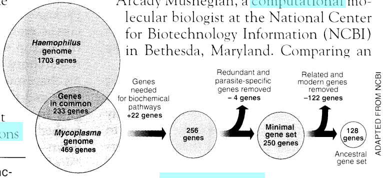
1. Introduction to topic modeling
2. Topic models for understanding history
3. The bigger picture: Using probability models to solve problems with data

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.

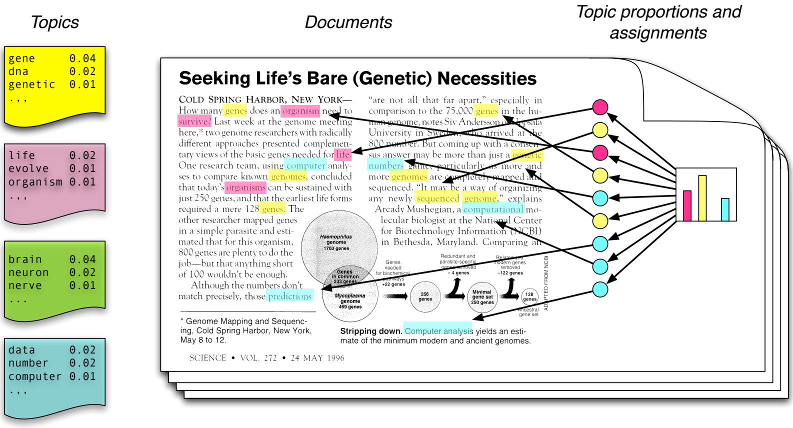
Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just 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



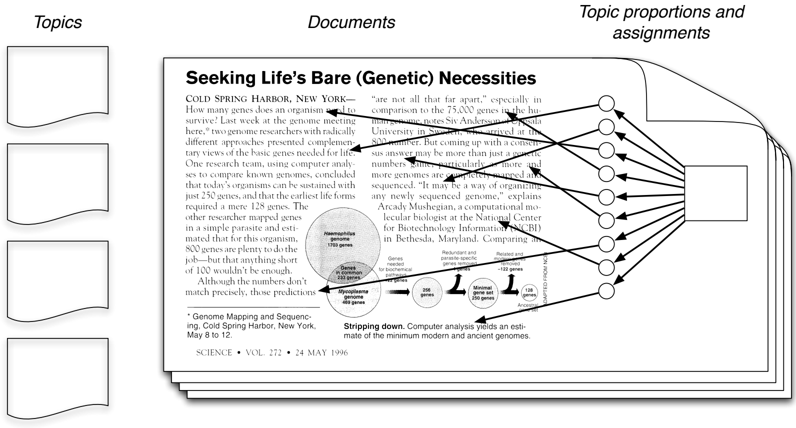
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

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

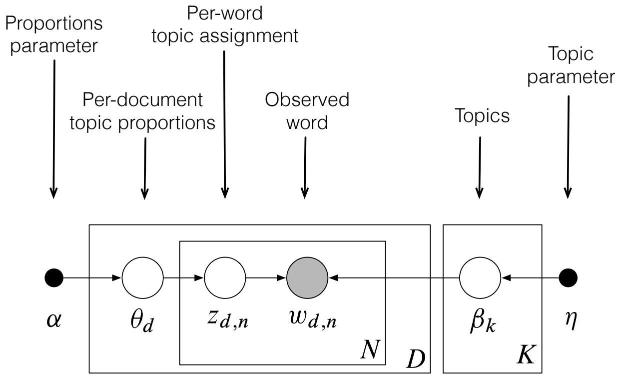
Documents exhibit multiple topics.



Latent Dirichlet Allocation

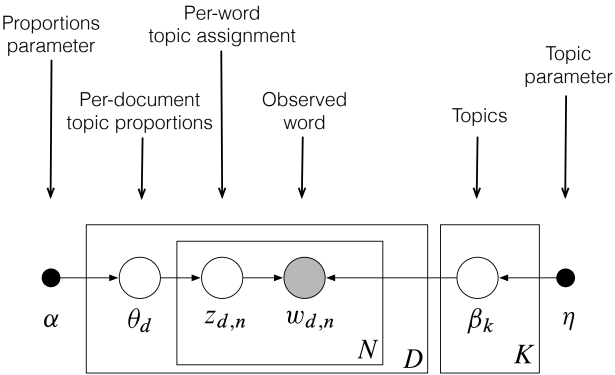


Latent Dirichlet Allocation



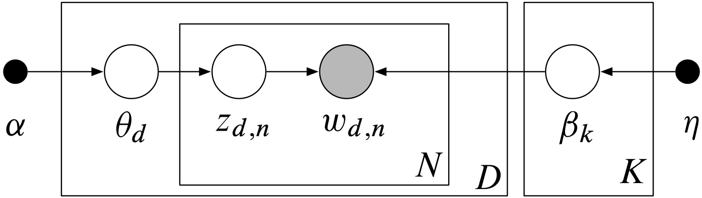
LDA as a graphical model

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



LDA as a graphical model

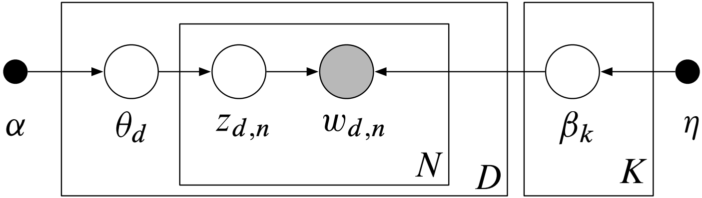
- Encodes independence assumptions about the variables
- Defines a factorization of the joint probability distribution
- Connects to algorithms for computing with data



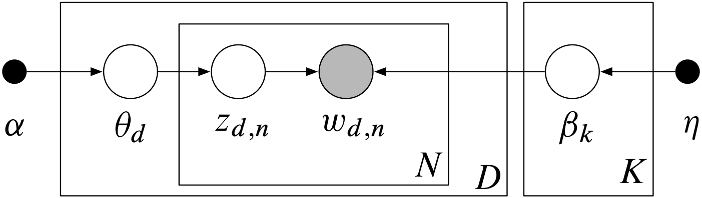
- The joint defines a posterior, $p(\theta, z, \beta \mid w)$.

- 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)



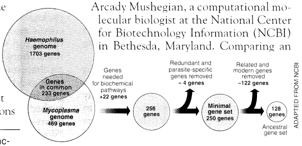
- LDA in R [https://cran.r-project.org/web/packages/lda/]
- GenSim [https://radimrehurek.com/gensim]
- Mallet [http://mallet.cs.umass.edu]
- Vowpal Wabbit [http://hunch.net/∼vw/]
- Apache Spark [http://spark.apache.org/]
- SciKit Learn [http://scikit-learn.org/]

- **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

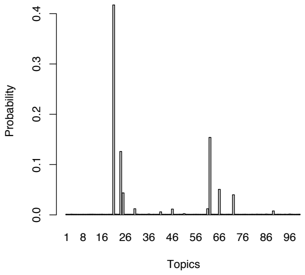
COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known 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

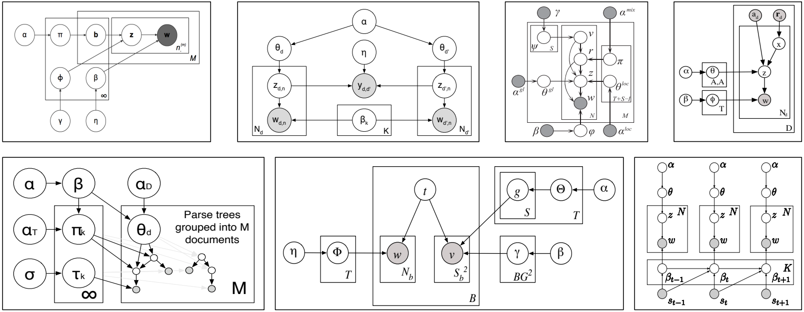
1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says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

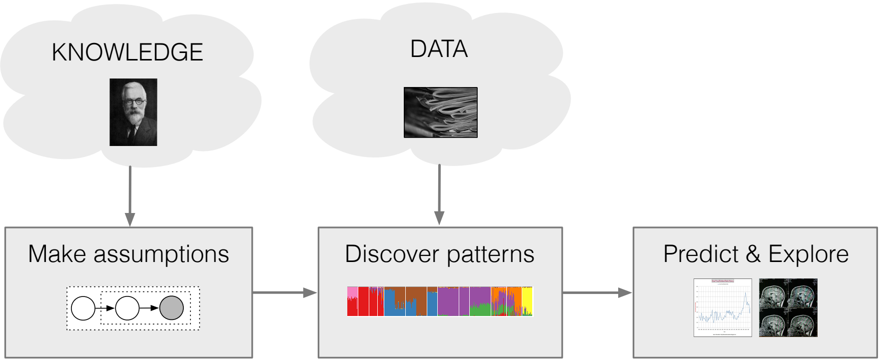
11	12	13	14	15
Children	Stock	Church	Art	Police
School	Percent	War	Museum	Yesterday
Women	Companies	Women	Show	Man
Family	Fund	Life	Gallery	Officer
Parents	Market	Black	Works	Officers
Child	Bank	Political	Artists	Case
Life	Investors	Catholic	Street	Found
Says	Funds	Government	Artist	Charged
Help	Financial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

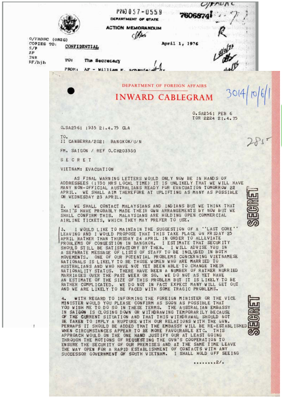


- Organizing and finding patterns in text is important in the sciences, humanities, industry, and culture.
- LDA is a simple building block that enables many applications. Topic modeling is an active field of research.
- Algorithmic improvements let us fit models to massive data.



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

Topic Models for Understanding History


P760057-0559
DEPARTMENT OF STATE
7606874
ACTION MEMORANDUM
R
O/FADRC (ORIG)
COPIES TO:
S/P
AF
INR
RF/hjb
CONFIDENTIAL
April 1, 1976
TO: The Secretary
FROM: AF - William E. Schaufele, Jr.

DEPARTMENT OF FOREIGN AFFAIRS
INWARD CABLEGRAM
3014/10/6/1

O.SA2561 PEB 6
TOR 2224 21.4.75

O.SA2561 1935 21.4.75 CLA

TO.
II CANBERRA/2021 BANGKOK/U/N

FM. SAIGON / REF O.CH203930

S E C R E T

VIETNAM: EVACUATION

AS FINAL WARNING LETTERS WOULD ONLY NOW BE IN HANDS OF ADDRESSEES (1730 HRS LOCAL TIME) IT IS UNLIKELY THAT WE WILL HAVE MANY NON-OFFICIAL AUSTRALIANS READY FOR EVACUATION TOMORROW 22 APRIL. WE SHALL AIM THEREFORE AT UPLIFTING AS MANY AS POSSIBLE ON WEDNESDAY 23 APRIL.

2. WE SHALL CONTACT MALAYSIANS AND INDIANS BUT WE THINK THAT THAI'S HAVE PROBABLY MADE THEIR OWN ARRANGEMENTS BY NOW BUT WE SHALL CONFIRM THIS. MALAYSIANS ARE HOLDING OPEN COMMERCIAL AIRLINE TICKETS, WHICH THEY MAY PREFER TO USE.

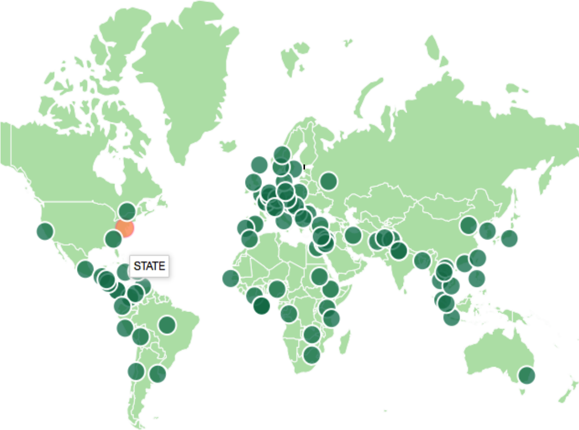
3. I WOULD LIKE TO MAINTAIN THE SUGGESTION OF A ''LAST CORE'' LEAVING AND I WOULD PROPOSE THAT THIS TAKE PLACE ON FRIDAY 25 APRIL RATHER THAN THURSDAY 24 APRIL IN ORDER TO ALLEVIATE PROBLEMS OF CONGESTION IN BANGKOK. I ESTIMATE THAT SECURITY SHOULD STILL BE SATISFACTORY BY THEN. I WILL ADVISE YOU IN A SEPARATE MESSAGE OF LISTS OF STAFF TO BE INCLUDED IN BOTH MOVEMENTS. ONE OF OUR POTENTIAL PROBLEMS CONCERNING VIETNAMESE NATIONALS IS LIKELY TO BE THOSE WOMEN WHO ARE MARRIED TO AUSTRALIANS AND WHO HAVE NOT YET BEEN ABLE TO CHANGE THEIR NATIONALITY STATUS. THERE HAVE BEEN A NUMBER OF RATHER HURRIED MARRIAGES OVER THE PAST WEEK OR SO. WE DO NOT AS YET HAVE AN ESTIMATE OF THE SIZE OF THIS PROBLEM BUT IT IS LIKELY TO BE RATHER COMPLICATED. WE DO NOT IN FACT EXPECT MANY WILL GET OUT AND WE ARE LIKELY TO BE FACED WITH SOME TRAGIC PROBLEMS.

4. WITH REGARD TO INFORMING THE FOREIGN MINISTER OR THE VICE MINISTER WOULD YOU PLEASE CONFIRM AS SOON AS POSSIBLE THAT YOU WISH ME TO DO SO IN THESE TERMS. THE AUSTRALIAN EMBASSY IN SAIGON IS CLOSING DOWN OR WITHDRAWING TEMPORARILY BECAUSE OF THE CURRENT SITUATION AND THAT THIS WITHDRAWAL SHOULD NOT BE TAKEN TO IMPLY A RUPTURE WITH OUR RELATIONS WITH THE GVN. PERHAPS IT SHOULD BE ADDED THAT THE EMBASSY WILL BE RE-ESTABLISHED WHEN CIRCUMSTANCES APPEAR TO BE MORE FAVOURABLE ETC. THIS APPROACH WOULD ON THE ONE HAND JUSTIFY OUR AT LEAST GOING THROUGH THE MOTIONS OF REQUESTING THE GVN'S COOPERATION TO ENSURE THE SECURITY OF OUR PREMISES AND AT THE SAME TIME LEAVE THE WAY OPEN FOR A RAPID ESTABLISHMENT OF CONTACTS WITH ANY SUCCESSOR GOVERNMENT OF SOUTH VIETNAM. I SHALL HOLD OFF SEEING

SECRET

SECRET

........2/.



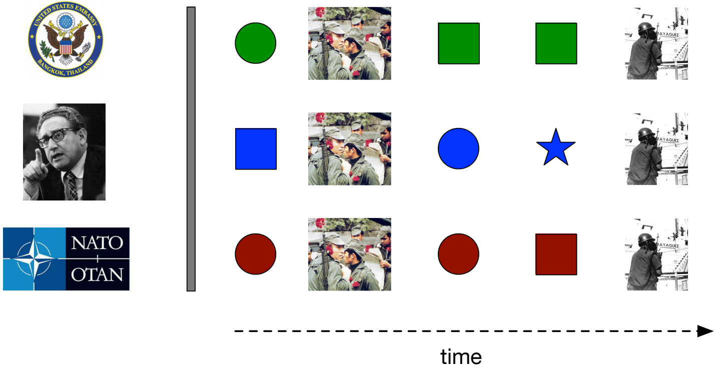
- Historians want to identify important events from primary sources.
- Example: Embassies send cables to each other during the 1970s
- Goal: Use topic models to find events in this data set

right	minister	bank	negotiation
human	depart	exchange	proposal
church	arrive	room	position
freedom	schedule	reserve	agreement
person	party	central	question
inquiry	hour	governor	agree
violation	departure	rate	point
religious	president	finance	negotiate
discrimination	reception	financial	issue
cooper	airport	single	western

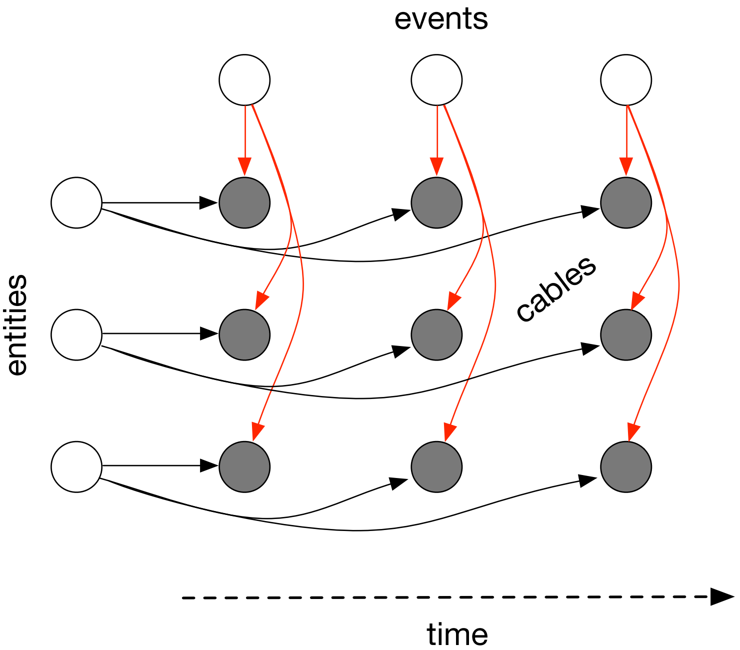
fish	university	build	refugee
bill	student	construction	status
vessel	health	facility	personnel
fishery	medical	plant	resettlement
zone	education	unit	name
airport	professor	extension	family
mile	school	supply	parole
fare	child	area	swiss
dote	american	work	check
water	care	cost	grantee

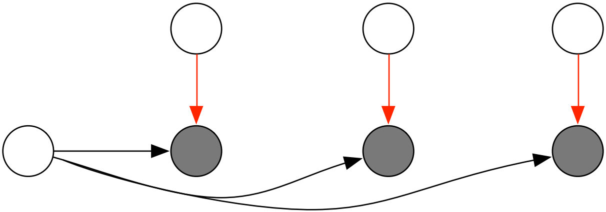
soviet	control	press	arab
moscow	narcotic	article	israeli
ussr	drug	story	east
union	traffic	news	talk
brezhnev	rangoon	interview	egyptian
detente	indonesian	medium	middle
russian	extradition	statement	peace
side	enforcement	coverage	minister
pravda	opium	carry	palestinian
gromyko	attorney	american	settlement

Topic models (by themselves) are a start. But they don't identify events.



- Embassies typically discuss their *usual business*
- When a cable is about an **event**:
  - It diverges from the usual business of the sender
  - Multiple embassies discuss it
- *Usual business* is framed in terms of topics;
  events are framed in terms of words.


events
entities
cables
time

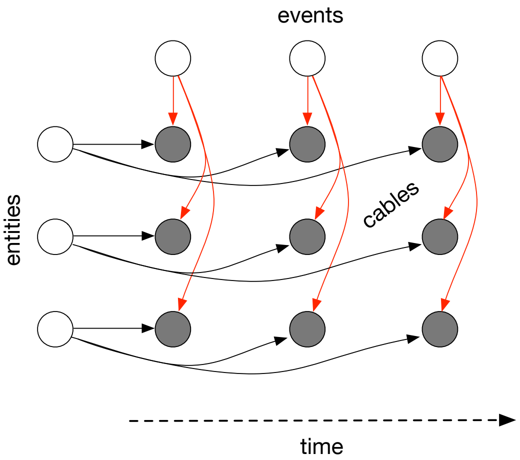


Hidden variables

- Topics
- Event description (per week)
- Topic description (per entity)
- Topic strength, event strength (per cable)

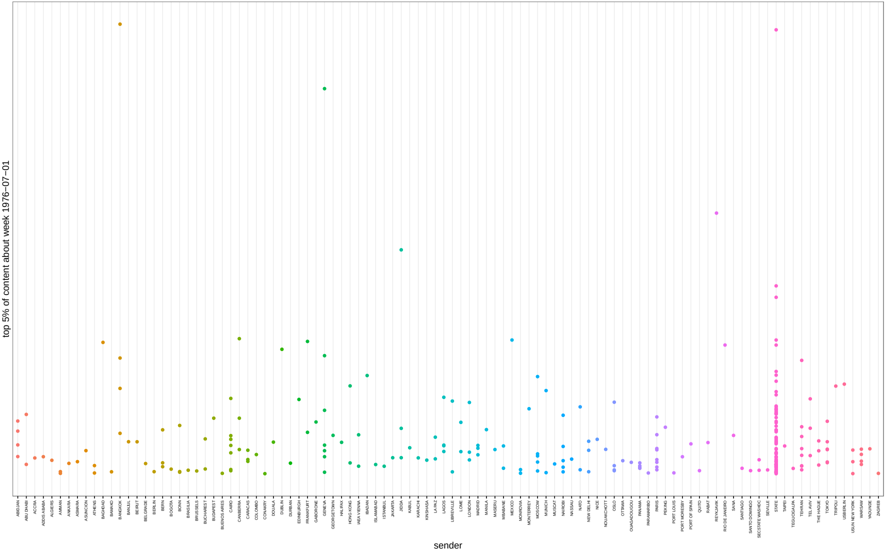
Observed variables

- Cables (per-week, per-entity)



To find events:

- Calculate the posterior of the hidden variables given the observed variables
- Examine cables where the event strength is high


top 5% of content about week 1976−07−01
sender
ABIDJAN
ABU DHABI
ACCRA
ADDIS ABABA
ALGIERS
AMMAN
ANKARA
ASMARA
ASUNCION
ATHENS
BAGHDAD
BAMAKO
BANGKOK
BANJUL
BEIRUT
BELGRADE
BERLIN
BERN
BOGOTA
BONN
BRASILIA
BRUSSELS
BUCHAREST
BUDAPEST
BUENOS AIRES
CAIRO
CANBERRA
CARACAS
COLOMBO
CONAKRY
DOUALA
DUBLIN
DURBAN
EDINBURGH
FRANKFURT
GABORONE
GENEVA
GEORGETOWN
HALIFAX
HONG KONG
IAEA VIENNA
IBADAN
ISLAMABAD
ISTANBUL
JAKARTA
JIDDA
KABUL
KARACHI
KINSHASA
LA PAZ
LAGOS
LIBREVILLE
LOME
LONDON
MADRID
MANILA
MASERU
MBABANE
MEXICO
MONROVIA
MONTERREY
MOSCOW
MUNICH
MUSCAT
NAIROBI
NASSAU
NATO
NEW DELHI
NICE
NOUAKCHOTT
OSLO
OTTAWA
OUAGADOUGOU
PANAMA
PARAMARIBO
PARIS
PEKING
PORT LOUIS
PORT MORESBY
PORT OF SPAIN
QUITO
RABAT
REYKJAVIK
RIO DE JANEIRO
SANA
SANTIAGO
SANTO DOMINGO
SECSTATE WASHDC
SEVILLE
STATE
TAIPEI
TEGUCIGALPA
TEHRAN
TEL AVIV
THE HAGUE
TOKYO
TRIPOLI
USBERLIN
USUN NEW YORK
WARSAW
YAOUNDE
ZAGREB

1. THE FOLLOWING IS A COMPOSITE REPORT ON SOME OF THE ASPECTS OF THE RECENT HIJACKING OF THE AIR FRANCE AIRBUS BASED ON CONVERSATIONS WITH A NUMBER OF THE AMERICANS WHO WERE RELEASED. THEIR ACCOUNTS WERE GENERALLY CONGRUENT AND COINCIDE WITH THE REPORTS FROM THE GERMAN EMBASSY IN KAMPALA (REF. A). SINCE MANY ASPECTS ARE A MATTER OF IMPRESSION RATHER THAN FIRM OBSERVATION, THERE ARE SOME DIFFERENCES IN THE ACCOUNTS, WHICH WE NOTE. WE GO OVER SOME OF THE SAME GROUND COVERED IN THE ACCOUNTS IN REFS B, C AND D. THERE IS SOME CONFUSION OVER NOMENCLATURE FOR THE THREE EIEMENTS INVOLVED IN THE HIJACKING. ALL THE AMERICANS REFERRED TO MEMBERS OF ALL THREE AS PLO PRESUMABLY SINCE ENGLISH SPEAKING GERMANS WITH WHOM THEY HAD MOST CONTACT SAID THEY WORKED FOR PLO. BUT THE ARABS ON THE GROUND WHO WERE APPARENTLY IN CHARGE OF THE WHOLE CHE GUEVARA COMMANDO GROUP MADE THE DISTINCTION AND SPECIFIED THEY WORKED FOR PFLP AND NOT PLO. GERMANS PERHAPS DID NOT UNDERSTAND THE DISTINCTION. 2. RETAINED AMERICANS. SOME OF THE AMERICANS HAD IDENTIFIED GEORGE AND RENE KARFUNKEL AS AMERICANS AND WERE AWARE THAT THEY HAD ONLY AMERICAN PASSPORTS AND HAD BEEN IN ISRAEL ONLY BRIEFLY. NO ONE HAD A SATISFACTORY EXPLANATION OF WHY THE KARFUNKELS HAD NOT BEEN RELEASED WITH THE OTHER AMERICANS. THE KARFUNKELS APPEARED VERY ORTHODOX, ATE ONLY KOSHER FOOD...

1. FRG INFORMED EMBASSY AT 2130 LOCAL THAT FRENCH GOVERNMENT HAS PASSED IT THE FOLLOWING MESSAGE: A. THE HIJACKERS REJECT ANY EXCHANGE OUTSIDE OF ENTEBBE AIRPORT. B. THE EXCHANGE MUST TAKE PLACE UNDER THE SUPERVISION OF AMIN (OR OTHER HIGH UGANDAN OFFICIAL), TWO FRENCH REPS, AND SOMALI AMBASSADOR. C. HIJACKERS NOT PREPARED TO DISTINGUISH BETWEEN THE PRISONERS HELD IN ISRAEL AND THOSE HELD IN OTHER COUNTRIES. D. ALL HOSTAGES MUST BE EXCHANGED AGAINST ALL THE PRISONERS. E. THE HIJACKERS EXPECT AN ANSWER FROM ALL FOUR COUNTRIES HOLDING PRISONERS, NOT ONLY ISRAEL. F. THE HIJACKERS REFER TO THE COMPLETE LIST OF 53 "COMRADES." G. THE HIJACKERS INSIST AGAIN ON A PACKAGE DEAL: 53 "COMRADES" AGAINST ALL THE HOSTAGES AT ENTEBBE AIRPORT.

2. THE FRENCH PASSED A SECOND MESSAGE, THIS ONE FROM AMIN: AMIN TOLD THE FRENCH AMBASSADOR THAT HE EXPECTS ALL FOUR COUNTRIES TO COMMUNICATE TO HIM THE FLIGHT NUMBERS AND ETA OF ALL AIRCRAFT BRINGING PRISONERS TO UGANDA BEFORE THE END OF THE ULTIMATUM. (HE DID NOT SPECIFY AN HOUR.)

3. THE FRG CRISIS CENTER TOLD US THE FRG IS CONSULTING WITH THE OTHER GOVERNMENTS INVOLVED AT THE HIGHEST LEVEL. IT HAS NOT RPT NOT REACHED A DECISION ON RELEASE OF PRISONERS. HILLENBRAND

1976-07-04 | TEL AVIV | STATE | U.S. CITIZENS FROM HIJACKED AIR FRANCE PLANE

1. ACCORDING TO AS YET PROVISIONAL INFORMATION OBTAINED FROM IDF SPOKESMAN AND PASSENGERS, THERE WERE FOUR AMERICAN CITIZENS AMONG PASSENGERS FROM HIJACKED AIR FRANCE FLIGHT BROUGHT TO ISRAEL ON JULY 4. THEY ARE GEORGE AND RENE GARFUNKEL OF NEW YORK CITY; MRS. JANETTE ALMOG (HUSBAND ESRA ALMOG IS ISRAELI CITIZEN) OF MADISON WISCONSIN; AND MOSHE PERES OF NEW HAVEN CONNECTICUT. ALL RETURNED PASSENGERS ARE REPORTED WELL IN TEL AVIV. THOSE IN TRANSIT ARE BEING HOUSED AT PLAZA HOTEL IN TEL AVIV.

2. PASSENGERS KILLED DURING LIBERATION WERE JEAN-JACQUES MIMOUNI REPORTEDLY OF FRENCH NATIONALITY AND MRS IDA BOROCHOWITZ, AN ISRAELI OF RUSSIAN ORIGIN. ONE ISRAELI SOLDIER DIES IN FIGHT. NINE PERSONS REQUIRING MEDICAL CARE WERE LEFT IN NAIROBI DURING BRIEF STOP-OVER.

3. ACCORDING TO THE GARFUNKELS, FRENCH PILOT SAID THAT HIJACKERS ENTERED EMBARKATION AREA WITH SIX PACKAGES CLAIMING THAT THEY CONTAINED CANDY. AT TIME OF THEIR ENTRY ELECTRICITY ALLEGEDLY WENT OUT AND STOPPED SCREENING DEVICES FROM WORKING. RATHER THAN DELAY THE PLANE FOR EXAMINATION OF PACKAGES, THE SIX WERE HURRIED ON BOARD.

4. GARFUNKELS ARE LEAVING JULY 5 ON EL AL...

The way it really happened~The tense, action~packed story of the raid that startled the world.

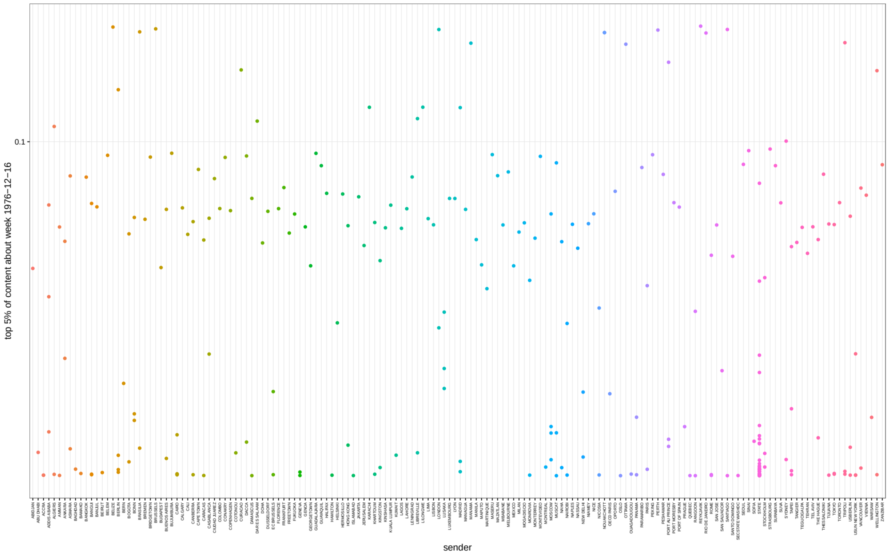
"OPERATION THUNDERBOLT" A

An INTER-OCEAN film "OPERATION THUNDERBOLT" A Golan-Globus Film of a G.S. Films Production
starring YEHORAM GAON · ASSAF DAYAN · KLAUS KINSKY SYBIL DANNING · ORI LEVY · ARIK LAVI and MARK HEATH as Idi Amin
Produced by MENAHEM GOLAN and YORAM GLOBUS Music by DOV SELTZER Directed by MENAHEM GOLAN Eastmancolour Distributed by EMI Films Limited

EMI

FROM THURS. OCT. 20th ABC 1 **Shaftesbury Ave** Tel: 836 8861 Licensed Bar

AND AT SELECTED ABC AND OTHER LEADING CINEMAS FROM OCT. 27th. SEE LOCAL PRESS FOR DETAILS


top 5% of content about week 1976−12−16
0.1
sender
ABIDJAN
ABU DHABI
ACCRA
ADDIS ABABA
ALGIERS
AMMAN
ANKARA
ASMARA
BAGHDAD
BAMAKO
BANGKOK
BANGUI
BANJUL
BEIRUT
BELEM
BELIZE
BERLIN
BERN
BOGOTA
BONN
BRASILIA
BREMEN
BRIDGETOWN
BRUSSELS
BUDAPEST
BUENOS AIRES
BUJUMBURA
CAIRO
CALGARY
CALI
CANBERRA
CAPE TOWN
CARACAS
CASABLANCA
CIUDAD JUAREZ
COLOMBO
CONAKRY
COPENHAGEN
COTONOU
CURACAO
DACCA
DAMASCUS
DAR ES SALAAM
DOHA
DUSSELDORF
EC BRUSSELS
FLORENCE
FRANKFURT
FREETOWN
FUKUOKA
GENEVA
GENOA
GEORGETOWN
GUADALAJARA
GUAYAQUIL
HALIFAX
HAMILTON
HELSINKI
HERMOSILLO
HONG KONG
ISLAMABAD
JAKARTA
JERUSALEM
KARACHI
KHARTOUM
KINGSTON
KINSHASA
KUALA LUMPUR
KUWAIT
LAGOS
LAHORE
LENINGRAD
LIBREVILLE
LILONGWE
LIMA
LISBON
LONDON
LUSAKA
LUXEMBOURG
LYON
MADRID
MANAGUA
MANAMA
MANILA
MAPUTO
MARTINIQUE
MASERU
MAZATLAN
MBABANE
MELBOURNE
MEXICO
MILAN
MOGADISCIO
MONROVIA
MONTERREY
MONTEVIDEO
MONTREAL
MOSCOW
MUSCAT
NAHA
NAIROBI
NAPLES
NASSAU
NEW DELHI
NIAMEY
NICE
NICOSIA
NOUAKCHOTT
OECD PARIS
OPORTO
OSLO
OTTAWA
OUAGADOUGOU
PANAMA
PARAMARIBO
PARIS
PEKING
PERTH
PESHAWAR
PORT AU PRINCE
PORT MORESBY
PORT OF SPAIN
PRAGUE
QUEBEC
RANGOON
REYKJAVIK
RIO DE JANEIRO
ROME
SAN JOSE
SAN SALVADOR
SANTIAGO
SANTO DOMINGO
SECSTATE WASHDC
SEOUL
SINAI
SOFIA
STATE
STOCKHOLM
STRASBOURG
SURABAYA
SUVA
SYDNEY
TAIPEI
TANGIER
TEGUCIGALPA
TEHRAN
TEL AVIV
THE HAGUE
THESSALONIKI
TIJUANA
TOKYO
TORONTO
TRIPOLI
USBERLIN
USUN NEW YORK
VANCOUVER
VIENNA
WARSAW
WELLINGTON
ZANZIBAR

1976-12-20 | STATE | ALL POSTS | GAMING

1. THE GENERAL ACCOUNTING OFFICE IS ACTING UPON A REQUEST, BY THE HOUSE GOVERNMENT OPERATIONS COMMITTEE, TO GATHER INFORMATION ON THE OPERATION OF GAMING DEVICES IN ALL EMBASSY AND CONSULATE FACILITIES. GAMING DEVICES INCLUDE SLOT MACHINES, WHEELS OF CHANCE, DICE GAMES, ETC. BUT SPECIFICALLY EXCLUDE BINGO.

2. WE NEED TO HAVE FOLLOWING INFORMATION BY PRIORITY CABLE NO LATER THAN DECEMBER 30. REPLIES SHOULD BE DIRECTED TO LEAMON R. HUNT, DEPUTY ASSISTANT SECRETARY FOR OPERATIONS. NEGATIVE RESPONSES ARE REQUIRED. A. NUMBER OF GAMING DEVICES, WHERE LOCATED, AND WHO MANAGES THEM.

B. OPERATING POLICIES.

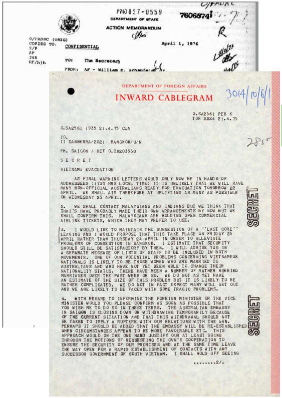
C. WHETHER LEASING OR CONCESSIONS ARE PERMITTED, TO INCLUDE IDENTIFYING MACHINES LEASED OR BELONGING TO CONCESSIONAIRES.

D. NUMBER OF DEVICES PURCHASED AND PROCUREMENT POLICIES.

E. AMOUNT OF PROFITS DERIVED.

3. WE APPRECIATE YOUR PROMPT ATTENTION TO THIS REQUEST. KISSINGER

- 1976-12-21 | BANJUL | STATE | GAMING DEVICES
  THERE ARE NO GAMING DEVICES IN EMBASSY FACILITIES.WYGANT
- 1976-12-22 | BREMEN | STATE | GAMING DEVICES
  NO RPT NO GAMING DEVICES OF ANY TYPE AT AMCONSUL BREMEN. LONGMYER
- 1976-12-21 | BANGUI | STATE | GAMING DEVICES
  THERE ARE NO RPT NO GAMING DEVICES IN EMBASSY FACILITIES. QUAINTON
- 1976-12-22 | BELIZE | STATE | GAMING DEVICES
  NEGATIVE RESPONSE. WALSH
- 1976-12-21 | ABIDJAN | STATE | GAMING DEVICES
  FOR: LEAMON R. HUNT, DEPUTY ASSISTANT SECRETARY FOR OPERATIONS NO GAMING DEVICES EXIST AT THIS POST. STEARNS
- 1976-12-21 | ASMARA | STATE | GAMING DEVICES
  CONSULATE GENERAL ASMARA NEITHER OPERATES OR OWNS ANY GAMING DEVICES. WAUCHOPE
- 1976-12-21 | AMMAN | STATE | GAMING DEVICES
  EMBASSY AMMAN SUBMITS NEGATIVE REPORT ON AVAILABILITY OF GAMING DEVICES WITHIN MISSION. PICKERING


P760057-0559

DEPARTMENT OF STATE

ACTION MEMORANDUM

7606874

O/FADRC (ORIG)
COPIES TO:
S/P
AF
INR
RF/hjb

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. Schaufele Jr.

DEPARTMENT OF FOREIGN AFFAIRS

INWARD CABLEGRAM

3014/10/6/1

O.SA2561 PEB 6
TOR 2224 21.4.75

O.SA2561 1935 21.4.75 CLA

TO.
II CANBERRA/2021 BANGKOK/U/N

FM. SAIGON / REF O.CH203930

S E C R E T

VIETNAM: EVACUATION

AS FINAL WARNING LETTERS WOULD ONLY NOW BE IN HANDS OF ADDRESSEES (1730 HRS LOCAL TIME) IT IS UNLIKELY THAT WE WILL HAVE MANY NON-OFFICIAL AUSTRALIANS READY FOR EVACUATION TOMORROW 22 APRIL. WE SHALL AIM THEREFORE AT UPLIFTING AS MANY AS POSSIBLE ON WEDNESDAY 23 APRIL.

2. WE SHALL CONTACT MALAYSIANS AND INDIANS BUT WE THINK THAT THAI'S HAVE PROBABLY MADE THEIR OWN ARRANGEMENTS BY NOW BUT WE SHALL CONFIRM THIS. MALAYSIANS ARE HOLDING OPEN COMMERCIAL AIRLINE TICKETS, WHICH THEY MAY PREFER TO USE.

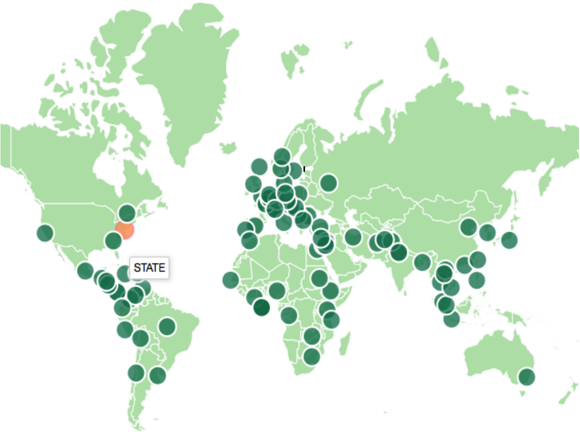
3. I WOULD LIKE TO MAINTAIN THE SUGGESTION OF A ''LAST CORE'' LEAVING AND I WOULD PROPOSE THAT THIS TAKE PLACE ON FRIDAY 25 APRIL RATHER THAN THURSDAY 24 APRIL IN ORDER TO ALLEVIATE PROBLEMS OF CONGESTION IN BANGKOK. I ESTIMATE THAT SECURITY SHOULD STILL BE SATISFACTORY BY THEN. I WILL ADVISE YOU IN A SEPARATE MESSAGE OF LISTS OF STAFF TO BE INCLUDED IN BOTH MOVEMENTS. ONE OF OUR POTENTIAL PROBLEMS CONCERNING VIETNAMESE NATIONALS IS LIKELY TO BE THOSE WOMEN WHO ARE MARRIED TO AUSTRALIANS AND WHO HAVE NOT YET BEEN ABLE TO CHANGE THEIR NATIONALITY STATUS. THERE HAVE BEEN A NUMBER OF RATHER HURRIED MARRIAGES OVER THE PAST WEEK OR SO. WE DO NOT AS YET HAVE AN ESTIMATE OF THE SIZE OF THIS PROBLEM BUT IT IS LIKELY TO BE RATHER COMPLICATED. WE DO NOT IN FACT EXPECT MANY WILL GET OUT AND WE ARE LIKELY TO BE FACED WITH SOME TRAGIC PROBLEMS.

4. WITH REGARD TO INFORMING THE FOREIGN MINISTER OR THE VICE MINISTER WOULD YOU PLEASE CONFIRM AS SOON AS POSSIBLE THAT YOU WISH ME TO DO SO IN THESE TERMS. THE AUSTRALIAN EMBASSY IN SAIGON IS CLOSING DOWN OR WITHDRAWING TEMPORARILY BECAUSE OF THE CURRENT SITUATION AND THAT THIS WITHDRAWAL SHOULD NOT BE TAKEN TO IMPLY A RUPTURE WITH OUR RELATIONS WITH THE GVN. PERHAPS IT SHOULD BE ADDED THAT THE EMBASSY WILL BE RE-ESTABLISHED WHEN CIRCUMSTANCES APPEAR TO BE MORE FAVOURABLE ETC. THIS APPROACH WOULD ON THE ONE HAND JUSTIFY OUR AT LEAST GOING THROUGH THE MOTIONS OF REQUESTING THE GVN'S COOPERATION TO ENSURE THE SECURITY OF OUR PREMISES AND AT THE SAME TIME LEAVE THE WAY OPEN FOR A RAPID ESTABLISHMENT OF CONTACTS WITH ANY SUCCESSOR GOVERNMENT OF SOUTH VIETNAM. I SHALL HOLD OFF SEEING

........2/.

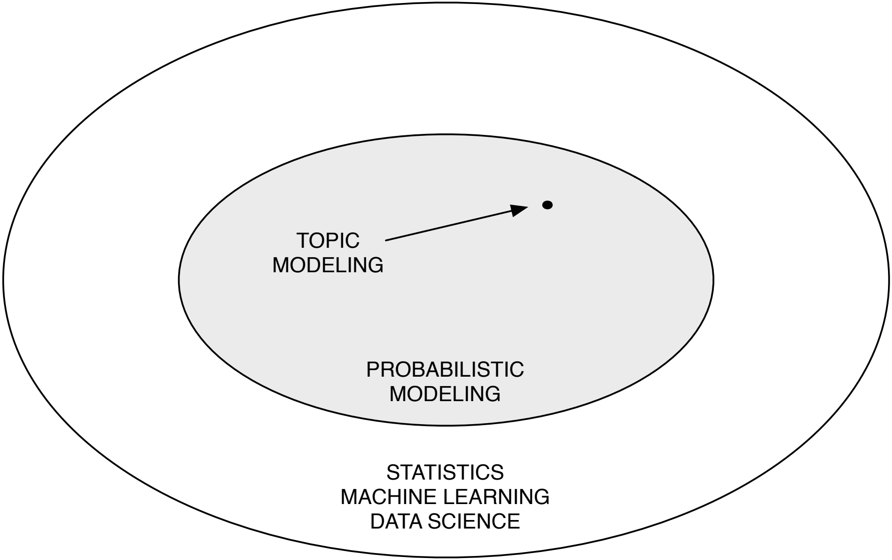
SECRET

SECRET

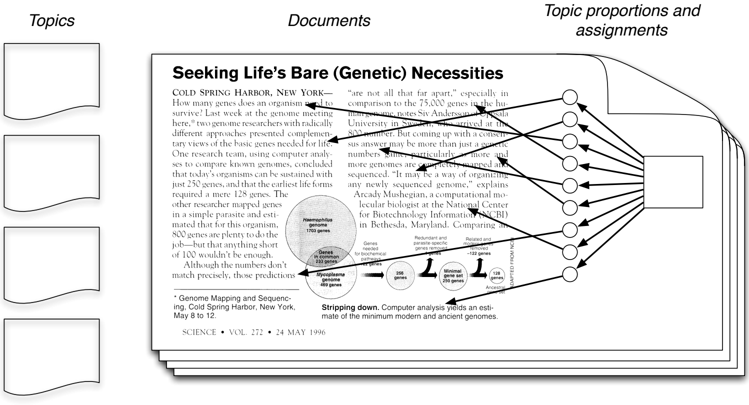


- Topic models can help us find events
- Extensions:
  - Network characteristics
  - Better characterize an event for fewer “false positives”
  - Word embeddings
  - Autocorrelated time series

Discussion: Modern Probabilistic Modeling


TOPIC
MODELING
PROBABILISTIC
MODELING
STATISTICS
MACHINE LEARNING
DATA SCIENCE

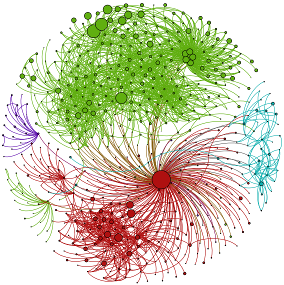
I. Assume our data come from a model with hidden patterns at work

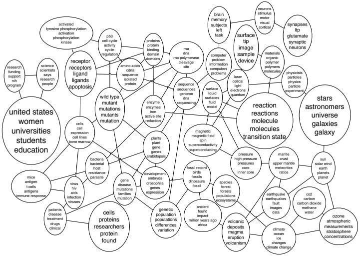


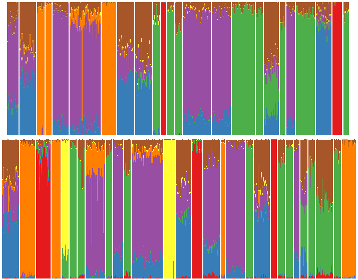
II. Discover those patterns from data

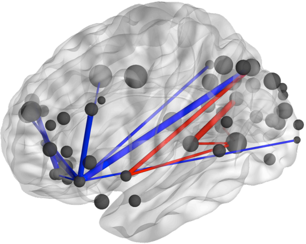
$$\boxed{\nu^* = \arg\max_{\nu} \ \mathbb{E}_q \left[\log p(x, z, \beta \mid \alpha)\right] + \mathbb{H}\left[q(z, \beta \mid \nu)\right]}$$

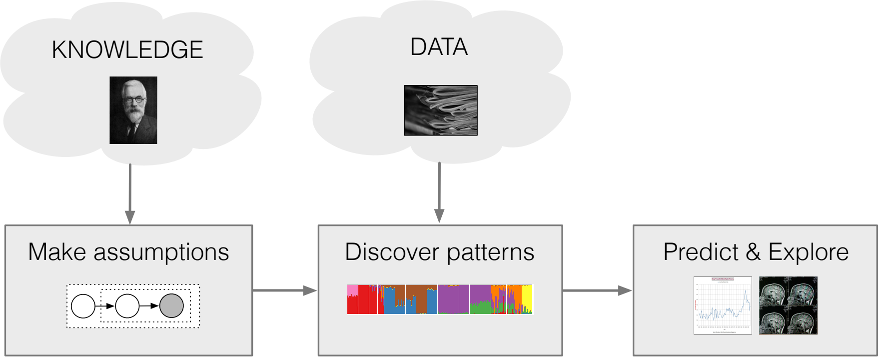
III. Use the discovered patterns to predict about and explore the data





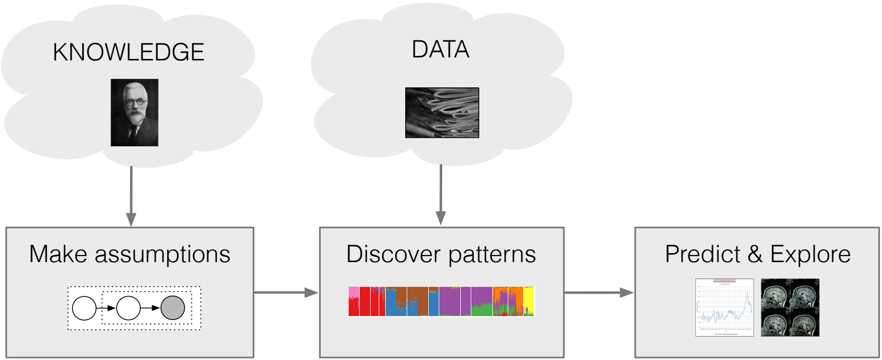






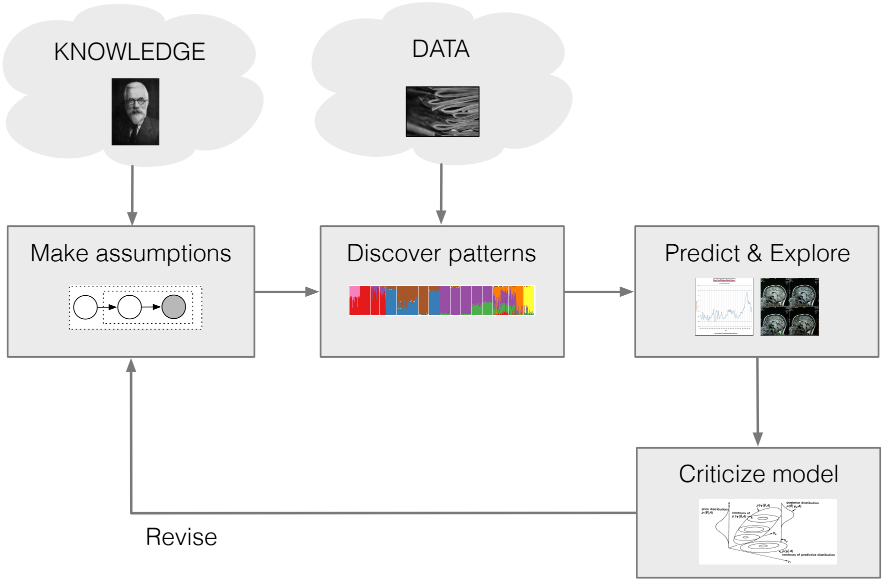
Our perspective:

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



What we need in probabilistic ML:

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


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

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)