

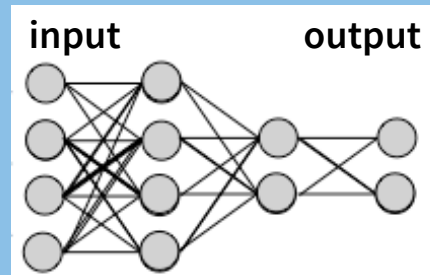
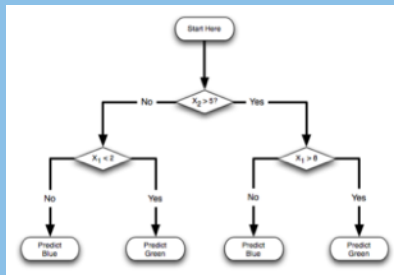
Detecting and Characterizing Events

Unsupervised Machine Learning for Social Science

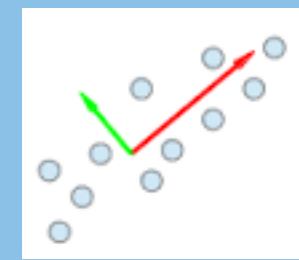
Allison J.B. Chaney
Princeton University

methods

supervised learning

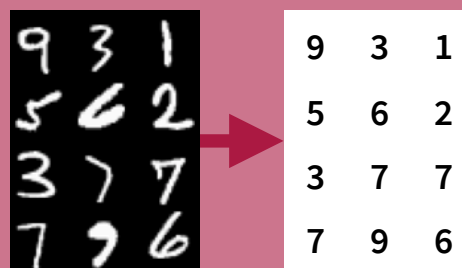


unsupervised learning

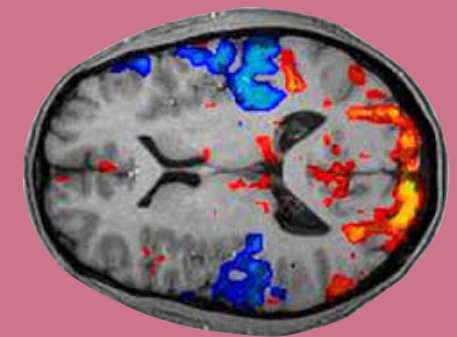


goals

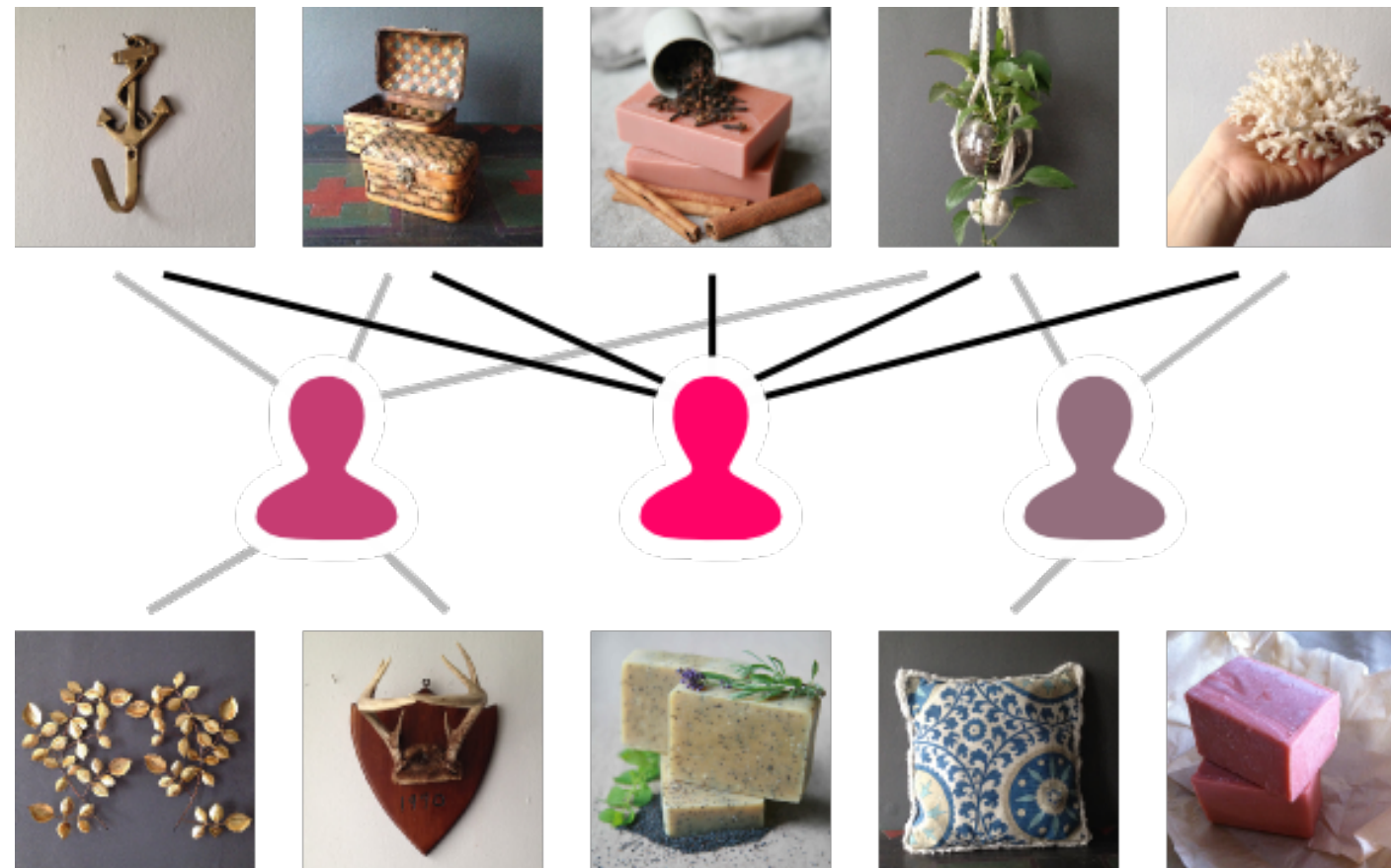
prediction



exploration



Examples of Unsupervised ML for Social Science: Recommendation Systems



A Probabilistic Model for Using Social Networks in Personalized Item Recommendation.

Chaney, Blei, and Eliassi-Rad. RecSys, 2015.

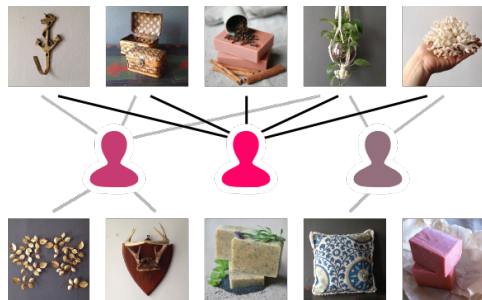
A Large-scale Exploration of Group Viewing Patterns.

Chaney, Gartrell, Hofman, Guiver, Koenigstein, Kohli, and Paquet. TVX, 2014.

How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility

Chaney, Stewart, Engelhardt. arXiv, 2017.

Examples of Unsupervised ML for Social Science: Recommendation Systems

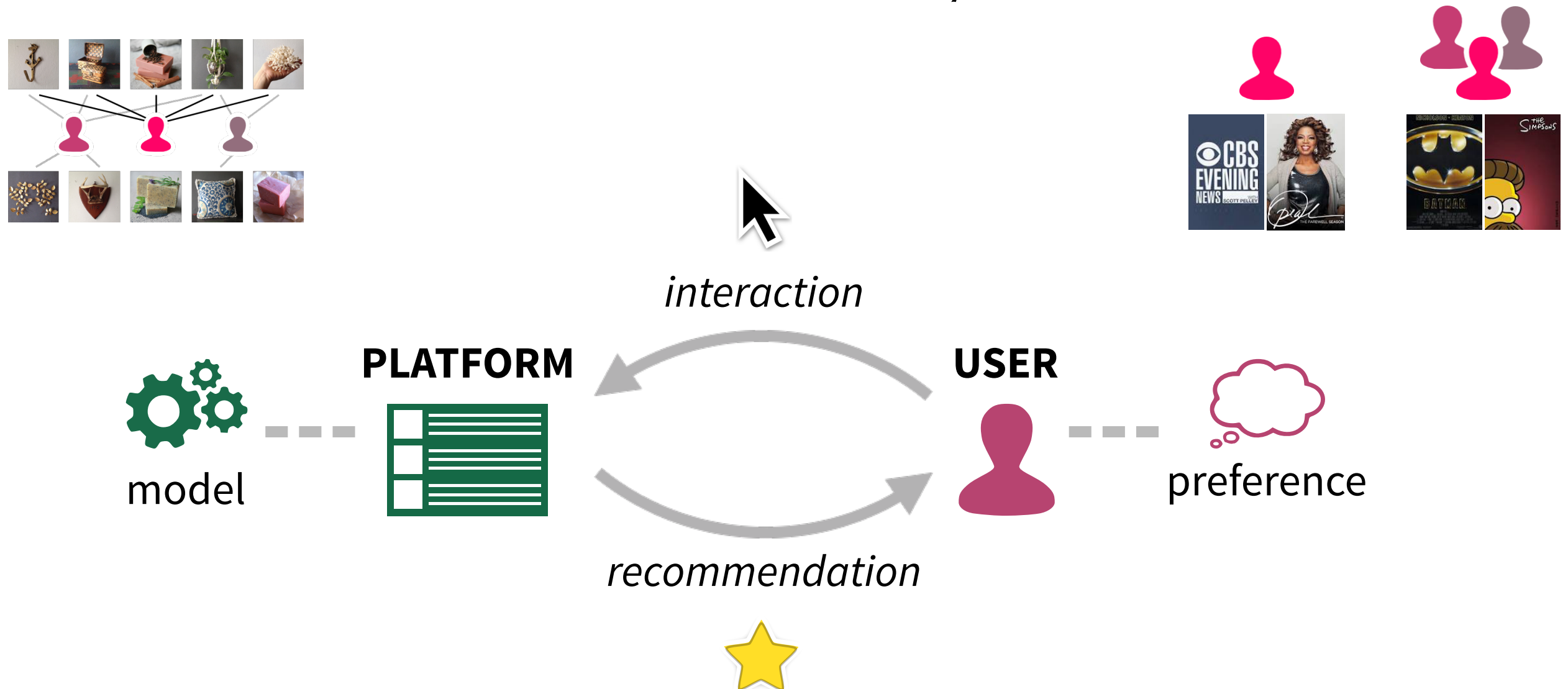


A Probabilistic Model for Using Social Networks in Personalized Item Recommendation.
Chaney, Blei, and Eliassi-Rad. RecSys, 2015.

A Large-scale Exploration of Group Viewing Patterns.
Chaney, Gartrell, Hofman, Guiver, Koenigstein, Kohli, and Paquet. TVX, 2014.

How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility
Chaney, Stewart, Engelhardt. arXiv, 2017.

Examples of Unsupervised ML for Social Science: Recommendation Systems

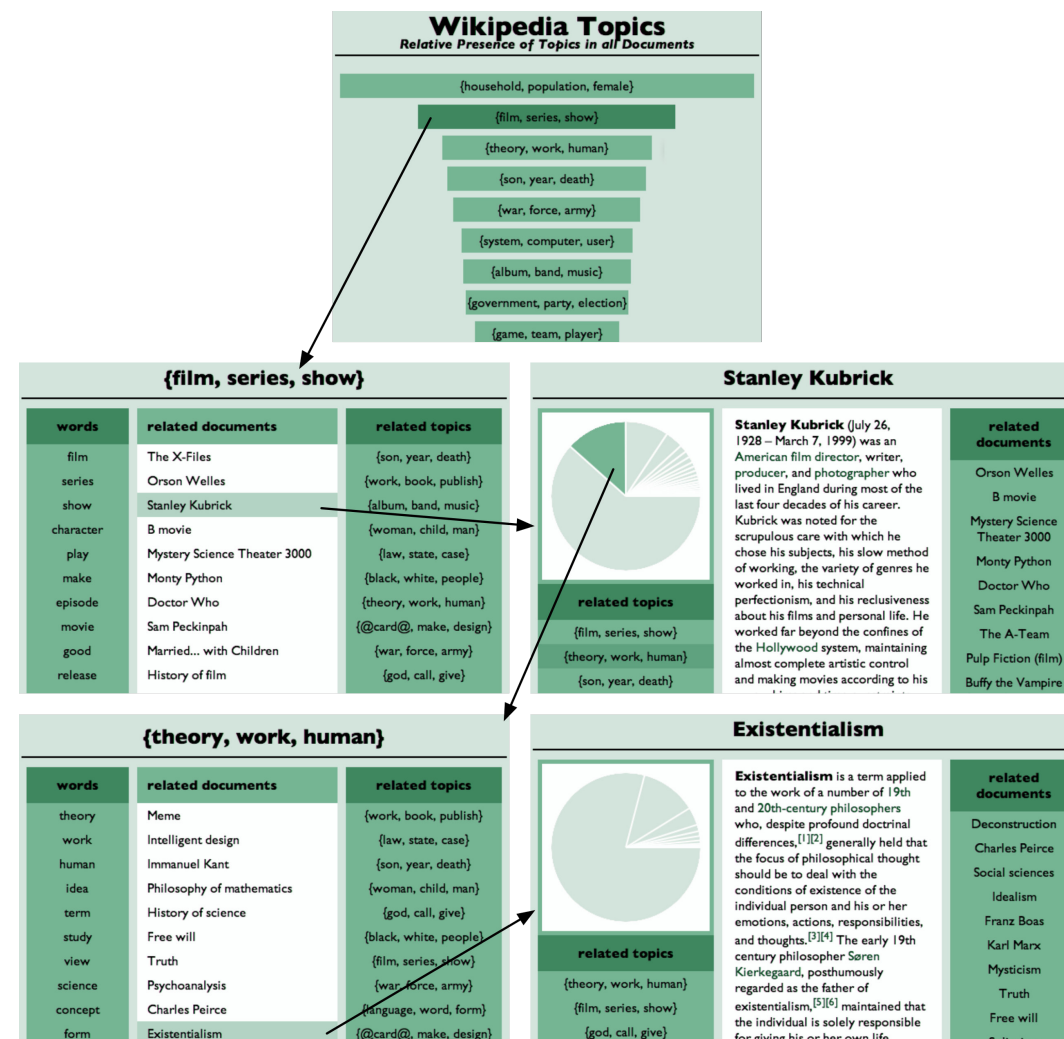


A Probabilistic Model for Using Social Networks in Personalized Item Recommendation.
Chaney, Blei, and Eliassi-Rad. RecSys, 2015.

A Large-scale Exploration of Group Viewing Patterns.
Chaney, Gartrell, Hofman, Guiver, Koenigstein, Kohli, and Paquet. TVX, 2014.

How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility
Chaney, Stewart, Engelhardt. arXiv, 2017.

Examples of Unsupervised ML for Social Science: Text Analysis / Topic Models



Visualizing topic models.

Chaney and Blei. ICWSM, 2012.

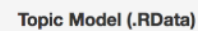
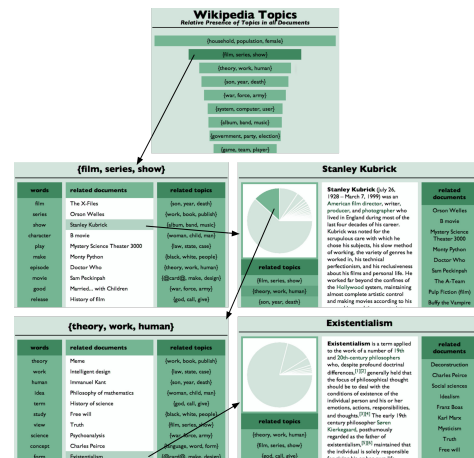
The Power of Aggregation for Topic Models Used For Measurement.

Chaney, Shiraito, Stewart. Text as Data, 2017.

Detecting and Characterizing Events.

Chaney, Wallach, Blei, and Connelly. EMNLP, 2016.

Text Analysis / Topic Models



Select poliblogs2008.K100.RData

Upload complete

Number of clusters

20

Download

Topic Summary

econom economi financi market wi

Following a nearly 3% rise in home sales in February that raised hopes that the housing crisis was b ...

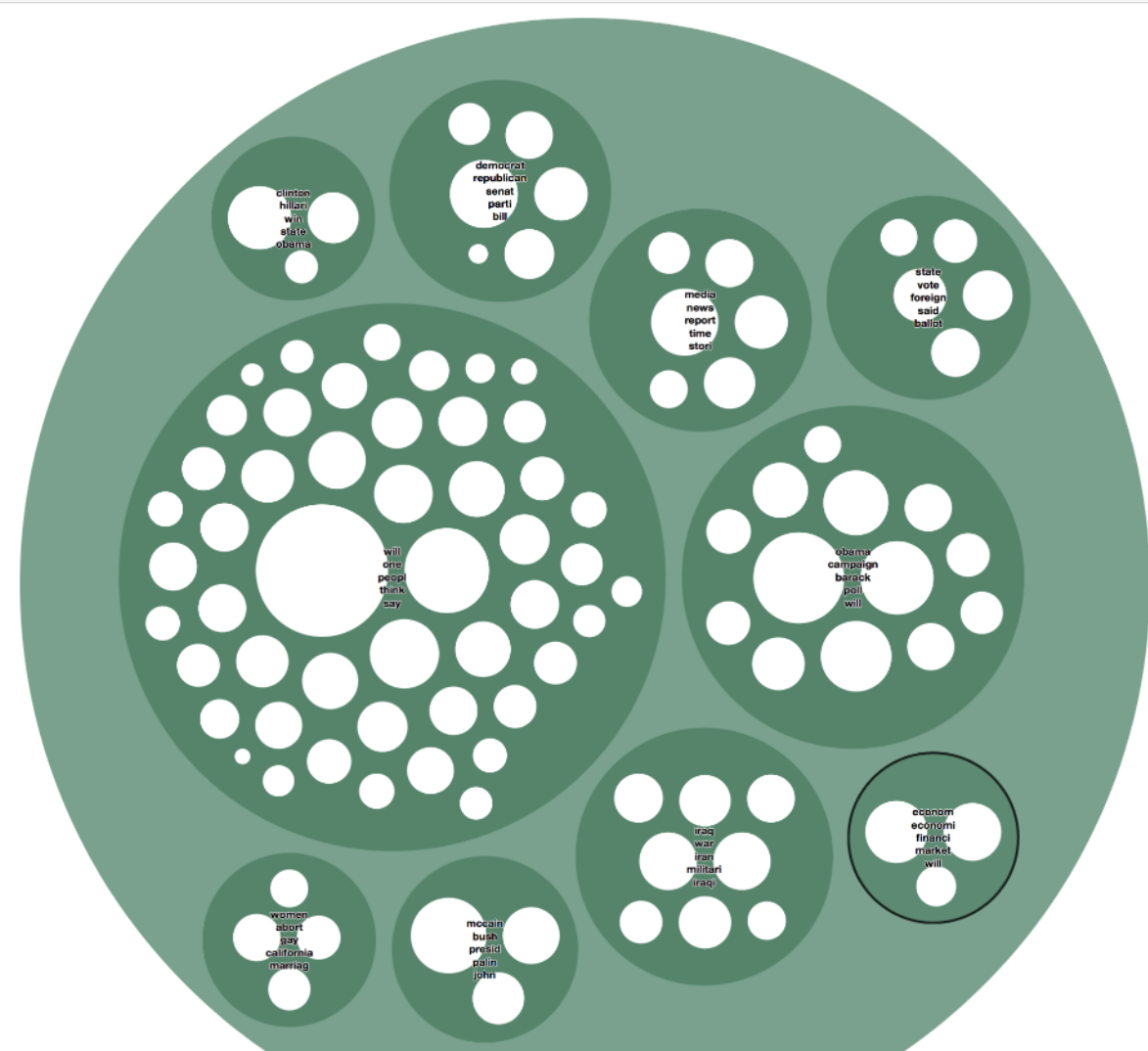
It would be the mother of all bailouts if it came to pass - hundreds of billions of taxpayer dollars ...

My syndicated column today updates what used to be one of my favorite Aesop's fables: "The Ant and t ...

Housing foreclosures are up 65% compared to last year indicating that the housing slowdown and credi ...

Not a catastrophe and there were some silver linings in the dark clouds on unemployment. But overall ...

Another month, another record drop in home prices and there appears to be no relief in sight:U.S. ho ...



Visualizing topic models.

Chaney and Blei. ICWSM, 2012.

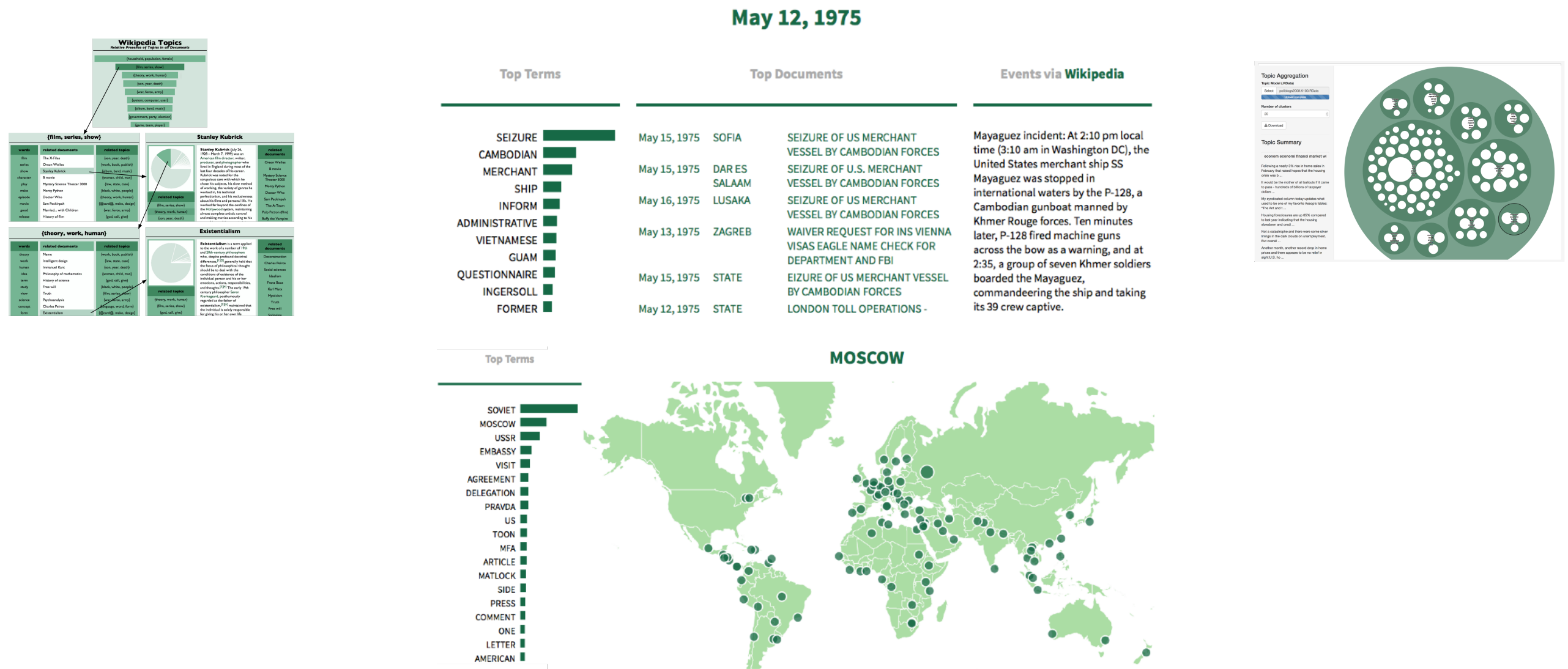
The Power of Aggregation for Topic Models Used For Measurement.

Chaney, Shiraito, Stewart. Text as Data, 2017.

Detecting and Characterizing Events.

Chaney, Wallach, Blei, and Connolly. EMNLP, 2016.

Examples of Unsupervised ML for Social Science: Text Analysis / Topic Models

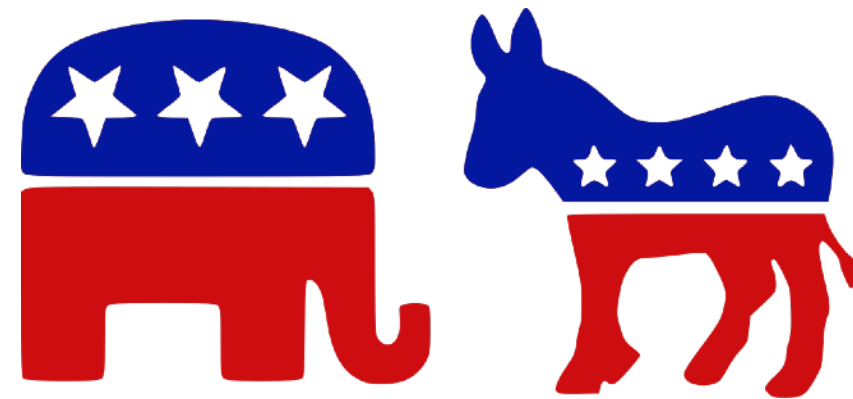


Visualizing topic models.
Chaney and Blei. ICWSM, 2012.

The Power of Aggregation for Topic Models Used For Measurement.
Chaney, Shiraito, Stewart. Text as Data, 2017.

Detecting and Characterizing Events.
Chaney, Wallach, Blei, and Connelly. EMNLP, 2016.

Why are events important?



Our Task:

Given a huge corpus of primary source documents,
identify events of potential interest and
characterize them with relevant words and sources.

(Or, make life easier for historians dealing with
millions of primary source documents.)



Matthew Connelly's History Lab at Columbia

U.S. State Department Cables

Message content

Date sent

Authoring entity

...

7606374

DEPARTMENT OF STATE
ACTION MEMORANDUM

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

April 1, 1976

TO: The Secretary
FROM: AF - William K. Schaufele, Jr.

Rhodesia - A Proposed Course of Action

The Problem

Any moves we make with respect to Rhodesia must be in concert with an over-all Southern African strategy which will consider our relationships with the Soviet Union, Cuba, our European allies, black African countries, China and South Africa. S/P will shortly present you with a paper on this broader conceptualization of a rationale of action regarding the Southern African problem. However, in terms of immediacy, we should focus on Rhodesia - on what we can do in the near future to preempt the Soviets and Cubans, improve our position in Africa, and possibly help avert widespread, intensified conflict in Rhodesia itself.

The Strategic Imperatives

The Soviet/Cuban intervention in Angola has drastically affected the determinants of our policies toward Rhodesia. Our essentially passive stance no longer is the most appropriate approach. Preclusion of further expansion of Soviet/Cuban presence and influence in Southern Africa will require us to become more directly involved there than we might prefer.

As you have pointed out, failure to try to counter Soviet/Cuban designs in Southern Africa would contribute to the belief that the U. S. is unwilling

CONFIDENTIAL
GDS

9/15/90



Matthew Connelly's History Lab at Columbia

U.S. State Department Cables

2,674,486 messages

sent between 1973 and 1978

34,204 unique sending entities

7606374

DEPARTMENT OF STATE
ACTION MEMORANDUM

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

April 1, 1976

TO: The Secretary
FROM: AF - William K. Schaufele, Jr.

Rhodesia - A Proposed Course of Action

The Problem

Any moves we make with respect to Rhodesia must be in concert with an over-all Southern African strategy which will consider our relationships with the Soviet Union, Cuba, our European allies, black African countries, China and South Africa. S/P will shortly present you with a paper on this broader conceptualization of a rationale of action regarding the Southern African problem. However, in terms of immediacy, we should focus on Rhodesia - on what we can do in the near future to preempt the Soviets and Cubans, improve our position in Africa, and possibly help avert widespread, intensified conflict in Rhodesia itself.

The Strategic Imperatives

The Soviet/Cuban intervention in Angola has drastically affected the determinants of our policies toward Rhodesia. Our essentially passive stance no longer is the most appropriate approach. Preclusion of further expansion of Soviet/Cuban presence and influence in Southern Africa will require us to become more directly involved there than we might prefer.

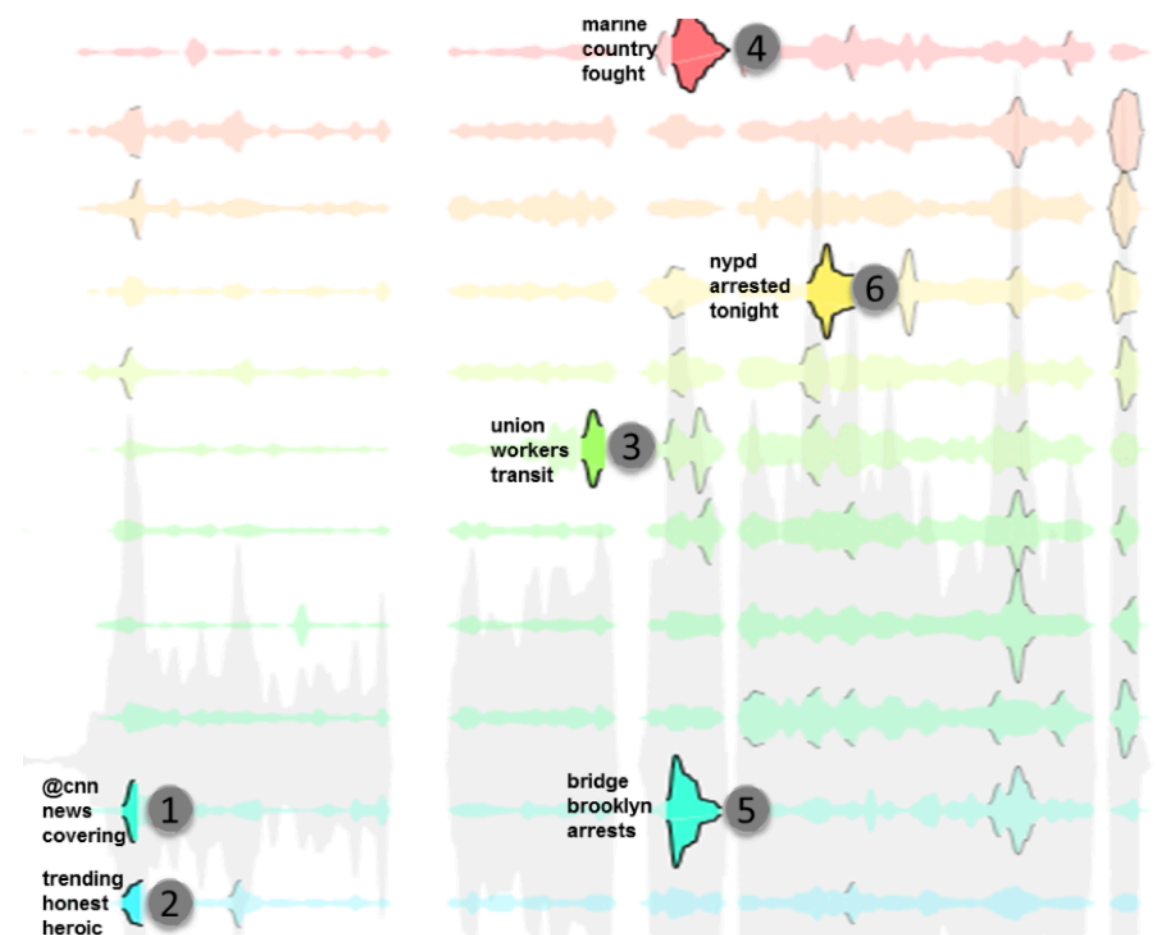
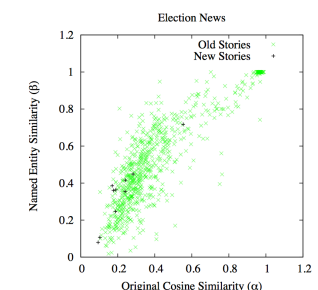
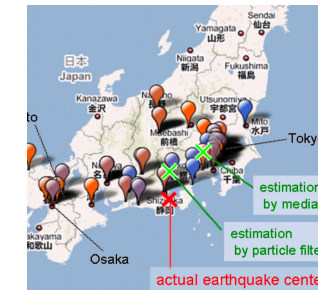
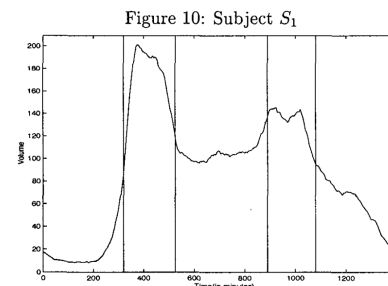
As you have pointed out, failure to try to counter Soviet/Cuban designs in Southern Africa would contribute to the belief that the U. S. is unwilling

CONFIDENTIAL
GDS

9/15/90

What is an event?

- **Event detection from time series data.**
Guralnik and Srivastava. KDD, 1999.
- **Earthquake shakes Twitter users: real-time event detection by social sensors.**
Sakaki, et al. WWW, 2010.
- **A study of retrospective and on-line event detection.** Yang, et al. SIGIR, 1998.
- **Text classification and named entities for new event detection.**
Kumaran and Allan. SIGIR, 2004.
- **A novel burst-based text representation model for scalable event detection.**
Zhao, et al. ACL, 2012.
- **Leadline: Interactive visual analysis of text data through event identification and exploration.** Dou, et al. VAST, 2012.



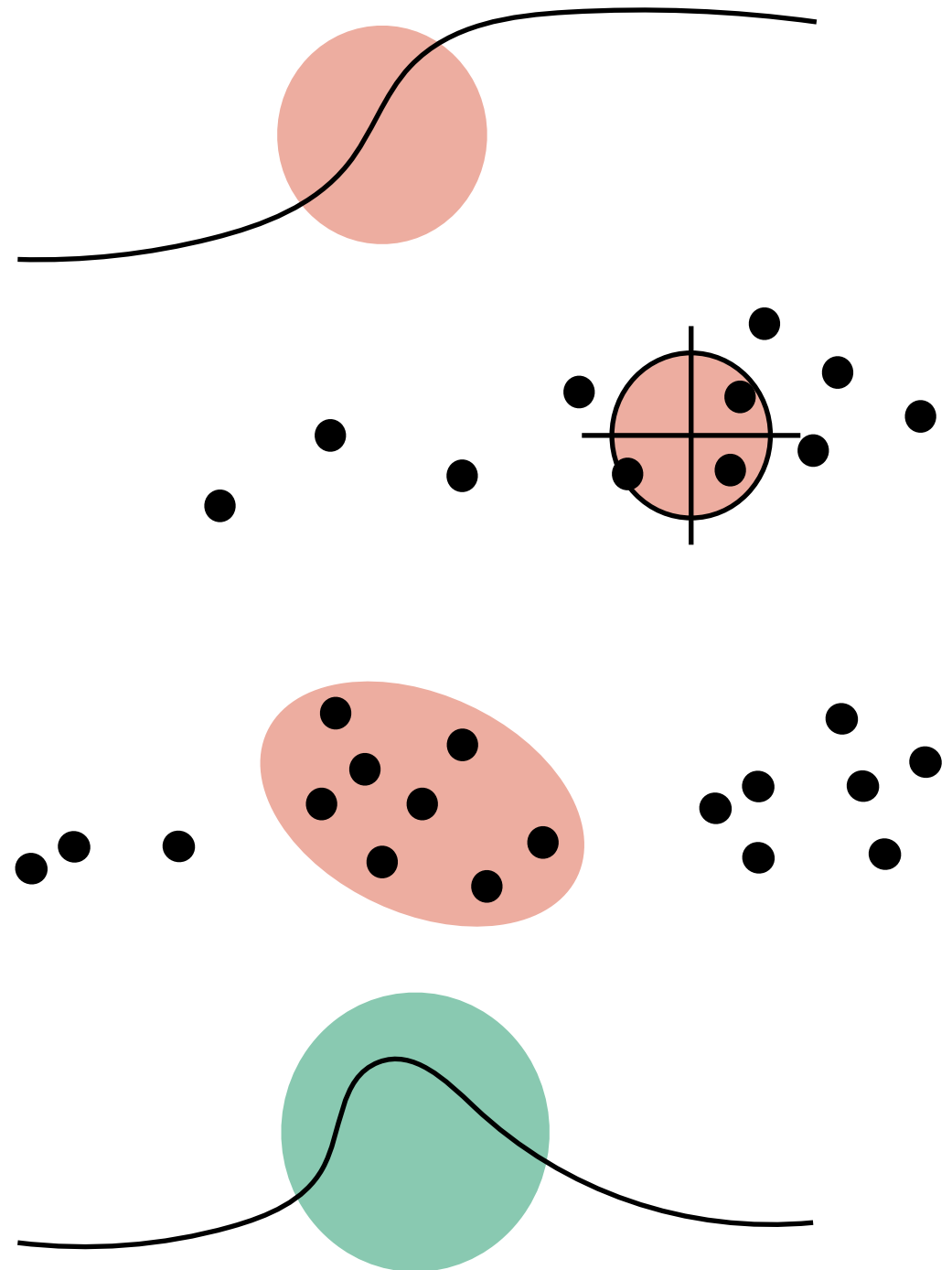
What is an event?

change point

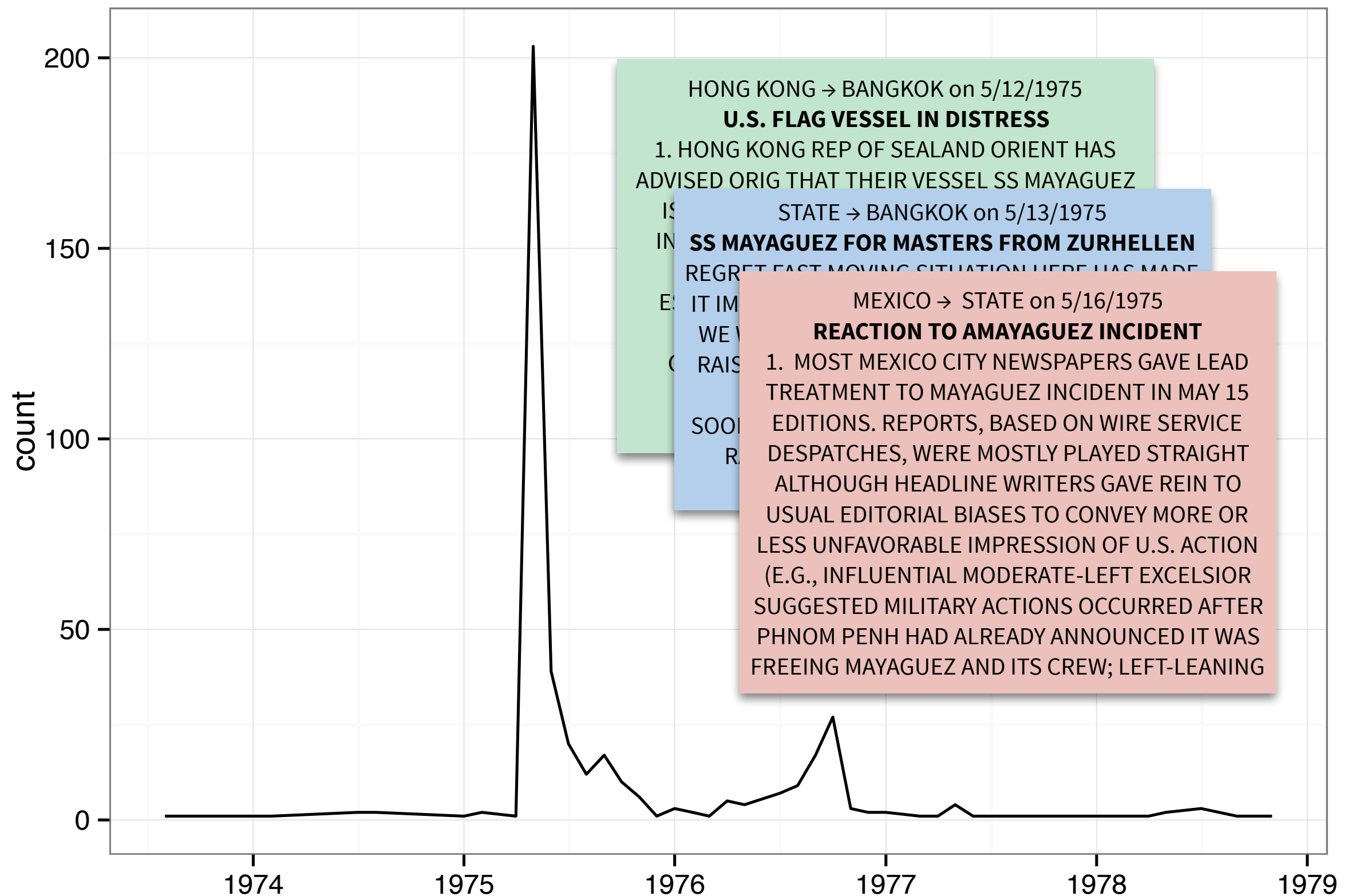
location

cluster of sources

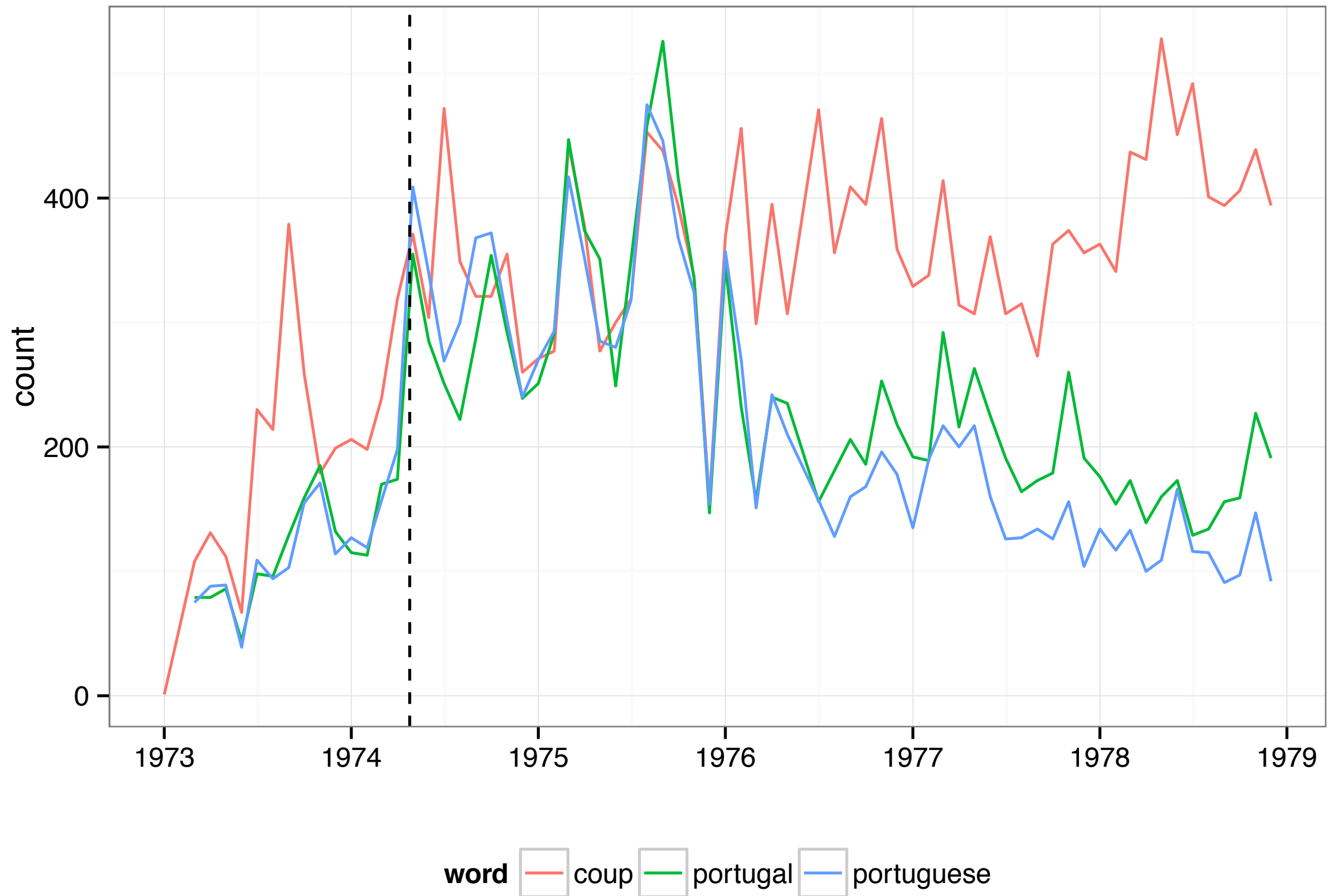
temporary deviation
from business-as-usual



Mayaguez Incident



Carnation Revolution



common themes

authoring entities

events



messages

HONG KONG → BANGKOK on 5/12/1975
U.S. FLAG VESSEL IN DISTRESS

1. HONG
ADVISED
IS REP
IN GULI
94PT
EST BY

STATE → BANGKOK on 5/13/1975
SS MAYAGUEZ FOR MASTERS FROM ZURHELLEN

REGRET
MADE I
INFORMED
MATTERS
DISCUSSIO
YOU SOON

MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT

1. MOST MEXICO CITY NEWSPAPERS GAVE THE
TREATMENT TO MAYAGUEZ INCIDENT IN MAY 15
EDITIONS. REPORTS, BASED ON WIRE SERVICE
DESPATCHES, WERE MOSTLY PLAYED STRAIGHT
ALTHOUGH HEADLINE WRITERS CANNOT REIN TO
USUAL EDITORIAL BIASES TO MAKE THEM MORE OR

modeling messages

May 2, 1974 SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES

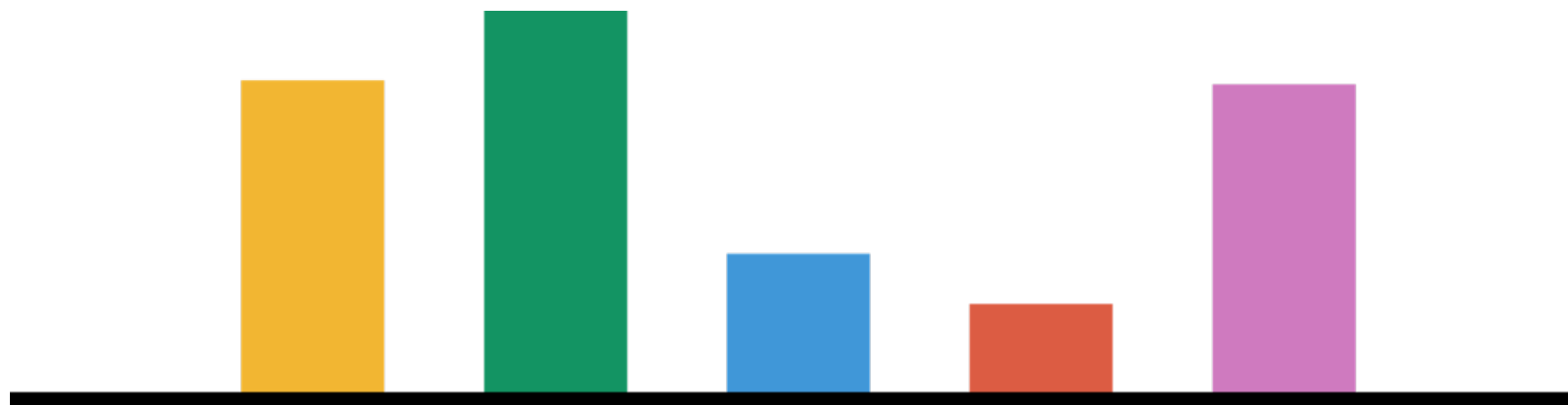
REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL: 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



modeling messages

May 2, 1974 **SITUATION** IN **PORTUGAL** AND ITS AFRICAN TERRITORIES

REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST **GOVERNMENT** **OFFICIALS** ON **SITUATION** IN **PORTUGAL**: 2. **ARMED** **FORCES** **INSURGENT** **MOVEMENT** APPEARS TO BE IN EFFECTIVE **CONTROL** OF **PORTUGAL**. EX-PRESIDENT THOMAZ AND EX- **PRIME MINISTER** **CAETANO** HAVE BEEN SENT INTO **EXILE** IN **MADEIRA**. 3. SEVEN-MAN **JUNTA** OF **NATIONAL** SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER **MILITARY** **OFFICERS** (2 **NAVY**, 2 **AIRFORCE** AND 1 **ARMY**) HAS BEEN SET UP TO ADMINISTER **GOVERNMENT**. **GENERAL** SPINOLA, **SPEAKING** ON BEHALF OF THE **JUNTA**, ANNOUNCED ON APRIL 26 THAT THE **JUNTA** WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF **OPINION** FULL **FREEDOM** OF **EXPRESSION** IN ORDER TO ACCELERATE FORMATION OF '**CIVIC** ASSOCIATIONS' ...

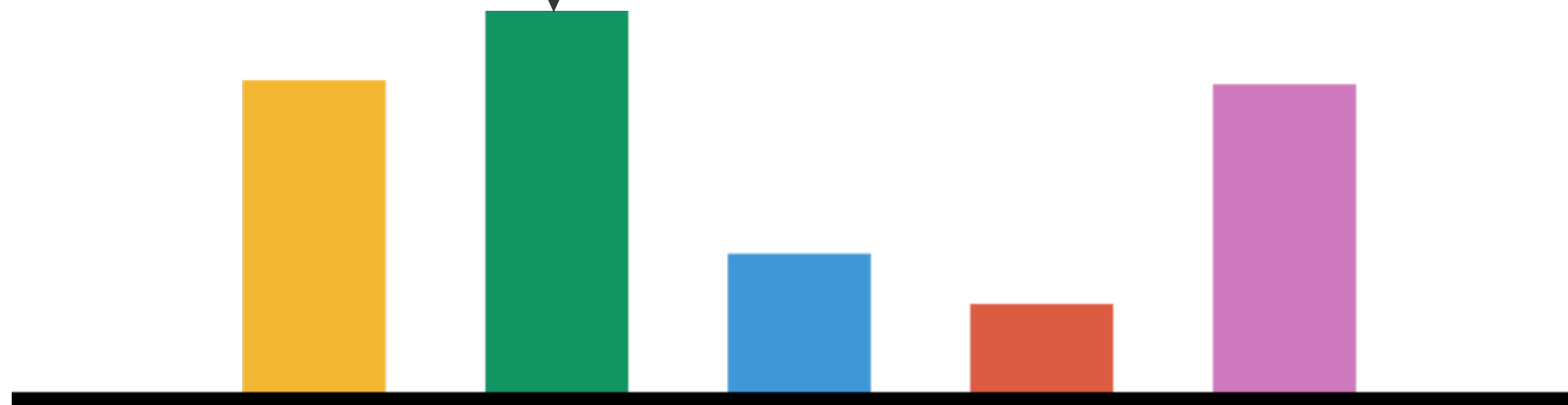


Latent Dirichlet allocation. Blei, Ng, and Jordan, 2003.

modeling messages

May 2, 1974 **SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES**

REFERENCE: LUSAKA **MILITARY** **FORCE OFFICER** **ARMY** **DEFENSE** **NAVY** **...** **GOVERNMENT** **OFFICIALS** **ON SITUATION IN PORTUGAL: 2. ARMED FORCES** **INSURGENT MOVEMENT** **APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-** **PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO** **HAVE BEEN SENT INTO** **EXILE IN MADEIRA. 3. SEVERAL** **OF NATIONAL SALVATION** **CONSISTING OF** **SPINOLA, COSTA GOMES, AND OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE** **AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA,** **SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA** **WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM** **OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...**

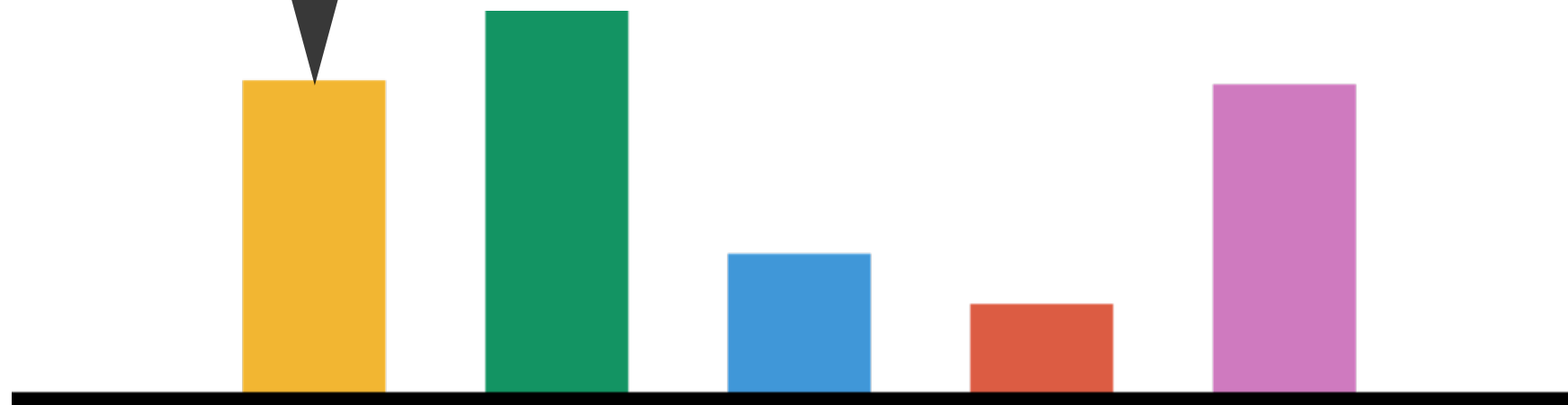


Latent Dirichlet allocation. Blei, Ng, and Jordan, 2003.

modeling messages

May 2, 1974 **SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES**

REFERENCE TO THE GOVERNMENT OF PORTUGAL MAY BE DRAWN UPON IF QUERIED
BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL: 2. **ARMED FORCES**
INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-
PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO
EXILE IN MADEIRA. A JUNTA OF NATIONAL SALVATION CONSISTING OF
SPINOLA, COSTA GOMES, AND FIVE OTHER **MILITARY OFFICERS** (2 **NAVY**, 2 **AIRFORCE**
AND 1 **ARMY**) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. **GENERAL** SPINOLA,
SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA
WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF **OPINION** FULL **FREEDOM**
OF **EXPRESSION** IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...

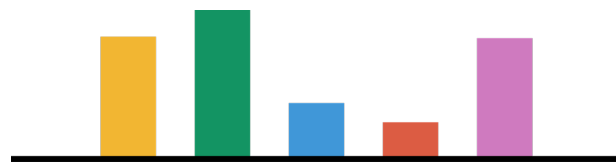


Latent Dirichlet allocation. Blei, Ng, and Jordan, 2003.

modeling messages

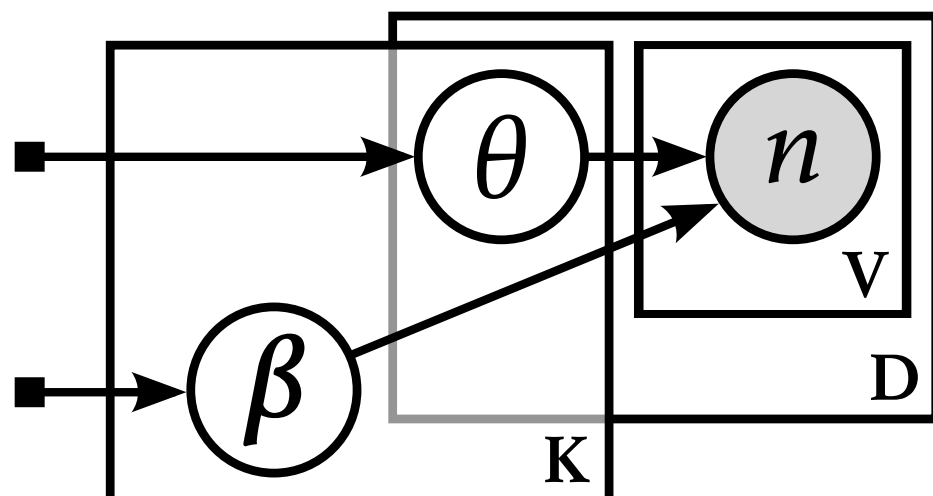
May 2, 1974 SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES

REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL: 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



Latent Dirichlet allocation.
Blei, Ng, and Jordan, 2003.

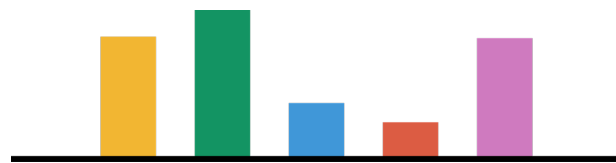
GaP: A Factor Model for Discrete Data. Canny, 2004.



modeling messages

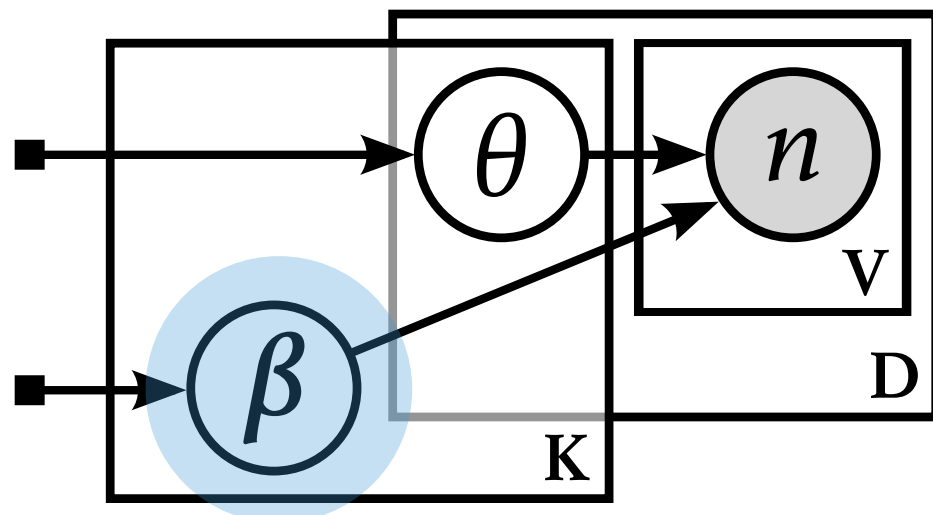
May 2, 1974 **SITUATION IN PORTUGAL** AND ITS AFRICAN TERRITORIES

REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST **GOVERNMENT OFFICIALS** ON **SITUATION IN PORTUGAL**: 2. **ARMED FORCES** **INSURGENT MOVEMENT** APPEARS TO BE IN EFFECTIVE **CONTROL** OF **PORTUGAL**. EX-PRESIDENT THOMAZ AND EX- **PRIME MINISTER CAETANO** HAVE BEEN SENT INTO **EXILE** IN **MADEIRA**. 3. SEVEN-MAN **JUNTA** OF **NATIONAL** SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER **MILITARY OFFICERS** (2 **NAVY**, 2 **AIRFORCE** AND 1 **ARMY**) HAS BEEN SET UP TO ADMINISTER **GOVERNMENT**. **GENERAL** SPINOLA, **SPEAKING** ON BEHALF OF THE **JUNTA**, ANNOUNCED ON APRIL 26 THAT THE **JUNTA** WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF **OPINION** FULL **FREEDOM** OF **EXPRESSION** IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



Latent Dirichlet allocation.
Blei, Ng, and Jordan, 2003.

GaP: A Factor Model for Discrete Data. Canny, 2004.



vocabulary words

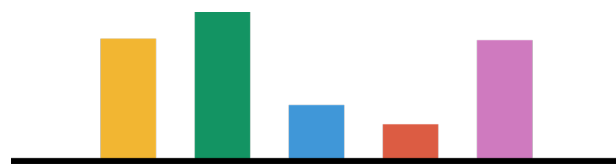
topics	0.1	0.3	0.1	0.0	0.5	0.0
	0.7	0.0	0.1	0.2	0.0	0.0
	0.0	0.1	0.2	0.4	0.0	0.3

$$\beta_k \sim \text{Dirichlet}(\cdot)$$

modeling messages

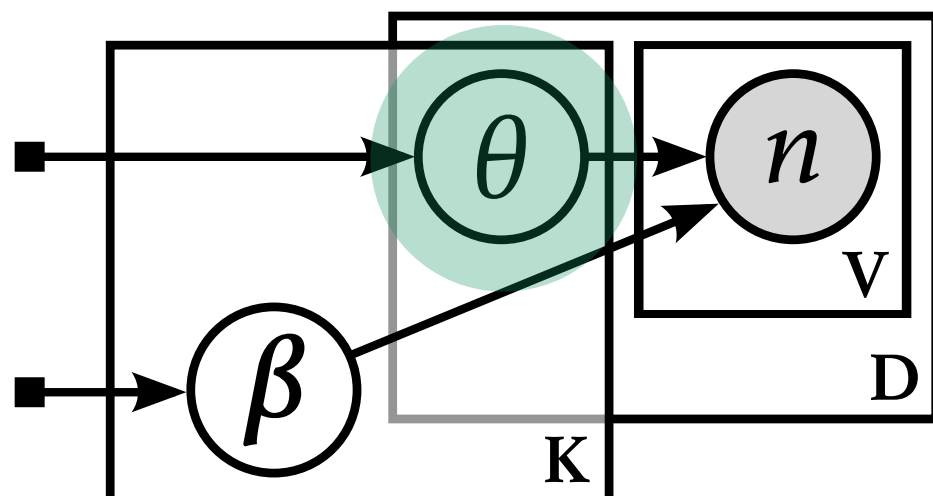
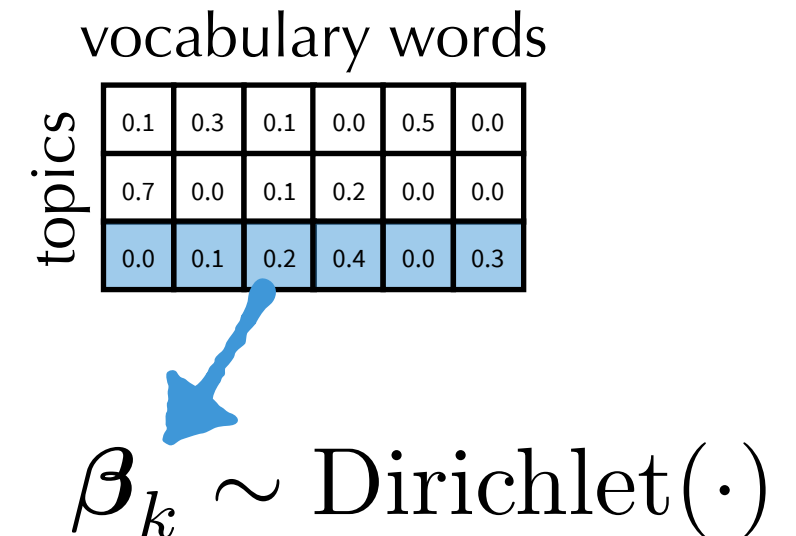
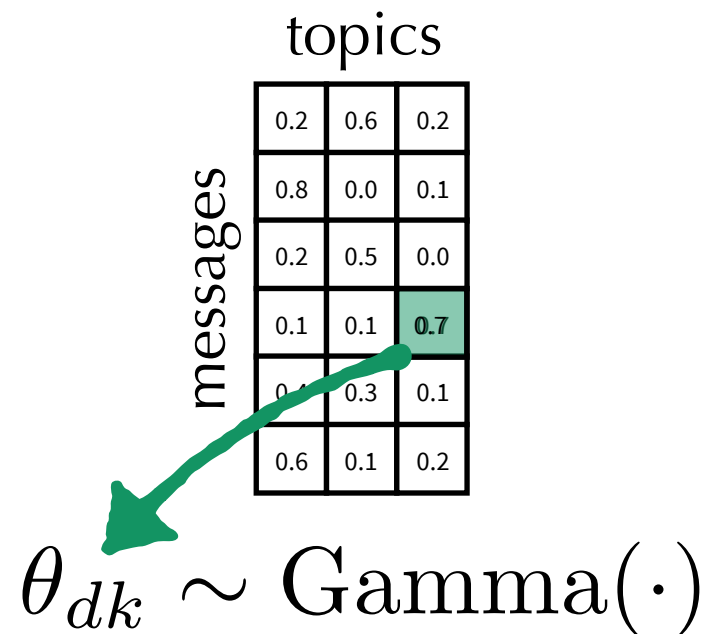
May 2, 1974 **SITUATION IN PORTUGAL** AND ITS AFRICAN TERRITORIES

REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST **GOVERNMENT OFFICIALS** ON **SITUATION IN PORTUGAL**: 2. **ARMED FORCES** **INSURGENT MOVEMENT** APPEARS TO BE IN EFFECTIVE **CONTROL** OF **PORTUGAL**. EX-**PRESIDENT** THOMAZ AND EX-**PRIME MINISTER CAETANO** HAVE BEEN SENT INTO **EXILE** IN **MADEIRA**. 3. SEVEN-MAN **JUNTA** OF **NATIONAL** SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER **MILITARY OFFICERS** (2 **NAVY**, 2 **AIRFORCE** AND 1 **ARMY**) HAS BEEN SET UP TO ADMINISTER **GOVERNMENT**. **GENERAL** SPINOLA, **SPEAKING** ON BEHALF OF THE **JUNTA**, ANNOUNCED ON APRIL 26 THAT THE **JUNTA** WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF **OPINION** FULL **FREEDOM** OF **EXPRESSION** IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



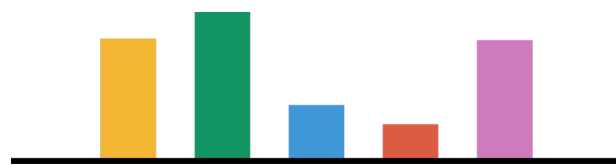
Latent Dirichlet allocation.
Blei, Ng, and Jordan, 2003.

GaP: A Factor Model for Discrete Data. Canny, 2004.



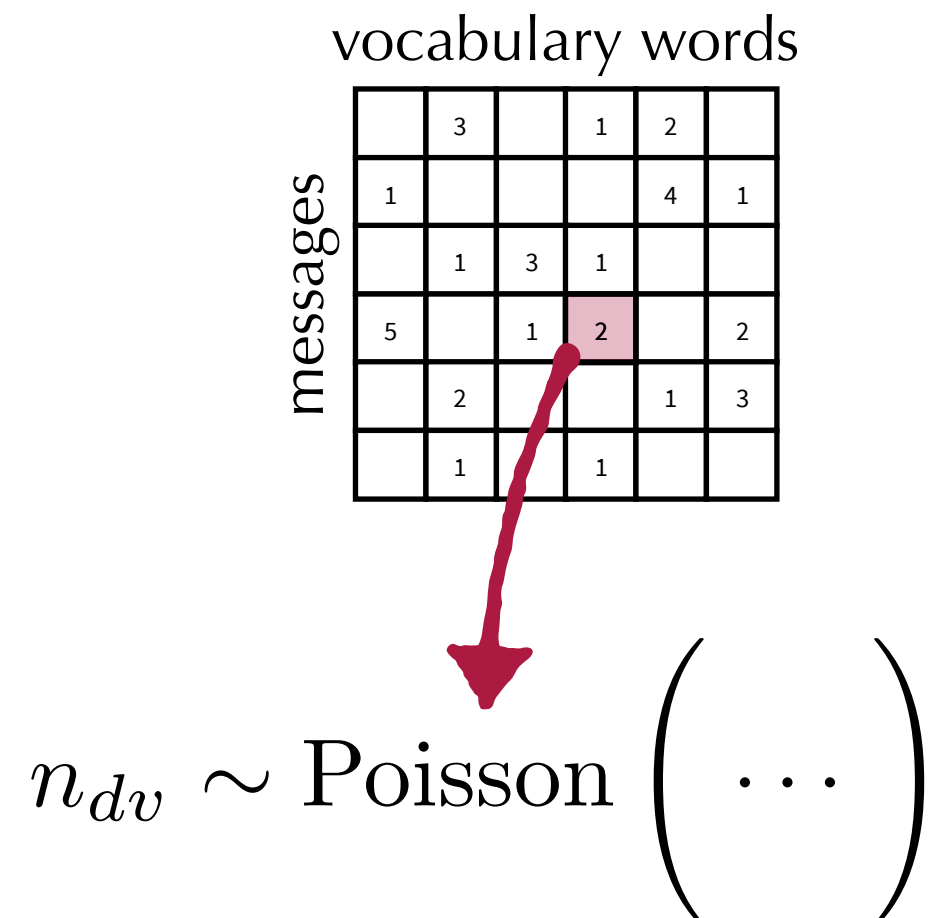
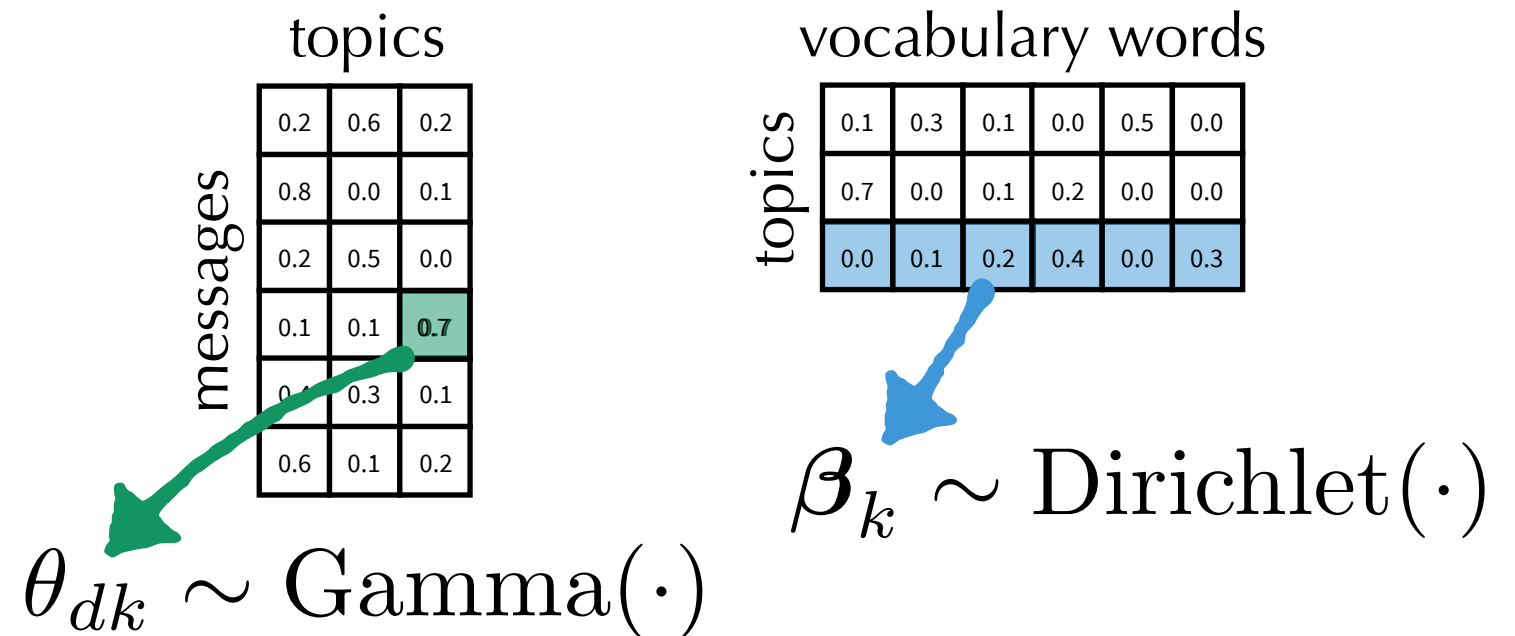
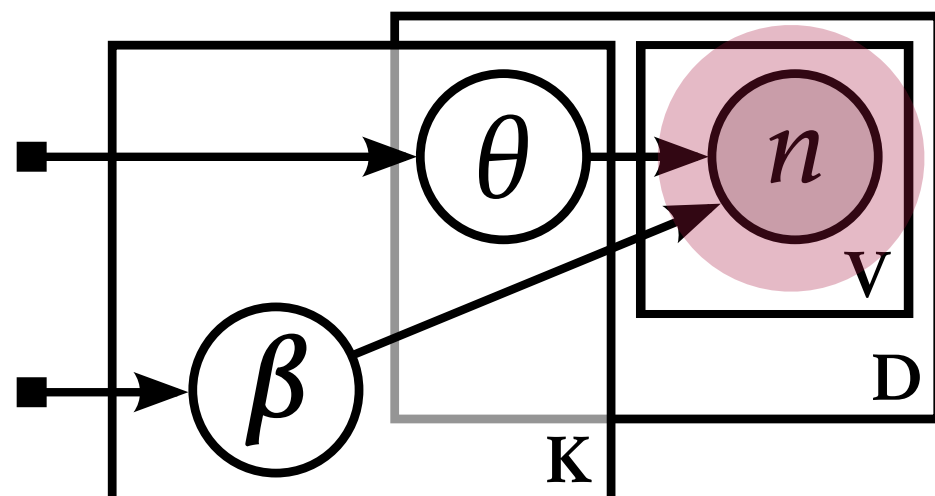
modeling messages

May 2, 1974 **SITUATION IN PORTUGAL** AND ITS AFRICAN TERRITORIES
 REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED
 BY HOST **GOVERNMENT OFFICIALS** ON **SITUATION IN PORTUGAL**: 2. **ARMED FORCES**
INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE **CONTROL** OF **PORTUGAL**. EX-
PRESIDENT THOMAZ AND EX- **PRIME MINISTER CAETANO** HAVE BEEN SENT INTO
EXILE IN **MADEIRA**. 3. SEVEN-MAN **JUNTA** OF **NATIONAL** SALVATION CONSISTING OF
 SPINOLA, COSTA GOMES, AND FIVE OTHER **MILITARY OFFICERS** (2 **NAVY**, 2 **AIRFORCE**
 AND 1 **ARMY**) HAS BEEN SET UP TO ADMINISTER **GOVERNMENT**. **GENERAL** SPINOLA,
SPEAKING ON BEHALF OF THE **JUNTA**, ANNOUNCED ON APRIL 26 THAT THE **JUNTA**
 WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF **OPINION** FULL **FREEDOM**
 OF **EXPRESSION** IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



Latent Dirichlet allocation.
 Blei, Ng, and Jordan, 2003.

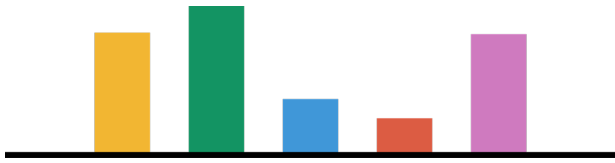
GaP: A Factor Model for Discrete Data. Canny, 2004.



modeling messages

May 2, 1974 SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES

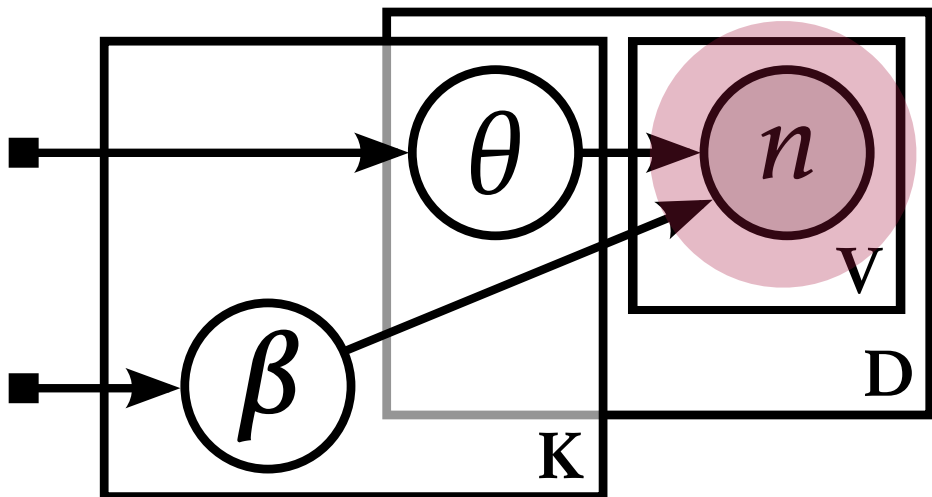
REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL: 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



Latent Dirichlet allocation.

Blei, Ng, and Jordan, 2003.

GaP: A Factor Model for Discrete Data. Canny, 2004.



	topics		
messages	0.2	0.6	0.2
	0.8	0.0	0.1
	0.2	0.5	0.0
	0.1	0.1	0.7
	0.4	0.3	0.1
	0.6	0.1	0.2

$$\theta_{dk} \sim \text{Gamma}(\cdot)$$

	vocabulary words					
topics	0.1	0.3	0.1	0.0	0.5	0.0
	0.7	0.0	0.1	0.2	0.0	0.0
	0.0	0.1	0.2	0.4	0.0	0.3

$$\beta_k \sim \text{Dirichlet}(\cdot)$$

vocabulary words

messages

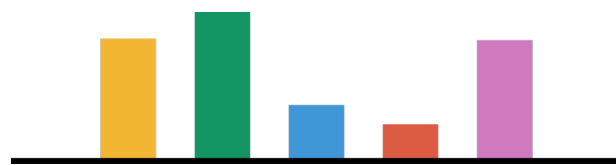
	3		1	2	
1				4	1
	1	3	1		
5		1	2		2
	2			1	3
	1		1		

$$n_{dv} \sim \text{Poisson} \left(\sum_{k=1}^K \theta_{dk} \beta_{kv} \right)$$

modeling messages

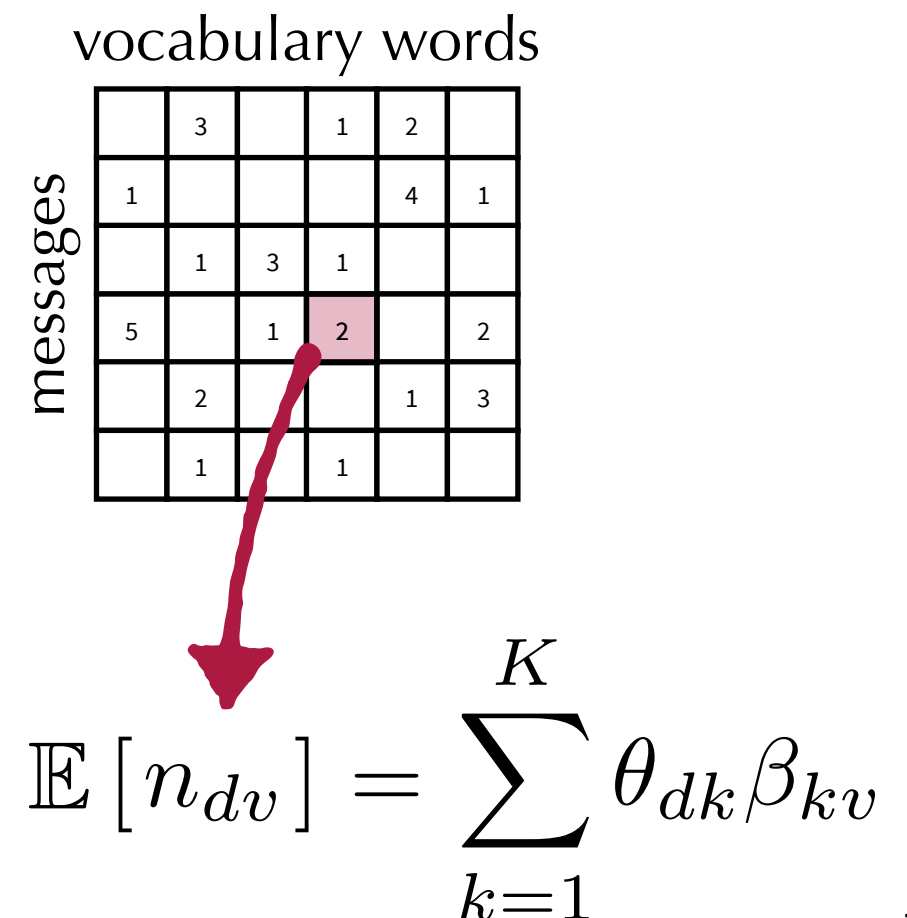
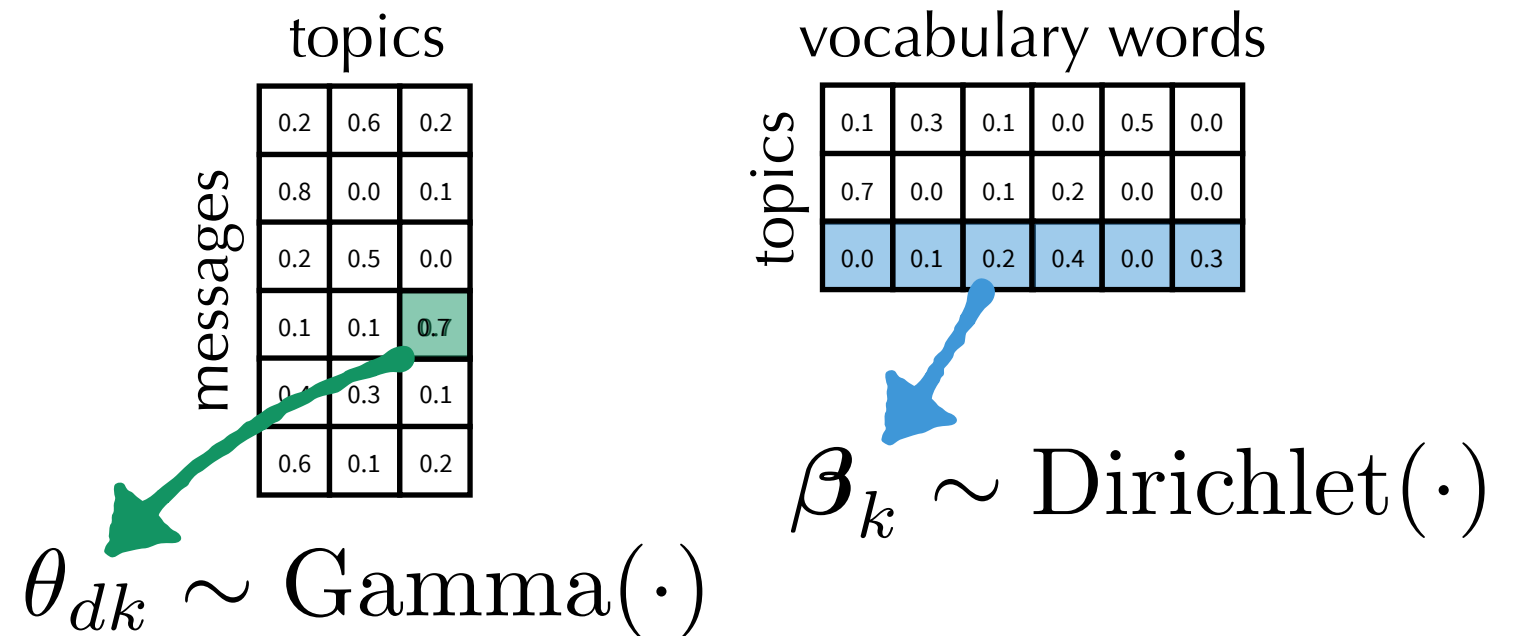
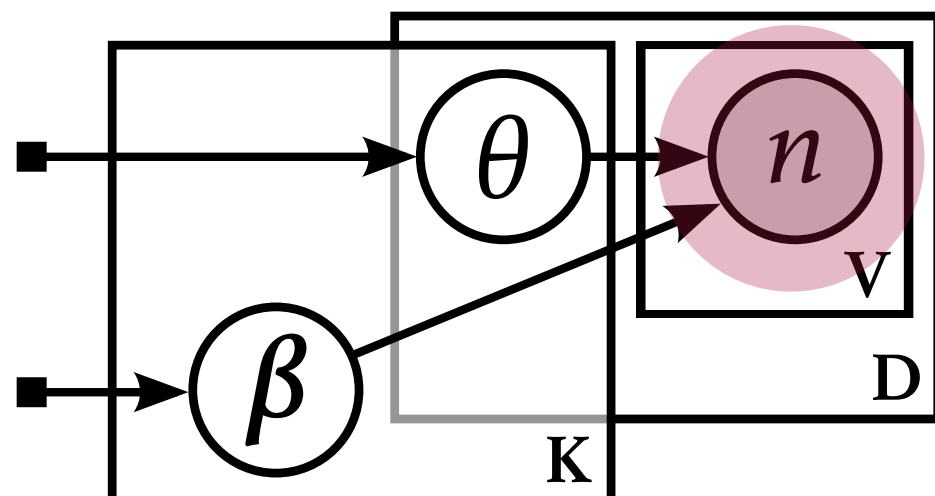
May 2, 1974 SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES

REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL: 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



Latent Dirichlet allocation.
Blei, Ng, and Jordan, 2003.

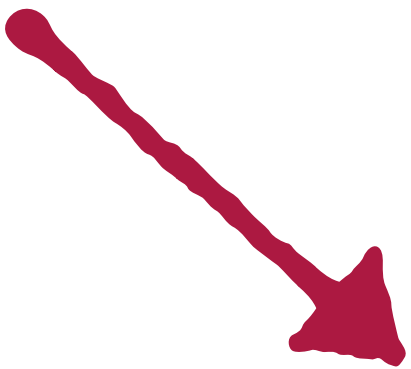
GaP: A Factor Model for Discrete Data. Canny, 2004.



common themes

authoring entities

events



messages

HONG KONG → BANGKOK on 5/12/1975
U.S. FLAG VESSEL IN DISTRESS

1. HONG KONG ADVISED
ADVISED
IS REPORTED
IN GULF OF TONKIN
94PT
EST BY

STATE → BANGKOK on 5/13/1975
SS MAYAGUEZ FOR MASTERS FROM ZURHELLEN

REGRET
MADE
INFORMED
MATTERS
DISCUSSIO
YOU SOON

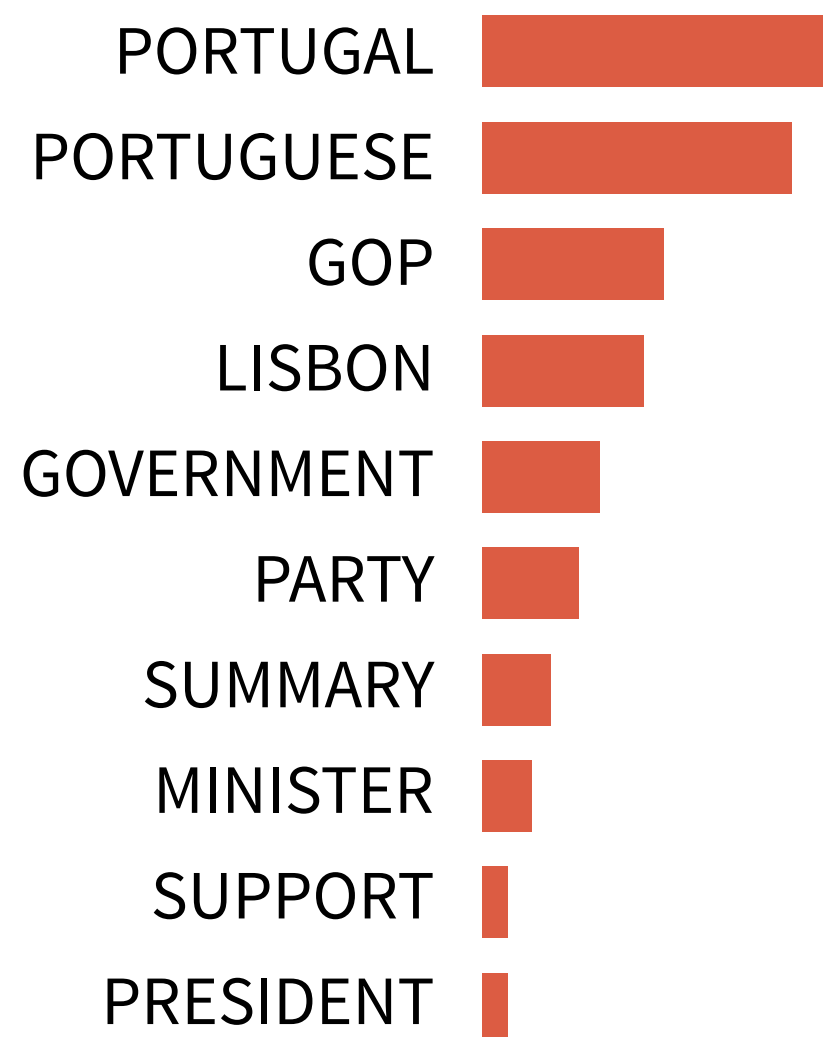
MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT

1. MOST MEXICO CITY NEWSPAPERS GAVE LEAD TREATMENT TO MAYAGUEZ INCIDENT IN MAY 15 EDITIONS. REPORTS, BASED ON WIRE SERVICE DESPATCHES, WERE MOSTLY PLAYED STRAIGHT ALTHOUGH HEADLINE WRITERS GAVE REIN TO USUAL EDITORIAL BIASES TO CONVEY MORE OR

modeling authoring entities



Lisbon Embassy



Lisbon Embassy



Bangkok Embassy



United Nations



OECD



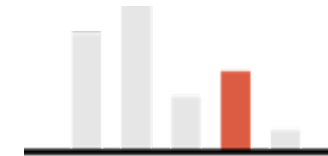
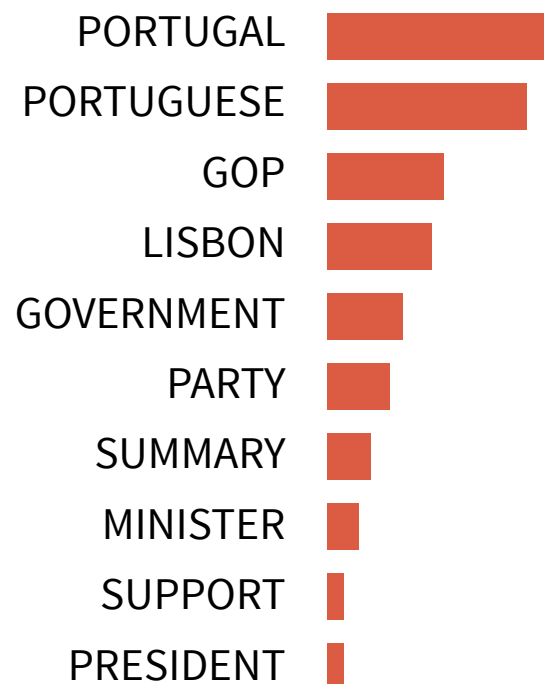
Lisbon Embassy

Documents sent by Lisbon

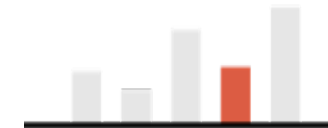
Self Interest



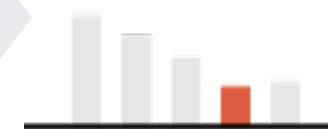
Entity Topic



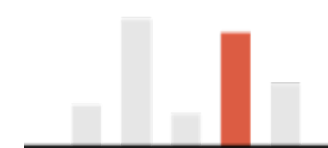
REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN **PORTUGAL**. 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF **PORTUGAL**. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER **CAETANO** HAVE BEEN SENT INTO EXILE IN **MADEIRA**. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



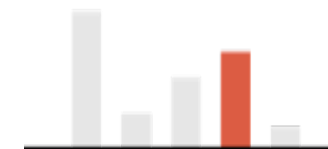
MEETING FOR THE WORLD ADMINISTRATIVE RADIO CONFERENCE (WARC), GENEVA, OCTOBER 23 TO NOVEMBER 17, 1978 1. PLEASE ACREDIT THE FOLLOWING DELEGATION TO THE SUBJECT MEETING: REPRESENTATIVE SAMUEL E. PROBST NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION DEPARTMENT OF COMMERCE **ALTERNATE** REPRESENTATIVES RICHARD E. SHRUM OFFICE OF INTERNATIONAL COMMUNICATIONS POLICY BUREAU OF ECONOMIC AND **BUSINESS** AFFAIRS DEPARTMENT OF STATE WILLIAM R. TORAK ASSISTANT CHIEF INTERNATIONAL AND OPERATIONS **DIVISION** FEDERAL COMMUNICATIONS COMMISSION SPECIAL ADVISER DEXTER ANDERSON TELECOMMUNICATIONS ATTACHE UNITED STATES MISSION GENEVA ADVISERS



SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT **OFFICIALS** ON SITUATION IN **PORTUGAL**. 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF **PORTUGAL**. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN **MADEIRA**. 3. SEVEN-MAN JUNTA OF NATIONAL **SALVATION** CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, ...



INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF **PORTUGAL** PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN **MADEIRA**. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND **FIVE** OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL **SPINOLA**, MEETING FOR THE WORLD ADMINISTRATIVE **RADIO** CONFERENCE (WARC), GENEVA, OCTOBER 23 TO **NOVEMBER** 17, 1978 1. PLEASE ACREDIT THE FOLLOWING x DEPARTMENT OF COMMERCE ALTERNATE REPRESENTATIVES RICHARD E. SHRUM OFFICE OF INTERNATIONAL COMMUNICATIONS POLICY BUREAU OF ECONOMIC AND ...



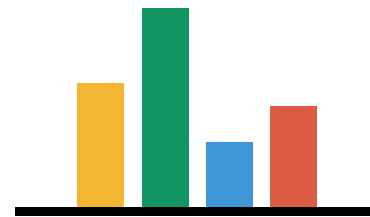
MEETING FOR THE WORLD ADMINISTRATIVE RADIO CONFERENCE (WARC), GENEVA, OCTOBER 23 **TO NOVEMBER** 17, 1978 1. PLEASE ACREDIT THE FOLLOWING DELEGATION TO THE SUBJECT MEETING: REPRESENTATIVE SAMUEL E. PROBST NATIONAL TELECOMMUNICATIONS AND **INFORMATION** ADMINISTRATION BUSINESS AFFAIRS DEPARTMENT OF **STATE** WILLIAM R. TORAK ASSISTANT INTERNATIONAL AND OPERATIONS **DIVISION** FEDERAL COMMUNICATIONS COMMISSION SPECIAL ADVISER DEXTER ANDERSON TELECOMMUNICATIONS ATTACHE UNITED **STATES** MISSION GENEVA ADVISERS FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN **PORTUGAL**. 2. ARMED FORCES



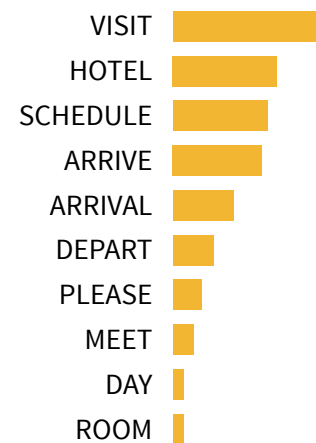
Lisbon Embassy

Documents sent by Lisbon

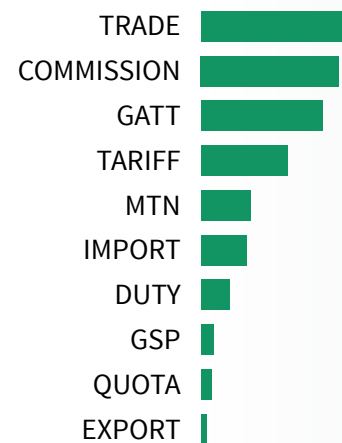
Entity Concerns



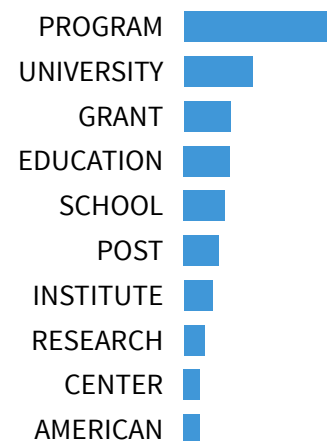
General Topic



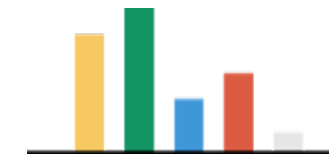
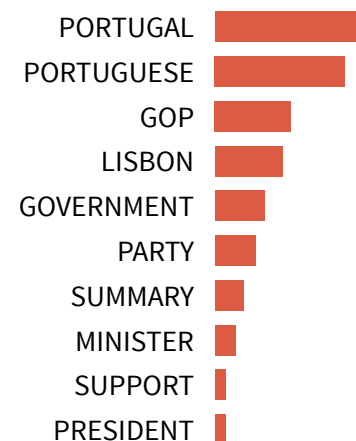
General Topic



General Topic



Entity Topic



REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL. 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...



MEETING FOR THE WORLD ADMINISTRATIVE RADIO CONFERENCE (WARC), GENEVA, OCTOBER 23 TO NOVEMBER 17, 1978 1. PLEASE ACREDIT THE FOLLOWING DELEGATION TO THE SUBJECT MEETING: REPRESENTATIVE SAMUEL E. PROBST NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION DEPARTMENT OF COMMERCE ALTERNATE REPRESENTATIVES RICHARD E. SHRUM OFFICE OF INTERNATIONAL COMMUNICATIONS POLICY BUREAU OF ECONOMIC AND BUSINESS AFFAIRS DEPARTMENT OF STATE WILLIAM R. TORAK ASSISTANT CHIEF INTERNATIONAL AND OPERATIONS DIVISION FEDERAL COMMUNICATIONS COMMISSION SPECIAL ADVISER DEXTER ANDERSON TELECOMMUNICATIONS ATTACHE UNIT: STATES MISSION GENEVA ADVISERS



SPEAKING ON BEHALF OF THE JUNTA, ANNOUNCED ON APRIL 26 THAT THE JUNTA WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF OPINION FULL FREEDOM OF EXPRESSION IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL. 2. ARMED FORCES INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. EX-PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, ...

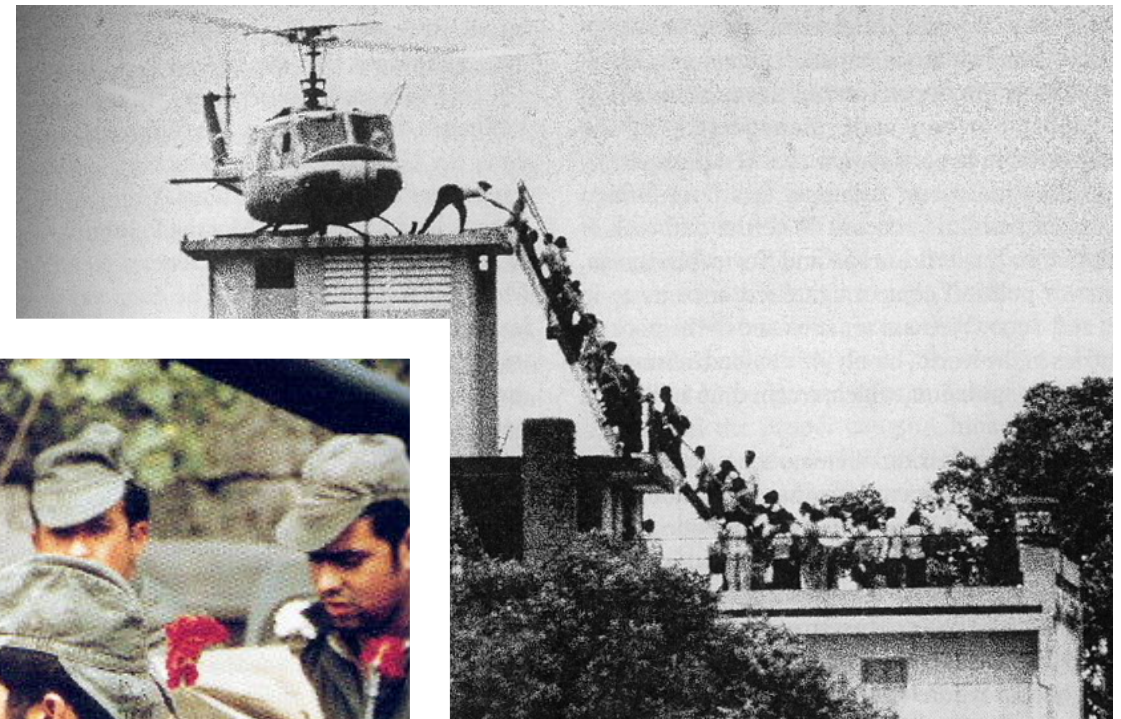


INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE CONTROL OF PORTUGAL. PRESIDENT THOMAZ AND EX-PRIME MINISTER CAETANO HAVE BEEN SENT INTO EXILE IN MADEIRA. 3. SEVEN-MAN JUNTA OF NATIONAL SALVATION CONSISTING OF SPINOLA, COSTA GOMES, AND FIVE OTHER MILITARY OFFICERS (2 NAVY, 2 AIRFORCE AND 1 ARMY) HAS BEEN SET UP TO ADMINISTER GOVERNMENT. GENERAL SPINOLA, MEETING FOR THE WORLD ADMINISTRATIVE RADIO CONFERENCE (WARC), GENEVA, OCTOBER 23 TO NOVEMBER 17, 1978 1. PLEASE ACREDIT THE FOLLOWING DELEGATION TO THE SUBJECT MEETING: REPRESENTATIVE SAMUEL E. PROBST NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION BUSINESS AFFAIRS DEPARTMENT OF STATE WILLIAM R. TORAK ASSISTANT INTERNATIONAL AND OPERATIONS DIVISION FEDERAL COMMUNICATIONS COMMISSION SPECIAL ADVISER DEXTER ANDERSON TELECOMMUNICATIONS ATTACHE UNITED STATES MISSION GENEVA ADVISERS FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL. 2. ARMED FORCES



MEETING FOR THE WORLD ADMINISTRATIVE RADIO CONFERENCE (WARC), GENEVA, OCTOBER 23 TO NOVEMBER 17, 1978 1. PLEASE ACREDIT THE FOLLOWING DELEGATION TO THE SUBJECT MEETING: REPRESENTATIVE SAMUEL E. PROBST NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION BUSINESS AFFAIRS DEPARTMENT OF STATE WILLIAM R. TORAK ASSISTANT INTERNATIONAL AND OPERATIONS DIVISION FEDERAL COMMUNICATIONS COMMISSION SPECIAL ADVISER DEXTER ANDERSON TELECOMMUNICATIONS ATTACHE UNITED STATES MISSION GENEVA ADVISERS FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED BY HOST GOVERNMENT OFFICIALS ON SITUATION IN PORTUGAL. 2. ARMED FORCES

What about **events**?



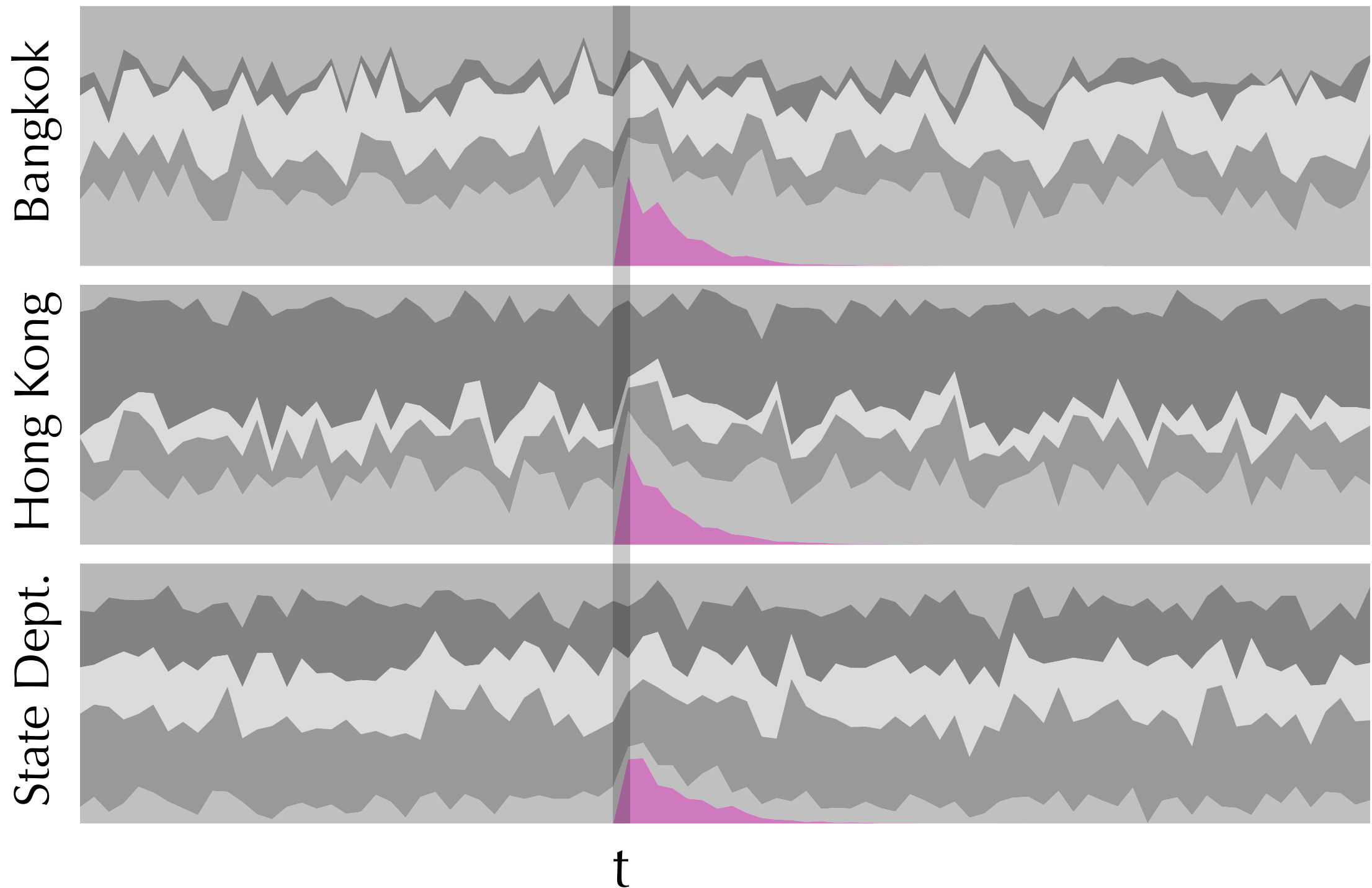
Events are temporary shifts away from **business as usual** in message content.

Thus far, our model of cables captures **business as usual**.



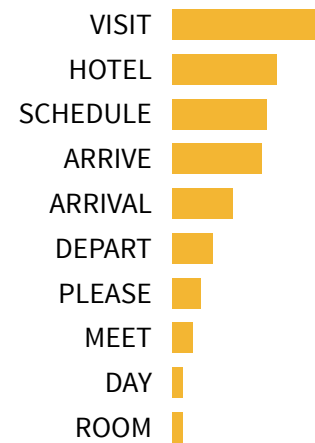
modeling events

stream of messages

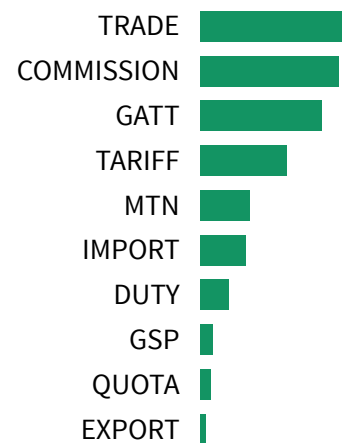




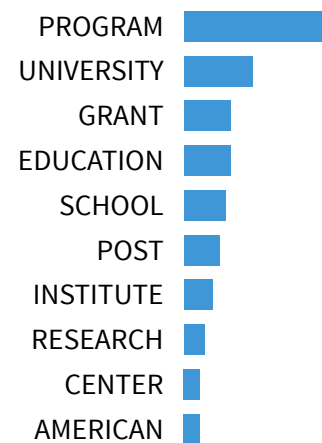
General Topic



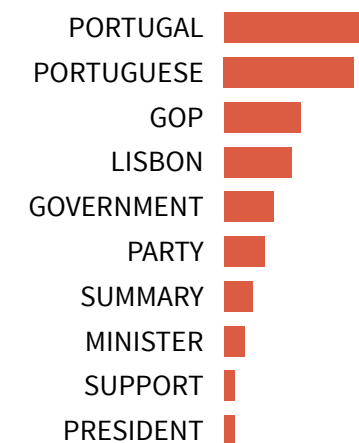
General Topic



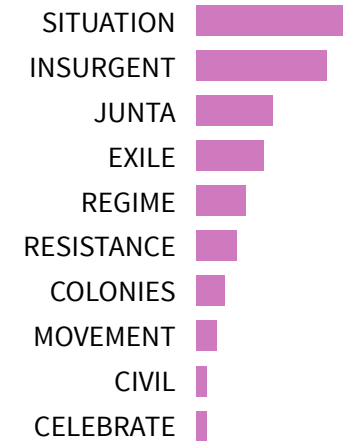
General Topic



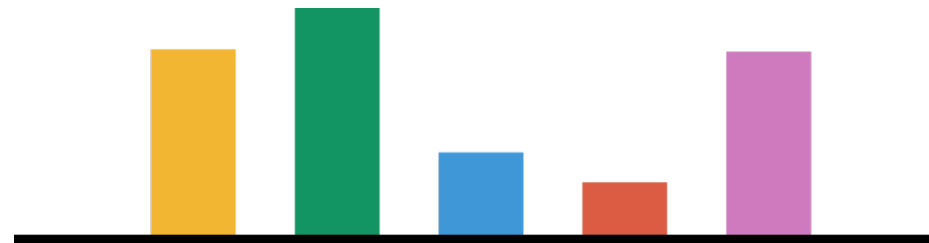
Entity Topic




Event Topic

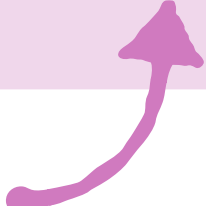


May 2, 1974 **SITUATION IN PORTUGAL AND ITS AFRICAN TERRITORIES**
 REFERENCE: LUSAKA 758 1. FOLLOWING ROUNDUP MAY BE DRAWN UPON IF QUERIED
 BY HOST **GOVERNMENT OFFICIALS** ON **SITUATION IN PORTUGAL**: 2. **ARMED FORCES**
INSURGENT MOVEMENT APPEARS TO BE IN EFFECTIVE **CONTROL OF PORTUGAL**. EX-
PRESIDENT THOMAZ AND EX- **PRIME MINISTER CAETANO** HAVE BEEN SENT INTO
EXILE IN MADEIRA. 3. SEVEN-MAN **JUNTA OF NATIONAL SALVATION** CONSISTING OF
 SPINOLA, COSTA GOMES, AND FIVE OTHER **MILITARY OFFICERS** (2 **NAVY**, 2 **AIRFORCE**
 AND 1 **ARMY**) HAS BEEN SET UP TO ADMINISTER **GOVERNMENT**. **GENERAL SPINOLA**,
SPEAKING ON BEHALF OF THE **JUNTA**, ANNOUNCED ON APRIL 26 THAT THE **JUNTA**
 WILL "IMMEDIATELY BEGIN PERMITTING ALL CURRENTS OF **OPINION** FULL **FREEDOM**
 OF **EXPRESSION** IN ORDER TO ACCELERATE FORMATION OF 'CIVIC ASSOCIATIONS' ...

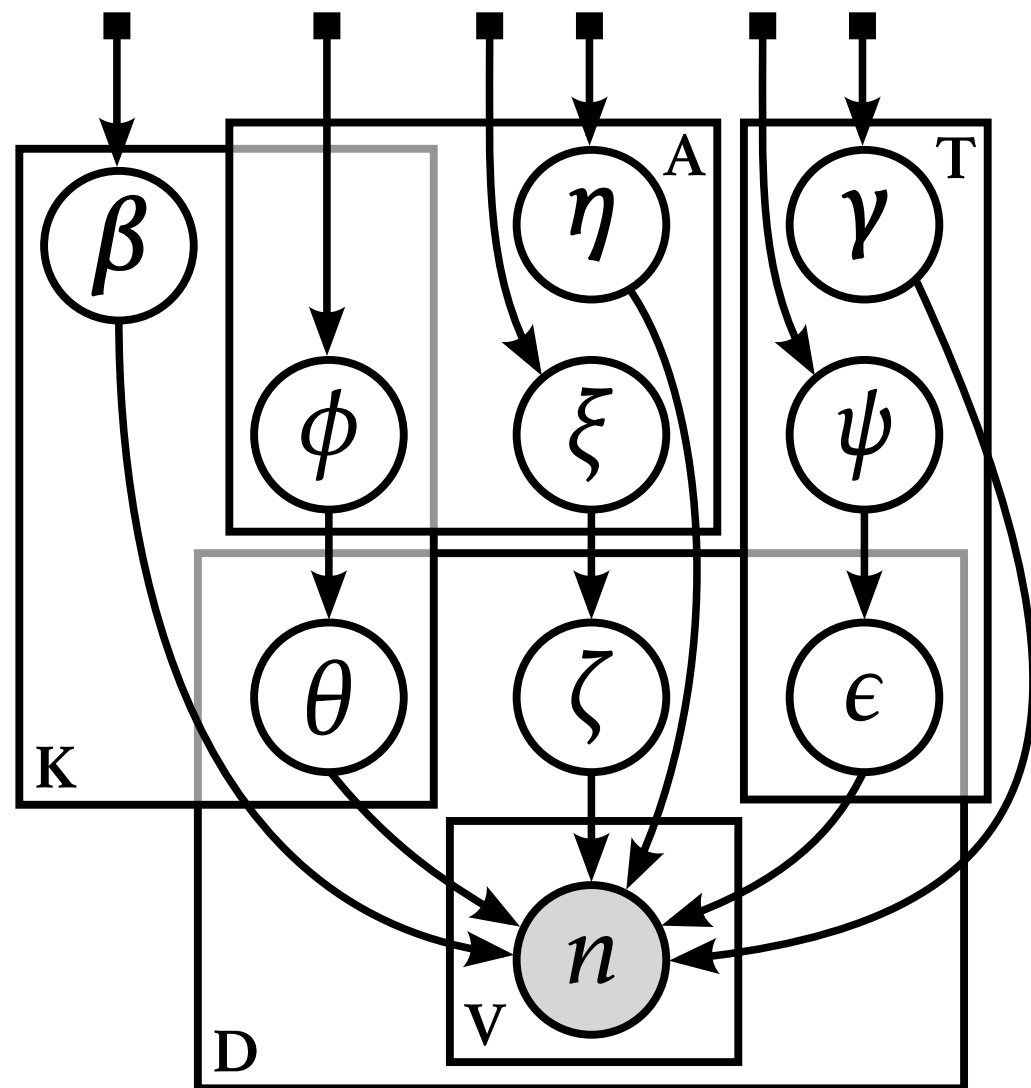


$$n_{dv} \sim \text{Poisson} \left(\sum_{k=1}^K \theta_{dk} \beta_{kv} + \zeta_d \eta_{a_d v} + \sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{tv} \right)$$

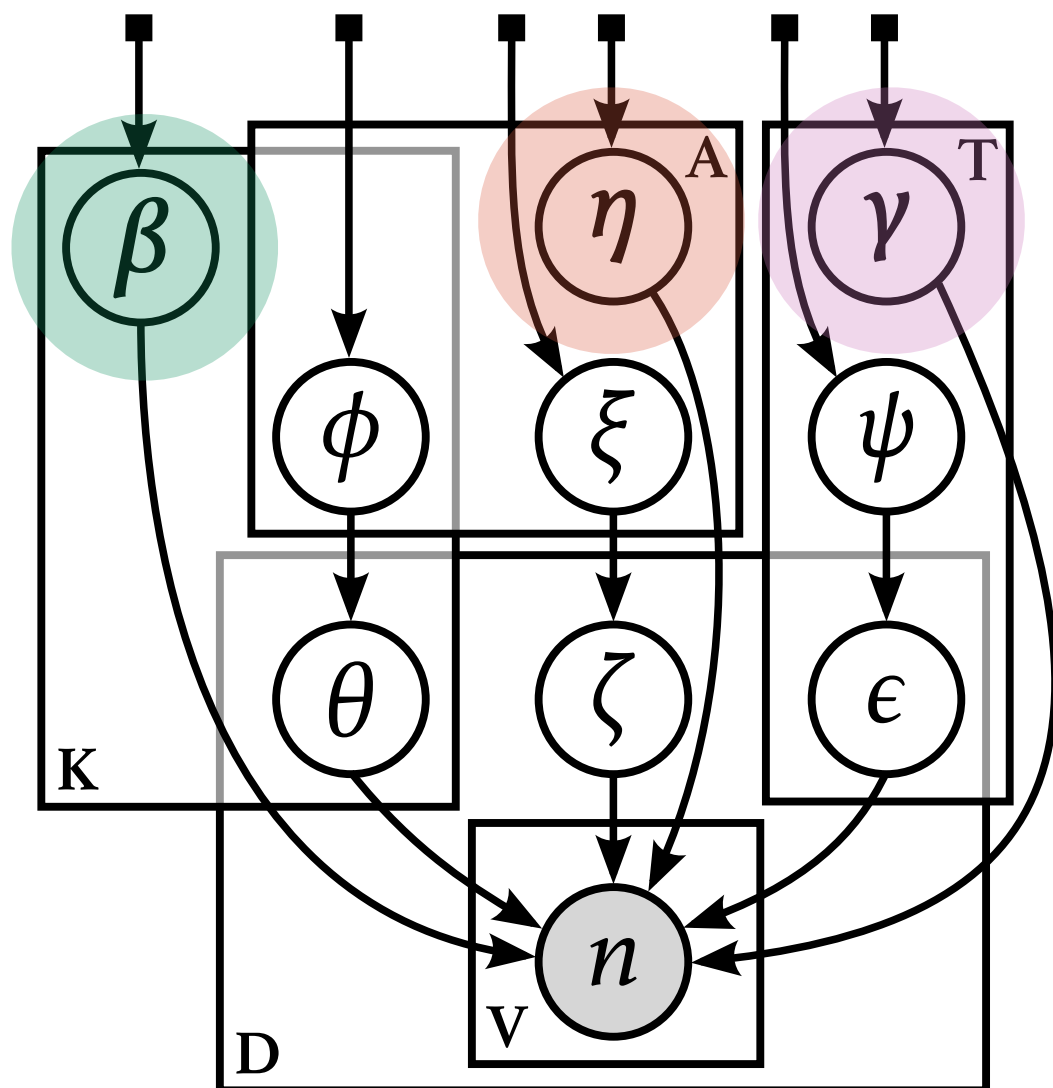
decay of relevancy 



$$n_{dv} \sim \text{Poisson} \left(\overset{\text{🌐}}{\sum_{k=1}^K \theta_{dk} \beta_{kv}} + \overset{\text{👤}}{\zeta_d \eta_{a_d v}} + \overset{\text{🕒}}{\sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{tv}} \right)$$



$$n_{dv} \sim \text{Poisson} \left(\overset{\text{🌐}}{\sum_{k=1}^K \theta_{dk} \beta_{kv}} + \overset{\text{👤}}{\zeta_d \eta_{a_d v}} + \overset{\text{🕒}}{\sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{tv}} \right)$$

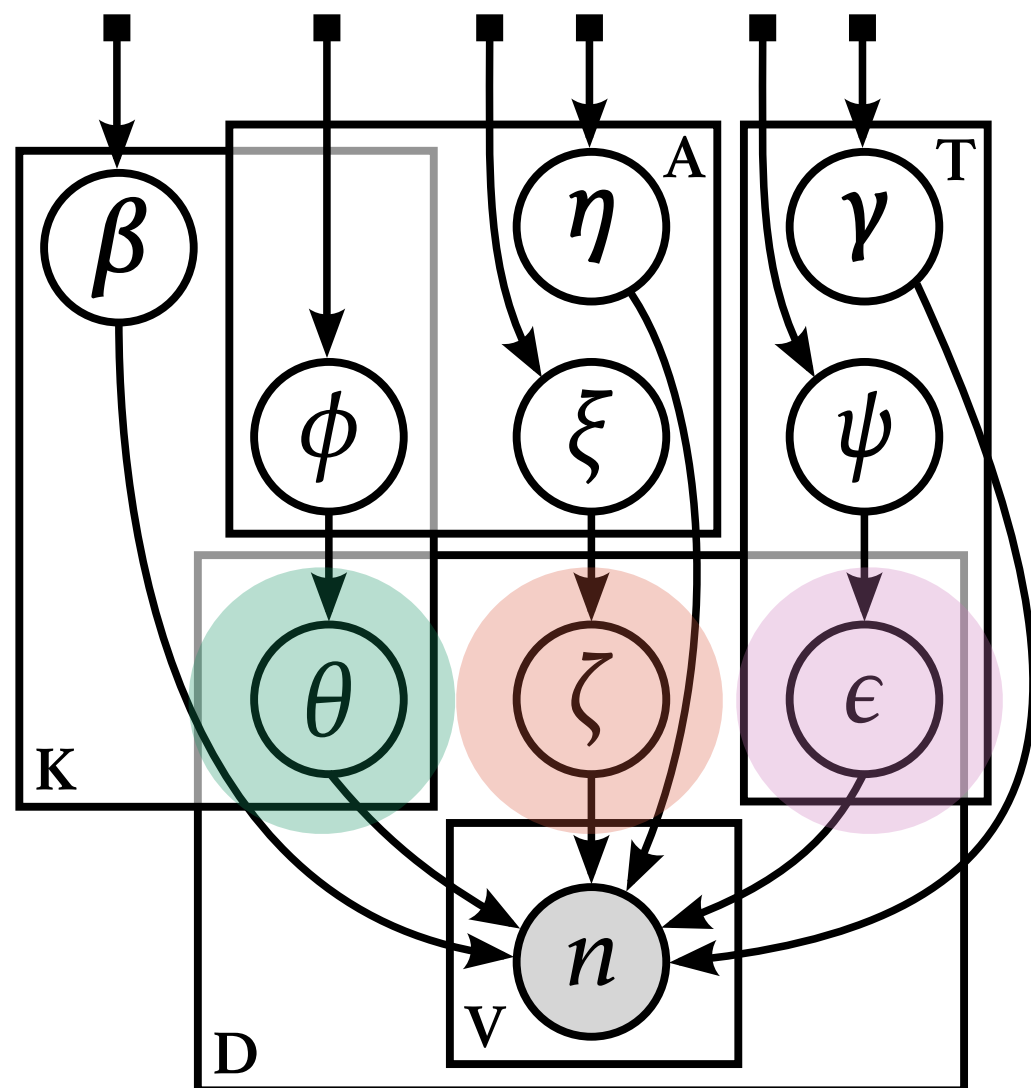


$$\beta_k \sim \text{Dirichlet}(\cdot)$$

$$\eta_a \sim \text{Dirichlet}(\cdot)$$

$$\gamma_t \sim \text{Dirichlet}(\cdot)$$

$$n_{dv} \sim \text{Poisson} \left(\overset{\text{🌐}}{\sum_{k=1}^K \theta_{dk} \beta_{kv}} + \overset{\text{👤}}{\zeta_d \eta_{a_d v}} + \overset{\text{🕒}}{\sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{tv}} \right)$$



$$\beta_k \sim \text{Dirichlet}(\cdot)$$

$$\eta_a \sim \text{Dirichlet}(\cdot)$$

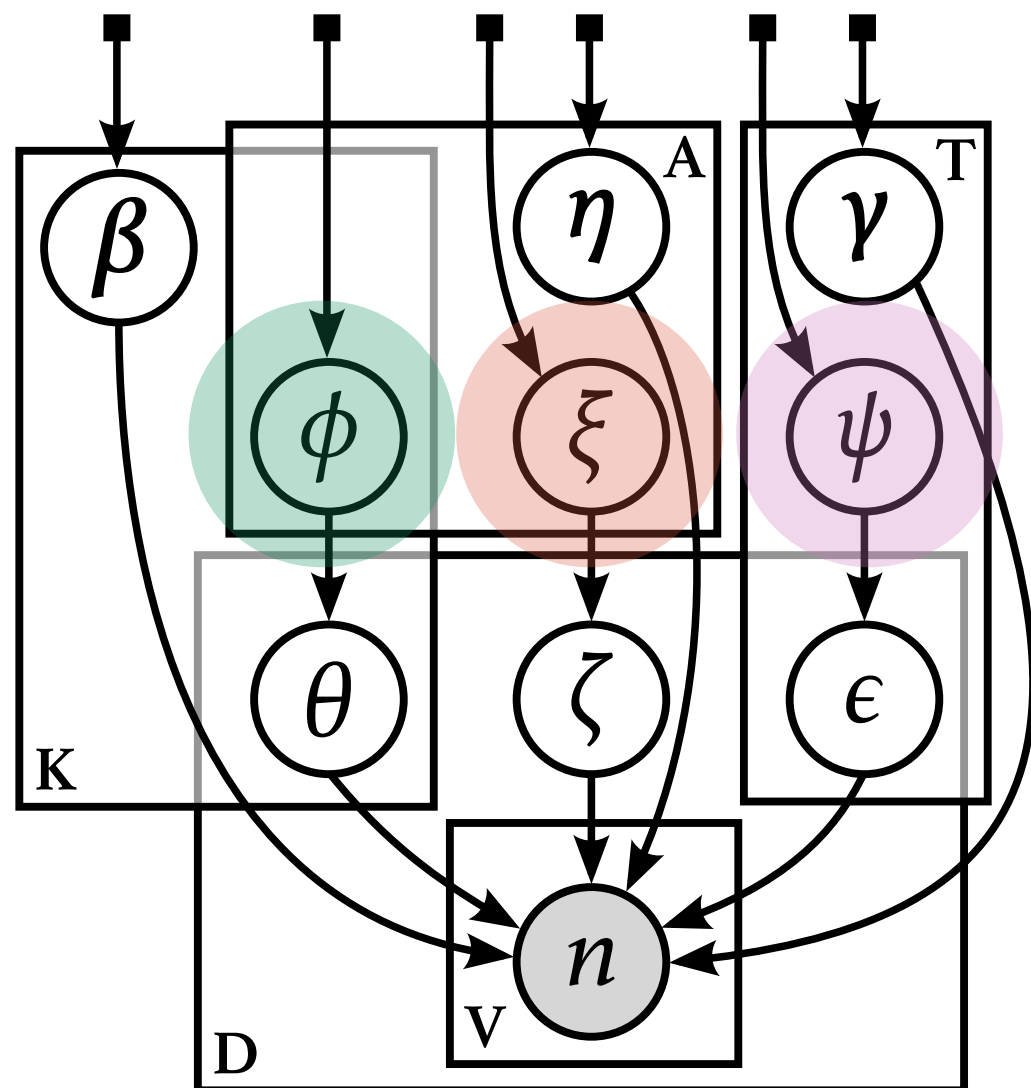
$$\gamma_t \sim \text{Dirichlet}(\cdot)$$

$$\theta_{dk} \sim \text{Gamma}(\dots)$$

$$\zeta_d \sim \text{Gamma}(\dots)$$

$$\epsilon_{dt} \sim \text{Gamma}(\dots)$$

$$n_{dv} \sim \text{Poisson} \left(\overset{\text{🌐}}{\sum_{k=1}^K \theta_{dk} \beta_{kv}} + \overset{\text{👤}}{\zeta_d \eta_{a_d v}} + \overset{\text{🕒}}{\sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{tv}} \right)$$



$$\beta_k \sim \text{Dirichlet}(\cdot)$$

$$\eta_a \sim \text{Dirichlet}(\cdot)$$

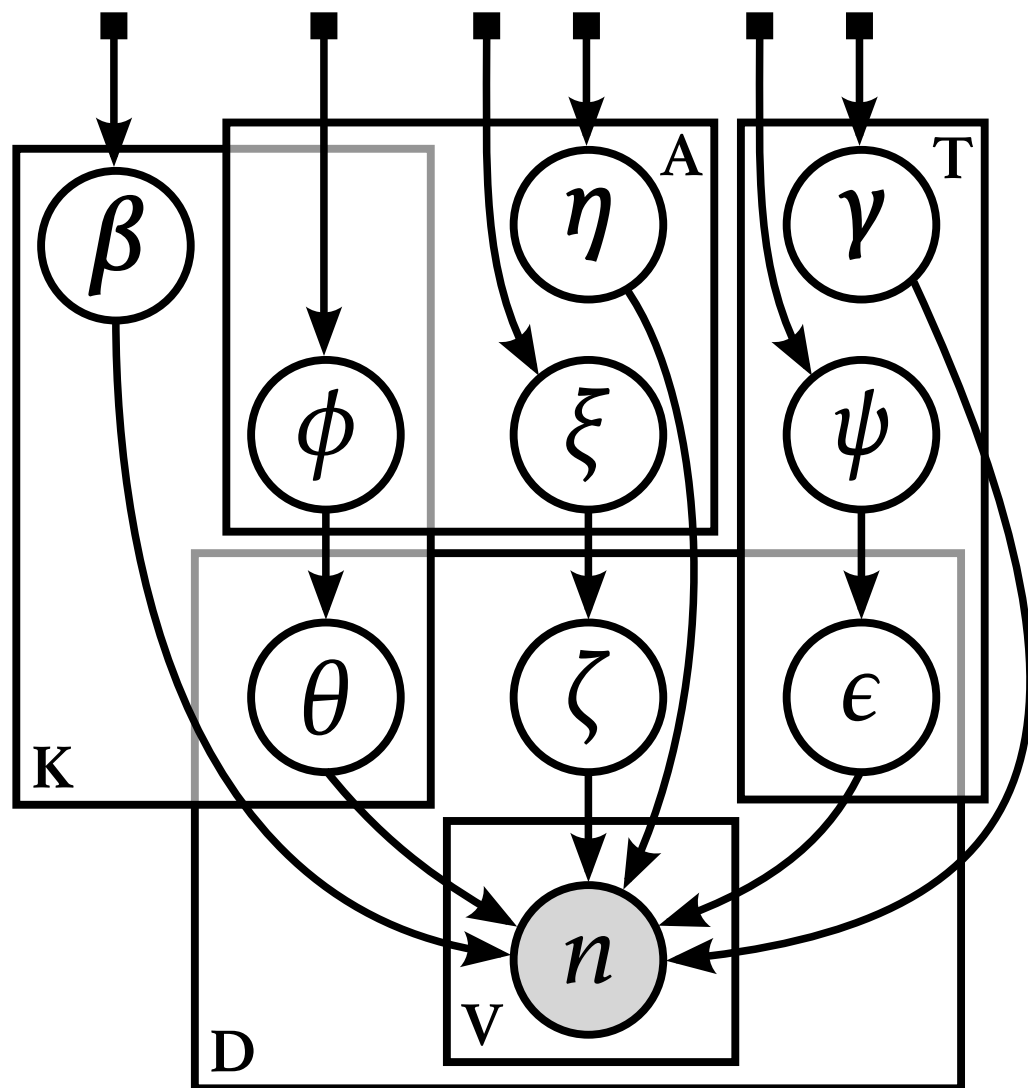
$$\gamma_t \sim \text{Dirichlet}(\cdot)$$

$$\theta_{dk} \sim \text{Gamma}(\dots)$$

$$\zeta_d \sim \text{Gamma}(\dots)$$

$$\epsilon_{dt} \sim \text{Gamma}(\dots)$$

$$n_{dv} \sim \text{Poisson} \left(\overset{\text{🌐}}{\sum_{k=1}^K \theta_{dk} \beta_{kv}} + \overset{\text{👤}}{\zeta_d \eta_{a_d v}} + \overset{\text{🕒}}{\sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{tv}} \right)$$



$$\beta_k \sim \text{Dirichlet}(\cdot)$$

$$\eta_a \sim \text{Dirichlet}(\cdot)$$

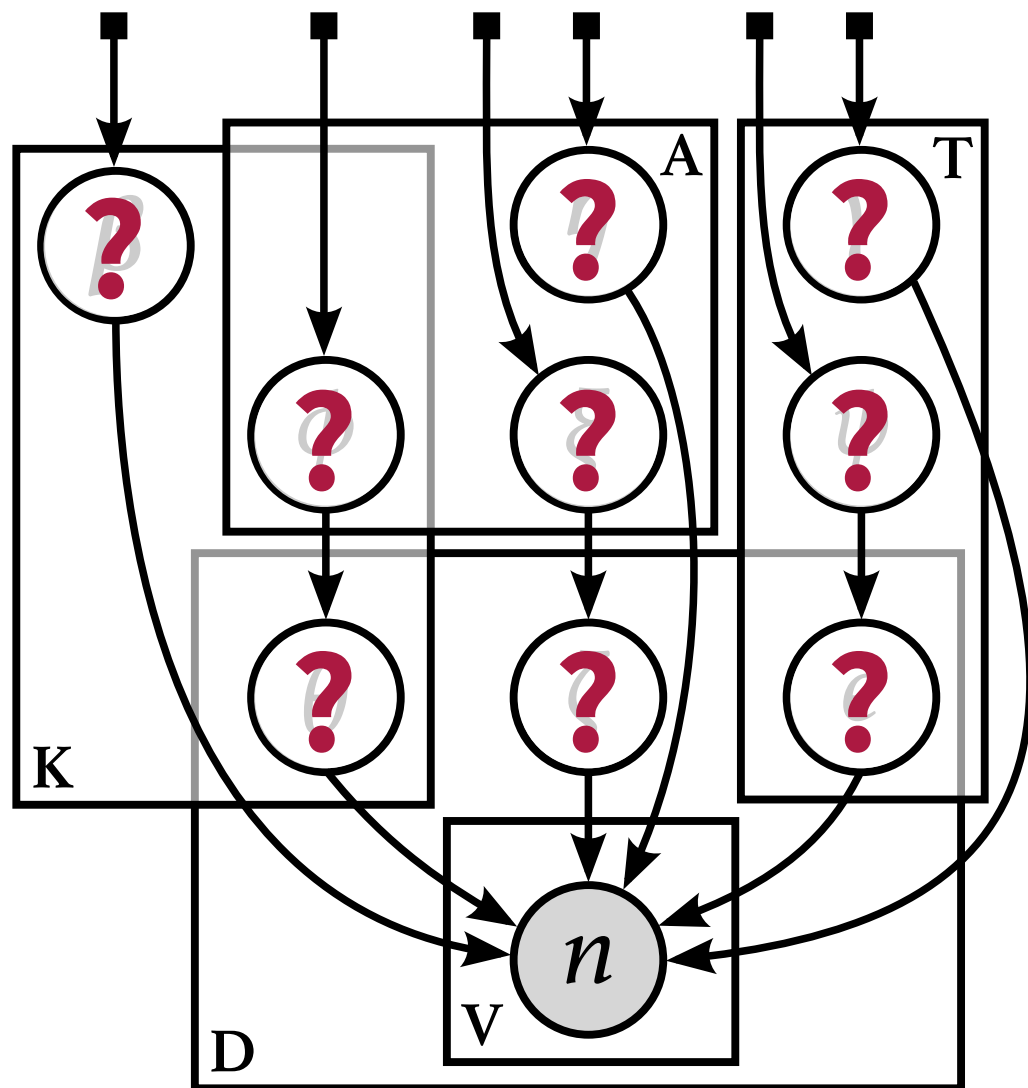
$$\gamma_t \sim \text{Dirichlet}(\cdot)$$

$$\theta_{dk} \sim \text{Gamma}(\dots)$$

$$\zeta_d \sim \text{Gamma}(\dots)$$

$$\epsilon_{dt} \sim \text{Gamma}(\dots)$$

$$n_{dv} \sim \text{Poisson} \left(\overset{\text{Globe}}{\sum_{k=1}^K \theta_{dk} \beta_{kv}} + \overset{\text{Person}}{\zeta_d \eta_{av}} + \overset{\text{Clock}}{\sum_{t=1}^T f(t_d, t) \epsilon_{dt} \gamma_{vt}} \right)$$



$$\beta_k \sim \text{Dirichlet}(\cdot)$$

$$\eta_a \sim \text{Dirichlet}(\cdot)$$

$$\gamma_t \sim \text{Dirichlet}(\cdot)$$

$$\theta_{dk} \sim \text{Gamma}(\dots)$$

$$\zeta_d \sim \text{Gamma}(\dots)$$

$$\epsilon_{dt} \sim \text{Gamma}(\dots)$$

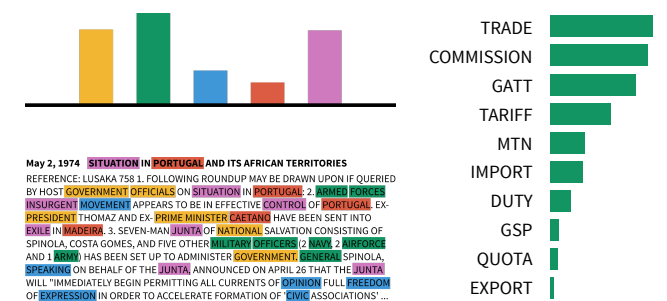
Data Analysis Process

observed data

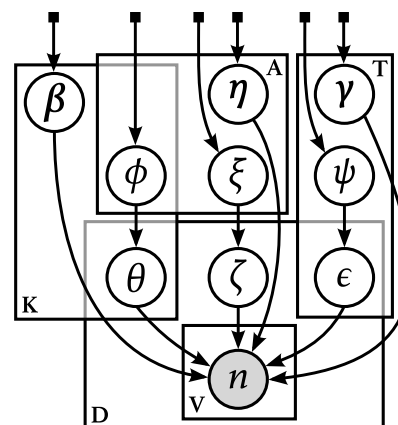
HONG KONG → BANGKOK on 5/12/1975
U.S. FLAG VESSEL IN DISTRESS
1. HONG KONG REP OF SEALAND
STATE → BANGKOK on 5/13/1975
ADVISE **SS MAYAGUEZ FOR MASTERS**
MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT
1. MOST MEXICO CITY NEWSPAPERS
GAVE LEAD TREATMENT TO
MAYAGUEZ INCIDENT IN MAY 15
EDITIONS. REPORTS, BASED ON WIRE
SERVICE DESPATCHES, WERE MOSTLY
PLAYED STRAIGHT ALTHOUGH

inference
algorithm

learned parameters



model assumptions



visualization &
exploration



Posterior Distribution

latent model parameters

easy to compute

$$p(\beta, \theta, \phi, \eta, \xi, \zeta, \gamma, \psi, \epsilon \mid \mathbf{N}, \alpha) = \frac{p(\beta, \theta, \phi, \eta, \xi, \zeta, \gamma, \psi, \epsilon, \mathbf{N} \mid \alpha)}{\int \cdots \int p(\beta, \theta, \phi, \eta, \xi, \zeta, \gamma, \psi, \epsilon, \mathbf{N} \mid \alpha) d\beta \cdots d\epsilon}$$

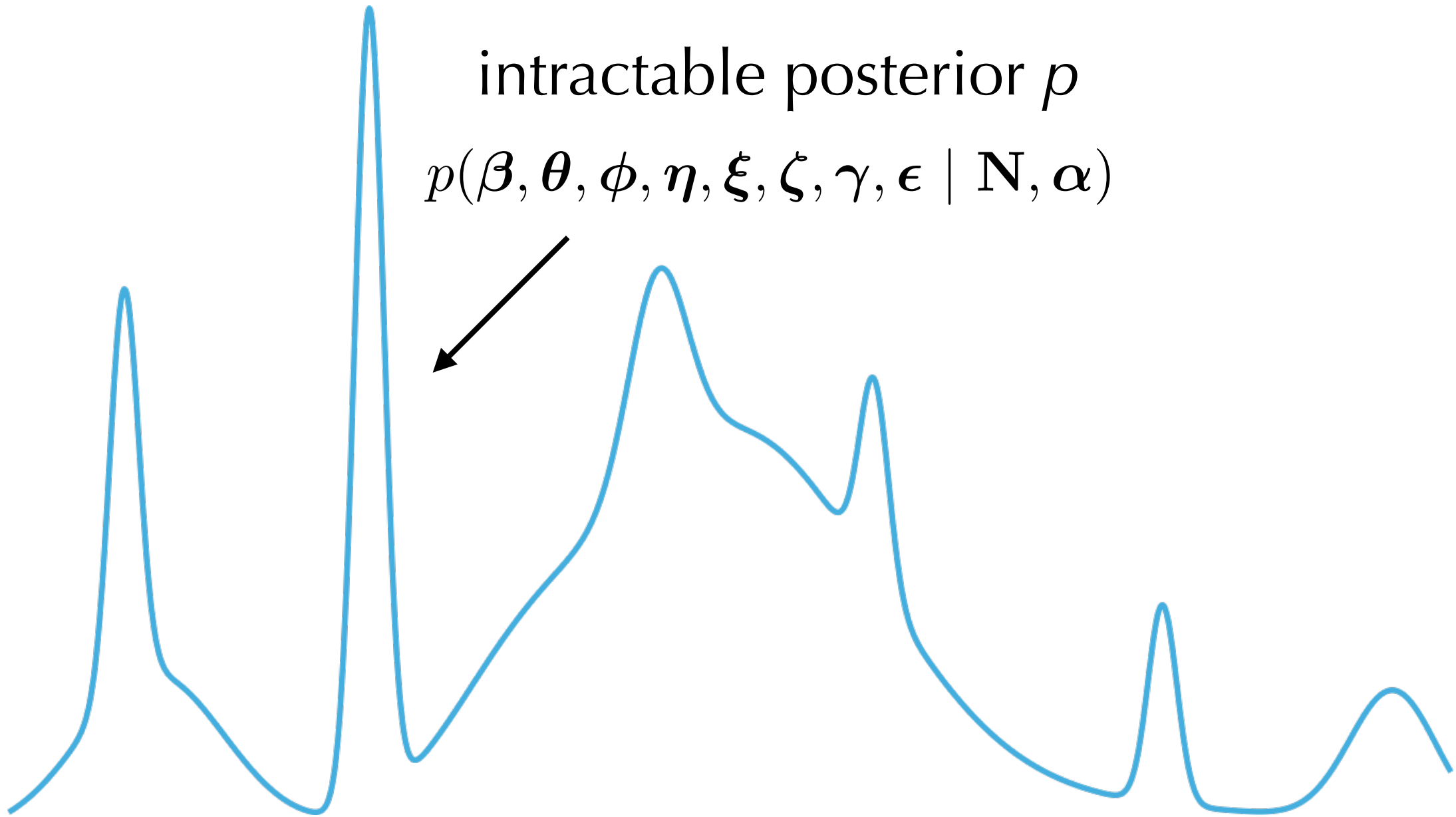
observed data

intractable

model hyperparameters

Variational Inference

intractable posterior p
 $p(\beta, \theta, \phi, \eta, \xi, \zeta, \gamma, \epsilon \mid \mathbf{N}, \alpha)$



Variational Inference

intractable posterior p

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon} \mid \mathbf{N}, \boldsymbol{\alpha})$$

easy to compute
approximation q

$$\approx q(\boldsymbol{\beta} \mid \lambda_{\boldsymbol{\beta}}) q(\boldsymbol{\theta} \mid \lambda_{\boldsymbol{\theta}}) q(\boldsymbol{\phi} \mid \lambda_{\boldsymbol{\phi}}) q(\boldsymbol{\xi} \mid \lambda_{\boldsymbol{\xi}}) q(\boldsymbol{\zeta} \mid \lambda_{\boldsymbol{\zeta}}) q(\boldsymbol{\gamma} \mid \lambda_{\boldsymbol{\gamma}}) q(\boldsymbol{\epsilon} \mid \lambda_{\boldsymbol{\epsilon}})$$

Variational Inference

coordinate ascent algorithm

$q(\boldsymbol{\beta} \mid \lambda_{\beta})$ update λ_{β} using estimates of other parameters

$q(\boldsymbol{\theta} \mid \lambda_{\theta})$ update λ_{θ} using estimates of other parameters

$q(\boldsymbol{\phi} \mid \lambda_{\phi})$ update λ_{ϕ} using estimates of other parameters

$q(\boldsymbol{\xi} \mid \lambda_{\xi})$ update λ_{ξ} using estimates of other parameters

$q(\boldsymbol{\zeta} \mid \lambda_{\zeta})$ update λ_{ζ} using estimates of other parameters

$q(\boldsymbol{\gamma} \mid \lambda_{\gamma})$ update λ_{γ} using estimates of other parameters

$q(\boldsymbol{\epsilon} \mid \lambda_{\epsilon})$ update λ_{ϵ} using estimates of other parameters

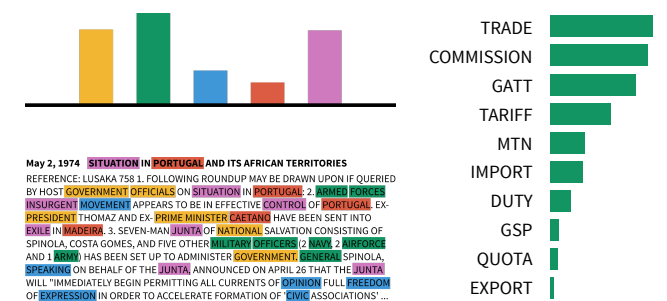
Data Analysis Process

observed data

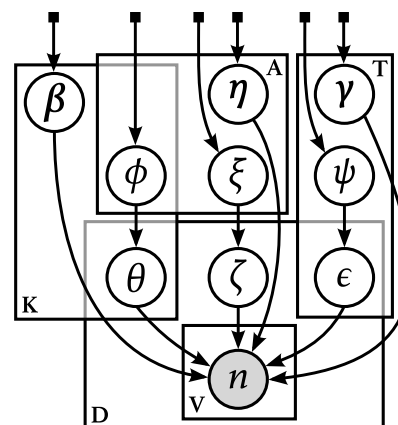
HONG KONG → BANGKOK on 5/12/1975
U.S. FLAG VESSEL IN DISTRESS
1. HONG KONG REP OF SEALAND
STATE → BANGKOK on 5/13/1975
ADVISE **SS MAYAGUEZ FOR MASTERS**
MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT
1. MOST MEXICO CITY NEWSPAPERS
GAVE LEAD TREATMENT TO
MAYAGUEZ INCIDENT IN MAY 15
EDITIONS. REPORTS, BASED ON WIRE
SERVICE DESPATCHES, WERE MOSTLY
PLAYED STRAIGHT ALTHOUGH

inference
algorithm

learned parameters



model assumptions



visualization &
exploration



Event Detection

Method	nDCG
Capsule (this work)	0.693
term-count deviation + tf-idf	0.652
term-count deviation	0.642
random	0.557
“event-only” Capsule	0.426

Document Recovery

THE ENCYCLOPEDIA OF **KIDNAPPINGS**

ERITREAN Liberation Front: terrorist kidnappers

A nationalist group dedicated to liberating the province of Eritrea from Ethiopia, the ELF made its debut with the bombing of an Ethiopian Airlines

. . .

after an unspecified payoff. Thirteen months later, on March 26, 1974, five Tenneco Oil employees were kidnapped after their helicopter was grounded by inclement weather near Massawa, Eritrea. The terrorists demanded Tenneco's help in freeing 75 political prisoners and finding an author to publicize ELF's story. They abducted a second pilot who came to pick up the hostages (at ELF's request) on May 27. The latest captive was released on June 23, and the survey team's pilot was freed three days later, with a written demand for \$1 million ransom.

Document Recovery

Our Model

June 2, 1974 STATE TENNECO SEVEN

June 1, 1974 ATHENS TENNECO SEVEN: PRESS STATEMENT
STATE FOR BRUCE ROGERS AF/ET

May 29, 1974 STATE TENNECO FIVE AND AMERICAN NURSE

May 30, 1974 STATE TENNECO SEVEN

June 8, 1974 ATHENS TENNECO SIX FREEMAN GAVE EMBASSY
ADVANCE COPY OF PRESS RELEASE HE

June 8, 1974 ATHENS TENNECO SIX

June 5, 1974 ATHENS TENNECO IN ETHIOPIA

June 3, 1974 STATE TENNECO SEVEN

June 7, 1974 STATE TENNECO IN ETHIOPIA

June 3, 1974 STATE TENNECO SEVEN

May 27, 1974 ADDIS
ABABA TENNECO FIVE

June 1, 1974 ATHENS TENNECO SEVEN FOR WELDON
GRAHAM WHO ARRIVES KHARTOUM

May 27, 1974 (random)

May 31, 1974 ABIDJAN IVORIAN TO HEAD UNIVERSITY

May 29, 1974 BREMEN VISAS IBEX NAME CHECK USSR M.S.
"MIKHAIL LERMONTOV" ARRIVING

May 28, 1974 STATE NONIMMIGRANT VISA CASE

May 28, 1974 BONN ADS REQUEST - E.E. FAIRCHILD
CORP., ROCHESTER, N.Y.

May 29, 1974 BUDAPEST HUNGARIAN-POLISH COOPERATION
PROTOCOL SIGNED

May 29, 1974 MOSCOW CAMPAIGN TO DISCOURAGE SOVIET
JEWISH EMIGRATION

May 31, 1974 MONTREAL PRIVATE TRADE OPPORTUNITY

May 31, 1974 WARSAW SHIPMENT OF SPECIAL EQUIPMENT

May 28, 1974 FRANKFURT NUCLEAR INSTRUMENTATION
EXHIBITION MARCH 18-21 1974

May 28, 1974 ARA THE REMOVAL OF FENCE BETWEEN
THE REPUBLIC OF PANAMA AND

June 1, 1974 LIBREVILLE GOG INTEREST IN DISCUSSING
FUTURE OF IRON ORE MINE WITH

May 27, 1974 WARSAW CULTURAL PRESENTATIONS:
MANHATTAN THEATRE PROJECT

Document Recovery: Cables

Fall of Saigon / End of Vietnam War



Document Recovery: Cables

Our Model

April 30, 1975	GEORGETOWN	RESETTLEMENT OF VIETNAMESE REFUGEES STATE FOR AMB. L. DEAN
May 1, 1975	JIDDA	FOREIGN RELATIONS OF PROVISIONAL REVOLUTIONARY GOVERNMENT OF SOUTH VIET-NAM
April 29, 1975	ASUNCION	RESETTLEMENT OF VIETNAMESE REFUGEES
May 1, 1975	LAGOS	RESETTLEMENT OF VIETNAMESE
April 28, 1975	GABORONE	RESETTLEMENT OF VIETNAMESE REGUGEEES
April 29, 1975	LA PAZ	RESETTLEMENT OF VIETNAMESE REFUGEES
April 28, 1975	WELLINGTON	RESETTLEMENT OF VIETNAMESE REFUGEES
April 29, 1975	BRIDGETOWN	RESETTLEMENT OF VIETNAMESE REFUGEES
April 28, 1975	COTONU	RESETTLEMENT OF VIETNAMESE REFUGEES
May 15, 1975	NATO	MBFR: FORM OF AGREEMENTS: SPC MEETING MAY 15
April 29, 1975	BANGUI	RESETTLEMENT OF VIETNAMESE REFUGEES
May 2, 1975	LAGOS	n/a
May 10, 1975	CARACAS	FOREIGN MINISTER TO BE QUESTIONED CONCERNING GOV FOREIGN POLICY TOWARD

April 28, 1975 (random)

April 28, 1975	PRAGUE	MCCLAIN FAMILY BAND
April 30, 1975	CANNON, HOWARD	DELEGATE SELECTION TO IWY CONFERENCE
May 1, 1975	TOKYO	CONSULTATION ON PROPOSED INTERNATIONAL FUND FOR
April 29, 1975	VO	REQUEST FOR EVACUATION / VISA ASSISTANCE FROM VIETNAM
May 2, 1975	USUN	BRITISH TO ASK TURKEY TO DELAY NEW YORK PUBLICATION OF TURK CYPRIOT
April 29, 1975	LAGOS	MERIDIAN HOUSE INVITEE
May 2, 1975	NATO BRUSSELS	NPG-DRAFT AGENDA FOR JUNE NPG MINISTERIAL MEETING
May 2, 1975	TUNIS	DISPLAY OF F5F AIRCRAFT
April 29, 1975	OECD PARIS	US EXCEPTION REQUESTS
April 30, 1975	STATE	PA/SFCP OFFICIAL VISIT DR. BAUMGARTNER; PROJECT 5-532-13
April 28, 1975	NEW DELHI	W/W-AIRLINE TICKET FOR ERIC JOHN VAN LIENDEN
April 29, 1975	STATE	EXECUTIVE ORDER 10422 FOLLOWING PROCESSED UNDER SECTION 8
May 2, 1975	USUN	SCIENTIFIC AND TECHNICAL NEW YORK SUBCOMMITTEE (OUTER SPACE)

Visualization for Exploration



princeton.edu/~achaney/capsule/

Application to Clinton Emails

From: Vinograd, Samantha

Sent: Wednesday,
September 12, 2012 06:41 PM

Subject: hey

— for S's awareness — we
have proposed the calls to
Libyan and Egyptian
Presidents. Early AM EDT.

Samantha Vinograd

Senior Advisor to the
National Security Advisor

From: Otero, Maria

Sent: Wednesday,
September 12, 2012 6:46 AM

Subject: I am so sorry

Hillary:

I'm just boarding plane to
Honduras and thinking of
you especially with this
painful tragedy in Libya.

Warmest,

Maria

From: Brose, Christian

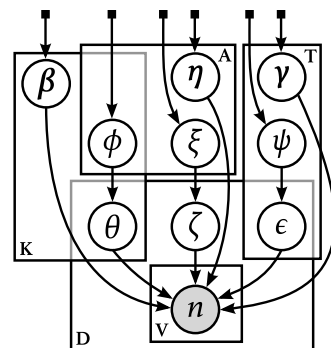
Sent: Wednesday,
September 12, 2012 10:09 AM

Subject: Wow

What a wonderful, strong
and moving statement by
your boss. please tell her
how much Sen. McCain
appreciated it. Me too

Summary

described an unsupervised model for
detecting and characterizing events
explored results on U.S. State Department cables
and Clinton emails, including demonstrating a
browser of the cables

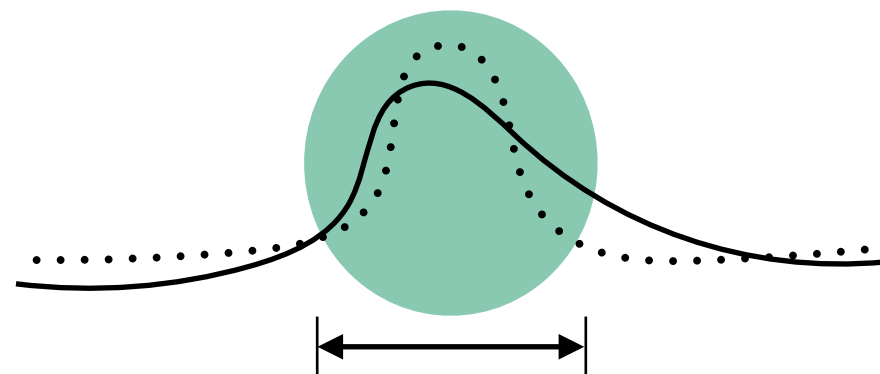
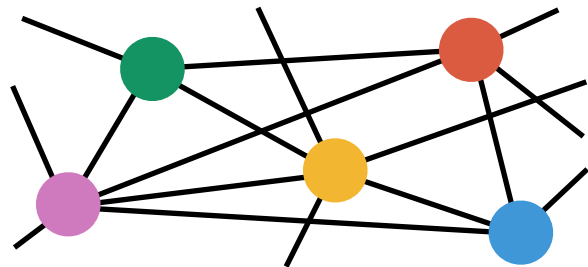


Next Steps

network dynamics
(relationships between entities)

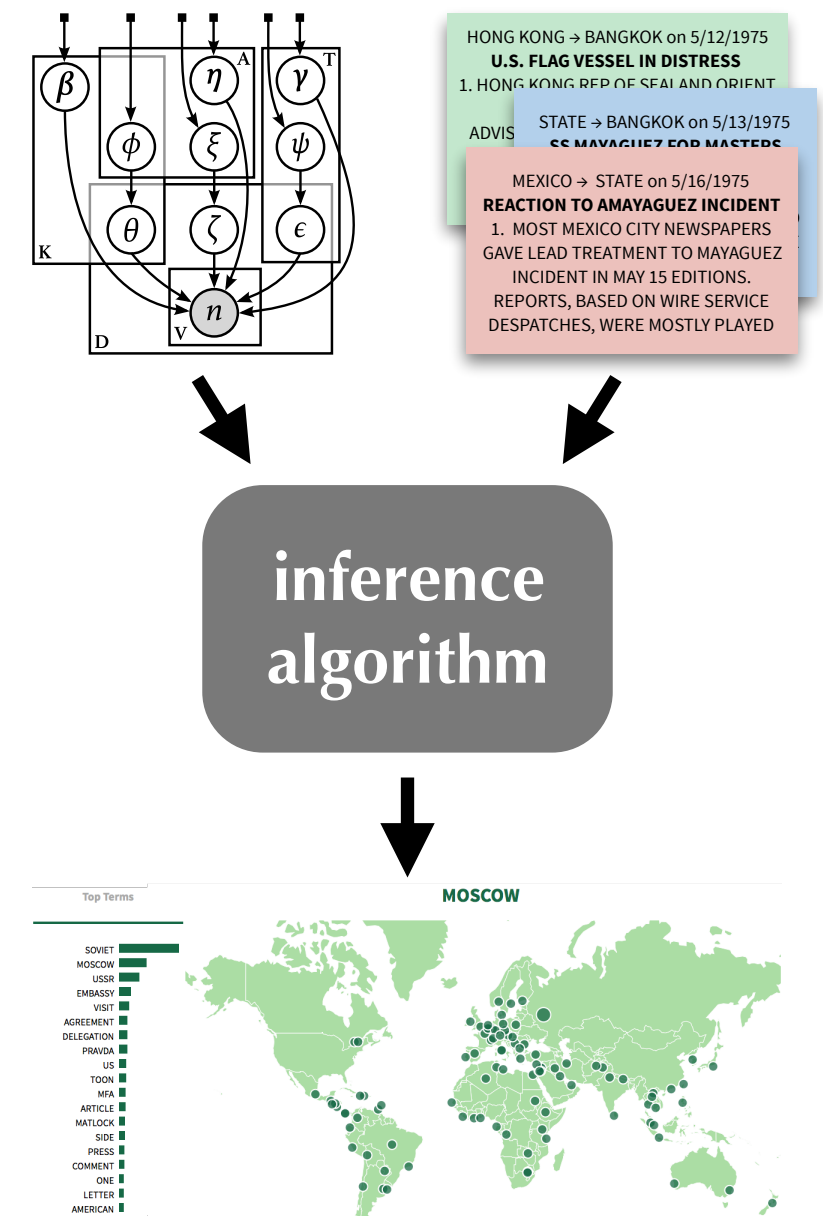
propagation of information along network

learning event duration or decay shape



Research Themes: Methodology

- work with domain experts to build latent variable models
- derive, implement, and apply scalable inference algorithms
- build tools to visualize, explore, and interpret model results

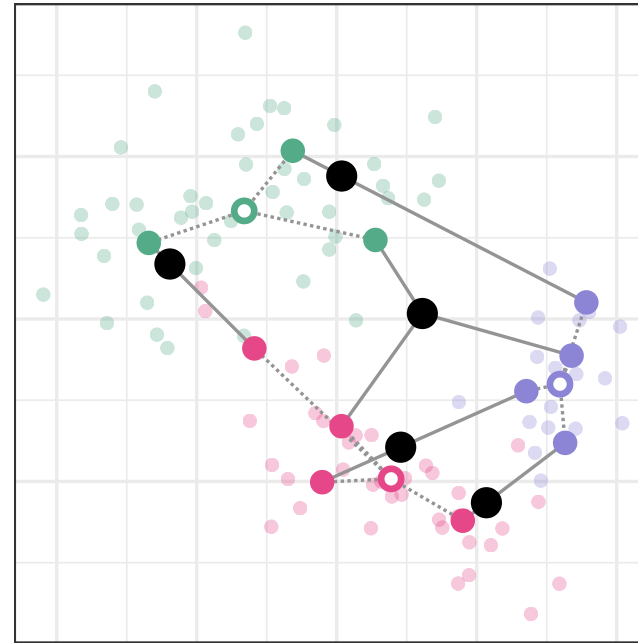


Research Themes: Applications

Goal: identify influences on human behavior

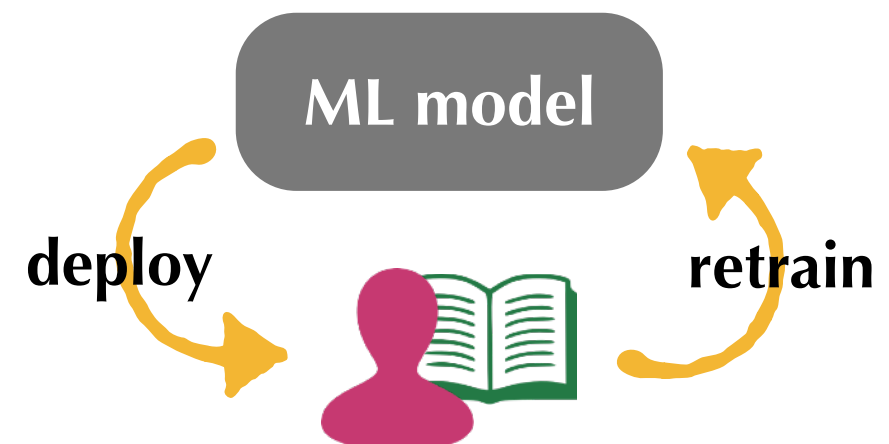
- convolved observations

- political voting
- demography
- financial markets
- gene expression



- algorithms as a source of influence

- news and media recommendation
- text correction and prediction
- financial loans
- criminal profiling
- medical intervention



Collaborators



Hanna Wallach



Matt Connelly



David Blei

Thank you!

ajbc.io/capsule

Postdoctoral Advisors Barbara Engelhardt & Brandon Stewart
Princeton Center for Information Technology Policy Seed Grant
IC Postdoctoral Research Fellowship Program (ORISE, DOE & ODNI)