A Bayesian Model for Detecting and Characterizing Events

Allison J.B. Chaney

David Blei, advisor

Pre-Final Public Oral
Department of Computer Science
Princeton University
February 24, 2016

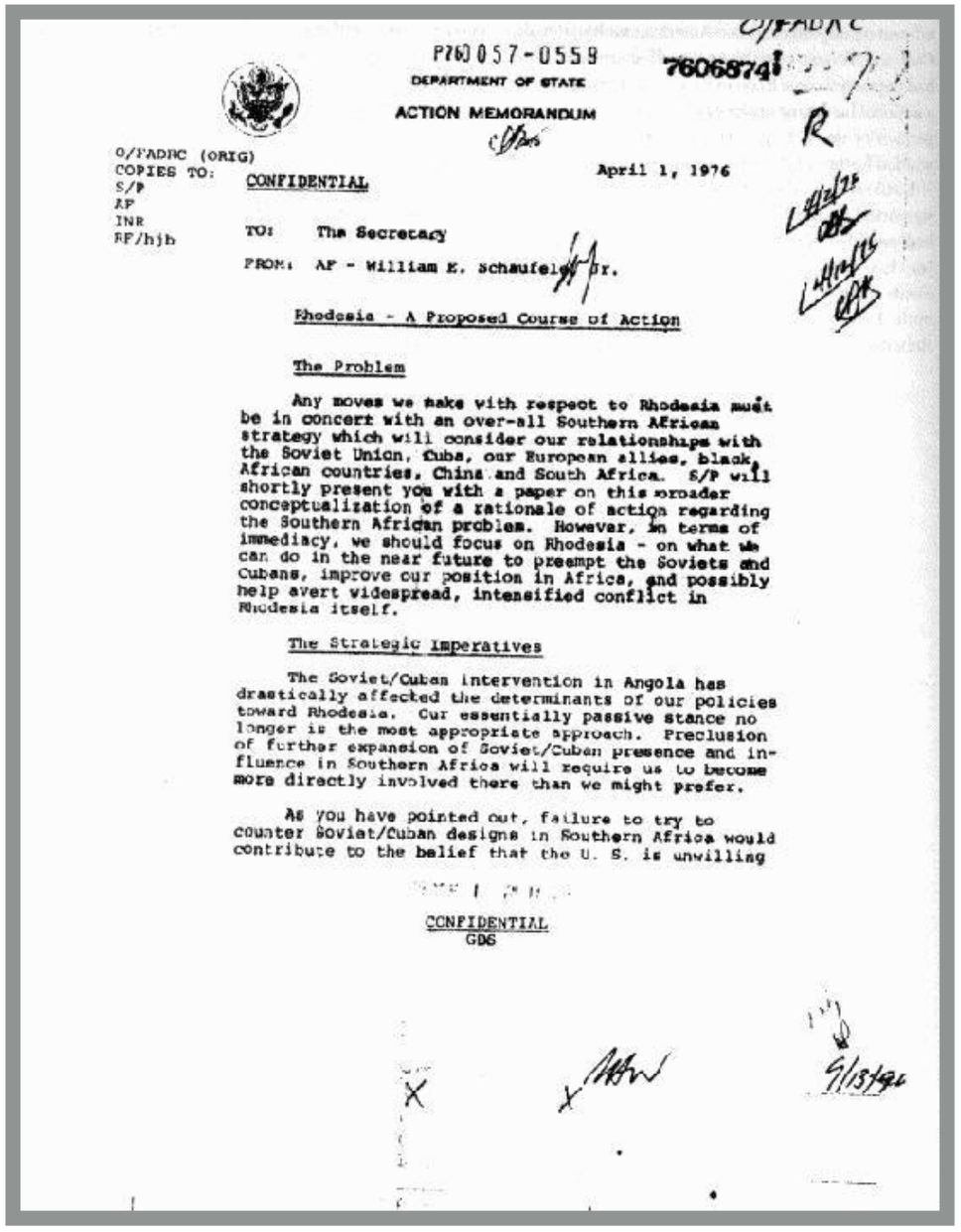
We can do nothing but scrutinize historical events themselves if we want to discover what they are.

– Dean W.R. Matthews, *What is an Historical Event?*

Matthew Connelly's History Lab at Columbia

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

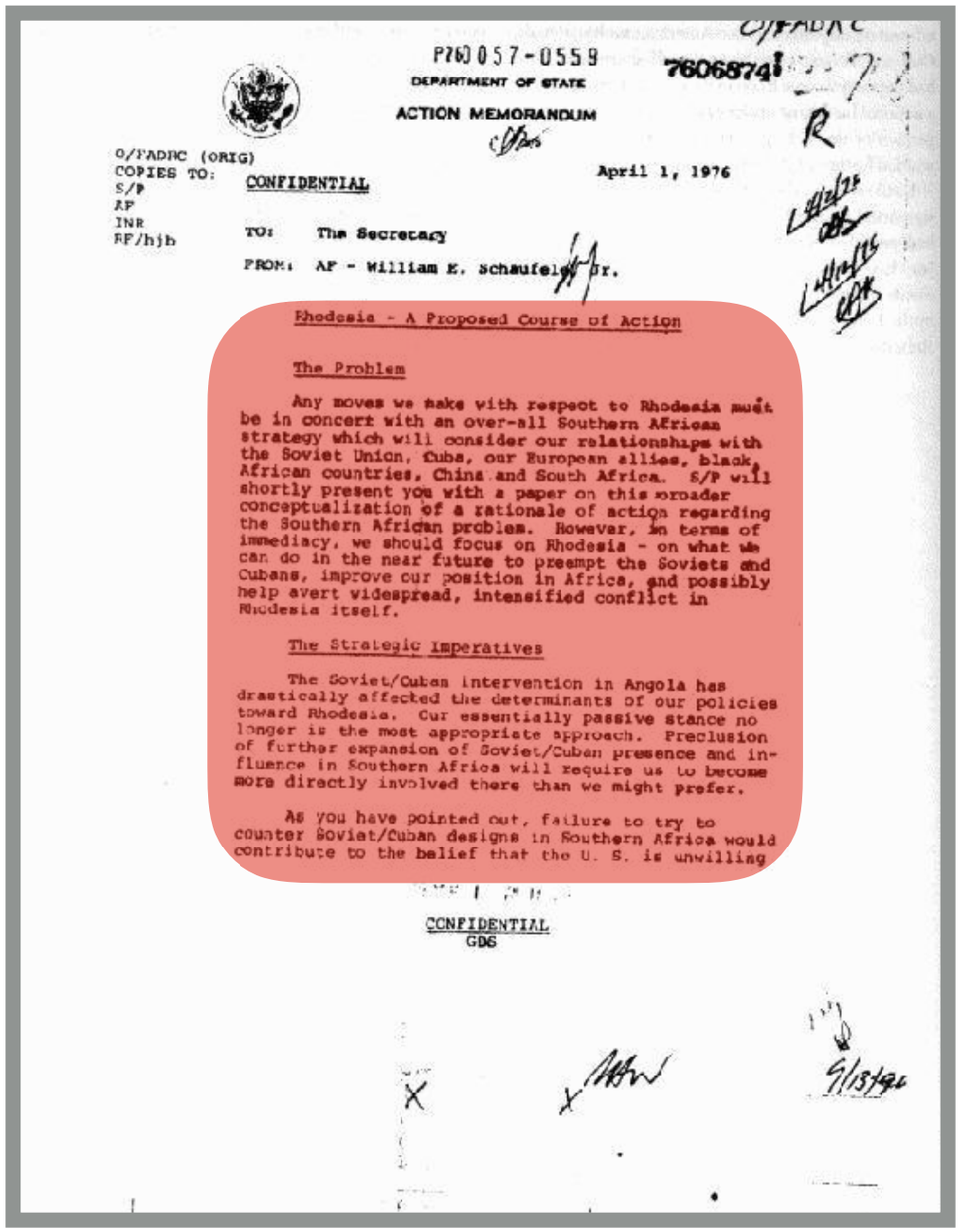
CONFIDENTIAL
GDS

9/13/94

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

Message content (text)



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

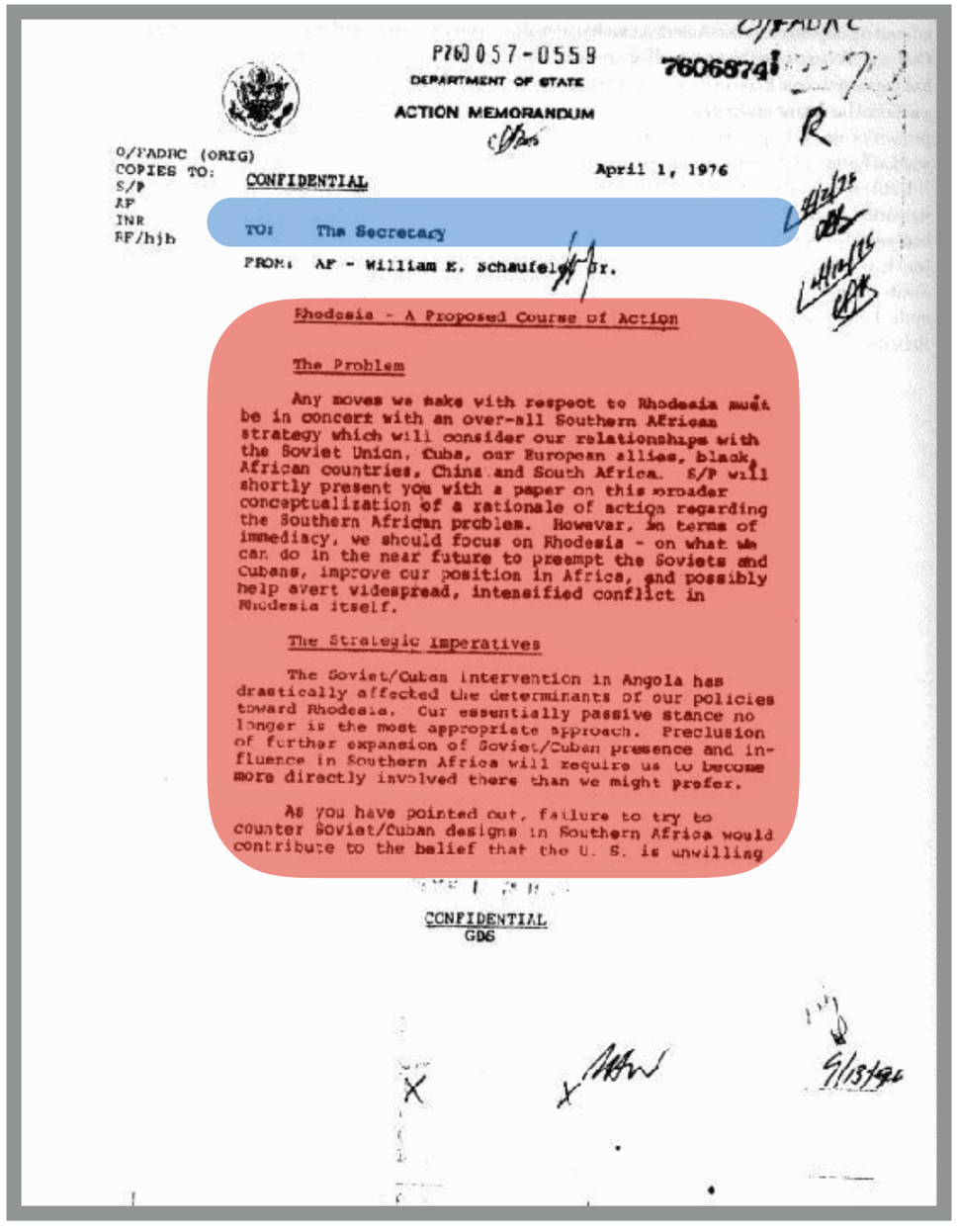
CONFIDENTIAL
GDS

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

Message content (text)

Sending entity (embassy, department, or individual)



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

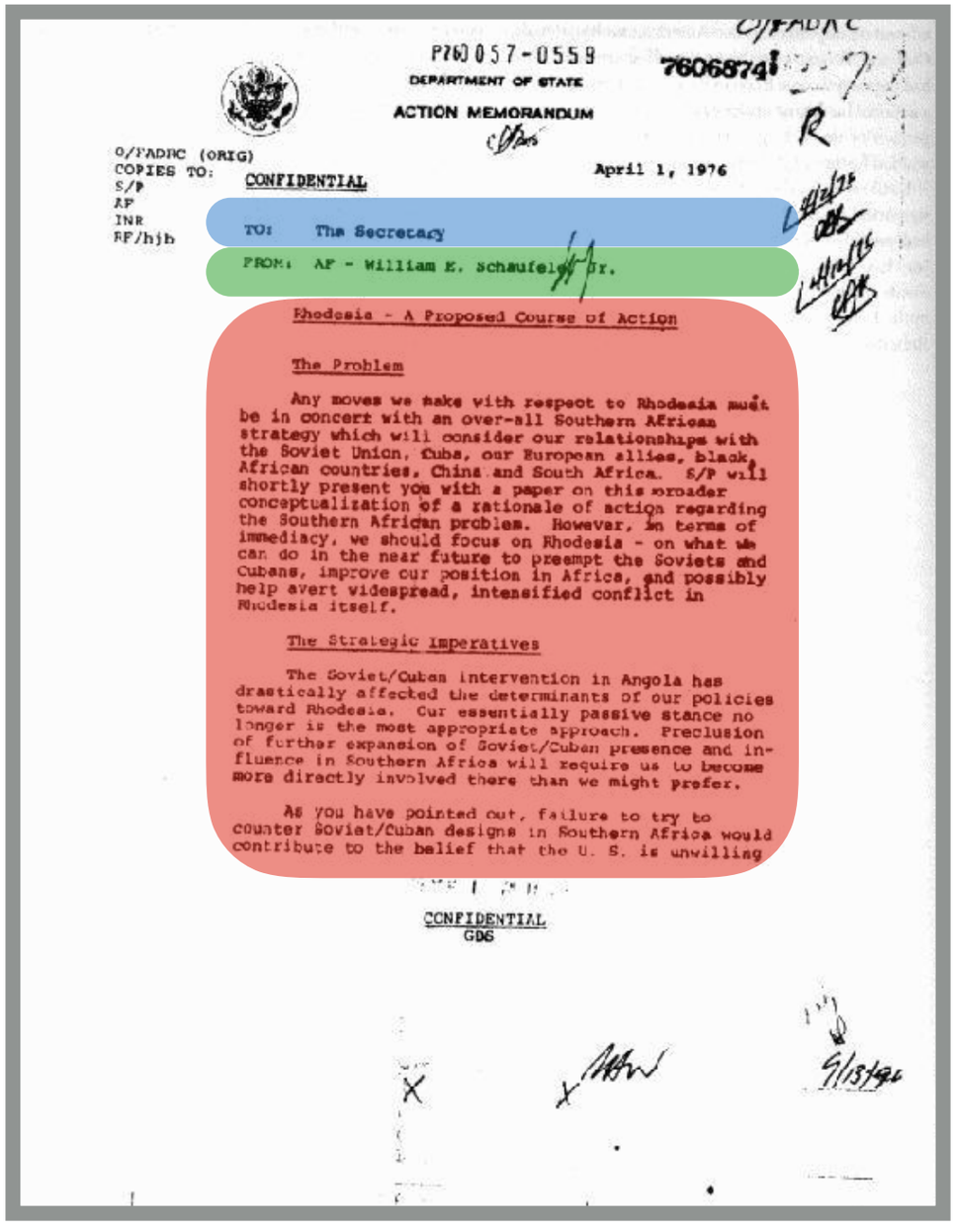
Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

Message content (text)

Sending entity (embassy, department, or individual)

One or more receiving entities



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

Matthew Connelly's
History Lab at Columbia

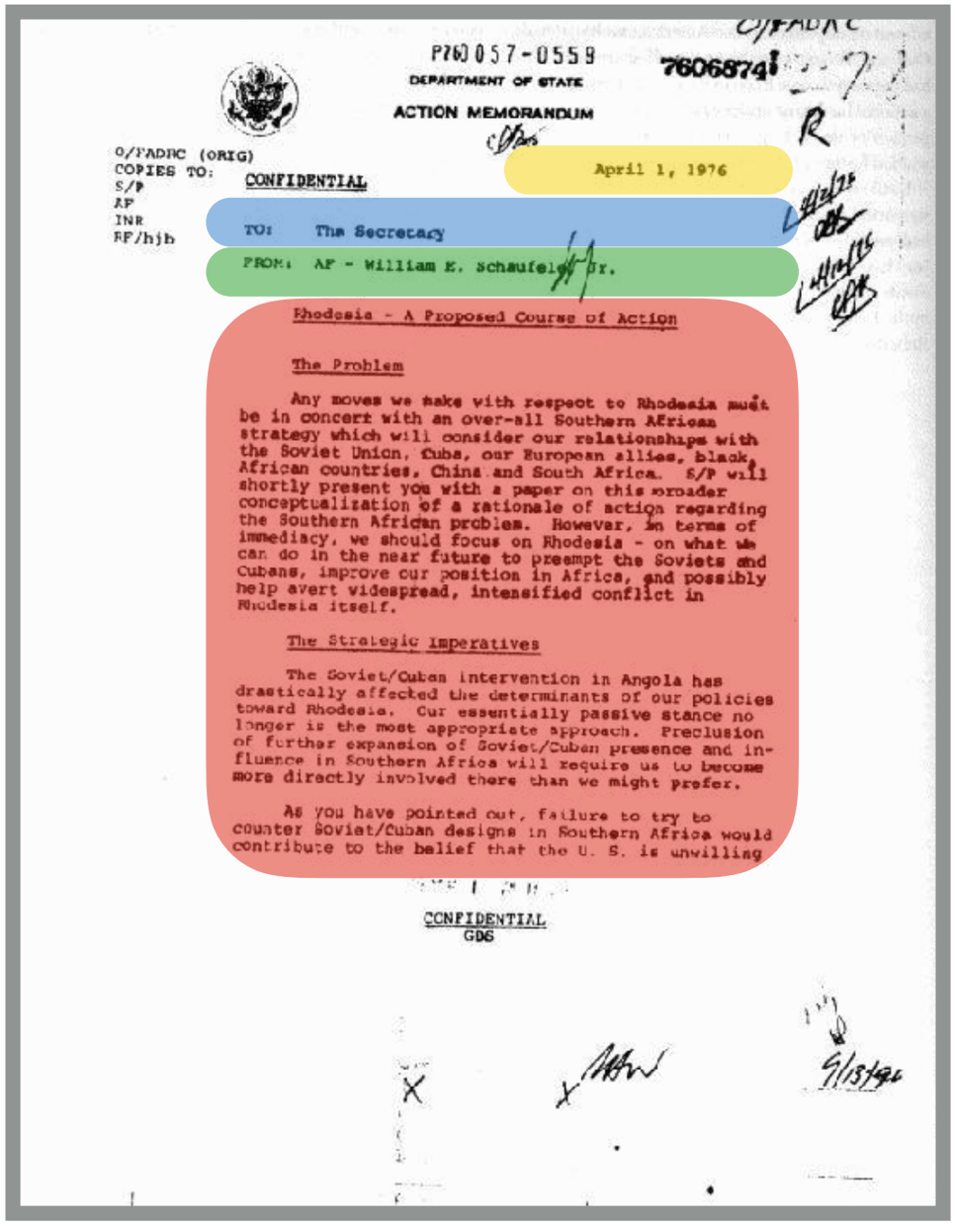
U.S. State Department Cables

Message content (text)

Sending entity (embassy, department, or individual)

One or more receiving entities

Send date



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

9/13/96

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

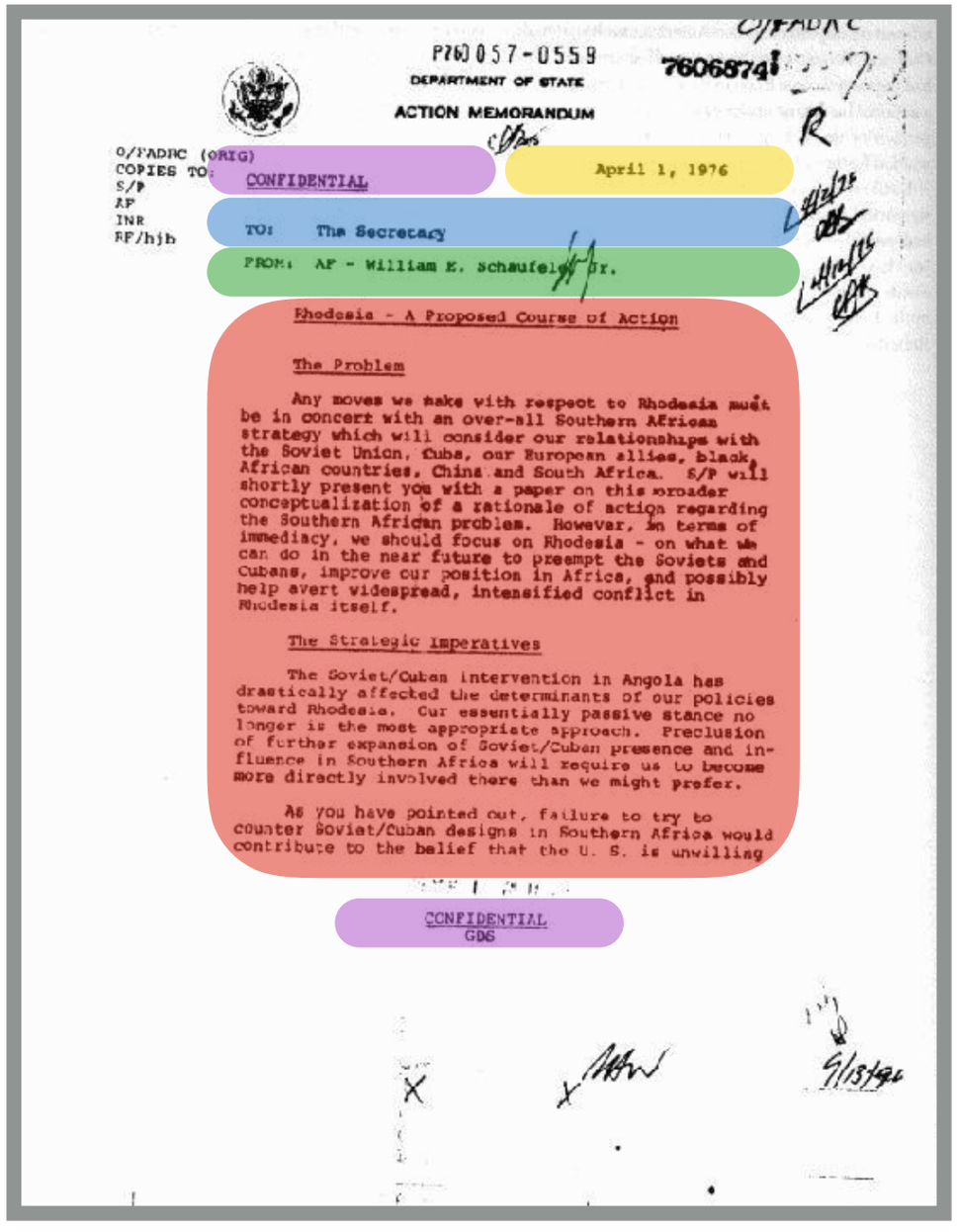
Message content (text)

Sending entity (embassy, department, or individual)

One or more receiving entities

Send date

Classification level



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

9/13/96

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

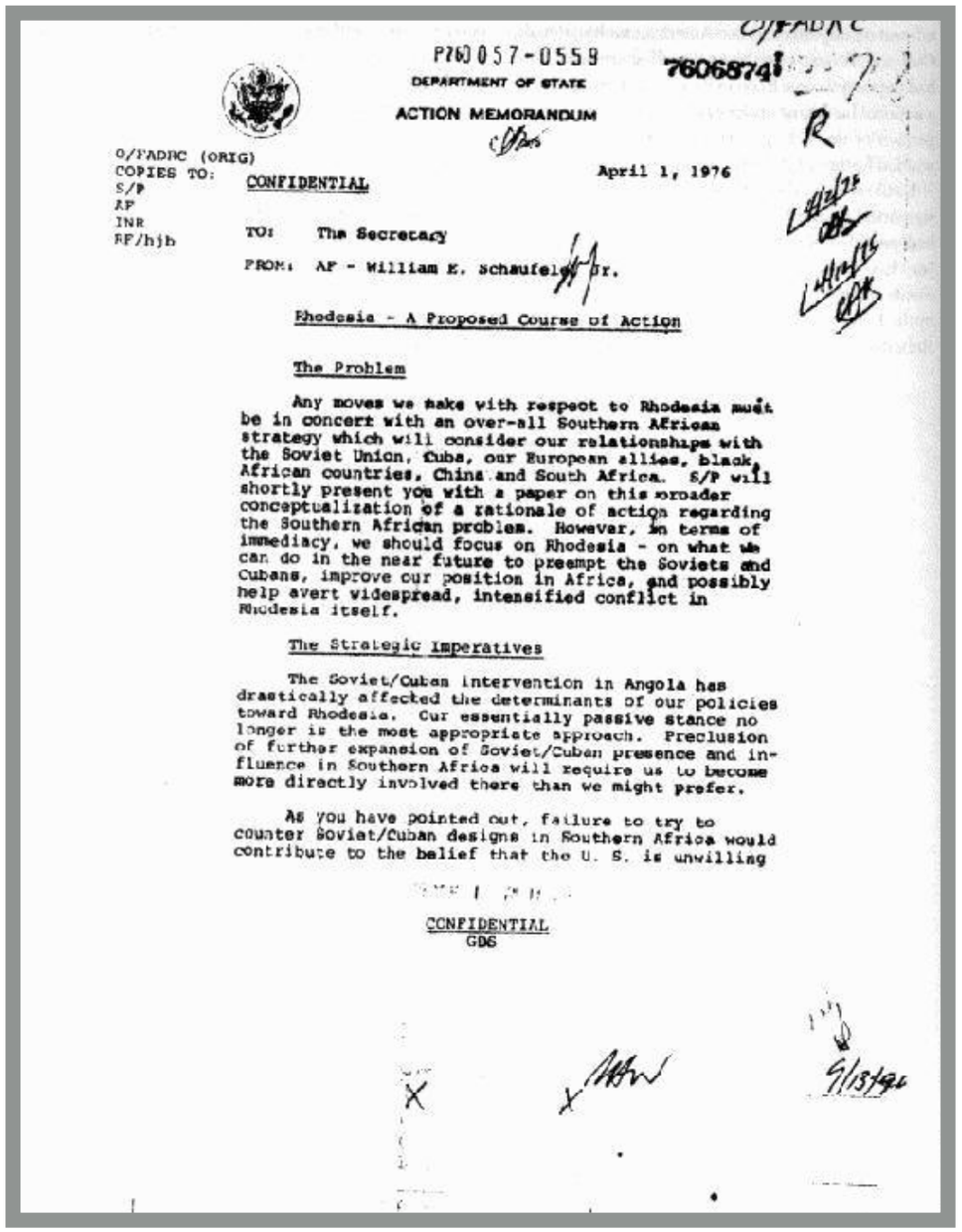
2,674,486 messages

sent between 1973 and 1978

34,204 unique sending entities

23.4% sent from State Dept.

41.6% sent to State Dept.



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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.

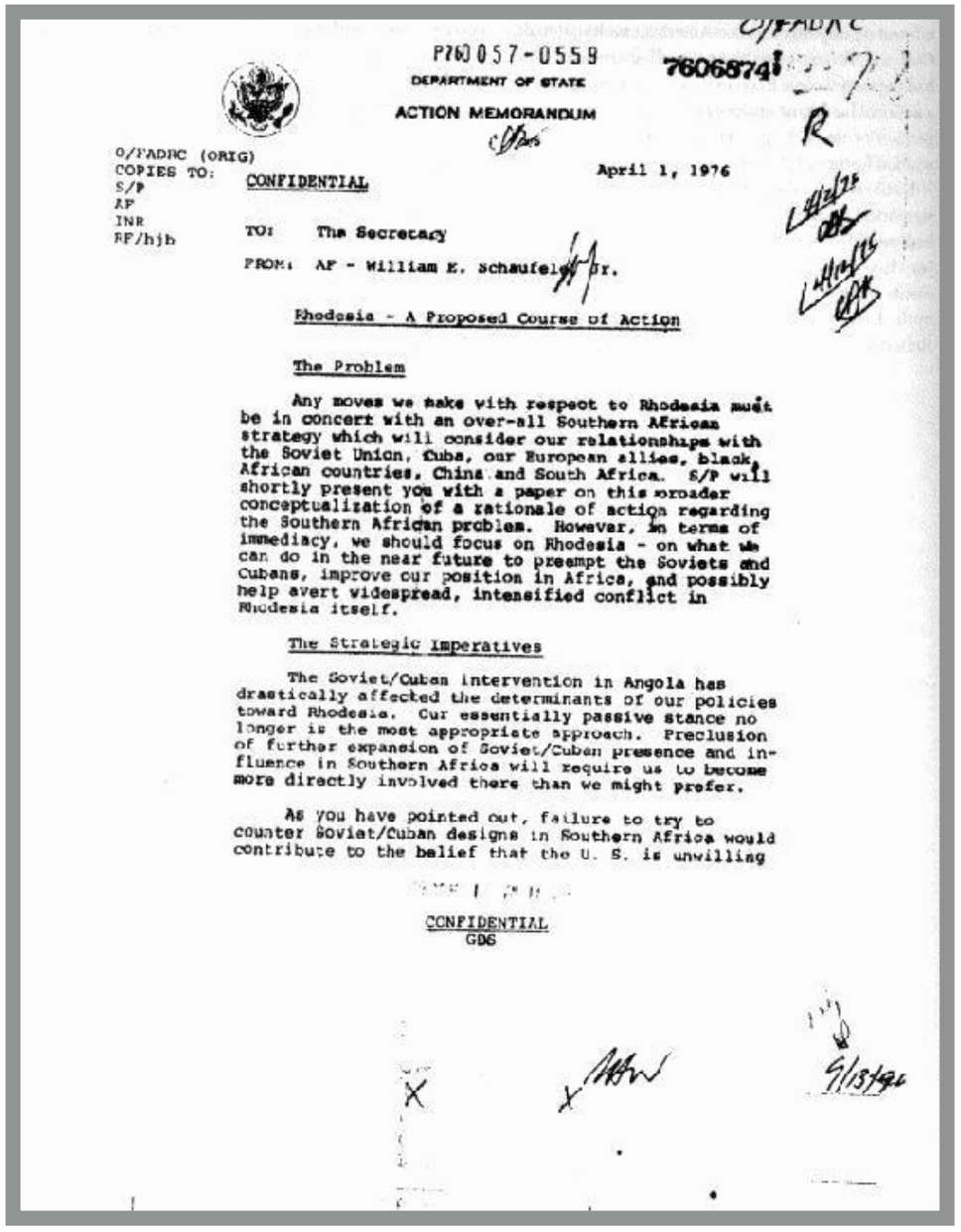
As you have pointed out, failure to try to counter Soviet/Cuban designs in Southern Africa would contribute to the belief that the U. S. is unwilling

CONFIDENTIAL
GDS

9/13/96

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

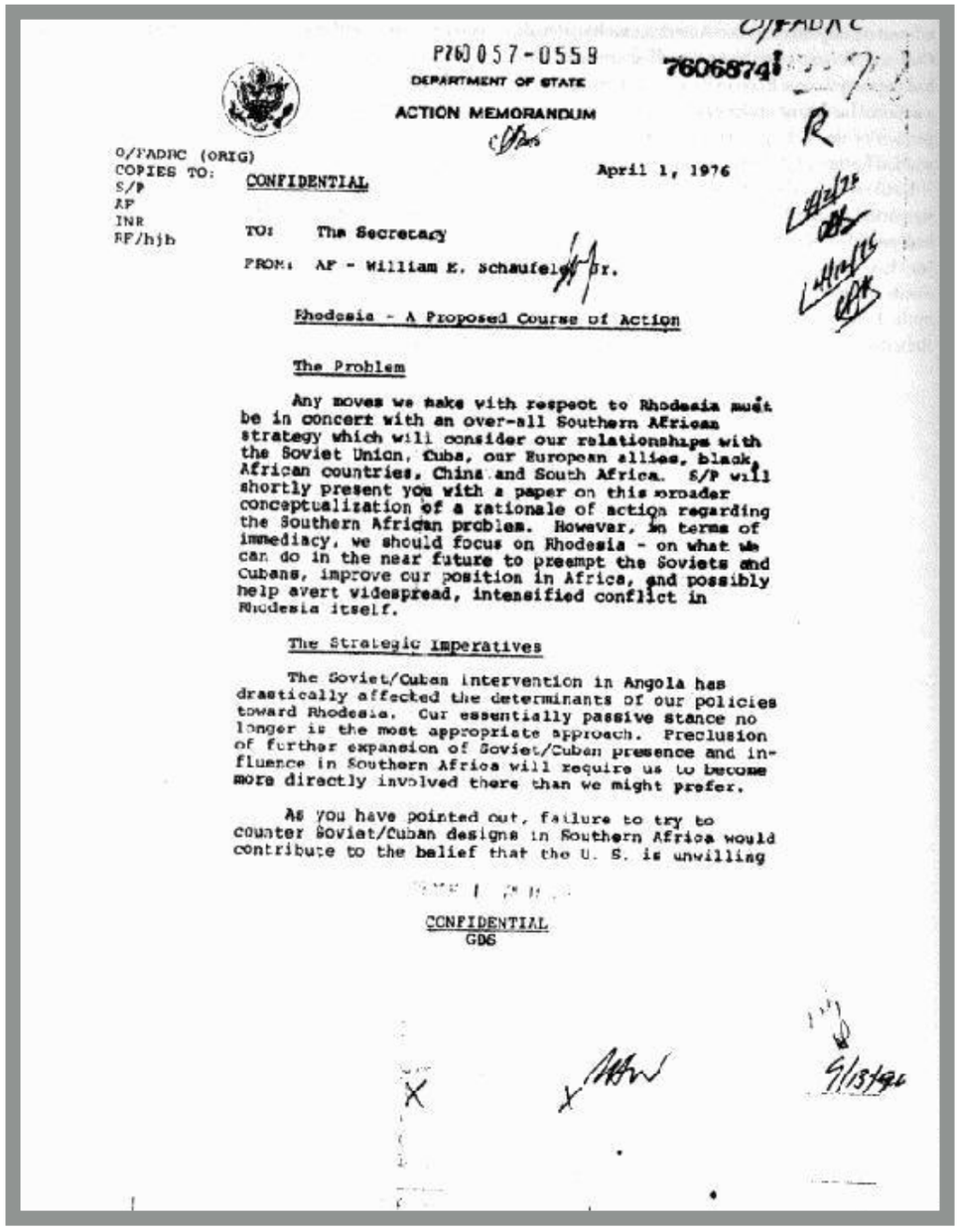
CONFIDENTIAL
GDS

9/13/94

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

What interesting events can be found in these messages?



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

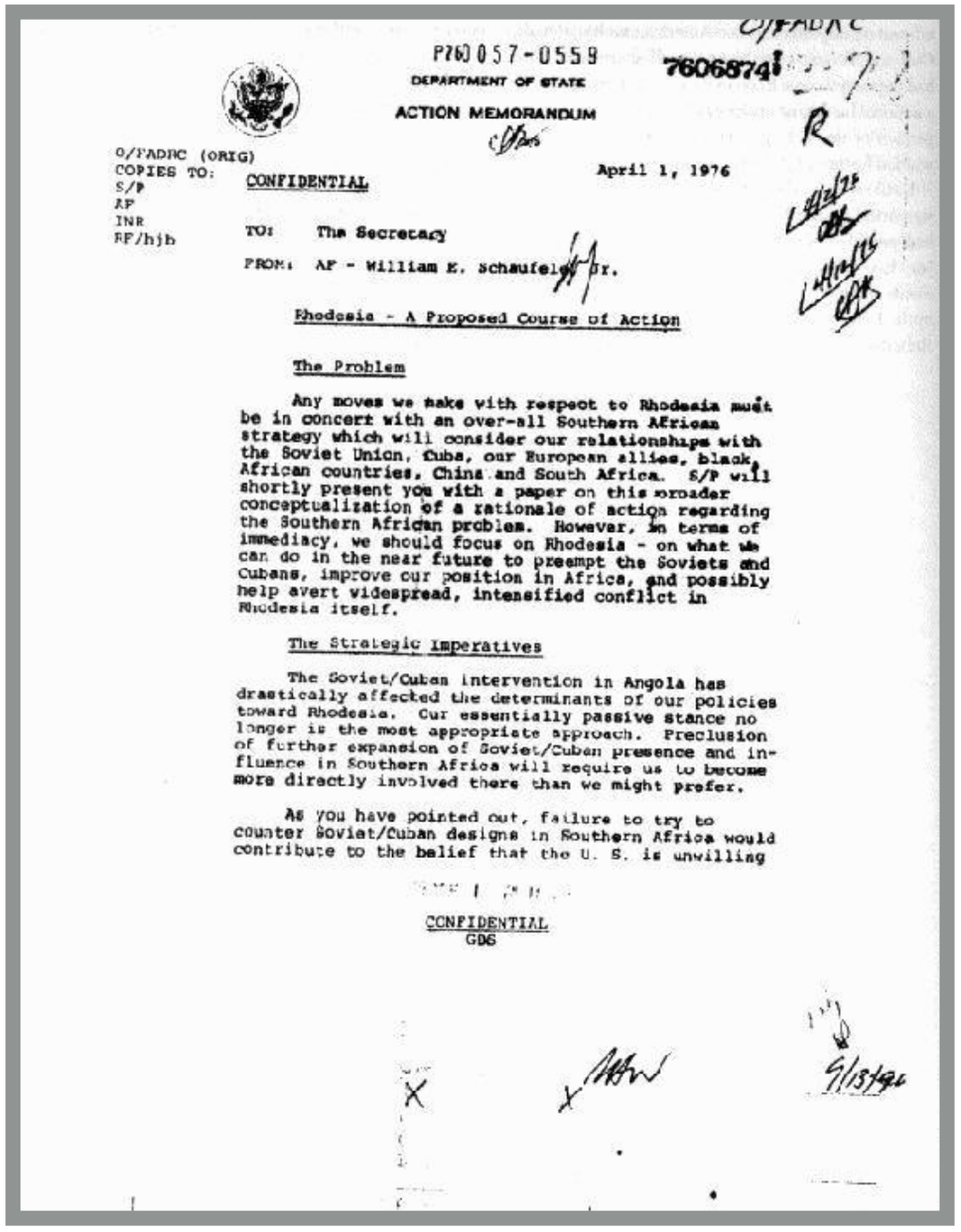
9/13/94

Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

What interesting events can be found in these messages?

How can we describe events?



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

9/13/94

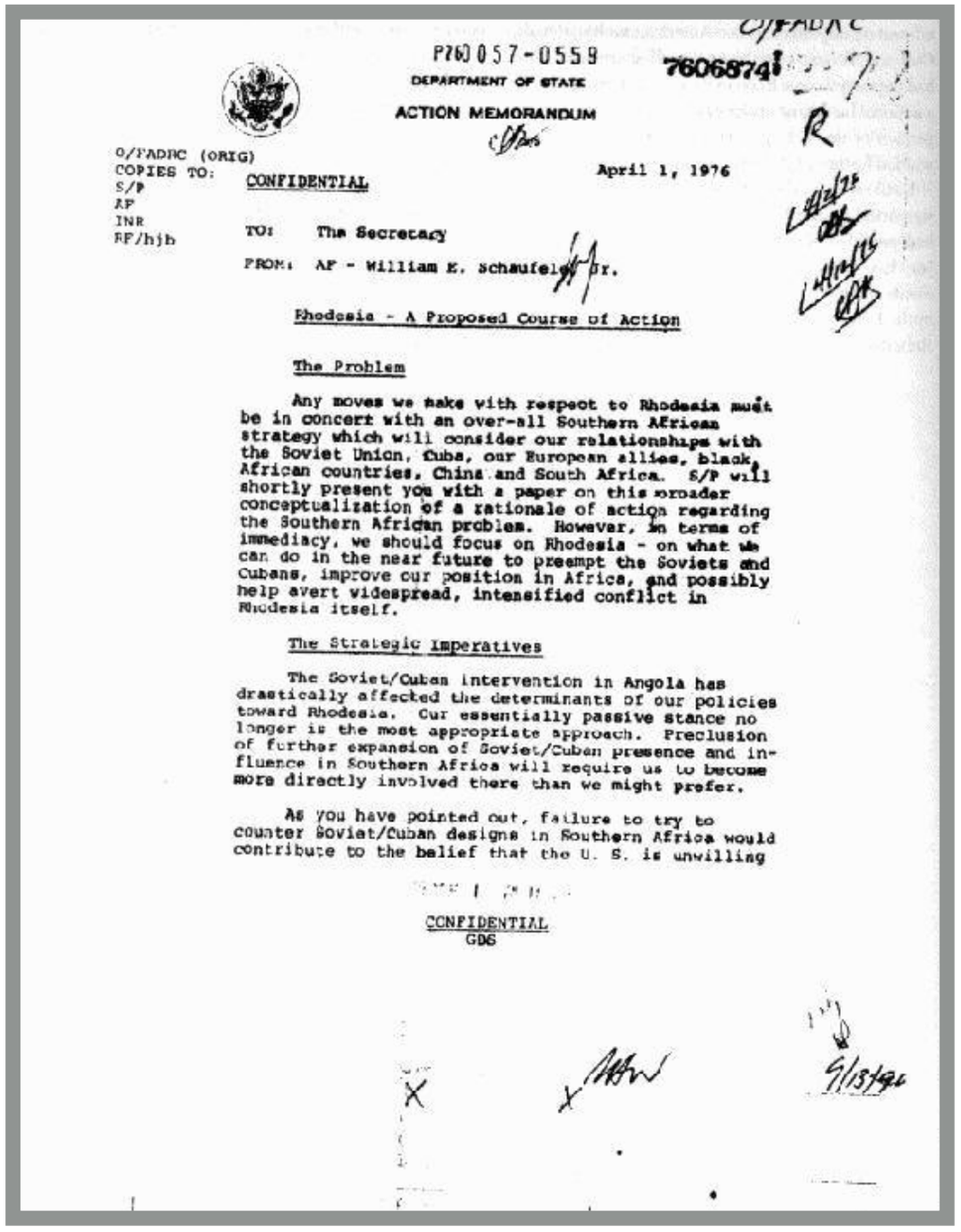
Matthew Connelly's
History Lab at Columbia

U.S. State Department Cables

What interesting events can be found in these messages?

How can we describe events?

What relationships do different entities have?



P760057-0559

DEPARTMENT OF STATE

7606874

ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

9/13/94

Matthew Connelly's
History Lab at Columbia

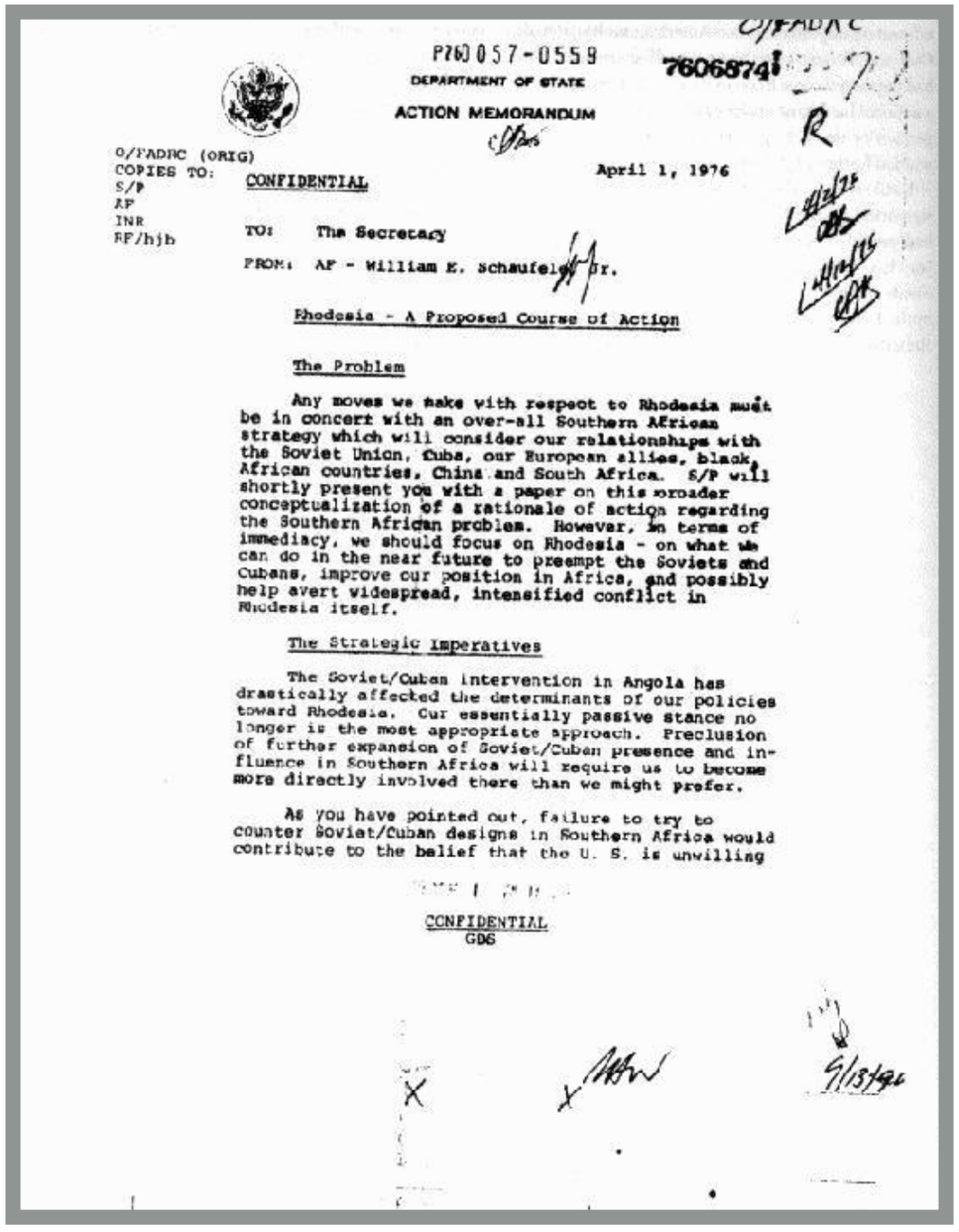
U.S. State Department Cables

What interesting events can be found in these messages?

How can we describe events?

What relationships do different entities have?

What are the typical concerns of different entities?



P760057-0559
DEPARTMENT OF STATE
7606874
ACTION MEMORANDUM

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

CONFIDENTIAL

April 1, 1976

TO: The Secretary

FROM: AF - William E. 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

9/13/96

Events are unobserved.

Events are unobserved.

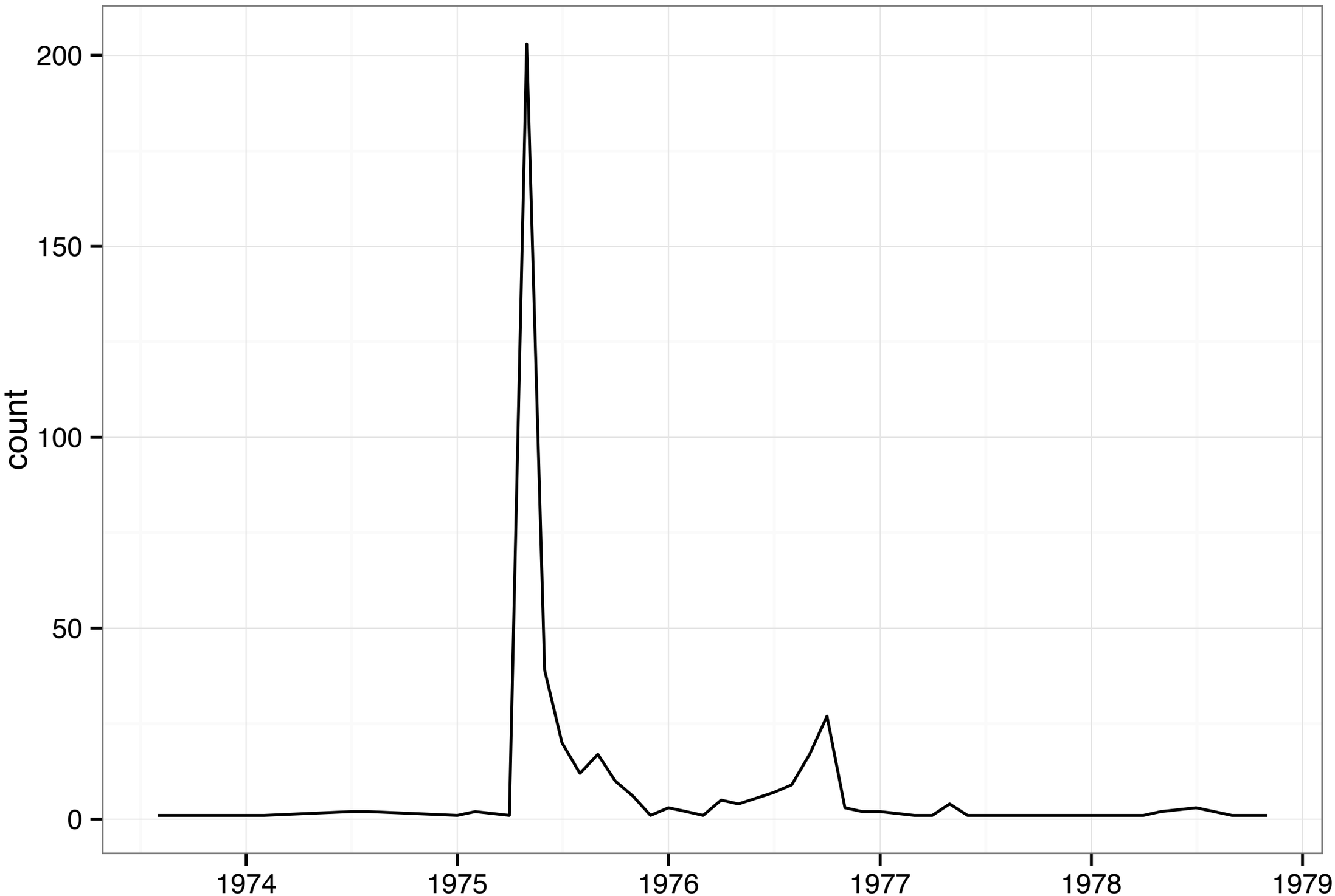
What are observed ways to characterize events?

Events are unobserved.

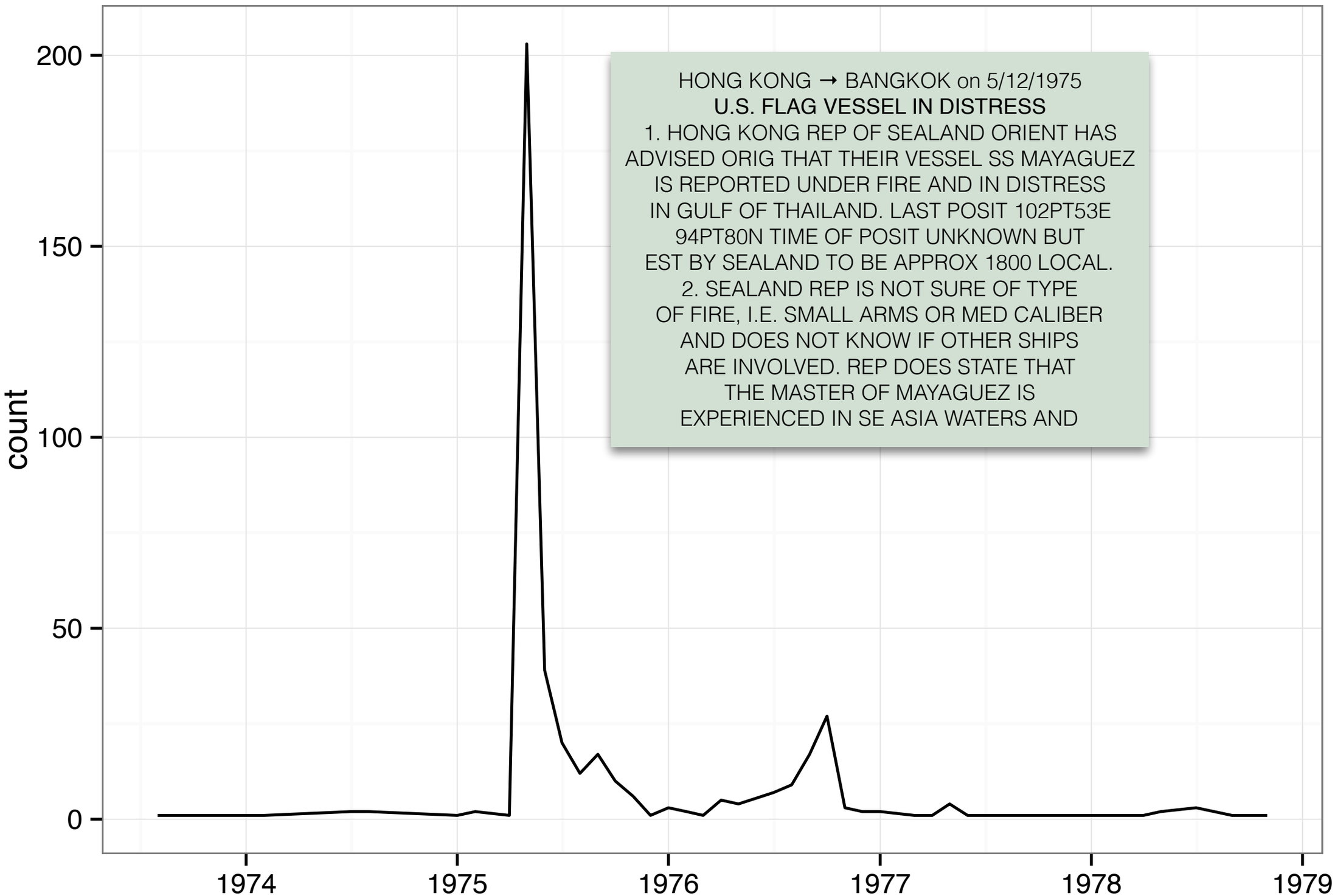
What are observed ways to characterize events?

Temporary shifts away
from business as usual
in message content

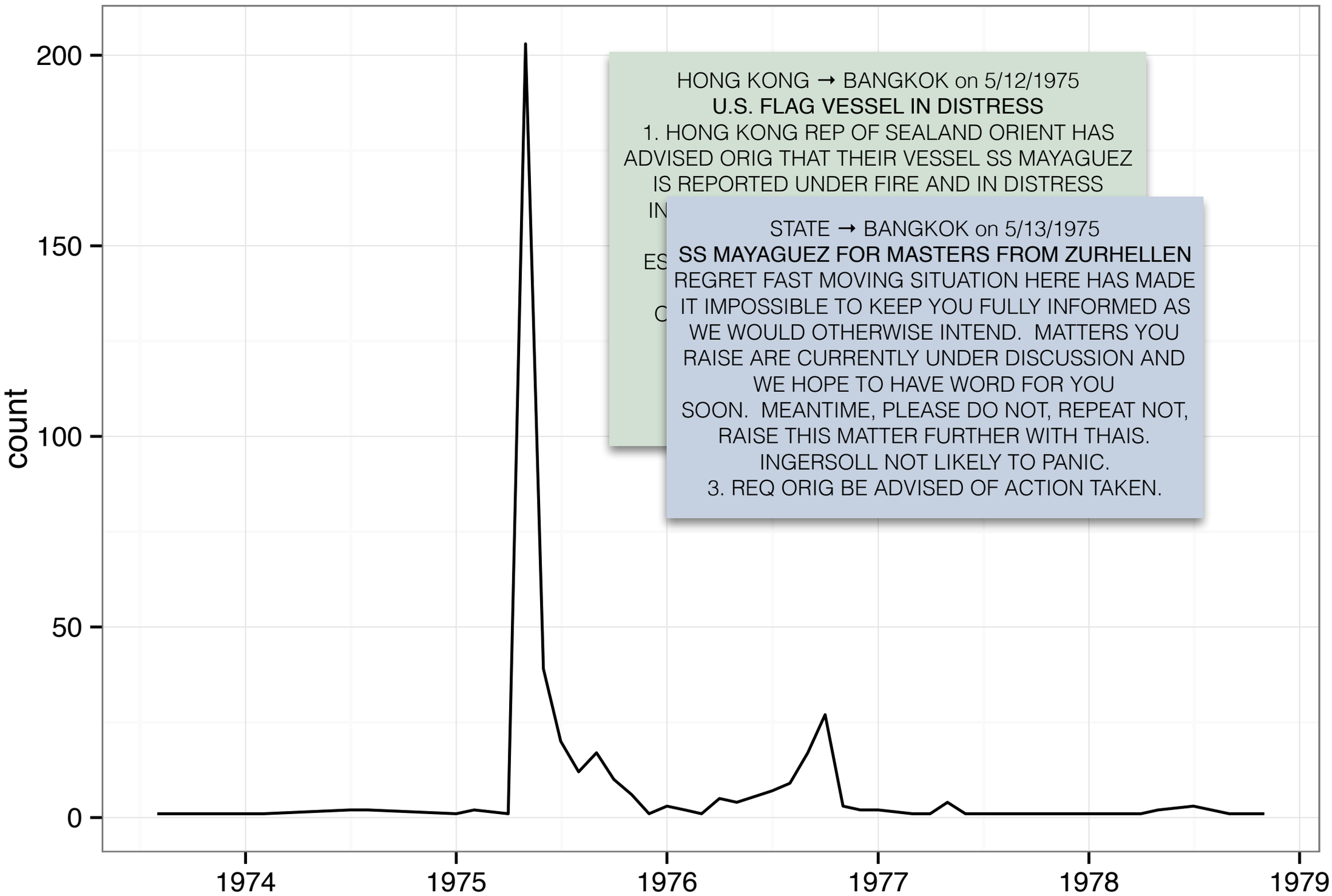
Mayaguez Incident



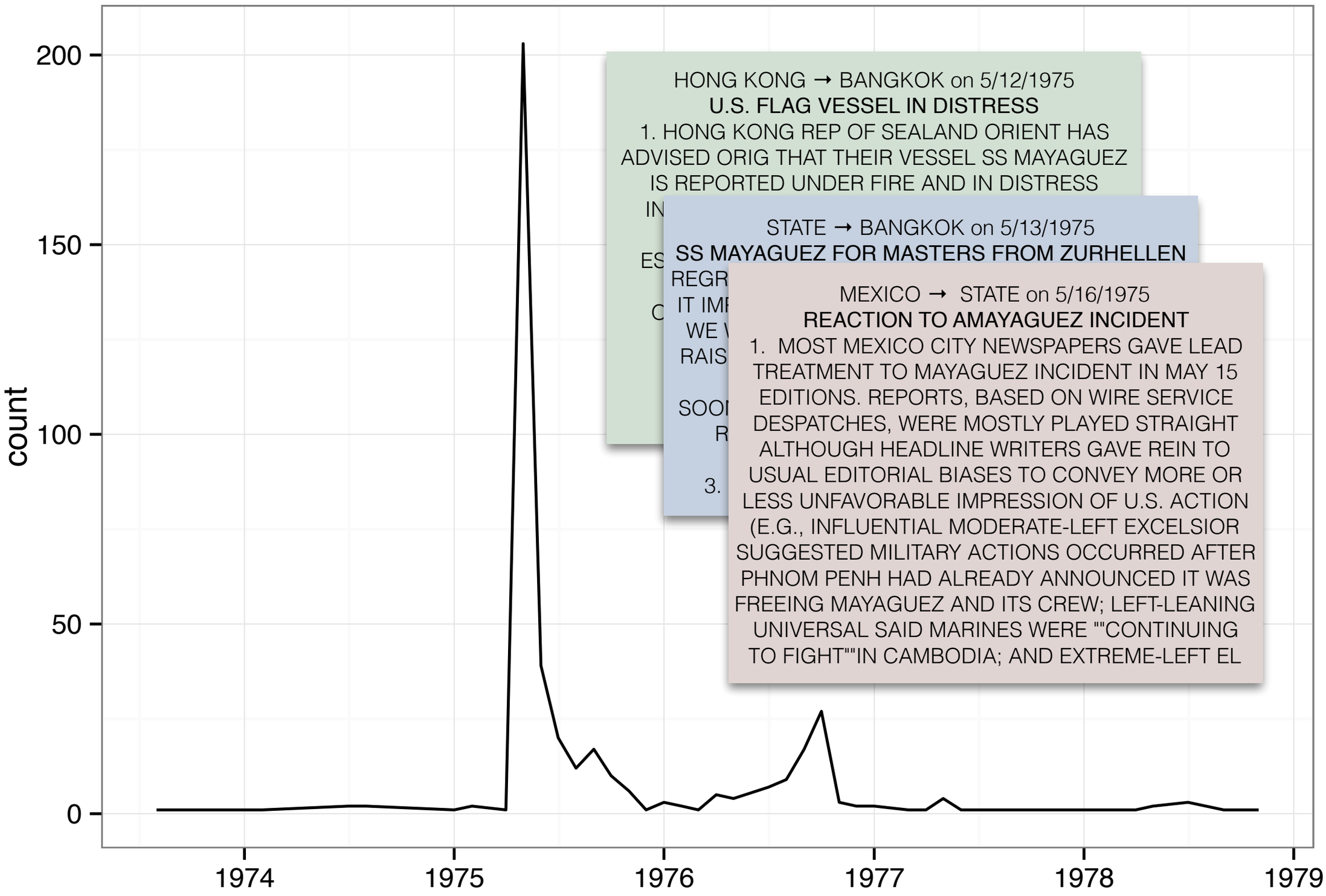
Mayaguez Incident



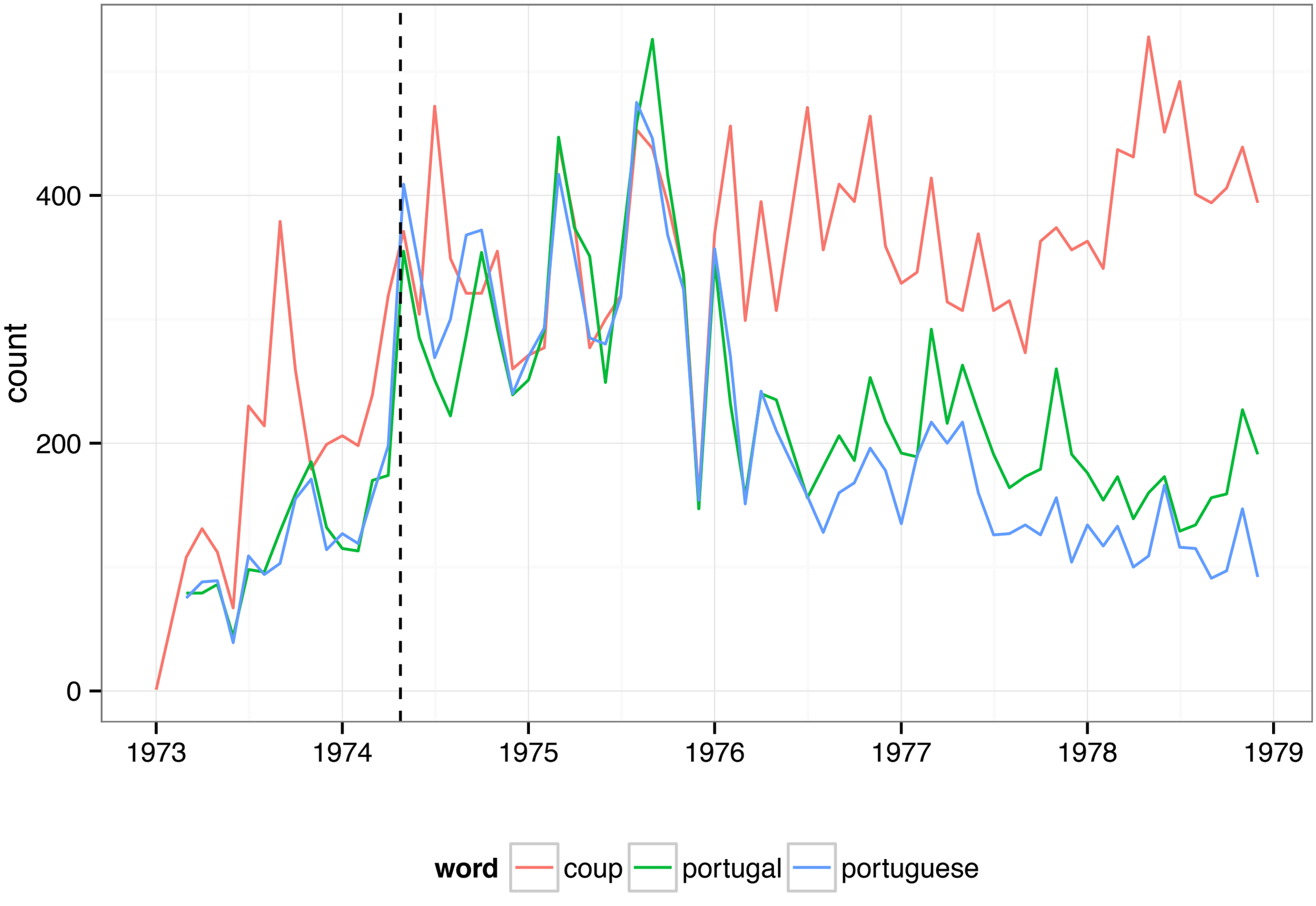
Mayaguez Incident

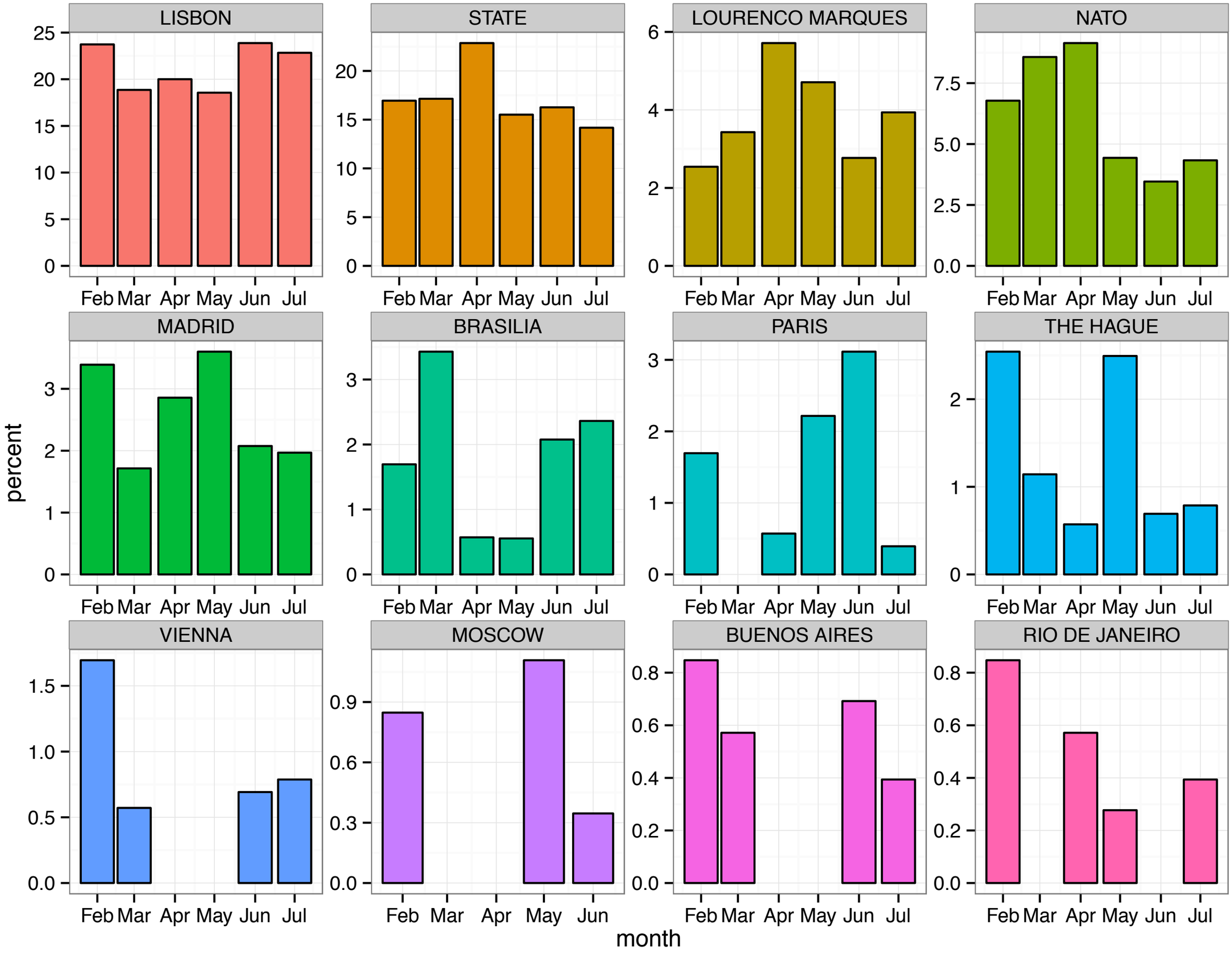


Mayaguez Incident



Carnation Revolution




LISBON
STATE
LOURENCO MARQUES
NATO
MADRID
BRASILIA
PARIS
THE HAGUE
VIENNA
MOSCOW
BUENOS AIRES
RIO DE JANEIRO
Feb Mar Apr May Jun Jul
percent
month

What are the key actors in constructing our model?

cables

HONG KONG → BANGKOK on 5/12/1975

U.S. FLAG VESSEL IN DISTRESS

1. HONG KONG REP OF SEALAND ORIENT HAS ADVISED ORIG THAT THEIR VESSEL SS MAYAGUEZ IS REPORTED UNDER FIRE AND IN DISTRESS
IN

ES

C

STATE → BANGKOK on 5/13/1975

SS MAYAGUEZ FOR MASTERS FROM ZURHELLEN

REGR
IT IMF
WE
RAIS

SOO
R

3.

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 LESS UNFAVORABLE IMPRESSION OF U.S. ACTION (E.G., INFLUENTIAL MODERATE-LEFT EXCELSIOR SUGGESTED MILITARY ACTIONS OCCURRED AFTER PHNOM PENH HAD ALREADY ANNOUNCED IT WAS FREEING MAYAGUEZ AND ITS CREW; LEFT-LEANING UNIVERSAL SAID MARINES WERE ""CONTINUING TO FIGHT""IN CAMBODIA; AND EXTREME-LEFT EL

entities

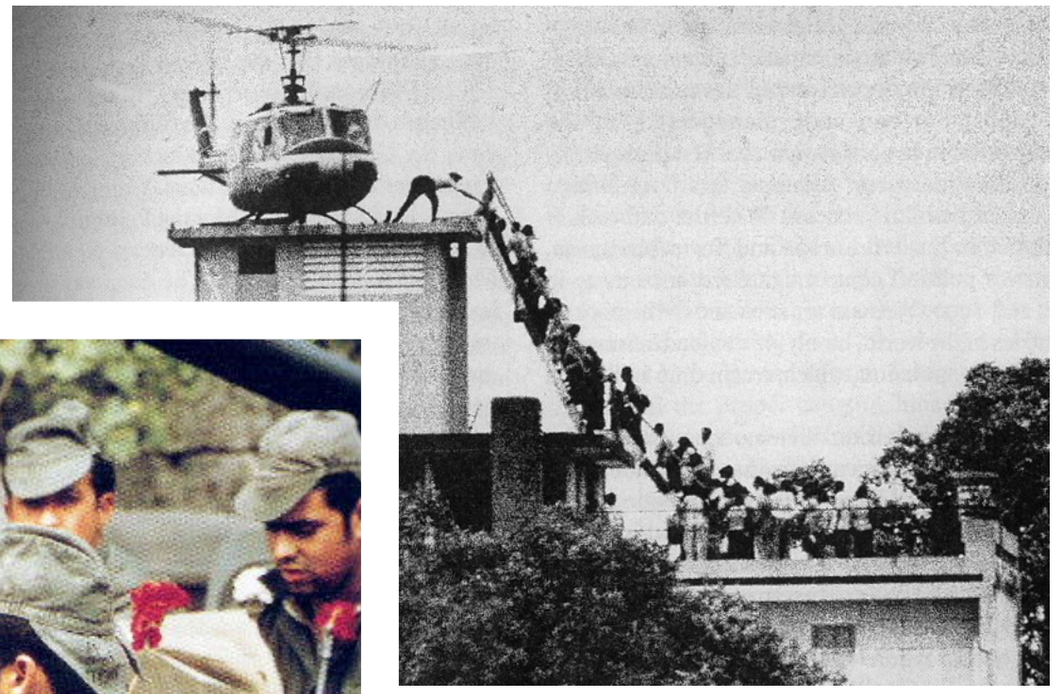
cables

HONG KONG → BANGKOK on 5/12/1975

STATE → BANGKOK on

MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT
1. MOST MEXICO CITY NEWSPAPERS GAVE LEAD TREATMENT TO

events



cables

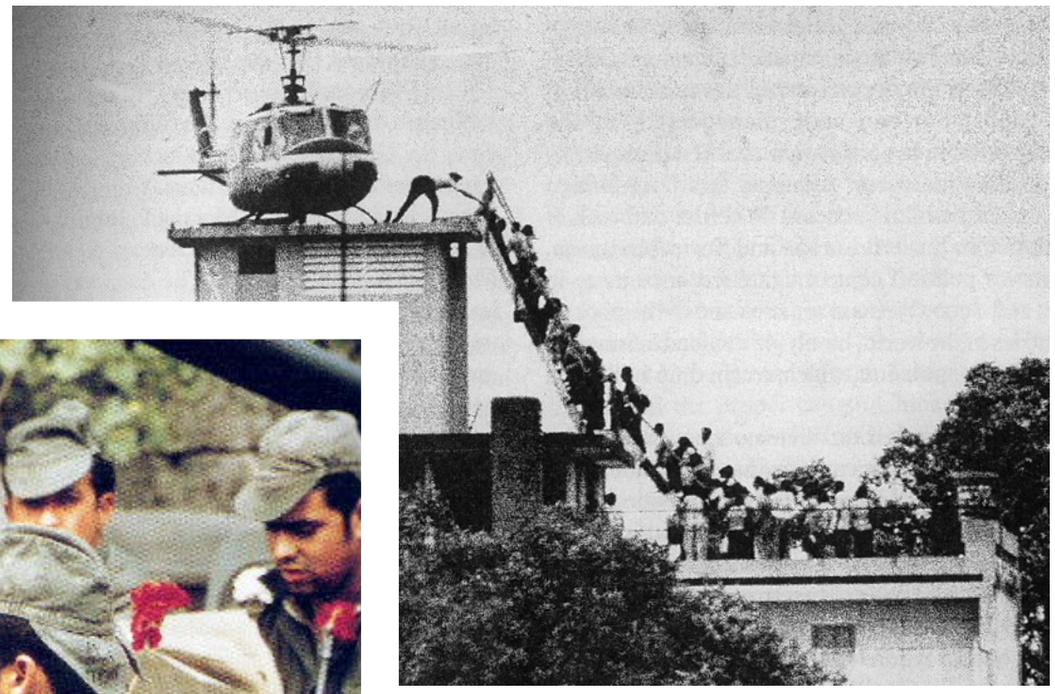
HONG KONG → BANGKOK on 5/12/1975

STATE → BANGKOK on

MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT
1. MOST MEXICO CITY NEWSPAPERS GAVE LEAD TREATMENT TO

entities

events



cables

HONG KONG → BANGKOK on 5/12/1975

STATE → BANGKOK on

MEXICO → STATE on 5/16/1975
REACTION TO AMAYAGUEZ INCIDENT
1. MOST MEXICO CITY NEWSPAPERS GAVE LEAD TREATMENT TO

entities

representing cables with topics

11/28/197: EUROPEAN MINISTERS OF EDUCATION CONFERENCE BUCHAREST SECOND AND THIRD PLENARY SESSIONS OF CONFERENCE CONTINUED NOVEMBER 26 WITH SPEECHES BY 18 ADDITIONAL DELEGATIONS AND EIGHT OBSERVER DELEGATIONS ON DEVELOPMENTS IN NATIONAL EDUCATION SINCE1967. OF THREE COUNTRIES REPRESENTED AS OBSERVERS (U.S., CANADA, ISRAEL), ONLY CANADIAN DELEGATION ADDRESSED PLENARY WITH DESCRIPTION OF MUTUAL CANADIAN-EUROPEAN EDUCATIONAL PROBLEMS AND CONCERNS. CANADIAN DELEGATE CONCLUDED WITH EXPRESSION OF CANADIAN DESIRE FOR FULL MEMBERSHIP IN EUROPEAN REGION. FOLLOWING DEPARTMENT INSTRUCTIONS, U.S. DEL DID NOT ADDRESS PLENARY ALTHOUGH DELEGATION PLANS PARTICIPATION AS APPROPRIATE IN WORKING SESSIONS BEGINNING NOVEMBER 27.

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

representing cables with topics

11/28/197: EUROPEAN MINISTERS OF EDUCATION CONFERENCE BUCHAREST SECOND AND THIRD PLENARY SESSIONS OF CONFERENCE CONTINUED NOVEMBER 26 WITH SPEECHES BY 18 ADDITIONAL DELEGATIONS AND EIGHT OBSERVER DELEGATIONS ON DEVELOPMENTS IN NATIONAL EDUCATION SINCE1967. OF THREE COUNTRIES REPRESENTED AS OBSERVERS (U.S., CANADA, ISRAEL), ONLY CANADIAN DELEGATION ADDRESSED PLENARY WITH DESCRIPTION OF MUTUAL CANADIAN-EUROPEAN EDUCATIONAL PROBLEMS AND CONCERNS. CANADIAN DELEGATE CONCLUDED WITH EXPRESSION OF CANADIAN DESIRE FOR FULL MEMBERSHIP IN EUROPEAN REGION. FOLLOWING DEPARTMENT INSTRUCTIONS, U.S. DEL DID NOT ADDRESS PLENARY ALTHOUGH DELEGATION PLANS PARTICIPATION AS APPROPRIATE IN WORKING SESSIONS BEGINNING NOVEMBER 27.

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

Advantage: Good for discovering general, interpretable themes

useful for representing *entities' typical concerns*

Disadvantage: Does not capture word-level shifts in language or subject

not useful for representing *event descriptions*

representing cables with words

11/28/197: EUROPEAN MINISTERS OF EDUCATION CONFERENCE BUCHAREST
SECOND AND THIRD PLENARY SESSIONS OF CONFERENCE CONTINUED NOVEMBER 26 WITH SPEECHES BY 18 ADDITIONAL DELEGATIONS AND EIGHT OBSERVER DELEGATIONS ON DEVELOPMENTS IN NATIONAL EDUCATION SINCE1967. OF THREE COUNTRIES REPRESENTED AS OBSERVERS (U.S., CANADA, ISRAEL), ONLY CANADIAN DELEGATION ADDRESSED PLENARY WITH DESCRIPTION OF MUTUAL CANADIAN-EUROPEAN EDUCATIONAL PROBLEMS AND CONCERNS. CANADIAN DELEGATE CONCLUDED WITH EXPRESSION OF CANADIAN DESIRE FOR FULL MEMBERSHIP IN EUROPEAN REGION. FOLLOWING DEPARTMENT INSTRUCTIONS, U.S. DEL DID NOT ADDRESS PLENARY ALTHOUGH DELEGATION PLANS PARTICIPATION AS APPROPRIATE IN WORKING SESSIONS BEGINNING NOVEMBER 27.

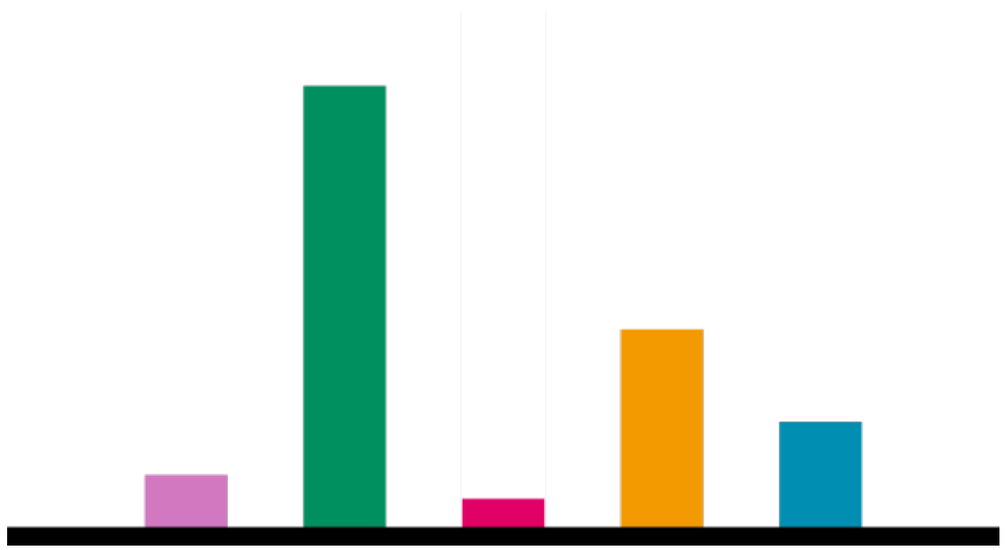
Advantage: Good for capturing shifts in specific subjects and terminology

useful for representing *event descriptions*

Disadvantage: harder to interpret, may cause scalability issues

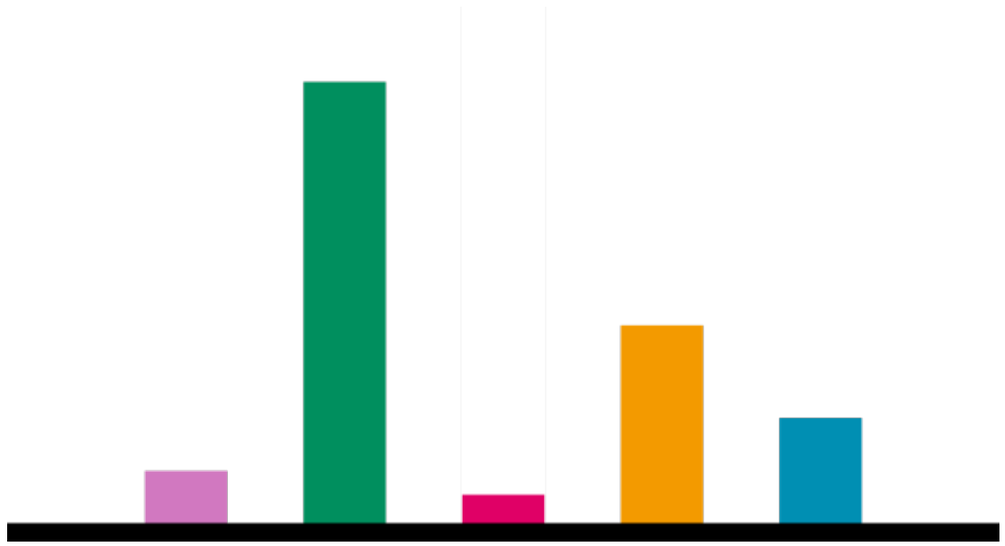
not useful for representing *entities' typical concerns*

modeling entities



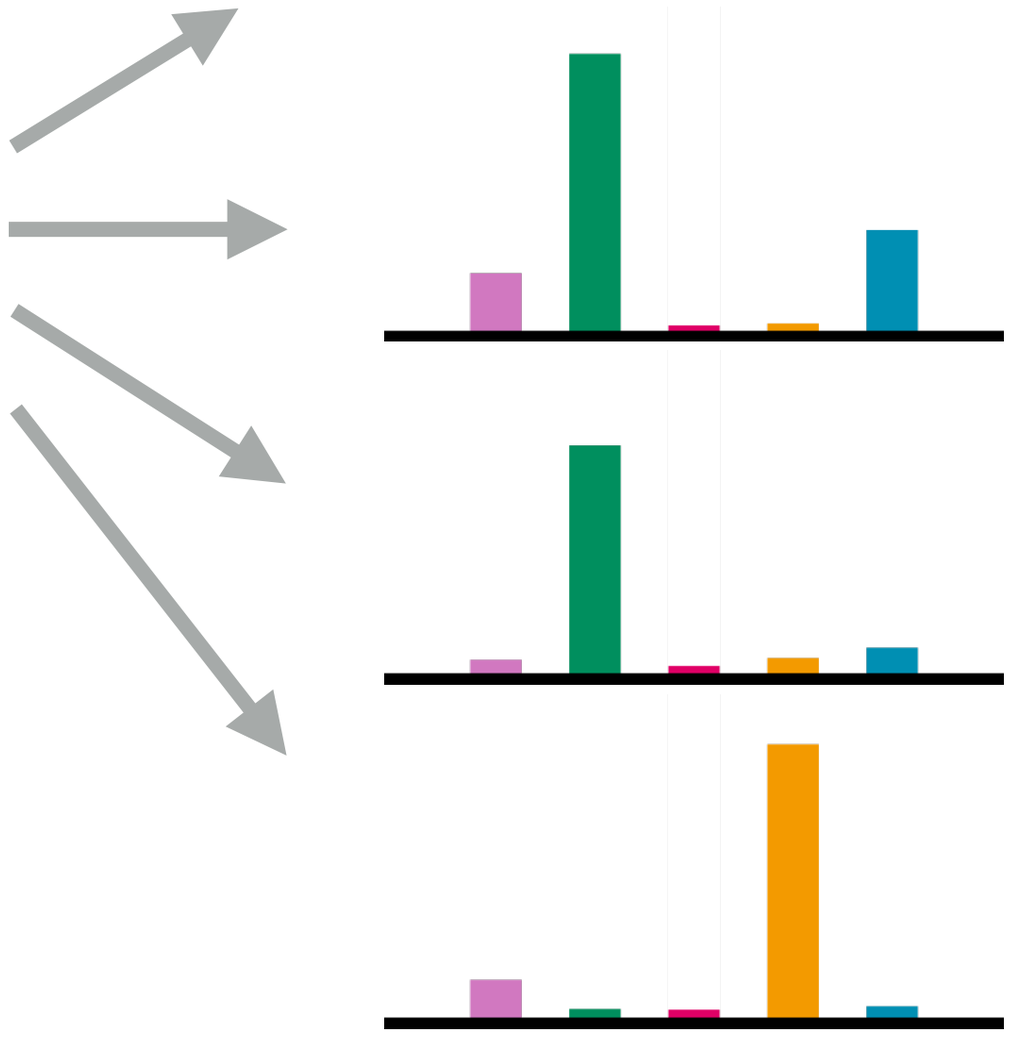
typical concerns
of the Bangkok Embassy

modeling entities



typical concerns
of the Bangkok Embassy

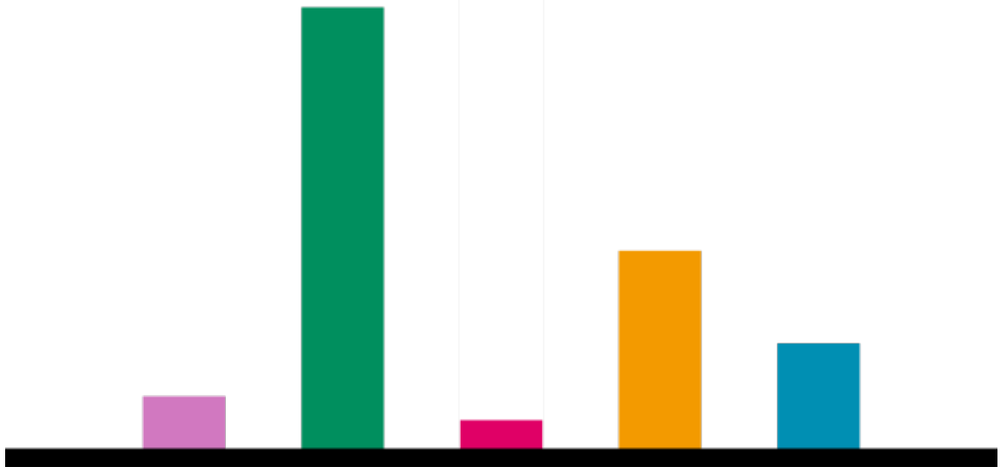
cables sent



modeling entities

for each entity *i* and topic *k*, draw typical concerns:

$$\phi_{i,k} \sim \mathrm{Gamma}(\alpha_\phi, \beta_\phi)$$



typical concerns
of entity *i*

entity-only model of cables

for each entity *i* and topic *k*, draw typical concerns:

$$\phi_{i,k} \sim \mathrm{Gamma}(\alpha_\phi, \beta_\phi)$$

entity-only model of cables

for each entity *i* and topic *k*, draw typical concerns:

$$\phi_{i,k} \sim \mathrm{Gamma}(\alpha_\phi, \beta_\phi)$$

for each topic *k* and vocabulary term *v*, draw topics:

$$\theta_{k,v} \sim \mathrm{Gamma}(\alpha_\theta, \beta_\theta)$$

entity-only model of cables

for each entity *i* and topic *k*, draw typical concerns:

$$\phi_{i,k} \sim \mathrm{Gamma}(\alpha_\phi, \beta_\phi)$$

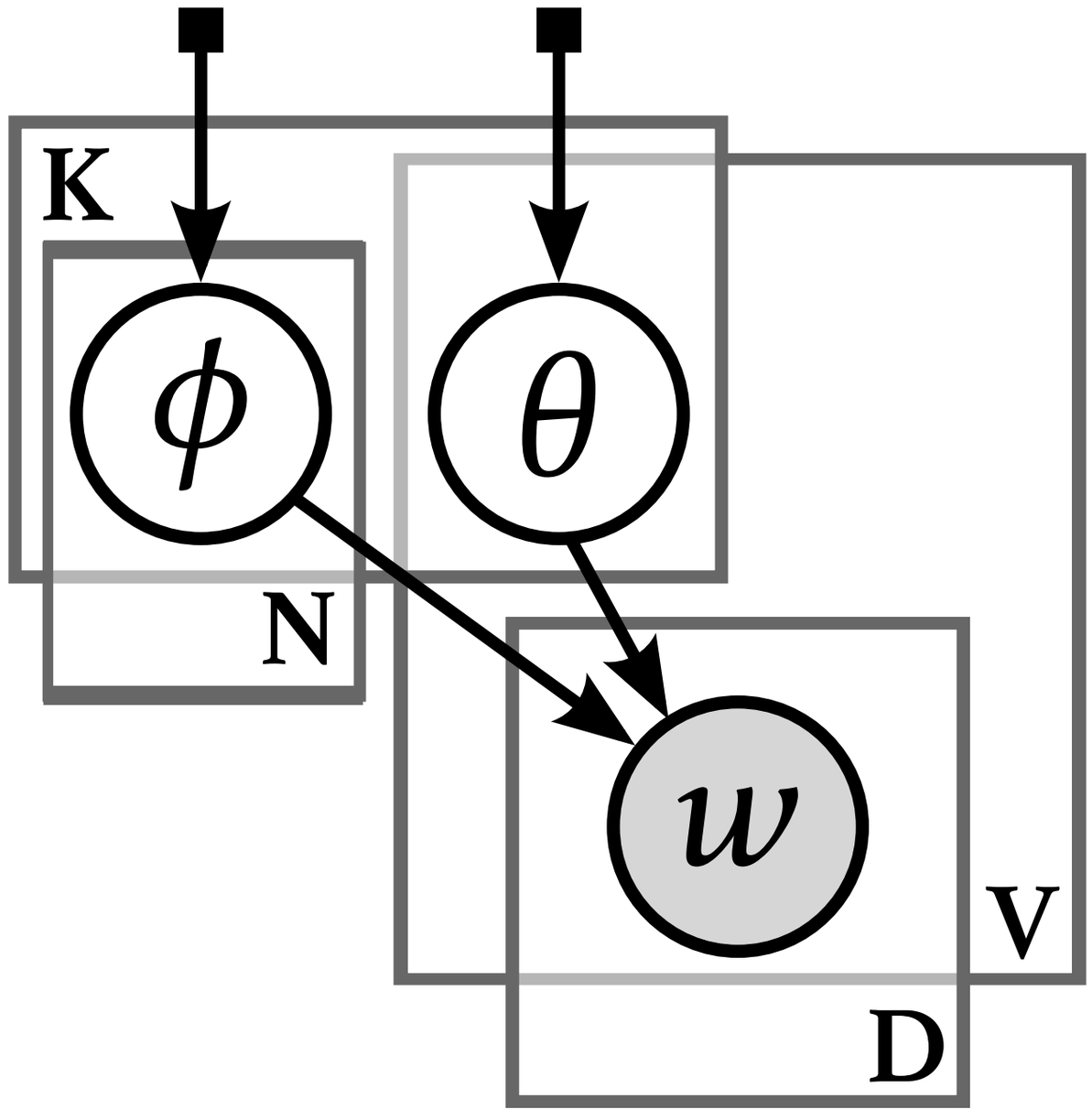
for each topic *k* and vocabulary term *v*, draw topics:

$$\theta_{k,v} \sim \mathrm{Gamma}(\alpha_\theta, \beta_\theta)$$

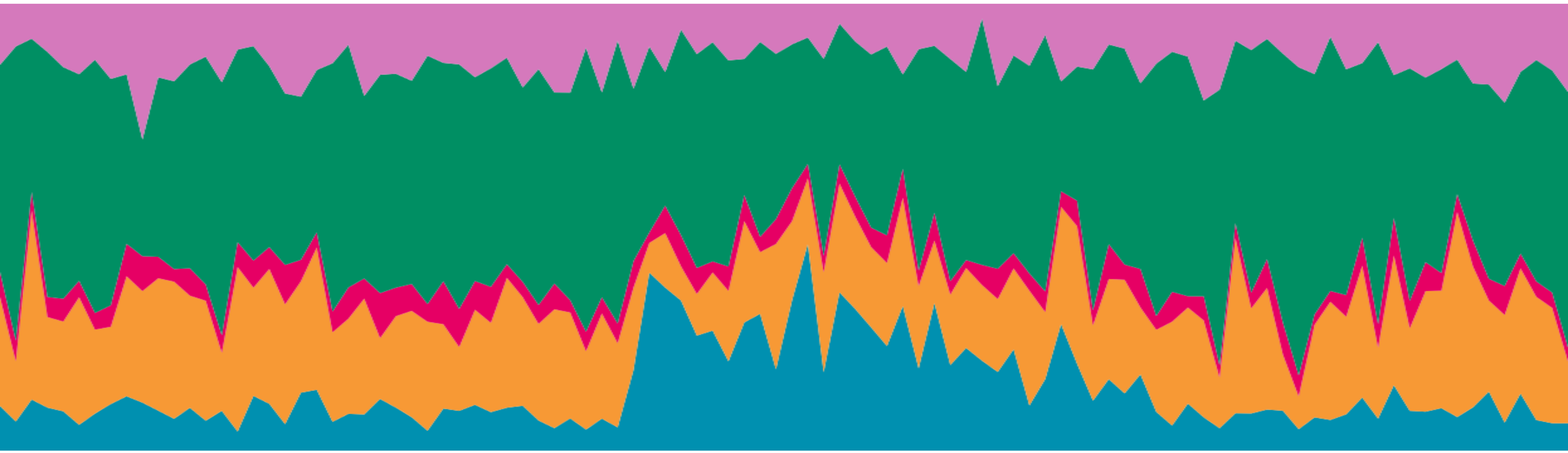
for each cable *n* (sent by entity *i*) and vocabulary term v:

$$w_{n,v} \sim \mathrm{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v}\right)$$

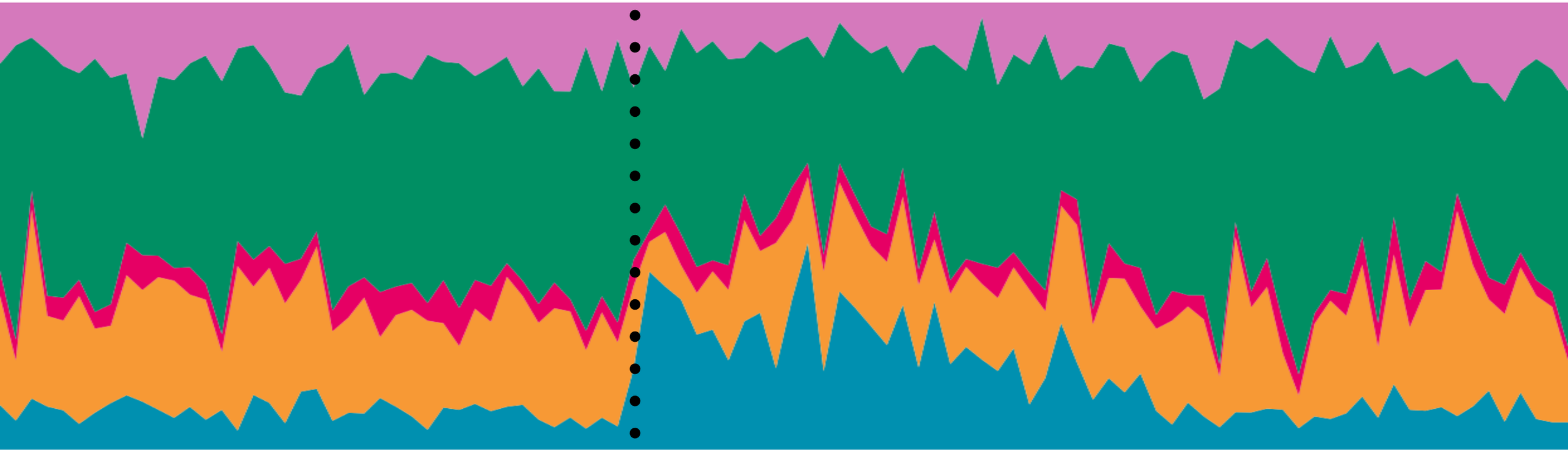
entity-only model of cables



modeling events

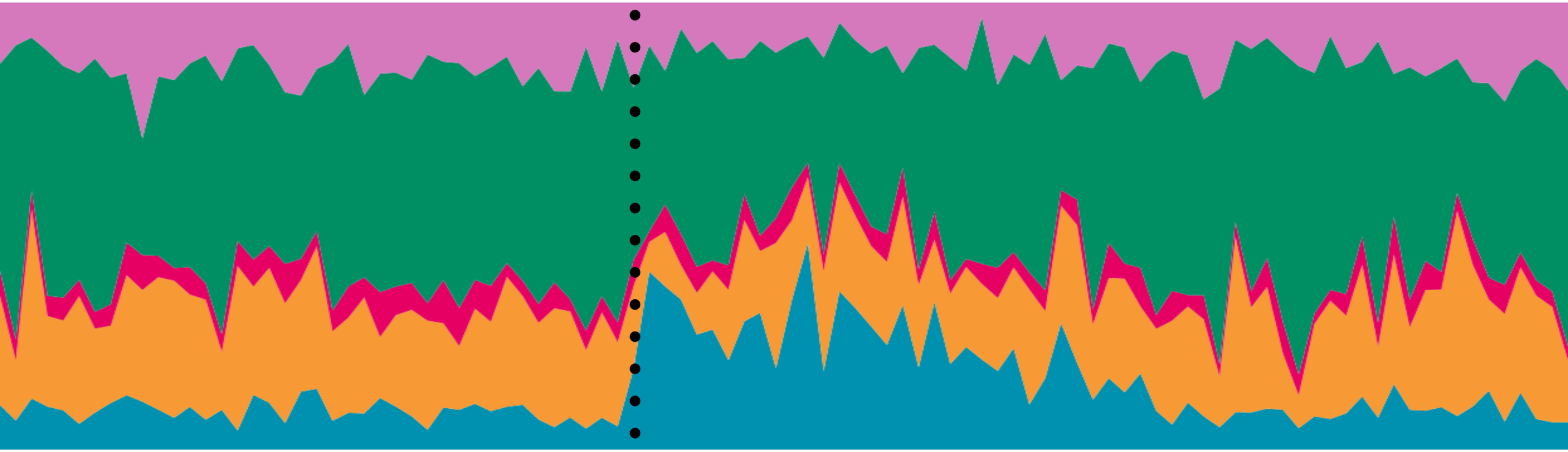


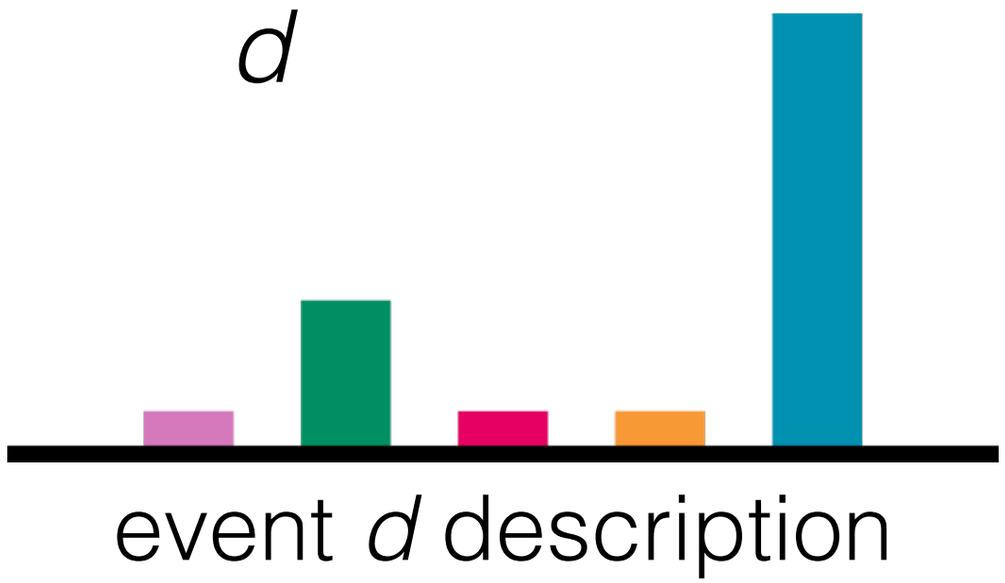
modeling events



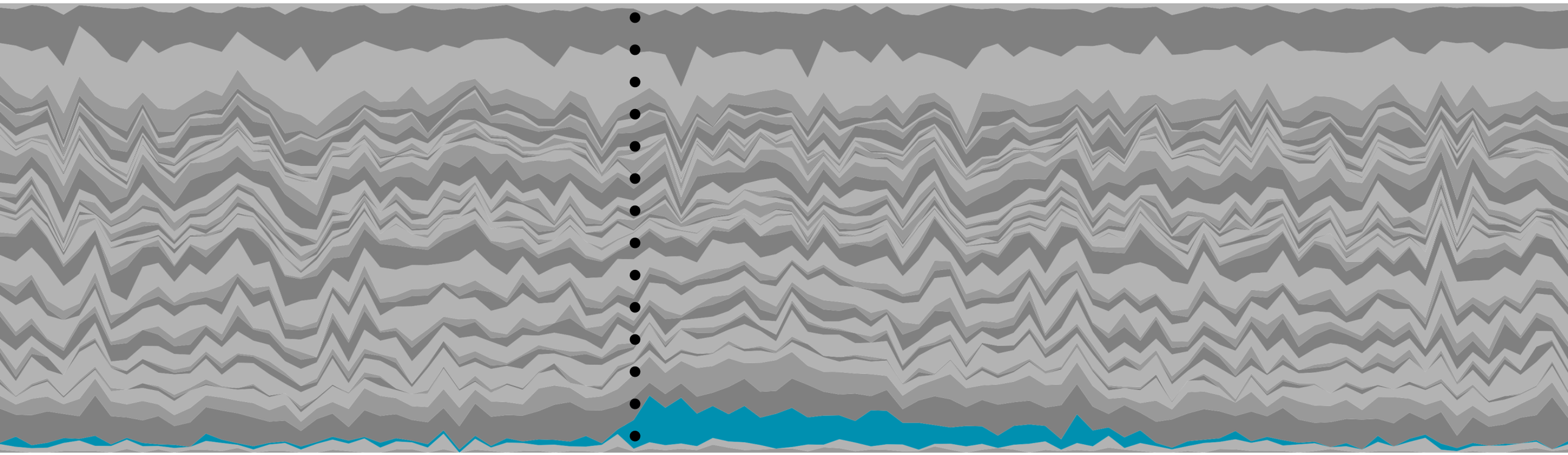
d

modeling events

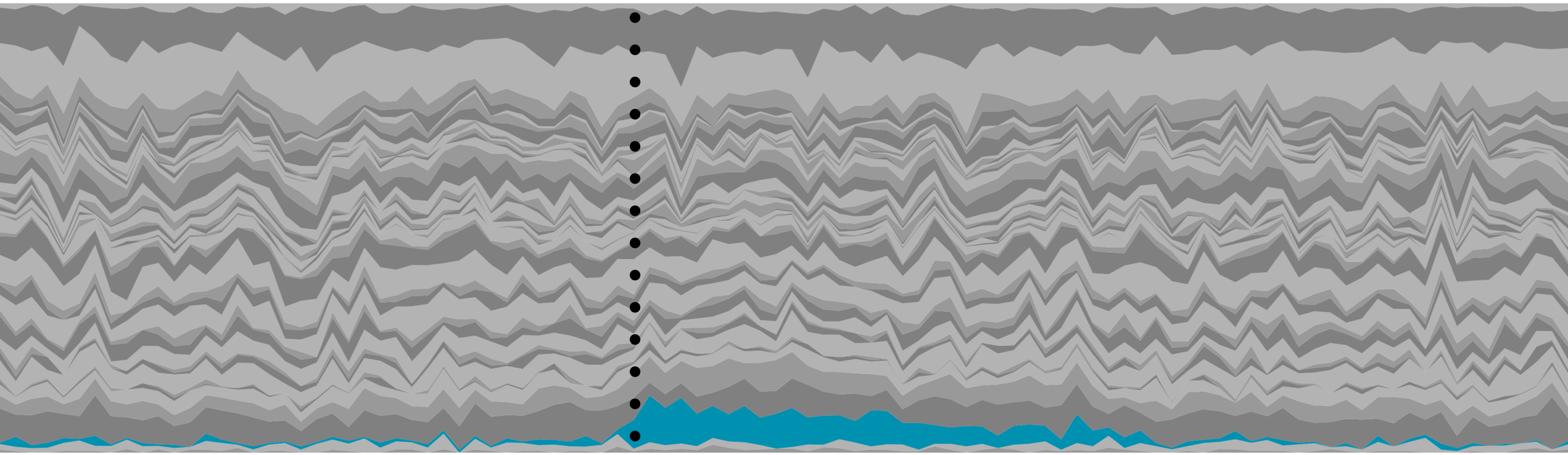


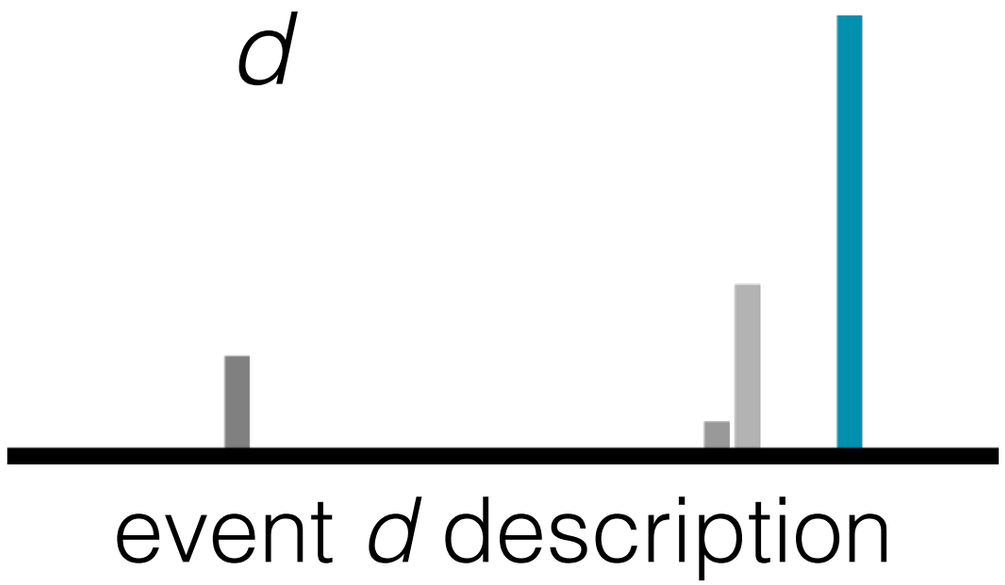


modeling events



modeling events





event-only model of cables

for each time *t*, draw event strength:

$$\epsilon_t \sim \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon)$$

event-only model of cables

for each time *t*, draw event strength:

$$\epsilon_t \sim \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon)$$

for each time *t* and vocab term *v*, draw event description:

$$\pi_{t,v} \sim \mathrm{Gamma}(\alpha_\pi, \beta_\pi)$$

event-only model of cables

for each time *t*, draw event strength:

$$\epsilon_t \sim \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon)$$

for each time *t* and vocab term *v*, draw event description:

$$\pi_{t,v} \sim \mathrm{Gamma}(\alpha_\pi, \beta_\pi)$$

for each cable *n* (sent at time *d*) and vocabulary term v:

$$w_{n,v} \sim \mathrm{Poisson}\left(\sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

event-only model of cables

for each time *t*, draw event strength:

$$\epsilon_t \sim \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon)$$

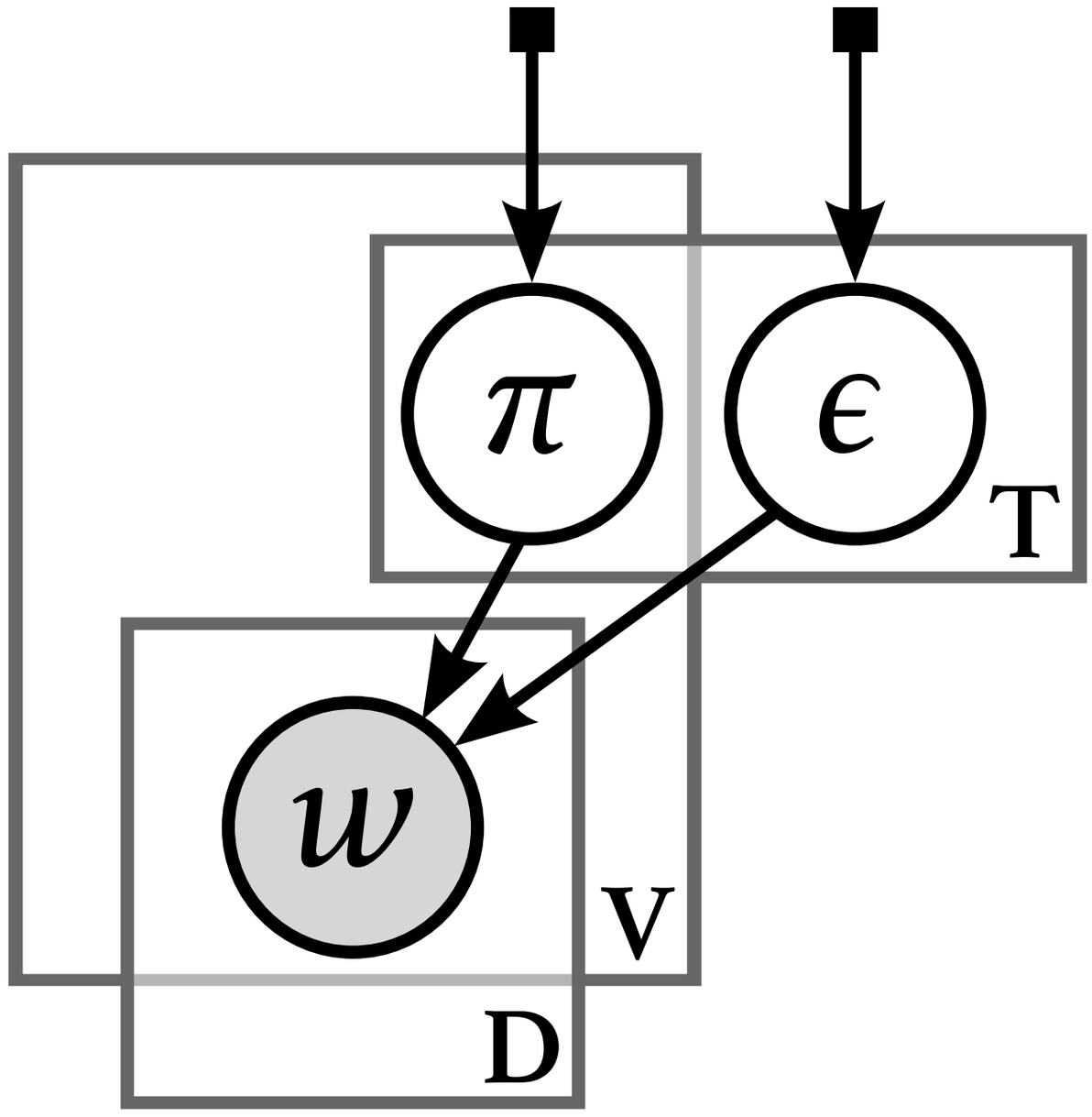
for each time *t* and vocab term *v*, draw event description:

$$\pi_{t,v} \sim \mathrm{Gamma}(\alpha_\pi, \beta_\pi)$$

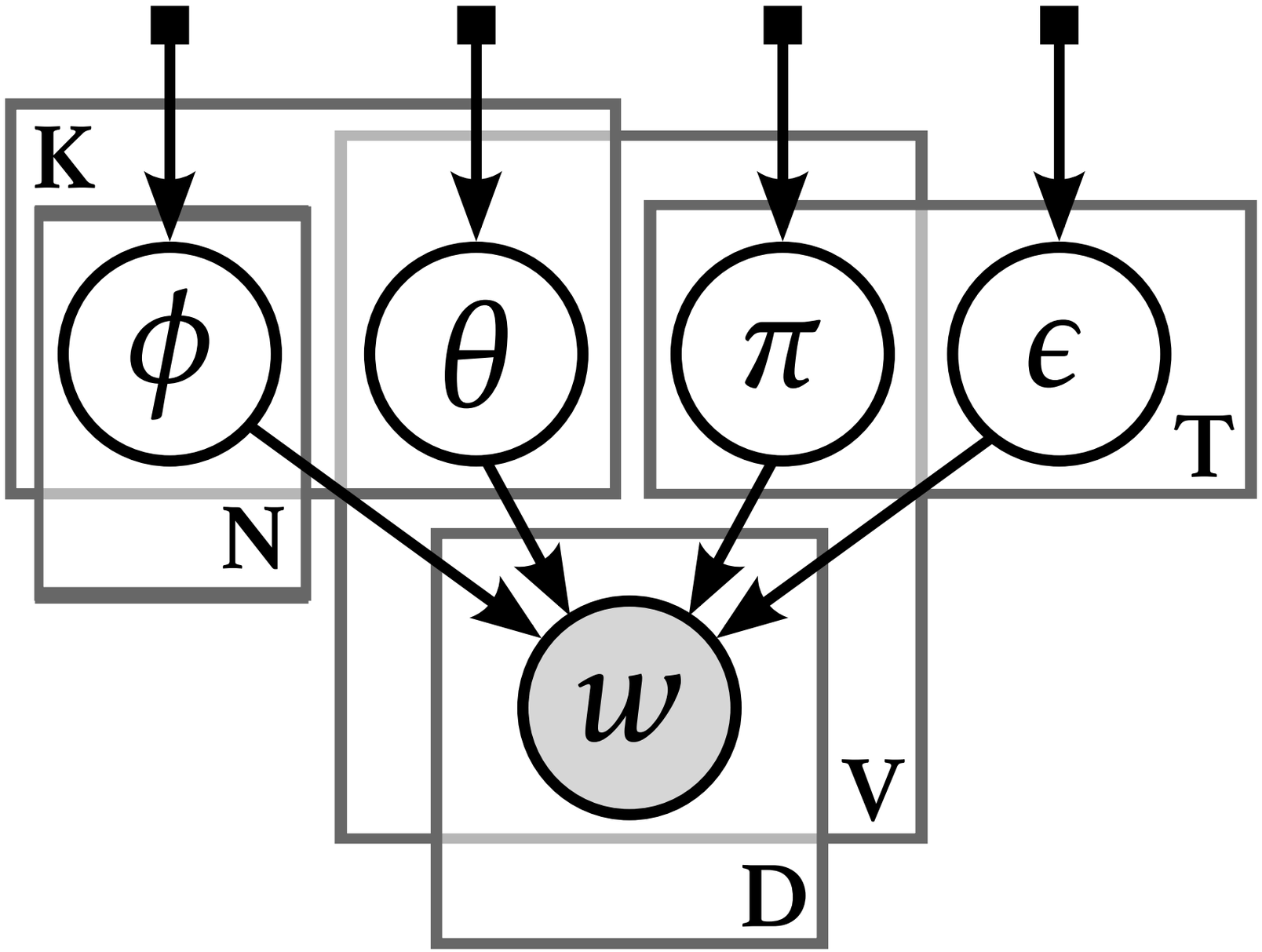
for each cable *n* (sent at time *d*) and vocabulary term v:

$$w_{n,v} \sim \mathrm{Poisson}\left(\sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

event-only model of cables



full model of cables



full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

sum over topics

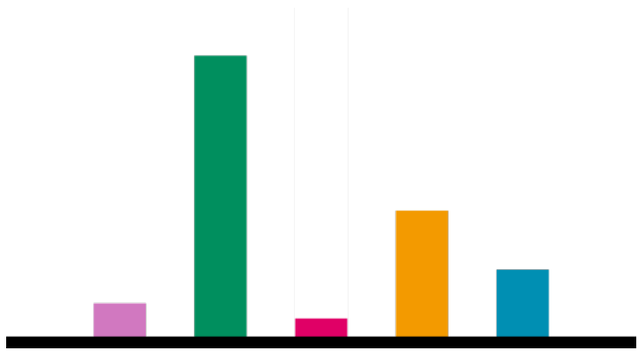
$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

sum over topics

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$



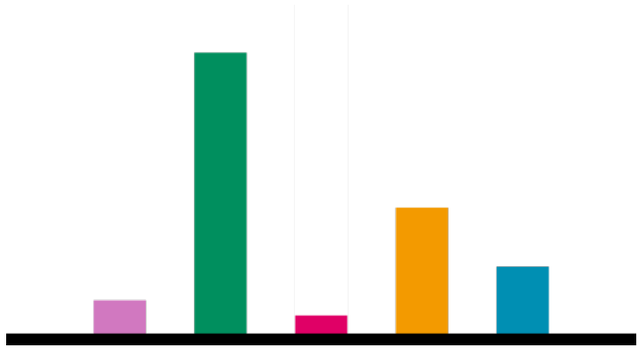
typical concerns of entity *i*

full model of cables

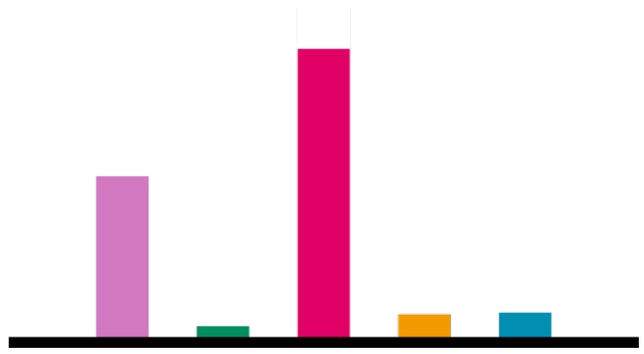
for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

sum over topics

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$



typical concerns of entity *i*



topics for word *v*

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

1973 1978

sum over all time steps

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

1973 1978

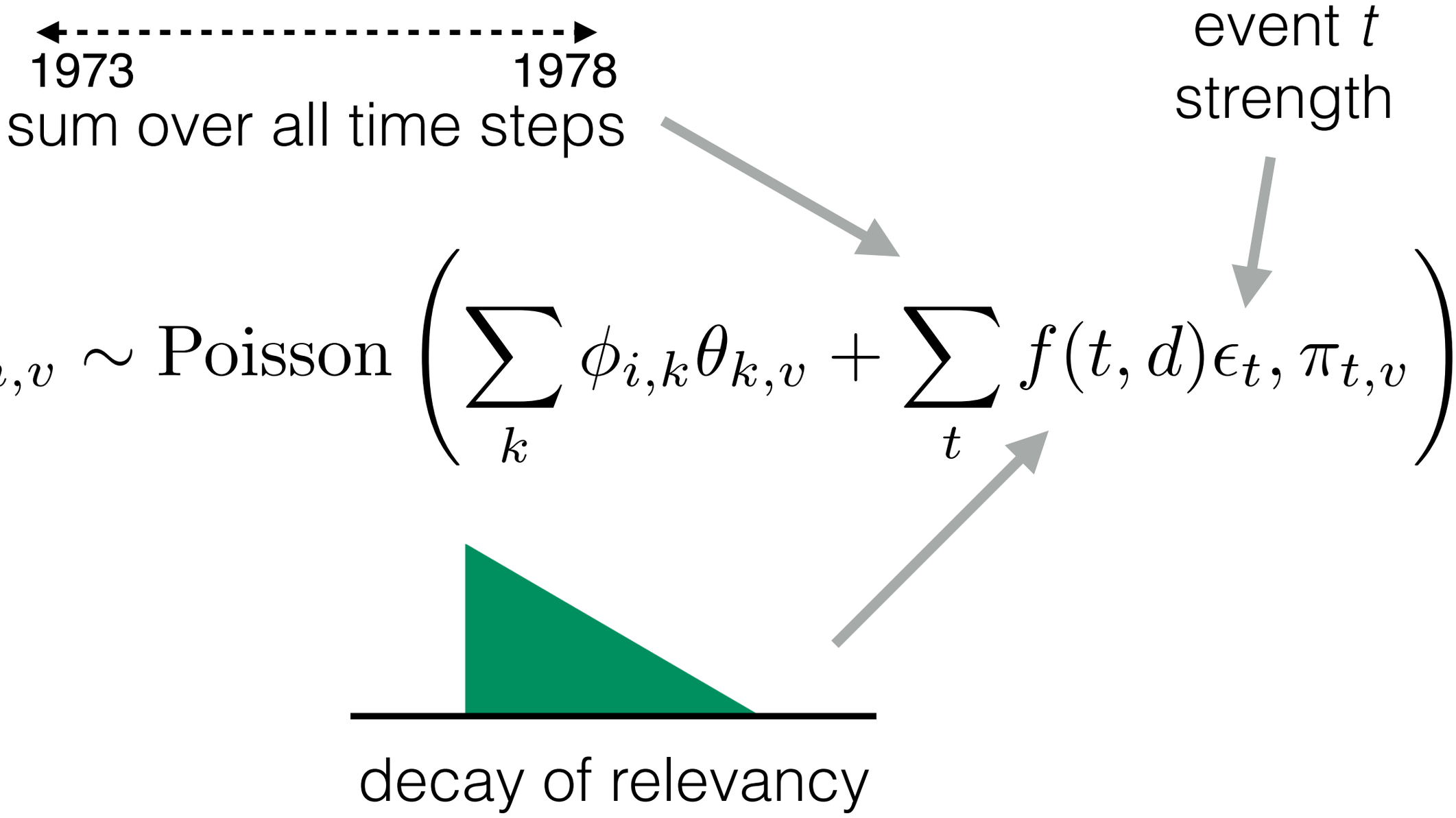
sum over all time steps

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

decay of relevancy

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:



full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

1973 — 1978

sum over all time steps

event *t* strength

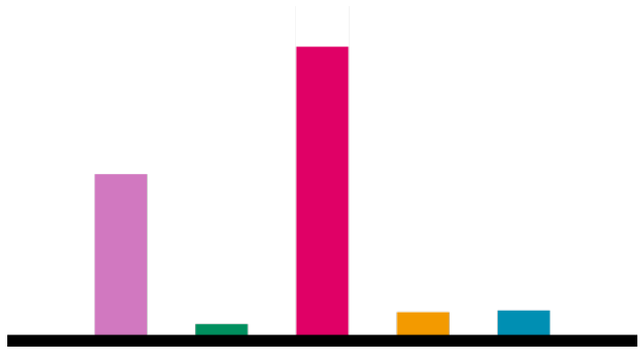
$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

decay of relevancy

event descriptions in vocabulary space

full model of cables

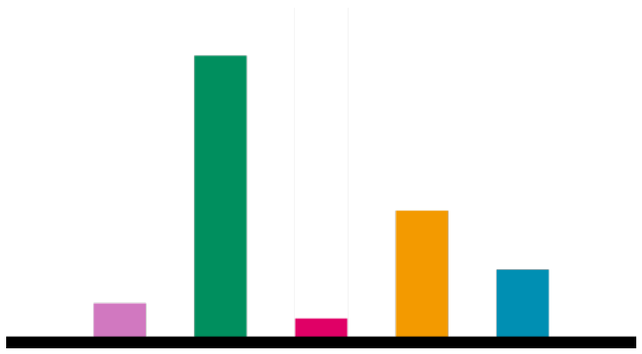
for each cable *n* (sent by entity *i* at time *d*) and vocab term v:



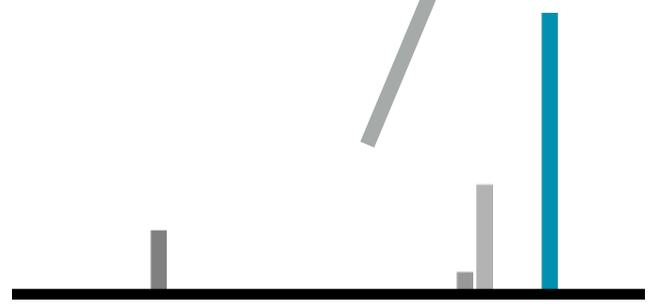
topics

event strengths

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$



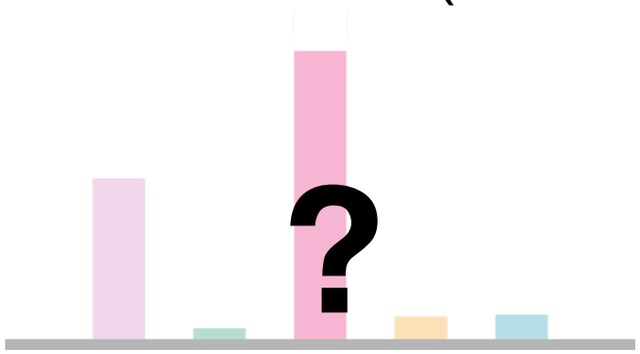
typical concerns



event descriptions

full model of cables

for each cable *n* (sent by entity *i* at time *d*) and vocab term v:

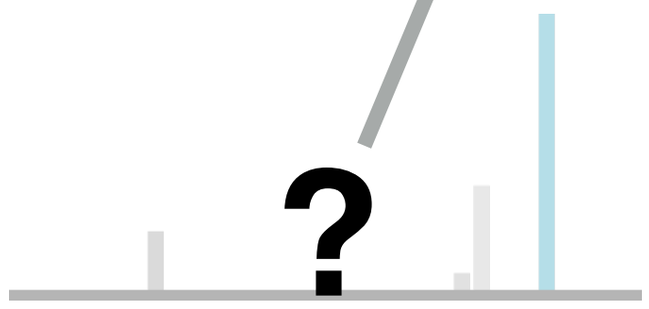


topics

event strengths

$$w_{n,v} \sim \text{Poisson}\left(\sum_k \phi_{i,k}\theta_{k,v} + \sum_t f(t,d)\epsilon_t, \pi_{t,v}\right)$$

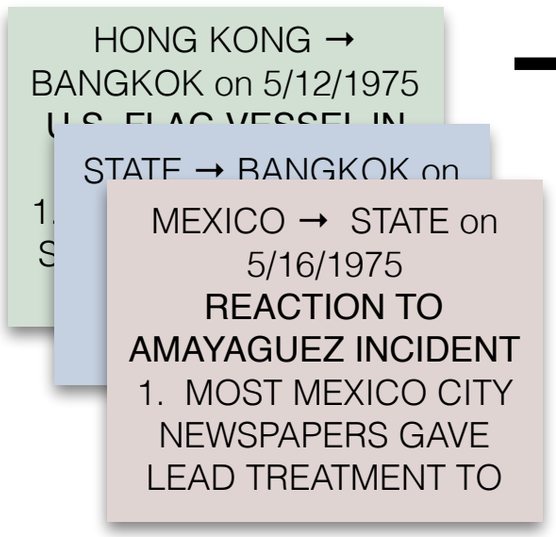
typical concerns



event descriptions

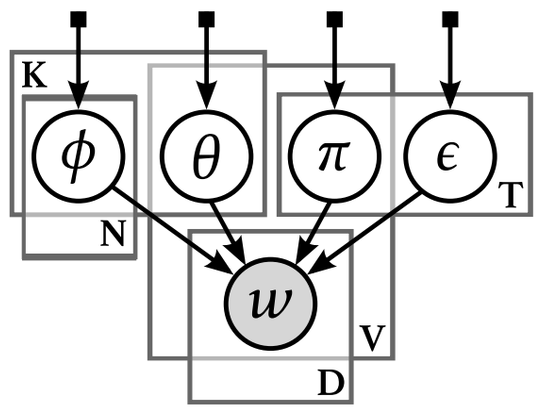
How do we find the values of the hidden parameters that best fit the data?

observed data



inference algorithm

model assumptions



learned parameters

entity typical concerns

topics

event strengths

event descriptions

visualization & exploration

Posterior Distribution

latent model parameters

easy to compute

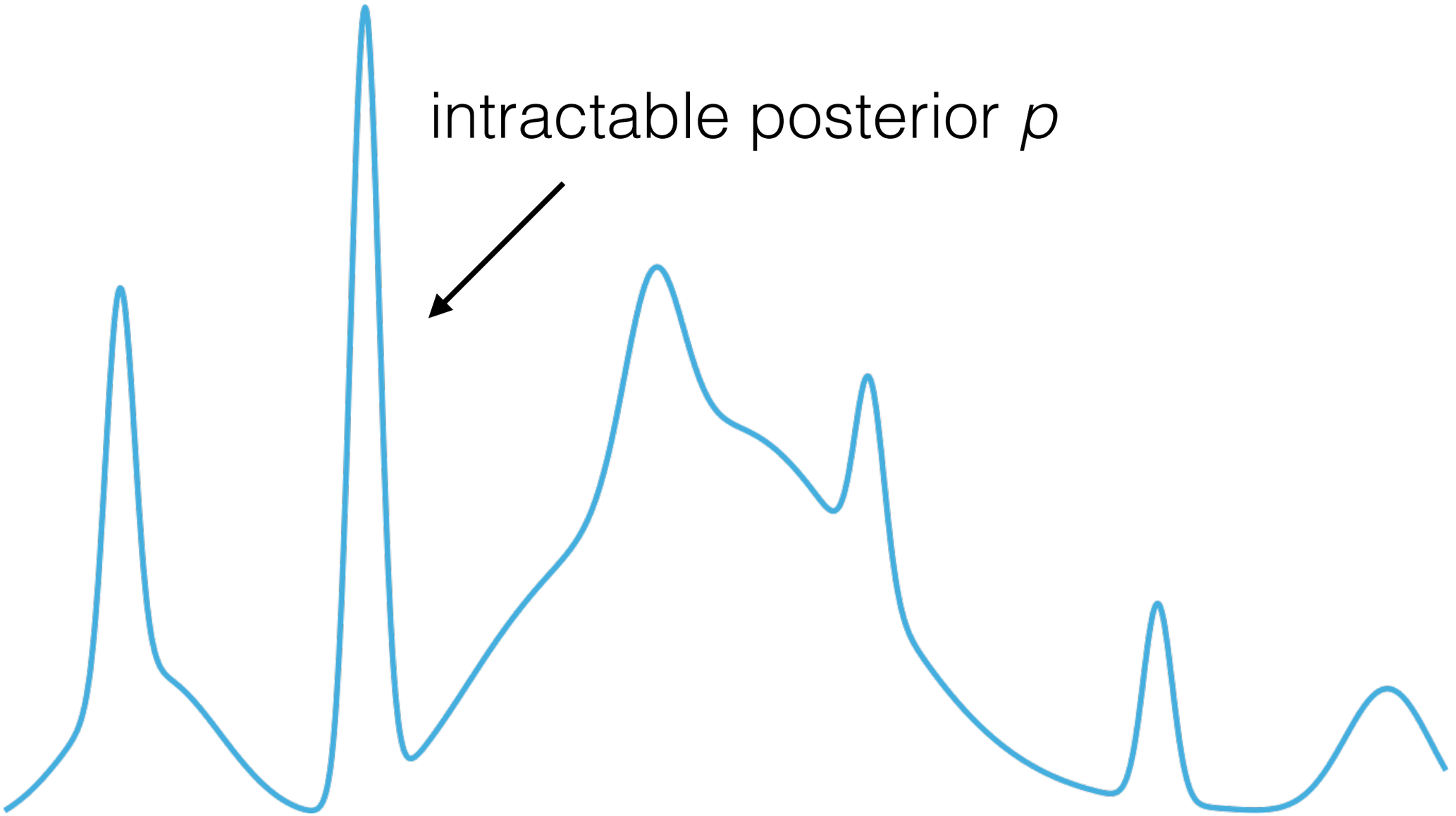
$$p(\phi, \theta, \epsilon, \pi \mid w, \alpha, \beta) = \frac{p(\phi, \theta, \epsilon, \pi, w \mid \alpha, \beta)}{\int_\phi \int_\theta \int_\epsilon \int_\pi p(\phi, \theta, \epsilon, \pi, w \mid \alpha, \beta)}$$

observed data

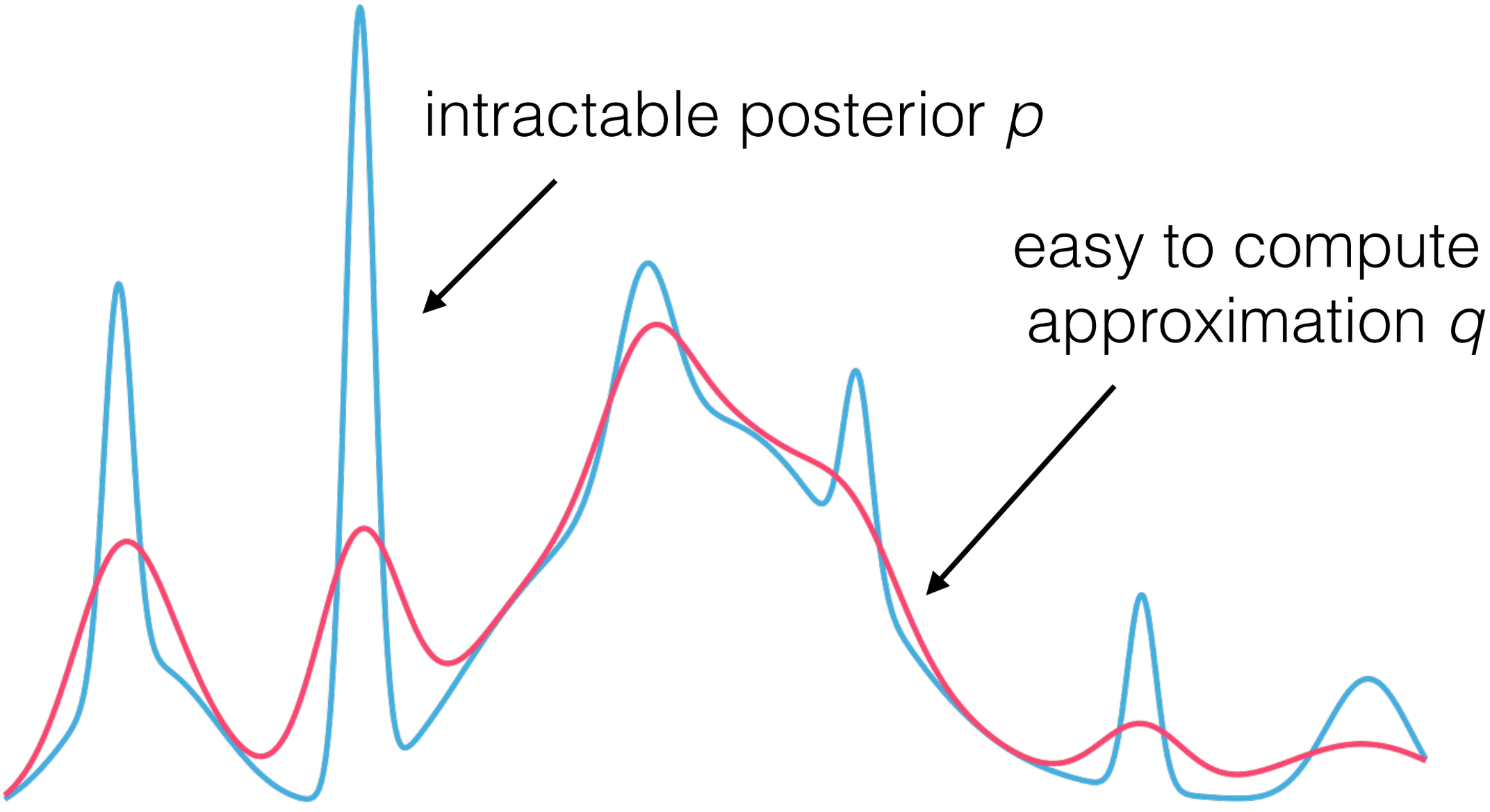
model hyperparameters

intractable

Variational Inference



Variational Inference



Variational Inference

Black box variational inference. Ranganath, Gerrish, and Blei, 2014.

- reduces model-specific mathematical derivations
- uses stochastic optimization
    - noisy gradient is computed from Monte Carlo samples from the variational distribution
    - downsides: looser fit, slower to converge

Standard conjugate variational inference

- model-specific mathematical derivations of closed-form updates
- uses coordinate ascent: iteratively updates each parameter while holding the others fixed

quick model comparison

current model

previous model

quick model comparison

current model	previous model
word counts are observations	fit LDA then hold document topics fixed as observations

quick model comparison

current model	previous model
word counts are observations	fit LDA then hold document topics fixed as observations
gamma-distributed event strengths	Poisson or Bernoulli-distributed event occurrences

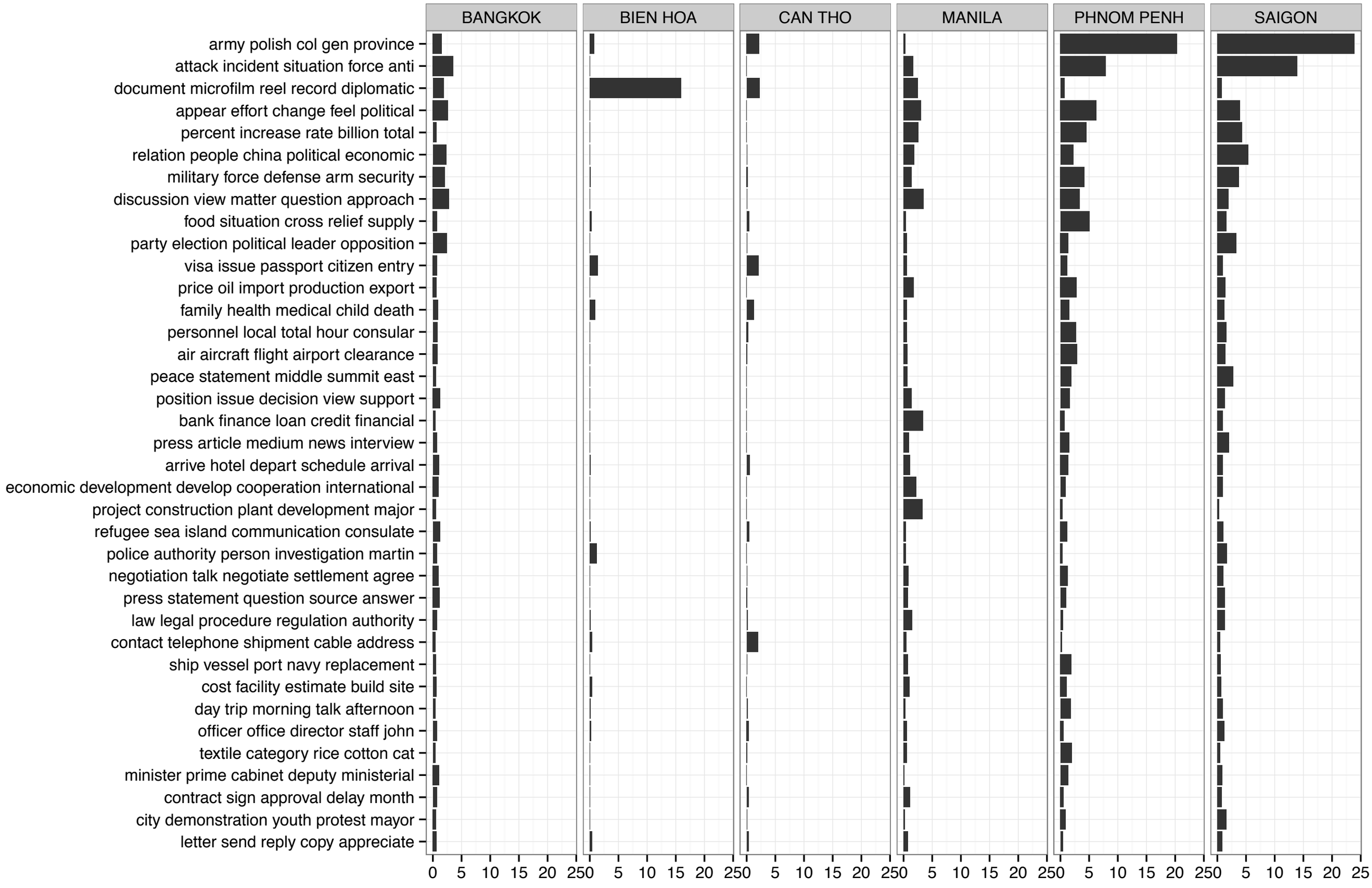
quick model comparison

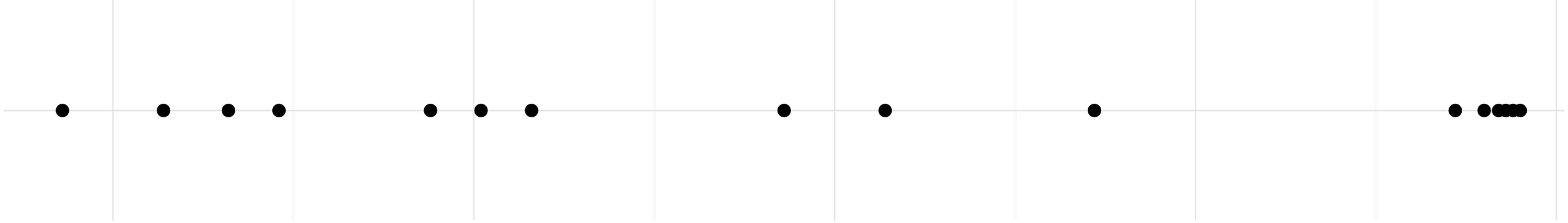
current model	previous model
word counts are observations	fit LDA then hold document topics fixed as observations
gamma-distributed event strengths	Poisson or Bernoulli-distributed event occurrences
event descriptions in vocabulary space	event descriptions in topic space

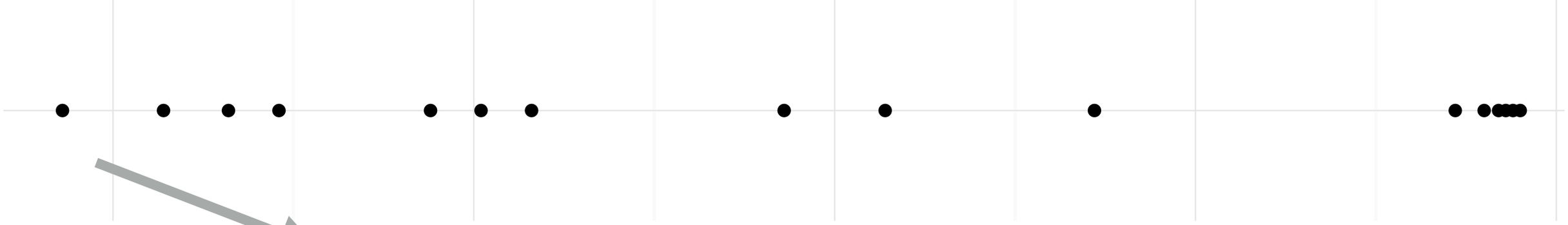
quick model comparison

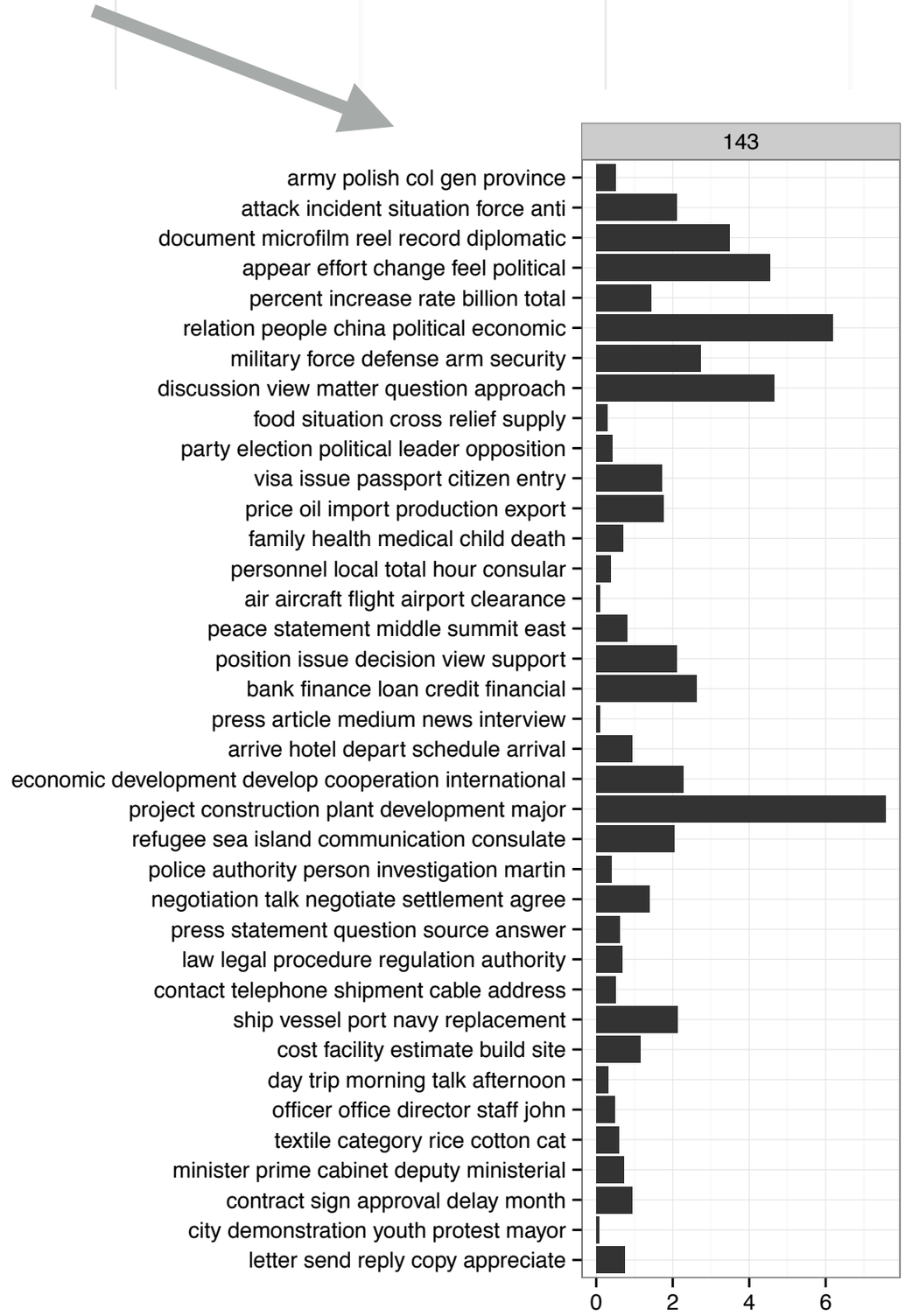
current model	previous model
word counts are observations	fit LDA then hold document topics fixed as observations
gamma-distributed event strengths	Poisson or Bernoulli-distributed event occurrences
event descriptions in vocabulary space	event descriptions in topic space
conjugate model, closed-form updates	non-conjugate model, BBVI

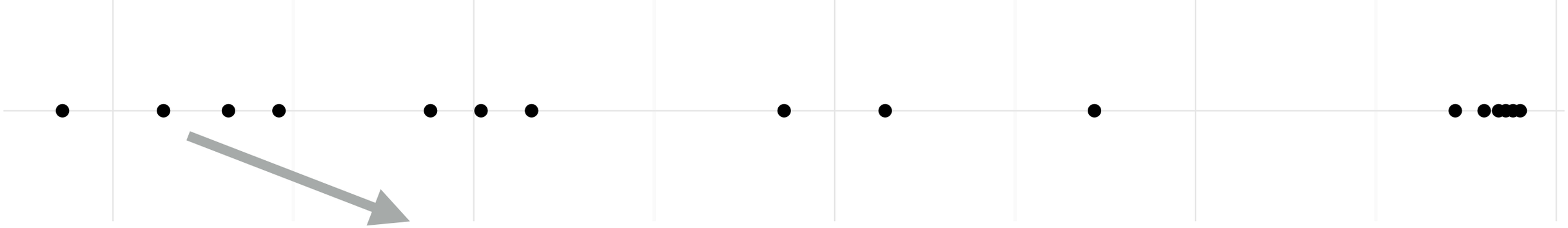
Exploration: Multiple Entities

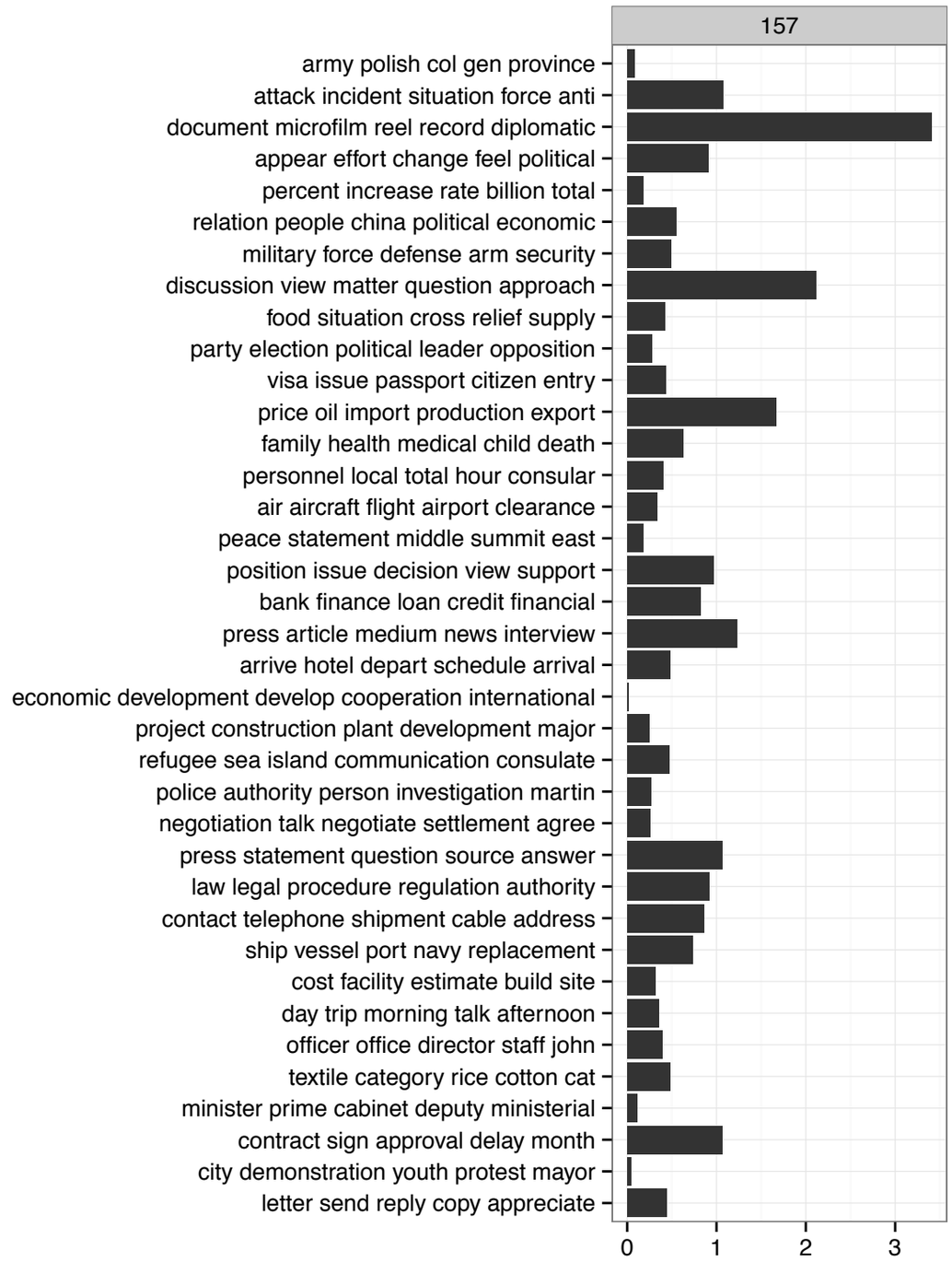


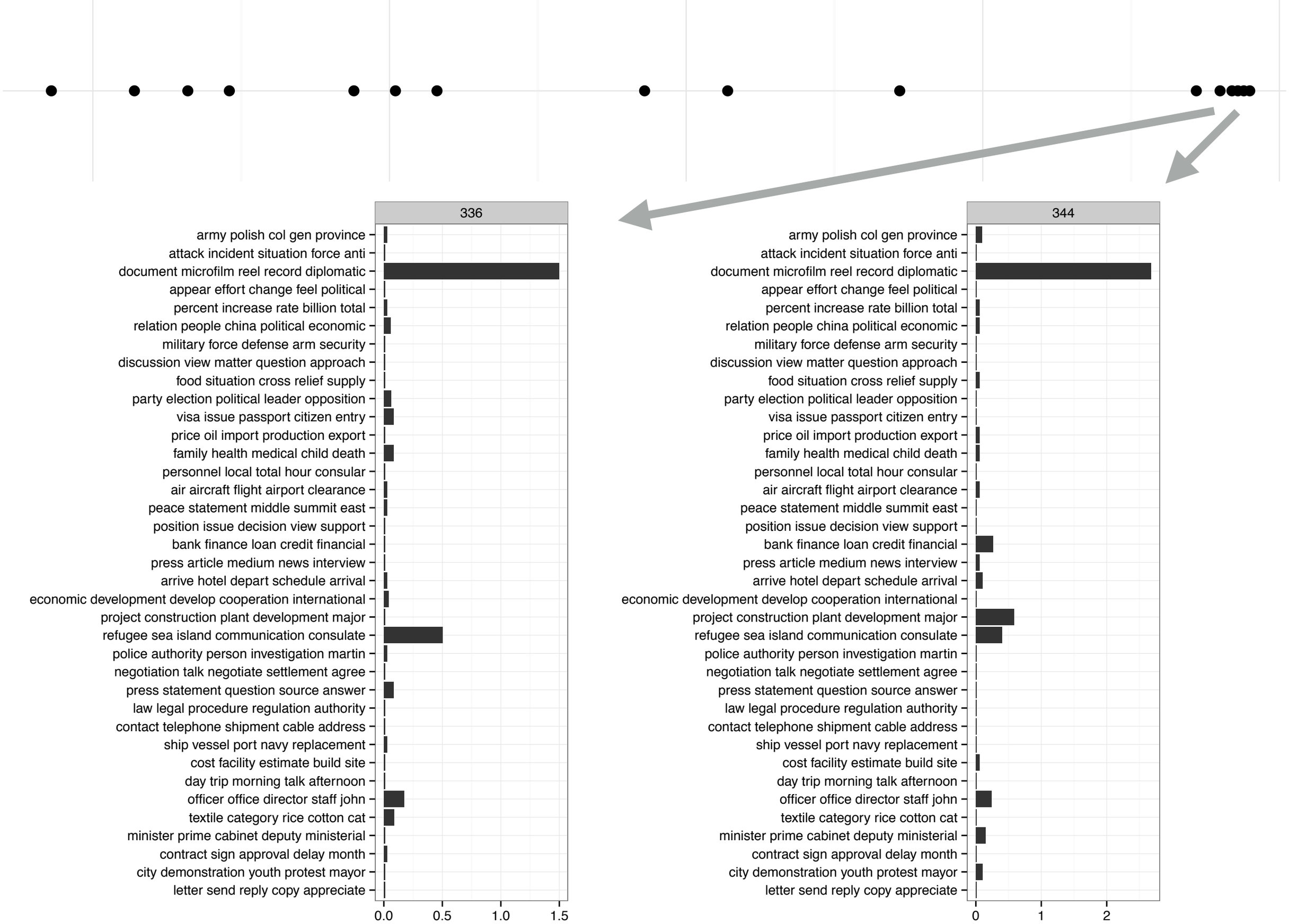





143
army polish col gen province
attack incident situation force anti
document microfilm reel record diplomatic
appear effort change feel political
percent increase rate billion total
relation people china political economic
military force defense arm security
discussion view matter question approach
food situation cross relief supply
party election political leader opposition
visa issue passport citizen entry
price oil import production export
family health medical child death
personnel local total hour consular
air aircraft flight airport clearance
peace statement middle summit east
position issue decision view support
bank finance loan credit financial
press article medium news interview
arrive hotel depart schedule arrival
economic development develop cooperation international
project construction plant development major
refugee sea island communication consulate
police authority person investigation martin
negotiation talk negotiate settlement agree
press statement question source answer
law legal procedure regulation authority
contact telephone shipment cable address
ship vessel port navy replacement
cost facility estimate build site
day trip morning talk afternoon
officer office director staff john
textile category rice cotton cat
minister prime cabinet deputy ministerial
contract sign approval delay month
city demonstration youth protest mayor
letter send reply copy appreciate
0
2
4
6




157
army polish col gen province
attack incident situation force anti
document microfilm reel record diplomatic
appear effort change feel political
percent increase rate billion total
relation people china political economic
military force defense arm security
discussion view matter question approach
food situation cross relief supply
party election political leader opposition
visa issue passport citizen entry
price oil import production export
family health medical child death
personnel local total hour consular
air aircraft flight airport clearance
peace statement middle summit east
position issue decision view support
bank finance loan credit financial
press article medium news interview
arrive hotel depart schedule arrival
economic development develop cooperation international
project construction plant development major
refugee sea island communication consulate
police authority person investigation martin
negotiation talk negotiate settlement agree
press statement question source answer
law legal procedure regulation authority
contact telephone shipment cable address
ship vessel port navy replacement
cost facility estimate build site
day trip morning talk afternoon
officer office director staff john
textile category rice cotton cat
minister prime cabinet deputy ministerial
contract sign approval delay month
city demonstration youth protest mayor
letter send reply copy appreciate
0
1
2
3


336
army polish col gen province
attack incident situation force anti
document microfilm reel record diplomatic
appear effort change feel political
percent increase rate billion total
relation people china political economic
military force defense arm security
discussion view matter question approach
food situation cross relief supply
party election political leader opposition
visa issue passport citizen entry
price oil import production export
family health medical child death
personnel local total hour consular
air aircraft flight airport clearance
peace statement middle summit east
position issue decision view support
bank finance loan credit financial
press article medium news interview
arrive hotel depart schedule arrival
economic development develop cooperation international
project construction plant development major
refugee sea island communication consulate
police authority person investigation martin
negotiation talk negotiate settlement agree
press statement question source answer
law legal procedure regulation authority
contact telephone shipment cable address
ship vessel port navy replacement
cost facility estimate build site
day trip morning talk afternoon
officer office director staff john
textile category rice cotton cat
minister prime cabinet deputy ministerial
contract sign approval delay month
city demonstration youth protest mayor
letter send reply copy appreciate
0.0
0.5
1.0
1.5
344
army polish col gen province
attack incident situation force anti
document microfilm reel record diplomatic
appear effort change feel political
percent increase rate billion total
relation people china political economic
military force defense arm security
discussion view matter question approach
food situation cross relief supply
party election political leader opposition
visa issue passport citizen entry
price oil import production export
family health medical child death
personnel local total hour consular
air aircraft flight airport clearance
peace statement middle summit east
position issue decision view support
bank finance loan credit financial
press article medium news interview
arrive hotel depart schedule arrival
economic development develop cooperation international
project construction plant development major
refugee sea island communication consulate
police authority person investigation martin
negotiation talk negotiate settlement agree
press statement question source answer
law legal procedure regulation authority
contact telephone shipment cable address
ship vessel port navy replacement
cost facility estimate build site
day trip morning talk afternoon
officer office director staff john
textile category rice cotton cat
minister prime cabinet deputy ministerial
contract sign approval delay month
city demonstration youth protest mayor
letter send reply copy appreciate
0
1
2

Validation

- compare discovered events to manually collected examples of known historical events (and corresponding cables)
    - How many of the known events are recovered?
    - How does the average distributions of the known cables compares to the discovered event distribution?
- present the discovered events (date, topic distribution, and entities involved) to an expert historian

Current Work

Current Work

- Currently verifying implementation of inference algorithm on simulated data

Current Work

- Currently verifying implementation of inference algorithm on simulated data
- Goals:

Current Work

- Currently verifying implementation of inference algorithm on simulated data

- Goals:
    - show that our model outperforms heuristic baselines on simulated data

Current Work

- Currently verifying implementation of inference algorithm on simulated data

- Goals:
    - show that our model outperforms heuristic baselines on simulated data
    - explore model sensitivity to hyperparameters

Current Work

- Currently verifying implementation of inference algorithm on simulated data

- Goals:
    - show that our model outperforms heuristic baselines on simulated data
    - explore model sensitivity to hyperparameters
    - explore real-world data with the model

Current Work

- Currently verifying implementation of inference algorithm on simulated data

- Goals:
    - show that our model outperforms heuristic baselines on simulated data
    - explore model sensitivity to hyperparameters
    - explore real-world data with the model

- Model extensions

Current Work

- Currently verifying implementation of inference algorithm on simulated data

- Goals:
    - show that our model outperforms heuristic baselines on simulated data
    - explore model sensitivity to hyperparameters
    - explore real-world data with the model

- Model extensions
    - include interactions between entities

Current Work

- Currently verifying implementation of inference algorithm on simulated data

- Goals:
    - show that our model outperforms heuristic baselines on simulated data
    - explore model sensitivity to hyperparameters
    - explore real-world data with the model

- Model extensions
    - include interactions between entities
    - learn event duration

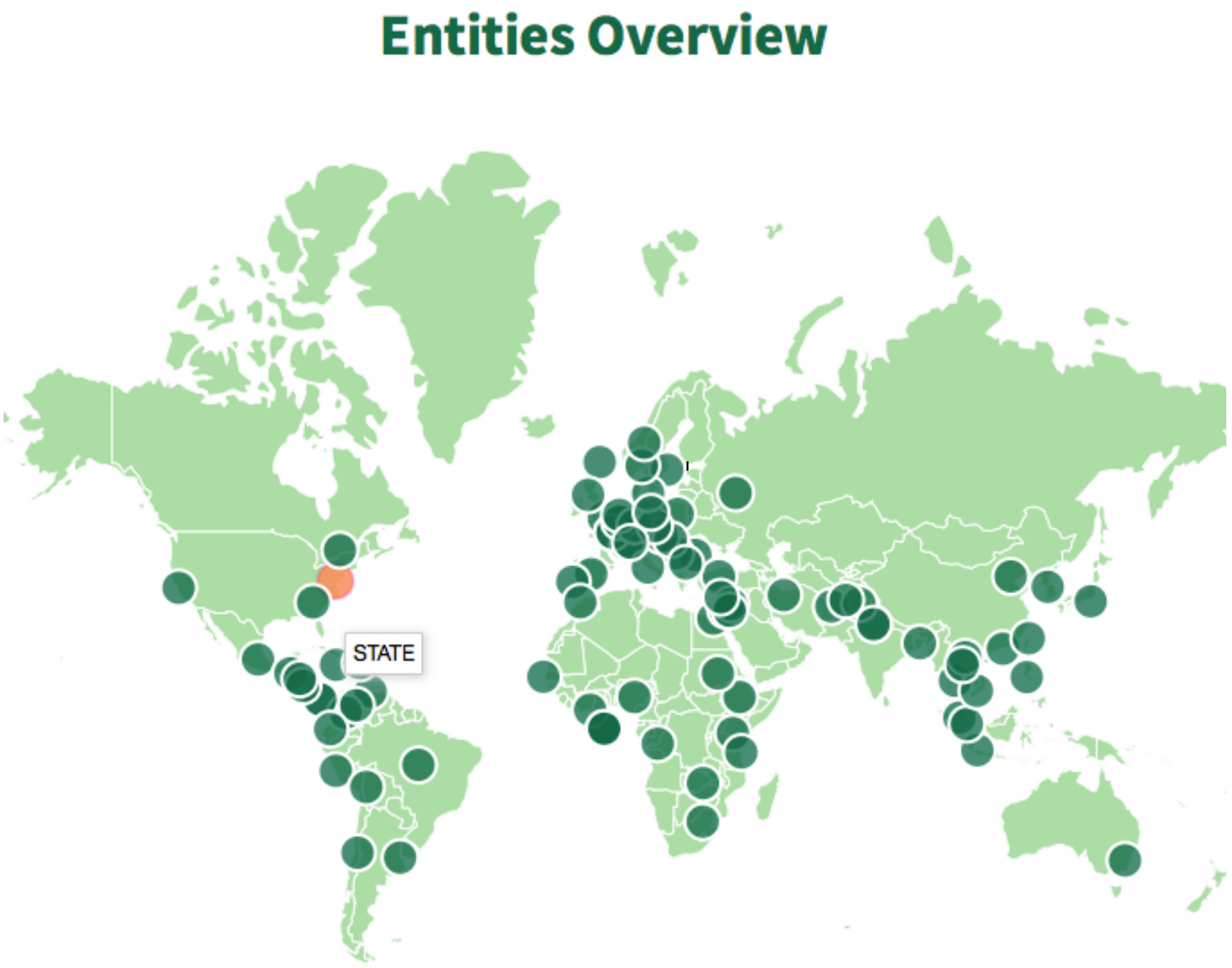
Current Work

- Currently verifying implementation of inference algorithm on simulated data

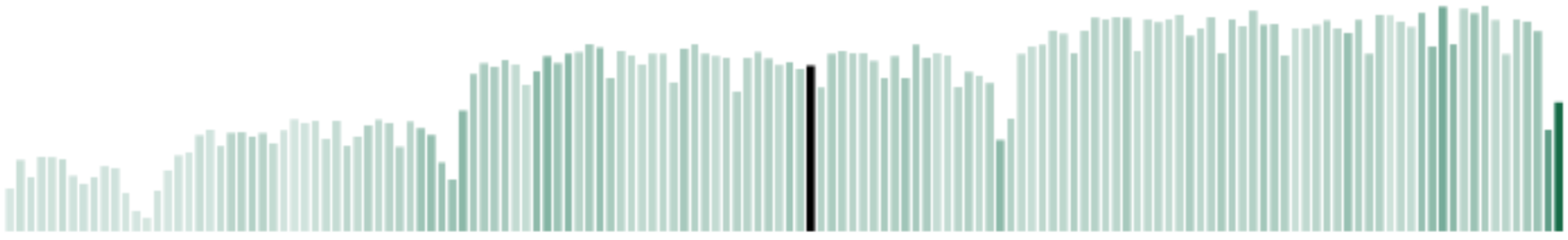
- Goals:
    - show that our model outperforms heuristic baselines on simulated data
    - explore model sensitivity to hyperparameters
    - explore real-world data with the model

- Model extensions
    - include interactions between entities
    - learn event duration
    - explore different event decay shapes

Visualization



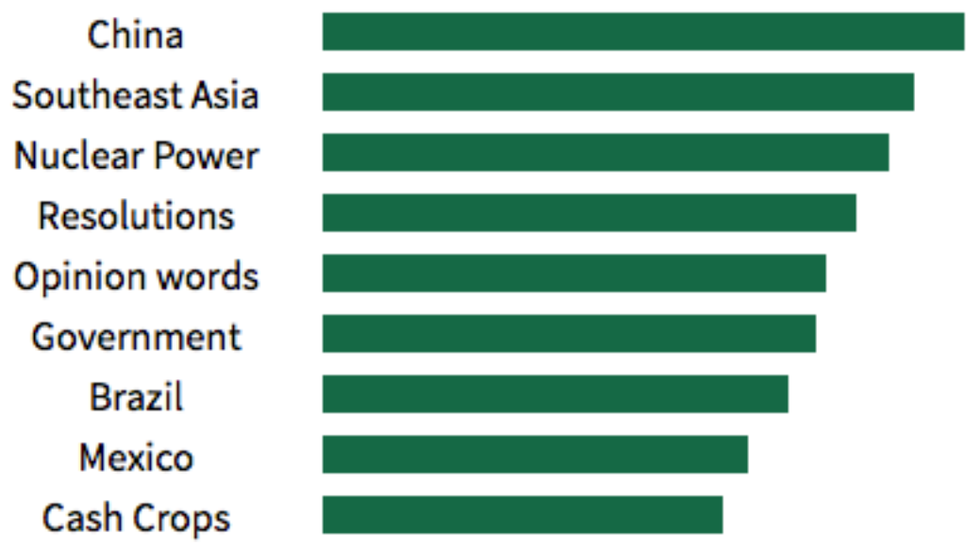
Events Overview



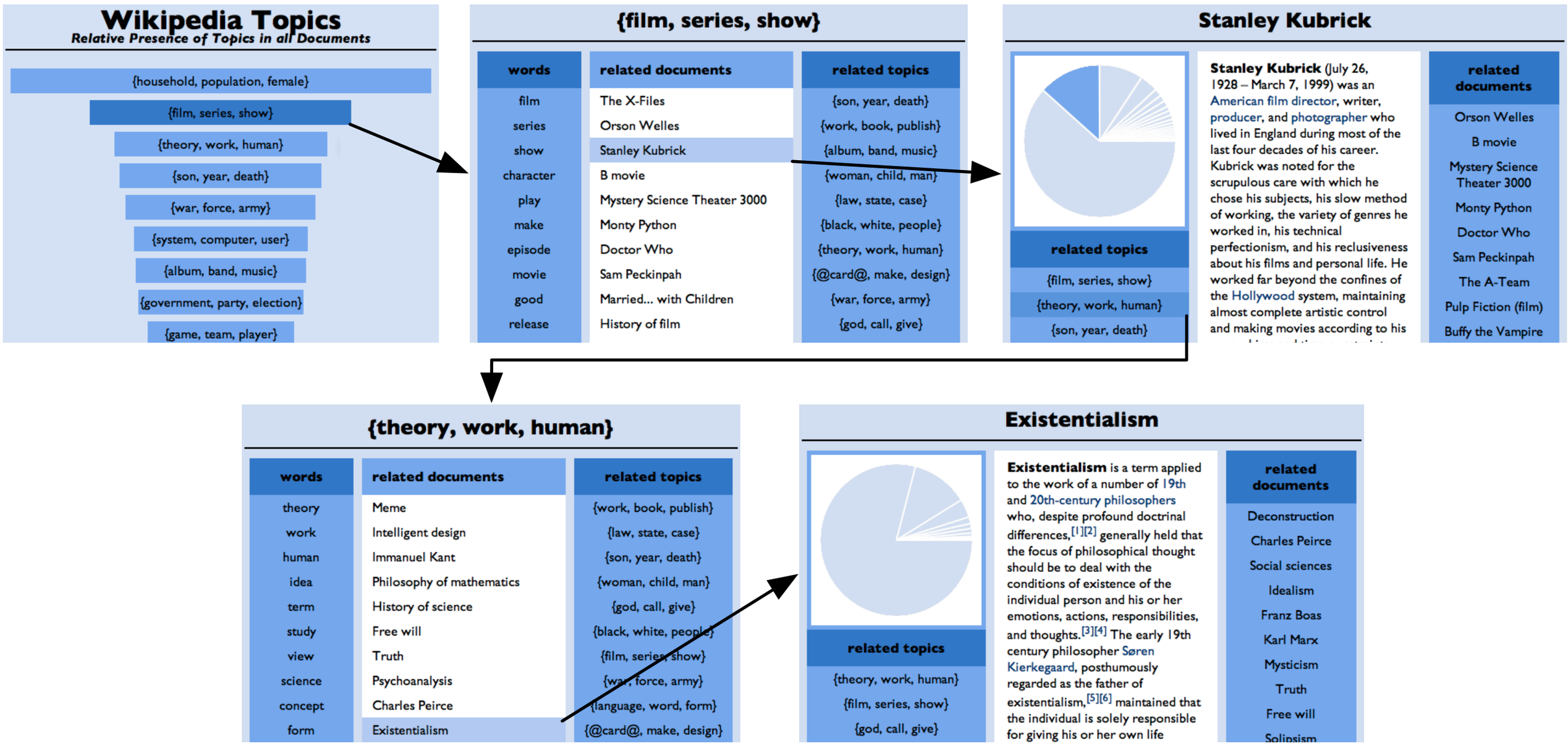
Aug. 21, 1974

8321 cables sent

Event strength: 0.085411

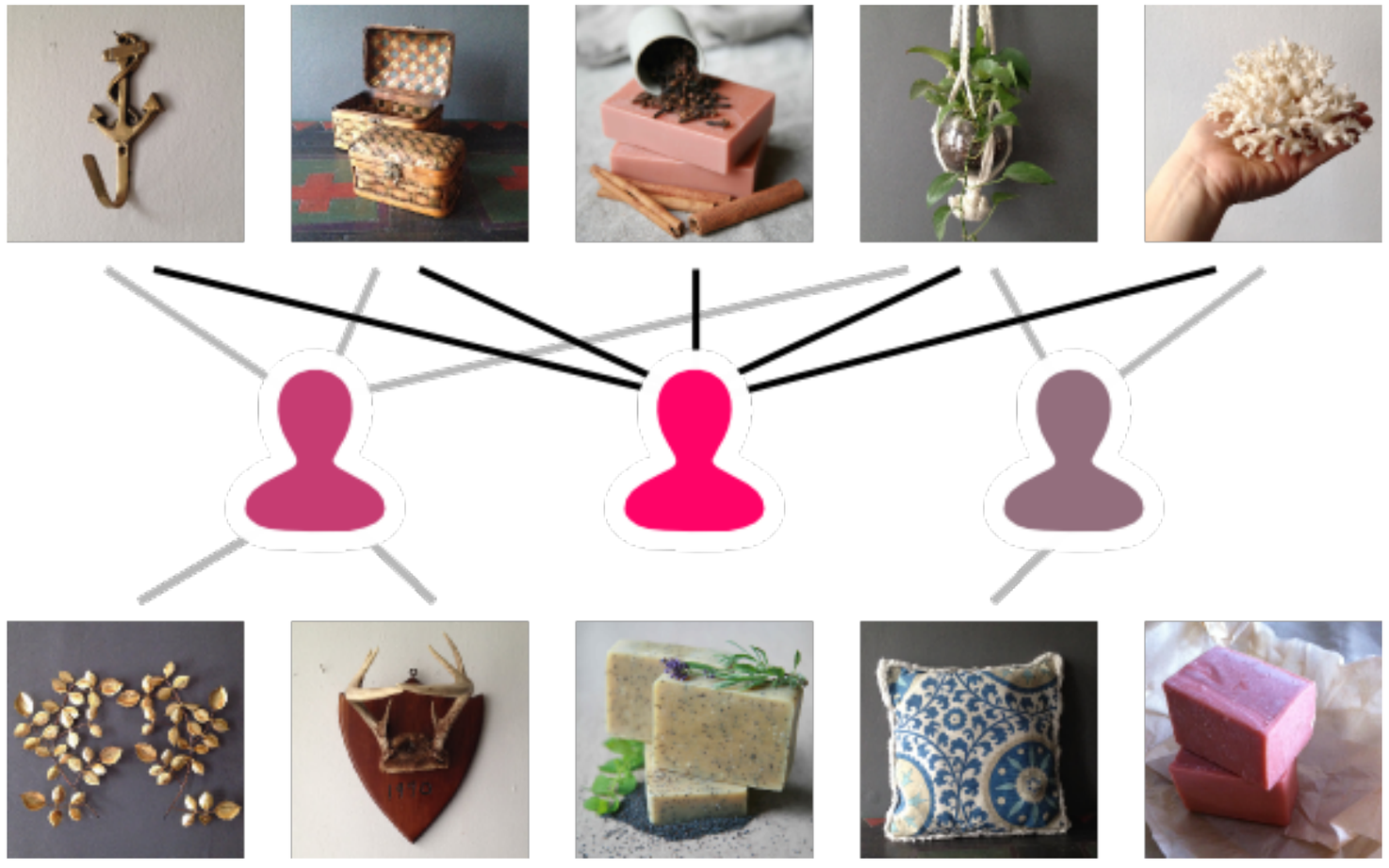


Visualizing Topic Models



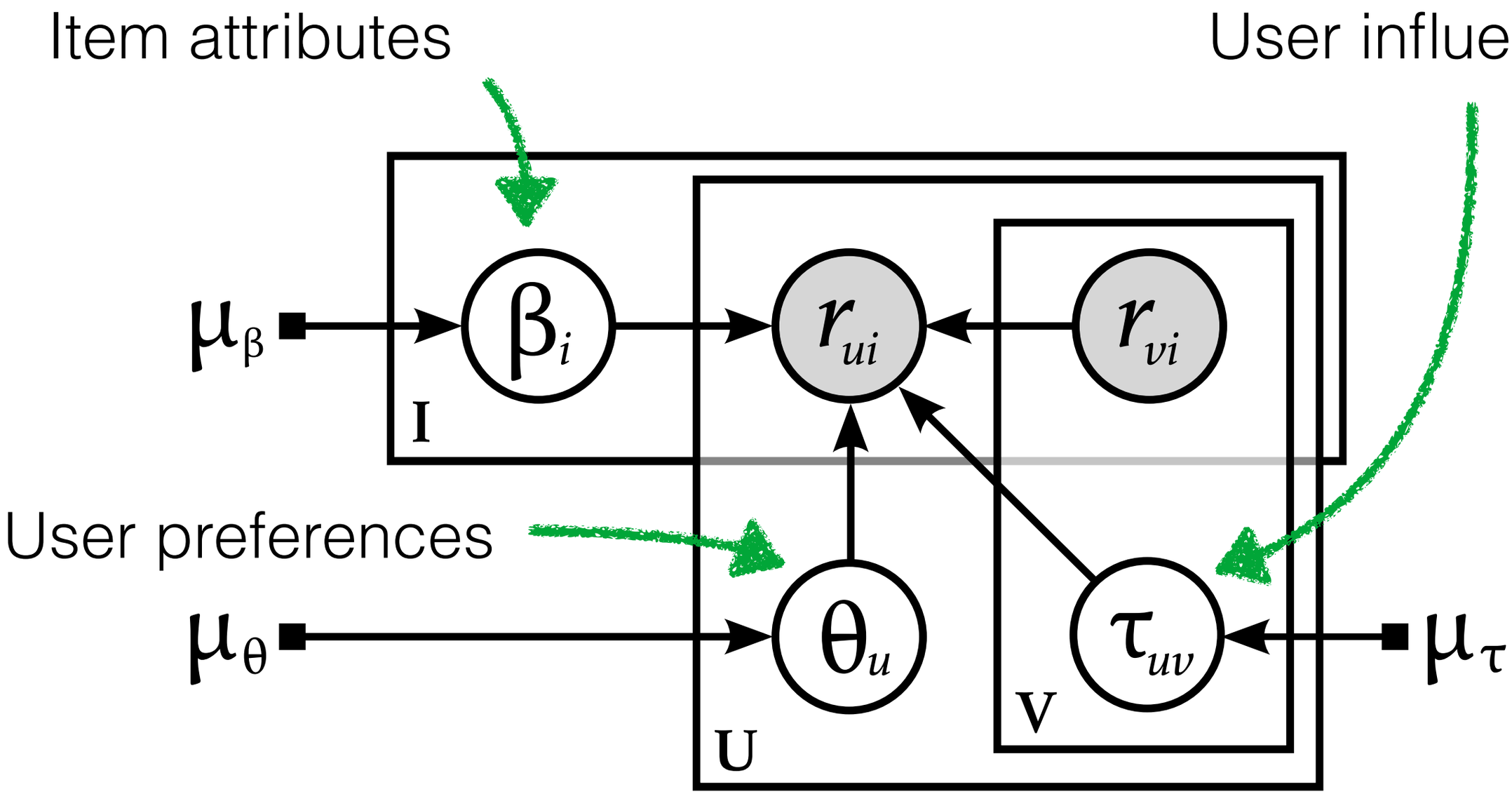
In collaboration with David Blei

Social Poisson Factorization



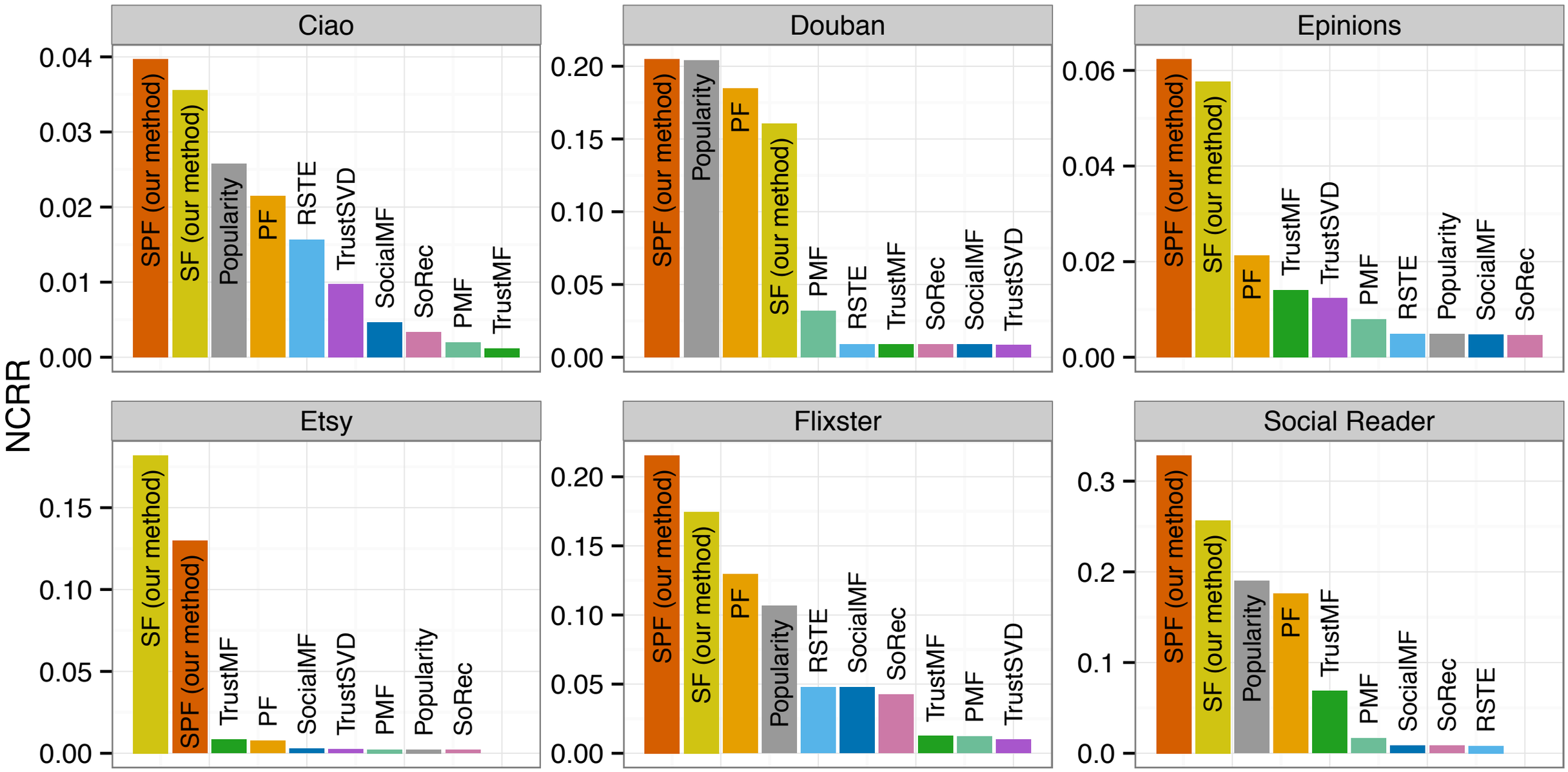
In collaboration with David Blei and Tina Eliassi-Rad

Social Poisson Factorization



In collaboration with David Blei and Tina Eliassi-Rad

Social Poisson Factorization



In collaboration with David Blei and Tina Eliassi-Rad

Dissertation Themes

Dissertation Themes

- Analysis of discrete human behavior data

Dissertation Themes

- Analysis of discrete human behavior data
- Poisson additive models for attribution

Dissertation Themes

- Analysis of discrete human behavior data
- Poisson additive models for attribution
- Visualization and exploration as first-class research problems

Thank you!

Questions and suggestions welcome.

Committee

David Blei (advisor)
Barbara Engelhardt
Elad Hazan
Brandon Stewart

Additional Collaborators

Hanna Wallach
Matthew Connelly
Rohan Shah

Core Dissertation Chapters

Social Poisson Factorization

Chaney, Allison J.B., David M. Blei, and Tina Eliassi-Rad. "A probabilistic model for using social networks in personalized item recommendation." *RecSys*. 2015.

Detecting and Characterizing Historical Events

Visualizing Topic Models

Chaney, Allison J.B., and David M. Blei. "Visualizing Topic Models." *ICWSM*. 2012.

Complete Conditionals

$$\phi_{i,k} \mid \theta, \epsilon, \pi, \mathbf{W} \sim \text{Gamma}\left(\alpha_\phi + \sum_{v,n \in N_i} z^{entity}_{n,v,k}, \beta_\phi + |N_i| \sum_v \theta_{k,v}\right)$$

$$\theta_{k,v} \mid \phi, \epsilon, \pi, \mathbf{W} \sim \text{Gamma}\left(\alpha_\theta + \sum_{v,n} z^{entity}_{n,v,k}, \beta_\theta + \sum_i |N_i| \phi_{i,k}\right)$$

$$\epsilon_t \mid \phi, \theta, \pi, \mathbf{W} \sim \text{Gamma}\left(\alpha_\epsilon + \sum_{v,n \in N_t} z^{event}_{n,v,t}, \beta_\epsilon + \sum_v \sum_{d=0}^{\delta} |N_{t+d}| f(t+d,t) \sum_k \pi_{t,k}\right)$$

$$\pi_{t,v} \mid \phi, \theta, \epsilon, \mathbf{W} \sim \text{Gamma}\left(\alpha_\pi + \sum_{v,n \in N_t} z^{event}_{n,v,t}, \beta_\pi + \sum_v \sum_{d=0}^{\delta} |N_{t+d}| f(t+d,t) \epsilon_d\right)$$