

Who, What, When, Where, and Why?

A Computational Approach to Understanding Historical Events Using State Department Cables

Allison J.B. Chaney
Princeton University

Hanna Wallach
Microsoft Research

David M. Blei
Columbia University



Matthew Connelly
History Lab at Columbia

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?*

Who?
What?
Where?
When?



observable ways
to characterize
unobservable events

Who?
What?
Where?
When?
Why?



*observable ways
to characterize
unobservable events*

data

- communications between the U.S. State Department and its embassies (“cables”)
- around two million cables
- sent between 1973 and 1977

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

O.SA2561 PEB 6
TOR 2224 21.4.75

O.SA2561 1935 21.4.75 CLA

TO.
II CANBERRA/2021 BANGKOK/U/N

FM. SAIGON / REF O.CH203930

SECRET

VIETNAM: EVACUATION

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

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

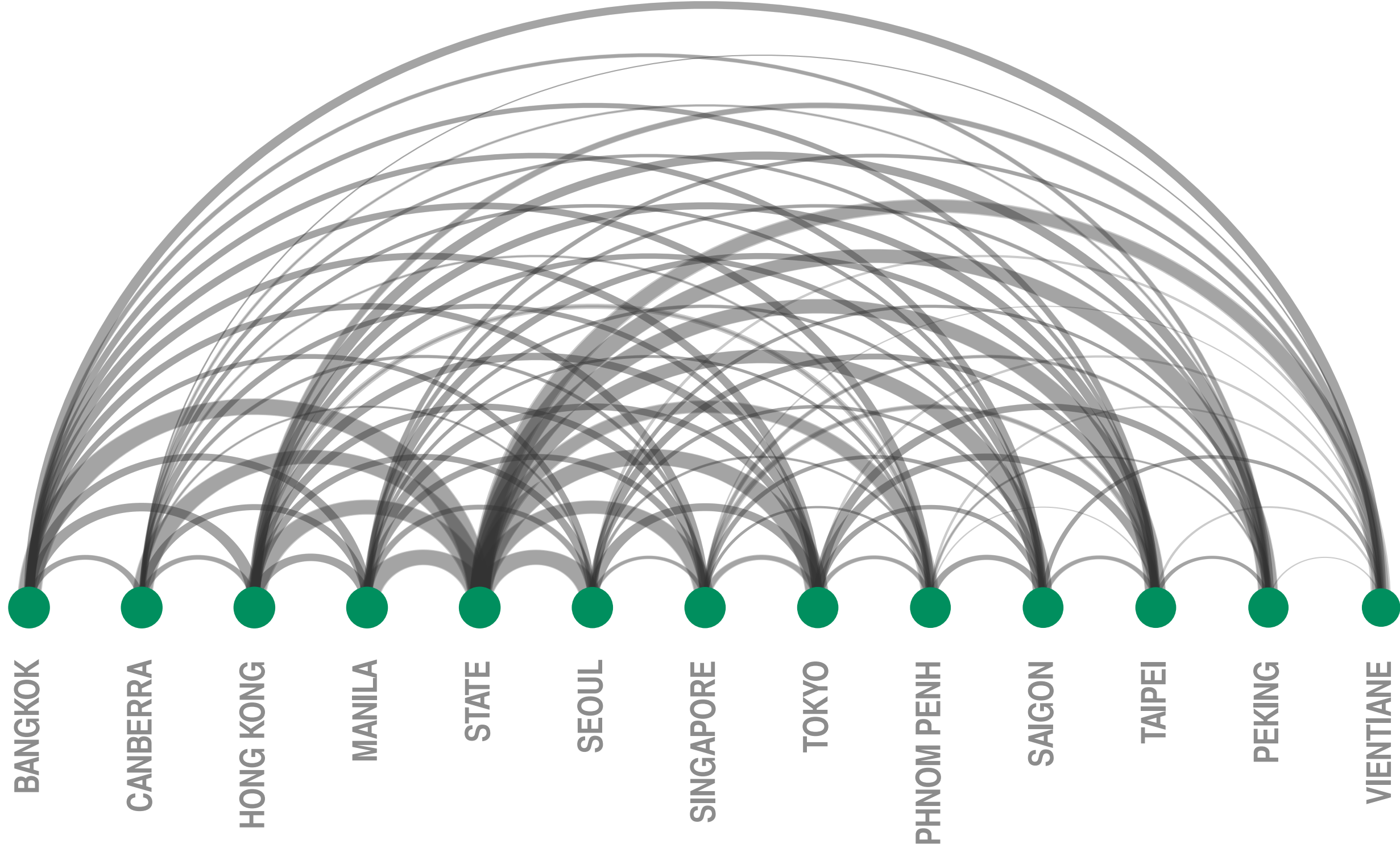
3. I WOULD LIKE TO MAINTAIN THE SUGGESTION OF A "LAST CORE" LEAVING AND I WOULD PROPOSE THAT THIS TAKE PLACE ON FRIDAY 25 APRIL RATHER THAN THURSDAY 24 APRIL IN ORDER TO ALLEVIATE PROBLEMS OF CONGESTION IN BANGKOK. I ESTIMATE THAT SECURITY SHOULD STILL BE SATISFACTORY BY THEN. I WILL ADVISE YOU IN A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

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

.....2/.

SECRET

SECRET



SAIGON

4/7/1975: LIST OF ORPHANS

INS UNABLE PROVIDE REQUESTED LIST FOR SEVERAL DAYS AND PAN AM DOES NOT HAVE LIST. FFAC SAN FRANCISCO ANTICIPATES PROVIDING DEPARTMENT COMPLETE LIST LATE ON APRIL 7.

4/9/1975: LIST OF ORPHANS

PRESIDIO SERGEANT WAS RIGHT. LIST IS OF MARGINAL VALUE. IT LISTS 78 DEAD CHILDREN WHEREAS FFAC SAIGON LIST GIVES 76. ALSO, NINE NAMES ON FFAC SAN FRANCISCO LIST NOT ON ...

STATE

4/4/1975: VIETNAMESE ORPHANS

CONGEN RECEIVING NUMEROUS INQUIRIES FROM LARGE AMERICAN MILITARY COMMUNITY HERE CONCERNING ADOPTION AND IMMIGRATION TO US OF VIETNAMESE ORPHANS. ...

NAHA

key actors

cables

4/4/1975: VIETNAMESE ORPHANS
CONGEN RECEIVING NUMEROUS
IN
AM
HI
AM
VI

4/7/1975: LIST OF ORPHANS
INS UNABLE PROVIDE REQUESTED
LIST FOR SEVERAL DAYS AND PAN
AM DOES NOT HAVE LIST. FFAC
SAN FRANCISCO ANTICIPATES
PROVIDING DEPARTMENT
COMPLETE LIST LATE ON APRIL 7.

entities



events



representing cables

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 SINCE 1967. 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.

representing cables

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 SINCE 1967. 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.

representing cables

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 SINCE 1967. 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.

representing cables

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 SINCE 1967. 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.

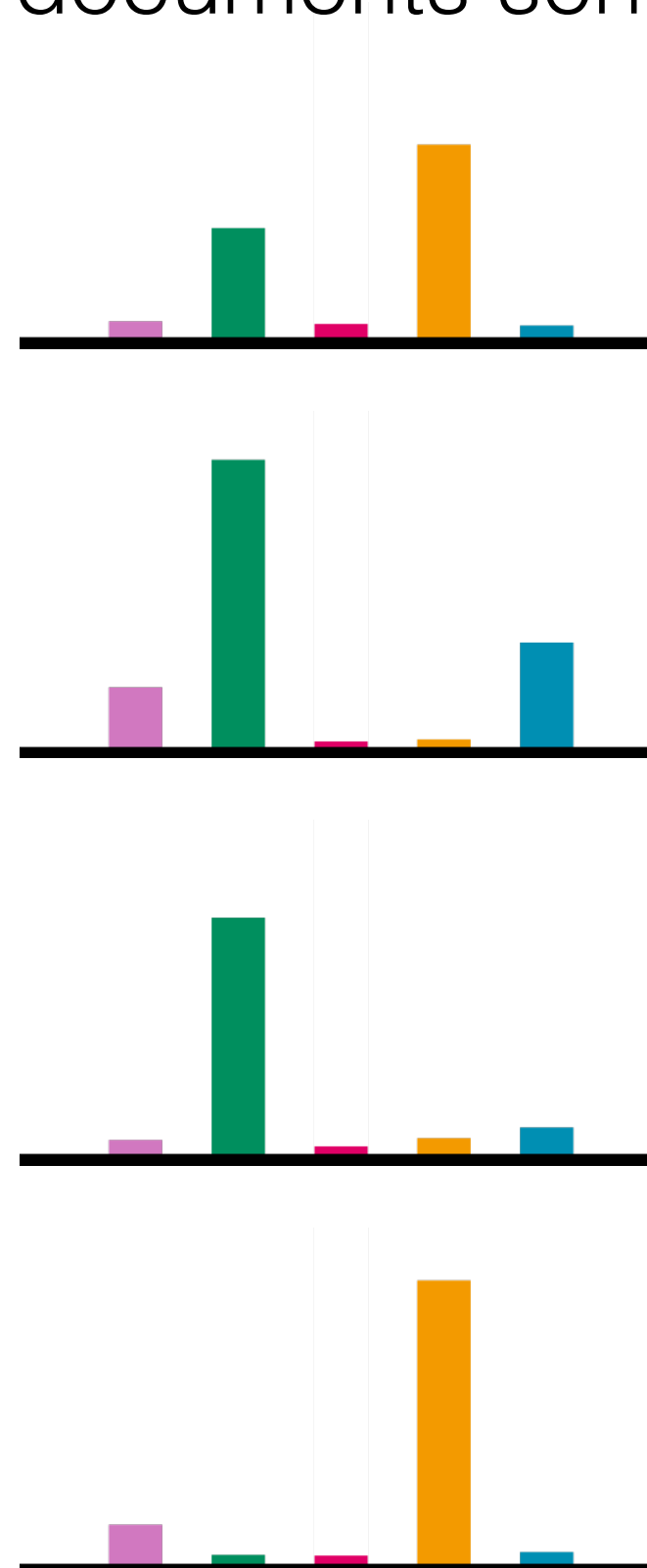


Latent Dirichlet allocation. Blei, Ng, and Jordan, 2003.



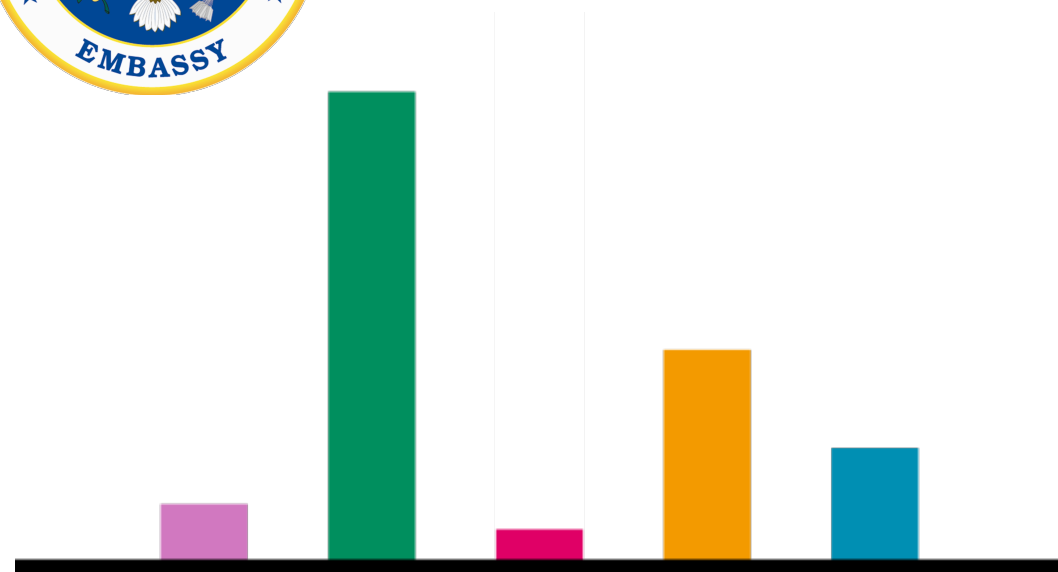
SAIGON

documents sent



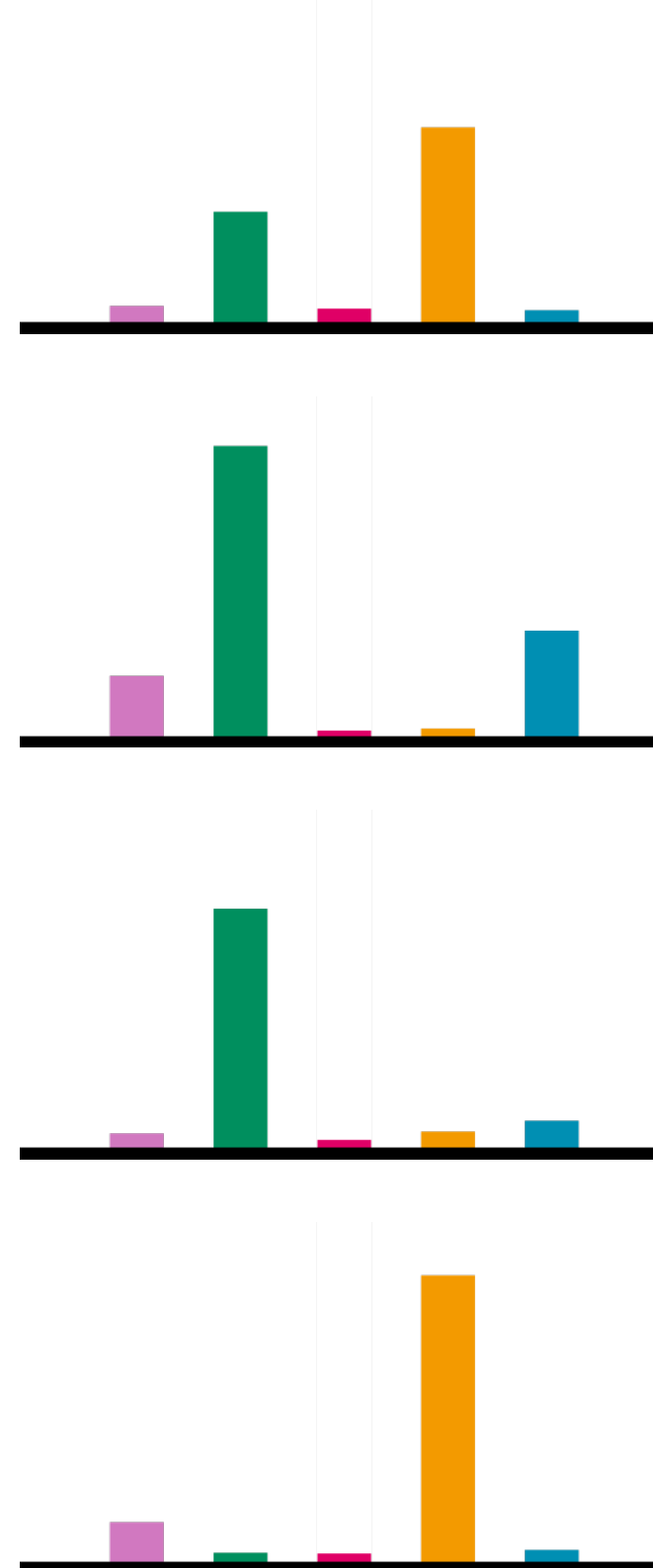


SAIGON



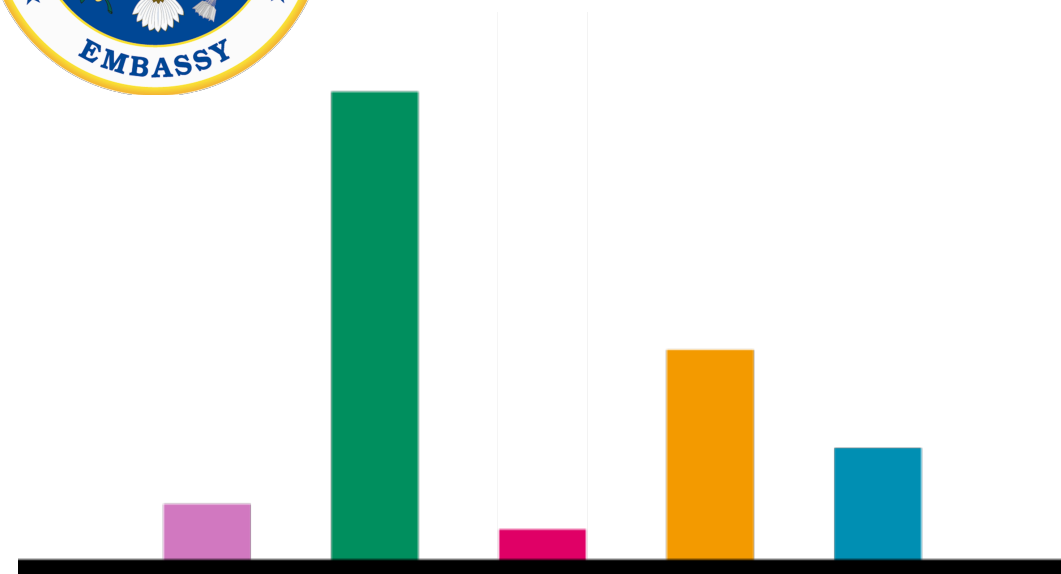
typical concerns

documents sent

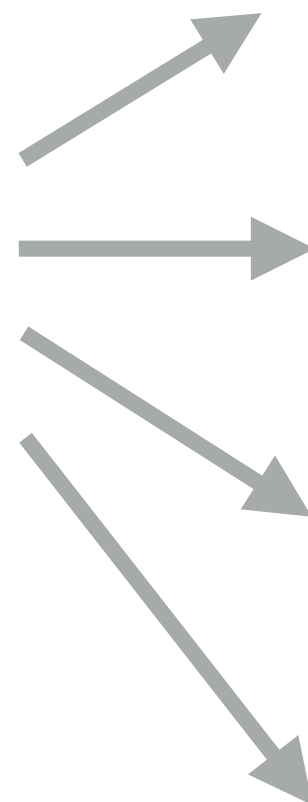




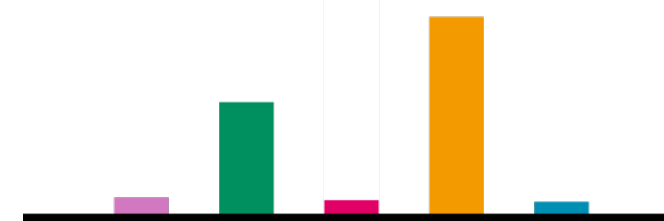
SAIGON



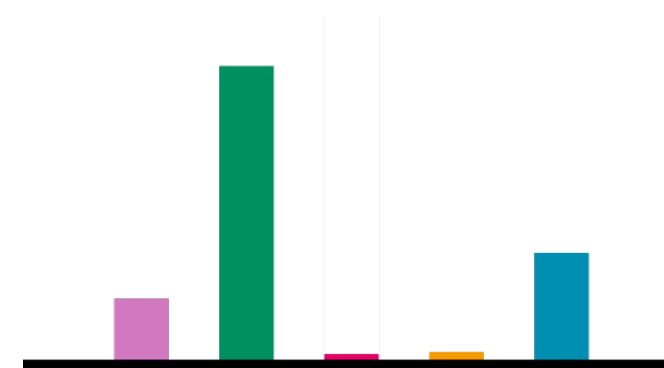
typical concerns



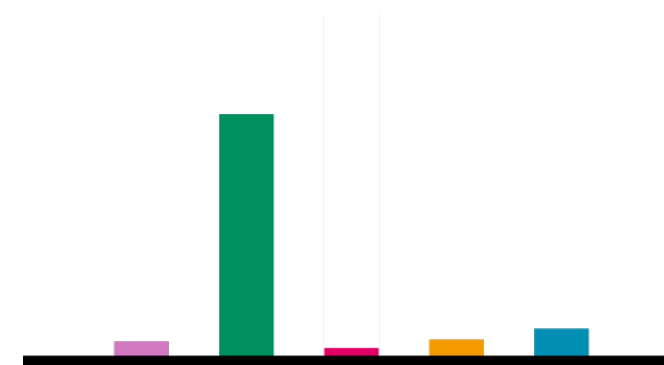
documents sent



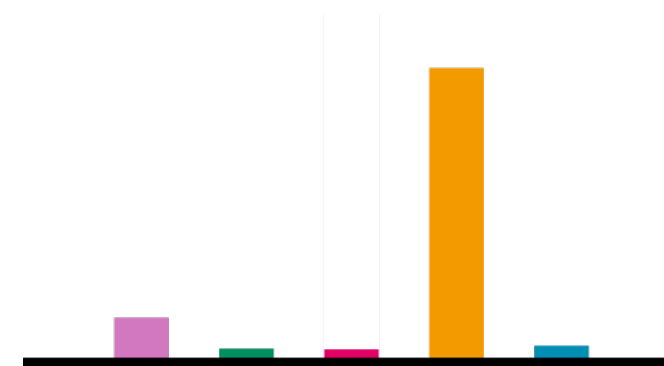
θ_1



θ_2



θ_3



\vdots

θ_d



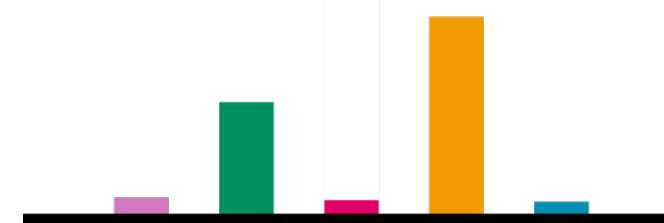
SAIGON



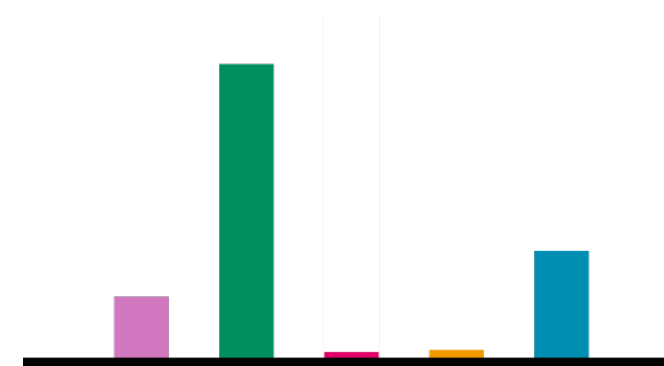
typical concerns

$$\phi_{0k} \sim \text{Gamma}(\alpha_\phi, \mu_\phi / \alpha_\phi)$$

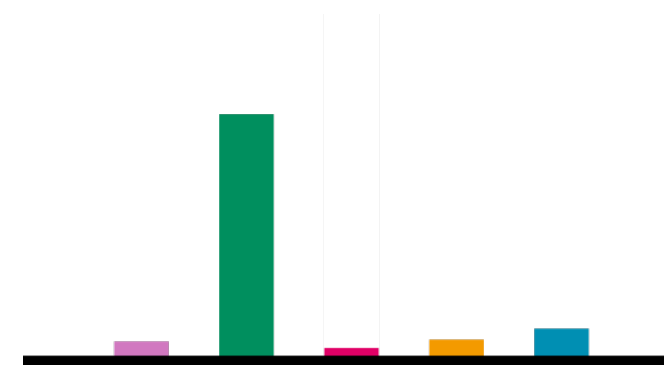
documents sent



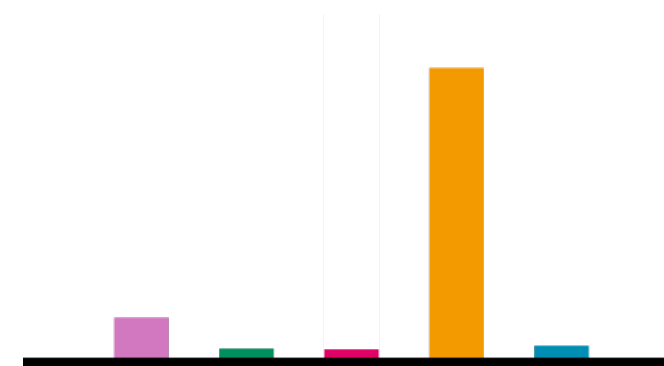
θ_1



θ_2



θ_3

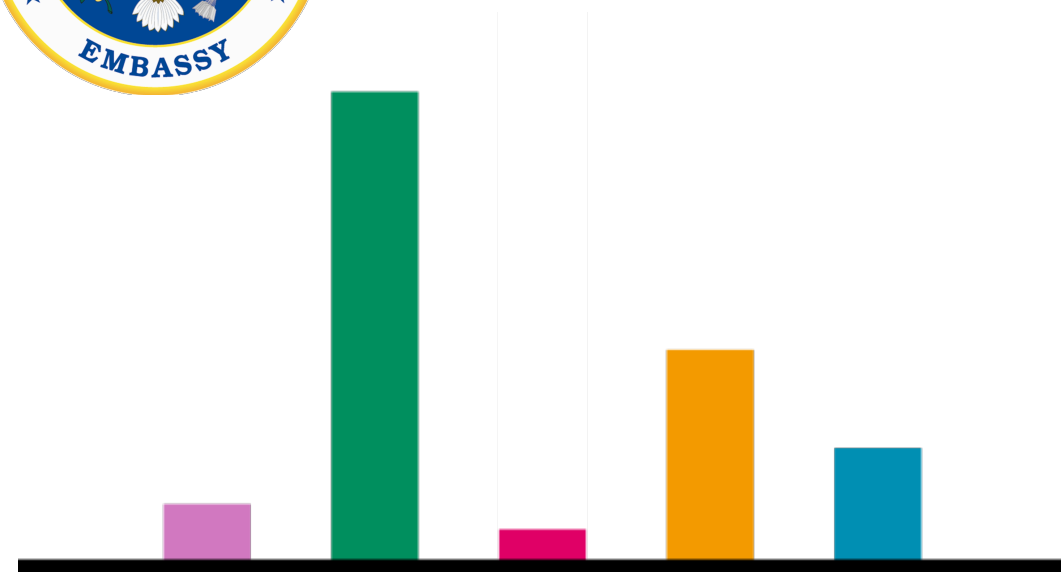


\vdots

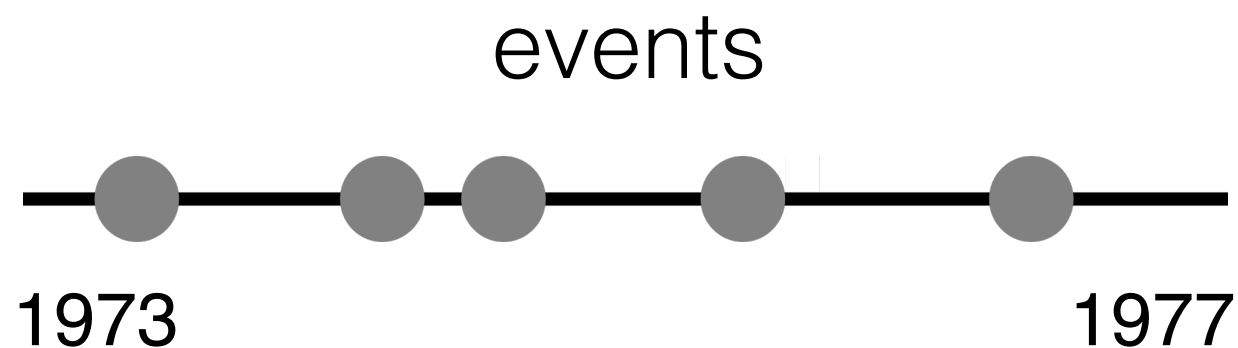
θ_d



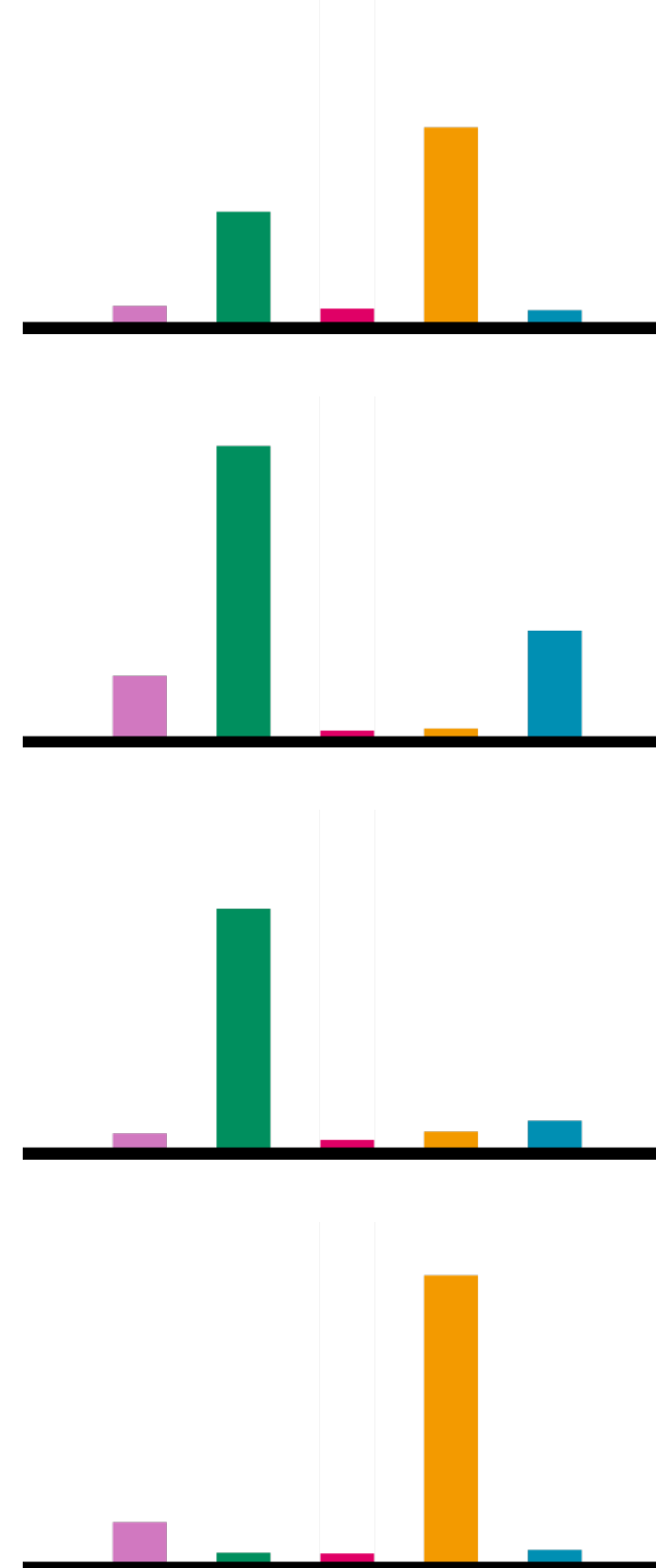
SAIGON



typical concerns



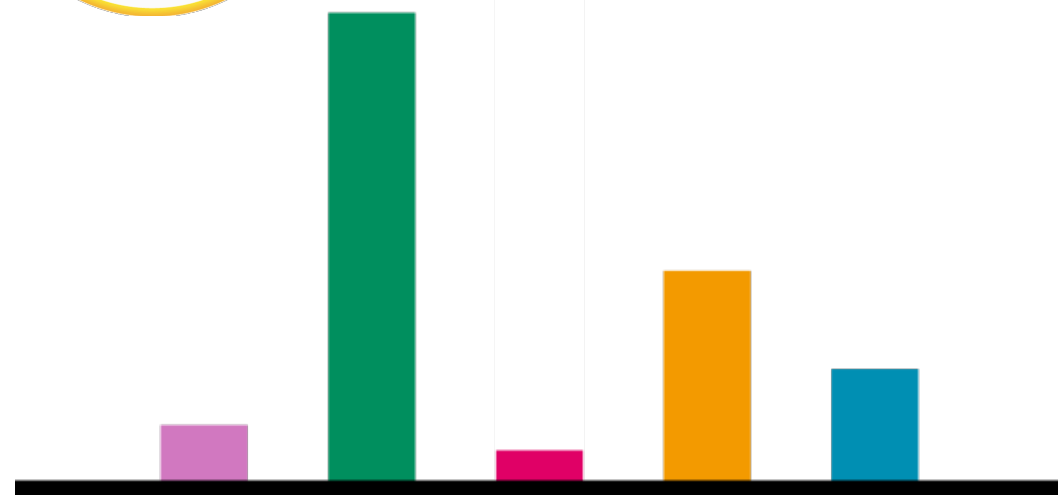
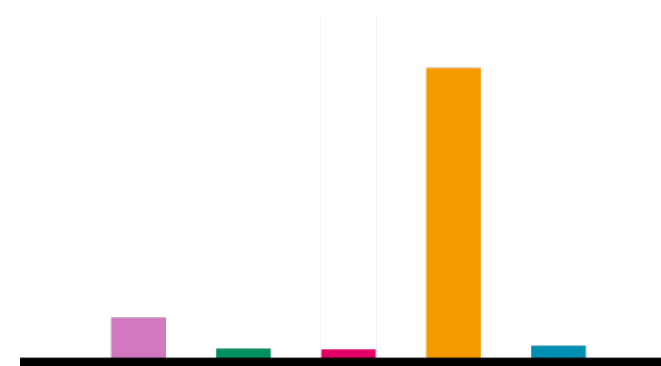
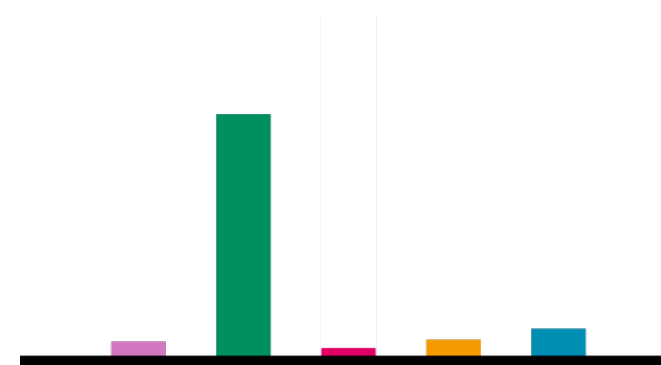
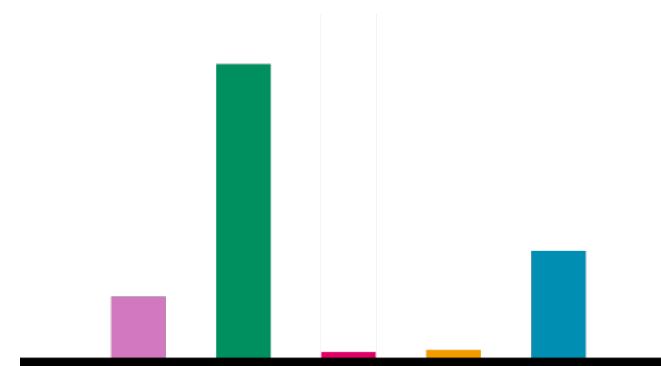
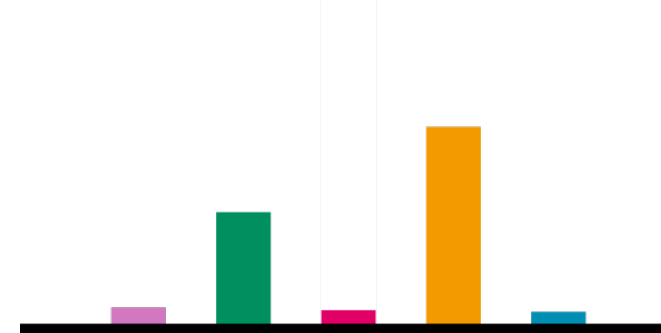
documents sent





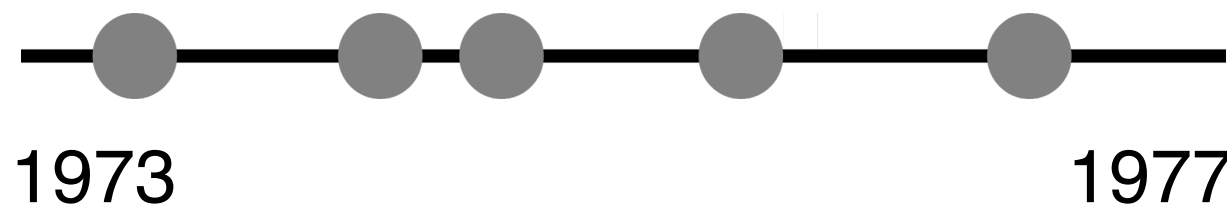
SAIGON

documents sent

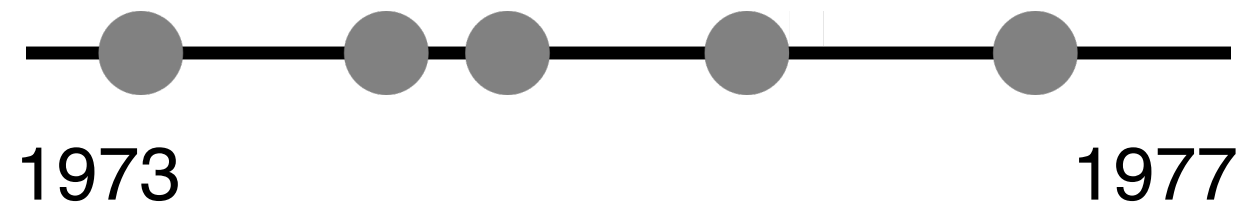


typical concerns

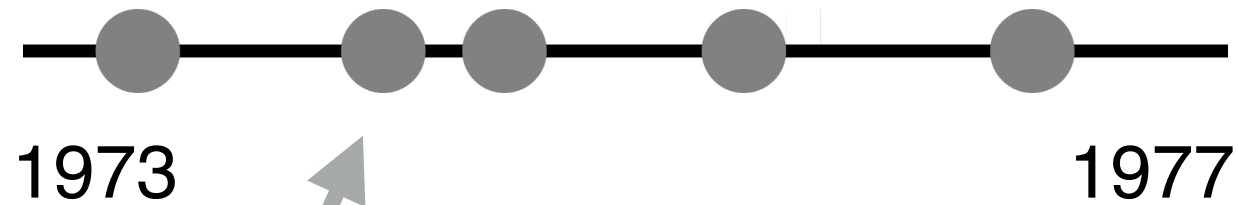
events



modeling events



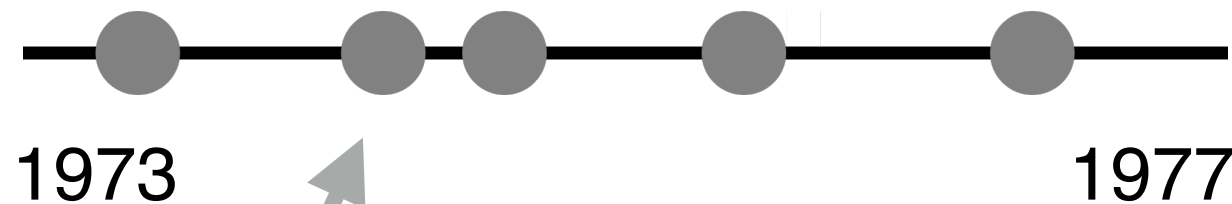
modeling events



WHEN?

$$\epsilon_i \sim \text{Poisson}(\eta_\epsilon)$$

modeling events



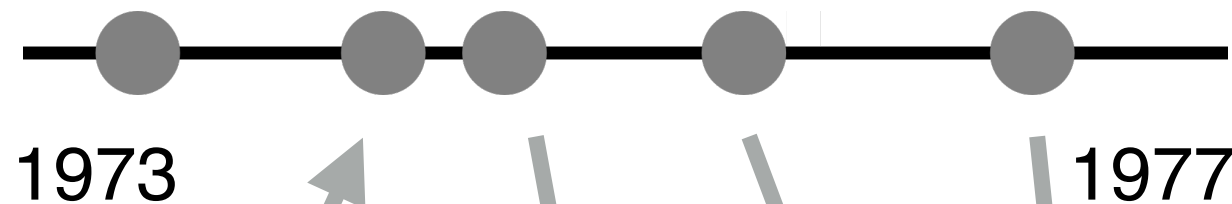
WHEN?

$$\epsilon_i \sim \text{Poisson}(\eta_\epsilon)$$

WHAT?

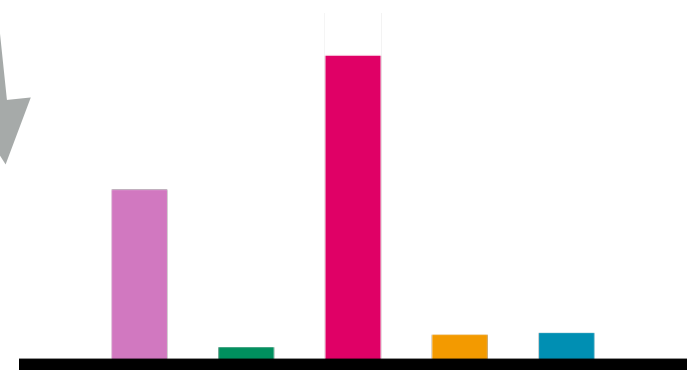
$$\pi_{ik} \sim \text{Gamma}(\alpha_\pi, \mu_\pi / \alpha_\pi)$$

modeling events



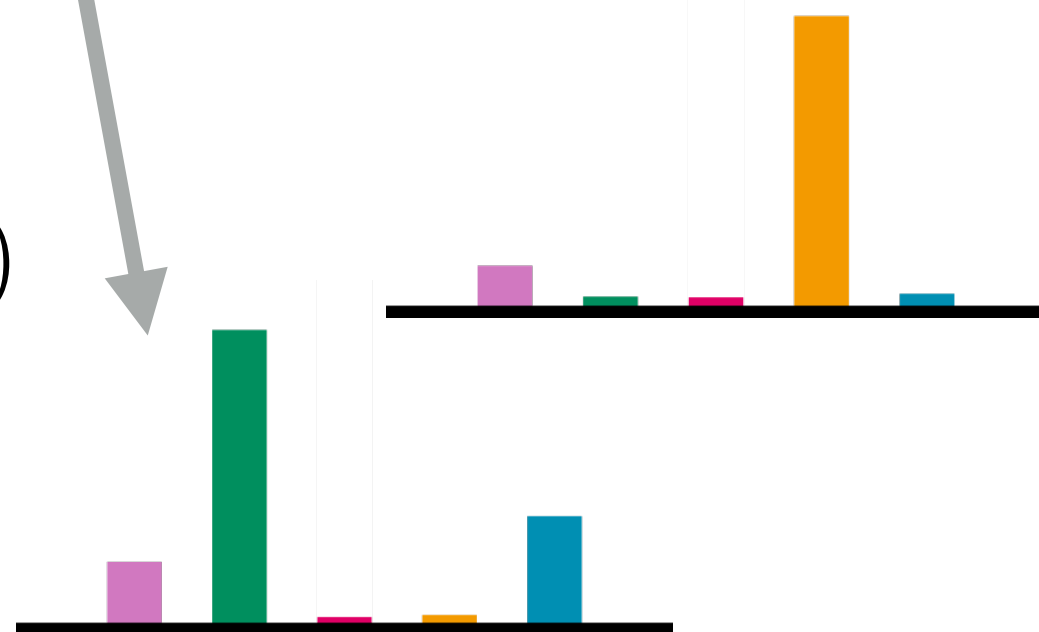
WHEN?

$$\epsilon_i \sim \text{Poisson}(\eta_\epsilon)$$

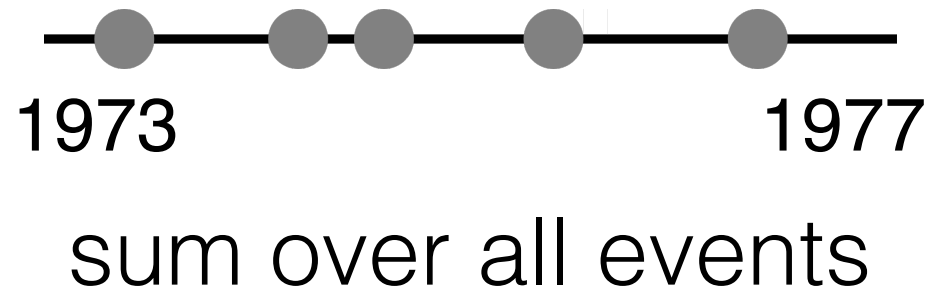
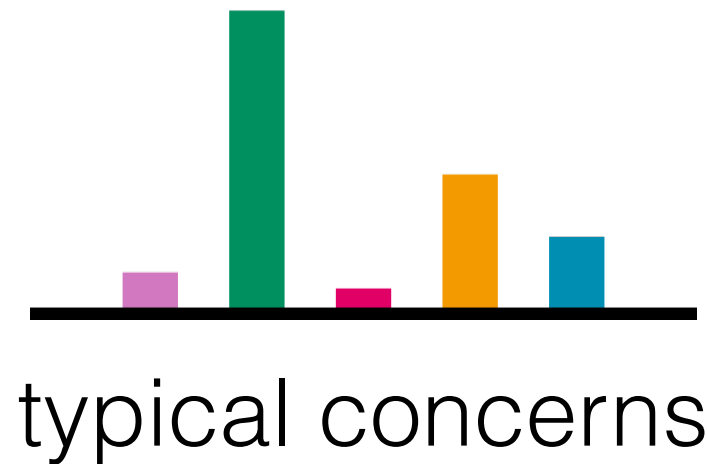


WHAT?

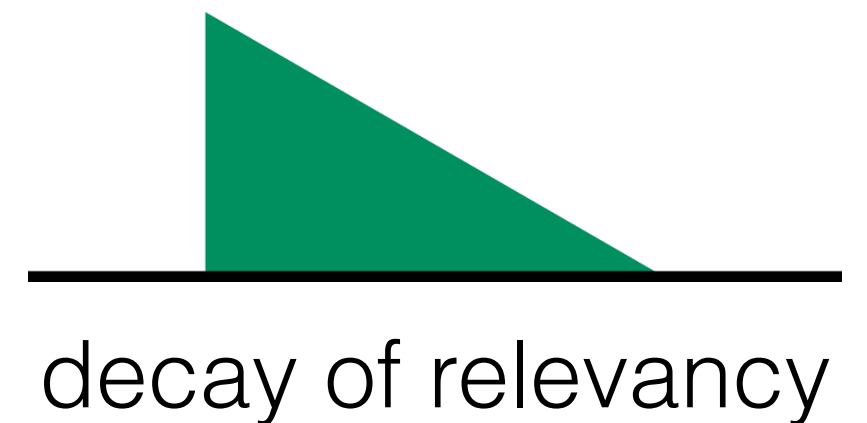
$$\pi_{ik} \sim \text{Gamma}(\alpha_\pi, \mu_\pi / \alpha_\pi)$$



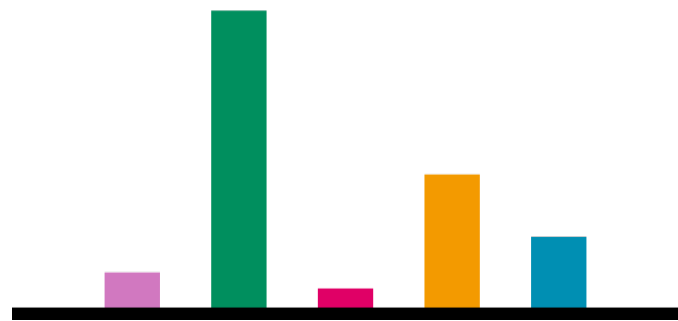
modeling cables



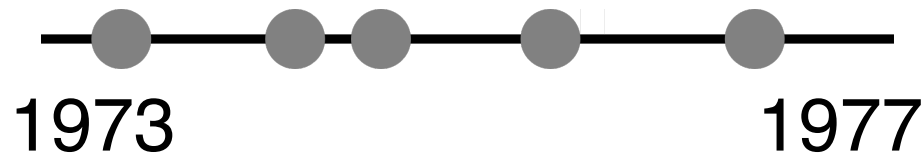
$$\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j) \pi_i$$



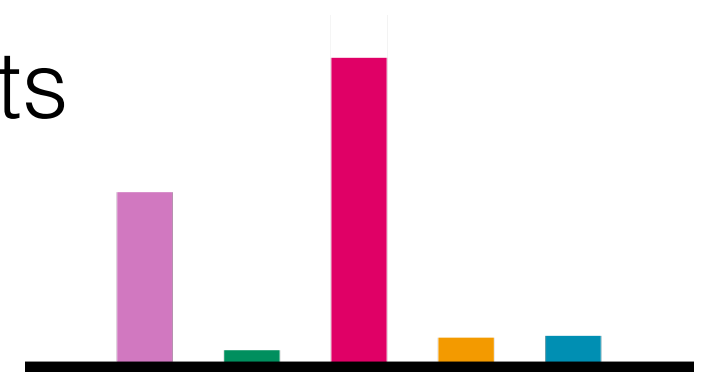
modeling cables



typical concerns



sum over all events



event description

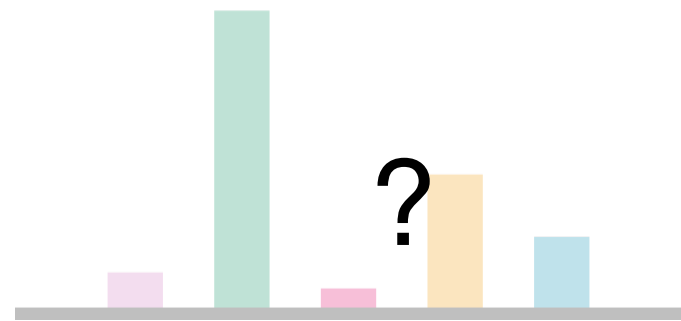


decay of relevancy

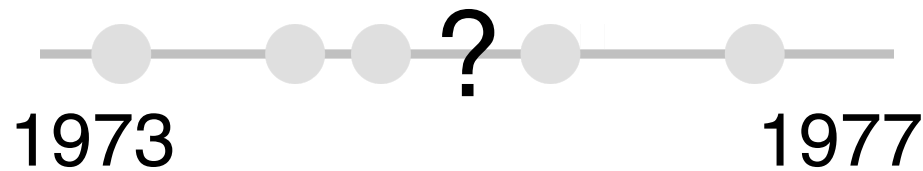
$$\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j) \pi_i$$

$$\theta_{jk} \sim \text{Gamma}(\alpha_\theta, \phi_{jk}/\alpha_\theta)$$

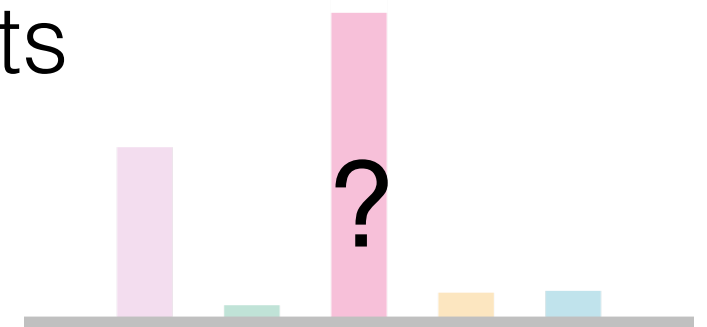
modeling cables



typical concerns



sum over all events



event description

$$\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j) \pi_i$$

$$\theta_{jk} \sim \text{Gamma}(\alpha_\theta, \phi_{jk}/\alpha_\theta)$$



decay of relevancy

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

observed data

4/4/1975: VIETNAMESE ORPHANS
CONGEN RECEIVING NUMEROUS
IN
AM
HI
AM
VI
4/7/1975: LIST OF ORPHANS
INS UNABLE PROVIDE REQUESTED
LIST FOR SEVERAL DAYS AND PAN
AM DOES NOT HAVE LIST. FFAC
SAN FRANCISCO ANTICIPATES
PROVIDING DEPARTMENT
COMPLETE LIST LATE ON APRIL 7.



**black box
variational
inference**



learned parameters

entity typical concerns

event occurrence

event content



exploration

model assumptions

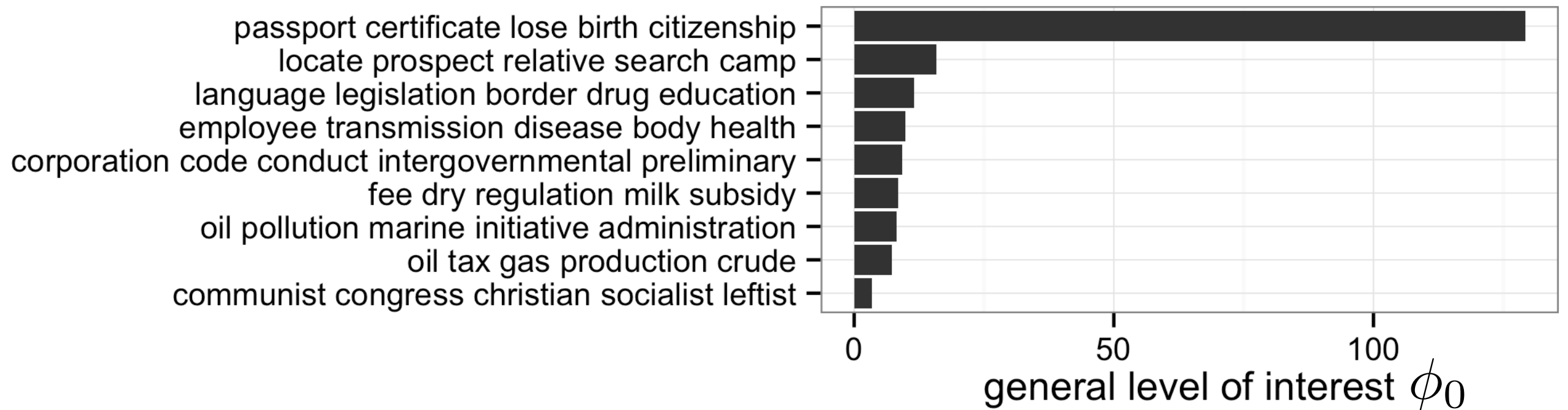


Black box variational inference. Ranganath, Gerrish, and Blei, 2014.

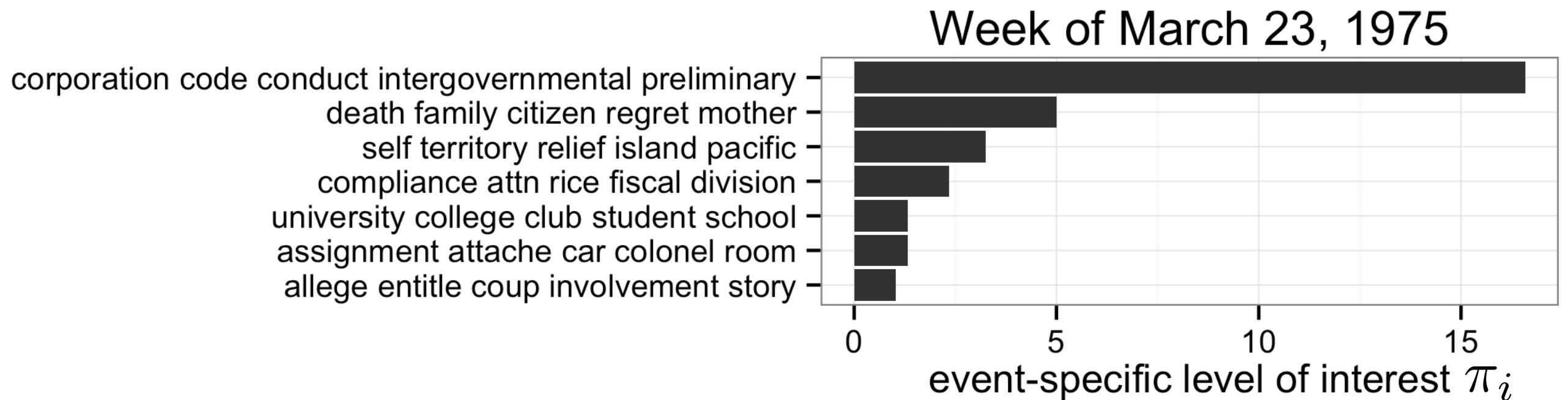
validation

- 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 topic distribution of the known cables compares to the discovered event distribution?
- present the discovered events (date, topic distribution, and entities involved) to an expert historian

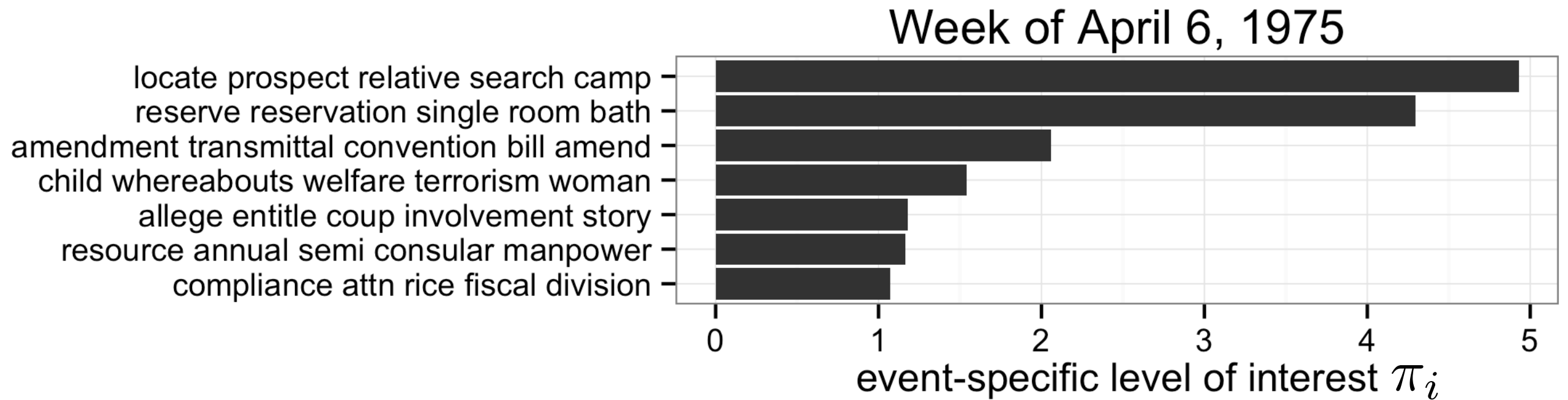
exploration: Saigon



exploration: Saigon



exploration: Saigon



results summary

- topic models can describe documents, but they cannot identify when events occur
- we explicitly model event occurrences and event content
- our model can be used to identify and explore events

future work

- Main next step: share events between entities
- Other areas:
 - include interactions between entities
 - learn event duration
 - explore different event decay shapes
 - thorough model validation

Thank you!
Questions and suggestions welcome.

