

# Topic Modeling the Global Climate Policy Debate

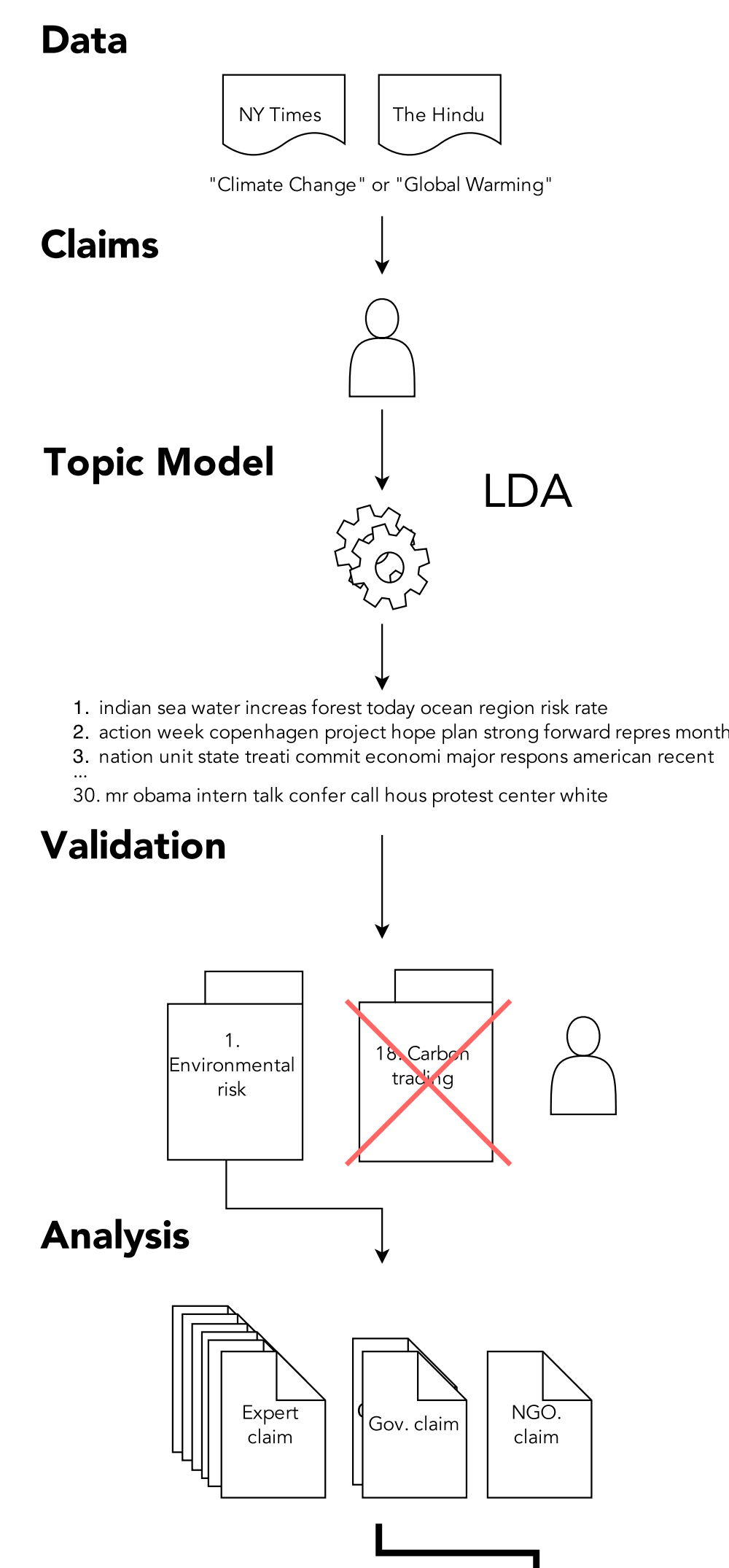
Veikko Eranti, Anna Kukkonen & Tuukka Ylä-Anttila (University of Helsinki)

## 1. Introduction

- Climate change is a **globally debated** issue in the **public sphere**
- Different **frames** are used to **interpret** it: it may be defined e.g. as an **environmental, economic, democratic, or scientific** problem
- These frames imply environmental, economic etc. **values**, which participant groups (Experts, Governments, NGOs) use to **justify** arguments
- We use **computational text analysis** with **theory of political argumentation** to see how political justifications can be **operationalized** and **found** in large text corpora
- Do participant groups speak about climate change in **different terms**?
- Do these terms form **frames** that are used to **justify political claims** and can they be **located with topic modeling**?
- Which topics **unite** and **divide** the participant groups in the debate?

justification: environmental  
 "Unchecked climate change will place the world's natural resources under incredible stress ... It will hit the poorest hardest and fastest, but none of the world's 6.8 billion people will be exempt. Science tells us that if we don't act now, an estimated 1 billion people will be uprooted because of climate change between now and 2050." – Kumi Naidoo, Executive Director, Greenpeace International, New York Times 10 Dec 2009  
 justification: equality  
 justification: scientific

## 2. Process



- We downloaded **articles** (n=677) in The New York Times and The Hindu mentioning *climate change* or *global warming* published max. 3 weeks before or after the **climate negotiations** of Kyoto (1997), Copenhagen (2009) and Durban (2011)
- We **hand-coded political claims** made by **speaker group** (Expert, Government or NGO) and **justification** used for this claim. *Who makes a statement and what does she base it on?*
- We output the data into files with one claim each, **stemmed** words with the Python Natural Language Toolkit and **removed stop words**
- Using MALLET, we ran a **Latent Dirichlet Allocation** (LDA) model on the data to produce **30 topics**. LDA groups together documents which use similar **vocabulary**
- We calculated the **standard deviation** of speaker groups in each topic to see if the topic **unites** or **divides** the groups

## 3. Validation

- Naming** and **validating** topics have to be done carefully, if we want to make **meaningful interpretations** of topics
- Based on a **qualitative reading** of the **top 10 words** and **documents** in each topic, we named 17 **semantically valid** and **relevant** topics out of 30
- Since the LDA algorithm uses **vocabulary** to group documents into topics, some topics contain documents where the **same words are used** to discuss completely **different issues**. We **discarded** the topics in which less than 8 of top 10 documents corresponded to the name we had given to them, resulting in **14 topics**

## 4. Topics

- These tables show the **7 most uniting** and **7 most dividing topics** (based on deviation of speaker groups) and their top 10 words.

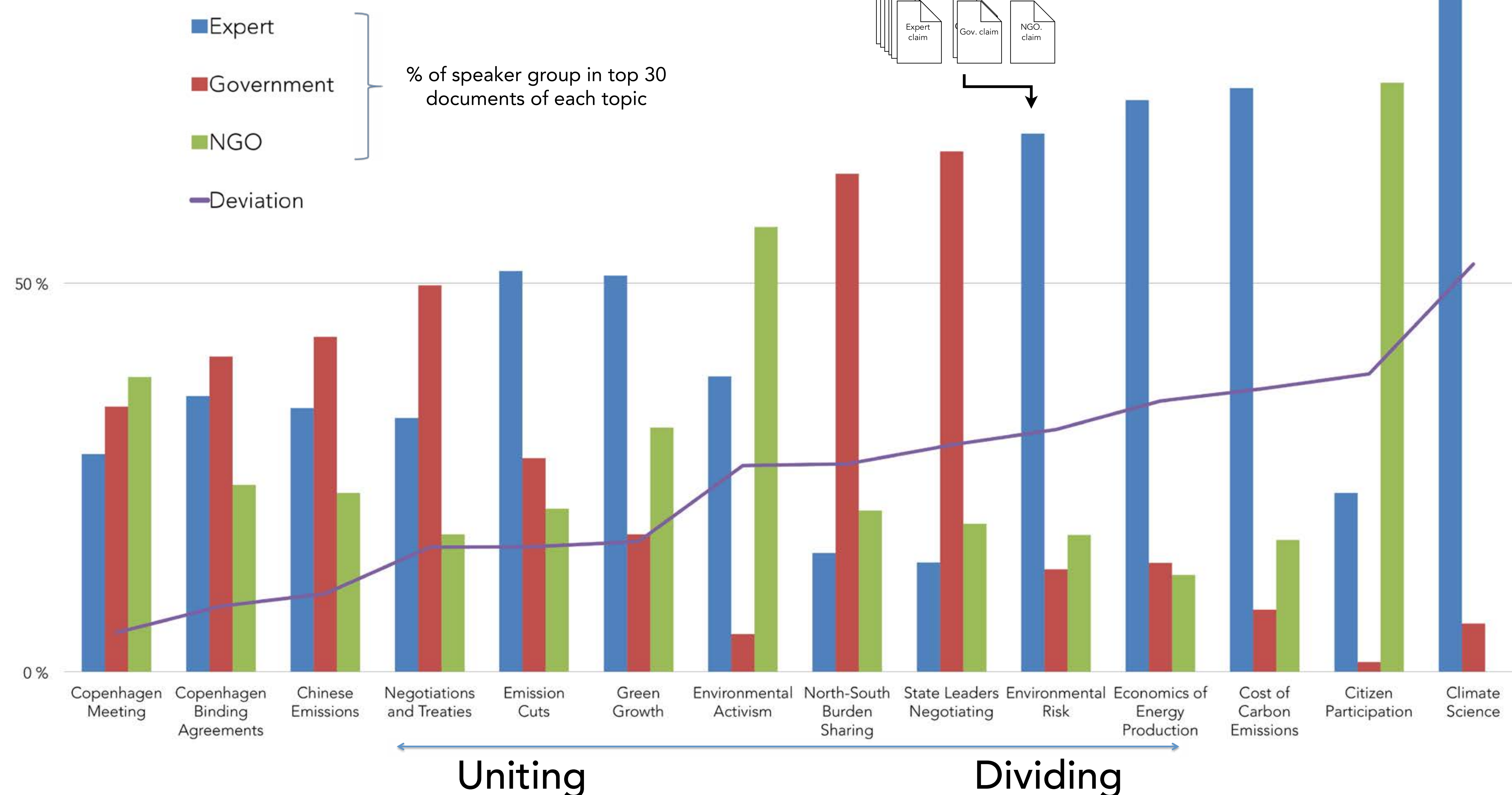
Copenhagen Meeting	Copenhagen Binding Agreement	Chinese Emissions	Negotiations and Treaties	Emission Cuts	Green Growth	Environmental Activism
action	copenhagen	china	nation	emiss	energi	peopl
week	leader	target	unit	cut	fund	govern
copenhagen	summit	reduct	state	greenhous	billion	environ
project	deal	chines	treati	industri	state	environment
hope	accord	growth	commit	gas	public	protect
plan	bind	intens	economi	gase	clean	campaign
strong	agreement	current	major	call	invest	tree
forward	end	reduc	respons	system	renew	speak
repres	legal	plan	american	adopt	creat	everi
month	told	oblig	recent	declar	govern	human

North-South Burden Sharing	State Leaders Negotiating	Environmental Risk	Economics of Energy Production	Cost of Carbon Emissions	Citizen Participation	Climate Science
countri	meet	indian	econom	carbon	part	warm
develop	minist	sea	technolog	emiss	make	global
talk	confer	water	cost	reduc	citi	scientist
commit	mr	increas	compani	dioxid	organis	research
provid	day	forest	energi	pollut	green	univers
financ	prime	today	fuel	trade	member	atmosph
demand	singh	ocean	power	cap	differ	caus
adapt	thursday	region	money	coal	number	studi
ensur	announc	risk	price	power	initi	effect
warsaw	attend	rate	mani	product	greenpeac	releas

## 5. Results

- Topic modeling can be used to analyse justifications used in debates, since **justifications are reflected in the vocabularies speakers use**
- However, even in data about a political issue, some topics cannot be interpreted as justifications (e.g. **descriptive** topics about the process of climate meetings: **Copenhagen Meeting**)
- Descriptive topics** tend to be **uniting**: discussed by all groups
- Topics that reflect **justifications** can be **uniting** or **dividing**
- For example, **Climate Science** is a topic where **experts** have strong **issue ownership** and **Citizen Participation** is discussed almost solely by **NGOs**
- However, **uniting topics can provide common ground in the debate**: e.g. **Green Growth** is discussed by all groups



politicsociology.org

veikko.eranti@helsinki.fi, anna.kukkonen@helsinki.fi, tuukka.yla-anttila@helsinki.fi