



## Digital Technologies and Value Creation

Dr. Philippe Blaettchen  
Bayes Business School (formerly Cass)

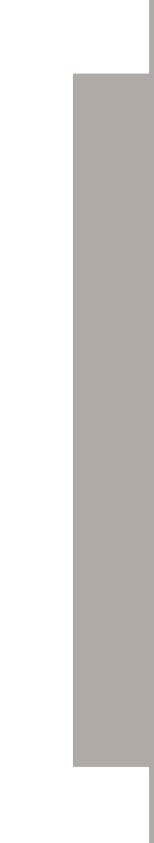
## Learning objectives of today

**Goals:** Understand cultural measurement with machine learning in more depth

- Why do we need to measure culture (recap)?
- What is the issue with traditional methods and how can we do better?
- What is topic modeling and how can it help us analyze text data?

**How will we do this?**

- We briefly recap the importance of culture in organizational design and the ways to measure it
- We will see the difficulty in judging culture without algorithmic tools
- We then go through a worked example of measuring culture using topic modeling



**A brief recap about culture and organizational design**

## An attempt at a definition

“a complex set of **values, beliefs, assumptions, and symbols** that define the way in which a firm conducts its business”

(Barney, 1986)

“a pattern of **shared** basic assumptions”

(Schein, 2010)

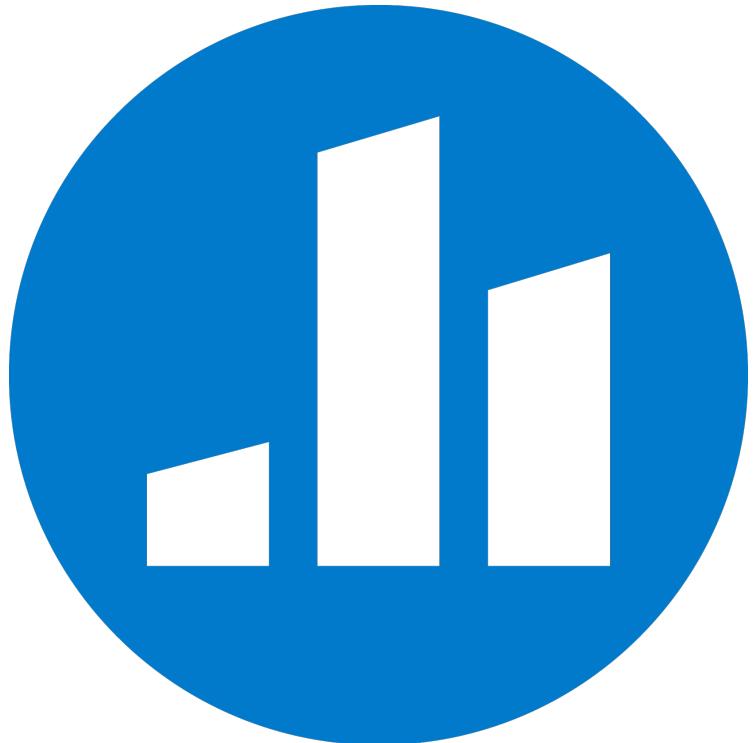


Reflects what people collectively believe & care about

## The importance of organizational culture to firms

- Effects on employees: satisfaction & turnover
- Effects on group dynamics: creativity & efficiency
- Effects on firm performance: profitability & post-merger integration

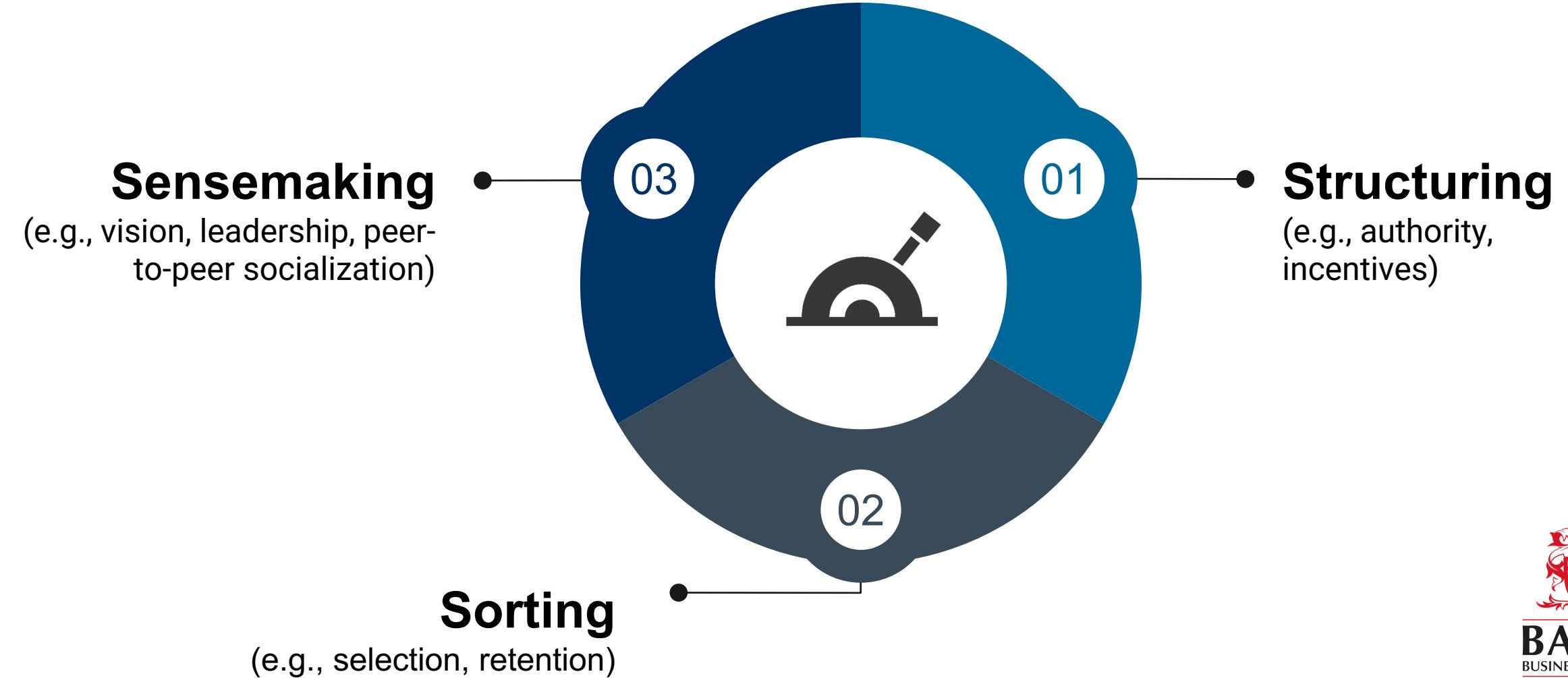
## Managing organizational culture



Drawing on your own experience, which types of initiatives do you consider most effective to create a high cultural fit of employees with their firm?

Go to [pollev.com/dtvc2021](https://pollev.com/dtvc2021)

## Three levers of organizational design



## Data sources to study organizational culture

To manage culture, we need to measure it, and to measure it, we need data

## Traditional and novel data sources to study organizational culture



- Employees' **surveys, interviews**
- **Qualitative feedback** by superiors as part of performance assessment

SURVEY ANALYSIS



- Text data produced by organizational members is ever-so **easily accessed**

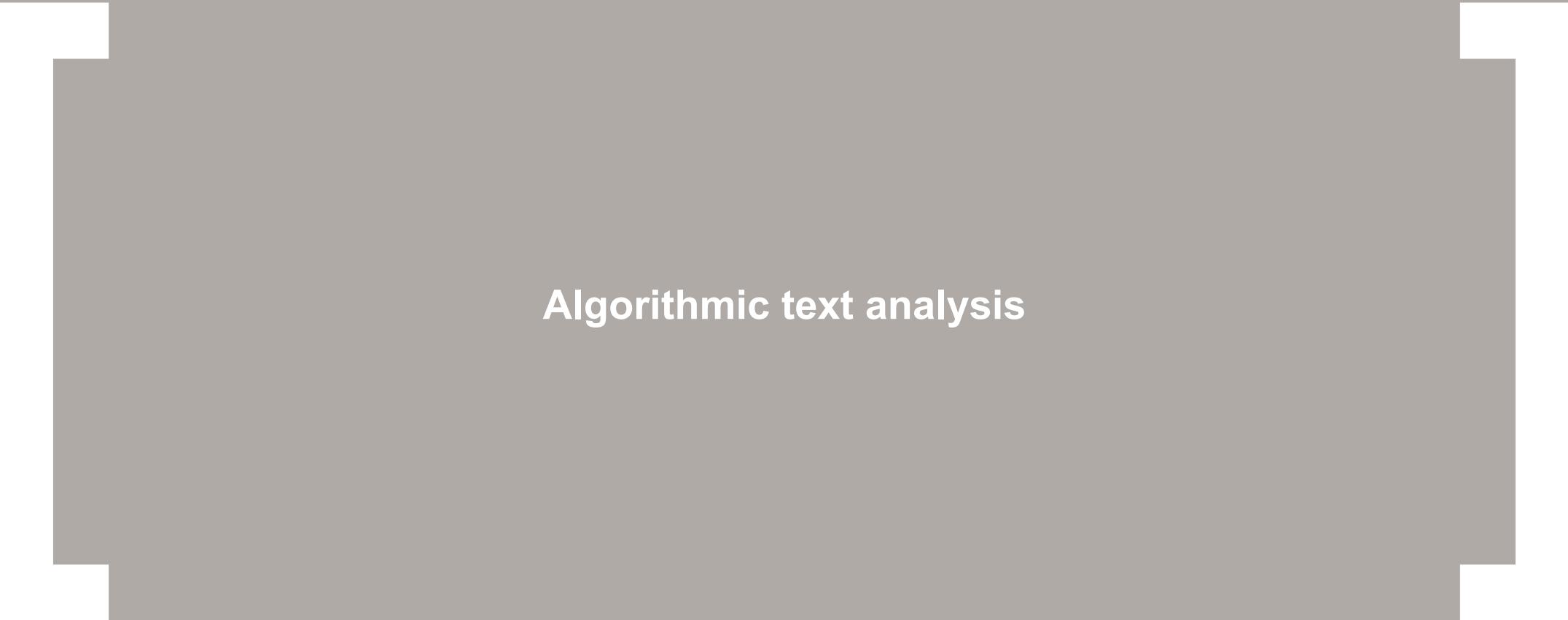
?

## Analyzing text to measure organizational culture

- You are a team of People Analytics experts just hired by Chimera to conduct a cultural assessment at the company. Before starting the job, you want to get a sense of what is going on with their org culture.
- You collect all the reviews that Chimera employees have posted on the website Glassdoor.com over time to understand the organizational culture at the firm.
- Use the spreadsheet containing the Chimera employees reviews (sheet “Pros”) and spend the next 15’ with your group to answer the following questions:
  1. What is the Chimera culture about (i.e., which values and beliefs the employees have)?
  2. Do employees agree in which values and beliefs they share?
- Be ready to discuss the two points with the rest of the class.

## Results

1. What is the Chimera culture about (i.e., which values and beliefs the employees have)?
  - ...
  
2. Do employees agree in which values and beliefs they share?
  - ...



**Algorithmic text analysis**

## Using algorithms instead

### ALGORITHMICALLY ASSISTED TEXT ANALYSIS

DICTIONARY  
METHODS

MACHINE LEARNING [for Natural Language  
Processing: NLP]

SUPERVISED ML

UNSUPERVISED ML

Topic  
Modelling

Embedding  
Models

## Dimensionality reduction as the main objective

- **High dimensionality** of documents **needs to be reduced**
- **Document:** list of words that create meaning based on their unique order

Document A

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

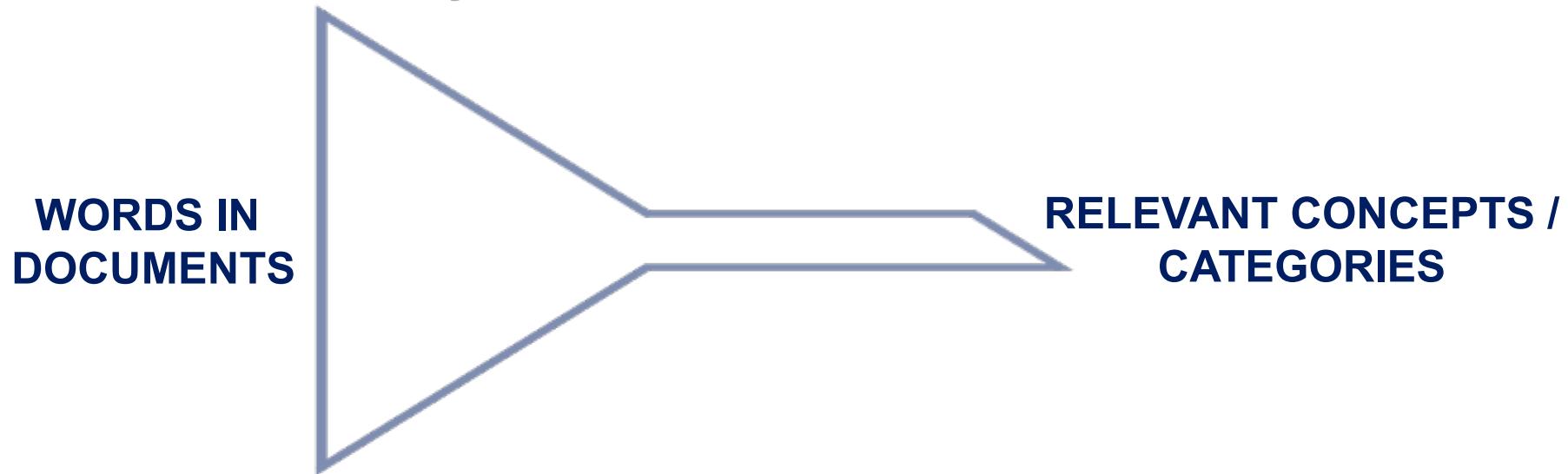


but *how?*

Document B

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## How to perform dimensionality reduction of text?



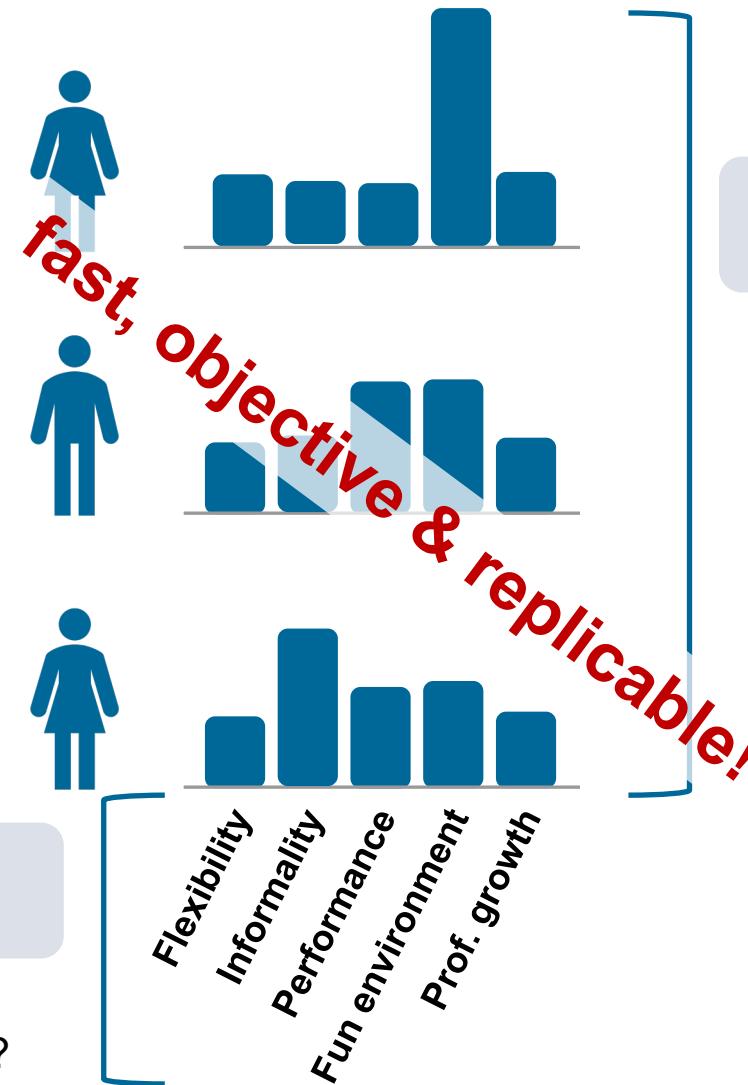
- Dimensionality reduction traditionally performed by “**manual coding**” (as in qualitative studies), e.g.,
  - Read each document to flag specific content
  - Read across all documents to identify relevant shared categories
- **Limitations:**
  - Not well suited for **large corpora** (time consuming and prone to errors)
  - Heavy **dependence on researcher pre-existing knowledge** of context
- Process can be **algorithmically assisted**

# Objective of using NLP techniques

Home		Insert	Draw	Page Layout	Formulas	Data	Review	View	Tell me
Paste	Cut	Format Cells	Text Box	Page Setup	Font	Font Size	Font Color	Font Style	Share
B	I	U	Text Box	Page Setup	Font	Font Size	Font Color	Font Style	Share
A1	X	V	Wrap Text	General	Conditional Formatting	Format as Table	Cell Styles	Insert	Delete
Review text-pros									
1	Review text-pros								
2	You will be working with the most talented ppl around.								
3	Freedoms and responsibility. You're treated like an adult. You're part of a pro team that is highly functioning. The company is well respected and has a super positive brand awareness -- I'm never able to go anywhere without getting peppered with raves and happy customer comments, and I love wearing company logo gear when I'm out for this reason.								
4	You matter at Chimera. There is no dead wood, everyone is doing something very important to the company or they wouldn't be there. What you are doing matters and makes a difference, and that feels good. You can really make a difference no matter what your role.								
5	The company often repeats that as an employee, you're not on an olympic team -- meaning, you don't do this 24/7 with no life, with all focus on the gold. We're more of a varsity team, playing to win, very good, but still have to balance with classes and learning. We need to have a good work/life balance and it's very important to the company that we do.								
6	I think the best thing - and subtle - is this no vacation/no holiday stuff. You work when you work, they seem to realize that everyone is working hard and all the time, nights, weekends often, a ton of hard work, so if you want to take time off for a vacation, a long holiday, a day off, whatever - it happens at your own discretion, you don't ask, you don't get permission, you just do what you have to. No one is keeping track. I've never heard of this kind of policy and you cannot imagine what it does for your morale, for feeling like you're being treated like a grown up. For that rule alone I think it stands beyond any other org.								
7	Great colleagues -- incredible really								
8	The upper management of Chimera really does seem to want to take a different approach with how they handle their employees, and accepts the fact that as an adult, you're able to get your job done. They appear to be very with it. The call center attempts to echo that sentiment, and provides many perks that make work an alright place to be. From the free coffee and refreshments in the cafeteria to the free rental plan, great benefits, and competitive pay, there is a great ideal that the company attempts to live up to - that the people populating the center are worth the extra buck or two. It was refreshing to work for a company that was capable of identifying problems, and instead of sweeping them under the rug, took the initiative to change the problem.								
9	The people there are fantastic, the service is great, the facilities are terrific as well as comfortable. Some of the perks were worth it (free Chimera). The training team is great, the support staff are incredible, all very hard working people, the representatives on the floor are, without a doubt, the best in the industry. And who doesn't want to work "in the movies"? The pay is pretty decent, the benefits are acceptable.								
10	Benefits are terrific and I have my perfect shift. Free movies are always wonderful. This is the perfect job for me because it's really low stress and it is a perfect avenue for my love of film, film history and general film knowledge.								
11	Transparent corporate culture and opportunities to learn about the new exciting development in the industry.								
12	You will be surrounded by brilliant, competent, mature, hard working peers. You will be challenged to exceed your past accomplishments. It's a big company that strives to maintain a small company feel. Office politics are 99% absent; you are expected to focus on your work and you can expect from your peers that you will not be distracted by gossip or personality clashes. Chimera is a place that is very aware of its culture and everyone inside works hard to maintain it. Everybody you know outside of work loves Chimera, and they will let you know when it comes up that you work there.								
13	Innovation, intelligent people, incredible brand, incredible business model								
14	Fast paced, dynamic, not afraid to try something different to make the company better. That means not everything is a win, but the company is honest about what hasn't worked out and doesn't hold on because of ego or prior investment.								
15	My favorite thing about Chimera is being surrounded by a ridiculously talented group of people who are also fun to work with. Management provides a clear compass about the priorities of the company, and that makes it easy for teams to prioritize their projects appropriately. The next few years are going to be really interesting, with the company making the transition from a really good DVD by mail service to a great way to get movies by mail or over the internet.								
16	You are surrounded by very bright people focused on creating a great product for our subscribers and a great environment for employees. It's exciting to be on the cutting-edge of media and technology. The company offers a lot of freedom, and the responsibility that comes with it. Employees are treated like the adults they are.								
17	From the top down, my colleagues at Chimera are among the most talented individuals in their class of trade. There is not a set vacation policy other than "take it when you need it". Maybe the best reason of all is the communication. Management goes to great lengths to make sure that everyone in the organization has a full understanding of the opportunities and challenges facing the business, as well as the strategic vision that the company is executing against.								
18	Great projects, high performance culture, incredible people to work with								
19	Smart, passionate, fun loving colleagues that keep the mood light and the pace fast. Flat organizations rule. Great facilities - clean, professional, uncluttered, with plenty of space and sunshine. A place where 'forgiveness' is valued over 'permission'. A place where 'take risks and recover quickly' is lived every day. Customers are raving fans of our product.								

## Culture Content

What is the culture about?



## Culture Configuration



Cultural Strength  
[0,1]

Do employees agree in which values and beliefs they share?



**BAYES**  
BUSINESS SCHOOL  
CITY UNIVERSITY OF LONDON



Topic modeling

# Data generation & algorithmic estimation



## How is a document created by an individual?

“What topics are important for me to write about?”

Topic distribution per document

“Which words to describe a topic?”

Word distribution per topic



## LDA Topic Modeling for parameter estimation

Determines which topics are discussed in each document

Which topics per document & in which proportion?

Sorts words into topics

Which words associated with which topic?

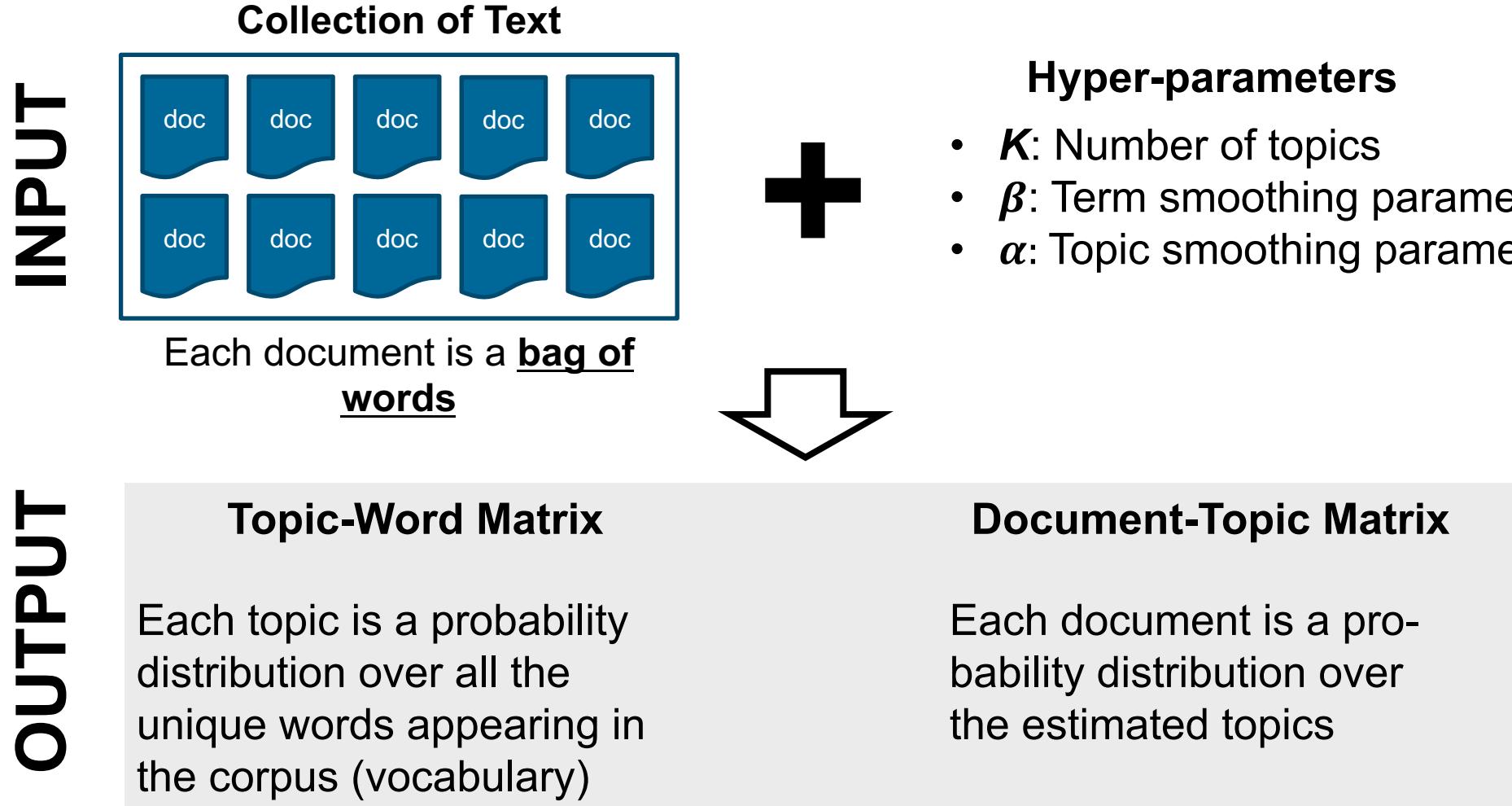
### Cultural Configuration

Do employees agree about which values and beliefs they share?

### Cultural Content

What is the culture about?

# Input and Output



## What is a topic?

- Any language construct that can be signified with a set of words
- For example:
  - Cultural values/norms
  - Technological or scientific domains/paths
  - Product/Market/Industry categories

## Available algorithms

- Latent Dirichlet Allocation (LDA)
- Hierarchical Topic Modeling
- Dynamics Topic Modeling
- Correlated Topic Modeling

## Available algorithms

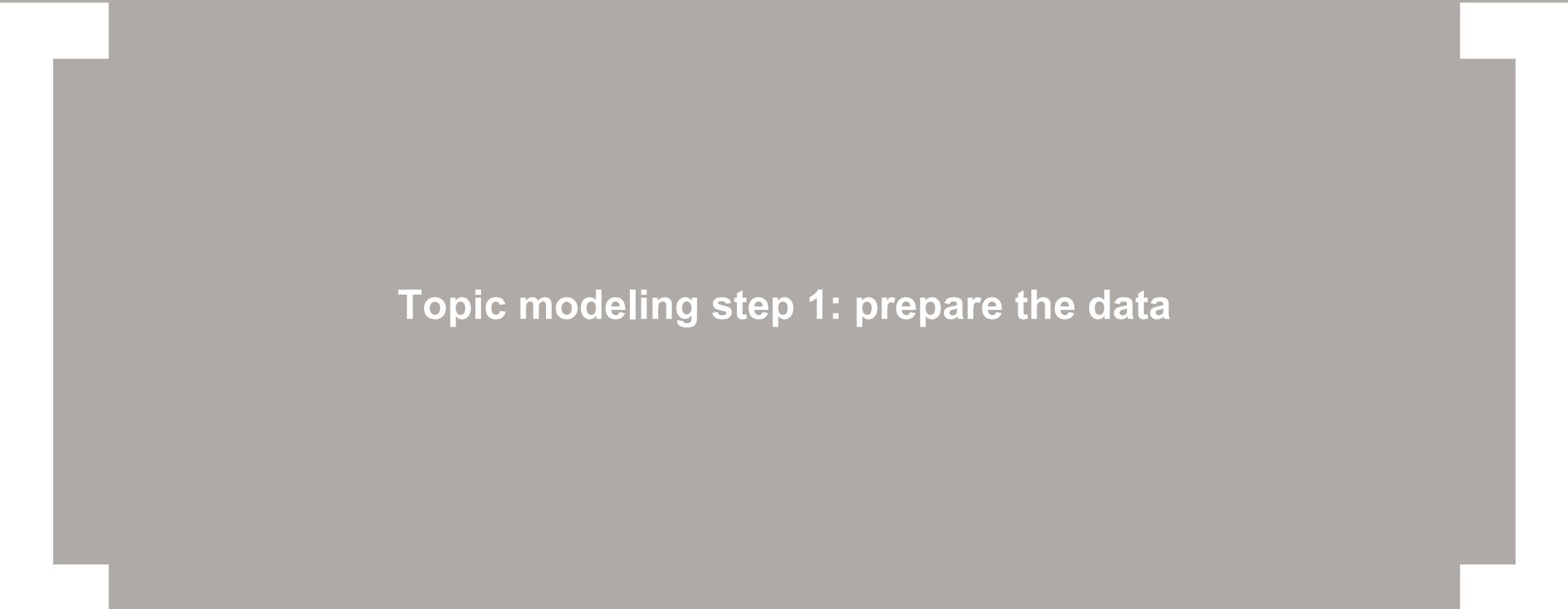
- **Latent Dirichlet Allocation (LDA)**

The most common method with minimum distributional assumptions

The focus of today's exercise

Python: the most versatile and advanced implementations

- Hierarchical Topic Modeling
- Dynamics Topic Modeling
- Correlated Topic Modeling



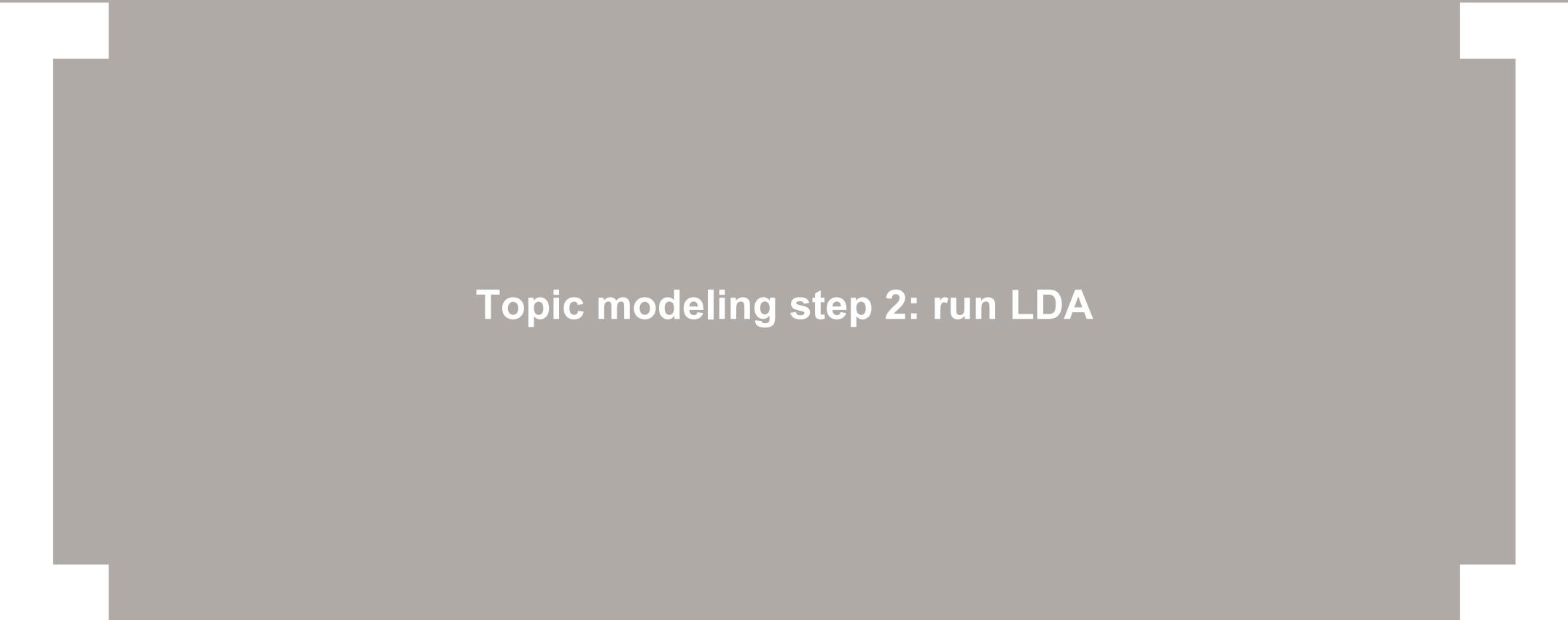
**Topic modeling step 1: prepare the data**

## Data collection and preparation

- **Collect the data**
- Put the data in the **right format** for the tool of your choice
- **Discard stop words:** language-specific stop word lists available via python libraries
- **Discard frequent / uninformative words**
- **Stem**, i.e., remove text dimensionality and reduce word to its root (e.g., “*families*”, “*family*”, “*familiar*” → “*famili*”)

Let's see this in practice





**Topic modeling step 2: run LDA**

## How to set the hyperparameters?

Need to perform **sensitivity analyses** of your results with different values of hyperparameters

$\alpha$ : Topic smoothing parameter

$\beta$ : Term smoothing parameter

$K$ : Number of topics

A range of approaches to tune  $K$

Manual tuning: trade-off meaning and nuance

Perplexity minimization: usually for predictive purposes, less for interpretation

Coherence maximization: usually for descriptive purposes

Let's run an LDA topic modeling algorithm



# Understanding LDA outputs

## TOPIC-WORD MATRIX

$$\Theta \in \mathbb{R}^{K \times V}$$

Columns = WORDS

Rows = TOPICS

$$\begin{bmatrix} \delta_{11} & \cdots & \delta_{1V} \\ \vdots & \ddots & \vdots \\ \delta_{K1} & \cdots & \delta_{KV} \end{bmatrix}$$

## DOCUMENT-TOPIC MATRIX

$$\Theta \in \mathbb{R}^{N \times K}$$

Columns = TOPICS

Rows = DOCUMENT

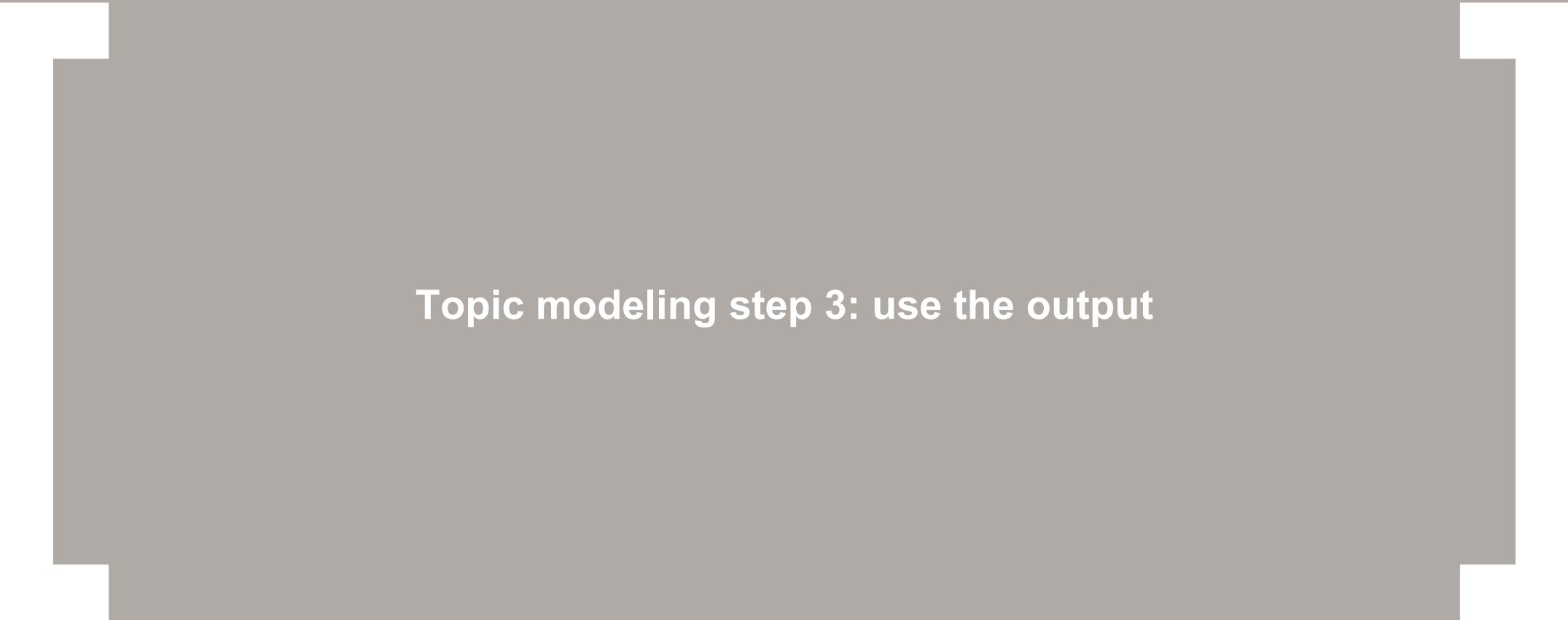
$$\begin{bmatrix} \vartheta_{11} & \cdots & \vartheta_{1K} \\ \vdots & \ddots & \vdots \\ \vartheta_{N1} & \cdots & \vartheta_{NK} \end{bmatrix}$$

- Shows the frequency of each term in each topic
- The top 10-20 may be used to interpret the topic
- Some topics may capture garbage

- Shows the weight of each topic in each document
- Each row (document) sums up to 1
- Some documents may signify a topic very strongly, some may represent multiple topics loosely

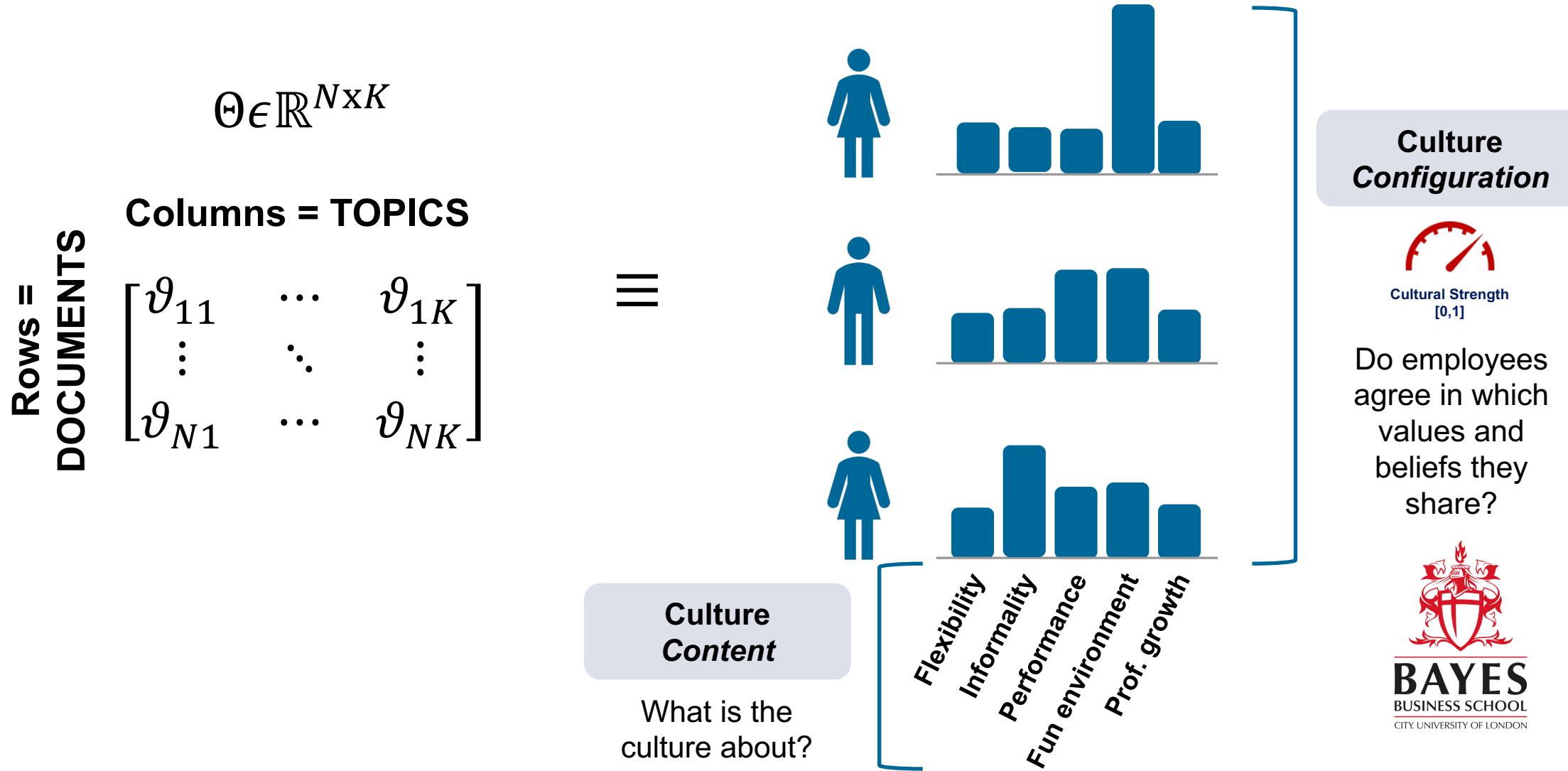
Let's see the output in our example





**Topic modeling step 3: use the output**

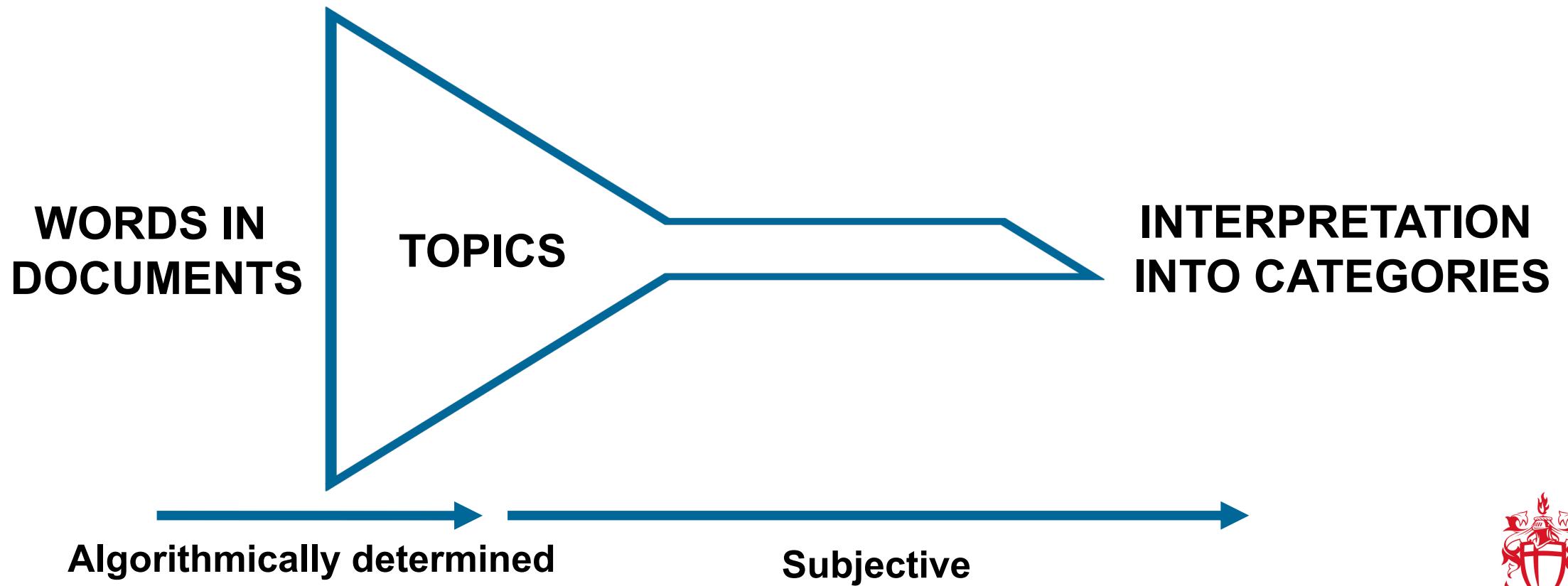
## Using the output of the LDA algorithm





**Topic modeling step 3.1: culture content**

## Interpreting cultural content



## “Manual” interpretation using the topic-over-word matrix

Topic	Words	What is it about?
1	environ, time, flexibl, technolog, new, hour, friendli, alway, chang, engin	<b>Work environment</b>
2	work, great, compani, peopl, manag, team, cultur, place, job, product	<b>Generic praise</b>
3	life, balanc, understand, mind, engag, attitud, explor, perfect, u, situat	<b>Work-life balance</b>
4	start, everyon, global, meet, cowork, will, system, share, though, across	<b>Collaborative culture</b>
5	good, pay, nice, growth, train, market, top, perform, process, organ	<b>Generic</b>
6	benefit, employe, salari, industri, compens, care, competit, health, better, program	<b>Compensation</b>
7	busi, within, advanc, atmospher, option, respect, plenti, stock, insur, gym	<b>Perks</b>
8	depend, locat, build, field, gain, hand, real, strategi, requir, fresher	?
9	show, less, much, need, still, reward, success, full, month, term	?
10	get, make, like, also, day, go, take, sale, everi, even	?
11	free, 2, 1, 3, 4, facil, polici, 5, food, week	<b>Numbers</b>
12	home, offic, schedul, etc, travel, class, peer, futur, may, seem	<b>Work schedule</b>
13	year, servic, profession, know, vacat, tri, famili, think, abl, money	<b>Generic</b>
14	opportun, lot, learn, project, career, intern, differ, train, support, interest	<b>Career growth</b>



Let's take a look at the “interpretation burden”



# Interpreting cultural is not that easy

The root causes of interpretation problems:

- **Synonyms:** “*cat*” and “*kitty*”

Lead to **redundant topics** s.t. some contain “*cat*”, others “*kitty*”

- **Homonyms:** “*bank*” (river) and “*bank*” (financial institution)

**Polysemy:** same words belong to multiple topics

Lead to **uninterpretable topics**, that contains words related to both “*river*” and “*cash*”

## Alternative approach based on “prototypical text”

### Which topics to interpret through prototypical text?

- The ones that are important in the corpus
- If each topic was equally important to describe the corpus, it would be present in ratio  $1/K$  in each document

Rank order topics in terms of how many documents they have  $>1/K$  presence in  
Pick the top  $K^*$  topics based on a scree plot ("best value for money")

review	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10
1	0.1	0.1	0.3	0.00	0.00	0.2	0.2	0.05	0.0	0.05
2	0.2	0.0	0.4	0.00	0.00	0.3	0.0	0.00	0.1	0.00
3	0.1	0.1	0.2	0.05	0.05	0.2	0.1	0.20	0.0	0.00
4	0.0	0.0	0.6	0.10	0.00	0.3	0.0	0.00	0.0	0.00



Which topics to interpret?

## Alternative approach based on “prototypical text”

**Which documents are prototypical for a topic?**

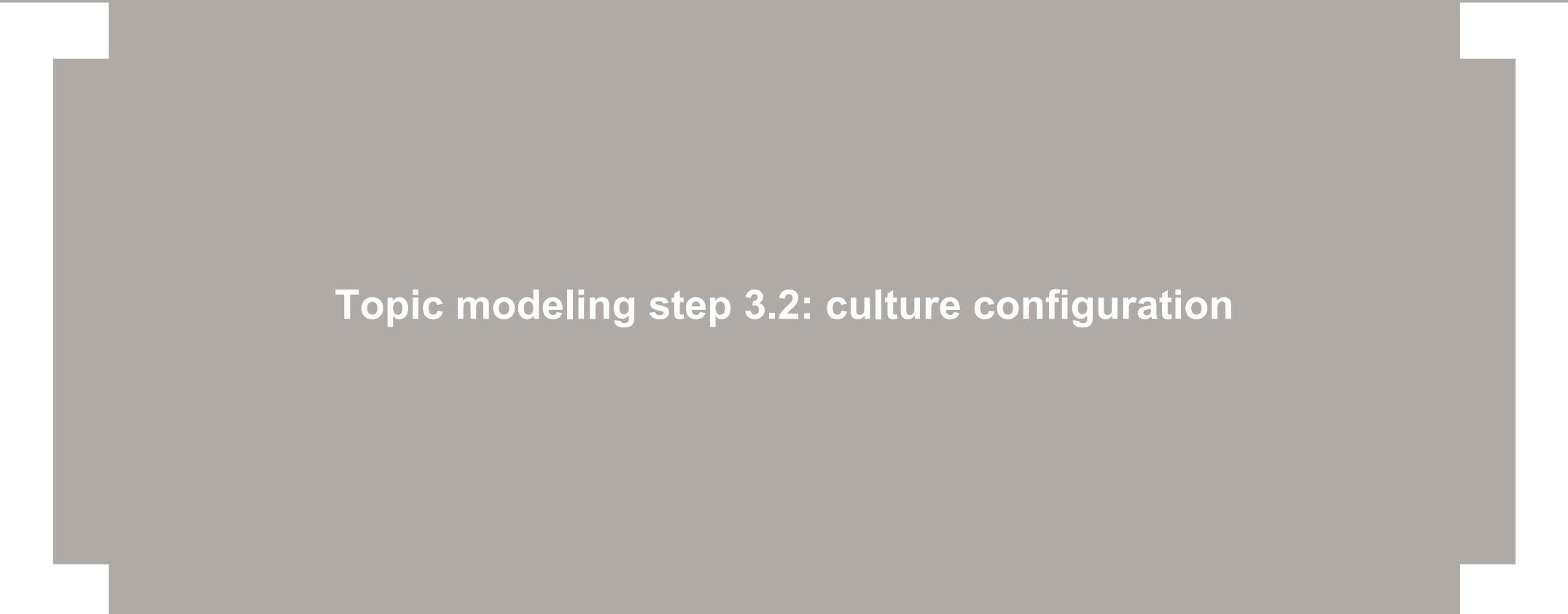
- Ideally one that contains **just one topic**
- If a topic constitutes x% of each document, we would need to read at least **1/x documents** to find common theme.

review	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10
1	0.1	0.1	0.3	0.00	0.00	0.2	0.2	0.05	0.0	0.05
2	0.2	0.0	0.4	0.00	0.00	0.3	0.0	0.00	0.1	0.00
3	0.1	0.1	0.2	0.05	0.05	0.2	0.1	0.20	0.0	0.00
4	0.0	0.0	0.6	0.10	0.00	0.3	0.0	0.00	0.0	0.00

Which reviews  
to focus on for  
interpretation?

Let's use instead prototypical reviews to interpret topics





**Topic modeling step 3.2: culture configuration**

# Using the document-over-topic matrix to measure cultural configuration

review	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10
1	0.1	0.1	0.3	0.00	0.00	0.2	0.2	0.05	0.0	0.05
2	0.2	0.0	0.4	0.00	0.00	0.3	0.0	0.00	0.1	0.00
3	0.1	0.1	0.2	0.05	0.05	0.2	0.1	0.20	0.0	0.00
4	0.0	0.0	0.6	0.10	0.00	0.3	0.0	0.00	0.0	0.00

**Individual similarity:**

Average reviews' (rows') distance to proxy for cultural similarity

**Individual focus:**  
Average focus of individual reviews (rows) on few vs. many topics to proxy for cultural focus

$$AJS = \frac{\sum_i \sum_j JS(\vartheta_i, \vartheta_j)}{D(D-1)} \quad \forall i, j \text{ s.t. } i < j$$

$$AHI = \frac{\sum_{i=1}^D \sum_{k=1}^K (\vartheta_i^k)^2}{D}$$



**Do we believe the findings?**

## Do we believe the findings?

How do we figure out if our analysis provides an accurate picture  
of the organizational culture?

# The Netflix culture

## Seeking Excellence

### Dream Team

### Freedom and Responsibility



1. encourage independent decision-making by employees
2. share information openly, broadly, and deliberately
3. are extraordinarily candid with each other
4. keep only our highly effective people
5. avoid rules

### Informed Captains

**Judgment:** You make wise decisions despite ambiguity. You identify root causes, and get beyond treating symptoms. You think strategically, and can articulate what you are, and are not, trying to do. You are good at using data to inform your intuition. You make decisions based on the long term, not near term.

**Communication:** You are concise and articulate in speech and writing. You listen well and seek to understand before reacting. You maintain calm poise in stressful situations to draw out the clearest thinking. You adapt your communication style to work well with people from around the world who may not share your native language. You provide candid, helpful, timely feedback to colleagues.

**Curiosity:** You learn rapidly and eagerly. You contribute effectively outside of your specialty. You make connections that others miss. You seek to understand our members around the world, and how we entertain them. You seek alternate perspectives.

**Courage:** You say what you think, when it's in the best interest of Netflix, even if it is uncomfortable. You make tough decisions without agonizing. You take smart risks and are open to possible failure. You question actions inconsistent with our values. You are able to be vulnerable, in search of truth.

**Passion:** You inspire others with your thirst for excellence. You care intensely about our members and Netflix's success. You are tenacious and optimistic. You are quietly confident and openly humble.

## Highly Aligned, Loosely Coupled

### Context Not Control

### Disagree Openly

**Selflessness:** You seek what is best for Netflix, rather than what is best for yourself or your group. You are open-minded in search of great ideas. You make time to help colleagues. You share information openly and proactively.

**Innovation:** You create new ideas that prove useful. You re-conceptualize issues to discover solutions to hard problems. You challenge prevailing assumptions, and suggest better approaches. You keep us nimble by minimizing complexity and finding time to simplify. You thrive on change.

**Inclusion:** You collaborate effectively with people of diverse backgrounds and cultures. You nurture and embrace differing perspectives to make better decisions. You are curious about how our different backgrounds affect us at work, rather than pretending they don't affect us. You recognize we all have biases, and work to grow past them. You intervene if someone else is being marginalized.

**Integrity:** You are known for candor, authenticity, transparency, and being non-political. You only say things about fellow employees that you say to their face. You admit mistakes freely and openly. You treat people with respect independent of their status or disagreement with you

**Impact:** You accomplish amazing amounts of important work. You demonstrate consistently strong performance so colleagues can rely upon you. You make your colleagues better. You focus on results over process.

# Matching our findings to the publicized culture

Netflix values	Categories identified by the authors through topic interpretation [“pros” reviews section]
<b>Context not control</b>	Performance-oriented
<b>Disagree openly</b>	—
	Growth opportunities
<b>Dream team</b>	Intelligent co-workers
	Talented workforce
	Freedom & responsibility
<b>Freedom and responsibility</b>	Freedom
	Good work conditions
<b>Highly aligned, loosely coupled</b>	Flexible
<b>Informed captains</b>	Good management
<b>Seeking excellence</b>	Innovative
	Good pay
<b>Not mapped onto Netflix values</b>	Good vacation policy
	Good pay, perks & benefits
	Good technology

Netflix values	Categories identified by the authors through topic interpretation [“cons” reviews section]
<b>Context not control</b>	Culture of fear
	Metric-driven environment
	Lack of structure
	Favoritism
	Limited growth opportunities
	Poor pay
	High work pressure
	Lack of consistency
<b>Disagree openly</b>	Toxic culture
	Lack of consistency
	Challenging work environment
<b>Dream team</b>	—
<b>Freedom and responsibility</b>	—
<b>Highly aligned, loosely coupled</b>	—
<b>Informed captains</b>	—
<b>Seeking excellence</b>	Metric-driven environment
	High work pressure
<b>Not mapped onto Netflix values</b>	Inadequate benefits
	Inadequate support to employees

## **Final remarks**

# The pros and cons of NLP on organizational text

## Pros

- Large-scale data (sample representativeness and almost “real time”)
- Higher representativeness of responses across hierarchical levels of organizations
- Possibility to uncover cultural dimensions that are unique to firms

## Cons

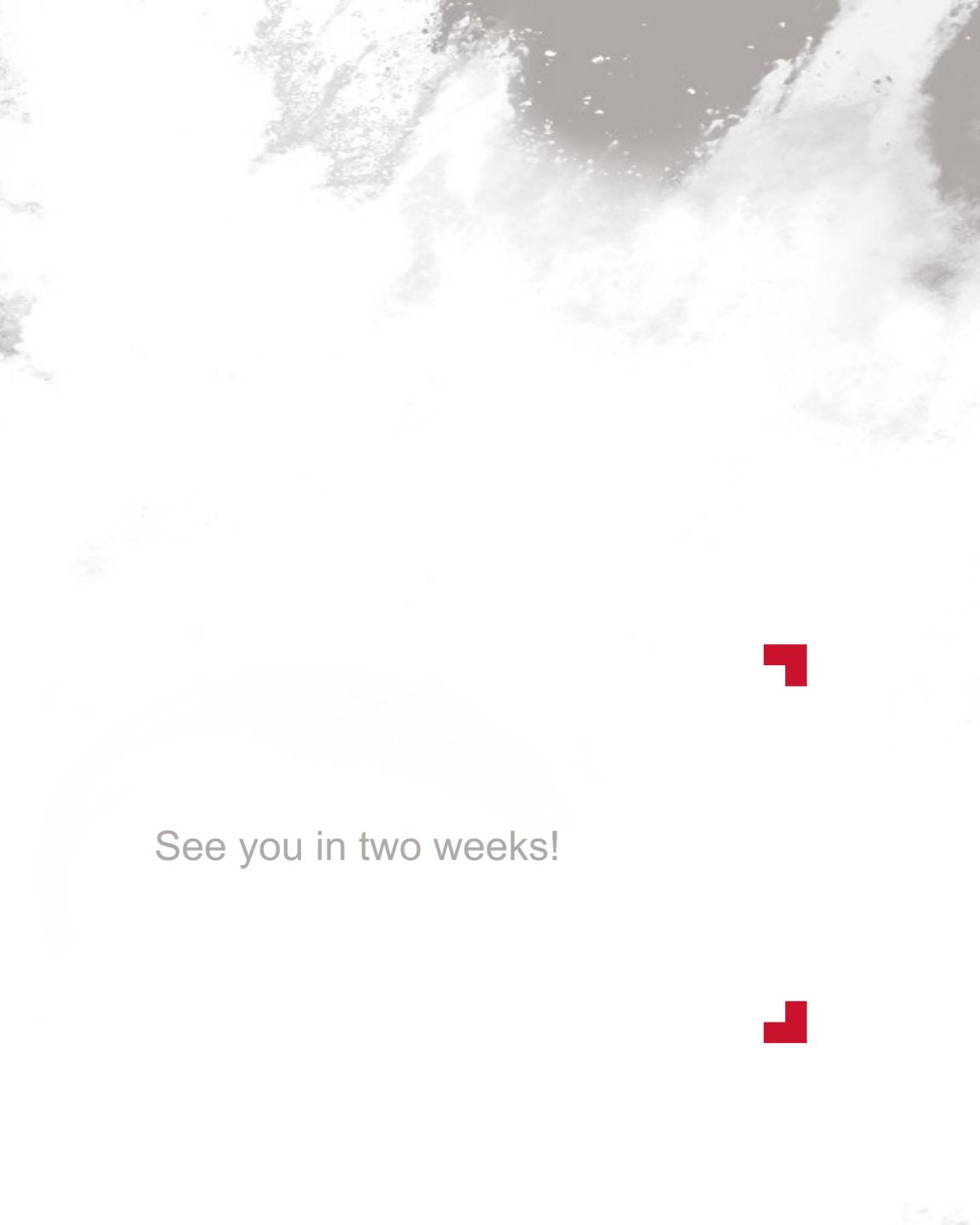
- How to parse out noise from data?
- How to calibrate findings in a way that is understandable by humans?

## Summary

- Organizational culture **matters for firms** because of implications for  
Individuals satisfaction, well-being & turnover  
Group dynamics  
Firm performance
- We have **multiple levers to design** (and maintain) an organizational culture
- **Data** useful to measure organizational culture  
Surveys / interviews  
Text produced at scale by organizational members
- **Methods** to measure organizational culture  
**Survey analysis**  
Unsupervised ML for **NLP**: topic modelling, word embedding (and sentiment analysis)  
An example of **topic modelling application**

## Useful reading

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning research*, 3(Jan), 993-1022.
- Corritore, M., Goldberg, A., & Srivastava, S. B. (2020). Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly*, 65(2), 359-394.
- Evans, J. A., & Aceves, P. (2016). Machine translation: mining text for social theory. *Annual Review of Sociology*, 42, 21-50.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 267-297.
- Marchetti, A., Puranam, P. (2020) Interpreting Topic Models Using Prototypical Text: From 'Telling' To 'Showing'. INSEAD Working Paper No. 2020/49/STR.  
<https://dx.doi.org/10.2139/ssrn.3717437>.
- All the articles in Poetics, Volume 41, Issue 6:  
<http://www.sciencedirect.com/science/journal/0304422X/41>



See you in two weeks!

