

EMPLOYEE REVIEWS

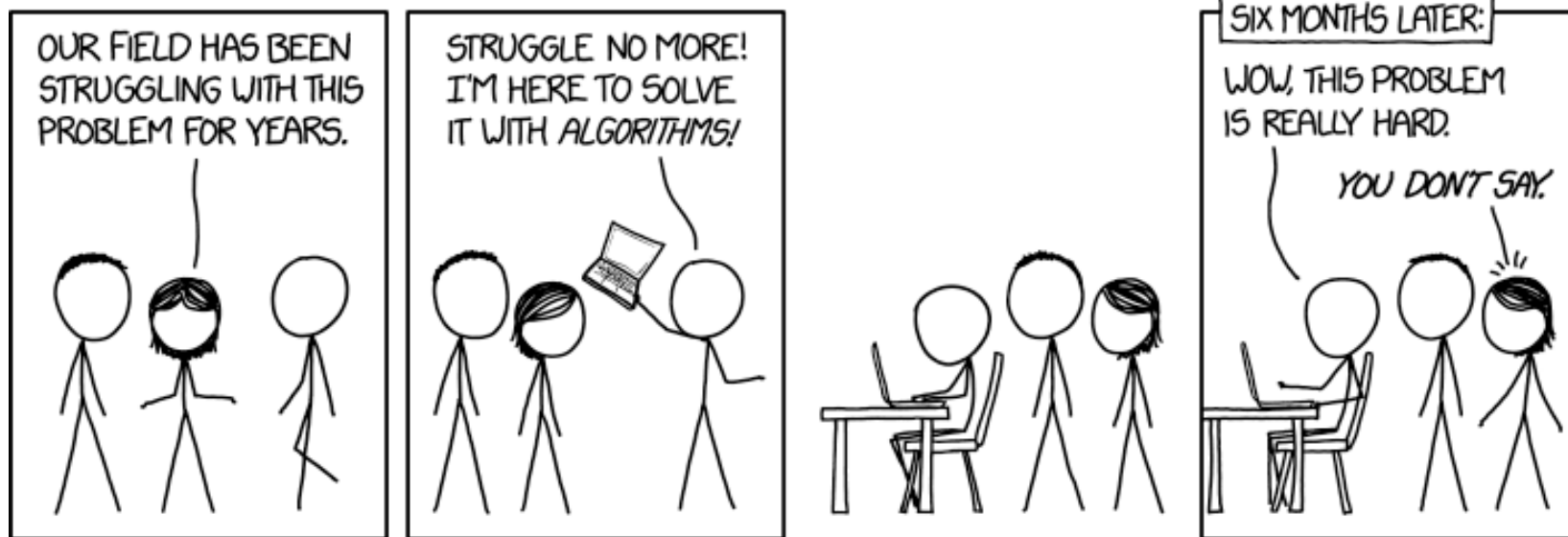
How to use Text Analytics to understand what they say

Andrew Marritt | Founder & CEO

andrew.marritt@organizationview.com | +41 78 905 0775

Specialist text analysis services

OrganizationView works with global 'sophisticated' clients to help them understand large volumes of employee (and increasingly customer) feedback.



Why text?

Text, analysed correctly can identify be a real driver for meaningful change. Our approach means we regularly find themes which are unlikely to have been included in a traditional survey.

From...

in the next survey

...To

“Change the music. It’s so moody and depressing and totally not our customer. We have had several complaints from customers and the team”

“Since the last survey, they have changed the music and it creates a better work place for staff and customers”

The Dataset - Pro Question

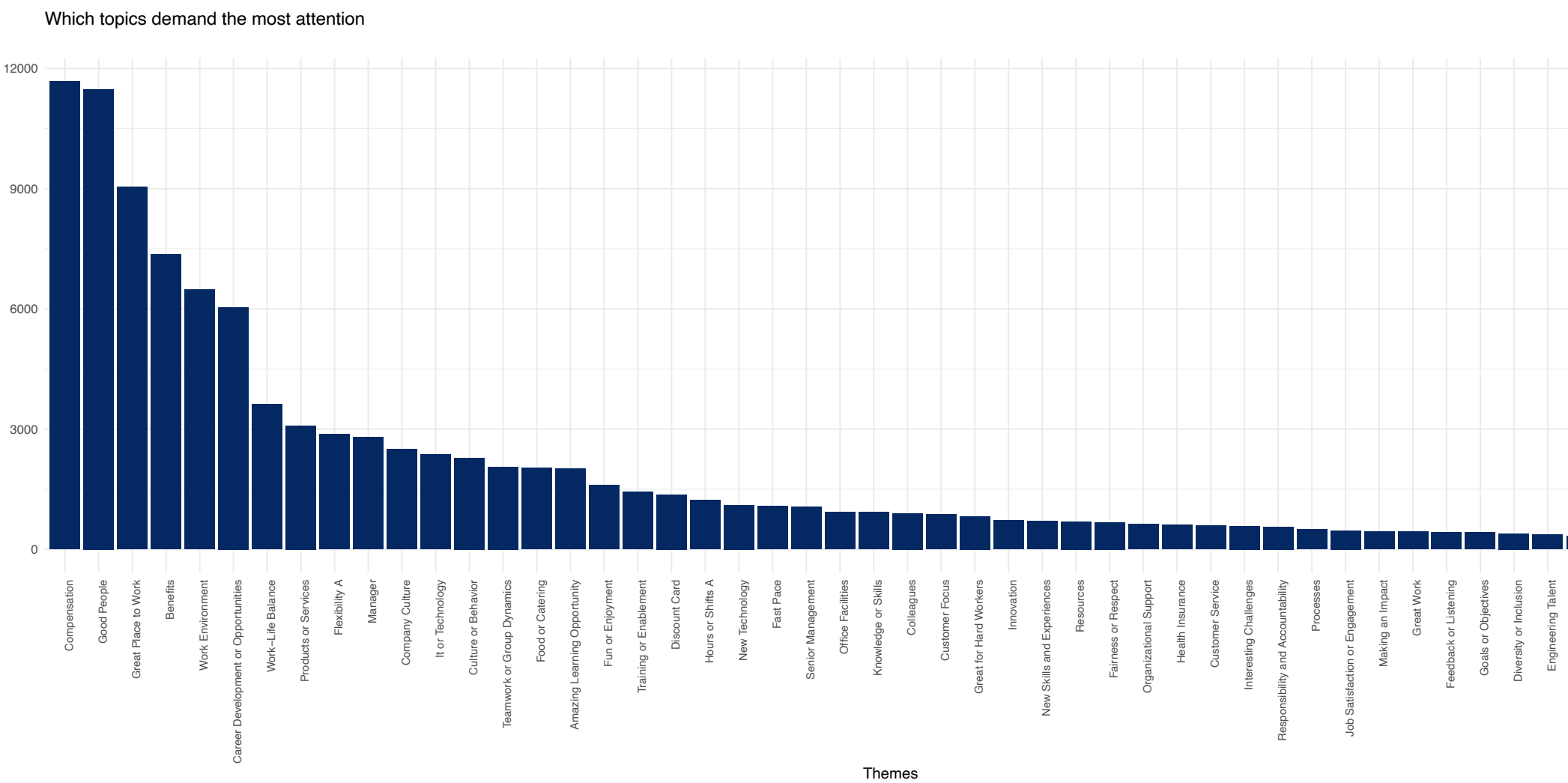
Number of Comments	Number of Themes	Number of Words	Mean Comment	Median Comment	Longest Comment
65,527	99	913,926	23.8 _{words}	14 _{words}	892 _{words}

DEMOGRAPHICS	TEXT	QUANT RATINGS
Company	Summary	Overall
Date	Pros	Work balance
Job Title	Cons	Culture / Values
	Advice to management	Snr Management...

COMPANIES					
Amazon	Apple	Facebook	Google	Microsoft	Netflix

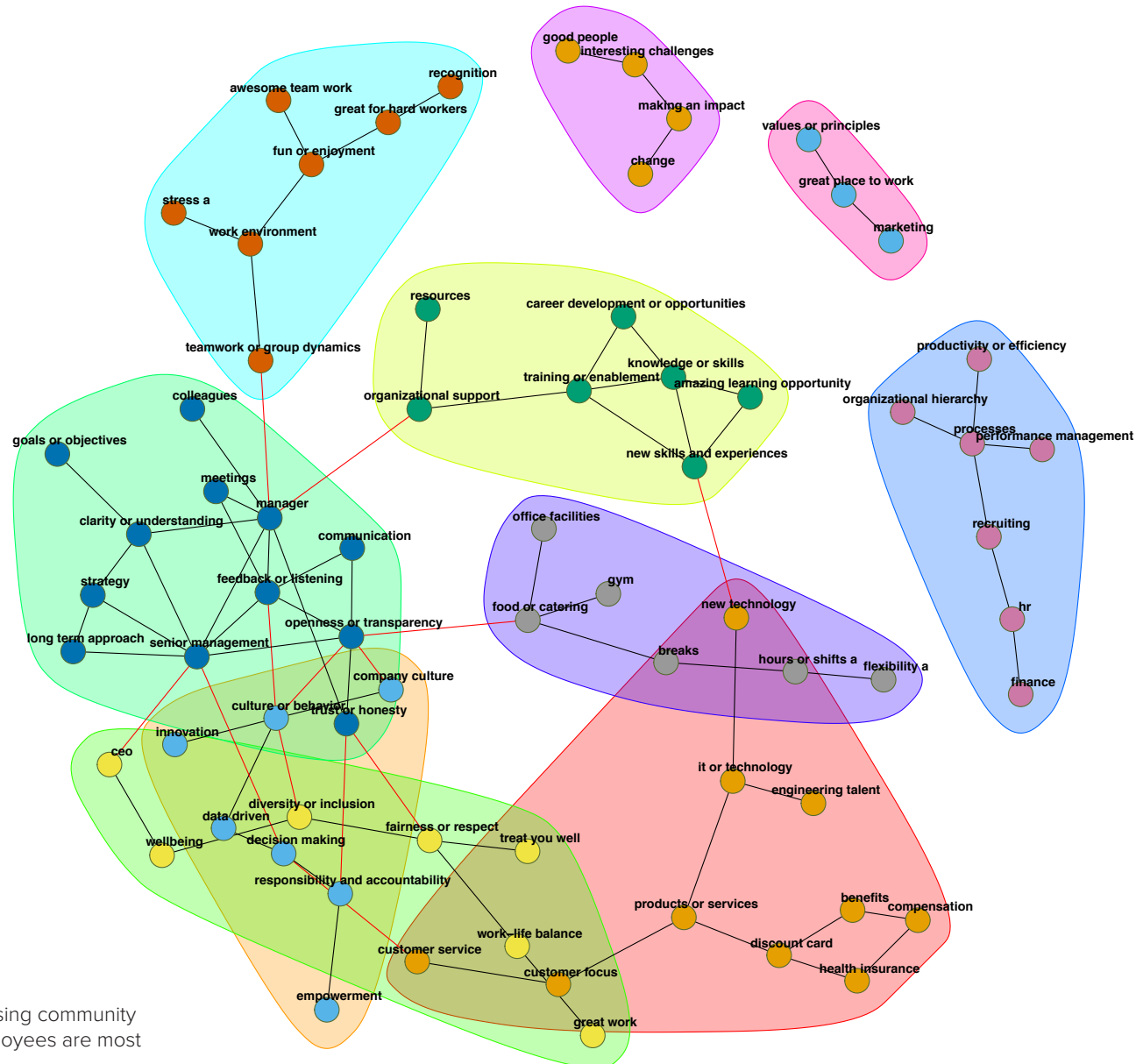
The topics in the dataset

The first place to start is looking at the overall topics



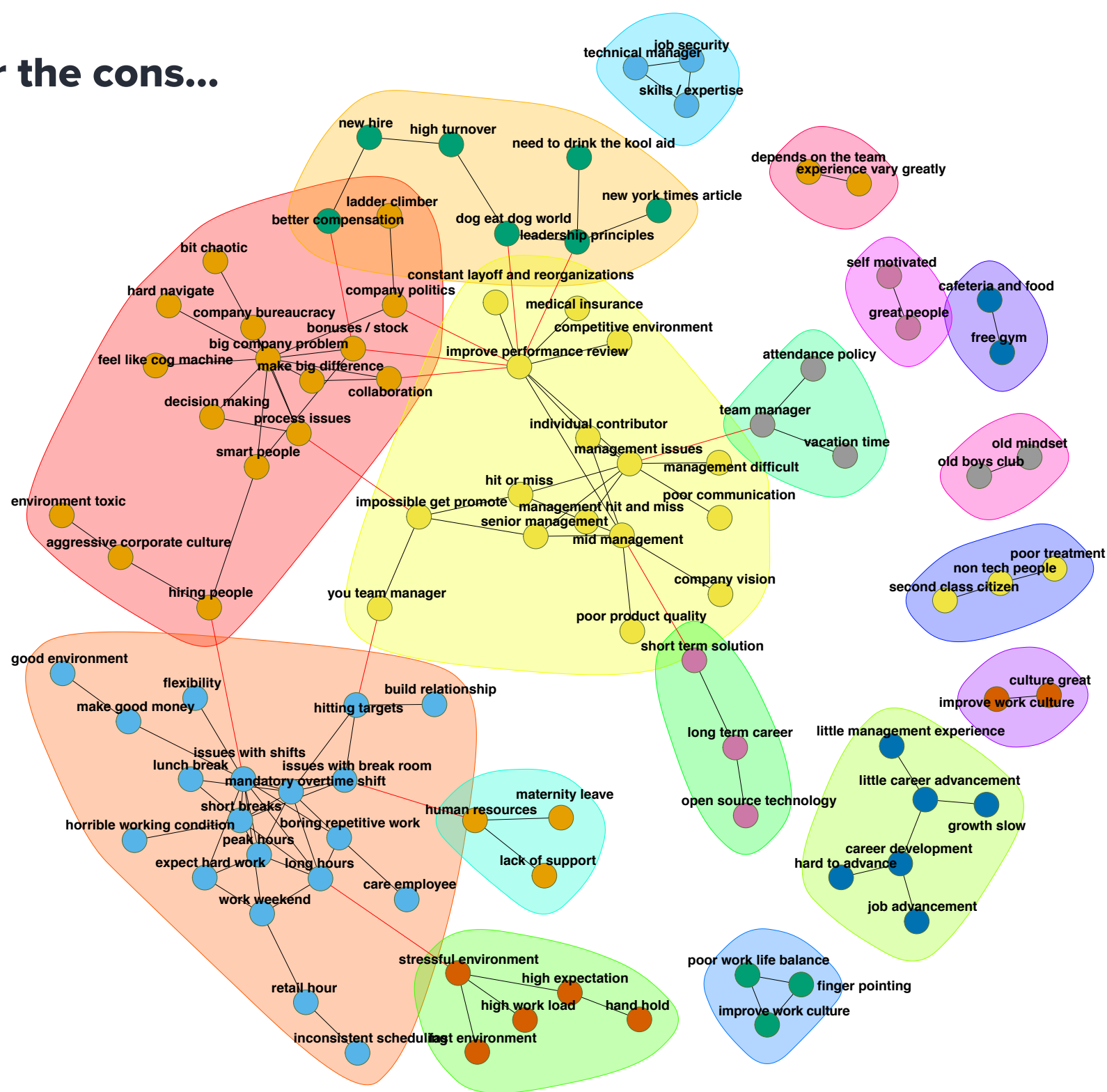
Clustering the comments by cooccurrence

Clustering of comments is a good first-step to developing personas

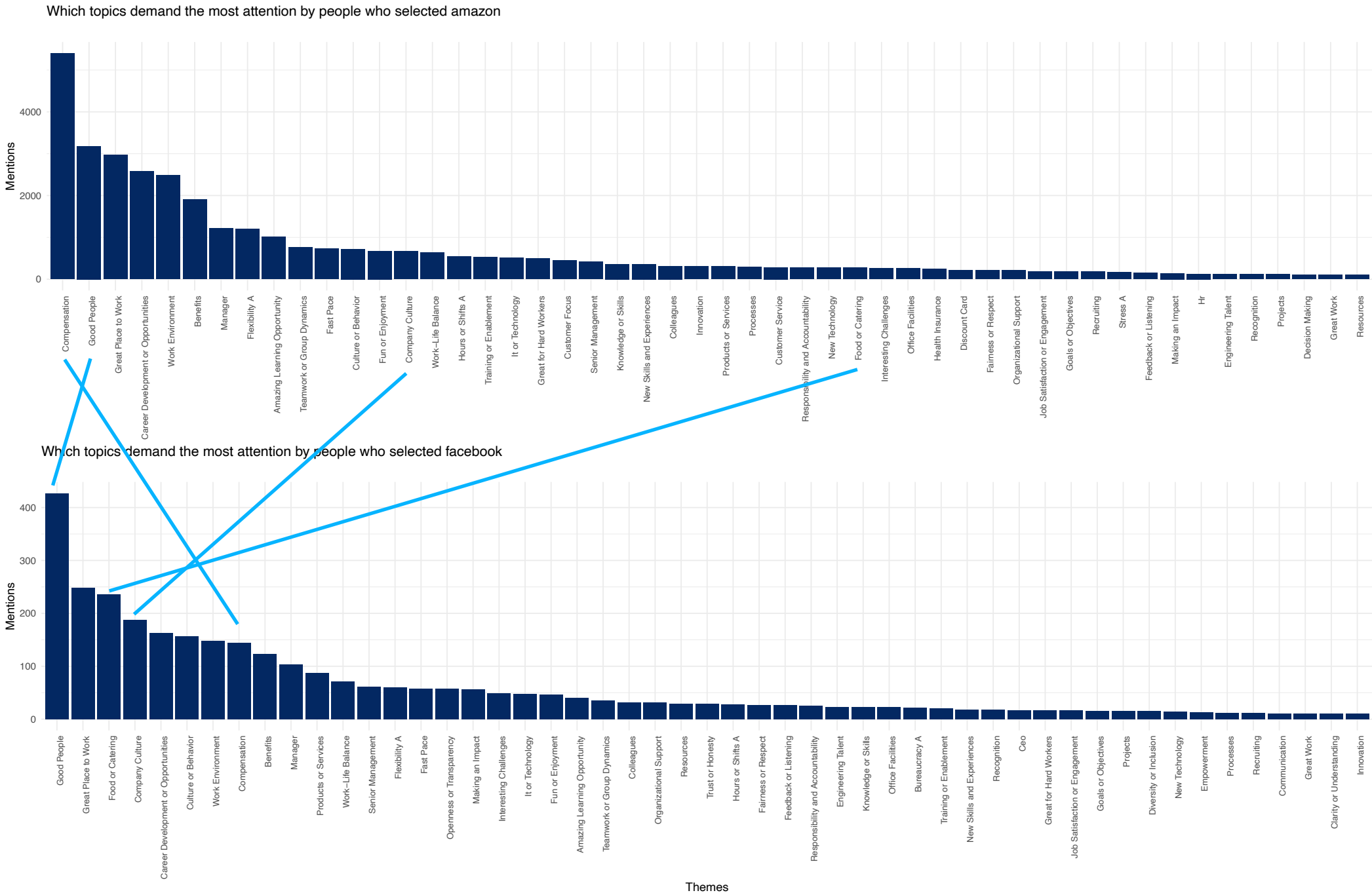


A probabilistic co-occurrence network using community detection to identify which themes employees are most likely to talk about together.

Same for the cons...

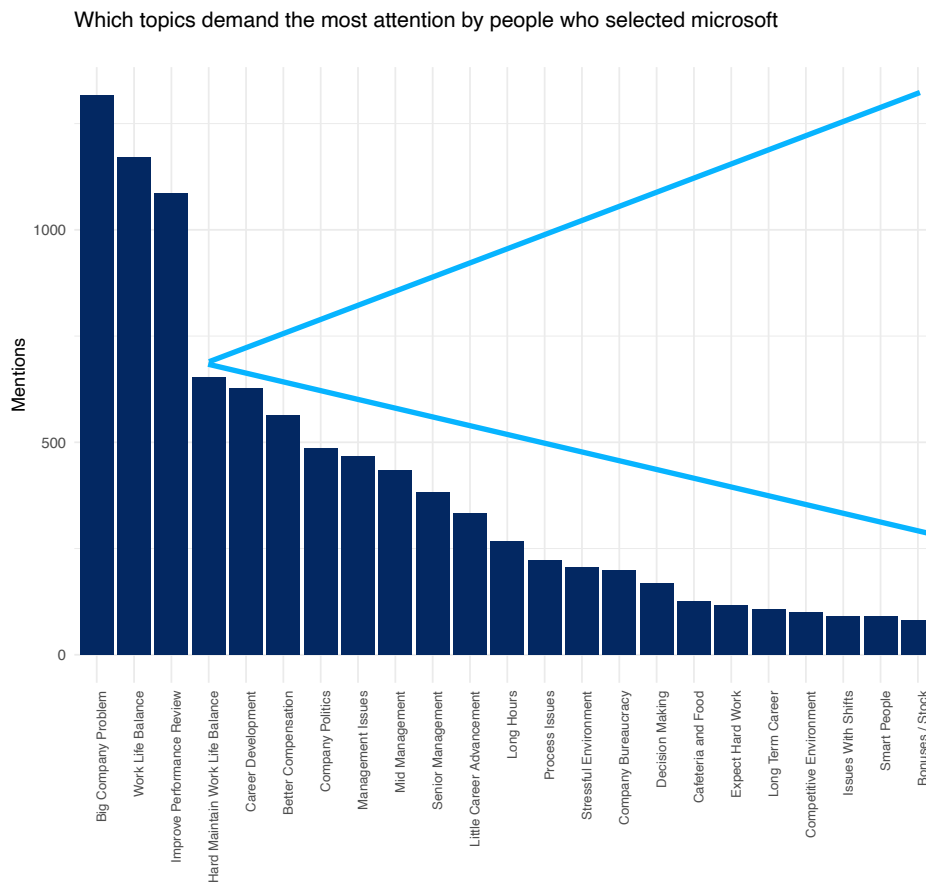


Comparing companies



Spotting externally recognised problems

The case of Microsoft and the forced-ranking



THE VERGE

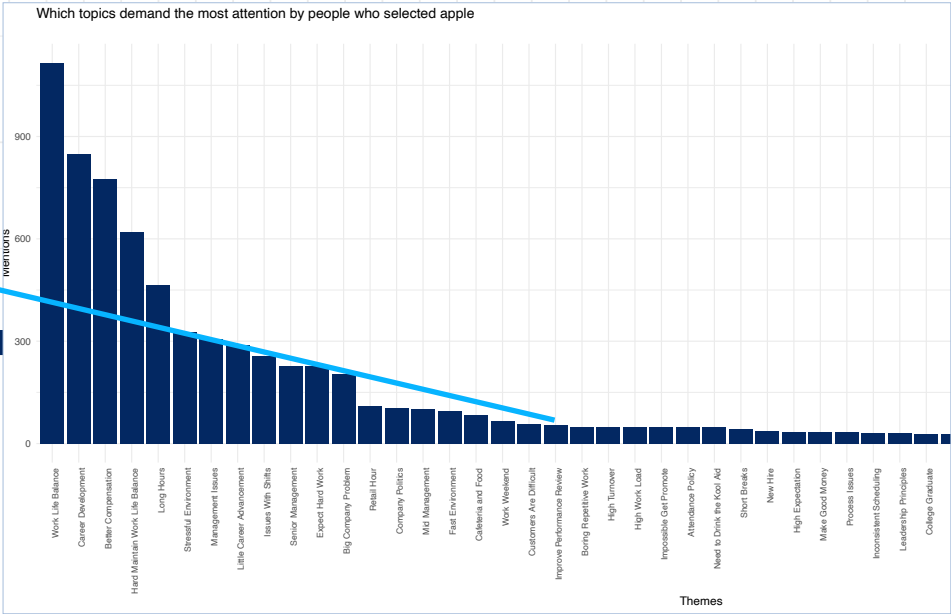
TECH ▾REVIEWS ▾SCIENCE ▾CREATORS ▾ENTERTAINMENT ▾VIDEO ▾MORE ▾

MICROSOFT / BREAKING / BUSINESS

Microsoft axes its controversial employee-ranking system

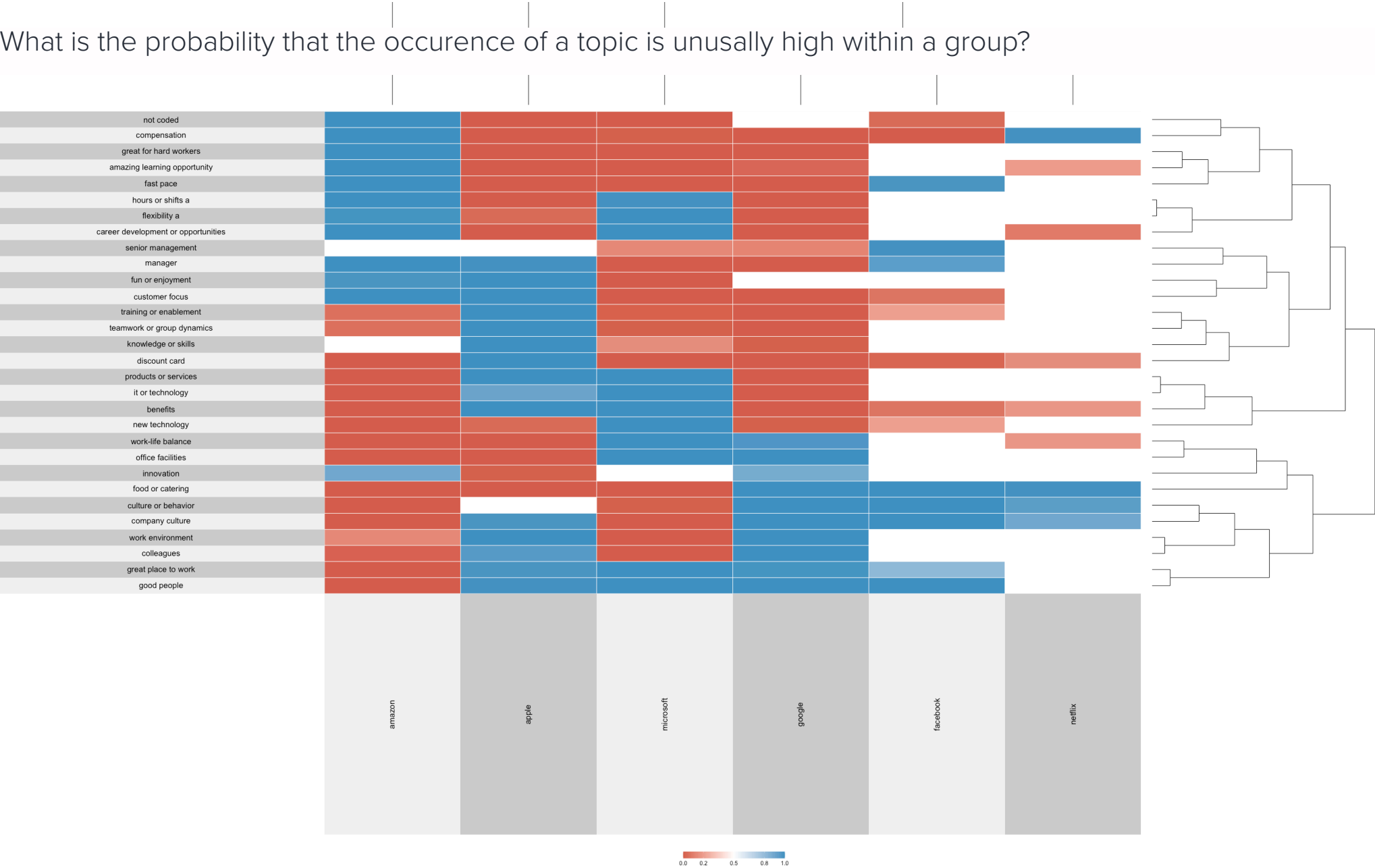
By Tom Warren | @tomwarren | Nov 12, 2013, 12:53pm EST

SHARE



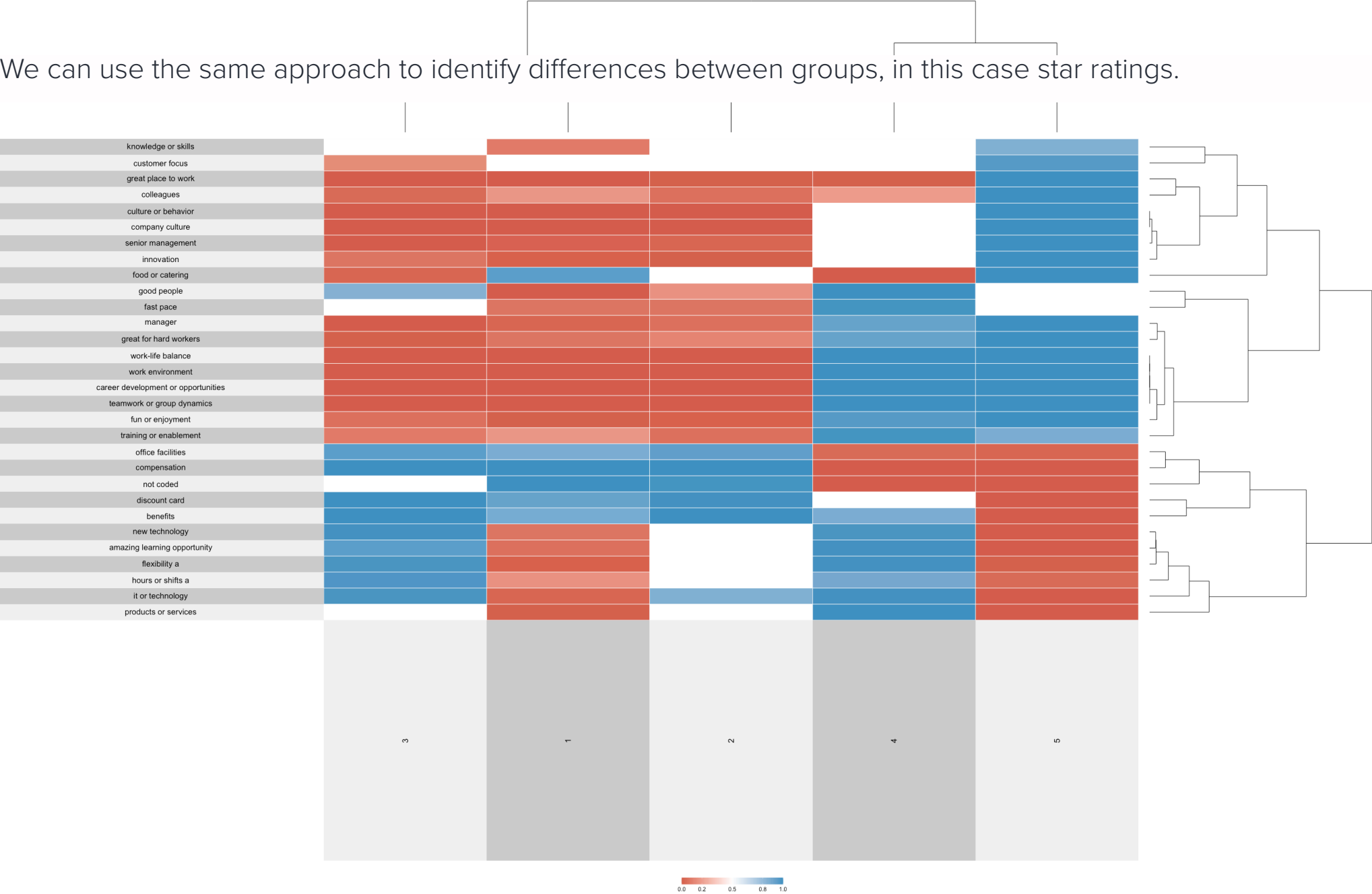
Using a probabilistic model to identify differences

What is the probability that the occurrence of a topic is unusually high within a group?



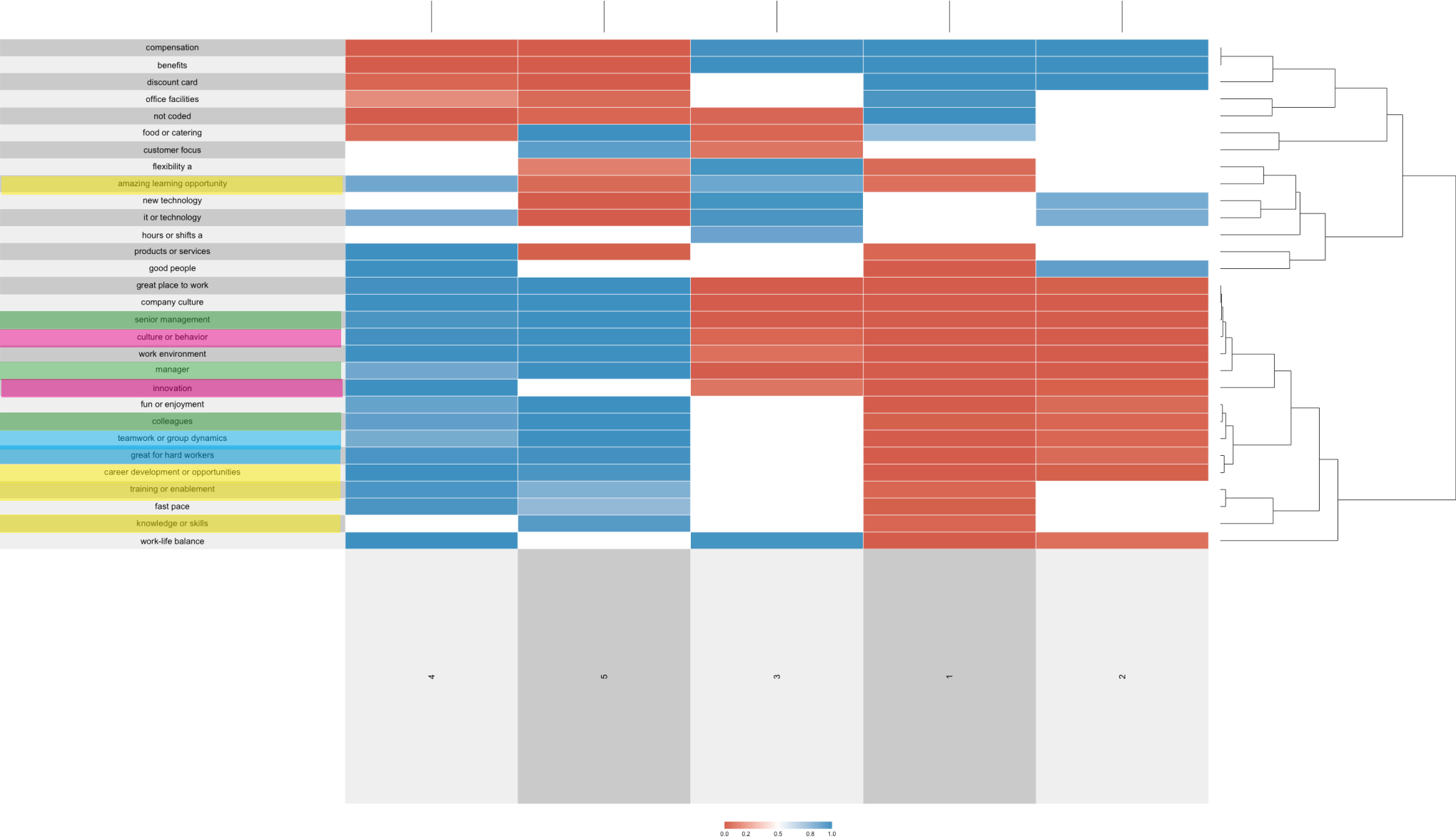
Looking at topics by the overall rating group

We can use the same approach to identify differences between groups, in this case star ratings.



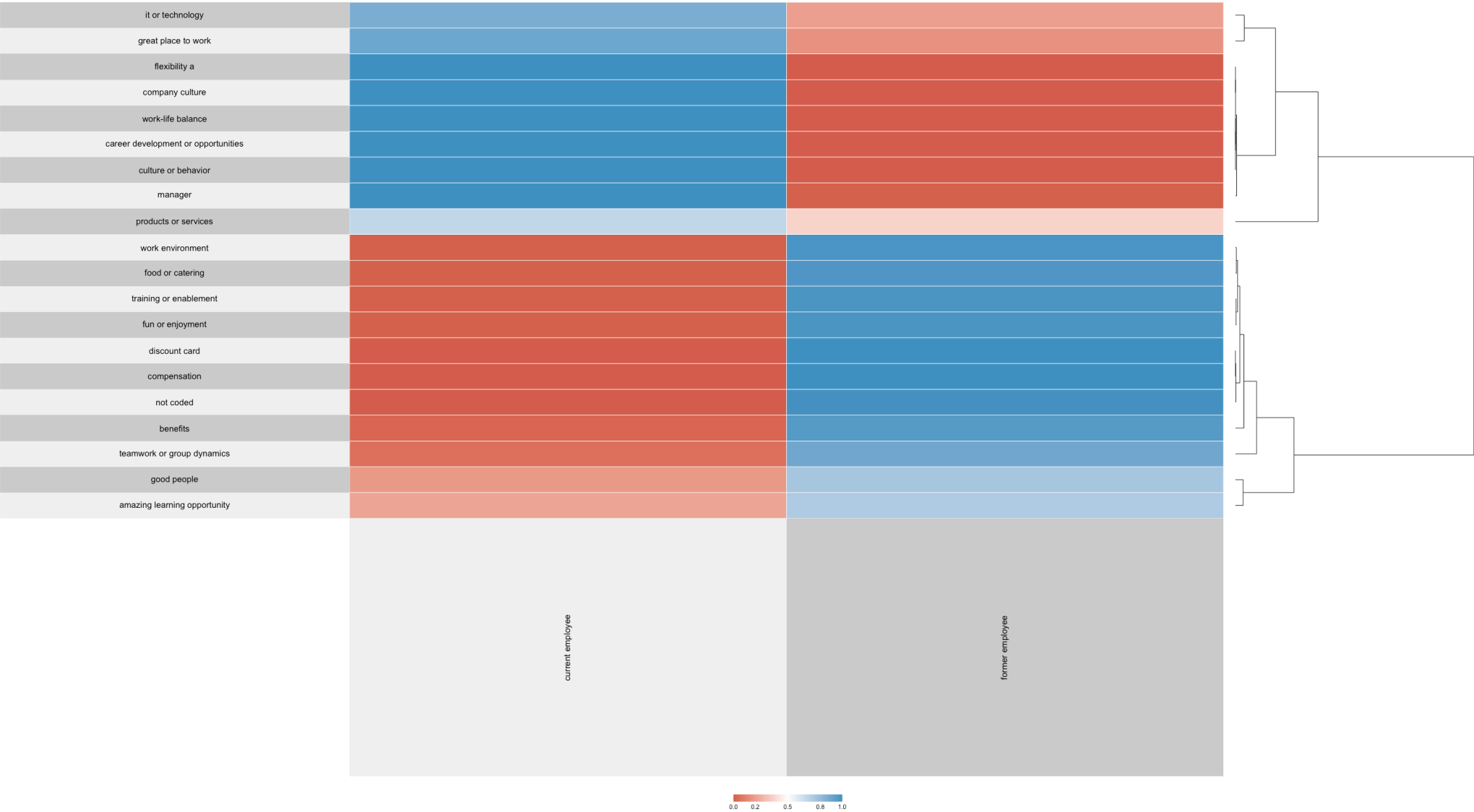
What about the other star categories? Management

Naturally topics related to management are rated for those who rate management as good.



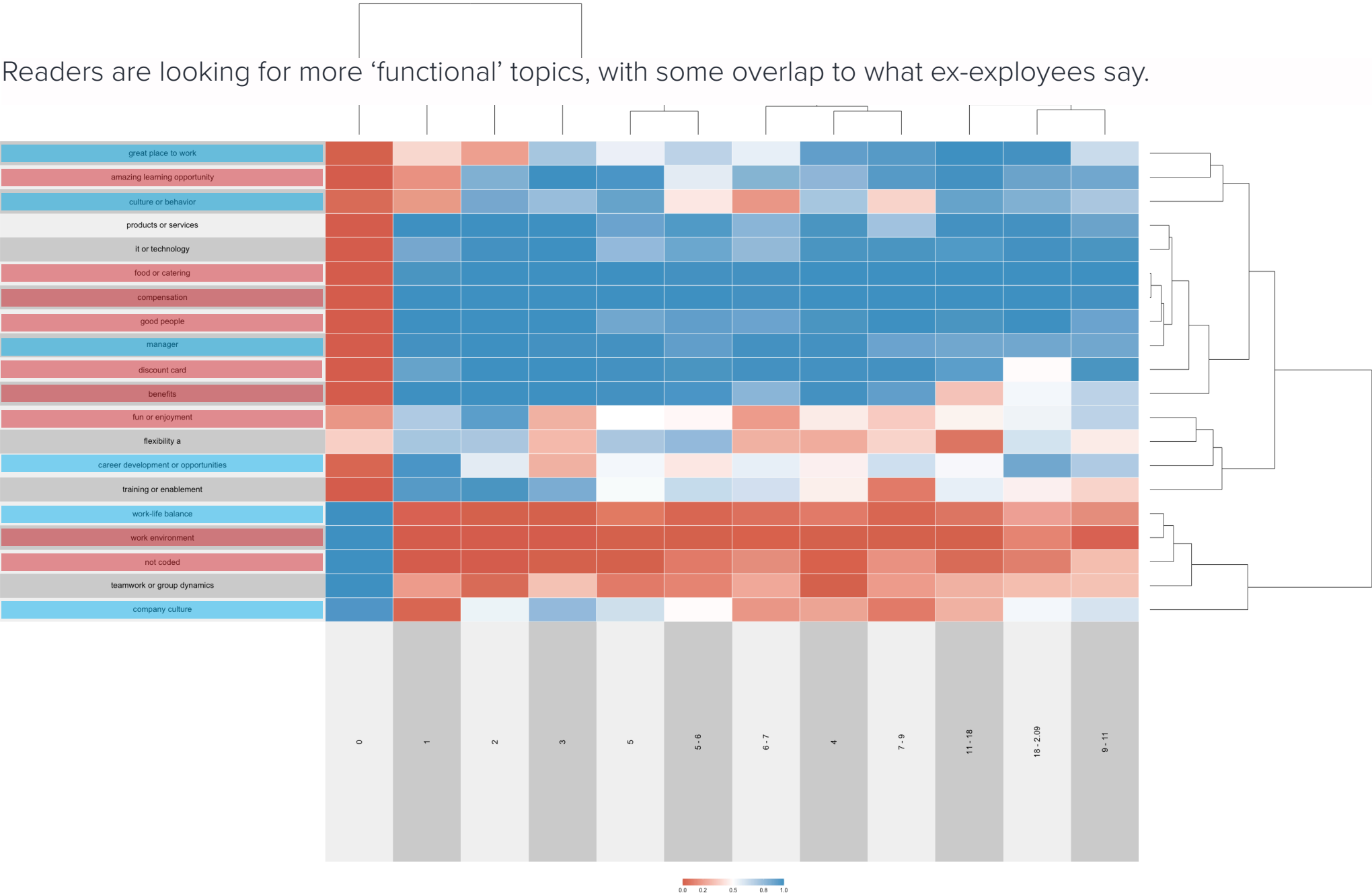
Where do current and former employees differ?

What is the probability that the occurrence of a topic is unusually high within a group?

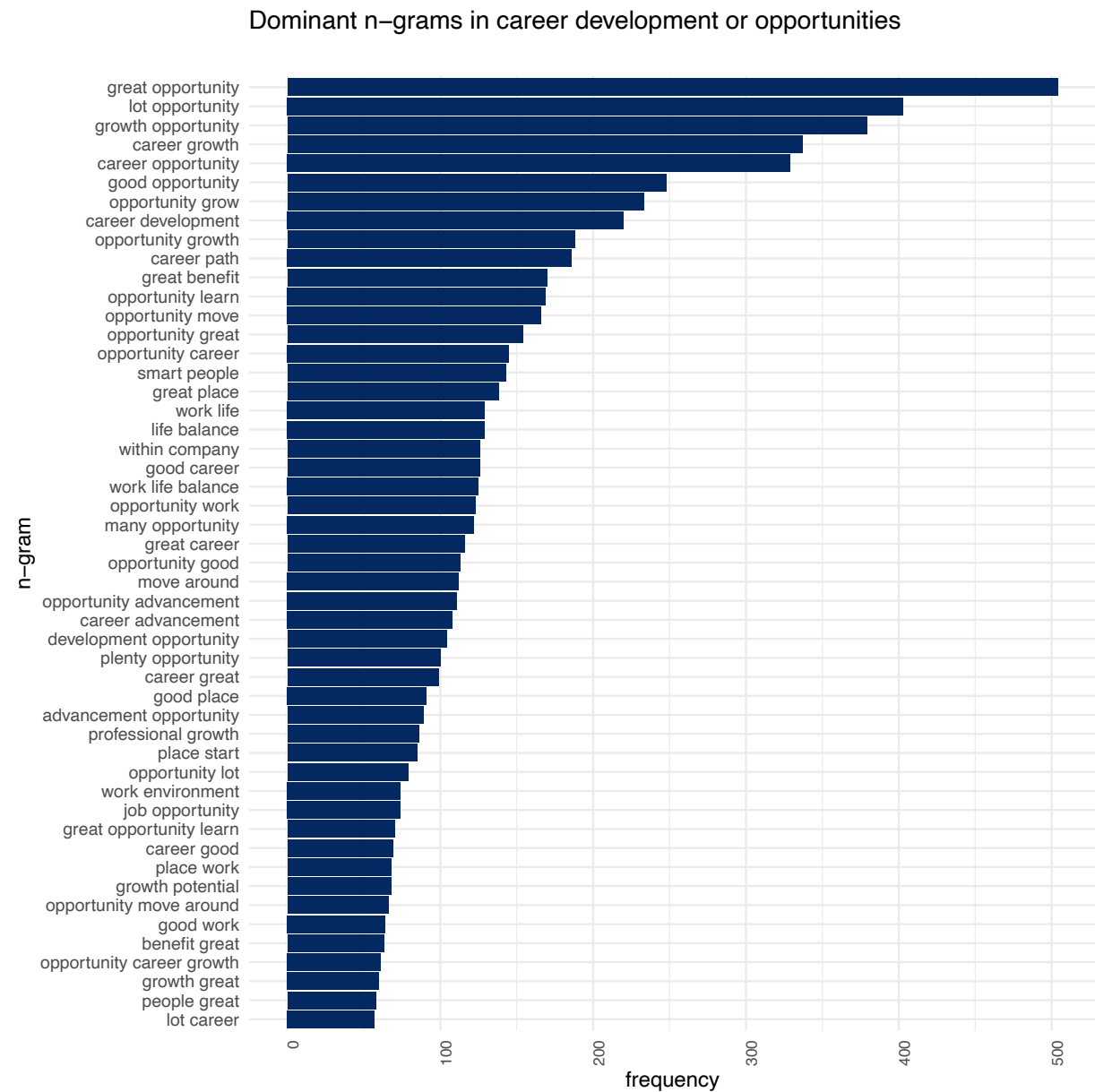


What differentiates a useful comment?

Readers are looking for more ‘functional’ topics, with some overlap to what ex-employees say.

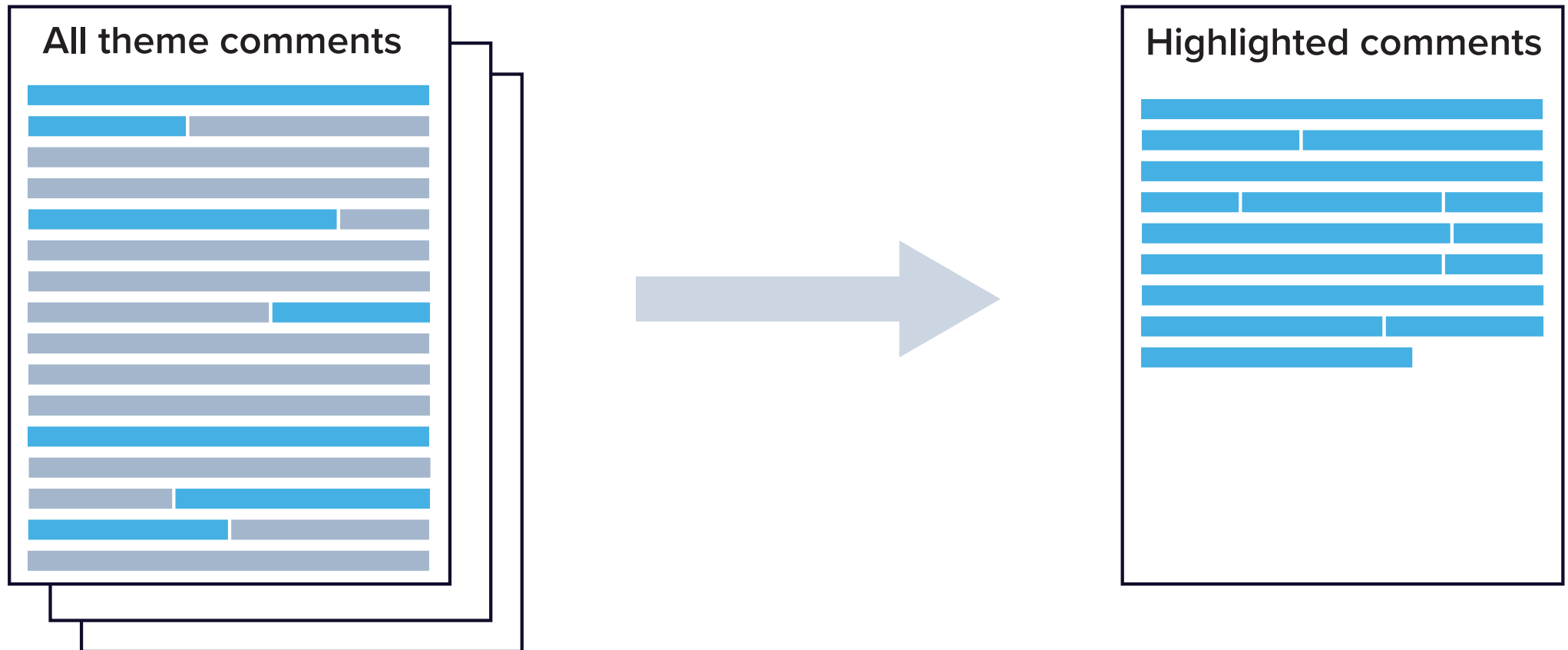


Reverse-engineering a topic to gain more understanding



Summarising underlying text

We use proprietary document summarisation algorithms to identify example sentences



TEXT ANALYTICS TECHNIQUES

Some of our lessons-learnt from the last 5 years

Key Challenges with analysing employee text comments

In most instances the real alternative to analysing text is “doing nothing.”



Sentence
fragments &
spelling

Multiple
languages

Medium-sized
datasets

Turnaround
needs to be
quick

Can't cost a
fortune

Specific Models



Generic Models

(Humans + Algorithms)

>

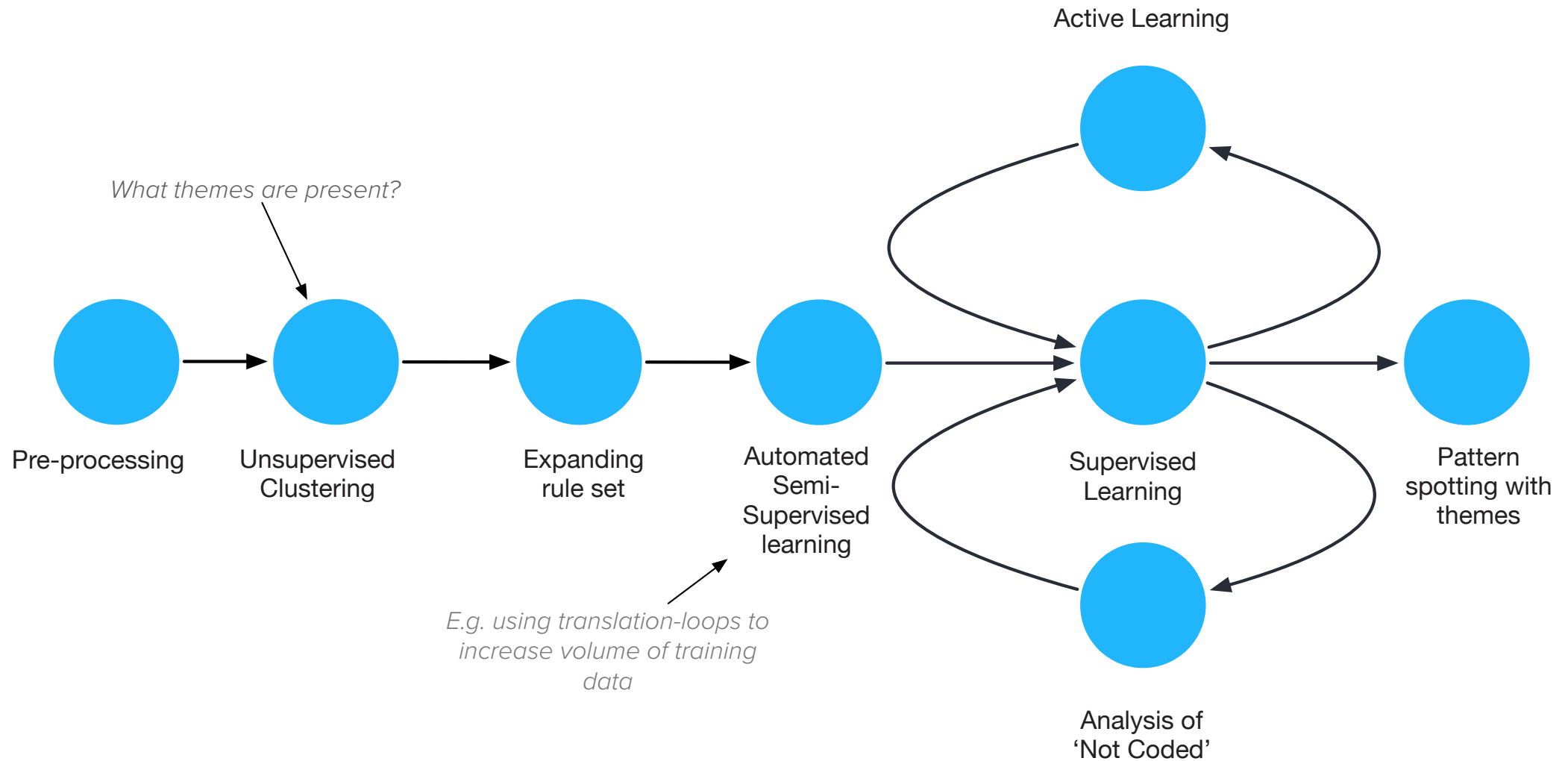
(Humans OR Algorithms)

**People in organizations develop
their own language**

{eg. Project Names, Acronyms, just plain misuse...}

Our Process

A highly-simplified representation of our overall text coding process.



History of text analytics

3 phases of text analytics. Data / Computing availability has been a key driver for promoting the adoption of new approaches



**NLP / Programmed
Grammatical
Approaches**

Statistical Approaches

**Deep Learning / Word
Embeddings**

“You shall know a word by the company it keeps!”

John Firth

Page 11, “A synopsis of linguistic theory 1930-1955“, in *Studies in Linguistic Analysis 1930-1955*, 1962

Word Embeddings

Dimensionality reduction can give you an idea of what is happening



‘Word-based NLP’ - Dimensionality Reduction techniques

Common techniques such as stop-word-removal, stemming and lemmatization all attempt to reduce dimensionality.

“My system keeps crashing! his crashed yesterday, ours crashes daily”

Stemming

**‘My system keep crash hi crash
yesterday, our crash daili’**

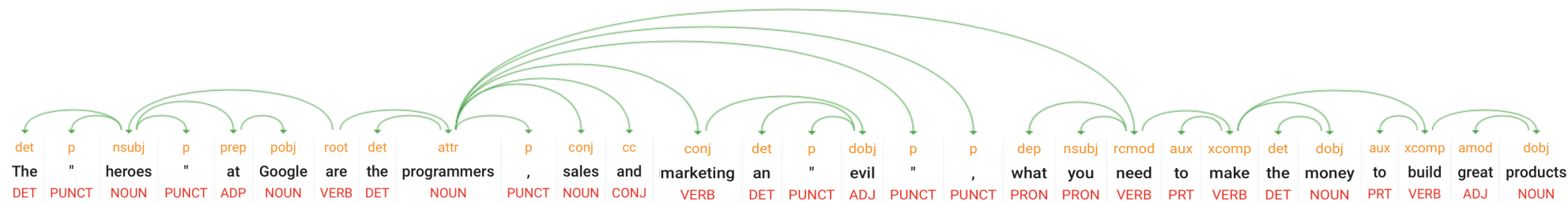
Lemmatization

**‘My system keep crash ! his crash
yesterday , ours crash daily’**

From removing stop words to emphasising important phrases

Common techniques such as stop-word-removal, stemming and lemmatization all attempt to reduce dimensionality. A better approach is to traverse a parsing tree.

“The “heroes” at Google are the programmers, sales and marketing an “evil”, what you need to make the money to build great products.”



Graph Traversal to identify potential candidates (short phrases)

Create weights on candidate phrases

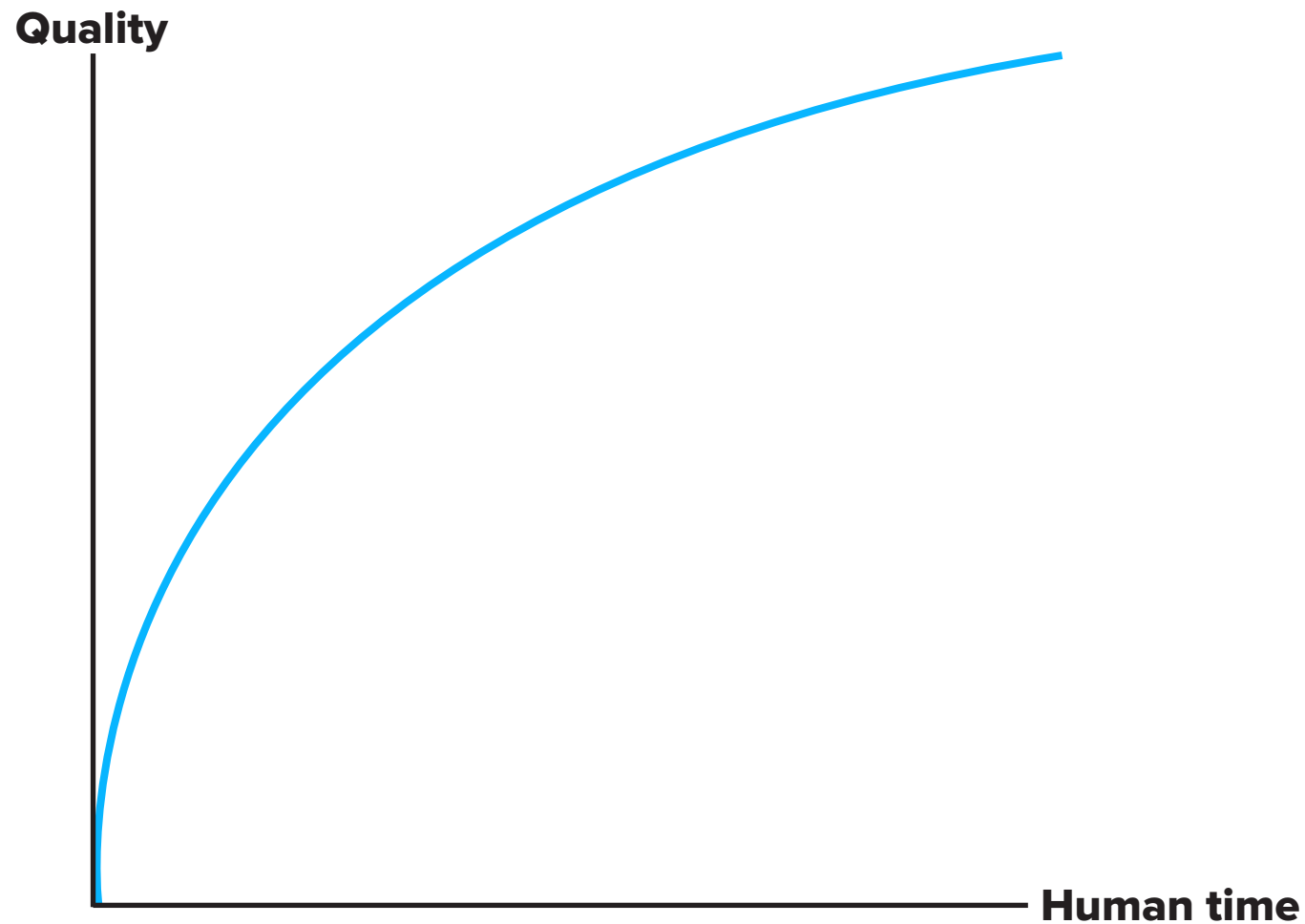
w_1 w_2 w_3 w_4

w_n

Cluster candidate phrases

In a commercial setting typically we're optimising on two variables

Given our constraints, there are two things that define our success



Building and maintaining a training set

We decided to build our own coding platform

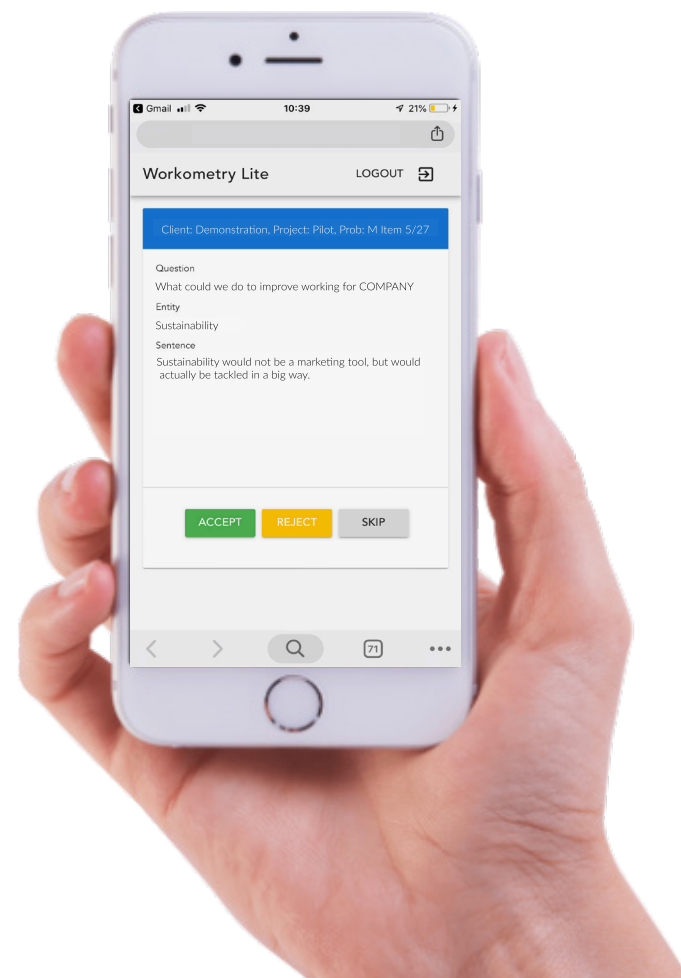
Provision domain experts to the right projects

Data Scientist out of the loop

Additional flexibility - eg visualisation

Measure productivity

Few scaling costs



Maximising human input

A few examples

Healthy

More work
needed



Overfit

Possibly linked
to another
concept

What has worked?

Some things have worked better than others

Domain-specific embeddings

Refine before automating

Experiment, experiment,
experiment

What next?

There's always too many ideas to try!

Improve automatic labelling

Improve tools for coders
(work on UI / UX)

Better models for uncertainty
(which to pass to human)

Building generic models for rare
but important themes

Further Ensemble methods

New techniques (ELMFIT, BERT)

Andrew Marritt | Founder & CEO

andrew.marritt@organizationview.com | +41 78 905 0775

Please subscribe: www.empiricalHR.com

Coming up... 4x4x4 Webinar – 1 July 2020

How ABN AMRO Designs and Delivers a Practical Approach for Continuous Employee Listening

Guest Speaker: Patrick Coolen, Global Head of People Analytics, Strategic Workforce Planning and HR Survey Management, ABN AMRO

Patrick Coolen will share best practises on ABN AMRO's continuous employee listening journey. He will talk about why ABN AMRO believes the voice of the employee is important and will share how their survey management and people analytics teams joined forces to measure, visualise and share the voice of the employee with senior management and the rest of their organisation.



ABN·AMRO

Date: Wednesday 1 July 2020

Time: 1600 BST | 1700 CEST
1100 EDT | 0800 PDT