

Text as Networks Polinode presentation Nov. 2020

Andrew Marritt | Founder & CEO

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

We code text

Workometry text coding is now available in 3 different versions to adapt to the needs of most text analysis requirements. All coding models can be automated for clients running ‘always-on’ surveys.

Custom

Our original service. Completely custom-built ‘inductive’ coding model.

When Custom is right:

- You want the highest level of quality, with the greatest coverage possible
- When the question you have asked is not standard
- If you require a proprietary set of themes, for example to match an existing methodology or company-specific products

Standard

An extensive, pre-built coding model for a selection of the most commonly-asked survey questions

When Standard is right:

- When you’ve asked a question similar to our standard set
- You have a smaller number of answers meaning insufficient data for a custom model
- You’re looking for a high-quality model at a very affordable price

Hybrid

Our Hybrid service combines many of the benefits of both Standard and Custom. We start with a Standard model and tune it for your data.

When Hybrid is right:

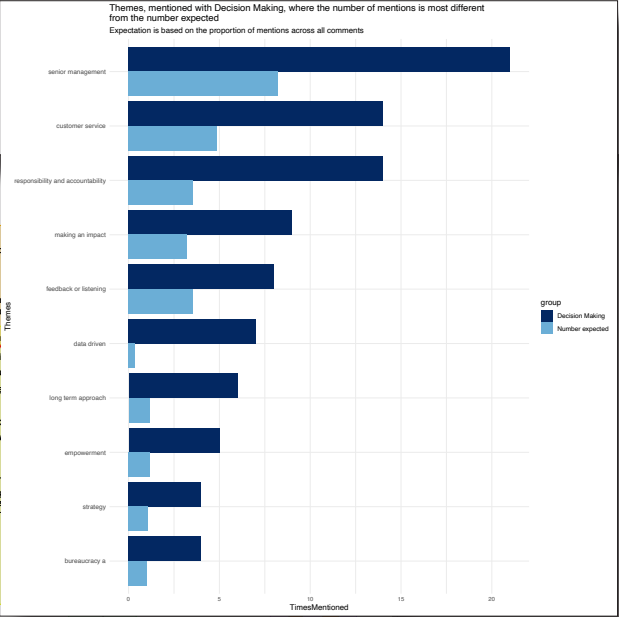
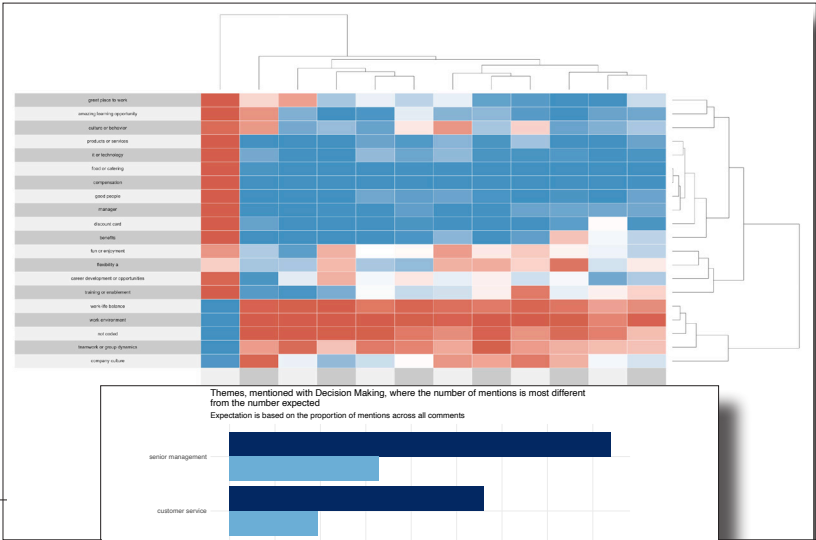
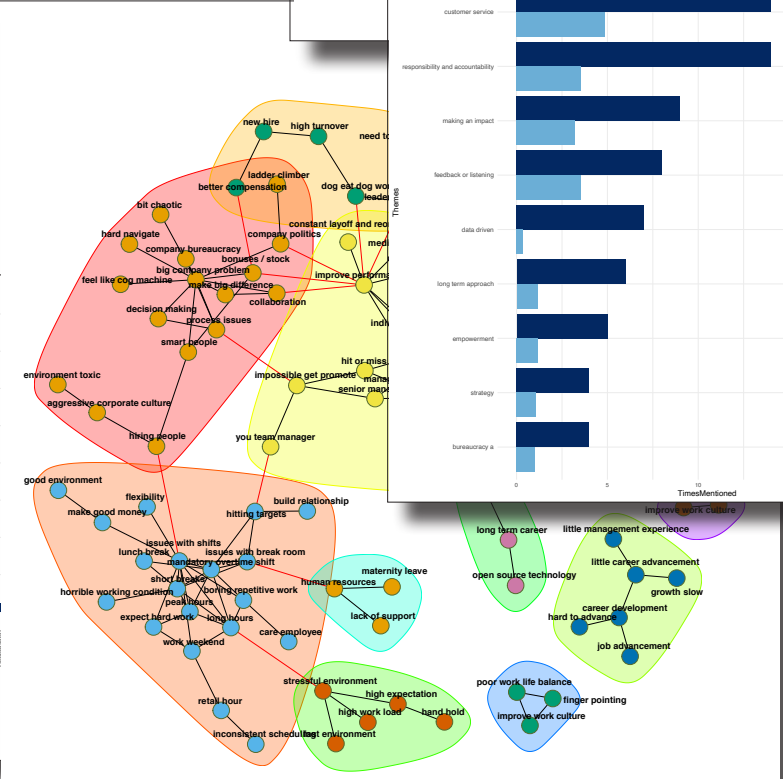
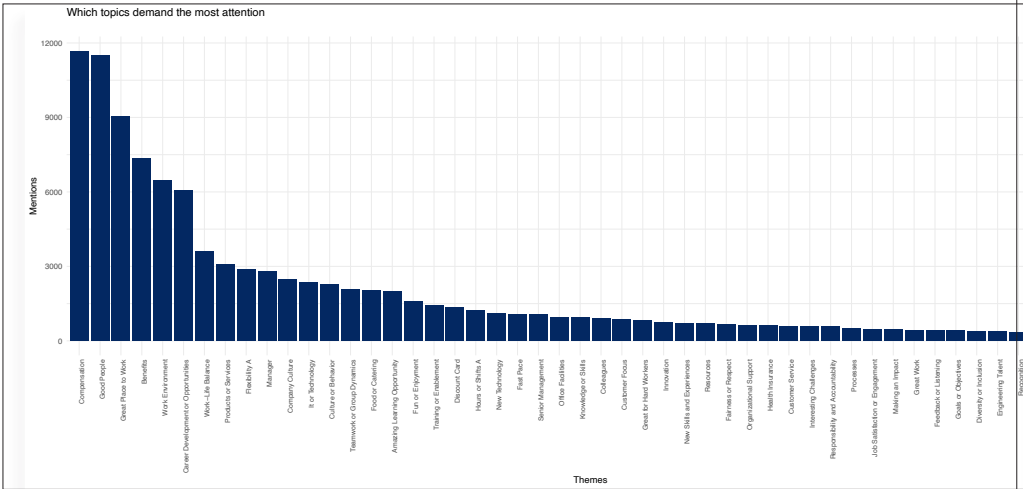
- When you’ve asked a question similar to our standard set
- When you want to ensure that company-specific themes are captured
- More cost effective than Custom

Visualisation to reveal insight and patterns

We've learned that text classification is only a small part of what is needed to drive great decision-making and effective action.

An objective of all text-analysis is to summarise the composition of comments, guiding the decision maker to areas and ideas that need action.

We've found that machine learning has a strong role in identifying which themes and ultimately comments have the most value. Visualising these models can show how ideas are linked, which ideas are dominant with different segments of employee and which themes are emergent.



Coding interface (Beta)

For those wanting maximum control of their text classification we can offer the coding interface our analysts use. Designed for domain-experts to maintain and refine a text model it offers a secure web-based platform that can be used on the desktop or on a mobile device.

Our coding interface provides a simple-UI to enable domain experts perform the most common tasks needed to build and maintain great training data.

- Reviewing and refining ambiguous examples
- Seeding new themes and creating training examples
- Identifying noisy / erroneous training examples and removing from training data
- Data augmentation

Models can be deployed to run automatically on new data or training sets can be downloaded to enable data-scientists to use proprietary models.

The screenshot displays the Coding interface (Beta) for text classification. The interface is divided into a sidebar on the left and a main grid of entity cards on the right.

Sidebar (Left): A list of entities for classification, including:

- Communication (Last revised: 2020-07-30 09:55:37, Training examples: 152, LOW: 22, MED: 21, HIGH: 72)
- Compensation (Last revised: 2020-07-30 09:55:37, Training examples: 14190, LOW: 1023, MED: 832, HIGH: 4442)
- Customer Focus (Last revised: 2020-07-30 09:55:37, Training examples: 172, LOW: 105, MED: 123, HIGH: 529)
- Corporate Social Responsibility (Last revised: 2020-07-30 09:55:37, Training examples: 14, LOW: 0, MED: 0, HIGH: 5)
- Customer Service (Last revised: 2020-07-30 09:55:37, Training examples: 593, LOW: 39, MED: 52, HIGH: 30)
- Decision Making (Last revised: 2020-07-30 09:55:37, Training examples: 172, LOW: 28, MED: 23, HIGH: 145)
- Discount Card (Last revised: 2020-07-30 09:55:37, Training examples: 1734, LOW: 46, MED: 74, HIGH: 671)
- Diversity Or Inclusion (Last revised: 2020-07-30 09:55:37, Training examples: 395, LOW: 31, MED: 34, HIGH: 178)
- Company Culture (Last revised: 2020-07-30 09:55:37, Training examples: 129, LOW: 122, MED: 129, HIGH: 1082)
- Company Performance (Last revised: 2020-07-30 09:55:37, Training examples: 53, LOW: 14, MED: 6, HIGH: 32)
- Culture Or Behavior (Last revised: 2020-07-30 09:55:37, Training examples: 2213, LOW: 307, MED: 278, HIGH: 1026)
- Data Driven (Last revised: 2020-07-30 09:55:37, Training examples: 67, LOW: 6, MED: 3, HIGH: 38)

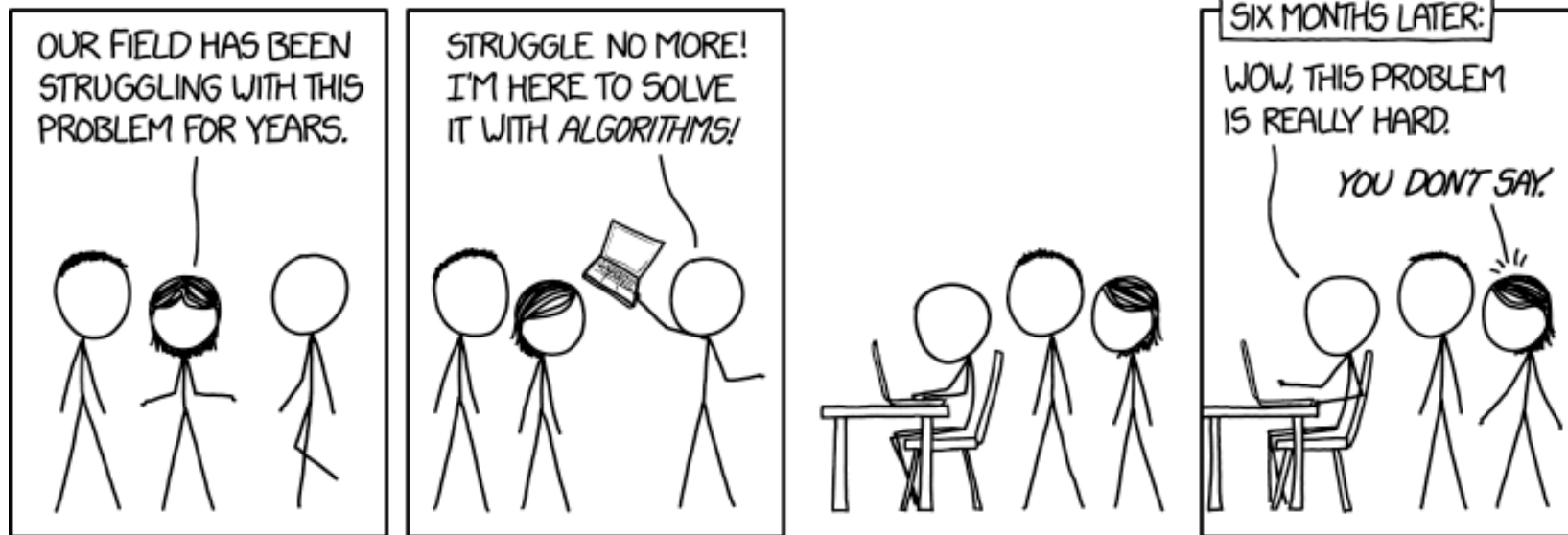
Main Grid (Right): A grid of entity cards, each displaying the entity name, last revised date, training examples count, and classification metrics (LOW, MED, HIGH). The cards are arranged in a 4x3 grid, with the last card in the row being empty.

Top Bar: Contains navigation links: NEW ENTITIES, ATTRIBUTE CHUNKS, UPLOAD FILES, and LOGOUT. It also includes a search bar with a "REFRESH" button and a "SEARCH AGAIN" button.

Bottom Bar: Displays "Page 2 of 9" and navigation buttons (Previous, Next).

There are no easy solutions

But we're getting good at solving specific problems



Old approach: Bag of Words

Old approach to text analysis was mostly context-free.

Identify words in a sentence

Pre-process (eg stemming, removal of stop words)

Count frequency of 'tokens'

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

J.R. Firth

Parsing trees to understand and segment sentences

Parse trees help you understand the structure of text in a graph (tree) format.

Try the API

Text Analytics is a key part of People Analytics

ANALYZE

[See supported languages](#)

Entities

Sentiment

Syntax

Categories

✓ Dependency

✓ Parse Label

✓ Part of Speech

✓ Lemma

✓ Morphology

nn

Text

NOUN

number=SINGULAR

proper=PROPER

nsubj

Analytics

NOUN

number=SINGULAR

proper=PROPER

root

is

be

VERB

mood=INDICATIVE

number=SINGULAR

person=THIRD

tense=PRESENT

det

a

DET

amod

key

ADJ

attr

part

NOUN

number=SINGULAR

prep

of

ADP

nn

People

NOUN

number=PLURAL

proper=PROPER

pobj

Analytics

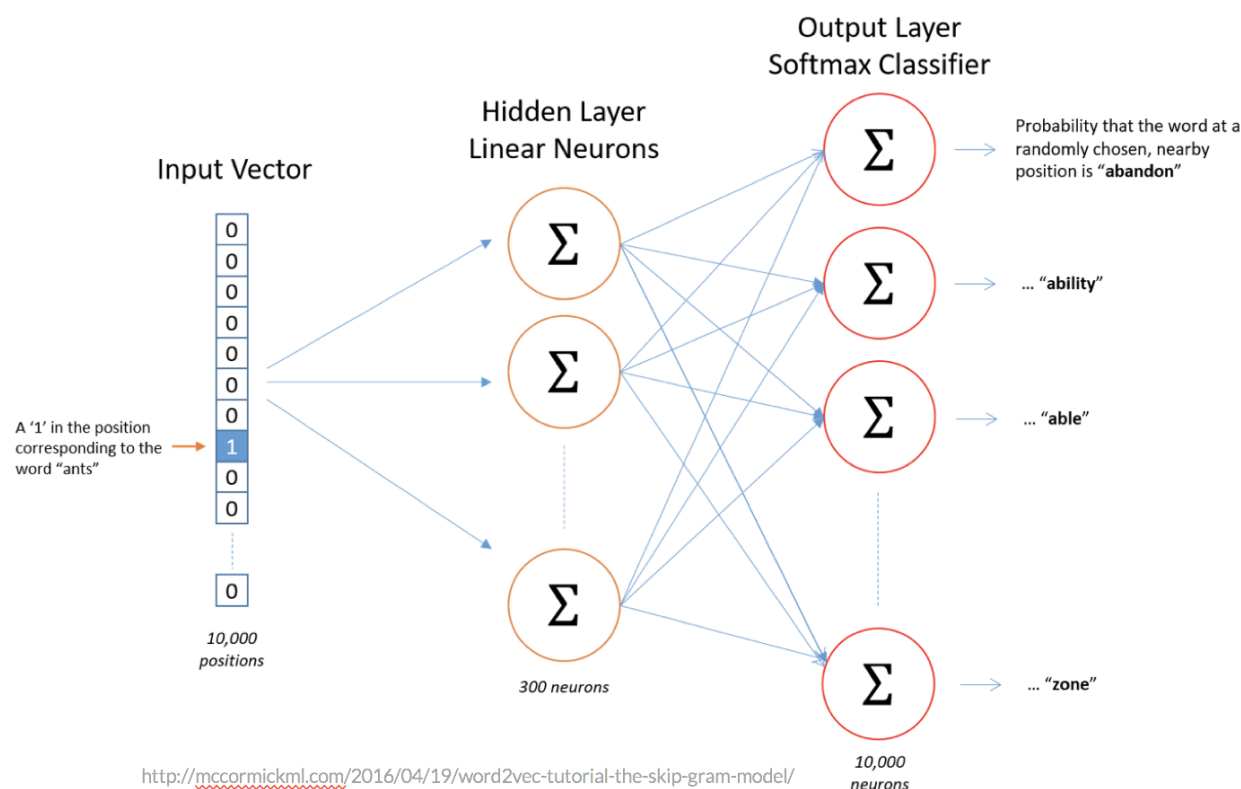
NOUN

number=SINGULAR

proper=PROPER

Word Embeddings: Driving most growth since 2012

Embeddings are created by a process of identifying relationships between words and their neighbours.



“Skip-Gram”
With one-hot encoded centre word, we can predict context words.

Hidden layer creates embeddings

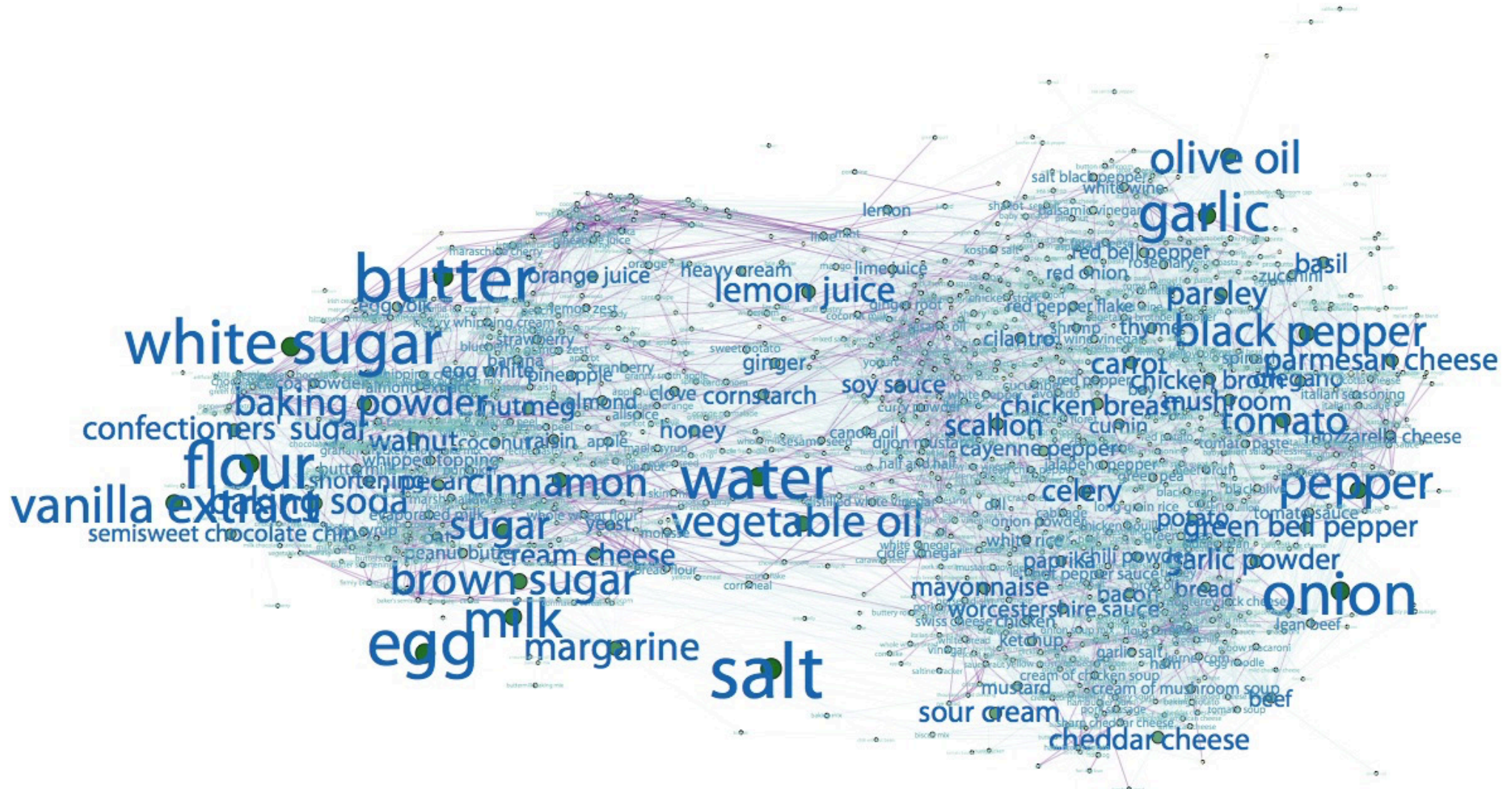
Semantically similar expressions inhabit a similar space

Dimensionality reduction lets you understand what is going on.



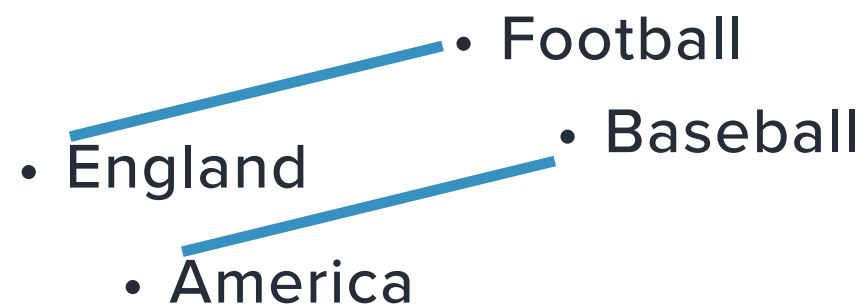
Ingredient networks produce a similar result

Graphs with a force layout produce a similar result with cooccurence networks



Position and direction provides meaning in embeddings

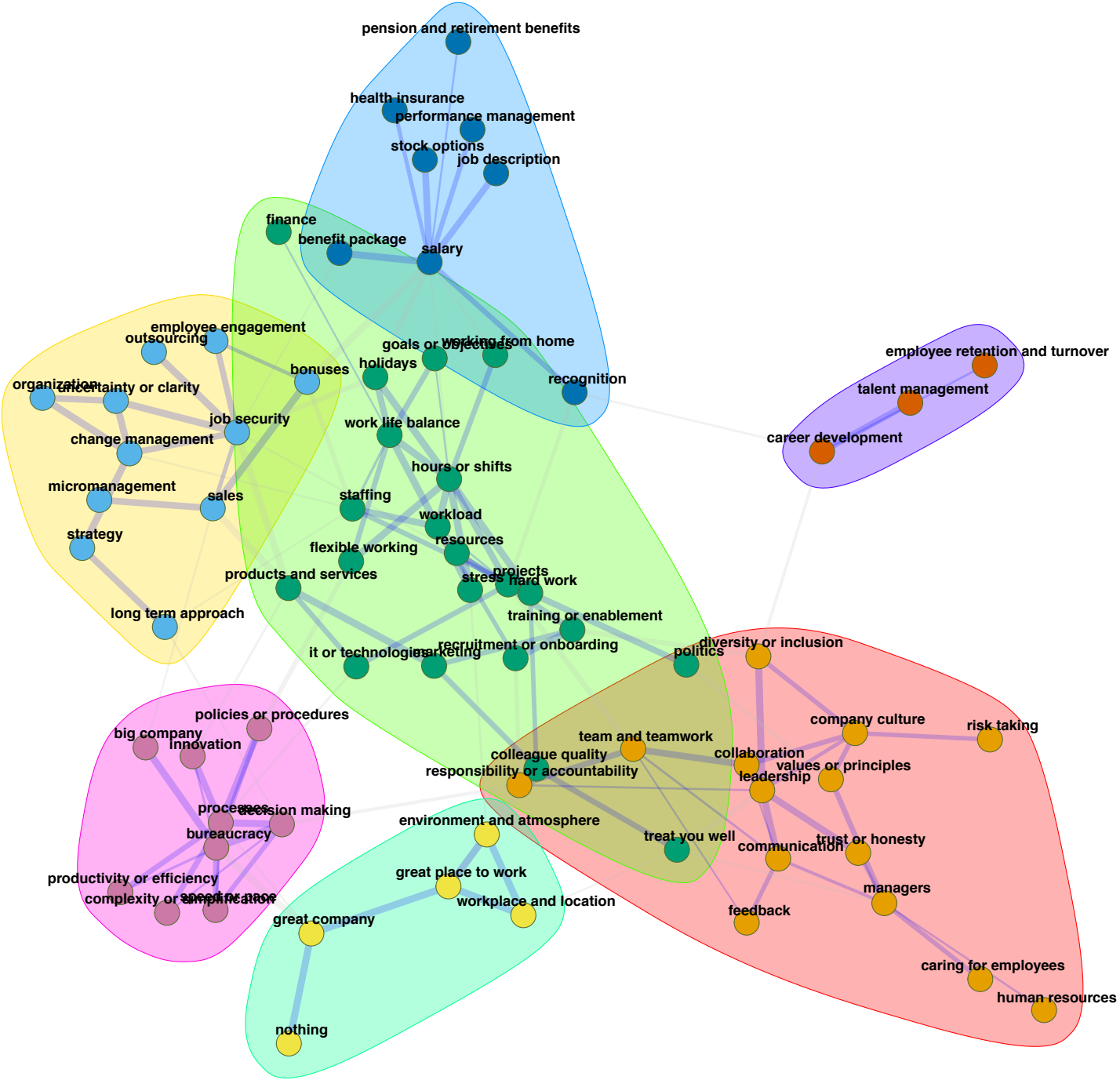
Subtracting vectors provides relationship between words



England	Football	America	Show	Top N: 2
Baseball America				

BMW	car	Boeing	Show	Top N: 2
airplane aircraft				

Abstracting further - cooccurrence networks



How we create these networks

Empirical Bayesian approach

Assume proportion
of people
mentioning theme
is constant (prior)

Use evidence of
co-occurrence to
update prior

Subtract
distributions to
create probability

The data we create

Resulting calculation creates a weighted edge-list

Given the 30 people who talk about alignment we'd expect 3.14 people to mention decision making. Instead we see 9. We think the chance of them being strongly related is 94%.

TimesMentioned	NumberExpected	ProbabilityHigher	Factor	entity	SizeOfFactor	ProportionWhoMention	multiple	factorPropAnswers
30	0.070191457	1	alignment	alignment	30	1	427.402439	0.0009567241764
0	0.014551887	0.1016	alignment	assignments	30	0	0	0.0009567241764
0	0.623163181	0.2708	alignment	benefit package	30	0	0	0.0009567241764
2	1.106799441	0.60955	alignment	big company	30	0.066666667	1.807012117	0.0009567241764
0	1.111079408	0.24525	alignment	bonuses	30	0	0	0.0009567241764
0	0.255086027	0.26595	alignment	brand and reputation	30	0	0	0.0009567241764
0	0.139526921	0.24575	alignment	breaks	30	0	0	0.0009567241764
2	1.63066739	0.51695	alignment	bureaucracy	30	0.066666667	1.226491689	0.0009567241764
3	6.593717009	0.20405	alignment	career development	30	0.1	0.45497858	0.0009567241764
0	0.307301624	0.26695	alignment	caring for employees	30	0	0	0.0009567241764
0	0.232830199	0.26465	alignment	certainty and clarity	30	0	0	0.0009567241764
1	1.591291694	0.3636	alignment	change management	30	0.033333333	0.628420298	0.0009567241764
2	0.325277485	0.84985	alignment	collaboration	30	0.066666667	6.148596491	0.0009567241764
0	1.802722059	0.2083	alignment	colleague quality	30	0	0	0.0009567241764
0	0.045367649	0.17985	alignment	colleagues help you	30	0	0	0.0009567241764
5	2.256398551	0.7865	alignment	communication	30	0.166666667	2.215920587	0.0009567241764
0	0.106999173	0.23365	alignment	commute	30	0	0	0.0009567241764
0	0.022255828	0.12875	alignment	company car	30	0	0	0.0009567241764
2	4.148999914	0.2677	alignment	company culture	30	0.066666667	0.482043876	0.0009567241764
0	0.104431192	0.23255	alignment	company performance	30	0	0	0.0009567241764
0	0.066767484	0.20625	alignment	competitive	30	0	0	0.0009567241764
0	0.207150398	0.26155	alignment	complexity or simplification	30	0	0	0.0009567241764
0	0.014551887	0.1016	alignment	corporate social responsibility	30	0	0	0.0009567241764
0	0.24053414	0.26355	alignment	customer focus	30	0	0	0.0009567241764
0	0.302165663	0.26635	alignment	data and digitalisation	30	0	0	0.0009567241764
9	3.144919679	0.9394	alignment	decision making	30	0.3	2.861758302	0.0009567241764
1	1.15987103	0.41795	alignment	diversity or inclusion	30	0.033333333	0.862164822	0.0009567241764
0	0.011983907	0.0892	alignment	dress code and uniform	30	0	0	0.0009567241764
1	0.520443975	0.5493	alignment	employee engagement	30	0.033333333	1.921436404	0.0009567241764

Andrew Marritt | Founder & CEO

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