

Text Visualization

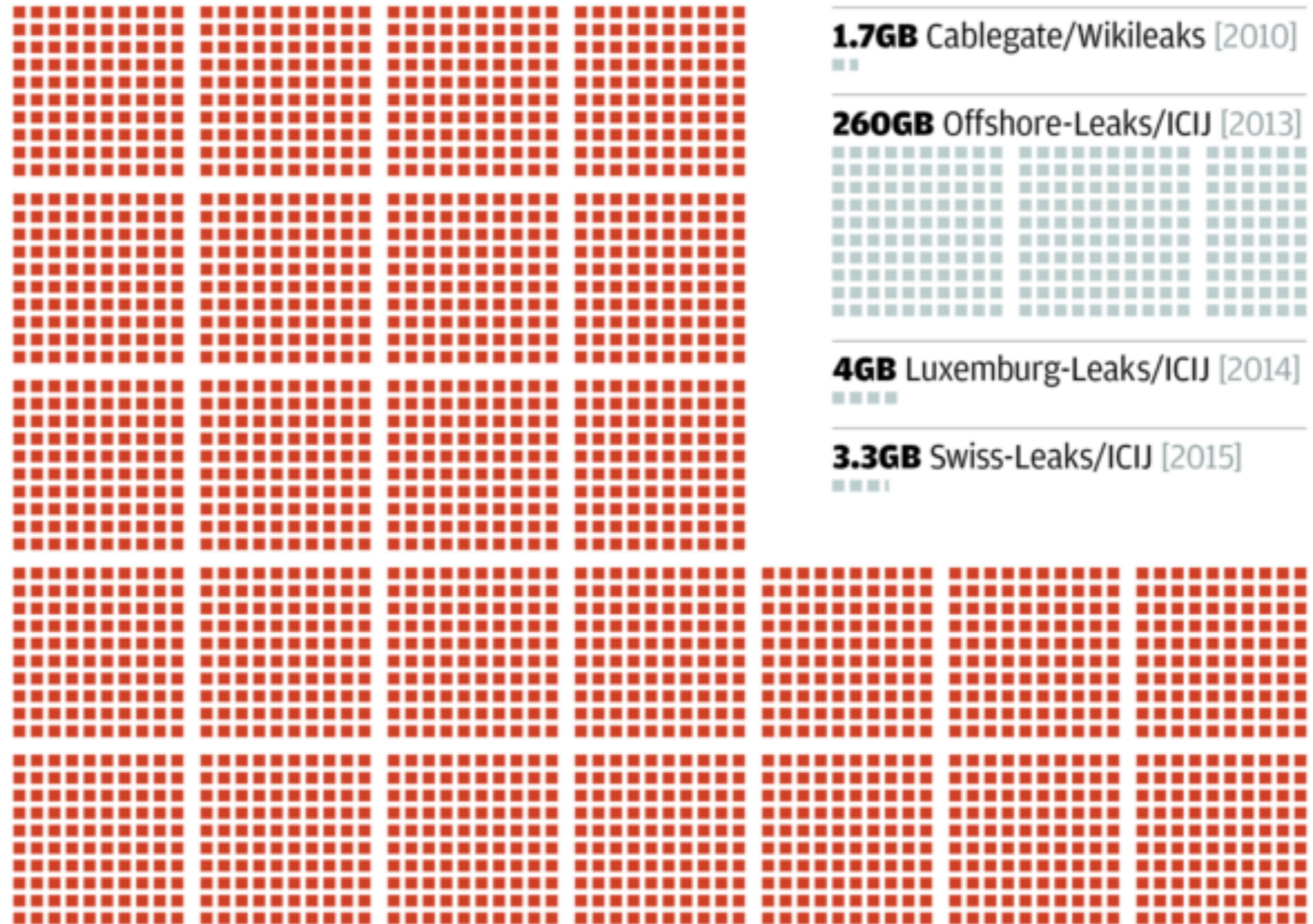
Lecture in: “Visualization
and Visual Data Analysis”

thanks @ Elena Sofie Rudkowsky

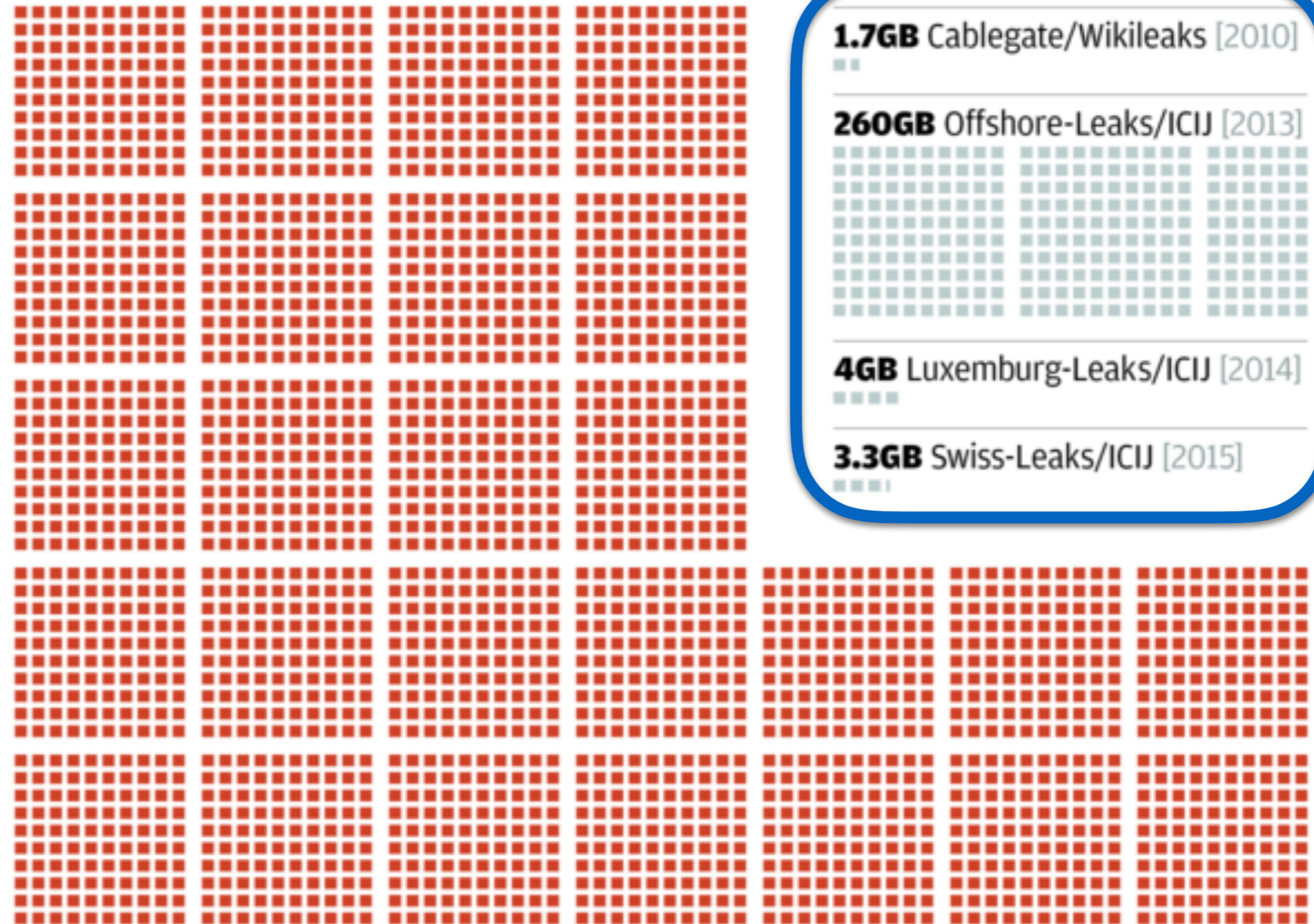


Introductory Example: Panama Papers

2.6TB The Panama Papers/ICIJ [2016]



<http://www.computerworld.com/article/3053601/security/consider-the-panama-papers-breach-a-warning.html>

Scale of the Leak
compared to others**2.6TB** The Panama Papers/ICIJ [2016]

<http://www.computerworld.com/article/3053601/security/consider-the-panama-papers-breach-a-warning.html>

Mossack Fonseca

- law office in Panama
- 214,000 letterbox companies in 21 tax havens

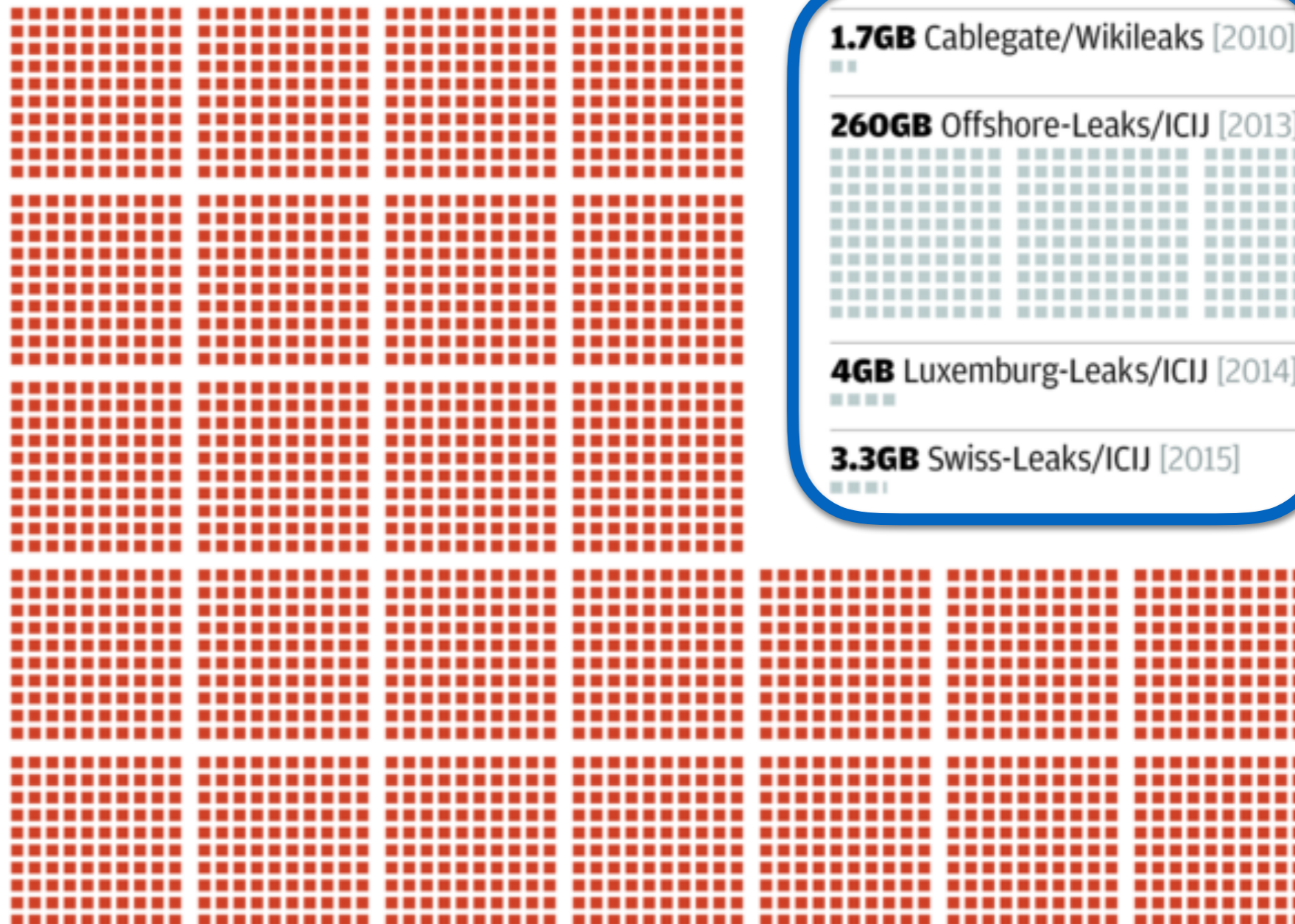
Letterbox Company

- address for tax purposes, money laundering...
- business carried on elsewhere/nowhere

Prominent Owners

- Iceland: Prime Minister Gunnlaugsson
- Russia: Circle around president Putin
- Syria: Network around president al-Assad
-

2.6TB The Panama Papers/ICIJ [2016]

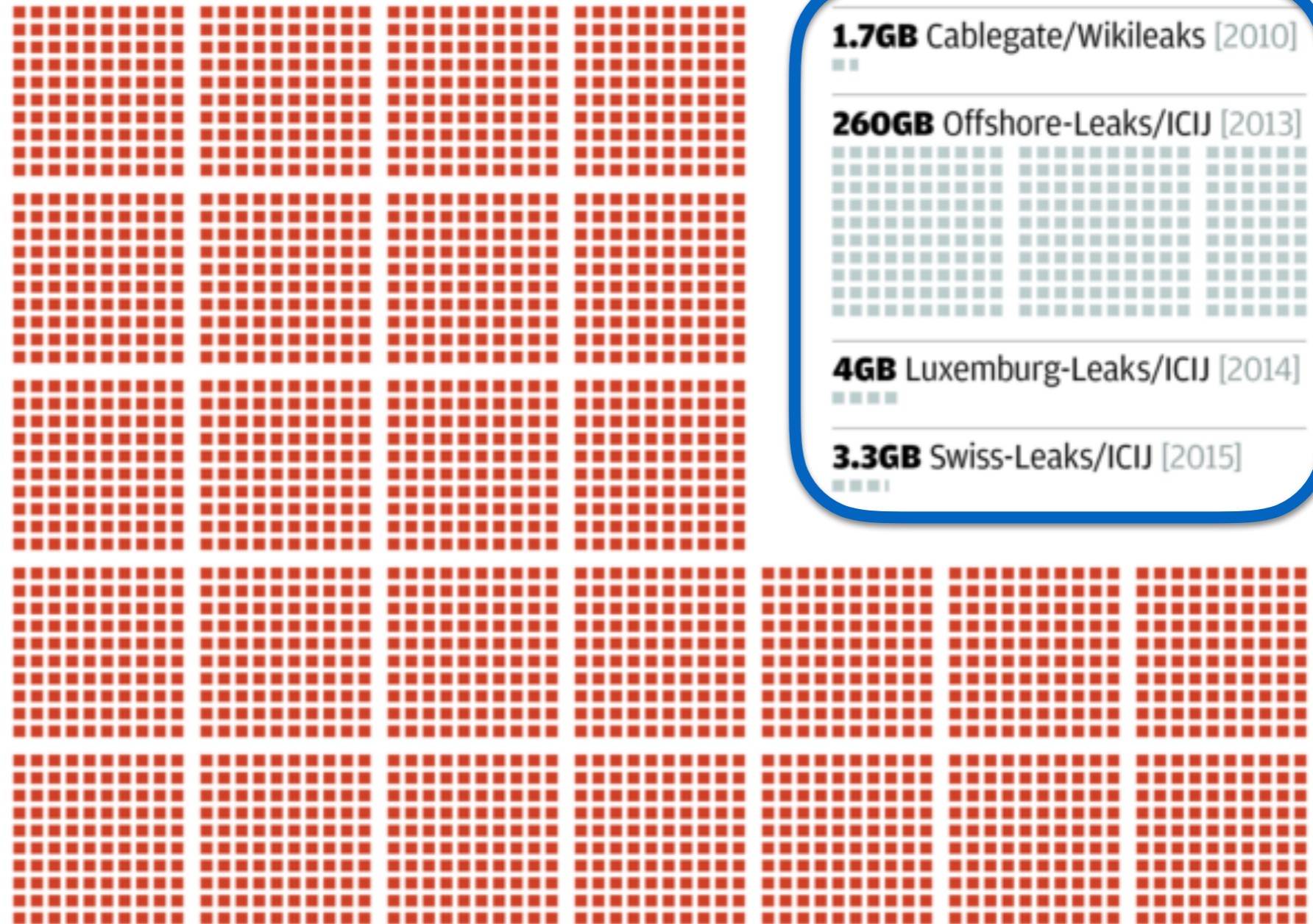


<http://www.computerworld.com/article/3053601/security/consider-the-panama-papers-breach-a-warning.html>

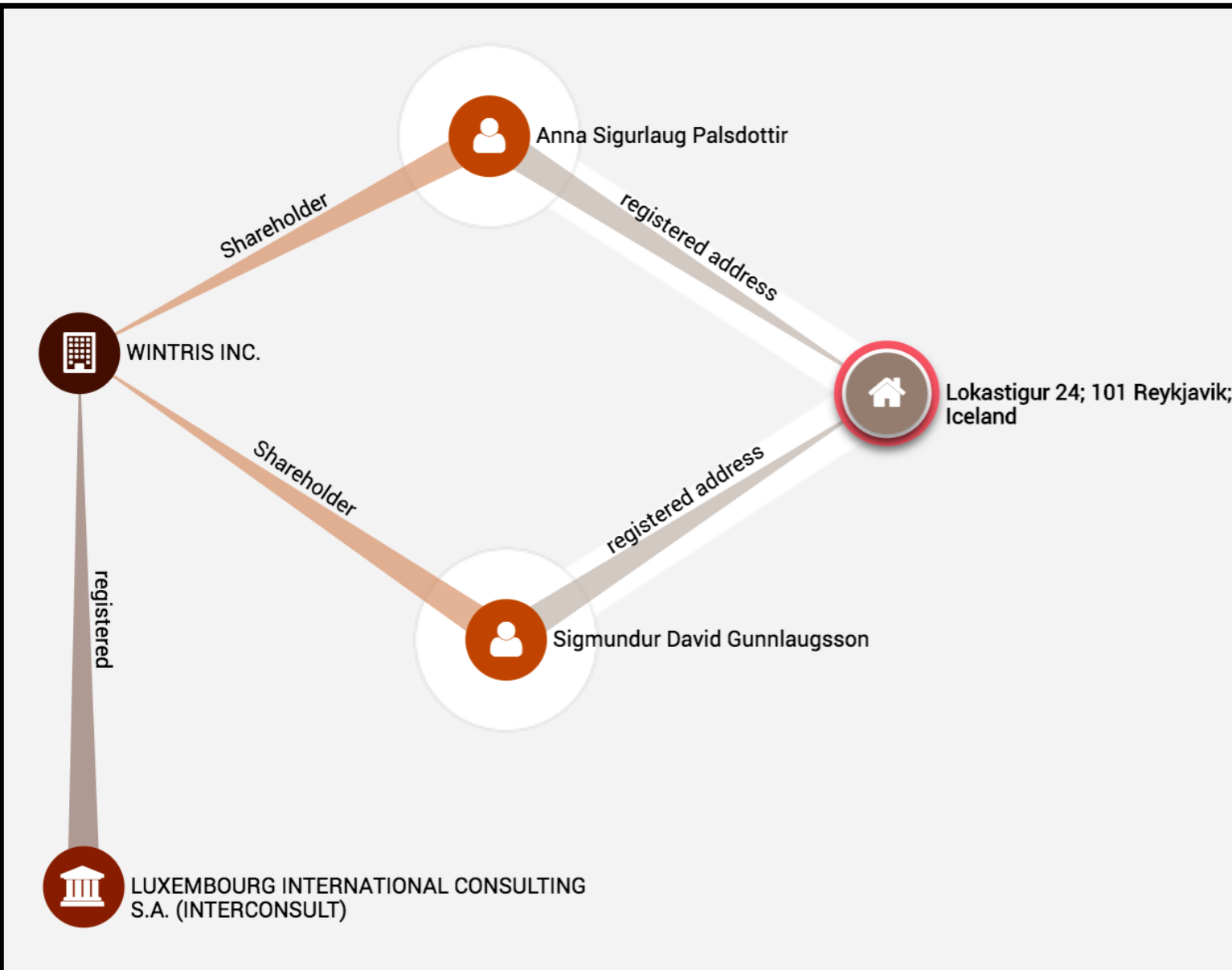
Analysis

- **400 journalists** from 80 countries
- **11.5 million documents** (PDFs, emails, text files, database format files, images...)
- **1 year research**
- **text analysis software** like Nuix

2.6TB The Panama Papers/ICIJ [2016]

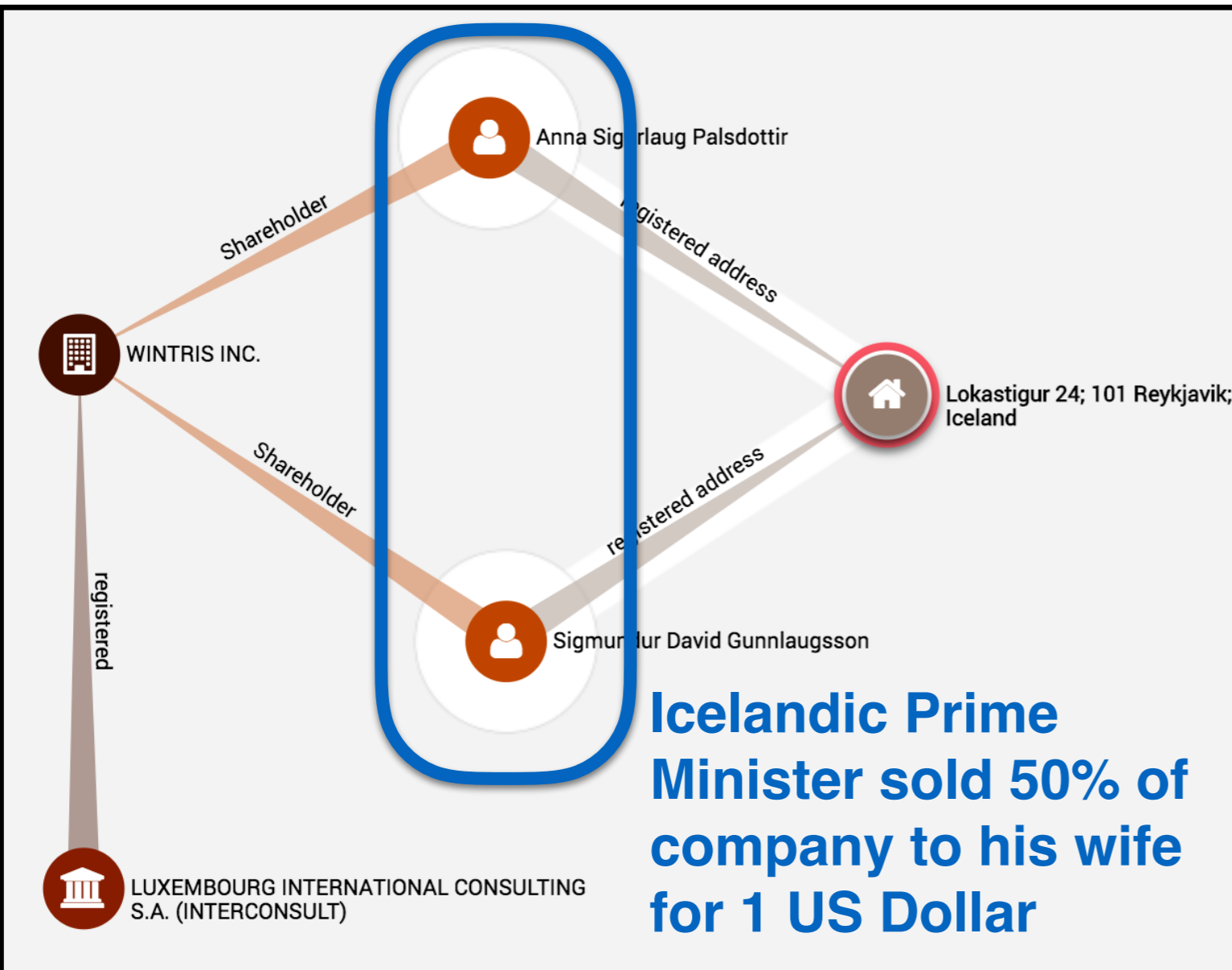


<http://www.computerworld.com/article/3053601/security/consider-the-panama-papers-breach-a-warning.html>



https://panamapapers.icij.org/the_power_players/

<http://panamapapers.sueddeutsche.de/articles/56fec0cda1bb8d3c3495adfc/>



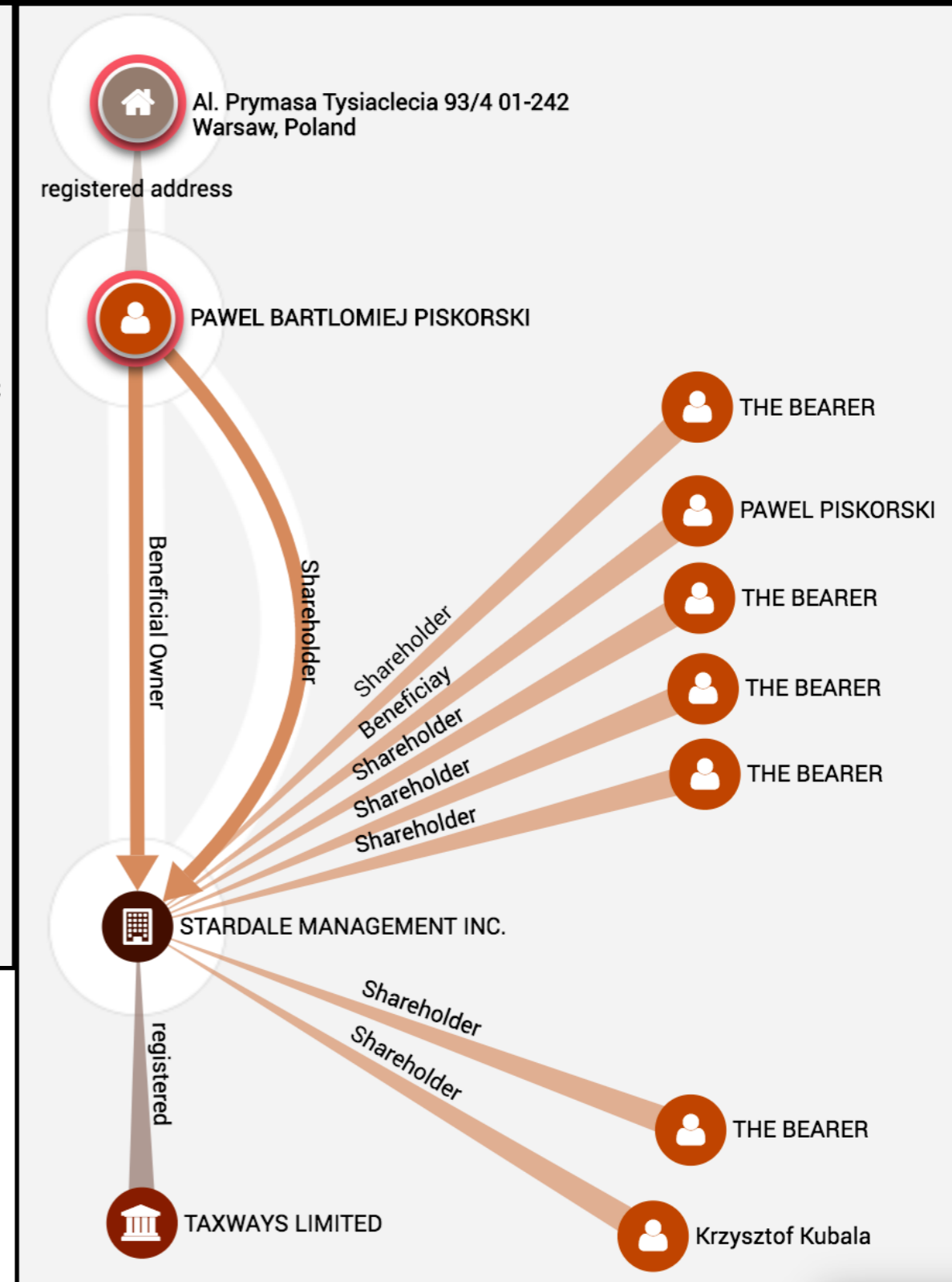
Gunnlaugsson neglects to disclose his Wintris Inc. shareholding	Gunnlaugsson sells 50 percent share of the company to his future wife for one US Dollar
25.4.2009	31.12.2009

https://panamapapers.icij.org/the_power_players/

<http://panamapapers.sueddeutsche.de/articles/56fec0cda1bb8d3c3495adfc/>



<p>Gunnlaugsson neglects to disclose his Wintris Inc. shareholding</p> <p>25.4.2009</p>	<p>Gunnlaugsson sells 50 percent share of the company to his future wife for one US Dollar</p> <p>31.12.2009</p>
---	--



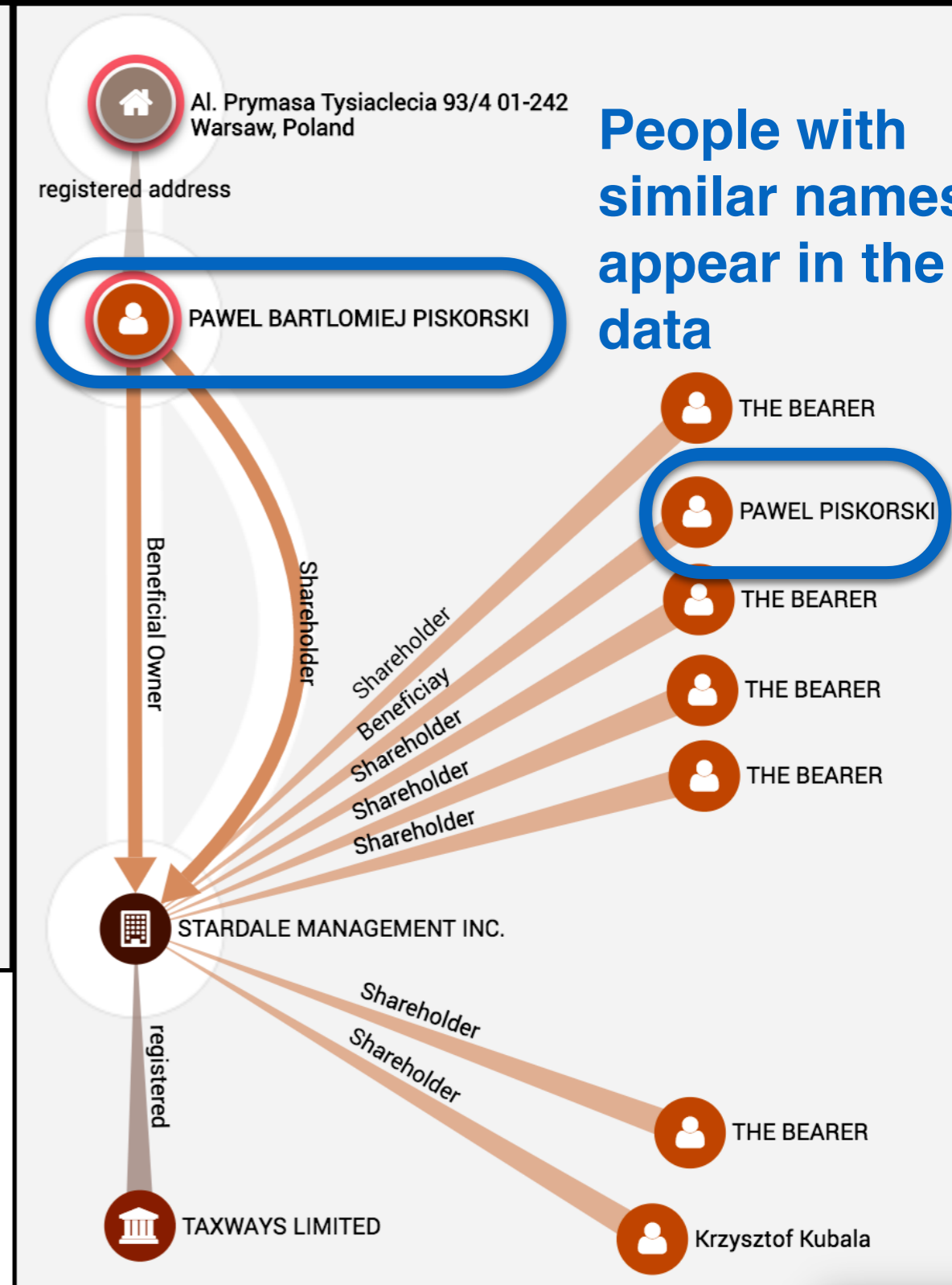
https://panamapapers.icij.org/the_power_players/

<http://panamapapers.sueddeutsche.de/articles/56fec0cda1bb8d3c3495adfc/>



Gunnlaugsson neglects to disclose his Wintris Inc. shareholding
25.4.2009

Gunnlaugsson sells 50 percent share of the company to his future wife for one US Dollar
31.12.2009



People with similar names appear in the data

https://panamapapers.icij.org/the_power_players/

<http://panamapapers.sueddeutsche.de/articles/56fec0cda1bb8d3c3495adfc/>

Question:

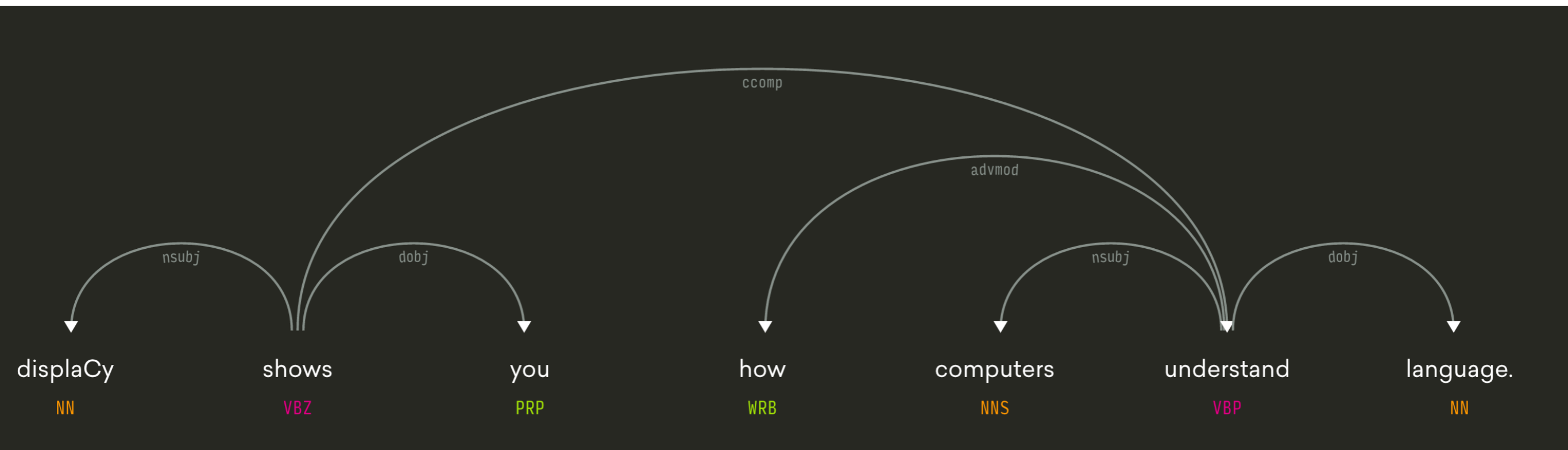
What can be extracted out of textual data?

Agenda

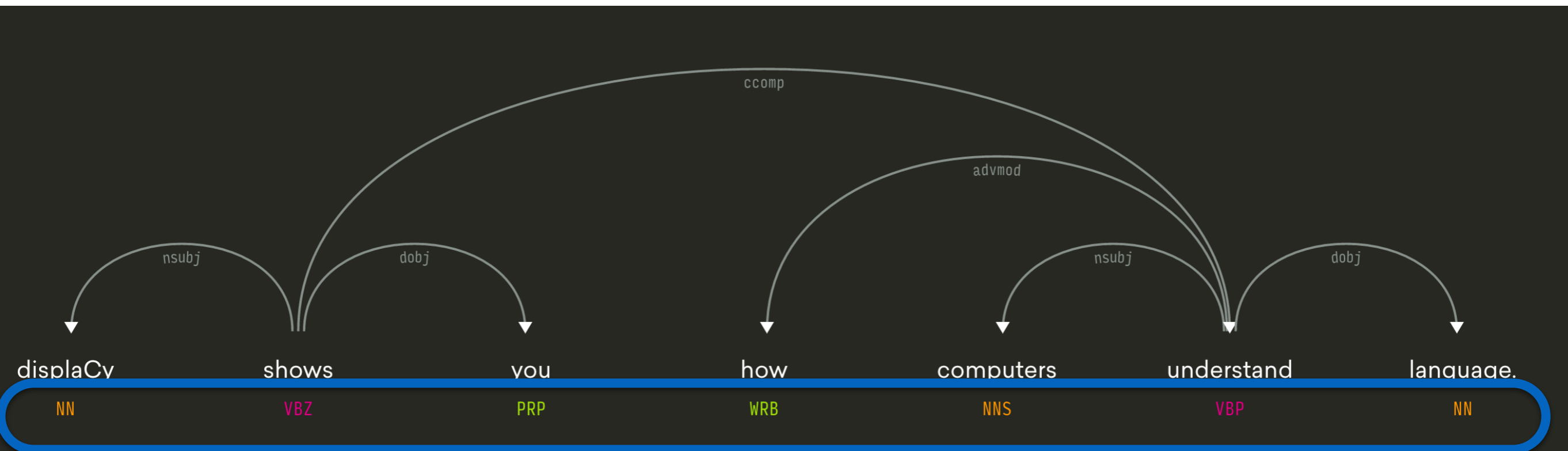
- From Text to Structure
 - POS Tags and Parsing Trees
 - Bag-of-Words
 - Distributed Word Embeddings
 - Topic Models
 - Named Entities
 - Sentiment Analysis
 - Temporal Events
 - Deep Learning
- Projects

From Text to Structure

POS Tags and Parsing Trees



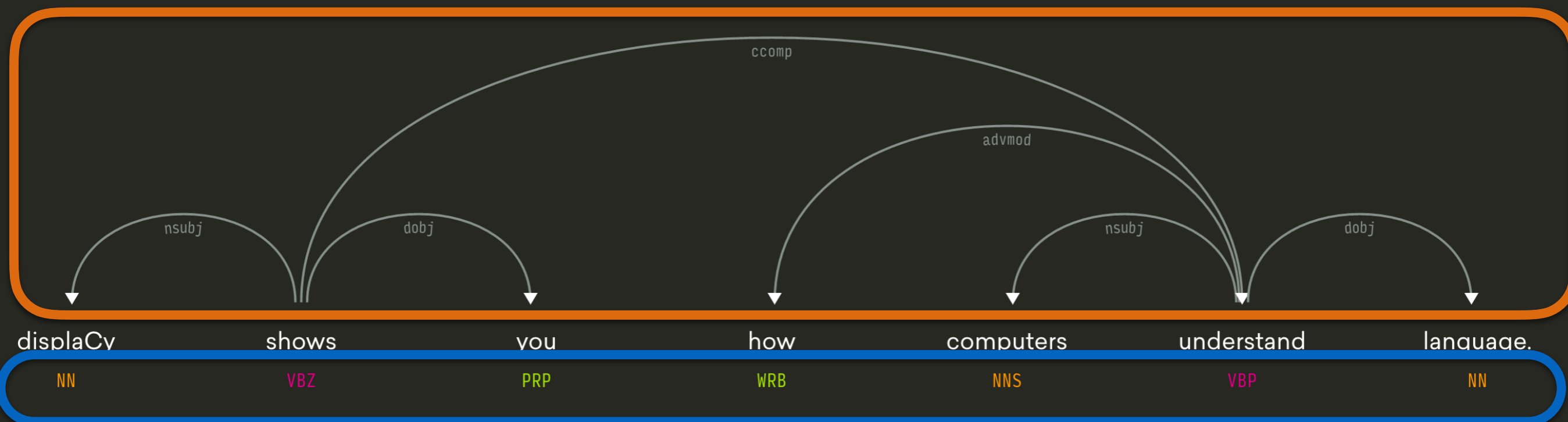
<https://demos.explosion.ai/displacy/>



Part-of-speech Tags (POS Tags)
mark verbs, nouns, adjectives etc.

<https://demos.explosion.ai/displacy/>

Parsing (or Syntax) Trees show the syntactical structure of language like object, subject...



Part-of-speech Tags (POS Tags)
mark verbs, nouns, adjectives etc.

<https://demos.explosion.ai/displacy/>

Question:

**Are POS Tags and Syntax Trees
interesting for Pattern Detection?**

Mostly not interesting for text mining (pattern detection) in document collections!

Mostly not interesting for text mining (pattern detection) in document collections!



Mostly not interesting for text mining (pattern detection) in document collections!



 **New York Magazine** ✓
March 31, 2016 · 🌐

Donald J. Trump has the **grammar of an 11-year-old.** That's not opinion.
That's research-proven.



**Only Bernie Sanders's
speeches went above
a 10th-grade level.**

LIVE
CNN
Intelligence
9:57 10:00

<https://www.facebook.com/NewYorkMag/videos/10154081648719826/>

Mostly not interesting for text mining (pattern detection) in document collections!



TABLE 1
SENTENCE SIZE

Number of sentences spoken by each speaker and sentence word count statistics. Number of words in a sentence is shown by average and 50%/90% cumulative values for all, stop and non-stop words.

speaker	number of sentences	sentence size					
		all		stop		non-stop	
Hillary Clinton	1,206	15.7*	21 43	9.0*	12 26	7.0*	9 19
Donald Trump	1,970	10.9*	15 36	6.6*	9 22	4.6*	6 16
total	3,176	14.8	19 40	9.5	11 25	7.5	9 19

Fields with * (e.g. 15.5*) link to data files and Wordles. Hover over the field to show these links. [See analysis.](#)

TABLE 2
PART OF SPEECH COUNT

Count of words categorized by part of speech (POS).

	part of speech									
	n+v+adj+adv		nouns (n)		verbs (v)		adjectives (adj)		adverbs (adv)	
Hillary Clinton	7,636*	2,173	3,581*	1,117	2,322*	799	1,306*	543	427*	97
	40.5%	28.5%	46.9%	31.2%	30.4%	34.4%	17.1%	41.6%	5.6%	22.7%
Donald Trump	8,158*	1,752	3,639*	953	2,375*	588	1,621*	550	523*	83
	37.9%	21.5%	44.6%	26.2%	29.1%	24.8%	19.9%	33.9%	6.4%	15.9%
total	15,794*	3,008	7,220*	1,635	4,697*	1,102	2,927*	886	950*	127
	39.1%	19.0%	45.7%	22.6%	29.7%	23.5%	18.5%	30.3%	6.0%	13.4%

Fields with * (e.g. 15.5*) link to data files and Wordles. Hover over the field to show these links. [See analysis.](#)

<http://mkweb.bcgsc.ca/debates2016/>

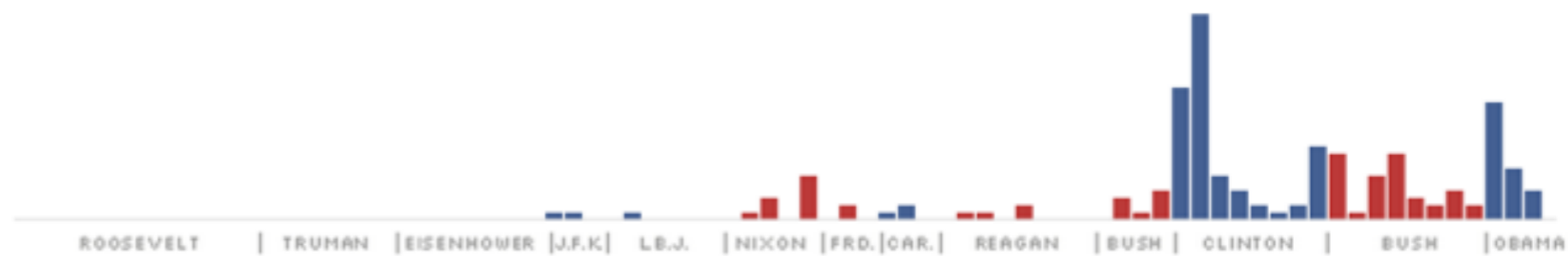
Bag-Of-Words (BOW)

Patterns of Speech: 75 Years of the State of the Union Addresses (Data Story, Washington Post 2011)

‘health care’

The expansion of health insurance coverage remains unpopular with nearly half the country, but Mr. Obama defended the health care law in his 2011 speech, though he added that he was willing with Republicans to improve it. In 1994, Mr. Clinton promoted his plan, which collapsed that year.

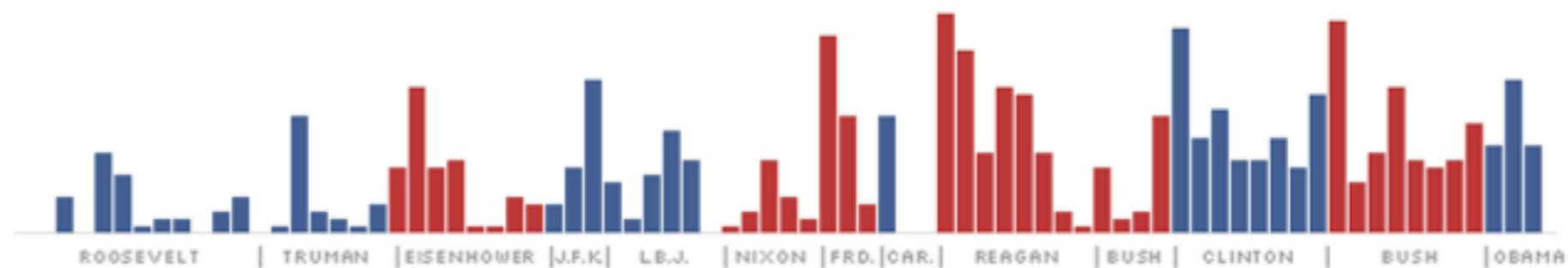
HEALTH CARE



‘tax’

Presidents have used the word every year since 1981, when Mr. Reagan uttered it 30 times, detailing his plan to reduce taxes and government spending.

TAX, TAXED, TAXES, TAXING



Patterns of Speech: 75 Years of the State of the Union Addresses (Data Story, Washington Post 2011)

'health care'

The expansion of health insurance coverage remains unpopular with nearly half the country, but Mr. Obama defended the health care law in his 2011 speech, though he added that he was willing with Republicans to improve it. In 1994, Mr. Clinton promoted his plan, which collapsed that year.

HEALTH CARE

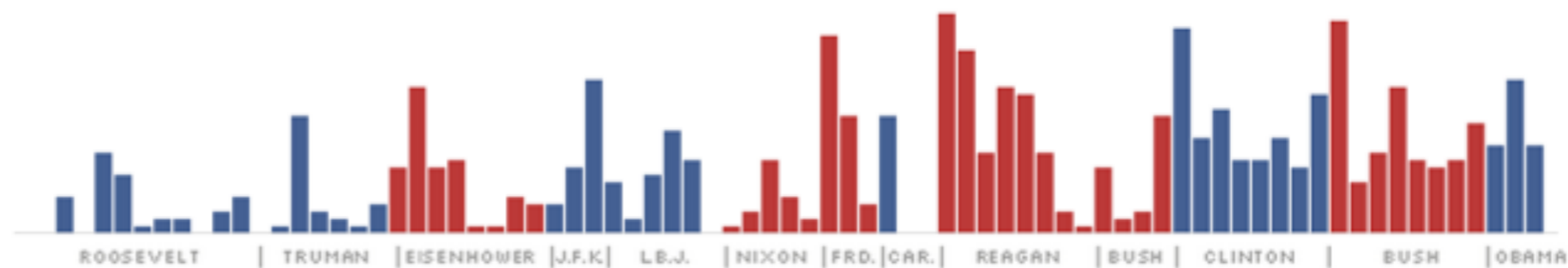


'health care'
became popular late

'tax'

Presidents have used the word every year since 1981, when Mr. Reagan uttered it 30 times, detailing his plan to reduce taxes and government spending.

TAX, TAXED, TAXES, TAXING



	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

← Word Vector
(Passage Vector)

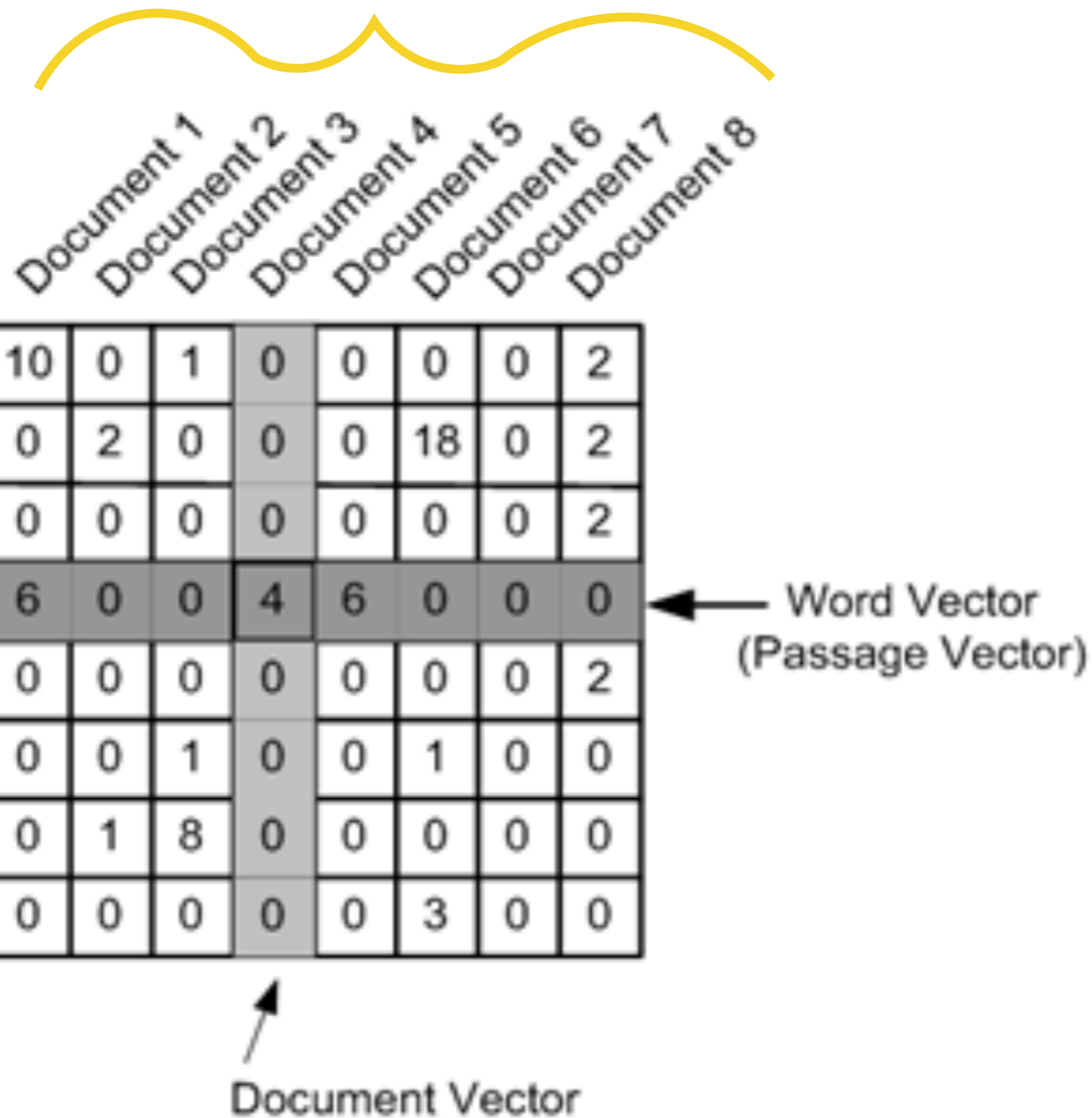
Document Vector

Bag of Words (BOW)

counts words per document

document corpus

vocabulary=
terms within
the corpus



	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

Word Vector (Passage Vector)

Document Vector

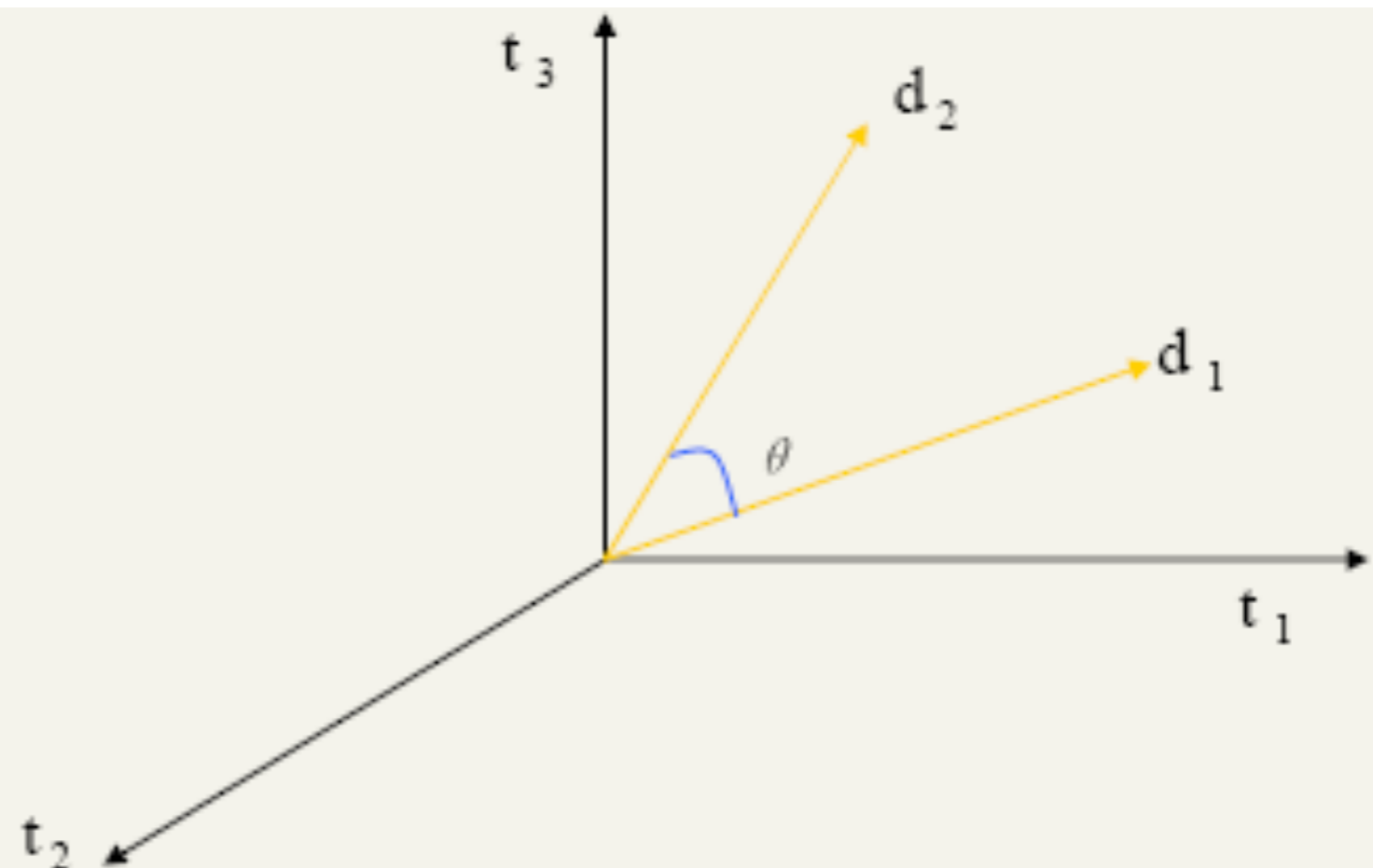
	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

← Word Vector
(Passage Vector)

Document Vector

	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

Document Vector



Document 1
Document 2
Document 3
Document 4
Document 5
Document 6
Document 7
Document 8

Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

Document Vector

t_2

t_3

d_2

d_1

t_1

θ

Document 1
Document 2
Document 3
Document 4
Document 5
Document 6
Document 7
Document 8

Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	3	0	0	0

Document Vector

t_2

t_3

d_2

d_1

θ

t_1

Cosine Similarity = θ (angle)
Distance between Document Vectors

Document 1 Document 2 Document 3 Document 4 Document 5 Document 6 Document 7 Document 8

Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	3	0	0	0

Document Vector

t_2

t_3

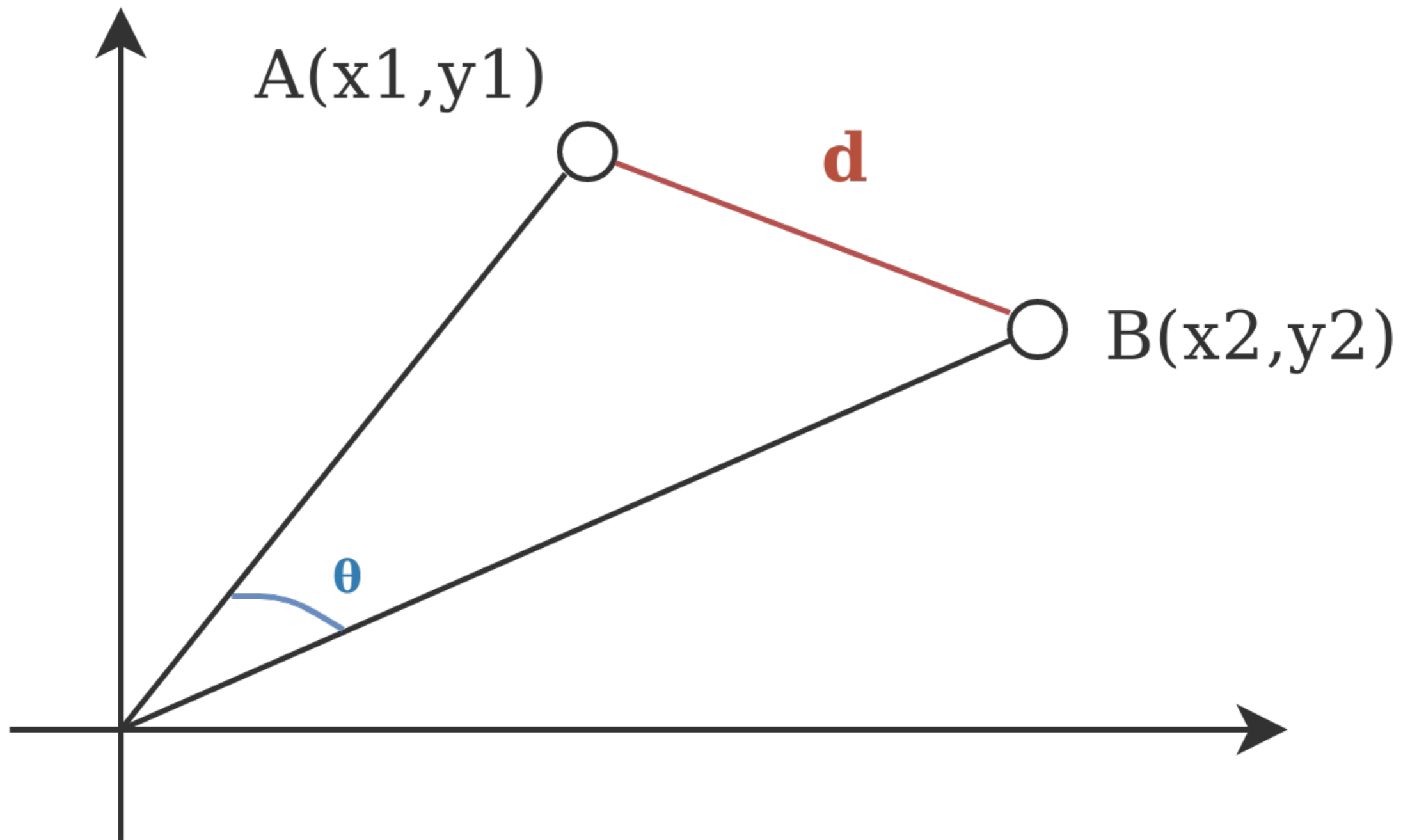
d_2

d_1

θ

t_1

Cosine Similarity - Why not using Euclidean Distance (Similarity)?



<https://cmry.github.io/notes/euclidean-v-cosine>

Is counting good enough?
Ideas for improvement?

Term Frequency - Inverse Document Frequency (TF-IDF)

We introduce a new weight
- consisting of two parts!

	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

Document Vector

term frequency (how often does the term appear in the current document?)
normalised according to document length

document frequency (how many documents contain the term?)

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

tf_{ij} = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

If a term appears in many documents
(term is not interesting & weight = low)

df = high
N/df = low
log(N/df) = low
weight = low

df = low
N/df = high
log(N/df) = high
weight = high

	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

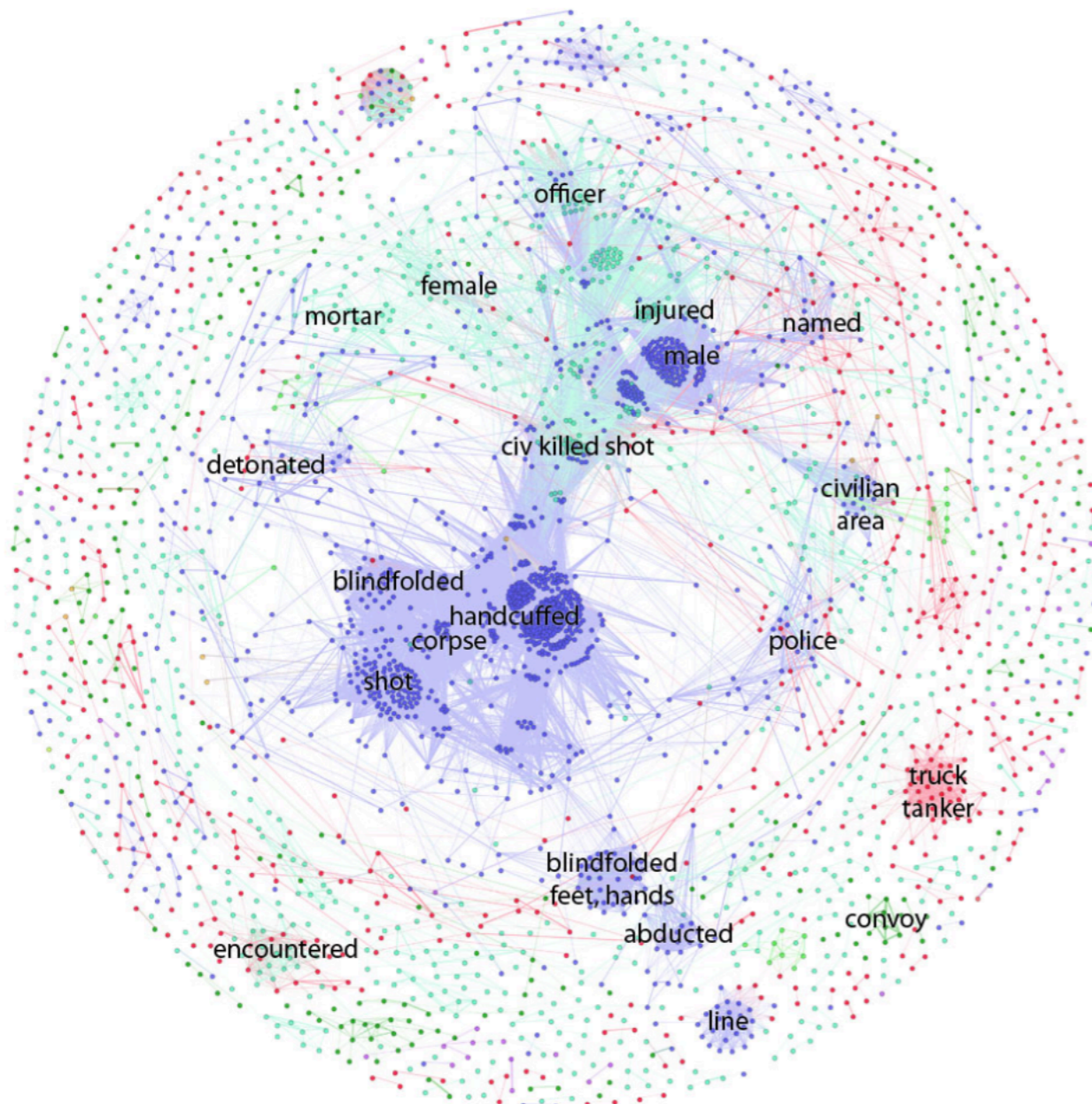
Document Vector

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

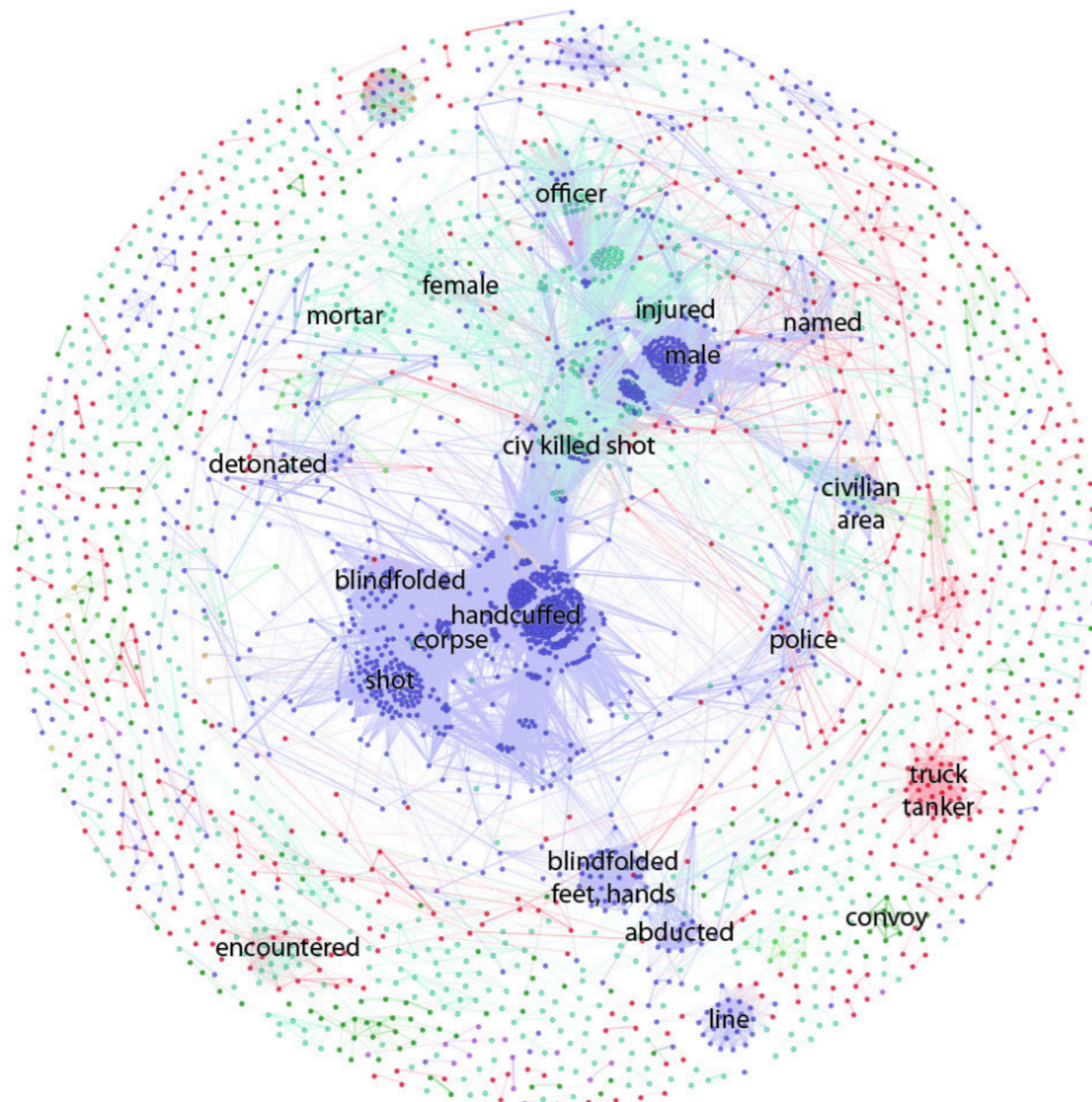
tf_{ij} = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

Document Term Matrix Visualization: WikiLeaks Iraq War Logs

How could we visualize such a matrix?



<http://jonathanstray.com/a-full-text-visualization-of-the-iraq-war-logs>



~ 400,000 war reports

**visualized as keyword clusters
according to TF-IDF
(cosine similarity)**

<http://jonathanstray.com/a-full-text-visualization-of-the-iraq-war-logs>

Can you think of any Limitations of the TF-IDF Modeling Approach?

Limitations of TF-IDF

- **no syntactic or semantic relationships** of words or passages
- **words** are treated independently and are **not comparable**
- **topics** (word concepts) are **not reflected** very well
- stop word removal needed (or other **preprocessing** steps) in order **to achieve meaningful results**

Distributed Word Embeddings

some versions:

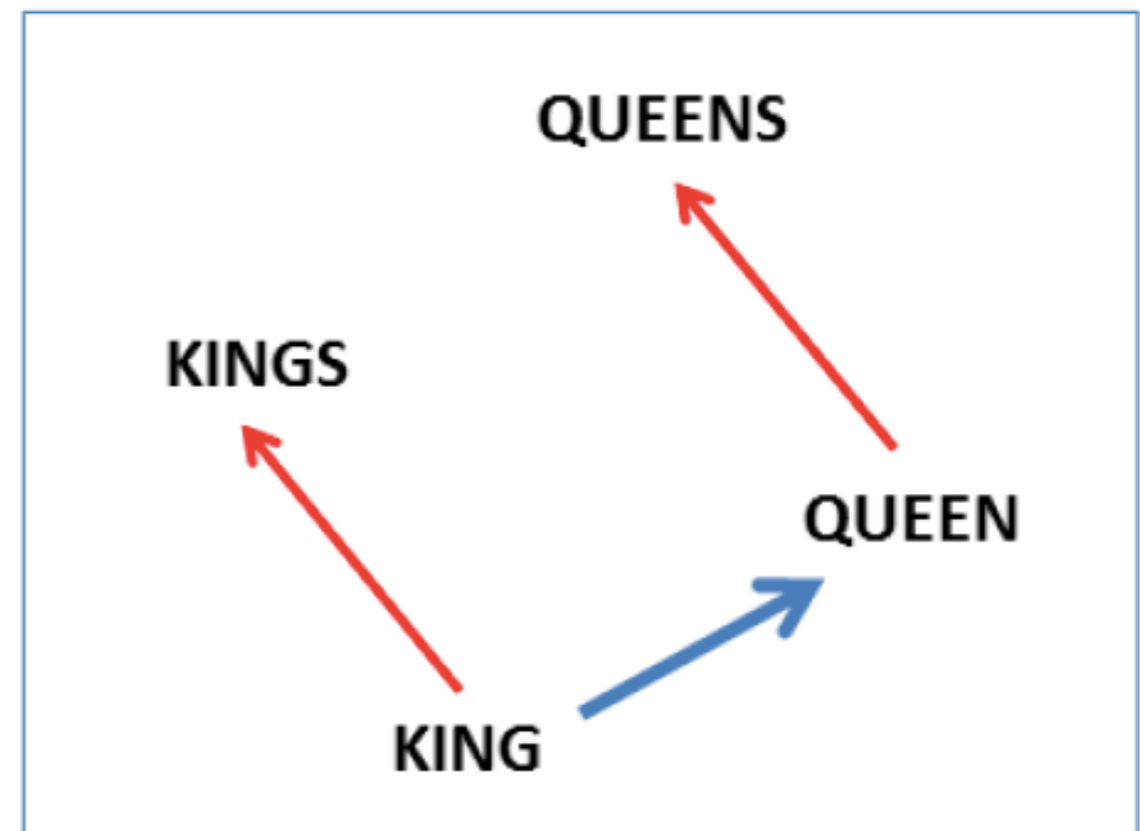
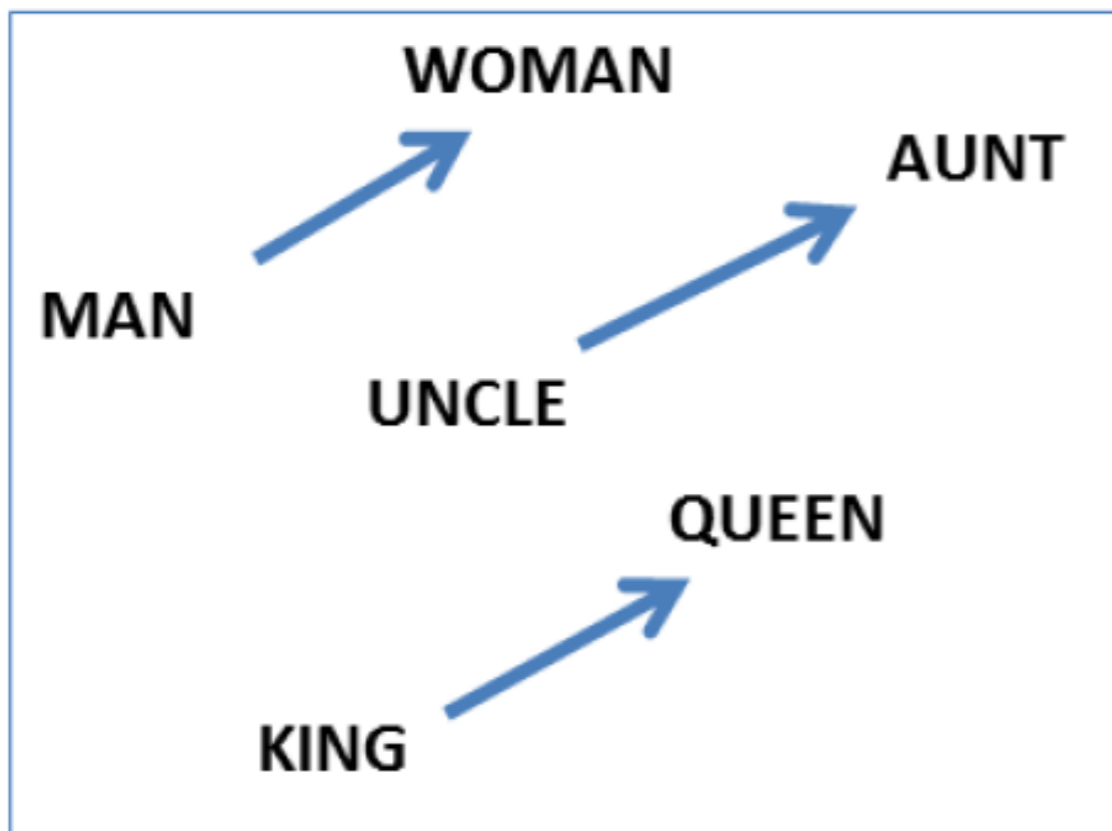
word2vec/glove/fasttext/bert/elmo/GPT-2

Words as distributions according to their context


- a word is a representation of all its context windows within a corpus (**context = N-gram** of which a word is part)
- similar **words** tend to have similar representations —> **comparable**
- **syntax** as well as **semantics** is reflected
- no document vectors are created, only word vectors

Efficient estimation of word representations in vector space

Tomas Mikolov et al., ICLR Workshop, 2013



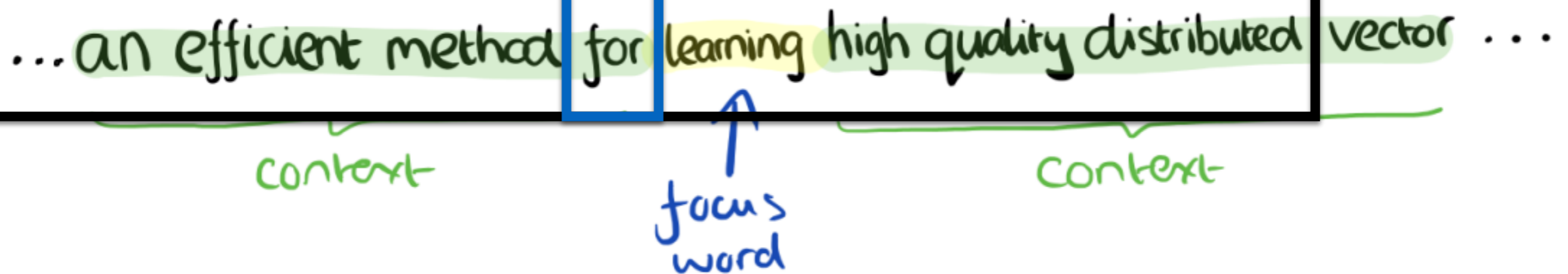
... an efficient method for learning high quality distributed vector ...



The diagram consists of a central blue arrow pointing upwards to the word 'learning'. Below the arrow is the text 'focus word'. To the left of the arrow is a green bracket under the phrase 'an efficient method for', with the word 'context' written below it. To the right of the arrow is another green bracket under the phrase 'high quality distributed vector', with the word 'context' written below it. The entire phrase 'an efficient method for learning high quality distributed vector' is highlighted in light green.

<https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

... an efficient method for learning high quality distributed vector ...

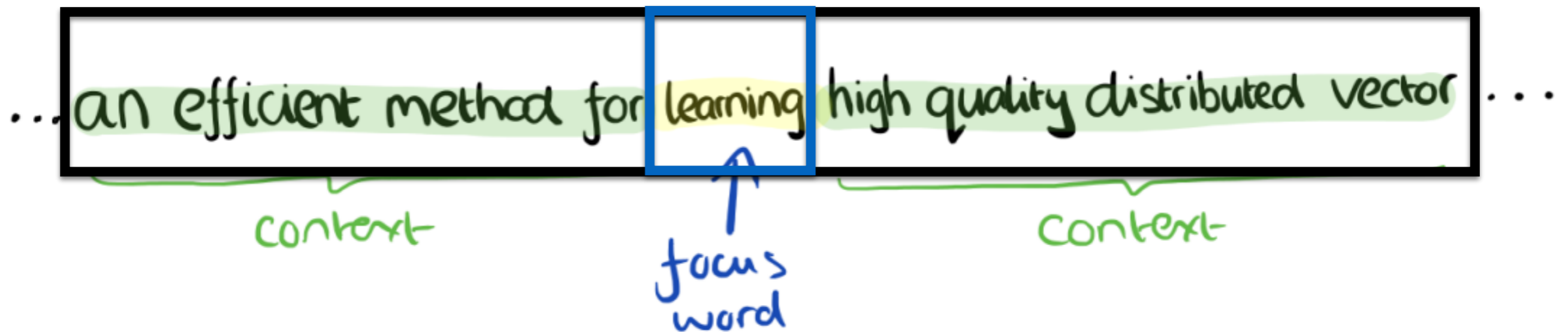


context

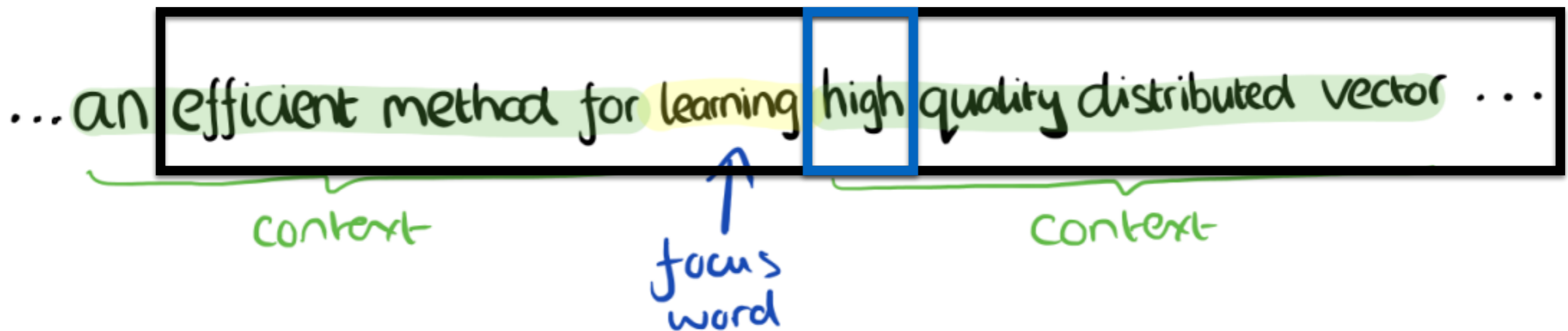
focus word

context

<https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

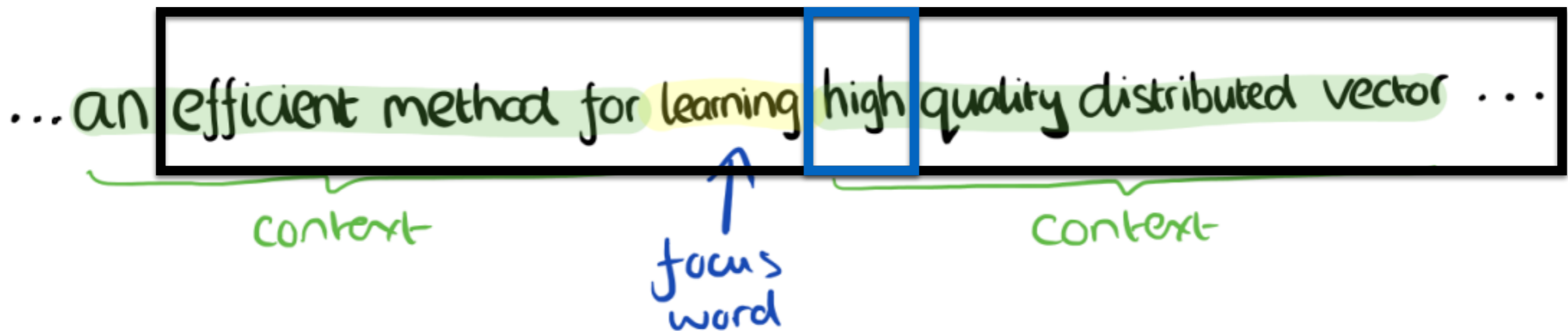


<https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>



<https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

We try to predict the current focus word with a Neural Network



<https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

First: Our **output reflects** the relation between all words within the vocabulary (**Co-Occurrences**).

After a dimensionality reduction (from maybe 10,000 words in vocabulary to 100 dimensions) we receive something related to “**concepts**” like.....

First: Our **output reflects** the relation between all words within the vocabulary (**Co-Occurrences**).

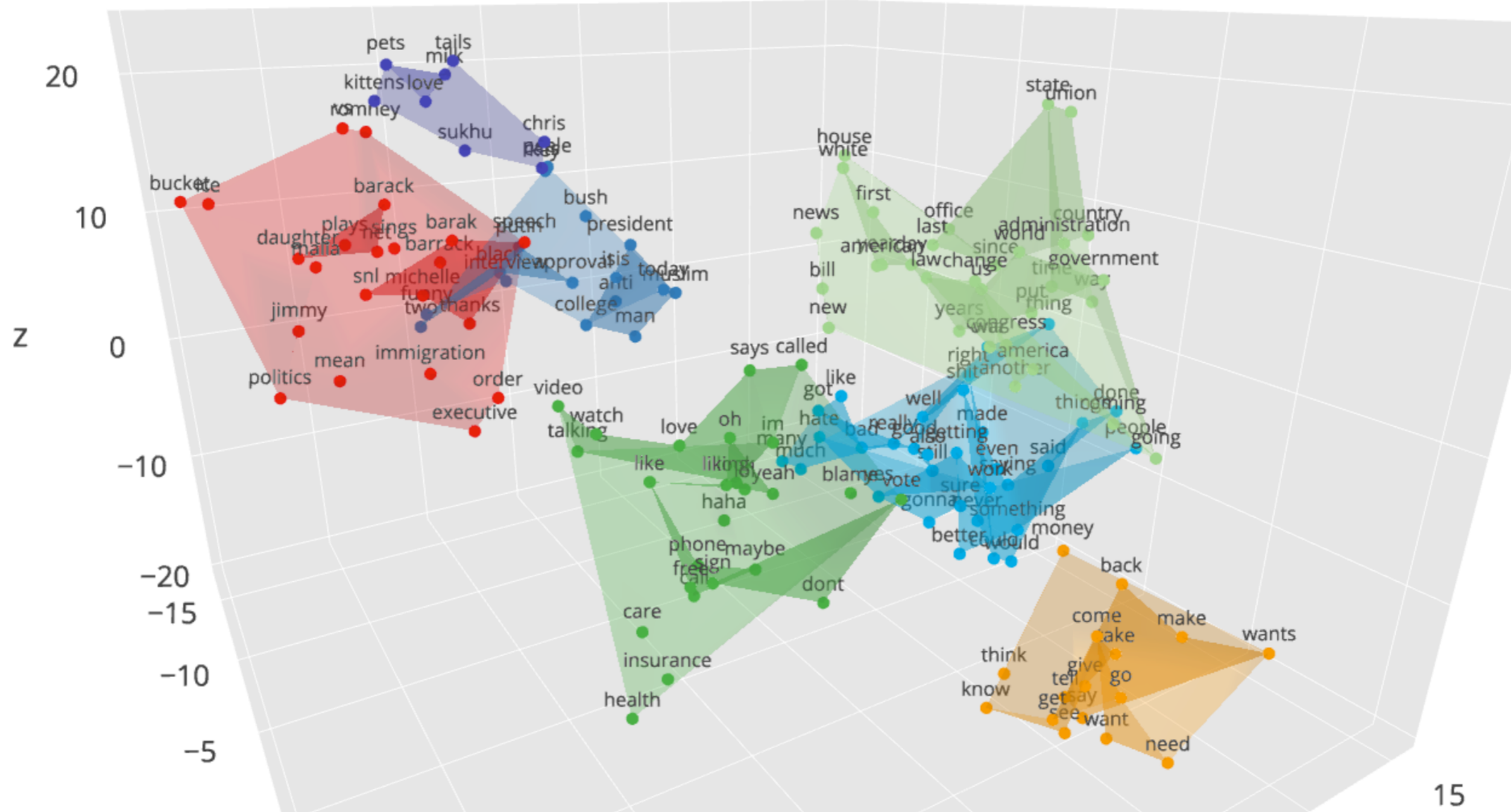
After a dimensionality reduction (from maybe 10,000 words in vocabulary to 100 dimensions) we receive something related to “**concepts**” like.....



<https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

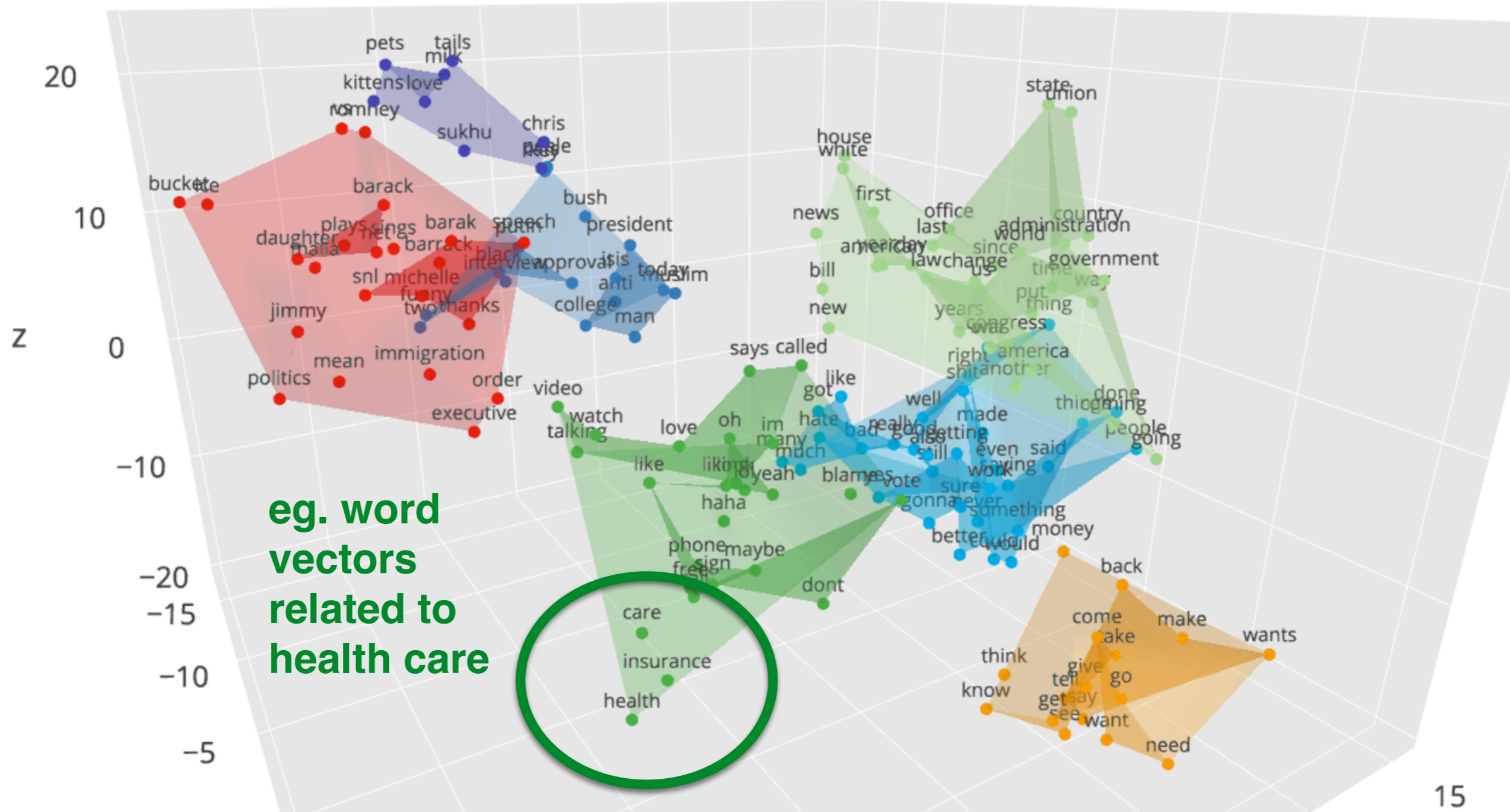
How could we visualize such vector spaces?

Obama Word2Vec Clustering

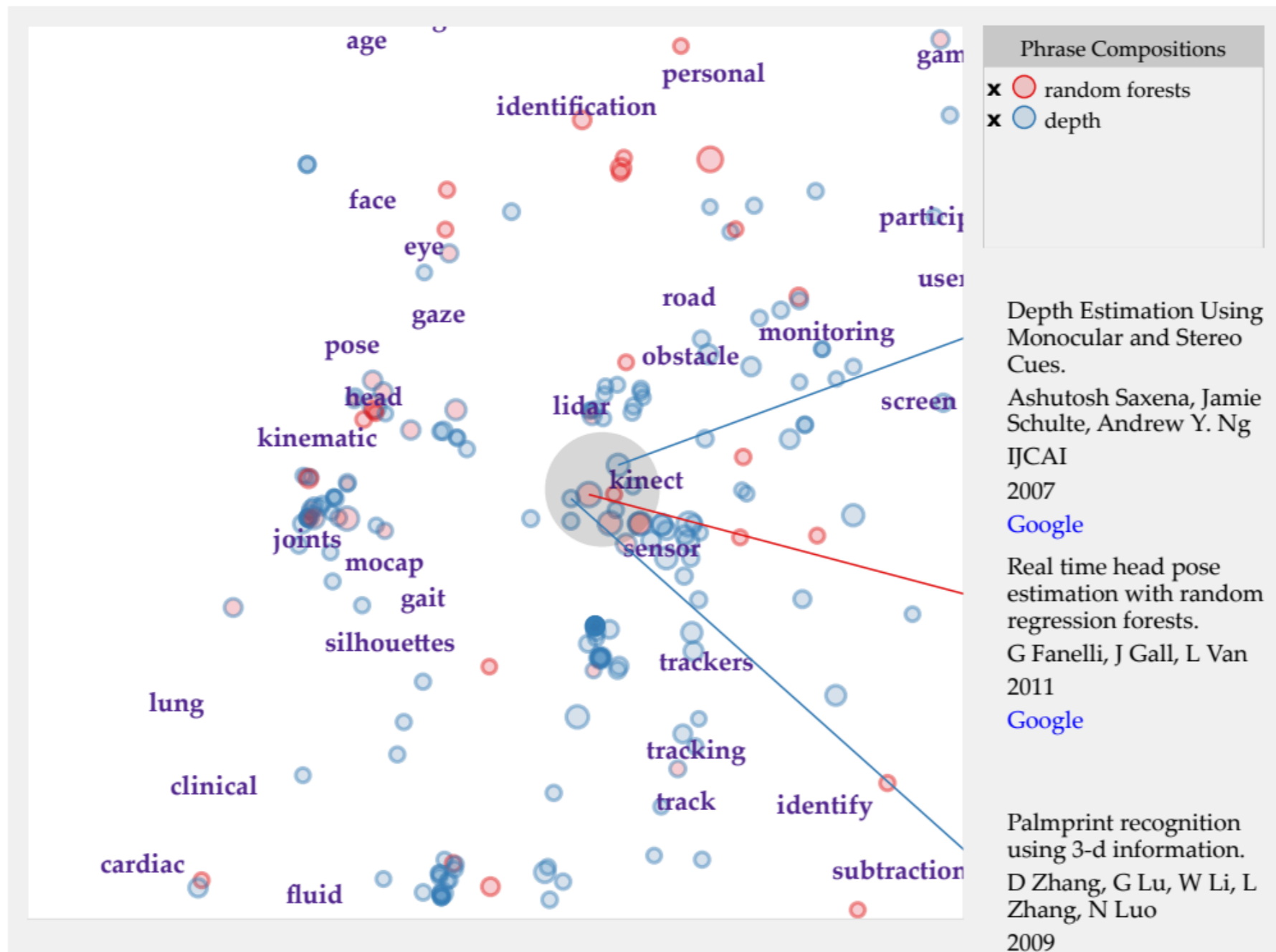


<https://plot.ly/~KevinAccount/18.embed>

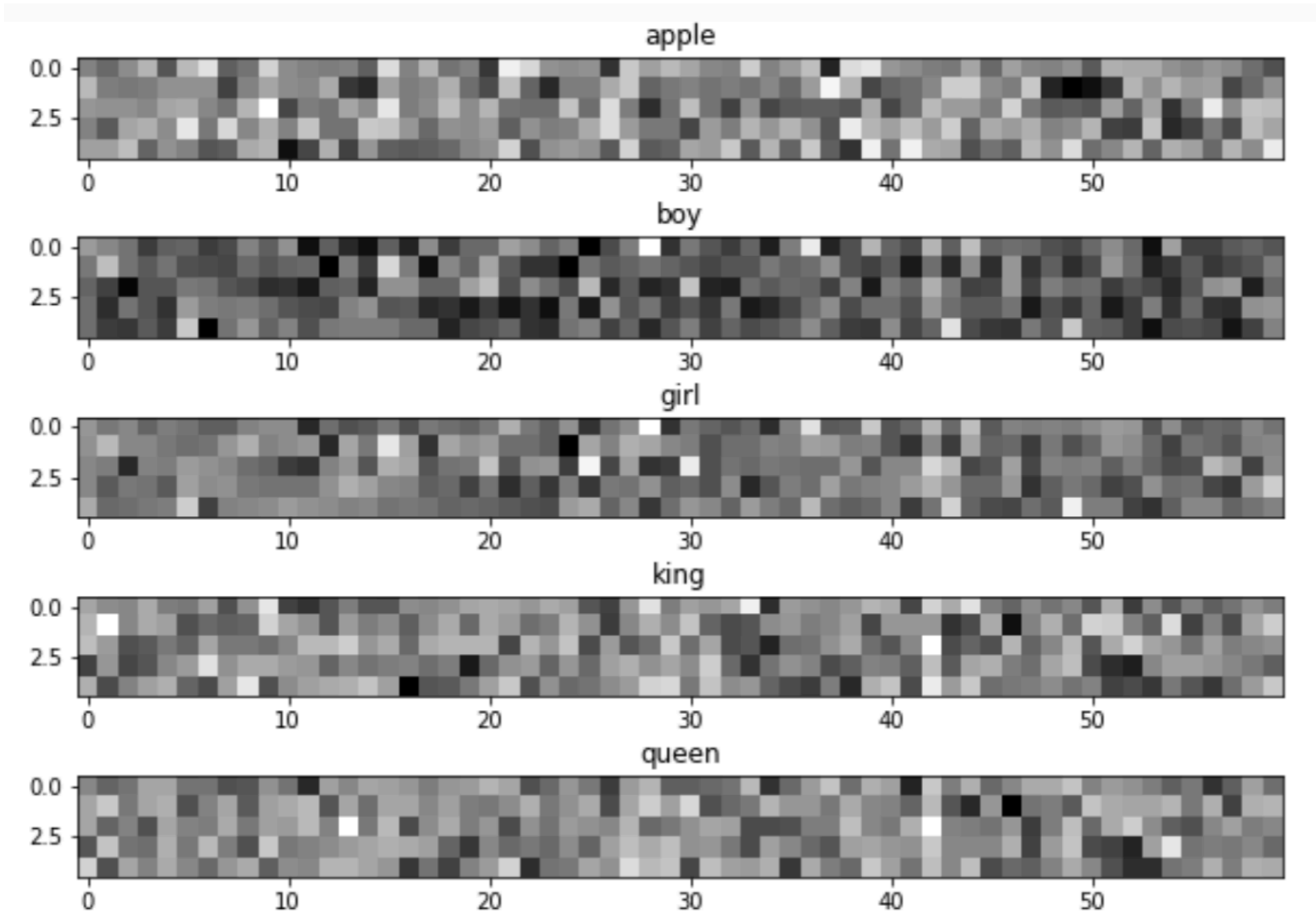
Obama Word2Vec Clustering



<https://plot.ly/~KevinAccount/18.embed>



Berger, Matthew, Katherine McDonough, and Lee M. Seversky. "cite2vec: Citation-driven document exploration via word embeddings." *IEEE transactions on visualization and computer graphics* 23.1 (2016): 691-700.



<https://towardsdatascience.com/visualisation-of-embedding-relations-word2vec-bert-64d695b7f36>

What are possible shortcomings of word2vec?

Limitations of Word 2 Vec

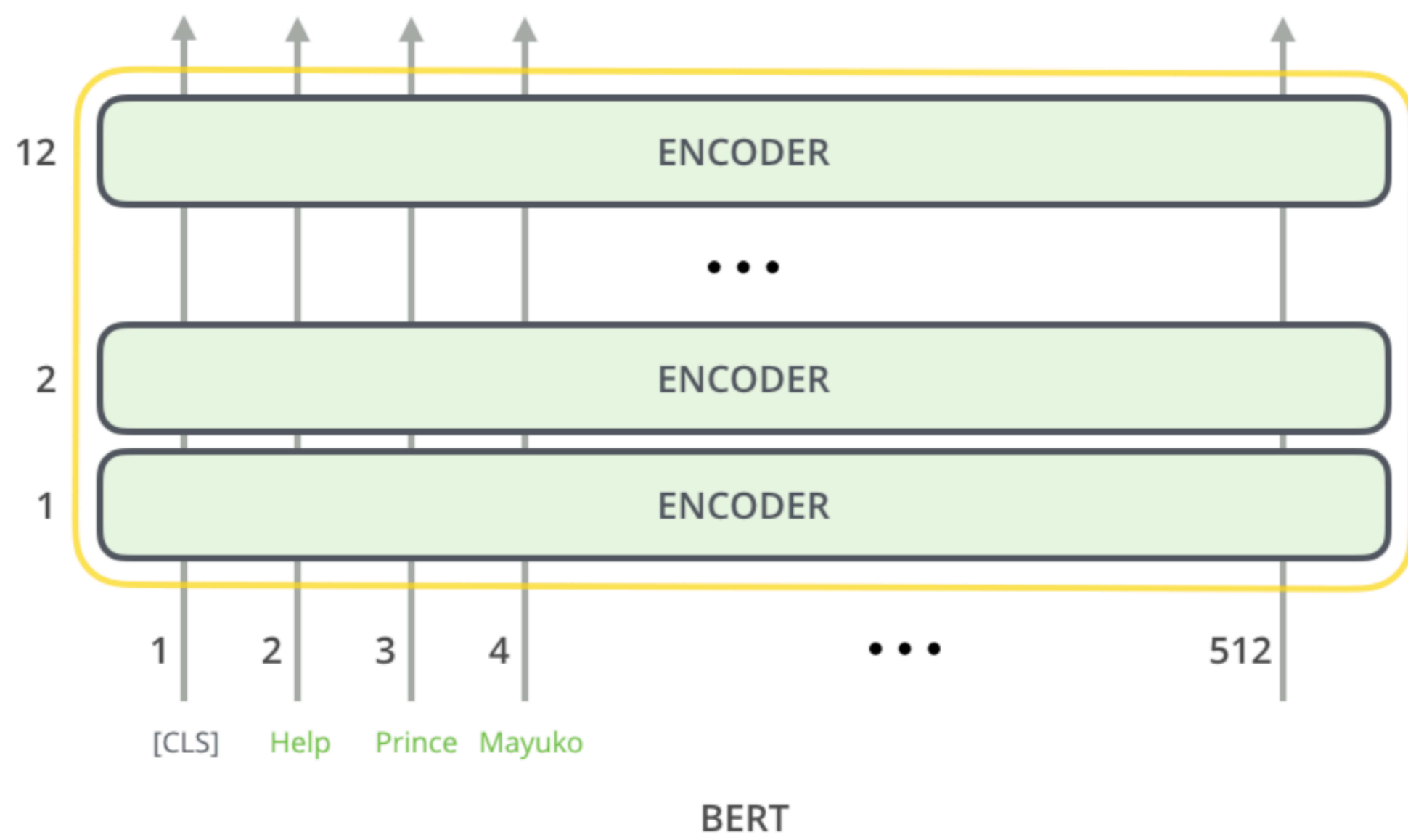
- different contexts get averaged

```
bank vault  
bank robber  
river bank
```

- how to represent documents? (doc2vec, ...)
- cannot handle unknown words

BERT (Bidirectional Encoder Representations
from Transformers)

ELMo (Embeddings from Language Models)



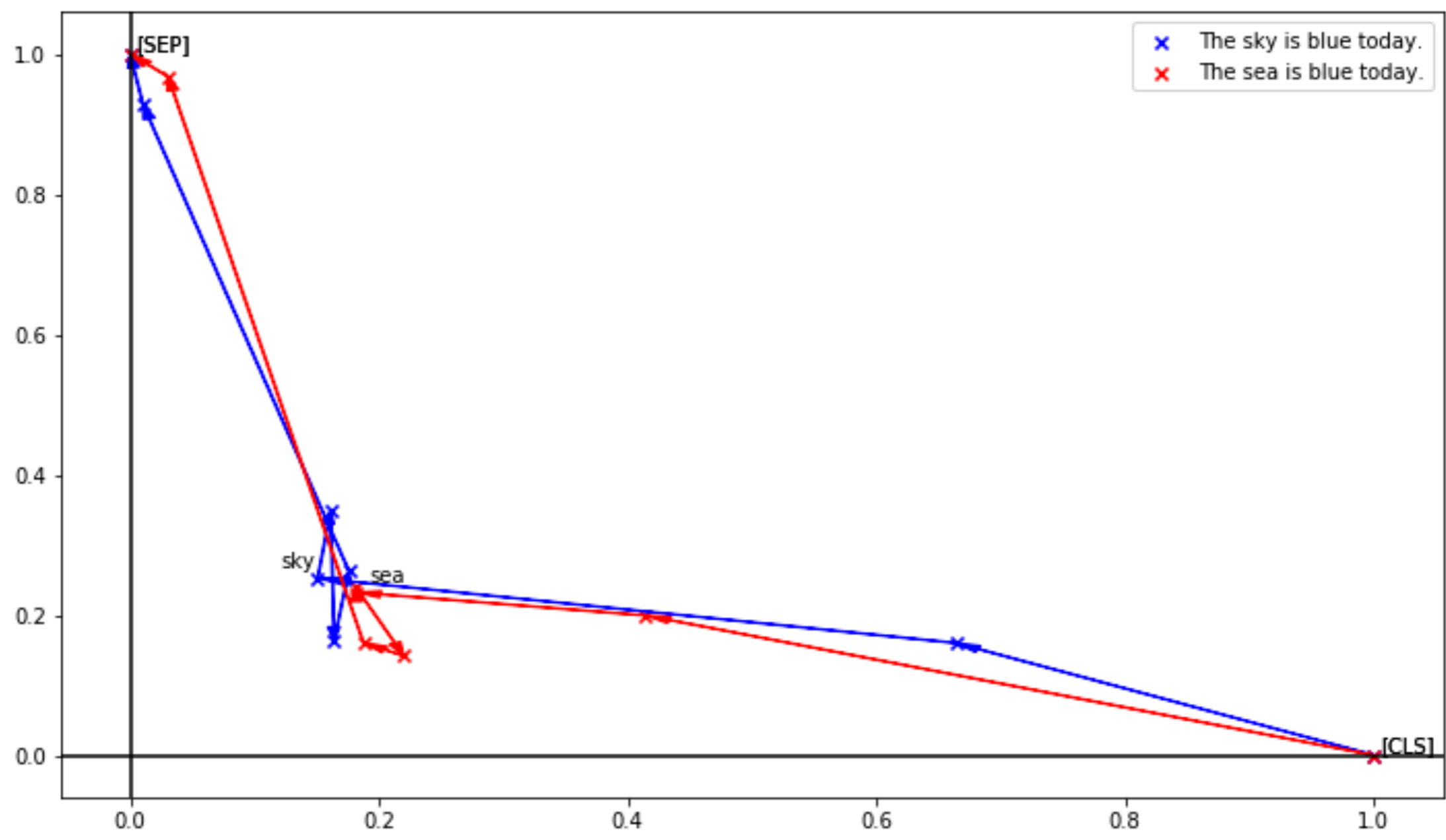
<https://jalammar.github.io/illustrated-bert/>

First 5 values for each meaning of "bank".

bank vault	tensor([2.1319, -2.1413, -1.6260, 0.8638, 3.3173])
bank robber	tensor([1.1868, -1.5298, -1.3770, 1.0648, 3.1446])
river bank	tensor([1.1295, -1.4725, -0.7296, -0.0901, 2.4970])

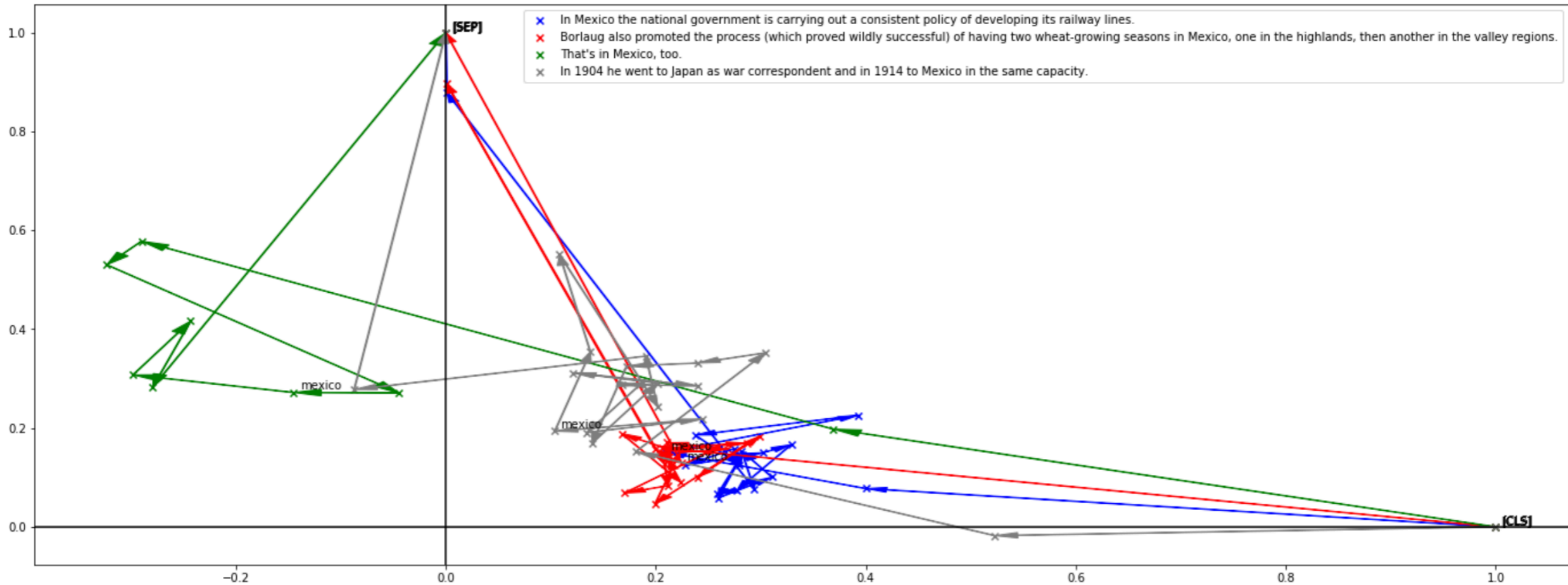
<https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/>

[CLS] This is a sentence example. [SEP]



<https://towardsdatascience.com/visualisation-of-embedding-relations-word2vec-bert-64d695b7f36>

[CLS] This is a sentence example. [SEP]



<https://towardsdatascience.com/visualisation-of-embedding-relations-word2vec-bert-64d695b7f36>

Topic Modeling

Probabilistic topic models

David Blei, Communications of the ACM 2012

Topics

gene 0.04
dna 0.02
genetic 0.01
...

life 0.02
evolve 0.01
organism 0.01
...

brain 0.04
neuron 0.02
nerve 0.01
...

data 0.02
number 0.02
computer 0.01
...

Documents

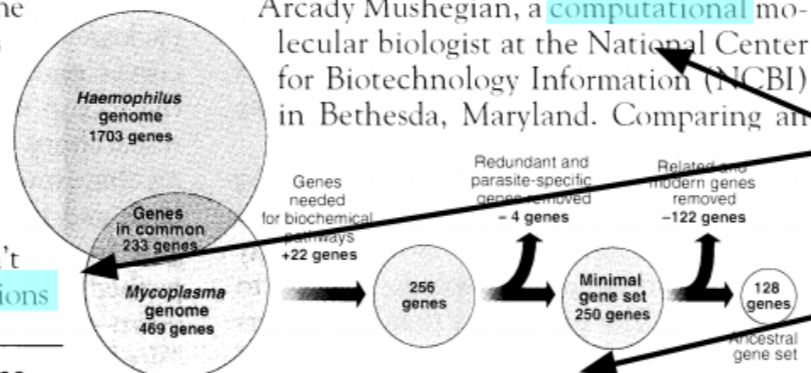
Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

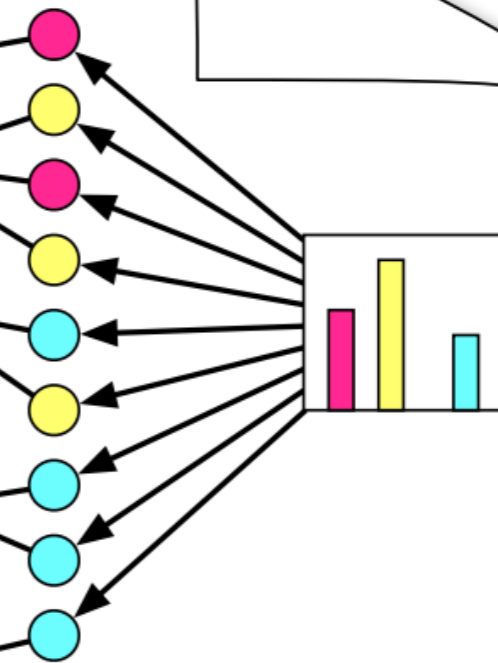
"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an



Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments



Input

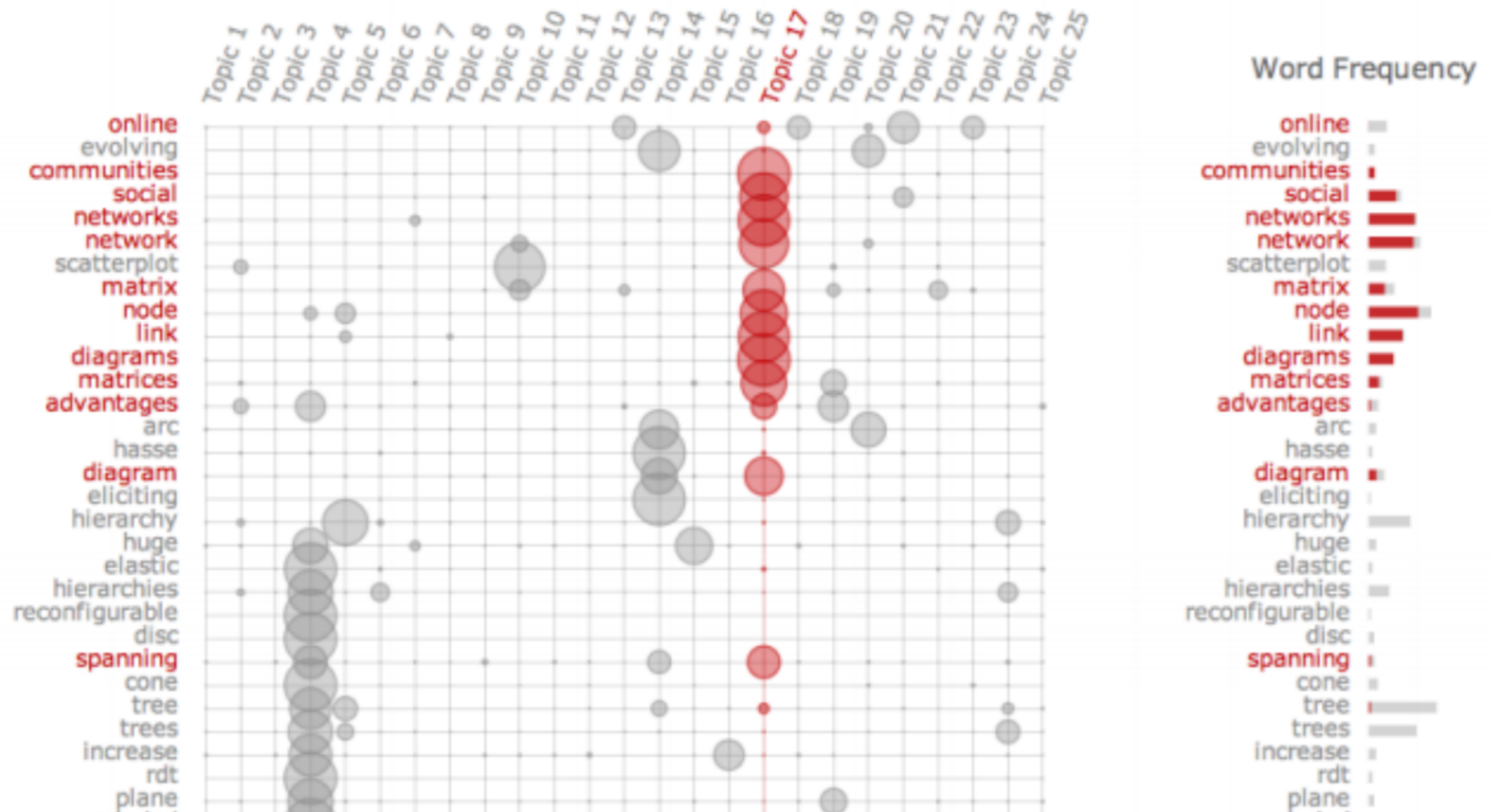
- Corpus = Text Sources (amount= M)
- Number of Topics (N) you want the model to cover

Output: two different kinds of distributions

- N Topic Distributions with probabilities for each word of the vocabulary
- M Text Distributions with probabilities for each of the N topics

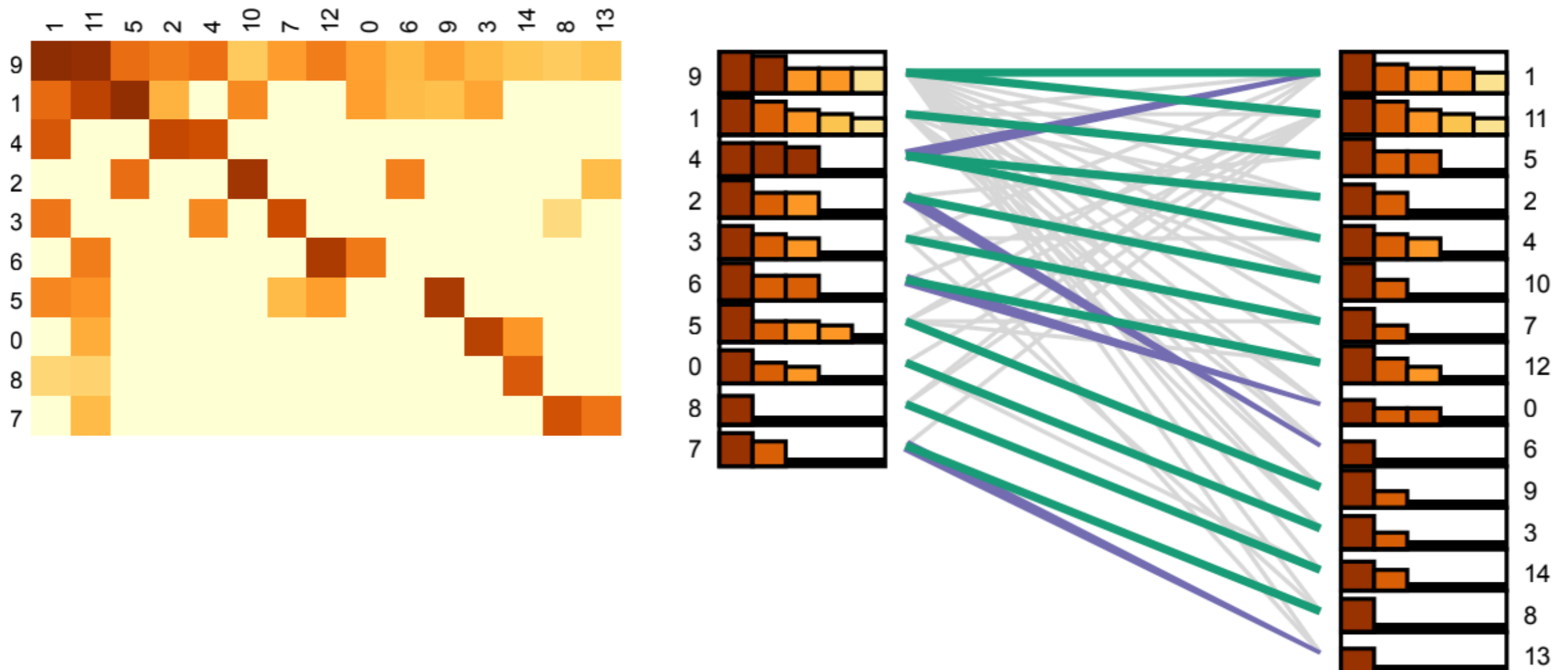
Termite: Visualization techniques for assessing textual topic models

Jason Chuang, Christopher D. Manning, Jeffrey Herr; Proceedings of the International Working Conference on Advanced Visual Interfaces 2012

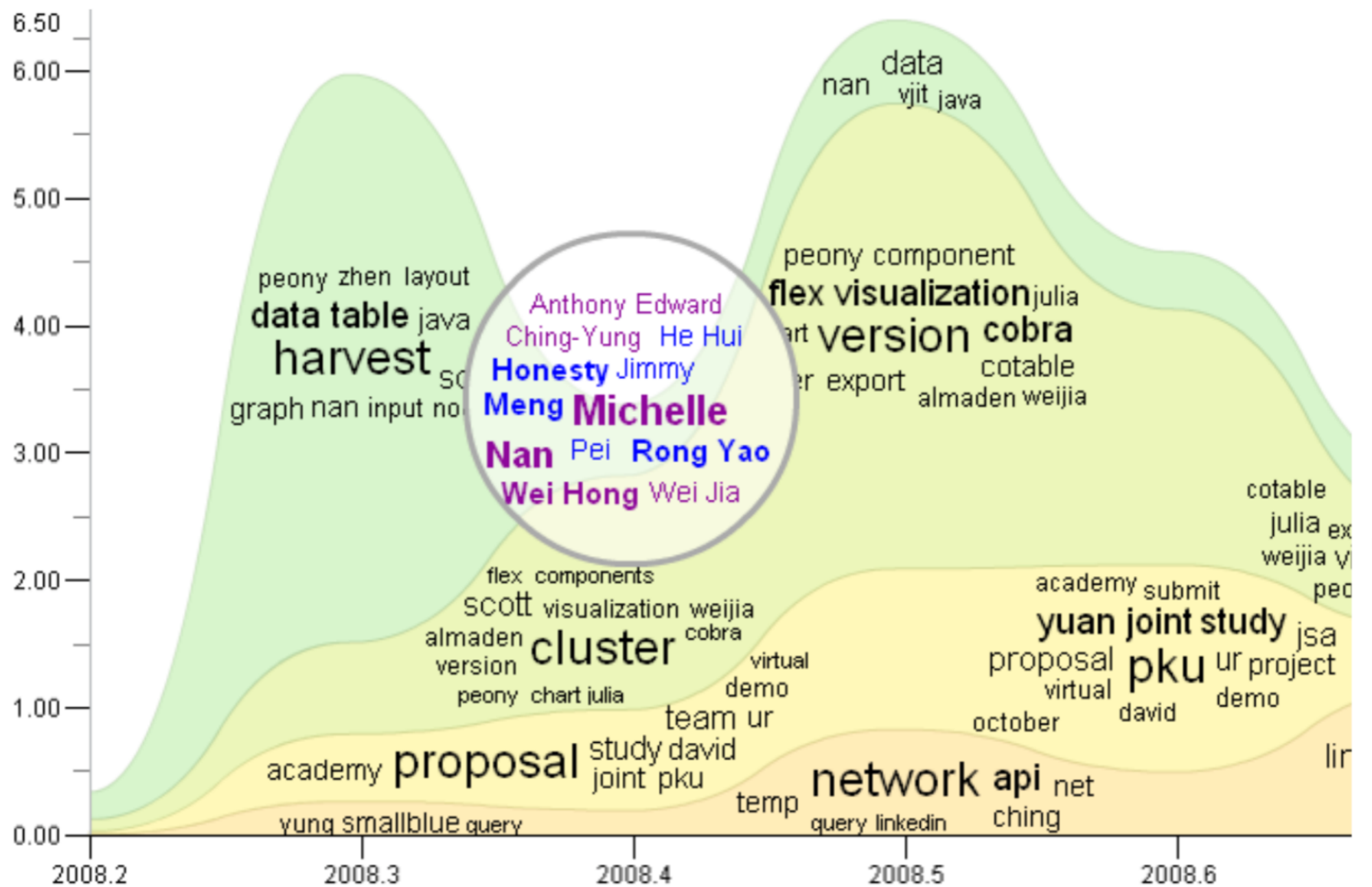


Serendip: Topic model-driven visual exploration of text corpora.

Alexander, E., Kohlmann, J., Valenza, R., Witmore, M., & Gleicher, M. (2014, October). In Visual Analytics Science and Technology (VAST), 2014

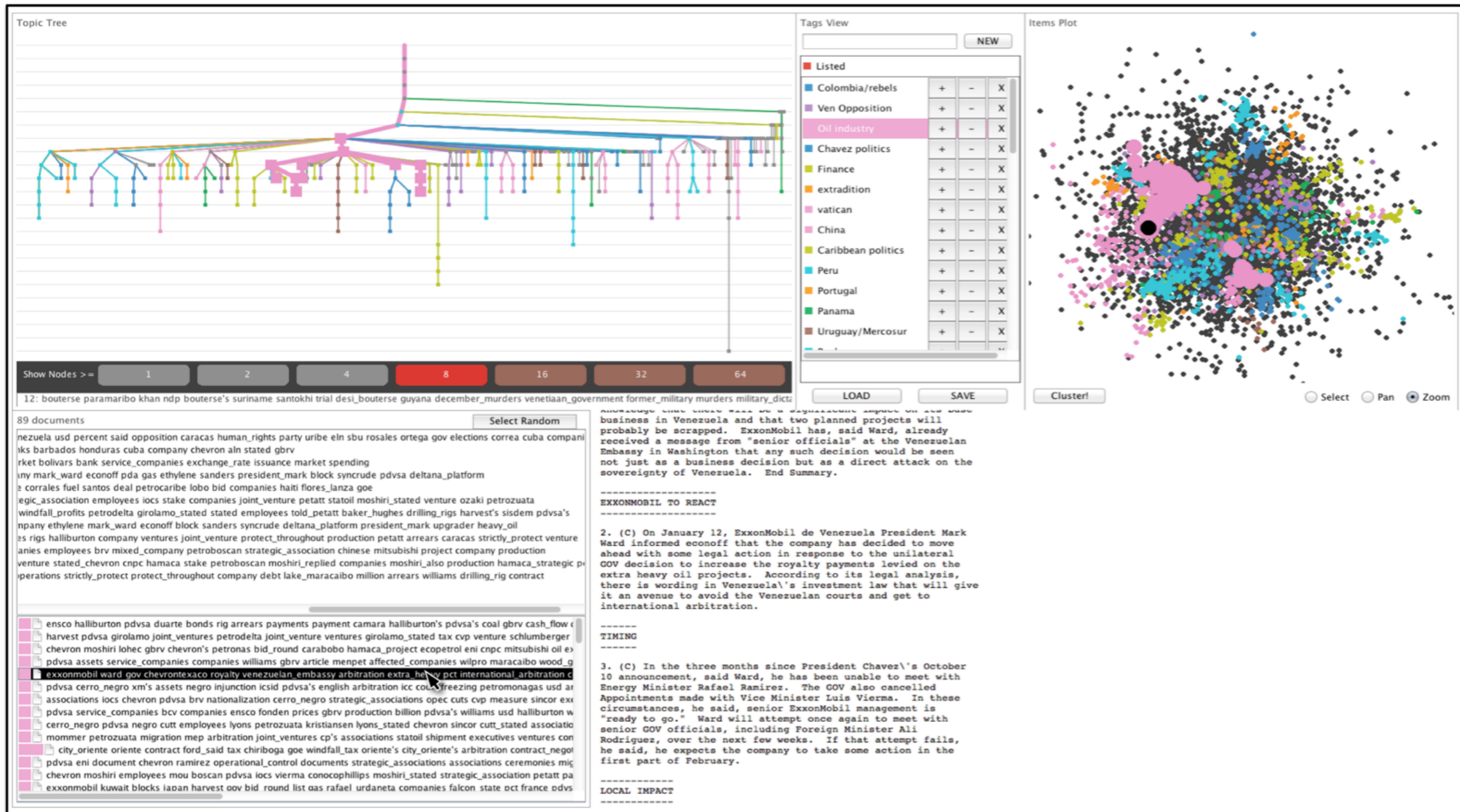


TIARA: A Visual Exploratory Text Analytic System, Wei et. al, KDD 2010)



Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists (Brehmer et al., IEEE InfoVis 2014)

old version of the software:



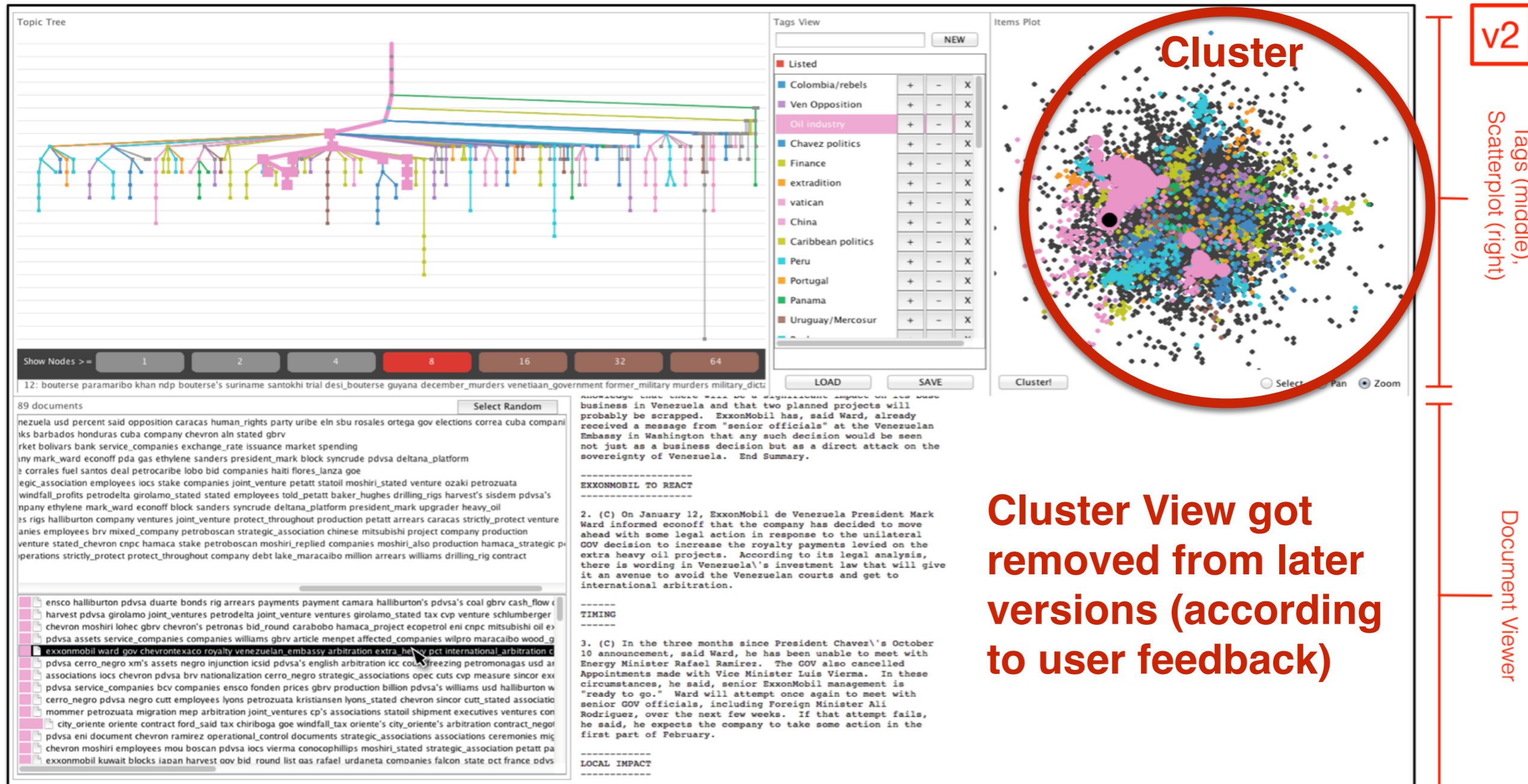
The screenshot displays the interface of the Visual Document Mining Tool, version 2 (v2). The interface is divided into several main sections:

- Topic Tree (left):** A hierarchical tree structure showing topics and their relationships. A control bar below it allows selecting the number of nodes to display (1, 2, 4, 8, 16, 32, 64).
- Tags View (middle):** A table listing various tags with columns for 'Listed', '+', '-', and 'X'. The 'Oil industry' tag is highlighted in pink.

Tag	Listed	+	-	X
Colombia/rebels		+	-	X
Ven Opposition		+	-	X
Oil industry		+	-	X
Chavez politics		+	-	X
Finance		+	-	X
extradition		+	-	X
vatican		+	-	X
China		+	-	X
Caribbean politics		+	-	X
Peru		+	-	X
Portugal		+	-	X
Panama		+	-	X
Uruguay/Mercosur		+	-	X
- Items Plot (right):** A scatterplot showing the distribution of items (documents) across the topic space. A 'Cluster!' button and 'Select', 'Pan', 'Zoom' options are visible.
- Cluster List (bottom left):** A list of 89 documents with a 'Select Random' button. The selected document is highlighted in pink.
- Document Viewer (bottom right):** A text viewer showing the content of the selected document, including sections like 'EXXONMOBIL TO REACT', 'TIMING', and 'LOCAL IMPACT'.

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists (Brehmer et al., IEEE InfoVis 2014)

old version of the software:



Topic Tree

Tags View

Tag	Listed	+	-	X
Colombia/rebels		+	-	X
Ven Opposition		+	-	X
Oil industry		+	-	X
Chavez politics		+	-	X
Finance		+	-	X
extradition		+	-	X
vatican		+	-	X
China		+	-	X
Caribbean politics		+	-	X
Peru		+	-	X
Portugal		+	-	X
Panama		+	-	X
Uruguay/Mercosur		+	-	X

Items Plot

Cluster

Cluster View got removed from later versions (according to user feedback)

Cluster List

89 documents

nezuela usd percent said opposition caracas human_rights party uribe eln sbu rosales ortega gov elections correa cuba compani
iks barbados honduras cuba company chevron ain stated gbrv
rket bolivars bank service_companies exchange_rate issuance market spending
ny mark_ward econoff pda gas ethylene sanders president_mark block syncrude pdvsa deltana_platform
e corrales fuel santos deal petrocaribe lobo bid companies haiti flores_lanza goe
egic_association employees iocs stake companies joint_venture petatt statoil moshiri_stated venture ozaki petrozuata
windfall_profits petrodelta girolamo_stated stated employees told_petatt baker_hughes drilling_rigs harvest's sisdem pdvsa's
npany ethylene mark_ward econoff block sanders syncrude deltana_platform president_mark upgrader heavy_oil
es rigs halliburton company ventures joint_venture protect_throughout production petatt arrears caracas strictly_protect venture
anies employees brv mixed_company petrobosc strategic_association chinese mitsubishi project company production
venture stated_chevron cnpc hamaca stake petrobosc moshiri_replied companies moshiri_also production hamaca_strategic_p
perations strictly_protect protect_throughout company debt lake_maraicao billion arrears williams drilling_rig contract

ensco halliburton pdvsa duarte bonds rig arrears payments payment camara halliburton's pdvsa's coal gbrv cash_flow
harvest pdvsa girolamo joint_ventures petrodelta joint_venture ventures girolamo_stated tax cvp venture schlumberger
chevron moshiri lohec gbrv chevron's petronas bid_round carabobo hamaca_project ecopetrol eni cnpc mitsubishi oil e
pdvsa assets service_companies companies williams gbrv article menpet affected_companies wilpro maracaibo wood_g
exxonmobil ward gov chevrontexaco royalty venezuelan_embassy arbitration extra_heavy_pct international_arbitration
pdvsa cerro_negro xm's assets negro injunction icisd pdvsa's english arbitration icc coul freezing petromonagas usd ar
associations iocs chevron pdvsa brv nationalization cerro_negro strategic_associations opec cuts cvp measure sincor ex
pdvsa service_companies bcv companies ensco fonden prices gbrv production billion pdvsa's williams usd halliburton w
cerro_negro pdvsa negro cutt employees lyons petrozuata kristiansen lyons_stated chevron sincor cutt_stated associatio
mommer petrozuata migration mep arbitration joint_ventures cp's associations statoil shipment executives ventures con
city_orient oriente contract ford_said tax chiriboga goe windfall_tax oriente's city_oriente's arbitration contract_nego
pdvsa eni document chevron ramirez operational_control documents strategic_associations associations ceremonies mic
chevron moshiri employees mou boscan pdvsa iocs vierma conocophillips moshiri_stated strategic_association petatt pa
exxonmobil kuwait blocks iaoran harvest oov bid round list oas rafael urdaneta companies falcon state oct france pdvs

Document Viewer

EXXONMOBIL TO REACT

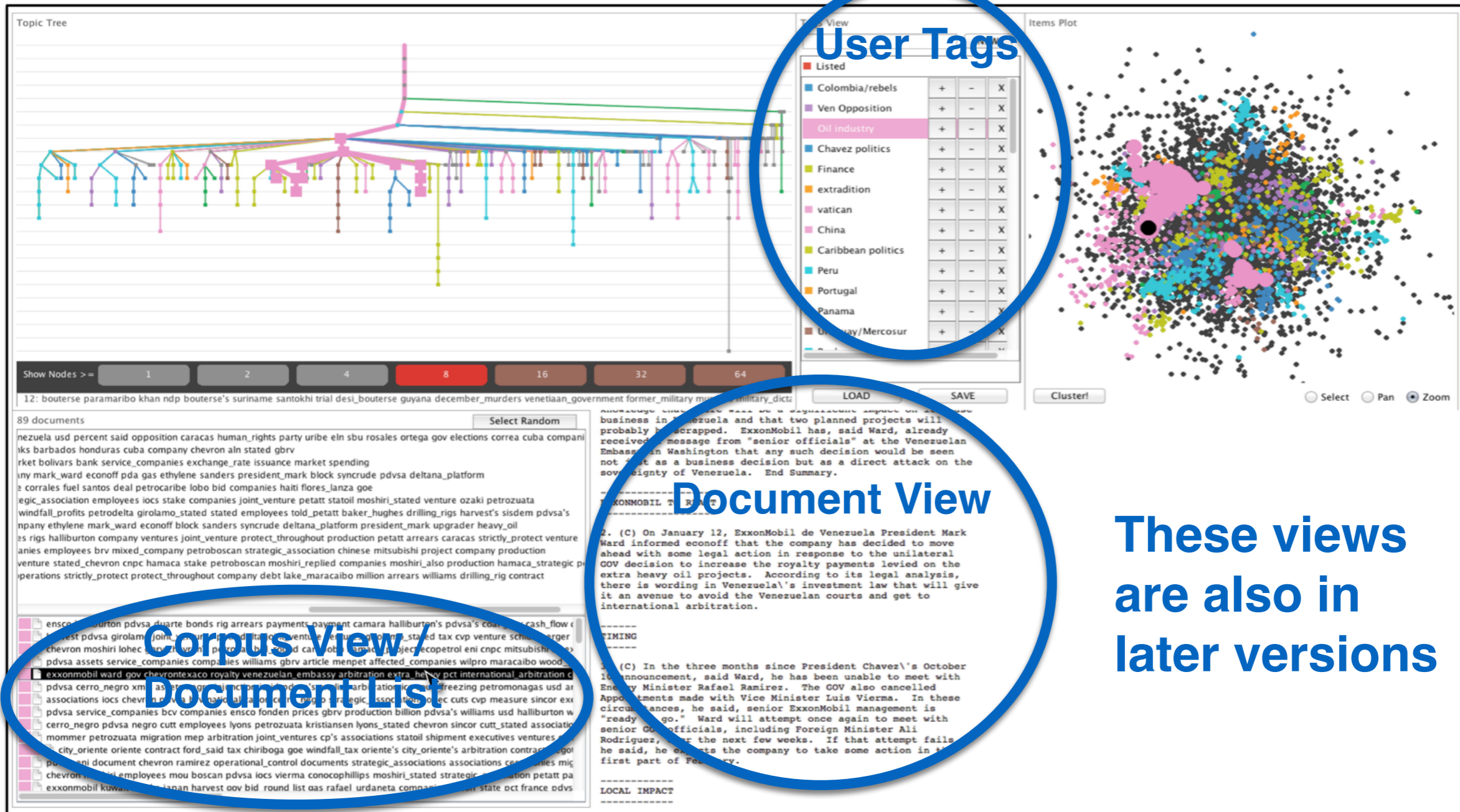
2. (C) On January 12, ExxonMobil de Venezuela President Mark Ward informed econoff that the company has decided to move ahead with some legal action in response to the unilateral GOV decision to increase the royalty payments levied on the extra heavy oil projects. According to its legal analysis, there is wording in Venezuela's investment law that will give it an avenue to avoid the Venezuelan courts and get to international arbitration.

3. (C) In the three months since President Chavez's October 10 announcement, said Ward, he has been unable to meet with Energy Minister Rafael Ramirez. The GOV also cancelled Appointments made with Vice Minister Luis Vierma. In these circumstances, he said, senior ExxonMobil management is "ready to go." Ward will attempt once again to meet with senior GOV officials, including Foreign Minister Ali Rodriguez, over the next few weeks. If that attempt fails, he said, he expects the company to take some action in the first part of February.

LOCAL IMPACT

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists (Brehmer et al., IEEE InfoVis 2014)

old version of the software:



The screenshot displays a complex interface for document mining. On the left, a 'Topic Tree' visualizes hierarchical relationships between topics. Below it, a 'Cluster List' shows 89 documents with a 'Select Random' button. At the bottom left, a 'Document List' provides a detailed view of document metadata. The central 'Document View' shows the full text of a document, with a blue circle highlighting a specific section. On the right, an 'Items Plot' scatterplot visualizes the distribution of documents across topics, with a blue circle highlighting a specific cluster. A 'User Tags' table is also visible, listing various topics and their associated tags.

Tag	+	-	X
Colombia/rebels			X
Ven Opposition			X
Oil industry			X
Chavez politics			X
Finance			X
extradition			X
vatican			X
China			X
Caribbean politics			X
Peru			X
Portugal			X
Panama			X
Uruguay/Mercosur			X

Document View

2. (C) On January 12, ExxonMobil de Venezuela President Mark Ward informed econoff that the company has decided to move ahead with some legal action in response to the unilateral GOV decision to increase the royalty payments levied on the extra heavy oil projects. According to its legal analysis, there is wording in Venezuela's investment law that will give it an avenue to avoid the Venezuelan courts and get to international arbitration.

3. (C) In the three months since President Chavez's October 10 announcement, said Ward, he has been unable to meet with Energy Minister Rafael Ramirez. The GOV also cancelled appointments made with Vice Minister Luis Vierma. In these circumstances, he said, senior ExxonMobil management is "ready to go." Ward will attempt once again to meet with senior GOV officials, including Foreign Minister Ali Rodriguez, over the next few weeks. If that attempt fails he said, he expects the company to take some action in the first part of February.

Document List

ensco halliburton pdvsa duarte bonds rig arrears payments payment camara halliburton's pdvsa's coal... cash_flow (

est pdvsa girolam joint venture... stated tax cvp venture sch... rger

chevron moshiri lohec... pdvsa assets service_companies companies williams gbrv article menpet affected_companies wilpro maracaibo wood...

exxonmobil ward gov chevron texaco royalty venezuelan embassy arbitration extra heavy pct international arbitration g

pdvsa cerro negro xm... pdvsa cerro negro xm... pdvsa cerro negro xm... pdvsa cerro negro xm...

associations iocs chevron... pdvsa service_companies bcv companies ensco fonden prices gbrv production billion pdvsa's williams usd halliburton w

cerro negro pdvsa negro cutt employees lyons petrozuata kristiansen lyons_stated chevron sincor cutt_stated associati

mommer petrozuata migration mep arbitration joint_ventures cp's associations statoil shipment executives ventures

city_orient oriente contract ford_said tax chiriboga goe windfall tax oriente's city_oriente's arbitration contract...

pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa pdvsa

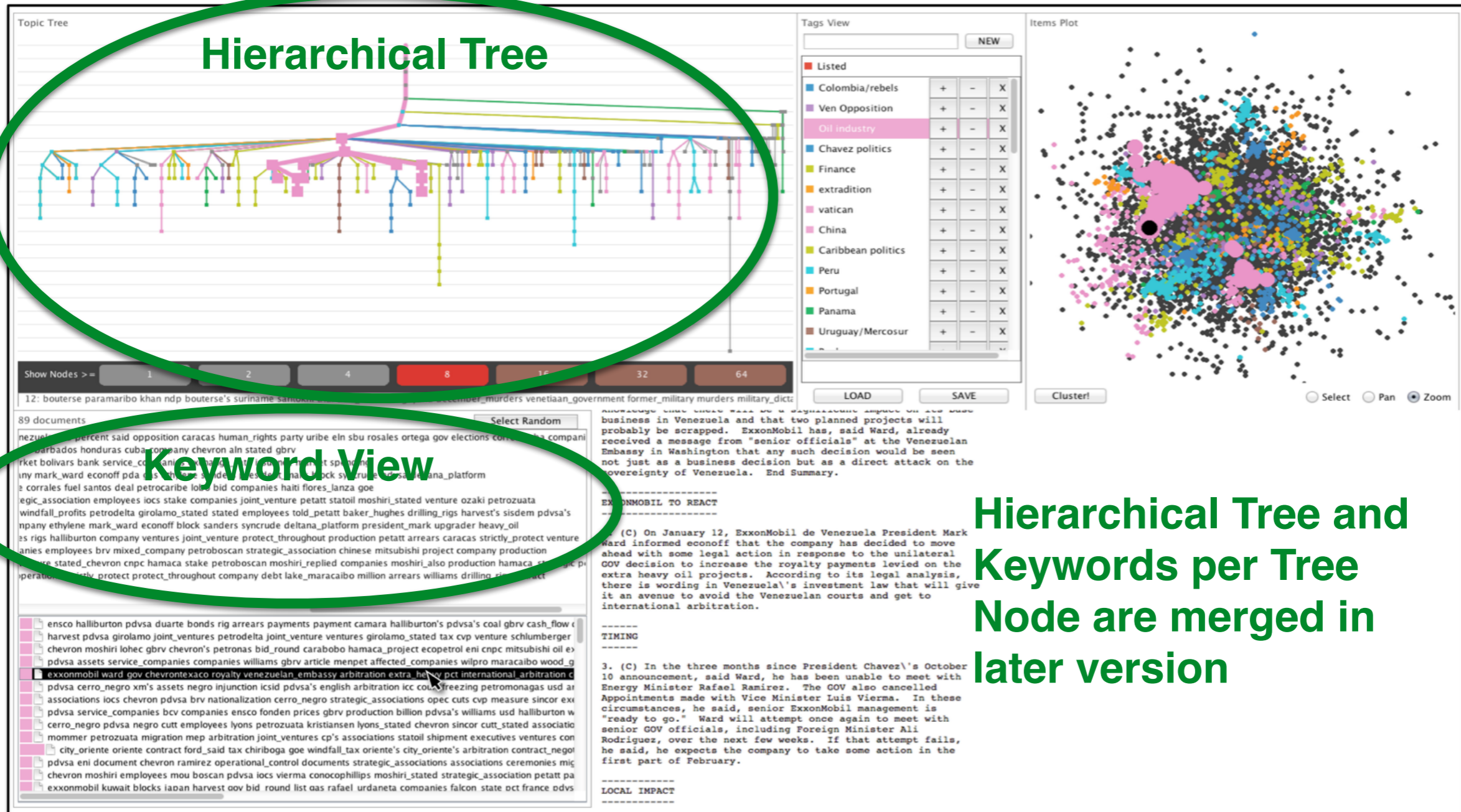
chevron moshiri employees mou boscan pdvsa iocs vierma conocophillips moshiri_stated strategic... pdvsa pdvsa

exxonmobil kuwa... japan harvest oov bid round list oas rafael urdaneta company... state oct france pdvs

These views are also in later versions

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists (Brehmer et al., IEEE InfoVis 2014)

old version of the software:



The screenshot displays the software interface with several key components:

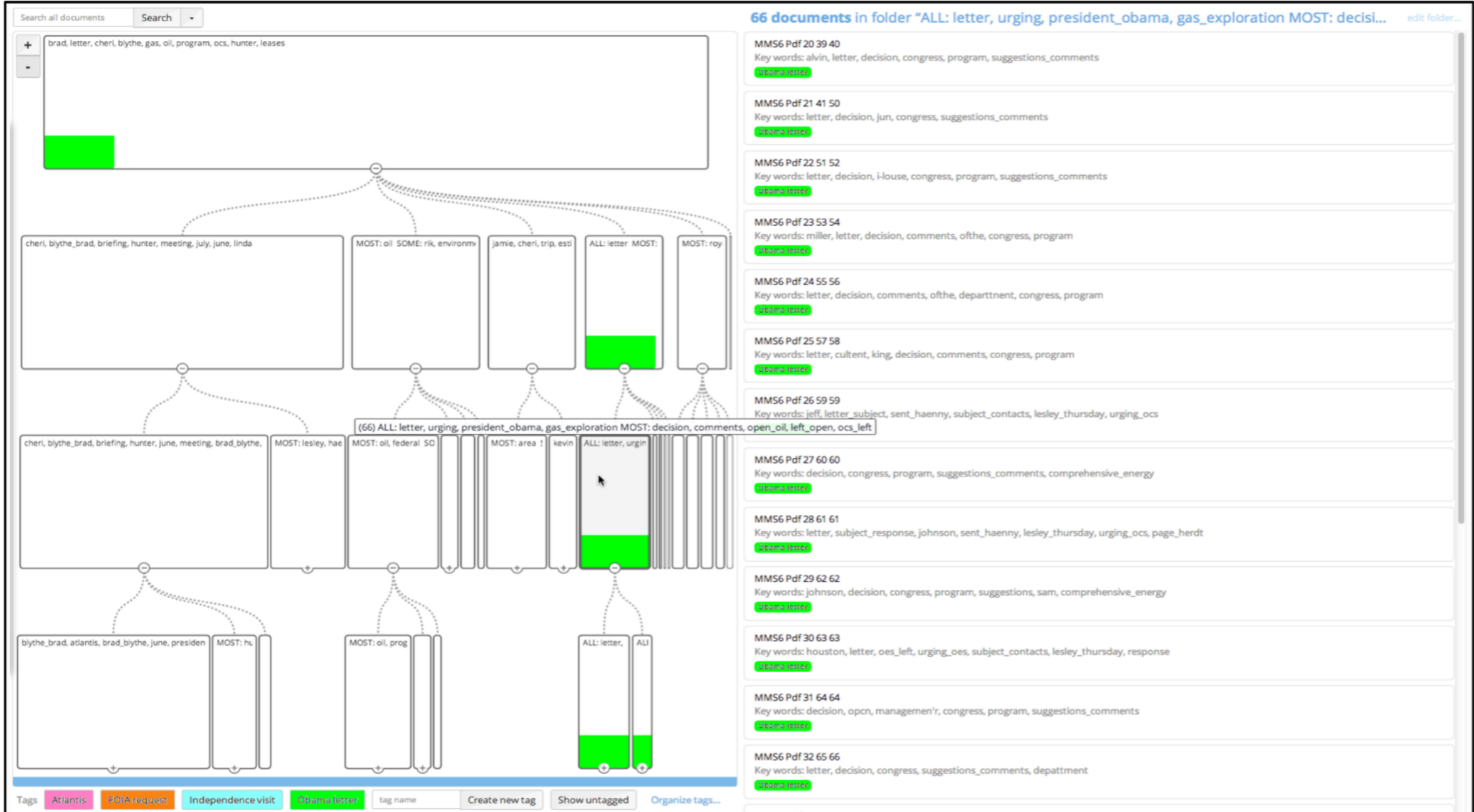
- Topic Tree:** A hierarchical tree structure on the left side, circled in green, showing relationships between topics.
- Keyword View:** A list of keywords at the bottom, also circled in green, corresponding to the nodes in the tree.
- Tags View:** A table in the middle-right showing a list of tags with columns for '+' and '-' signs, and a 'NEW' button.
- Items Plot:** A scatterplot on the right side showing a dense cluster of points, with a 'Cluster!' button below it.

Red annotations on the right side of the image identify these components: 'v2' in a red box, 'Tags (middle), Scatterplot (right)', and 'Document Viewer'. On the left side, red annotations identify 'Topic Tree', 'Cluster List', and 'Document List'.

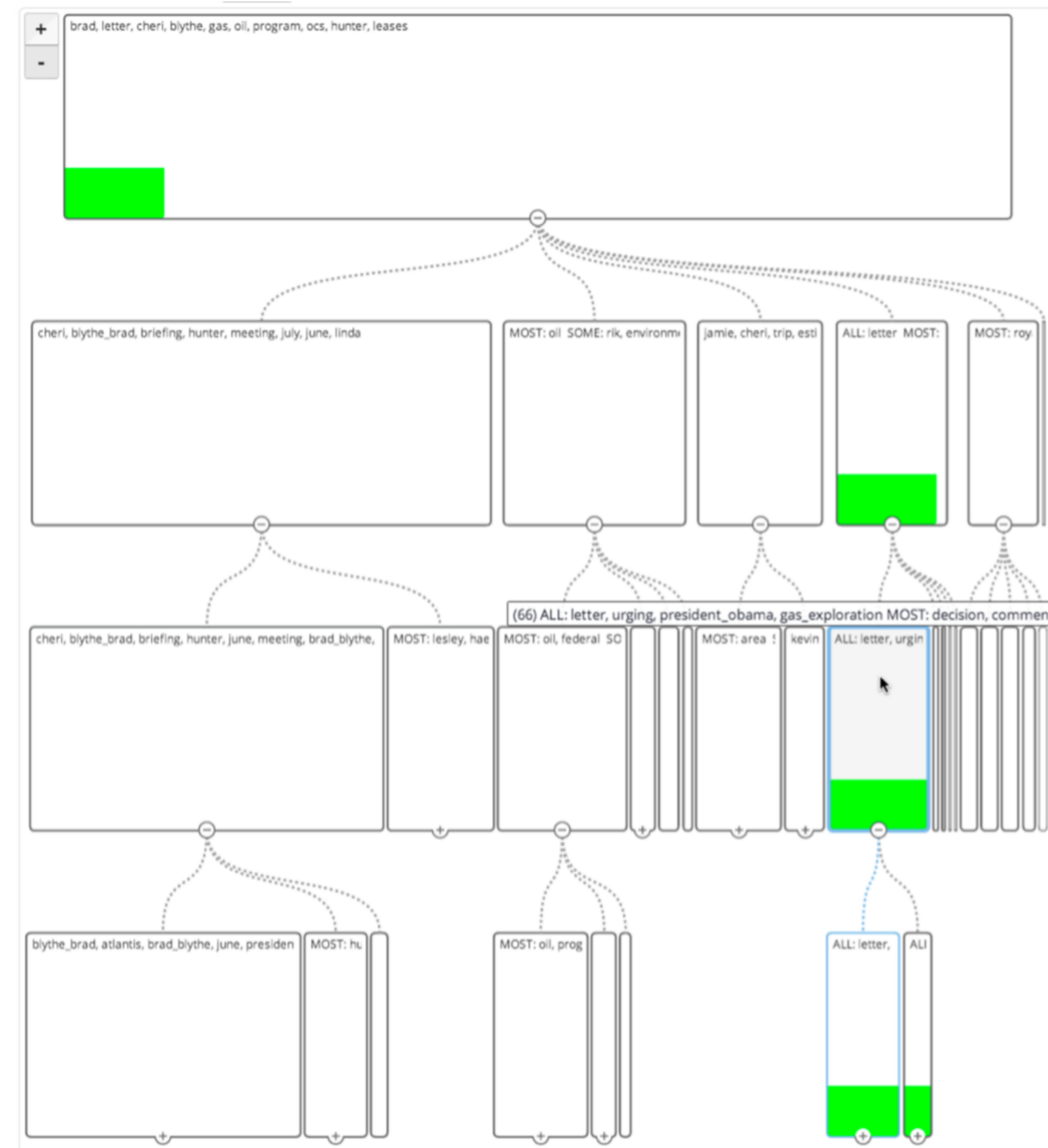
Hierarchical Tree and Keywords per Tree Node are merged in later version

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists (Brehmer et al., IEEE InfoVis 2014)

new version focuses on hierarchical tree according to cosine similarity (TF-IDF) of documents:



The screenshot displays a complex interface for document mining. On the left, a vertical red line is labeled 'Keyword Search' at the top and 'Topic Tree' below it. A search bar at the top left contains the text 'brad, letter, cheri, blythe, gas, oil, program, ocs, hunter, leases'. Below the search bar is a hierarchical tree structure. The root node is a box containing the search terms. It branches into several child nodes, each representing a topic. One node is highlighted in green. Below the tree, there is a 'Tags' section with several colored buttons: 'Atlantis', 'FOIA request', 'Independence visit', and 'Obama/leslie'. On the right side, a vertical red line is labeled 'Document List / Viewer'. This section contains a list of 66 documents. Each document entry includes a title (e.g., 'MMS6 Pdf 20 39 40'), key words, and a green button labeled 'View Document'. A red box with 'v4' is located in the top right corner of the interface.



- **Size of nodes** related to number of documents
- **Keywords located** within the corresponding nodes
- **Tags color-coded** according to percentage of fill level within node
- **Tree expandable**

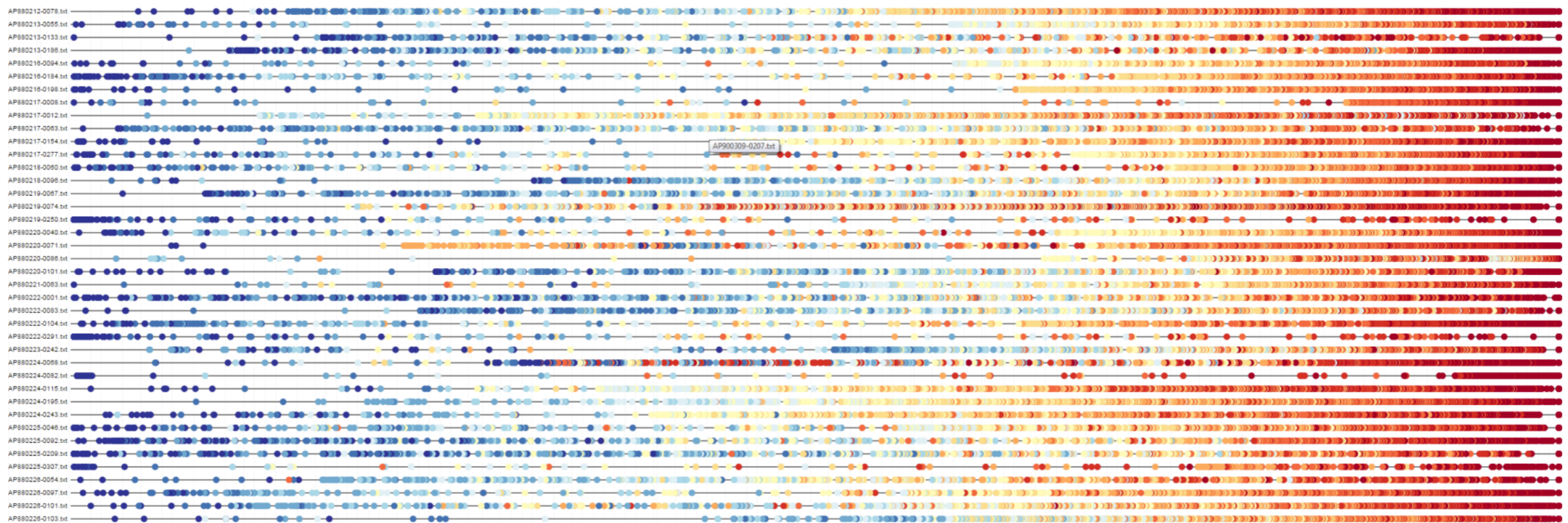


Fig. 1: Buddy plots show consistency of document relationships across topic models by encoding similarity with respect to individual documents. In this figure, each row represents a document, with the rest of the corpus encoded as circular glyphs along the row. Distance from the row's document in one model is encoded using horizontal position, while distance in a second model is encoded using color. This combination of encodings lets us see similarities from two models within one row of glyphs. Deviations in similarity between the two models can be identified as breaks from a smooth gradient. Though the two models seem to correlate well with documents at either extreme (blue documents to the left, red documents to the right), we see dramatic shifts between different classifications for documents in between, identified by breaks in the blue-to-red gradient structure.

Alexander, Eric, and Michael Gleicher. "Task-driven comparison of topic models." *IEEE transactions on visualization and computer graphics* 22.1 (2015): 320-329.

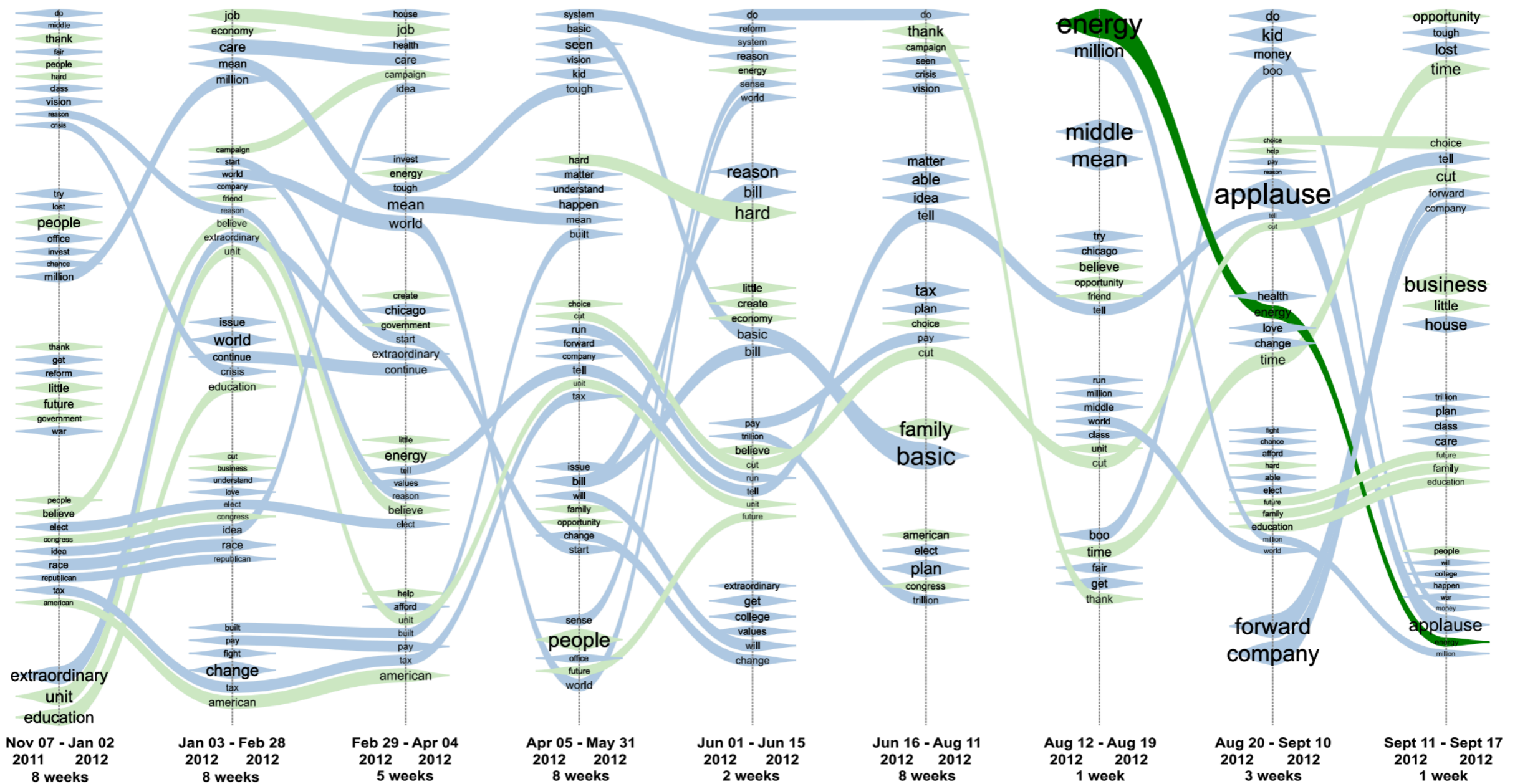


Fig. 1. ThemeDelta visualization for Barack Obama's campaign speeches during the U.S. 2012 presidential election (until September 10, 2012). Green lines are shared terms between Obama and Romney. Data from the "The American Presidency Project" at UCSB (<http://www.presidency.ucsb.edu/>).

Cui, Weiwei, et al. "Textflow: Towards better understanding of evolving topics in text." *IEEE transactions on visualization and computer graphics* 17.12 (2011): 2412-2421.

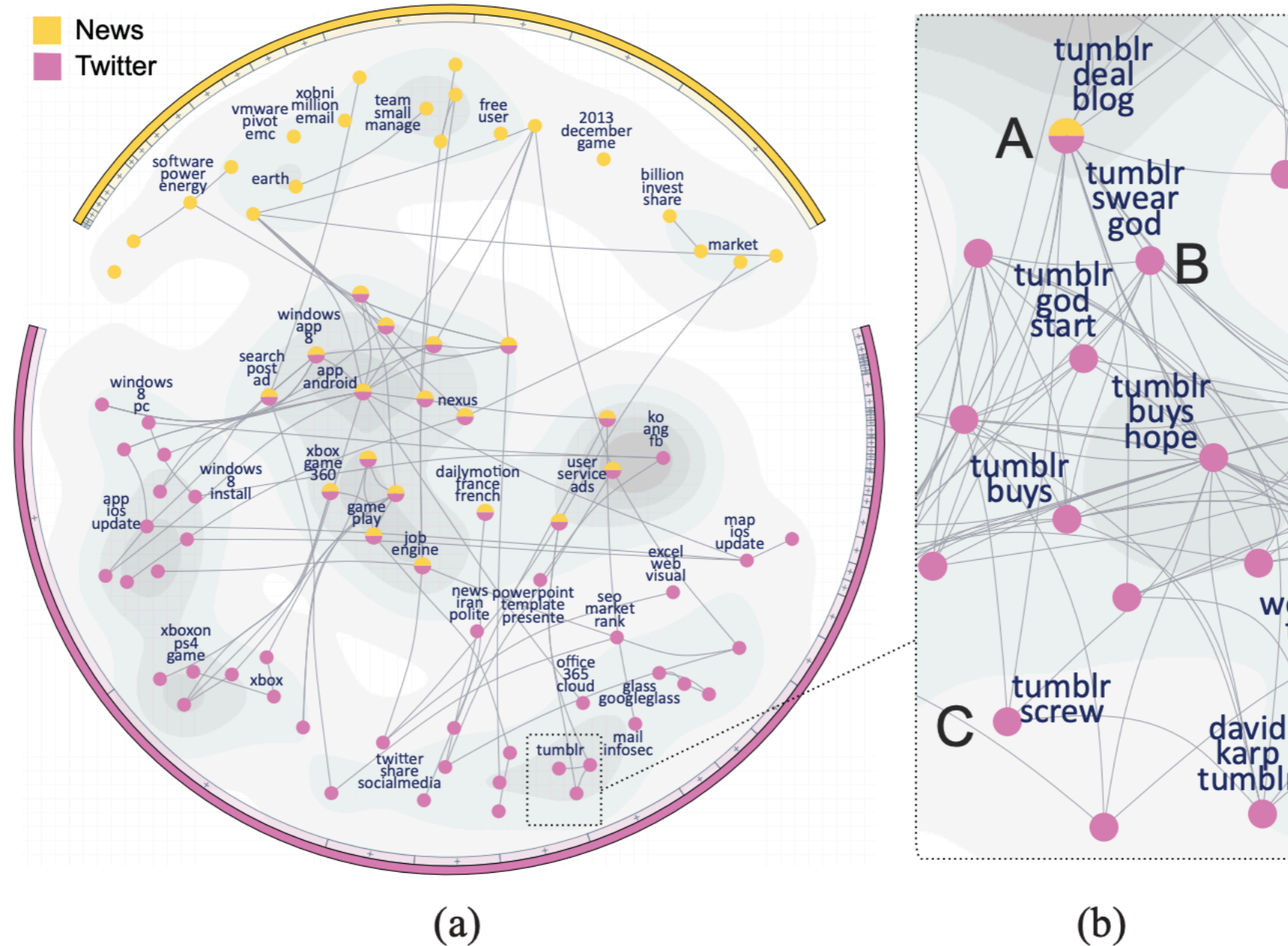


Figure 12. Matching the news corpus with the Twitter corpus: (a) overview; (b) comparison of Tumblr related topics.

Liu, Shixia, et al. "Topicpanorama: A full picture of relevant topics." *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2014.

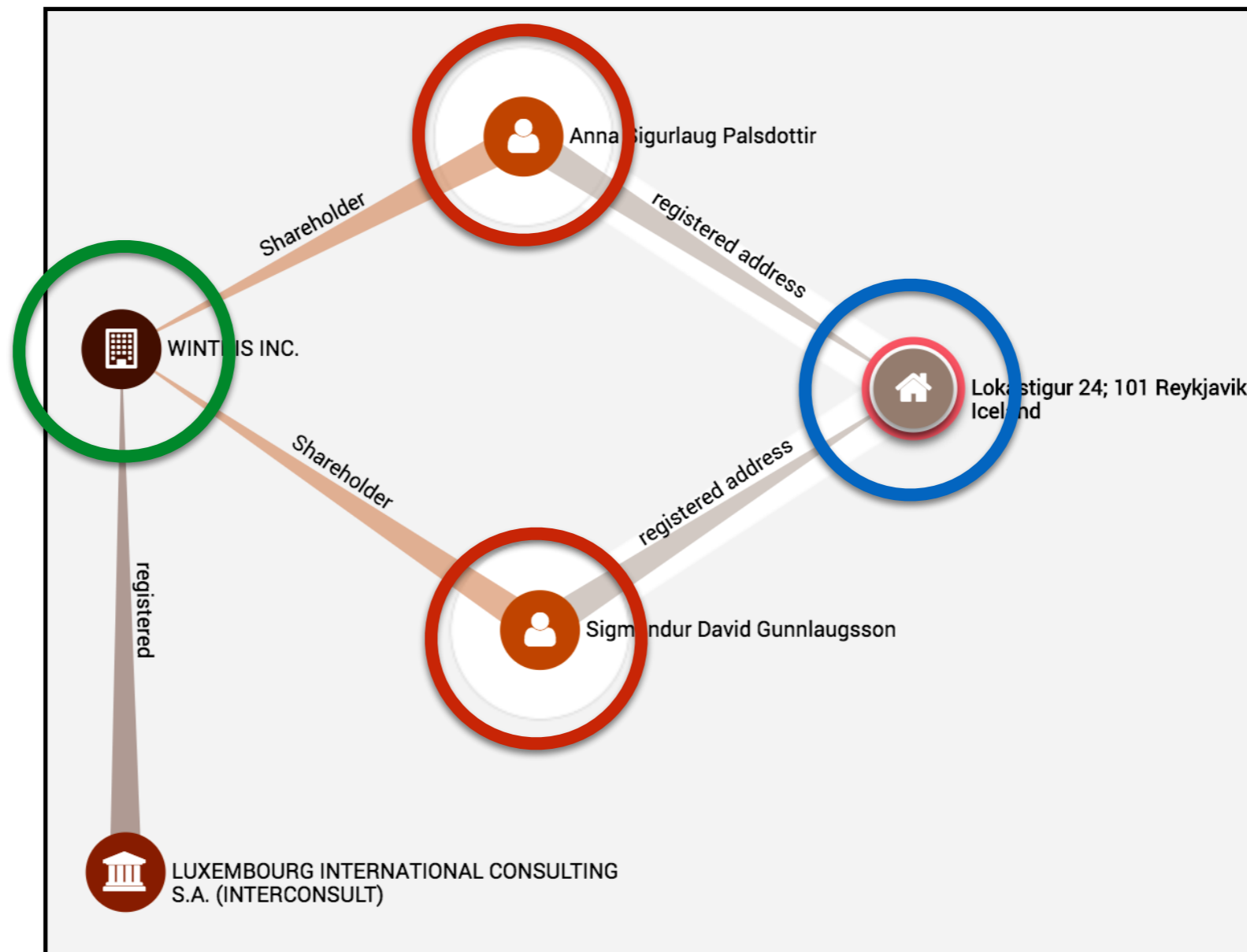
Named Entities

Named Entities are:

- People
- Organizations/Companies
- Locations
- Time Events

Panama Papers Networks use different visual encodings for:

People
Organizations
Locations



What is possible?

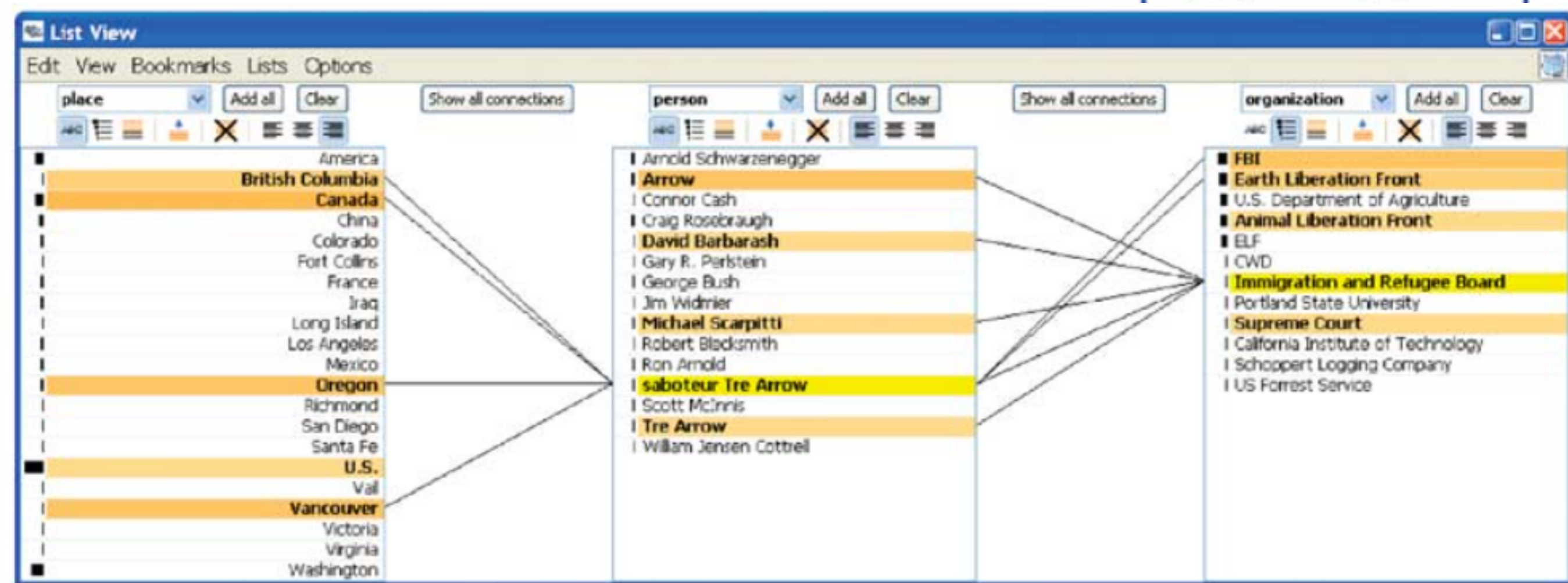
⟨Google⟩₁, headquartered in ⟨Mountain View⟩₆, unveiled the new ⟨Android⟩₄

⟨phone⟩₃ at the ⟨Consumer Electronic Show⟩₇. ⟨Sundar Pichai⟩₅ said in his

<p>1. Google</p> <p>Sentiment: Score 0 Magnitude 0</p> <p>Wikipedia Article</p> <p>Saliency: 0.26</p> <p>ORGANIZATION</p>	<p>2. users</p> <p>Sentiment: Score 0.4 Magnitude 0.9</p> <p>Saliency: 0.15</p> <p>PERSON</p>
<p>3. phone</p> <p>Sentiment: Score 0 Magnitude 0</p> <p>Saliency: 0.13</p> <p>CONSUMER GOOD</p>	<p>4. Android</p> <p>Sentiment: Score 0.1 Magnitude 0.2</p> <p>Wikipedia Article</p> <p>Saliency: 0.12</p> <p>CONSUMER GOOD</p>
<p>5. Sundar Pichai</p> <p>Sentiment: Score 0 Magnitude 0.1</p> <p>Wikipedia Article</p> <p>Saliency: 0.11</p> <p>PERSON</p>	<p>6. Mountain View</p> <p>Sentiment: Score 0 Magnitude 0</p> <p>Wikipedia Article</p> <p>Saliency: 0.10</p> <p>LOCATION</p>

<https://cloud.google.com/natural-language/>

Jigsaw: Supporting investigative analysis through interactive visualization, Stasko et. al, Information Visualization, 2008






Sentiment Analysis

What is Sentiment Analysis?

<https://cloud.google.com/natural-language/>

What is Sentiment Analysis?

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

 My experience so far has been fantastic! POSITIVE	 The product is ok I guess NEUTRAL	 Your support team is useless NEGATIVE
---	---	---

<https://monkeylearn.com/sentiment-analysis/>

Example Sentence:

Google, headquartered in Mountain View, unveiled the new Android phone at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

<https://cloud.google.com/natural-language/>

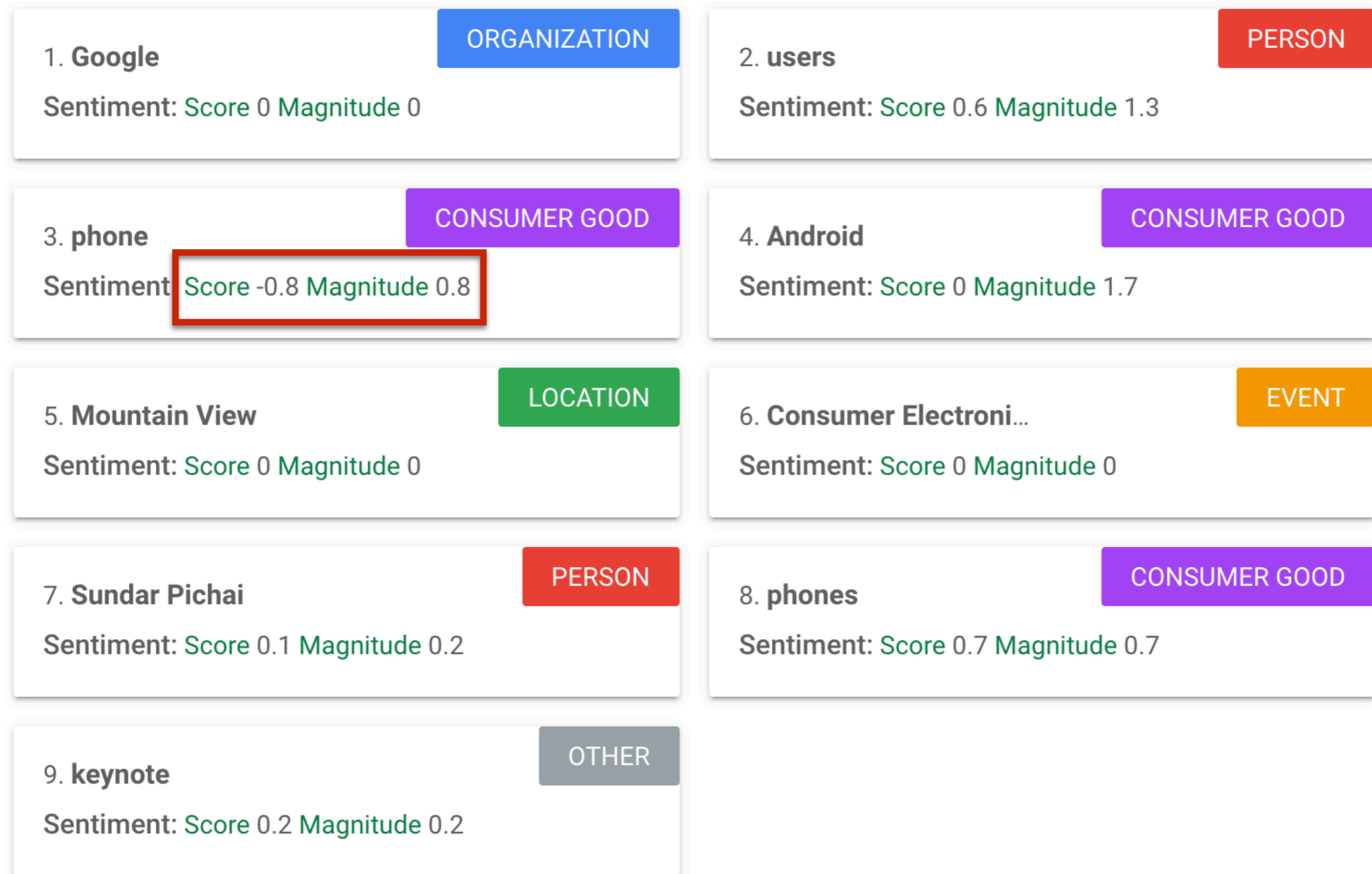
1. Google Sentiment: Score 0 Magnitude 0	ORGANIZATION	2. users Sentiment: Score 0.6 Magnitude 1.3	PERSON
3. phone Sentiment: Score 0 Magnitude 0	CONSUMER GOOD	4. Android Sentiment: Score 0.5 Magnitude 1	CONSUMER GOOD
5. Sundar Pichai Sentiment: Score 0.1 Magnitude 0.2	PERSON	6. Mountain View Sentiment: Score 0 Magnitude 0	LOCATION
7. Consumer Electroni... Sentiment: Score 0.1 Magnitude 0.1	EVENT	8. phones Sentiment: Score 0.7 Magnitude 0.7	CONSUMER GOOD
9. keynote Sentiment: Score 0.2 Magnitude 0.2	OTHER		

<https://cloud.google.com/natural-language/>

Updated Sentence:

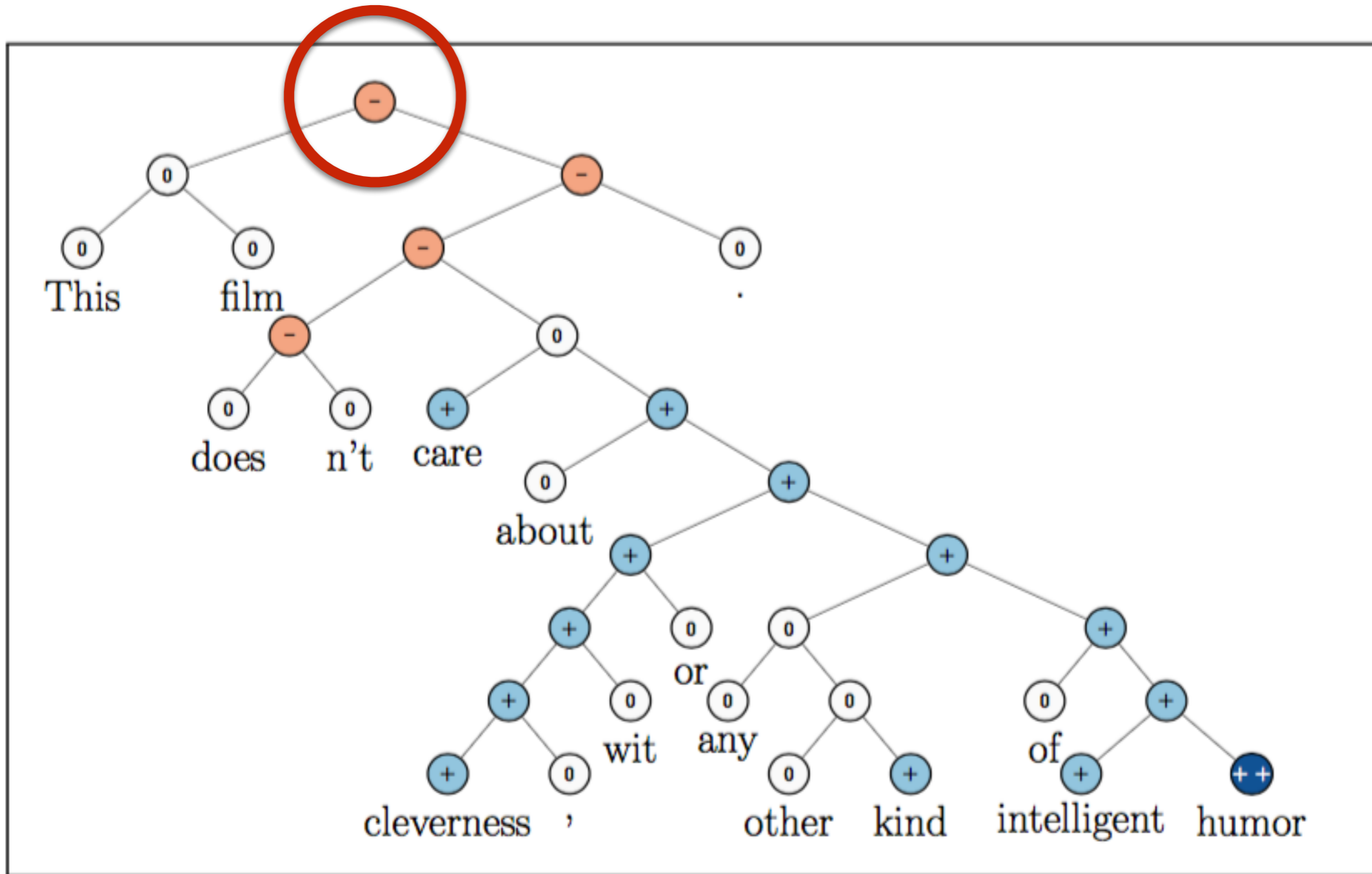
Google, headquartered in Mountain View, unveiled the new **and very expensive** Android phone at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

<https://cloud.google.com/natural-language/>

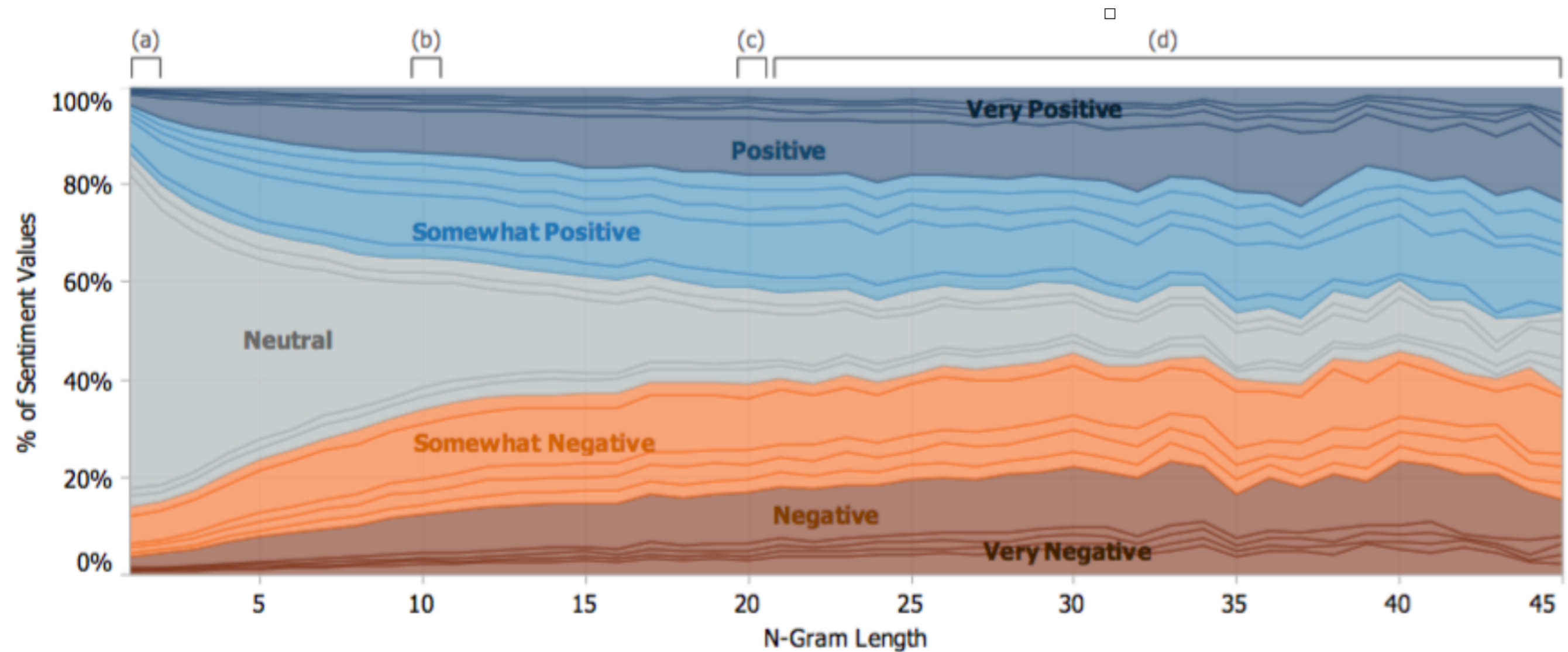


<https://cloud.google.com/natural-language/>

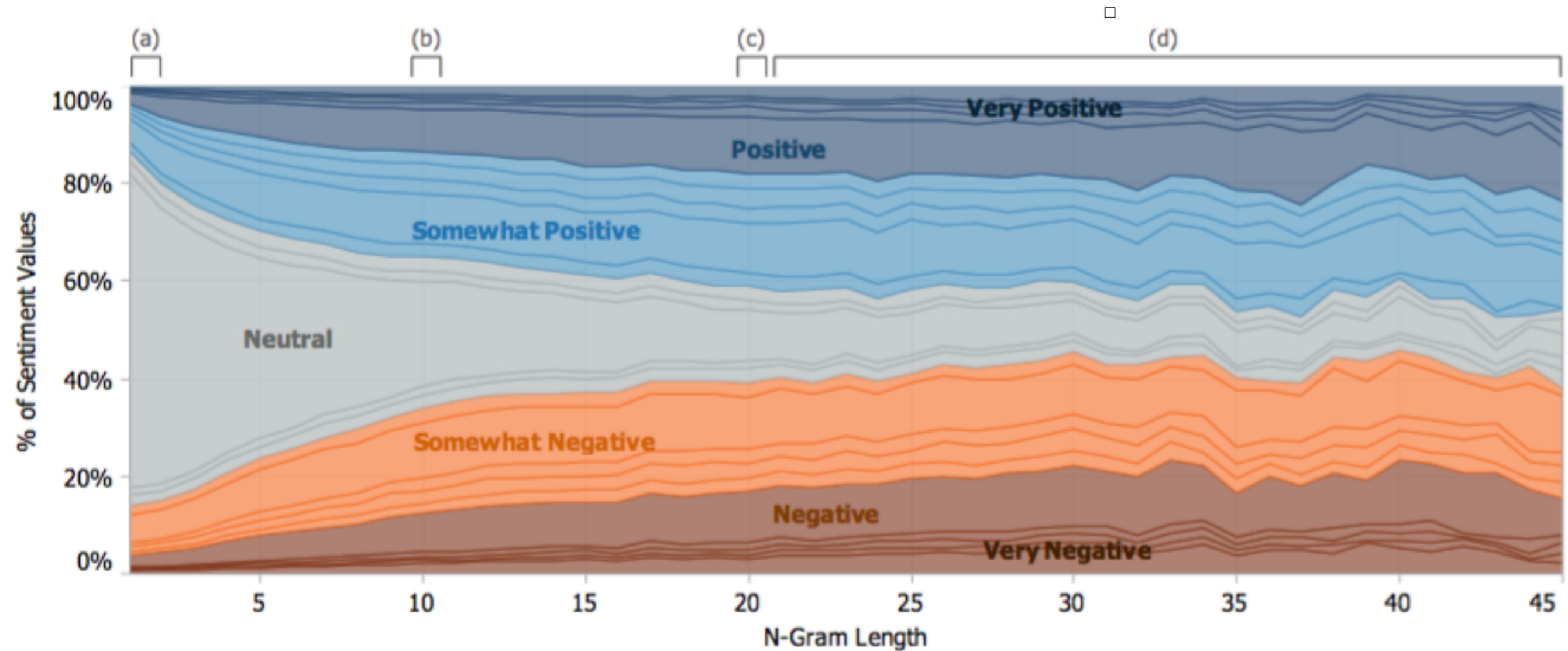
Although more nodes within this sentence have a **positive meaning**, the **sentence is labeled negative!**



Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, Socher et al., EMNLP 2013



The longer a phrase gets, the more likely it is either positive or negative!



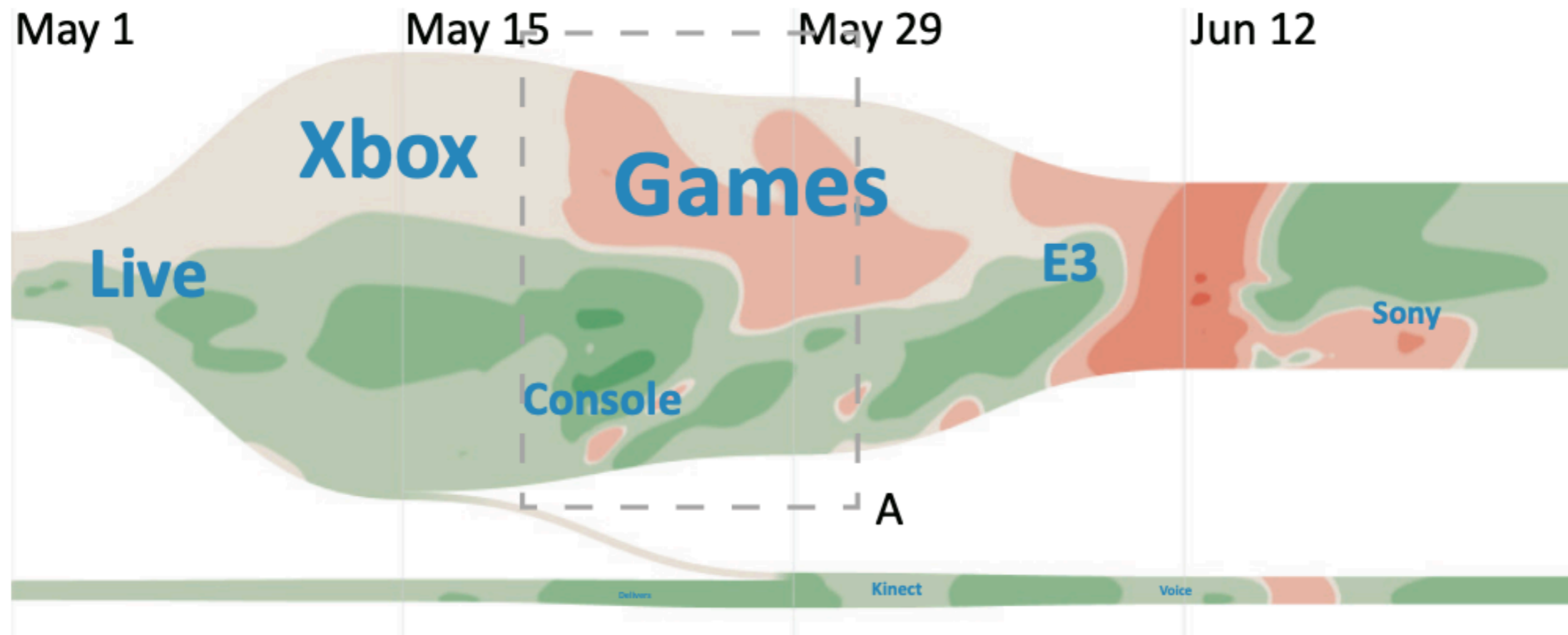
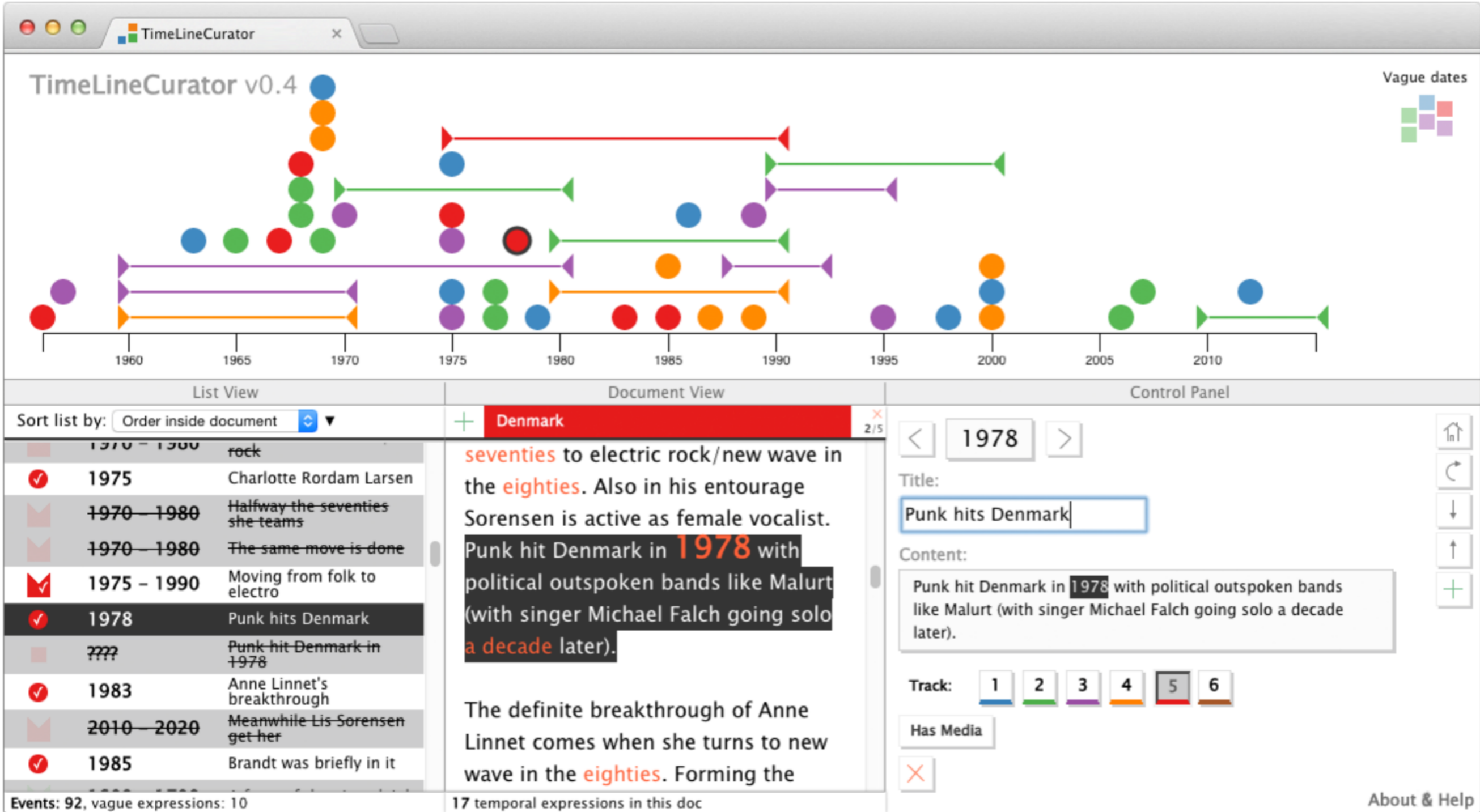


Fig. 7. Opinion diffusion on the Xbox topic from the period between May 15 to 29 when Xbox One was announced.

Wu, Yingcai, et al. "Opinionflow: Visual analysis of opinion diffusion on social media." *IEEE transactions on visualization and computer graphics* 20.12 (2014): 1763-1772.

Temporal Events

TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text; Johanna Fulda, Matthew Brehmer, Tamara Munzner; IEEE VAST 2015



The screenshot displays the TimeLineCurator v0.4 interface, which is divided into three main sections: List View, Document View, and Control Panel.

List View: Shows a list of events with their start and end dates and descriptions. The events are:

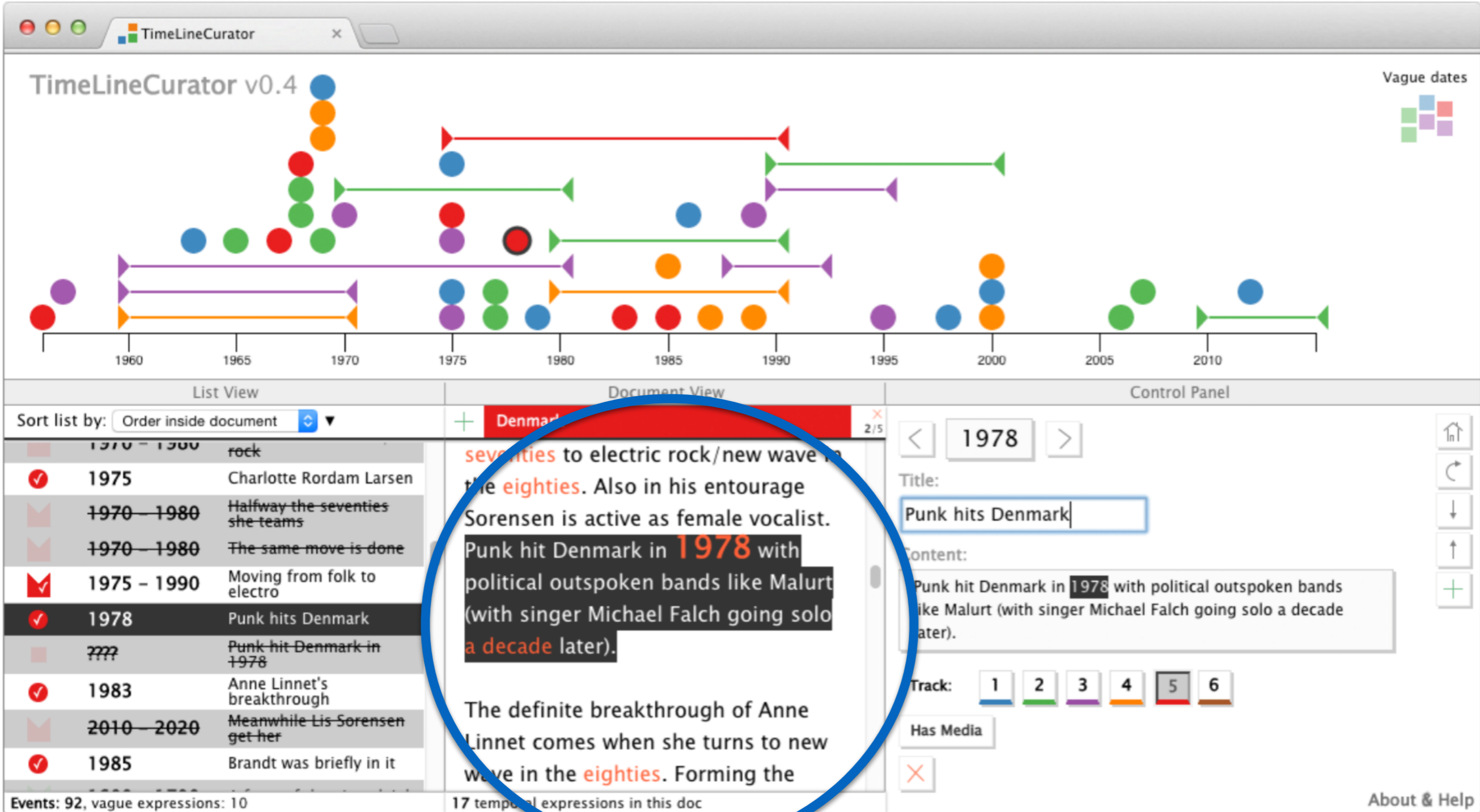
Start Date	End Date	Description
1970 - 1980		rock
1975		Charlotte Rordam Larsen
1970 - 1980		Halfway the seventies she teams
1970 - 1980		The same move is done
1975 - 1990		Moving from folk to electro
1978		Punk hits Denmark
????		Punk hit Denmark in 1978
1983		Anne Linnet's breakthrough
2010 - 2020		Meanwhile Lis Sorensen get her
1985		Brandt was briefly in it

Document View: Shows the selected event "Denmark" (2/5) with its content: "seventies to electric rock/new wave in the eighties. Also in his entourage Sorensen is active as female vocalist. Punk hit Denmark in 1978 with political outspoken bands like Malurt (with singer Michael Falch going solo a decade later). The definite breakthrough of Anne Linnet comes when she turns to new wave in the eighties. Forming the

Control Panel: Shows the selected event "Punk hits Denmark" with the year 1978. The title is "Punk hits Denmark" and the content is "Punk hit Denmark in 1978 with political outspoken bands like Malurt (with singer Michael Falch going solo a decade later).". The track is set to 5. There are also buttons for "Has Media" and "About & Help".

At the bottom of the interface, it shows "Events: 92, vague expressions: 10" and "17 temporal expressions in this doc".

Time Events get extracted from an input text source



The screenshot displays the TimeLineCurator v0.4 interface. At the top, a timeline visualization shows various colored dots and arrows representing time events and durations from 1960 to 2010. Below this, the interface is divided into three main sections: List View, Document View, and Control Panel.

List View: A table listing extracted events with checkboxes for selection.

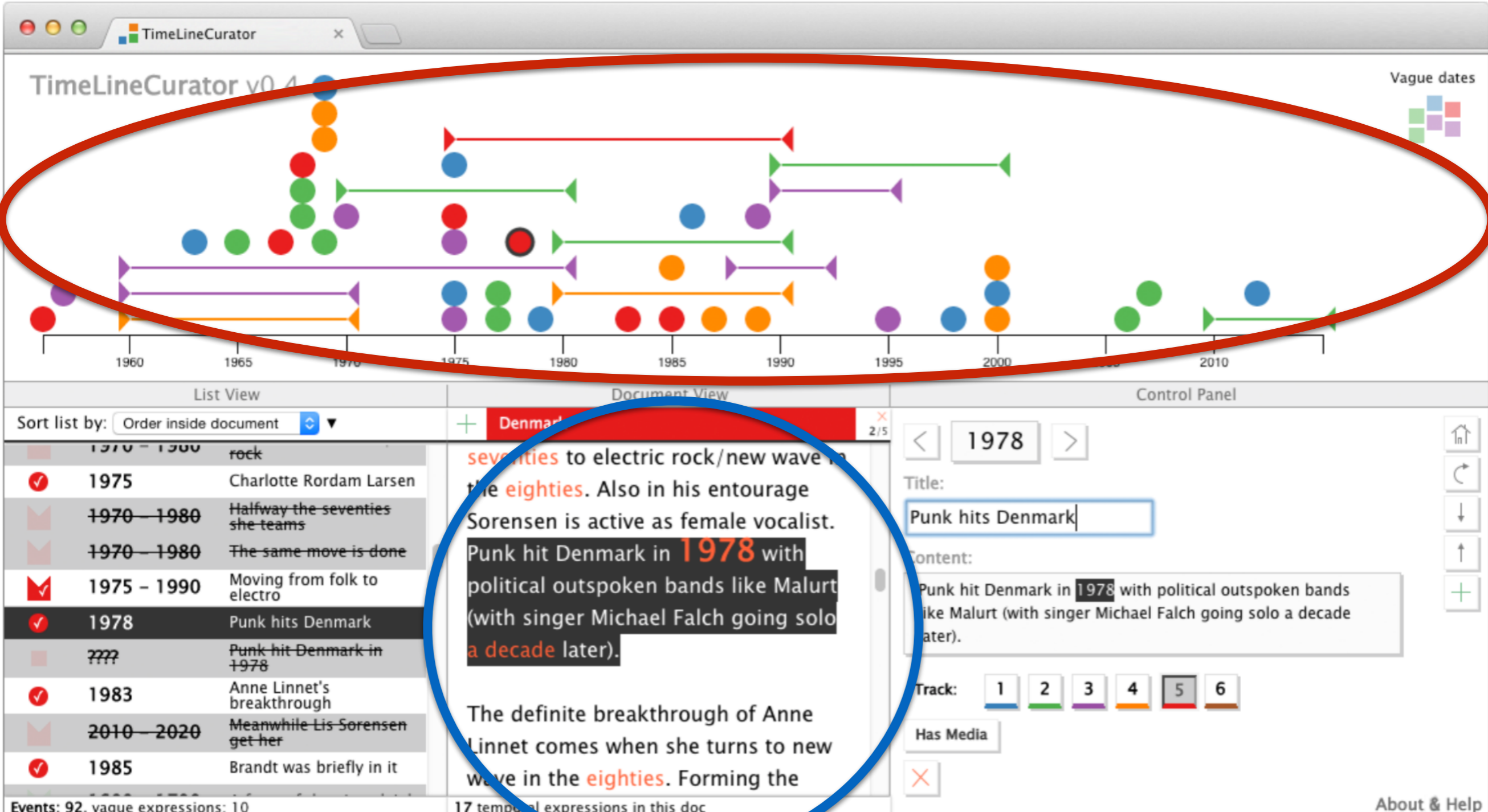
Year	Description
1970 - 1980	rock
1975	Charlotte Rordam Larsen
1970 - 1980	Halfway the seventies she teams
1970 - 1980	The same move is done
1975 - 1990	Moving from folk to electro
1978	Punk hits Denmark
????	Punk hit Denmark in 1978
1983	Anne Linnet's breakthrough
2010 - 2020	Meanwhile Lis Sorensen get her
1985	Brandt was briefly in it

Document View: Shows a text document with a blue circle highlighting a specific event. The text includes: "seventies to electric rock/new wave in the eighties. Also in his entourage Sorensen is active as female vocalist. Punk hit Denmark in 1978 with political outspoken bands like Malurt (with singer Michael Falch going solo a decade later). The definite breakthrough of Anne Linnet comes when she turns to new wave in the eighties. Forming the".

Control Panel: Provides navigation and editing tools. It shows a current year of 1978, a title field containing "Punk hits Denmark", and a content field with the corresponding text snippet. There are also track selection buttons (1-6) and a "Has Media" checkbox.

At the bottom left, it states "Events: 92, vague expressions: 10". At the bottom right, it says "17 temporal expressions in this doc" and "About & Help".

Time Events get extracted from an **input text source** and **visualised as Timeline**



The screenshot displays the TimeLineCurator v0.4.1 interface. The top section shows a timeline visualization with various colored dots and arrows representing events and their durations. A red oval highlights this visualization. Below the timeline, the interface is divided into three main panels: List View, Document View, and Control Panel.

List View: Shows a list of events sorted by 'Order inside document'. The event '1978 Punk hits Denmark' is selected and highlighted.

Year	Event
1970 - 1980	rock
1975	Charlotte Rordam Larsen
1970 - 1980	Halfway the seventies she teams
1970 - 1980	The same move is done
1975 - 1990	Moving from folk to electro
1978	Punk hits Denmark
????	Punk hit Denmark in 1978
1983	Anne Linnet's breakthrough
2010 - 2020	Meanwhile Lis Sorensen get her
1985	Brandt was briefly in it

Document View: Shows the text from the selected document. A blue oval highlights the text: "Punk hit Denmark in 1978 with political outspoken bands like Malurt (with singer Michael Falch going solo a decade later).".

Control Panel: Shows the event details for 'Punk hits Denmark' in 1978. The title is 'Punk hits Denmark' and the content is 'Punk hit Denmark in 1978 with political outspoken bands like Malurt (with singer Michael Falch going solo a decade later)'. The track is set to 5.

At the bottom, the status bar indicates: "Events: 92, vague expressions: 10" and "17 temporal expressions in this doc".

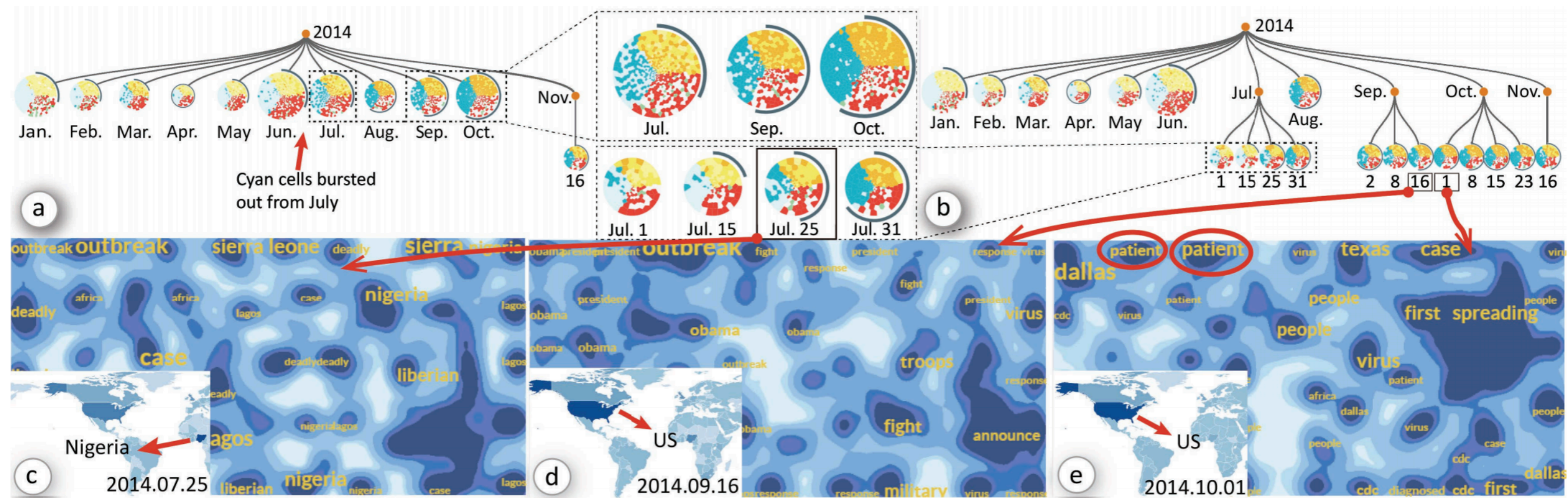


Fig. 8. Case study of the Ebola outbreak: (a) visual tree of the past subevents with respect to the Ebola outbreak in 2014; (b) visual tree created by expanding the nodes of July, September, and October in the original tree shown in (a); (c)-(e) geographic maps showing the geospatial changes of three subevents highlighted in (b) as well as the topic maps showing the topical changes of the three same subevents.

Wu, Yingcai, et al. "StreamExplorer: A multi-stage system for visually exploring events in social streams." *IEEE transactions on visualization and computer graphics* 24.10 (2018): 2758-2772.

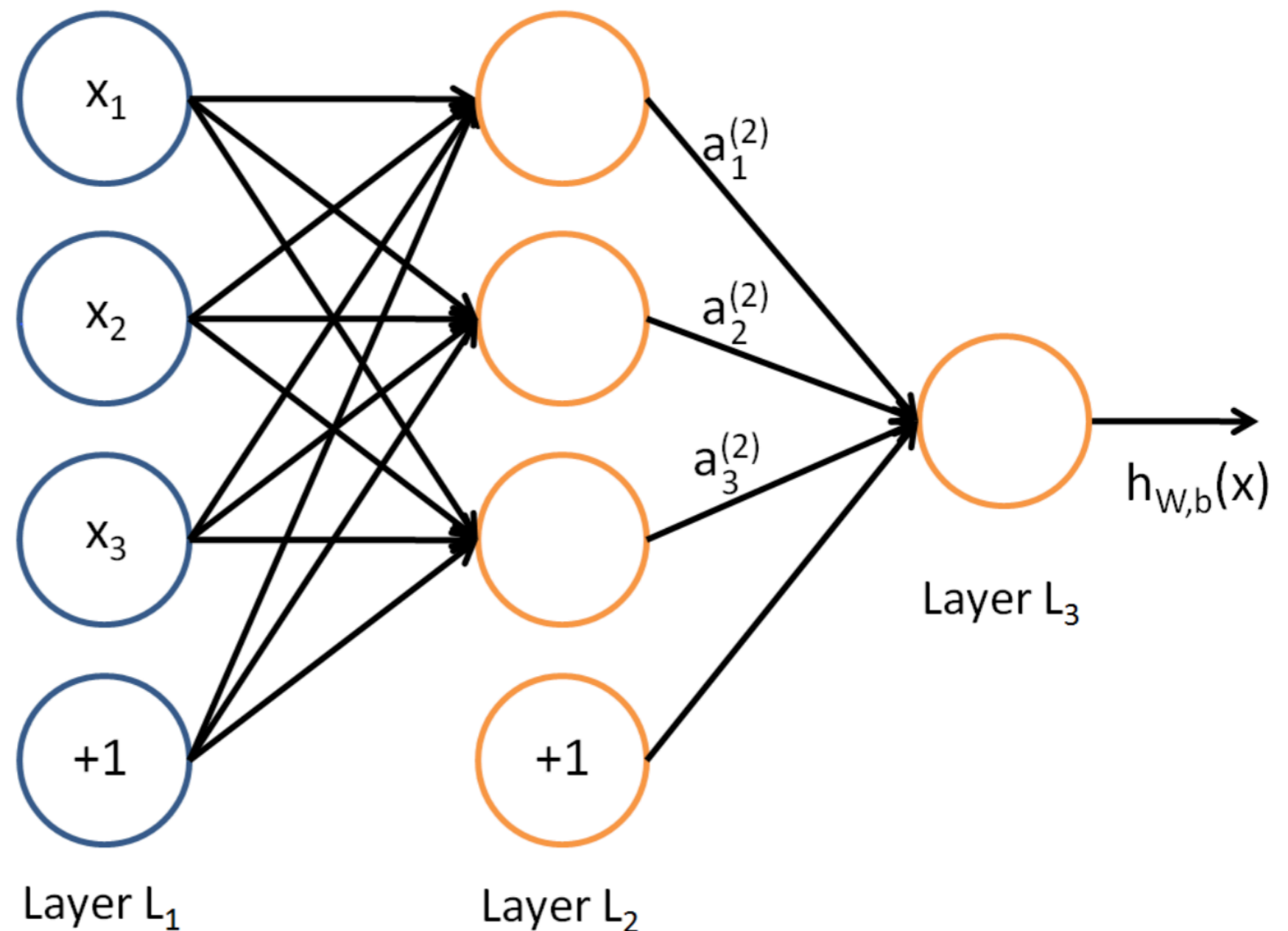
Deep Learning

What is deep learning?

<https://cloud.google.com/natural-language/>

Starting from a Neural Network (NN):

Google, headquartered in Mountain View, unveiled the new

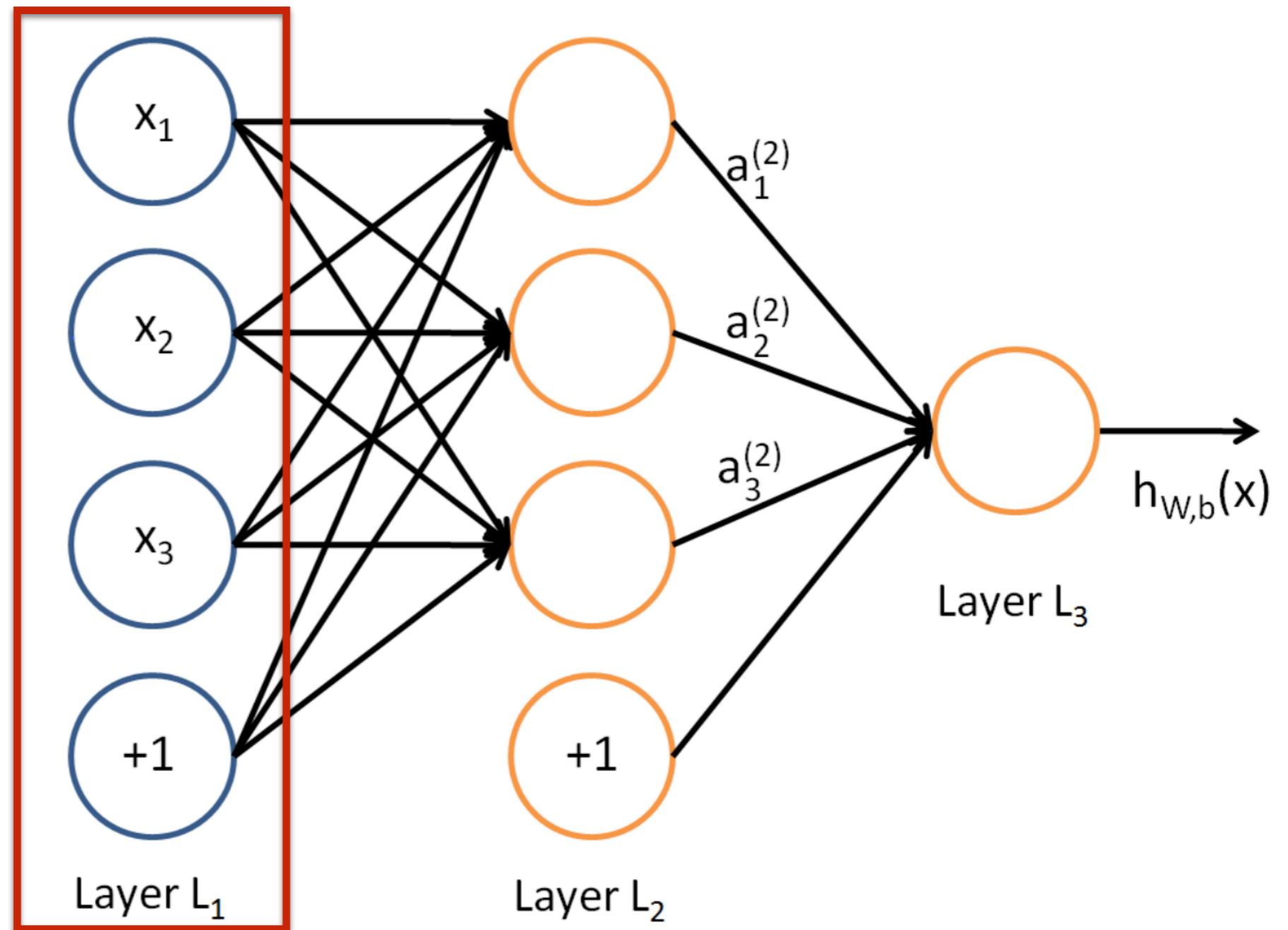


<https://towardsdatascience.com/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90>

Input Layer

Starting from a Neural Network (NN):

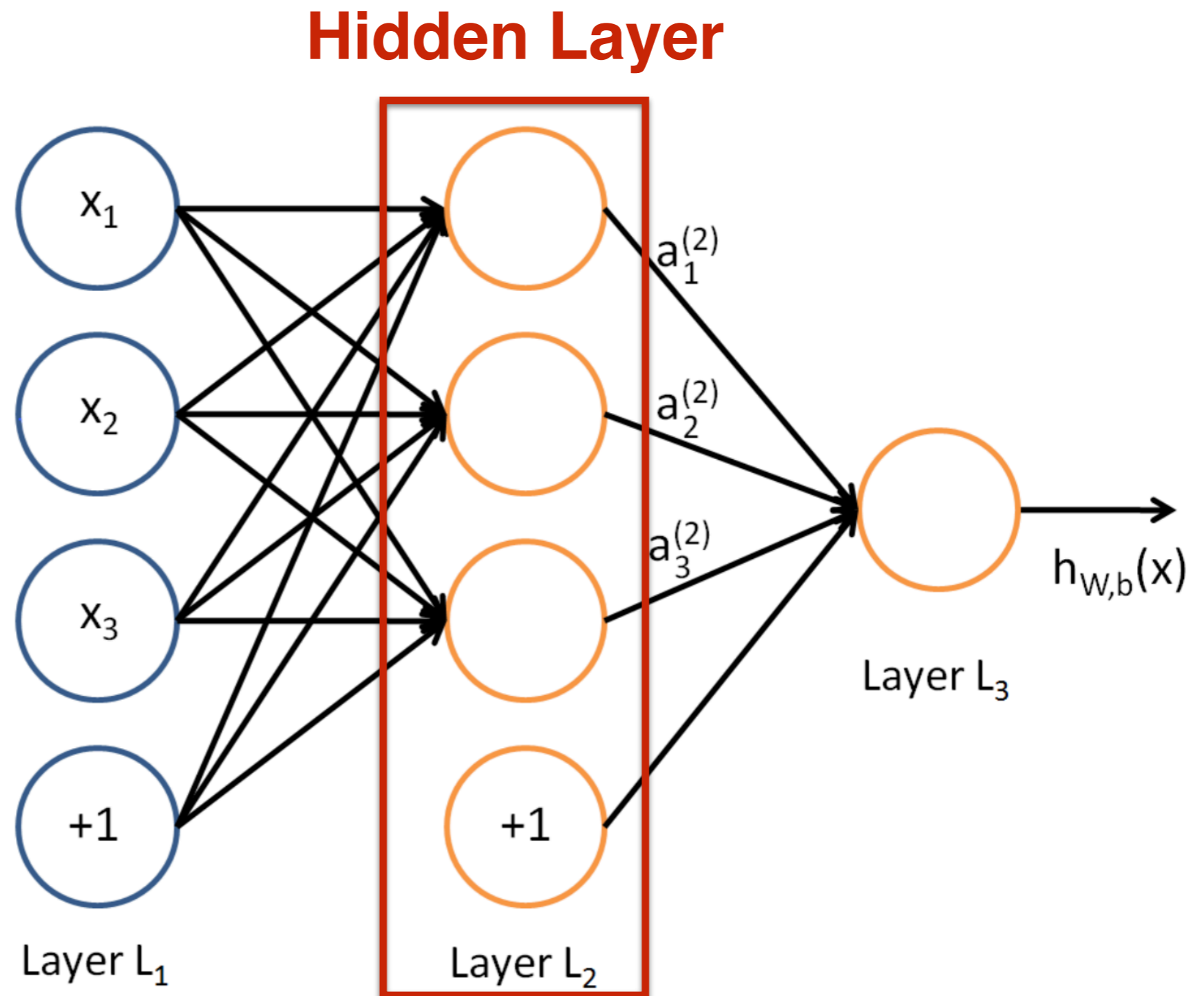
Google, headquartered in Mountain View, unveiled the new



<https://towardsdatascience.com/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90>

Starting from a Neural Network (NN):

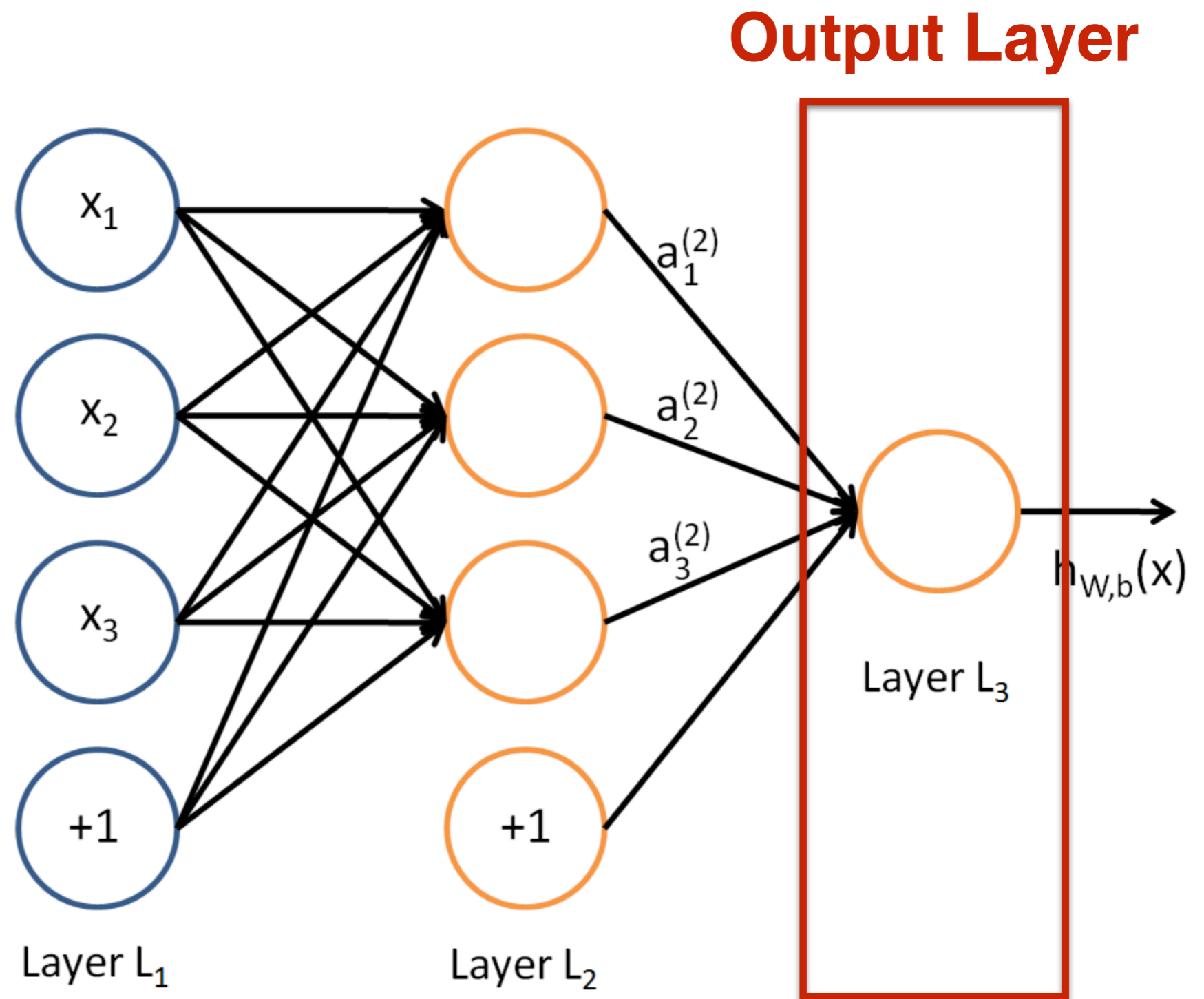
Google, headquartered in Mountain View, unveiled the new



<https://towardsdatascience.com/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90>

Starting from a Neural Network (NN):

Google, headquartered in Mountain View, unveiled the new



<https://towardsdatascience.com/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90>

Deep Neural Network:
More than one hidden layer.

<https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d>

Deep Neural Network:

More than one hidden layer.

Deep Learning:

Representation learning instead of feature engineering.

<https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d>

Deep Neural Network:

More than one hidden layer.

Deep Learning:

Representation learning instead of feature engineering.

Types:

Convolutional (CNN), Recurrent (RNN), Variational Autoencoders (VAE), Generative Adversarial Networks (GAN), ...

<https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d>

Tasks:

POS tagging

Parsing

Named-Entity Recognition

Semantic Role Labeling

Sentiment Classification

Machine translation

Question answering

Dialogue systems

Contextual Embeddings

...

Young, Tom, et al. "Recent trends in deep learning based natural language processing." *IEEE Computational Intelligence Magazine* 13.3 (2018): 55-75.

TABLE VII: Machine translation (Numbers are BLEU scores)

Paper	Model	WMT2014 English2German	WMT2014 English2French
Cho et al. [82]	Phrase table with neural features		34.50
Sutskever et al. [74]	Reranking phrase-based SMT best list with LSTM seq2seq		36.5
Wu et al. [162]	Residual LSTM seq2seq + Reinforcement learning refining	26.30	41.16
Gehring et al. [163]	seq2seq with CNN	26.36	41.29
Vaswani et al. [113]	Attention mechanism	28.4	41.0

BLEU Score:

"the closer a machine translation is to a professional human translation, the better it is" – this is the central idea behind BLEU.

<https://en.wikipedia.org/wiki/BLEU>

Young, Tom, et al. "Recent trends in deep learning based natural language processing." *iee Computational intelligence magazine* 13.3 (2018): 55-75.

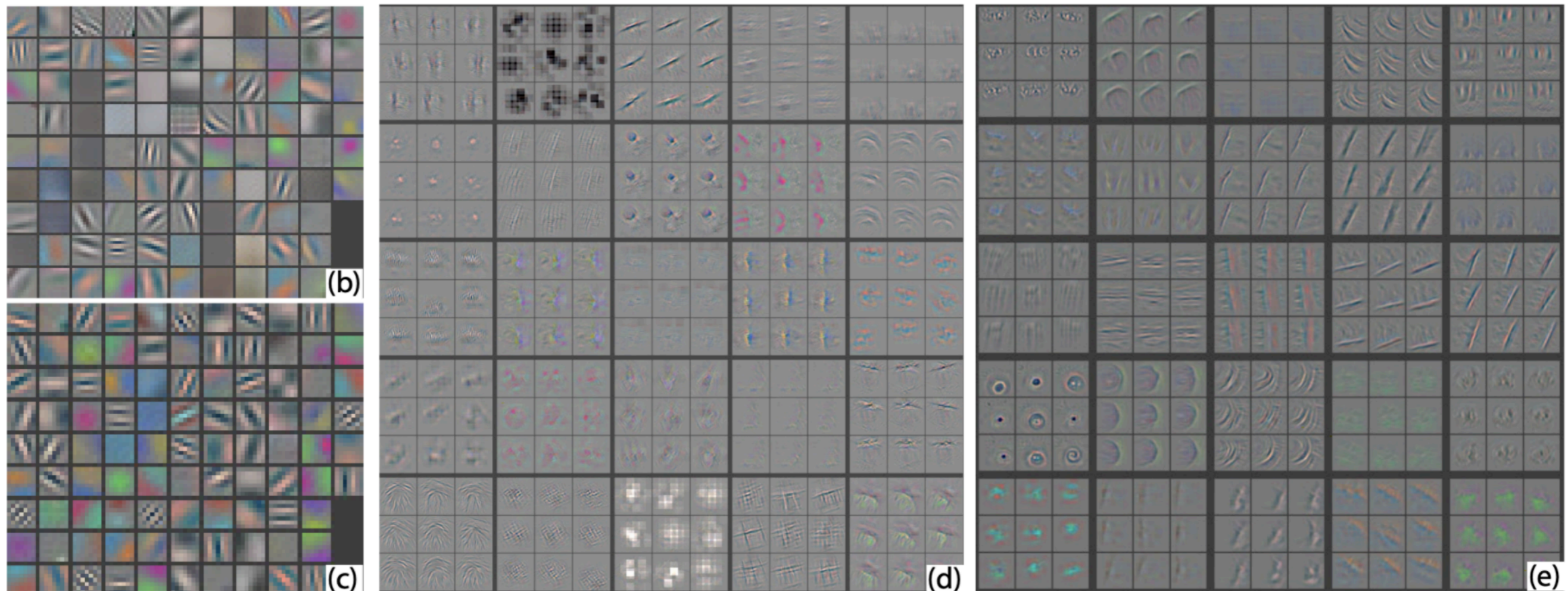
TABLE X: Comparison of ELMo + Baseline with the previous state of the art (SOTA) on various NLP tasks. The table has been adapted from [41]. SOTA results have been taken from [41]; SQUAD [166]: QA task; SNLI [178]: Stanford Natural Language Inference task; SRL [153]: Semantic Role Labelling; Coref [179]: Coreference Resolution; NER [180]: Named Entity Recognition; SST-5 [4]: Stanford Sentiment Treebank 5-class classification;

Task	Previous SOTA	Previous SOTA Results	Baseline	ELMo + Baseline	Increase (Absolute/Relative)
SQuAD	Liu et al. [181]	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Qian et al. [182]	88.6	88.0	88.70 \pm 0.17	0.7 / 5.8%
SRL	Luheng et al. [183]	81.7	81.4	84.6	3.2 / 17.2%
Coref	Kenton et al. [184]	67.2	67.2	70.4	3.2 / 9.8%
NER	Matthew et al. [185]	91.93 \pm 0.19	90.15	92.22 \pm 0.10	2.06 / 21%
SST-5	Bryan et al. [186]	53.7	51.4	54.7 0.5	3.3 / 6.8%

Task	BiLSTM+ ELMo+Attn	BERT
QNLI	79.9	91.1
SST-2	90.9	94.9
STS-B	73.3	86.5
RTE	56.8	70.1
SQuAD	85.8	91.1
NER	92.2	92.8

TABLE XI: QNLI [187]: Question Natural Language Inference task; SST-2 [4]: Stanford Sentiment Treebank binary classification; STS-B [188]: Semantic Textual Similarity Benchmark; RTE [189]: Recognizing Textual Entailment; SQUAD [166]: QA task; NER [180]: Named Entity Recognition.

Young, Tom, et al. "Recent trends in deep learning based natural language processing." *ieee Computational intelligence magazine* 13.3 (2018): 55-75.

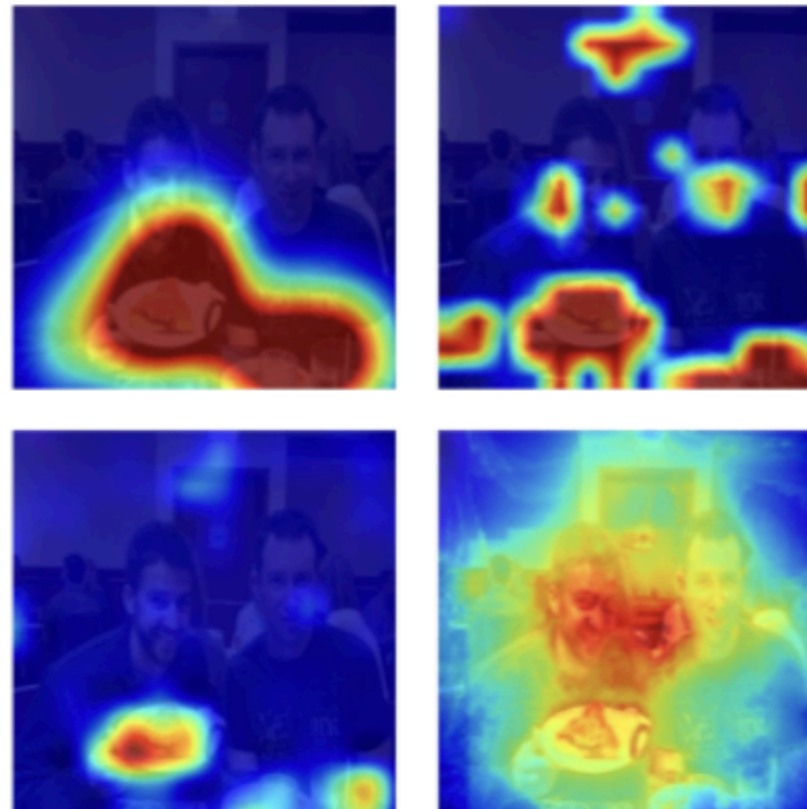


A figure from "Visualizing and Understanding Convolutional Networks" by Zeiler and Fergus, 2013, shows early results for a technique called **feature visualization** that visualizes the learned features in intermediate hidden layers of a deep learning model.

<https://medium.com/multiple-views-visualization-research-explained/visualization-in-deep-learning-b29f0ec4f136>

Attention

Das, Agrawal, et al. 2016



<https://medium.com/multiple-views-visualization-research-explained/visualization-in-deep-learning-b29f0ec4f136>

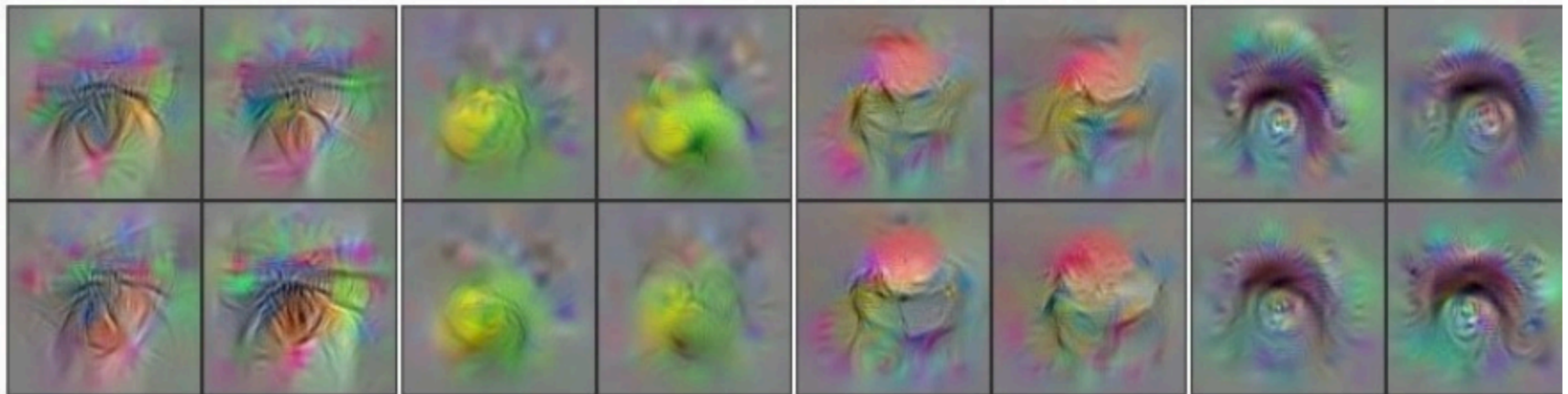
Saliency

Smilkov, et al. 2017

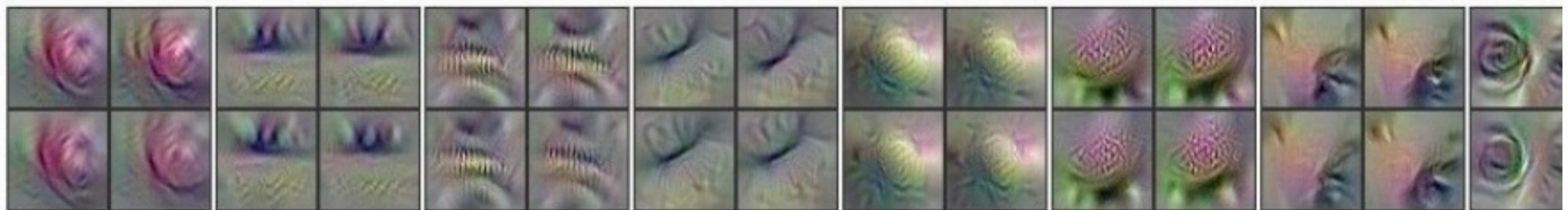


<https://medium.com/multiple-views-visualization-research-explained/visualization-in-deep-learning-b29f0ec4f136>

Layer 3



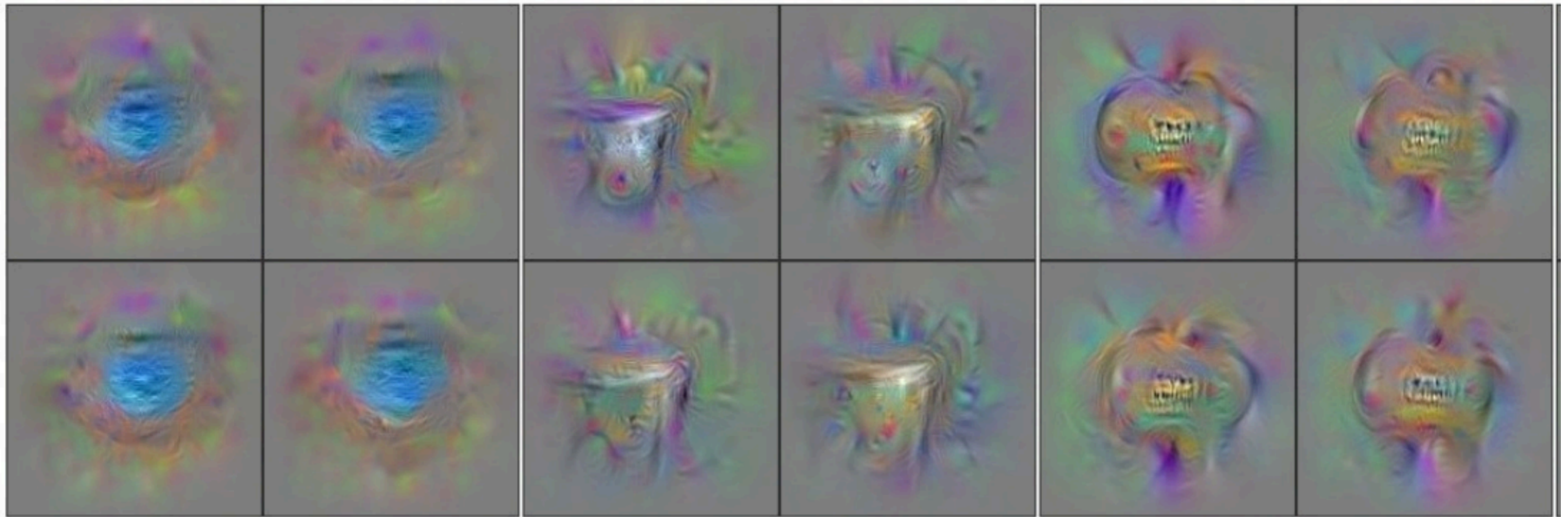
Layer 2



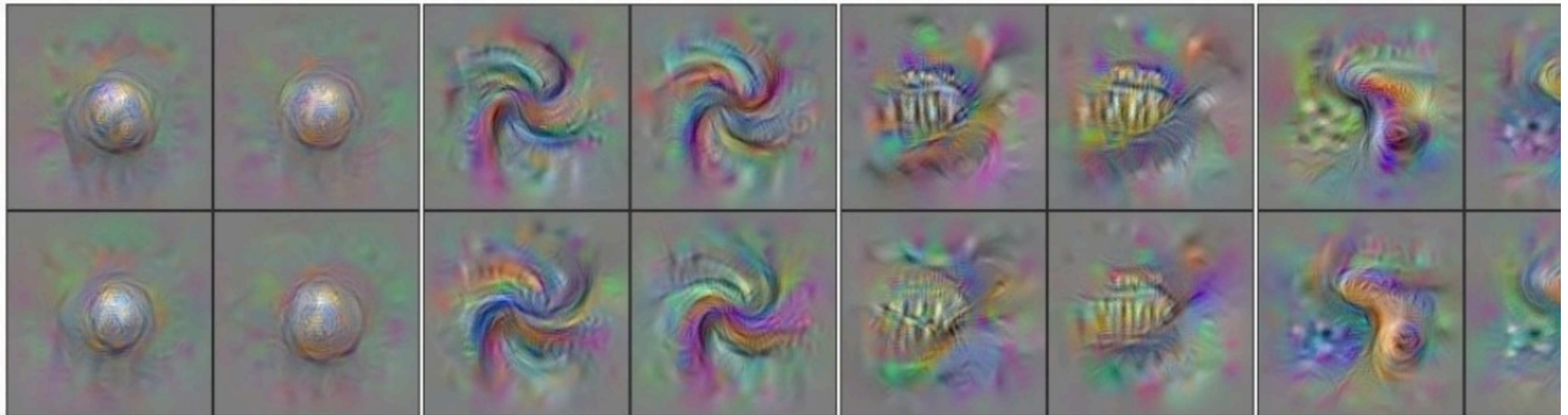
Layer 1

Yosinski, Jason, et al. "Understanding neural networks through deep visualization." *arXiv preprint arXiv:1506.06579* (2015).

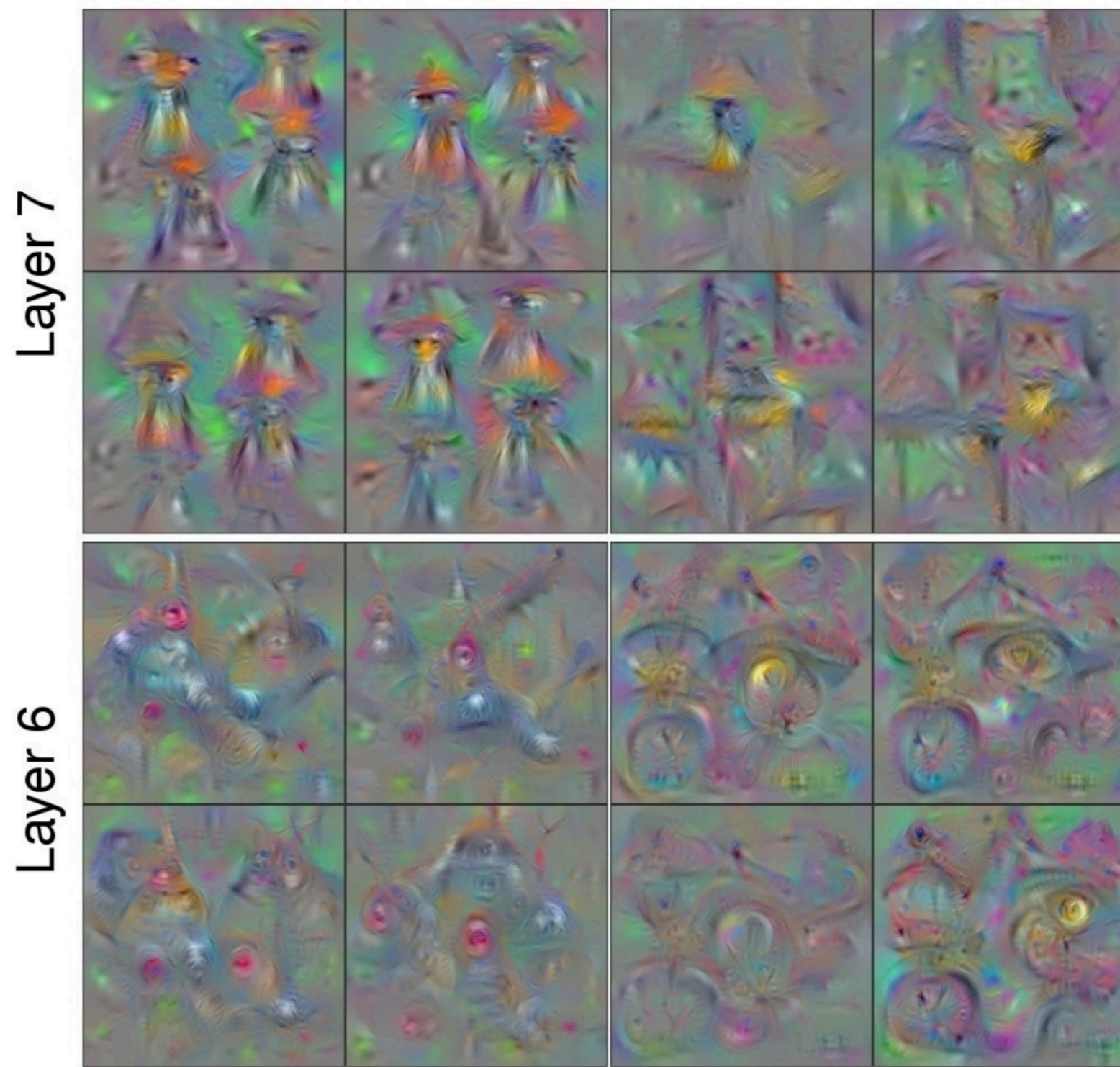
Layer 5



Layer 4

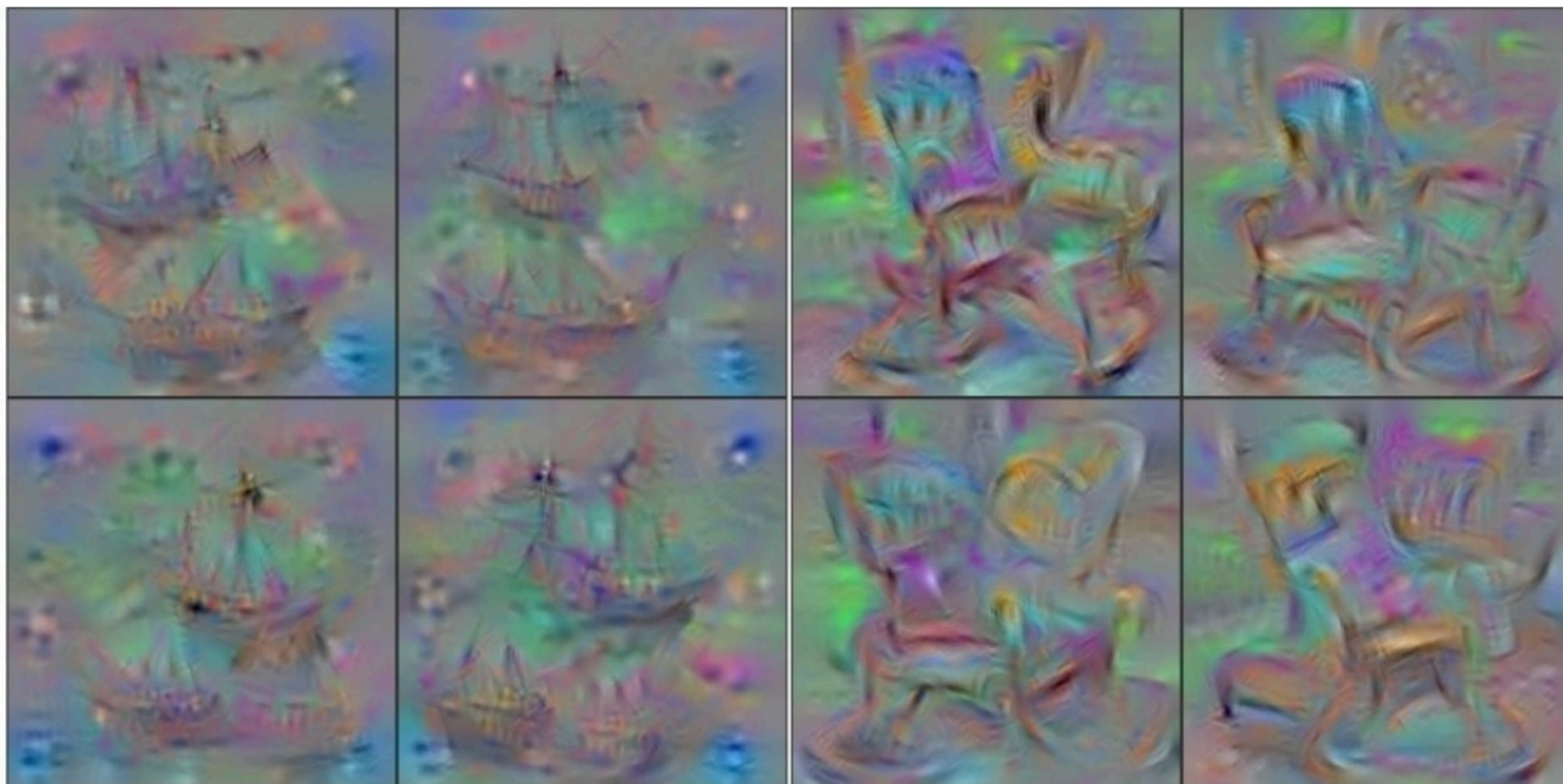


Yosinski, Jason, et al. "Understanding neural networks through deep visualization." *arXiv preprint arXiv:1506.06579* (2015).



Yosinski, Jason, et al. "Understanding neural networks through deep visualization." *arXiv preprint arXiv:1506.06579* (2015).

Layer 8



Pirate Ship

Rocking Chair

Yosinski, Jason, et al. "Understanding neural networks through deep visualization." *arXiv preprint arXiv:1506.06579* (2015).

Projects