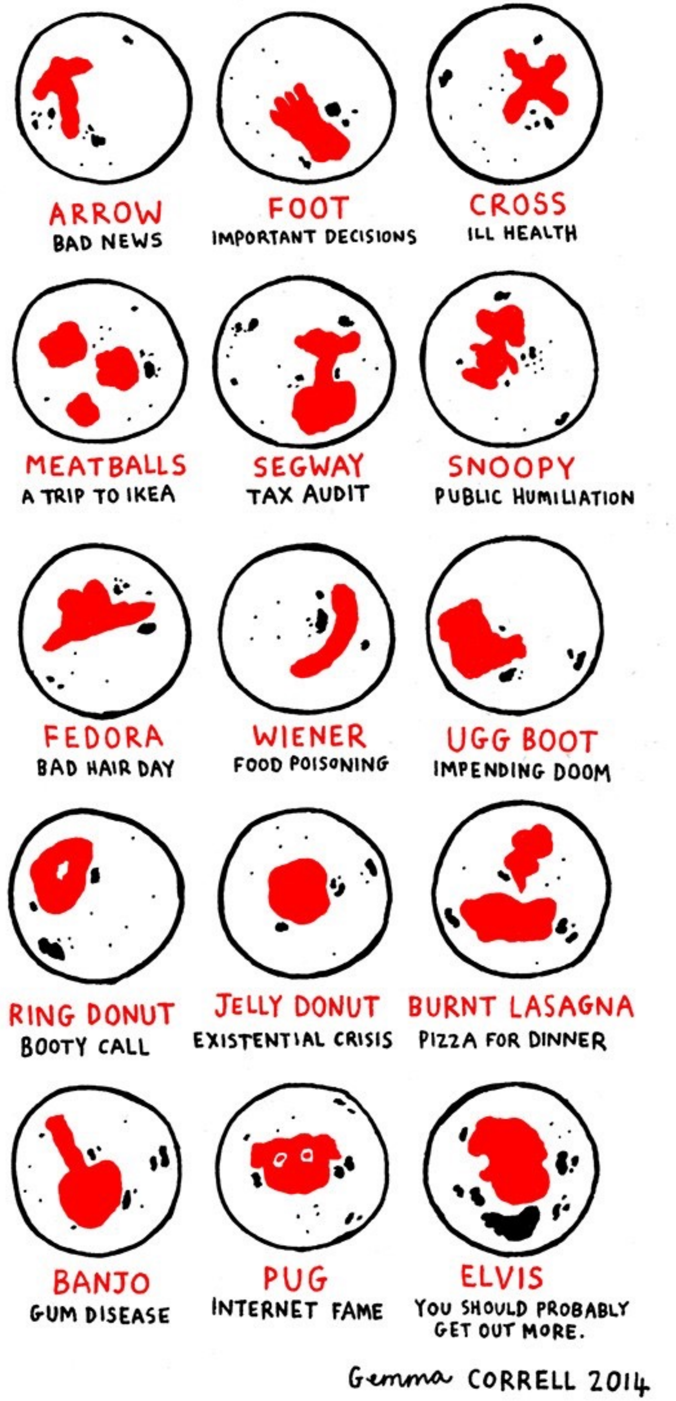
Decoding Quarterly Earnings Calls

Peter Goff https://www.linkedin.com/in/peter-t-goff/

A Tale of Sorrow and Woe: NLP, Clustering, & Me



Transcripts of Quarterly Earnings Calls

- Does the content of these calls relate to subsequent changes in stock price?

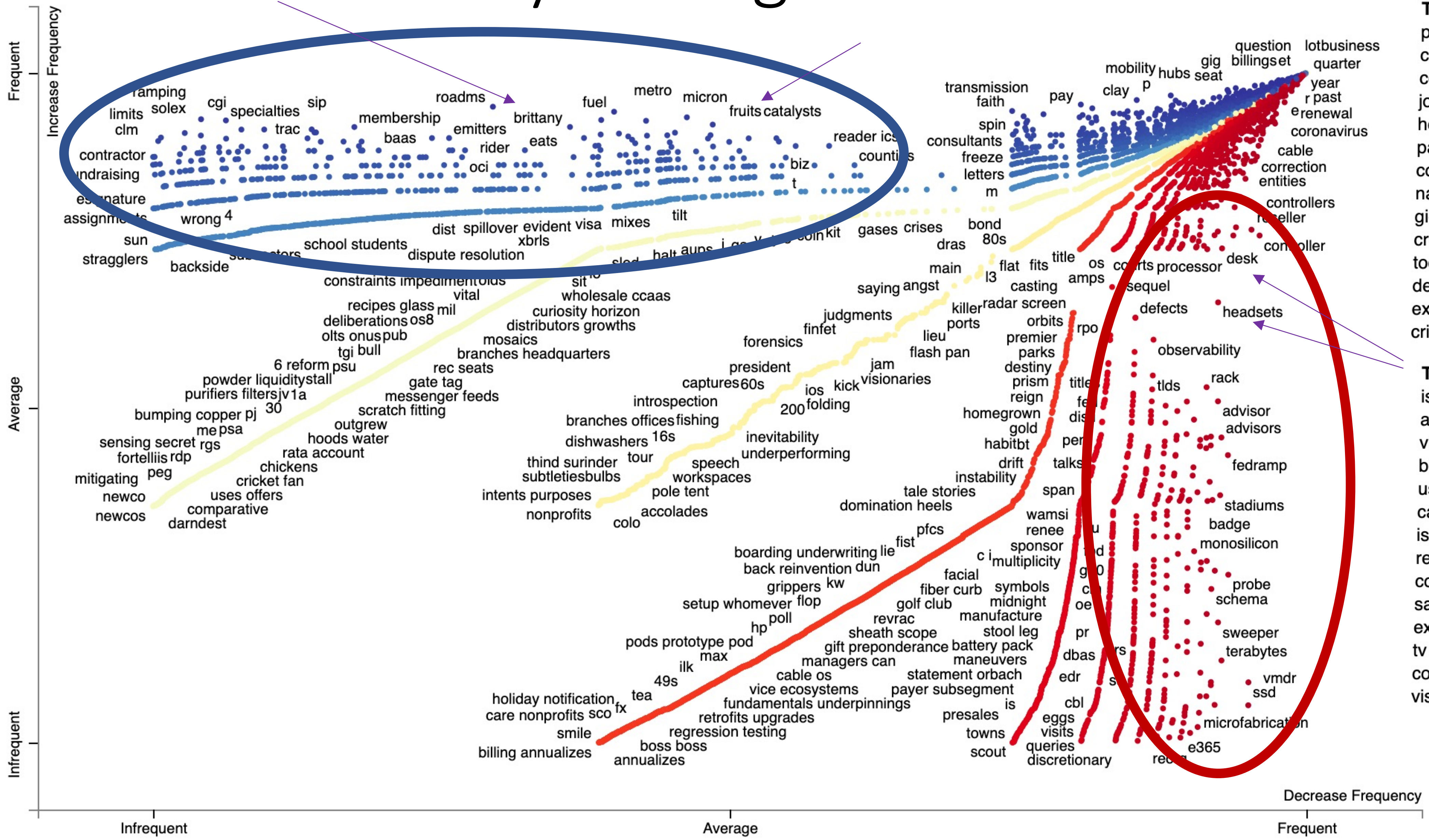
- 4,440 transcripts from 350 technology stocks

- Question-and-Answer:
  - Raw: 4,000 words
  - **Clean: 1,927 words per transcript**

Exploration

- **What** are they talking about? Nouns
- **How** are they talking about things? Adjectives

What are they talking about?

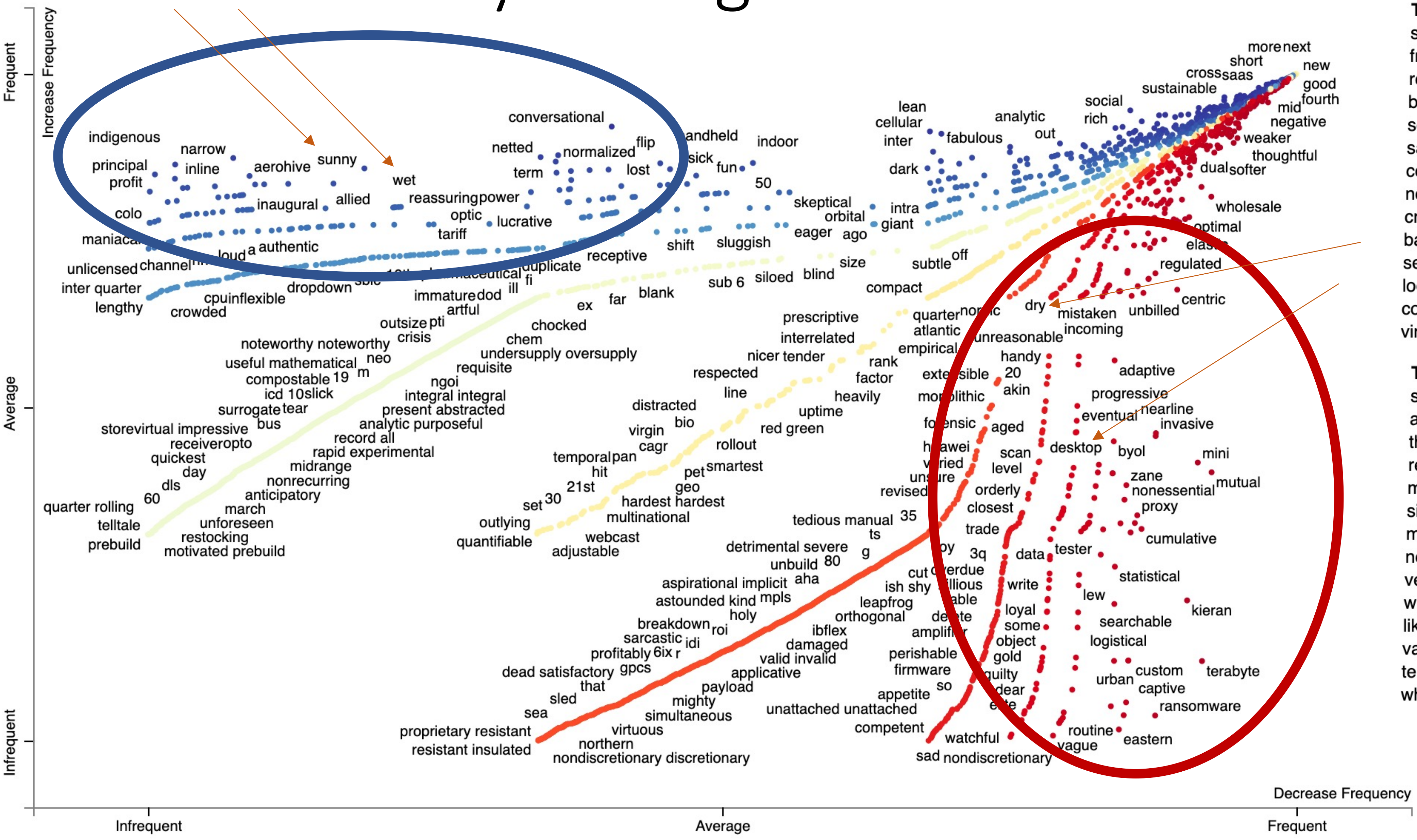


Top Increase
payments
center
congratulations
job
home
payment
congrats
nanometer
gig
credit
tools
delivery
expenses
crisis

Top Decrease
issues
assets
video
billings
users
carrier
issue
renewal
coronavirus
saas
execution
tv
component
vision

Characteristic
headwind
seasonality
headwinds
verticals
covid
uptick
backlog
tailwind
capex
margins
cetera
nanometer
margin
quarter
midpoint
coronavirus
standpoint
synergies
geographies
congrats
quarters
oems
tailwinds
cloud
headcount
saas
operator
churn
skus
workloads

How are they talking about it?



Top Increase
- sustainable
- front
- retail
- broad
- social
- safe
- covid
- normal
- cross
- back
- seasonal
- local
- corporate
- virtual

Characteristic
- indiscernible
- covid
- saas
- accretive
- flattish
- hyperscale
- upsell
- lumpy
- multiyear
- ratable
- addressable
- dilutive
- webscale
- incremental
- oled
- pandemic
- impactful
- gross
- ransomware
- bifacial
- nearline
- bullish
- uptick
- curious
- transactional
- onboarding
- recompete
- sequential
- unbilled
- terabyte

Top Decrease
- saas
- annual
- thoughtful
- regional
- macro
- simple
- mutual
- negative
- vertical
- weaker
- likely
- vast
- terabyte
- wholesale

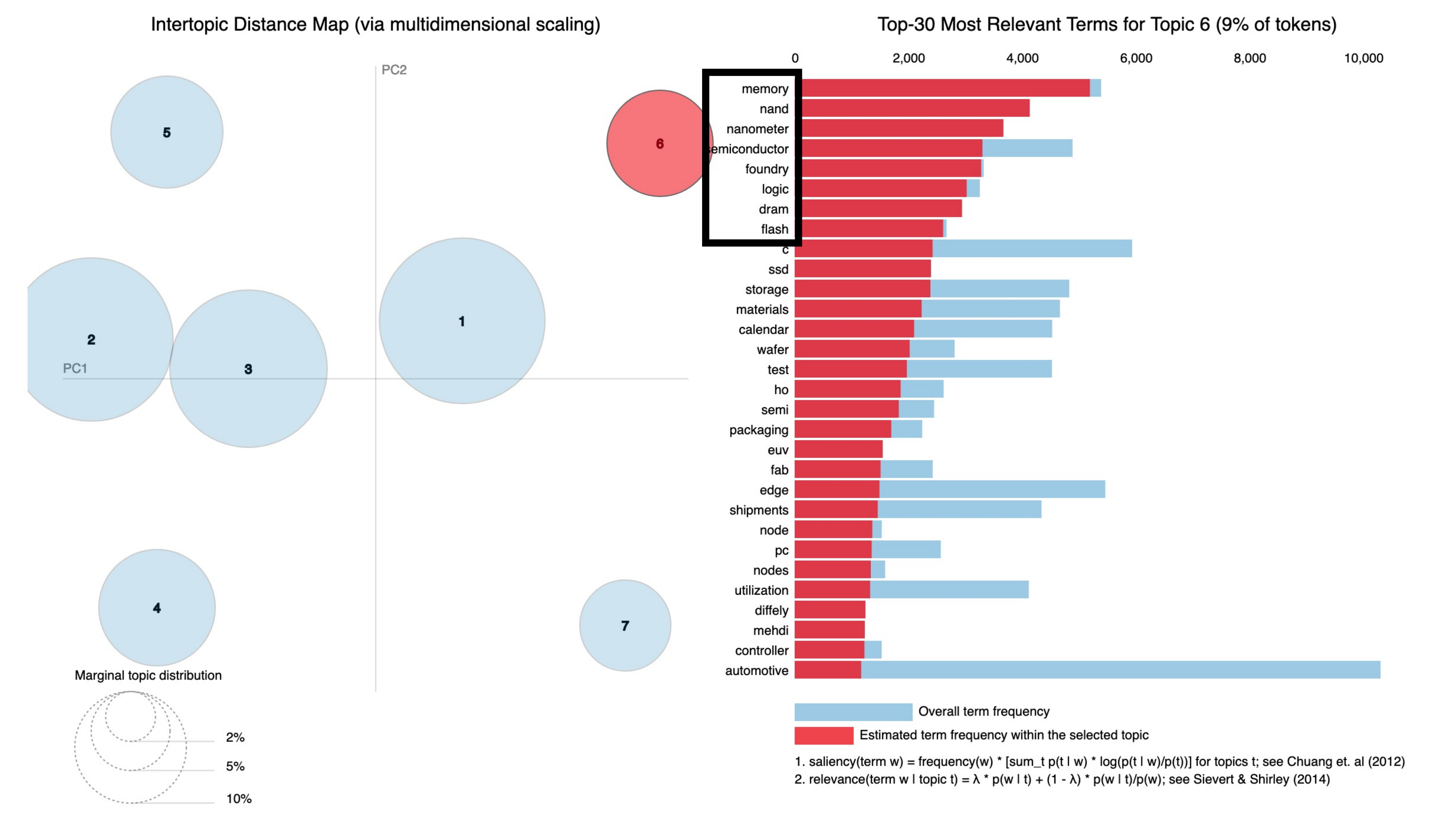
Sentiment Analysis

- Loughran-McDonald Sentiment Word Lists - finance specific dictionary.
  - Stocks that **increased** following the call: Polarity = 0.0375
  - Stocks that **declined** following the call: Polarity = 0.0100

  - Subjectivity was equal across the groups (0.0689 versus 0.0692)

Let’s build some topic models!

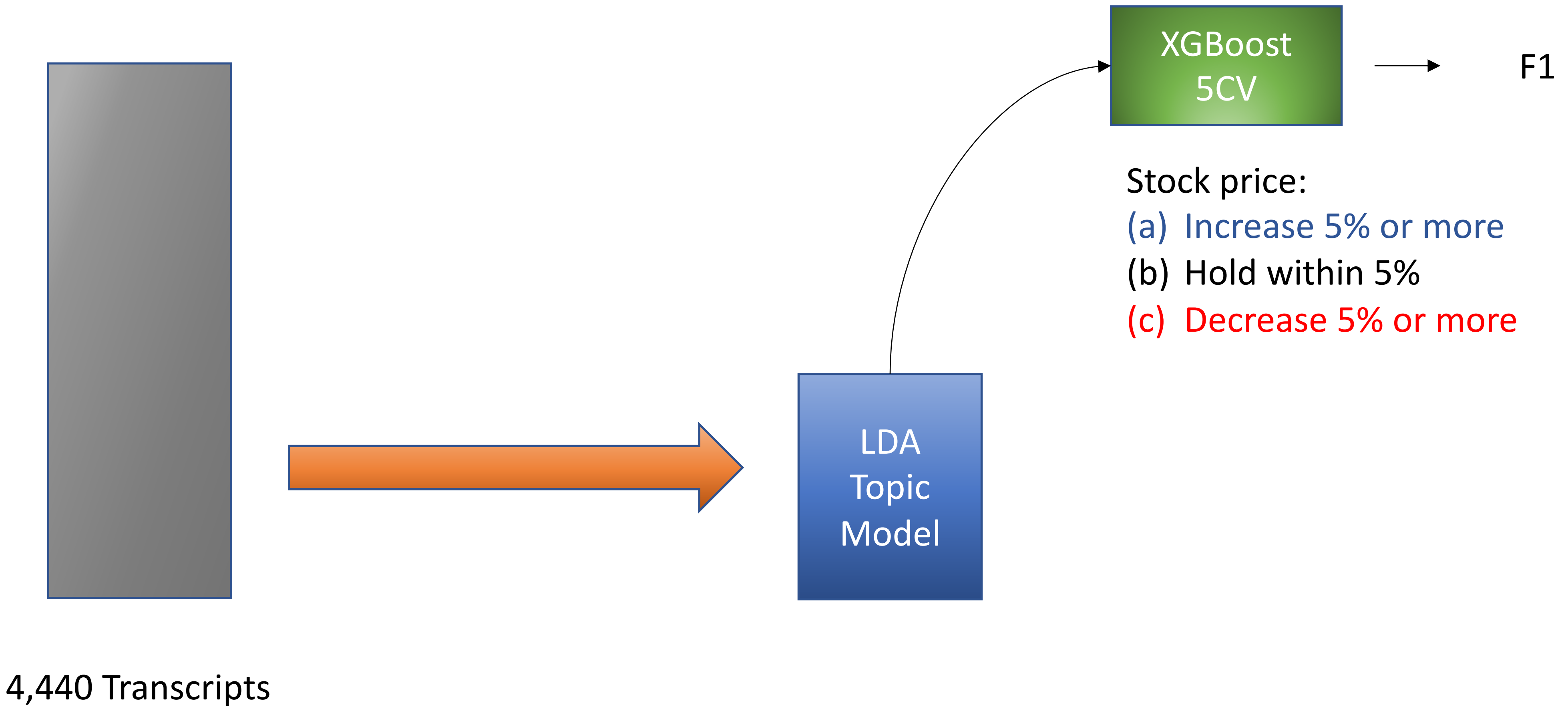
… but how build evidence for the topic structure?



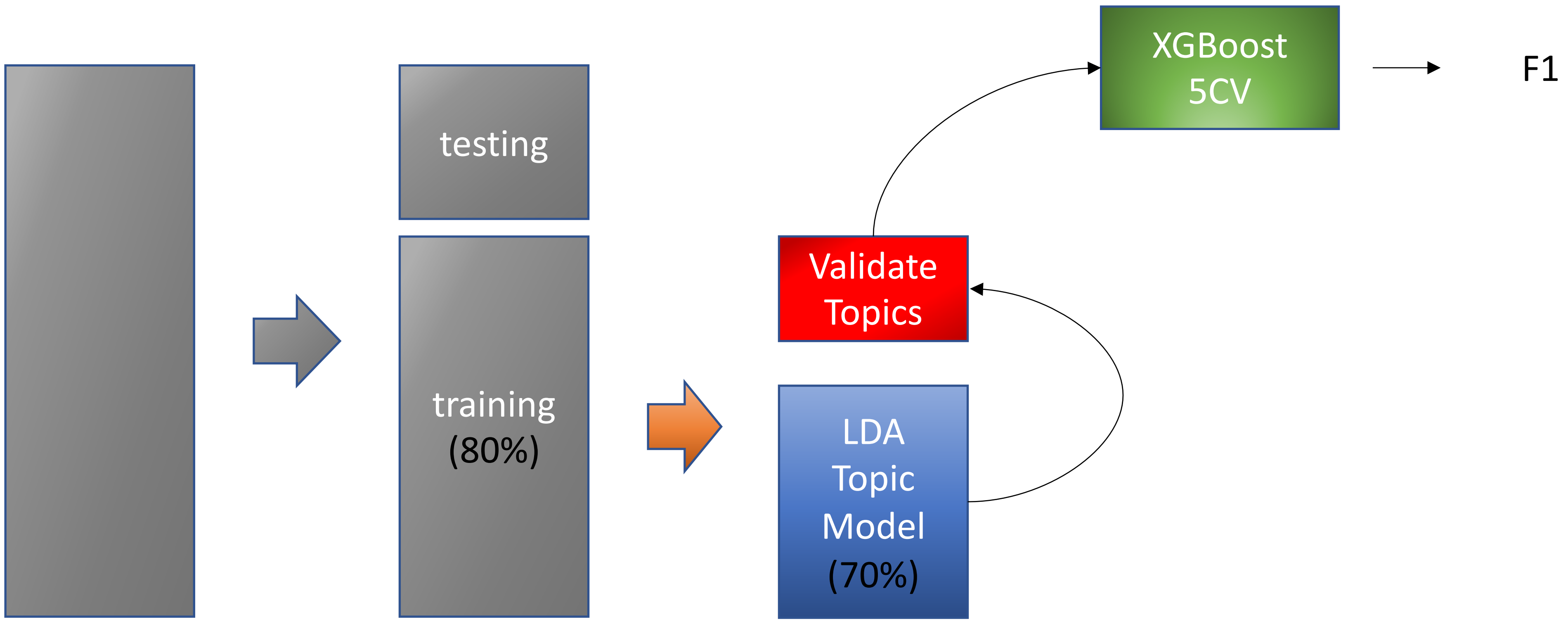
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 6 (9% of tokens)
0
2,000
4,000
6,000
8,000
10,000
memory
nand
nanometer
semiconductor
foundry
logic
dram
flash
c
ssd
storage
materials
calendar
wafer
test
ho
semi
packaging
euv
fab
edge
shipments
node
pc
nodes
utilization
diffely
mehdi
controller
automotive
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Methods: Substantive topics will have predictive validity.

Methods: Substantive topics will have predictive validity.

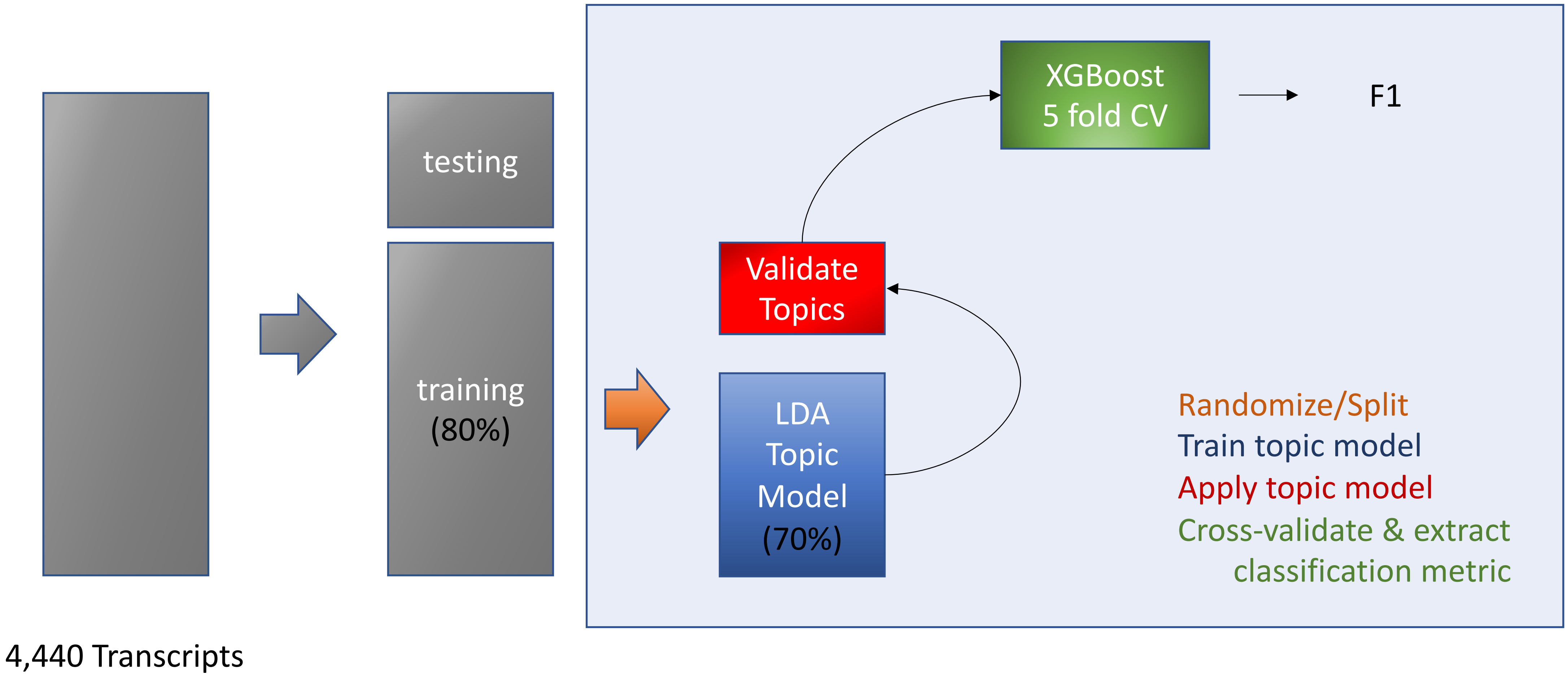


Methods: Substantive topics will have predictive validity.

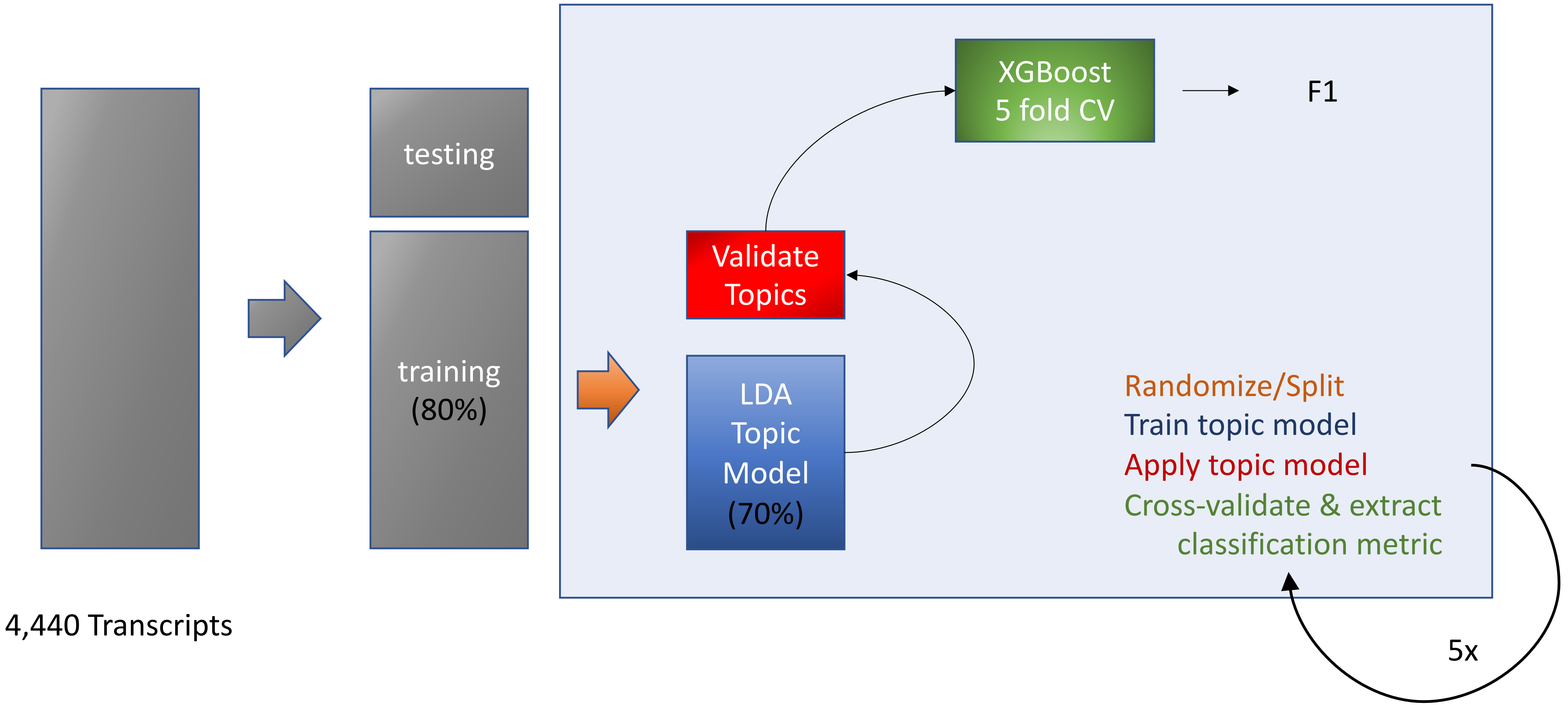


4,440 Transcripts

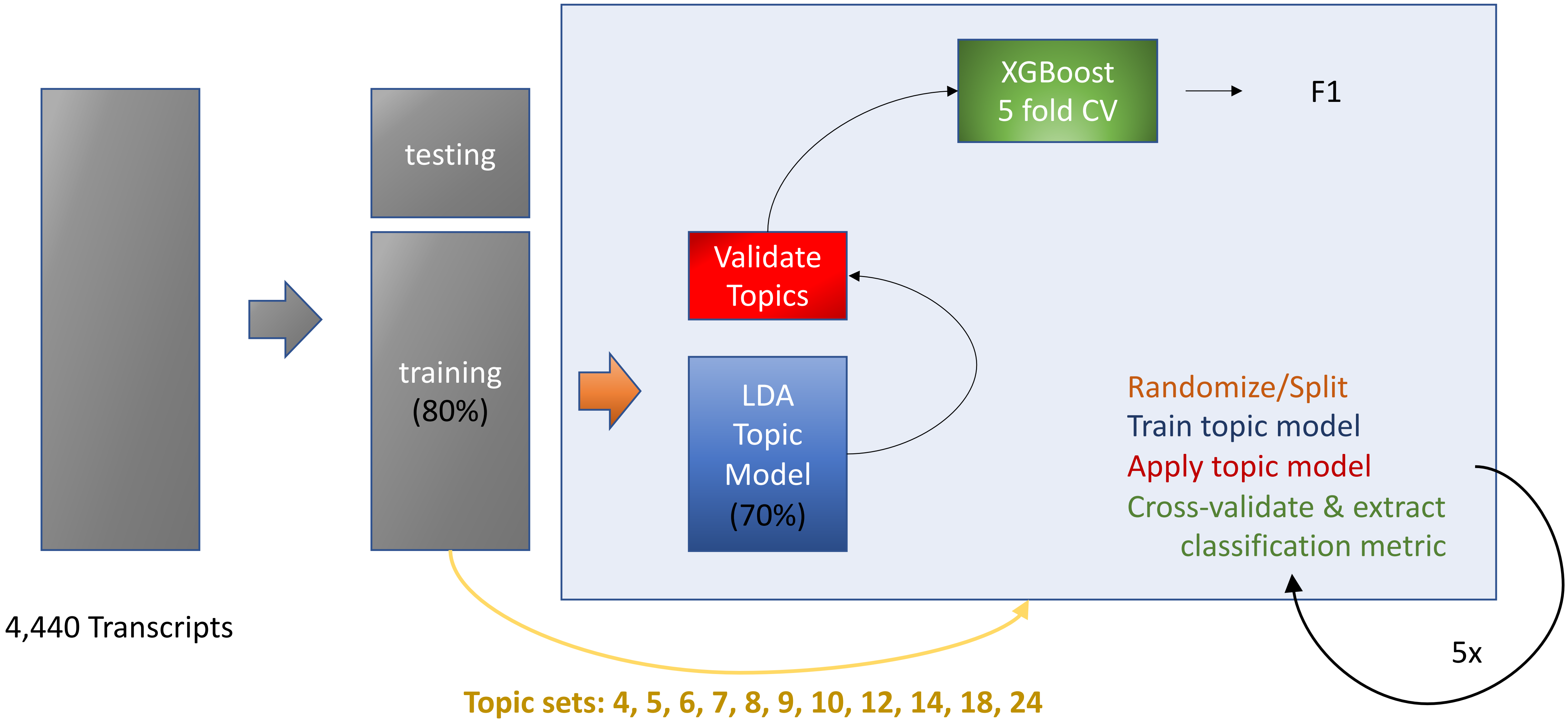
Methods


testing
training
(80%)
XGBoost
5 fold CV
F1
Validate
Topics
LDA
Topic
Model
(70%)
Randomize/Split
Train topic model
Apply topic model
Cross-validate & extract
classification metric
4,440 Transcripts

Methods

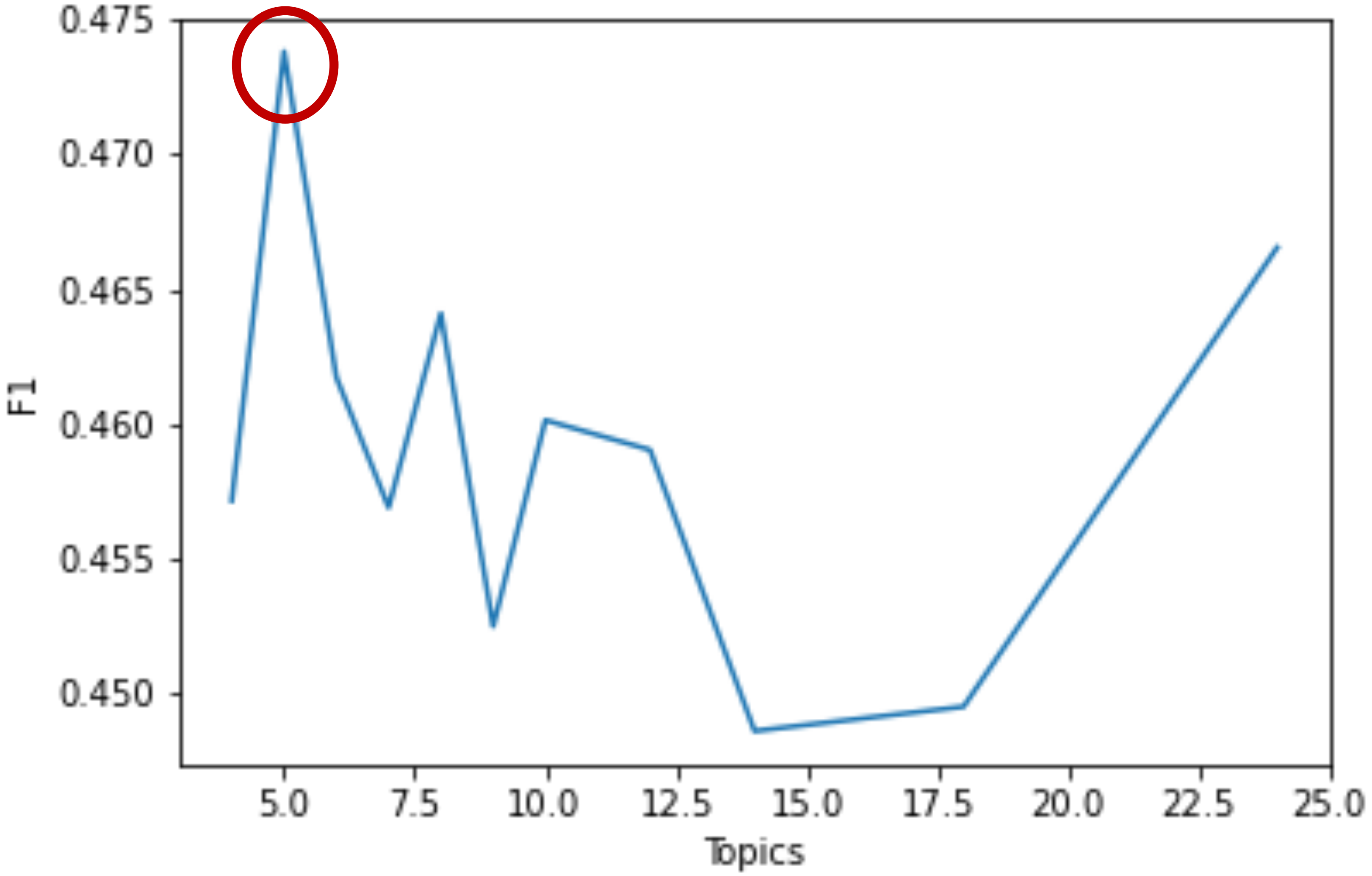

testing
training
(80%)
XGBoost
5 fold CV
F1
Validate
Topics
LDA
Topic
Model
(70%)
Randomize/Split
Train topic model
Apply topic model
Cross-validate & extract
classification metric
4,440 Transcripts
5x

Methods



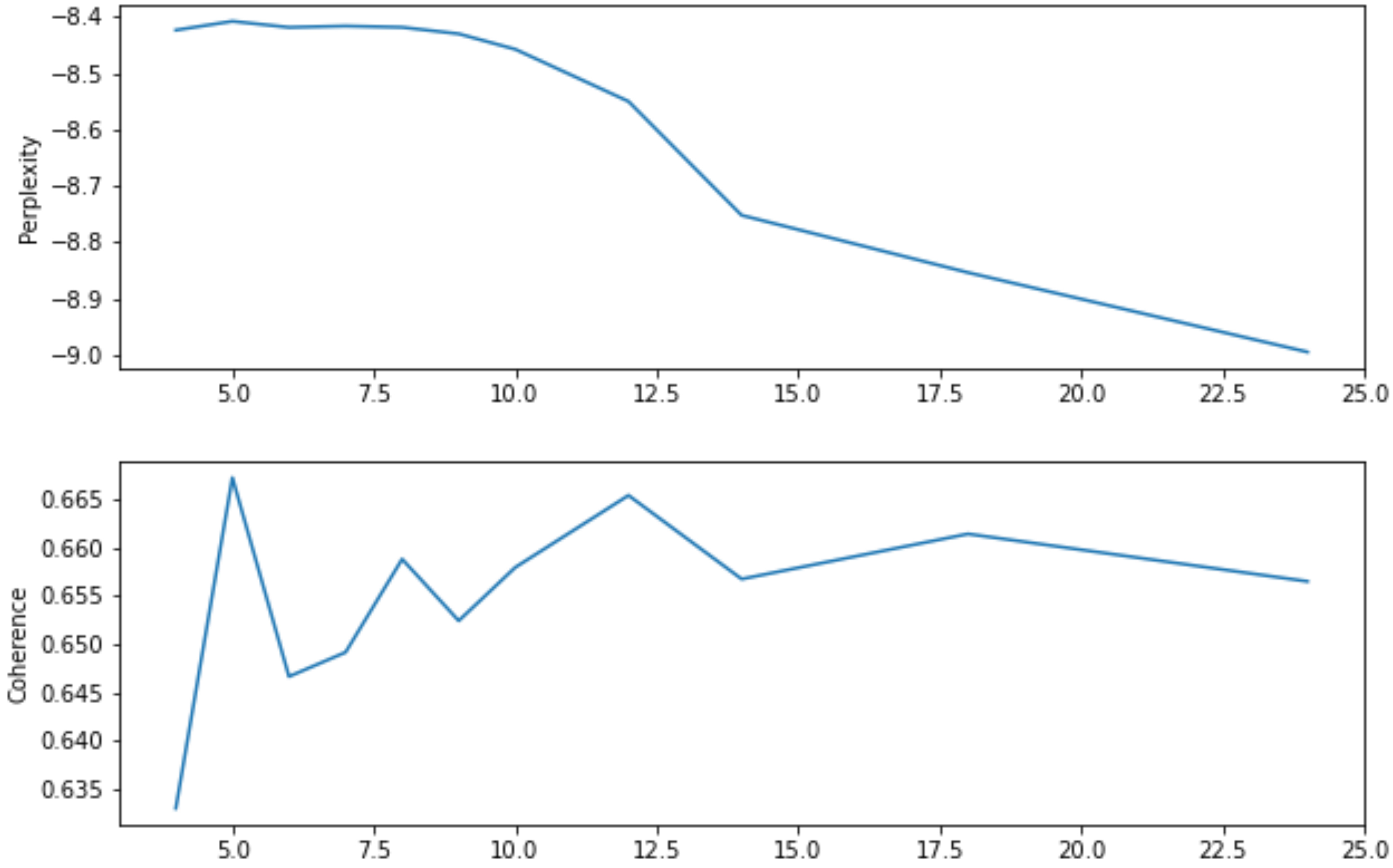
testing
training
(80%)
4,440 Transcripts
LDA
Topic
Model
(70%)
Validate
Topics
XGBoost
5 fold CV
F1
Randomize/Split
Train topic model
Apply topic model
Cross-validate & extract
classification metric
5x
Topic sets: 4, 5, 6, 7, 8, 9, 10, 12, 14, 18, 24

Outcome Metrics: F1



This is **not** the evidence you are looking for

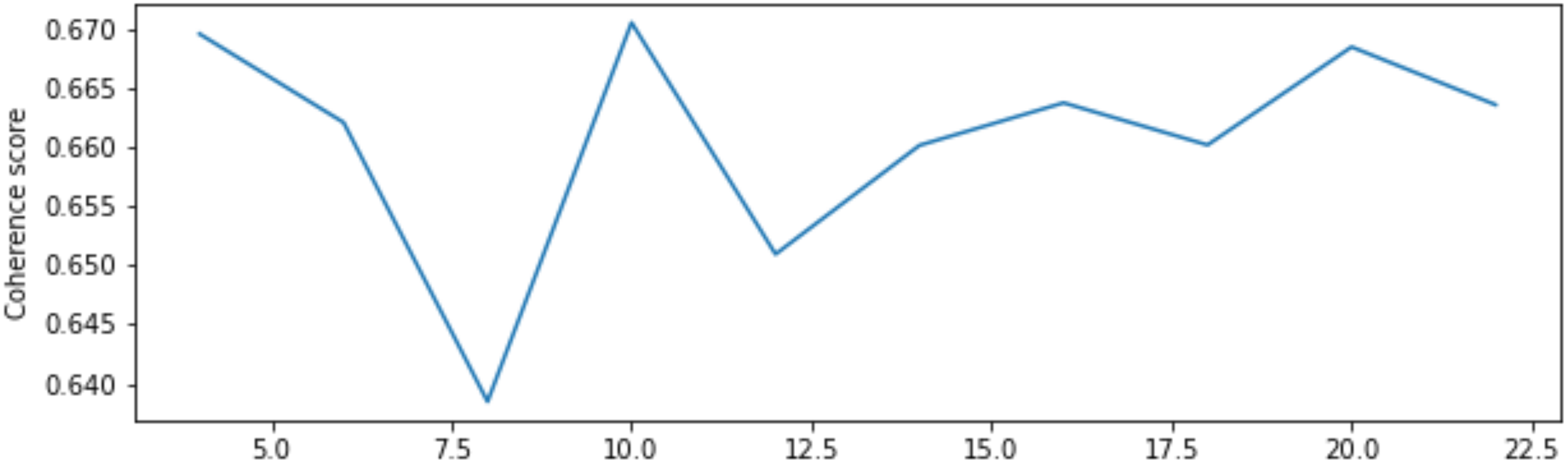
Outcome Metrics: Perplexity & Coherence



This is **not** the evidence you are looking for

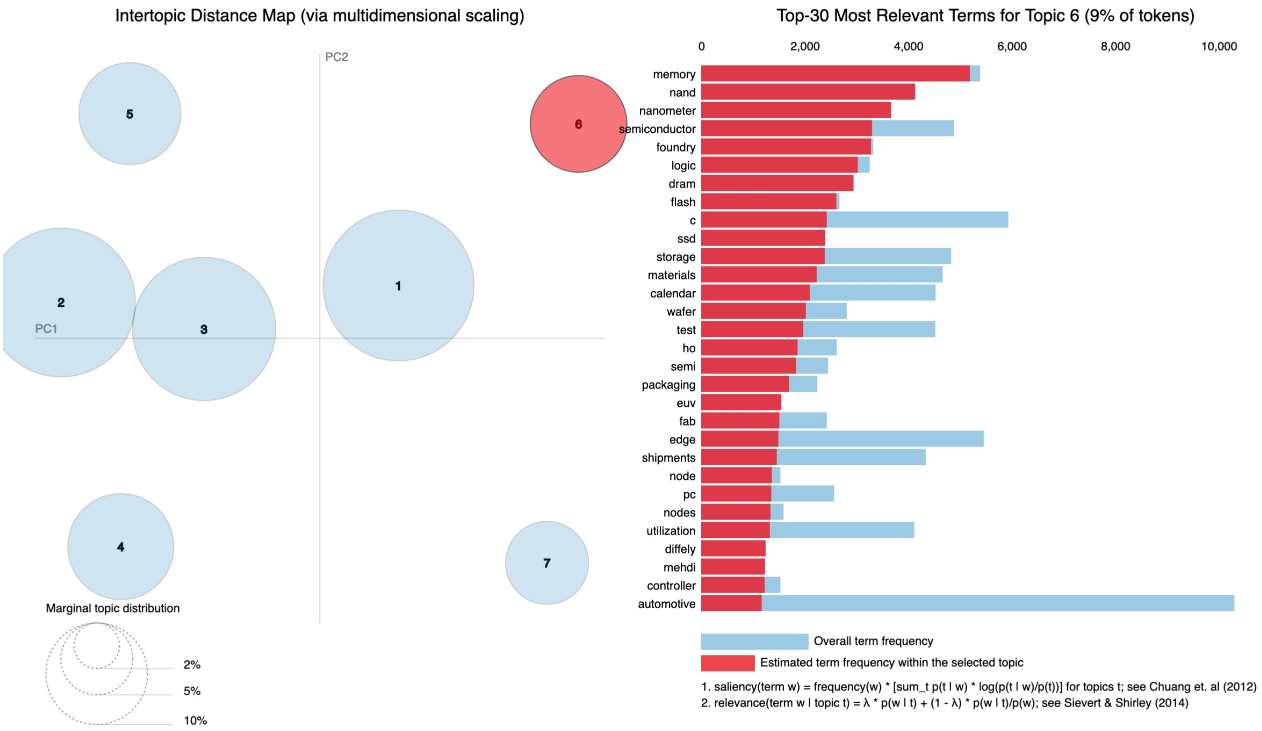
Gensim LDA -> MAchine Learning for LanguagE Toolkit

- Gensim's Variational Bayes sampling is fast.
- MALLET's Gibbs sampling is more precise.

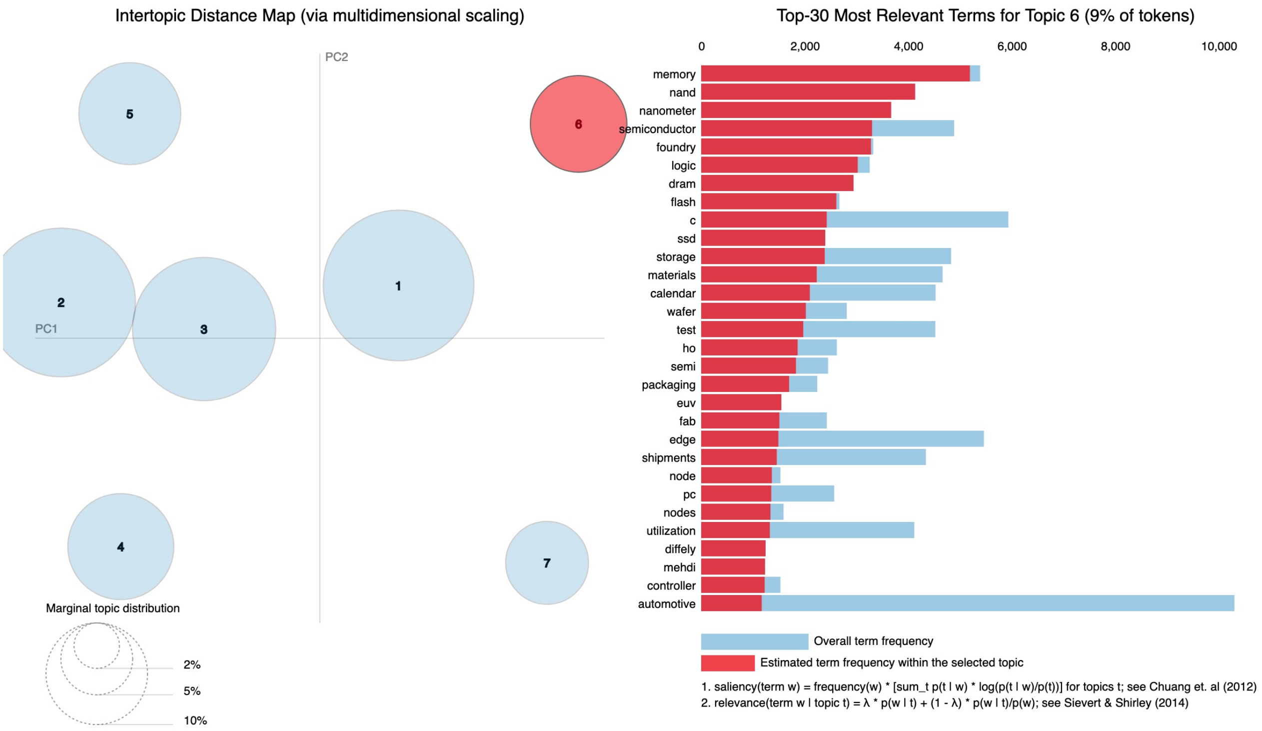


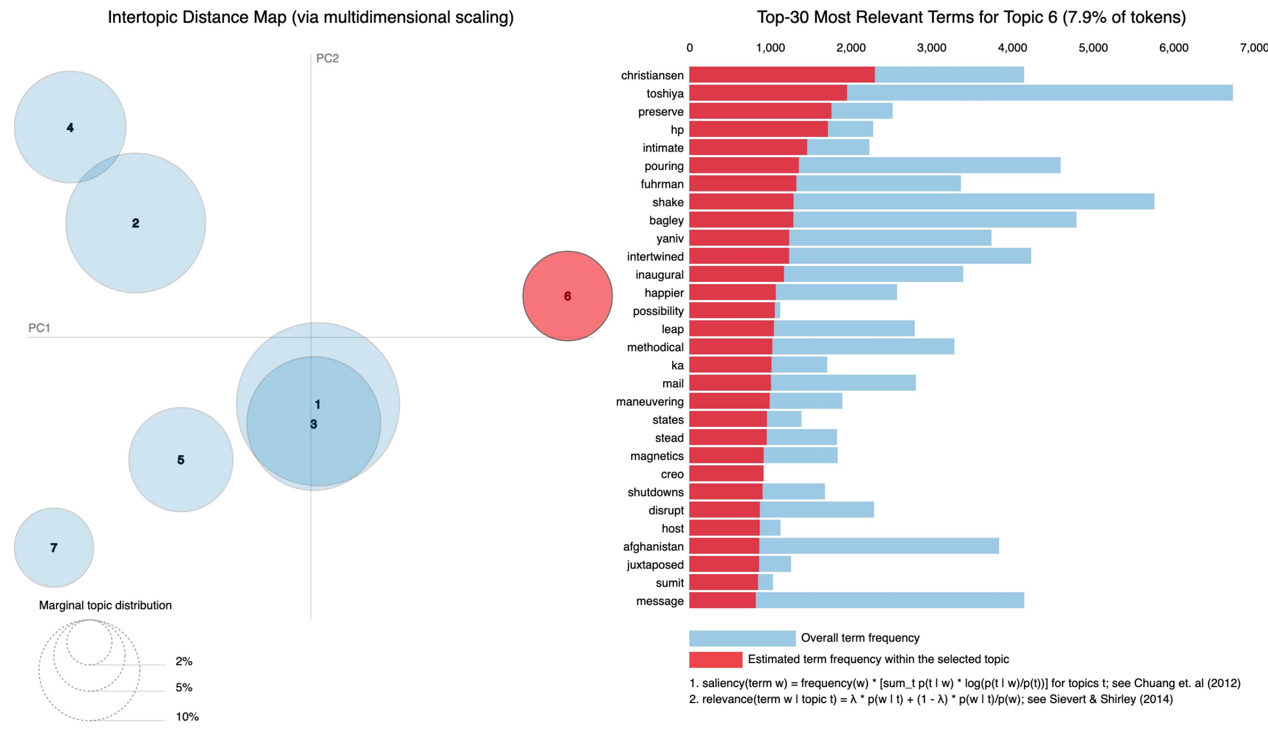
This is **not** the evidence you are looking for

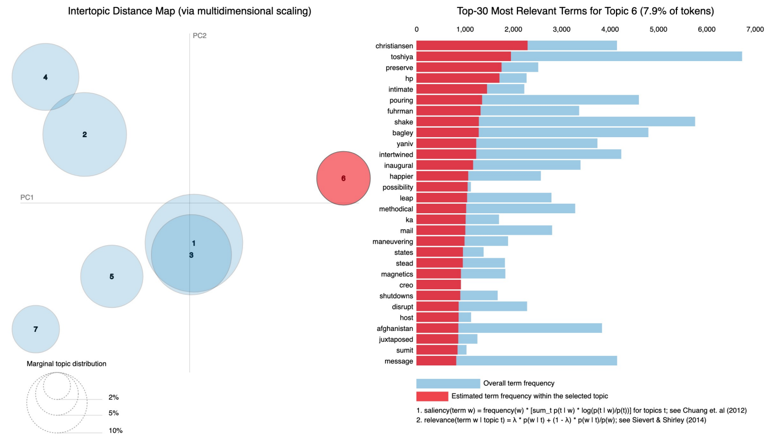
Me: I should use that initial 7-topic model to examine connections to other features – time trends, company affiliation and such.

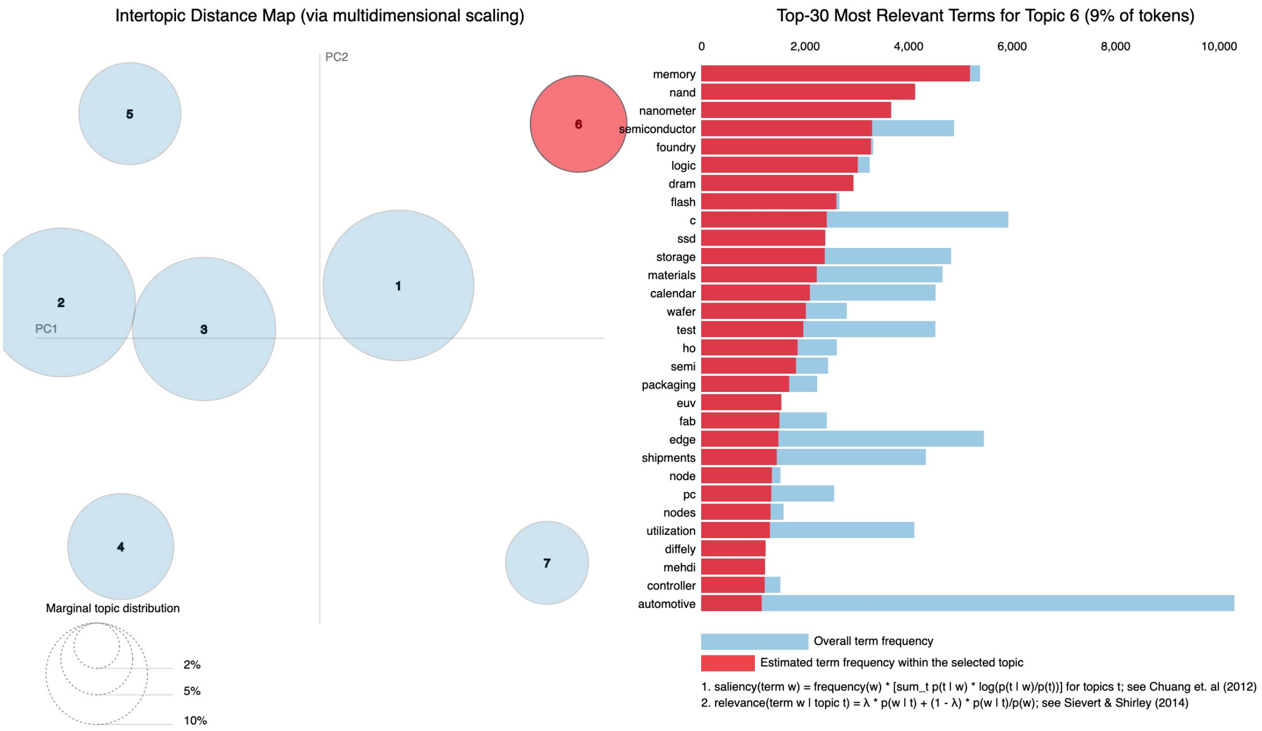


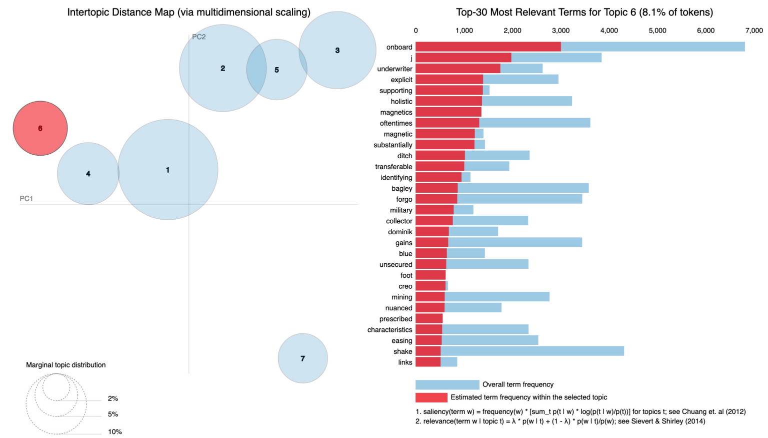
Me: I should use that initial 7-topic model to examine connections to other features – time trends, company affiliation and such.

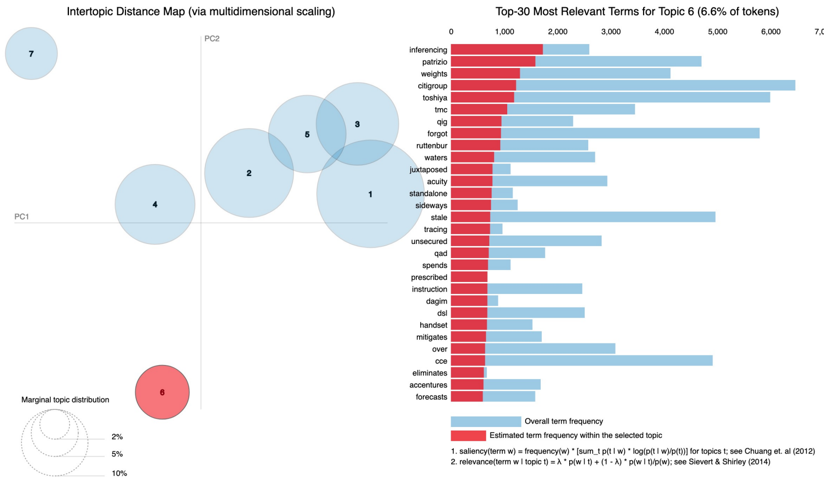





Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
4
2
6
1
3
5
7
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 6 (7.9% of tokens)
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
christiansen
toshiya
preserve
hp
intimate
pouring
fuhrman
shake
bagley
yaniv
intertwined
inaugural
happier
possibility
leap
methodical
ka
mail
maneuvering
states
stead
magnetics
creo
shutdowns
disrupt
host
afghanistan
juxtaposed
sumit
message
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
5
6
2
3
1
4
7
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 6 (9% of tokens)
0
2,000
4,000
6,000
8,000
10,000
memory
nand
nanometer
semiconductor
foundry
logic
dram
flash
c
ssd
storage
materials
calendar
wafer
test
ho
semi
packaging
euv
fab
edge
shipments
node
pc
nodes
utilization
diffely
mehdi
controller
automotive
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
2
5
3
6
4
1
7
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 6 (8.1% of tokens)
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
onboard
j
underwriter
explicit
supporting
holistic
magnetics
oftentimes
magnetic
substantially
ditch
transferable
identifying
bagley
forgo
military
collector
dominik
gains
blue
unsecured
foot
creo
mining
nuanced
prescribed
characteristics
easing
shake
links
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
7
5
3
2
4
1
6
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 6 (6.6% of tokens)
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
inferencing
patrizio
weights
citigroup
toshiya
tmc
qig
forgot
ruttenbur
waters
juxtaposed
acuity
standalone
sideways
stale
tracing
unsecured
qad
spends
prescribed
instruction
dagim
dsl
handset
mitigates
over
cce
eliminates
accentures
forecasts
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

This is **not** the topic structure you are looking for

Me: Could I use the terms from the initial 7-topic model to seed a Corex model and extract topics from there?

Corex: Sure thing!

Me: Could I use the terms from the initial 7-topic model to seed a Corex model and extract topics from there?

Corex: Sure thing!
… but I’ll give most observations a 99% chance of being grouped into 3 or more topics.

Me: Could I use the terms from the initial 7-topic model to seed a Corex model and extract topics from there?

Corex:

Me:

But I could create a set of topics, say 50 or so, and then I could use clustering to reduce them to a meaningful set, right?

Me: But I could create a set of topics, say 50 or so, and then I could use clustering to reduce them to a meaningful set, right?

DBSCAN: Heck yeah!

Me: But I could create a set of topics, say 50 or so, and then I could use clustering to reduce them to a meaningful set, right?

DBSCAN: Heck yeah!
… but I’m going to cluster 90% of all observations into **a single cluster**, no matter how to you tweak epsilon and neighbors.

Stable/reliable construction of topics: Failed
Semi-supervised topic models: Failed
Clustering of Topic models: Failed

Validation on the outcome: Failed
Validation on topic model metrics: Failed
Improved modeling approach: Failed