

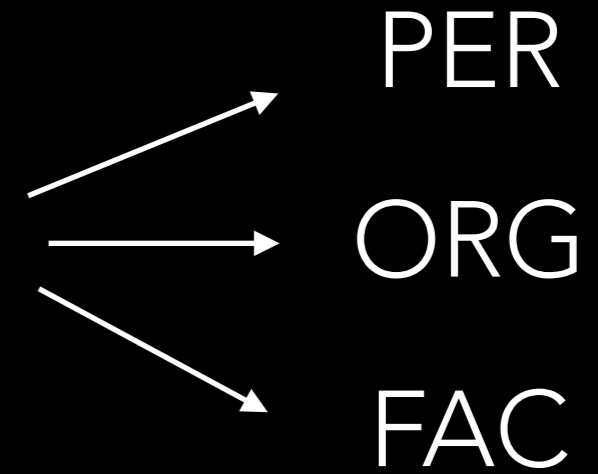
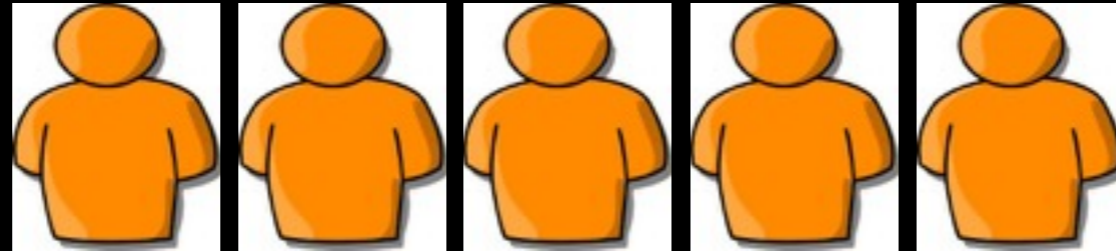
DISTRIBUTED NON-PARAMETRIC  
REPRESENTATIONS FOR VITAL FILTERING:  
UW @ TREC KBA 2014

IGNACIO CANO - SAMEER SINGH - CARLOS GUESTRIN

TEXT RETRIEVAL CONFERENCE- NOVEMBER 2014

# VITAL FILTERING (VF)

entities



documents



each doc

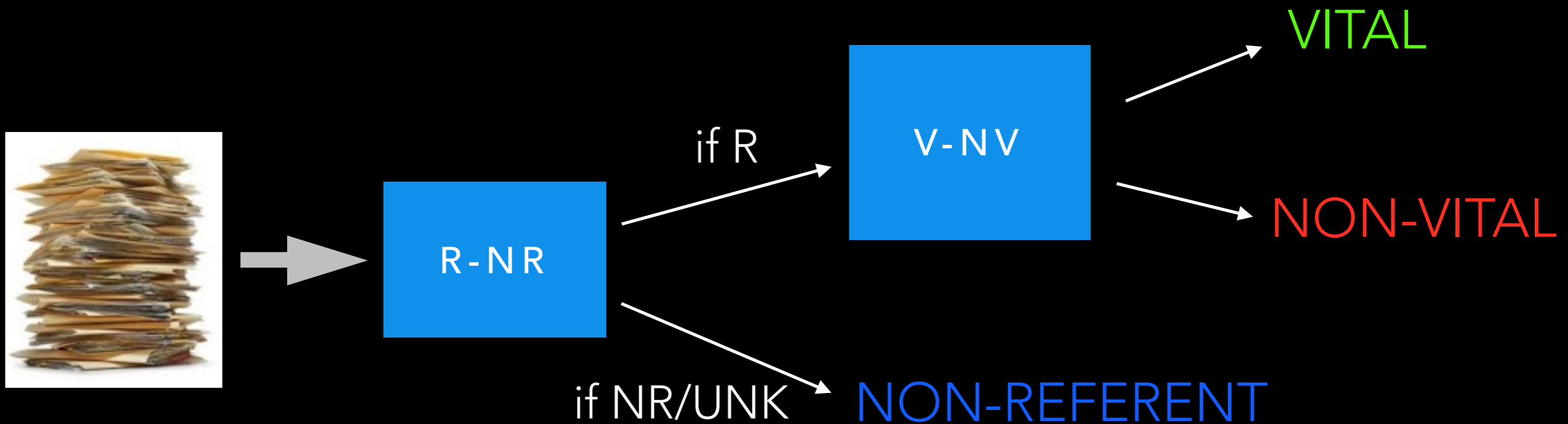


+



# VITAL FILTERING (VF)

We use two classifiers in cascade.



# Agenda

- **METHOD**
- EXPERIMENTS
- ACCELERATE & CREATE
- CONCLUSION

# METHOD

- We need to predict whether a document contains **novel information** for an entity
- If we can "**summarize**" the contents seen for an entity
- We can compare the document to the entity's summary
  - If similar, we have probably seen the information
  - If dissimilar, it is novel -> predict VITAL
- How can we **accurately and efficiently** "summarize" all the documents that we have seen for an entity?

# BAG OF WORDS

+ word counts are simple to implement,  
often work quite well

- sparsity

“Barack was elected president today”

“Obama has won the election”

- size of representation grows over time

# WORD EMBEDDINGS

- Vector representations of words
  - fixed length array of real values

$$V_{\text{france}} = (0.002, 0.2, 0.4\dots)$$

$$V_{\text{paris}} = (0.005, 0.18, 0.43\dots)$$

# WORD EMBEDDINGS

x2

france  
austria      belgium  
germany

president  
minister

leader  
chief      head

four  
two  
three  
six

those  
these

blue      black  
red

x1



# DOCUMENT EMBEDDING

e = Ted Sturdevant

d = ... So, it comes as good news that state officials are making significant progress in determining which chemicals pollute Puget Sound and in identifying where they come from. A report last week from the Department of Ecology casts a wide net over culprits. Most toxic chemicals are used in some way by all of us, said Ecology Director Ted Sturdevant. They are in our homes and gardens...

# DOCUMENT EMBEDDING

e = Ted Sturdevant

d = ... So, it comes as good news that state officials are making significant progress in determining which chemicals pollute Puget Sound and in identifying where they come from. A report last week from the Department of Ecology casts a wide net over culprits. **Most toxic chemicals are used in some way by all of us, said Ecology Director Ted Sturdevant.** They are in our homes and gardens...

# DOCUMENT EMBEDDING

e = Ted Sturdevant

Most toxic chemicals are used in some way by all of

RBS JJ NNS VBP VBN IN DT NN IN DT IN

us, said Ecology Director Ted Sturdevant

PRP VBD NNP NNP NNP NNP

# DOCUMENT EMBEDDING

e = Ted Sturdevant

chemicals are used

way

NNS VBP VBN

NN

said Ecology Director Ted Sturdevant

VBD NNP NNP NNP NNP

# DOCUMENT EMBEDDING

$e = \text{Ted Sturdevant}$

$W = [\text{chemicals, are, used, way, said, ecology, director, } \del{\text{ted\_sturdevant}}]$

$v_{\text{chemicals}} = f(\text{chemicals})$

$v_{\text{are}} = f(\text{are})$

$v_{\text{said}} = f(\text{said})$

$v_{\text{used}} = f(\text{used})$

$v_{\text{ecology}} = f(\text{ecology})$

$v_{\text{way}} = f(\text{way})$

$v_{\text{director}} = f(\text{director})$

$\mathbf{v}_d = \frac{1}{|W|} ( \mathbf{v}_{\text{chemicals}} + \mathbf{v}_{\text{are}} + \mathbf{v}_{\text{used}} + \mathbf{v}_{\text{way}} + \mathbf{v}_{\text{said}} + \mathbf{v}_{\text{ecology}} + \mathbf{v}_{\text{director}} )$

# METHOD

- So far we have a representation for documents.
- To summarize what we have seen of an entity

$$v_e = \frac{1}{|D_e|} \sum_{d \in D_e} v_d$$

- Compare embedding of doc to entity's
  - If close -> predict NON-VITAL

e=Obama

x2

senator



x1

e=Obama

x2

senator



x1



e=Obama

x2

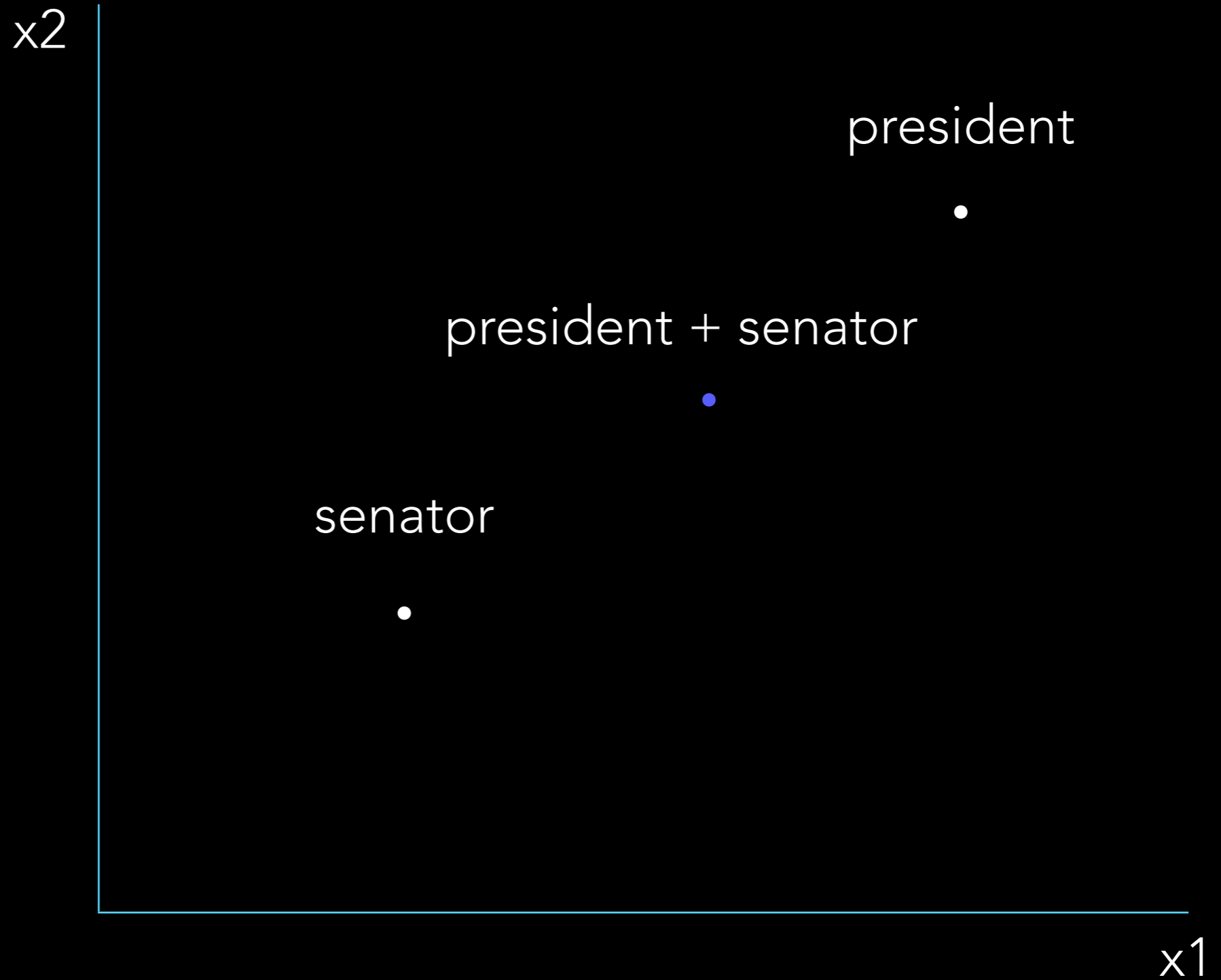
president

senator

x1



e=Obama



# MULTIPLE EMBEDDINGS FOR ENTITIES

- Use multiple embeddings to capture the different clusters of an entity.
- We propose an algorithm that:
  - assigns each document to a single cluster,
  - adaptively grows the number of clusters
  - is efficient

e=Obama

x2

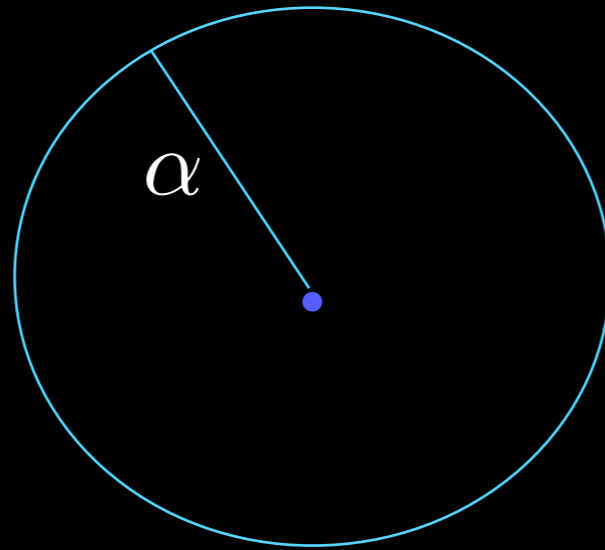


x1

e=Obama

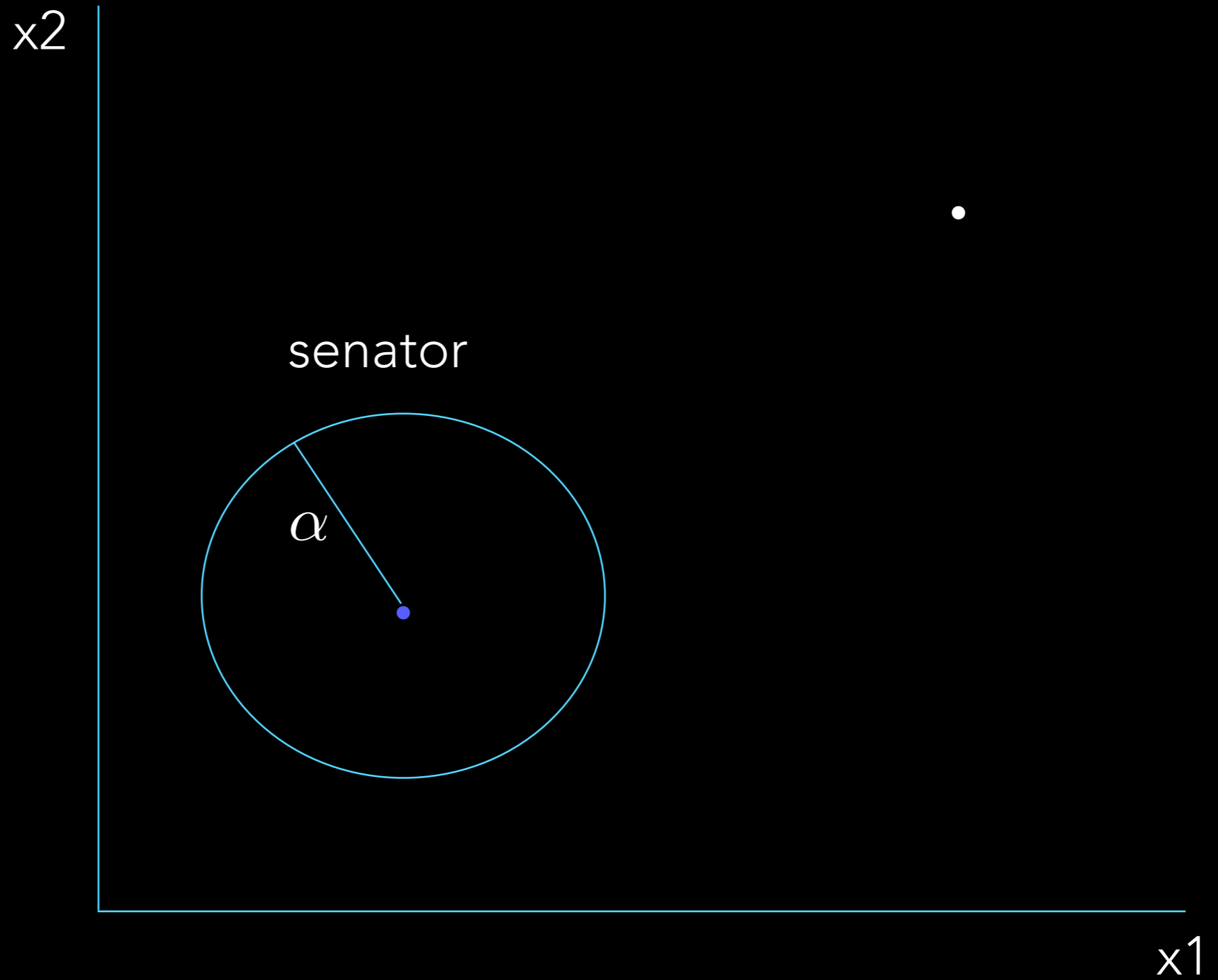
x2

senator

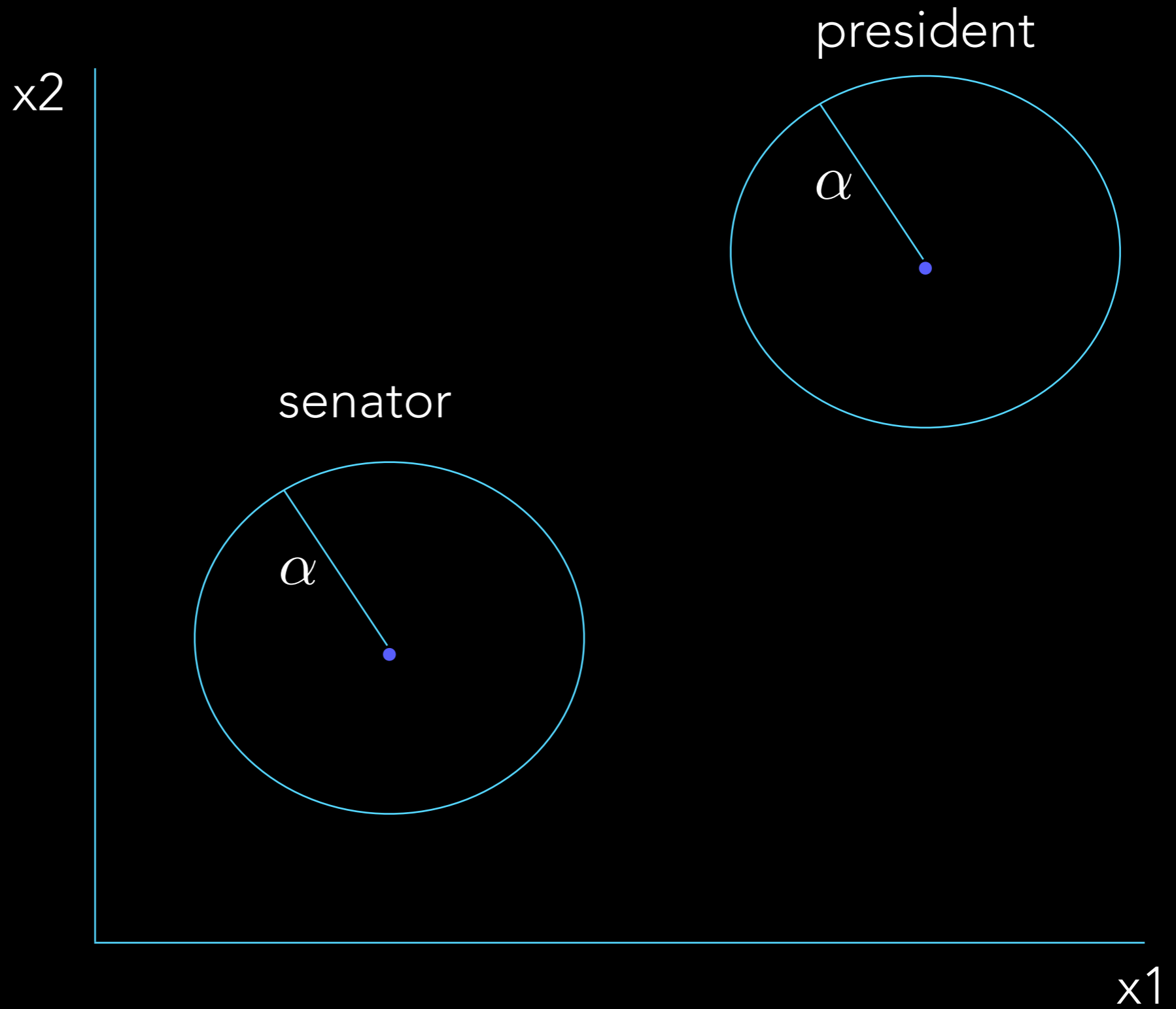


x1

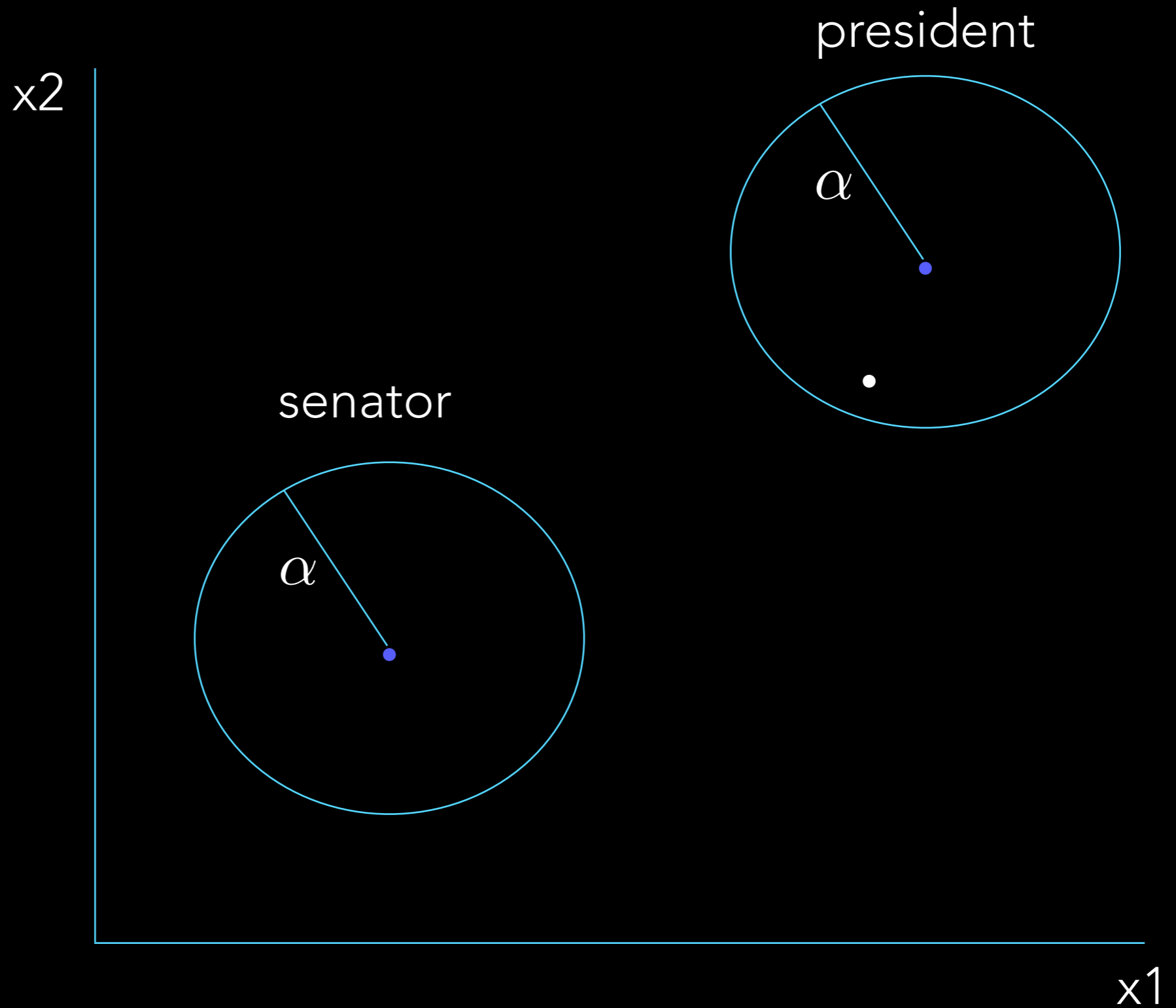
e=Obama



e=Obama

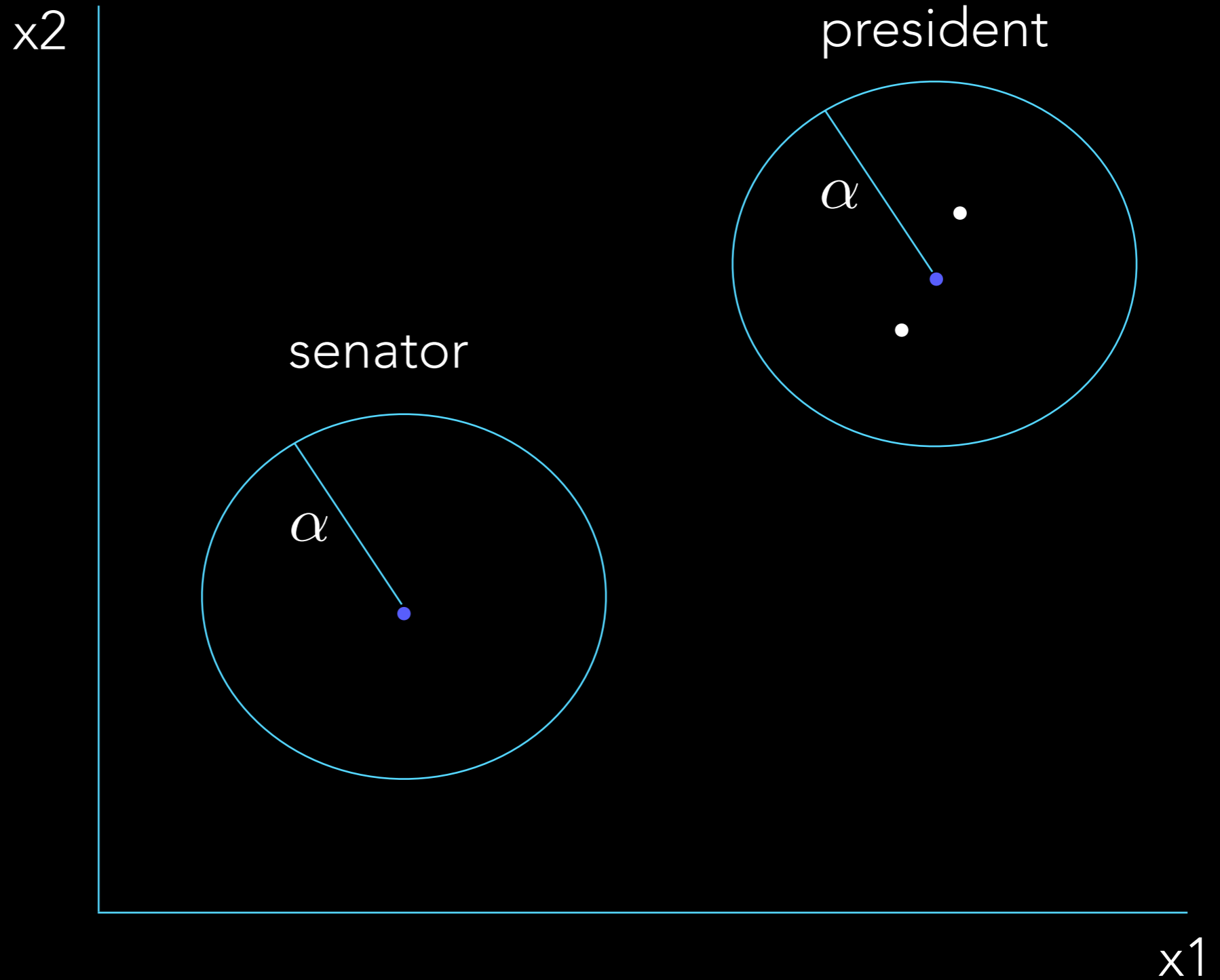


e=Obama





e=Obama



# METHOD

- Take embedding of document
- Compare to all cluster vectors of an entity
  - If close to ANY -> we have seen similar document
  - If far from ALL -> document contains novel info

# METHOD

- Unfortunately this is not accurate

“Obama wins the presidential election” (2008)

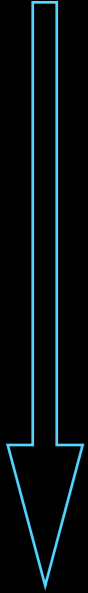
“Obama wins his second presidential election” (2012)

# METHOD

- Propose a staleness measure
  - low staleness -> novel, freshness, vital
  - high staleness -> old, stale, non-vital

e=Obama

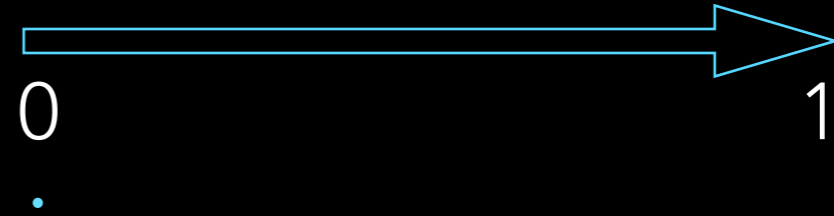
time



Obama Wins Election - NYTimes.com - The New York Times  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.

STALENESS

p(vital)



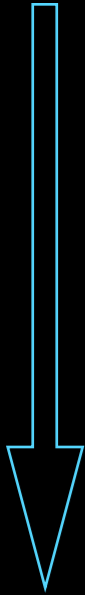
high

e=Obama

STALENESS

p(vital)

time



Obama Wins Election - NYTimes.com - The New York Times  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.

Obama Wins Election - NYTimes.com - The New York Times  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.

0

1

high

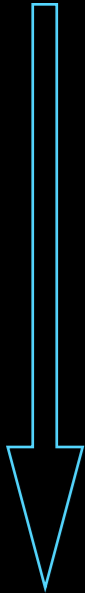
high

e=Obama

STALENESS


p(vital)

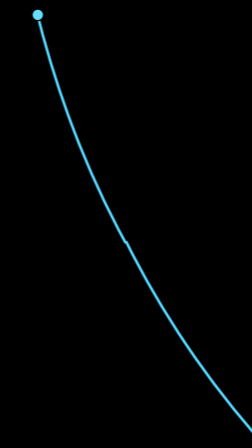
time



**Obama Wins Election - NYTimes.com - The New York Times**  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein **Obama** was elected the 44th **president** of the United States on Tuesday, as the country chose him as its **first** black chief executive. ... Monitor live results all night on The Times's **Election 2008** Dashboard.

**Obama Wins Election - NYTimes.com - The New York Times**  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein **Obama** was elected the 44th **president** of the United States on Tuesday, as the country chose him as its **first** black chief executive. ... Monitor live results all night on The Times's **Election 2008** Dashboard.

**CNN Announces Obama as the Next President Elect ...**  
 [www.youtube.com/watch?v=9dKAKII1bUE](http://www.youtube.com/watch?v=9dKAKII1bUE)  
Nov 5, 2008 - Uploaded by GettingtotheTruth2  
Lies upon lies comming from the shit hole which is **obamas** mouth.  
.... **Election 2008** - United States **Presidential** Inauguration by Marc Schulz ...



high

high

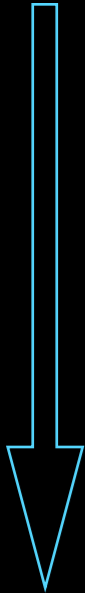
high

e=Obama

STALENESS


p(vital)

time



**Obama Wins Election - NYTimes.com - The New York Times**  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.

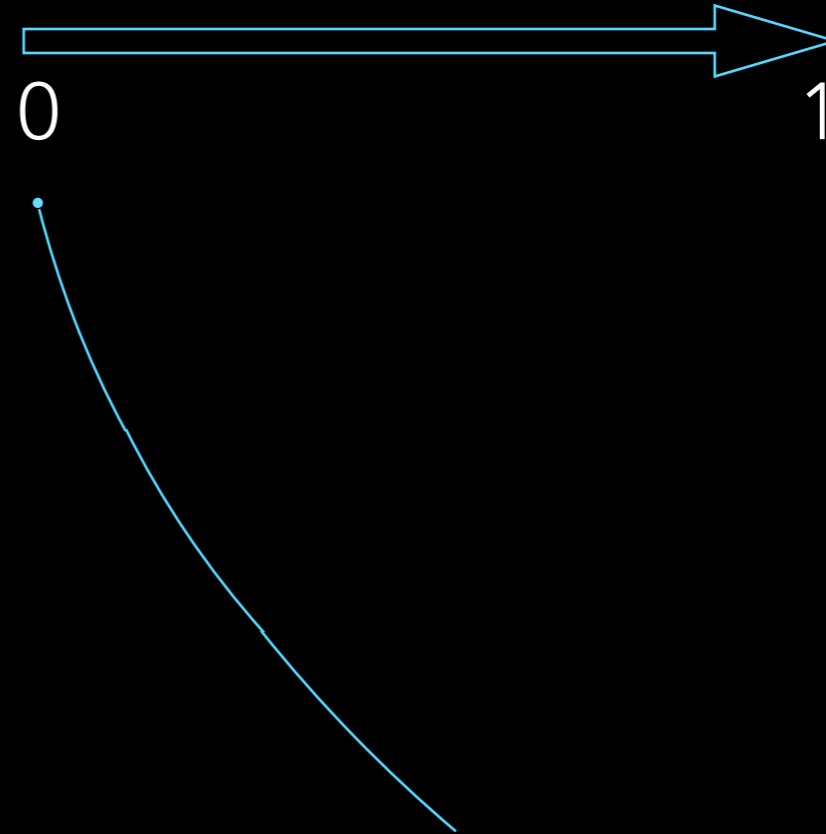
**Obama Wins Election - NYTimes.com - The New York Times**  
[www.nytimes.com/2008/11/05/campaign.html](http://www.nytimes.com/2008/11/05/campaign.html) ... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.

**CNN Announces Obama as the Next President Elect ...**  
 [www.youtube.com/watch?v=9dKAKII1bUE](http://www.youtube.com/watch?v=9dKAKII1bUE)  
Nov 5, 2008 - Uploaded by GettingtotheTruth2  
Lies upon lies comming from the shit hole which is obamas mouth.  
.... Election 2008 - United States Presidential Inauguration by Marc Schulz ...

**American President: Barack Obama: Campaigns and ...**  
[millercenter.org/president/obama/essays/](http://millercenter.org/president/obama/essays/) ... Miller Center of Public Affairs  
Obama's election to the Senate instantly made him the highest-ranking ..... His margin of victory was down slightly from 2008, making him the first President since ...

0

1



high

high

high

high

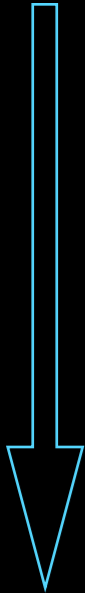



e=Obama

STALENESS

p(vital)

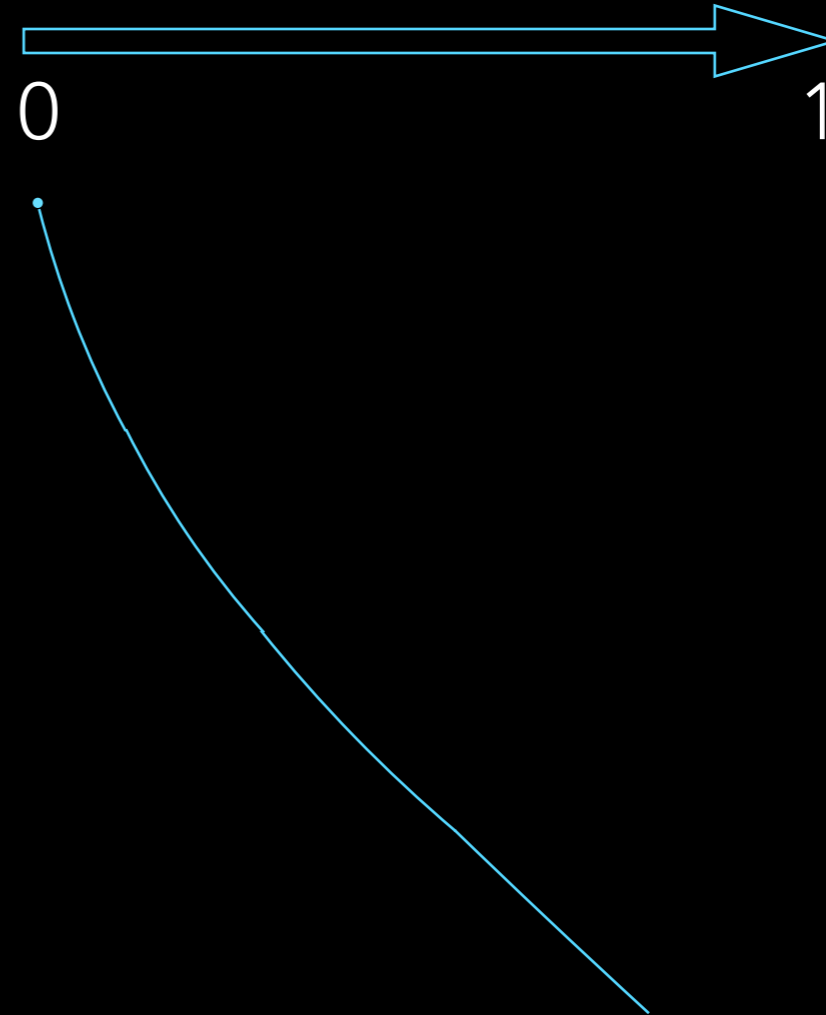
time



- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- CNN Announces Obama as the Next President Elect ...**  
 www.youtube.com/watch?v=9dKAKII1bUE  
Nov 5, 2008 - Uploaded by GettingtotheTruth2  
Lies upon lies comming from the shit hole which is obamas mouth.  
.... Election 2008 - United States Presidential Inauguration by Marc Schulz ...
- American President: Barack Obama: Campaigns and ...**  
millercenter.org/president/obama/essays/... Miller Center of Public Affairs  
Obama's election to the Senate instantly made him the highest-ranking ..... His margin of victory was down slightly from 2008, making him the first President since ...
- Obama elected nation's 44th president - USATODAY.com**  
usatoday30.usatoday.com/news/.../election2008/president.ht... USA Today  
Nov 1, 2008 - Democrat Barack Obama has been elected the 44th president of the United ... making history by becoming the first African American to win the nation's highest office. ... Updated 11/5/2008 2:03 PM | Comment | Recommend, E-mail | Print ... in Phoenix after Barack Obama won the presidential election.

0

1



high

high

high

high

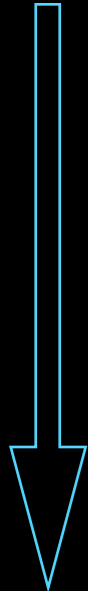
low

e=Obama

STALENESS

p(vital)

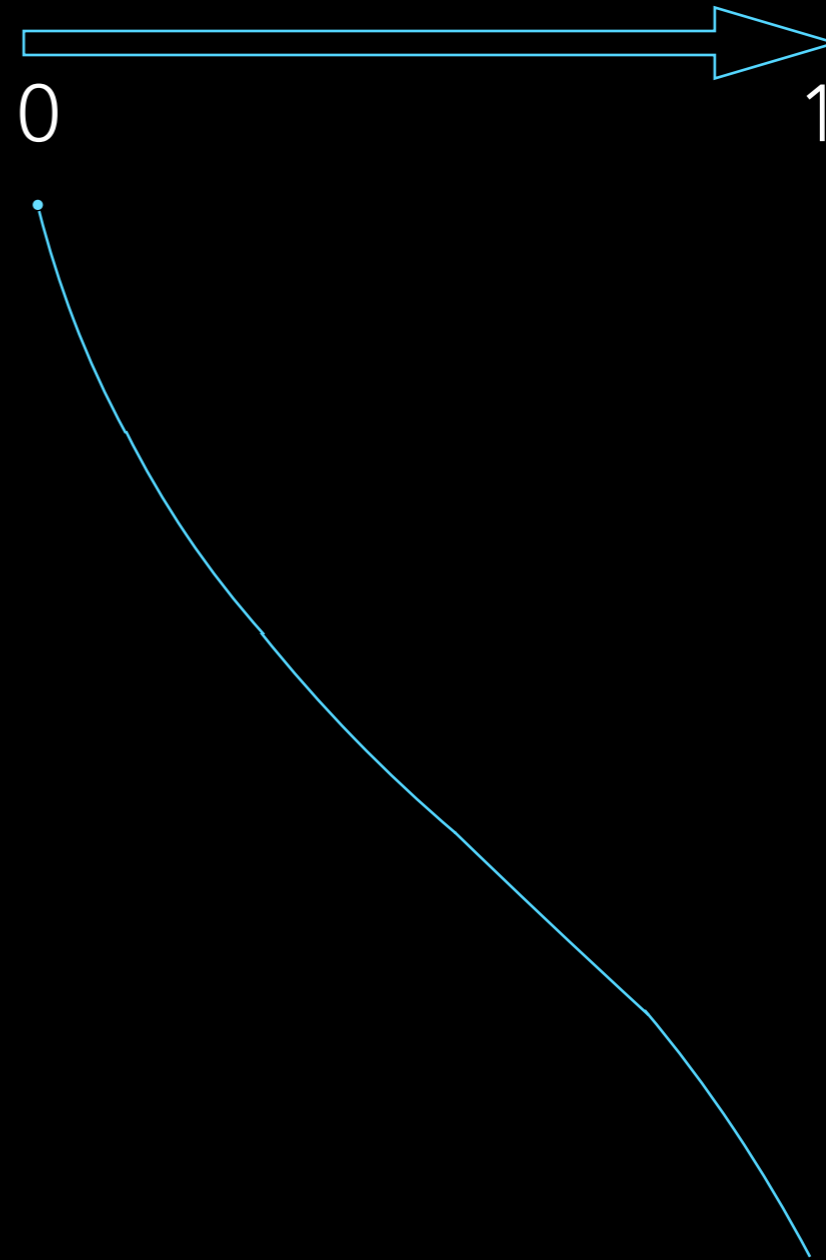
time



- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- CNN Announces Obama as the Next President Elect ...**  
 www.youtube.com/watch?v=9dKAKII1bUE  
Nov 5, 2008 - Uploaded by GettingtotheTruth2  
Lies upon lies comming from the shit hole which is obamas mouth.  
.... Election 2008 - United States Presidential Inauguration by Marc Schulz ...
- American President: Barack Obama: Campaigns and ...**  
millercenter.org/president/obama/essays/... Miller Center of Public Affairs  
Obama's election to the Senate instantly made him the highest-ranking ..... His margin of victory was down slightly from 2008, making him the first President since ...
- Obama elected nation's 44th president - USATODAY.com**  
usatoday30.usatoday.com/news/.../election2008/president.ht... USA Today  
Nov 1, 2008 - Democrat Barack Obama has been elected the 44th president of the United ... making history by becoming the first African American to win the nation's highest office. ... Updated 11/5/2008 2:03 PM | Comment | Recommend, E-mail | Print ... in Phoenix after Barack Obama won the presidential election.
- Presidential Election of 2008 - 270toWin.com**  
www.270towin.com/2008\_Election/ 270toWin.com  
Results of the presidential election of 2008, won by Barack H. Obama with 365 electoral ... It was also the first time two sitting senators ran against each other.

0

1



high

high

high

high

low

low

e=Obama

STALENESS

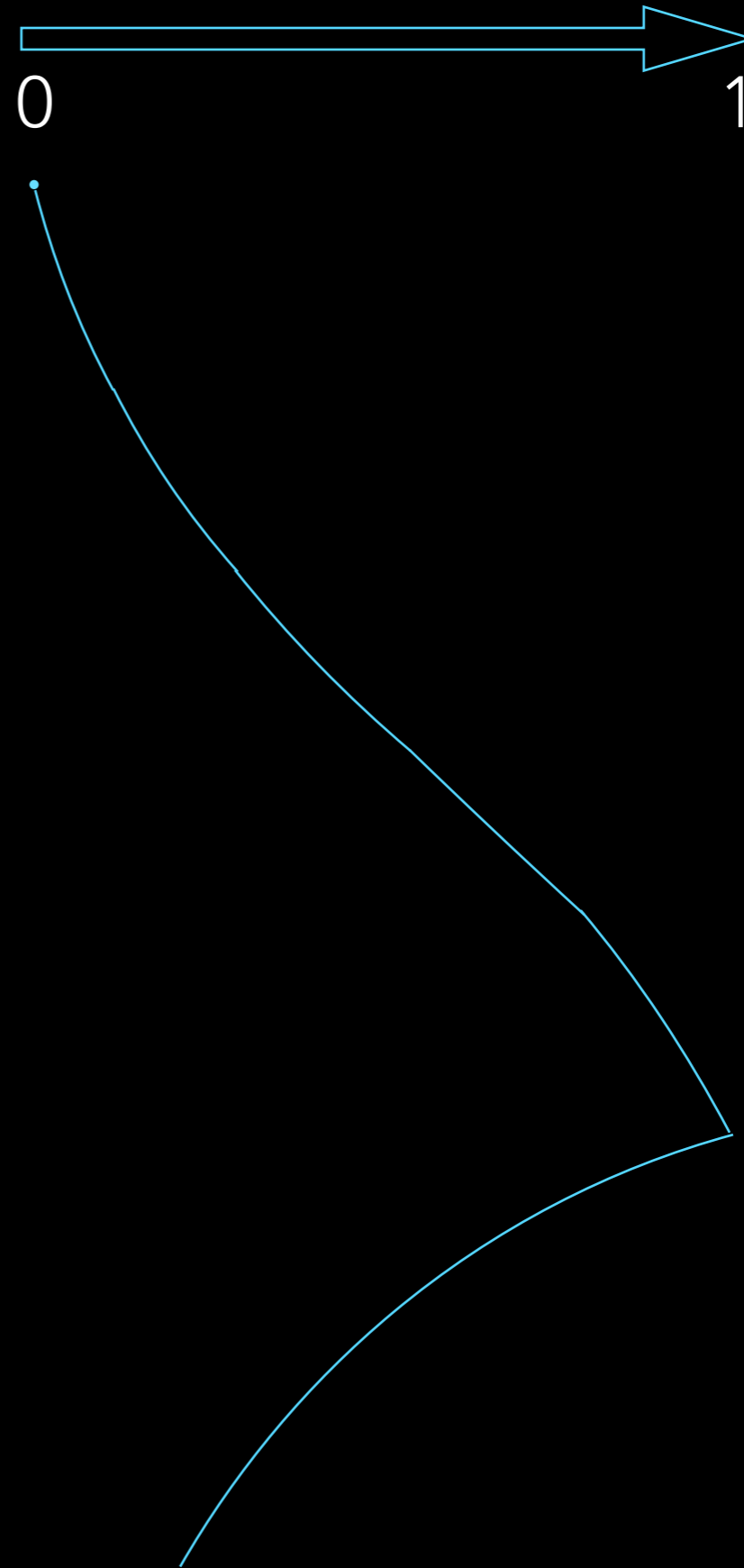
p(vital)

time



- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- CNN Announces Obama as the Next President Elect ...**  
 www.youtube.com/watch?v=9dKAKII1bUE  
Nov 5, 2008 - Uploaded by GettingtotheTruth2  
Lies upon lies comming from the shit hole which is obamas mouth.  
.... Election 2008 - United States Presidential Inauguration by Marc Schulz ...
- American President: Barack Obama: Campaigns and ...**  
millercenter.org/president/obama/essays/... Miller Center of Public Affairs  
Obama's election to the Senate instantly made him the highest-ranking ..... His margin of victory was down slightly from 2008, making him the first President since ...
- Obama elected nation's 44th president - USATODAY.com**  
usatoday30.usatoday.com/news/.../election2008/president.ht... USA Today  
Nov 1, 2008 - Democrat Barack Obama has been elected the 44th president of the United ... making history by becoming the first African American to win the nation's highest office. ... Updated 11/5/2008 2:03 PM | Comment | Recommend, E-mail | Print ... in Phoenix after Barack Obama won the presidential election.
- Presidential Election of 2008 - 270toWin.com**  
www.270towin.com/2008\_Election/ 270toWin.com  
Results of the presidential election of 2008, won by Barack H. Obama with 365 electoral ... It was also the first time two sitting senators ran against each other.

...



high

high

high

high

low

low

e=Obama

STALENESS

p(vital)

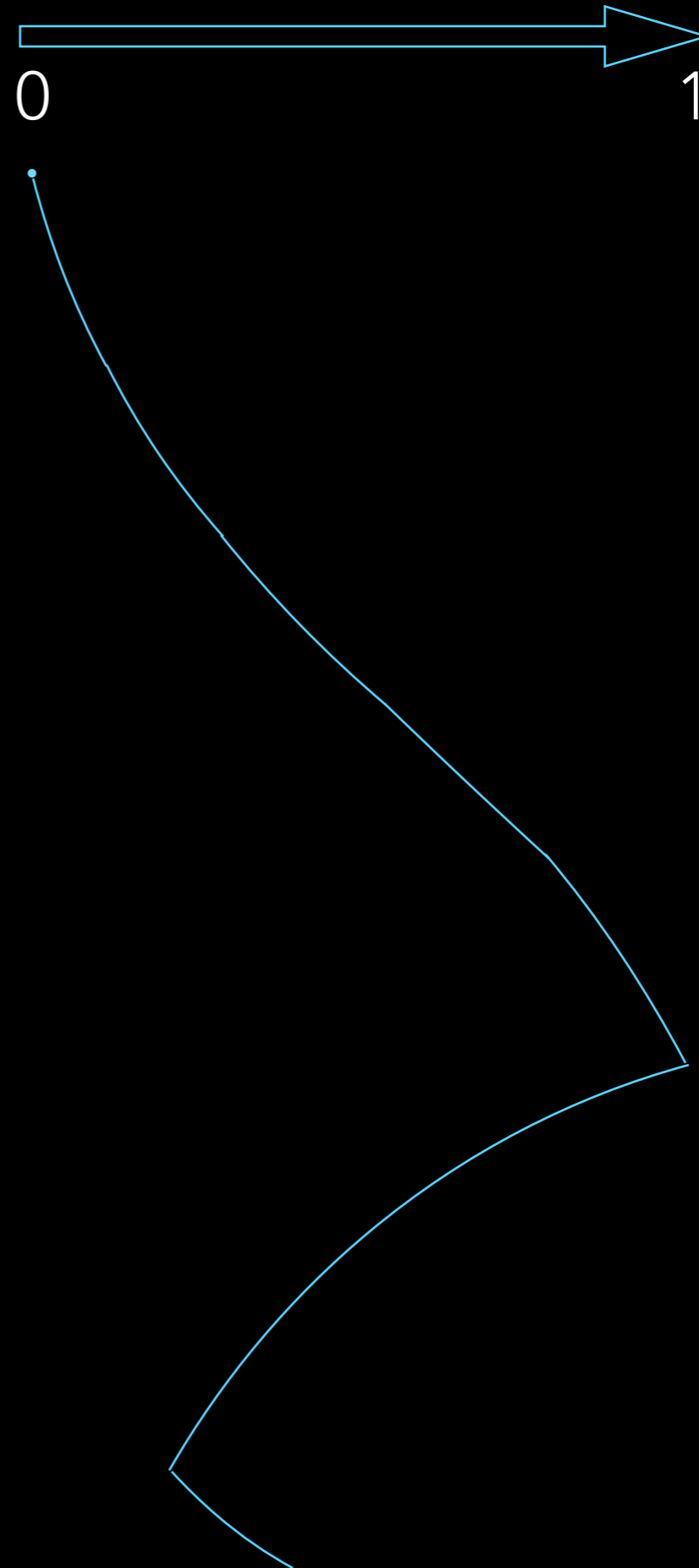
time



- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- Obama Wins Election - NYTimes.com - The New York Times**  
www.nytimes.com/2008/11/.../05campaign.html?... The New York Times  
Nov 4, 2008 - Barack Hussein Obama was elected the 44th president of the United States on Tuesday, as the country chose him as its first black chief executive. ... Monitor live results all night on The Times's Election 2008 Dashboard.
- CNN Announces Obama as the Next President Elect ...**  
 www.youtube.com/watch?v=9dKAKII1bUE  
Nov 5, 2008 - Uploaded by GettingtotheTruth2  
Lies upon lies coming from the shit hole which is obamas mouth.  
.... Election 2008 - United States Presidential Inauguration by Marc Schulz ...
- American President: Barack Obama: Campaigns and ...**  
millercenter.org/president/obama/essays/... Miller Center of Public Affairs  
Obama's election to the Senate instantly made him the highest-ranking ..... His margin of victory was down slightly from 2008, making him the first President since ...
- Obama elected nation's 44th president - USATODAY.com**  
usatoday30.usatoday.com/news/.../election2008/president.ht... USA Today  
Nov 1, 2008 - Democrat Barack Obama has been elected the 44th president of the United ... making history by becoming the first African American to win the nation's highest office. ... Updated 11/5/2008 2:03 PM | Comment | Recommend, E-mail | Print ... in Phoenix after Barack Obama won the presidential election.
- Presidential Election of 2008 - 270toWin.com**  
www.270towin.com/2008\_Election/ 270toWin.com  
Results of the presidential election of 2008, won by Barack H. Obama with 365 electoral ... It was also the first time two sitting senators ran against each other.
- ...
- Election 2012: Obama re-elected president - NY Daily News**  
www.nydailynews.com/.../election-2012-america-decides-lon... Daily News  
Nov 6, 2012 - President Barack Obama won re-election Tuesday night despite a ... that encumbered his first term and crimped the middle class dreams of ...

0

1



high

high

high

high

low

low

high

# STALENESS

- Take embedding of document
- Compare to all cluster vectors of an entity
  - If far from ALL -> document contains novel info
  - If close to ANY and LOW staleness -> VITAL
  - If close to ANY and HIGH staleness -> NON-VITAL

# Agenda

- VITAL FILTERING
- METHOD
- **EXPERIMENTS**
- ACCELERATE & CREATE
- CONCLUSION



# FEATURES

- **Baseline**
  - Document Level
    - log length and source of the document
  - Document-Entity Level
    - number and position of occurrences of entities in the document, etc.

# OUR FEATURES

- **Embedding**

- embedding of document

- **Clustering**

- min and average distances between embeddings of document and clusters

- **Temporal**

- staleness of the entity
- staleness of the assigned cluster



# MODEL ABLATION RESULTS

MODEL	P	R	<b>F1</b>
Baseline	0.475	0.238	0.317
Baseline Multi-task	0.367	0.405	0.385

# MODEL ABLATION RESULTS

MODEL	P	R	F1
Baseline	0.475	0.238	0.317
Baseline Multi-task	0.367	0.405	0.385

MODEL	P	R	F1
Embedding	0.449	0.376	0.409

# MODEL ABLATION RESULTS

MODEL	P	R	F1
Baseline	0.475	0.238	0.317
Baseline Multi-task	0.367	0.405	0.385

MODEL	P	R	F1
Embedding	0.449	0.376	0.409
Embedding-Temporal	0.444	0.358	0.397

# MODEL ABLATION RESULTS

MODEL	P	R	F1
Baseline	0.475	0.238	0.317
Baseline Multi-task	0.367	0.405	0.385

MODEL	P	R	F1
Embedding	0.449	0.376	0.409
Embedding-Temporal	0.444	0.358	0.397
Embedding-Clustering	0.439	0.372	0.403

# MODEL ABLATION RESULTS

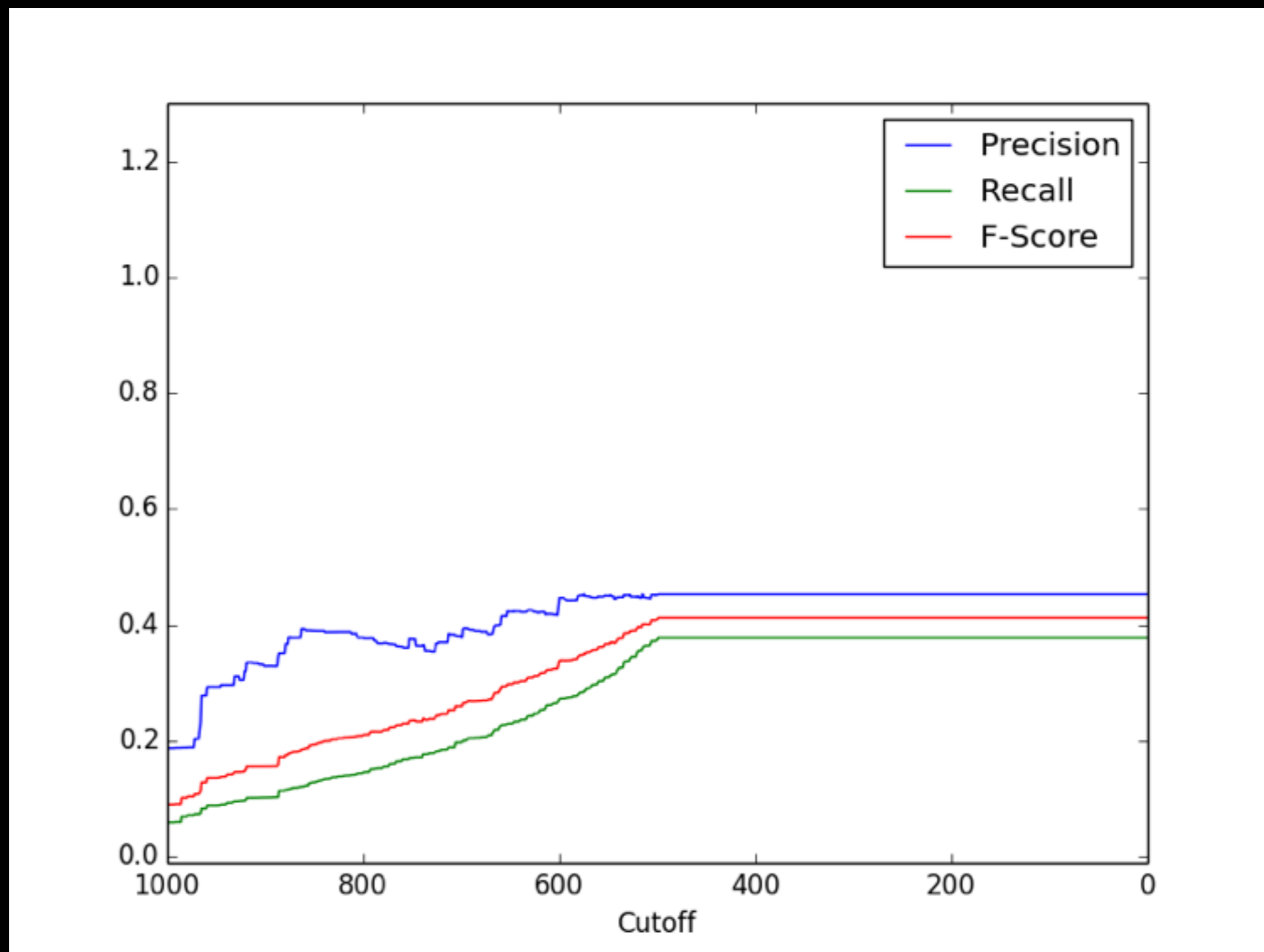
MODEL	P	R	F1
Baseline	0.475	0.238	0.317
Baseline Multi-task	0.367	0.405	0.385

MODEL	P	R	F1
Embedding	0.449	0.376	0.409
Embedding-Temporal	0.444	0.358	0.397
Embedding-Clustering	0.439	0.372	0.403
Embedding-Temporal-Clustering	0.452	0.378	<b>0.412</b>

# OFFICIAL VITAL RESULTS

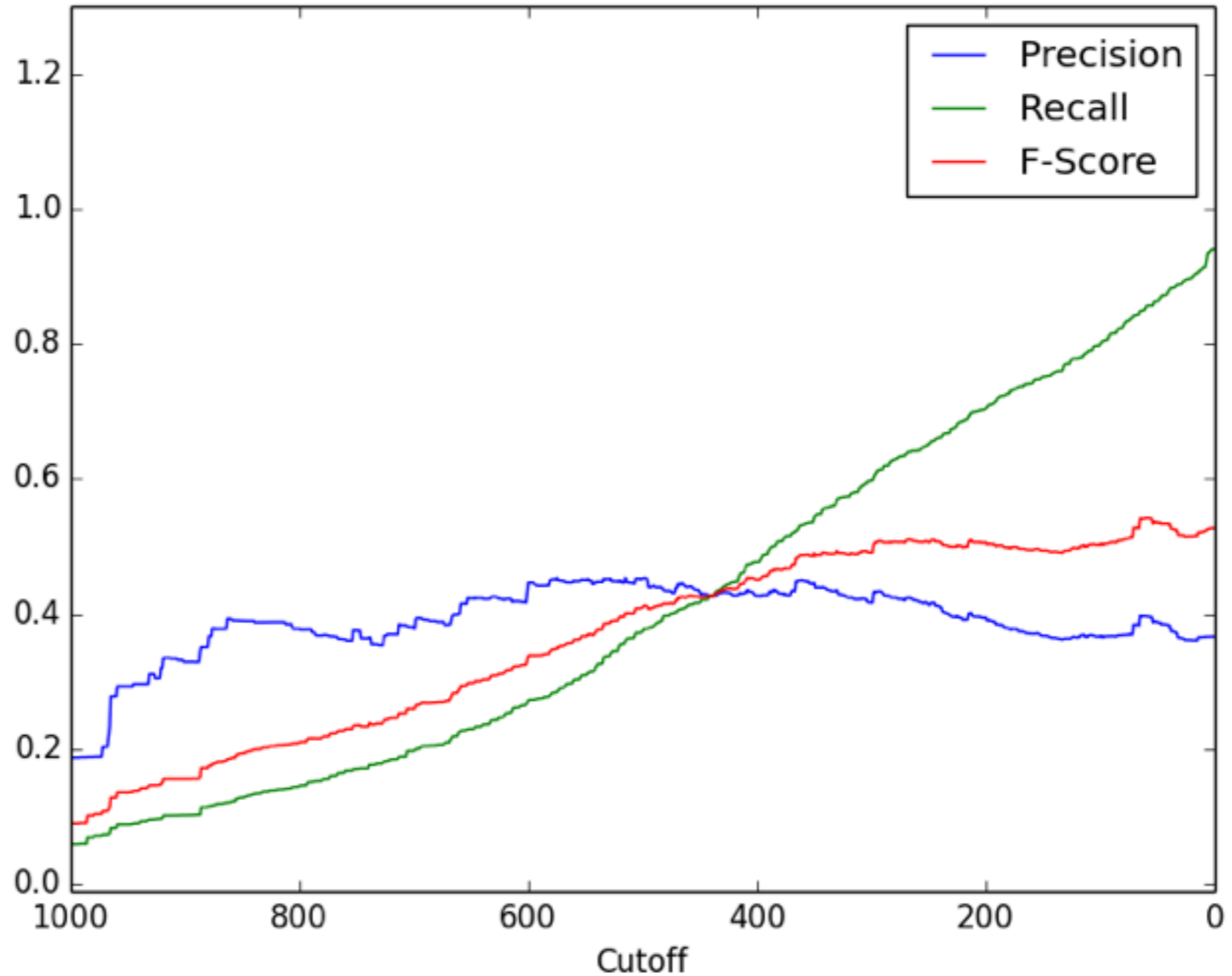
P = 2nd

F1 = 8th



# UNOFFICIAL VITAL RESULTS

F1 = 2nd



# Agenda

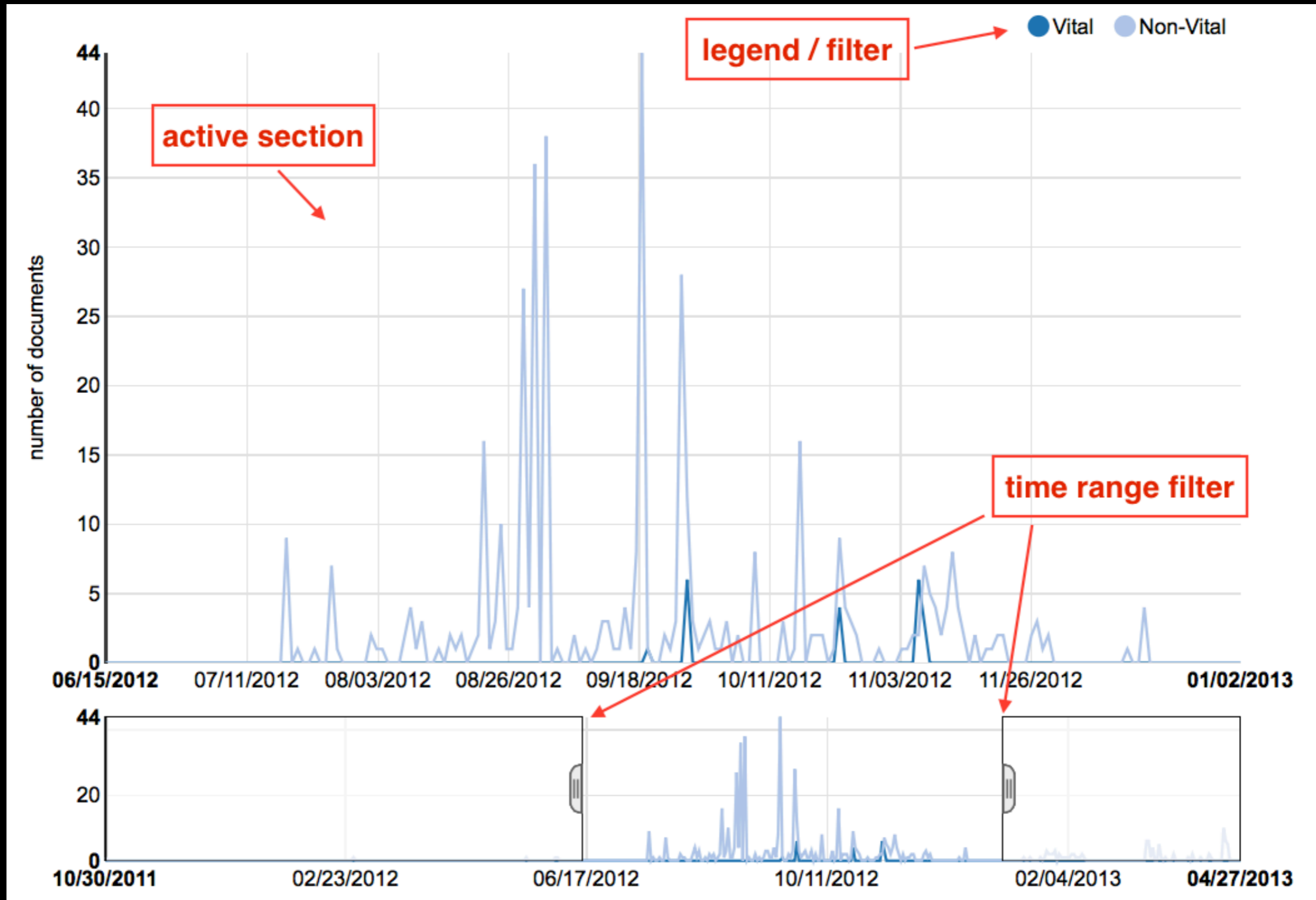
- VITAL FILTERING
- METHOD
- EXPERIMENTS
- **ACCELERATE & CREATE**
- CONCLUSION & FUTURE WORK



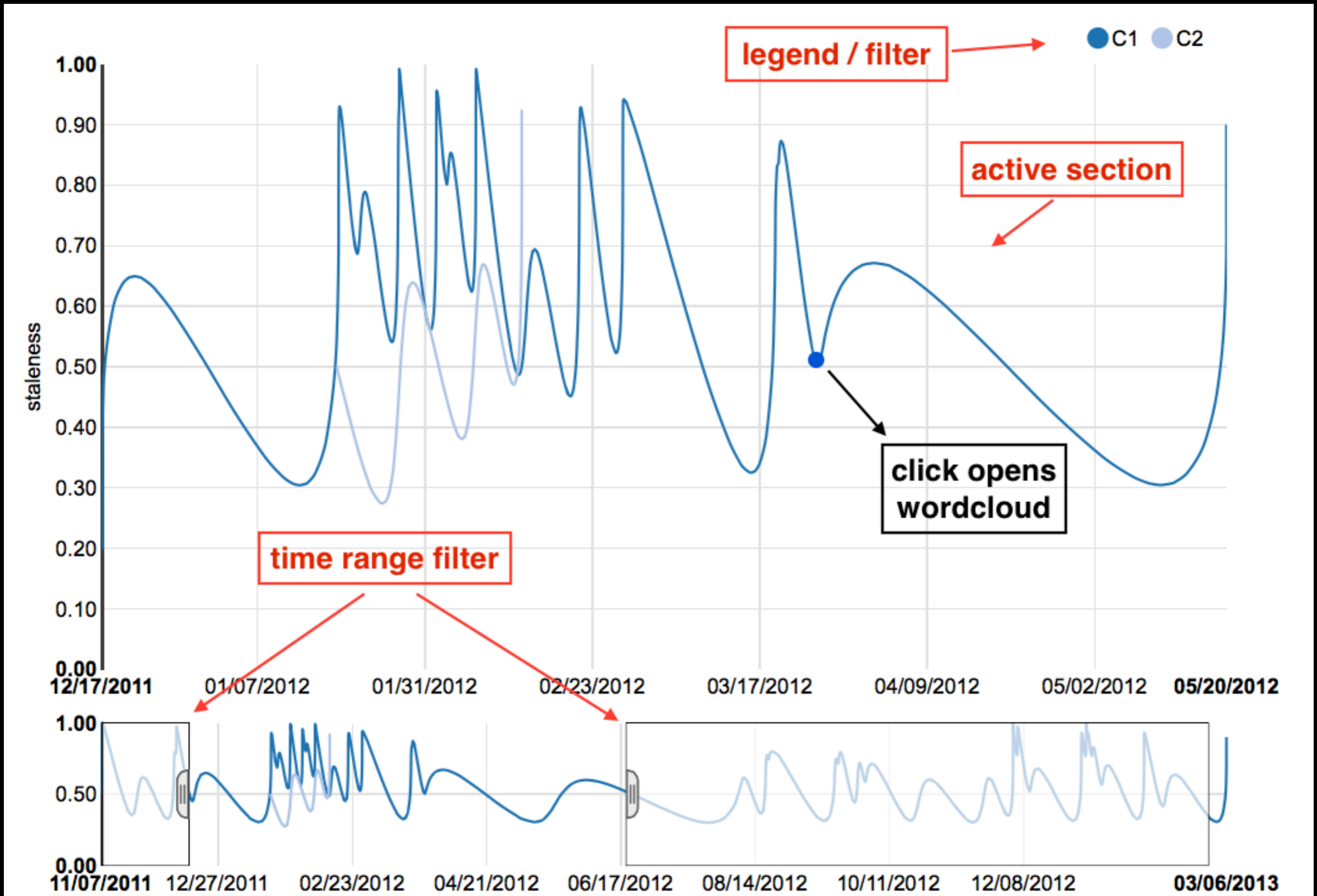
# VISUALIZATION

- Browser-based visualization prototype with interactive widgets
- Mainly for debugging purposes

# DOCUMENT VIEW



# CLUSTER VIEW



# Agenda

- VITAL FILTERING
- METHOD
- EXPERIMENTS
- ACCELERATE & CREATE
- **CONCLUSION**

# CONCLUSION

- **Semi-supervised** model for document filtering tasks
  - Low dimensional **vector based semantics** is useful for VF
  - Non-parametric clustering can capture **multiple aspects of entities**
  - Simple decay based **temporal dynamics** indicate novelty
- Official results show we are **2nd** in Precision and **3rd** in SU
  - Follow up analysis suggests we are **2nd** in F1 as well

THANK\_YOU

KUDOS

THANK

WELL\_DONE

**THANKS**

THANKS\_TO

GOOD\_LUCK

BOUQUETS

APPRECIATED

GRATEFUL