

# Gold standard for Enron Organizational Hierarchy

**Apoorv Agarwal**

July 9th 2012

Columbia University

in collaboration with

Adinoyi Omuya (Wisdom), Aaron Harnly, Owen Rambow

# Enron Email Corpus

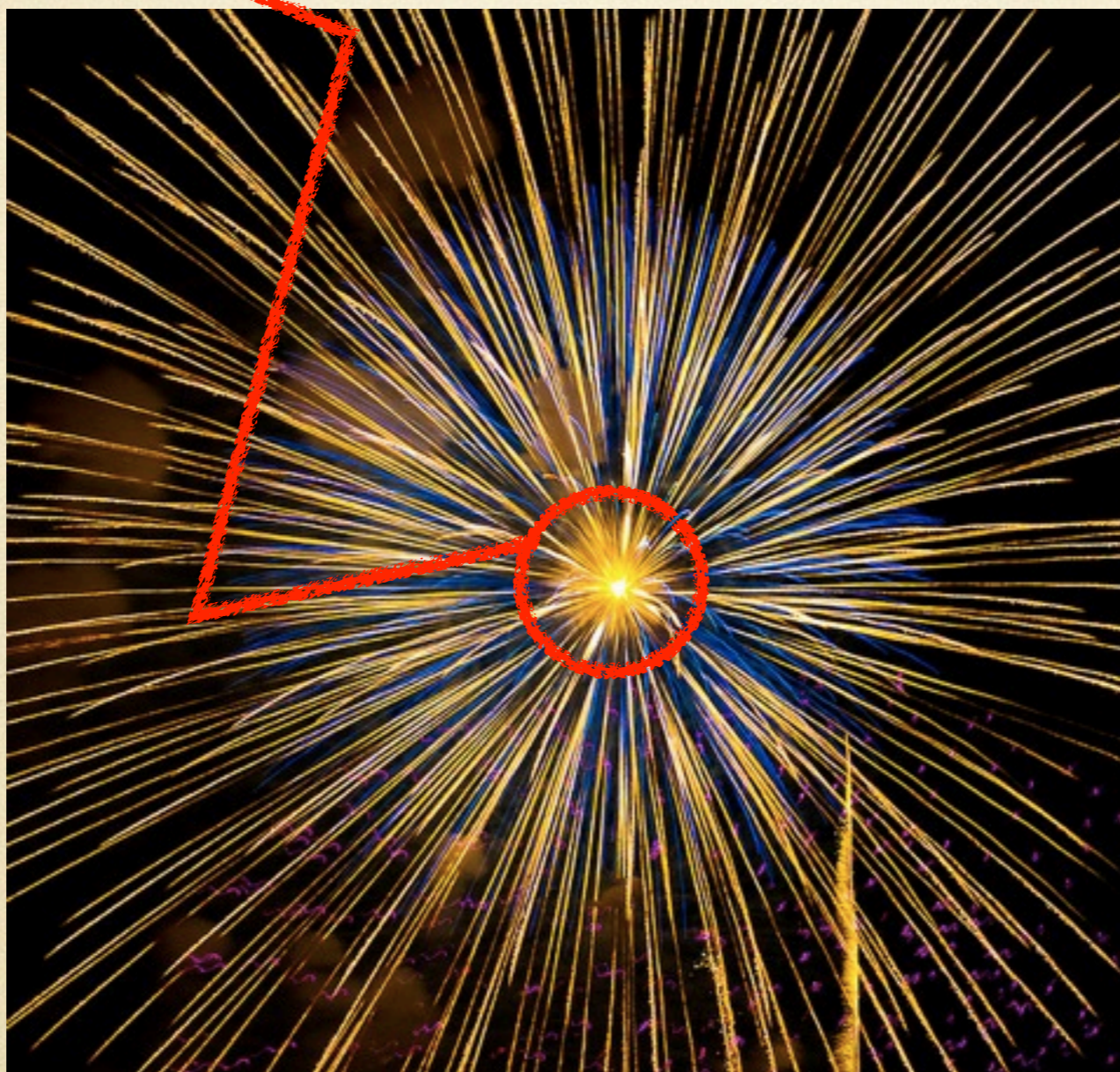
Klimt & Yang '04



# Enron Email Corpus

Klimt & Yang '04

158 "core" people

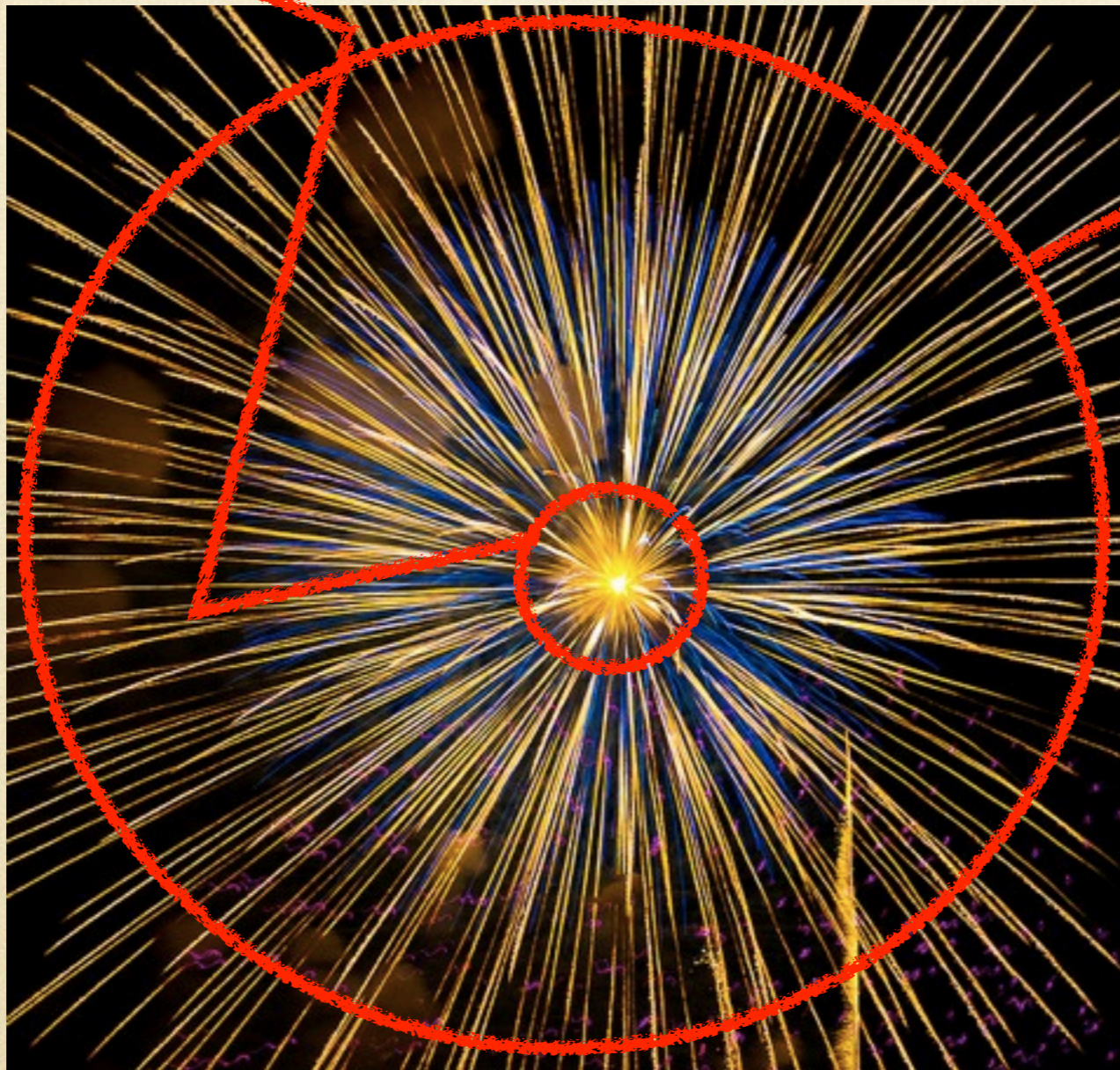


# Enron Email Corpus

Klimt & Yang '04

158 "core" people

90K "non-core"  
people



# Task: Hierarchy Prediction

Given the Enron email corpus, predict the organizational hierarchy of Enron employees.

# Task: Hierarchy Prediction

Given the Enron email corpus, predict the organizational hierarchy of Enron employees.

SNA



Rowe et al. '07



# Task: Hierarchy Prediction

Given the Enron email corpus, predict the organizational hierarchy of Enron employees.

SNA



Rowe et al. '07



No quant.  
evaluation!

# Task: Hierarchy Prediction

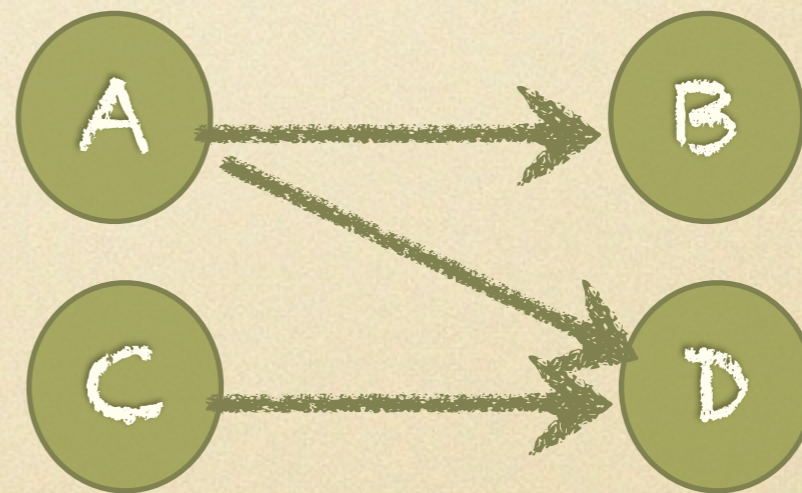
Given the Enron email corpus, predict the organizational hierarchy of Enron employees.

SNA

NLP

Rowe et al. '07

Bramsen et al. '11  
Gilbert '12



No quant.  
evaluation!



# Task: Hierarchy Prediction

Given the Enron email corpus, predict the organizational hierarchy of Enron employees.

SNA



Rowe et al. '07

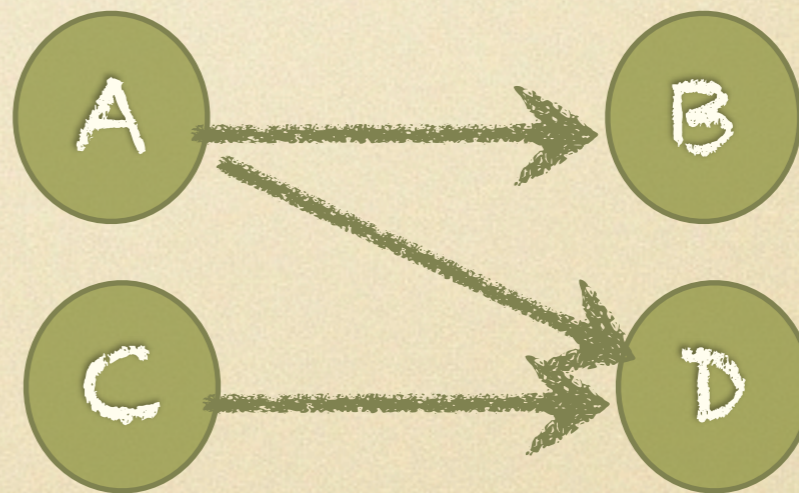


No quant. evaluation!

NLP



Bramsden et al. '11  
Gilbert '12



Dominance prediction not hierarchy!

# Our Contributions

# Our Contributions

1. Present a new Gold Standard for hierarchy prediction

# Our Contributions

1. Present a new Gold Standard for hierarchy prediction
2. Present a baseline social network analysis (SNA) based system that out-performs a recent NLP based system by Gilbert '12

# Gold Standard creation

# Gold Standard creation

- search emails for "org-chart"

# Gold Standard creation

- search emails for "org-chart"
- manually convert excel or visio documents into MongoDB

# Gold Standard creation

- search emails for "org-chart"
- manually convert excel or visio documents into MongoDB

Hierarchy not just dominance!



# Gold Standard creation

- search emails for "org-chart"
- manually convert excel or visio documents into MongoDB

Hierarchy not just dominance!

- 1518 employees (previous: 158 emp.)

# Gold Standard creation

- search emails for "org-chart"
- manually convert excel or visio documents into MongoDB

Hierarchy not just dominance!



- 1518 employees (previous: 158 emp.)
- "core" and "non-core" employees (previous: only "core")

# Compare ...

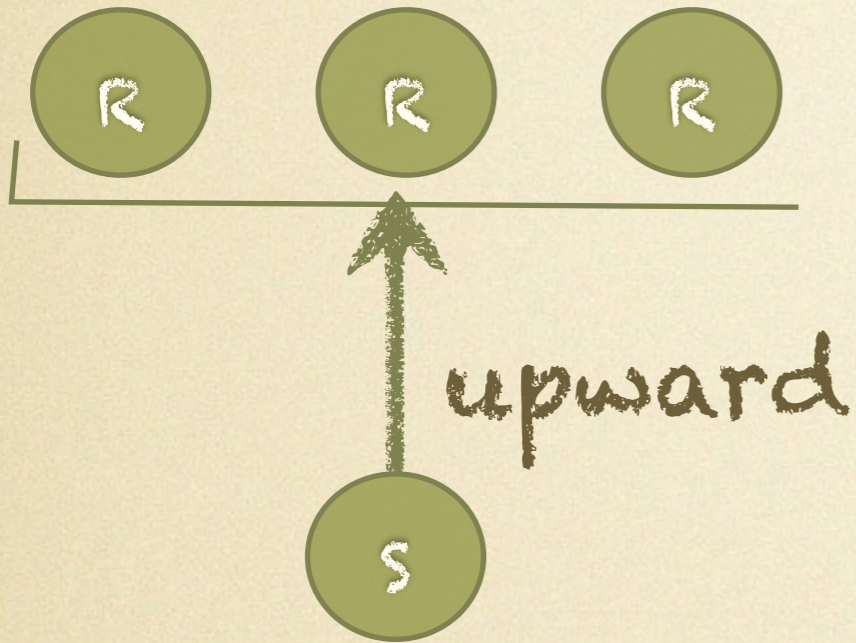
- Gilbert '12 : NLP based approach for dominance prediction
- Our SNA based baseline

Gilbert '12

Emails of 132 employees

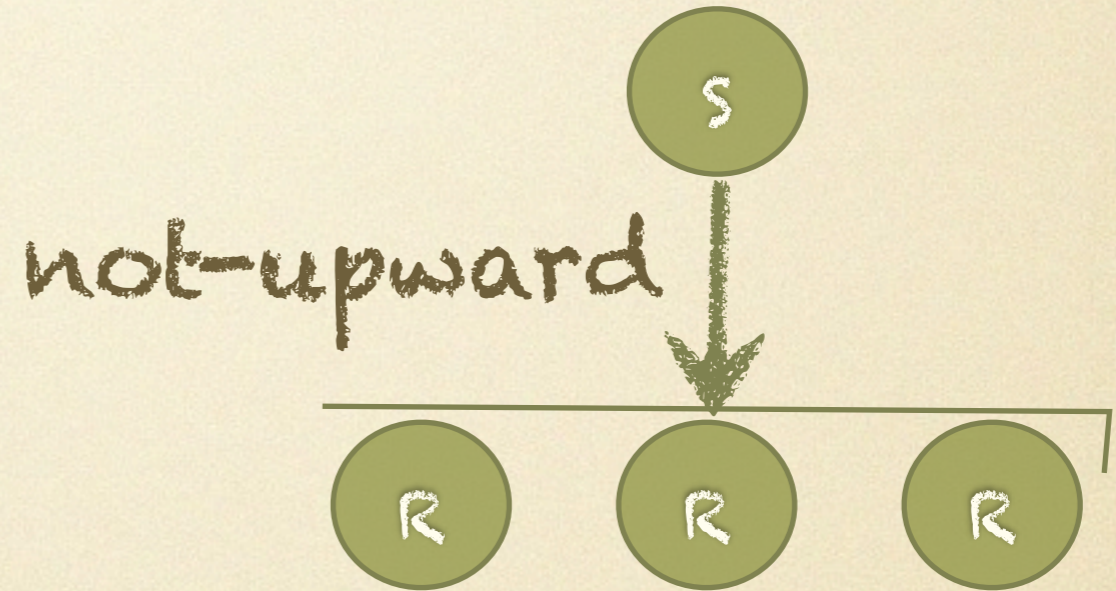
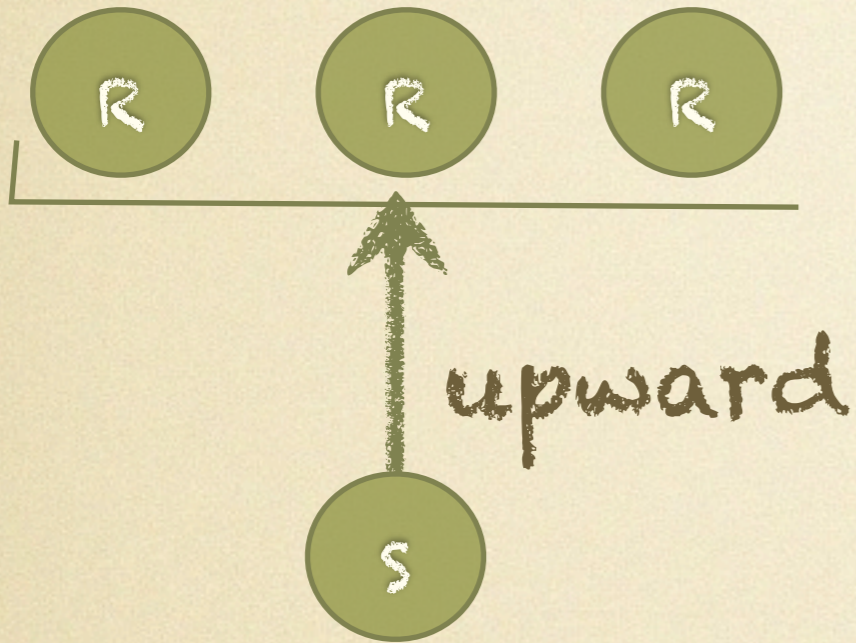
Gilbert '12

Emails of 132 employees



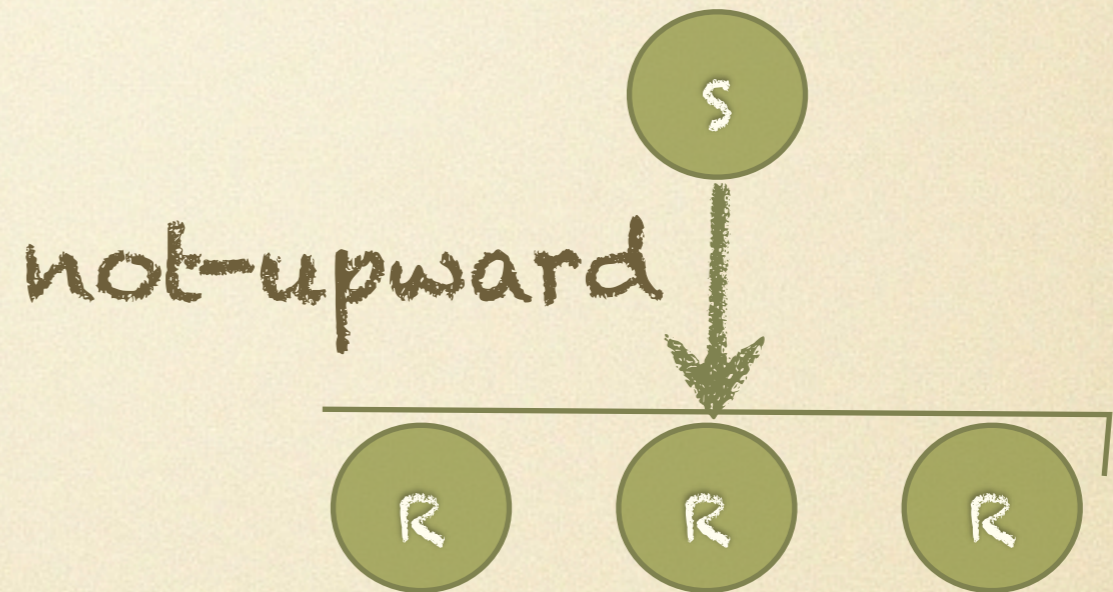
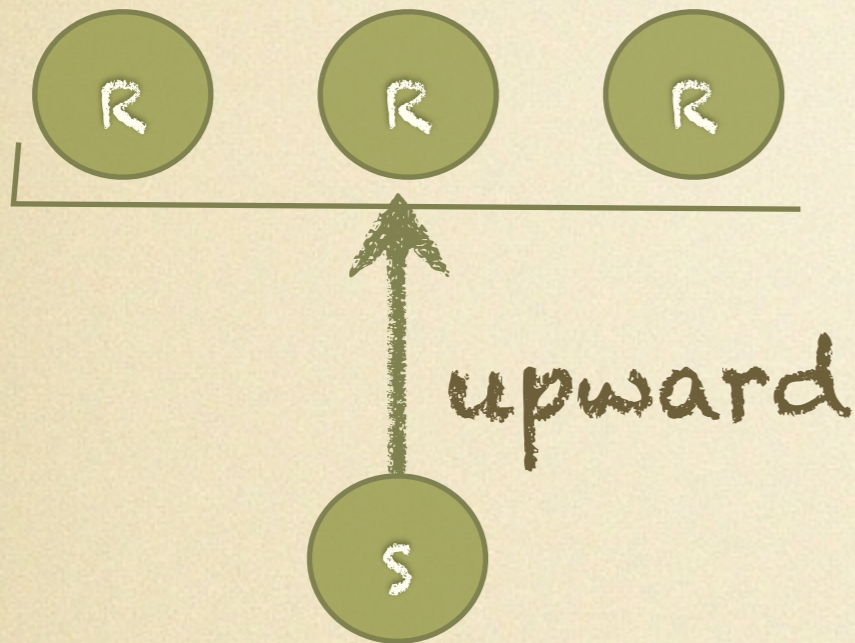
# Gilbert '12

Emails of 132 employees



# Gilbert '12

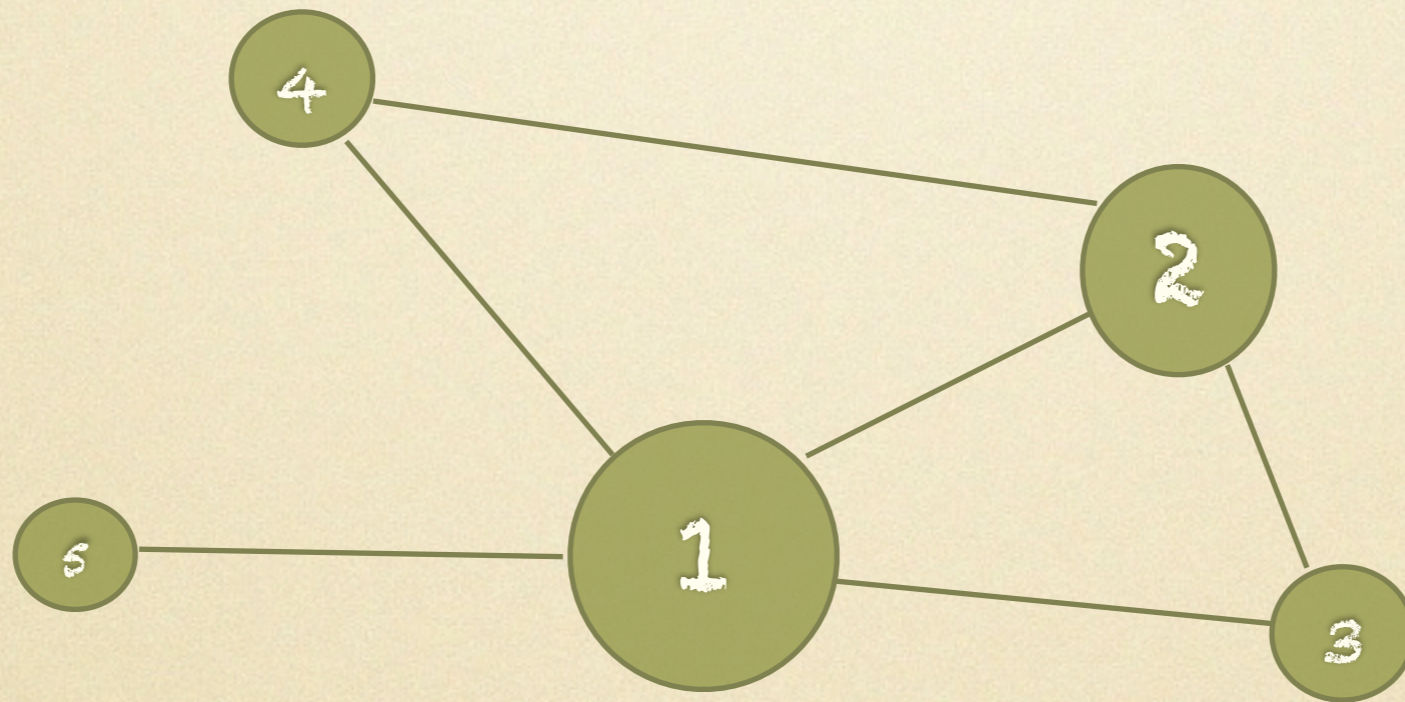
Emails of 132 employees



- Feature space: N-gram + feature selection
- SVM with 3-fold cross-validation

# Our SNA approach

- Sort based on degree centrality



$$1 > 2 > \{3 = 4\} > 5$$



# Evaluation and Results

e11	e12
e21	e22
...	...
em1	em2
....	...
en1	en2

Gold Standard

# Evaluation and Results

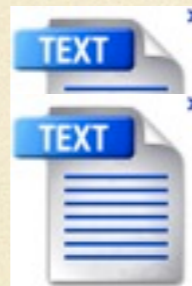
e11	e12
e21	e22
...	...
em1	em2
...	...
en1	en2

Our SNA approach

Gilbert '12

Gold Standard

# Evaluation and Results



e11	e12
e21	e22
...	...
em1	em2
...	...
en1	en2

Our SNA approach

Gilbert '12

Gold Standard

# Evaluation and Results

2640 (19%)



e11	e12
e21	e22
...	...
em1	em2
...	...
en1	en2

Our SNA approach

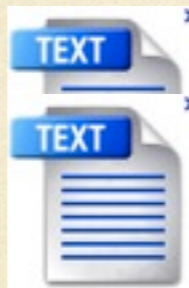
Gilbert '12

Gold Standard

# Evaluation and Results

2640 (19%)

82.37%



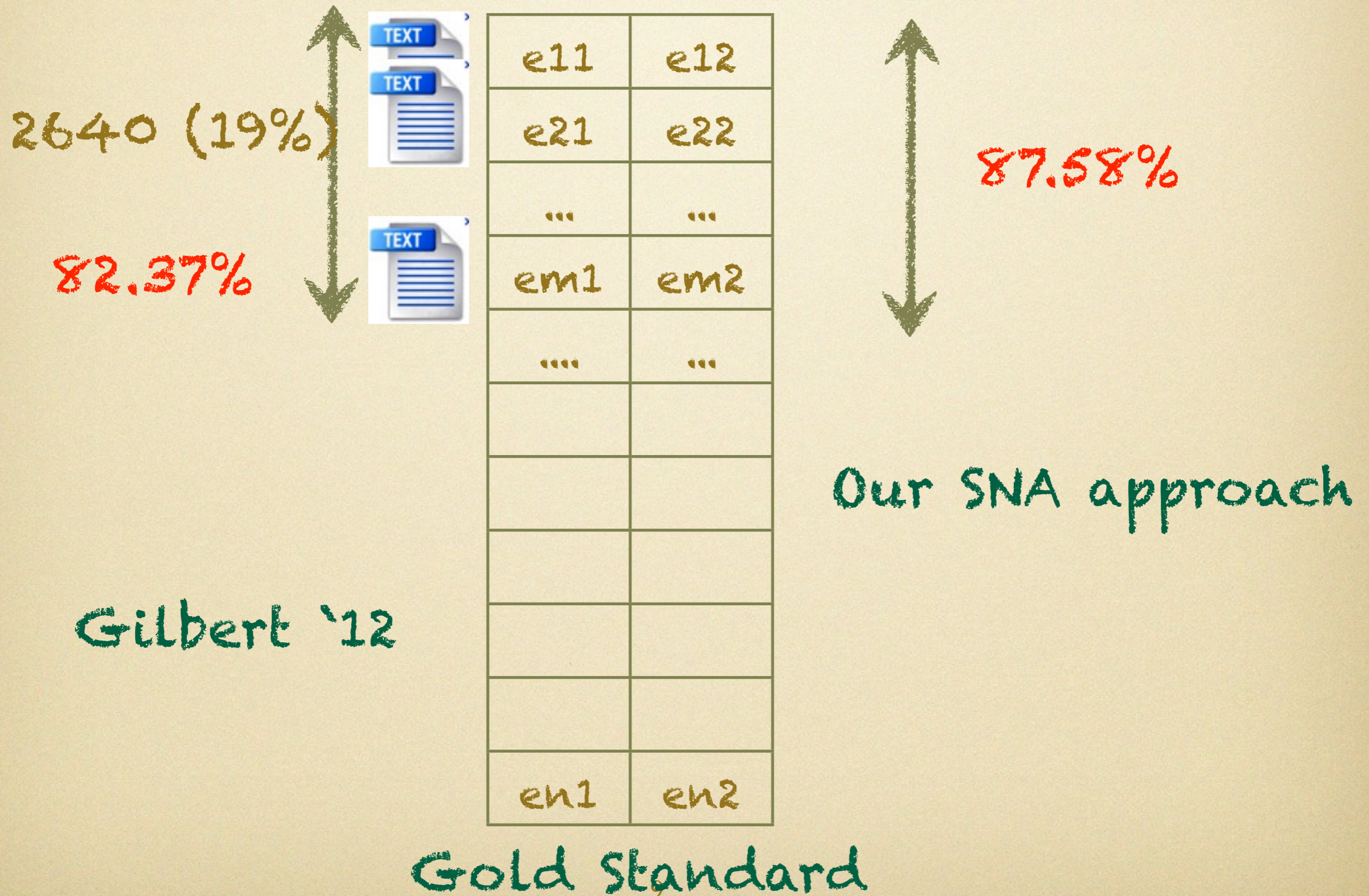
e11	e12
e21	e22
...	...
em1	em2
....	...
en1	en2

Our SNA approach

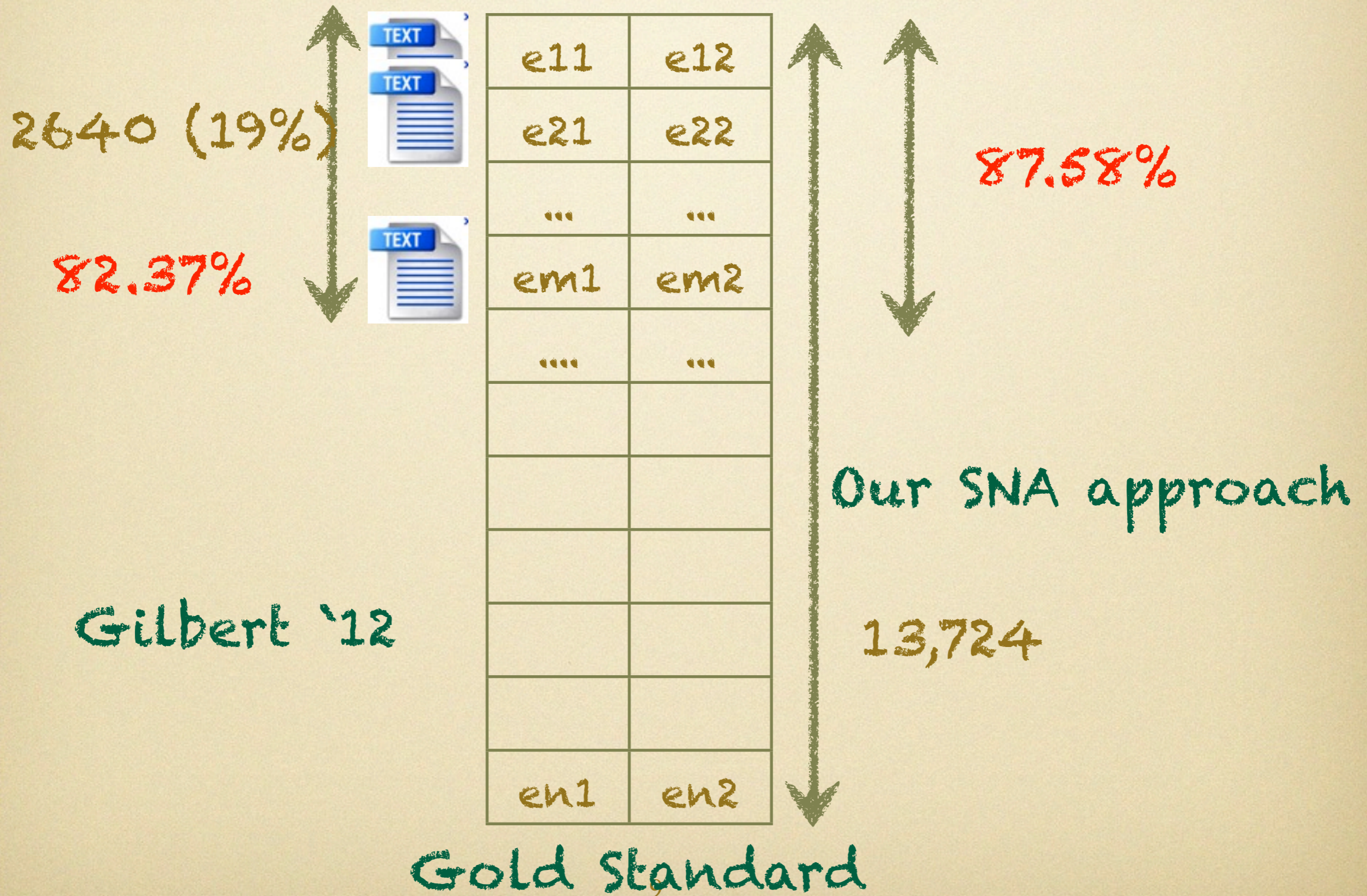
Gilbert '12

Gold Standard

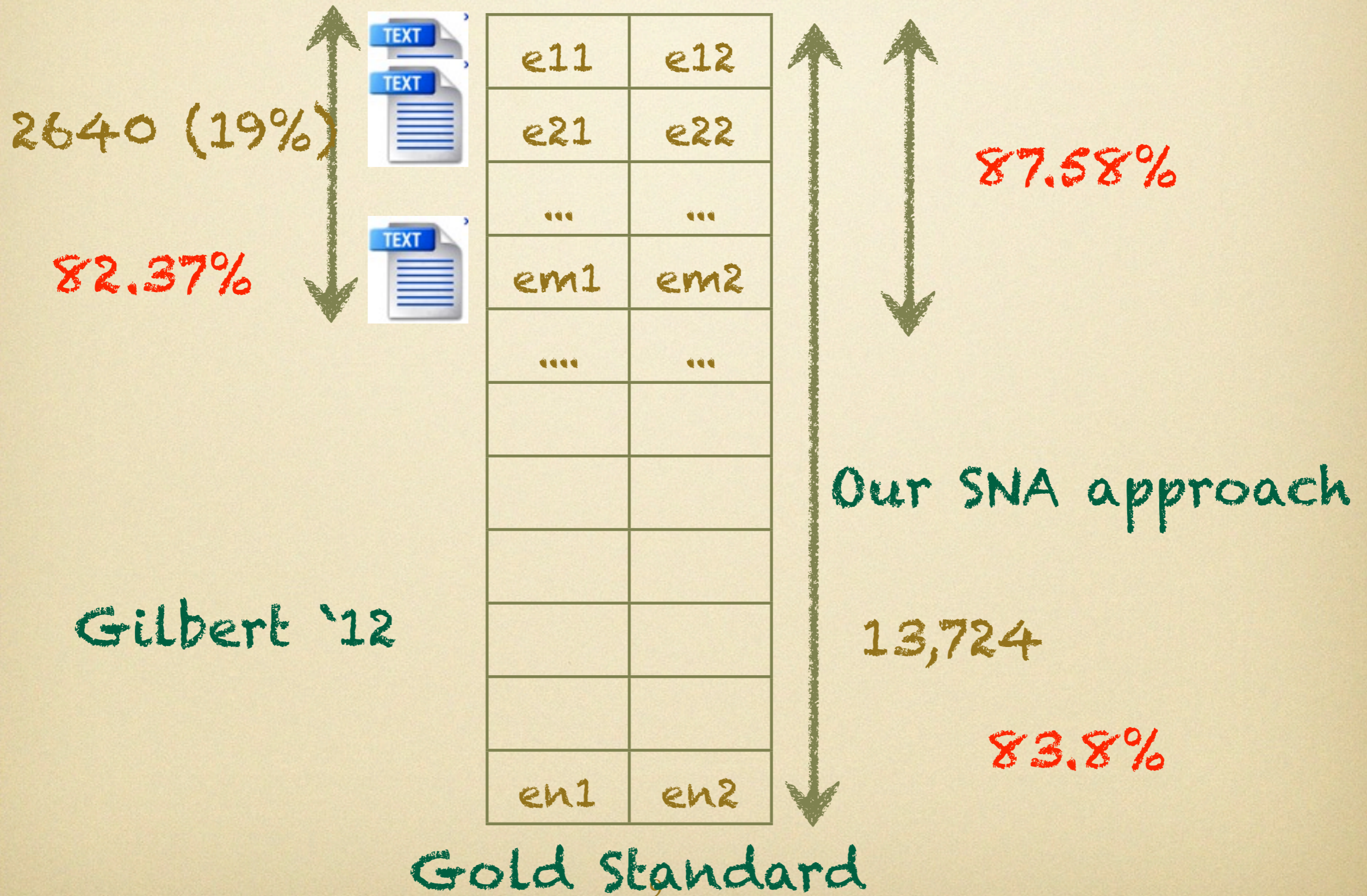
# Evaluation and Results



# Evaluation and Results



# Evaluation and Results





# Conclusion

# Conclusion

- Introduce a large gold standard for hierarchy prediction of Enron employees (contact [apoorv@cs.columbia.edu](mailto:apoorv@cs.columbia.edu))

# Conclusion

- Introduce a large gold standard for hierarchy prediction of Enron employees (contact [apoorv@cs.columbia.edu](mailto:apoorv@cs.columbia.edu))
- Show for our gold standard, a simple SNA based approach outperforms a recent NLP based approach

# Conclusion

- Introduce a large gold standard for hierarchy prediction of Enron employees (contact [apoorv@cs.columbia.edu](mailto:apoorv@cs.columbia.edu))
- Show for our gold standard, a simple SNA based approach outperforms a recent NLP based approach

Thanks!

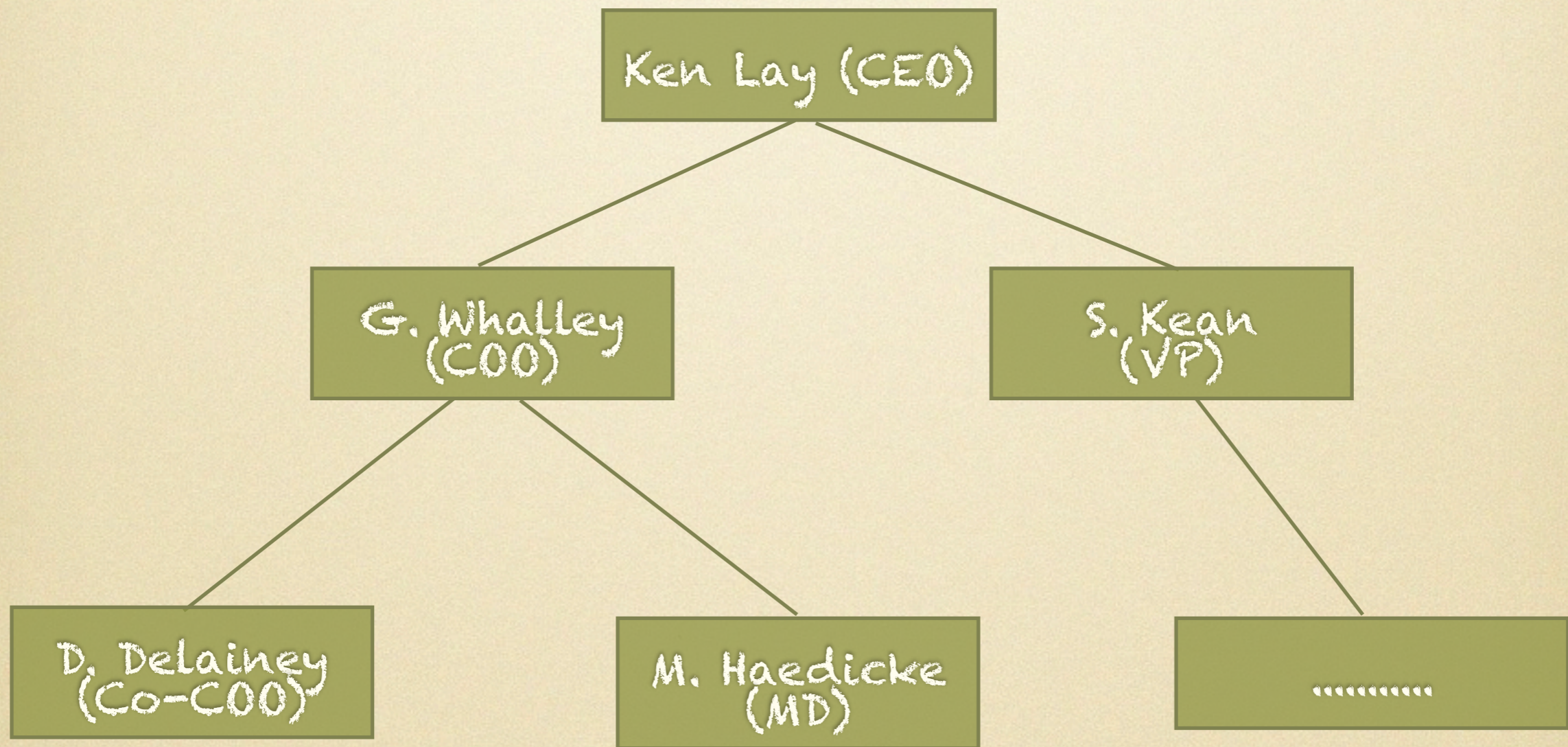
# UB-NLP

- $\langle A, B \rangle$  in Gold: 13,724 (G)
- $\langle A, B \rangle$  where A emails B: 2604 (T)
- UB-NLP:  $(2604 + 5560) / 13724 = 59.6\%$

# Task: Hierarchy Prediction

Rowe et al. '07

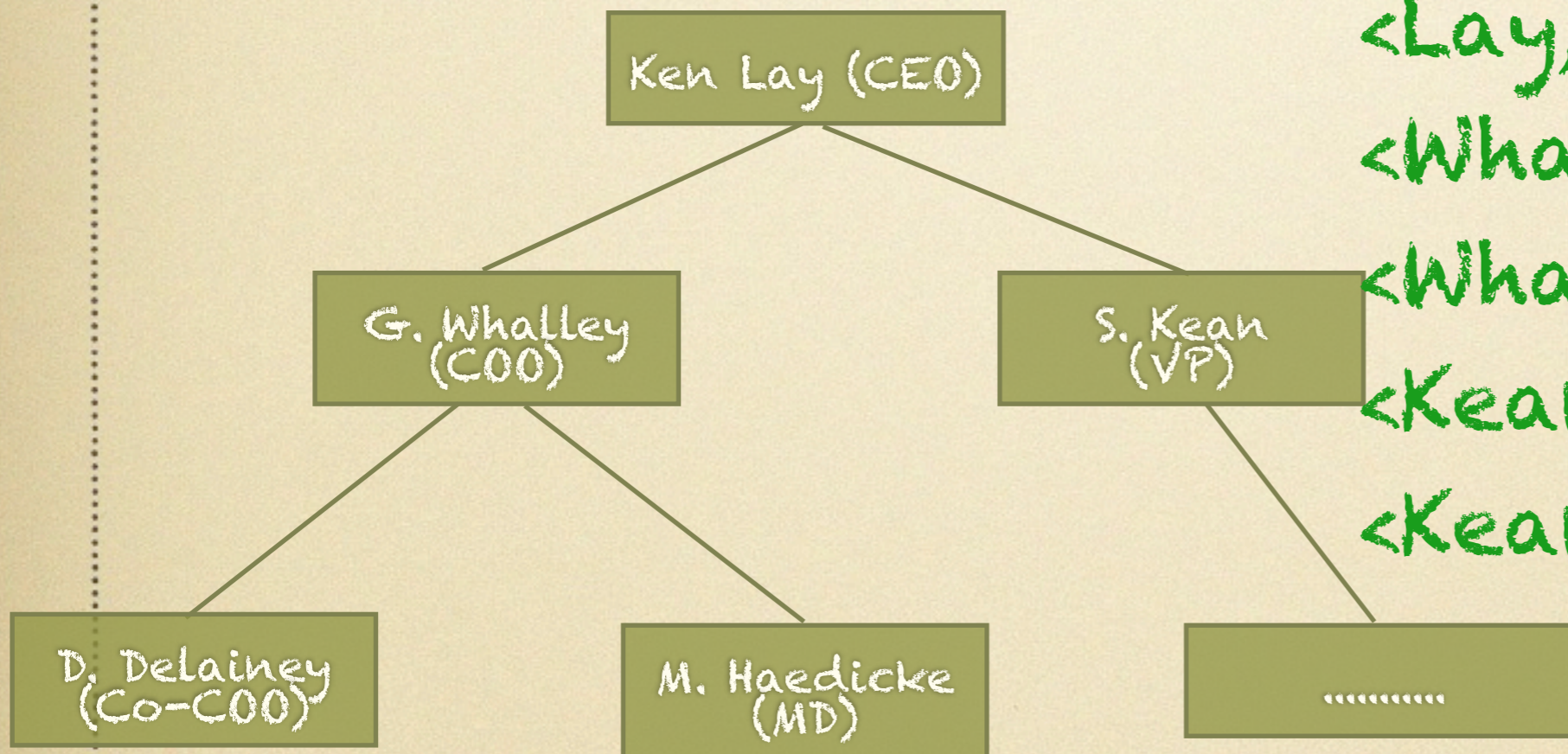
Bramsén et al. '11  
Gilbert '12



# Task: Hierarchy Prediction

Rowe et al. '07

Bramsen et al. '11  
Gilbert '12

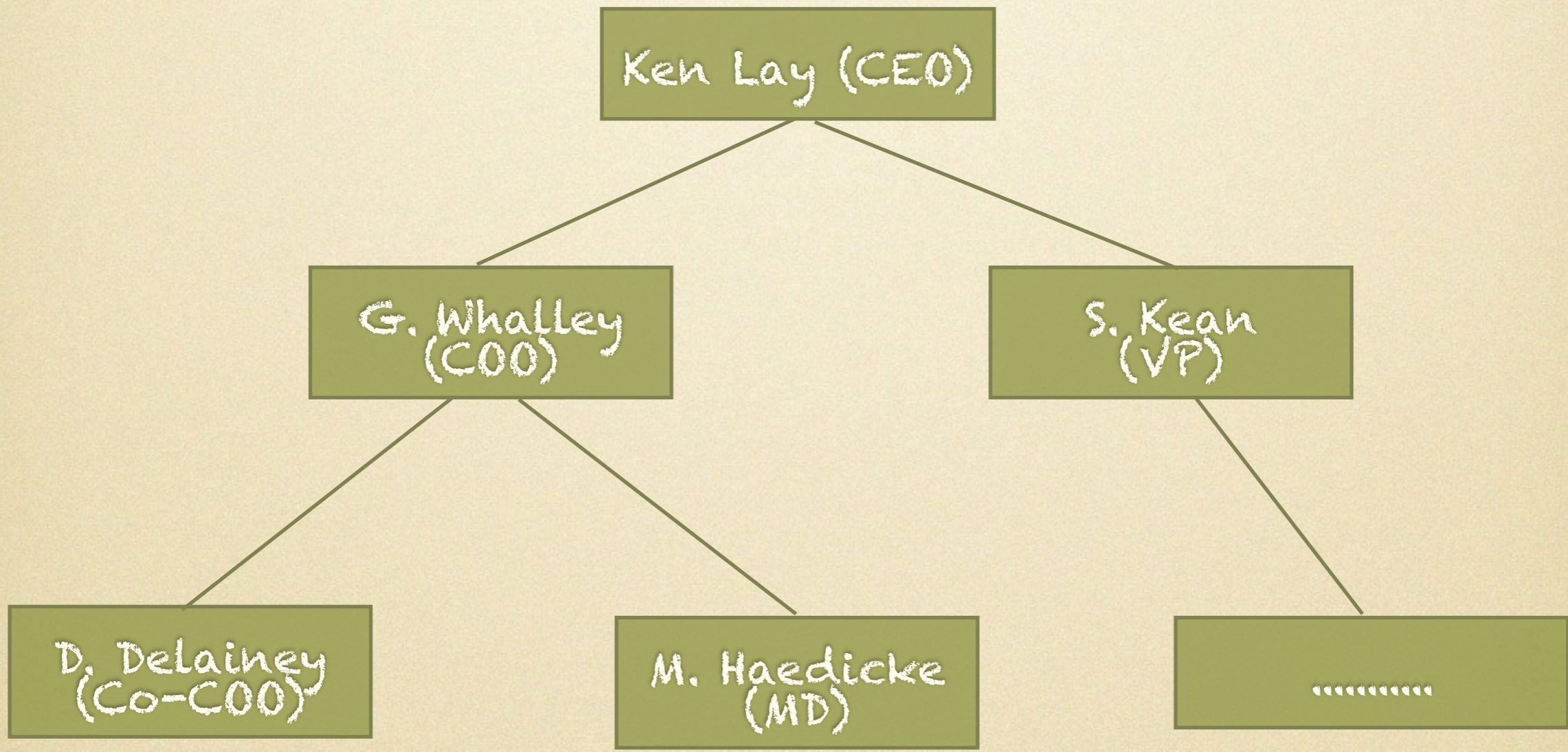


<Lay, Whalley>  
<Whalley, Delainey>  
<Whalley, Haedicke>  
<Kean, Delainey>  
<Kean, Haedicke>

# Task: Hierarchy Prediction

Rowe et al. '07

Bramsen et al. '11  
Gilbert '12





# Our Contributions

## New Gold Standard

- 1518 employees
- 13K dominance pairs ( $\langle A, B \rangle$  tuples):  
 $\langle \text{Lay, Whalley} \rangle$ ;  $\langle \text{Whalley, Delainey} \rangle$ ;  
 $\langle \text{Lay, Kean} \rangle$  ...
- both "core" and "non-core"

# Our Contributions

## New Gold Standard

- 1518 employees
- 13K dominance pairs ( $\langle A, B \rangle$  tuples):  
 $\langle \text{Lay, Whalley} \rangle$ ;  $\langle \text{Whalley, Delainey} \rangle$ ;  
 $\langle \text{Lay, Kean} \rangle$  ...
- both "core" and "non-core"

## SNA system out-performs NLP system

- Current state-of-the-art NLP based system (Gilbert '12)
- Social network analysis based system (SNA)