

ThemeDelta: Dynamic Segmentations over Temporal Topic Models

Paper By: Samah Gad, Waqas Javed, Sohaib Ghani, Niklas Elmqvist, Tom Ewing, Keith N. Hampton, and Naren Ramakrishnan

Published: IEEE Transactions on Visualization and Computer Graphics 21(5) 2015

Presentation By: Yasha Pushak



a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

What: Text Dataset with Timestamps

“I can’t believe I have so many racist friends...”

“These are the faces of the Syrian refugees. Men, women, and children who’s homes were destroyed and were forced to flee.”

“Canada extends it’s condolences to France.”

“Shouldn’t we help our homeless before refugees?”

“Terror in Paris...”

“Diversity is our strength. We strongly condemn the acts aimed at certain Canadians after the Paris attacks.”

“Canada is full. Say No to terrorists.”

“My great grand parents immigrated from Russia to escape violence... No one told them ‘We’re full.’”

“Canada stands with Paris.”

Why: Identify Scatter/Gather Relationships

"These are the faces of the Syrian refugees.
Men, women, and children who's homes
were destroyed and were forced to flee."

"I can't believe I have so
many racist friends..."

"My great grand parents immigrated
from Russia to escape violence...
No one told them 'We're full'."

"Diversity is our strength. We strongly
condemn the acts aimed at certain
Canadians after the Paris attacks."

"Canada extends it's
condolences to France."

"Terror in Paris..."

"Shouldn't we help our
homeless before refugees?"

"Canada is full. Say
No to terrorists."

"Canada stands with Paris."

Processing the bags of words

Latent Dirichlet Allocation (LDA)

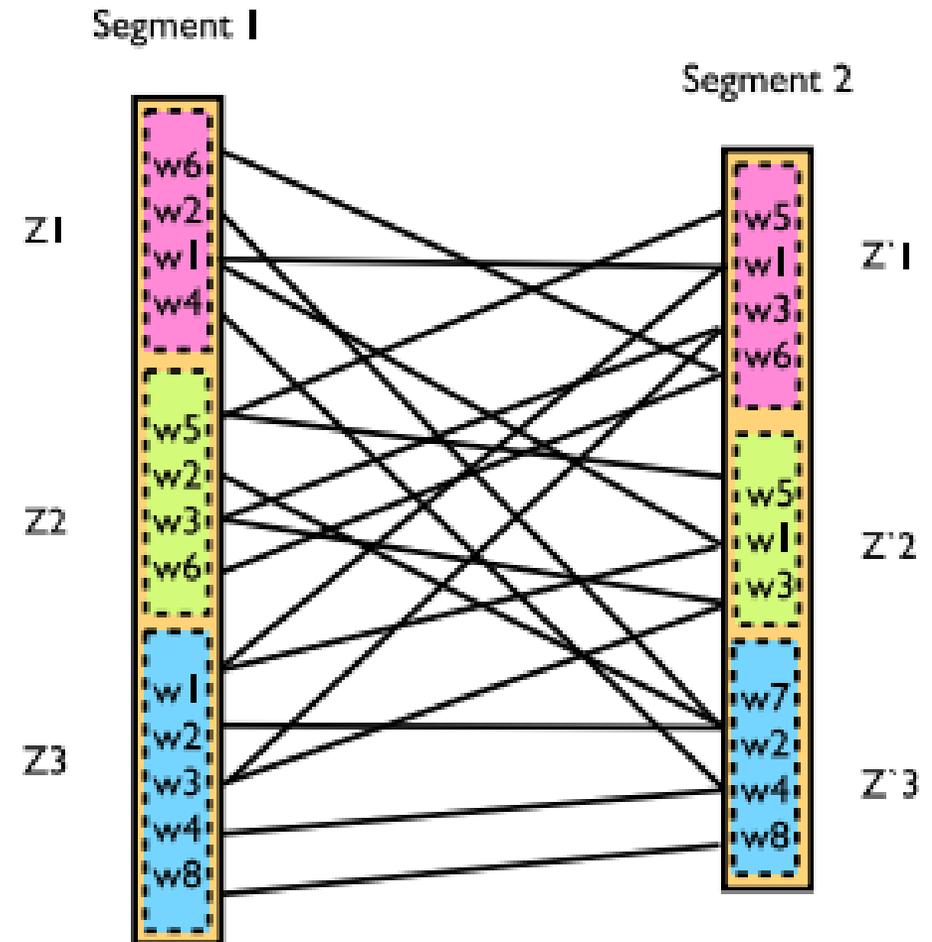
Input: Bag of Words over time

Output: Topics (Groups of keywords at a specific point in time)

Timeline Segmentation

Input: Topics

Output: Optimal time intervals containing groups of topics



What? Why? How?

What: Data

- Timestamped text dataset

What: Derived

- Bag of Words over time

- Topics (Groups of keywords at a specific point in time)

- Time intervals containing groups of topics

Why: Tasks

- Identify changes in topics over time

- Identify scatter/gather relationships

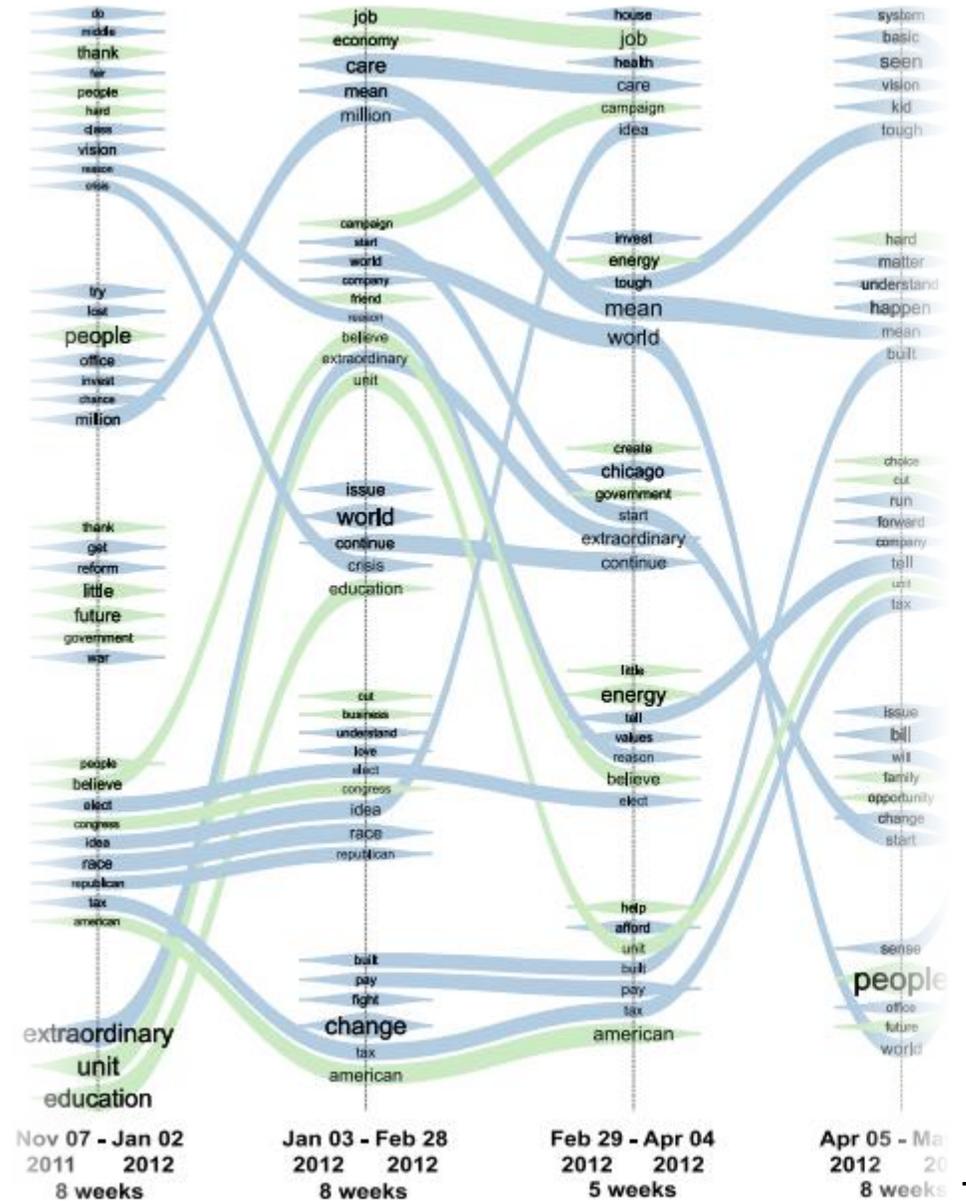
What? Why? How?

How: Encode

- Parallel axes for time segments
- Spatially partition topics along a segment
- Label keywords within topics
- Linked keywords across time intervals
- Segment labels for dates and duration

How: Encode (Free Channels)

- Size of labels for quantitative data
- Width of links for quantitative data
- Link colour for categorical or ordered data



What? Why? How?

How: Encode (Free Channels)

Size of labels for quantitative data

Width of links for quantitative data

Link colour for categorical or ordered data

How: Manipulate

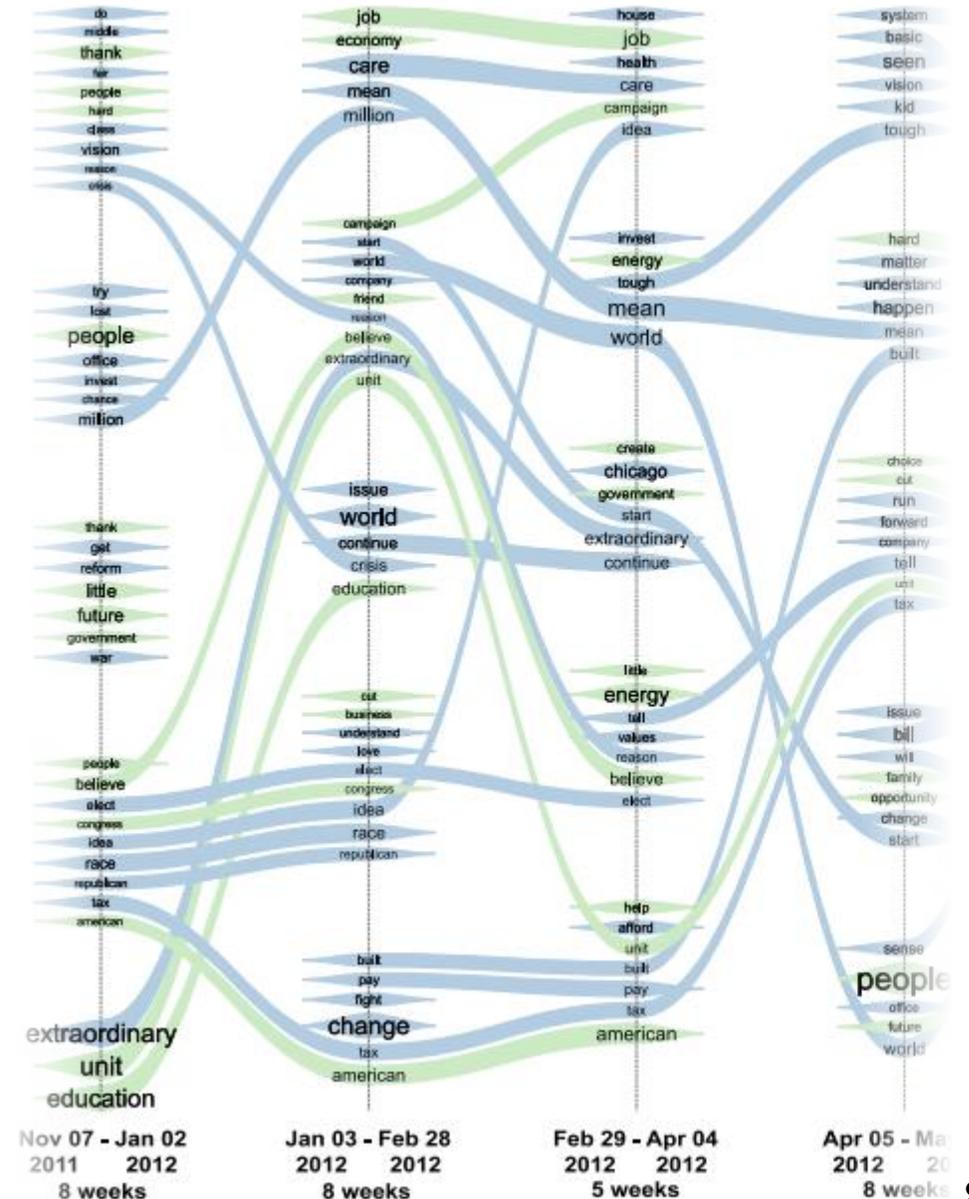
Navigate: geometric zooming and panning

Select: highlight keywords

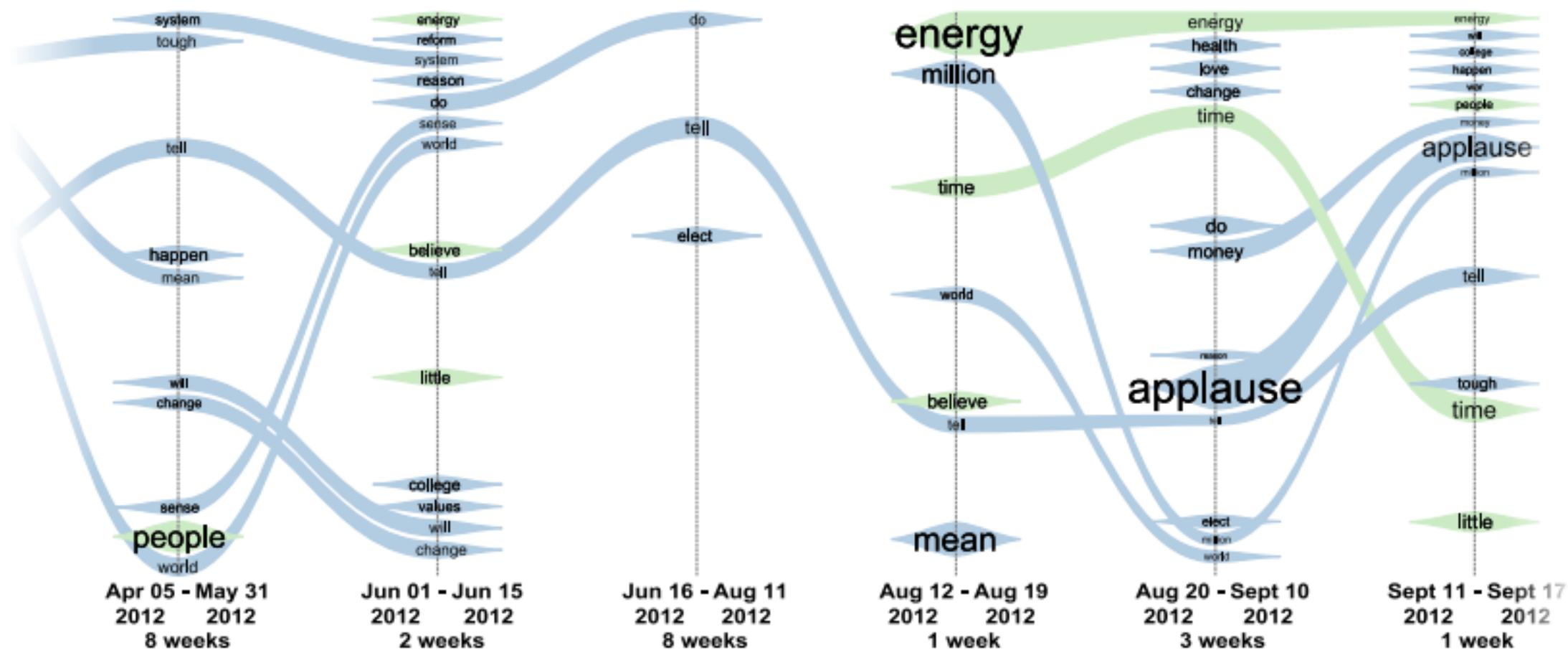
Search: Select keywords by searching

How: Reduce

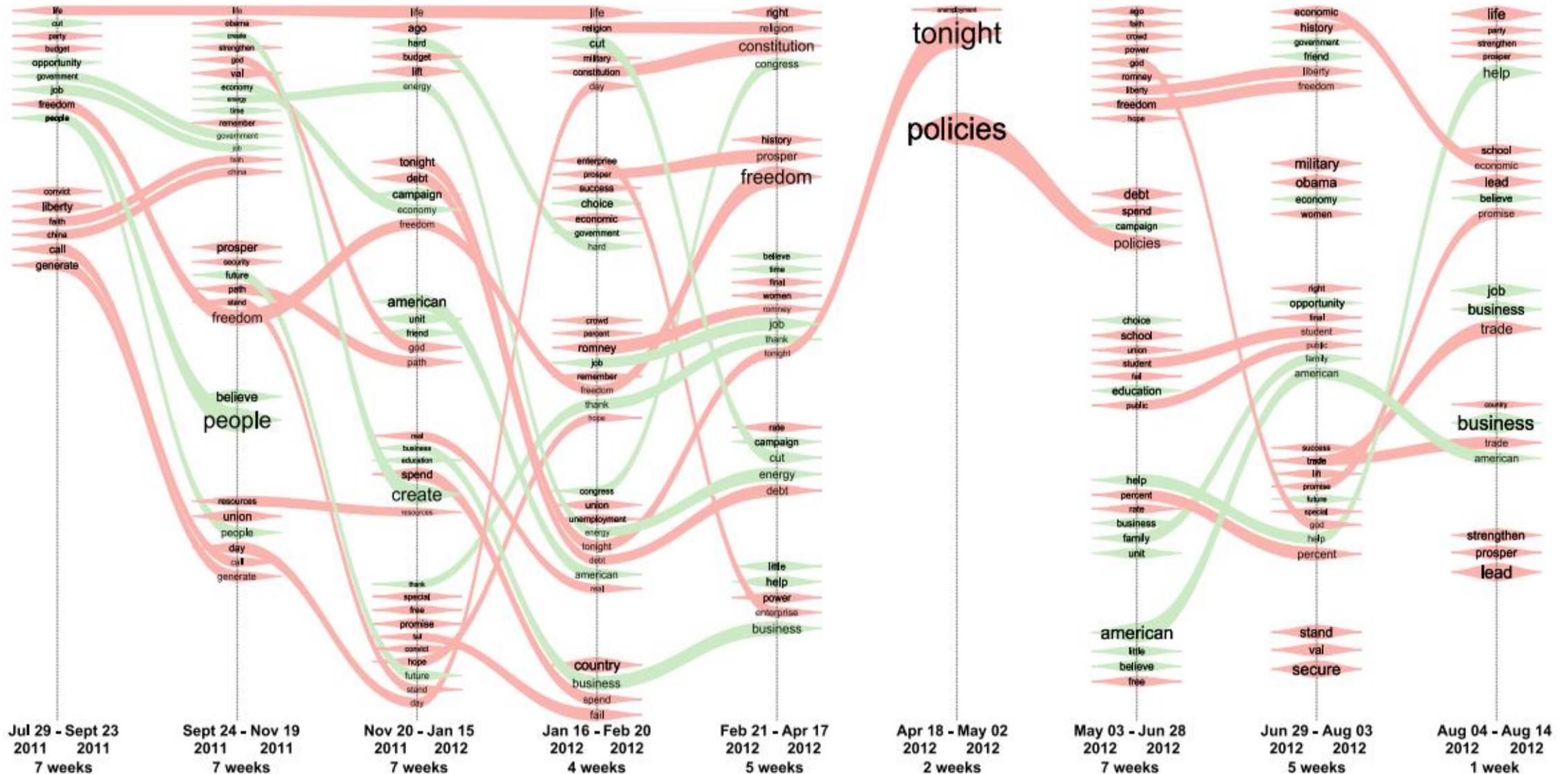
Filter: by selected keywords and resort



Filtering on "Energy"



Example: US Presidential Election 2012 - Mitt Romney



Spanish Flu in the News

Sep 09 – Oct 09

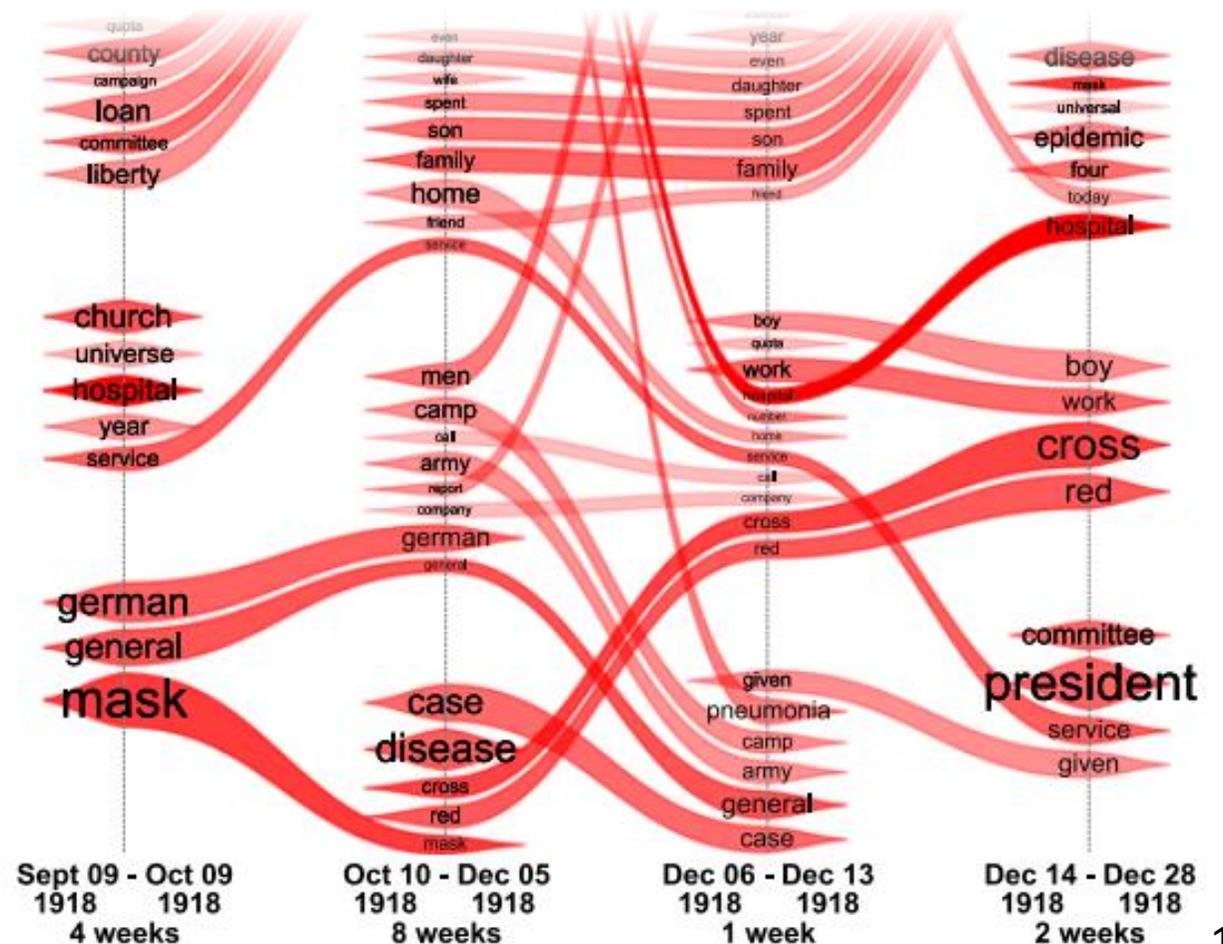
“German”, “mask” – advisories from the first world war to wear a mask

Oct 10 – Dec 05

“home”, “family”, “son”, “daughter” – men from the army were allowed to return home

Dec 06 – Dec 13

“German” disappears – The war was won on November 11.



Expert User Study on Spanish Flu Data

Changed-Focused Questions

How did the newspapers describe the spread of influenza?

How does the description of the pandemic change over time?

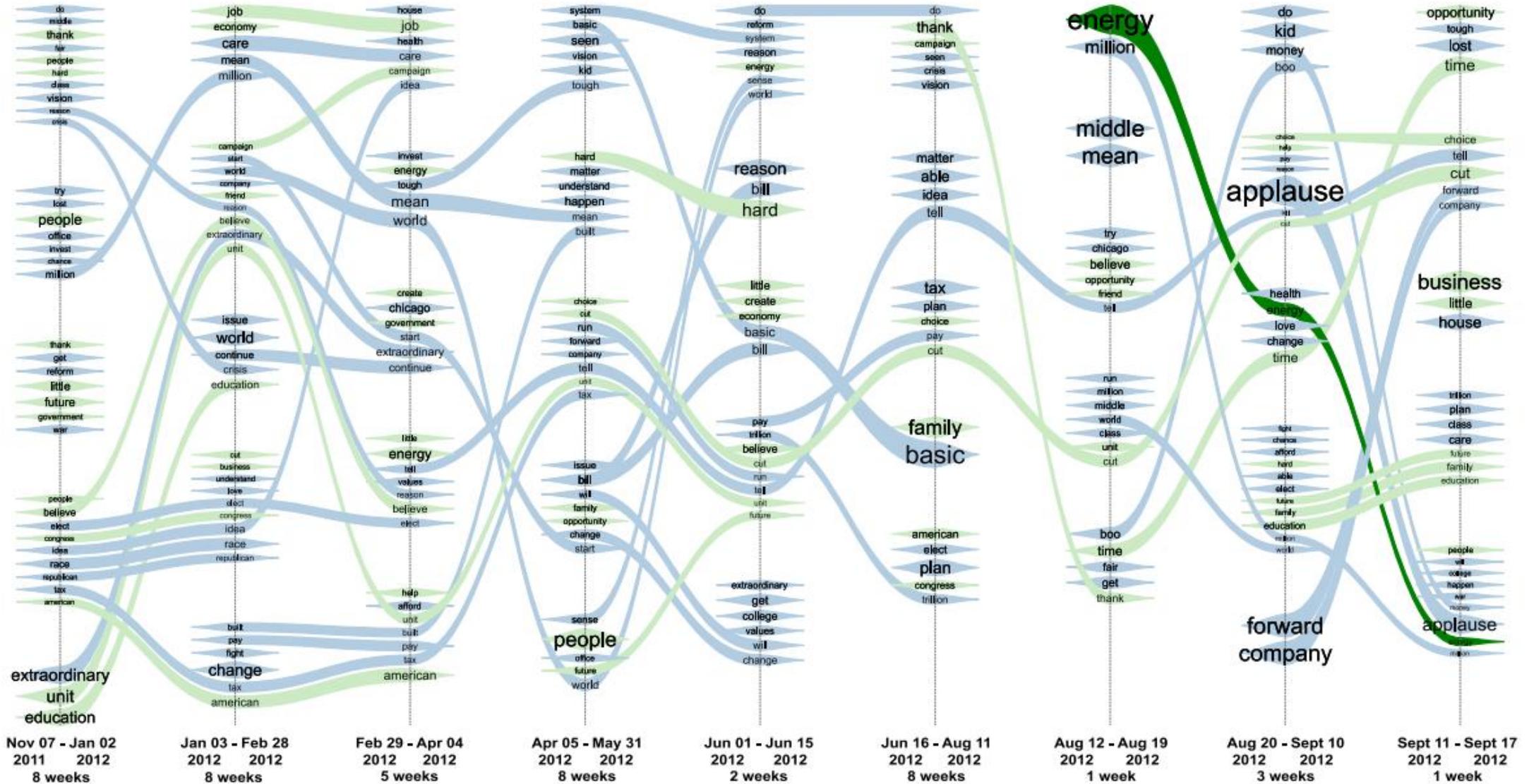
Are there different times when the influenza pandemic becomes less important? What are those time periods?

Connection-Focused Questions

What are the categories that appear to be associated with influenza in different newspapers?

Was there a specific feeling that surrounded the influenza reporting in the newspapers?

Scalability Limits



Thank You

References

ThemeDelta: Dynamic Segmentations over Temporal Topic Models, by Samah Gad, Waqas Javed, Sohaib Ghani, Niklas Elmqvist, Tom Ewing, Keith N. Hampton, Naren Ramakrishnan, in IEEE Transactions on Visualization and Computer Graphics 21(5) 2015.

Visualization Analysis and Design, by Tamara Munzner, A K Peters Visualization Series, CRC Press, 2014.

What? Why? How?

What: Data

- Timestamped text dataset

What: Derived

- Bag of Words over time

- Topics (Groups of words)

- Time intervals containing groups of topics

Why: Tasks

- Identify changes in topics over time

- Identify scatter/gather relationships

How: Encode

- Parallel axes for time segments

- Spatially partition topics along a segment

- Label keywords within topics

- Linked keywords across time intervals

- Segment labels for dates and duration

How: Encode (Free Channels)

- Size of labels for quantitative data

- Width of links for quantitative data

- Link colour for categorical or ordered data

How: Manipulate

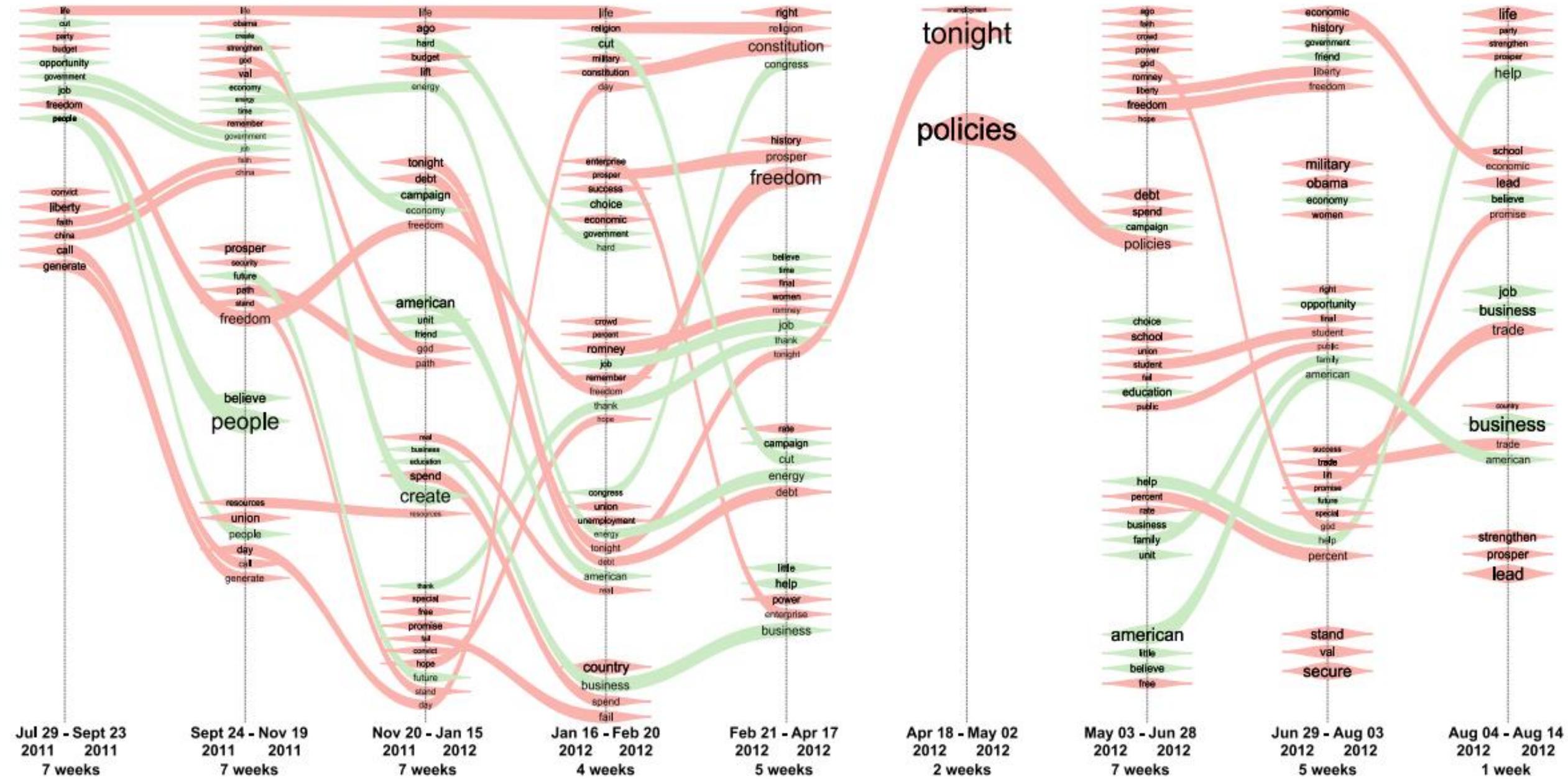
- Navigate: geometric zooming and panning

- Select: highlight keywords

- Search: Select keywords by searching

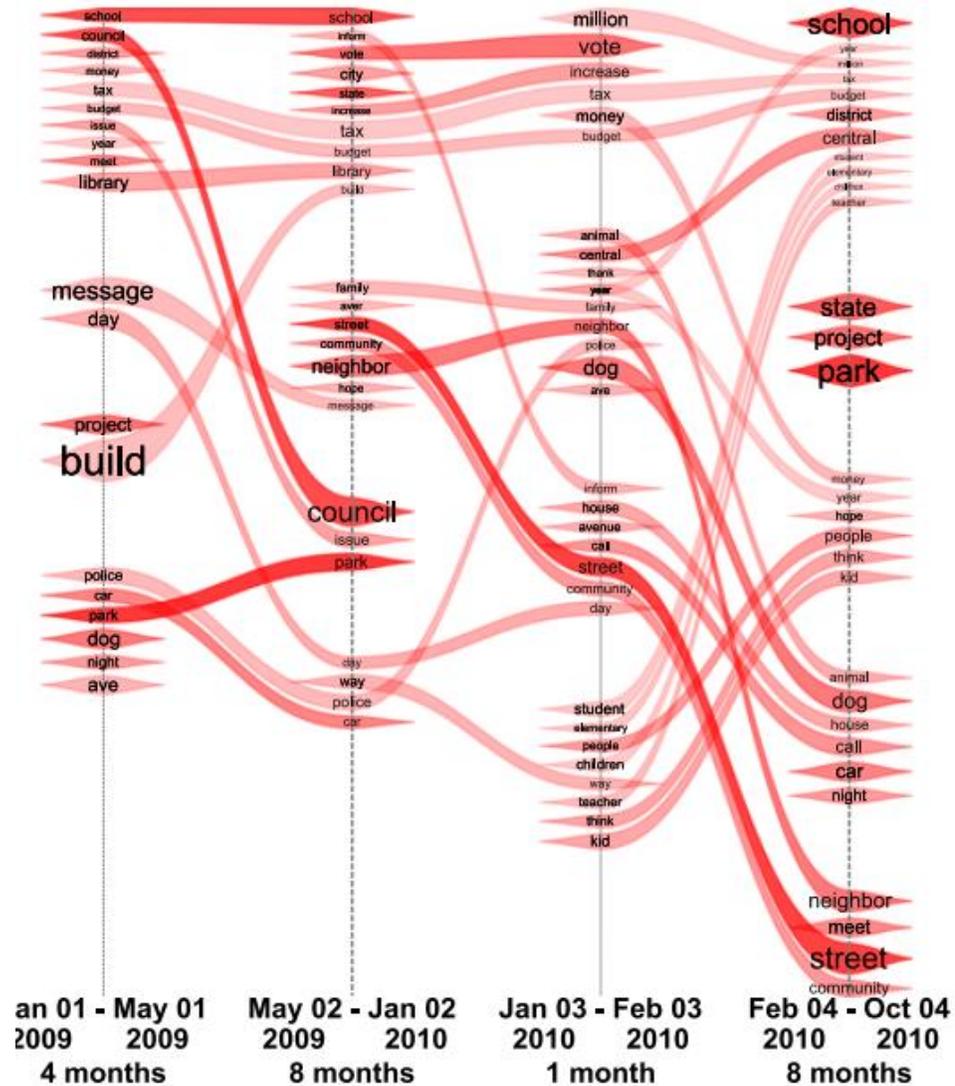
How: Reduce

- Filter: by selected keywords and resort

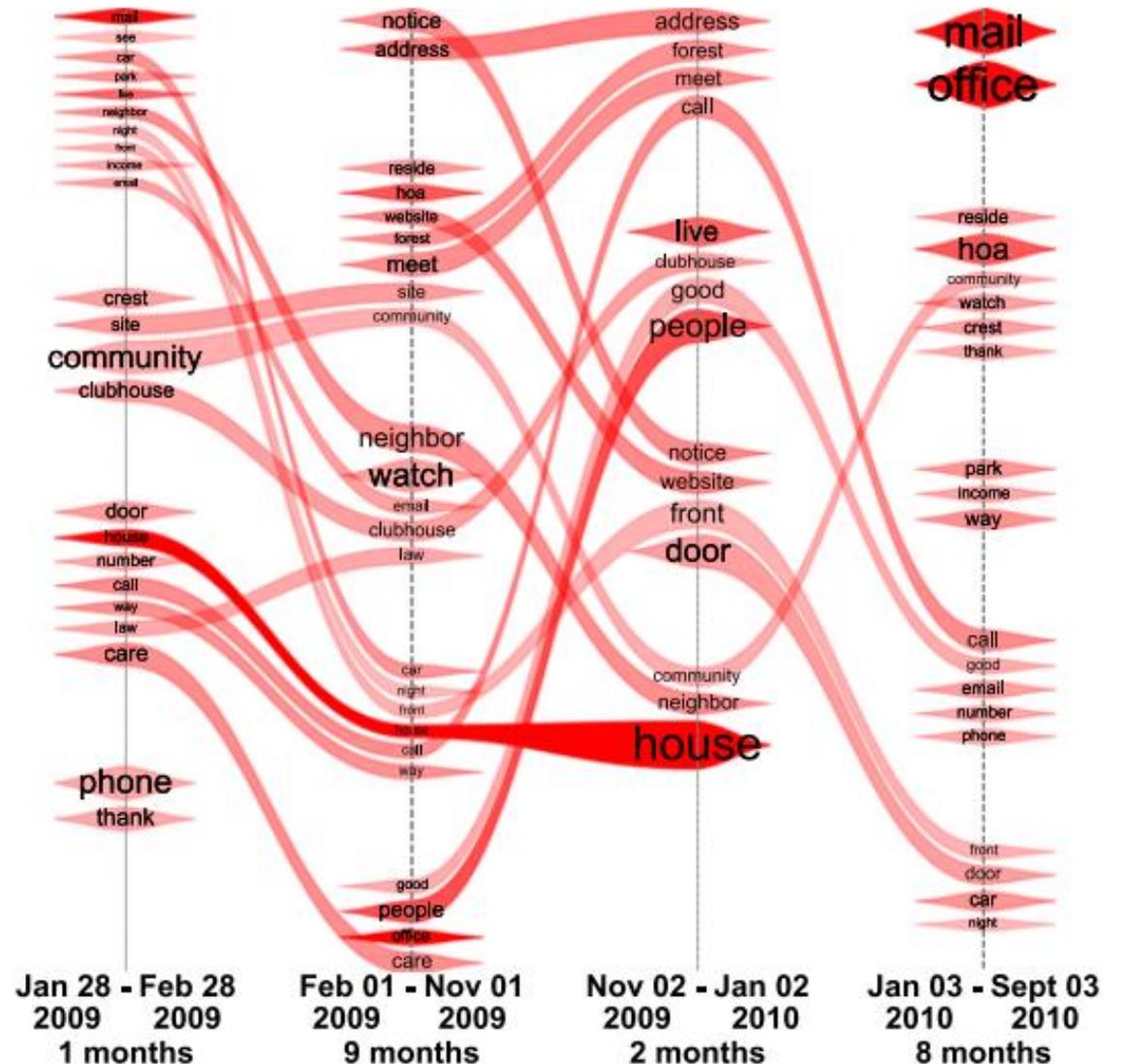


Low vs High Income Neighbourhood Discussions

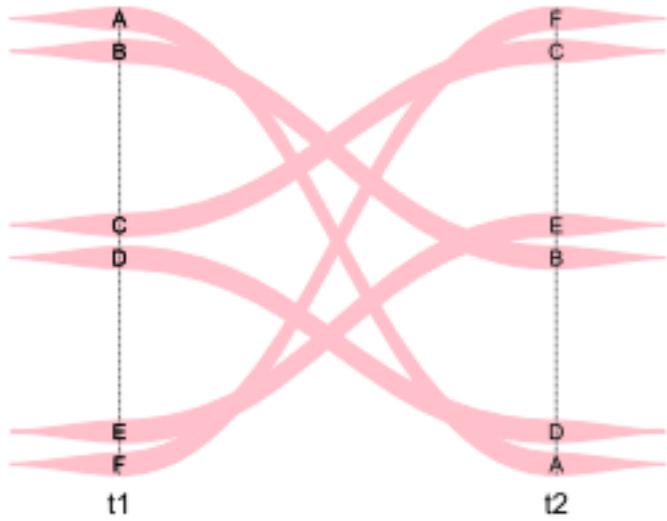
Low Income



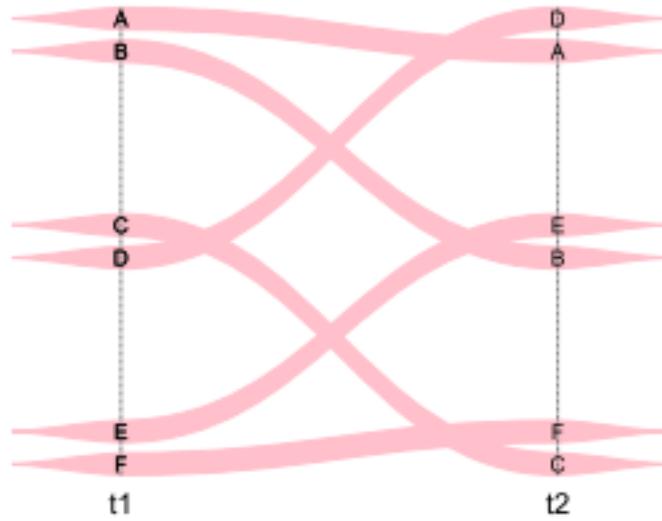
High Income



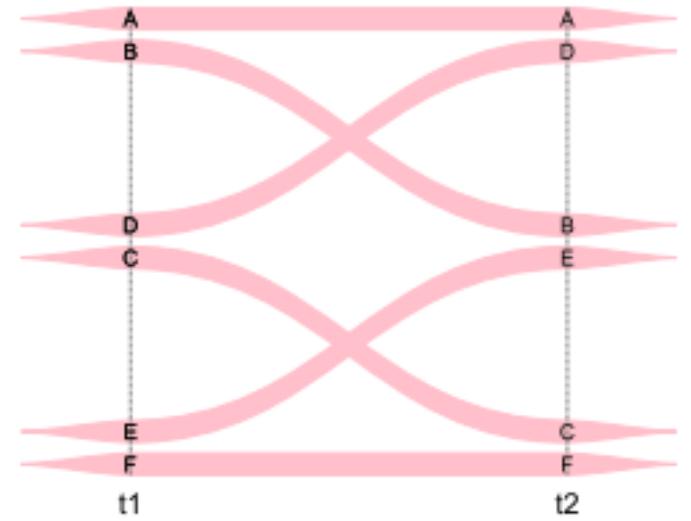
Layout Sorting



(a) Without sorting.



(b) Topic sorting only.



(c) Topic plus term sorting.