

Text visualization

Lucas Rizoli
CPSC 533C, November 2006

Web–pages

Email

Instant messages

Digitized books, articles

Why use text in visualization?

Bleed

Reading is fast



[from <http://en.wikipedia.org/wiki/Image:Statue-Of-Liberty.jpg>]

Text can be a dense representation
Text can be inexact

Fast
Dense
Inexact

Difficulties of using text

Space
Arrangement
Orientation
Legibility
Meaning

afterlife agnostic america anthropology atheism belief bible blog
book bradlaugh buddhism bunting bush carneades cartoons carvaka characters
christianity darwin dawkins death dennett derrida
disbelief dostoyevsky evolution extremism faith freethinking
god greece harris history holbach hume huxley images india intelligent_design
islam jesus jews mill morality moses nietzsche politics positive_atheism prayer
preliterate readings reason religion revival rushdie russell
science shelley skepticism tolerance travels wieseltier woolf writing

This tag occurs 6 times.

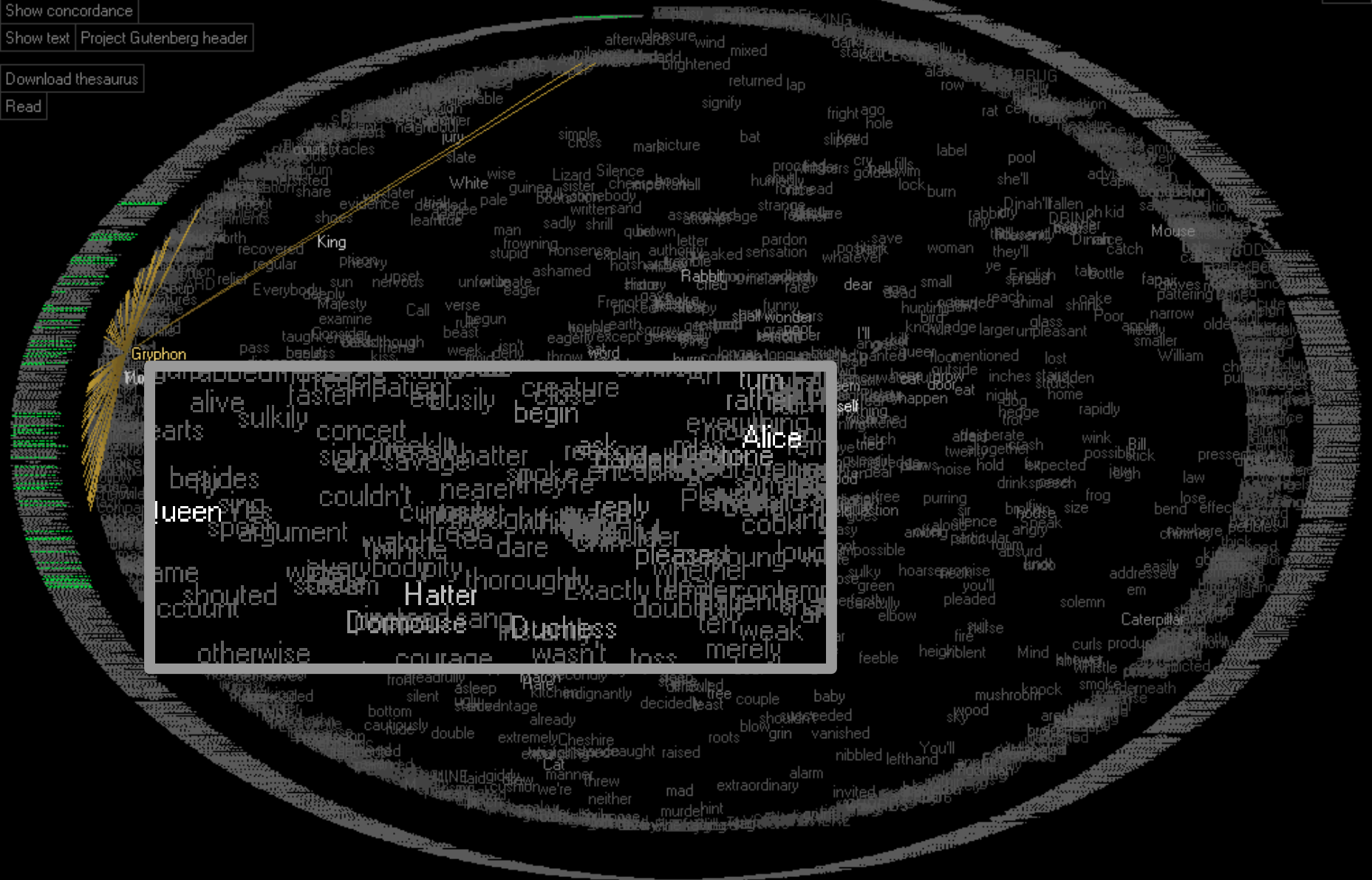
[from <http://www.futureofthebook.org/mitchellstephens/>]

Alice's Adventures In Wonderland

Help
About

They very soon came upon a Gryphon, lying fast asleep in the

- Show concordance
- Show text Project Gutenberg header
- Download thesaurus
- Read



[from <http://www.textarc.org/>]

Alice's Adventures In Wonderland



Help
About

Text

Show only KWIC index (Key Word In Context)

Show

Hide text Project Gutenberg header

Download thesaurus

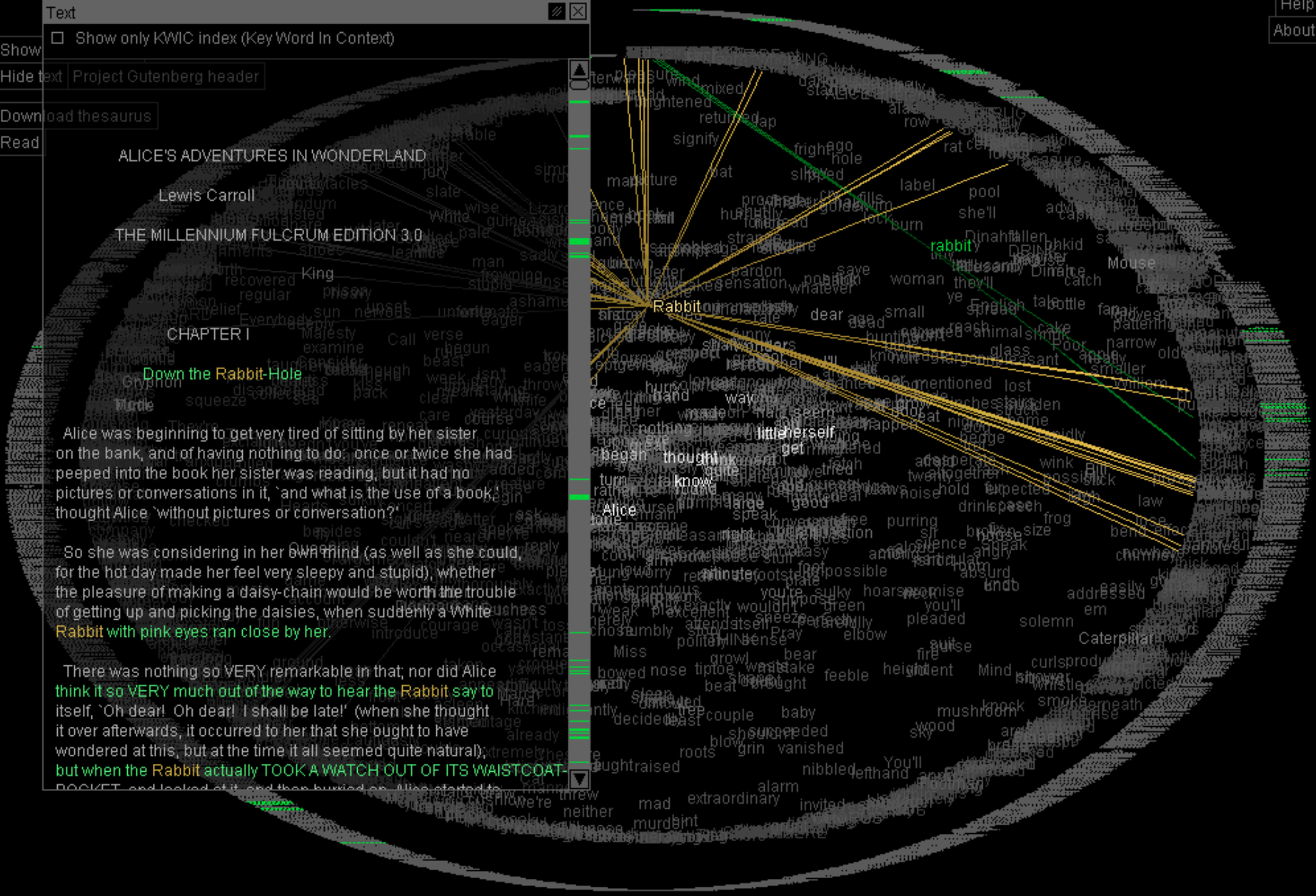
Read

ALICE'S ADVENTURES IN WONDERLAND
Lewis Carroll
THE MILLENNIUM FULCRUM EDITION 3.0
CHAPTER I
Down the Rabbit-Hole

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?'

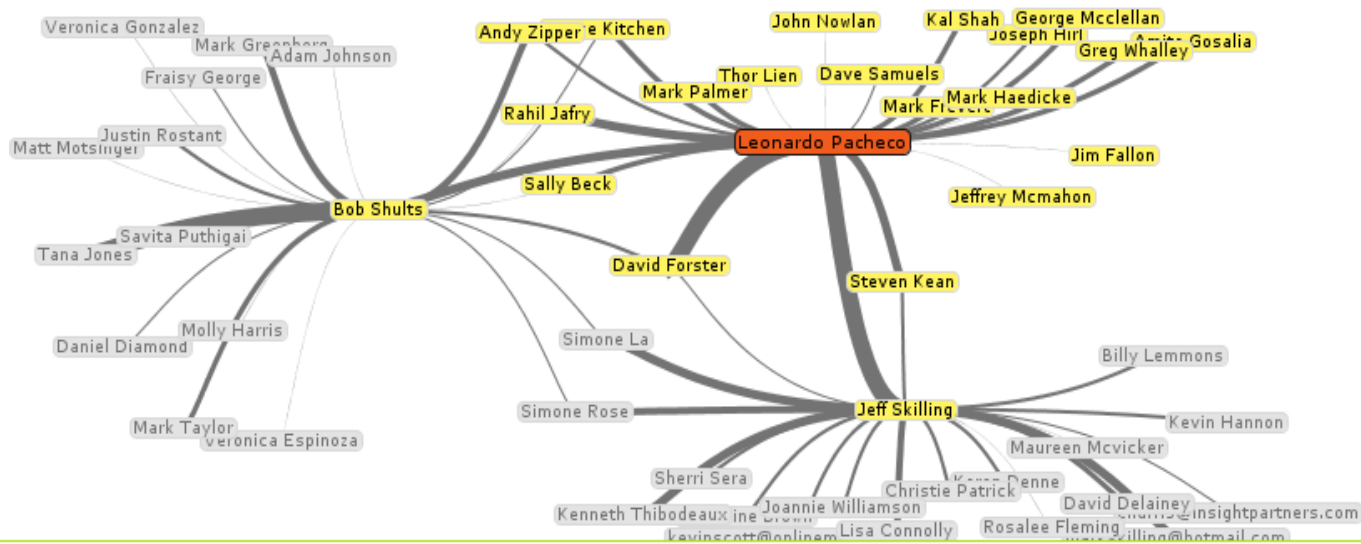
So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a white Rabbit with pink eyes ran close by her.

There was nothing so VERY remarkable in that; nor did Alice think it so VERY much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually TOOK A WATCH OUT OF ITS WAISTCOAT-POCKET, and looked at it, and then hurried on, Alice started to



[from <http://www.textarc.org/>]

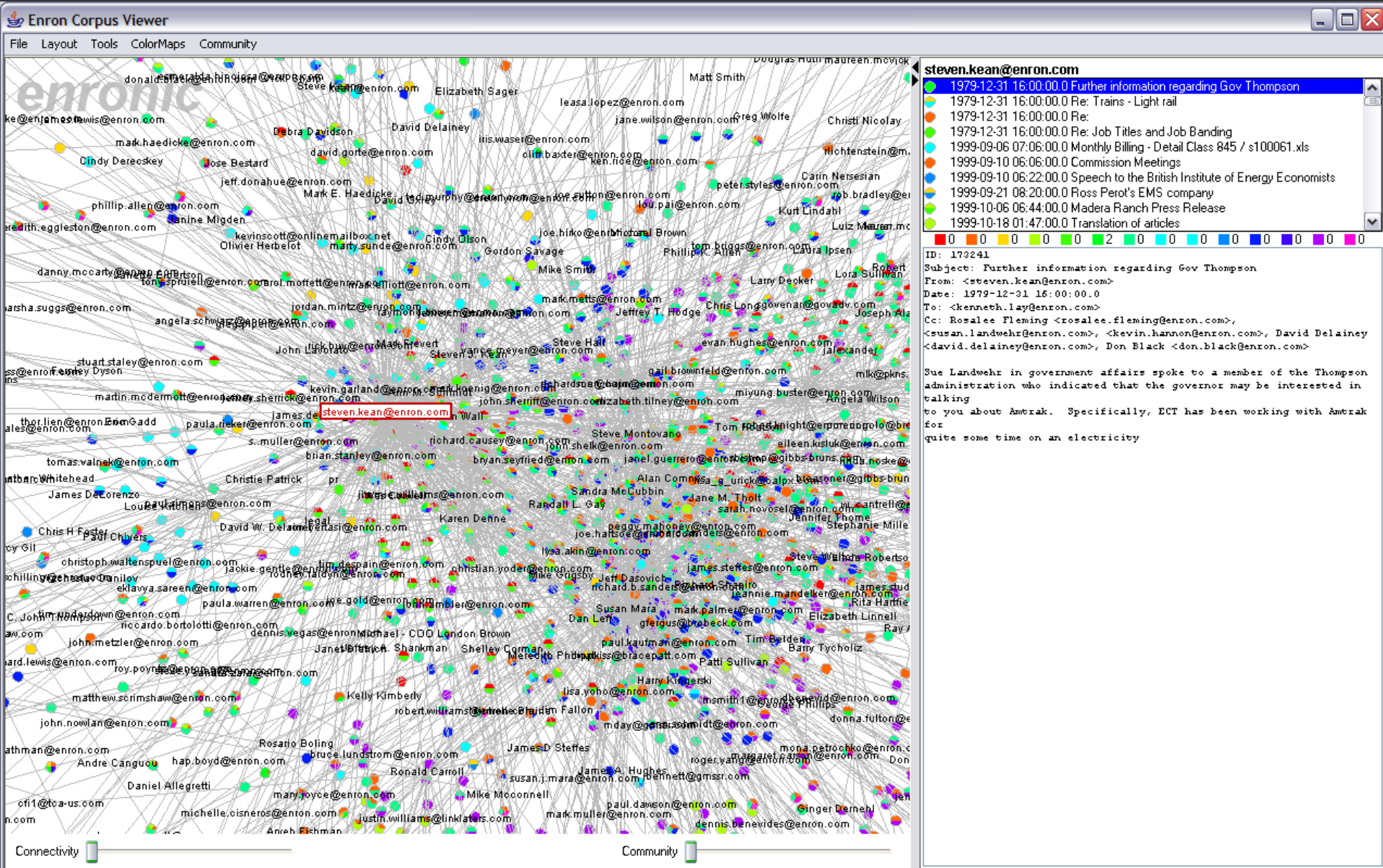
Index
Searching
Explicit in data



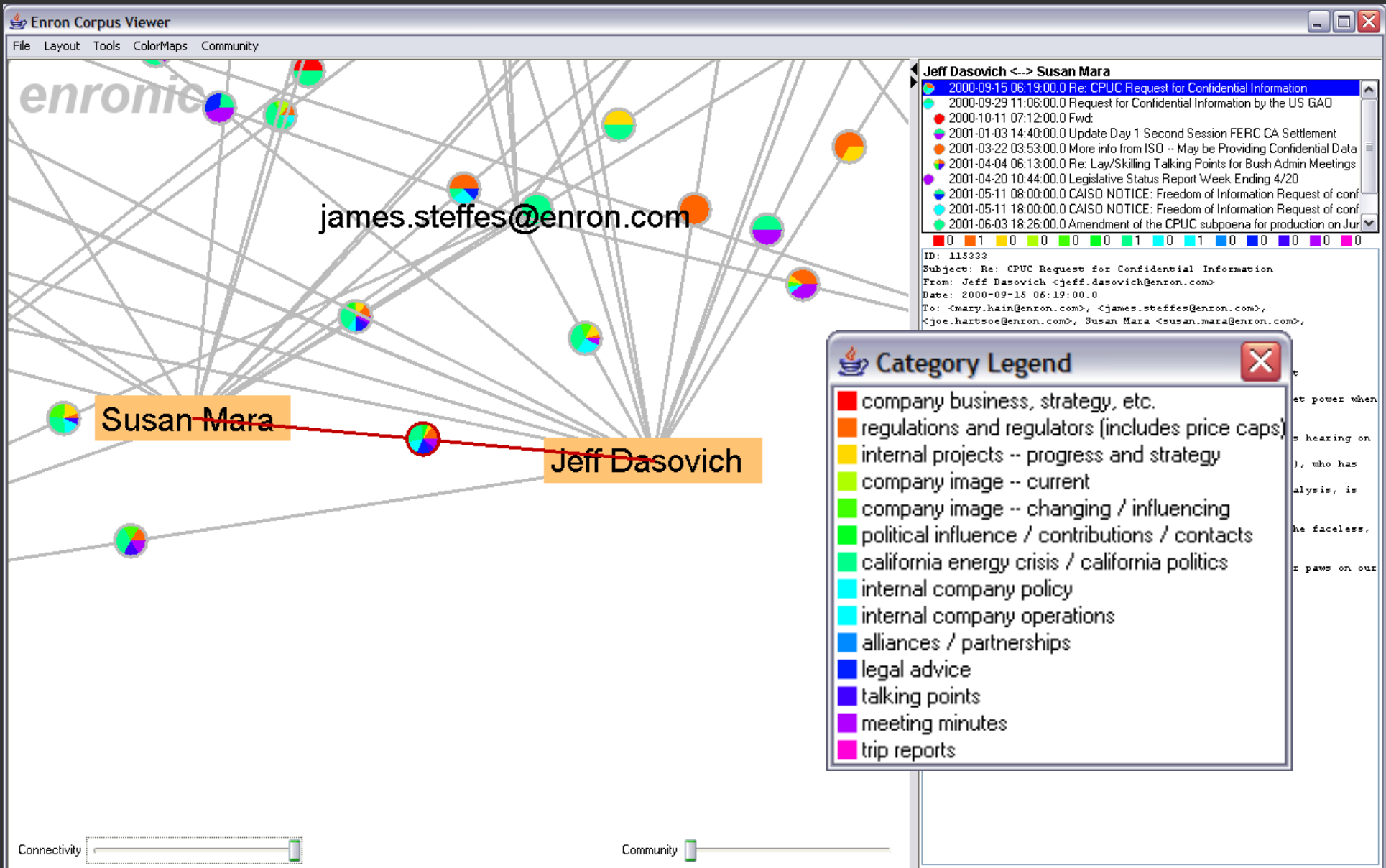
Leonardo Pacheco (326 emails, 20 themes and 20 people)

| emails | | themes | | |
|--|-----------|---|--|--------|
| Trading list Tammie Schoppe | 21 Feb 02 | Average Deal Count Brown Bag Capacity Options Conf Rm Contact Info Electronic Trading Enron | Executive Summary Global Markets Enron Online Gas Trading Global Products | |
| Filed 2/15/02 - Visa Karen Buckley | 19 Feb 02 | Lite Bytz Management Report New Product Open Season Public Report Report Through December Rod Nelson Status Report Year End | | |
| UBSW 401 K information Julie Clyatt | 13 Feb 02 | view all themes | | |
| UBS Warburg Energy - Holiday Schedule Julie Clyatt | 13 Feb 02 | <th>people</th> | | people |
| Immigration: I539 (spouse) Karen Buckley | 11 Feb 02 | Sally Beck Jim Fallon David Forster Mark Frevert Amita Gosalia Mark Haedicke Joseph Hirt Rahil Jafry Steven Kean Louise Kitchen Thor Lien George McClellan Jeffrey McMahon John Nowlan Mark Palmer Dave Samuels Kal Shah Bob Shults Greg Whalley Andy Zipper | | |
| Trader Presentations - 2/7 @ 2:30 p.m. (CST) in Conf... Liz Taylor | 07 Feb 02 | | | |

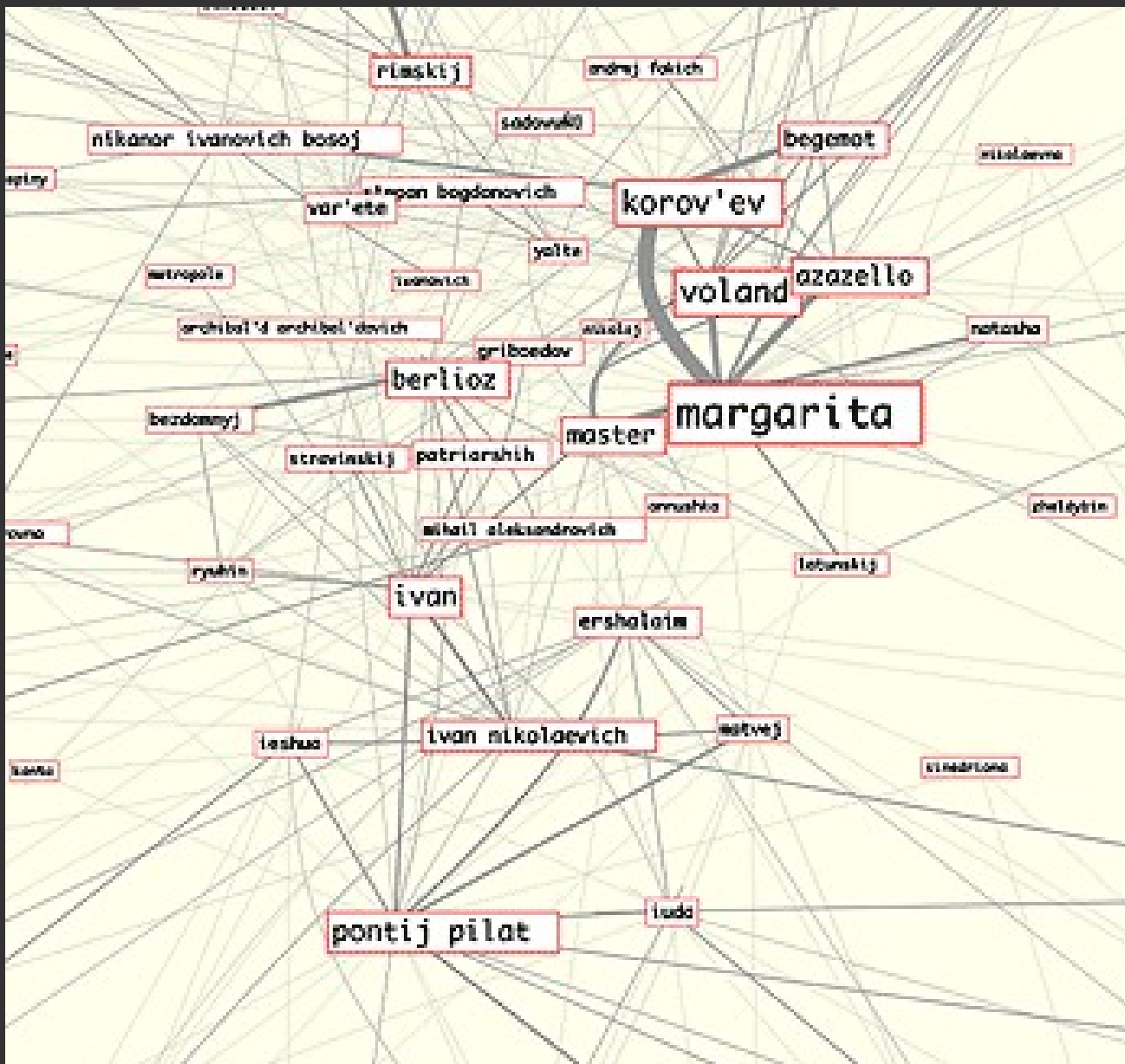
[from <http://enron.trampolinesystems.com/>]



[from <http://jheer.org/enron/>]



[from <http://jheer.org/enron/>]



[from http://www.idlewords.com/2004/03/your_literary_masterpiece_was_delicious.htm]

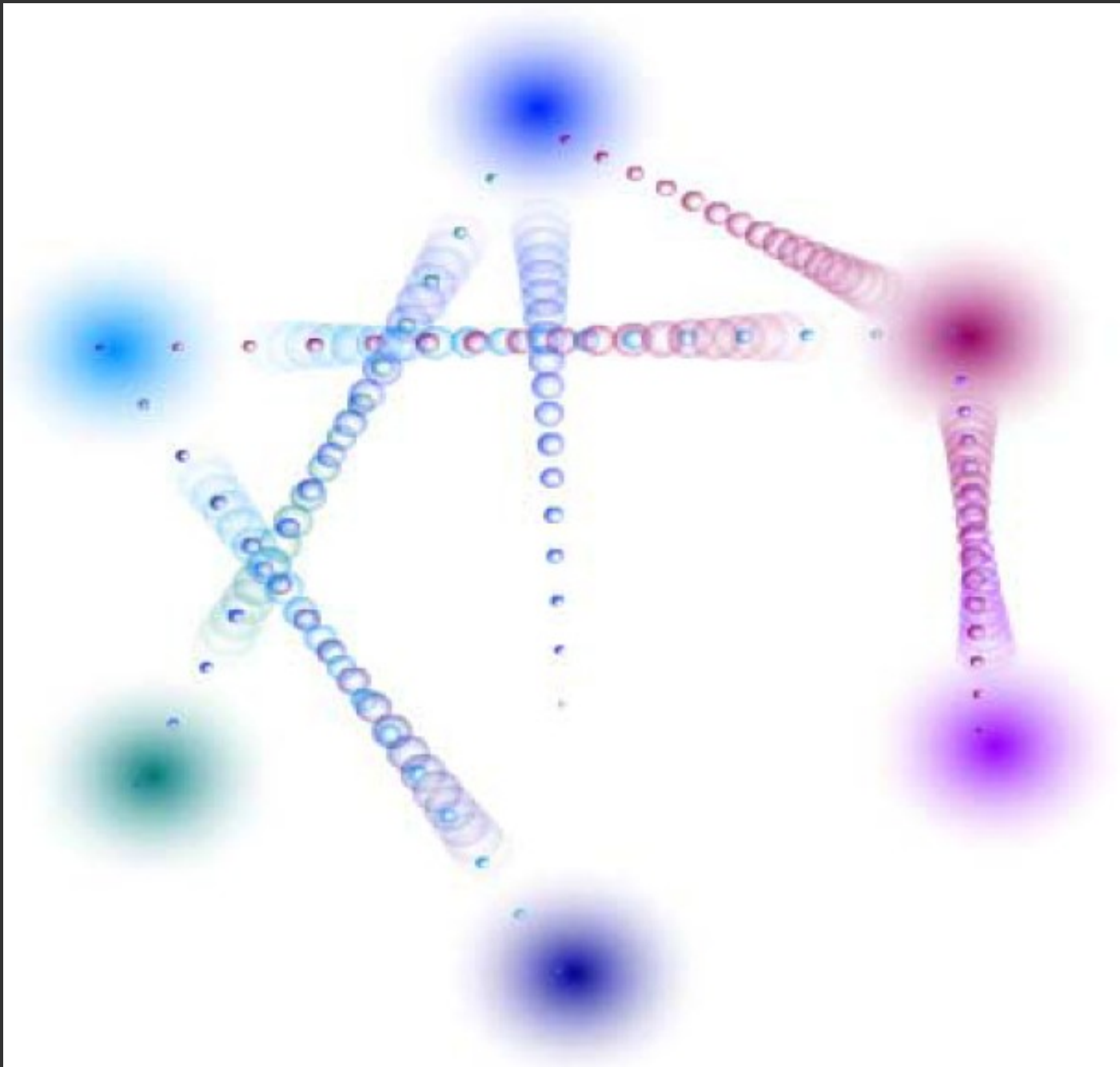
Graph

Analyzing

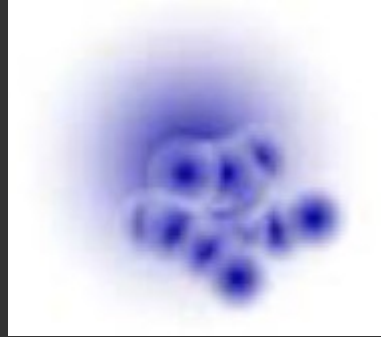
Derived from data

Human supervision of automated processes

Reliance on meta-data
Says little about content



[from Tat, A., & Carpendale, M. S. T. (2002)]



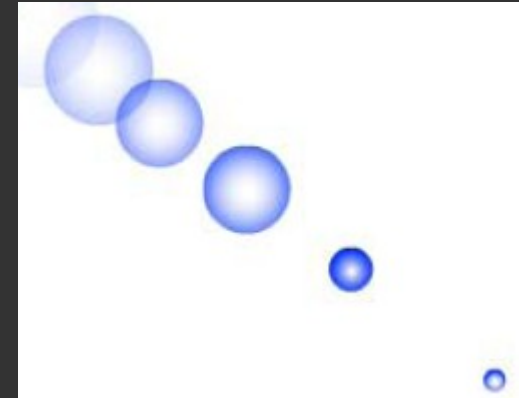
Wordiness



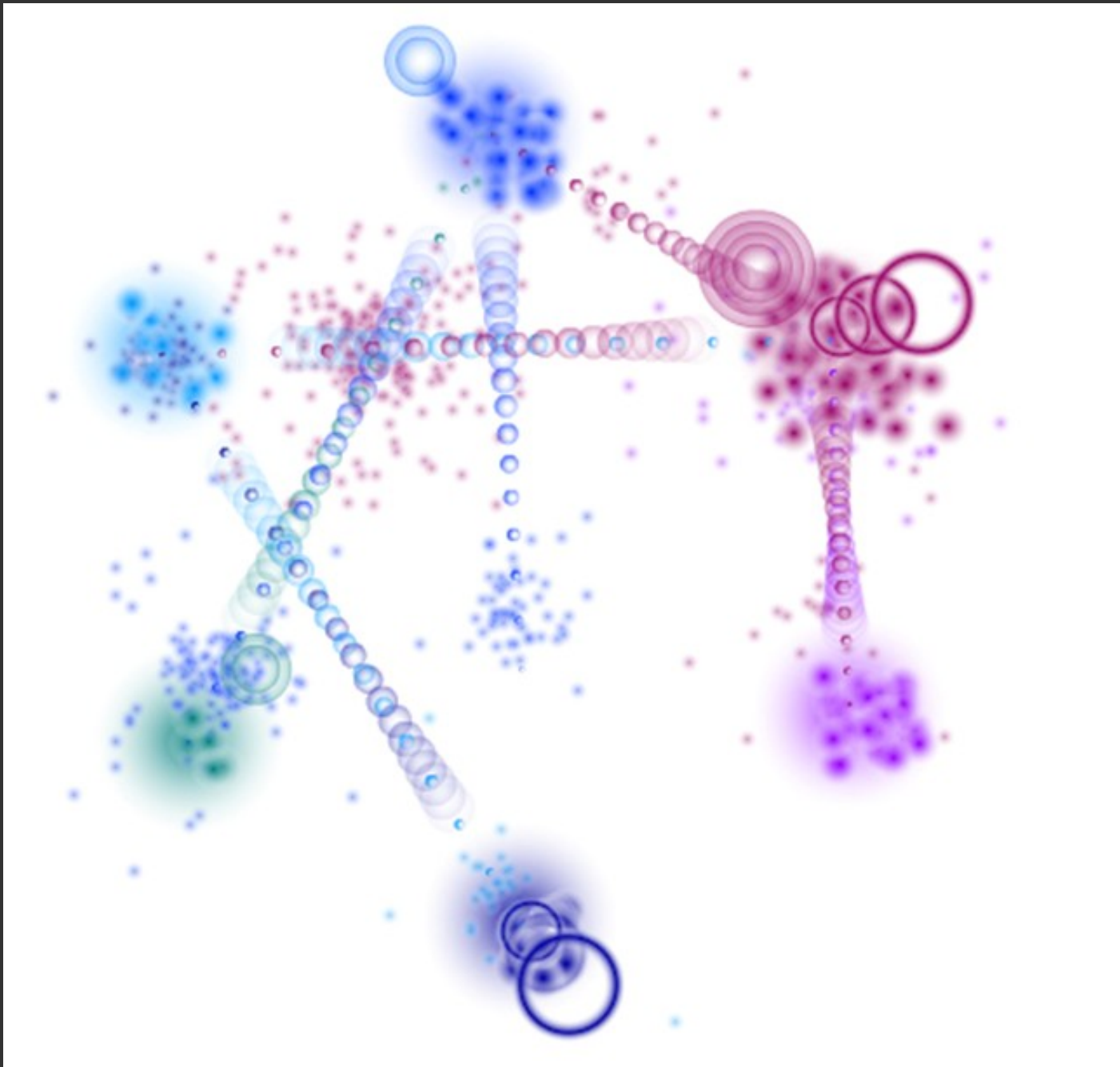
CAPS



Exclamations



Direction of
conversation



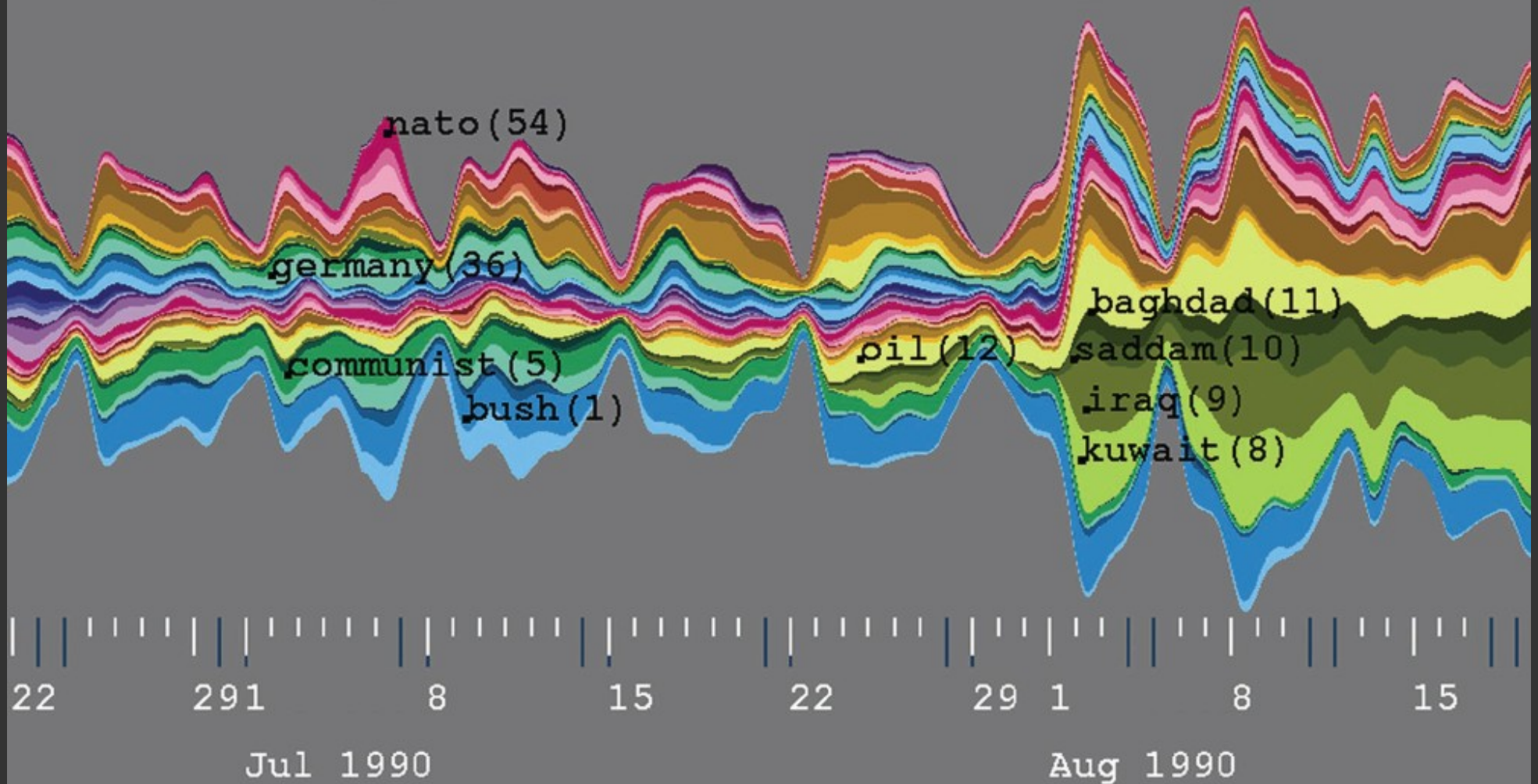
[from Tat, A., & Carpendale, M. S. T. (2002)]

German economic
and monetary union

OPEC agrees to
raise oil price

NATO to redefine
military strategy

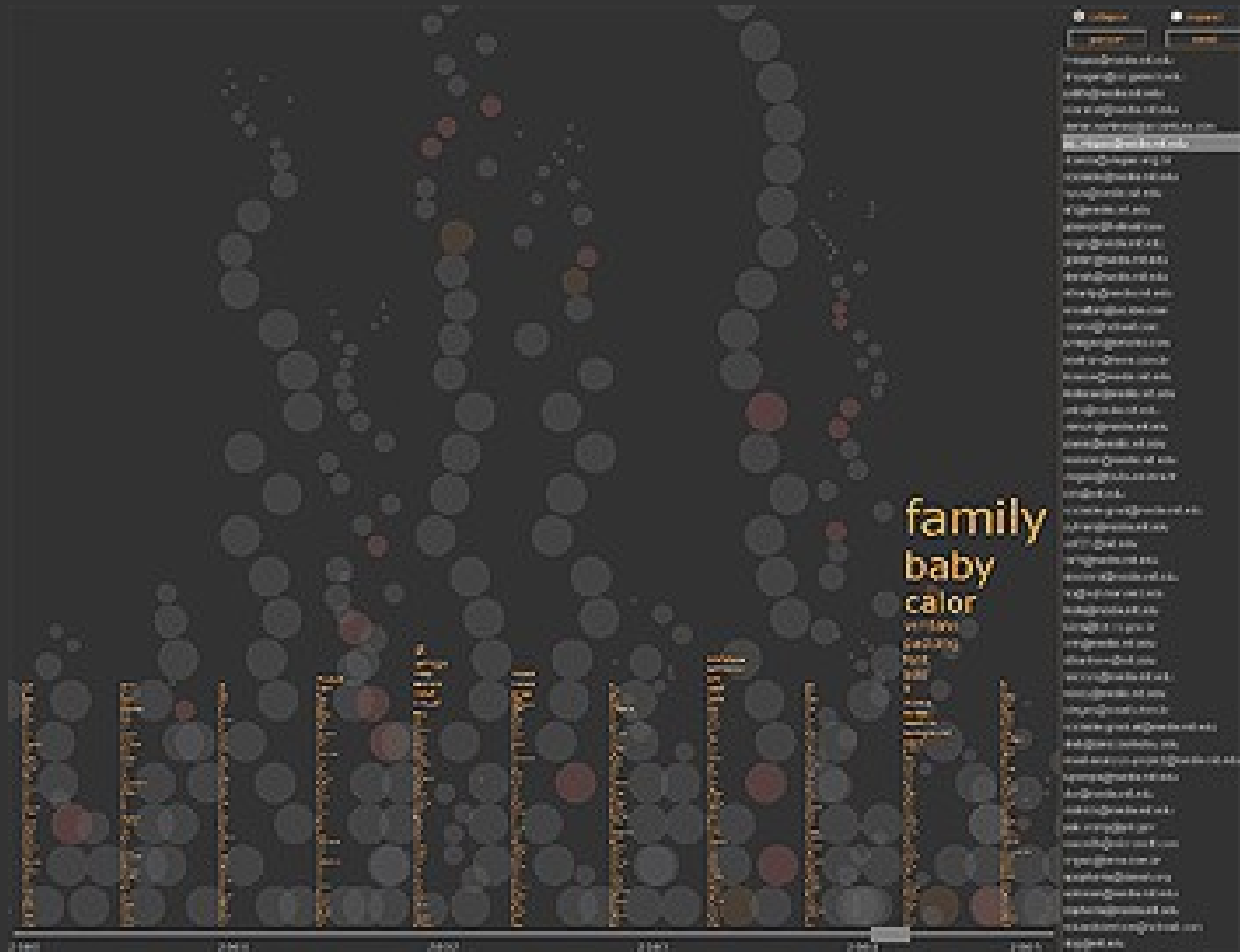
Iraq invades
Kuwait



[from Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002)]



[from Viégas, F. B., Golder, S., & Donath, J. (2006)]



[from <http://alumni.media.mit.edu/~fviegas/projects/themail/study/index.htm>]



[from Viégas, F. B., Golder, S., & Donath, J. (2006)]

Unique visual representation
Exploration
Derived from data
Increasingly semantic
Greater reliance on human users

Trouble pre-processing data
Many assumptions made

Finding meaning in text is difficult

Adjusting for word frequency
Full semantic processing

Take-home lessons

Text in visualization

Fast, dense, inexact
Complicated to apply

Visualizing text

Range of levels and methods

Meta-data adds structure

Pre-processing is hard, important

- Ceglowski, M. (2004). *Your Literary Masterpiece Was Delicious*. Retrieved November 6, 2006 from http://www.idlewords.com/2004/03/your_literary_masterpiece_was_delicious.htm
- Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). ThemeRiver: Visualizing Thematic Changes in Large Document Collections. *Visualization and Computer Graphics, IEEE Transactions*, 8, 9–20.
- Heer, J. (2004). *Exploring Enron: Visualizing ANLP Results*. Retrieved November 6, 2006 from <http://jheer.org/enron/v1/>
- Paley, W. B. (2002). TextArc: Showing Word Frequency and Distribution in Text. In Wong, P. C., & Keith Andrews (Eds.), *Proceedings of the IEEE Symposium on Information Visualization (Infovis '02) Poster Compendium*. Los Alamitos, CA, USA: IEEE Press.
- Paley, W. B. (n.d.). *TextArc.org Home*. Retrieved November 6, 2006 from <http://textarc.org/>
- Tat, A., & Carpendale, M. S. T. (2002). Visualizing Human Dialog. *Proceedings of the IEEE Conference on Information Visualization (Infovis '02)*. London, UK: IEEE Press.
- Trampoline Systems (n.d.). *Trampoline Enron Explorer*. Retrieved November 6, 2006 from <http://enron.trampolinesystems.com/>
- Viégas, F. B., Golder, S., & Donath, J. (2006). Visualizing Email Content: Portraying Relationships from Conversational Histories. *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '06)*. Montréal, Québec, Canada: ACM Press.