

Condor Tips & Tricks



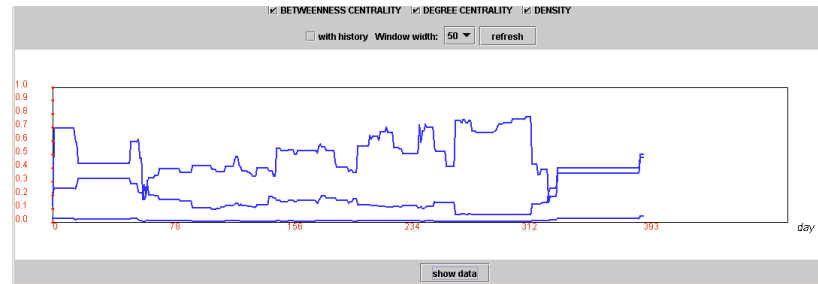
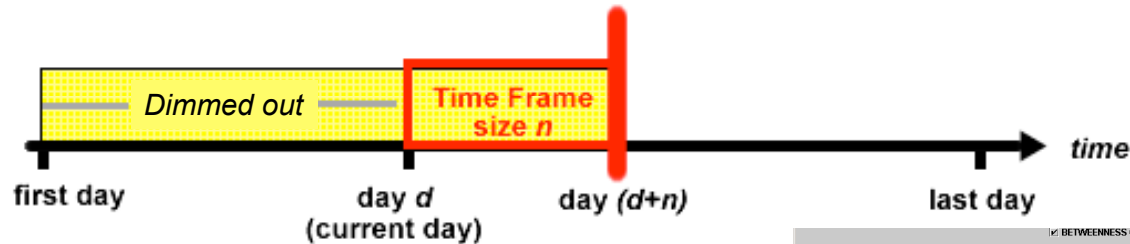
1/10/2008 Peter Gloor
pgloor@mit.edu

Obtain academic version from

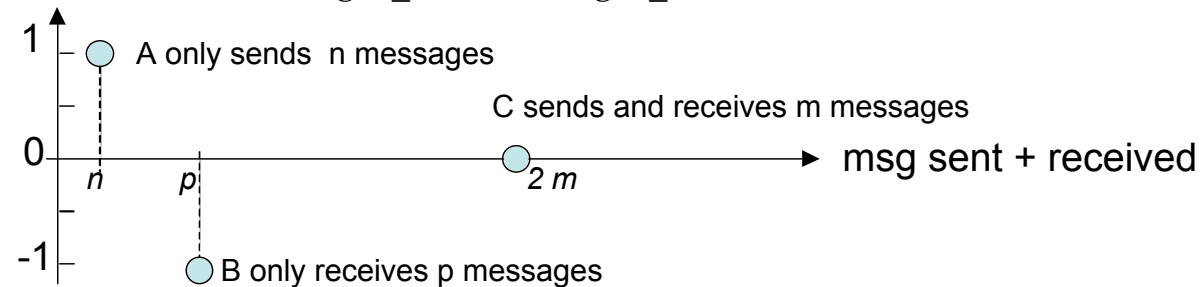
- <http://tiger.soberit.hut.fi/virtualbrownbag>
- Need to register first
- Go to file galleries
 - Condor
 - Condor support tools
- Needs monthly authorization code, obtainable from pgloor@mit.edu

Condor Components

With history:



Contribution Index $\frac{\text{messages_sent} - \text{messages_received}}{\text{messages_sent} + \text{messages_received}}$



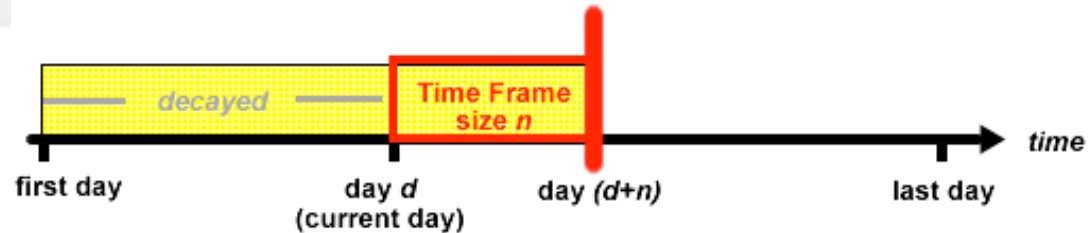
fillia.makedon@dartmouth.edu,
 yan.zhao@dartmouth.edu,
 jamie.ford@dartmouth.edu

} devlab@cs.dartmouth.edu

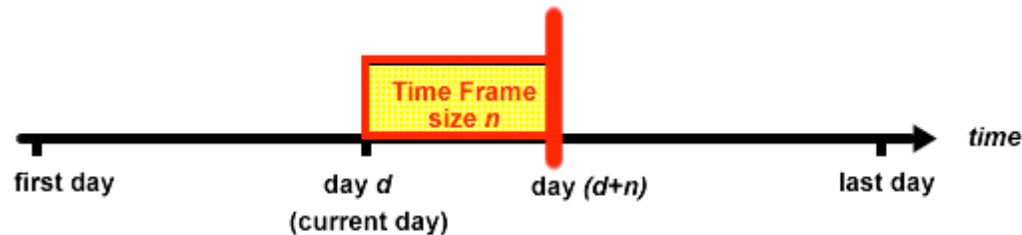
Temporal Visualization by a Sliding Time Frame



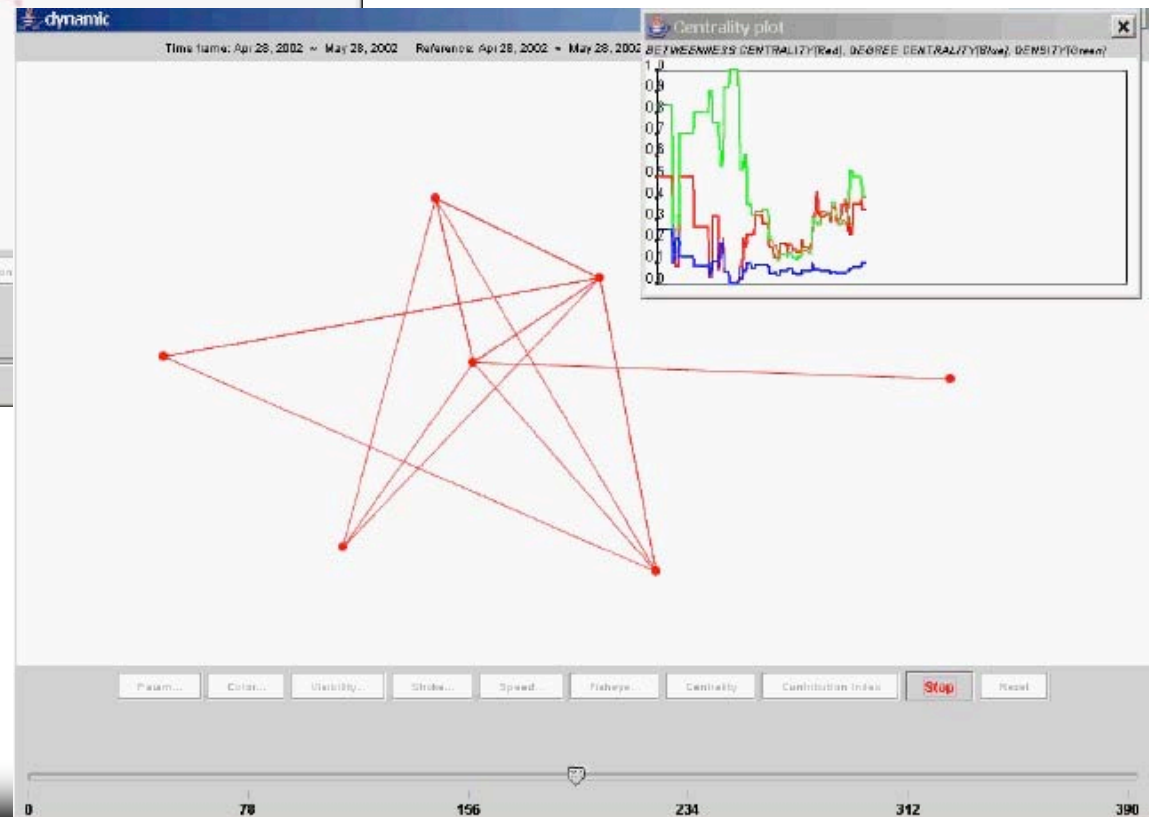
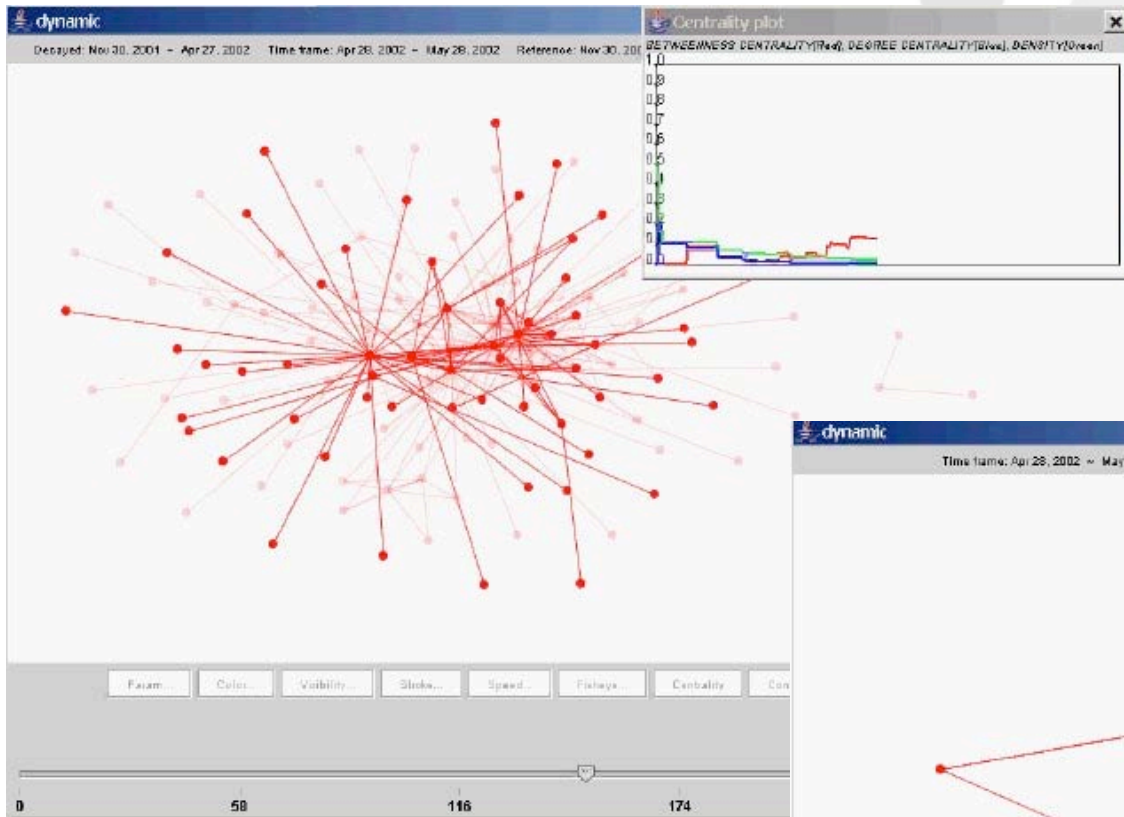
With history:



No history:



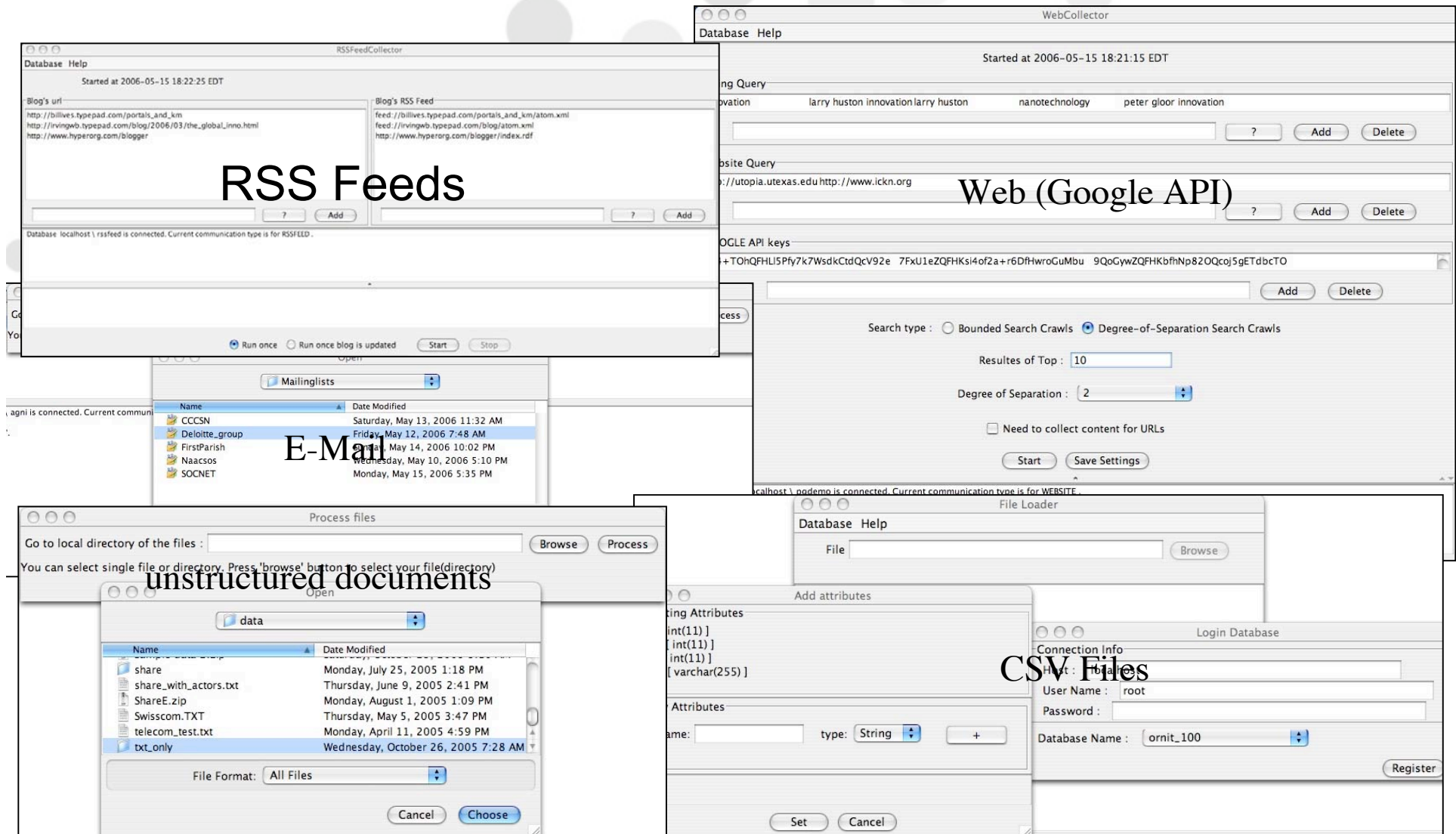
With and without history



Importing E-Mail Data

- Convert Outlook to Eudora (converter.exe)
- Load Eudora *.mbx files directly
- Collect Weblogs
- Load Excel spreadsheets (Fileloader)
- Converting gmail: configure Eudora as pop client
- Converting Yahoo mail to pop: YPOPS to Outlook Express
- Spider online mailinglists (online.jar)

Condor takes 5 different input formats



Condor database structure

datasetId	mailAddress	organization	domain
3	schwartz jeff	UNKNOWN	us - atlanta
3	indorf kaplan nalir	UNKNOWN	us - new york
4	lamert steven	UNKNOWN	us - san francisco
4	schmalberger thor	UNKNOWN	ch - basel
4	kmwiig@krii.com	krii	com

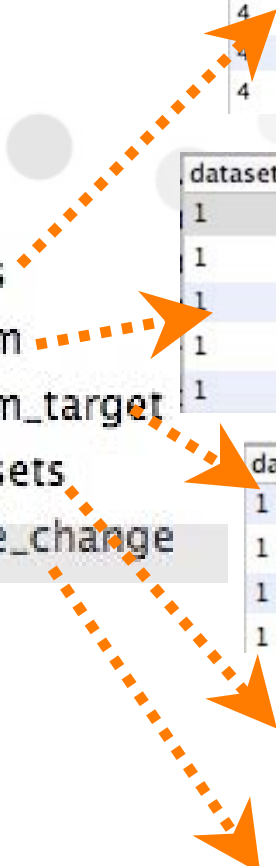
datasetId	comm_id	comm_from	comm_subject	comm_time	comm_content	timestamp
1	1	schmalberger thor	CKN Webinar upda	2002-03-20 11:4		2005-07-25 13:2
1	2	jooste adriaan	Accepted: CKN We	2002-03-21 00:0		2005-07-25 13:2
1	3	schmalberger thor	Dry-run CKN Webi	2002-05-03 11:0		2005-07-25 13:2
1	4	schmalberger thor	Updated: Dry-run	2002-05-15 03:2		2005-07-25 13:2
1	5	schmalberger thor	Updated: Invitatio	2002-05-29 08:4		2005-07-25 13:2

datasetId	comm_id	comm_to	tag
1	1	gloor peter	To
1	1	jooste adriaan	To
1	1	athey robin	To
1	1	gauton anne	To

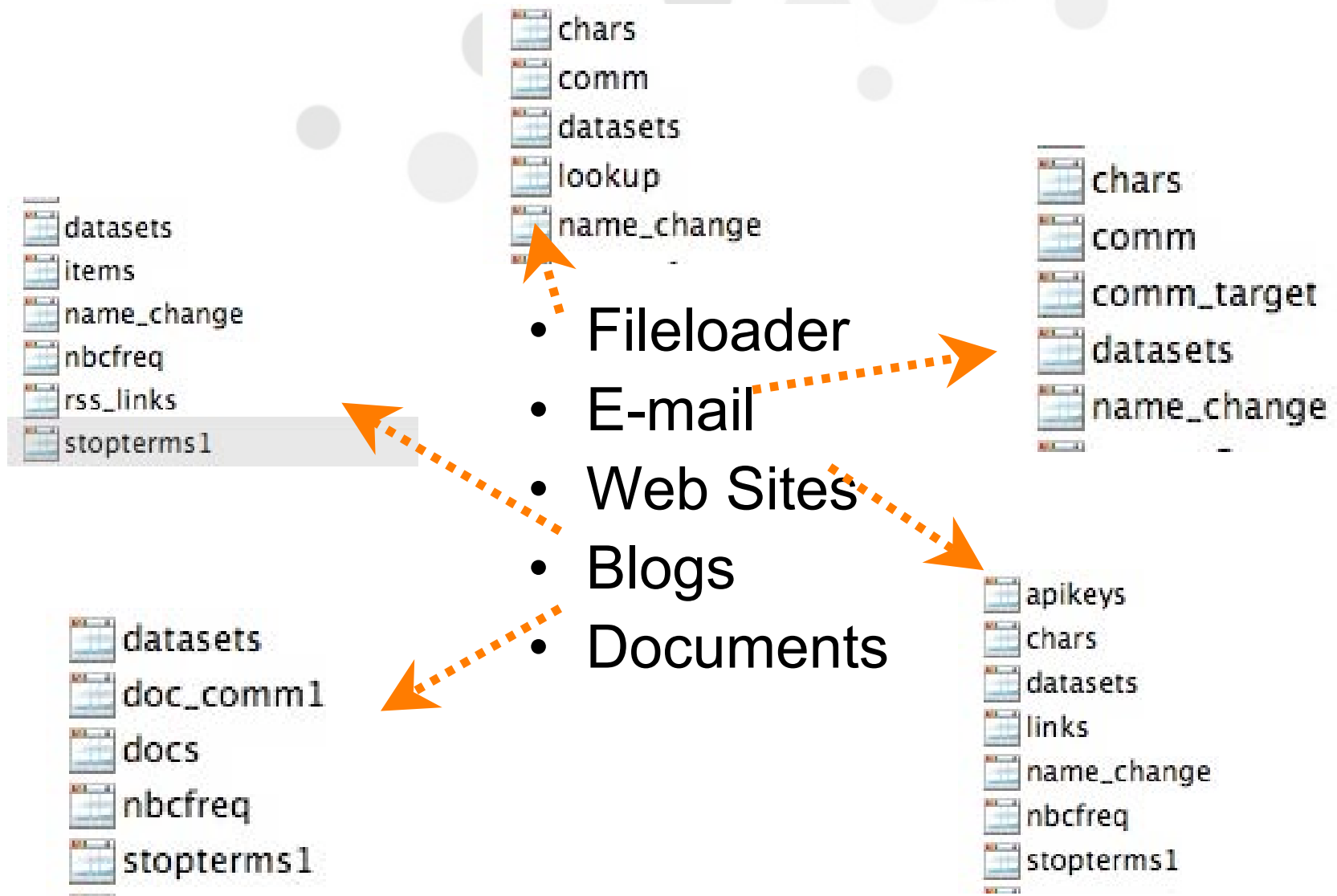
datasetId	dataset
1	webinar
2	firstpa
3	comblid
4	dataset

old_name	new_name
dcame-lstk-4v3pt	
nico@nicokleyn.co	kleyn nico
nikleyn@dc.com	kleyn nico

chars
comm
comm_target
datasets
name_change



5 types of databases



Condor Scalability

- 500-5000 actors
- 2000-100,000 messages (links,ties)

Four ways of scaling down:

1. “Want to set time constraints” (uses less memory)

set time interval and content for dataset combld

Time

from 2001-10-16 01:50:12 to 2002-11-13 06:35:30

☐ all time - or - ☐ time before: begin time: 2003 Jan 1

☒ time between: end time: 2004 Jan 1

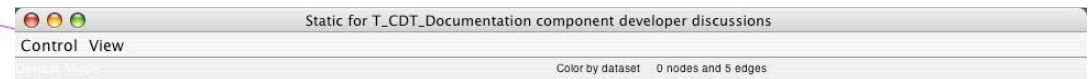
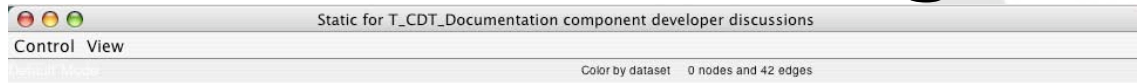
☐ time since:

☒ all contents - or - with keyword(s):

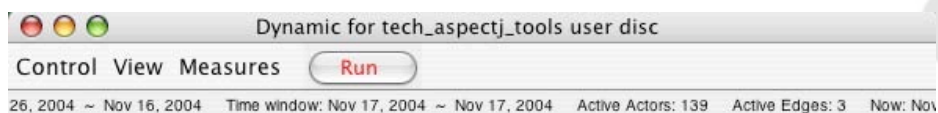
& means AND, I means OR, (specifies the precedence. " encloses single word. e.g. (houseI"real estate")&sale

2. Merge addresses (uses less memory)
 3. Reduce communication frequency
 4. Run without history
- Reducing number of actors saves most memory

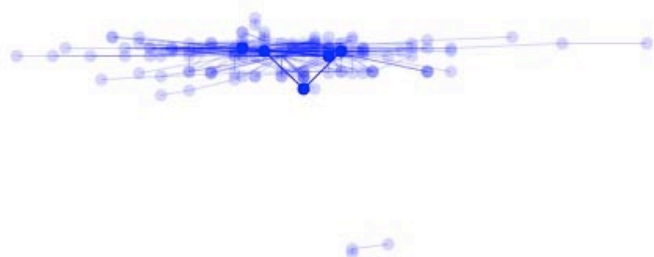
Delete mailing list addresses



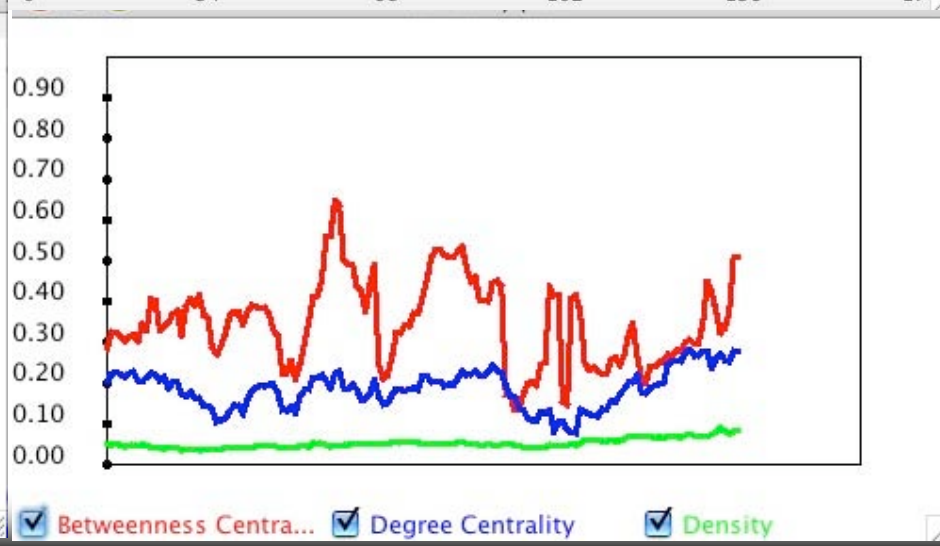
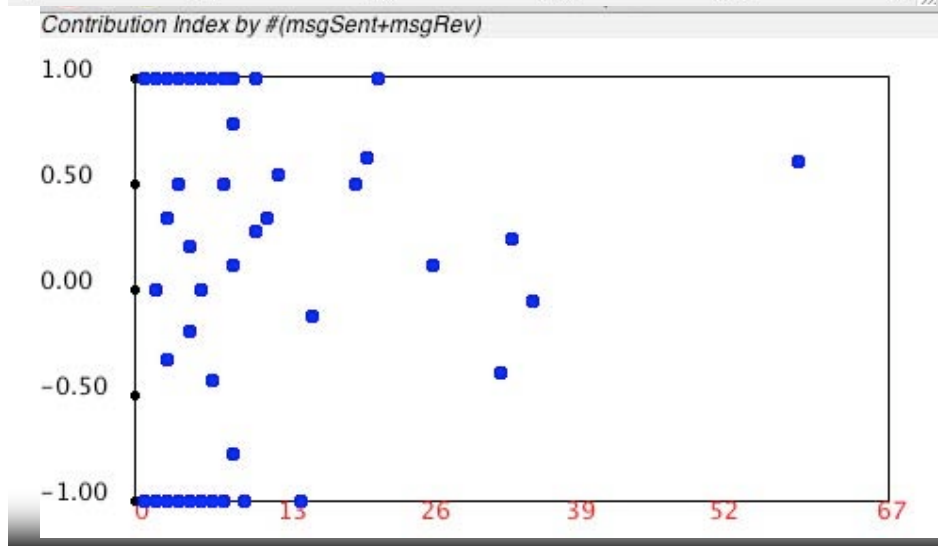
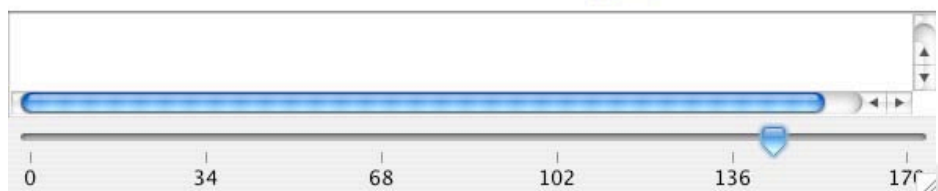
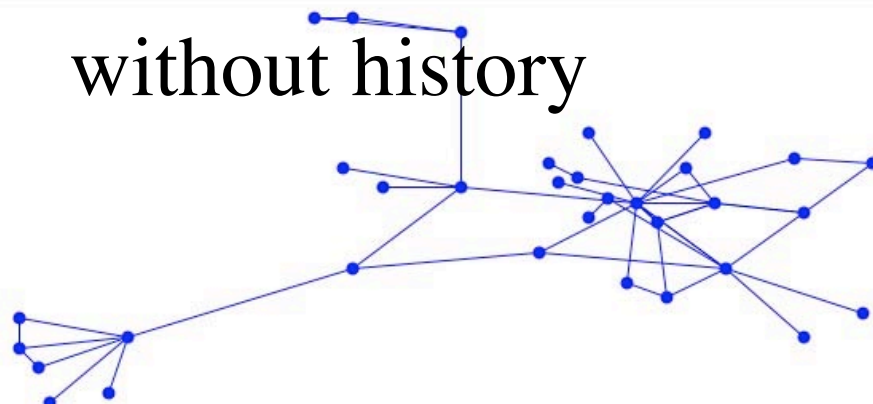
2. Discover COINs



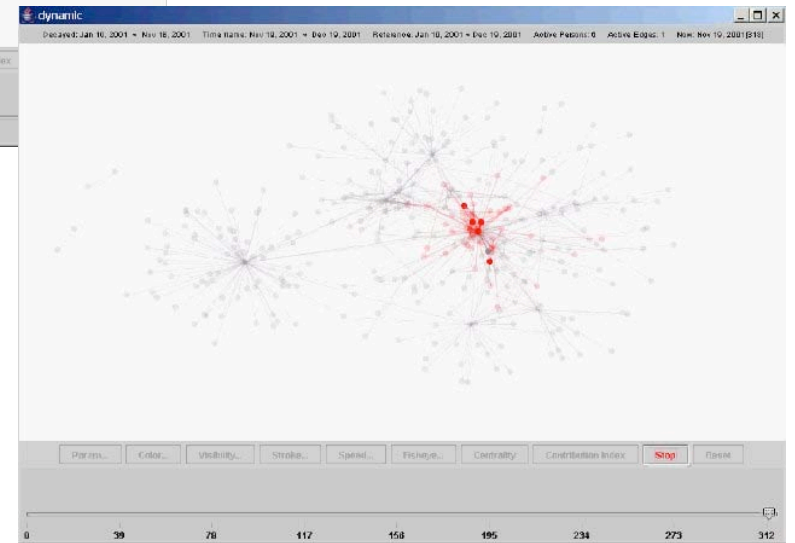
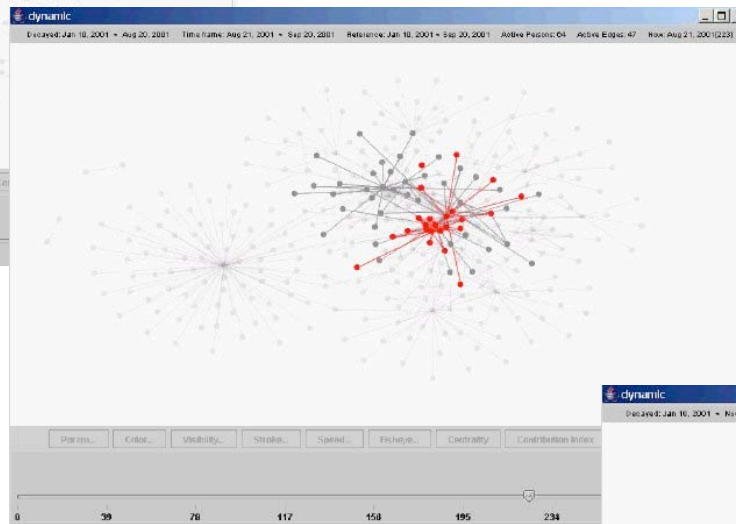
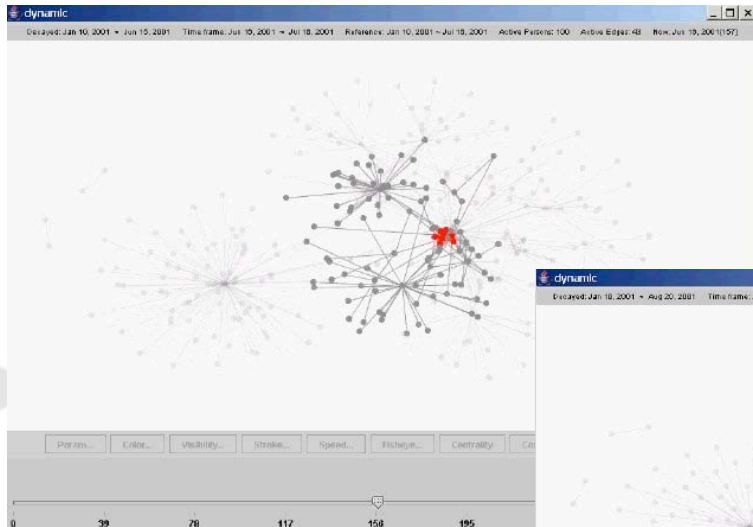
with history



without history



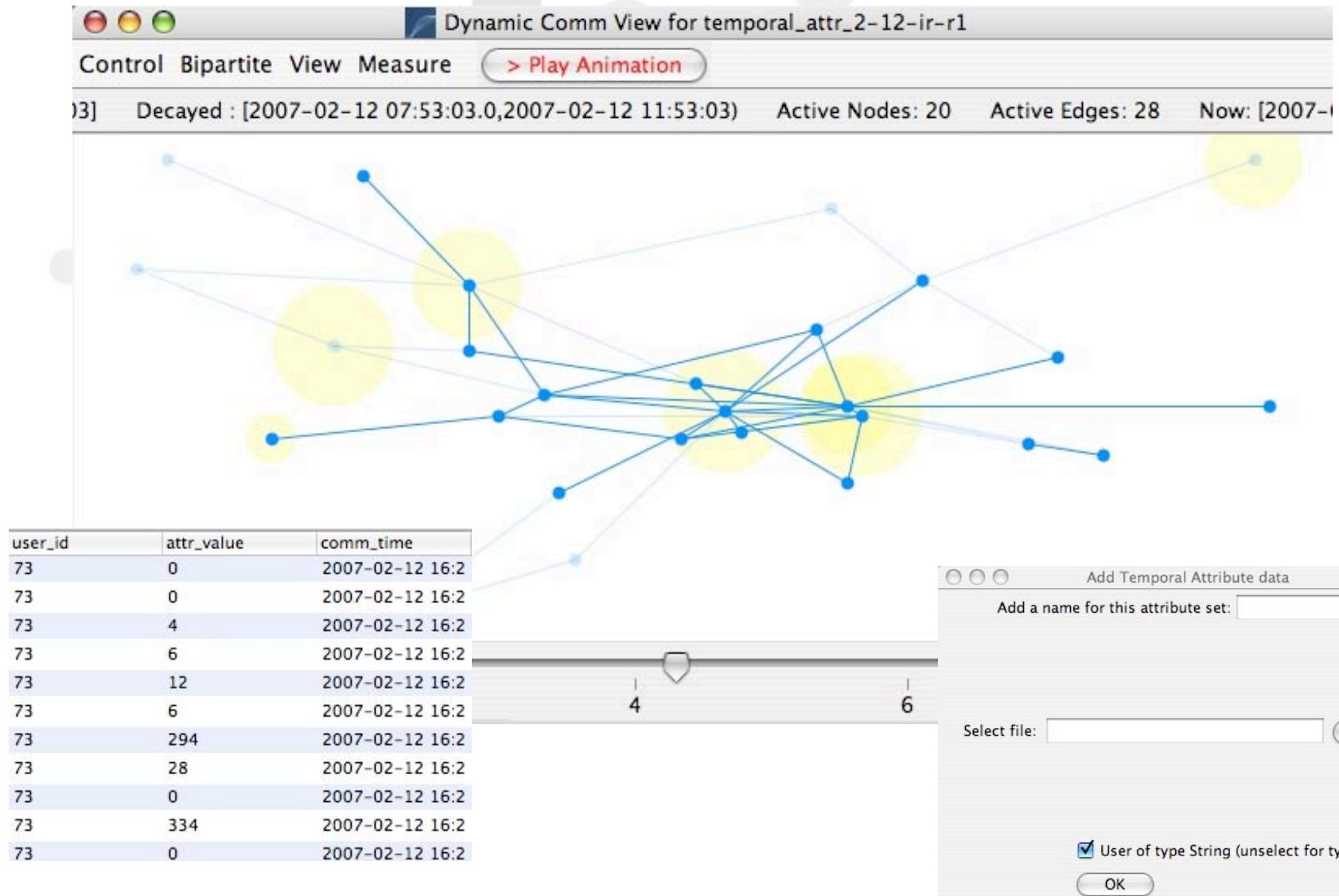
Discover COIN/CLN/CIN



Advanced Feature list

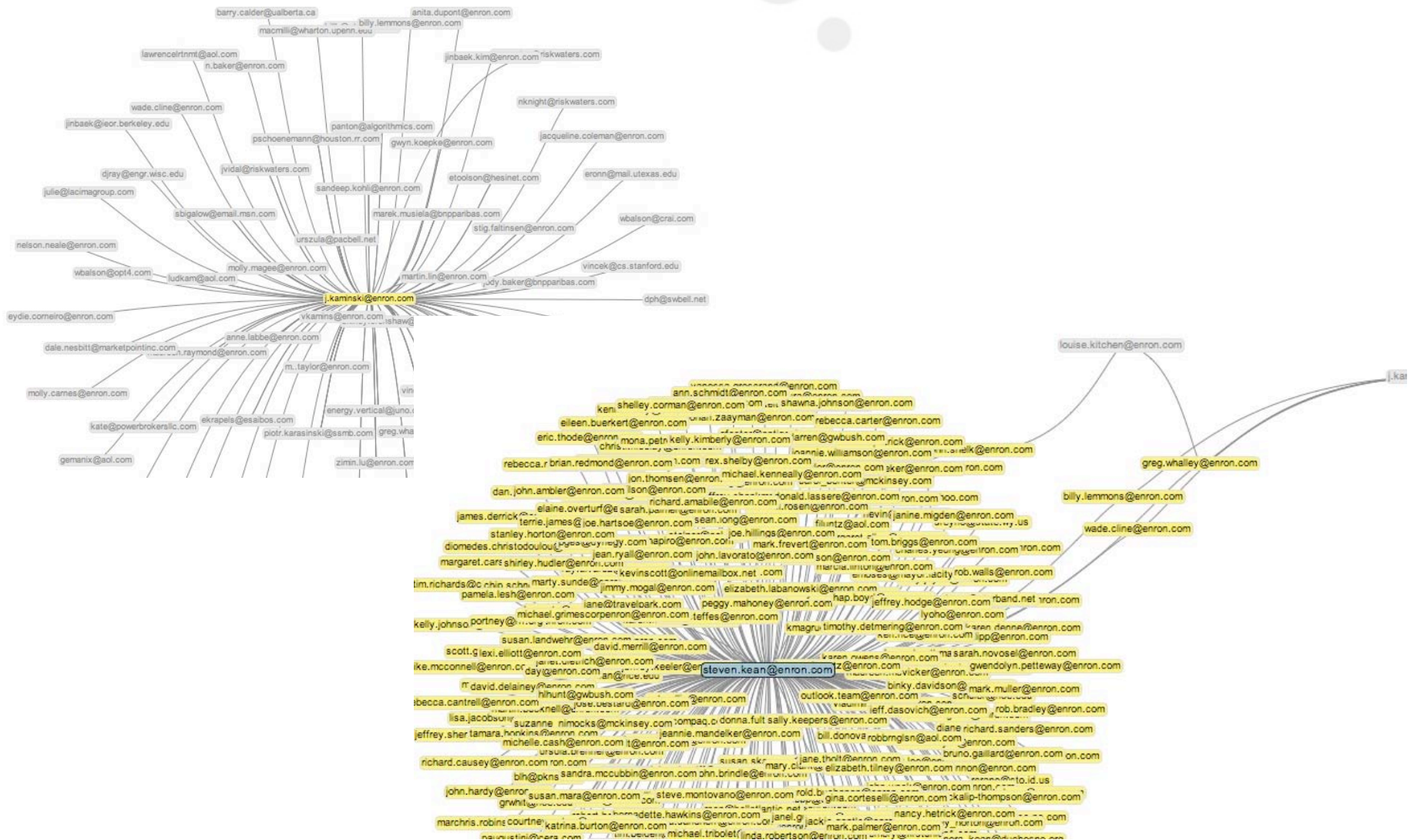
- Adding temporal attributes
- Local view of context (nearest neighbors view)
- Google Blog parser
- Google Scholar parsing
- Google News parsing
- Term View
- galaxymap
- Exporting functions (awvci)
- ART (average response time)

Adding temporal attributes



Local view of context (alpha)

Local Context Comm View for enron_small



Google Blog Collector

Database Help

API

☒ use Google Blog Search
☐ use MSN Live WebSearch
☐ use Google News Search
☐ use Google Scholar Search

Query

Text Query
- for exact phrase search, enclose queries with "..." (e.g.: "coca c
- separate multiple queries with commas (e.g.: coca cola, virgin
barack obama, hillary clinton, john mccain, john edwards, rudy

URL Query
- enter the websites you want to search (e.g.: http://www.cocac
- separate multiple queries with commas

Options

Results of Top Degree of Separation :

Connected to database localhost:pgtest . Database of type WEBSITE .

Connected to database "pgtest"

Preview results for query barack

Select relevant results for query barack obama (20 results in total):

☐

1. **Barack Obama** is no JFK (<http://www.punditreview.com/2007/07/19/barack-obama-is-no-jfk/>)
When I see stories like these two below about **Barack Obama** and the other Dems running for president, I can't help but think that they have no ide

☒

2. **Barack Obama**, sex-ed, and kindergarteners (<http://www.thecarpetbaggerreport.com/archives/11513.html>)
Sex education for kindergarteners had become an issue in his race against Keyes because of **Obama's** work on the issue as chairman of the health
EDT 2007)

☒

3. Dinner with **Barack** (http://my.barackobama.com/page/-/podcasts/20070718_Podcast2.mp4)
On July 14th, four supporter sat down and shared an intimate dinner and conversation with **Barack Obama** -- watch what happened and see **Barack** i

☒

4. How Democratic Elite, Not **Barack Obama**, Are Driving Down the ... (<http://thinkonthesethings.wordpress.com/2007/07/19/how>
Barack Obama is not on some mission to avoid calling himself a Democrat. Anyone who has read The Audacity of Hope knows this. He has also given

☒

5. IMG_0542.jpg (<http://www.flickr.com/photos/barackobamadotcom/854729426/>)
Barack Obama posted a photo.: IMG_0542.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

6. IMG_0536.jpg (<http://www.flickr.com/photos/barackobamadotcom/854725070/>)
Barack Obama posted a photo.: IMG_0536.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

7. IMG_0551 (<http://www.flickr.com/photos/barackobamadotcom/854735332/>)
Barack Obama posted a photo.: IMG_0551.(Thu Jul 19 18:01:05 EDT 2007)

☒

8. IMG_0545.jpg (<http://www.flickr.com/photos/barackobamadotcom/854731214/>)
Barack Obama posted a photo.: IMG_0545.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

9. IMG_0539.jpg (<http://www.flickr.com/photos/barackobamadotcom/853868677/>)
Barack Obama posted a photo.: IMG_0539.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

10. IMG_0547.jpg (<http://www.flickr.com/photos/barackobamadotcom/854733076/>)
Barack Obama posted a photo.: IMG_0547.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

11. IMG_0530.jpg (<http://www.flickr.com/photos/barackobamadotcom/854721082/>)
Barack Obama posted a photo.: IMG_0530.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

12. IMG_0522.jpg (<http://www.flickr.com/photos/barackobamadotcom/853857871/>)
Barack Obama posted a photo.: IMG_0522.jpg.(Thu Jul 19 18:01:05 EDT 2007)

☒

13. IMG_0538.jpg (<http://www.flickr.com/photos/barackobamadotcom/854726090/>)
Barack Obama posted a photo.: IMG_0538.jpg.(Thu Jul 19 18:01:05 EDT 2007)

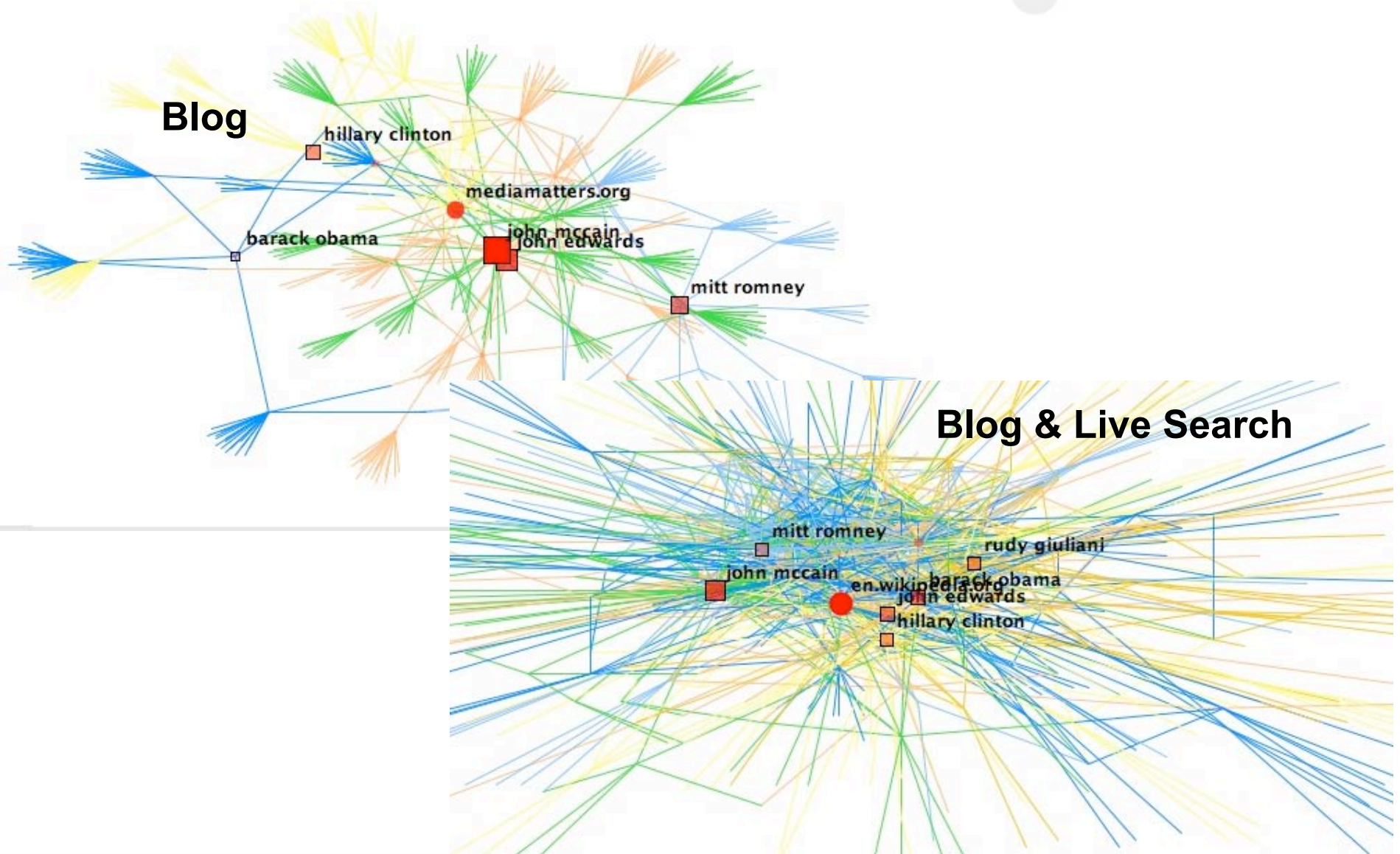
☒

14. Charleston Interns (<http://www.flickr.com/photos/barackobamadotcom/853873657/>)
Barack Obama posted a photo.: Charleston Interns.(Thu Jul 19 18:01:05 EDT 2007)

☒

15. IMG_0535.jpg (<http://www.flickr.com/photos/barackobamadotcom/854723852/>)
Barack Obama posted a photo.: IMG_0535.jpg.(Thu Jul 19 18:01:05 EDT 2007)

Google Blog Parser



Term view index

Term processing

☒ Use **existing index** (uncheck to create a new index)

Terms Selection Strategy

☒ Preview terms

☐ Upload list of terms

Indexing

☐ Include **numbers** (whether numbers should be indexed)

☒ Use Porter **stemming** (whether words should be reduced to their stem)

☐ Set constraints on document date

Stopwords

Stopwords are highly frequent words like "the" or "for". Stopwords are skipped during...

☐ use **minimal** stopwords list

☒ use **default** stopwords list

☐ use **extended** stopwords list

☐ use **blog** stopwords list

☐ use **custom** stopwords list

(one word or phrase per line, no commas, lowercase words)

Term view index - 2

Term Selection

This dialog lets you select which terms will be used for the term view
Click the words or phrases from the left column to add it to the right selection

Add Words
word (norm. frequency) (click to add)

- v6.00.2800.1165 (56.0)
- mimeole (56.0)
- priority (56.0)
- microsoft (56.0)
- msmail (56.0)
- charset (32.0)
- iso-8859-1 (32.0)
- http (29.0)

Add 10 from the top

Add Phrases
phrase (norm. frequency) (click to add) -->

- microsoft before (56.0)
- microsoft very (56.0)
- msmail before (56.0)
- msmail very (56.0)
- charset very (32.0)
- charset before (32.0)
- http before (29.0)
- http very (29.0)

Add 10 from the top

Selected Terms
(click item to remove)

- produced (56.0)
- survey (36.0)
- thomas (31.0)
- mobile (30.0)
- deloitte (28.0)
- nico (28.0)
- ckn (28.0)
- consulting (28.0)
- sent (26.0)
- schmalberger (26.0)
- web (26.0)
- please (26.0)
- sent before (26.0)
- sent very (26.0)
- web before (26.0)
- web very (26.0)
- before mailing (26.0)
- very mailing (26.0)
- 2002 (25.0)
- thurgauerstrasse (24.)
- converted (24.)
- pgloor (24.)
- original (24.)
- zurich (24.)
- users (24.)
- boundary (24.)
- people (23.0)

☒ sort terms by norm. frequency
☐ sort terms by name

☒ Factor actor's weight in when computing communication weight
☒ Index dates (necessary for dynamic termview)

Keep top % of cooc. 0.5

All your previous terms selections are automatically stored.
Use previous selections

Only Export selected terms

OK, collect communications ...

Exporting Functions

Average weighted variance of contribution index (AWVCI),
to calculate the balancedness of a team's communication.
See Gloor et Al, 2006, "Finding Collaborative Innovation Networks 1
Correlating Performance with Social Network Structure"

```
*****  
AWVCI = 0.3710078  
For 21 chars on DAY 0 to 114.  
*****
```

day	m	var(CI)
0	2	2.0
1	4	0.25333333
2	9	0.1729025
3	9	0.11303161
4	9	0.11303161
5	9	0.12104606
6	9	0.12529045
7	9	0.12529045
8	9	0.12529045
9	9	0.12529045
10	9	0.12529045
11	9	0.10454648
12	9	0.114305556
13	9	0.114305556
14	9	0.11735481
15	9	0.11735481
16	9	0.11735481
17	9	0.12942052
18	9	0.12942052

Individual BC, timeperiod from 0 to 114 [use tab as delimiter].
athey robin -1.0000 -1.0000 0.0000 0.1071 0.1071 0.1071 0.1071
barber meshel -1.0000 -1.0000 -1.0000 -1.0000 -1.0000 -1.0000 -1.0000
bier jessica l -1.0000 -1.0000 0.2500 0.1071 0.1071 0.1071 0.1071

Minimum value, timeperiod from 0 to 114 [use tab as delimiter].
athey robin -1.0000
barber meshel -1.0000
bier jessica l -1.0000

Maximum value, timeperiod from 0 to 114 [use tab as delimiter].
athey robin 0.2662
barber meshel 0.0000
bier jessica l 0.5823

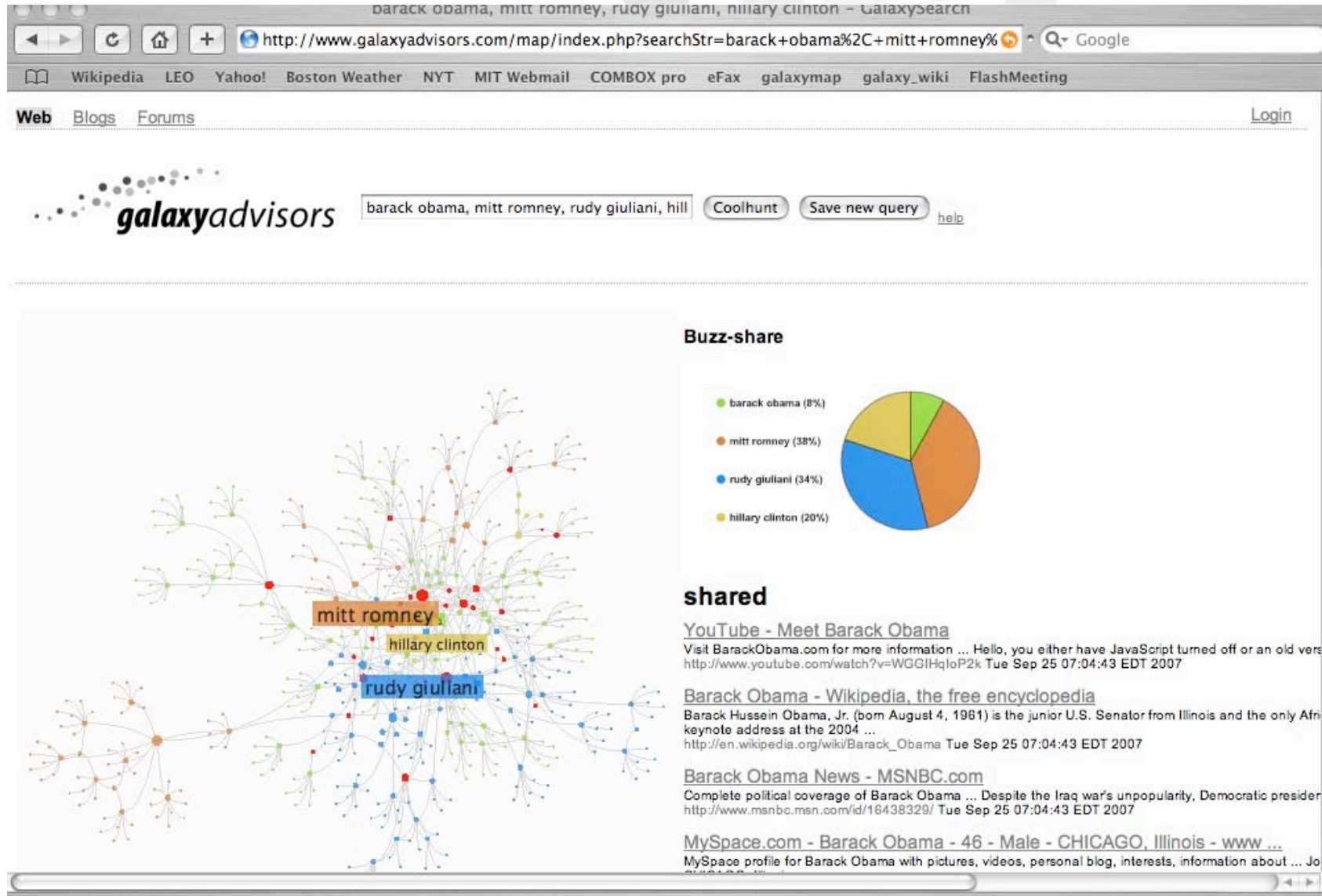
Mean, timeperiod from 0 to 114 [use tab as delimiter].
athey robin 0.0952
barber meshel -0.2000
bier jessica l 0.2947

Standard dev, timeperiod from 0 to 114 [use tab as delimiter].
athey robin 0.1509
barber meshel 0.4018
bier jessica l 0.2144

Deviation 1st degree, timeperiod from 0 to 114 [use tab as delimiter].
athey robin -1.0000 0.0000 1.0000 0.1071 0.0000 0.0000
barber meshel -1.0000 0.0000 0.0000 0.0000 0.0000 0.0000
bier jessica l -1.0000 0.0000 1.2500 -0.1429 0.0000 0.0000

Deviation 2nd degree, timeperiod from 0 to 114 [use tab as delimiter].
athey robin -1.0000 1.0000 1.0000 -0.8929 -0.1071 0.0000 0.0000

Condorview (galaxymap)



Wish list

To come

- Weighting by betweenness for term view
- Sentiment analysis
- Universal data model
- Load Wikipedia, US patent database,

Privacy is a primary concern

- Buy-in of study population crucial
- Identity of individuals has to be protected
 - Replace names by identifiers
 - Results only reported in anonymous form
- Proper measures to safeguard data
 - Only anonymized data goes outside of company
 - Content is encrypted
- Open & honest communication with users
 - Share results with study population