Condor
Tips & Tricks

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

Obtain academic version from

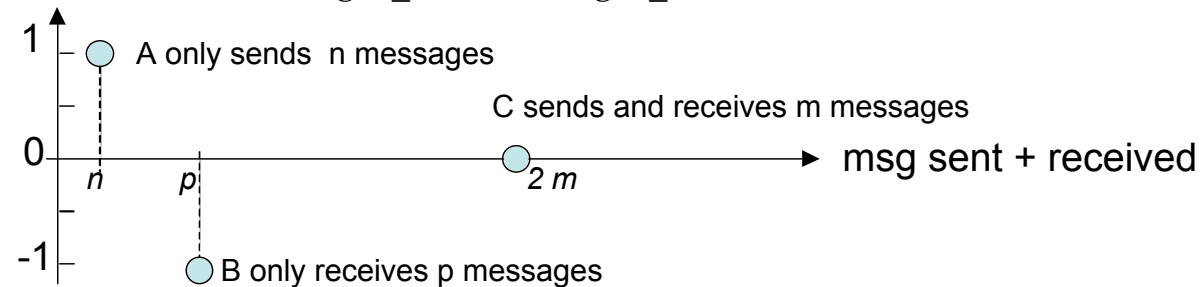- http://tiger.soberit.hut.fi/virtualbrownbag
- Need to register first
- Go to file galleries
  - Condor
  - Condor support tools
- Needs monthly authorization code, obtainable from pgloor@mit.edu

Condor Components

With history:

first day — day *d* (current day) — day *(d+n)* — last day — time

Time Frame size *n*
Dimmed out
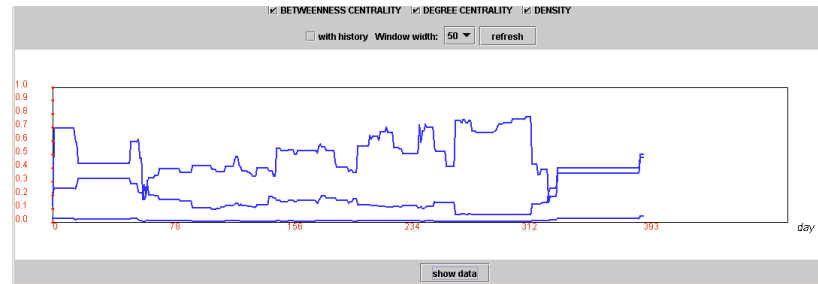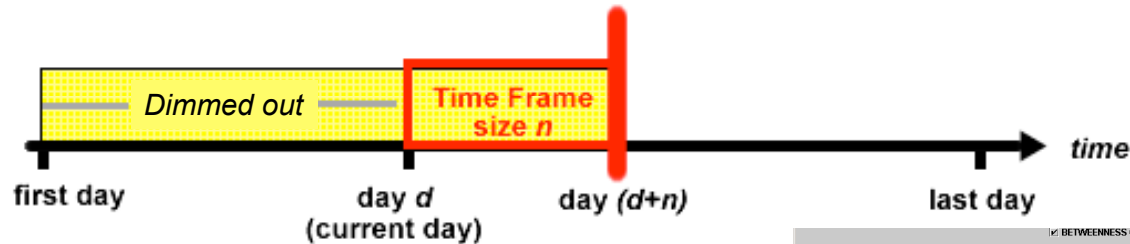
Contribution Index $\dfrac{messages\_sent - messages\_received}{messages\_sent + messages\_received}$

A only sends n messages

C sends and receives m messages

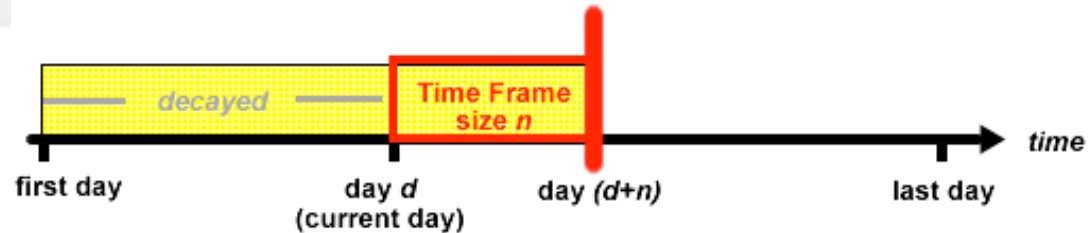msg sent + received

B only receives p messages

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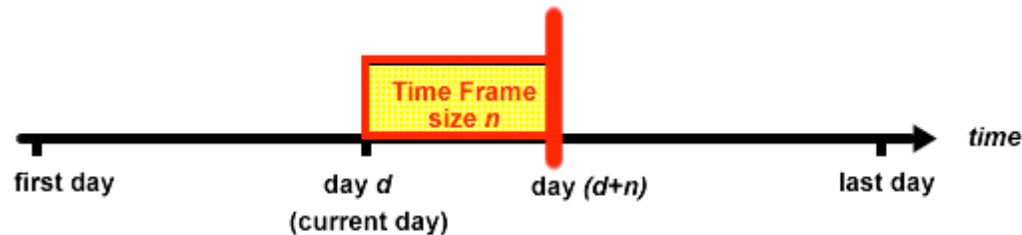jamie,ford@dartmouth.edu

devlab@cs.dartmouth.edu

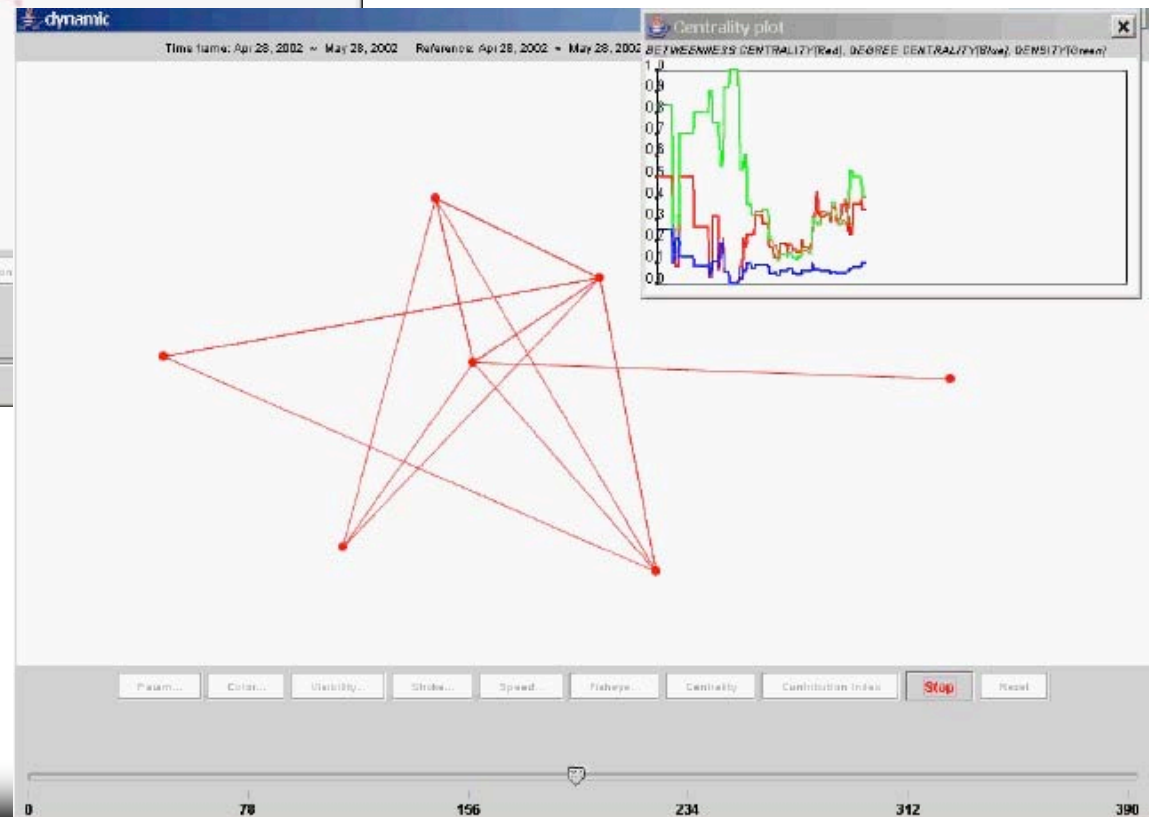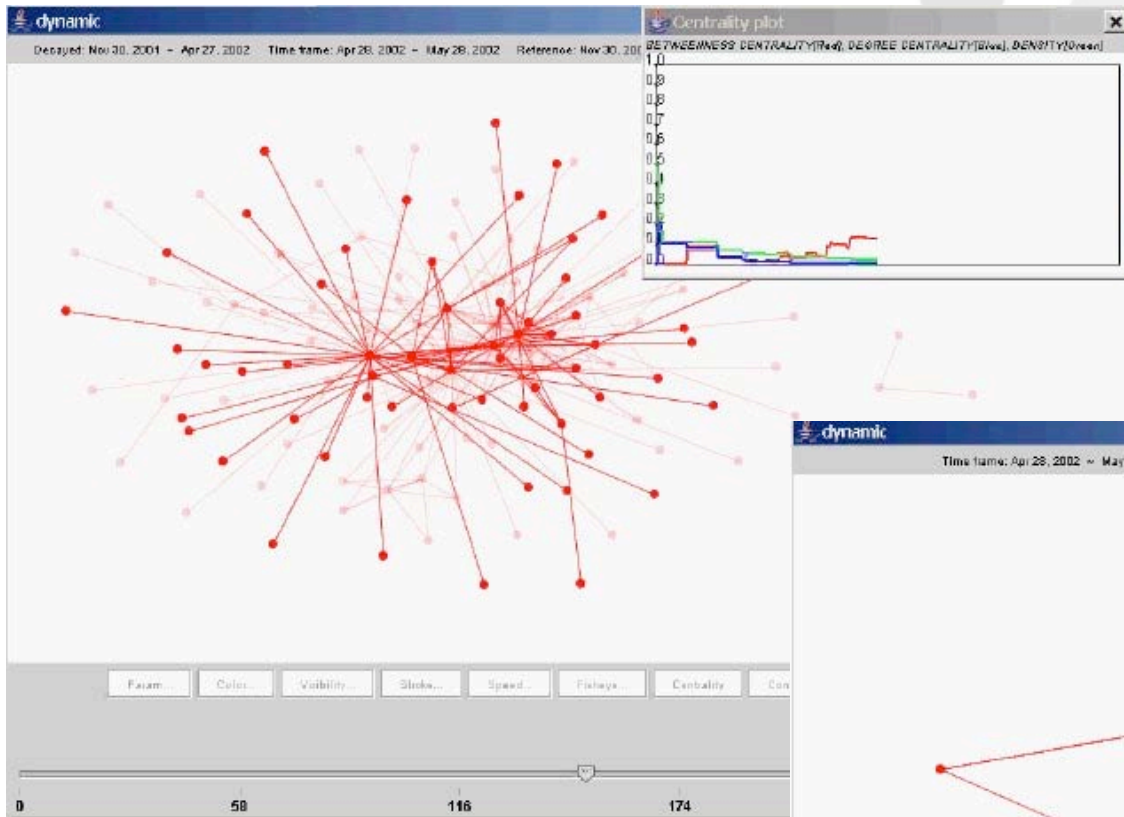# Temporal Visualization by a Sliding Time Frame

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

RSS Feeds

Web (Google API)

E-Mail

unstructured documents

CSV Files
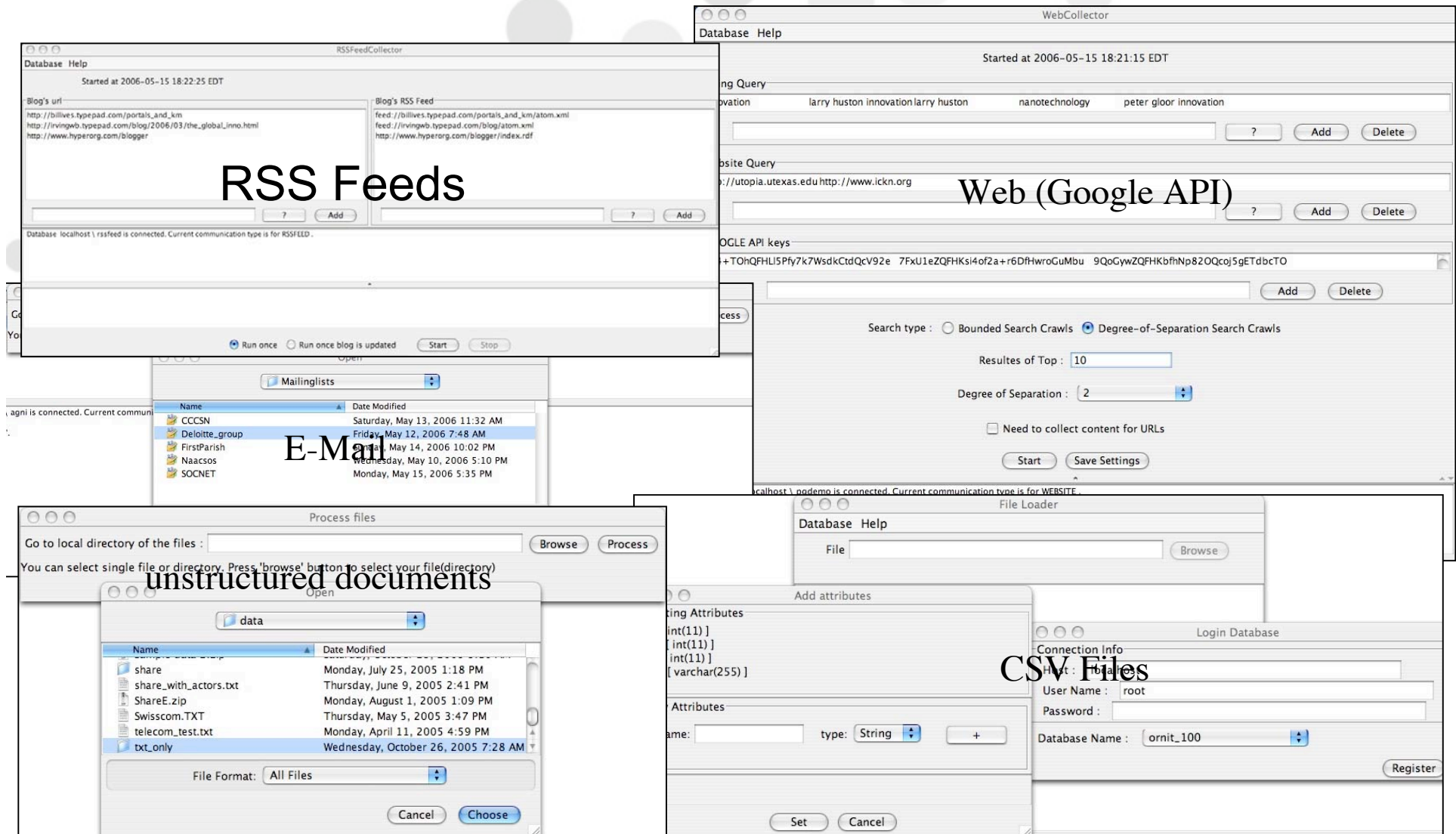
Condor database structure

datasetId	mailAddress	organization	domain
3	schwartz jeff	UNKNOWN	us – atlanta
3	indorf kaplan nalir	UNKNOWN	us – new york
4	lambert 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: Invitatior	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

- chars
- comm
- comm_target
- datasets
- name_change

datasetId	dataset
1	webinar
2	firstpa
3	combId

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

5 types of databases

chars
comm
datasets
lookup
name_change

datasets
items
name_change
nbcfreq
rss_links
stopterms1

chars
comm
comm_target
datasets
name_change

- Fileloader
- E-mail
- Web Sites
- Blogs
- Documents

datasets
doc_comm1
docs
nbcfreq
stopterms1

apikeys
chars
datasets
links
name_change
nbcfreq
stopterms1
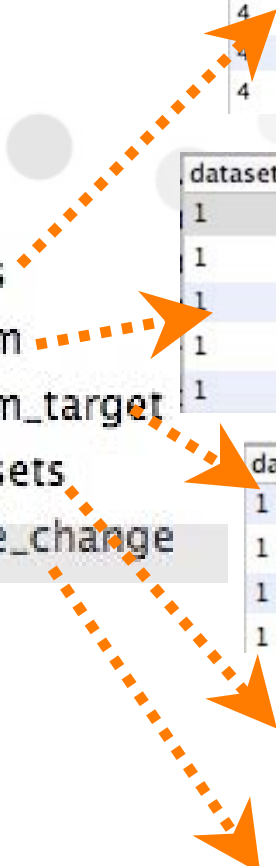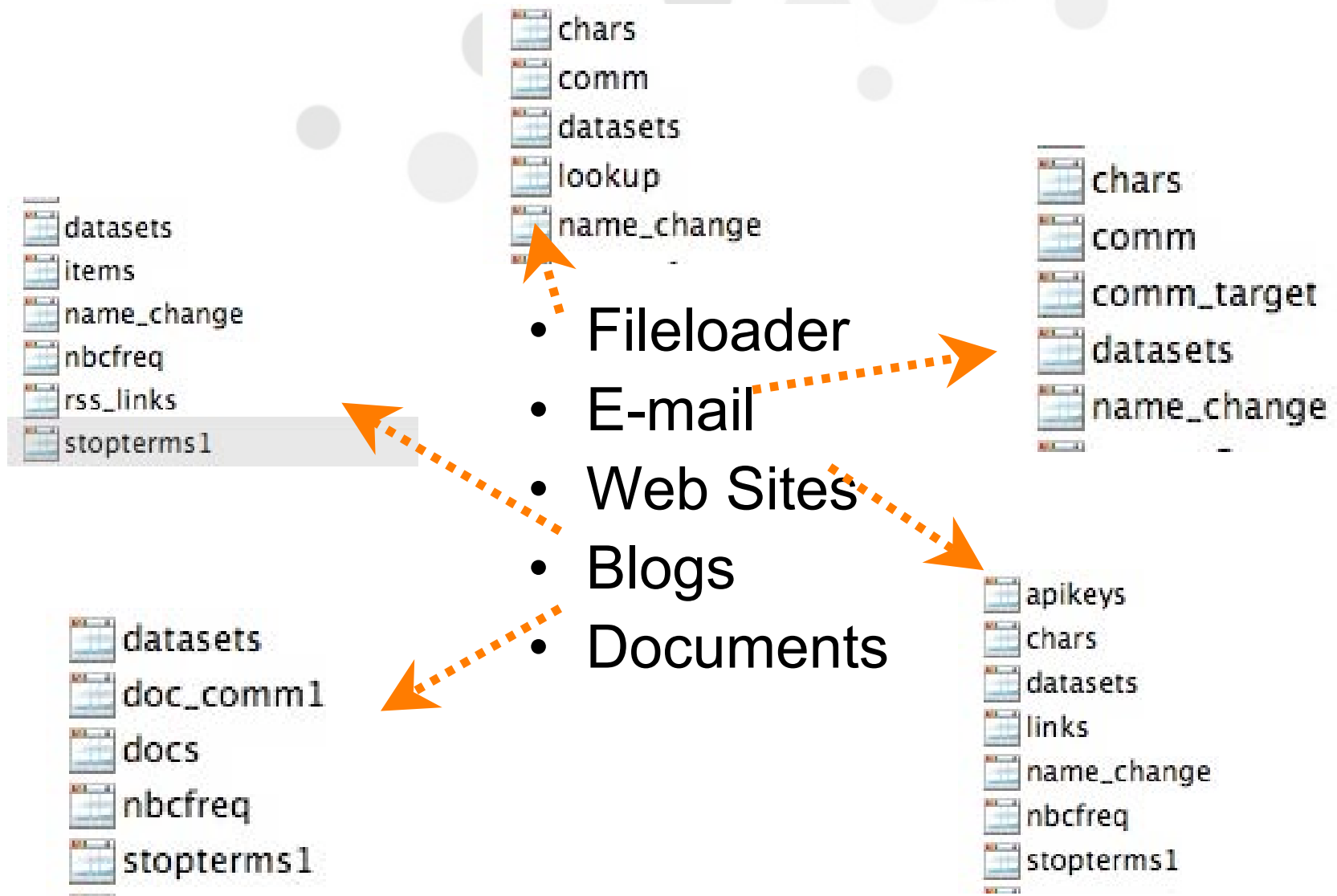
Condor Scalability

- 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:
                    time since: end time: 2004   Jan   1

all contents   – or –   with keyword(s):

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

2. Merge addresses (uses less memory)
3. Reduce communication frequency
4. Run without history

- Reducing number of actors saves most memory
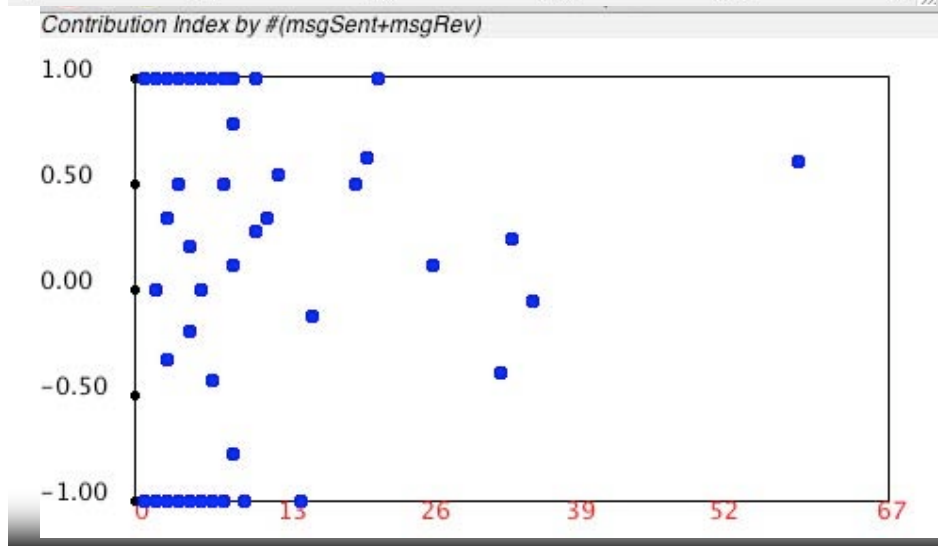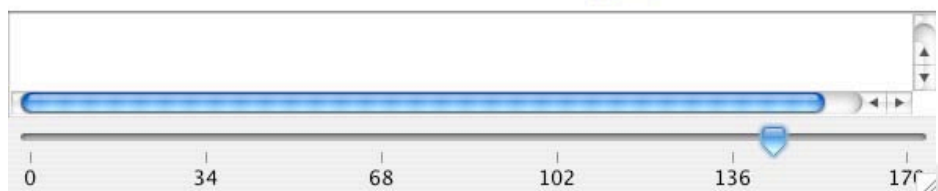
Delete mailing list addresses

Static for T_CDT_Documentation component developer discussions

Control View

Color by dataset   0 nodes and 42 edges

Static for T_CDT_Documentation component developer discussions

Control View

Color by dataset   0 nodes and 5 edges
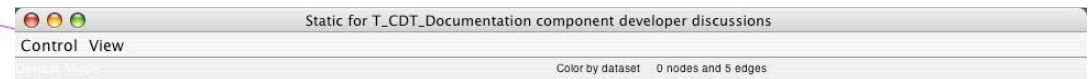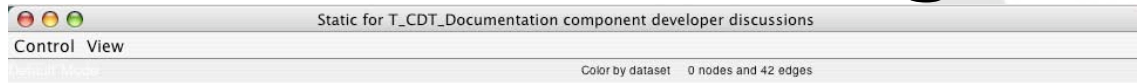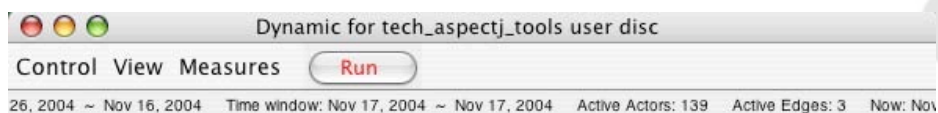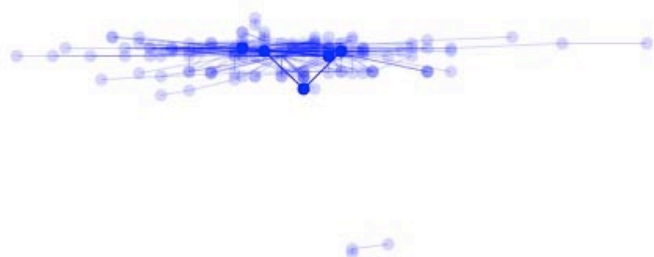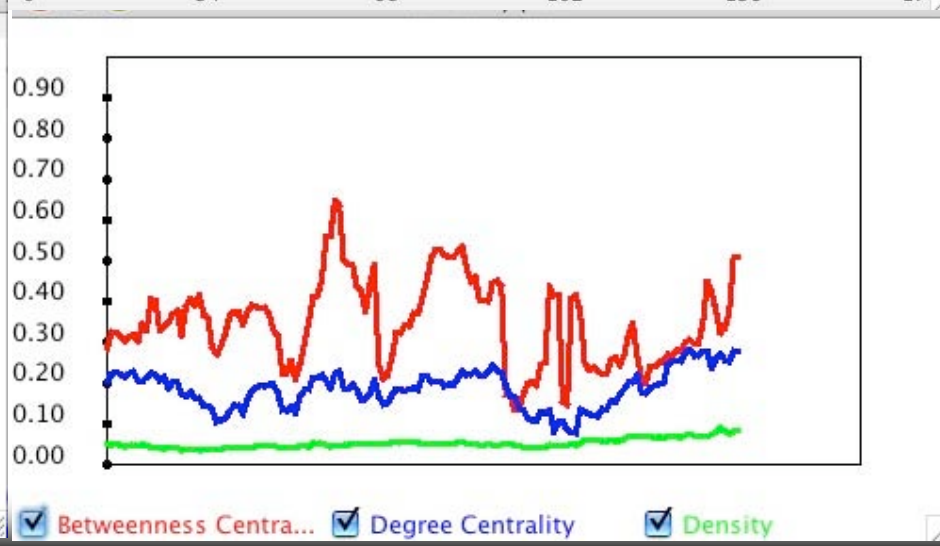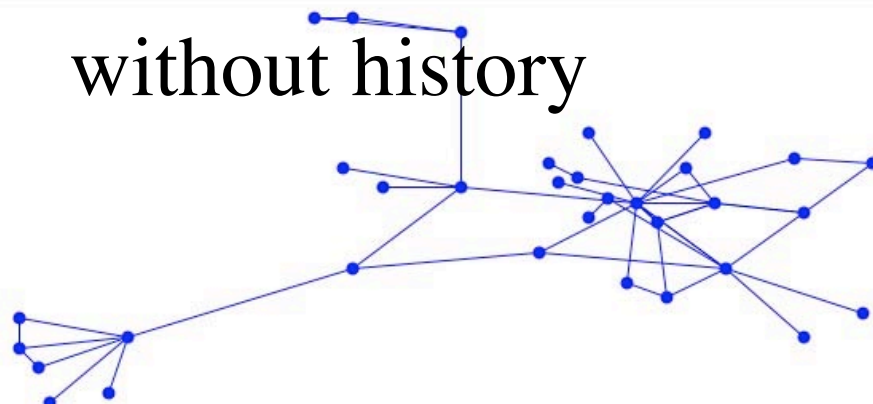
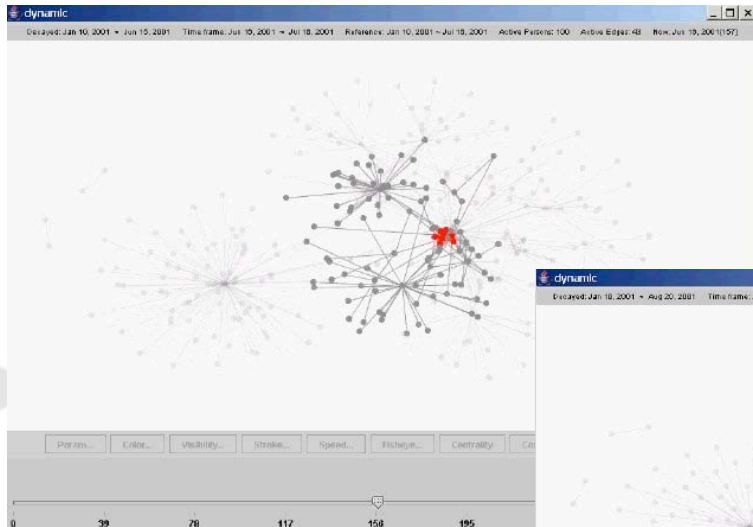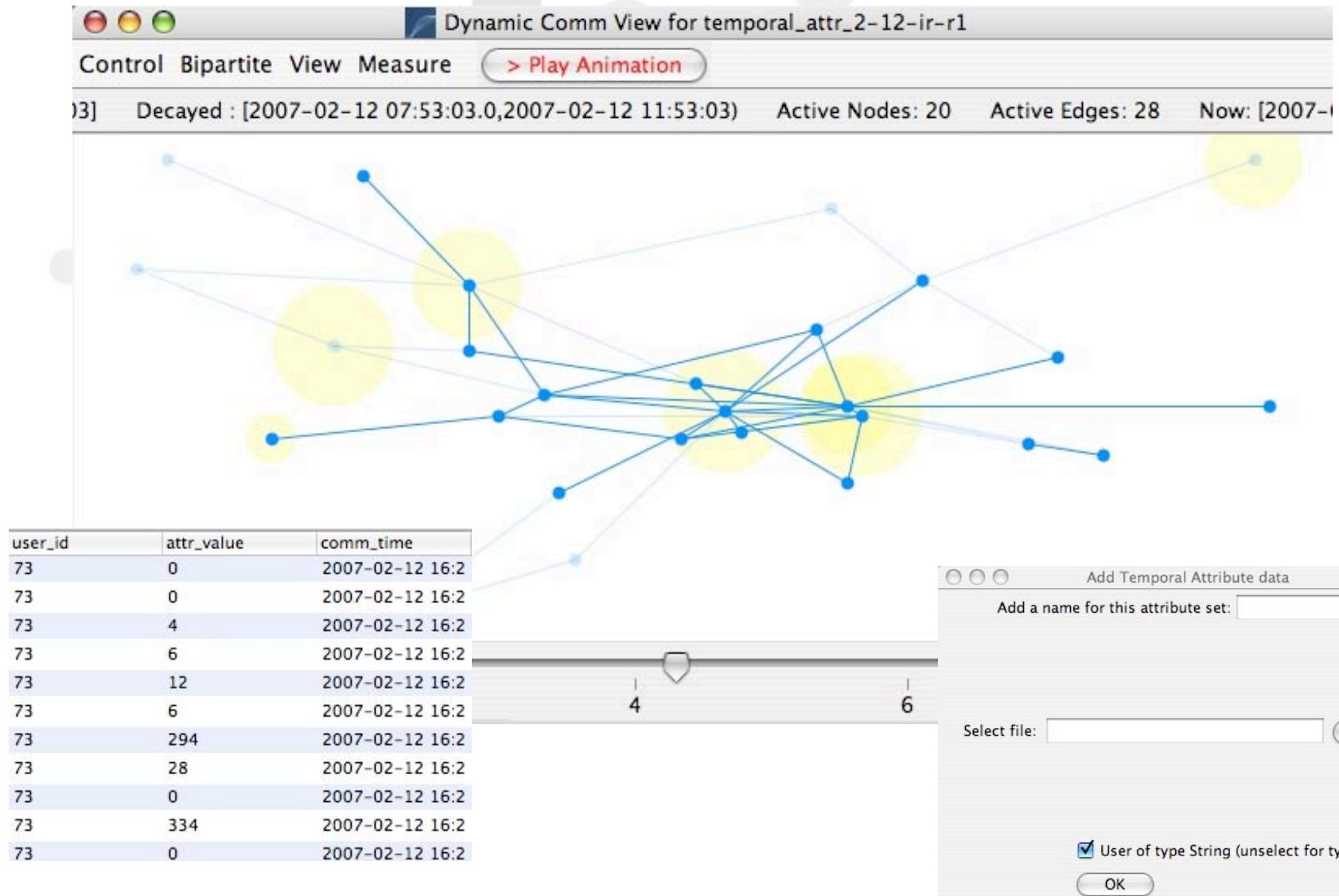# 2. Discover COINs
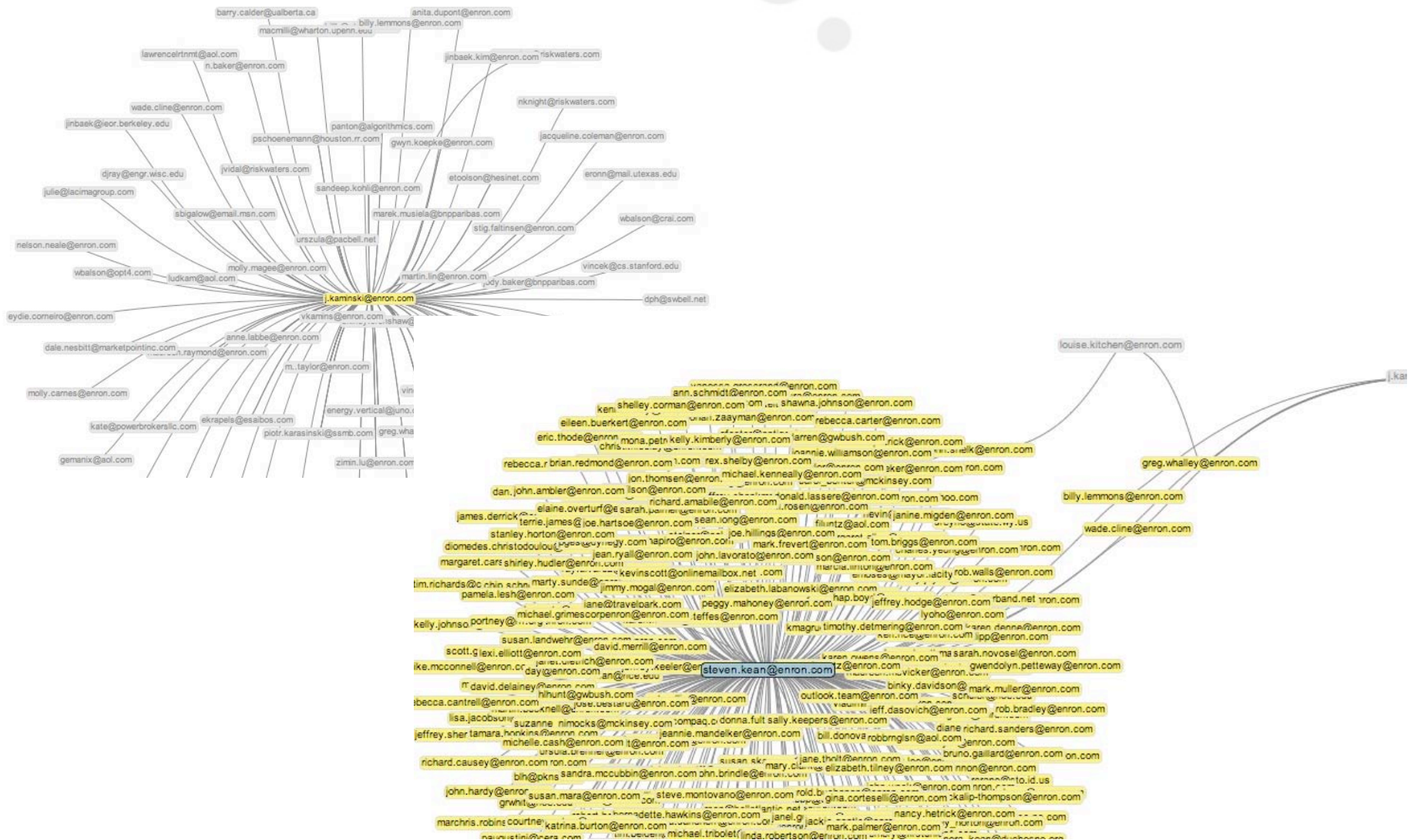
Discover COIN/CLN/CIN
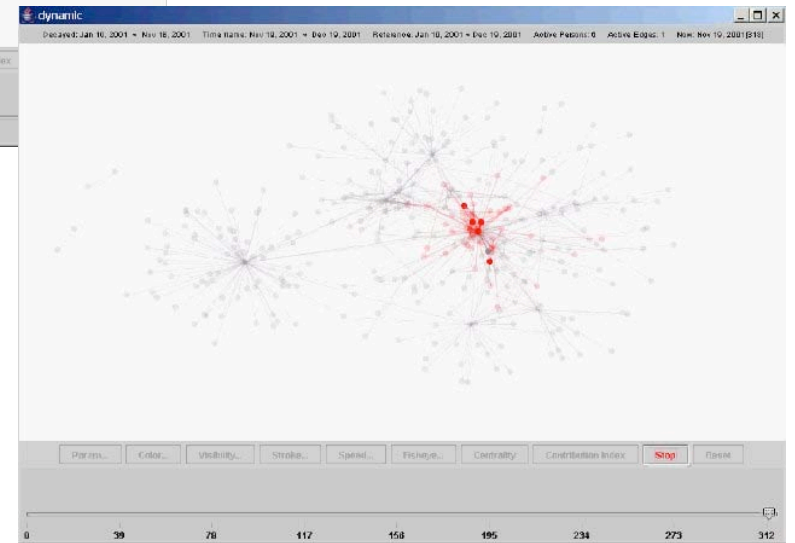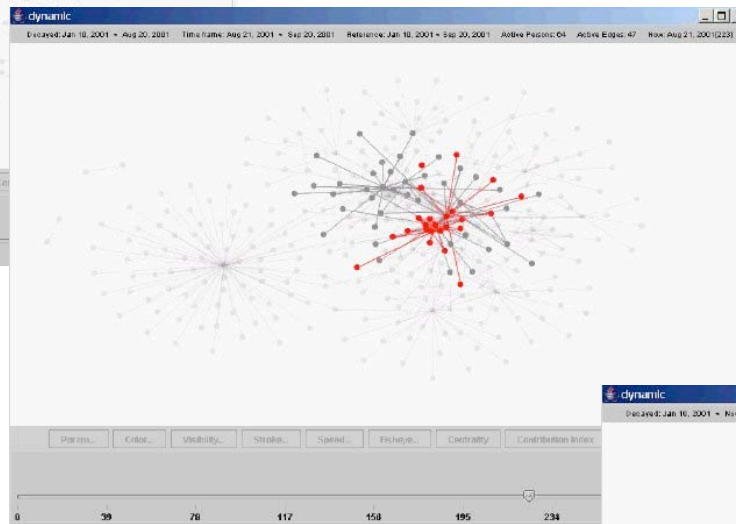
Advanced Feature list

- 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)

Google Blog Collector

Google Blog Parser

Term view index

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)

Add your own term

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.)*

○ 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 balanced'ness 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
```
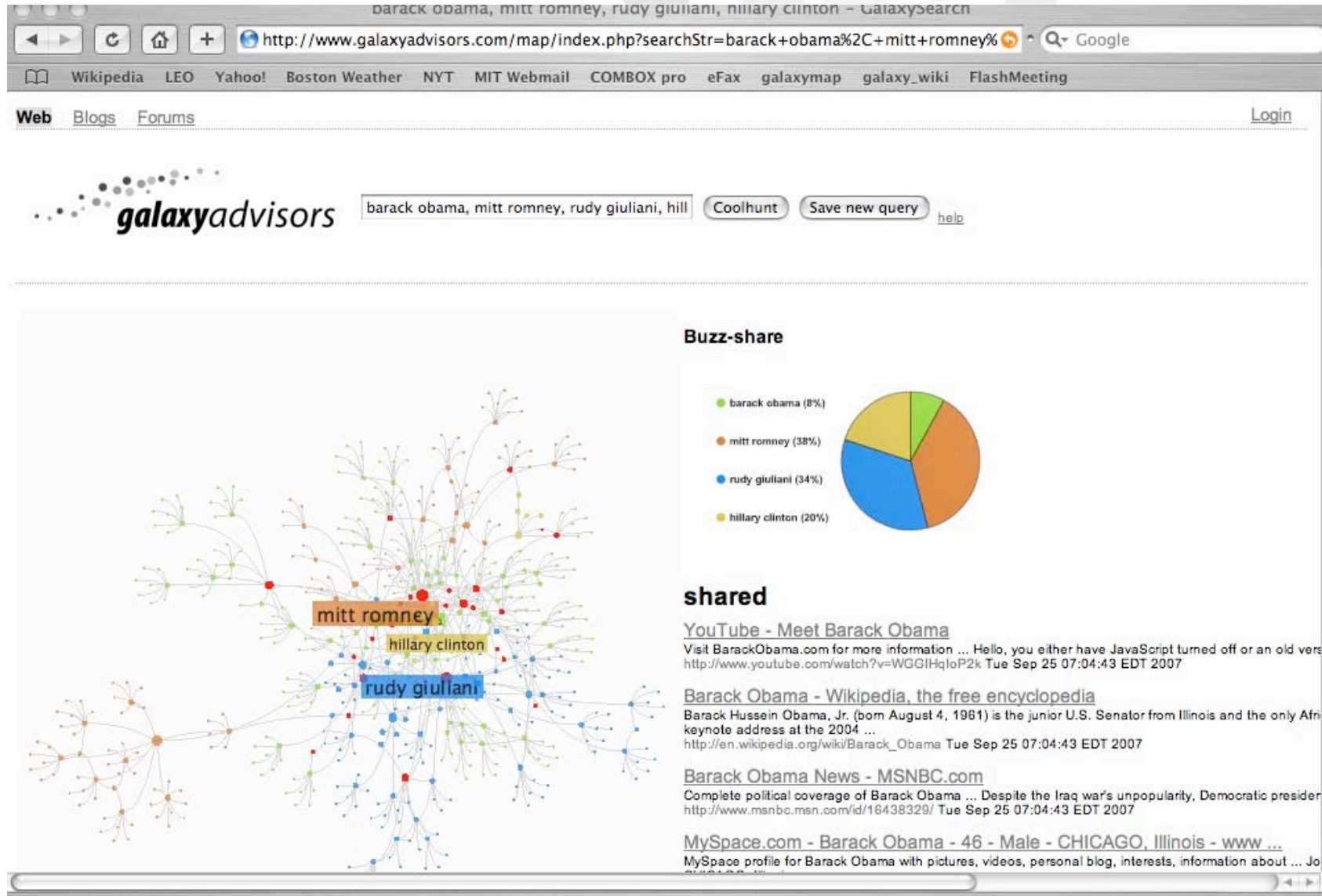
Condorview (galaxymap)

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