

Analyzing Trends in Science & Technology Innovation

Cody Dunne, Pengyi Zhang, Chen Huang, Jia Sun,
Ben Shneiderman, Ping Wang & Yan Qu

{cdunne, ben}@cs.umd.edu

{pengyi, chuang, jsun, pwang, yanqu}@umd.edu

<http://stick.ischool.umd.edu>

28th Annual Human-Computer Interaction Lab Symposium

May 25-26, 2011

College Park, MD

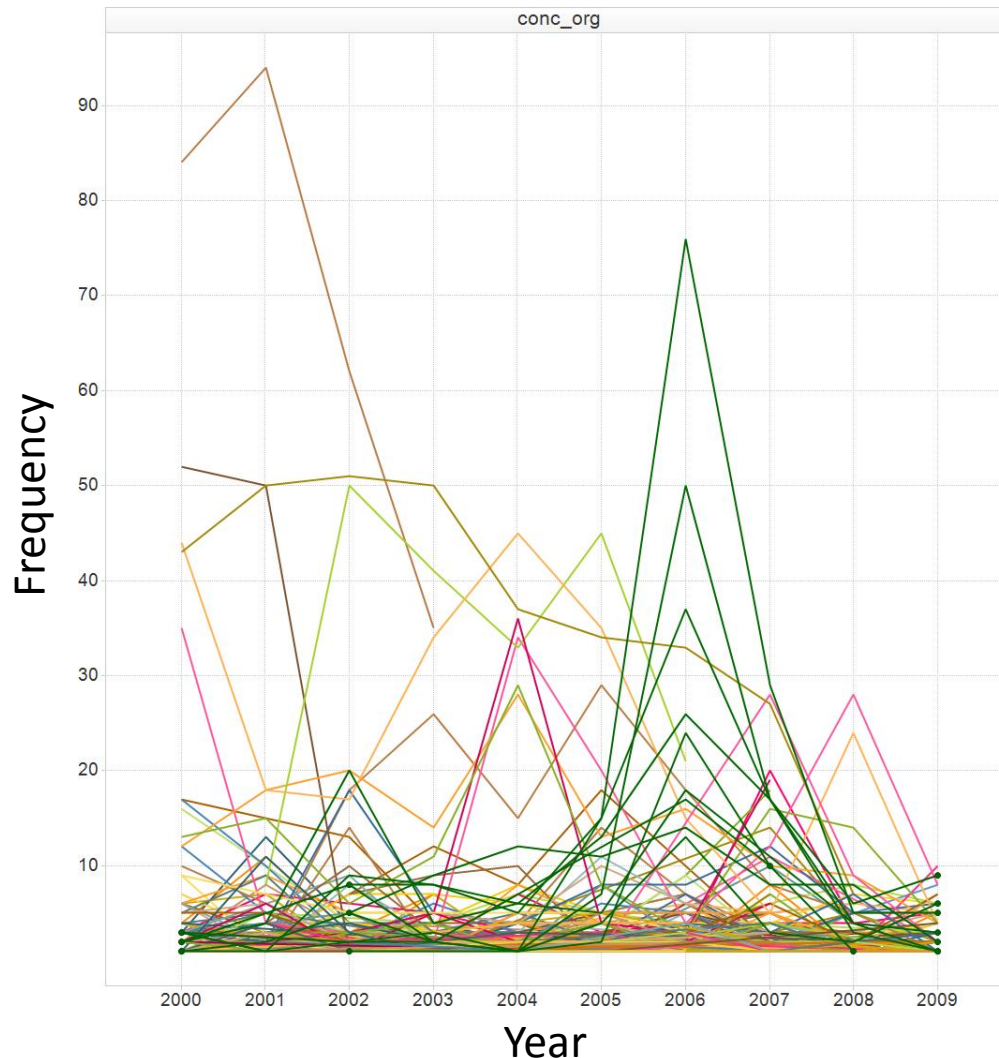


UNIVERSITY OF
MARYLAND



Business Intelligence 2000-2009

2006 Peak: Concept-Entity Co-Occurrence



Data Mining

- National Security Agency
- NSA
- White House
- FBI
- AT&T
- American Civil Liberties Union
- Electronic Frontier Foundation
- Dept. of Homeland Security
- CIA



Business Intelligence 2000-2009 Matrix showing Co-Occurrence of concepts and entities

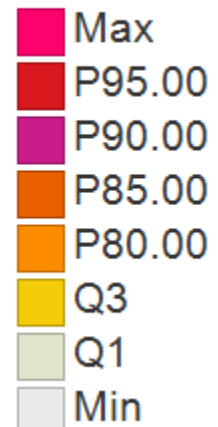
Colors:

- Max
- P95.00
- P90.00
- P85.00
- P80.00
- Q3
- Q1
- Min



Business Intelligence 2000-2009: (subset)

Colors:





Business Intelligence 2000-2009: Data mining

- NSA
- CIA
- FBI
- White House
- Pentagon
- DOD
- DHS
- AT&T
- ACLU
- EFF
- Senate Judiciary Committee



Business Intelligence 2000-2009:

Tech1

- Google
- Yahoo
- Stanford
- Apple

Tech2

- IBM, Cognos
- Microsoft
- Oracle

Finance

- NASDAQ
- NYSE
- SEC
- NCR
- MicroStrategy



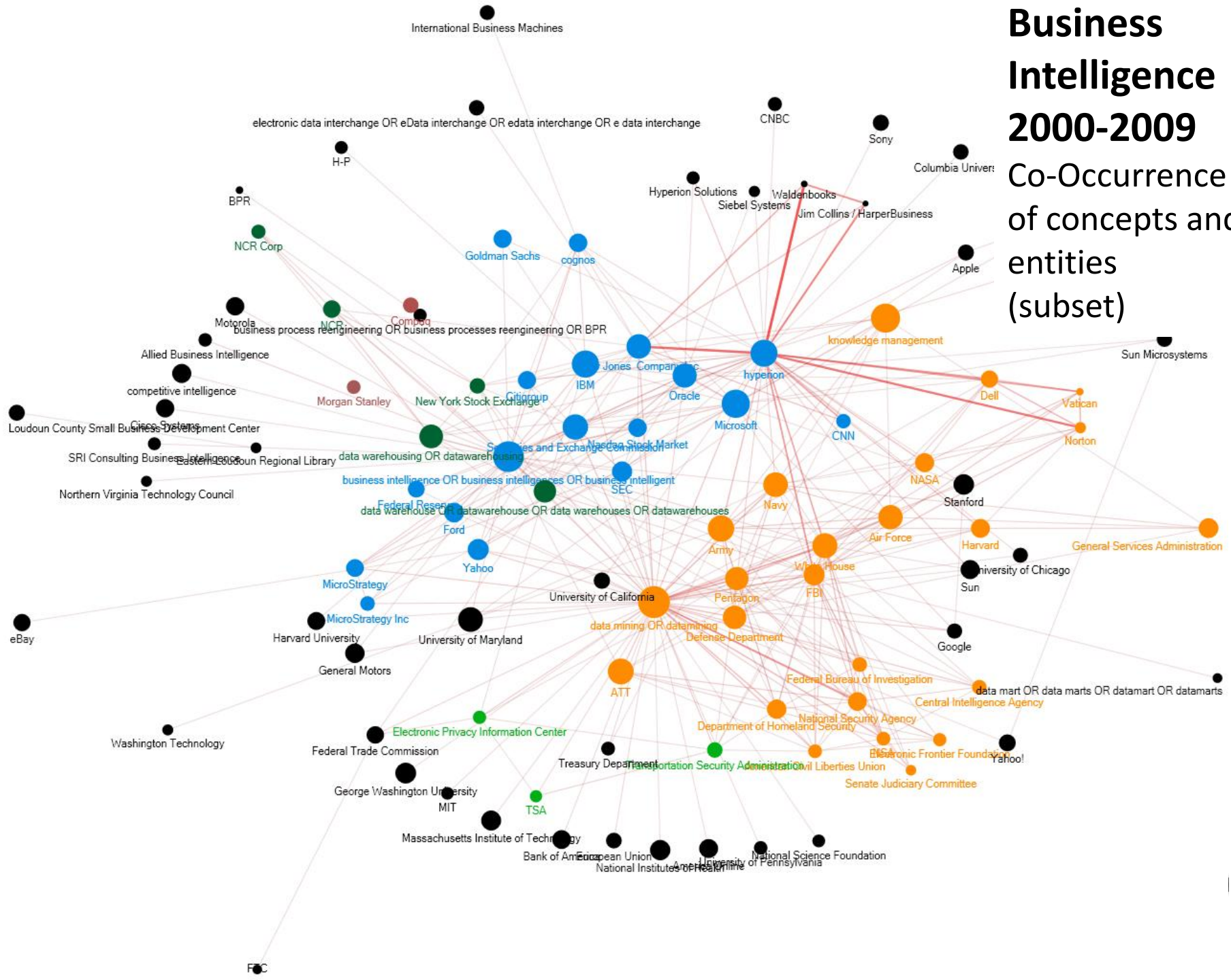
Business Intelligence

2000-2009:

- Air Force
- Army
- Navy
- GSA
- UMD*



Co-Occurrence of concepts and entities (subset)



The STICK Project

- NSF SciSIP Program
 - Science of Science & Innovation Policy
 - Goal: Scientific approach to science policy
- The STICK Project
 - Science & Technology Innovation Concept Knowledge-base
 - Goal: Monitoring, Understanding, and Advancing the (R)Evolution of Science & Technology Innovations

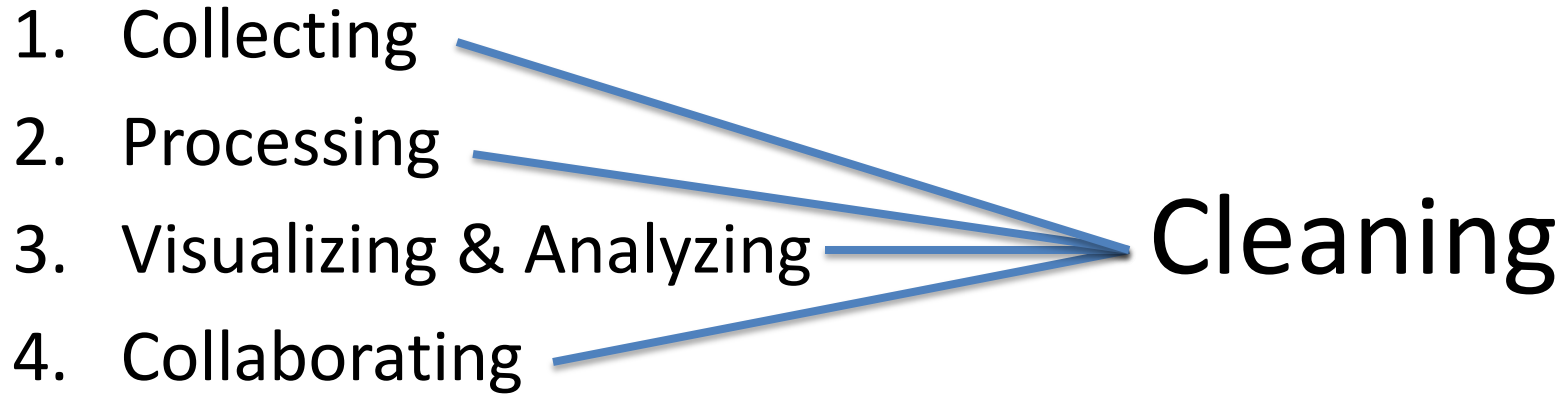


STICK Contribution

- **Scientific**, data-driven way to track innovations
 - Vs. current expert-based, time consuming approaches (e.g., Gartner's Hype Cycle, tire track diagrams)
- Includes both **concept** and **product** forms
 - Study relationships between
- Study the innovation ecosystem
 - **Organizations & people**
 - Both those producing & using innovations



Process



Collecting

Identify Concepts

- Begin with target concepts
 - Business Intelligence
 - Health IT
 - Cloud Computing
 - Customer Relationship Management
 - Web 2.0
- Develop 20-30 sub concepts from domain experts, wikis

Data Sources

- **News**



- Dissertation

- Academic



- Patent



- Blogs

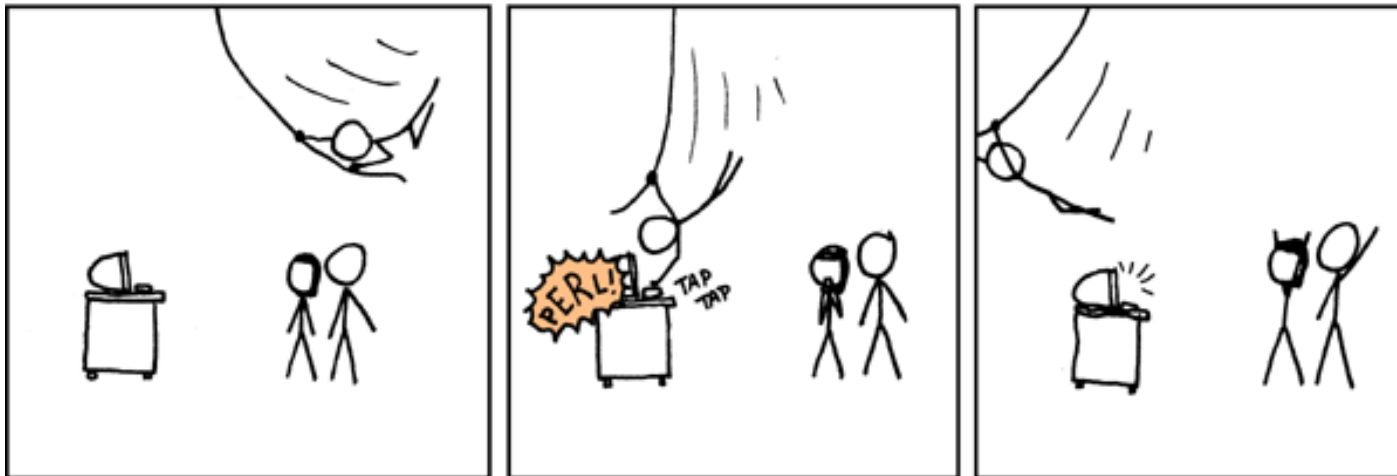


Collecting (2)

- Form & Expand Queries

```
ABS (  
  "customer relationship management" OR  
  "customers relationship management" OR  
  "customer relation management"  
) OR TEXT (...) OR SUB (...) OR TI (...)
```

- Scrape Results



Source: <http://xkcd.com/208>



```
<?xml version="1.0"?>  
<quiz>  
  <question>  
    Who was the forty-second  
    president of the U.S.A.?  
  </question>  
  <answer>  
    William Jefferson Clinton  
  </answer>  
  <!-- Note: We need to add  
    more questions later. -->  
</quiz>
```

XML

Processing

Automatic Entity Recognition

- BBN IdentiFinder

Trackle was founded by <ENAMEX TYPE="LOCATION" end="1342" start="1329">Silicon Valley</ENAMEX> serial entrepreneur <ENAMEX TYPE="PERSON" end="1374" start="1364">Pavan Nigam</ENAMEX> and technology executive <ENAMEX TYPE="PERSON" end="1412" start="1401">Naveen Saxea</ENAMEX>. Nigam was the founder of <ENAMEX TYPE="GPE" end="1449" start="1441">Healtheon</ENAMEX> (a health care start-up that merged with <ENAMEX TYPE="ORGANIZATION" end="1496" start="1492">WebMD</ENAMEX> in <TIMEX TYPE="DATE" end="1504" start="1501">1999</TIMEX>) and Cendura (a start-up that was acquired by <ENAMEX TYPE="ORGANIZATION" end="1570" start="1552">Computer Associates</ENAMEX> in <TIMEX TYPE="DATE" end="1578" start="1575">2006</TIMEX>). Trackle, who has received seed funding by <ENAMEX TYPE="ORGANIZATION" end="1627" start="1625">NEA</ENAMEX> (<ENAMEX TYPE="ORGANIZATION"

Crowd-Sourced Verification

- Extract most frequent 25%
- Assign to CrowdFlower
 - Workers check organization names and sample sentences

Is AAT an organization based on the sentence below?

0 Investment Grade Bond Funds 1838 Bd-Deb (BDF) N NA NA NA NA All-American Tm (AAT)-c N 13

Choose one (required)

- ☐ Unknown
☐ Yes
☐ No

Is ABA an organization based on the sentence below?

Financial Information - Includes (but is not limited to) financial institution, American Banking Association (ABA) account number, remittance address, lock box number, automated clearing house (ACH) information, and credit

Is ACH an organization based on the sentence below?

Financial Information - Includes (but is not limited to) financial institution, American Banking Association (ABA) account number, remittance address, lock box number, automated clearing house (ACH) information, and credit

Processing (2)

- Compute Co-Occurrence Networks
 - Overall edge weights
 - Slice by time to see network evolution
- Output



CSV

SRC	TAR	type	weight	date
IBM	cognos	org_conc	10	13-Nov-07
cognos	IBM	conc_org	10	13-Nov-07
data mining OR datami	National Security Agen	conc_org	9	12-May-06
National Security Agen	data mining OR datami	org_conc	9	12-May-06
NSA	data mining OR datami	org_conc	8	12-May-06
NSA	National Security Agen	org_org	8	12-May-06
data mining OR datami	NSA	conc_org	8	12-May-06
National Security Agen	NSA	org_org	8	12-May-06
NSA	data mining OR datami	org_conc	7	19-May-06
data mining OR datami	NSA	conc_org	7	19-May-06
Oracle	hyperion	org_conc	6	2-Mar-07
hyperion	Oracle	conc_org	6	2-Mar-07
Wall Street Journal	cognos	org_conc	6	13-Nov-07
Wall Street Journal	IBM	org_org	6	13-Nov-07
NSA	National Security Agen	org_org	6	19-May-06
IBM	Wall Street Journal	org_org	6	13-Nov-07
data mining OR datami	National Security Agen	conc_org	6	19-May-06
cognos	Fortune Brands	conc_org	6	13-Nov-07

GraphML

```
<?xml version="1.0" encoding="utf-8"?><graphml xmlns=
"http://graphml.graphdrawing.org/xmlns" xmlns:xsi=
"http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation=
"http://graphml.graphdrawing.org/xmlns
http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
...
<key attr.name="type" attr.type="string" for="node" id="d2" />
<key attr.name="Title" attr.type="string" for="node" id="d1" />
<key attr.name="Date" attr.type="string" for="node" id="d0" />
<graph edgedefault="directed" id="">
  <node id="PQ69133991">
    <data key="d0">Jan 11, 2001</data>
    <data key="d1">A MONITOR GUIDE TO THE BESTSELLERS</data>
    <data key="d2">doc</data>
    <data key="d3">Nara Schoenberg, Tribune Staff Writer</data>
  </node>
  <node id="PQ54016055">
    <data key="d0">May 23, 2000</data>
    <data key="d1">Technology Stocks, Led by Sun, Adobe,
Recover From Day's Lows</data>
    <data key="d2">doc</data>
```

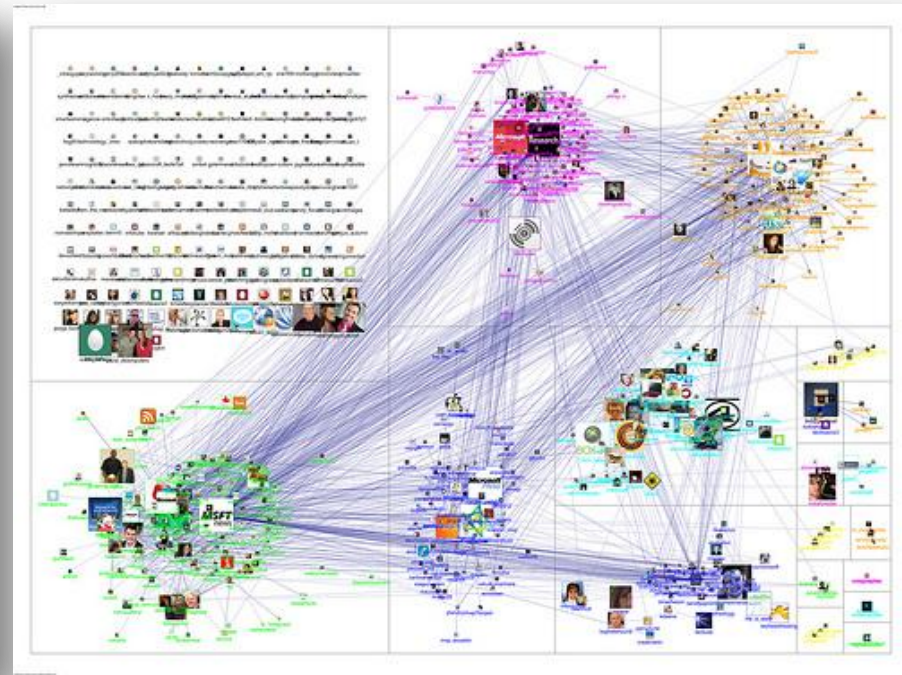
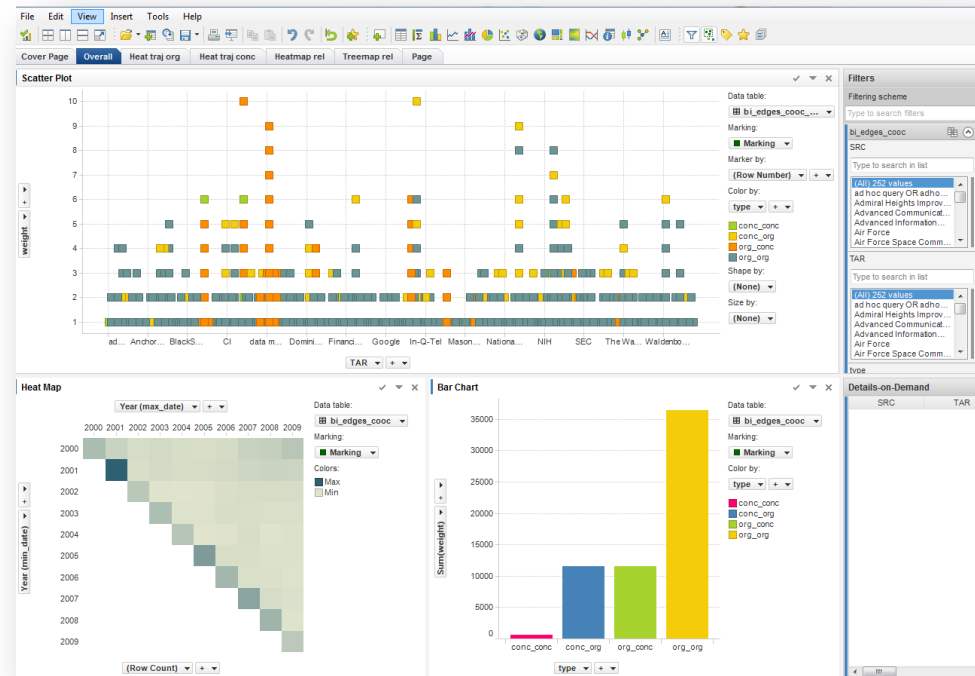
Visualizing & Analyzing

Spotfire

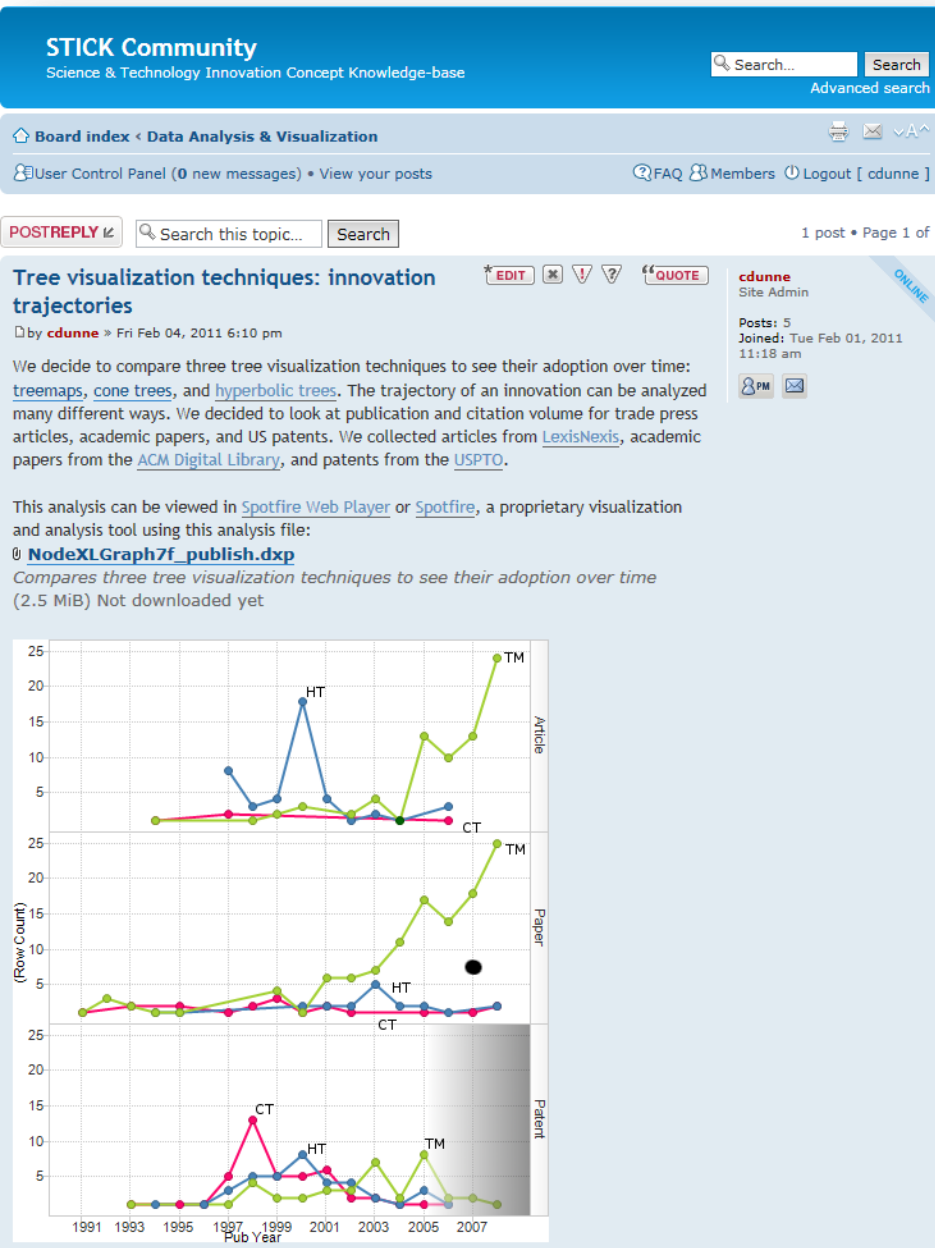
- Import CSV, Database
- Standard charts
- Multiple coordinated views
- Highly scalable

NodeXL

- CSV, Spigots, GraphML
- Automate feature
 - Batch analysis & visualization
- Excel 2007/2010 template



Collaborating



- Online Research Community
- Share data, tools, results
 - Data & analysis downloads
 - Spotfire Web Player
- Communication
- Co-creation, co-authoring



Ongoing Work

- Collecting:** Additional data sources and queries
- Processing:** Improving entity recognition accuracy
- Visualizing & Analyzing:** Visualizing network evolution
- Co-occurrence network sliced by time
- Collaborating:** Develop the STICK Community site
- Motivate user participation
 - Improve the resources available
 - Local testing
 - Invitation-only testing



Take Away Messages

- Easier scientific, data-driven innovation analysis:
 - Automatic collection & processing of innovation data
 - Easy access to visual analytic tools for finding clusters, trends, outliers
 - Communities for sharing data, tools, & results



Analyzing Trends in Science & Technology Innovation

Cody Dunne, Pengyi Zhang, Chen Huang, Jia Sun,
Ben Shneiderman, Ping Wang & Yan Qu

`{cdunne, ben}@cs.umd.edu`

`{pengyi, chuang, jsun, pwang, yanqu}@umd.edu`

<http://stick.ischool.umd.edu>

Thanks to: National Science Foundation grant SBE-0915645



UNIVERSITY OF
MARYLAND

