



---

# Applications and Benefits of Network Analysis in Technology-Enhanced Learning

---

**H. Ulrich Hoppe**

COLLIDE Research Group

University of Duisburg-Essen

Department of Computer Science and Applied Cognitive Science

ICSCS 2011, Chiang Mai (Thailand), Nov/Dec 2011



# Motivation

---



General aspects of network analysis and network models are increasingly used to understand/support online communities.

*Illustrate relevance for TEL!*

---

WE'RE A TERRIBLE MATCH.  
BUT IF WE SLEEP TOGETHER,  
IT'LL MAKE THE LOCAL  
HOOKUP NETWORK A  
SYMMETRIC GRAPH.

I CAN'T ARGUE  
WITH THAT.



**The persuasiveness  
of network analysis**



# Outline

---

- Network analysis
  - general foundations / origins
  - analysis of social networks
- Case study 1:  
Detection of innovators/trends in online (learning) communities
- Case study 2:  
Concept mapping study
- Outlook



# Part 1

---

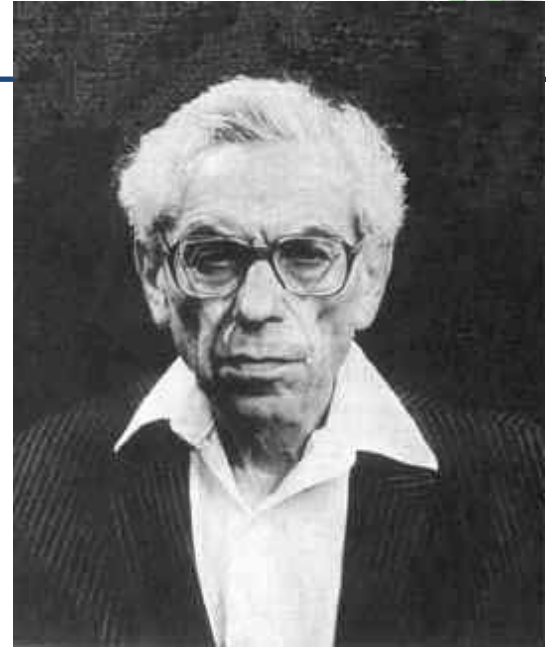
## Network Analysis



# Guess who ...

---

Paul Erdős, 1913-1996  
Hungarian mathematician



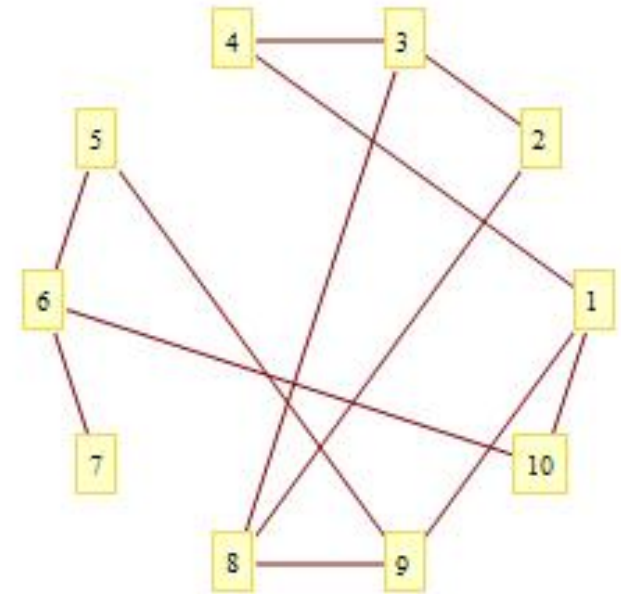
Who said of himself that he was a machine that converted coffee into mathematical theorems ?

# ... and why is he of interest?

## Theory of “random graphs” or “random networks”

A random network, is a graph which contains links that are chosen at random with equal probability for inter-linking. Random links typically result in shortcuts to remote nodes. For example, random links between nodes #6 and #10 or nodes #4 and #1 serve to reach to clusters on the opposite side of the network. This shortening of path length tends to increase connectivity.

Random Network (Probability = 30%)



# Another Hungarian ...

---



## ***Albert-László Barabási***

born 1967, scientist (esp. physics)

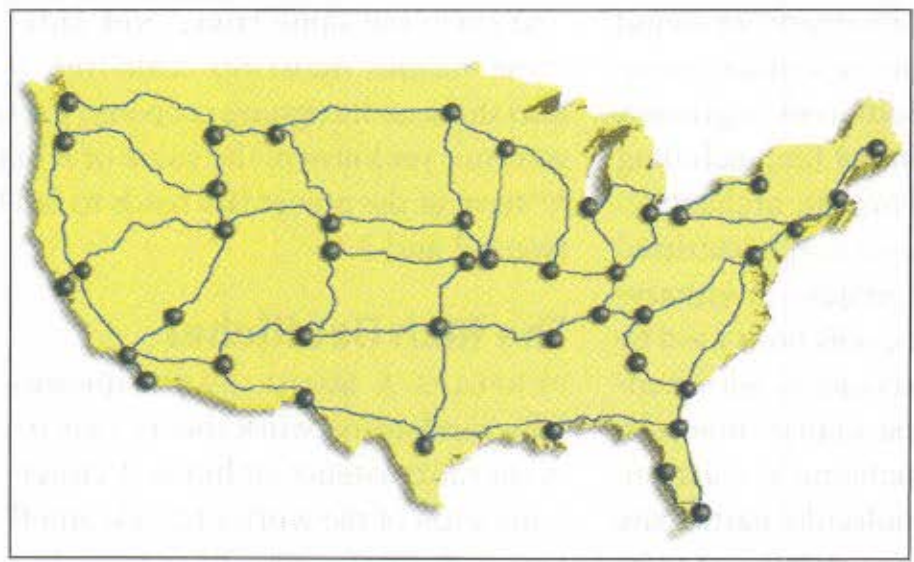
now at Northeastern University (Boston)

discovered that many (most) networks in reality evolve differently in that they show a more uneven distribution of links

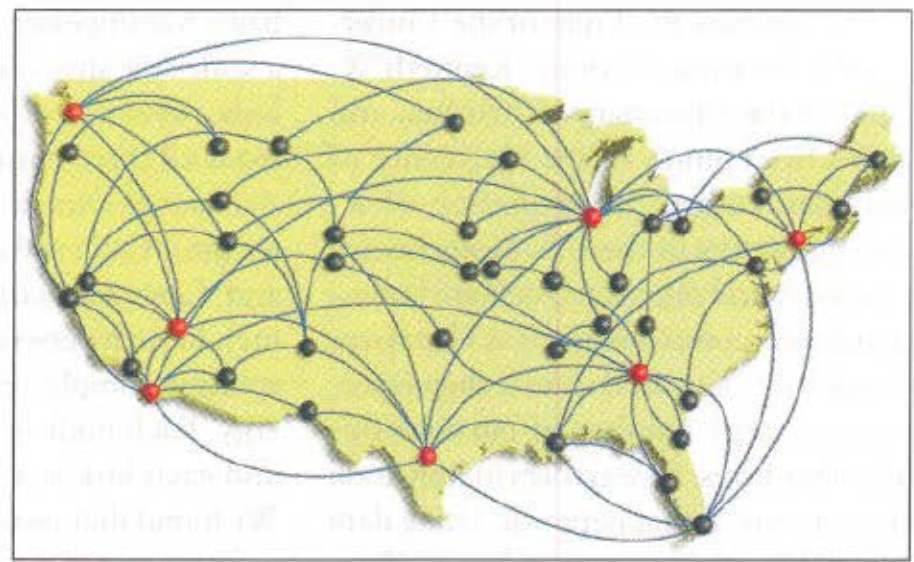


# Random vs. Scale-free

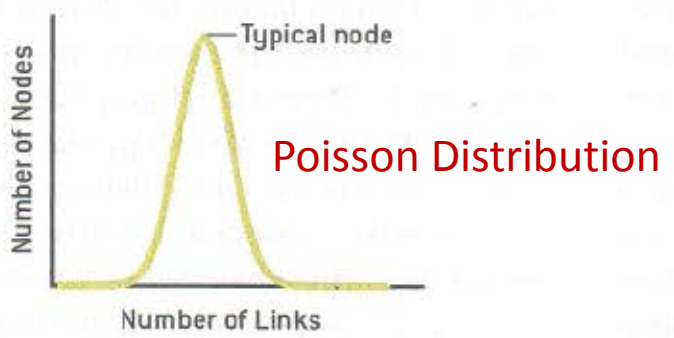
Random Network



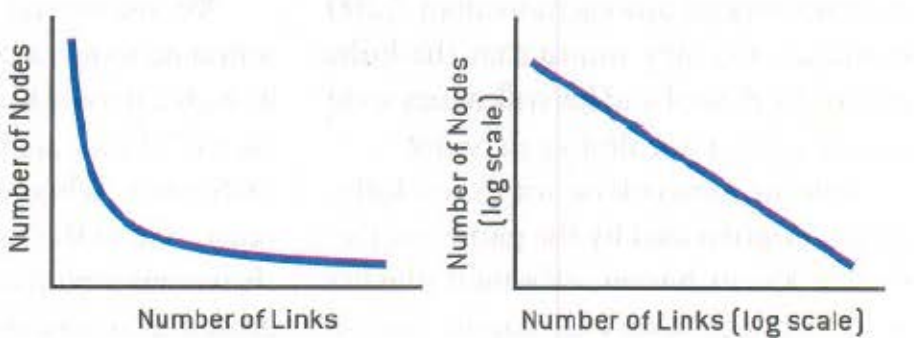
Scale-Free Network



Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages

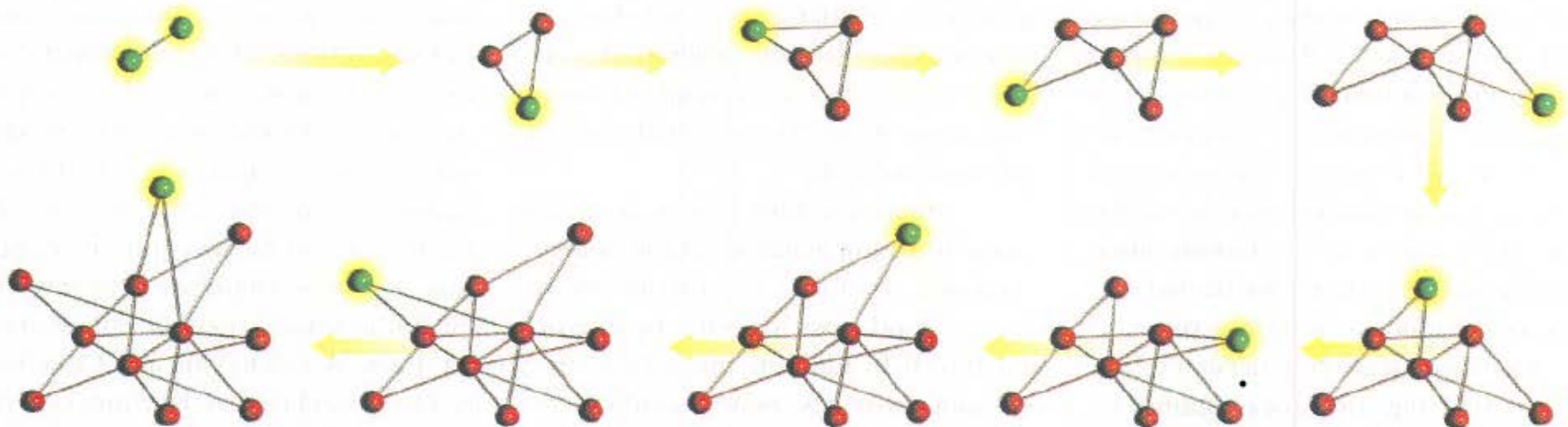


# Preferential attachment



## BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



Barabási, Albert-László; Bonabeau, Eric (May 2003). "Scale-Free Networks".  
*Scientific American* **288** (5): 60–9.



# Simulation (NetLogo)

Preferential Attachment with cycles - NetLogo (C:\Users\hoppe\Documents\Lehre\WiSe10\_11\EComSocNet2010\Examples)

File Edit Tools Zoom Tabs Help

Interface Information Procedures

Edit Delete Add  | normal speed |  view updates | on ticks | Settings...

On plot?
  On layout?

On layout?

**Degree Distribution**

**Degree Distribution (log-log)**

ticks: 412 3D

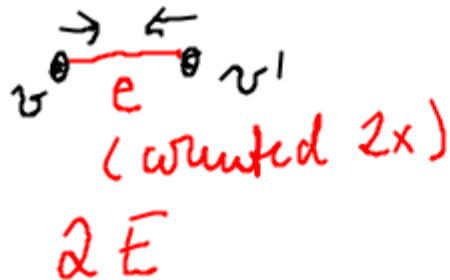


# Network Density

Problem: What is the average degree  $d(G)$  for a given density?

$$|V| = N$$

$$d(G) = \frac{1}{N} \sum_{v \in V} d(v)$$



$$d(G) = \frac{2 \cdot |E|}{N}$$

$$S = \frac{2|E|}{N(N-1)} = \frac{d(G)}{N-1}$$

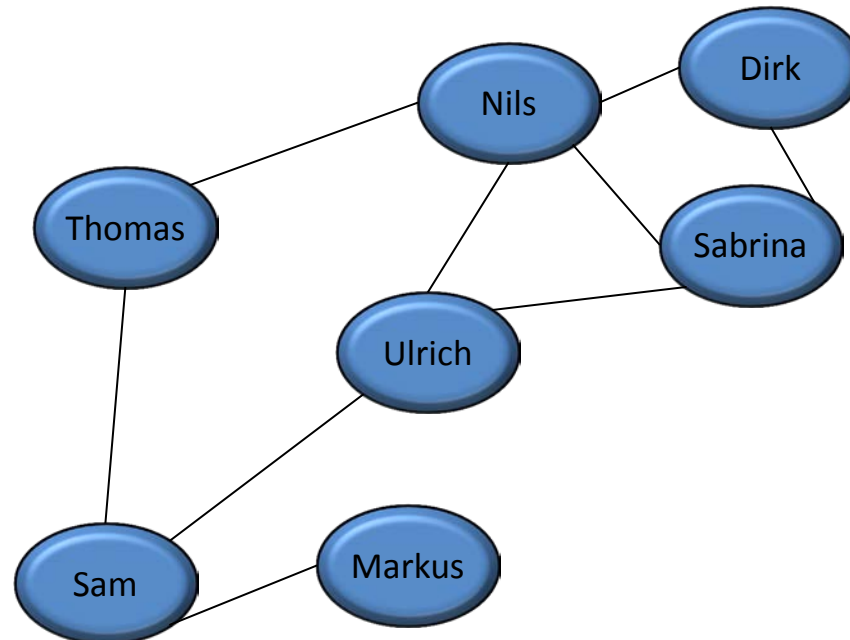


# Social Networks

---

## Simple social networks (“sociograms”)

- Observed relations between actors

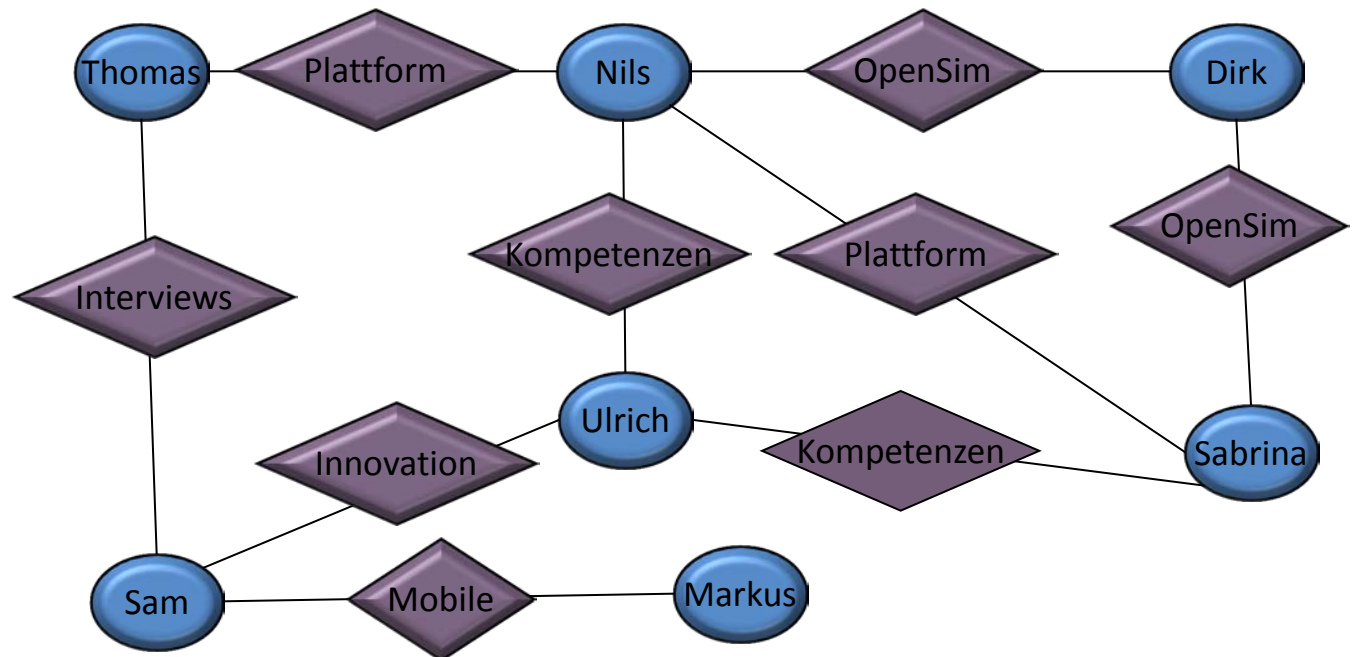




# Social Networks

## Bipartite “Affiliation Networks”

- Mediating objects (between actors)

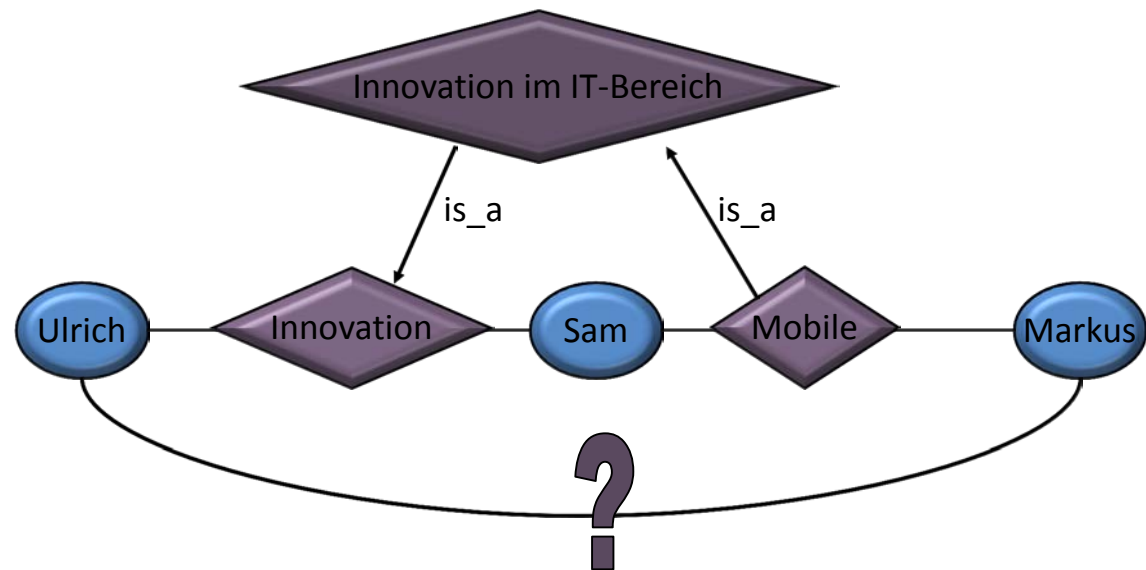




# Hybrid Networks

## “Multimodal Networks”

- Objects mediate between actors
- Semantic relations mediate between objects



See also:

Ogata et al., 1997

„Knowledge awareness maps“



# Enrichment of Semantic Networks

---

Approach based on tagging [Mika, 2004]:

1. Form tag - actor network
  - tags representing concepts; strength dependent on number of occurrences
2. “Fold away” actors =>
  - Semantic tag - tag relations
3. Add directions (B subconcept of A):

$$B \subseteq A \Leftrightarrow A \cap B = B, \frac{|B|}{|A|} < k.$$

$k$  dep. on network characteristics



# Inference Matrix

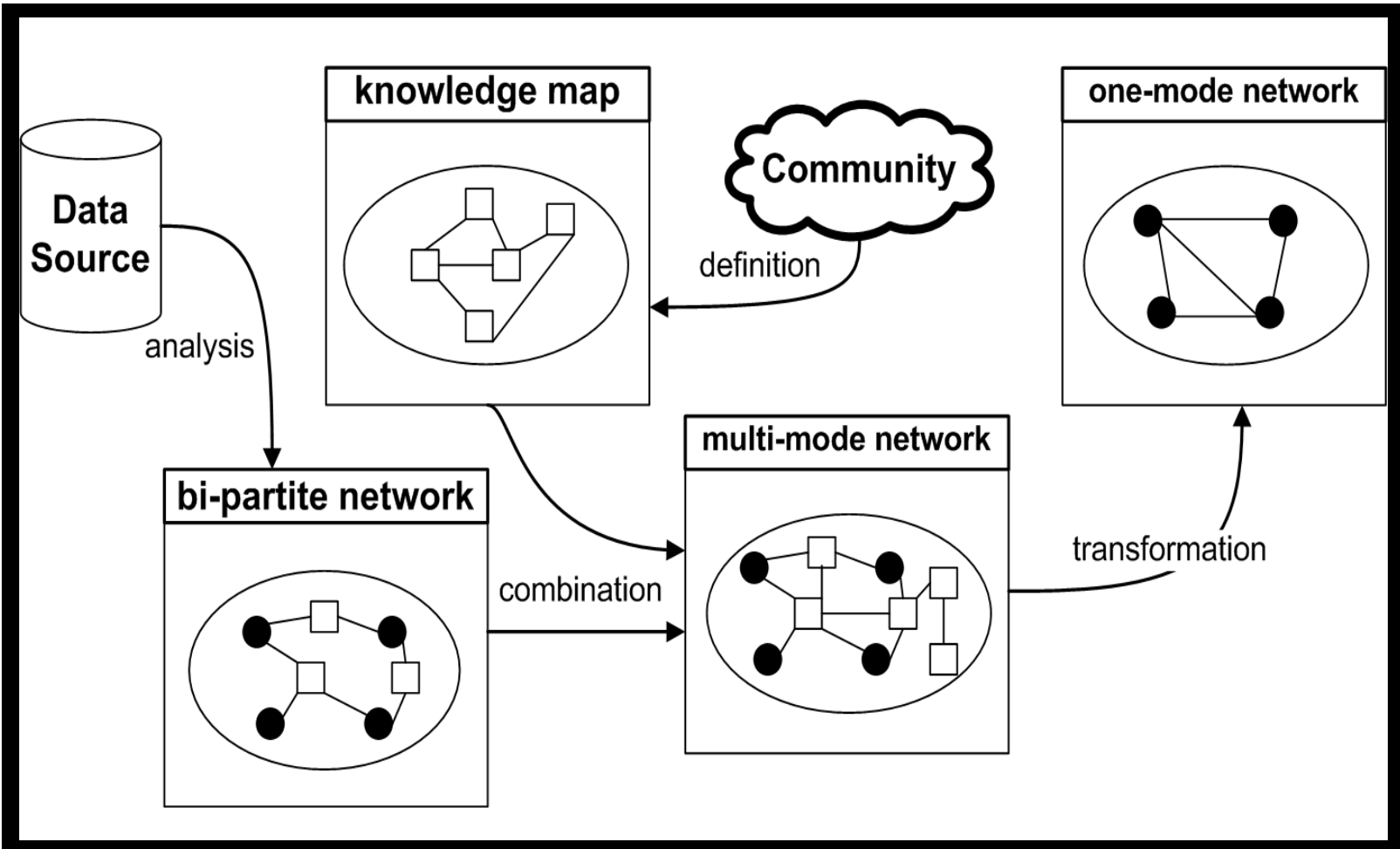
## Semantic Net / Ontology

Social Network

Relation?	yes	no
yes	normal	Check relation: <b>Trend?</b>
no	Check: Possible synergy/coop?	normal



# Semantic - Social Network Analysis



# Part 2

---



## Case Study: „Productive Online Communities“

# Open source projects as learning communities with free data sources

A Newton's cradle with five silver spheres. The word "COLLIDE" is written in green, semi-transparent letters across the spheres. The 'C' is on the first sphere, 'O' on the second, 'L' on the third, 'L' on the fourth, 'I' on the fifth, 'D' on the sixth, and 'E' on the seventh. The spheres are in motion, with the rightmost one having just struck the others.

- Students engaged in OSPs form communities of practice across universities
- Most data (log files from SVN/CVS repositories as well as mailing lists) freely available
- Different roles and skills of actors have to be considered (hierarchies, commit rights)
- Data has to be cleaned (e.g. merging name duplicates)

# Case: OpenSimulator



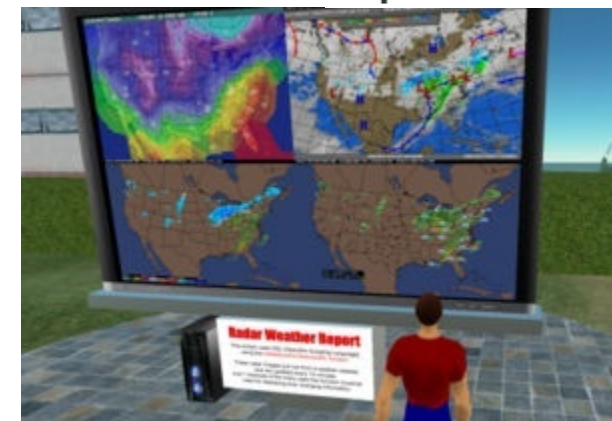
OpenSimulator:

- Open source server platform for hosting virtual worlds
- Compatible with the client for Second Life
- Can host alternative worlds



OpenSimulator

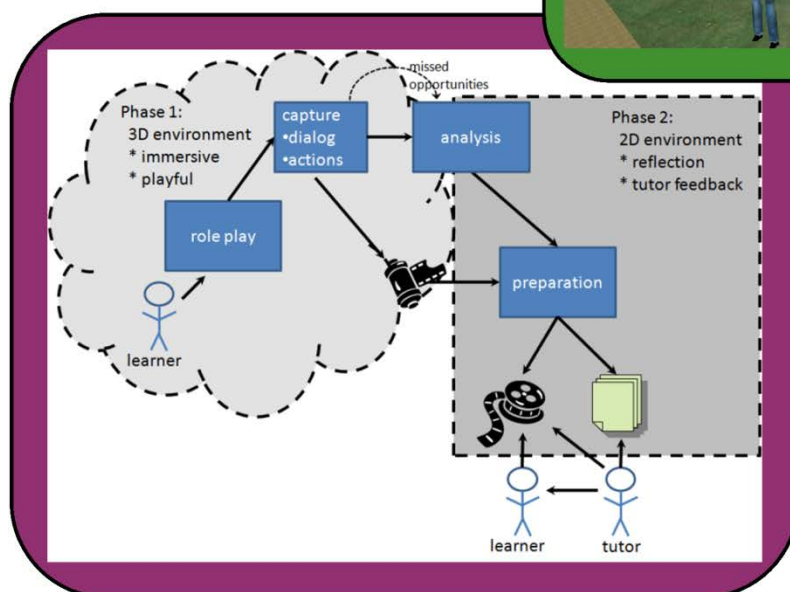
Used within KoPIWA research project  
(German project on Open Innovation  
and competence management in IT)



# Student project with OpenSimulator

Software project carried out by students of Computer Science at UDE:

*Serious game  
on training for  
start-up consulting*



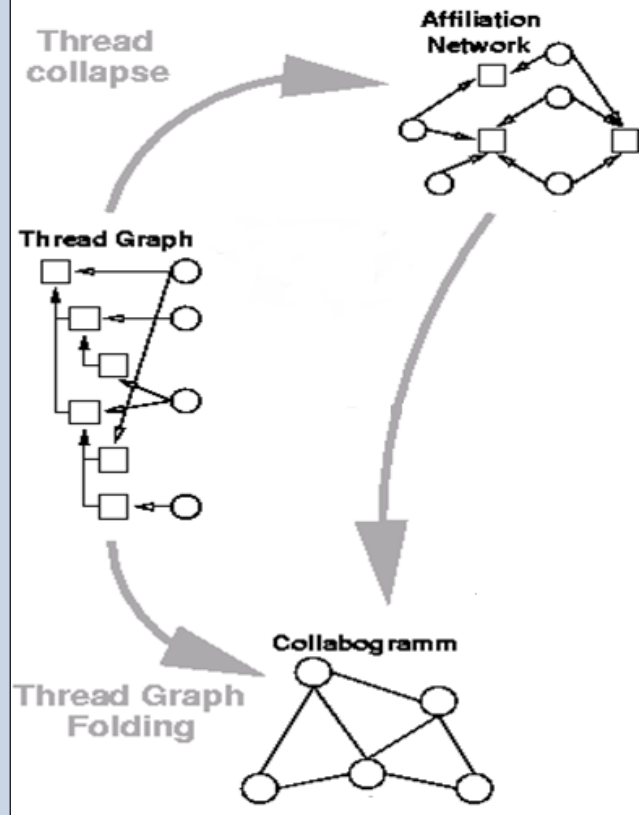
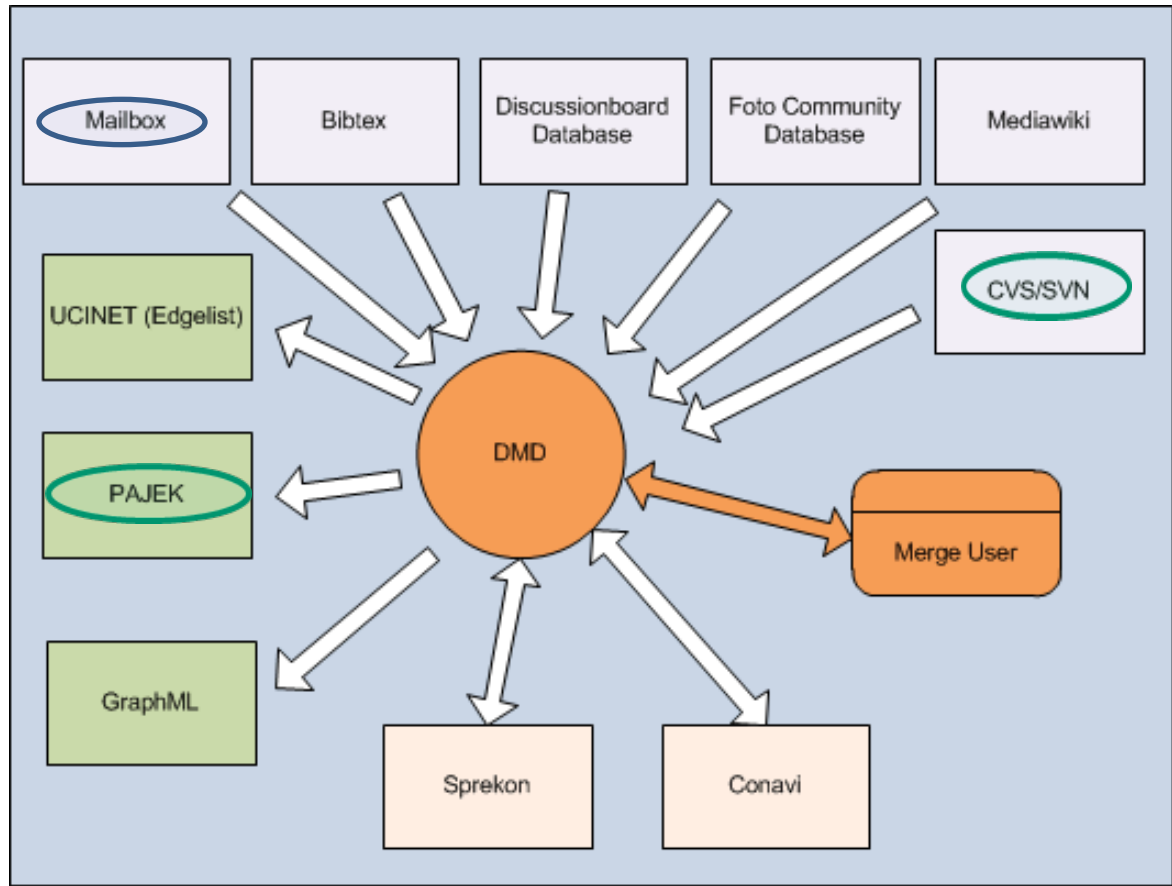
The screenshot shows a 2D virtual interview environment. A character is seated at a table in a meeting room. A chat window on the right displays the following text:

Marvin Amiri(00:02): Hello Mr. Buhmes.  
 Buhmes(00:10): Hello  
 Missed opportunity: You missed the opportunity to **shake hands with your partner**.  
 Marvin Amiri(00:12): I have been waiting for you. Please come in and have a seat. Did you have a good trip?  
 Buhmes(00:15) sitting down...  
 Buhmes(00:27): Yes. Everything was well.  
 Marvin Amiri(00:29): Let's start immediately. I read your application. Have you ever been interviewed before, Mr. Buhmes?  
 Buhmes(00:53): Yes. I have taken part in two or three interviews.  
 Part: Application and career choice  
 Marvin Amiri(00:55): What do you expect to do for us?  
 Buhmes(01:18): I don't know. I hope to learn a lot of things.  
 Remarks: You should take into account that your knowledge about the enterprise will be tested. Do some research to show interest.  
 Marvin Amiri(01:20): What do you know about our products?  
 Buhmes(01:36): I've heard only the best.  
 Marvin Amiri(01:38):



# Data extraction using DMD

Collide Data-Multiplexer-Demultiplexer Approach



# Data extraction using DMD



## Reading in from mailing lists

The screenshot shows the DMD software interface with two main panes:

- Left Pane (Preferences):**
  - Input:** Mbox, File: .\g07\dmd\OMD0107-0807.mbox-header, Browse
  - Options:** GhostNodes:  Decrypt:
  - Output:** Sprekon, File: .\g07\Neu\OMD0107-0807temp.spreml, Browse
  - Convert:** Convert
  - Log:** Exporting to Pajek, #### Output complete ####, #### Input starts ####, Importing Mbox, #### Input complete ####, #### Merging starts ####
- Right Pane (Merge Persons):**
  - Compare:** Name (selected), Email, log, Distance: 0-10, Set, Reset
  - Comparison Results:**
    - Entry 1: ID: 998, Name: Michael 'Mickey' Lauer, Email: mickey@vanille.de, Groups: 21
    - Entry 2: ID: 115, Name: Michael 'Mickey' Lauer, Email: mickey@openmoko.org, Groups: 115
  - Filter:** lauer, 1 user(s) found
  - Buttons:** Merge, OK, Cancel

From sean at dague.net Sat Sep 8 14:25:15 2007

From: sean at dague.net (Sean Dague)

Date: Sat, 8 Sep 2007 08:25:15 -0400

Subject: Opensim-dev Asset Storage Interface

In-Reply-To: <20070904212524.GQ29833@dague.net>

Message-ID: <20070908122515.GE1471@dague.net>

# Data extraction using DMD



## Getting logs from code repositories

Revision 17 - Directory Listing

Modified Thu Apr 2 20:09:56 2009 UTC (2 years, 2 months ago)

by sdague

updates to support a much more reasonable recovery system

----

Revision 16 - Directory Listing

Modified Tue Feb 10 17:35:46 2009 UTC (2 years, 4 months ago)

by sdague

merges from fixes in IBM Grid

----

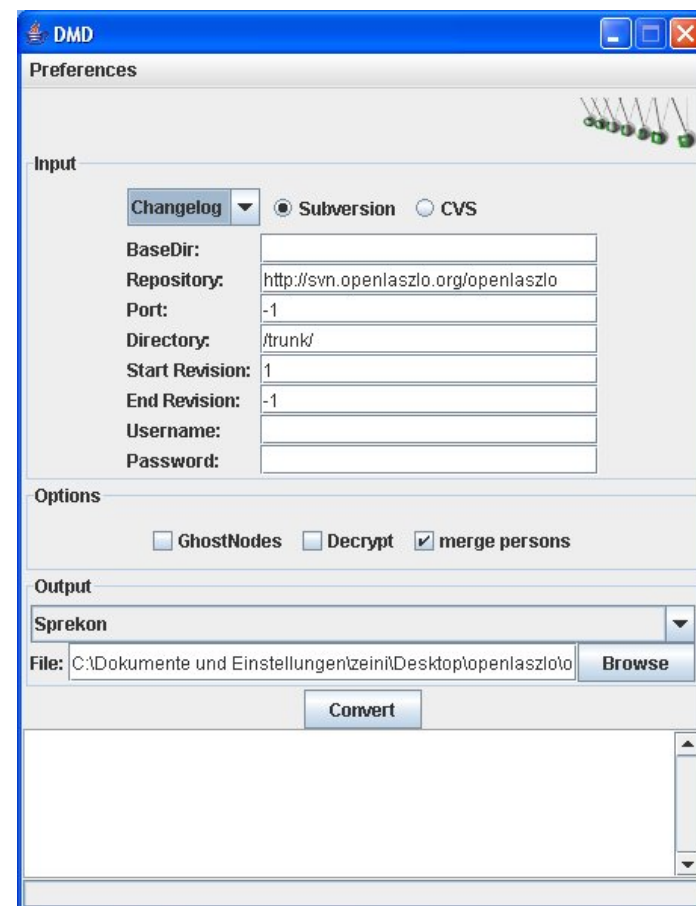
Revision 15 - Directory Listing

Modified Tue Feb 3 14:43:53 2009 UTC (2 years, 4 months ago)

by sdague

fail faster

----





# Source Data

---

OpenSim Developer mailing list and SVN source code repository between Sept. 2007 and Feb. 2009

- Cleaned by merging duplicate names in mailing lists
- Mapping real names from mailing lists to SVN nicknames
- SVN contains 26 users and 6011 objects (classes) based on 32867 objects including revisions
- Developer mailing list contains 197 users and 1184 topics based on 5505 emails.

Dataset from the SVN repository and the discussion board of the student subproject from October 2009 to March 2010

- SVN data from the students subproject contains 4 users and 235 objects based on 326 revisions.
- Discussion board hosted in a liferay based portal contain 12 students discuss 78 threads based on 243 messages (in German)



# Analysis using Pajek

The screenshot shows the Pajek software interface with two network graphs. The main window displays '10. Network N1 in time 278 (20)', and a smaller window shows '22. Network N13 in time 278 (7)'. The left sidebar contains a list of networks, partitions, vectors, permutations, clusters, and hierarchies.

**Network N1 (Time 278):** A complex network graph with nodes including:
 

- Opensim-dev build issues wrt native libs (redux)\_<5662>
- Opensim-dev Memleak in OpenSim?\_<5644>
- Opensim-dev Memleak in OpenSim?\_<5645>
- Adam Frisby
- Tedd Hansen
- Opensim-dev ScriptServer communication\_<5656>
- Stefan Andersson
- Opensim-dev build issues wrt native libs (redux)\_<5670>
- Opensim-dev Usage of trimeshes, was: Open Mantis issue,\_<5668>
- Opensim-dev Assembly version\_<5666>
- Dünnebeil Gerhard
- Opensim-dev In World Office Hours Redux (was Re: How to\_<5663>
- Sean Dague
- Anton Lauridsen
- Opensim-dev Memleak in OpenSim?\_<5658>
- Opensim-dev In World Office Hours Redux (w
- Charles Krinke
- Opensim-dev build issues wrt native libs (redux)\_<5664>
- Opensim-dev How to contribute to OpenSim?\_<5658>

**Network N13 (Time 278):** A smaller network graph with nodes including:
 

- /trunk/OpenSim/Framework/Data.SQLite/SQLiteInventory
- /trunk/bin/ode.dll\_<33088>
- Dan Miller
- Darren Guard
- /trunk/prebuild.xml\_<32871>
- /trunk/OpenSim/Framework/General/Types/InventoryItemBase.cs\_<32961>
- /trunk/OpenSim/Framework/Data.MySQL/MySQLInventoryData.cs\_<32935>

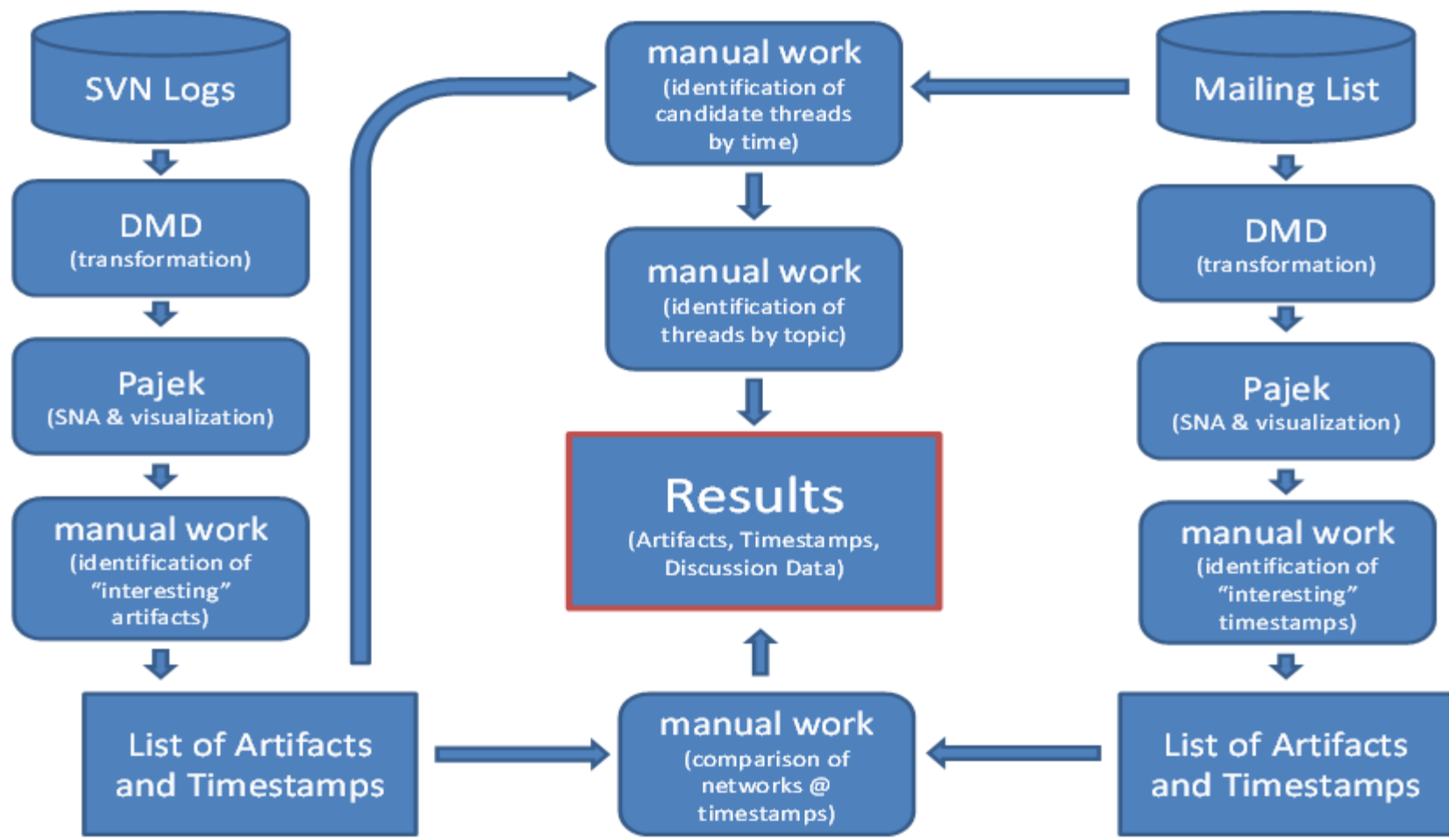
Mailing list at day 278

SVN at day 278



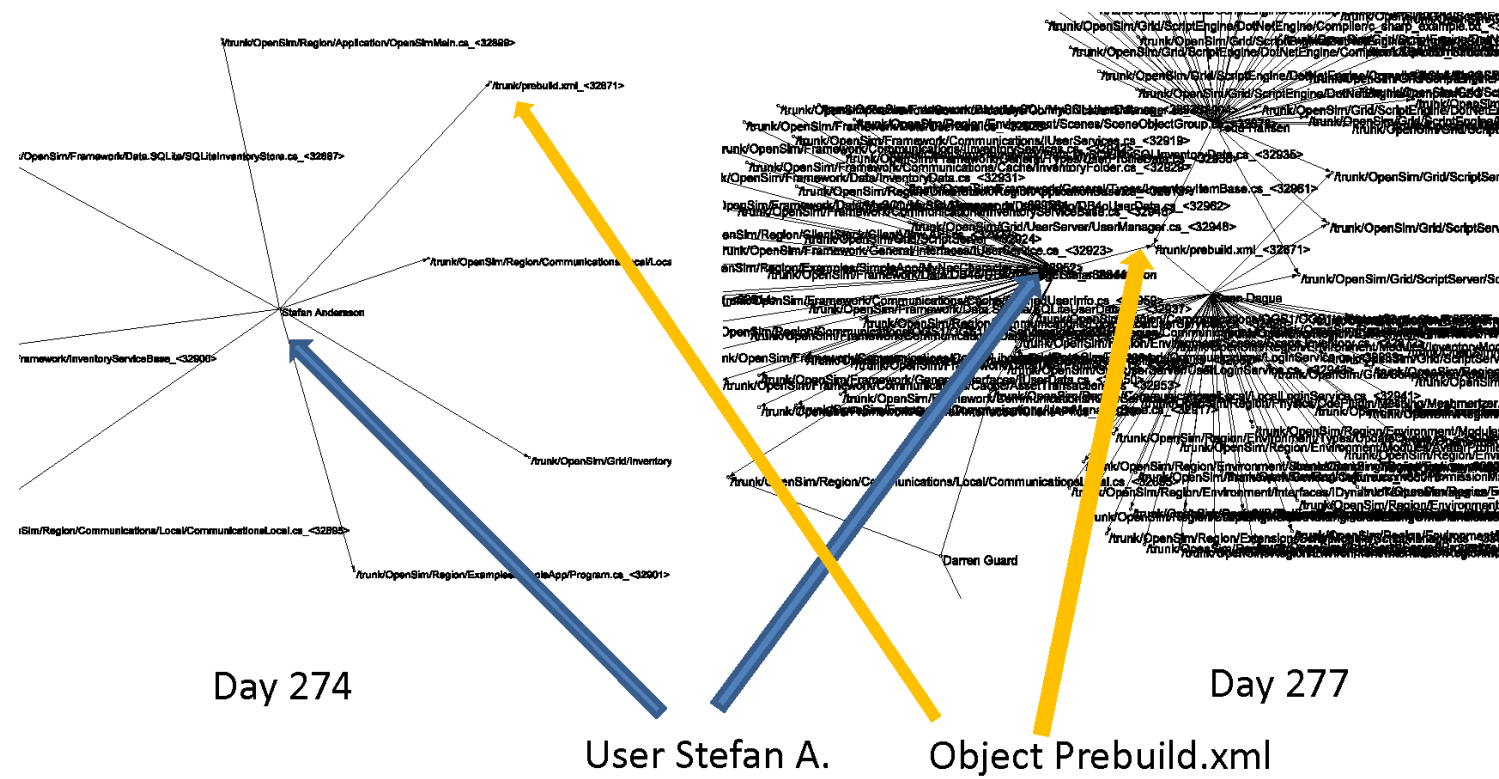


# Analysis workflow





# First observation



- Burst of activity in SVN
- Restructuring of the network (bridging between components)

# Comparison of centrality dynamics

Name	Delta T1/T2	Delta T2/T3	Delta T3/T4	Delta T1/T4
Adam F.	0	0,125	-0,125	0
Aldon H.	0	0,083333	-0,083333	0
Anton L.	0,1875	0,104167	0,333333	0,625
Brian M.	-0,125	0	0	-0,125
Charles K.	0,0625	0,020833	-0,083333	0
Chris D.	0	0	0	0
Dalien T.	0,1875	-0,354167	-0,020833	-0,1875
Dan M.	0,4375	-0,291667	0,166667	0,3125
Darok K.	0	0	0	0
Gerhard D.	0	0,083333	-0,083333	0
Izumi N.	0	0	0	0
Jeff A.	0,125	0	-0,125	0
Mic B.	0	0,125	-0,125	0
Michael W.	0,125	-0,145833	0,395833	0,375
Sean D.	0,375	-0,416667	0,229167	0,1875
Stefan A.	0,625	-0,479167	0,291667	0,4375
Tedd H.	0,125	-0,1875	0,1875	0,125
	Rank 4	Rank 3	Rank 2	Rank 1

*T= Time slices with length of 10 days*

*T1= Day 250-259*

*T2= Day 260-269*

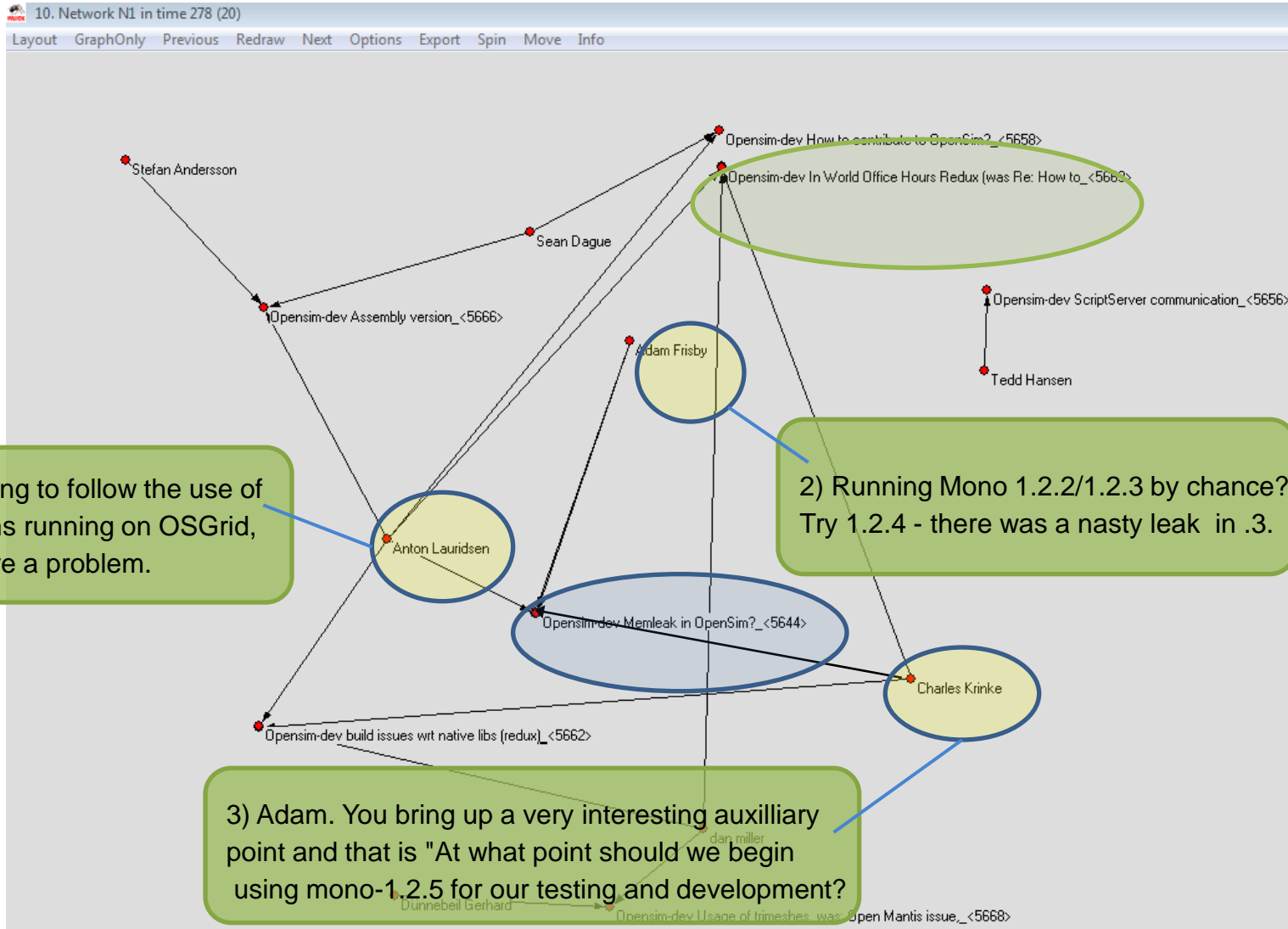
*T3= Day 270-279*

*T3= Day 280-289*

- Interesting time point according to interaction increase
- Actor “Tleiades” (Anton L.) identified in communication network at same time point by his increase of centrality (shooting star)



# Communication (mailing list)



1) I have been trying to follow the use of memory in my sims running on OSGGrid, and I think we have a problem.

2) Running Mono 1.2.2/1.2.3 by chance? Try 1.2.4 - there was a nasty leak in .3.

3) Adam. You bring up a very interesting auxilliary point and that is "At what point should we begin using mono-1.2.5 for our testing and development?"

# “Story” behind example case



- First indicator: high activity followed by network restructuring
- Second indicator: degree gradient points to specific actor “Tleiaades” and his entry to the core of the network  
( in the end, centrality moved to 0.0505 in the developer mailing list – rank 47 out of 199)
- Semantic background: T. initiated relevant discussions on XML-RPC as well as on group standards for build system



# Conclusion

---

- Open Source development communities are learning arenas crossing boundaries of universities and schools
- Network indicators can serve as detectors of relevant changes in the work of the community
- Our approach compares a social network based on coding actions with communication patterns over time
- ... and can be complemented with content-based semantic analyses

# Part 3

---



## Case Study: „Structural Quality Measures for Concept Maps“

# Concept Mapping Study



## Context:

ongoing EU project SCY (“Science Created by You”) on supporting collaborative inquiry learning through sharing learner-generated objects (ELOs)

=>

“SCY missions” use concept mapping as a tool for early conceptualization activities in a new domain: students create concept maps from given resource documents

# Concept Mapping Study

---



## Participants:

37 high school students aged between 16-18

## Procedure:

45 min - introduction to concept mapping  
(+ tool "SCY Mapper")

45 min - reading/highlighting of a two page text on  
global warming followed by concept map construction

experimental group received adaptive suggestions  
control group could only access a dictionary



# Analysis & Scaffolding

Request concept help | Request relation help | Lexicon

```

graph TD
    GW((global warming)) -- "has influenc..." --> MEW((more extreme weather))
    GW -- "increases" --> MG((melting glaciers))
    GW -- "increases" --> AO((acidification of oceans))
    GW -- "has influenc..." --> GE((greenhouse effect))
    GE -- "increases" --> CO2((co2))
    
```

**Concept Suggestions**

Maybe you consider adding the following:

- + fossil fuel [Concept already exists!](#)
- + sea level rise [Concept already exists!](#)
- + temperature [Concept already exists!](#)

Close

# Ontology Support / Matching



Autoplay Agents Status: Agent running

**Text**

Global **warming** refers to the increase observed in recent decades the **average temperature** of the lower **atmosphere** and the **oceans** and their expected future **warming**. Between 1906 and 2005, the **average temperature** near the ground by 0.74 ° C (+ / - 0.18 ° C) has increased. The decade 2000-2009 was by far the warmest **on** record, followed by the 1990s, which in turn were warmer than the 1980s. According to current scientific understanding of this is 'very likely' the reinforcement of the **natural greenhouse effect of human influence** cause. The man-made **warming** caused by **burning fossil fuels, deforestation** and world-embracing agriculture and livestock. As a result, the **greenhouse gas carbon dioxide (CO2)** and other **greenhouse gases** like methane and nitrous oxide accumulates in the **atmosphere** so that less **heat radiation** emitted by the **Earth's** surface into space. By far the largest **part** of the past and expected anthropogenic **warming** is due to the recent and still **growing** increase of the **greenhouse gas carbon dioxide**. Through strong feedback processes is the direct **heating effect of the carbon dioxide** but with significantly less likely than expected, resulting from the **heating, warming** also secondary effects. By the **year 2100**, depending on future **CO2 emissions** and the actual response of the **climate system** to ensuring expect a **warming** of 1.1 to 6.4 ° C. This would have a **number** of consequences: increased **glacier** melt, **sea level rise, ocean acidification, changes in precipitation** patterns, **more extreme weather** events, including With regard to the projected **rise in sea level**, many questions remain unanswered. The measured bandwidth of the **rise in sea level** by the end of the 21 Century is, depending on the scenario, 18 to 59 centimeters. The scenarios take into account but not the polar ice dynamics (eg, the collapse of large ice sheets in Antarctica) and uncertainties in climate-carbon cycle feedbacks: In a warmer **climate** will **reduce** the capacity of the **oceans** and the land surface **of man-made CO2** because in warmer **water** less **CO2** is dissolved and remove the **CO2** at higher

**Ontology-generated Concept Map**

**Student Concept Map**



# Expert Assessment

---

Four experts assessed the concept maps produced by the students

based on two criteria inspired by D. Jonassen:

- Completeness regarding the concepts used
- Connectedness regarding the relations introduced (adequacy, missing rel's)



# Correlations

---

Correlations of structural measures  
with expert judgments

	Criterion 1 (completeness)	Criterion 2 (connectedness)
# nodes	$r = 0.79, p = 0.000$	-
# edges	$r = 0.74, p = 0.000$	$r = 0.63, p = 0.000$
density	<b><math>r = -0.53, p = 0.001</math></b>	$r = -0.26, p = 0.11$



# Correlations

---

Correlations of structural measures  
with expert judgments

	Criterion 1 (completeness)	Criterion 2 (connectedness)
# nodes	$r = 0.79, p = 0.000$	-
# edges	$r = 0.74, p = 0.000$	$r = 0.63, p = 0.000$
# edges / # nodes	$r = 0.20, p = 0.2$	$r = 0.68, p = 0.000$



# Explanation

---

## ***Concept maps are scale-free networks!***

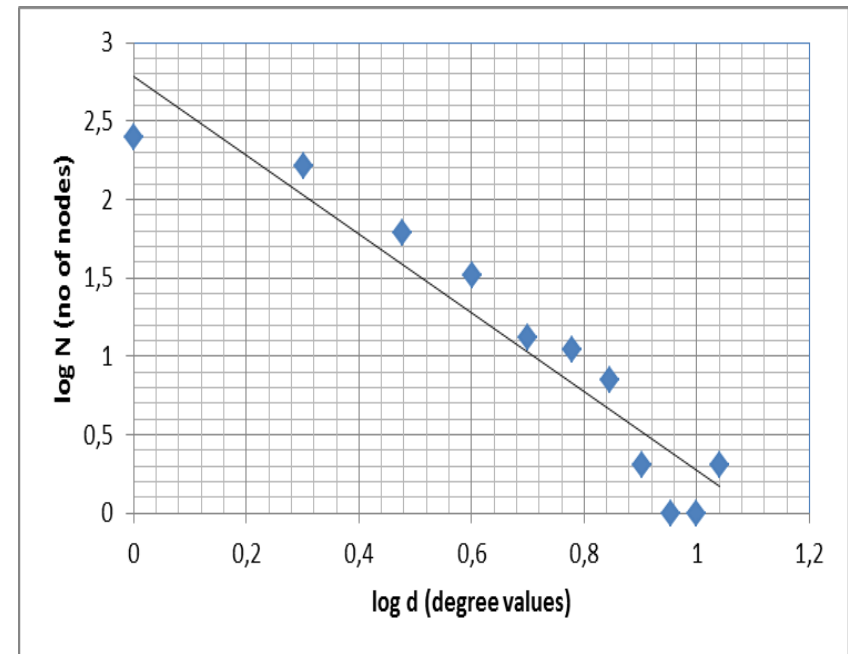
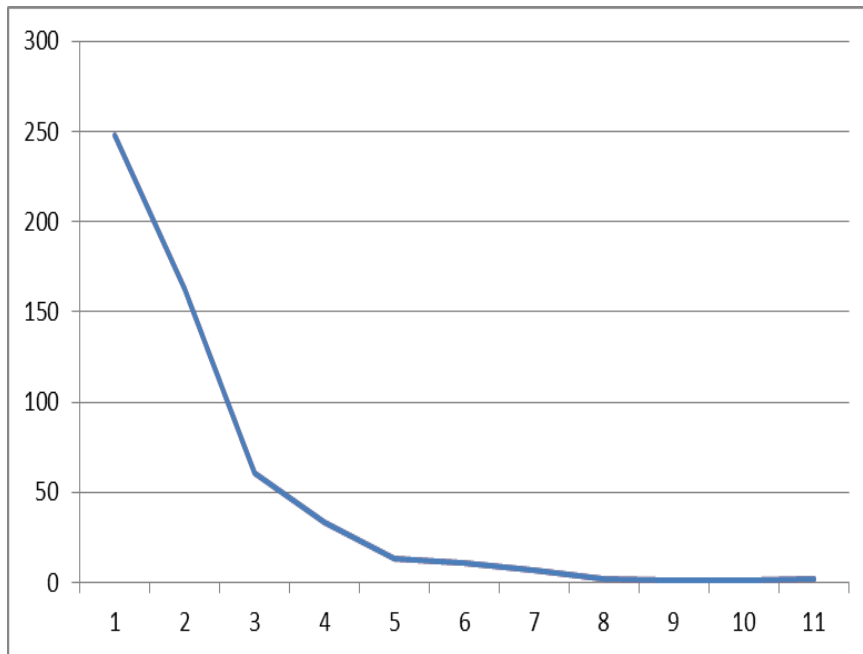
=> evolution of concept maps is governed by “preferential attachment”;

density decreases with growing networks

*M. Jacobson & M.Kapur (2010):  
“Ontologies as scale free networks –  
implications for theories of conceptual change”  
(panel paper at ICLS 2010)*



# The Litmus test



Aggregated degree distribution (from 37 maps)

left: number of nodes per degree value, right: log-log graph



# Prospects

---

Indicators for **conceptual change**:

- Identification of “semantic hubs”
- Changes in centrality (degree)
- New concepts appearing in the “core”
- ...



Questions ?