Small-scale P2P File-sharing for the Academic Environment

Julita Vassileva
University of Saskatchewan
Computer Science Department
MADMUC Lab

Contents

- Motivation
 - for the need of P2P filesharing in Academia
 - for the need of user motivation
- Context: Comtella P2P Filesharing System
 - Two Comtella versions
- Ensuring Presence
- Motivating New Contributions
- Motivating High Quality Contributions
- Design of Mechanisms for Self-Sustainable Online Community

Context I: in a Research Group

Need a search engine for locally stored papers

- Web links disappear, protected sites

- Hard disks too large to find stuff easily

Why P2P?

 Harvest the resources of a community of researchers

 Advantages of a distributed approach vs centralized

maranGraphics Inc.

Context II: In the Classroom

- Example: the "Ethics and IT" class
 - Themes: 1-ethics and IT, 2-privacy, 3-freedom of speech, 4-intellectual property, 5-wiretapping and encryption, 6-computer crime and security, 7workplace issues, 8-can we trust computers, 9-general social issues and 10professionalism
 - Textbook: good, but provides mainly U.S. context
 - Previous solution: students find web-materials and post them on their websites; instructor visits the personal websites, selects the best links and posts them on the class website.
 - Laborious, delayed effect, subjective (instructor picks)
- Why P2P solution?
 - Users are producers, maintainers and consumers: "power to the people"
 - Shared files become immediately available to everyone
 - Protocol allows simple search
 - Easy to develop, cheap to maintain

2 Contexts – 2 Versions of Comtella

Comtella 1

- For sharing academic papers (PDF files)
- Papers classified and searched by category (requires peers to share a category list)
- Applied for 3 months in several research groups at the CS Dept., ~20 users

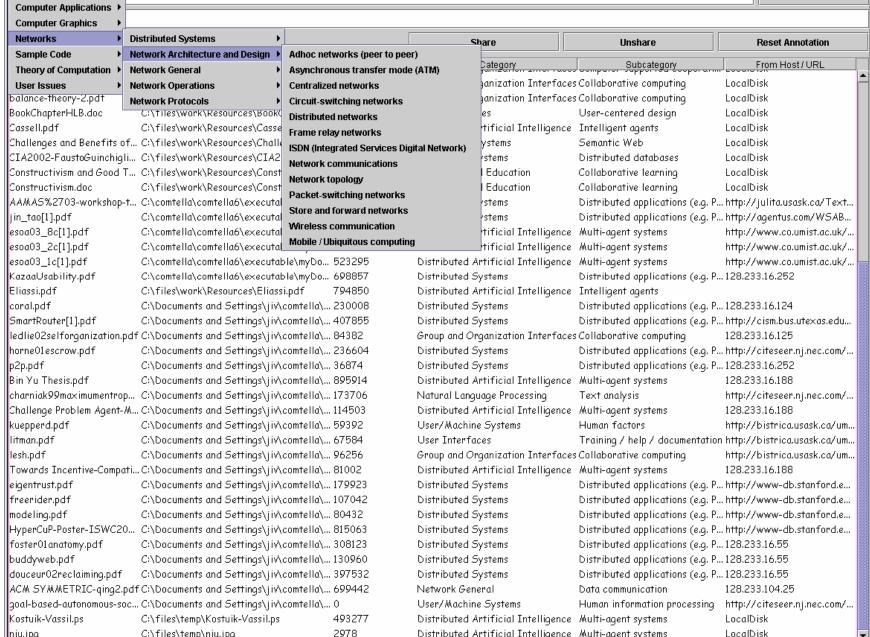
Comtella 2

- For sharing classrelated web-links (popular magazine articles etc.)
- Links classified according to the curriculum topics (usually weekly topics)
- Applied for 3 months in a 4th year class on Ethics and IT, 35 users



- Used by graduate students to search, store and share academic papers
 - Where did I save that file??
 - Who else may want to read this paper?
 - Did any of my colleagues find anything in this area?
- Allows both local and global search
- Client based on Gnutella 0.6 protocol (uses Jtella)
- Applied experimentally across the CS Dept for 3 months.

Comtella File Sharing System - MADMUC Lab, University of Saskatchewan Community Transferring Search Share Select Categories... Distributed Artificial Intelligence = Multi-agent systems **Browse** Computer Applications Computer Graphics **Distributed Systems** Networks Share Unshare **Reset Annotation** Sample Code Network Architecture and Design Adhoc networks (peer to peer) From Host / URL Subcategory Theory of Computation **Network General** Asynchronous transfer mode (ATM) anization Interfaces Collaborative computing LocalDisk **Network Operations User Issues** Centralized networks balance-theory-2.pd1 anization Interfaces Collaborative computing LocalDisk **Network Protocols** Circuit-switching networks BookChapterHLB.doc C:\files\work\Resources\Book(User-centered design LocalDisk Distributed networks Cassell.pdf C:\files\work\Resources\Casse tificial Intelligence Intelligent agents LocalDisk Frame relay networks Challenges and Benefits of... C:\files\work\Resources\Chal ystems Semantic Web LocalDisk ISDN (Integrated Services Digital Network) CIA2002-FaustoGuinchigli... C:\files\work\Resources\CIA2 rstems Distributed databases LocalDisk **Network communications** Constructivism and Good T... C:\files\work\Resources\Cons Education Collaborative learning LocalDisk Network topology Constructivism.doc C:\files\work\Resources\Cons Education Collaborative learning LocalDisk Packet-switching networks AAMAS%2703-workshop-t... C:\comtella\comtella6\executo rstems Distributed applications (e.g. P... http://julita.usask.ca/Text... Store and forward networks jin_tao[1].pdf Distributed applications (e.g. P., http://agentus.com/WSAB., C:\comtella\comtella6\executo Wireless communication esoa03_8c[1].pdf C:\comtella\comtella6\executa rtificial Intelligence Multi-agent systems http://www.co.umist.ac.uk/...





Why not share more files?

Your contribution

to the community

will be rewarded

services by your

downloaded are

C:\Documents

Settings\jiV\comtel

la\myDownloads,

and are set to be

shared by default

explicitly unshare

unless vou

them.

with better

peers.

All files

saved in:

and

Problems

- User Participation
 - "critical mass" needed
 - most users are free-riders
- Question 1: How to prevent free riding?
- Question 2: Why do people contribute?
 - doesn't cost (effort, money, inconvenience)
 - there is some incentive (usefulness, glory, money)
 - serves a greater cause (e.g. cancer research, SETI@home, etc.)

Approaches to Prevent Free Riding

- Sledge-hammer approaches
 - Impose a minimum participation level in order to participate at all (ensures sharing)
 - Make it hard to quit the application (ensures infrastructure)
- Economy-based (incentive-based) approaches
 - Define "participation points" that can be earned by users
 - Reward users with high participation point scores with
 - Digital cash (Mojo-nation, I-Help)
 - Better quality of service (KaZaA, eMule, Comtella1)
 - Reputation / power (Slashdot.org, Comella1,2)
- Socio-economic approaches (Comtella)
 - Based on social psychology (persuasion theory / captology techniques from advertisement and CRM

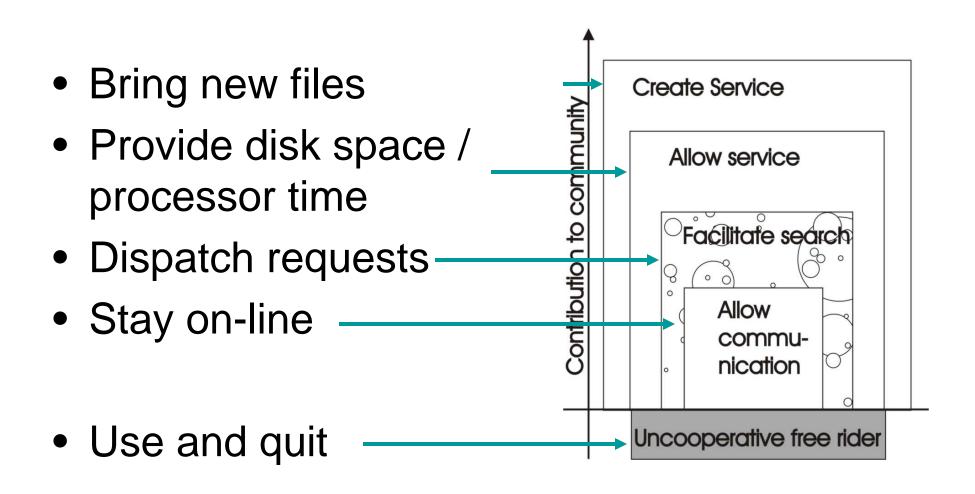
Important Condition: System Must be Useful

- Allow searching own files
 - Any file stored on disk can be found with Comtella
 - Shared files can be stored anywhere on disk
- Integration with other tools
 - With Browser (e.g. IE, Netscape, Mozilla, etc.)
 - allows viewing files directly from Comtella
 - prompts the user to share papers when opens a PDF file



- With Word Processor (e.g. MS Word)
 - generating lists of references automatically
- Additional functionality
 - Adding annotations and ratings

Levels of Participation



Ensuring Presence

Comtella 1

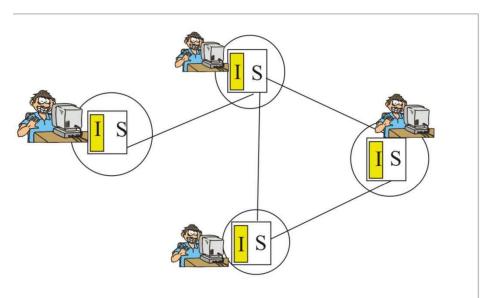
- Low presence was the greatest problem in Comtella 1
- Student activated their client only when they needed a resource and quitted afterwards
- Negative feedback loop

Comtella 2

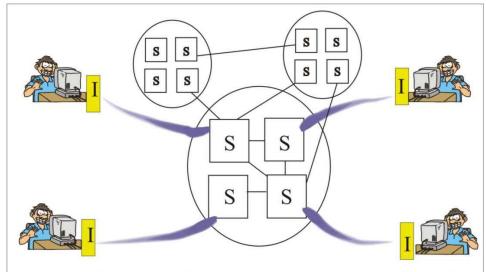
- Moved all clients on a server machine where they run all the time > the infrastructure is ensured, the shared files are available all the time
- Create another client a
 GUI which allows users
 to log in their client on the
 server

Solution = centralization?

Ensuring Presence: Architecture



Interface (I) is part of the Servent (S) which resides on the user machine. S interact through Gnutella.

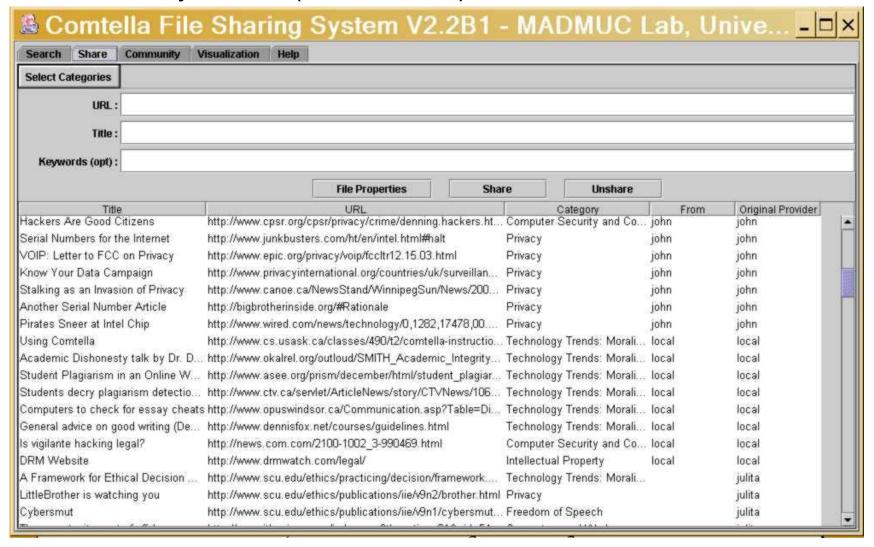


The Interface (I) is split from the servent (S); I resides on the user machine, S resides on a server; S interact among themselves with Gnutella; I interacts with S through socket.

Comtela 1 Comtella 2

omtella 2 in the Ethics Class

Share only URLs (bookmarks), not files



Ensuring New Contributions

Low-effort sharing

Comtella 1

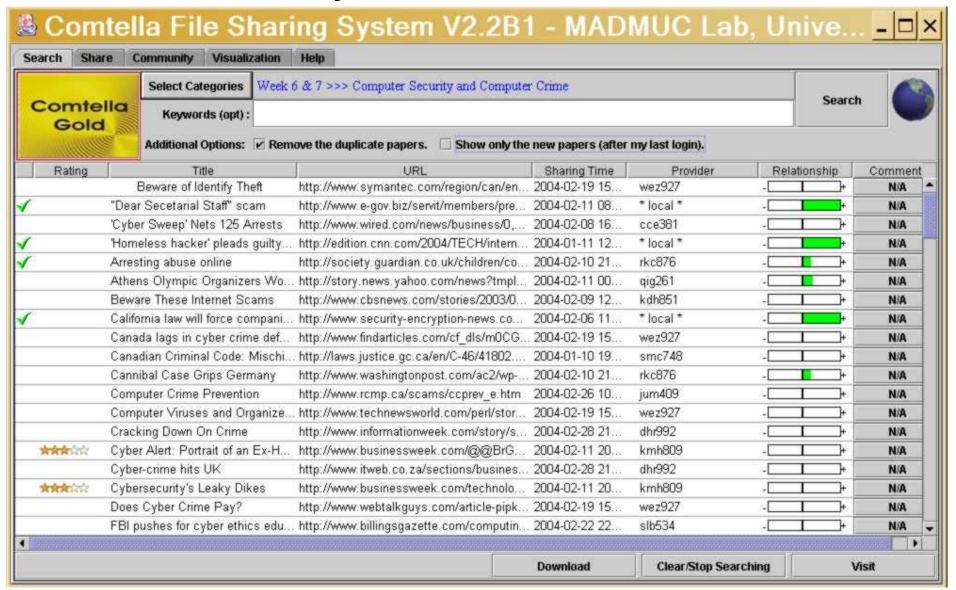
- One-click file sharing
- Automatic suggestion of category (text analysis of article)
- Automatic reminder to share if a PDF or PS file is opened in browser

Comtella 2

- Category selection is easy, typically users search and share for one given category each week
- Impossible to make the browser to automatically remind
- Students were required to share files for the class (participation mark)

Solution = integrate systems, prompt, help with paper annotation OR make it a policy to share

Search – by the theme of the week



Ensuring New Contributions

Rewarding Contributions

Comtella 1

- Count different forms of contributions
 - Number of new files brought in the system
 - Number of shared files
 - Disk space of shared files
 - Cooperative Uncooperative acts during downloads
 - Number of downloads vs number of uploads
- Reward in terms of better QoS (bandwidth → download speed)
- Result reward was not important really; speed was very good anyway

Comtella 2

- Count contributions:
 - Number of new links
 - Number of downloaded and shared links
 - Number of comments
 - Frequency of being on-line
- Reward in terms of access to more powerful search options
 - Remove duplicate results
 - See only new results
 - Sort results by various criteria
- Result seems to have been useful; about 64% usage of extra options

Solution = define appropriate metrics for participation; choose rewards that the students really care about

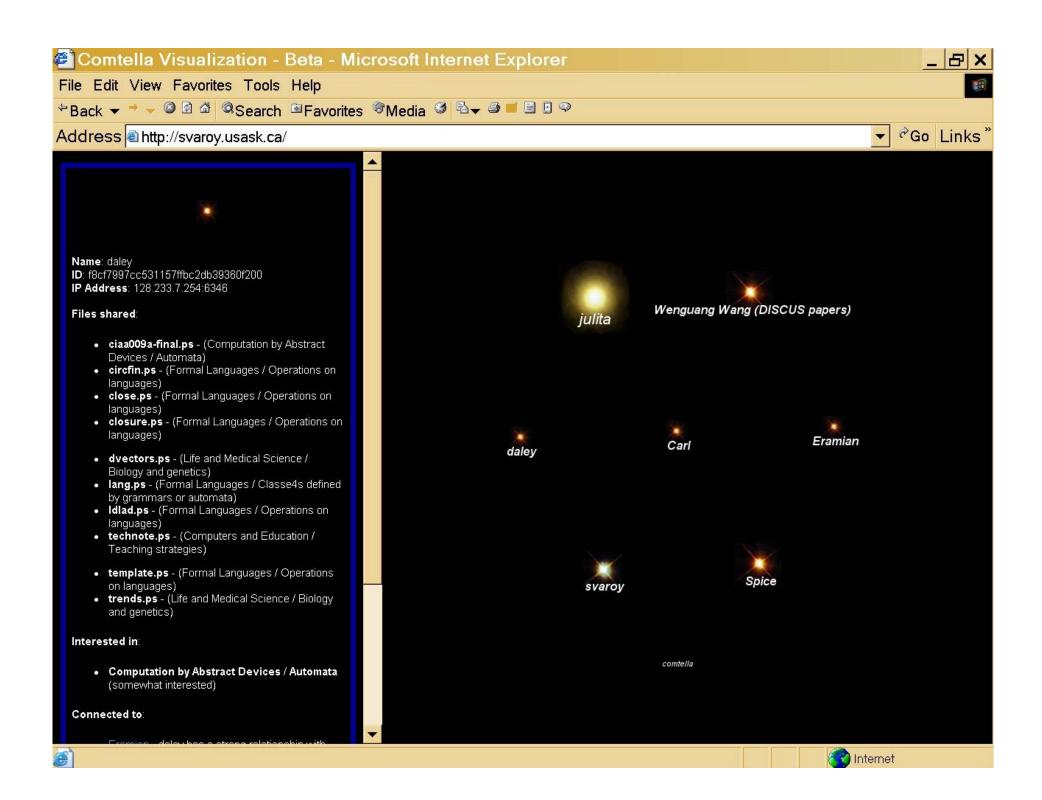
Social awareness

In isolation, selfishness is logical

To gain perspective, users need feedback about their social environment

In cities, the sidewalks provide the right kinds and numbers of interactions from which neighborhoods emerge.





Ensuring New Contributions

Social Visibility

Comtella 1

Show a visualization of community using <u>night-sky metaphor</u>.

- Only users who are currently active are shown
- Category of interest is not shown
- Colour, size and position of a star show the user's cooperativeness, level of contribution and groups of users exchanging frequently files.
- Size depends on who is currently on line (mistake!)
- Low level of participation leads to too few stars on the sky which is not motivating -- a negative feedback loop
- Misrepresentation? (Erickson)

Comtella 2

Show the whole community, the peers as nodes

- Those that are active at the moment – filled with colour
- Colour depends on user's rank (a combined participation metric)
- Size depends on the number of new contributed links
- Can be generated to show the contributions for a particular week (topic)
- The size can be used to show different aspects of participation (# new files, # total shared files, frequency of log-in)
- Students used the visualization, but only the default view

Solution = visualization that provides both community information and facilitates social comparison

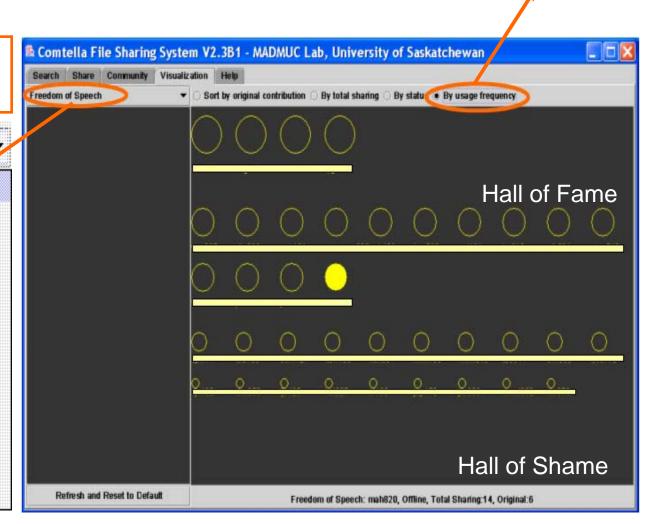
Ensuring New Contributions Social <u>Visibility</u>

Sorting Criteria Bar ->

Sort by original contribution \(\cap \) By total sharing \(\cap \) By status \(\cap \) By usage frequency.

List of all the available interest areas:

General View Technology Trends: Morality and Ethics Privacy Freedom of Speech Intellectual Property Wiretapping and Encryption Computer Security and Computer Crime Computers and Work Broader Social Issues Can we trust the Computer? Ethics and Professionalism



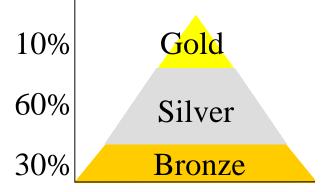
Ensuring New Contributions

Introducing Status in Comtella 2









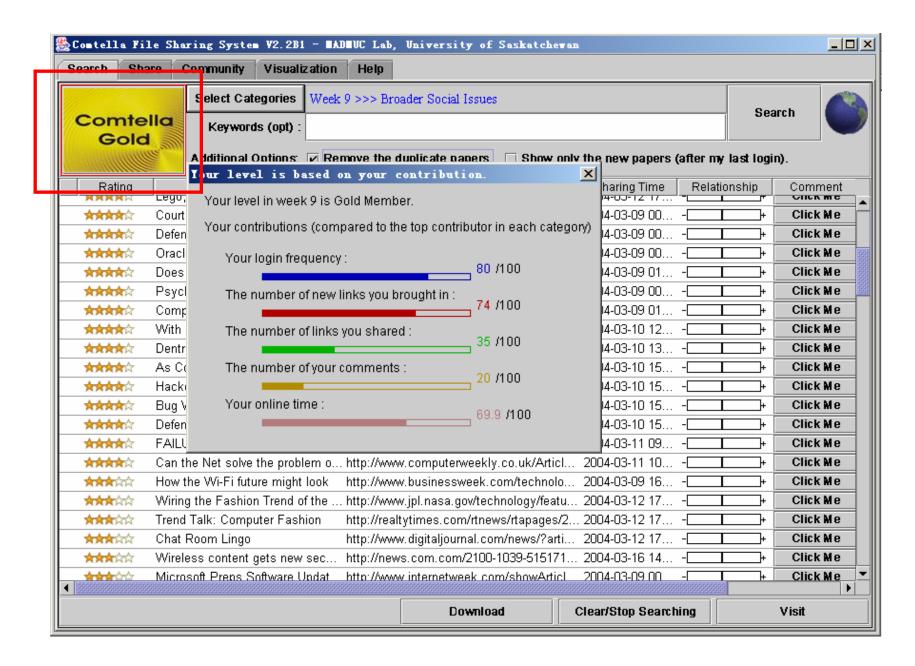
- Persuasion strategy used in CRM
 - Examples: Club memberships, Air Miles etc.
 - Based on the theory of discrete emotions (fear) effective persuasion strategy

Status

- based on a combination of participation metrics
- visualized as a card (gold, silver or bronze) example
- High-status users are rewarded with
 - Visibility in the community (visualization by status)
 - Better search options for gold and silver members

Solution = introduce a notion of social status, combined with visualization and rewards

Ensuring New Contributions Rewarding Contributions



Experiments

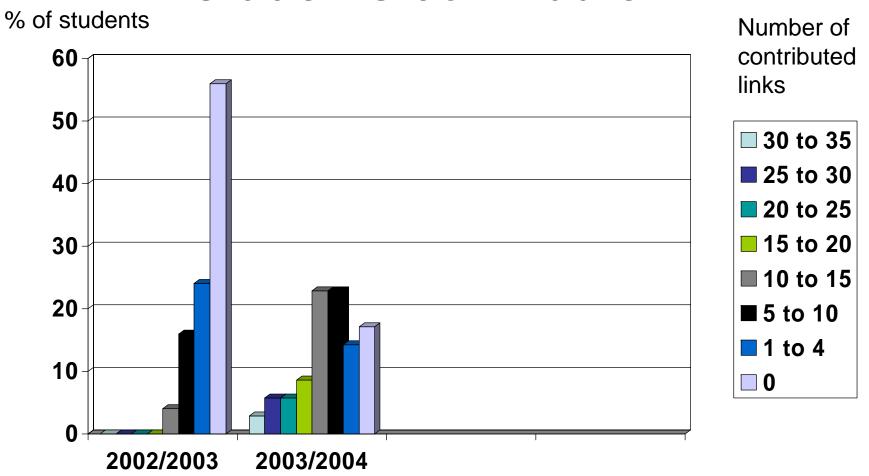
- 1. Is P2P useful in the classroom?
 - Compare two offerings of the Ethics in CS class, in 2003 and 2004, one without and one with Comtella 2.
 - Metrics: # papers brought in by students, regularity of contributions
- 2. Are the methods for encouraging participation effective?
 - Compare the number of contributed papers when using a baseline version and motivational version of Comtella 2.
 - The same student group, first without and then with treatment.

Question 1. Is Comtella useful in the classroom?

 Comparing two offerings of the class: one using personal websites and one using Comtella 2 (Jan 04 – Feb 29, 2004)

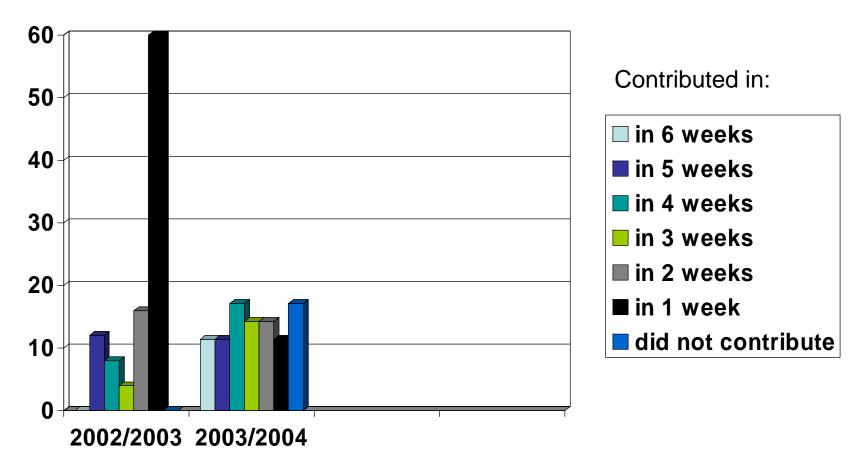
\ Class offering Parameters \	2002/03 - first 6 themes (using personal websites)	2003/04 - first 6 themes (using Comtella)
Nr of students in class:	25	35
Total Nr. of contributed links:	51	334
Average Nr. of new links / person (from those who contributed)	4.63	11.52
Nr of students who did not contribute at all:	14 (56% of the class)	6 (17% of the class)
% links contributed by the top five (5) students:	78% of all the contributed links	39% of all contributed links

Results: how many new links did students contribute



Results: how regularly did students contribute?

% of students



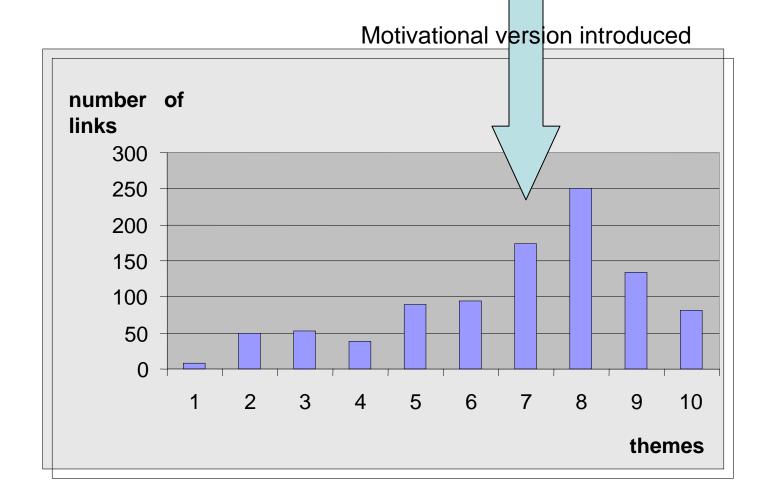
Discussion

- Comtella made it much easier
 - to share links
 - to find links shared by others and share on the go while viewing the already shared links
- The search facility was very useful
- Timeliness was important
 - shared papers are immediately available

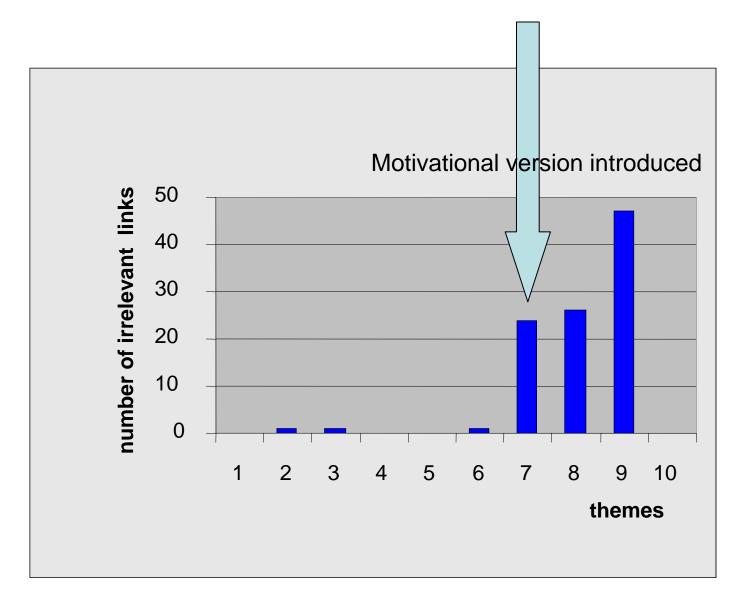
Question 2: Are our methods for encouraging participation effective?

- Compare student contributions before and after introducing the motivational version (with status, extra functionality for gold and silver, and social visualization)
 - Number of new links shared
 - Number of irrelevant links shared (gaming)

Results



Total number of links shared by all students over the 10 themes (weeks)



Number of irrelevant links contributed (cheats).

How to ensure good quality of contributions?

- Comtella 3 (first deployment this fall in a 4th year class on MAS, 15 st.)
- Entirely centralized: web-application: JSP, mySQL
- Measure quality of papers
 - Track "cheaters": let people flag irrelevant papers
 - Track how many people view a contribution and how they rate it (like "Impact factor")
 - Encourage users to rate contributions
 - Compute and display average ratings of papers
 - Compute a reputation of a user as a contributor and as a rater
 - User status depends both on the number of contributions and on reputation
- Reward quality users with high reputation are moderators (~Slashdot)
 - Higher status users have more rates to give away
 - Users gain status mostly based on their reputation
- Encourage users to rate papers
 - Each rating of a paper earns a c-unit that can be invested to increase the visibility of a contributed paper in the default view of search results

Conclusions

- P2P file-sharing can support collaborative collection of a shared class resource repository
 - Finally, a legal application of P2P!
 - Success (confirmed by students in a questionnaire after the end of class)
 - Students suggested that the system is used widely in many CS and humanities classes
- Persuasion techniques can be effective in designing incentive mechanisms
 - Rewarding students (reciprocation)
 - Hierarchical membership (theory of fear)
 - Visualization (social comparison)
- Be careful!
 - Every incentive system can be and will be gamed
 - If no attempts of gaming, it is not effective

More info: at http://bistrica.usask.ca/madmuc/peer-motivation.htm