



# Process and Community Analysis



**Chris Jensen and Walt Scacchi**  
**Institute for Software Research**  
**School of Information and Computer**  
**Science**

**University of California, Irvine**  
**{cjensen, wscacchi}@ics.uci.edu**

25th May 2004  
Edinburgh,  
Scotland, UK

*ICSE-MSR 2004*

# Current Research Areas



- **Types of artifacts**
  - **Source repositories**
  - **Issue/Defect repositories**
  - **Communication archives**
  - **Others**
- **Types of analysis**
  - **Social network analysis**
  - **Collaboration Awareness**
  - **Process discovery**

# Social Network Analysis



- **Work by some folks here**
  - **Committer networks and Module networks (Lopez-Fernandez, Robles, and Gonzalez-Barahona)**
  - **Artifact and activity relationships (Froehlich)**
- **Work by some folks elsewhere**
  - **Developer and project networks (Madey)**

# **Social Network Analysis and CVS Repositories**



- **Promising potential**
  - **Highlights relevant information**
  - **May give insight into software development team organization and structure**
  - **May allow to explain further how libre software development takes place**
  - **SNA is a topic where already a vast literature exists**

# **Social Network Analysis and CVS Repositories**



- **Salient challenges**
  - **Some interactions cannot be tracked through software repositories as they only contain some information of the whole development process**
  - **Not all software development projects use CVS in the same way (which will make difficult to compare results)**
  - **Ethical issues (anonymity)**

Courtesy of GSyC Luis Lopez, Jesus M. Gonzalez Barahona and Gregorio Robles





# Social Network Analysis and CVS Repositories



- **Augur**
  - **Promising potential**
    - **Unifies a semantic understanding of source code with activity context**
  - **Salient challenges**
    - **Inability to monitor relationships across organizations/repositories**

# Collaboration Awareness

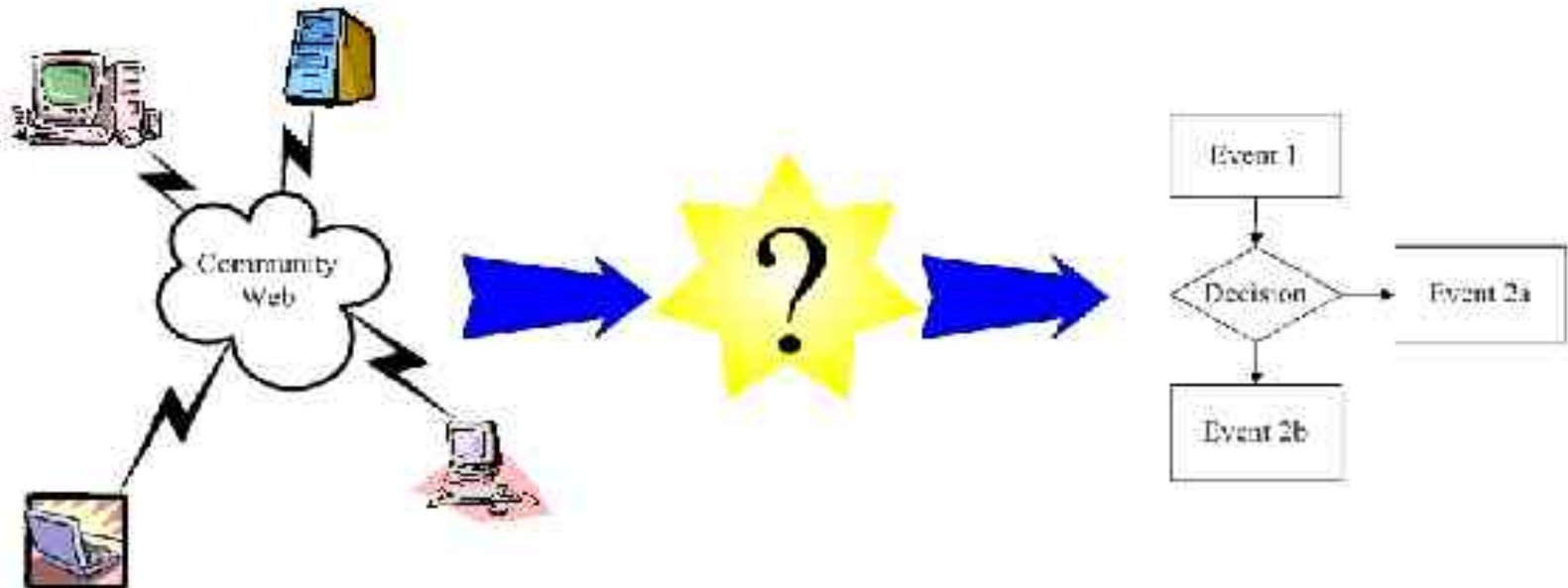


- **Project Watcher (Schneider, Gutwin, Penner, Paquette)**
  - **Distributed collaboration awareness in code repositories**
- **Promising potential**
  - **Awareness of group interaction patterns provides opportunities for increased activity coordination**
- **Salient challenges**
  - **Limited ability to make assertions about the development process**



# Process Discovery

- Correlation of repository artifacts that evidence the software process that produces/consumes them to infer activity sequences



# Process Discovery



- **Promising potential**
  - Reduce the entry barrier to process participation
  - Enables analysis, recovery, and redesign of processes in repositories too vast for manual discovery
- **Salient challenges**
  - How can we bring together analysis tools such as those presented here for more complex analysis?
  - How do we correlate data across multiple artifact types/formats?

# Selected Paper



**Lopez-Fernandez, Robles, and Gonzalez-Barahona “Applying Social Network Analysis to the Information in CVS Repositories”**

# Research Questions



- **Where do we go from here?**
- **What types of questions might we ask?**
- **What types of answers might practitioners be looking for?**
- **How can this work be brought to practitioners?**