Predicting Build Failures using Social Network Analysis on Developer Communication

Timo Wolf
Adrian Schröter
Daniela Damian
Thanh Nguyen

SIEMENS AG
Corporate Technology Software & Engineering

Software Engineering Global interAction Lab – SEGAL
Computer Science Department
University of Victoria, BC Canada
Introduction
Introduction

Build Failure
- Compile Error
- Test Failure
The damn build broke again! WHY?

Build Failure
- Compile Error
- Test Failure
The damn build broke again! WHY?

DB changed!
But that’s transparent.
Analyses takes time!
We should have talked earlier!

Build Failure
- Compile Error
- Test Failure
• Collaboration essential in software development

• Communication problems lead to coordination and integration failures

Build Failure
- Compile Error
- Test Failure

The damn build broke again! WHY?

DB changed!
But that’s transparent.
Analyses takes time!
We should have talked earlier!
Collaboration essential in software development

Communication problems lead to coordination and integration failures

Do specific communication behaviors enable effective and successful coordination?
Communication for collaboration
Communication for collaboration

- Communication:
- Task related content
Communication for collaboration

- Communication:
  - Task related content
  - Structure
Communication for collaboration

• Communication:
  • Task related content
  • Structure

Our focus: Communication Structure and relationship to coordination outcome
Study Settings: IBM Jazz
Study Settings: IBM Jazz

Ready for IBM Rational software & IBM Research
Study Settings: IBM Jazz

151 developers

47 functional teams
Study Settings: IBM Jazz

7 development sites in USA, Canada, Europe
Study Settings: IBM Jazz
Coordination & Integration in Jazz
Coordination & Integration in Jazz
Coordination & Integration in Jazz
Coordination & Integration in Jazz
Coordination & Integration in Jazz

Team A

Contributor 1

Contributor 2

Contributor 3

Team A Stream

Build 1

Team B

Build 1

Team B Stream

Contributor 4

Contributor 5

Contributor 6

commits changes

Change Sets

Contributor 4

Contributor 5

Contributor 6
Coordination & Integration in Jazz

Team A
Contributor 1
Contributor 2
Contributor 3

Team A Stream
Build 1

Project Integration Stream
Project Build 1
publish changes
get changes

Team B
Contributor 4
Contributor 5
Contributor 6

Team B Stream
Build 1
commits changes
Change Sets

8
Coordination & Integration in Jazz

Coordination Outcome: Build Result
Research Questions
Research Questions

![Diagram showing network connections and files with 'OK' and 'ERROR' labels.](image-url)
Research Questions

- SNA Measures: e.g. Density = 0.8

- SNA Measures: e.g. Density = 0.4

OK

ERROR

measure ➔ predict

measure ➔ predict
Research Questions

- **RQ1**: Can individual measures of communication structure predict integration failure?
Research Questions

- **RQ1**: Can individual measures of communication structure predict integration failure?
- **RQ2**: Can the combination of communication structure measures predict integration failure?
Constructing Social Networks in Jazz
Constructing Social Networks in Jazz

Previous Build

Build N Failed
Constructing Social Networks in Jazz

Previous Build

Build N Network

Build N Failed

C1 -> C2 -> C3

C1 -> C4

C2 <-> C3

C2 <-> C4

C1 <-> C4
Constructing Social Networks in Jazz

Build N Network

Previous Build

Build N Failed

C1

C2

C3

C4
Constructing Social Networks in Jazz
Constructing Social Networks in Jazz

Previous Build

Build N Network

Build N Failed
Constructing Social Networks in Jazz

Social Network Analysis (SNA)

quantitative measures of communication structure

Build N Network

Previous Build

Build N Failed
SNA Measures

Density

\[
\frac{6}{6} = 1
\]

\[
\frac{4}{6} = 0.66
\]

Degree Centrality

Structural Holes

Betweeness Centrality
SNA Measures

Density

\[
\frac{6}{6} = 1
\]

\[
\frac{4}{6} = 0.66
\]
SNA Measures

Degree Centrality

Density

\[
\frac{6}{6} = 1
\]

\[
\frac{4}{6} = 0.66
\]

Betweeness Centrality

Structural Holes
SNA Measures

Betweenness Centrality

Density

\[ \frac{6}{6} = 1 \]

\[ \frac{4}{6} = 0.66 \]

Degree Centrality

Structural Holes
SNA Measures

Structural Holes

Density
6/6 = 1
4/6 = 0.66

Degree Centrality

Betweenness Centrality
Data Collection & Selection

- 1288 Build Results
- 25713 Work Items
- 71019 Comments

- Selected Build Results of 5 Teams
- Selected 3 Integration Builds (nightly, weekly, beta)
RQ1: individual measure prediction?
RQ1: individual measure prediction?
RQ1: individual measure prediction?

Social Network Measures

- Density for Completeness
- Degree Centrality for Activity
- Betweenness Centrality for Control
- Structural Holes for Redundancy
RQ1: individual measure prediction?

Social Network Measures
- Density for Completeness
- Degree Centrality for Activity
- Betweenness Centrality for Control
- Structural Holes for Redundancy

No statistically significant difference!
RQ2: multi measures prediction?
RQ2: multi measures prediction?
RQ2: multi measures prediction?

1. Train with all measures

Bayesian Classifier
RQ2: multi measures prediction?

1. Train with all measures
2. Input

Bayesian Classifier
RQ2: multi measures prediction?

1. Train with all measures
2. Input
3. Predict

Bayesian Classifier
RQ2: multi measures prediction?

1. Train with all measures

2. Input

3. Predict

Bayesian Classifier
Prediction Results
# Prediction Results

<table>
<thead>
<tr>
<th>ERROR</th>
<th>Team Builds</th>
<th>Project Builds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Recall</td>
<td>55%</td>
<td>75%</td>
</tr>
<tr>
<td>Precision</td>
<td>52%</td>
<td>50%</td>
</tr>
</tbody>
</table>
# Prediction Results

<table>
<thead>
<tr>
<th>ERROR</th>
<th>Team Builds</th>
<th>Project Builds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Recall</td>
<td>55%</td>
<td>75%</td>
</tr>
<tr>
<td>Precision</td>
<td>52%</td>
<td>50%</td>
</tr>
</tbody>
</table>
# Prediction Results

<table>
<thead>
<tr>
<th>ERROR</th>
<th>Team Builds</th>
<th>Project Builds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Recall</td>
<td>55%</td>
<td>75%</td>
</tr>
<tr>
<td>Precision</td>
<td>52%</td>
<td>50%</td>
</tr>
</tbody>
</table>

- Recall:
  - 62 of 100 failed **builds** are classified correctly
# Prediction Results

<table>
<thead>
<tr>
<th>ERROR</th>
<th>Team Builds</th>
<th>Project Builds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Recall</td>
<td>55%</td>
<td>75%</td>
</tr>
<tr>
<td>Precision</td>
<td>52%</td>
<td>50%</td>
</tr>
</tbody>
</table>

- **Recall:**
  - 62 of 100 failed **builds** are classified correctly

- **Precision:**
  - 75 of 100 failure **predictions** are actual failures
  - 75% confidence that a prediction is correct
Summary
Summary

- Quantitative empirical evidence: Communication structures make a difference!
Summary

• Quantitative empirical evidence: Communication structures make a difference!

• Communication structures:
  • impact integration build results
  • are key for successful collaboration
Summary

- Quantitative empirical evidence: Communication structures make a difference!
- Communication structures:
  - impact integration build results
  - are key for successful collaboration
- Single measures do not predict build results
Summary

• Quantitative empirical evidence: Communication structures make a difference!

• Communication structures:
  • impact integration build results
  • are key for successful collaboration

• Single measures **do not** predict build results

• Combination measures **do** predict build results
Practical Implications

- Communication assessment for future builds
- Notification system for build failures
- Communication structure improvement
Future Work

- Factor analysis of combined measures
- Include technical information
- Prediction effectiveness case study
Questions?

Timo Wolf
timowolf@siemens.com

Daniela Damian
danielad@cs.uvic.ca

Adrian Schröter
schadr@uvic.ca

Thanh Nguyen
duy thanh@uvic.ca

SIEMENS AG
Corporate Technology
Software & Engineering

http://segal.uvic.ca
Backup Slides
Descriptive Statistics 1

<table>
<thead>
<tr>
<th></th>
<th>Team Level Builds</th>
<th>Project Level Builds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td># Builds</td>
<td>60</td>
<td>48</td>
</tr>
<tr>
<td># ERRORS</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td># OKs</td>
<td>40</td>
<td>32</td>
</tr>
</tbody>
</table>

| # Contributors:       |     |     |     |     |     |     |     |     |
| Min                   | 3   | 9   | 6   | 5   | 13 | 43   | 37   | 55   |
| Median                | 6   | 16.5| 18  | 15  | 20 | 55   | 57   | 69.5 |
| Mean                  | 12.68| 18.02| 20.15| 17.98| 22.87| 57.93| 52.27| 67.81|
| Max                   | 58  | 31  | 64  | 61  | 52 | 75   | 75   | 79   |

| # Directed Connections: |     |     |     |     |     |     |     |     |
| Min                   | 0   | 1   | 2   | 0   | 11 | 81   | 56   | 144  |
| Median                | 13  | 39.5| 95  | 36  | 74 | 236  | 149  | 280  |
| Mean                  | 51.58| 53.4 | 87.78 | 63  | 88.35| 253.1| 171.9| 285.8|
| Max                   | 361 | 139 | 355 | 401 | 300| 434  | 496  | 446  |
# Descriptive Statistics II

<table>
<thead>
<tr>
<th></th>
<th>Team Level Builds</th>
<th></th>
<th>Project Level Builds</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td># Change Sets:</td>
<td>Min</td>
<td>1</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>10</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>10.83</td>
<td>44.38</td>
<td>42.65</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>33</td>
<td>101</td>
<td>91</td>
</tr>
<tr>
<td># Work Items:</td>
<td>Min</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>6.5</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>16.43</td>
<td>15.56</td>
<td>23.07</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>131</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>