Talking About Concerns . . .

James D. Herbsleb
School of Computer Science
Carnegie Mellon University
What is Modularity?

- Thanks, Mary!
- Thanks, Dick!
Why Modularity?

- **Software** modularity does not matter at all.
- Except...
  - To the extent it modularizes **work**
- **Work** modularity aids human understanding
- **Work** modularity simplifies coordinating people and teams
Parnas: Expected Benefits of Modularity

• Reduce need for coordination
  • “separate groups would work on each module with little need for communication”

• Simplify comprehension
  • “it should be possible to study the system one module at a time”

• These effects lower the cost of change
  • “it should be possible to make drastic changes to one module without a need to change others”

Vision . . .

- “a vivid mental image; ‘he had a vision of his own death’” *
- “an Explanation of Life Founded upon the Writings of Giraldus and upon Certain Doctrines Attributed to Kusta Ben Luka” *
- “a thought, concept, or object formed by the imagination” **
- “direct mystical awareness of the supernatural“ **

*wordnetweb.princeton.edu/perl/webwn  **Merriam-Webster Dictionary
Proportion of dependencies that cross-cut

Cognition and coordination problems

100%  

Number of language-based modularizing mechanisms
Proportion of dependencies that cross-cut

100%

Traditional modularity

Cognition and coordination problems

Number of language-based modularizing mechanisms
Proportion of dependencies that cross-cut

100%

Number of language-based modularizing mechanisms

Aspects
Cognition and coordination problems
Proportion of dependencies that cross-cut

100%

Cognition and coordination problems

Number of language-based modularizing mechanisms
Proportion of dependencies that cross-cut

100%

Traditional modularity

Cognition and coordination problems

Number of language-based modularizing mechanisms

Consensus view at Recife
Proportion of dependencies that cross-cut

100%

Aspects

Cognition and coordination problems

Number of language-based modularizing mechanisms

Consensus view at Recife
Proportion of dependencies that cross-cut

100%

Cognition and coordination problems

Dystopian vision:
Modularity alone will never fix the problem.

My view (mildly exaggerated)
Approaching the Gray Area . . .

- Organizational design, work assignment, and tools set up to bring the right dependencies to the attention of the right people so they can act appropriately
Two Examples . . .

• Organizational design and work assignment
  – Lessons from feature-driven development

• Using information from the environment
  – Learning from human activity
Feature-Driven Development

• Unit of functionality in end-user terms
• Feature is the unit of development managed by a project
• Features tend to cut across traditional software entities
• Work often overseen by “feature manager”
• Developers associated with component, assigned to work on particular features
The Study

• Setting
  – Software for automotive navigation system
  – 1195 features
  – 32 months of activity
  – 179 engineers in 13 teams
  – 1.5 M LOC, 6789 source files, 107 architectural components
  – Organization had 5 years of prior experience with feature-driven development

• Architects prepare feature development specification
## What Causes Integration Failure?

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.992*</td>
<td>0.990*</td>
<td>0.990*</td>
<td>0.989*</td>
</tr>
<tr>
<td>Average Component Experience (log)</td>
<td>0.487*</td>
<td>0.984+</td>
<td>0.741+</td>
<td>0.754</td>
</tr>
<tr>
<td>Changed LOCs</td>
<td>1.021</td>
<td>1.089</td>
<td>1.063</td>
<td></td>
</tr>
<tr>
<td>Concentration of Changed LOCs</td>
<td>1.045</td>
<td>1.028</td>
<td>1.036</td>
<td></td>
</tr>
<tr>
<td>Number of Dependencies (log)</td>
<td>1.107*</td>
<td>1.091*</td>
<td>1.091*</td>
<td></td>
</tr>
<tr>
<td>Concentration of Number of Dependencies</td>
<td>1.032**</td>
<td>1.046**</td>
<td>1.078**</td>
<td></td>
</tr>
<tr>
<td>Number of Groups</td>
<td></td>
<td></td>
<td>1.101*</td>
<td>1.051*</td>
</tr>
<tr>
<td>GSD</td>
<td></td>
<td></td>
<td></td>
<td>13.924**</td>
</tr>
<tr>
<td>Feature Owner Belongs to Highly Changed Component</td>
<td>0.789</td>
<td>0.396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature Owner Belongs to Highly Coupled Component</td>
<td>0.839**</td>
<td>0.819**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration of Changed LOCs X F. Owner Belongs to Highly Changed Component</td>
<td>1.032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration of Number of Dependencies X F. Owner Belongs to Highly Coupled Component</td>
<td>0.977**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSD X Feature Owner Belongs to Highly Changed Component</td>
<td>3.736</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSD X Feature Owner Belongs to Highly Coupled Component</td>
<td>0.926</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deviance of the Model: 755.2  639.0  458.4  412.2
Deviance Explained: 11.7%  25.3%  46.4%  51.8%

(+ p < 0.1; * p < 0.05; ** p < 0.01)

Odds Ratios from Regression Assessing Factors Driving Feature Integration Failures

Ownership Matters!

## Destructive Feature Interaction

<table>
<thead>
<tr>
<th></th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td>0.981**</td>
<td>0.971**</td>
<td>0.964*</td>
</tr>
<tr>
<td><strong>Failures in the Past 5 Weeks</strong></td>
<td>2.127**</td>
<td>1.125*</td>
<td>1.011*</td>
</tr>
<tr>
<td><strong>Changed LOCs</strong></td>
<td>1.371**</td>
<td>1.201**</td>
<td>1.203**</td>
</tr>
<tr>
<td><strong>Average Component Experience (log)</strong></td>
<td>0.837+</td>
<td>0.997</td>
<td>0.908</td>
</tr>
<tr>
<td><strong>Number of Groups</strong></td>
<td>3.006**</td>
<td>4.037**</td>
<td>6.345**</td>
</tr>
<tr>
<td><strong>Overlap Among Groups</strong></td>
<td>0.943**</td>
<td>0.919**</td>
<td>0.901**</td>
</tr>
<tr>
<td><strong>Same Feature Owner</strong></td>
<td>0.876**</td>
<td>0.871**</td>
<td>0.852**</td>
</tr>
<tr>
<td><strong>GSD</strong></td>
<td>4.501**</td>
<td>2.509**</td>
<td>4.895**</td>
</tr>
<tr>
<td><strong>Number of Cross-Feature Dependencies (log)</strong></td>
<td>2.911**</td>
<td></td>
<td>4.938**</td>
</tr>
<tr>
<td><strong>Number of Groups X Number of Cross-Feature Dependencies</strong></td>
<td></td>
<td>0.607</td>
<td></td>
</tr>
<tr>
<td><strong>GSD X Number of Cross-Feature Dependencies</strong></td>
<td></td>
<td></td>
<td>0.799**</td>
</tr>
<tr>
<td><strong>Deviance of the Model</strong></td>
<td>12873.9</td>
<td>9413.1</td>
<td>8043.1</td>
</tr>
<tr>
<td><strong>Deviance Explained</strong></td>
<td>33.4%</td>
<td>51.3%</td>
<td>58.4%</td>
</tr>
</tbody>
</table>

(+ p < 0.1; * p < 0.05; ** p < 0.01)

Odds Ratios from Regression Assessing the Impact of Cross-Feature Interactions on Integration Failures

Co-location Doesn’t Scale

Broader Lessons

- Organizational arrangements matter!
- Effects can be quite large
- Effects often are not commonsensical
Inferring Dependencies from Traces of Human Activity

• Prior work
  • Use files changed together as measure of dependencies
  • Can generate a measure of coordination requirements
  • Validated in a number of settings

• Can we generalize from “files changed together” to “entities discussed together”?
A Brief Digression/Analogy

Google labs Books Ngram Viewer

Graph these case-sensitive comma-separated phrases: machine learning, symbol system
between 1950 and 2008 from the corpus English with smoothing of 3.

Search for books:


Run your own experiment! Raw data is available for download here.
Text Analysis: Field Robotics

- Project
  - Lunar X Prize competition
Text Analysis: Field Robotics

- Project
  - Lunar X Prize competition
- No automatically collected version or change data
- Constantly shifting component boundaries and interfaces
- Can we use text analysis to derive dependencies?
Steps

• Collected data
  • 25 all-hands meetings
  • About 10,000 words each

• Developed code book
  • 6 field robotics articles
## Code Book

<table>
<thead>
<tr>
<th>Component</th>
<th>Brief description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications</td>
<td>Communications external to the robot, for control or mission, including operator interface.</td>
</tr>
<tr>
<td>External Relations</td>
<td>Acquiring external resources (incl. funding, parts, &amp; purchasing), publicity, investor &amp; media relations.</td>
</tr>
<tr>
<td>Internal Relations</td>
<td>Project/program management, HR, task assignments, training, collaboration tools, clarifying norms &amp; expectations.</td>
</tr>
<tr>
<td>Mobility Effectors / Actuators</td>
<td>Effectors and actuators that propel the entire robot: e.g. tracks, wheels, shocks, &amp; motors with associated firmware.</td>
</tr>
<tr>
<td>Mission Specific Effectors / Actuators</td>
<td>All other motors, gears, &amp; moving parts that don’t move the robot as a whole, e.g. camera mast rotation motor.</td>
</tr>
<tr>
<td>Perception software / computing</td>
<td>Software, and any dedicated hardware, for: terrain mapping, environmental modeling, and/or object detection. Camera/lens zoom, shutter, and focus control software.</td>
</tr>
<tr>
<td>Planning software / computing</td>
<td>Mission task planning, including the overall mission plan and computing resources for semi-autonomous execution.</td>
</tr>
<tr>
<td>Power</td>
<td>Includes batteries, solar cells, switches, power cables &amp; controls.</td>
</tr>
<tr>
<td>Sensors</td>
<td>Camera; thermal, ultrasonic, tactile, radar/sonar range sensors; Inertial Measurement Unit, GPS, &amp; any wiring or processing going from sensors to controls.</td>
</tr>
<tr>
<td>Shared / general computing</td>
<td>Includes general purpose processors / onboard computers (e.g. avionics box). Abbreviated “gpp.”</td>
</tr>
<tr>
<td>Structure</td>
<td>Chassis, fasteners (e.g. Frangibolt, weld joints), radiator, payload, paints, reflectors.</td>
</tr>
</tbody>
</table>
Steps

- Collected data
  - 25 all-hands meetings
  - About 10,000 words each

- Developed code book
  - 6 field robotics articles

- Manual coding of decision discussions
  - Tested inter-rater reliability
  - QAP correlations .80
## Text Pre-Processing

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Removed contractions (e.g. changing “what’s” to “what is”).</td>
</tr>
<tr>
<td>2</td>
<td>Applied a Krovetz (dictionary-based) stemmer to covert terms into morphemes.</td>
</tr>
<tr>
<td>3</td>
<td>Removed common English terms (e.g. ‘the’), replacing them with placeholders (‘xxx’).</td>
</tr>
<tr>
<td>4</td>
<td>Removed punctuation.</td>
</tr>
<tr>
<td>5</td>
<td>Turned meaningful bigrams into unigrams (e.g. ‘solar cells’ became ‘solar_cells’).</td>
</tr>
</tbody>
</table>
Steps

- Collected data
  - 25 all-hands meetings
  - About 10,000 words each
- Developed code book
  - 6 field robotics articles
- Manual coding of decision discussions
  - Tested inter-rater reliability
    - QAP correlations .80
- Created thesaurus
Link Identification

- Co-occurrence of terms
- Human coding: same decision
- Selected sliding window size
  - Size 15 had best agreement with hand coding
  - Threshold established
- QAP correlations comparable to human-human agreement (~.8)
- Sets of links based on topics
Optics

- External relations
- Structure
- Sensors
- Planning software
- Requirements
- Mission specific effectors
- Mobility effectors
- Perception software
- Communications
- Testing
Thermal

Thermal system

Mission specific effectors

Power

Thermal models

Structure

Prototype fabrication

Structural models
Avionics

- Mission operations
- Mobility effectors
- Mission-specific effectors
- Structure
- Propulsion
- Power
- Thermal system
- Lander
- Launch vehicle
- Perception software
- Shared/general computing
- Prototype fabrication
- Planning software
Concluding Vision

- The gray area – work that cross-cuts language constructs – is here to stay
- Use organizational tactics
- Use computations over artifacts generated by development activities
- Explore new data sources, including documents and conversation
  - Activities reveal knowledge
  - Analysis can often make it actionable