Complex System Simulation: Interactions of NOM Molecules, Mineral Surfaces, and Microorganisms in Soils

Greg Madey
Computer Science & Engineering (CSE)
College of Engineering
University of Notre Dame
Overview

- Modeling of Complexity
- Stochastic Simulation of Natural Organic Matter (NOM) and Environmental/Microbial Interactions
- Another study (NSF - CISE/IIS)
  - Agent-Based Model of the Open Source Software (OSS) Development Phenomenon
Simulation of NOM and Microbial-Environmental Interactions

- Funded in part by NSF-ITR (DEB)
  - Co-PIs
    - Robert Wershaw, USGS-Denver
    - Jerry Leenheer, USGS-Denver
    - Patricia Maurice, Geological Sciences, ND
      - USGS - 5 years, New Jersey
    - Steve Cabaniss, Chemistry, UNM
    - Robert Wetzel, Biology, UNC-Chapel Hill
  - Students
    - Yingping Huang, CSE, (MS Thesis)
    - Xiaorong Xiang, CSE, (MS Thesis)
    - Eric Chanwich, CSE, REU
Complexity refers to the dynamic web of interrelationships within physical, biological, geochemical, hydrological, environmental, ecological, social, economic, engineered systems, etc.

The study of complexity includes systems that range from molecular to global in scale, and exhibit properties that depend not only on the individual actions of their components, but also the interactions among those components.
Properties of Complex Systems

- Entities (Often Many & Heterogeneous)
- Entity Behaviors
- Interactions Between Entities Including Feedback (Possibly Nonlinear)
- Often Sensitive Dependence to Initial Conditions
- Self-Organization (Self-Assembly?)
- Emergence
Understanding Complex Systems

- Parts versus the Whole
  - Limitations of the Reductionist Approach
- Sensitive Dependence on Initial Conditions
  - Limits to Predictability
- Understand the Properties of the System
  - Stability
  - Periodicity
  - Chaotic
  - Bounded/Unbounded
  - System Behaviors, Processes, Mechanisms, and Interactions
Modeling Complex Systems

- Mathematical Analysis
  - Stochastic Analysis
- Computer Simulation (Computational “x”)
  - Iteration/Recursion
  - Numerical Methods
- Computer Simulation (Bottom-up Modeling)
  - Discrete-Event
  - Agent-Based Modeling (includes CA)
- Very Large Computer Databases
  - Sensor Arrays, NanoSensors, Data Warehouses
  - Data Mining and Knowledge Discovery
Agent-Based Modeling

- **Object-Oriented Paradigm**
  - Entities are Objects (Agents)
  - Objects have: Attributes (data) & Behaviors (methods)
  - Classes of Objects (heterogeneous)
  - Inheritance/Polymorphism

- **Simulation Process**
  - Model Entities with Classes: Attributes & Behaviors
  - Instantiate (and destroy) Objects (Agents)
  - Model the Environment of the Objects (Agents)
  - Object Behaviors generate Interactions with Environment and other Objects (Agents)
  - Store State Information in Database
  - Post-Simulation Analysis (Data Mining/Knowledge Discovery)
Agent-Based Modeling Tools

- Object Oriented Languages: C++, Java, Objective-C, SmallTalk
- Simulation Libraries (Class Packages)
  - SWARM
  - RePast
- Simulation Environments
  - Starlogo, StarLogoT, NetLogo
  - Agent Sheets
  - AScape
  - Integrated Modeling Toolkit (IMT)
SWARM

- Agent-Based Modeling Library
- Open Source / Started at Santa Fe Institute - Chris Langton, A-Life
- ObjectiveC and Java
- Swarms
  - Collections of Agents
  - Swarms can be modeled hierarchically
  - Sub-Swarms
Meeting Announcement

Seventh Annual Swarm Users/Researchers Meeting

Notre Dame, Indiana USA

April 13 - 15, 2003

- The Registration Page will be available shortly.
- Mark your calendars!
- Call for papers, tutorials, and posters!

The Department of Computer Science & Engineering at the University of Notre Dame, and the Development Group (SDG) are pleased to announce the Seventh Annual SWARM Users/Researchers Meeting to be held on the campus of the University of Notre Dame, on April 13-15, 2003.

At Swarm 2003, scientists, modellers, and programmers working in a wide variety of domains have an opportunity to share their research, knowledge and experience with multi-agent modelling including, but not restricted to) the Swarm simulation system. The Swarm Development Group also uses the meeting to determine future development priorities, so come and tell us what's on your mind! The meeting is a section of the community, historically including content suitable for both inexperienced and experienced users and researchers.
Simulation of NOM and Microbial-Environmental Interactions

- NSF - ITR - Division of Environmental Biology
- Interdisciplinary project
  - Chemist
  - Geomicrobiologist
  - Biologist
  - Ecologist
  - Computer Scientist
- Stochastic Simulation of Environmental Transformations of Natural Organic Matter
  - In soil
  - In solution
Natural Organic Matter

- Ubiquitous in terrestrial, aquatic and marine ecosystems
- Important role in compositional evolution and fertility of soil
- Impacts mobility and transport of pollutants, e.g., trace metals, radionuclides and hydrophobic organic compounds
- Impacts availability of nutrients for microorganisms and plant communities
- Impacts growth and dissolution of minerals
Natural Organic Matter (cont)

Hardwood Swamp
Natural Organic Matter (cont)

Open Channel
Natural Organic Matter (cont)

Cedar Swamp
Focus of the Modeling

- Soil
- Ground Water
- Surface
- NOMs - Microbes
- Water
- Ground Water
Compositional evolution of NOM is an interesting problem important aspect of predictive environmental modeling. Prior modeling work is often too simplistic to represent the heterogeneous structure of NOM and its complex behaviors in ecosystems (e.g., carbon cycling models), also … Prior modeling work is often too compute-intensive to be useful for large-scale environmental simulations (e.g., molecular models employing connectivity maps or electron densities). Hence, a Middle Computational Approach is taken …
Project Goals

- Stochastic model of NOM evolution — middle computational approach
  - Algorithms
  - Parameters
- Represent individual molecules as discrete objects with
  - Specified elemental and functional group composition
  - Size/weight
  - Reactivity
- Model the evolution of NOM from biological precursor compounds
  - Lignins
  - Polysaccharides
  - Proteins
- Deploy web-based simulation for testing, feedback and usage
- Generate experimentally testable predictions about NOM evolution and properties - validation of simulation
Generate Experimentally Testable Predictions about NOM Evolution and Properties

Molecular weight is a key property controlling the reactivity of NOM.

(From Cabaniss et al., 2000)
Modeling

- Molecules and microbes are objects
- Molecules and microbes have attributes
  - Heterogeneous, distributions
  - Currently 1,000 objects, preferably 10,000 or more
- Molecules have behaviors (reactions)
  - Molecules in simulation are a representative sample of the larger population
  - Behaviors are stochastically determined
  - Dependent on the:
    - Attributes (intrinsic parameters)
    - Reaction rates
    - Environment (extrinsic parameters)
Modeling (cont)

- Objects of interest
  - Macromolecular precursors
    - Polysaccharides
    - Proteins
    - Polynucleotide, tannin, lignin, polyterpene, cutin
  - Smaller molecules
    - Phospholipids
    - Sugars
    - Amino acids
    - Flavinoids
    - Quinones
  - Microbes
Modeling (cont)

- Attributes
  - More specific than “percent carbon” but less detailed than a molecular connectivity map
  - Elemental composition
    - Number of C, H, O, N, S and P atoms in molecule
  - Functional group counts
    - Double-bonds
    - Ring structures
    - Phenyl groups
    - Alcohols
    - Phenols, ethers, esters, ketones, aldehydes, acids, aryl acids, amines, amides, thioethers, thiols, phosphoesters, phosphates
  - The time the molecule entered the system
  - Precursor type of molecule
Modeling (cont)

- Behaviors (reactions and processes)
  - Physical reactions
    - Adsorption to mineral surfaces
      - Initial adsorption
      - Surface migration to high-energy sites
      - Hemi-micelle formation at high coverage (cooperative, hydrophobicity dependent)
    - Aggregation/micelle formation (e.g., metal cation-induced aggregation) - flocs
    - Transport downstream (surface water)
    - Transport through porous media
    - Volatilization
AFM Image of NOM Adsorption

Maurice, 1999
Modeling (cont)

- Behaviors (reactions and processes)
  - Chemical reactions
    - Abiotic bulk reactions
      - Hydrolysis
      - Hydration
      - Ester condensation
      - Thermal decarboxylation
    - Abiotic surface reactions
  - Direct photochemical reactions
  - Indirect photochemical reactions
  - Extracellular enzyme reactions on large molecules
    - Bacteria
    - Fungi
    - Algae
  - Microbial uptake by small molecules
Modeling (cont)

- Environmental parameters
  - Temperature
  - pH
  - Light intensity
  - Metal concentrations (e.g., Al(III) and Fe)
  - Bacterial activity
  - Water flow rate/pressure gradient
  - Surface area
NOM 1.0

- GUI Version - Stand Alone
  - Simulation and Animation of Molecules
- Web-Based Collaboratory
  - Standard Browser Interface
    - HTML Forms / JSP
    - Java Servlets
    - JDBC - Oracle Database
    - Oracle Forms and Reports
- Model Development
- Shared Data and Simulations
- Collaboration Support
GUI Animation

Black - No Adsorption
Greys - Levels of Adsorption
Red - Lignins
Blue - Proteins
Green - Cellulosic
Yellow - Reacted
Orange - Adsorbed
Welcome to NOM Research Group!

You must sign in to use the simulator!

Existing Users
Enter your userid and password to sign in

Userid: [ ]
Password: [ ]

Sign In

New users? Sign up now
NOM Simulator

Welcome to NOM Research Group! X Y

NOM Simulator: Reports

Currently, you have the following sessions invoked. The first one is your most recent session. You can view reports for each session by clicking the following links. To start a new simulation, click here. To cleanup terminated sessions, click here.

- **Session 115**: Terminate Session
  - Reactions Reports

- **Session 114**: Terminate Session
  - Reactions Reports

- **Session 113**: TERMINATED
Welcome to NOM Research Group! X Y

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Environment</th>
<th>Molecules</th>
<th>Summary</th>
</tr>
</thead>
</table>

**NOM Simulator: Introduction**

To properly use the simulator, we need to gather data for environment and molecule types.

The wizard will walk you through several tasks:

- Provide environment variables. If you provided environment variables before, we will retrieve your information to let you edit.
- Provide molecule types and number of molecules of this type. You can also edit and delete your saved molecule information.
- Invoke the simulation
**NOM Simulator**

Welcome to NOM Research Group! X Y

---

**Introduction**

**Environment**

**Molecules**

**Summary**

---

**NOM Simulator: Environment**

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Microbe Density:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time(days):</td>
<td>0.0010</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fungal Density:</th>
<th>pH Value:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0010</td>
<td>7.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature:</th>
<th>PKW:</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>14.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Oxygen:</th>
<th>Light Density:</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0E-4</td>
<td>4.0E-6</td>
</tr>
</tbody>
</table>

**Environment Information**

Please provide the environment variables for your simulation. You may also edit your environment variables here. Before submitting the form, please ensure that all the fields are filled in, whether they be integers or doubles. If you have already provided environment variables, you may choose to skip this step.

**Skip & Next**

step 2 of 4 **Save & Next**
# NOM Simulator: Molecule

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Cellulose</th>
<th>Lignin</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Atom) C</td>
<td>360</td>
<td>400</td>
<td>240</td>
</tr>
<tr>
<td>(Atom) H</td>
<td>602</td>
<td>322</td>
<td>332</td>
</tr>
<tr>
<td>(Atom) N</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>(Atom) O</td>
<td>301</td>
<td>81</td>
<td>76</td>
</tr>
<tr>
<td>(Atom) S</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(Atom) P</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Double Bond</td>
<td>60</td>
<td>199</td>
<td>59</td>
</tr>
<tr>
<td>Total Ring</td>
<td>60</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>Structures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenyl Groups</td>
<td>0</td>
<td>40</td>
<td>5</td>
</tr>
</tbody>
</table>

**Known Molecule Information**

There are three types of already defined Molecule, please give the percentage of each. Give a value 0 for percentage if you don't want to include this molecule type in your simulation. If you do not want to include any of these three types of molecules, you may click the Skip & Next button, otherwise, please click the Save & Next button. Default values are 0.
<table>
<thead>
<tr>
<th>Phenyl Groups</th>
<th>Alcohols</th>
<th>Phenols</th>
<th>Ethers</th>
<th>Esters</th>
<th>Ketones</th>
<th>Esters</th>
<th>Aldehydes</th>
<th>Acids</th>
<th>Aryl Acid</th>
<th>Amines</th>
<th>Ring N</th>
<th>Amides</th>
<th>Thioethers</th>
<th>Thiols</th>
<th>Phosphoesters</th>
<th>H-phosphoesters</th>
<th>Phosphates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>182</td>
<td>0</td>
<td>119</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>1</td>
<td>118</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Values are 0.
**NOM Simulator: Molecule**

<table>
<thead>
<tr>
<th>Molecule Name</th>
<th>Percentage: 29.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Atom) C: 0</td>
<td>(Atom) H: 0</td>
</tr>
<tr>
<td>(Atom) N: 0</td>
<td>(Atom) O: 0</td>
</tr>
<tr>
<td>(Atom) S: 0</td>
<td>(Atom) P: 0</td>
</tr>
</tbody>
</table>

**Molecule Information**
Please provide molecule's name, percentage, number of atoms, and molecules for your simulation. Please remember, except "Molecule Name", all fields should be integers or doubles. "Percentage" should be between 0 and 100.

<table>
<thead>
<tr>
<th>Doublebond: 0</th>
<th>Arylacids: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rings: 0</td>
<td>Amines: 0</td>
</tr>
<tr>
<td>Phenyl: 0</td>
<td>RingN: 0</td>
</tr>
<tr>
<td>Alcohols: 0</td>
<td>Amides: 0</td>
</tr>
<tr>
<td>Phenols: 0</td>
<td>Thioethers: 0</td>
</tr>
<tr>
<td>Ethers: 0</td>
<td>Thiols: 0</td>
</tr>
<tr>
<td>Esters: 0</td>
<td>Phosphoesters: 0</td>
</tr>
<tr>
<td>Ketones: 0</td>
<td>HPhosphoesters: 0</td>
</tr>
<tr>
<td>Aldehydes: 0</td>
<td>Phosphates: 0</td>
</tr>
<tr>
<td>Acids: 0</td>
<td></td>
</tr>
</tbody>
</table>

**Functional Groups**
Please provide a number for each functional group. Default value is 0.
NOM Simulator: Summary

We have gathered all information we need, you may invoke your simulation now.

**Invoke Simulation**

---

### ENVIRONMENT INFORMATION

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation Time</td>
<td>2.0</td>
</tr>
<tr>
<td>Microbe Density</td>
<td>0.0010</td>
</tr>
<tr>
<td>Fungal Density</td>
<td>0.0010</td>
</tr>
<tr>
<td>pH Value</td>
<td>7.0</td>
</tr>
<tr>
<td>Temperature</td>
<td>300.0</td>
</tr>
<tr>
<td>PKW</td>
<td>14.0</td>
</tr>
</tbody>
</table>

### MOLECULE INFORMATION

<table>
<thead>
<tr>
<th>Molecule Name</th>
<th>Percentage Edit or Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>34.0 Delete</td>
</tr>
<tr>
<td>Cellulose</td>
<td>33.0 Delete</td>
</tr>
<tr>
<td>Lignin</td>
<td>33.0 Delete</td>
</tr>
<tr>
<td>moleculeA</td>
<td>0.0 Delete</td>
</tr>
</tbody>
</table>
NOM Simulator
Welcome to NOM Research Group! X Y

Your simulation has been invoked. Please click here to go to the reports page.
Welcome to NOM Research Group! X Y

NOM Simulator: Reports

Currently, you have the following sessions invoked. The first one is your most recent session. You can view reports for each session by click the following links. To start a new simulation, click here. To cleanup terminated sessions, click here.

- Session 116: Terminate Session
  - Reactions Reports

- Session 115: Terminate Session
  - Reactions Reports
NOM Simulator

Welcome to NOM Research Group! X Y

NOM Simulator: Reports

Currently, you have the following sessions invoked. The first one is your most recent session. You can view reports for each session by clicking the following links. To start a new simulation, click here. To cleanup terminated sessions, click here.

- Session 116: TERMINATED
  - Reactions Reports
Agent-Based Modeling approach
Swarm and other tools
Natural Organic Matter
Molecules as Agents (objects)
Web-based collaboratory
NOM 1.0 - Status/Plans

- Distributed System
  - Two Simulation Servers
  - Three Database Servers
  - Reports/Forms Server
  - Load Balancing / Fail Over
- Continue to Work on Core Simulation
- Get Researchers to Test and Provide Feedback
- Ad Hoc Queries to the Database
- Add Data Mining Interface/Capabilities
  - Clustering, Classification, Association Rules, Model Selection
Agent-Based Modeling of OSS

- NSF - CISE - Division of Information & Intelligent Systems
- Develop an Understanding of the Open Source Software (OSS) Development Phenomenon
  - Self-organized
  - Decentralized
  - Emergent Properties
  - Complex Adaptive Process
- Research Activity
  - Data collection
  - Social Network Models
  - SWARM-Based Simulation
Open Source Software (OSS)

- Free …
  - to view source
  - to modify
  - to share
  - of cost

- Examples
  - Apache
  - Perl
  - GNU
  - Linux
  - Sendmail

- Development
  - Mostly volunteer
  - Global teams
  - Virtual teams
  - Self-organized
  - Self-managed
  - Often large numbers of developers, testers, support help
  - Rapid, frequent releases
  - Mostly unpaid
Open Source Software (OSS): Significance

Contradicts traditional wisdom:
- Software engineering
- Coordination, large numbers
- Motivation of developers
- Quality
- Security
- Business strategy

Significant component of e-Business infrastructure

Little Research Done to Date

Great Research Opportunity
- Almost all activity in online
- Much of activity is archived
- SourceForge Repository

Research issues:
- Understanding motives
- Understanding processes
- Intellectual property
- Digital divide
- Self-organization
- Government policy
- Impact on innovation
- Ethics
- Economic models
- Cultural issues
Open Source Software (OSS)

- Major Component of e-Technology Infrastructure with major presence in
  - e-Business
  - e-Science
  - e-Government
  - e-Learning
- Apache has over 60% market share of Internet Web servers
- Linux on over 7 million computers
- Most Internet e-mail runs on Sendmail
- Tens of thousands of quality products
- Part of product offerings of companies like IBM
  - Apache in Websphere, Linux on mainframe
  - Corporate employees participating on OSS projects
Open Source Software (OSS)

- Seems to challenge traditional economic assumptions
- Model for software engineering
- New business strategies
  - Cooperation with competitors
  - Beyond trade associations, shared industry research, and standards processes — shared product development!
- Virtual, self-organizing and self-managing teams
- Intellectual property issues
- Government policy issues
Related Research

- Feller and Fitzgerald (ICIS, 2000)
  - Research framework and analysis of the OSS phenomenon
- Hars and Ou (HICSS 2001)
  - Survey of OSS developers
  - Reported on motivations of developers
- Scacchi (IEE Proceedings - Software, 2002)
  - Study of socio-technical processes associated with OSS development practices
- Wolf, Lakhani, and Bates (BCG/MIT Sloan, 2002)
  - Survey of Source Forge Developers
- Hann, Roberts, Slaughter, and Fielding (ICSE, 2002)
  - Survey of Apache developers - economic incentives
Self-Organizing Systems

- Large numbers of locally interacting agents
- Simple rules for agent behavior
- Heterogeneous agents: attributes and behaviors
- Often unexpected and difficult to predict global properties emerge
- Emergence
- Complexity: hence, need for agent-based simulation
Modeling Social Systems

- Agent-Based Approach
  - Schelling (1978)
    - Micro and Macro Motives, racial housing patterns
  - Axelrod (1984)
    - Iterated prisoners dilemma, strategies
  - Epstein and Axtell (1996)
    - Political behavior

- Goal is not to predict, but to understand the processes that lead to emergent behavior

- Swarm Simulation: Agent Based Modeling

- Social Network Theory
OSS as a Social Network

- Social Network Theory
  - Agents are nodes on a graph (developers)
  - Edges are relationships (project participation)
  - Growth of network: random or preferential attachment, formation of clusters
  - Network attributes: diameter, average degree
Collaborative Social Networks

- Six degrees of separation
  - Small World Phenomenon
  - MS => 1 degree of Separation
- Research on joint authorship
- Movie Co-Stars
  - Kevin Bacon Game
  - IMDB.com
- Open Source Software Development
Data Collection — Monthly

- Web crawler (scripts)
  - Python
  - Perl
  - AWK
  - Sed
- Monthly
- Since Jan 2001
- Project ID
- User ID
- Anonymized
Sample Summary Data

- SourceForge.net (May 2002)
  - Statistical data
    - 56,144 developers
    - 39,025 projects
  - Structural data
    - 43,871 developers on only one project
    - 16,821 isolated developers (no collaborative links)
  - Pareto distributions
    - Project sizes
    - Developer participation (number of projects a member of)
  - Trend data
Regression: Number of projects that developers are on

<table>
<thead>
<tr>
<th># projects</th>
<th># of developers on that many projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21488</td>
</tr>
<tr>
<td>2</td>
<td>3688</td>
</tr>
<tr>
<td>3</td>
<td>1086</td>
</tr>
<tr>
<td>4</td>
<td>413</td>
</tr>
<tr>
<td>5</td>
<td>177</td>
</tr>
<tr>
<td>6</td>
<td>76</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ y = 10.6905 - 3.70892 \times \]

\[ R^2 = 0.979906 \]
More Empirical Data

- Log number of projects (log # of developers) relationship
  - Project = 4.66
  - R-Square = 0.9

- Linear regression for total number of developers
  - Total Number of Developers = 17316.33 + 2251.29 * relmonth
  - R-Square = 1.00

Graphs showing empirical data with linear regression analysis.
Prototype Simulation

- Java Swarm / JDBC
- Developer class
- Each simulated developer is an instance of “Developer” with random attributes and behaviors
  - Local decision logic
  - Simulates self-organization
  - Create new projects
  - Join existing projects
  - Abandon a project
Prototype Simulation

- Simulating individual agent behavior stochastically
- Tuning simulation to fit empirical data

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Description of Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>probCreateInitially</td>
<td>Probability that a developer creates a project during his first time slice</td>
</tr>
<tr>
<td>probCreateEachPeriod</td>
<td>Probability that developer creates a project during any time slice after his first</td>
</tr>
<tr>
<td>probJoinEachPeriod</td>
<td>Probability that developer joins a project during any time slice after his first</td>
</tr>
<tr>
<td>probAbandonEachPeriod</td>
<td>Probability that developer abandons a project during any time slice after his first</td>
</tr>
<tr>
<td>developersPerPeriod</td>
<td>The number of developers introduced to the network each period</td>
</tr>
<tr>
<td>endTime</td>
<td>The duration of the simulation in time slices (days)</td>
</tr>
</tbody>
</table>
Prototype Simulation

- Oracle Tables
  - DEVELOPERS
  - PROJECTS
  - LINKS
    - Format identical to SourceForge empirical DB
    - Reuse of scripts for analysis of simulation data
Prototype currently models high level empirical statistics of SourceForge
Currently simulating random attachment
Plan to implement preferential attachment
Plan to fit simulation with lower level statistics and structural data from SourceForge, e.g., Pareto Distribution
Interpretation
Survey instruments to collect additional data on developer behavior
Discussion