

Community Building in Open Source Software Ecosystems

Patrick Adam Wagstrom
Department of Engineering and Public Policy
Carnegie Mellon University
Advisors: Jim Herbsleb and Kathleen Carley

October 2004

Overview

- What is Free and Open Source Software (F/OSS)?
- Previous Work
- Research Questions
- Modeling Software Development
- Sources of Validation Data
- Preliminary Results
- Virtual Experiment

What is F/OSS?

- Utilize copyright law to protect the rights of the user
- Term “Open Source” was coined in 1998
- “Free” software started by Richard Stallman in 1984
- Free for any use
- Free to redistribute and modify
- Communities are dynamic and driven by merit
- Increasing amounts of commercial interest

Previous Work

- Developer motivation (Ghosh 2003, Lakhini et al 2002, Lakhini 2003, Shah 2003)
- Economic Basis for Development (Lerner and Tirole 2002, Schiff 2002)
- Social network overview (Xu and Madey 2004, Sandusky et al 2004)
- Simulation of communities (Gao et al 2003)

Research Questions

- Can we predict adoption of open source projects?
- Can we predict development of open source projects?
- In what ways does the community effect the development of projects?
- What happens to a project when new users or developers join the project?

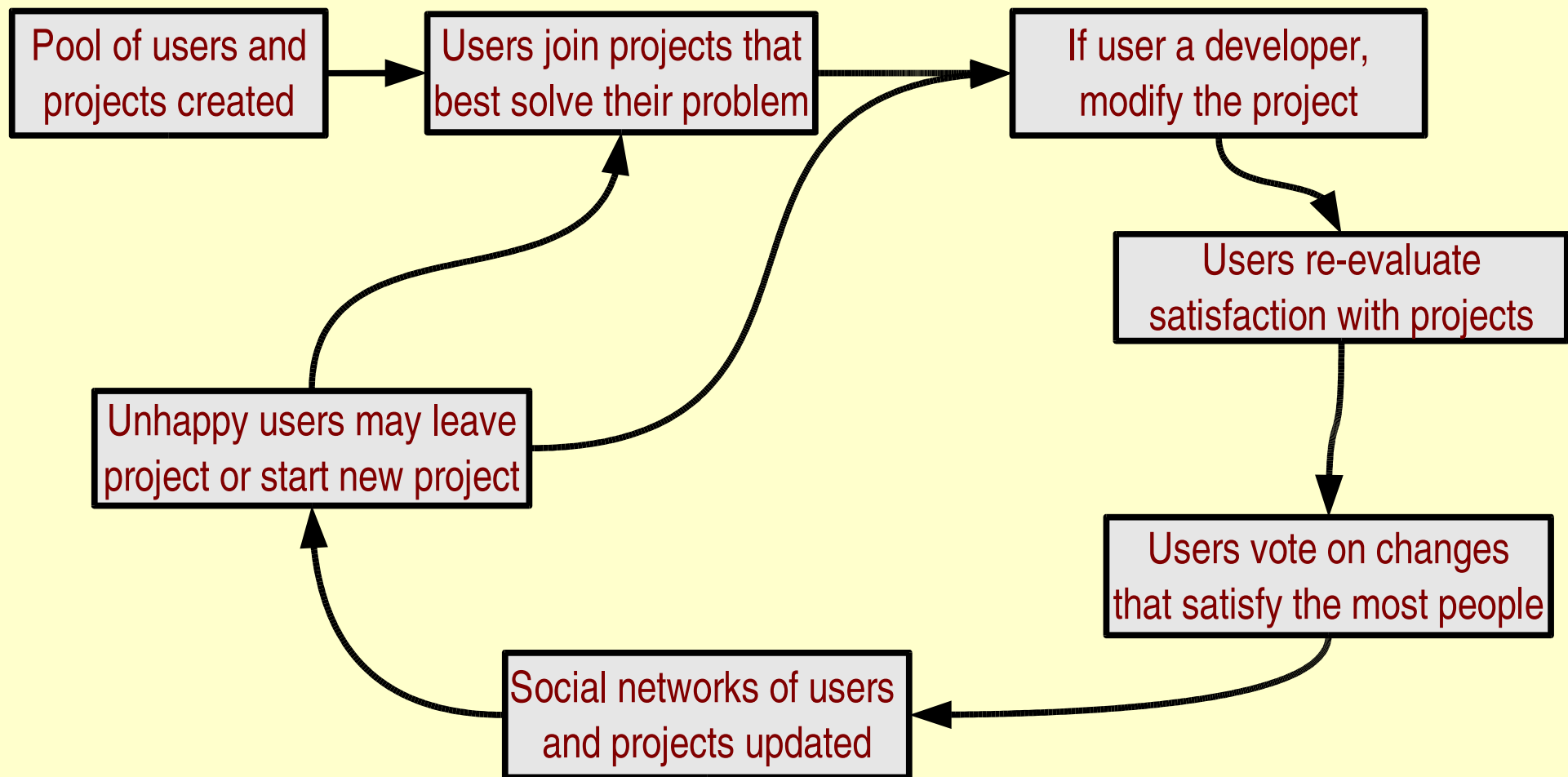
Overview

- What is Free and Open Source Software (F/OSS)?
- Previous Work
- Research Questions
- Modeling Software Development
- Sources of Validation Data
- Preliminary Results
- Virtual Experiment

Modeling Software Development

- Multi-Agent simulation called **OSSim** (pronounced **AWESOME!!**)
- Developers have differing **skill levels**, **problem sets**, and **motives**
- Find the best **project** that solves each **problem**
- Some agents choose to **modify** the project to solve problems better

OSSim Overview



OSSim Input Parameters

| Name | Value | Source |
|--------------------------|---------------------------------|--|
| Number of users | 50 | Arbitrary for virtual experiment |
| Number of projects | 2 | Arbitrary for virtual experiment |
| Percentage of developers | 0.1-0.3 uniform distribution | Based on Advogato.org and bug Information from Tigris.org |
| Developer skill | 0-100 custom distribution | Bimodal distribution with lognormal like Tail; from Advogato.org and Ghosh |
| Social network size | 0-50 uniform distribution | Advogato.org, Sandusky 2004, Xu 2004 |
| Project network size | 2-20 uniform distribution | Conjecture based on Ghosh data |
| Agent Attention Span | normal distribution (m=10, s=2) | My expert estimate |
| Problem model | NK | Kaufman 1993 |
| Problem complexity | N=15, K=2 | Experiments to provide results similar to observed values |

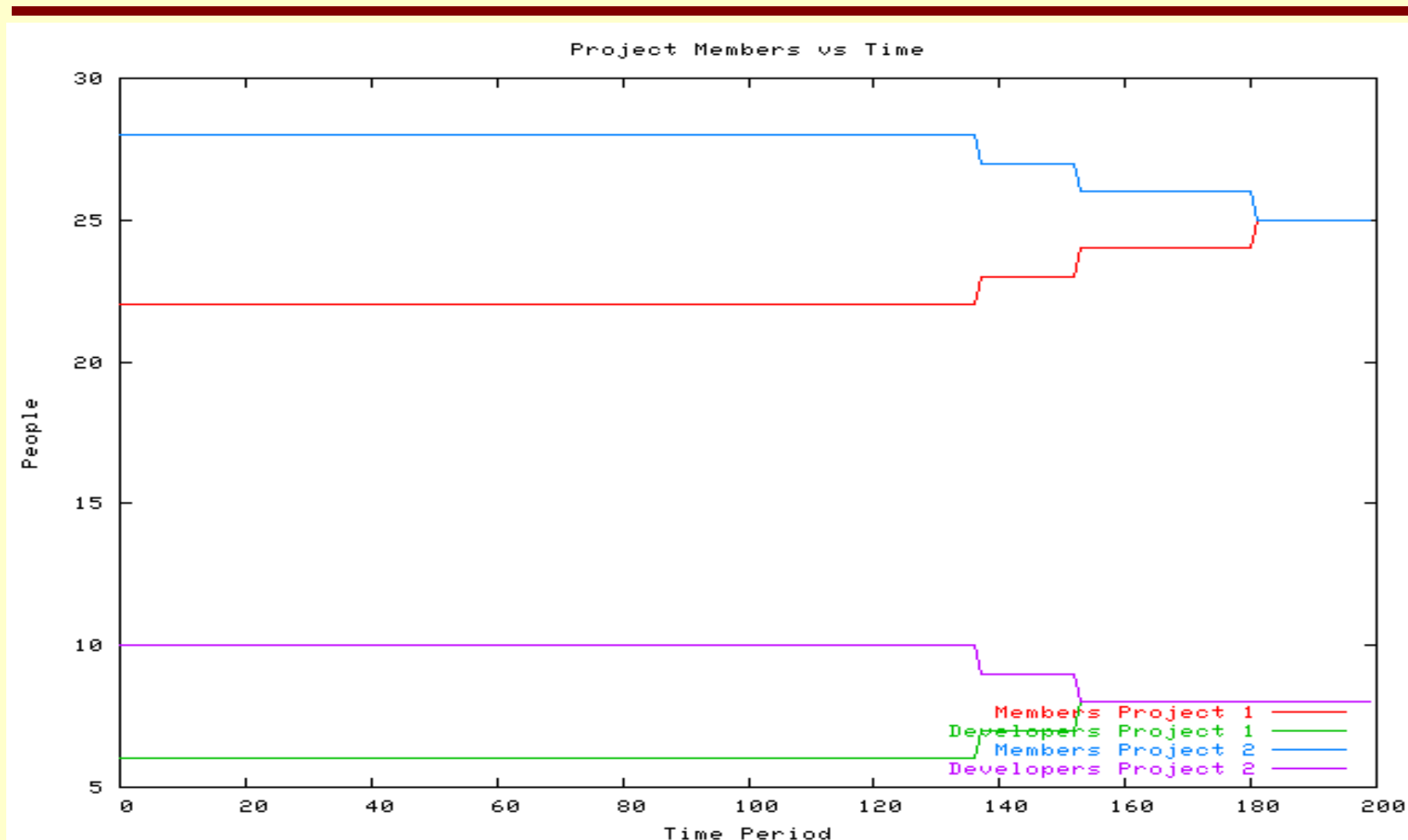
OSSim Output Values

- Project members
- Social network information
- Project code growth information
- Meta data to determine project direction

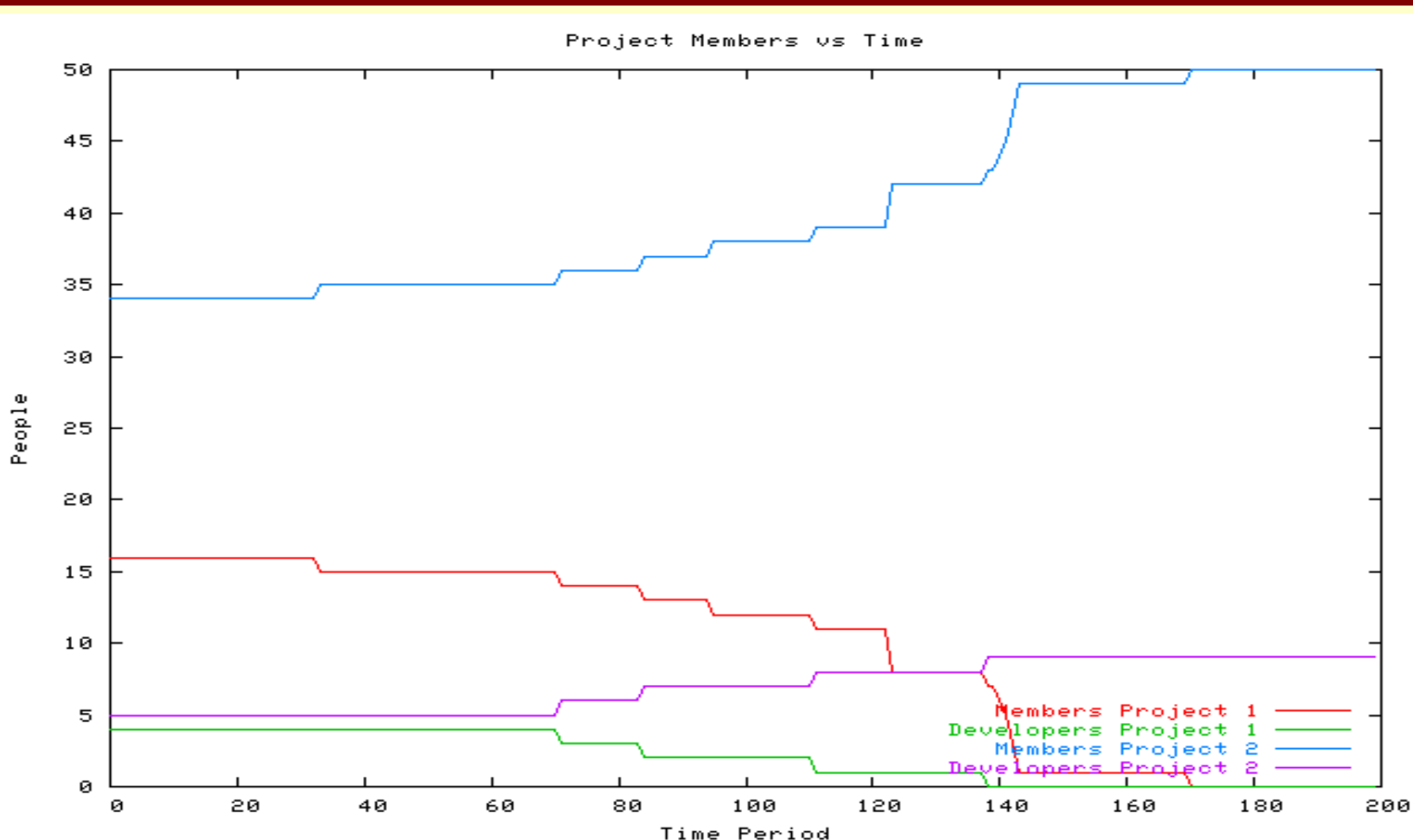
Validation Data Sources

- Advogato.org
 - Free/Open source developer site
 - Only tracks social aspects of interactions
- Tigris.org
 - Free/Open source hosting site
 - Does not explicitly track social interactions
- SourceForge.net
 - Similar to Tigris.org, but much larger

Preliminary Results



Preliminary Results (2)



Overview

- What is Free and Open Source Software (F/OSS)?
- Previous Work
- Research Questions
- Modeling Software Development
- Sources of Validation Data
- Preliminary Results
- Virtual Experiment

Virtual Experiment

- Medium scale open source project (5-10 volunteer developers)
- Corporate interest spawns new contributors
 - Vary number of contributors (1-2, 3-5, 8-10)
 - Vary skill and motivation of contributor
- Observe progress of project (users, overall fitness)
- What happens to volunteer developers

Virtual Experiment Results

- Note: These are like WMD's in Iraq...still **Imaginary** results
- When compared to the BATIK project, OSSim showed **similar results 72%** of the time
- When compared to SpamAssassin, OSSim showed **similar results 82%** of the time
- Corporate influence frequently **skews projects away** from their original volunteer hacker nature

Major Contributions

- New analysis of the social network structure of Free/Open Source Projects using new data streams
- First model of software development focusing exclusively on free/open source software engineering
- Results from OSSim have been shown to be close to those of real projects

Questions, comments, and large amounts
of currency are now welcome

References

NK Model

Agents have a genotype bit string of length N

0 1 1 1 0 0 1 1 0 1 1 1 0 1 1

N=15

Each bit is associated with K neighbors to create alleles

| | | | |
|-----|-----|-----|-----|
| 011 | 011 | 110 | |
| 111 | 110 | 101 | |
| 110 | 101 | 011 | K=2 |
| 100 | 011 | 110 | |
| 001 | 111 | 101 | |

Each allele has is evaluated against the randomly generated fitness landscape for that position. Overall fitness is the average of these values.

0.15 + 0.19 + 0.20 + 0.46 + 0.34 + 0.85 + 0.67 + 0.12 + 0.77 + 0.91 + 0.85 + 0.55 + 0.72 + 0.06 + 0.59

15

0.4913

Mutation in the NK Model

Selects a bit at random, flip it, and evaluate new fitness

0 1 1 1 **1** 0 1 1 0 1 1 1 0 1 1

| | | |
|------------|-----|-----|
| 011 | 011 | 110 |
| 111 | 110 | 101 |
| 111 | 101 | 011 |
| 110 | 011 | 110 |
| 101 | 111 | 101 |

K=2

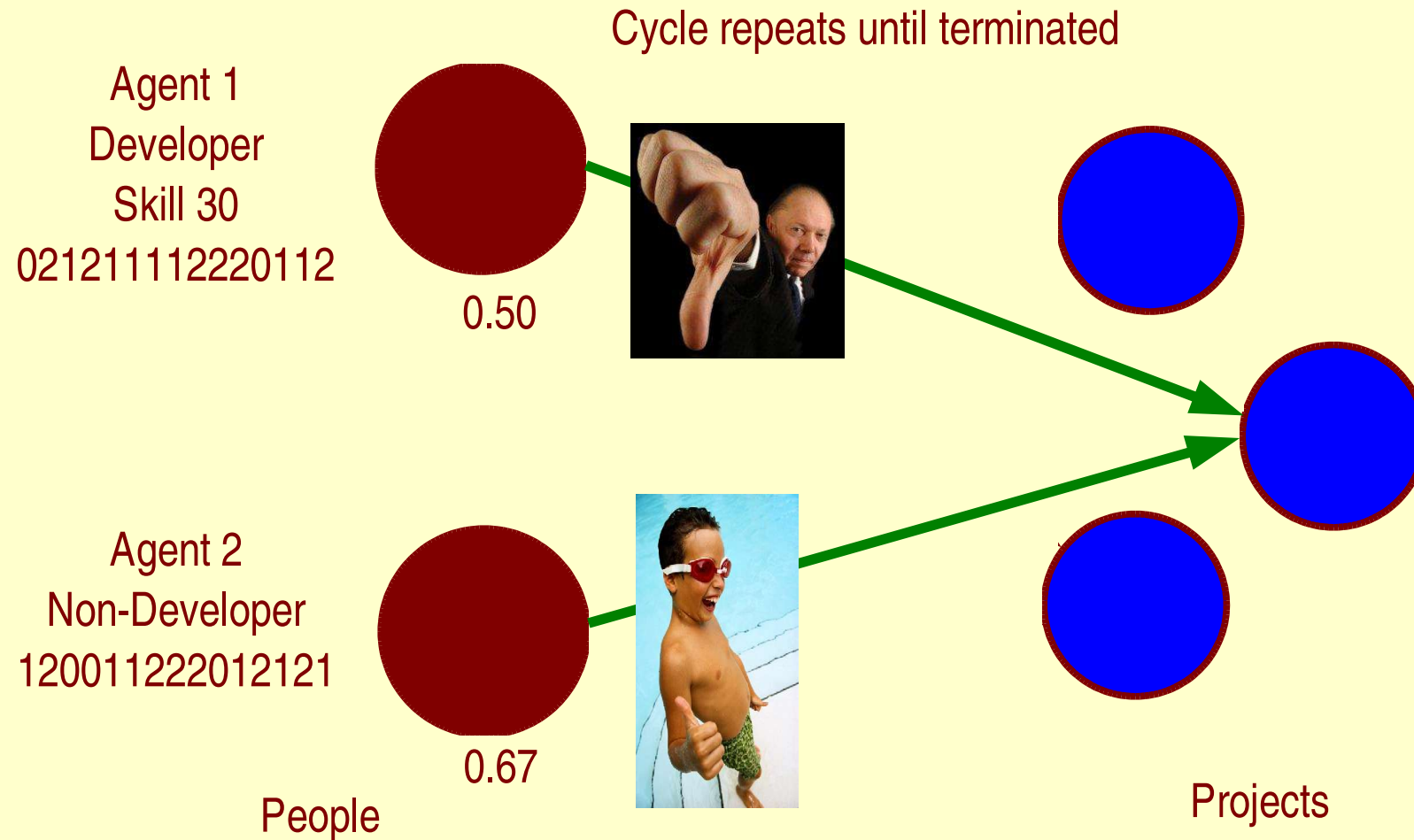


0.15 + 0.19 + **0.82 + 0.31 + 0.74** + 0.85 + 0.67 + 0.12 + 0.77 + 0.91 + 0.85 + 0.55 + 0.72 + 0.06 + 0.59

15

0.5493

The OSSim NK Process



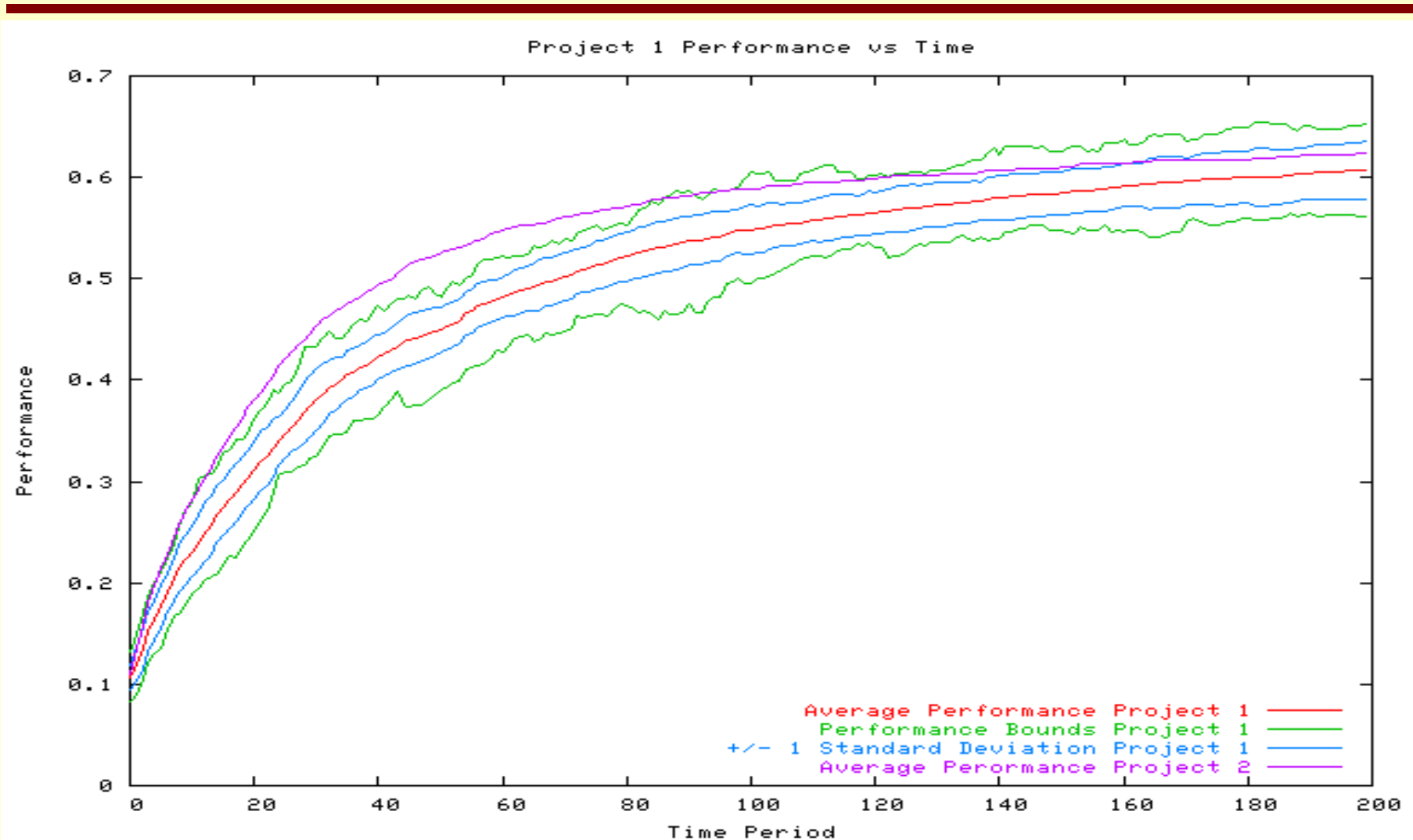
Agent Properties

- Set of **problems** the agent is encountering
- Attempt to match problems to **projects** (solutions)
- If the agent is a **developer** or is a **free rider**
- **Skill** if the agent is a developer
- Focus of the agent is controlled by **attention-span**
- Who an agent knows – **social network**
- Projects an agent knows – **project network**

Project Properties

- A set of **developers** who create the project and **users** who utilize the project
- **Social norms** that characterize the interaction process
- A **walled server** that controls access to the project resources and mediates communication
- A **fitness landscape** that can be used to evaluate **problems** of agents

Preliminary Results



Preliminary Results (2)

