Applying Biological Evolution to Software Ecosystems
A Case study with GNOME

Maëlick Claes

Software Engineering Lab
Maëlick Claes, Tom Mens, Alexander Serebrenik (TKe),
Mathieu Goeminne & Philippe Grosjean (ECONUM, UMONS)

BENEVOL 2012
Plan

Introduction

Evolutionary theories

Experiments on GNOME
Plan

Introduction

Evolutionary theories

Experiments on GNOME
Ecosystems

Biological ecosystem

Living species

Interactions

Soil, air, water, light, climat, etc

Shared habitat
Ecosystems

Software ecosystem

Software projects

Interactions

Software & hardware tools, users & developers communities

Shared development environment
Analogy

Software project vs. Biological species

- Evolution speed
- Mechanisms driving evolution
- Open source ecosystems: exhaustive information
  - Version control system
  - Mailing lists
  - Bug tracking systems
  - ...
Evolutionary theories

Plan

Introduction

Evolutionary theories

Experiments on GNOME
Darwinism

Phylogenetic tree
Reticulate evolution

Hypothetical evolution of Scleractinia corals
Evolutionary theories

Reticulate evolution

Horizontal gene transfer & Hybrid speciation

- Debian
- Ubuntu
- Crunchbang
- Maemo
- Moblin
- Meego

Timeline:
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
Plan

Introduction

Evolutionary theories

Experiments on GNOME
Transfer of knowledge

Migration patterns

- Migration (one-way transfer)
- Exchange (bidirectional transfer)
- Period transfer (cyclic transfer)
- Parallel work
## GNOME’s subsystems - Nautilus & Evolution

<table>
<thead>
<tr>
<th></th>
<th>GNOME</th>
<th>Evolution</th>
<th>Nautilus</th>
</tr>
</thead>
<tbody>
<tr>
<td># projects</td>
<td>1327</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td># projects with at least 2 days of commits</td>
<td>1263</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td># projects with coding activity</td>
<td>1268</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td># projects with coding activity and at least 2 days of commits</td>
<td>1222</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td># commits</td>
<td>1169251</td>
<td>66687</td>
<td>25860</td>
</tr>
<tr>
<td></td>
<td>(5.7%)</td>
<td>(2.21%)</td>
<td></td>
</tr>
<tr>
<td># coding commits</td>
<td>606300</td>
<td>44150</td>
<td>13133</td>
</tr>
<tr>
<td></td>
<td>(7.28%)</td>
<td>(2.17%)</td>
<td></td>
</tr>
<tr>
<td># authors with at least 1 commit</td>
<td>3841</td>
<td>867</td>
<td>707</td>
</tr>
<tr>
<td></td>
<td>(22.6%)</td>
<td>(18.4%)</td>
<td></td>
</tr>
<tr>
<td># coders (authors involved in coding)</td>
<td>2892</td>
<td>494</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>(17.1%)</td>
<td>(10.4%)</td>
<td></td>
</tr>
<tr>
<td># coders in at least 2 projects and 2 consecutive years</td>
<td>1392</td>
<td>341</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td>(24.5%)</td>
<td>(19.7%)</td>
<td></td>
</tr>
<tr>
<td># full years of activity</td>
<td>15</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>considered period of coding activity</td>
<td>1998</td>
<td>1998</td>
<td>1999</td>
</tr>
</tbody>
</table>
Experiments on GNOME

GNOME, Nautilus & Evolution - Metrics

\[
\begin{align*}
Dev_P(t) & = \{ \text{contributors} \in P \text{ who made at least two code commits during period } t \} \\
stayers_S(t) & = | \{ a | \exists P_1 \in S, \ a \in Dev_P(t-1) \land \exists P_2 \in S, \ a \in Dev_P(t) \} | \\
joinGlobal_S(t) & = | \{ a | \exists P \in S, \ a \in Dev_P(t) \land \forall P \in G, \ a \notin Dev_P(t-1) \} | \\
joinLocal_S(t) & = | \{ a | \exists P \in S, \ a \in Dev_P(t) \land \forall P \in S, \ a \notin Dev_P(t-1) \land \exists P \in G, \ a \in Dev_P(t-1) \} | \\
joiners_S(t) & = joinGlobal_S(t) + joinLocal_S(t) \\
leaveGlobal_S(t) & = | \{ a | \exists P \in S, \ a \in Dev_P(t-1) \land \forall P \in G, \ a \notin Dev_P(t) \} | \\
leaveLocal_S(t) & = | \{ a | \exists P \in S, \ a \in Dev_P(t-1) \land \forall P \in S, \ a \notin Dev_P(t) \land \exists P \in G, \ a \in Dev_P(t) \} | \\
leavers_S(t) & = leaveGlobal_S(t) + leaveLocal_S(t) \\
\text{attractivity}_S(t) & = \frac{joiners_S(t)}{stayers_S(t) + joiners_S(t)} \\
\text{repulsivity}_S(t) & = \frac{leavers_S(t)}{stayers_S(t) + leavers_S(t)}
\end{align*}
\]
Experiments on GNOME

Attractivity & repulsivity

Figure: Attractivity and repulsivity of GNOME (black circles) and its subsystems Evolution (red triangles) and Nautilus (blue squares).
Joiners & leavers origin

**Figure:** Number of coders locally (dotted lines) or globally (dashed lines) joining and leaving Evolution (red triangles) and Nautilus (blue squares) from 1998 till 2011. Triangles or squares are filled if global exceeds local.
Conclusion

Summary

- Analogy between software and biological ecosystems
- Adaptation of theory from biology to software: reticulate evolution
- Transfers in GNOME: Nautilus & Evolution have different behaviours

Future works

- Migration patterns
- More data: bug trackers and mailing lists
- Time window
- Other ecosystems (KDE) or other subsystems (Gedit, GIMP)
- Lower granularity (files vs. commits)
- Different activities (translation)
Thanks for your attention!

Questions?
Applying biology to computer science

- Genetic algorithms, neural networks, ant colony, . . .
- In software engineering & software evolution: