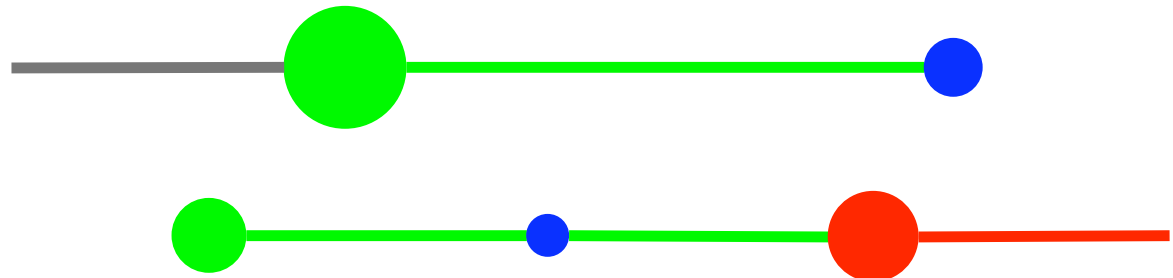




How developers drive software evolution

Tudor Gîrba, **Adrian Kuhn**, Mauricio Seeberger, Stéphane Ducasse

Software Composition Group
University of Bern
Switzerland





Context: Understanding software systems must go beyond the code

Software code might tell you what happened, but it cannot tell you why it happened

A reengineering pattern says to “chat with maintainers”

[Demeyer et al .2002]

Conway’s law

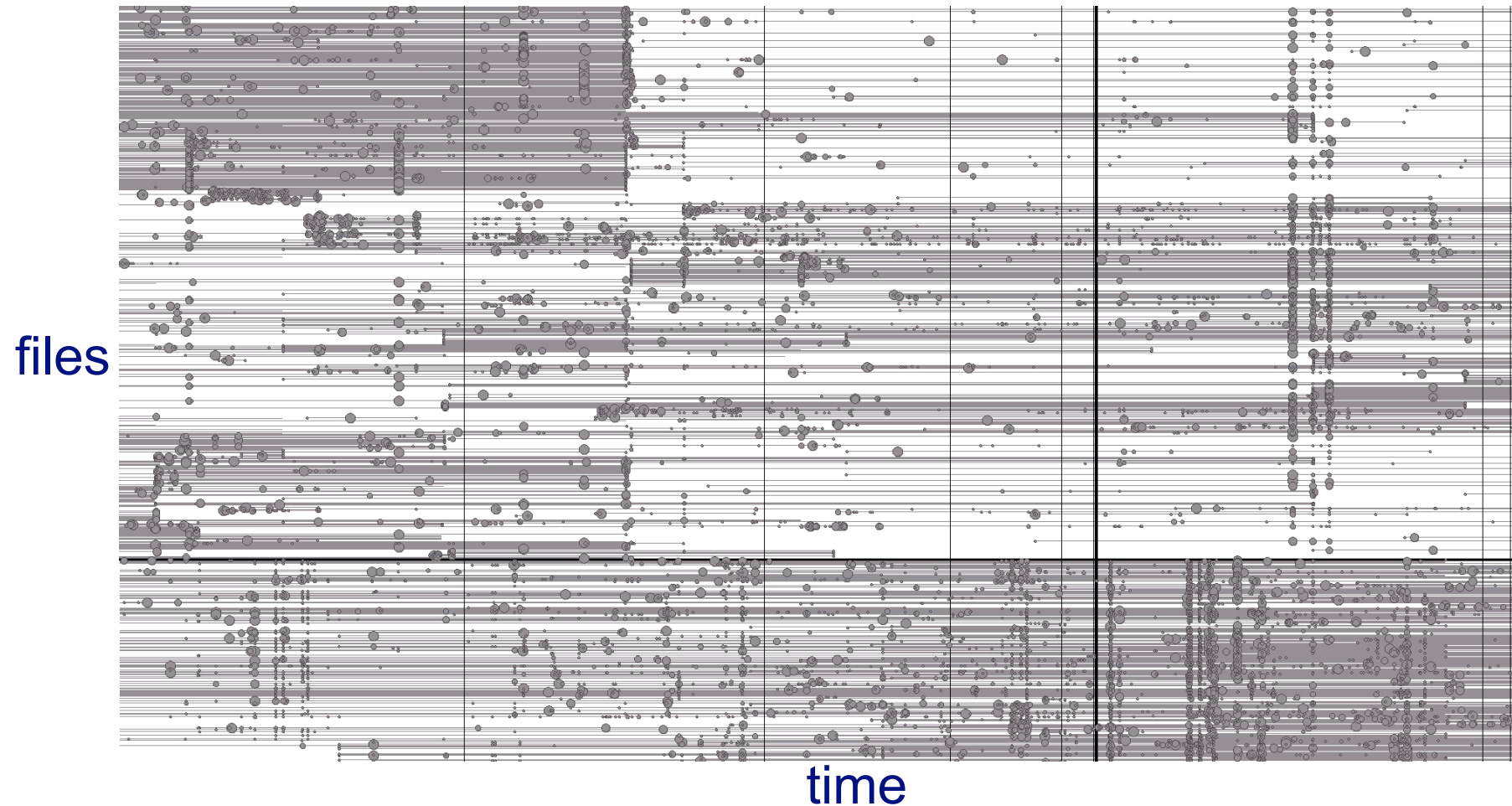
The shape of the organization reflects on the shape of the system

[Conway 67]



The commit history shows what happened

[Rysselberghe, Demeyer 2004]

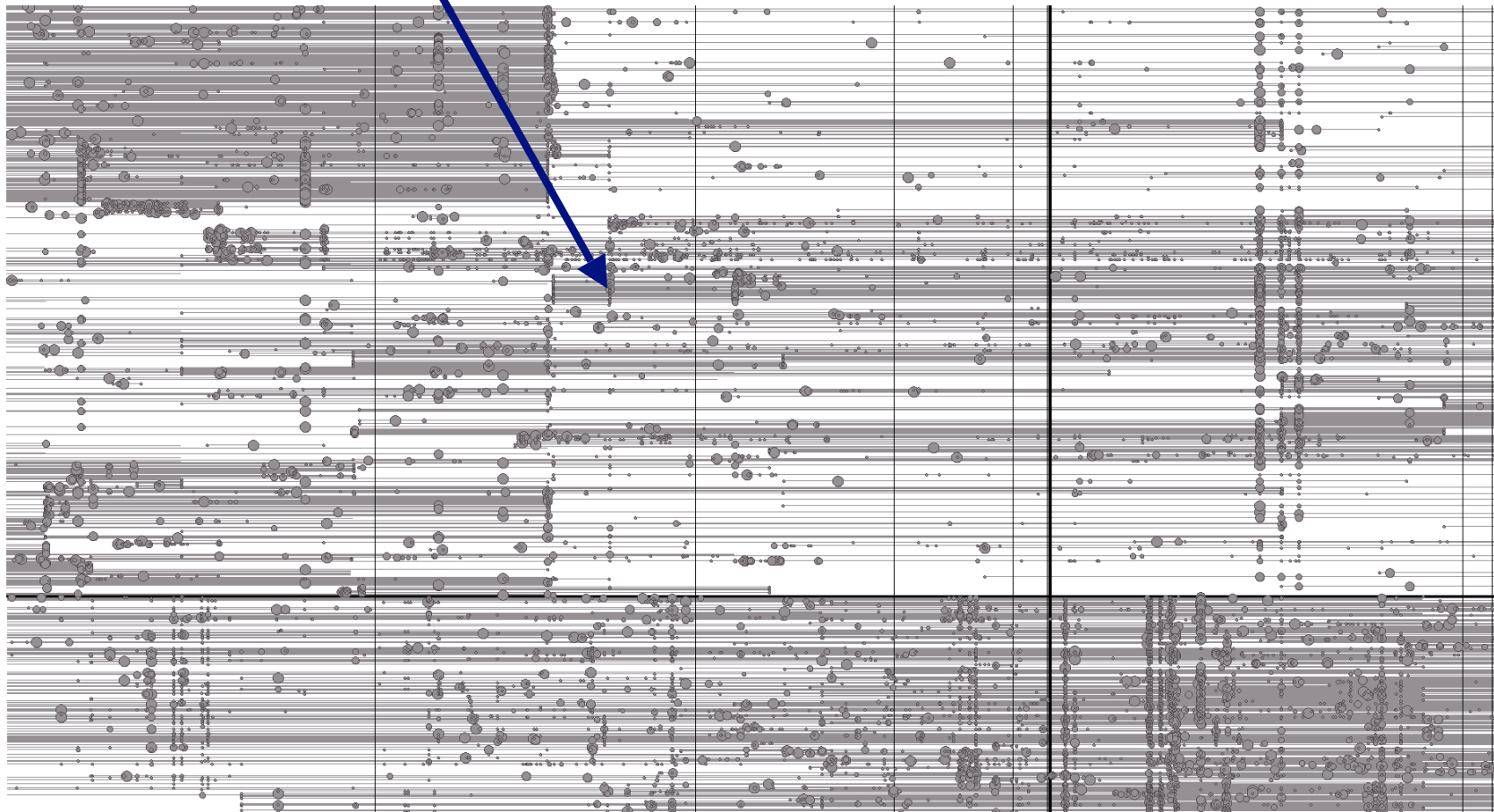




Problem: Who is responsible for this?

[Rysselberghe, Demeyer 2004]

files



time



Our approach: a file is owned by the developer that wrote the most lines

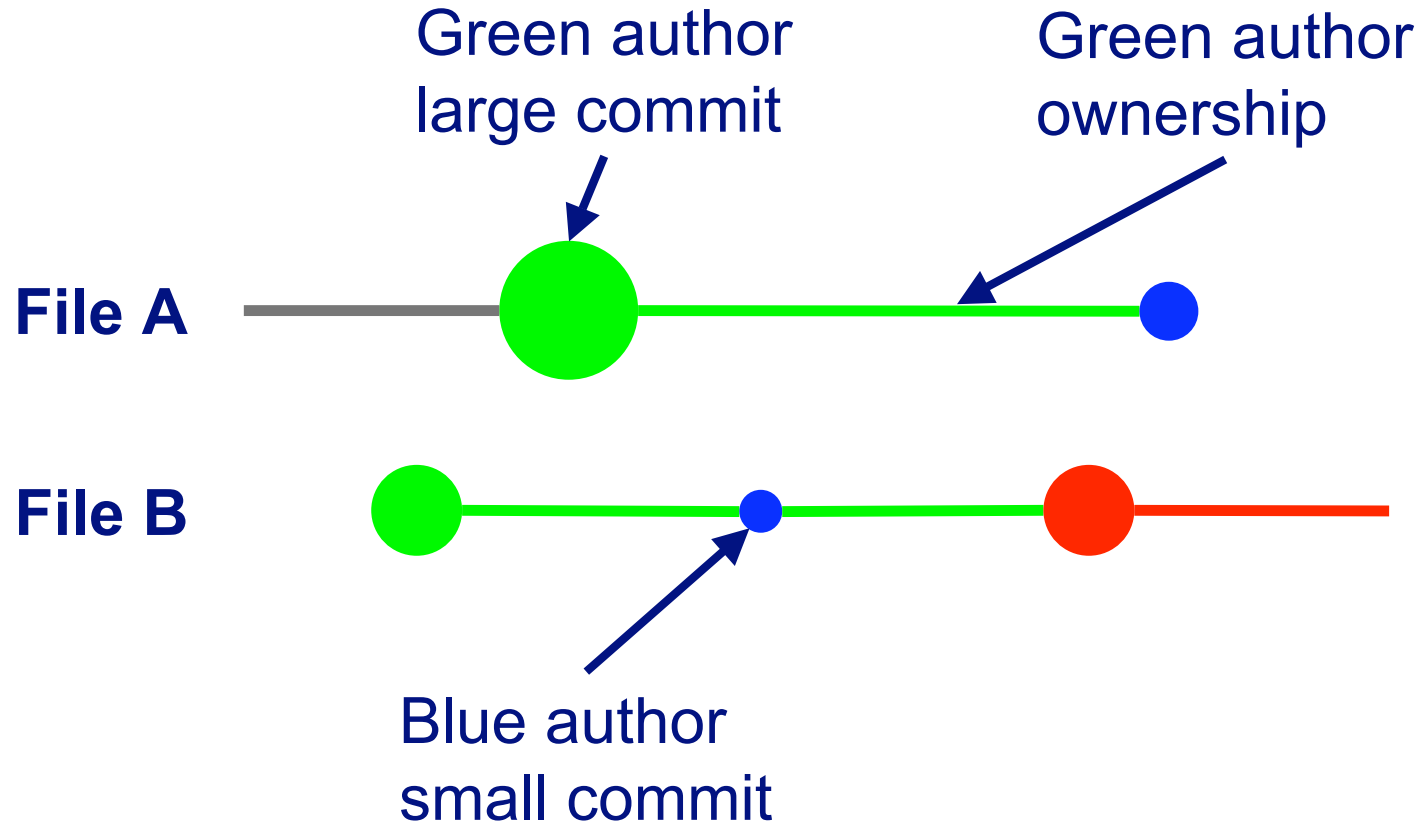
We use **cv**s log

```
revision 1.38 date: 2005/04/20 13:11:24;  
author: seeberge;  
state: Exp;  
lines: +36 -11 added implementation section  
-----  
revision 1.37 date: 2005/04/20 11:45:22;  
author: akuhn;  
state: Exp;  
lines: +4 -5 fixed errors in ownership formula
```



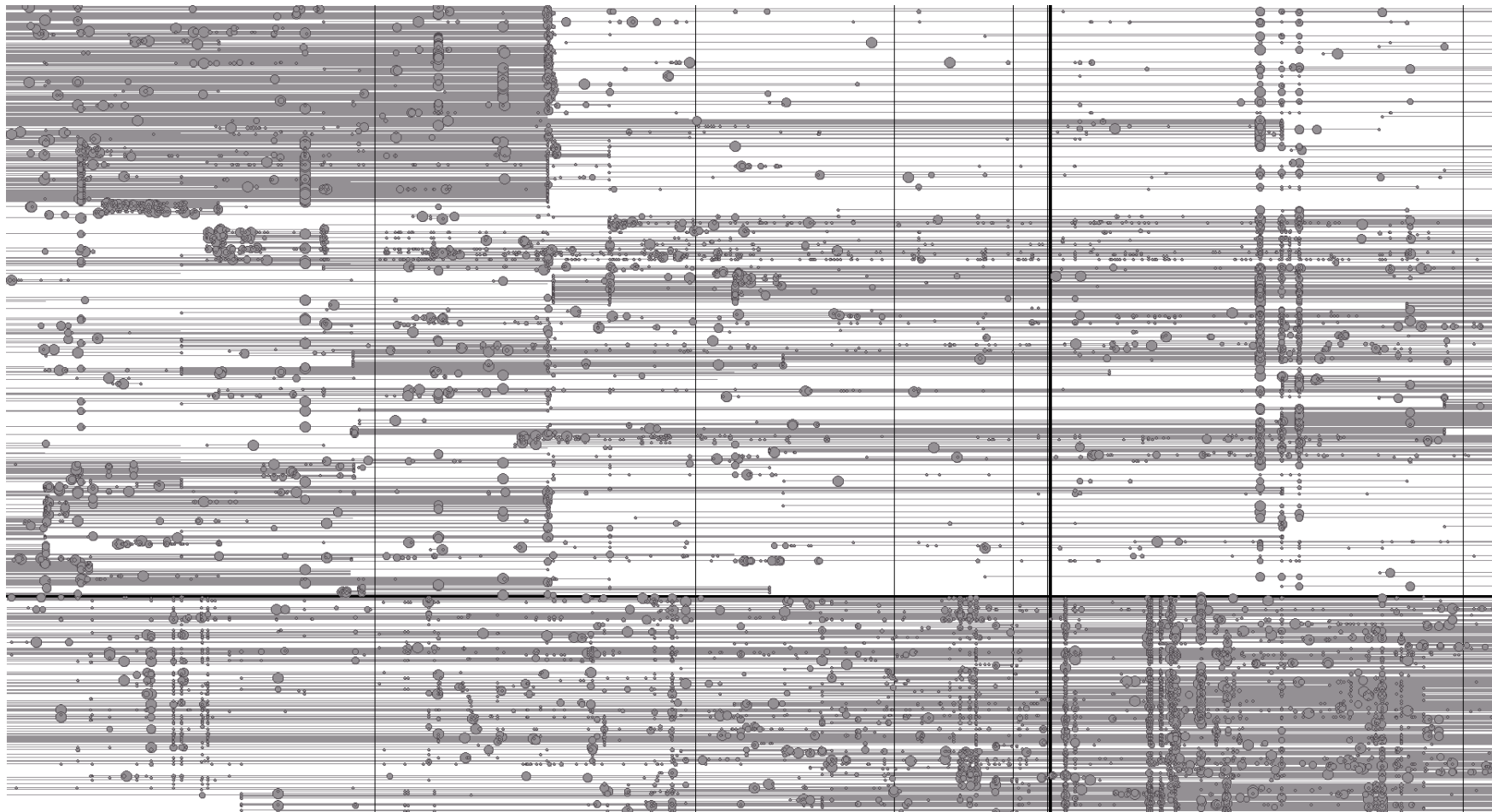


We color the lines to show which author owned which files in which period



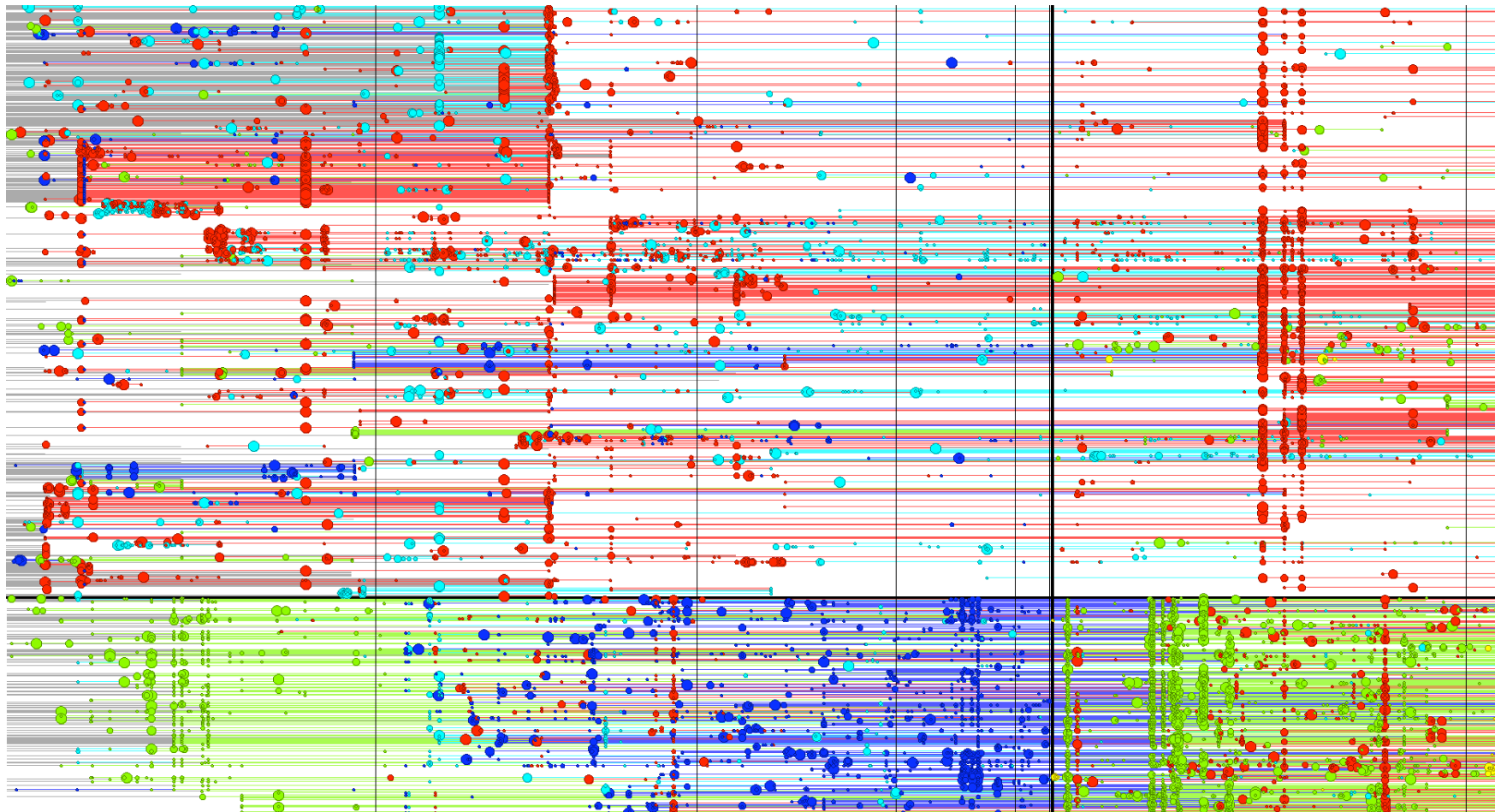


The commit history shows what happened





Ownership Map shows which author owned which files in which period

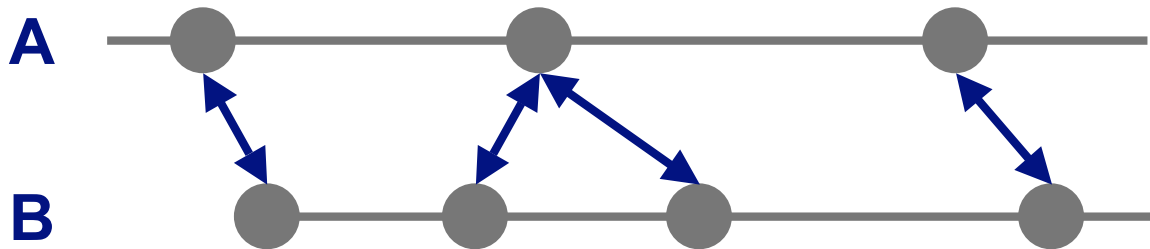




We cluster the files to favor colored blocks inside each module

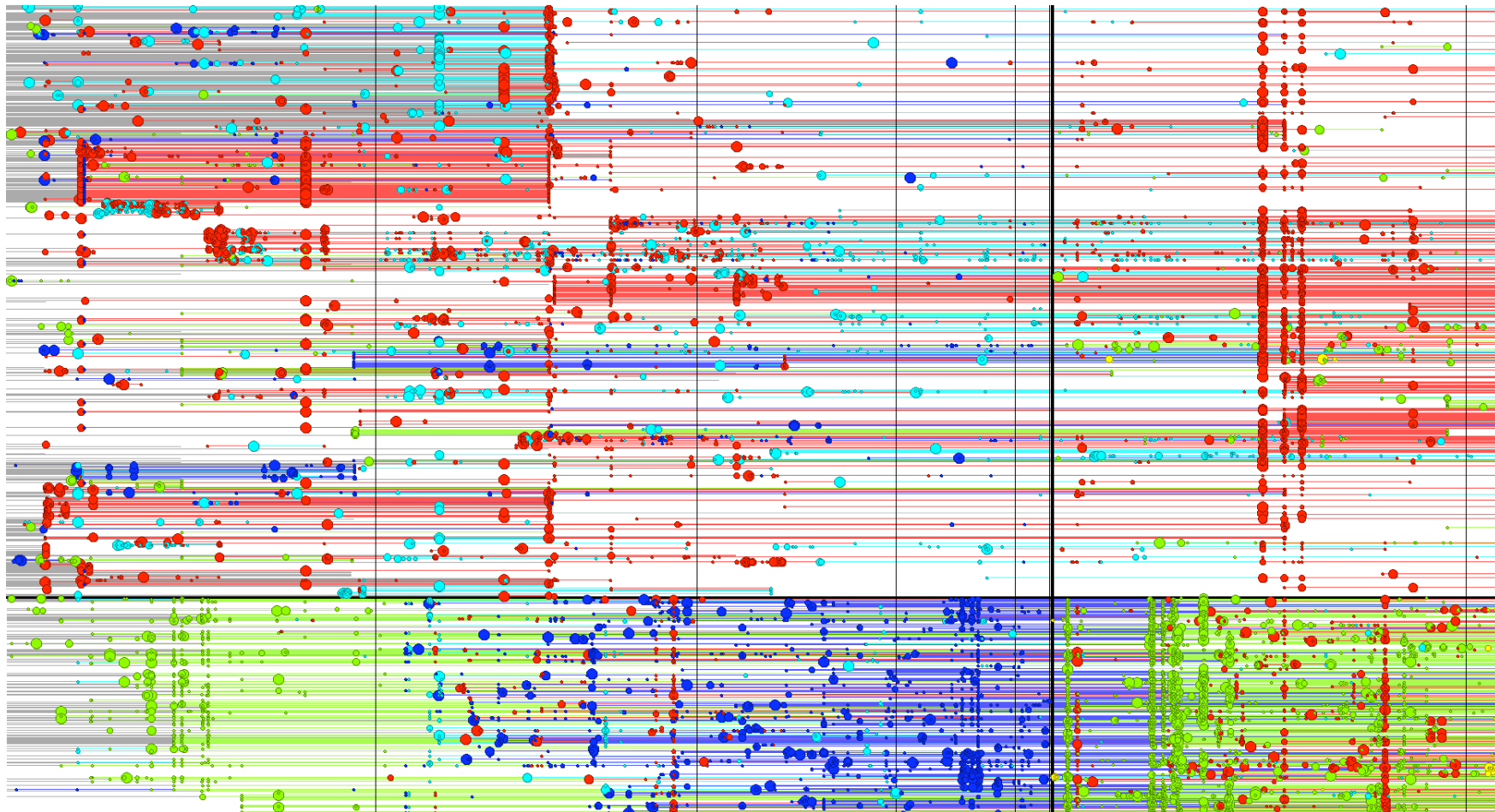
We use the Hausdorff distance between the commit timestamps

$$d(A,B) = \sum_{a \in A} \min^2\{ |a - b| \mid b \in B \}$$



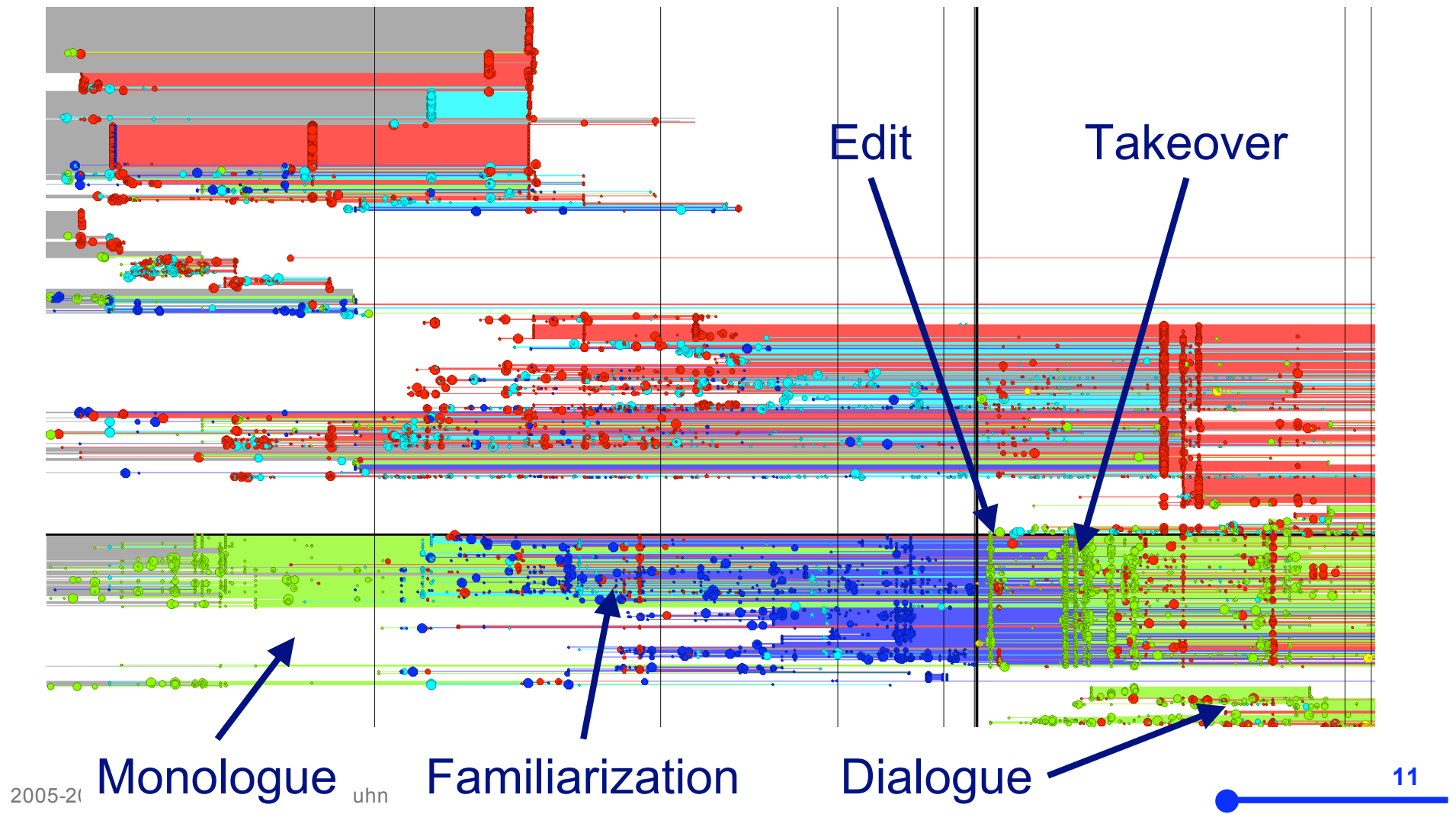


Ownership Map on alphabetically ordered files is not very useful, but ...





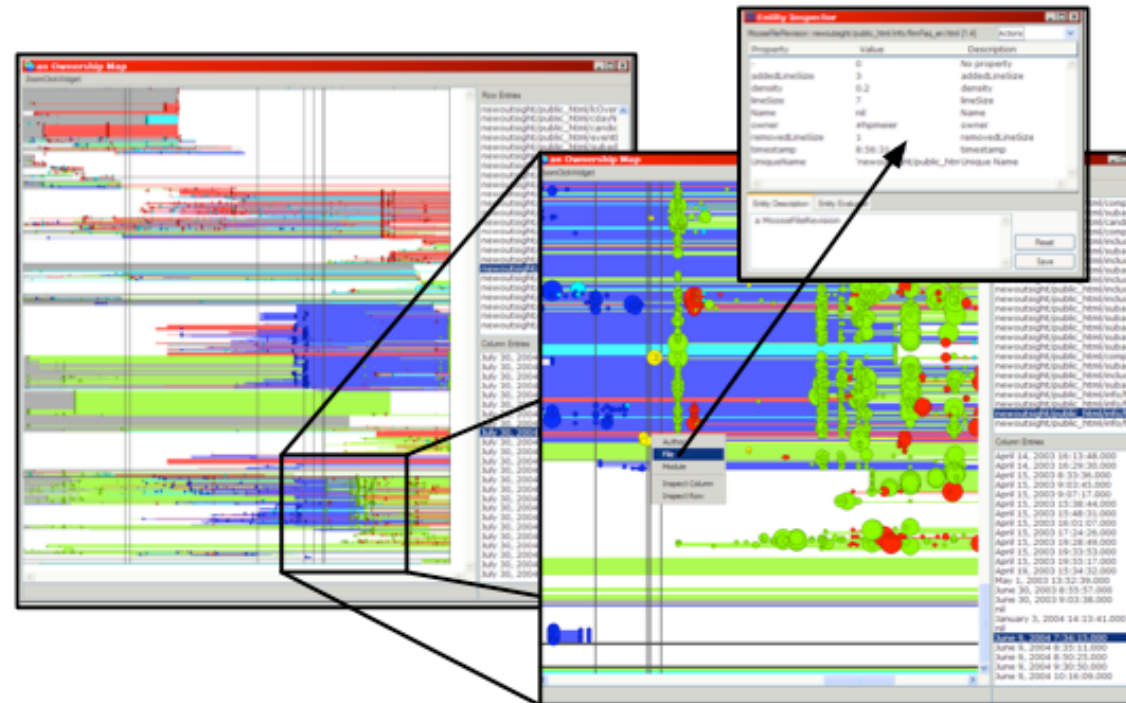
The ordered Ownership Map reveals developer patterns



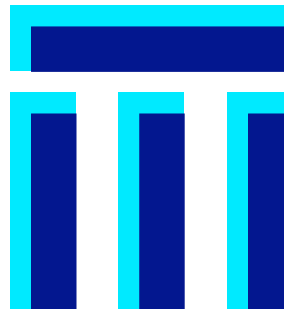


Moose and Chronia: Interaction is crucial when dealing with large amounts of data

Chronia



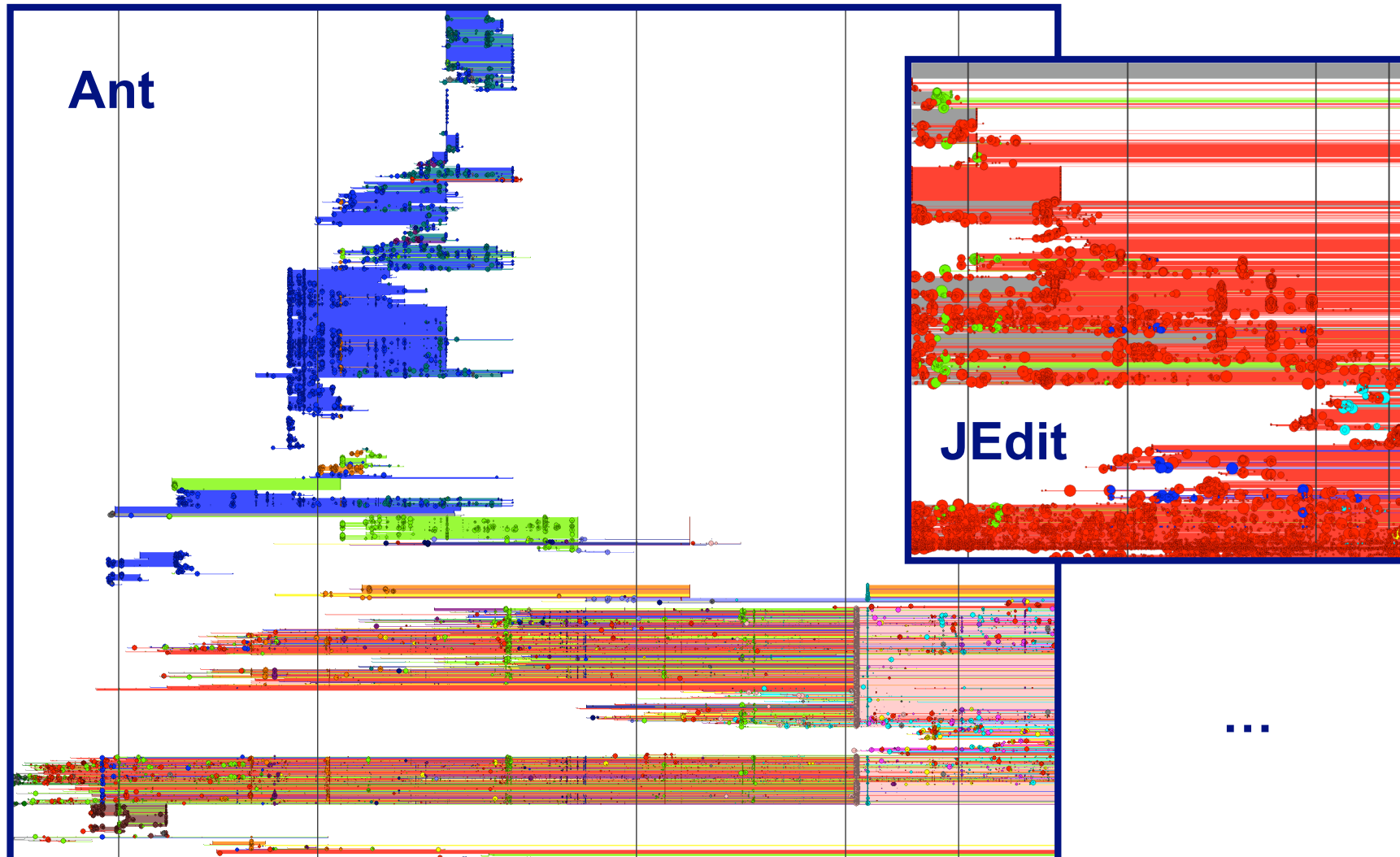
Moose



You will hear about Moose tomorrow



The Ownership Map was applied on several case studies

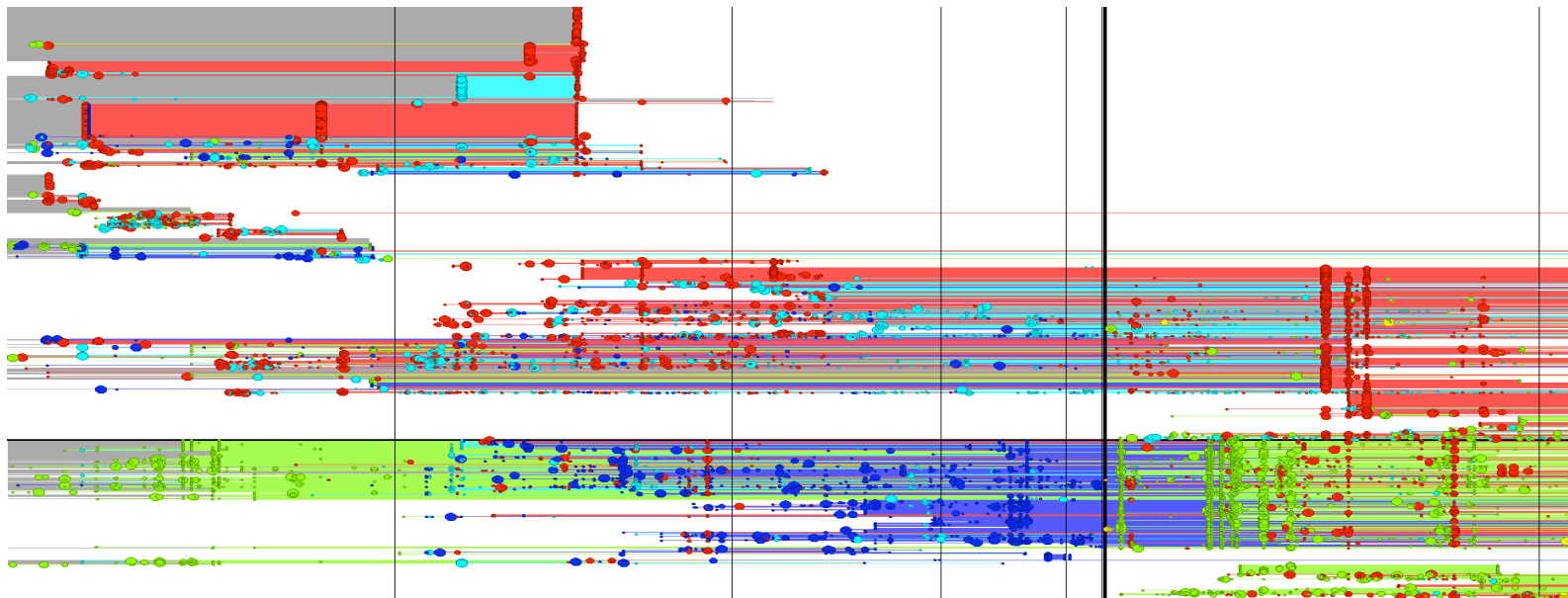




Conclusions: Ownership Map reveals aspects of how developers drive software evolution

Who developed which part of the system?

What was the behavior of the developers?



Questions?





A file is owned by the developer that wrote the most lines

The most straight forward way would be to count exactly what happened but **cvs annotate**, and **cvs diff** are **very slow**

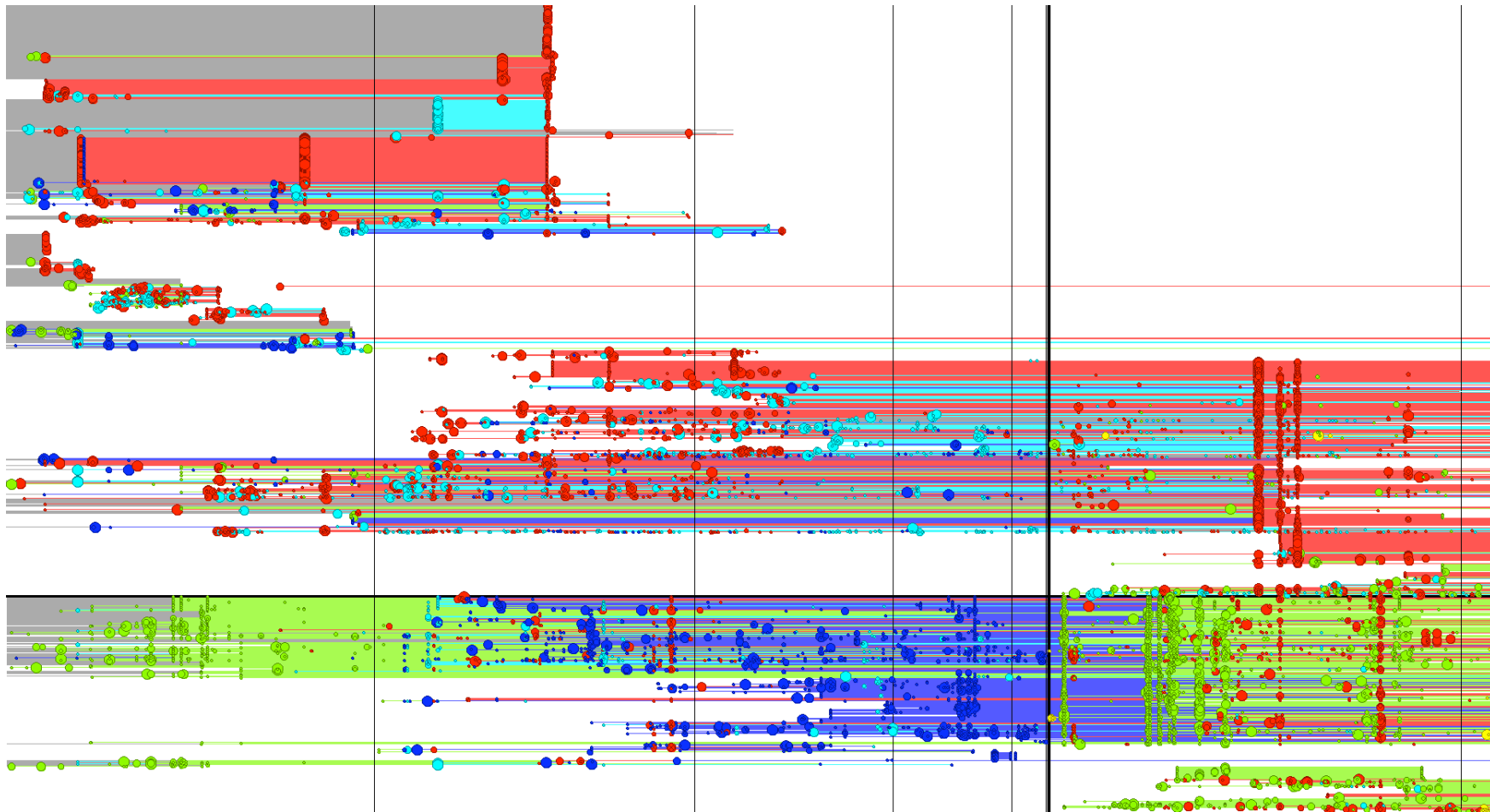
cvs log is fast, but less accurate

```
revision 1.38 date: 2005/04/20 13:11:24;
author: girba;
state: Exp;
lines: +36 -11 added implementation section
-----
revision 1.37 date: 2005/04/20 11:45:22;
author: akuhn;
state: Exp;
lines: +4 -5 fixed errors in ownership formula
```





Ownership map based on CVS log





Ownership map based on CVS annotate

