# Bioinformatics Data Skills

## Git and Version Control in data science

**Ταμπόσης Ιωάννης**

Researcher & Software Development Engineer

PhD Candidate, University of Thessaly

Department of Computer Science and Biomedical Informatics

# Τι θα μάθουμε

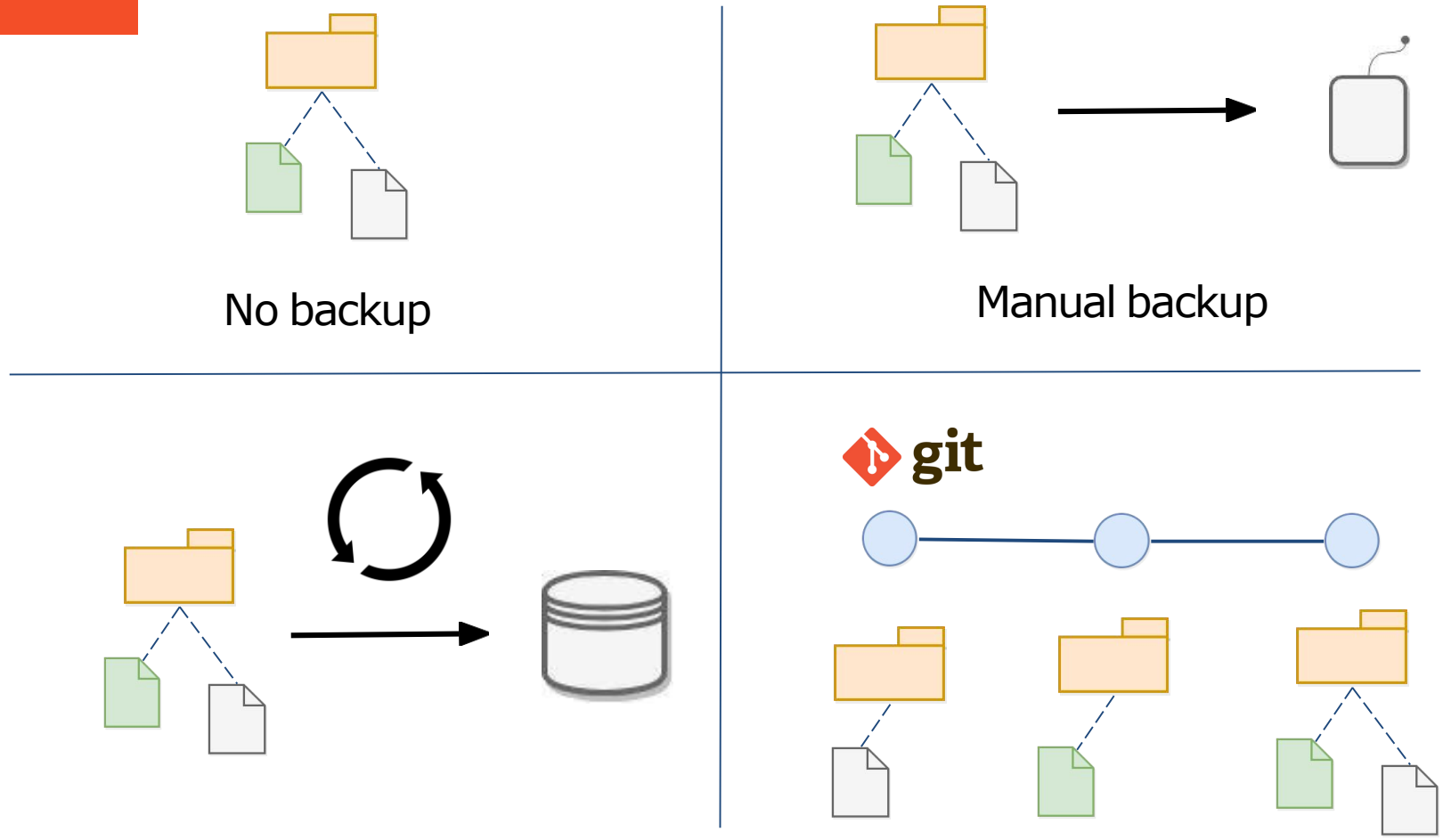
- Τι είναι το git
- Βασική χρήση git
- Δουλεύοντας τοπικά με git
- Δουλεύοντας απομακρυσμένα με git
- Συνεργασία μέσω git{,hub}

# Γιατί Version Control;

- Ως προγραμματιστές έχουμε ανάγκες
  - Νέος κώδικας μερικές φορές είναι buggy
  - Δουλεύουμε πολλοί ταυτόχρονα στον ίδιο κώδικα
  - Διαγράφουμε κώδικα που μπορεί να χρειαστεί ξανά
  - Χρειαζόμαστε back-ups για τη δουλειά μας

- Πώς κρατάμε πολλές εκδόσεις ενός αρχείου;
- Πώς επιστρέφουμε σε μία παλιά έκδοση;

# Γιατί Version Control;



# Γιατί Version Control;

- κρατάμε εκδόσεις στα αρχεία
- κάνουμε undo αλλαγές
- συνεργαζόμαστε με άλλους
- κρατάμε backups των αρχείων μας
- μοιραζόμαστε εύκολα τον κώδικα με την ομάδα
- ξέρουμε ποια είναι η «τελευταία» έκδοση


| | |
|---|---|
| 26th sept | 1 file, 2 mb |
| 27 sept | 1 file, 2 mb |
| 28 sept | 2 files, 3 mb |
| 29 sept | 2 files, 3.5 mb |
| 30 sept | 2 files, 3.5 mb |
| 1 oct | 3 files, 3.8 mb |
| 2 oct | 5 files, 5.2 mb |
| 3 oct | 5 files, 5.2 mb |
| 4 oct | 5 files, 5.5 mb |
| 5 oct | 5 files 5.5 mb |

| | | |
|---|---|---|
| 26th sept | Add first file | Tags |
| 26 sept | Make important edits | |
| 29 sept | Include second analysis | v2.0 Analysis 2 |
| 30 sept | Extend second analysis | |
| 30 sept | Add visualizations for first analysis | |
| 2 oct | Found errors in ANOVA, corrected | |
| 2 oct | Preparing third analysis | |
| 4 oct | Start third analysis | v3.0 Analysis 3 |

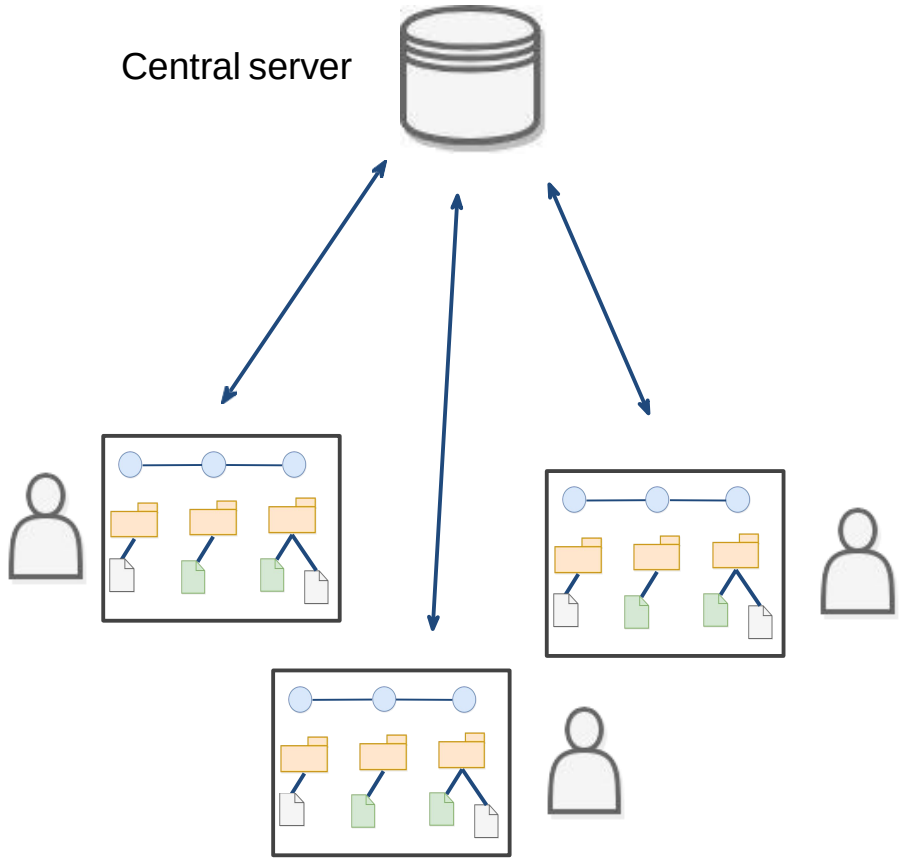
# Ιστορικά

- CVS – 1990
  - Από τα πρώτα πλήρη version control systems
- Subversion (SVN) – 2000
  - Διορθωμένο CVS για project-wide management
- git – 2005
  - Distributed version control system
- GitHub – 2008
  - Συνεργατικό περιβάλλον version control

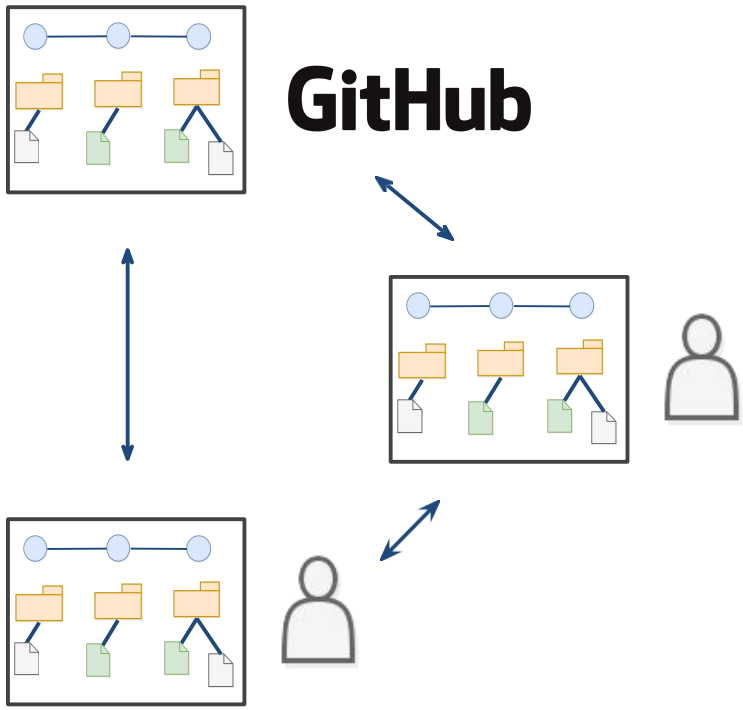
# Γιατί git;

## Centralized version control

SVN


## Distributed version control

# Τι είναι το git?

- Πρόγραμμα που τρέχεις στον υπολογιστή σου
- Εργαλείο στο command line
- Το αφήνεις να χειριστεί τον κώδικά σου

# Γιατί git;


Google Trends
Compare
git
Search term
svn
Search term
mercurial
Search term
Add comparison
Worldwide
2004 - present
All categories
Web Search
Interest over time
100
75
50
25
Note
Average
Jan 1, 2004
Nov 1, 2009
Sep 1, 2015

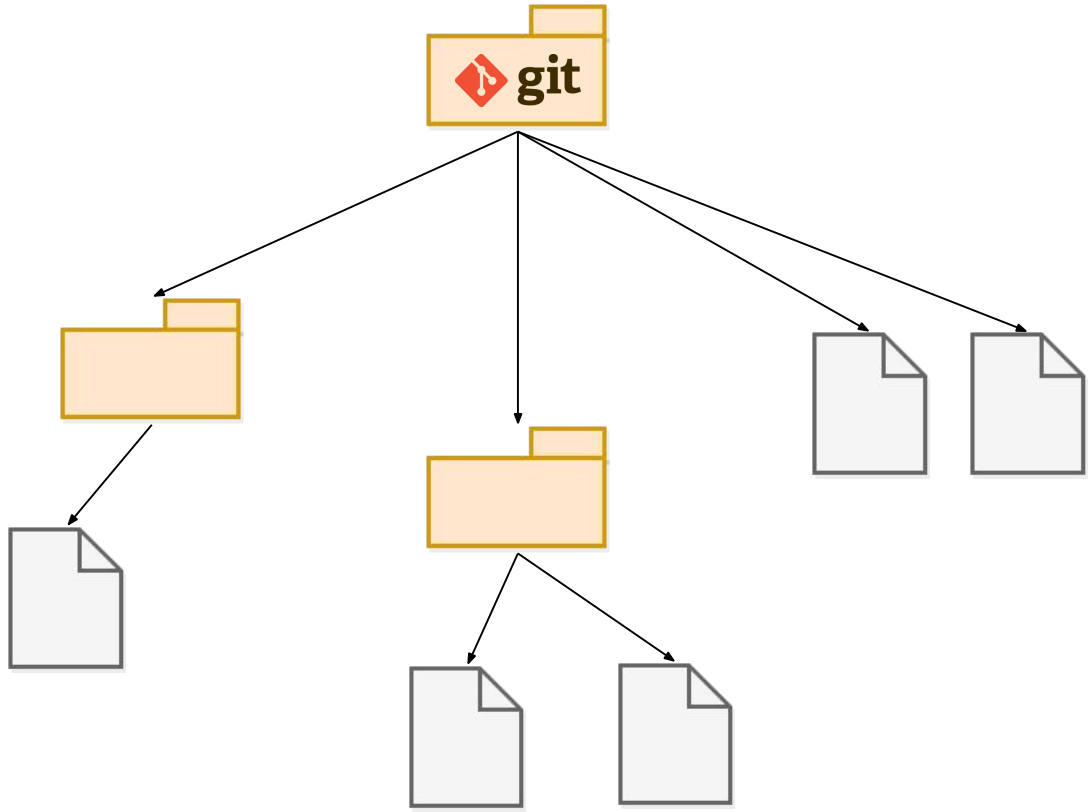
# Εγκατάσταση

- Linux (Debian, Ubuntu)
  - `apt-get install git`
- Mac
  - Τρέξε git και ακολούθα οδηγίες εγκατάστασης
- Windows
  - Κατέβασμα από το https://git-scm.com/download
  - Εγκατάσταση
  - Τρέξε το git CMD

# Τι κάνει το Git

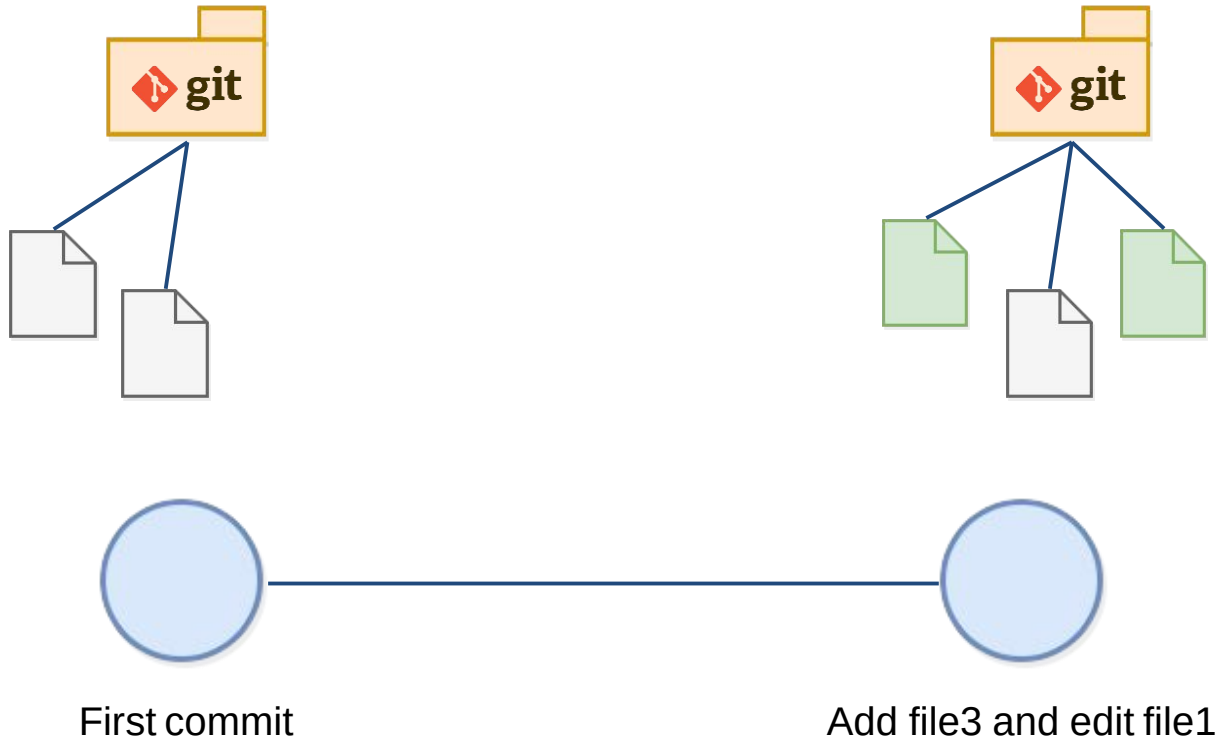
- Προσδιορίζει τι άλλαξε, ποιος το άλλαξε και γιατί
- Οργανωμένος τρόπος συνεργασίας στον κώδικα
- Τρόπος παρουσίασης κώδικα
- επιτρέπεται η πλοήγηση στο ιστορικό αρχείων

# The ‘repository’ and the ‘file tree’



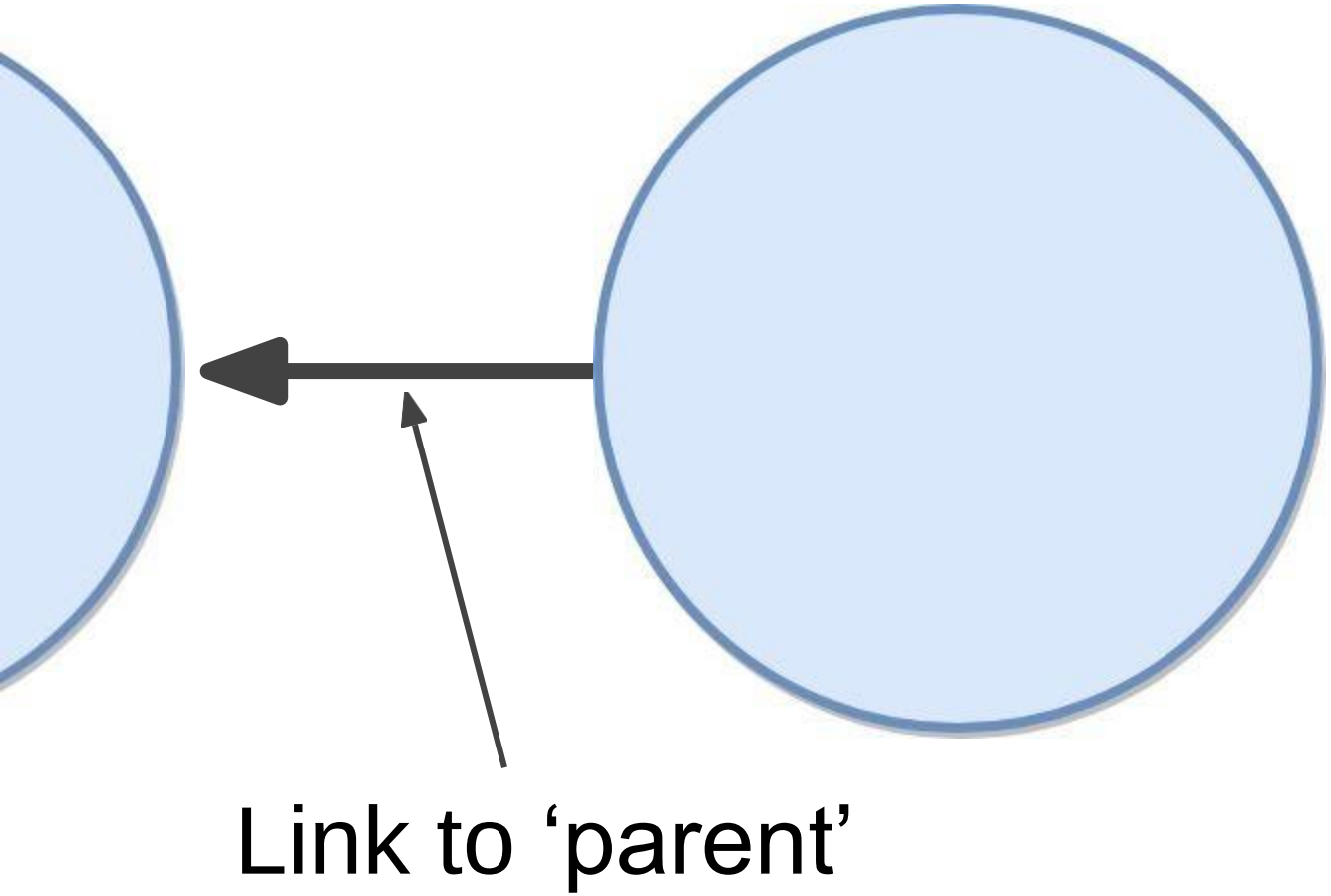

The ‘repository’ παρακολουθεί τις αλλαγές στο ‘file tree’

# What is a commit?




Ένα στιγμιότυπο συγκεκριμένης κατάστασης στο file tree

# What is a commit?

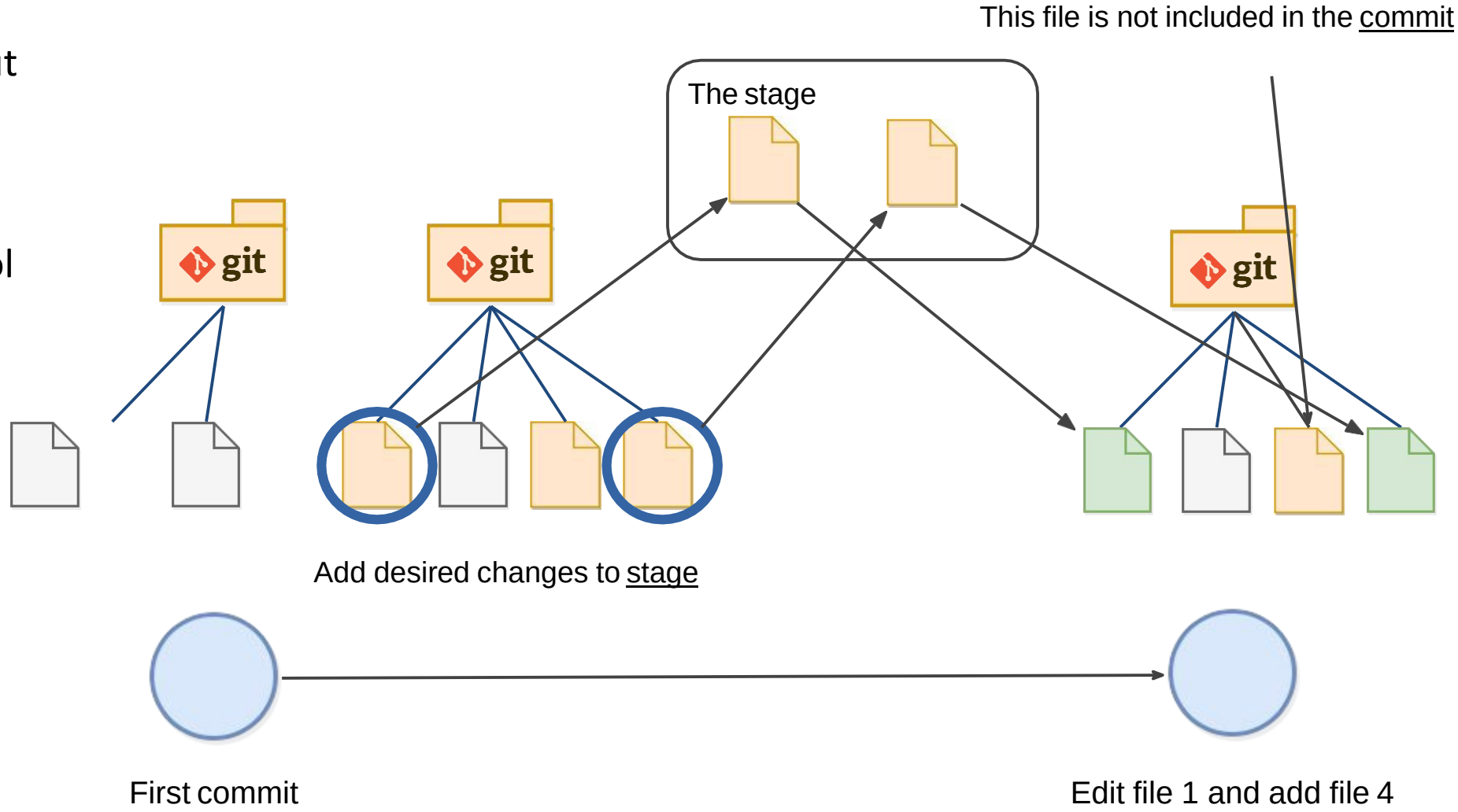


ID
Message
File changes
Author
Date

Link to ‘parent’

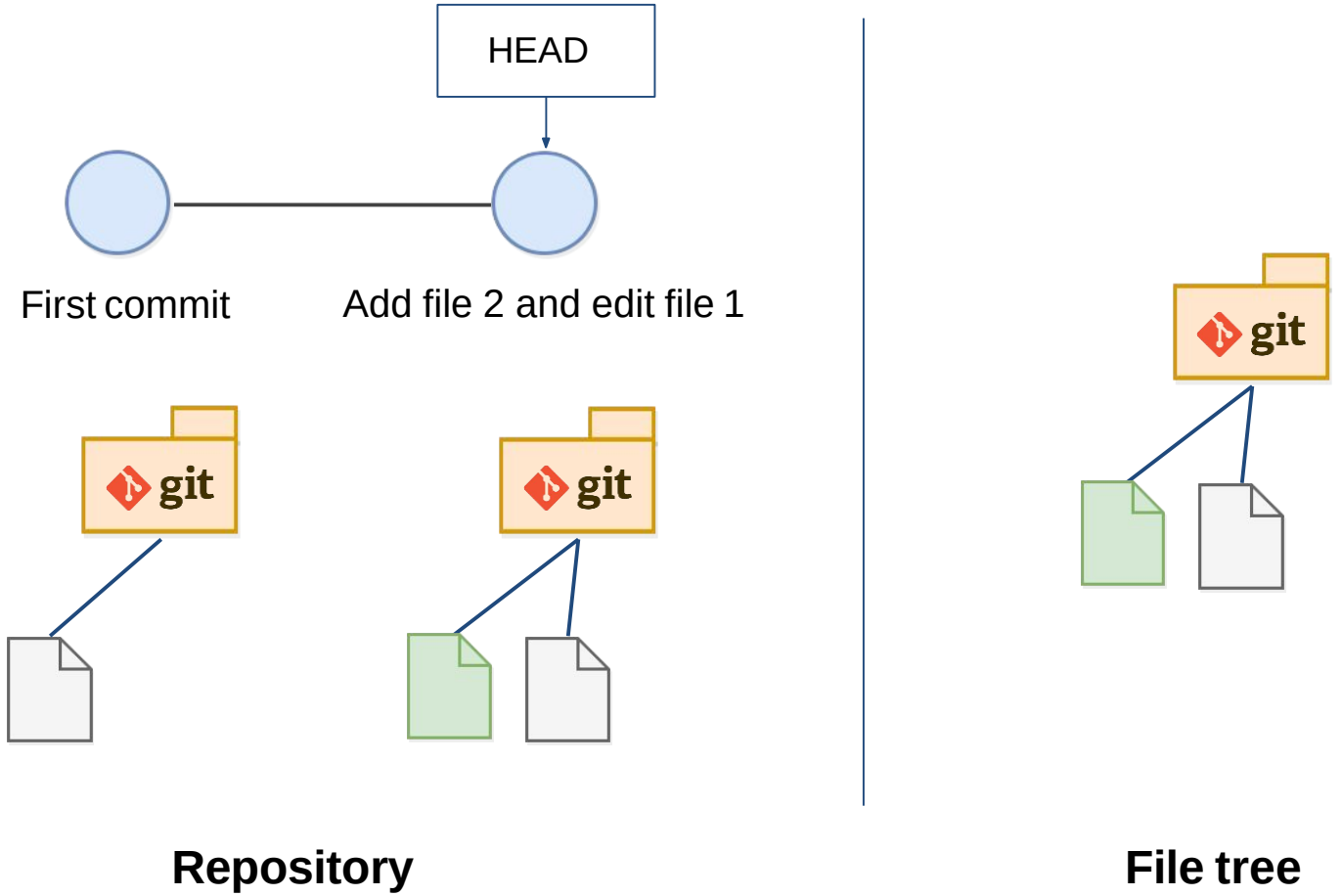
# What is the ‘stage’?

- Untracked: the file exists, but is not part of git's version control
- Staged: the file has been added to git's version control but changes have not been committed
- Committed: the change has been committed

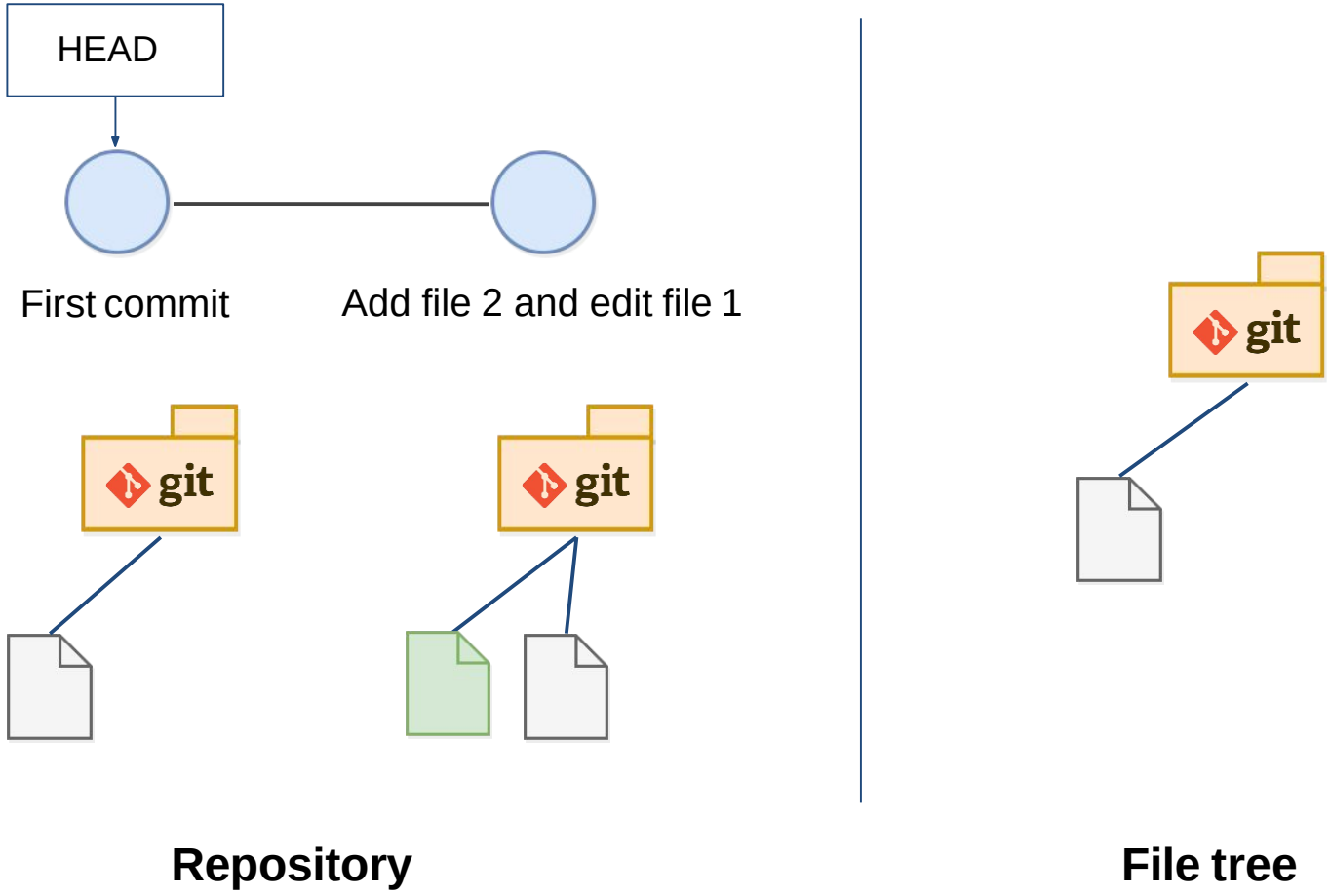



>>git status

# What is the ‘HEAD’?



# What is the ‘HEAD’?




By moving the HEAD, the file tree changes

# Comparing changes

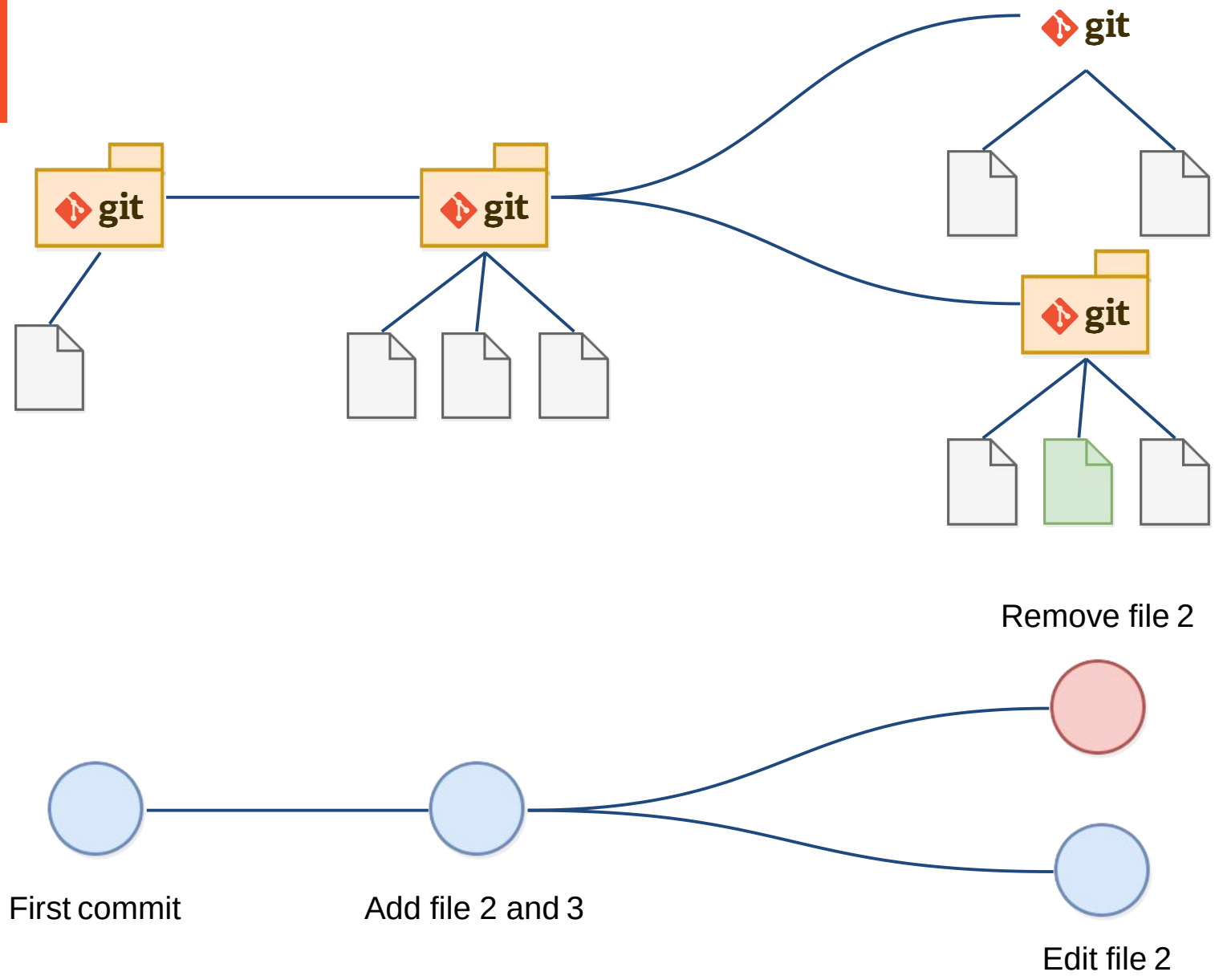

Differences between commits or branches can be easily visualized

Diff

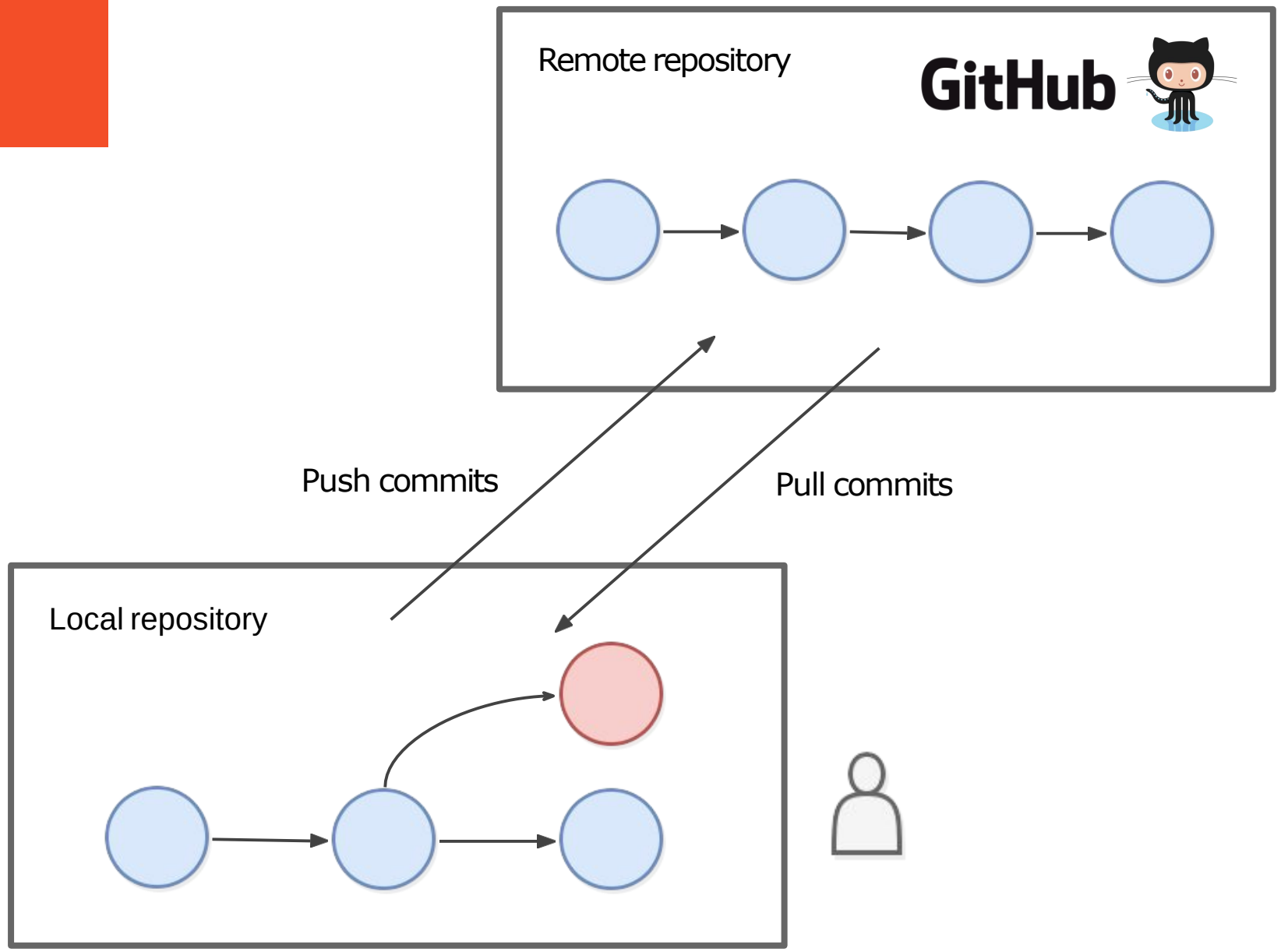
37 matching.R → algorithm.R — View

```
...  ...  @@ -1,6 +1,6 @@
1    1
2    2
3         - update_character_stats <- function(session, dict, character_stats, cur_ind) {
     3    + update_character_stats <- function(session, dict, character_stats, cur_ind, debug=T) {
4    4
5    5        input_english <- session$input$english
6    6        input_pinying <- session$input$pinying
     @@ -9,7 +9,8 @@ update_character_stats <- function(session, dict, character_stats, cur_ind) {
9    9            dict[[cur_ind]],
10   10           input_pinying,
11   11           input_english,
12        -       type=session$input$practice_type)
     12   +       type=session$input$practice_type,
     13   +       debug=debug)
13   14
```

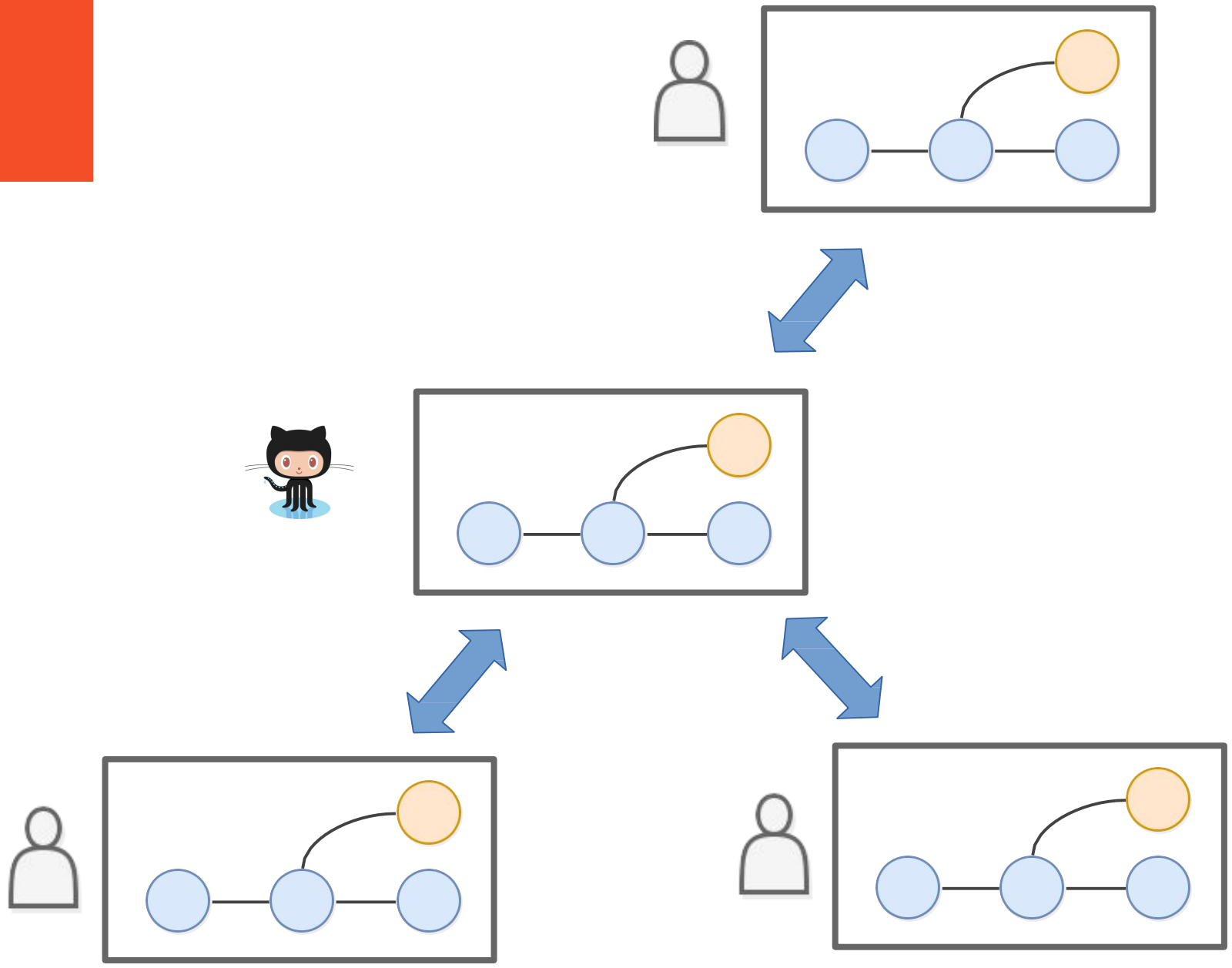
# What is a branch?



# The remote


Remote repository
GitHub
Push commits
Pull commits
Local repository

# Multiple users



# Remote repositories

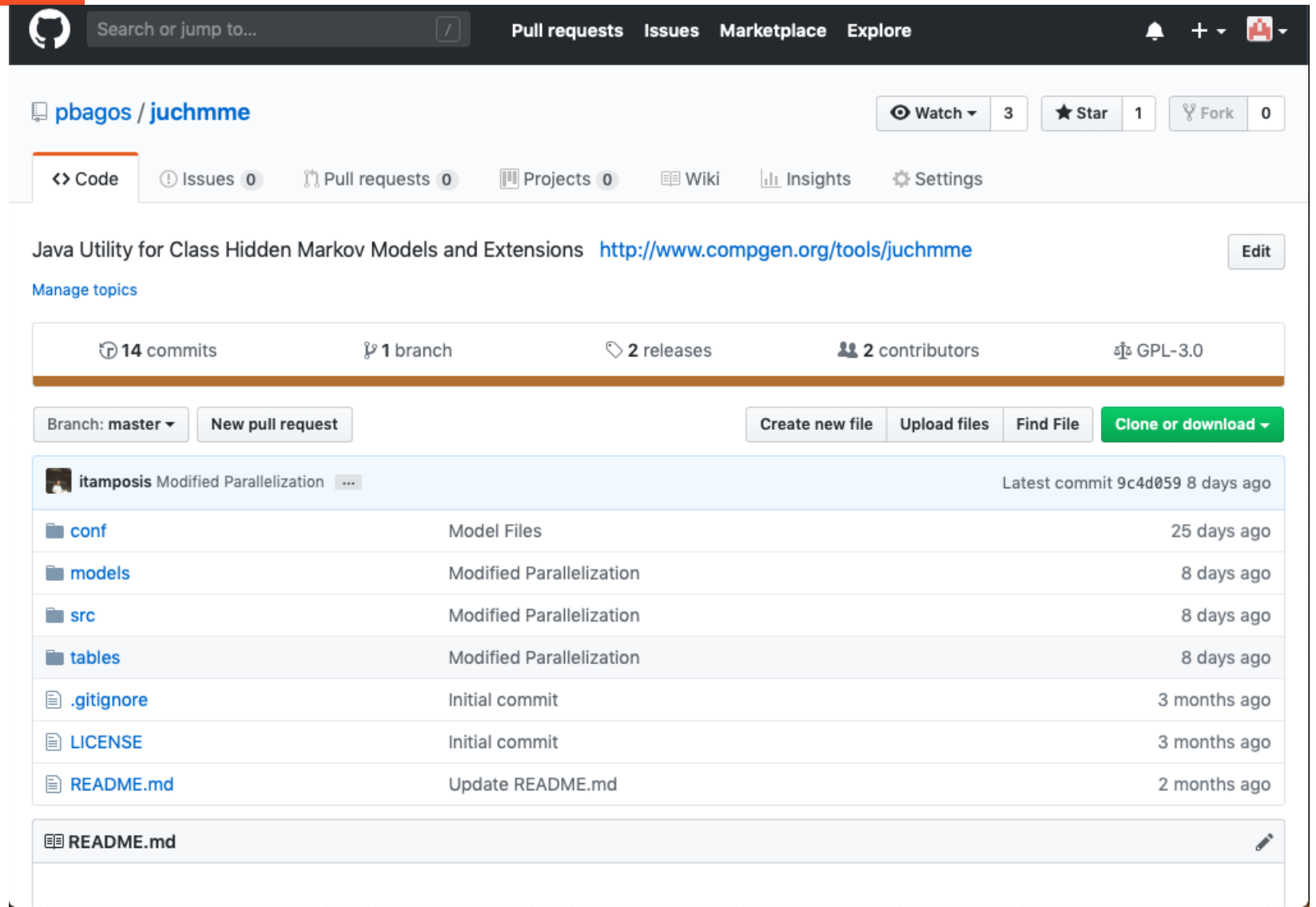
GitHub

- Κοινωνική πλατφόρμα για κώδικα
- Κοινός τρόπος για να γίνει δημόσιος ο κώδικας
- Επιτρέπει την αλληλεπίδραση με τον κώδικα άλλων λαών

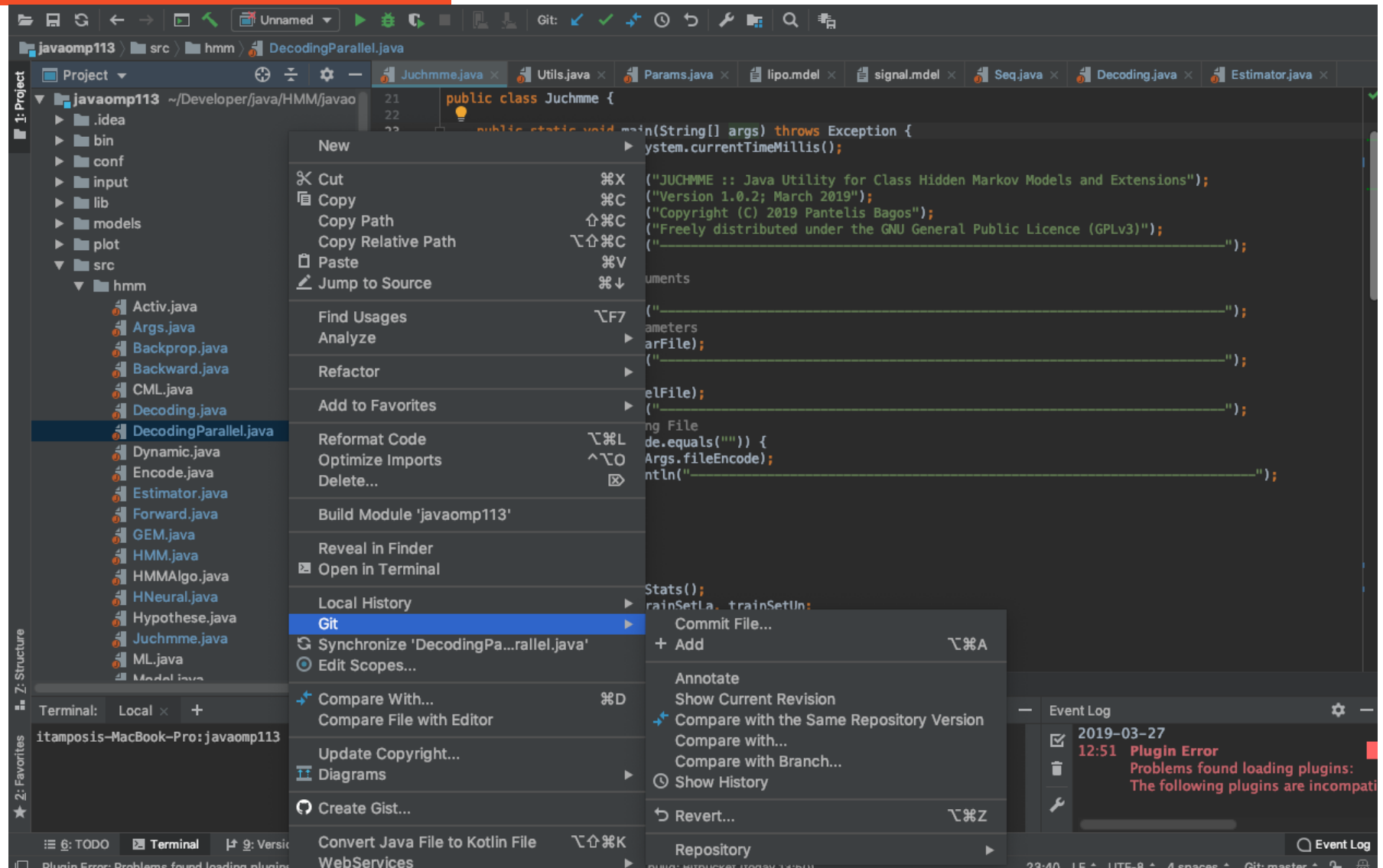
GitLab

Bitbucket

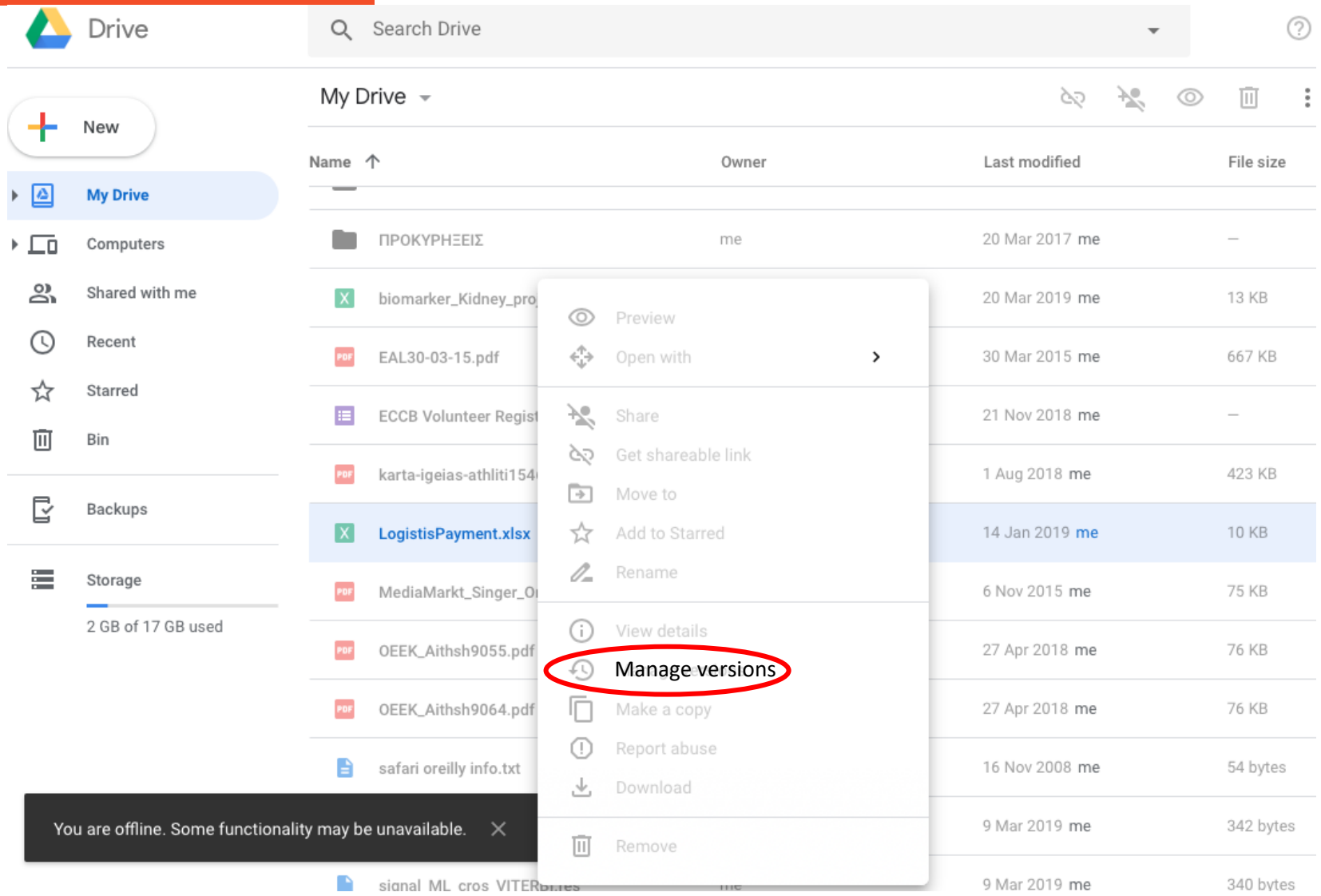
# Remote repositories



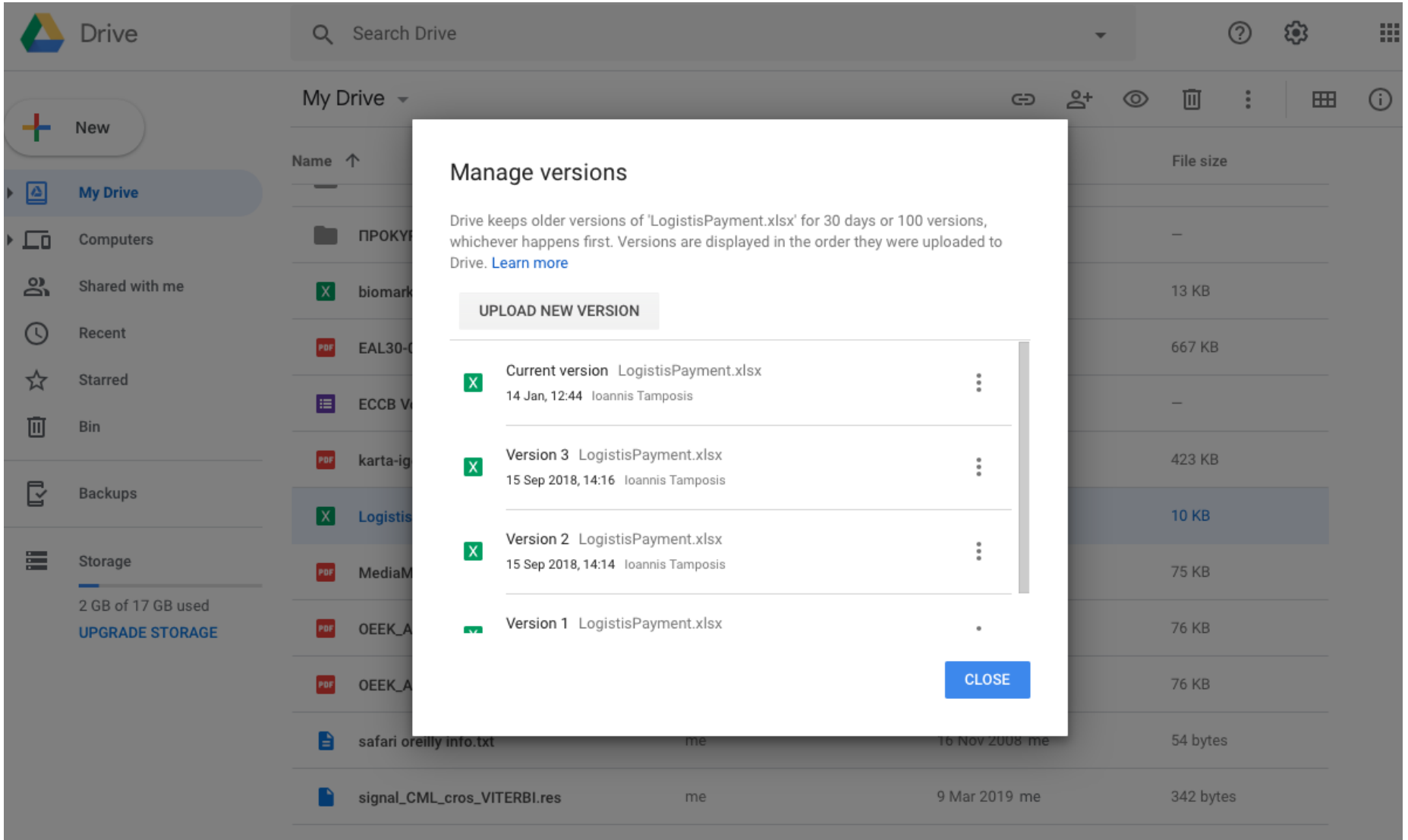
# IDE


javaomp113 src hmm DecodingParallel.java
Project
javaomp113 ~/Developer/java/HMM/javao
.idea
bin
conf
input
lib
models
plot
src
hmm
Activ.java
Args.java
Backprop.java
Backward.java
CML.java
Decoding.java
DecodingParallel.java
Dynamic.java
Encode.java
Estimator.java
Forward.java
GEM.java
HMM.java
HMMAlgo.java
HNeural.java
Hypothese.java
Juchmme.java
ML.java
Juchmme.java
Utils.java
Params.java
lipo.mdel
signal.mdel
Seq.java
Decoding.java
Estimator.java
public class Juchmme {
("JUCHMME :: Java Utility for Class Hidden Markov Models and Extensions");
("Version 1.0.2; March 2019");
("Copyright (C) 2019 Pantelis Bagos");
("Freely distributed under the GNU General Public Licence (GPLv3)");
New
Cut ⌘X
Copy ⌘C
Copy Path ⇧⌘C
Copy Relative Path ⌥⇧⌘C
Paste ⌘V
Jump to Source ⌘↓
Find Usages ⌥F7
Analyze
Refactor
Add to Favorites
Reformat Code ⌥⌘L
Optimize Imports ^⌥O
Delete...
Build Module 'javaomp113'
Reveal in Finder
Open in Terminal
Local History
Git
Synchronize 'DecodingPa...rallel.java'
Edit Scopes...
Compare With... ⌘D
Compare File with Editor
Update Copyright...
Diagrams
Create Gist...
Convert Java File to Kotlin File ⌥⇧⌘K
WebServices
Commit File...
Add ⌥⌘A
Annotate
Show Current Revision
Compare with the Same Repository Version
Compare with...
Compare with Branch...
Show History
Revert... ⌥⌘Z
Repository
Terminal: Local
itamposis-MacBook-Pro:javaomp113
Event Log
2019-03-27
12:51 Plugin Error
Problems found loading plugins:
The following plugins are incompati
6: TODO
Terminal
Event Log

# Google Drive



# Google Drive



# Getting started

**Πώς να το χρησιμοποιήσετε**

Από γραμμή εντολών:
https://git-scm.com

Από γραφικό περιβάλλον:
https://desktop.github.com

**Πώς να το μάθετε**

Codecademy - Interactive online tutorial (free):
https://www.codecademy.com/learn/learn-git