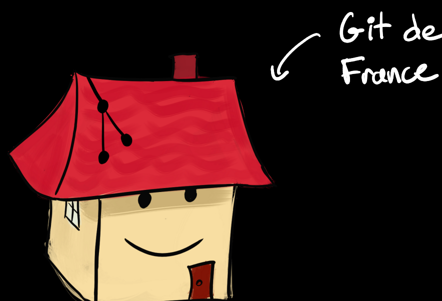
Which Distribution for Git Graphs

(ongoing work)

Julien COURTIEL (Université de Caen Normandie)
with Martin PEPIN (Université de Caen Normandie)

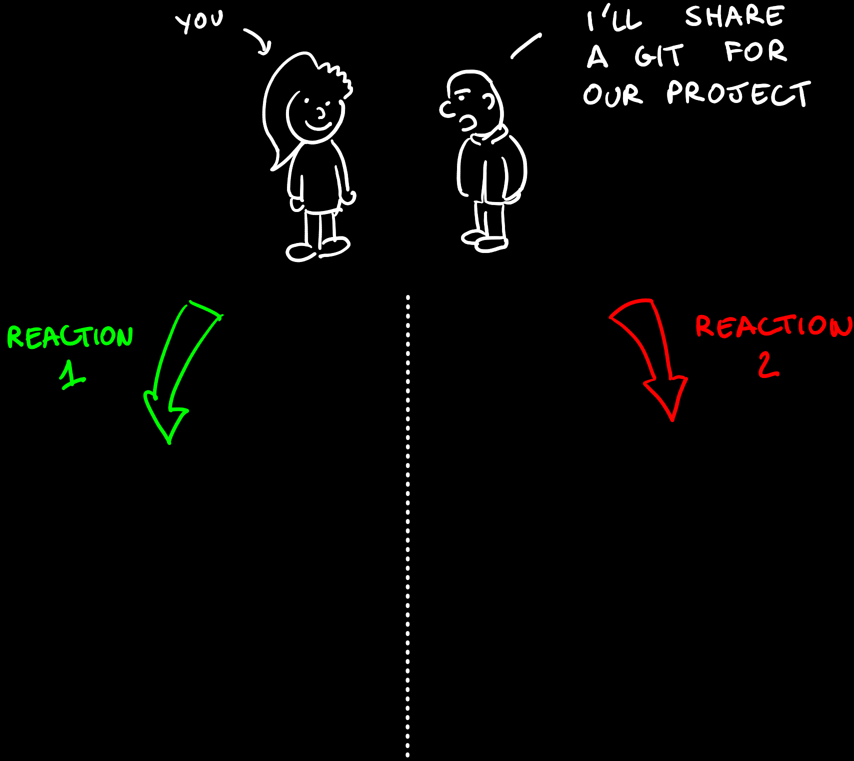


MAD Days 2025 (Rouen)

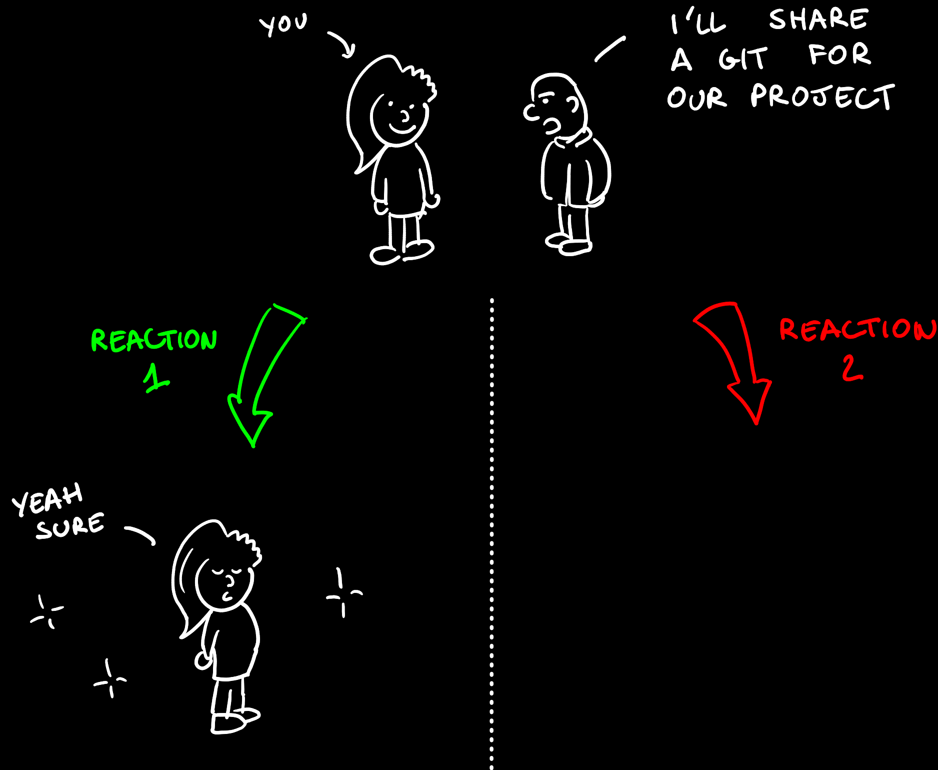
INTRODUCTION

HOW GIT WORKS

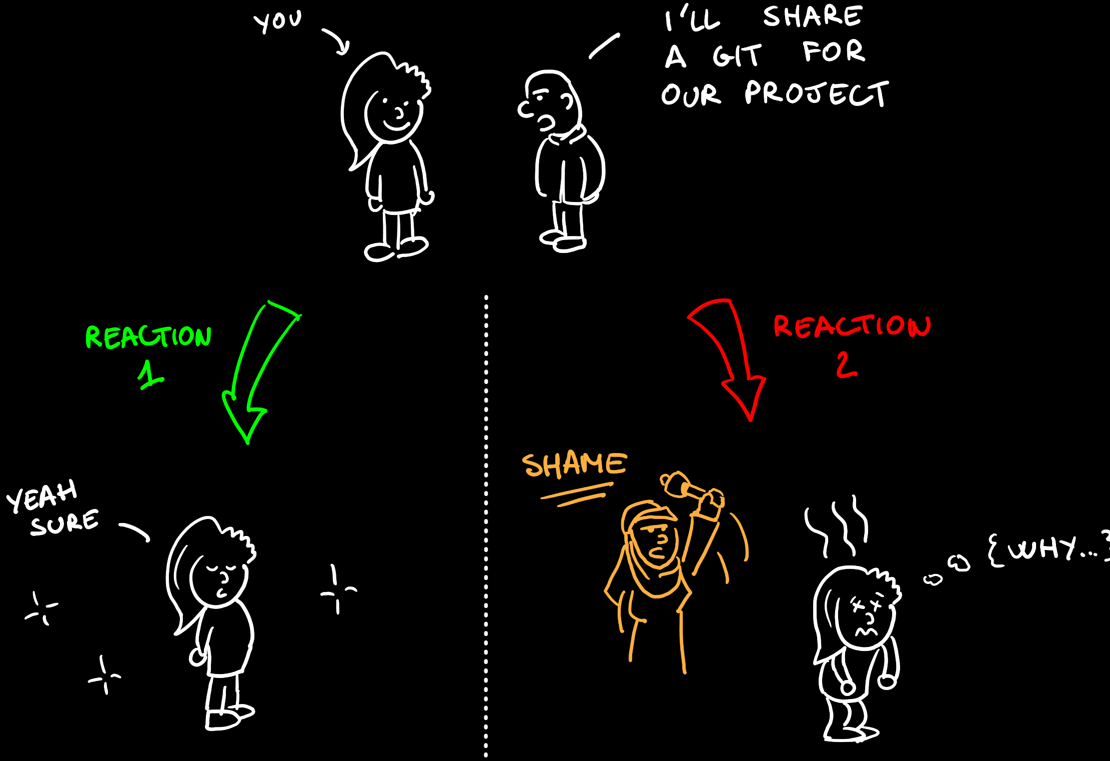
SITUATION PLAY



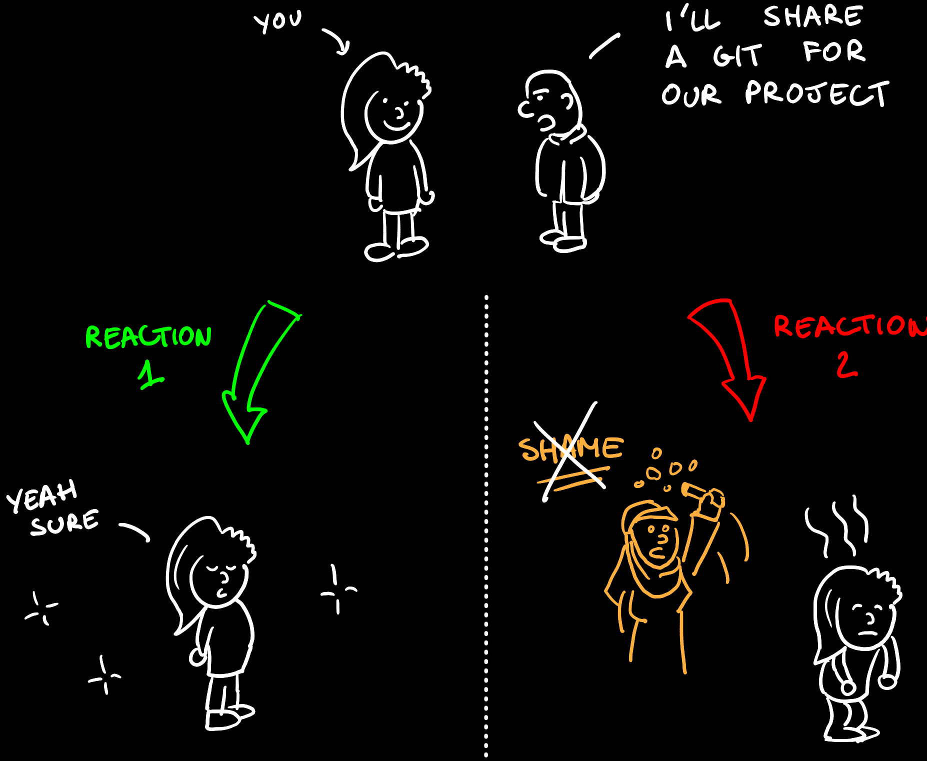
SITUATION PLAY



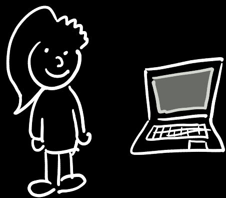
SITUATION PLAY



SITUATION PLAY



WORKING TOGETHER ON AN ARTICLE WITH GIT



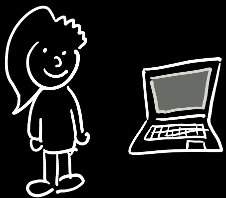
YOU

Situation play 2

You write an article
with an obscure coauthor

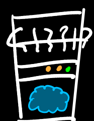
OBSCURE
COAUTHOR

WORKING TOGETHER ON AN ARTICLE WITH GIT



YOU

REMOTE SERVER

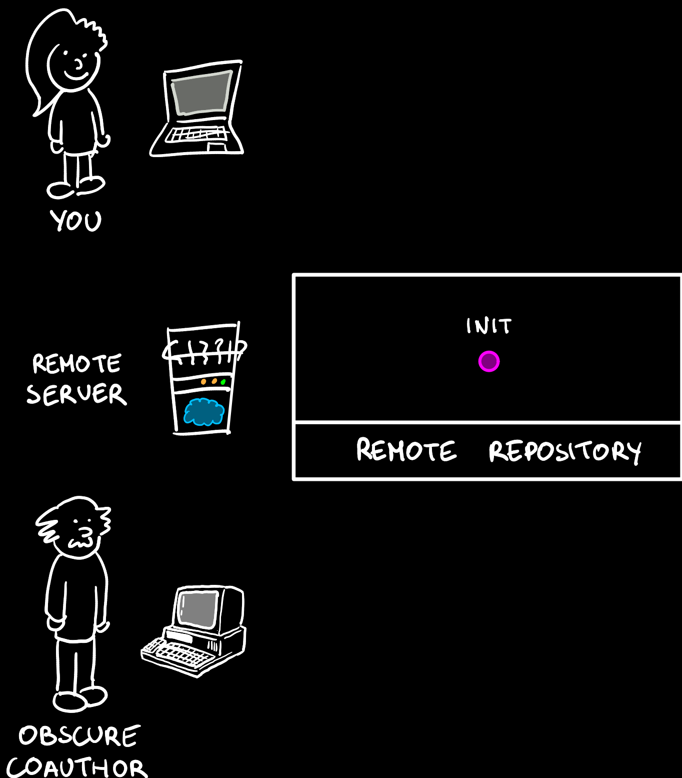


OBSCURE COAUTHOR

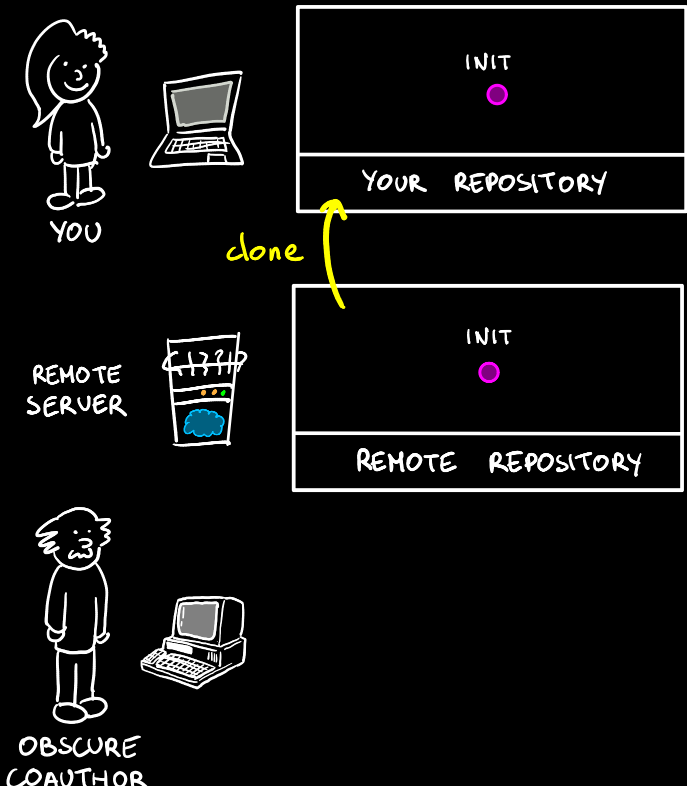
Situation play 2

You write an article with an obscure coauthor

WORKING TOGETHER ON AN ARTICLE WITH GIT

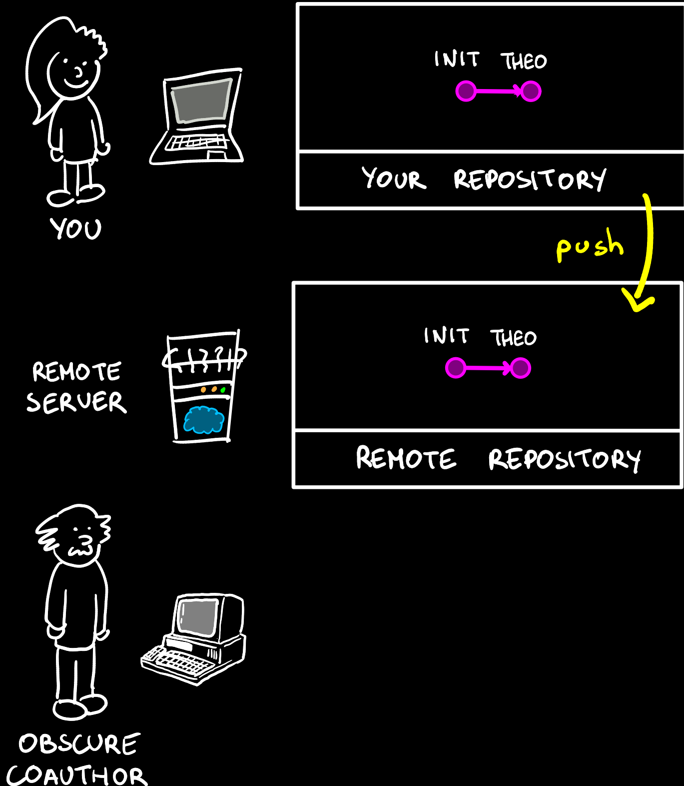


WORKING TOGETHER ON AN ARTICLE WITH GIT

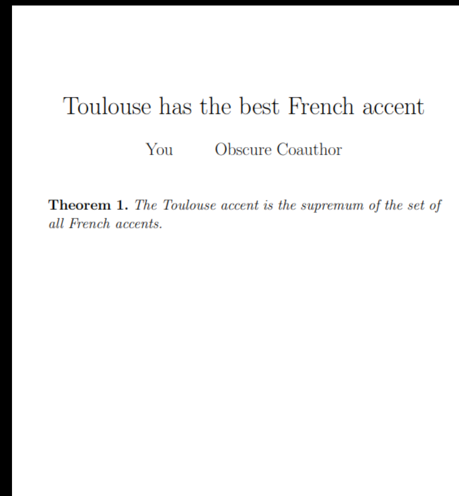


WORKING TOGETHER ON AN ARTICLE WITH GIT

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf

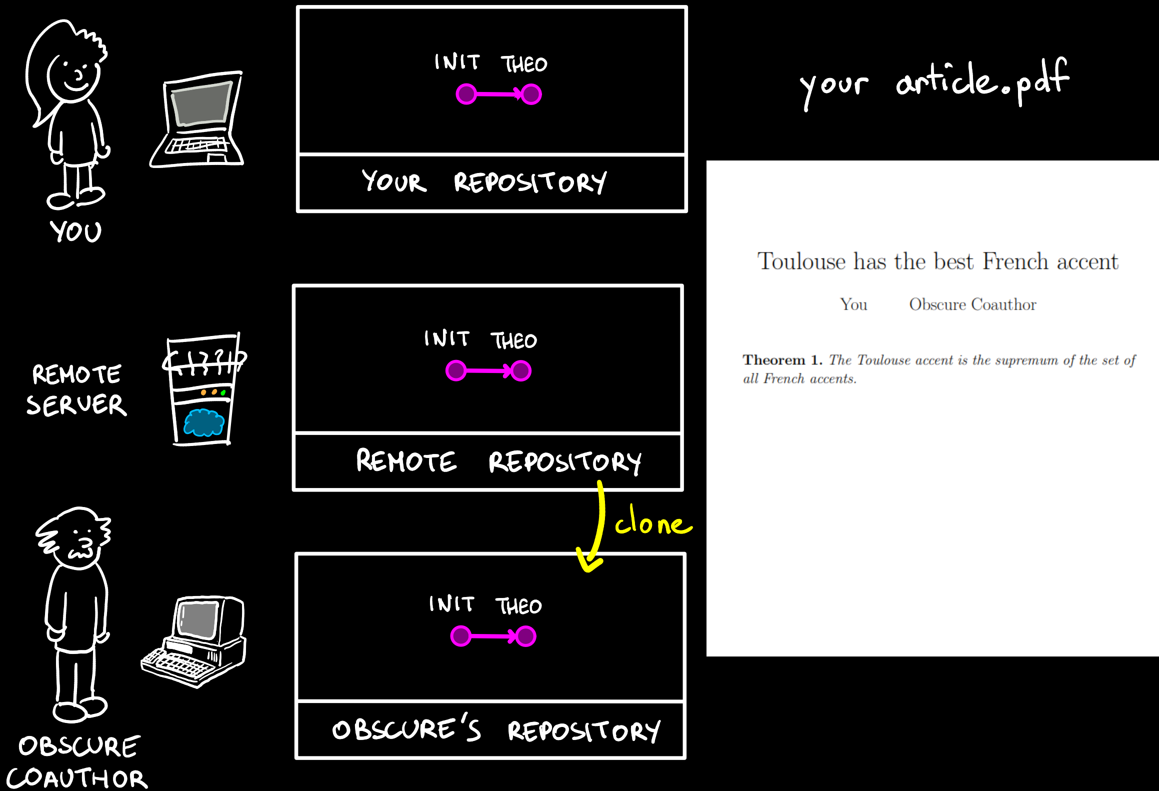


Toulouse has the best French accent

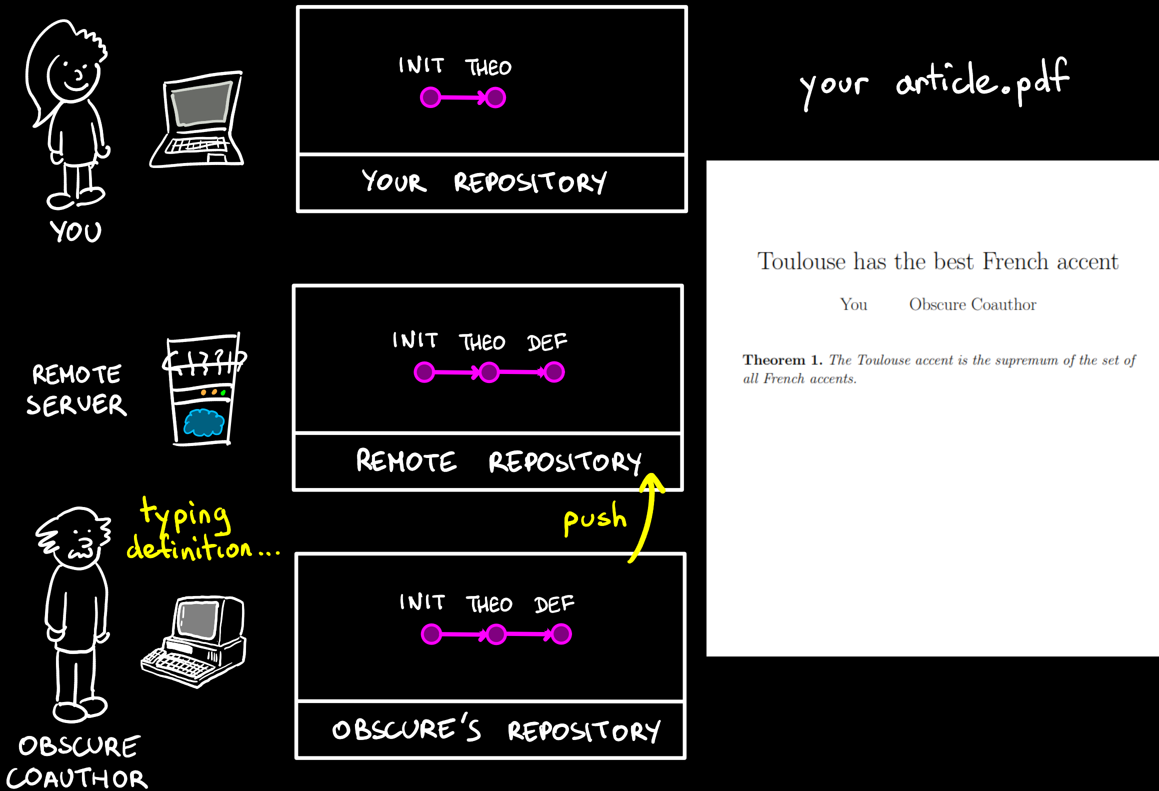
You Obscure Coauthor

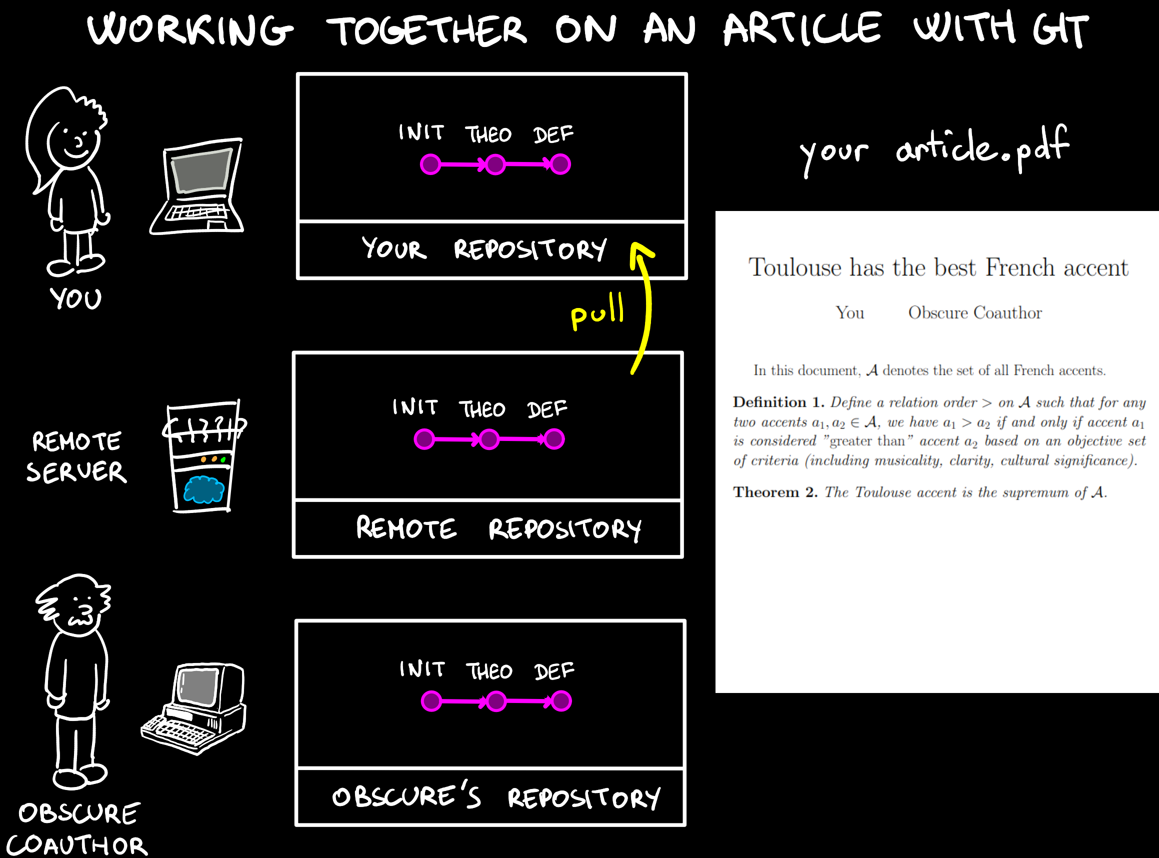
Theorem 1. *The Toulouse accent is the supremum of the set of all French accents.*

WORKING TOGETHER ON AN ARTICLE WITH GIT

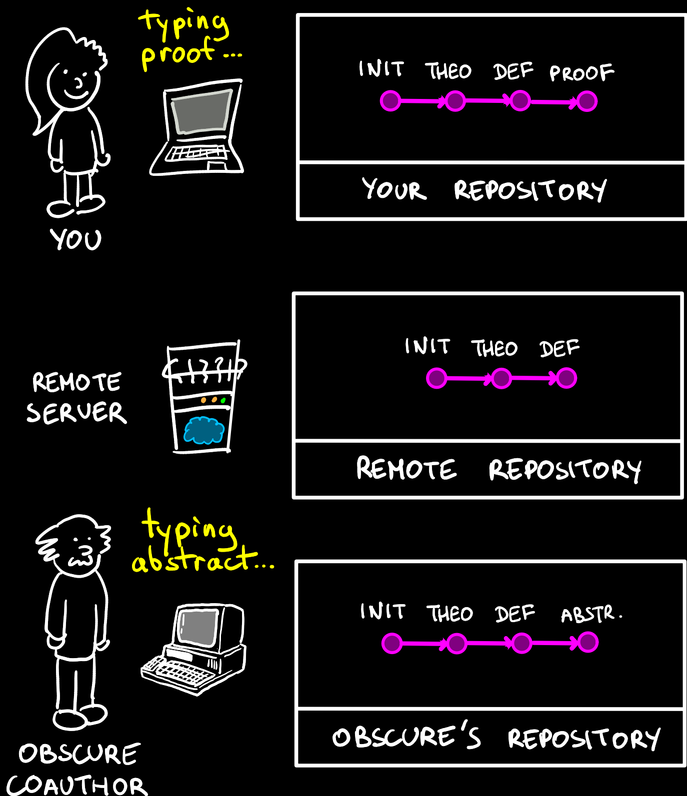


WORKING TOGETHER ON AN ARTICLE WITH GIT

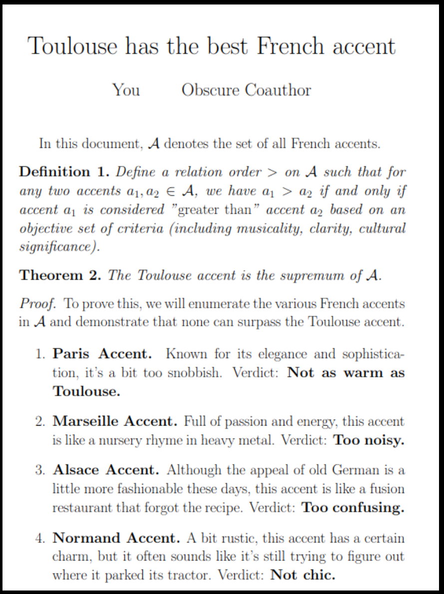



WORKING TOGETHER ON AN ARTICLE WITH GIT
YOU
INIT THEO DEF
YOUR REPOSITORY
your article.pdf
Toulouse has the best French accent
You Obscure Coauthor
In this document, $\mathcal{A}$ denotes the set of all French accents.
Definition 1. *Define a relation order > on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered* "greater than" *accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*
Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*
pull
REMOTE SERVER
INIT THEO DEF
REMOTE REPOSITORY
OBSCURE COAUTHOR
INIT THEO DEF
OBSCURE'S REPOSITORY

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf



Toulouse has the best French accent

You Obscure Coauthor

In this document, $\mathcal{A}$ denotes the set of all French accents.

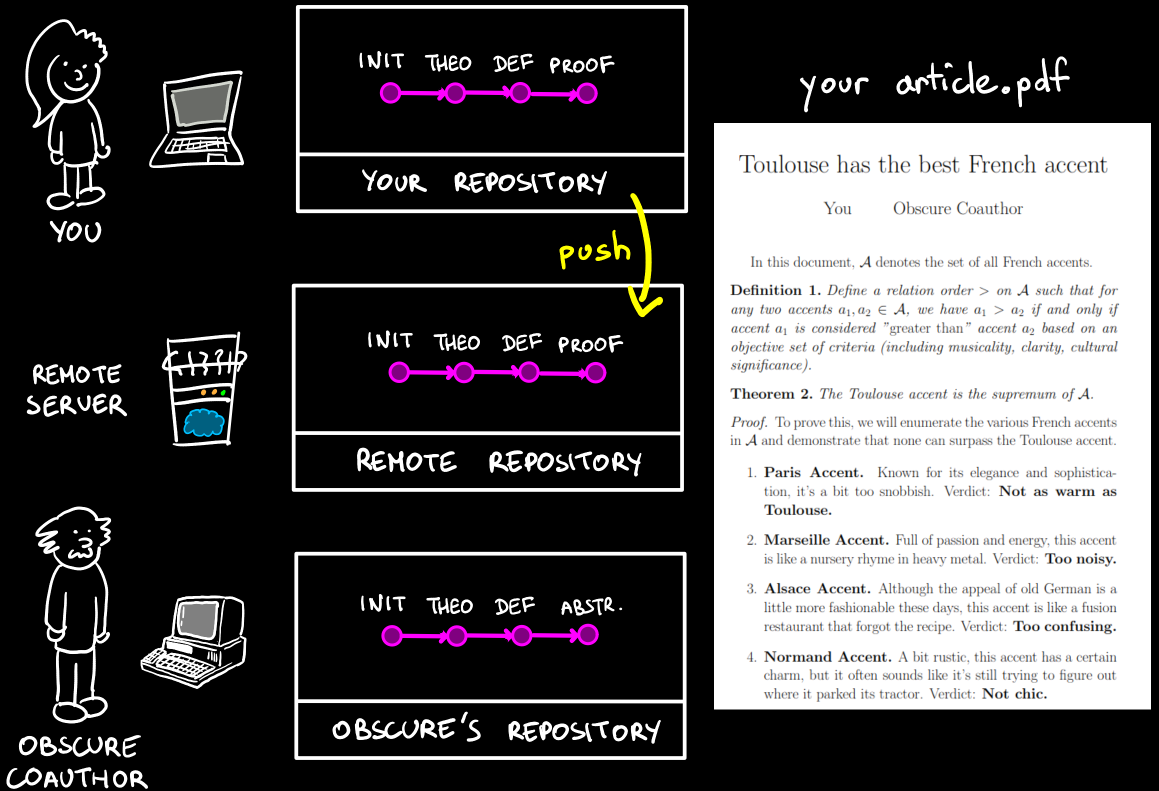
Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered* "greater than" *accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

Proof. To prove this, we will enumerate the various French accents in $\mathcal{A}$ and demonstrate that none can surpass the Toulouse accent.

1. **Paris Accent.** Known for its elegance and sophistication, it's a bit too snobbish. Verdict: **Not as warm as Toulouse.**
2. **Marseille Accent.** Full of passion and energy, this accent is like a nursery rhyme in heavy metal. Verdict: **Too noisy.**
3. **Alsace Accent.** Although the appeal of old German is a little more fashionable these days, this accent is like a fusion restaurant that forgot the recipe. Verdict: **Too confusing.**
4. **Normand Accent.** A bit rustic, this accent has a certain charm, but it often sounds like it's still trying to figure out where it parked its tractor. Verdict: **Not chic.**

WORKING TOGETHER ON AN ARTICLE WITH GIT



Toulouse has the best French accent

You Obscure Coauthor

In this document, $\mathcal{A}$ denotes the set of all French accents.

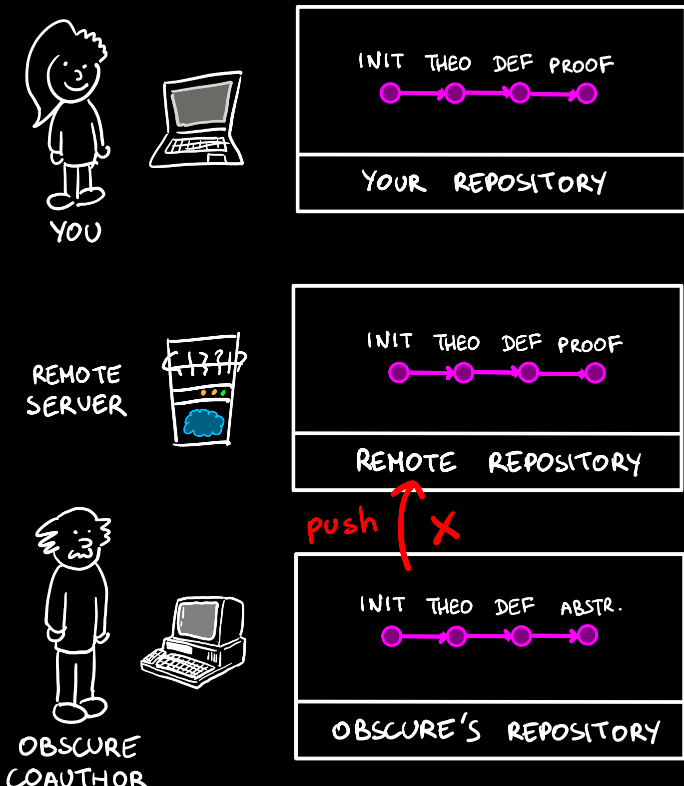
Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered* "greater than" *accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

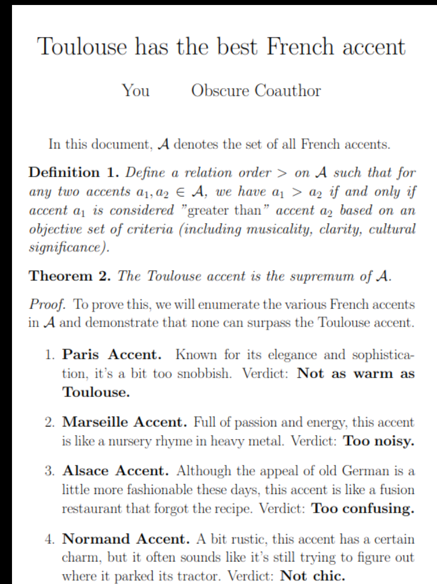
Proof. To prove this, we will enumerate the various French accents in $\mathcal{A}$ and demonstrate that none can surpass the Toulouse accent.

1. **Paris Accent.** Known for its elegance and sophistication, it's a bit too snobbish. Verdict: **Not as warm as Toulouse.**
2. **Marseille Accent.** Full of passion and energy, this accent is like a nursery rhyme in heavy metal. Verdict: **Too noisy.**
3. **Alsace Accent.** Although the appeal of old German is a little more fashionable these days, this accent is like a fusion restaurant that forgot the recipe. Verdict: **Too confusing.**
4. **Normand Accent.** A bit rustic, this accent has a certain charm, but it often sounds like it's still trying to figure out where it parked its tractor. Verdict: **Not chic.**

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf



Toulouse has the best French accent

You Obscure Coauthor

In this document, $\mathcal{A}$ denotes the set of all French accents.

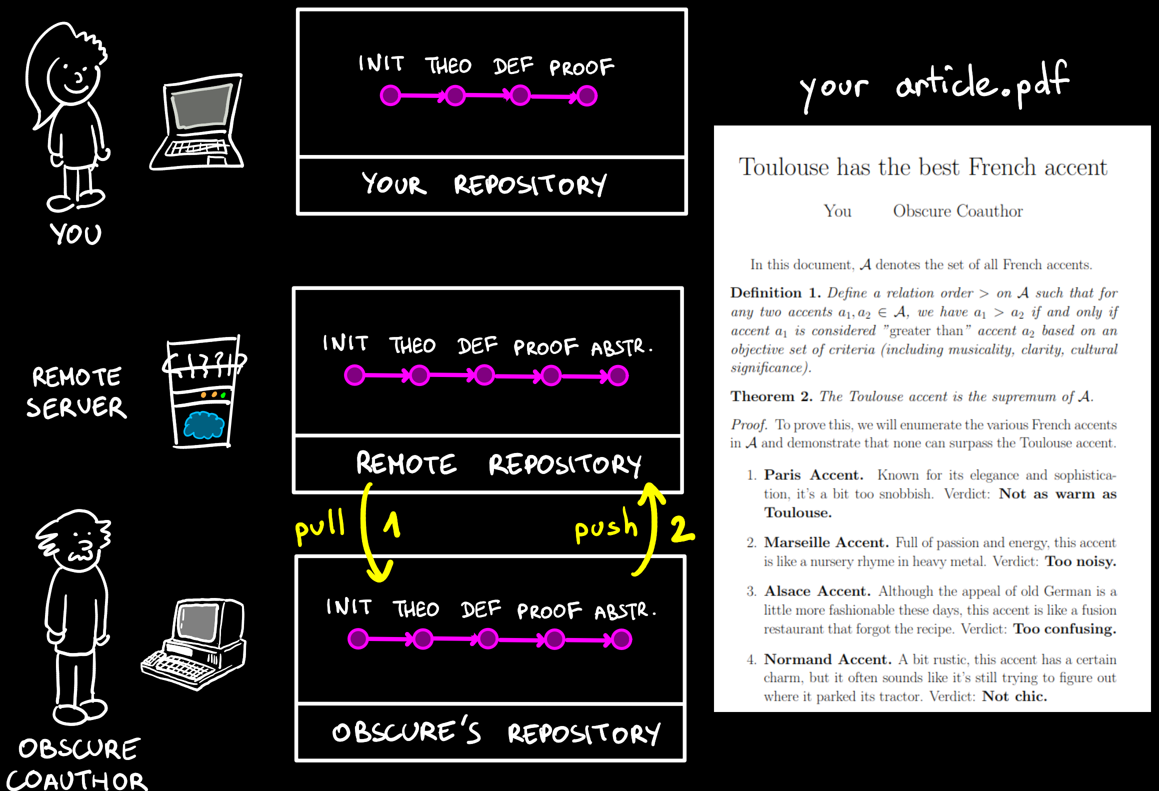
Definition 1. *Define a relation order* $>$ *on* $\mathcal{A}$ *such that for any two accents* $a_1, a_2 \in \mathcal{A}$*, we have* $a_1 > a_2$ *if and only if accent* a_1 *is considered* "greater than" *accent* a_2 *based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of* $\mathcal{A}$.

Proof. To prove this, we will enumerate the various French accents in $\mathcal{A}$ and demonstrate that none can surpass the Toulouse accent.

1. **Paris Accent.** Known for its elegance and sophistication, it's a bit too snobbish. Verdict: **Not as warm as Toulouse.**
2. **Marseille Accent.** Full of passion and energy, this accent is like a nursery rhyme in heavy metal. Verdict: **Too noisy.**
3. **Alsace Accent.** Although the appeal of old German is a little more fashionable these days, this accent is like a fusion restaurant that forgot the recipe. Verdict: **Too confusing.**
4. **Normand Accent.** A bit rustic, this accent has a certain charm, but it often sounds like it's still trying to figure out where it parked its tractor. Verdict: **Not chic.**

WORKING TOGETHER ON AN ARTICLE WITH GIT



Toulouse has the best French accent

You Obscure Coauthor

In this document, $\mathcal{A}$ denotes the set of all French accents.

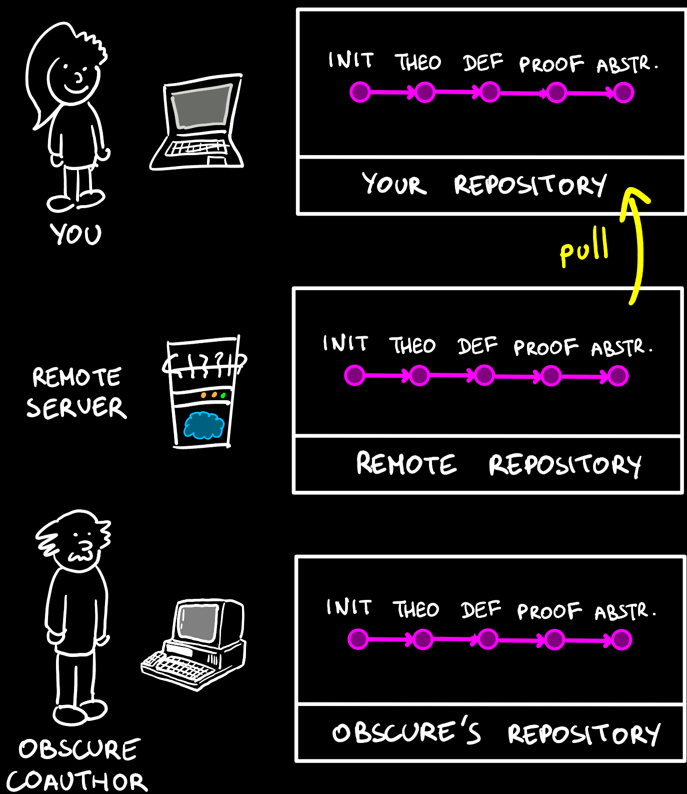
Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered* "greater than" *accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

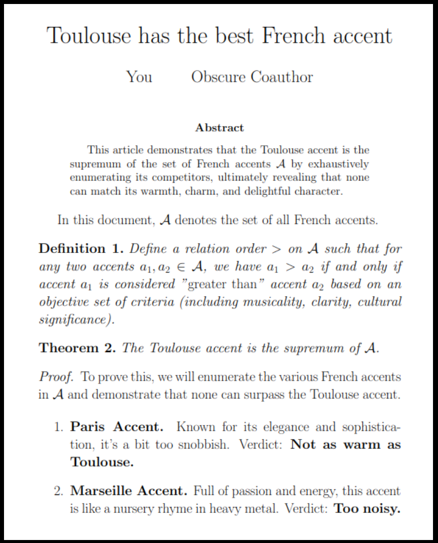
Proof. To prove this, we will enumerate the various French accents in $\mathcal{A}$ and demonstrate that none can surpass the Toulouse accent.

1. **Paris Accent.** Known for its elegance and sophistication, it's a bit too snobbish. Verdict: **Not as warm as Toulouse.**
2. **Marseille Accent.** Full of passion and energy, this accent is like a nursery rhyme in heavy metal. Verdict: **Too noisy.**
3. **Alsace Accent.** Although the appeal of old German is a little more fashionable these days, this accent is like a fusion restaurant that forgot the recipe. Verdict: **Too confusing.**
4. **Normand Accent.** A bit rustic, this accent has a certain charm, but it often sounds like it's still trying to figure out where it parked its tractor. Verdict: **Not chic.**

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf



Toulouse has the best French accent

You Obscure Coauthor

Abstract

This article demonstrates that the Toulouse accent is the supremum of the set of French accents $\mathcal{A}$ by exhaustively enumerating its competitors, ultimately revealing that none can match its warmth, charm, and delightful character.

In this document, $\mathcal{A}$ denotes the set of all French accents.

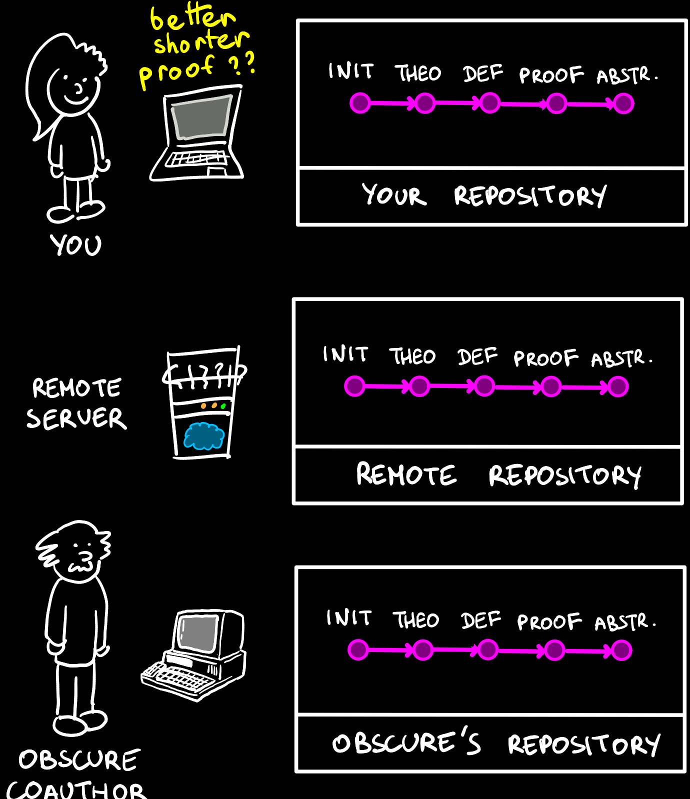
Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered "greater than" accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

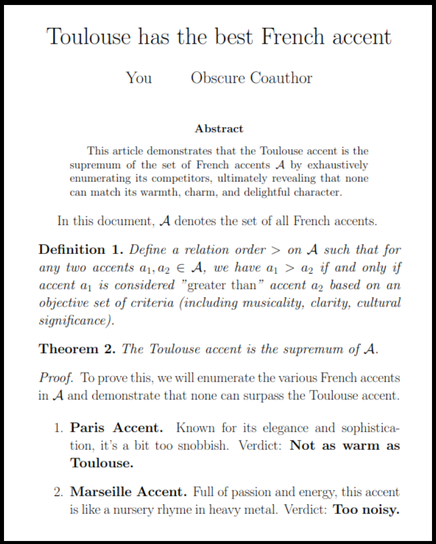
Proof. To prove this, we will enumerate the various French accents in $\mathcal{A}$ and demonstrate that none can surpass the Toulouse accent.

1. **Paris Accent.** Known for its elegance and sophistication, it's a bit too snobbish. Verdict: **Not as warm as Toulouse.**
2. **Marseille Accent.** Full of passion and energy, this accent is like a nursery rhyme in heavy metal. Verdict: **Too noisy.**

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf



Toulouse has the best French accent

You Obscure Coauthor

Abstract

This article demonstrates that the Toulouse accent is the supremum of the set of French accents $\mathcal{A}$ by exhaustively enumerating its competitors, ultimately revealing that none can match its warmth, charm, and delightful character.

In this document, $\mathcal{A}$ denotes the set of all French accents.

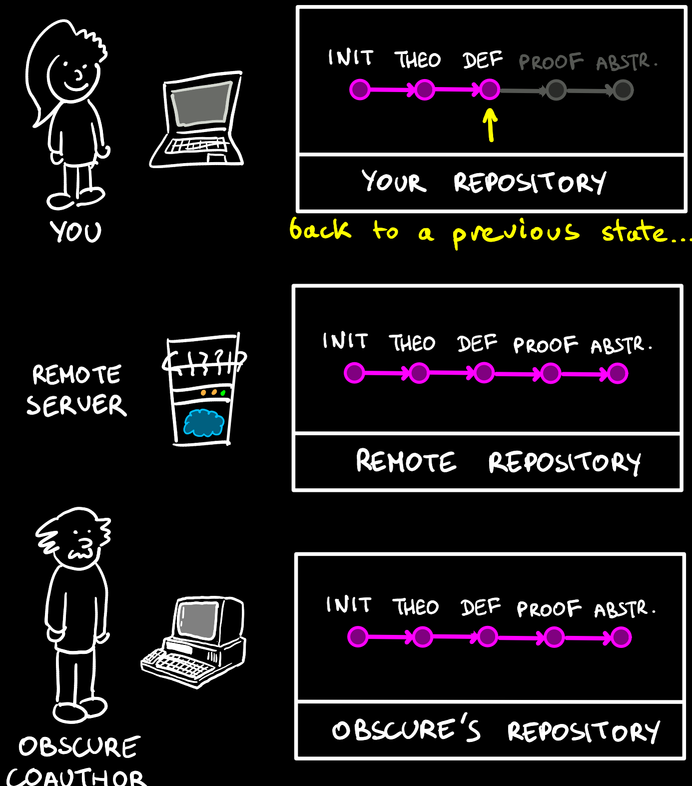
Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered "greater than" accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

Proof. To prove this, we will enumerate the various French accents in $\mathcal{A}$ and demonstrate that none can surpass the Toulouse accent.

1. **Paris Accent.** Known for its elegance and sophistication, it's a bit too snobbish. Verdict: **Not as warm as Toulouse.**
2. **Marseille Accent.** Full of passion and energy, this accent is like a nursery rhyme in heavy metal. Verdict: **Too noisy.**

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf

Toulouse has the best French accent

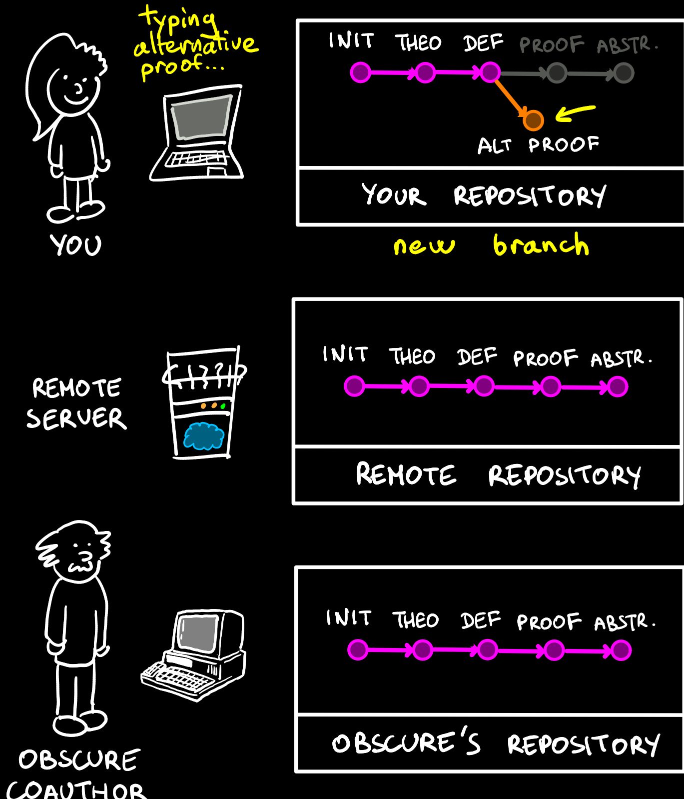
You   Obscure Coauthor

In this document, $\mathcal{A}$ denotes the set of all French accents.

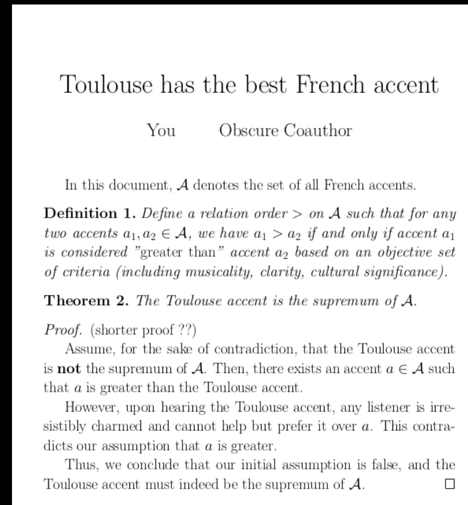
Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered* "greater than" *accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf



Toulouse has the best French accent

You Obscure Coauthor

In this document, $\mathcal{A}$ denotes the set of all French accents.

Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered* "greater than" *accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

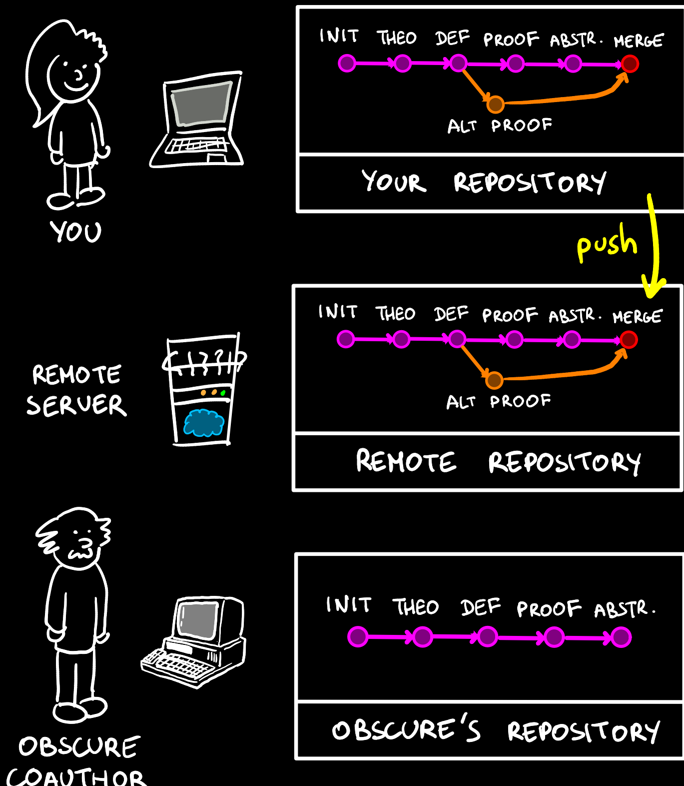
Proof. (shorter proof ??)

Assume, for the sake of contradiction, that the Toulouse accent is **not** the supremum of $\mathcal{A}$. Then, there exists an accent $a \in \mathcal{A}$ such that a is greater than the Toulouse accent.

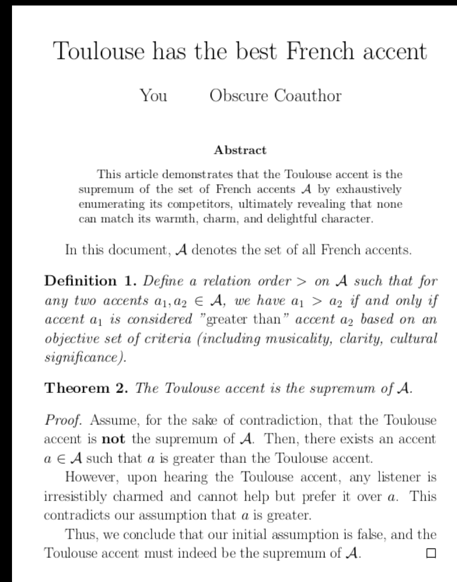
However, upon hearing the Toulouse accent, any listener is irresistibly charmed and cannot help but prefer it over a. This contradicts our assumption that a is greater.

Thus, we conclude that our initial assumption is false, and the Toulouse accent must indeed be the supremum of $\mathcal{A}$. □

WORKING TOGETHER ON AN ARTICLE WITH GIT



your article.pdf



Toulouse has the best French accent

You Obscure Coauthor

Abstract

This article demonstrates that the Toulouse accent is the supremum of the set of French accents $\mathcal{A}$ by exhaustively enumerating its competitors, ultimately revealing that none can match its warmth, charm, and delightful character.

In this document, $\mathcal{A}$ denotes the set of all French accents.

Definition 1. *Define a relation order $>$ on $\mathcal{A}$ such that for any two accents $a_1, a_2 \in \mathcal{A}$, we have $a_1 > a_2$ if and only if accent a_1 is considered "greater than" accent a_2 based on an objective set of criteria (including musicality, clarity, cultural significance).*

Theorem 2. *The Toulouse accent is the supremum of $\mathcal{A}$.*

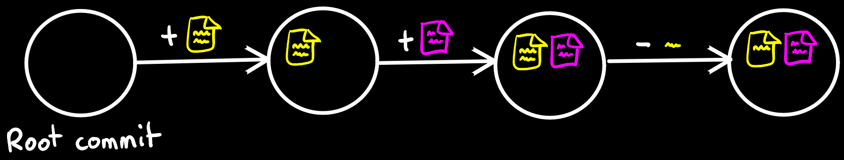
Proof. Assume, for the sake of contradiction, that the Toulouse accent is **not** the supremum of $\mathcal{A}$. Then, there exists an accent $a \in \mathcal{A}$ such that a is greater than the Toulouse accent.

However, upon hearing the Toulouse accent, any listener is irresistibly charmed and cannot help but prefer it over a. This contradicts our assumption that a is greater.

Thus, we conclude that our initial assumption is false, and the Toulouse accent must indeed be the supremum of $\mathcal{A}$. □

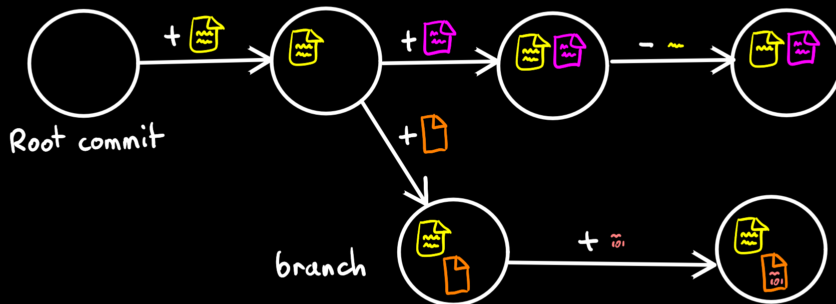
IN SUMMARY...

git is a Version Control System (VCS):
it stores all the project states over time.



IN SUMMARY...

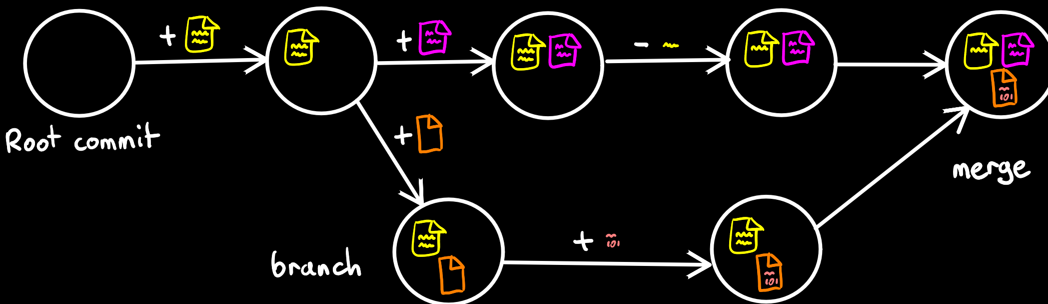
git is a Version Control System (VCS): it stores all the project states over time.



Git lets you create parallel development branches...

IN SUMMARY...

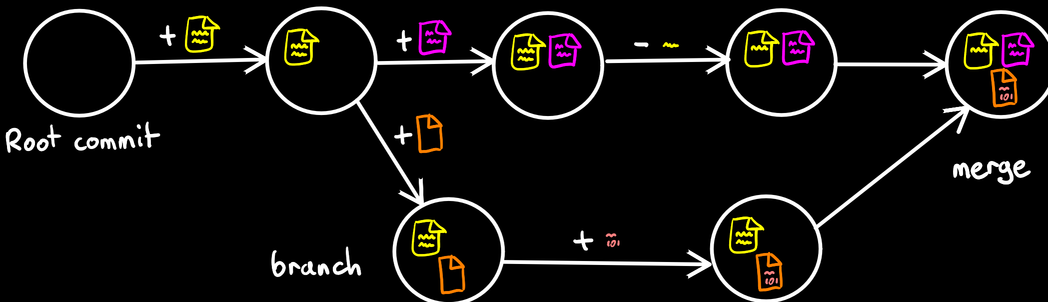
git is a Version Control System (VCS): it stores all the project states over time.



Git lets you create parallel development branches that can be integrated later.

IN SUMMARY...

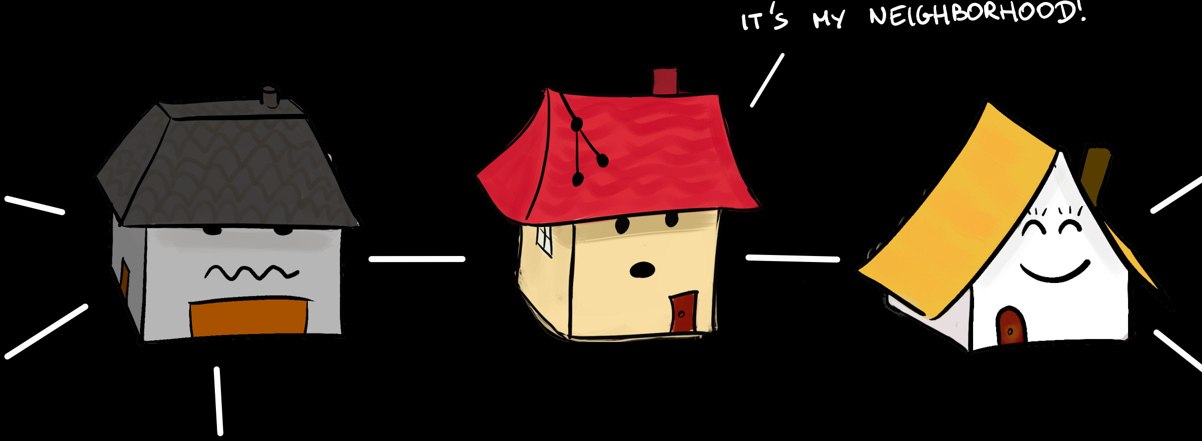
git is a Version Control System (VCS): it stores all the project states over time.



Git lets you create parallel development branches that can be integrated later.

This forms a Directed Acyclic Graph (DAG), where the vertices are the project states, also named <u>commits</u>.

RANDOM GIT GRAPHS



OUR MOTIVATION

Objectives:
- Define a probabilistic model for graphs of commits
- Design random generators for such a distribution

①

Why?

②

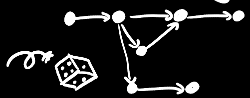
OUR MOTIVATION

Objectives:
- Define a probabilistic model for graphs of commits
- Design random generators for such a distribution

Why?

① VCS unit testing

use random graphs



to check program correctness

```
void test_remove_last_commit() {

 srand(time(NULL));

 for (int i = 0; i < NUM_TESTS; i++) {

  GitGraph graph1 = create_random_git_graph();
  size_t n1 = graph1.num_commits;

  GitGraph graph2 = remove_last_commit(graph1);
  size_t n2 = graph2.num_commits;

  assert(n1 == n2 + 1);

 }
}
```

②

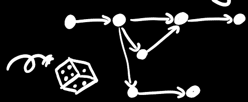
OUR MOTIVATION

Objectives:
- Define a probabilistic model for graphs of commits
- Design random generators for such a distribution

Why?

① VCS unit testing

use random graphs



to check program correctness

```
void test_remove_last_commit() {

 srand(time(NULL));

 for (int i = 0; i < NUM_TESTS; i++) {

  GitGraph graph1 = create_random_git_graph();
  size_t n1 = graph1.num_commits;

  GitGraph graph2 = remove_last_commit(graph1);
  size_t n2 = graph2.num_commits;

  assert(n1 == n2 + 1);

 }
}
```

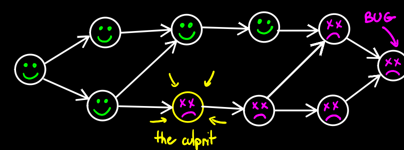
② Average-case complexity of VCS algorithms

+ benchmark

e.g.

→ **Regression Search Problem**

Find the commit that introduces a bug



Analysis of git bisect
[C. Dorbec Lecoq 2023]

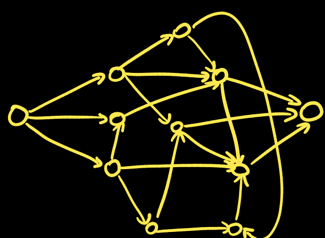
→ **reachability / label synchronisation Problem**
[Bulteau David Horn Tran-Girard]

WHICH GRAPHS TO CONSIDER ?

In git, every DAG

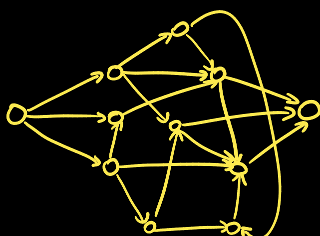
without restriction

can be generated...

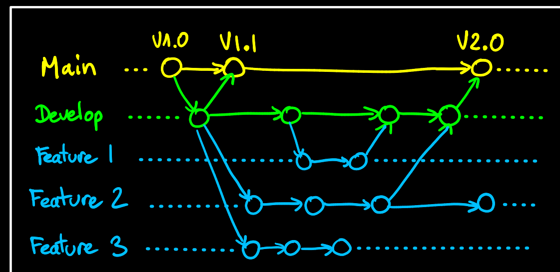


WHICH GRAPHS TO CONSIDER ?

In git, every DAG without restriction can be generated...

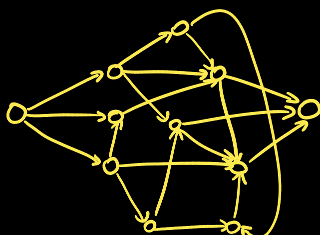


... but many projects follow a workflow

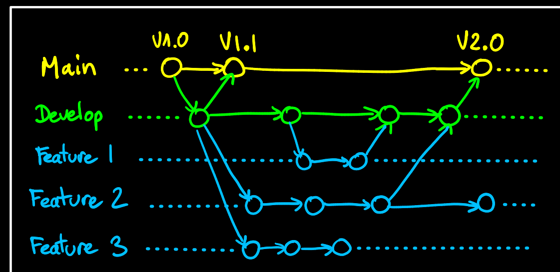


WHICH GRAPHS TO CONSIDER ?

In git, every DAG without restriction can be generated...



... but many projects follow a workflow



In the following, we consider a simple workflow but widely used in industry: the feature branch workflow

GIT GRAPH

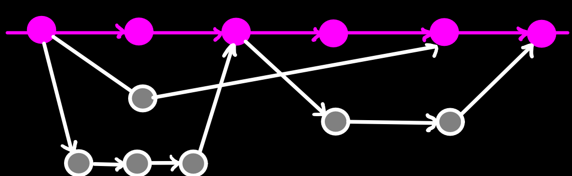
DEFINITION

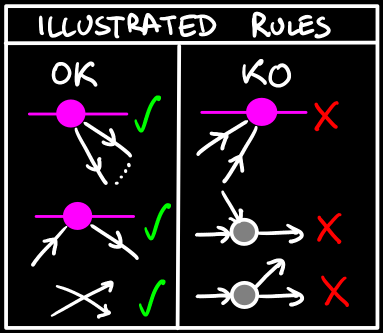
(feature branch) Git graph = DAG with

- a main branch (path of magenta vertices)
- 0, 1 or several feature branches, paths of ≥ 1 white vertices starting and ending on magenta vertices
- indegree ≤ 2 for all vertices

previously defined in [Lecoq 2024]

e.g.





GIT GRAPH

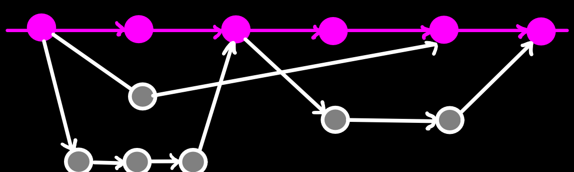
DEFINITION

(feature branch) Git graph = DAG with

- a main branch (path of magenta vertices)
- 0, 1 or several feature$^+$ branches, paths of ≥ 1 white vertices starting and ending on magenta vertices
- indegree ≤ 2 for all vertices

previously defined in [Lecoq 2024]

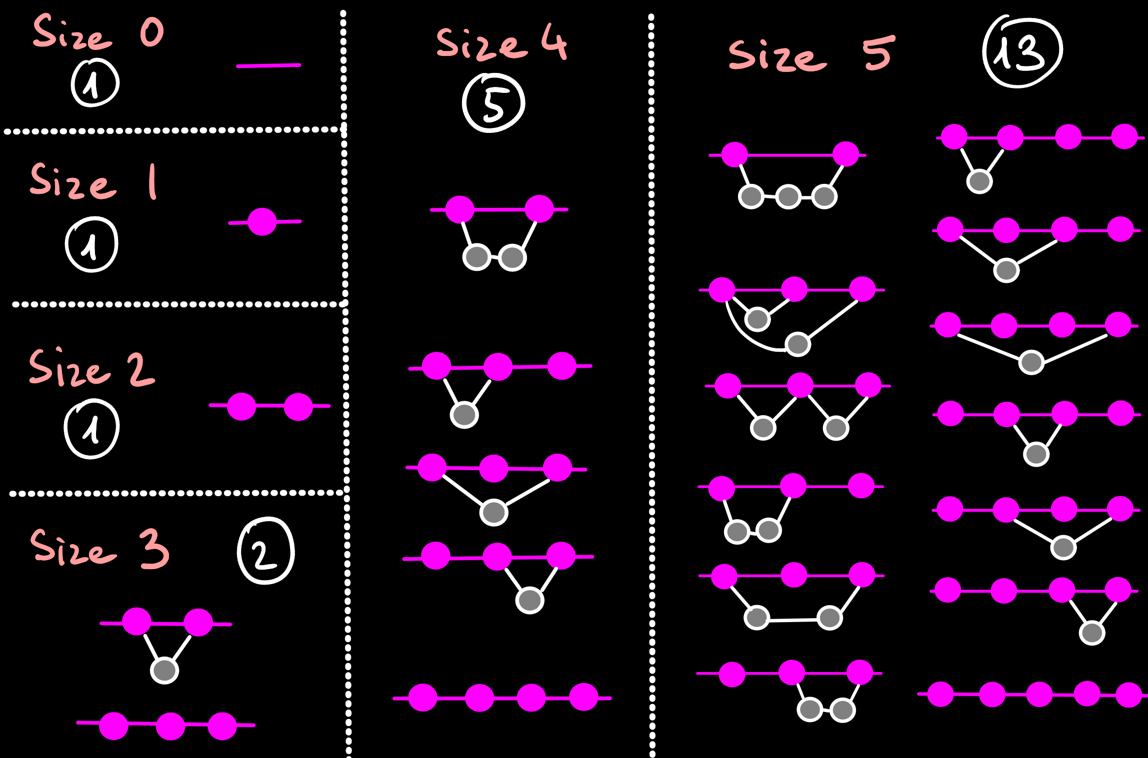
e.g.



GOALS

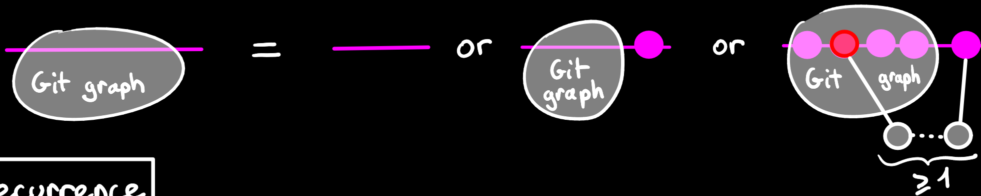
- Counting Git graphs (exactly or asymptotically)
- Random generation of Git graphs

ALL SMALL GIT GRAPHS



RECURSIVE DECOMPOSITION

Decomposition



Recurrence

$$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,\, k-1} \quad \text{for } n \geq 1$$

where $g_{n,k}$:= number of Git graphs with n vertices, k of them being magenta

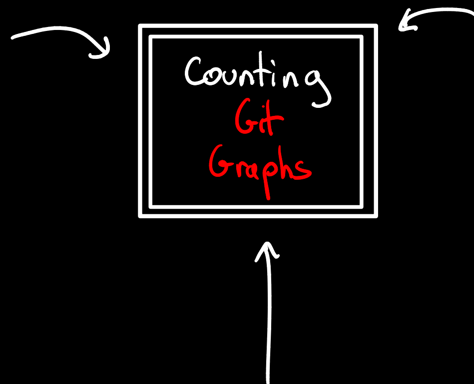
Differential Equation for the Generating Function

$$G(z,u) = 1 + zu\, G(z,u) + \frac{z^2 u^2}{1-z} \frac{\partial G}{\partial u}(z,u)$$

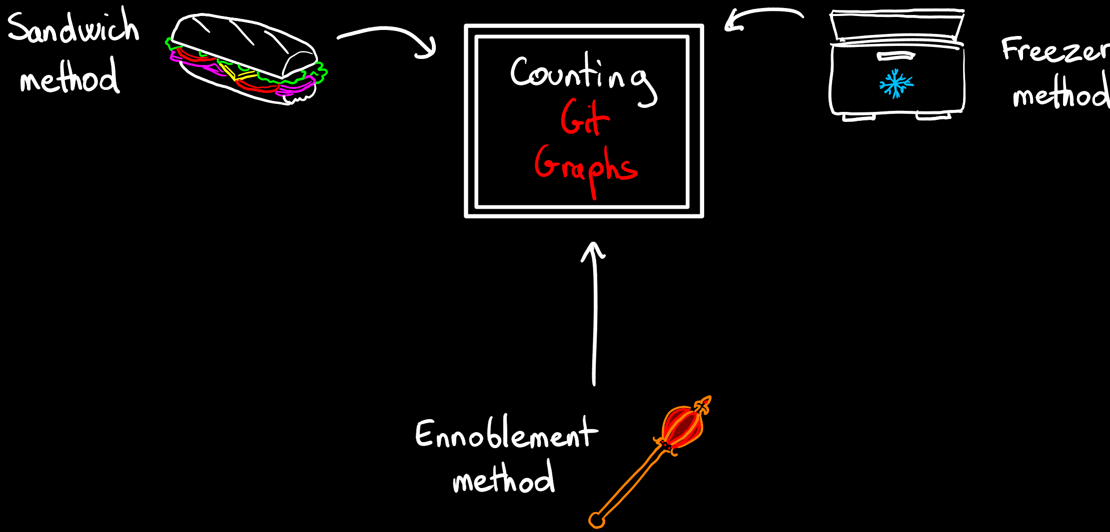
where $G(z,u) = \sum_{n \geq 0} \sum_{k \geq 0} g_{n,k}\, z^n u^k$

⚠ $G(z,u)$ is not analytic.

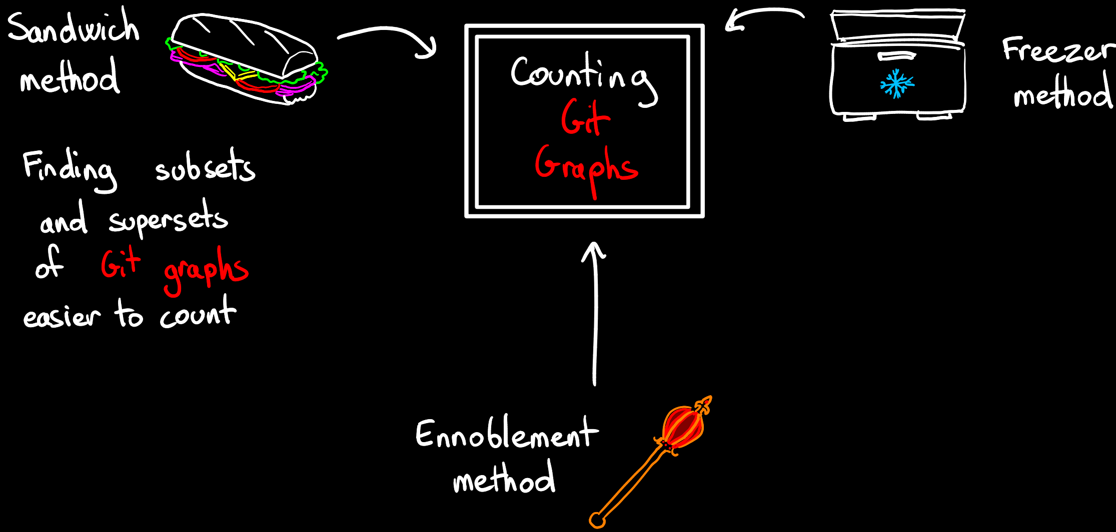
THREE ANGLES OF ATTACK



THREE ANGLES OF ATTACK

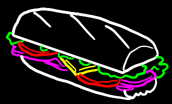


THREE ANGLES OF ATTACK



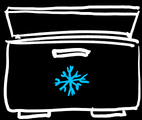
THREE ANGLES OF ATTACK

Sandwich method



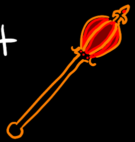
Finding subsets and supersets of Git graphs easier to count

Counting Git Graphs



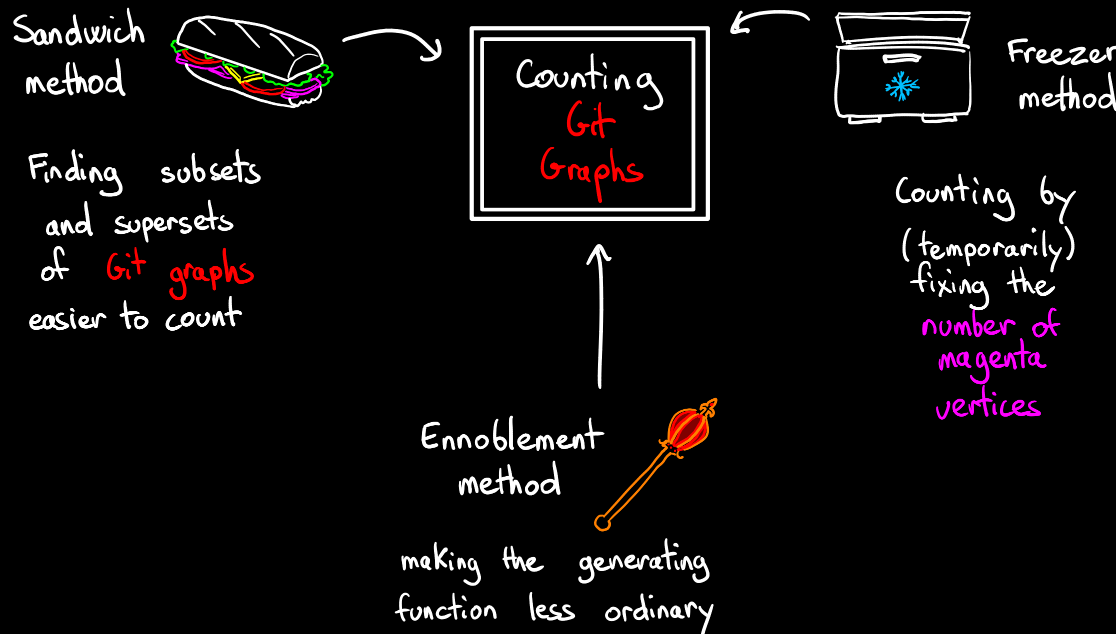
Freezer method

Ennoblement method



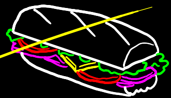
making the generating function less ordinary

THREE ANGLES OF ATTACK



THREE ANGLES OF ATTACK

~~Sandwich method~~



Finding subsets and supersets of Git graphs easier to count

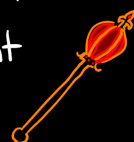
not developed today

Example of result

The random variable nb magenta vertices / n converges in probability to 1/2.

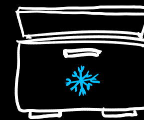
Counting Git Graphs

Ennoblement method



making the generating function less ordinary

Freezer method

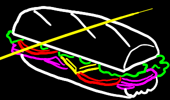


Counting by (temporarily) fixing the number of magenta vertices

THREE ANGLES OF ATTACK

Counting Git Graphs

~~Sandwich method~~



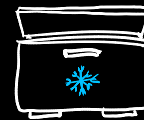
Finding subsets and supersets of Git graphs easier to count

not developed today

Example of result

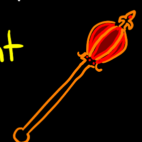
The random variable nb magenta vertices/ n converges in probability to 1/2.

Freezer method



Counting by (temporarily) fixing the number of magenta vertices

Ennoblement method



making the generating function less ordinary

Ennoblement Method

Recurrence	$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,k-1}$
Differential Equation	$G(z,u) = 1 + zu\,G(z,u) + \frac{z^2u^2}{1-z}\frac{\partial G}{\partial u}(z,u)$

Usual trick:

Ordinary Generating Function

$$\underbrace{\sum_{n,k \geq 0} g_{n,k} z^n u^k}_{G(z,u)},$$

not analytic ✗

ENNOBLEMENT METHOD

Recurrence	$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,k-1}$
Differential Equation	$G(z,u) = 1 + zu\, G(z,u) + \frac{z^2 u^2}{1-z} \frac{\partial G}{\partial u}(z,u)$

Usual trick:

Borel transform

Ordinary Generating Function $\underbrace{\sum_{n,k \geq 0} g_{n,k} z^n u^k}_{G(z,u),}$ not analytic ✗

Exponential Generating Function $\underbrace{\sum_{n,k \geq 0} \frac{g_{n,k}}{n!} z^n u^k}$ analytic, but no pretty equation ✗

ENNOBLEMENT METHOD

Recurrence	$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,k-1}$
Differential Equation	$G(z,u) = 1 + zu\,G(z,u) + \frac{z^2u^2}{1-z}\frac{\partial G}{\partial u}(z,u)$

Usual trick:

Ordinary Generating Function $\underbrace{\sum_{n,k\geq 0} g_{n,k} z^n u^k}_{G(z,u),}$ not analytic ✗

Borel transform →

Exponential Generating Function $\underbrace{\sum_{n,k\geq 0} \frac{g_{n,k}}{n!} z^n u^k}$ analytic, but no pretty equation ✗

Borel transform on u ↓

$\tilde{G}(z,u) = \sum_{n,k\geq 0} \frac{g_{n,k}}{k!} z^n u^k$ analytic ✓

and

Differential Equation for $\tilde{G}$
$\frac{\partial \tilde{G}}{\partial u} = z\tilde{G} + \frac{z^2u}{1-z}\frac{\partial \tilde{G}}{\partial u}$

✓

ENNOBLEMENT METHOD

Recurrence	$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,k-1}$
Differential Equation	$G(z,u) = 1 + zu\,G(z,u) + \frac{z^2u^2}{1-z} \frac{\partial G}{\partial u}(z,u)$

$$\tilde{G}(z,u) = \sum_{n,k \geq 0} \frac{g_{n,k}}{k!} z^n u^k$$

analytic ✓

and

Differential Equation for $\tilde{G}$

$$\frac{\partial \tilde{G}}{\partial u} = z\,\tilde{G} + \frac{z^2 u}{1-z} \frac{\partial \tilde{G}}{\partial u}$$

✓

ENNOBLEMENT METHOD

Recurrence	$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,\, k-1}$
Differential Equation	$G(z,u) = 1 + zu\, G(z,u) + \dfrac{z^2 u^2}{1-z} \dfrac{\partial G}{\partial u}(z,u)$

$$\tilde{G}(z,u) = \sum_{n,k \geq 0} \frac{g_{n,k}}{k!} z^n u^k$$

analytic ✓

and

Differential Equation for $\tilde{G}$

$$\frac{\partial \tilde{G}}{\partial u} = z\tilde{G} + \frac{z^2 u}{1-z} \frac{\partial \tilde{G}}{\partial u}$$ ✓

this can be solved!

Theorem

$$\tilde{G}(z,u) = \left(1 - \frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}}$$

ENNOBLEMENT METHOD

Recurrence	$g_{n,k} = g_{n-1,k-1} + \sum_{\ell \geq 1} (k-1)\, g_{n-1-\ell,k-1}$
Differential Equation	$G(z,u) = 1 + z u\, G(z,u) + \frac{z^2 u^2}{1-z} \frac{\partial G}{\partial u}(z,u)$

$$\tilde{G}(z,u) = \sum_{n,k \geq 0} \frac{g_{n,k}}{k!} z^n u^k$$

analytic ✓

and

Differential Equation for $\tilde{G}$

$$\frac{\partial \tilde{G}}{\partial u} = z\, \tilde{G} + \frac{z^2 u}{1-z} \frac{\partial \tilde{G}}{\partial u}$$

✓

this can be solved!

Theorem

$$\tilde{G}(z,u) = \left(1 - \frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}}$$

How can it be exploited?

DIFFERENT PERSPECTIVE

Is there a combinatorial explanation for the formula

$$\tilde{G}(z,u) = \left(1 - \frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}} \; ?$$

DIFFERENT PERSPECTIVE

Is there a combinatorial explanation for the formula

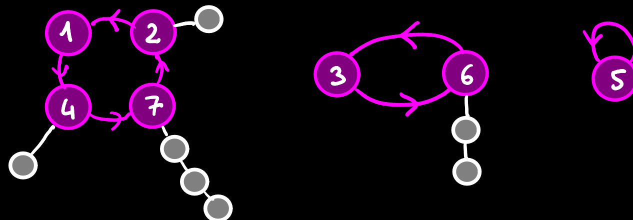
$$\tilde{G}(z,u) = \left(1 - \frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}} = \exp\left(\frac{1}{\frac{z}{1-z}} \ln\left(\frac{1}{1 - uz\frac{z}{1-z}}\right)\right) ?$$

DIFFERENT PERSPECTIVE

Definition

cyclarium = set of cycles of magenta vertices labeled from 1 to k where a chain of white unlabeled vertices is attached to each magenta vertex, except to the ones having the smallest label in their cycles.

e.g.:



Is there a combinatorial explanation for the formula

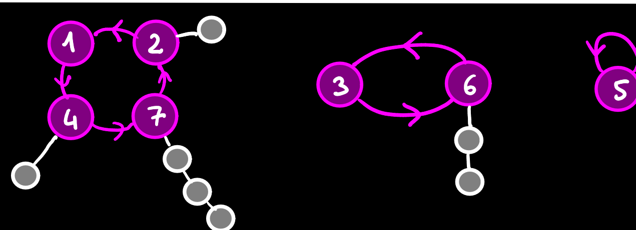
$$\tilde{G}(z,u) = \left(1-\frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}} = \exp\left(\frac{1}{\frac{z}{1-z}} \ln\left(\frac{1}{1-uz\frac{z}{1-z}}\right)\right) \, ?$$

Different Perspective

Definition

cyclarium = set of cycles of magenta vertices labeled from 1 to k where a chain of white unlabeled vertices is attached to each magenta vertex, except to the ones having the smallest label in their cycles.

e.g.:



Is there a combinatorial explanation for the formula

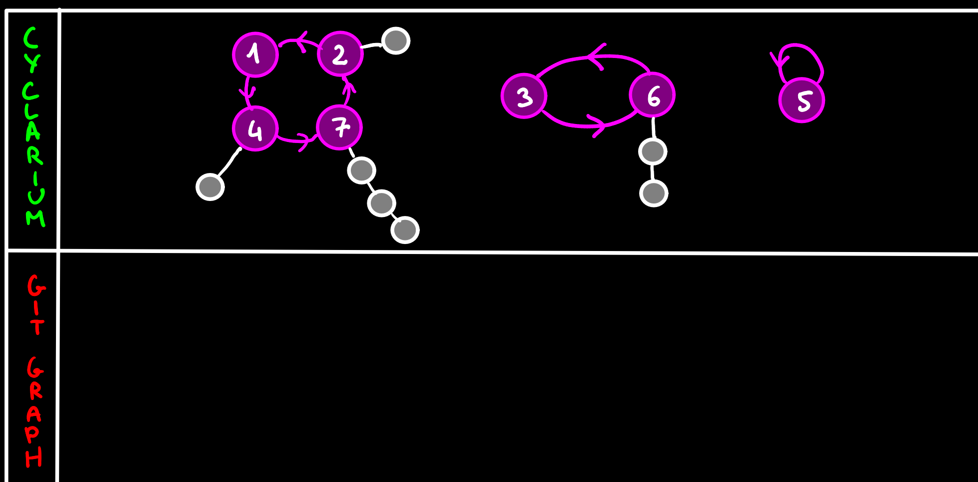
$$\tilde{G}(z,u) = \left(1 - \frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}} = \exp\left(\frac{1}{\frac{z}{1-z}} \ln\left(\frac{1}{1-uz\frac{z}{1-z}}\right)\right) ?$$

It's the generating function of cyclariums!

BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

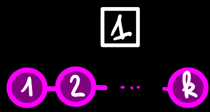


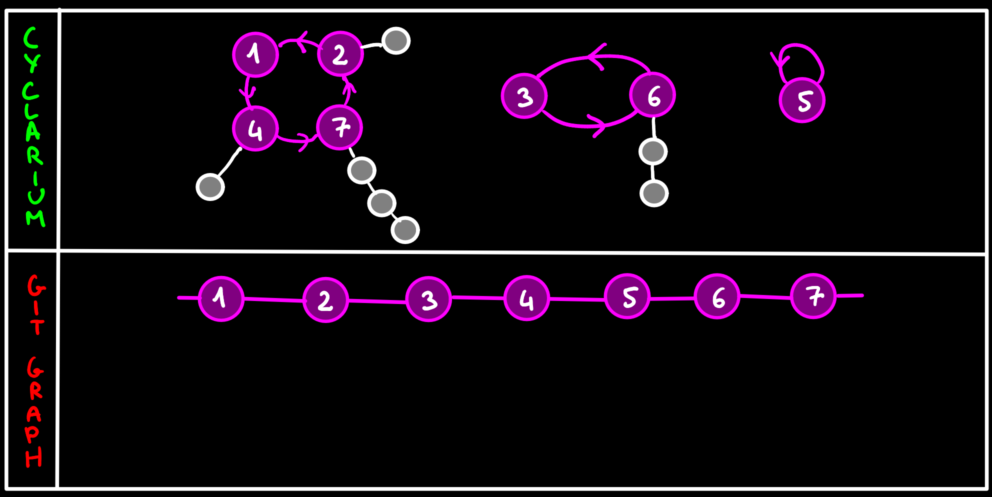
BIJECTION

Proposition

There is a bijection from cycloriums to Git graphs:

STEPS

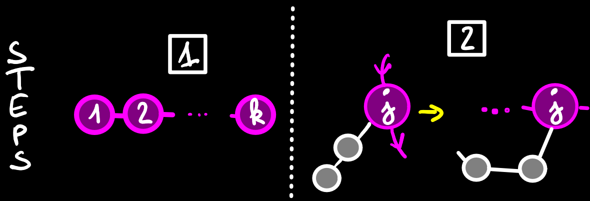


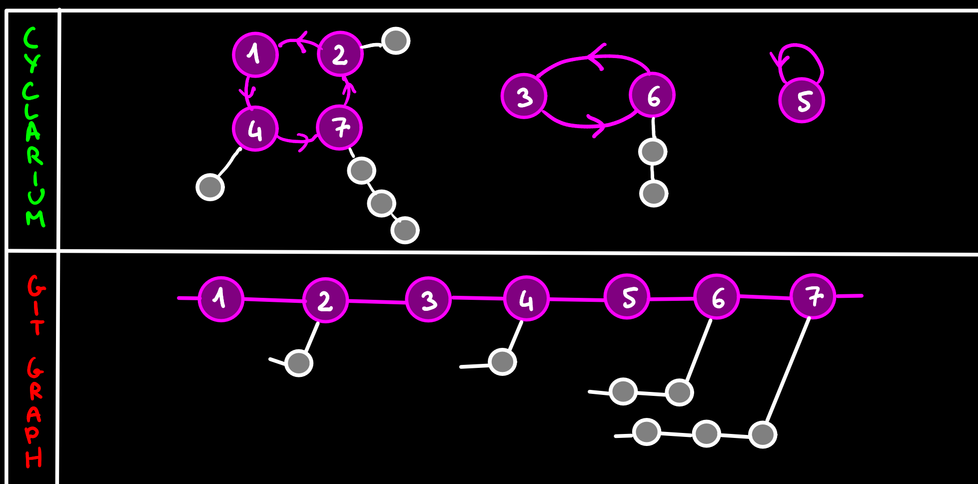


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

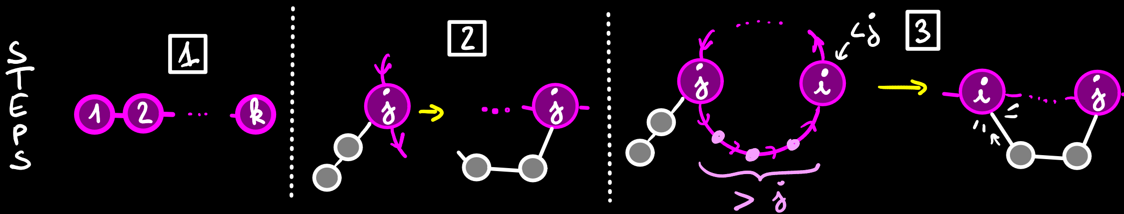


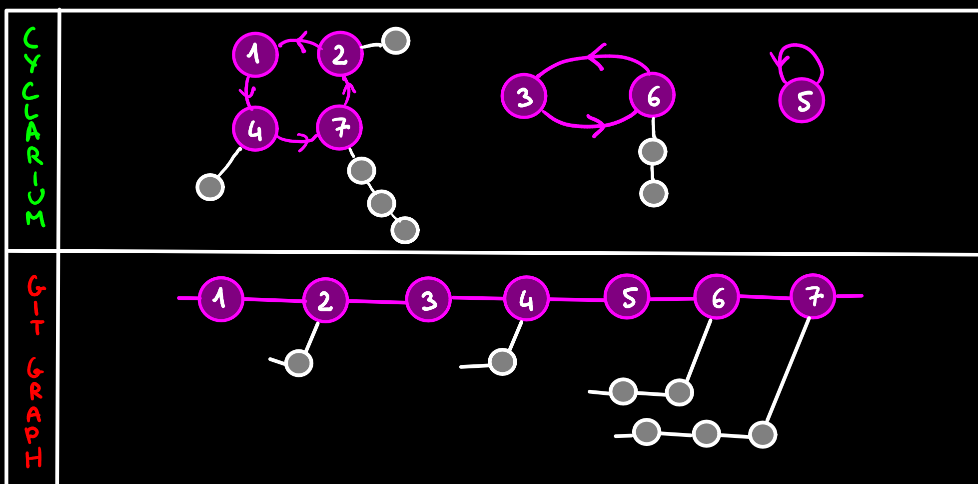


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

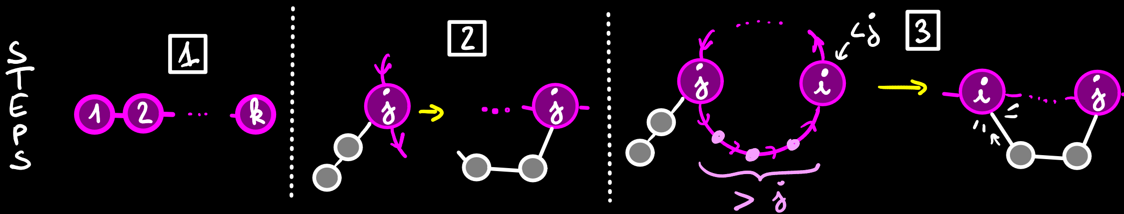


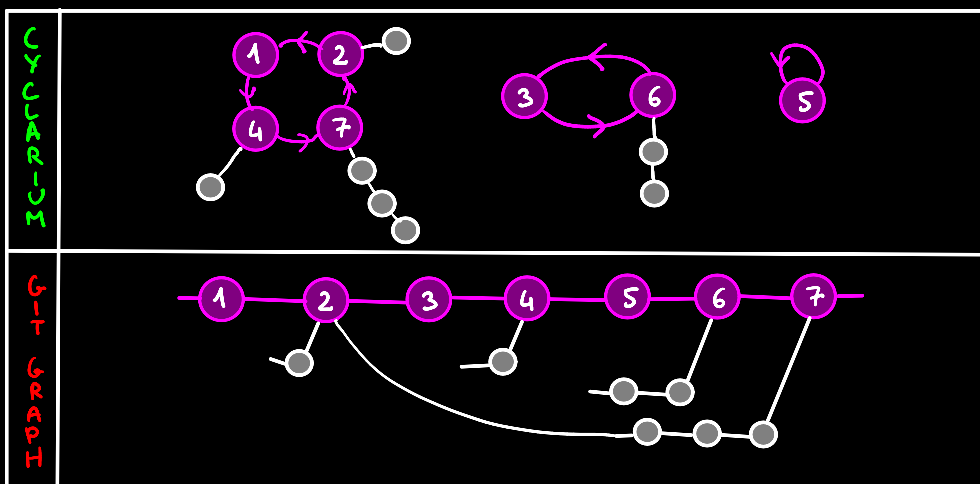


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

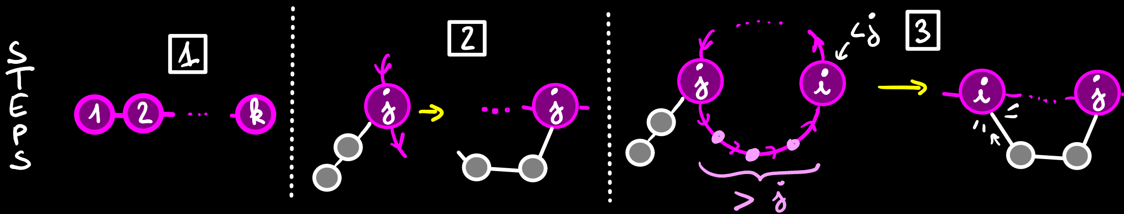


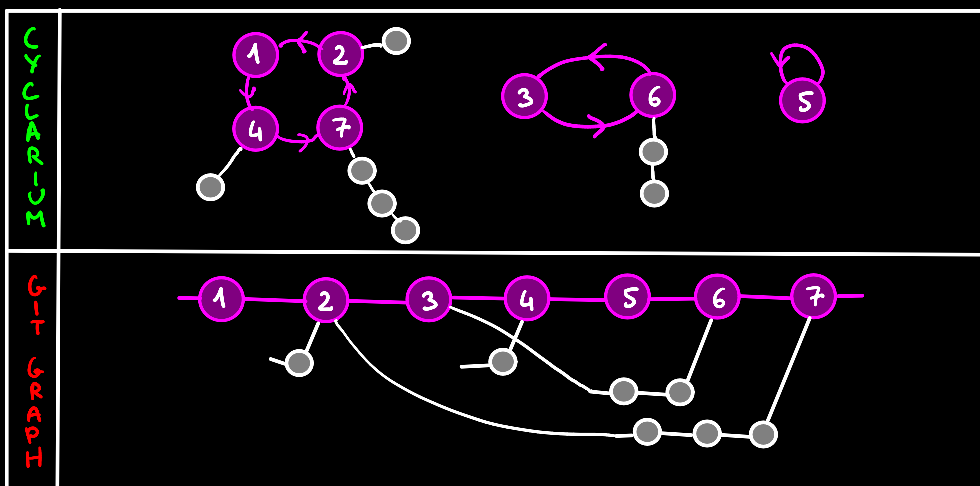


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

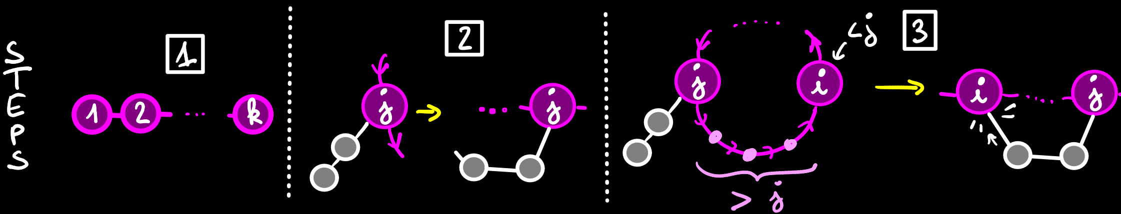


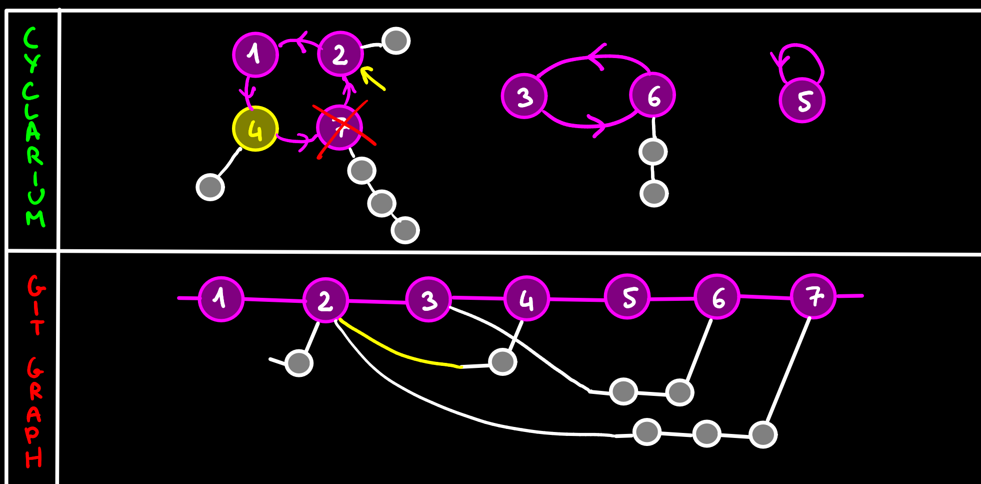


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

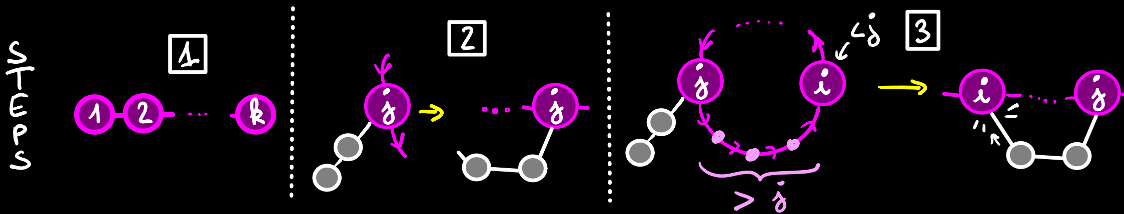


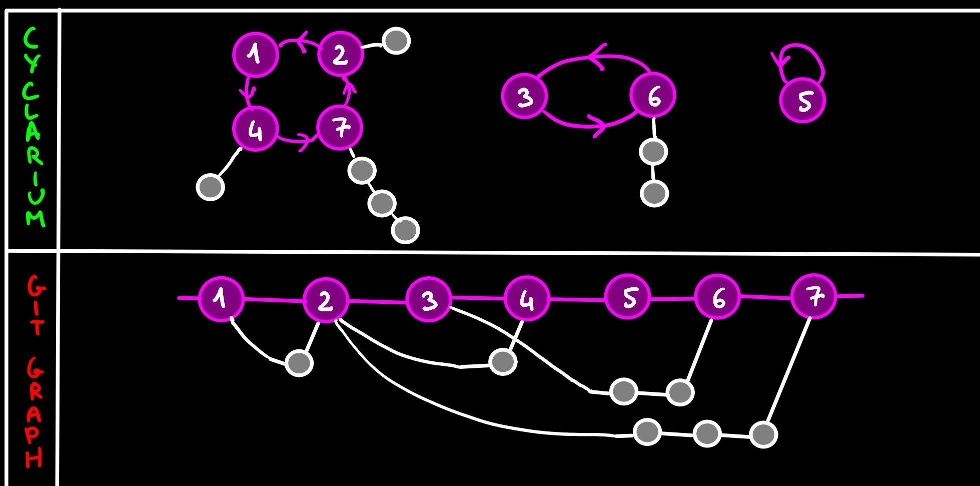


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:

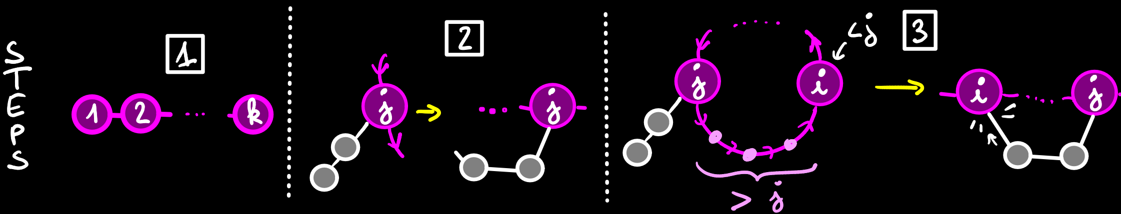


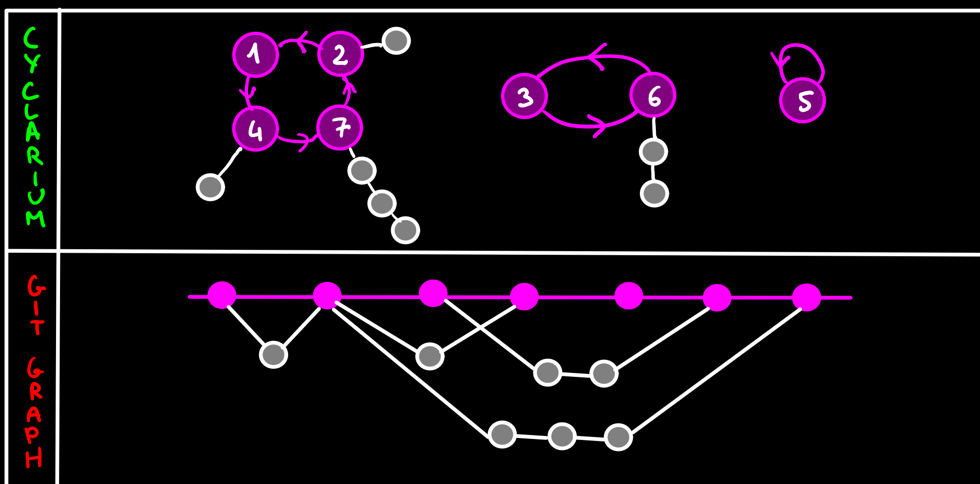


BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs:





BIJECTION

Proposition

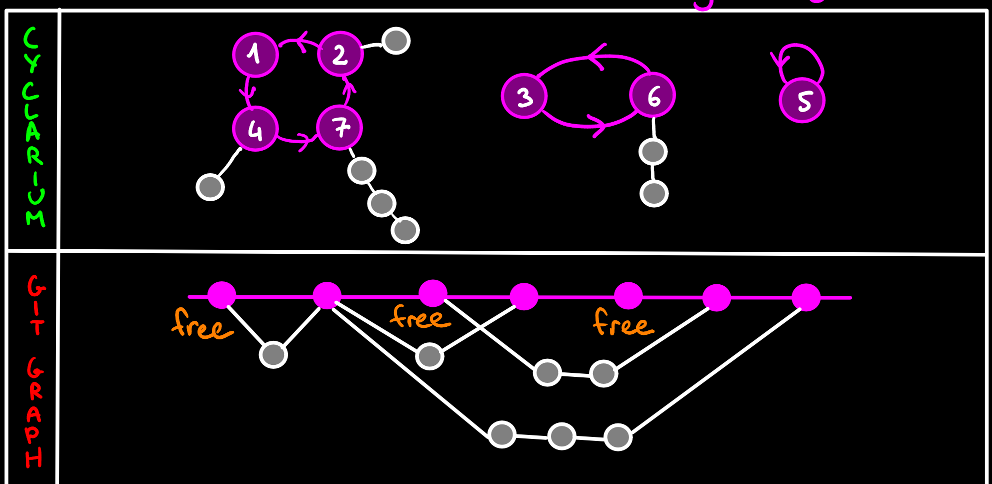
There is a bijection from cyclariums to Git graphs sending

vertices ⟶ vertices

magenta vertices ⟶ magenta vertices

cycles ⟶ free vertices
i.e magenta vertices of indegree ≤ 1

cycle lengths ⟶ magenta vertices in the connected components when the magenta edges are erased.



BIJECTION

Proposition

There is a bijection from cyclariums to Git graphs sending

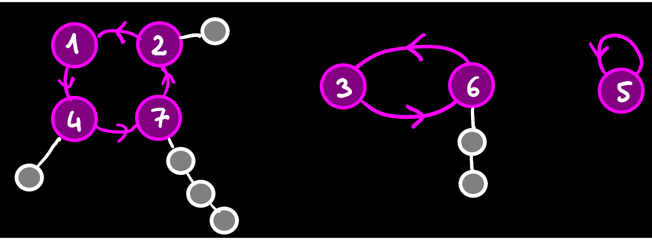
vertices ⟶ vertices

magenta vertices ⟶ magenta vertices

cycles ⟶ free vertices i.e magenta vertices of indegree ≤ 1

cycle lengths ⟶ magenta vertices in the connected components when the magenta edges are erased.

CYCLARIUM



Corollary

$$g_{n,k} = \sum_{f=1}^{k-1} \left[\begin{matrix} k \\ f \end{matrix} \right] \binom{n-k-1}{k-f-1} \qquad (k < n)$$

where $g_{n,k}$ = number of Git graphs counted by vertices & magenta vertices

and $\left[\begin{matrix} \cdot \\ \cdot \end{matrix} \right]$ = (unsigned) Stirling number of 1st kind

RANDOM GENERATION

THE RANDOM GENERATOR WE WANT

Inputs: Size n and number of magenta vertices k

Output: A random Git graph of size $\simeq n$ and $\simeq k$ magenta vertices

(uniformly when conditioned to have size n and k magenta vertices)

RANDOM MODEL

"Boltzmann model" (exponential in u, ordinary in z)

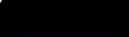
Fix $z > 0$ and $u > 0$.

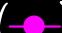
We wish to draw a Git graph γ with a weight proportional to

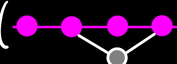
$$z^{\#\text{vertices in } \gamma} \; \frac{u^{\#\text{magenta vertices in } \gamma}}{(\#\text{magenta vertices in } \gamma)!}$$

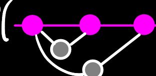
(Size is not fixed)

Examples

$\mathbb{P}($  $) \propto 1$

$\mathbb{P}($  $) \propto z\,u$

$\mathbb{P}($  $) \propto z^5 \frac{u^4}{24}$

$\mathbb{P}($  $) \propto z^5 \frac{u^3}{6}$

RANDOM MODEL

"Boltzmann model" (exponential in u, ordinary in z)

Fix $z > 0$* and $u > 0$.*

We wish to draw a Git graph γ with a weight ~~proportional~~ equal to

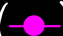
$$\frac{z^{\#\text{vertices in }\gamma}}{\tilde{G}(z,u)} \frac{u^{\#\text{magenta vertices in }\gamma}}{(\#\text{magenta vertices in }\gamma)!}$$

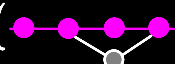
(Size is not fixed)

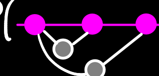
where $\tilde{G}(z,u) = \sum_{n,k \geq 0} \frac{g_{n,k}}{k!} z^n u^k = \left(1 - \frac{z^2 u}{1-z}\right)^{-\frac{1-z}{z}}.$

Examples

$\mathbb{P}($  $) = \frac{1}{\tilde{G}(z,u)}$

$\mathbb{P}($  $) = \frac{z\,u}{\tilde{G}(z,u)}$

$\mathbb{P}($  $) = \frac{z^5}{\tilde{G}(z,u)} \frac{u^4}{24}$

$\mathbb{P}($  $) = \frac{z^5}{\tilde{G}(z,u)} \frac{u^3}{6}$

*: in the disk of convergence of $\tilde{G}$

HOW TO SAMPLE GIT GRAPHS

Inputs : Size n and number of magenta vertices k

Step 1 Solve numerically z and u such that :

$$\mathbb{E}_{\text{boltz}(z,u)}(\text{size}) = n \qquad \mathbb{E}_{\text{boltz}(z,u)}(\#\text{mag}) = k$$

Step 2 Use a Boltzmann sampler for cyclariums for parameters z and u

Step 3 Bijection to Git graphs

HOW TO SAMPLE GIT GRAPHS

Inputs: Size n and number of magenta vertices k

Step 1 Solve numerically z and u such that:

$$\mathbb{E}_{\text{boltz}(z,u)}(\text{size}) = n \qquad \mathbb{E}_{\text{boltz}(z,u)}(\#\text{mag}) = k$$

$$\mathbb{E}_{\text{boltz}(z,u)}(\text{size}) = \underbrace{z \frac{\partial}{\partial z} \ln \tilde{G}(z,u)}_{\text{computable}} \qquad \mathbb{E}_{\text{boltz}(z,u)}(\#\text{mag}) = \underbrace{u \frac{\partial}{\partial u} \ln \tilde{G}(z,u)}_{\text{computable}}$$

Step 2 Use a Boltzmann sampler for cyclariums for parameters z and u

Step 3 Bijection to Git graphs

HOW TO SAMPLE GIT GRAPHS

Inputs : Size n and number of magenta vertices k

Step 1 Solve numerically z and u such that:

$$\mathbb{E}_{\mathrm{boltz}(z,u)}(\mathrm{size}) = n = \underbrace{z \frac{\partial}{\partial z} \ln \tilde{G}(z,u)}_{\text{computable}}$$

$$\mathbb{E}_{\mathrm{boltz}(z,u)}(\#\mathrm{mag}) = k = \underbrace{u \frac{\partial}{\partial u} \ln \tilde{G}(z,u)}_{\text{computable}}$$

Step 2 Use a Boltzmann sampler for cyclariums for parameters z and u

Step 3 Bijection to Git graphs

Fun fact

Let $\alpha = k/n$

if $\alpha < \frac{1}{2}$, $z \to \mathrm{fct}(\alpha)$ $u \to \mathrm{fct}(\alpha)$

if $\alpha > 1/2$, $z \sim \frac{\mathrm{fct}(\alpha)}{n}$, $u \sim \mathrm{fct}(\alpha)\, n^2$

ASYMPTOTIC ESTIMATE

ASYMPTOTIC ESTIMATE

Theorem The number of Git graphs with n vertices is asymptotically equivalent to

$$g_n \sim \frac{e^{1/8}}{2}\left(\frac{n}{2e}\right)^{\frac{n}{2}} \exp\left(\frac{1}{2}\ln\left(\frac{n}{2}\right)\sqrt{\frac{n}{2}} + \sqrt{2n} + \frac{\left(\ln\frac{n}{2}\right)^2}{32}\right)\left(\frac{n}{2}\right)^{-\frac{3}{8}}$$

Disclaimer: not totally proved

ASYMPTOTIC ESTIMATE

Theorem The number of Git graphs with n vertices is asymptotically equivalent to

$$g_n \sim \frac{e^{1/8}}{2} \left(\frac{n}{2e}\right)^{\frac{n}{2}} \exp\left(\frac{1}{2}\ln\left(\frac{n}{2}\right)\sqrt{\frac{n}{2}} + \sqrt{2n} + \frac{\left(\ln\frac{n}{2}\right)^2}{32}\right)\left(\frac{n}{2}\right)^{-\frac{3}{8}}$$

Disclaimer: not totally proved

How to get such a result?

ASYMPTOTIC ESTIMATE

Theorem The number of Git graphs with n vertices is asymptotically equivalent to

$$g_n \sim \frac{e^{1/8}}{2} \left(\frac{n}{2e}\right)^{\frac{n}{2}} \exp\left(\frac{1}{2}\ln\left(\frac{n}{2}\right)\sqrt{\frac{n}{2}} + \sqrt{2n} + \frac{\left(\ln\frac{n}{2}\right)^2}{32}\right)\left(\frac{n}{2}\right)^{-\frac{3}{8}}$$

Disclaimer: not totally proved

How to get such a result?

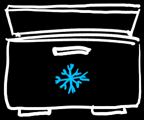
Freezer method !

IT'S EASIER WHEN YOU FIX k

$$G_k(z) := \sum_{n \geqslant 0} g_{n,k} z^n$$

= Generating Function of Git graphs where the number k of magenta vertices is fixed

Freeze k



IT'S EASIER WHEN YOU FIX k

$$G_k(z) := \sum_{n \geq 0} g_{n,k} \, z^n$$

= Generating Function of Git graphs where the number k of magenta vertices is fixed

Freeze k

Claim

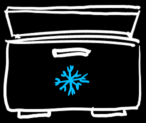
$$G_k(z) = z^k \prod_{j=0}^{k-1} \left(1 + \frac{j z}{1-z}\right)$$

IT'S EASIER WHEN YOU FIX k

$$G_k(z) := \sum_{n \geq 0} g_{n,k} \, z^n$$

= Generating Function of Git graphs where the number k of magenta vertices is fixed

Freeze k



Claim

$$G_k(z) = z^k \prod_{j=0}^{k-1} \left(1 + \frac{j z}{1-z}\right)$$

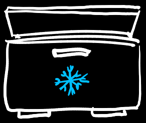
$$= \frac{z^{2k}}{(1-z)^k} \frac{\Gamma\left(k + (1-z)/z\right)}{\Gamma\left((1-z)/z\right)}$$

IT'S EASIER WHEN YOU FIX k

$$G_k(z) := \sum_{n \geq 0} g_{n,k}\, z^n$$

= Generating Function of Git graphs where the number k of magenta vertices is fixed

Freeze k



Claim

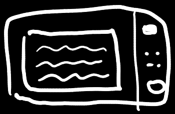
$$G_k(z) = z^k \prod_{j=0}^{k-1} \left(1 + \frac{j z}{1-z}\right)$$

$$= \frac{z^{2k}}{(1-z)^k} \frac{\Gamma\left(k + (1-z)/z\right)}{\Gamma\left((1-z)/z\right)}$$

Cauchy integral formula

$$g_{n,k} = \frac{1}{2\pi i} \oint \frac{z^{2k-n-1}}{(1-z)^k} \frac{\Gamma\left(k + (1-z)/z\right)}{\Gamma\left((1-z)/z\right)}\, dz$$

Unfreeze k

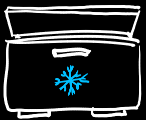


IT'S EASIER WHEN YOU FIX k

$$G_k(z) := \sum_{n \geq 0} g_{n,k} z^n$$

= Generating Function of Git graphs where the number k of magenta vertices is fixed

Freeze k



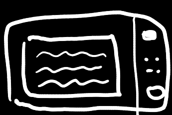
Claim

$$G_k(z) = z^k \prod_{j=0}^{k-1} \left(1 + \frac{j z}{1-z}\right)$$

$$= \frac{z^{2k}}{(1-z)^k} \frac{\Gamma(k + (1-z)/z)}{\Gamma((1-z)/z)}$$

Cauchy integral formula

Unfreeze k



$$g_{n,k} = \frac{1}{2\pi i} \oint \frac{z^{2k-n-1}}{(1-z)^k} \frac{\Gamma(k + (1-z)/z)}{\Gamma((1-z)/z)} dz$$

→ invitation to saddle-point ...

SADDLE POINT ANALYSIS

$$g_{n,k} = \frac{1}{2\pi i} \oint \frac{z^{2k-n-1}}{(1-z)^{k}} \frac{\Gamma(k + z/(1-z))}{\Gamma(z/(1-z))} dz$$

We use saddle-point method on g_{n,k_n} where $k_n = \frac{n}{2} + \alpha_n \sqrt{\frac{n}{2}}$

and α_n sub-polynomial

(an educated guess about the location of the maximum)

$\downarrow$

horrible formula for g_{n,k_n}

SADDLE POINT ANALYSIS

$$g_{n,k} = \frac{1}{2\pi i} \oint \frac{z^{2k-n-1}}{(1-z)^k} \frac{\Gamma\left(k + z/(1-z)\right)}{\Gamma\left(z/(1-z)\right)} dz$$

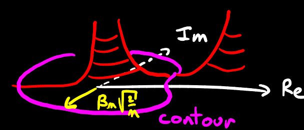
We use saddle-point method on g_{n,k_n} where $k_n = \frac{n}{2} + \alpha_n \sqrt{\frac{n}{2}}$

and α_n sub-polynomial

(an educated guess about the location of the maximum)

$$g_{n,k_n} \sim \left(\frac{n}{2e}\right)^{\frac{n}{2}} \exp\left(\lambda_n(\beta_n)\sqrt{\frac{n}{2}} + \mu_n(\beta_n)\right) \left(\frac{2}{n}\right)^{\frac{1}{4}} \sqrt{\frac{\beta_n}{\pi \ln\left(\frac{n}{2}\right)}}$$

where $\beta_n \sqrt{\frac{2}{n}}$ is the saddle-point



$$\lambda_n(x) = \frac{1}{2x}(1+x)\ln\left(\frac{n}{2}\right) + \frac{1}{x} + x + \frac{\ln(x)}{x} - x\ln(x) + \frac{\ln(2)}{x^2}$$

$$\mu_n(x) = \frac{\ln\left(\frac{n}{2}\right)^2}{32x^2} + \left(\frac{1}{4x^2} - \frac{5}{8} + \frac{3}{8}\frac{\ln(x)}{x^2} - \frac{1}{4}\ln(x)\right)\ln\left(\frac{n}{2}\right) + \frac{1}{2x^2} - \frac{1}{2} + \frac{x}{8} + \frac{5}{8}\frac{\ln(x)^2}{x^2} - \frac{\ln(x)^2 x}{2} + \frac{3}{2}\frac{\ln(x)^2}{x^2} - \frac{1}{2}\ln(x)^2 - \frac{1}{4}\ln(x)$$

ASYMPTOTIC ESTIMATE

Theorem The number of Git graphs with n vertices is asymptotically equivalent to

$$g_n \sim \frac{e^{1/8}}{2} \left(\frac{n}{2e}\right)^{\frac{n}{2}} \exp\left(\frac{1}{2}\ln\left(\frac{n}{2}\right)\sqrt{\frac{n}{2}} + \sqrt{2n} + \frac{\left(\ln \frac{n}{2}\right)^2}{32}\right) \left(\frac{n}{2}\right)^{-\frac{3}{8}}$$

Disclaimer: not totally proved

SOME LIMIT LAWS WE SHOULD OBTAIN UNDER UNIFORM DISTRIBUTION

→ Number of magenta vertices converges to a Gaussian distribution of mean $\sim \frac{n}{2} + \frac{1}{4}\ln\left(\frac{n}{2}\right)\sqrt{\frac{n}{2}} - \sqrt{2n}$ (local limit)

Variance to be determined...

→ Number of free vertices converges to a Gaussian distribution of mean $\sim \frac{1}{2}\sqrt{\frac{n}{2}}\ln\left(\frac{n}{2}\right)$ (distribution)

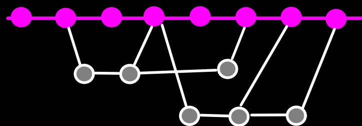
and variance $\sim \frac{1}{2}\sqrt{\frac{n}{2}}\ln\left(\frac{n}{2}\right)$

PERSPECTIVES ABOUT RANDOM GIT GRAPHS

→ Other random models

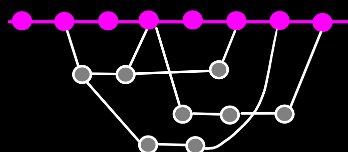
- Other graph models ▷ Collaboration with Clement + Maréchal + Pépin

Phoenix graphs



merged branches can be reborn

Fork Anywhere graphs



branches can be born anywhere but must be merged into main

More involved workflows

- more probabilistic models [Nicaud Peyraud-Magnin Rotondo]

Thank you!