

## Editorial

La lettre de ce mois de la rentrée met en lumière les personnes, acteurs et actrices de la recherche au GREYC, dont les idées et les tendances façonneront les mois à venir.

Le focus personne est consacré à de l'**Équipe IMAGE**.

Le focus scientifique est consacré à une thématique de l'**Équipe AMACC**.

La liste des séminaires des différentes équipes est disponible sur le site du GREYC : [séminaires](#).

## Actualités des actions du GT-Parité du GREYC - <https://www.greyc.fr/laboratoire/parite/>



**Vendredi 10 novembre 2023:** Nous avons invité Mathilde Saliou, journaliste et autrice, spécialisée dans le numérique, à présenter une conférence intitulée « Comment le numérique aggrave les inégalités ». Cette conférence a réuni plus de 30 personnes parmi les membres du GREYC, du LMNO et des commissions parité-égalité de l'UNICAEN et de l'ENSICAEN.



**Vendredi 13 octobre 2023 :** Comme chaque année depuis 2021, des membres du GT-parité ont présenté des conférences sur le thème « Osez les Sciences du Numérique » lors de la fête de la science devant des classes de collèges et lycées.

**Mercredi 11 octobre 2023:** B. Vallée a réalisé une intervention sur le sujet : « Non mixité dans le numérique et stéréotypes de genre » dans le cadre de conférences du M1 Info de l'UNICAEN.

## Sécurité informatique

Nouvelle campagne de phishing par QR Code. Les experts en cybersécurité ont repéré une campagne d'hameçonnage intégrant un QR code. Pour en savoir plus, [consultez ici](#) :

Véronique ROBERT Administrateur Systèmes et Réseaux / **CSSI**

## Nouveaux arrivants

**Geoffrey Laforest** a rejoint l'équipe MAD en novembre pour commencer une thèse, financée par le projet ANR Epi-RL. Elle s'intitule « Representation of Epistemic States for Epistemic Reinforcement Learning » et est encadrée par **Bruno Zanuttini** et **Alexandre Niveau**, ainsi que par **Olivier Buffet** (LORIA).

**Équipe MAD**

## Diffusion logicielle



Nous avons appris que le logiciel libre **G'MIC**, développé dans l'équipe IMAGE, est utilisé par l'ESA (Agence Spatiale Européenne) en partie pour le traitement des images provenant du JWST (James Webb Space Telescope), qui contiennent souvent un bruit fréquentiel qui se manifeste sous forme de bandes latérales qui vient dégrader les images. Le logiciel a par exemple été utilisé pour débruiter l'image de la proto-étoile HH211, illustrée en couverture du magazine Nature, du 5 octobre 2023 ([Vol. 622, Issue 7981](#)) :

**Équipe IMAGE**

**Tanguy Gernot** est recruté comme Ingénieur de Recherche CNRS dans l'équipe SAFE. Il va travailler en biométrie (conception et attaque de protection de données biométriques) et en forensique (analyse de traces numériques, plateforme d'investigation numérique G'DIP).

**Équipe SAFE**

Nous avons accueilli deux nouveaux doctorants en octobre 2023:

- **Tagi TAGIYEV**, contrat PEPR Spintronique sur le sujet: « Optimisation de la sensibilité des capteurs magnétorésistifs anisotropes réalisés dans des couches minces de manganite à l'aide de concentrateurs de flux magnétiques » sous la co-direction de **S. Flament** et **L. Méchin**
- **Luis Saluden**, bourse de thèse de l'ENS Saclay sur le sujet: « Contribution à l'hybridation de sources de champ électrique-magnétique et à leur contrôle. Mise en œuvre et applications. » sous la direction de **M. Denoual**.

**Équipe ELEC**

## Départs

L'**équipe CODAG** souhaite une bonne continuation à **Pierre Beust** à l'occasion de son recrutement sur un poste de professeur à l'Université de Rennes, depuis le 1<sup>er</sup> novembre. Pierre reste malgré tout parmi nous en tant que chercheur associé au GREYC.

**Equipe CODAG.**

## Projets

Le PEPR Cybersécurité "sécurité multimédia" a été accepté. Le projet se nomme **COMPROMIS** (COntenus MultiMédia PROtégés par Machines IntelligentES). Le projet COMPROMIS se base sur une vision moderne de la protection des données multimédia avec en son centre l'apprentissage profond. Ce projet défend l'idée que la protection des données multimédia doit nécessairement être associée à la sécurité des outils qui analysent ces données, c'est-à-dire de nos jours l'Intelligence Artificielle. Le constat est simple : La protection des données multimédia est assurément le domaine de la cybersécurité qui a le plus profité de l'IA., mais il a négligé la vérification du niveau de sécurité de ce nouvel outil. L'IA est devenue l'un des maillons faibles de la protection des données multimédia. Pourtant, la communauté protection des données multimédia est la plus à même pour étudier la sécurité de l'apprentissage profond en calquant ses valeurs cardinales (l'intégrité, la confidentialité et l'identification) aux données de l'apprentissage profond (données d'entraînement, données de test, modèle appris). Cette mise en parallèle de la protection des données multimédia et de la sécurité de l'apprentissage profond crée une fertilisation croisée vertueuse assurant la sécurité de bout en bout, des données jusqu'aux algorithmes qui les traitent. Les verrous scientifiques concernent ainsi tant les applications classiques de la protection des données multimédia (analyse forensique, tatouage, biométrie, deepfakes, stéganographie, chiffrement sélectif) que le domaine émergent de l'apprentissage profond. Ces problèmes scientifiques répondent directement aux préoccupations des agences de sécurité, des industriels, et des producteurs de contenus multimédia.

Le projet sur 5 ans comporte 20 thèses, 96 h.m postdoc et 12 h.m ingénieurs, 12 journées nationales, 1 école d'été, et 1 plateforme. Il rassemble 7 laboratoires (IRISA, Cristal, LIRMM, GREYC, LAMSADE, EURECOM, CEA).

**Équipe SAFE**

L'International Research Project (IRP) du CNRS AI & CYBER (Artificial Intelligence & Cybersecurity) avec la Norvège est créé. L'objectif principal du projet AI & Cyber est de mener des activités de recherche conjointes impliquant l'intelligence artificielle (IA) pour la cybersécurité. L'IA a un impact important sur de nombreux domaines de la science numérique et même plus généralement dans toutes les disciplines. Les algorithmes d'apprentissage peuvent aider à définir des systèmes de cybersécurité (authentification biométrique, détection d'une attaque de phishing...), d'analyser des traces numériques (forensique, supervision de système...) et de créer de nouvelles attaques (attaque de présentation en biométrie, déni de service distribué...). Nous considérons dans ce projet différents axes de recherche se concentrant sur des tâches spécifiques de cybersécurité. Ces sujets seront abordés dans le cadre de thèses de master et de doctorat conjointes (co-encadrées entre les deux universités), des projets collaboratifs (Horizon Europe), de l'organisation de conférences.... Des ateliers scientifiques (en ligne et présentiel impliquant des étudiants de Master) seront régulièrement organisés.

Le projet sur 5 ans implique le GREYC, l'Université d'Oslo (UiO), SINTEF, l'Université norvégienne des sciences et technologies (NTNU), le Norwegian Computing Center (NCR).

**Équipe SAFE et Équipe IMAGE**

## Thèses

**Mina Rafla** a soutenu sa thèse le 6 novembre, intitulée « A Bayesian Approach for Uplift Modeling: Application on Biased Data » et dirigée par **Bruno Crémilleux**.

**Équipe CODAG.**

**Simone Burin-Chu** a soutenu sa thèse le 12 décembre, intitulée « Conception, développement et évaluation d'un dispositif de réalité virtuelle pour la rééducation d'enfants et d'adolescents atteints de paralysie cérébrale : le projet RV-REEDUC » et dirigée par **Nicolas Benguigui**.

**Équipe CODAG.**

**Yris Brice Wandji Piugie** a soutenu sa thèse le 28/11 dont le titre est "Performance and Security Evaluation of Behavioral Biometric Systems" (encadré par **Christophe Rosenberger**, **Christophe Charrier** et **Joël Di Manno** de FIME).

**Équipe SAFE**

**Abdrahamane Wone** a soutenu sa thèse le 29/11 dont le titre est "Contribution to the certification of fingerprint systems: towards the reproducibility of the evaluation" (dirigé par **Christophe Rosenberger**, **Christophe Charrier** et **Joël Di Manno** de FIME).

**Équipe SAFE**

**Raphaël Strenaer** a soutenu sa thèse le 14 décembre 2023, intitulée « Caractérisations physique et thermique de transistors HEMTs de la filière GaN polarisés en régimes continu et pulsé » et dirigée par **Yannick Guhel** et **Bertrand Boudart** au sein de l'équipe cherbourgeoise du GREYC.

**Équipe ELEC**

## Animations

**Nicolas Benguigui** est passé dans l'émission « C'est Pas Faux » de Radio Phénix sur le thème des « datas dans le sport ».

« Appelées aussi data, les données ont pris ces dernières années une place prépondérante dans le sport au point d'être aujourd'hui devenues incontournables pour de nombreux clubs comme pour les sportifs amateurs ». Plus d'infos [ici](#).

**Équipe CODAG.**

**Djawad Bekkoucha** et **Abdelkader Ouali** ont présenté un séminaire intitulé « Extraction de Motifs d'Intervalles Fermés avec la Programmation Par Contraintes » le 23 novembre devant l'équipe TASC à l'IMT-Atlantique, Nantes.

**Équipe CODAG.**

## Résultat majeur

« Inductive definitions in logic versus programs of real-time cellular automata », un article fleuve de 60 pages de **Etienne Grandjean, Théo Grente** et **Véronique Terrier** a été définitivement accepté au journal Theoretical Computer Science.  
**Équipe AMACC**

## Responsabilités

**Gaële Simon** a été élue en septembre 2023 au Conseil académique de Normandie Université au titre du collège des personnels pour l'Université Le Havre Normandie.  
**Équipe MAD**

**Christophe Rosenberger** est renouvelé comme membre nommé au CNU 27.  
**Équipe SAFE**

**Adeline Roux-Langlois**, directrice de recherche CNRS au GREYC, a rejoint le bureau du GdR CNRS Sécurité Informatique en tant que directrice adjointe.  
**Équipe AMACC**

**Pascal Vanier** est membre du comité d'organisation du mois thématique « Mathématiques discrètes et informatique : mots, complexité, dynamique, groupes » se déroulant du 29 janvier 2024 au 1er mars 2024. Ce mois aura lieu au CIRM (Centre de international de rencontres mathématiques) à Luminy. Plusieurs membres de l'équipe AMACC participent à cette manifestation.

Ce mois thématique propose une promenade le long des mathématiques discrètes et de la théorie du calcul pour explorer à la fois une série d'objets mathématiques (groupes, systèmes dynamiques symboliques, mots...) et une série de notions de l'informatique (information, aléatoire, calculabilité, complexité...) qui partagent de nombreux liens étroits. Afin d'encourager les participants (en particulier les plus jeunes) à assister à l'ensemble du mois et de favoriser les interactions en dehors de la zone d'expertise de chacun, la première semaine consistera en une école de recherche avec une introduction approfondie aux objets et notions clés de l'ensemble du mois.

URL: [cirm-math](http://cirm-math)

**Équipe AMACC**

## Vulgarisation scientifique

**Grégory Bonnet** a été invité par la Corpo Sciences Caen – association des étudiants en sciences de Caen - à donner une conférence de vulgarisation de ses travaux le 17 octobre 2023, dans le cadre du cycle de conférences « Nos enseignants chercheurs à l'honneur ».  
**Équipe MAD**

**Adeline Roux-Langlois** (avec **Corentin Jeudy**) a écrit un article de vulgarisation « Cryptographie reposant sur les réseaux euclidiens » dans Techniques de l'ingénieur. URL: [h5216](https://doi.org/10.1051/technique/2023016)  
**Équipe AMACC**

## Focus personnel : Julien Rabin (équipe IMAGE)

Aujourd'hui, notre attention se porte sur **Julien Rabin**, sympathique MC27 (ENSICAEN) de l'équipe IMAGE du laboratoire GREYC, qui a soutenu son HDR en décembre 2022, et qui est spécialiste de l'IA pour le traitement et la génération des images numériques.



« **Qu'est-ce qui te passionne le plus dans ta recherche en traitement d'images et IA?** »

Le traitement des images est un domaine assez vaste qui constitue un terrain de jeu idéal pour la mise en oeuvre de divers outils méthodologiques (formulations variationnelles, approches statistiques, apprentissage automatique, modèles génératifs ...), ce qui le rend particulièrement attrayant. C'est la combinaison de ces différentes techniques couplées à des réseaux de neurones profonds qui ont notamment permis des résultats spectaculaires ces dernières années, notamment en génération automatique d'images, récemment popularisée sous le terme de « IA générative ». A ce titre, il est frappant de voir que les mêmes formalismes peuvent être utilisés pour des tâches en apparence éloignées, comme le traitement automatique des langues.

« **Y a-t-il un projet récent sur lequel tu as travaillé et dont tu es particulièrement fier? Qu'est-ce qui le rend spécial?** »

Dans le cadre de différents projets, nous nous sommes penchés sur la question de la synthèse de textures, qui est un cas particulier de génération d'images où l'on réduit le jeu des données d'entraînement à une seule image d'exemple. Les modèles génératifs de l'état de l'art donnant les résultats les plus réalistes s'appuient sur des grands modèles reposant sur des centaines de millions de paramètres pré-entraînés sur de grandes collections d'images et raffinés sur le jeu de données. Nous avons montré avec des collègues et des étudiants qu'il était possible dans ce cadre de rivaliser avec ces approches en considérant des modèles beaucoup plus modestes s'appuyant sur des techniques de génération par recopies de motifs locaux, exploitant le formalisme du transport optimal. Le défi auquel nous nous attaquons actuellement avec plusieurs doctorants est de généraliser ce type de modèles pour les petits ensembles de données.

« **En tant que chercheur, quels sont les défis les plus intéressants (ou frustrants) que tu rencontres actuellement dans ton domaine?** »

Les modèles ultra-performants apparus ces dernières années ne sont pas sans soulever un certain nombre de questions ou de défis qui méritent notre attention, et dont la communauté scientifique s'est déjà emparé. Parmi celles-ci, nous travaillons dans l'équipe sur les questions du contrôle du résultat. Les modèles génératifs profonds sont certes capables de produire des résultats désormais photo-réalistes, mais trop souvent au prix d'une trop grande liberté lorsqu'il s'agit de respecter les données observées pour des tâches de restauration d'images, ou encore le texte de description pour de la génération conditionnelle multi-modale. Un autre axe de recherche prometteur est celui de l'interprétabilité et de l'évaluation des modèles génératifs sur lequel nous travaillons actuellement.

« **En parlant de l'IA, il y a beaucoup de discussions sur l'éthique. Comment prends-tu en compte ces aspects dans ton travail?** »

En tant qu'enseignant, les différents succès des modèles de langages actuels doivent nous amener à prendre en compte ces évolutions dans notre manière d'enseigner et d'évaluer le travail des étudiants. Pour ce qui est de notre activité de recherche, nous pouvons d'ores et déjà témoigner de l'existence de biais dans les jeux de données utilisés pour apprendre des modèles et de l'importance de traiter ce type de problématique.

« **Quelles tendances penses-tu voir émerger dans ton domaine dans les années à venir?** »

Ces dernières années, le développement de matériel et de méthodes de calculs à grande échelle combinés à la disponibilité de quantité massive de données de qualité ont permis l'émergence de giga-modèles. Cette croissance des modèles va certainement atteindre ses limites, et n'est souvent pas souhaitable pour de multiples raisons que l'on a déjà évoquée pour certaines (impact environnemental à la fois pour l'apprentissage et pour l'inférence, interprétabilité, jeux de données limités ou coûteux ...). Dans ce contexte, concevoir des modèles offrant à la fois des performances de haut niveau tout en garantissant certaines propriétés (robustesse, généralisation, etc) reste un défi.

« **Conseils pour les Nouveaux Chercheurs : Si tu pouvais remonter le temps et donner un conseil à ton jeune moi cherchant dans le traitement d'images, quel serait-il?** »

De postuler au GREYC bien évidemment !

« **Merci Juju !** »

# FOCUS scientifique AMACC : Étude théorique de l'algorithme git bisect (1)

Git est un logiciel de gestion de versions très répandu aujourd'hui (voir [wiki Git](#)). Dans un article long à paraître dans la revue Algorithmica (et tiré de la version courte déjà « best paper award » de la conférence LATIN 2022), Julien Courtiel, Paul Dorbec et Romain Lecoq font une étude théorique de l'algorithme bisect utilisé par git pour détecter les bogues. Julien Courtiel nous propose de mieux comprendre leurs travaux.

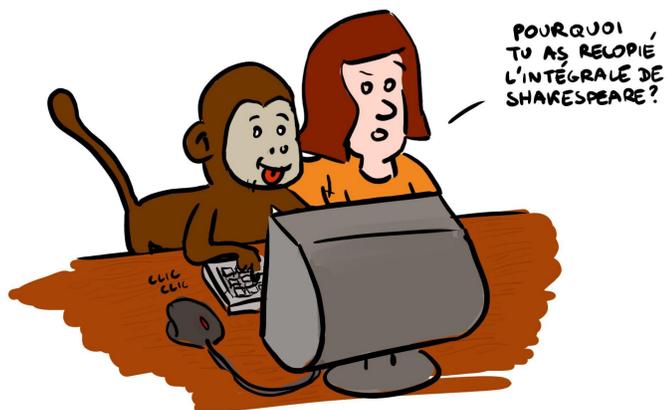
APRÈS AVOIR CORRIGÉ 200 COPIES DE LA, RÉDIGÉ UNE DEMANDE D'AVIS AU FSD, MONTÉ UN DOSSIER ANR, SOUMIS UN DOSSIER POUR VOTRE RIPEC, PARTICIPÉ À LA FÊNO, CRÉÉ VOTRE 3<sup>ÈME</sup> START-UP EN MOINS DE 10 ANS, VOUS N'AVEZ PLUS TROP DE TEMPS POUR LA RECHERCHE...



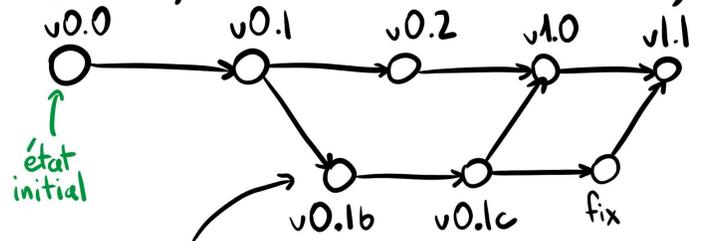
TOUTEFOIS, VOUS AVEZ EU UNE IDÉE DE GÉNIE: VOUS AVEZ APPRIVOISÉ UN SINGE POUR VOUS SECONDER DANS VOS PROJETS DE RECHERCHE (ÇA COÛTE MOINS CHER QU'UN DOCTORANT)



AFIN DE SUIVRE LES PROGRÈS DE VOTRE SINGE, VOUS UTILISEZ **GIT**, UN LOGICIEL DE GESTION DE VERSIONS (TRÈS CONNU SI VOUS ÊTES DANS L'INFORMATIQUE)



**GIT** SAUVEGARDE TOUTES LES MODIFICATIONS ENTRE LES VERSIONS D'UN PROJET (CE SONT LES "COMMIT", SYMBOLISÉS PAR DES FLÈCHES)



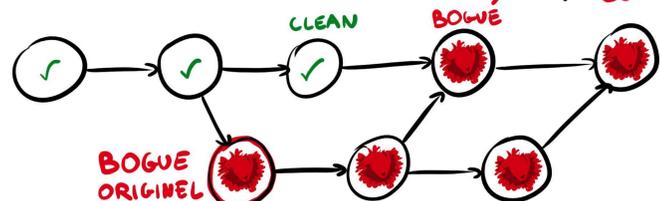
**GIT** PERMET AUSSI DE BIFURQUER DE LA TRAME PRINCIPALE DU PROJET VIA DES **BRANCHES** (QU'ON FUSIONNERA ULTÉRIEUREMENT)



MALHEUREUSEMENT LA DERNIÈRE VERSION DU PROJET COMPORTE UN MÉCHANT **BOGUE** DÙ AU CODE DU SINGE. VOUS VOUS DEMANDEZ DANS QUELLE VERSION CE BOGUE EST APPARU POUR LA 1<sup>ÈRE</sup> FOIS

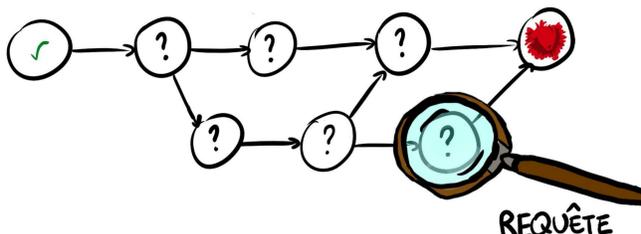


EN TERMES DE GRAPHE, LE BOGUE S'EST PROPAGÉ DE LA VERSION OÙ IL A ÉTÉ INTRODUIT AUX VERSIONS QUI SUIVENT. UNE VERSION EST DONC SOIT **CLEAN** SOIT **BOGÉE**



LE BUT DU JEU EST DE RETROUVER LA VERSION OÙ SE TROUVE CE "BOGUE ORIGINEL".

TOUTEFOIS ON NE CONNAÎT PAS LE STATUT DE TOUTES LES VERSIONS À L'AVANCE. POUR CE FAIRE, ON PEUT EFFECTUER UNE **REQUÊTE** SUR UNE VERSION PRÉCISE POUR SAVOIR SI ELLE EST **CLEAN** OU **BOGÉE**



SI LA REQUÊTE DÉVOILE UN **BOGUE**, ALORS LE BOGUE ORIGINEL EST LA VERSION REQUÊTÉE OU UNE AVANT ELLE:

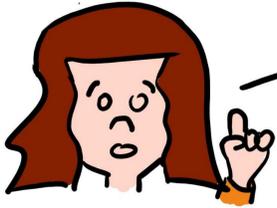
SI LA VERSION EST **CLEAN**, ALORS LE BOGUE ORIGINEL EST DANS L'AUTRE PARTIE DU GRAPHE

SAVOIR SI UN BOGUE EST PRÉSENT OU NON PEUT ÊTRE (TRÈS) LONG.

C'EST POURQUOI ON VEUT FAIRE LE MINIMUM DE REQUÊTES POSSIBLE POUR TROUVER LE BOGUE ORIGINEL

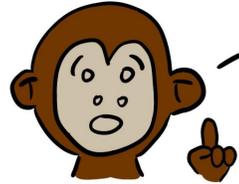
# FOCUS scientifique AMACC : Étude théorique de l'algorithme git bisect (2)

SI VOTRE DÉPÔT GIT NE COMPORTE PAS DE BRANCHES, LA STRATÉGIE EST SIMPLE



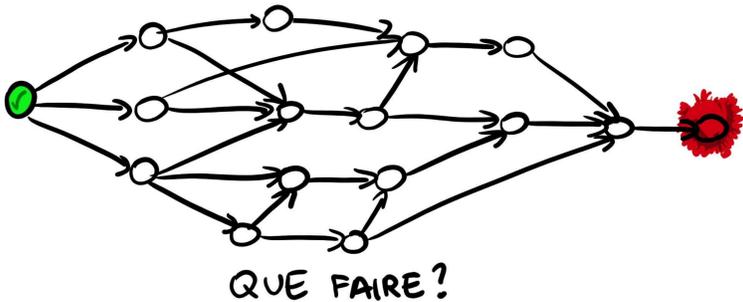
IL SUFFIT DE REQUÊTER LE MILIEU DE LA CHAÎNE. ON SE RETROUVE AVEC DEUX FOIS MOINS DE SOMMETS, QUE LA REQUÊTE DÉVOILE UN BOGUE OU NON.

LE PROCESSUS EST RÉPÉTÉ : C'EST LA RECHERCHE DICHOTOMIQUE ( $\lceil \log_2(n) \rceil$  REQUÊTES AU PIRE OÙ  $n$  EST LE NOMBRE DE VERSIONS)



HOUHI HOUHAHA

MAIS LE DÉPÔT DE VOTRE SINGE RESSEMBLE À :



QUE FAIRE ?

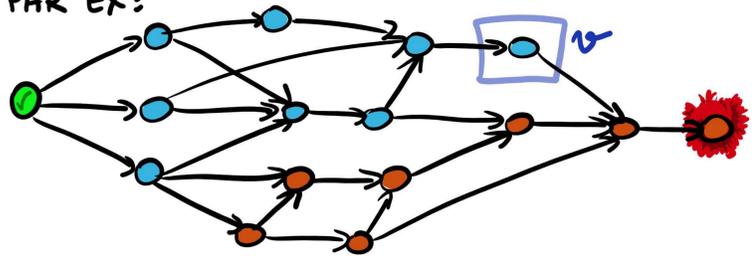


KIKI AIME LES BRANCHES

IL EXISTE UN ALGORITHME AU SEIN DE GIT QUI RÉSOUT LE PROBLÈME EN S'INSPIRANT DE LA RECHERCHE DICHOTOMIQUE: GIT BISECT!

LE PRINCIPE DE GIT BISECT EST HEURISTIQUE : IL CHOISIT LA REQUÊTE QUI PARTAGE LE GRAPHE DE LA MANIÈRE LA PLUS ÉQUILIBRÉE QUE POSSIBLE. POUR CELA, IL CALCULE UN SCORE POUR CHAQUE VERSION ET REQUÊTE CELLE AU SCORE MAXIMAL :  $score(v) = \min(\text{nbr ancêtres de } v, \text{nbr de non ancêtres de } v)$

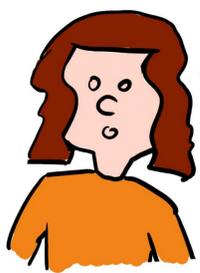
PAR EX :



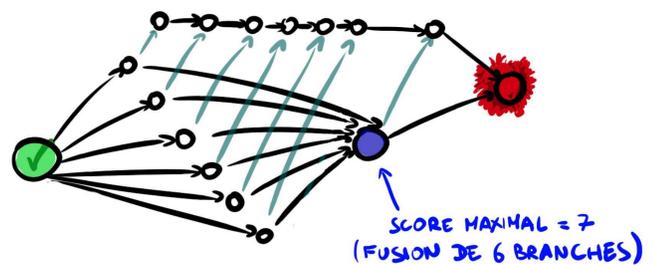
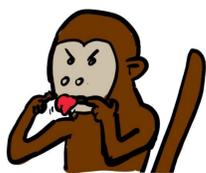
$score(v) = \min(8, 7) = 7$  (ON PEUT PAS FAIRE MEUX)

LA STRATÉGIE DE GIT BISECT EST-ELLE OPTIMALE ? C'EST LA QUESTION QUE SE SONT POSÉS JULIEN COURTIEL, PAUL DORBEC, ROMAIN LECOQ TROIS CHERCHEURS DE L'ÉQUIPE AMACC DU GREYC DANS UN ARTICLE AYANT REÇU LE "BEST PAPER AWARD" À LATIN22.

ILS ONT DÉCOUVERT DES EXEMPLES PATHOLOGIQUES, OÙ GIT BISECT PEUT UTILISER JUSQU'À  $2^k - 1$  REQUÊTES MAIS OÙ UNE STRATÉGIE OPTIMALE FAIT  $k + 1$  REQUÊTES DANS TOUS LES CAS. PAR EXEMPLE, POUR  $k = 3$  :



L'ALGO N'AVAIT JAMAIS ÉTUDIÉ DE MANIÈRE THÉORIQUE AVANT...



SCORE MAXIMAL = 7 (FUSION DE 6 BRANCHES)



C'EST TOUT ?

NAN. CES EXEMPLES PATHOLOGIQUES SONT TRÈS PEU VRAISEMBLABLES DANS LA MESURE OÙ ON NE FUSIONNE RAREMENT PLUS DE 2 BRANCHES SIMULTANÉMENT.

AU PROCHAIN BOGUE JE T'INSTALLE SAFARI COMME NAVIGATEUR WEB !



LES AUTEURS ONT POUSSÉ L'ÉTUDE AU CAS OÙ LES FUSIONS DE + DE 2 BRANCHES SONT INTERDITES. ILS ONT ALORS PROUVÉ QUE GIT BISECT ÉTAIT UNE  $\approx 1.71$ -APPROXIMATION. ILS ONT ENSUITE DÉCRIT UN ALGORITHME PLUS PERFORMANT THÉORIQUEMENT, BASÉ SUR LE NOMBRE D'OR, BAPTISÉ GOLDEN BISECT ( $\approx 1.44$ -APPROXIMATION) ↓ MAIS IL N'Y A PLUS DE PLACE POUR DÉTAILLER ↓