

Soutenance de thèse, Paris le 9 novembre 2018

Archives, fragments Web et diasporas

Pour une exploration désagrégée de corpus d'archives Web liées aux représentations en ligne des diasporas



Quentin Lobbé (LTCI Télécom ParisTech & Inria Paris)

Sous la direction de Pierre Senellart et Dana Diminescu

Plan de résolution

> Vers les représentations en ligne des diasporas

1. Les représentations en ligne des diasporas
2. Archiver le Web
3. Introduction au fragment Web
4. Les archives de la blogosphère marocaine
5. Les archives de yabiladi.com
6. Bilan et conclusion

Les représentations en ligne des diasporas

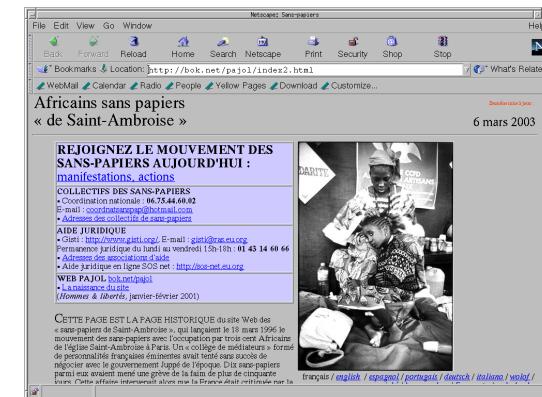
> Des populations dispersées mais présentes sur la toile



> Ordinateur personnel d'un couple de travailleurs philippins à Paris, Diminescu D, 2005



> mondeberbere.com, Maroc, 2002



> bok.net/pajol, France, 1996

Au tournant des années 2000, les multiples traces numériques laissées par les migrations deviennent l'objet de recherches sociologiques dédiées.

L'atlas e-Diasporas (1/2)

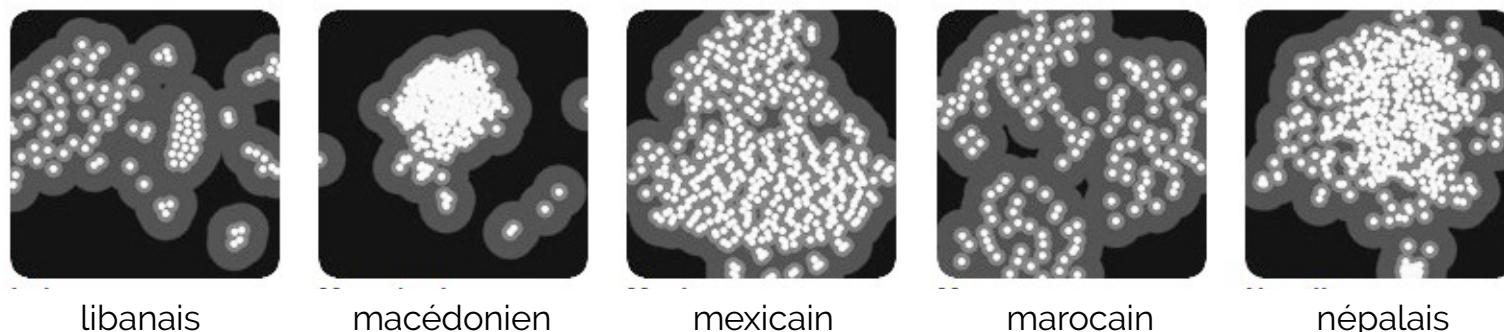
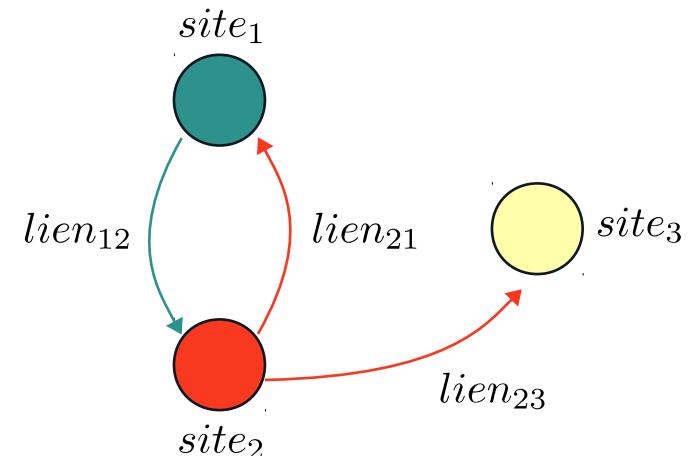
> Une initiative unique pour cartographier et étudier des collectifs migrants en ligne

Un **site Web migrant** est un site Web créé ou maintenu par une ou plusieurs personnes migrantes ou qui est en lien direct avec ces dernières.

Une **e-Diaspora** est un réseau dirigé de sites Web migrants liés entre eux par des liens de citations hypertextes.

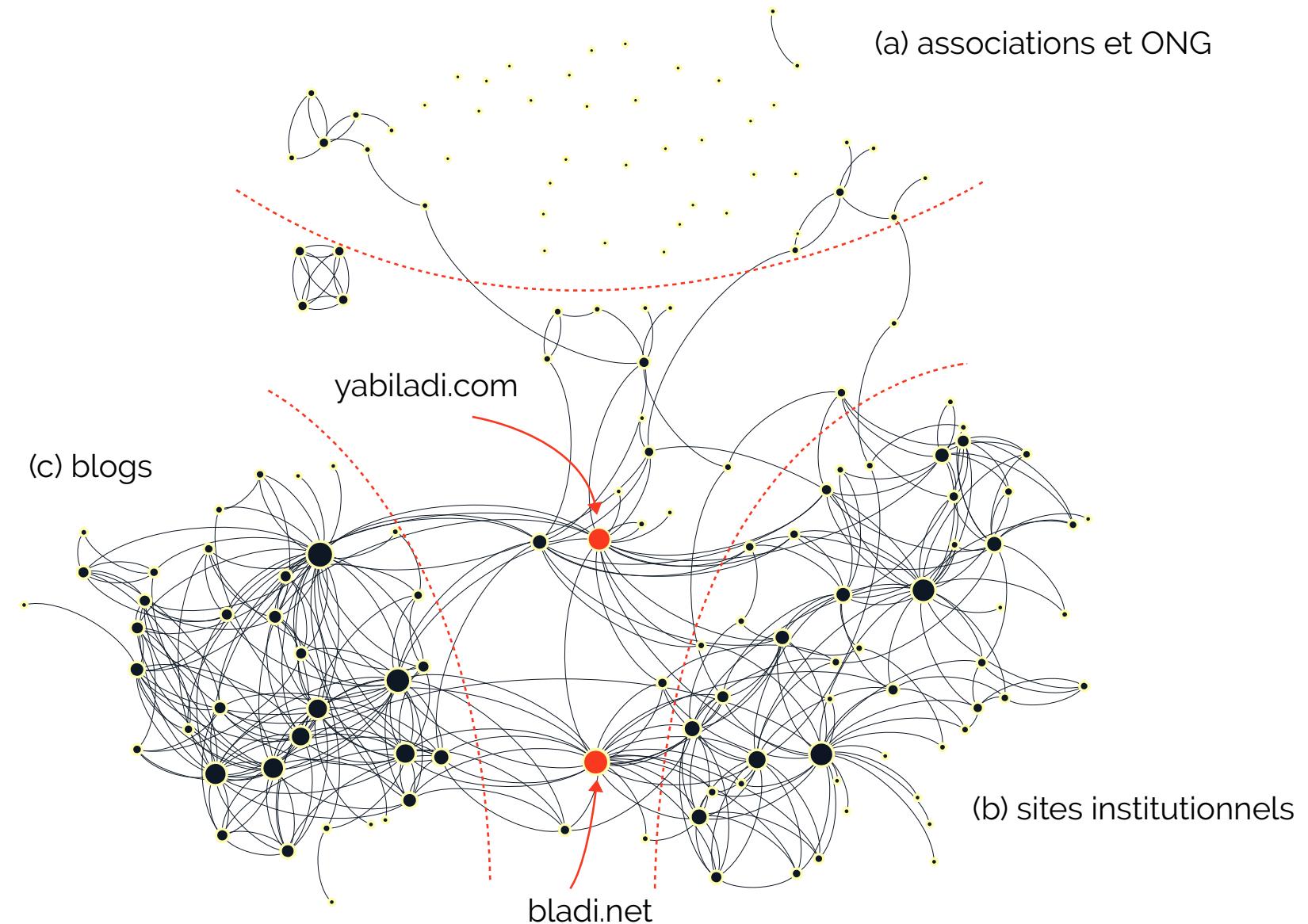
Une e-Diaspora est une entité qui se pense à la fois **en ligne** et **hors ligne**.

10 000 sites Web migrants observés, catégorisés et organisés parmi 30 e-Diasporas.



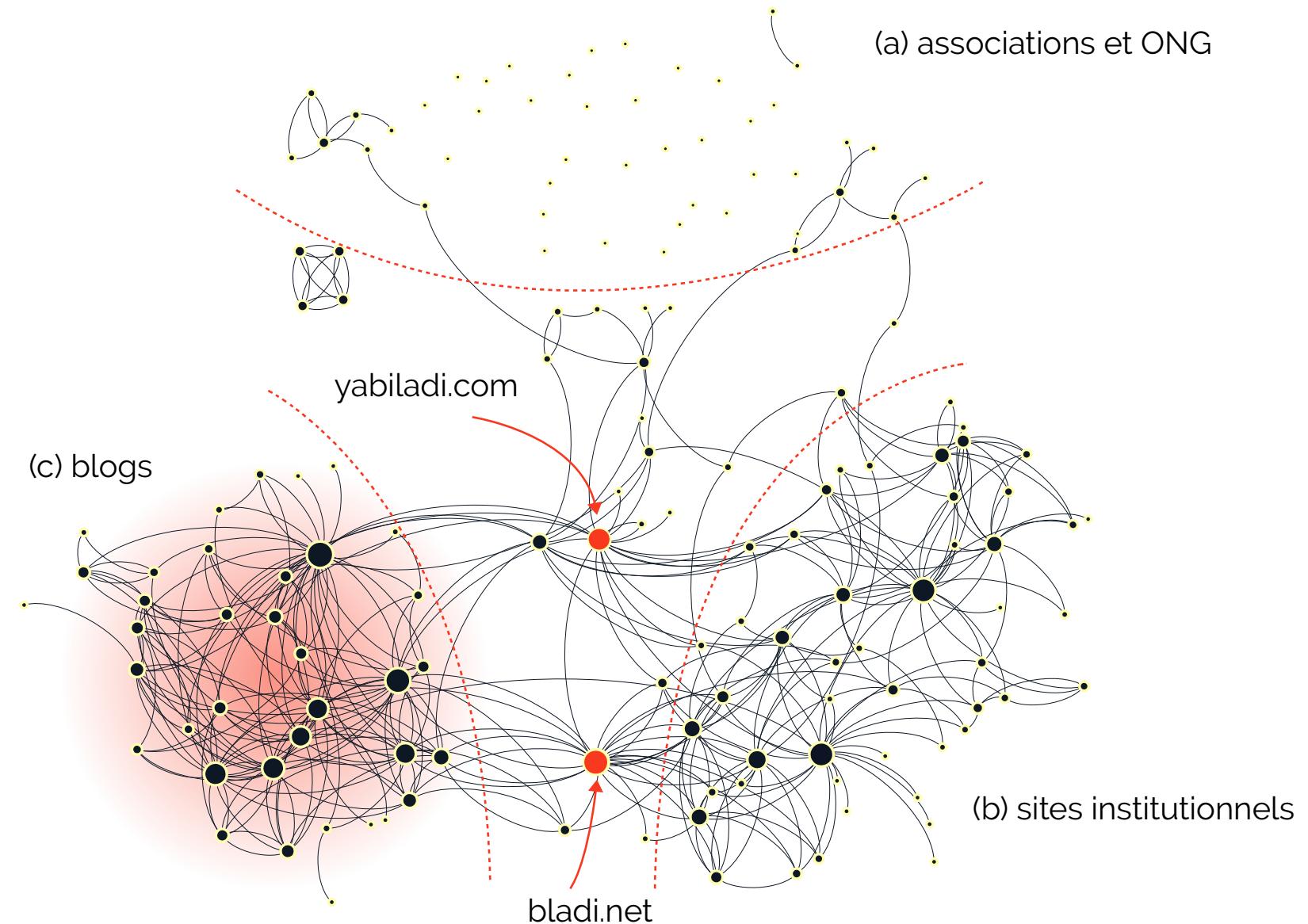
L'atlas e-Diasporas (2/2)

> Lire et comprendre une e-Diaspora



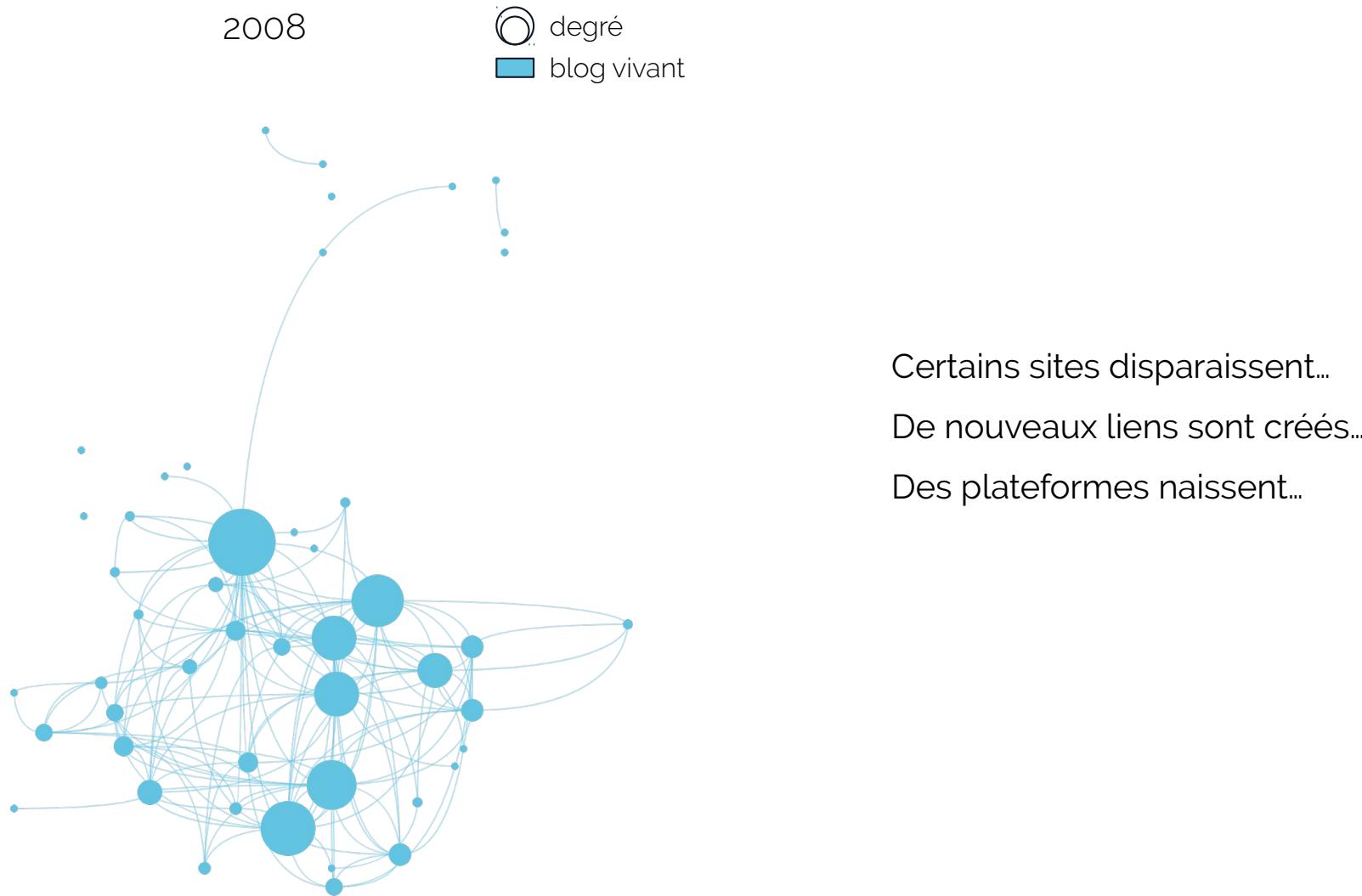
L'atlas e-Diasporas (2/2)

> Lire et comprendre une e-Diaspora



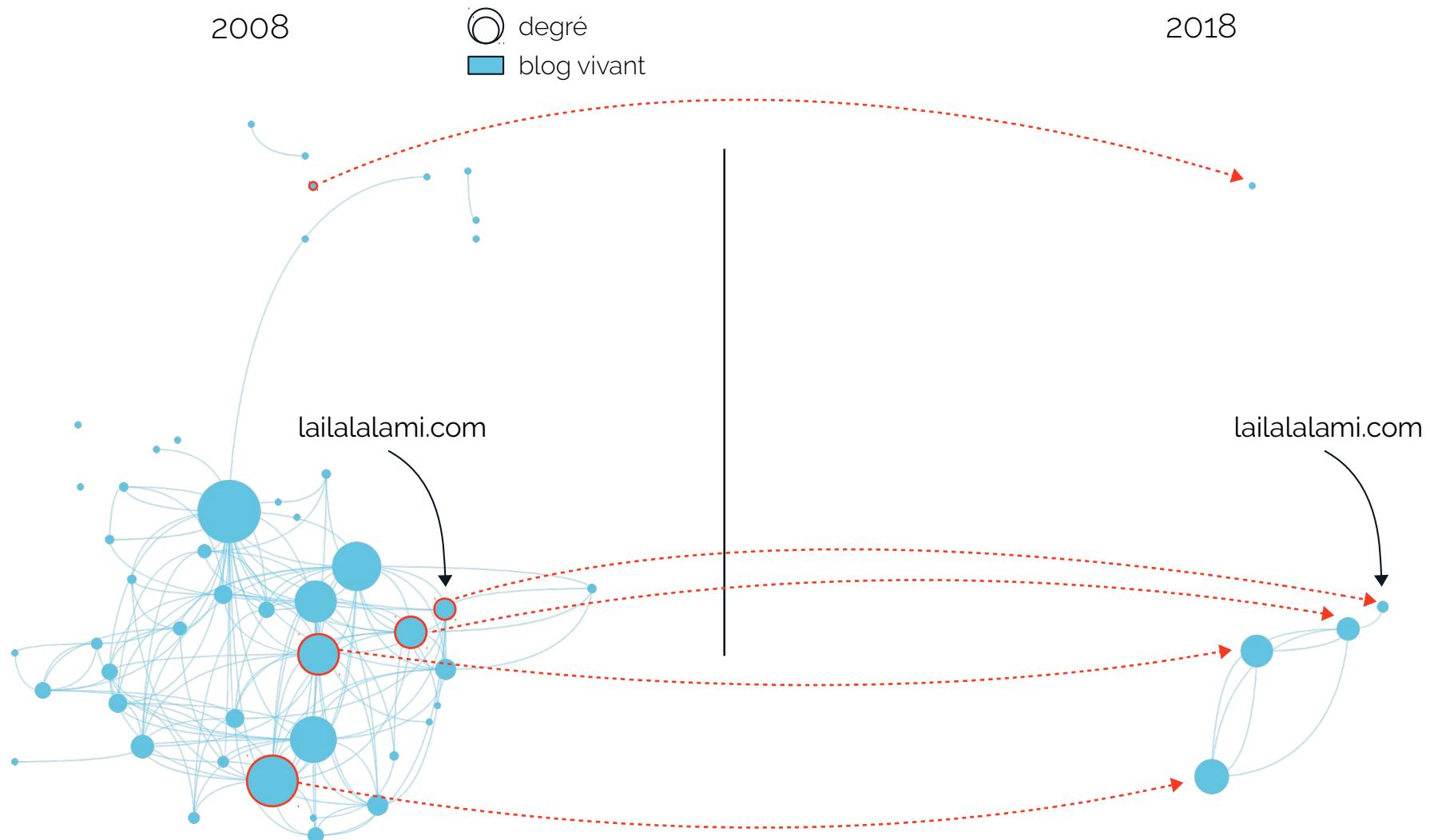
Le Web vivant est un espace en perpétuelle évolution

> La question des collectifs migrants éteints



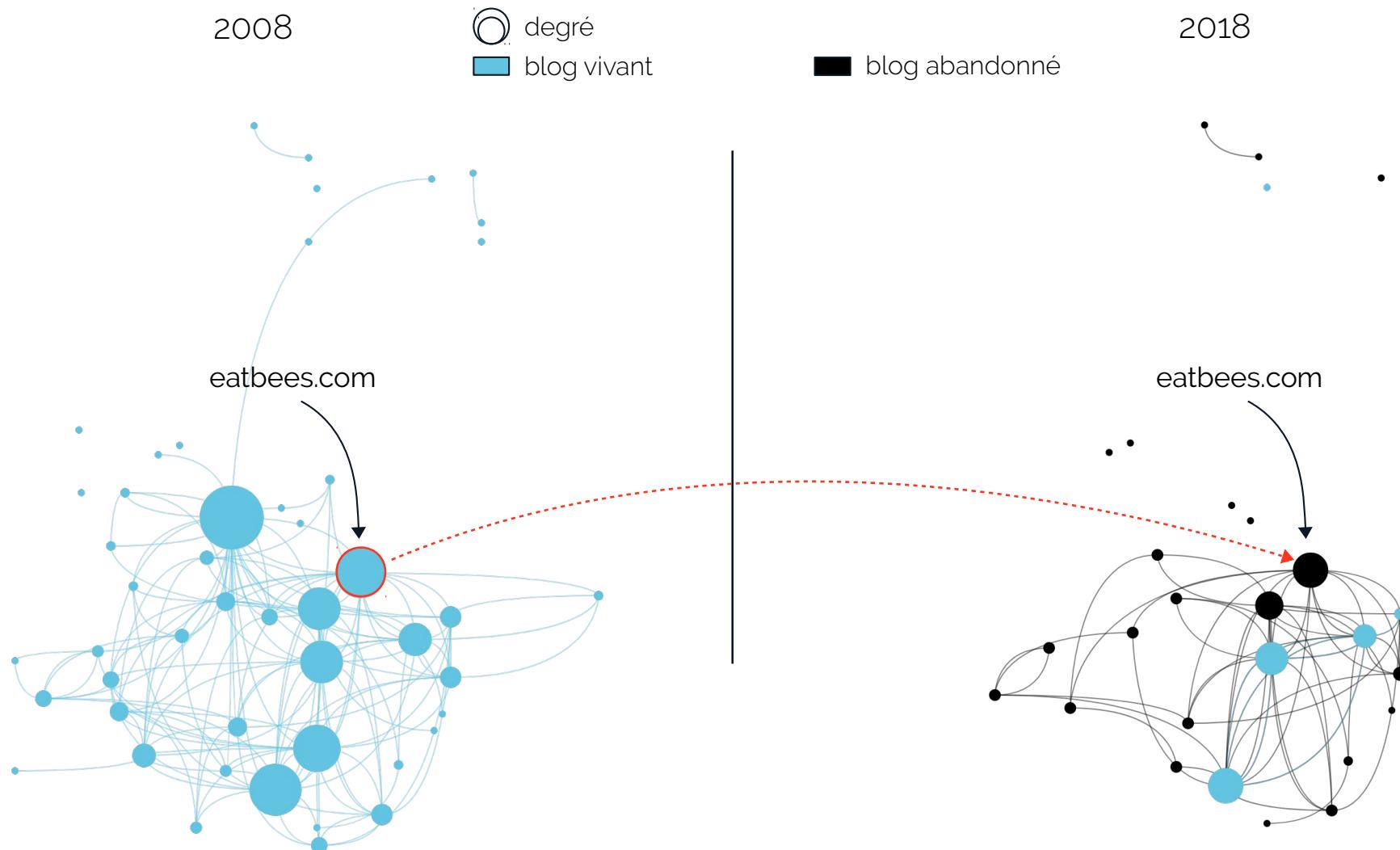
Le Web vivant est un espace en perpétuelle évolution

> La question des collectifs migrants éteints



Le Web vivant est un espace en perpétuelle évolution

> La question des collectifs migrants éteints



Le Web vivant est un espace en perpétuelle évolution

> La question des collectifs migrants éteints

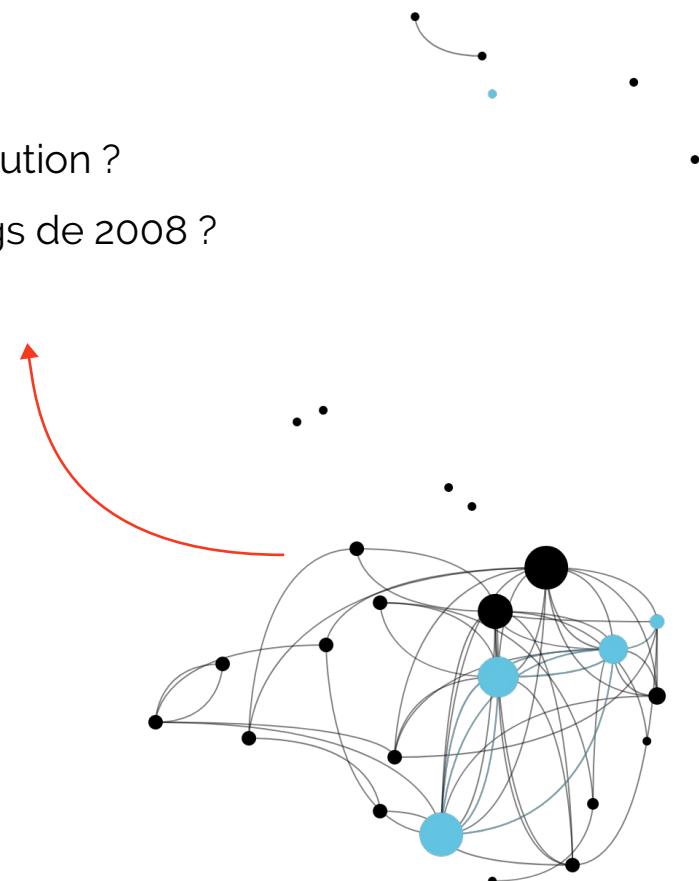
2018

Qu'est-il arrivé à la blogosphère marocaine ?

Quels enjeux hors ligne ont pu influencer son évolution ?

Que sont devenus les auteurs et lecteurs des blogs de 2008 ?

Comment retracer l'histoire d'une e-Diaspora ?



Le Web vivant est un espace en perpétuelle évolution

> La question des collectifs migrants éteints

Qu'est-il arrivé à la blogosphère marocaine ?

Quels enjeux hors ligne ont pu influencer son évolution ?

Que sont devenus les auteurs et lecteurs des blogs de 2008 ?

Comment retracer l'histoire d'une e-Diaspora ?



Pourquoi ne pas interroger des archives Web ?

Plan de résolution

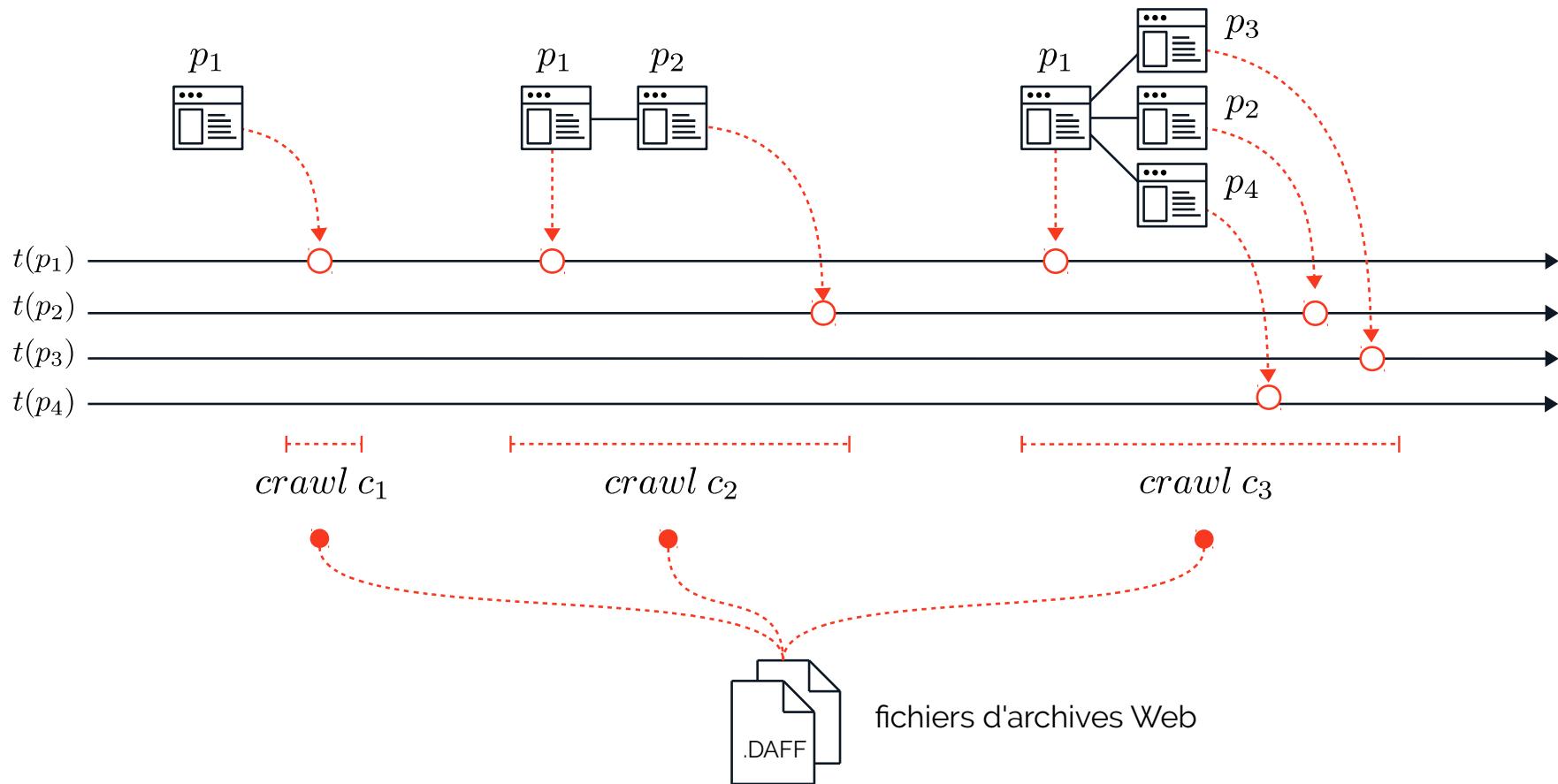
> Vers les archives Web

1. Les représentations en ligne des diasporas
- 2. Archiver le Web**
3. Introduction au fragment Web
4. Les archives de la blogosphère marocaine
5. Les archives de yabiladi.com
6. Bilan et conclusion

Archiver le Web ? (1/3)

> Collecter et préserver notre héritage numérique

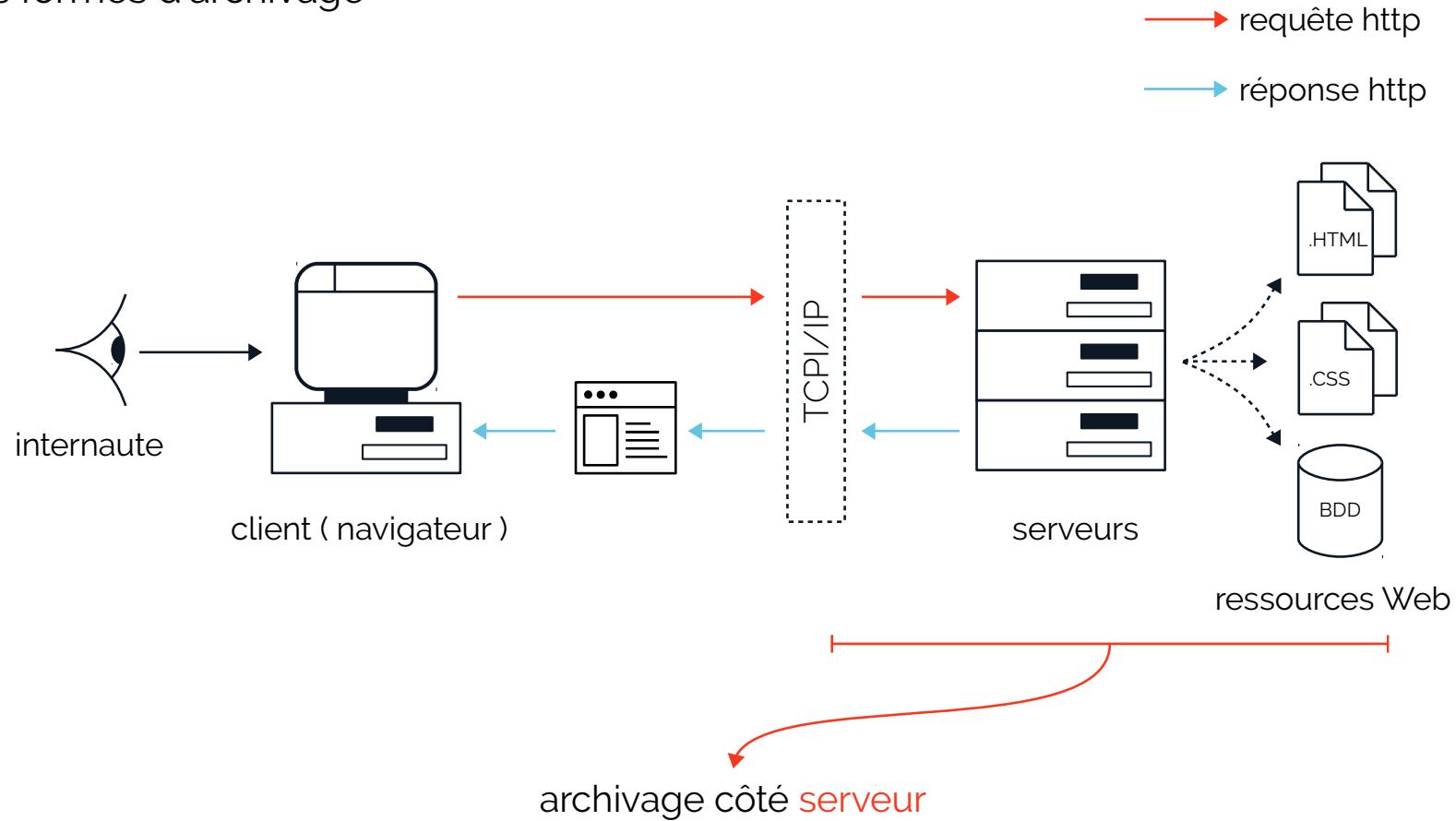
Faire du Web vivant et continu...



... un corpus discret d'archives Web

Archiver le Web ? (2/3)

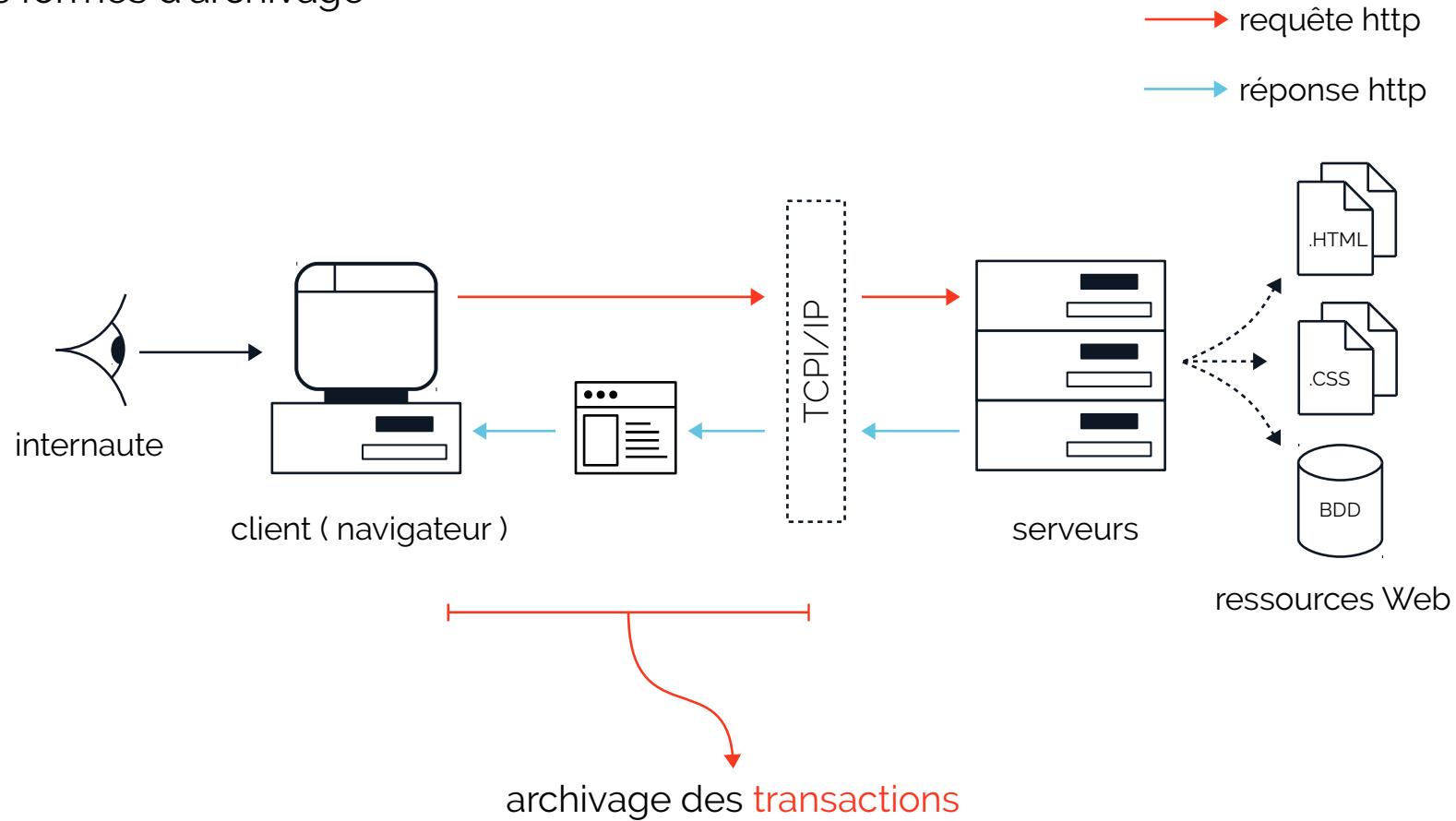
> Trois formes d'archivage



Étude en collaboration avec l'association Singa, autour du projet Calm

Archiver le Web ? (2/3)

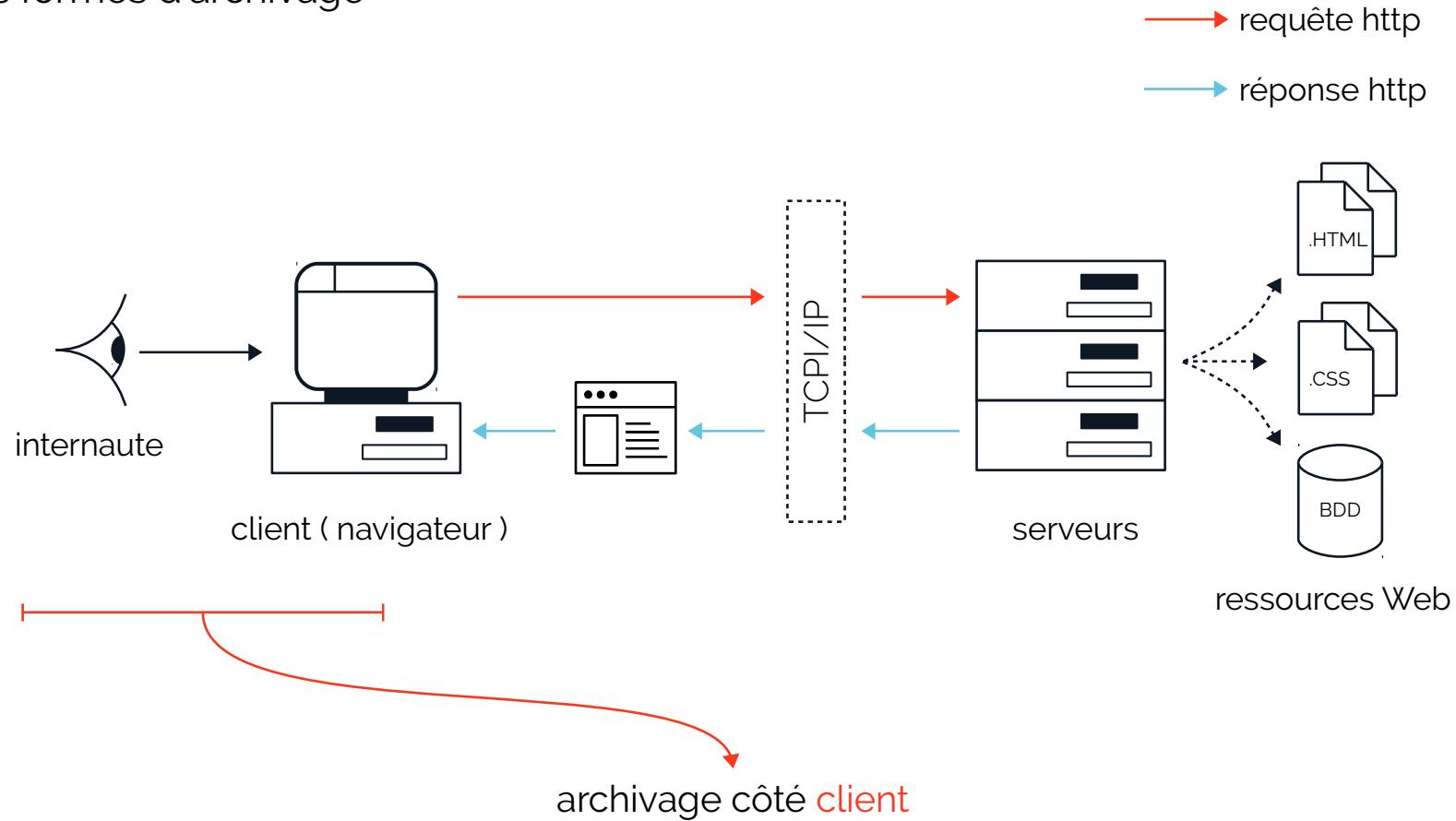
> Trois formes d'archivage



Étude en collaboration avec la Bibliothèque Publique d'Information du Centre Pompidou

Archiver le Web ? (2/3)

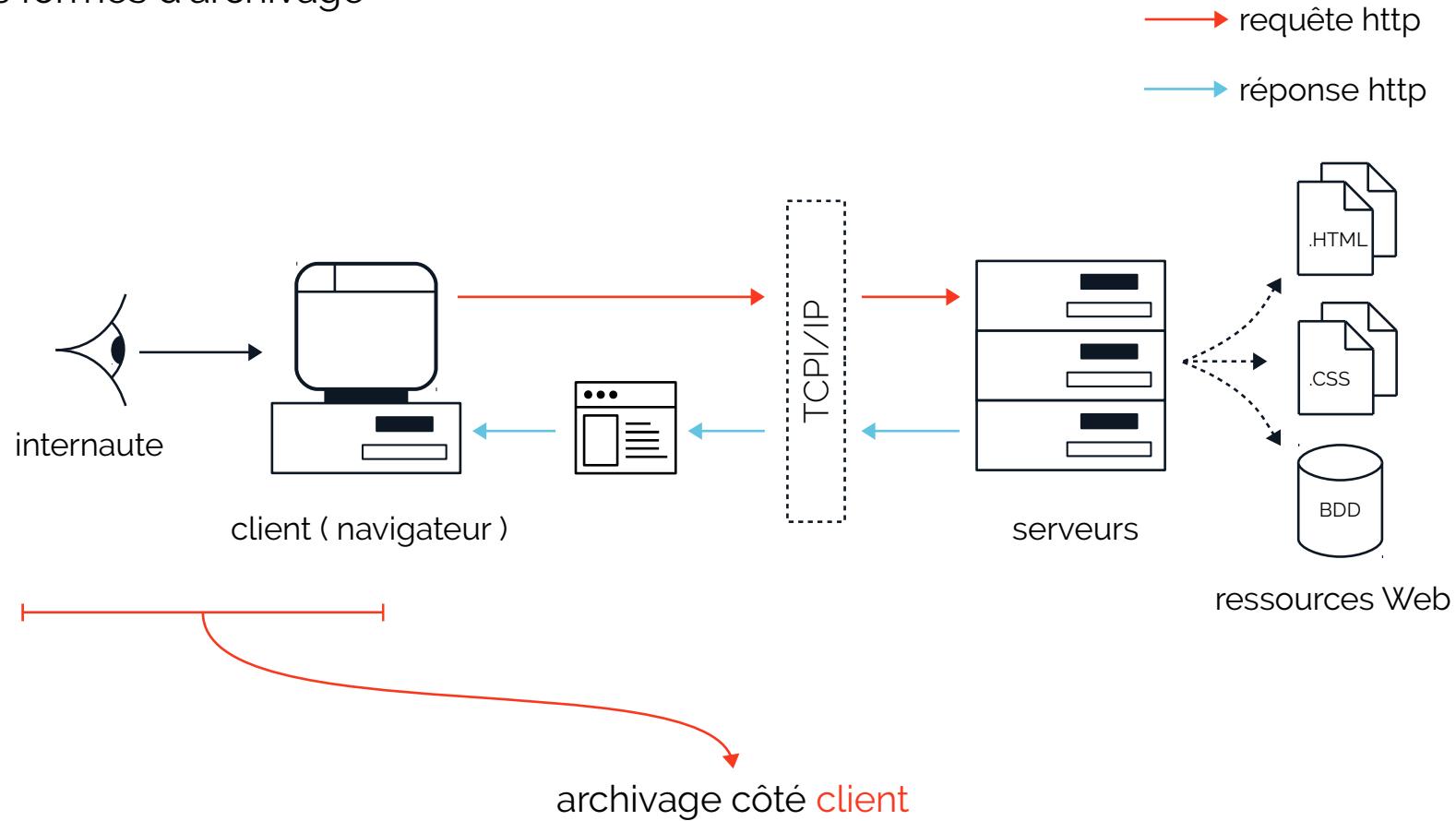
> Trois formes d'archivage



Étude sur les archives de l'Atlas e-Diasporas

Archiver le Web ? (2/3)

> Trois formes d'archivage



L'atlas e-Diasporas est aussi un corpus d'**archives Web** !

70 To de données

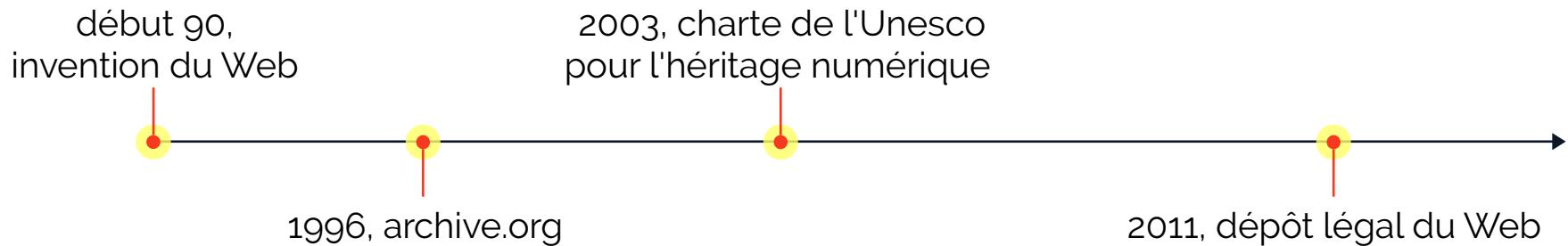
1030 M de pages Web

Capturées toutes les semaines ou mois (2010-2014)

Archivage par et au sein de l'INA

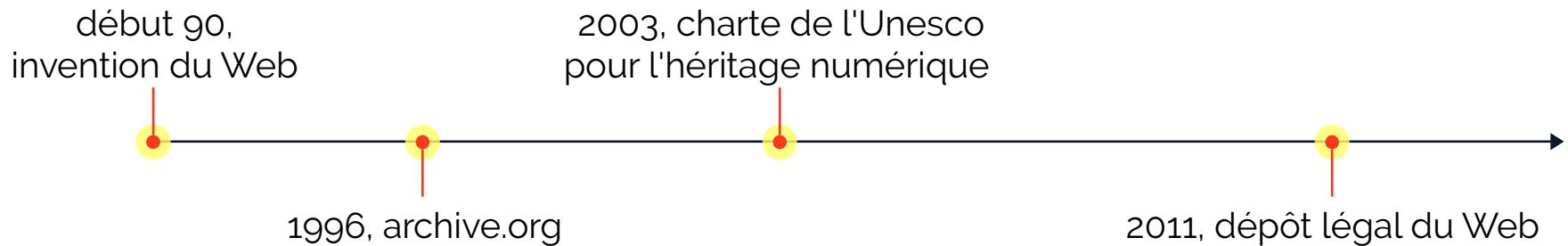
Archiver le Web ? (3/3)

> Les archives Web sont aussi vieilles que le Web lui même



Archiver le Web ? (3/3)

> Les archives Web sont aussi vieilles que le Web lui même

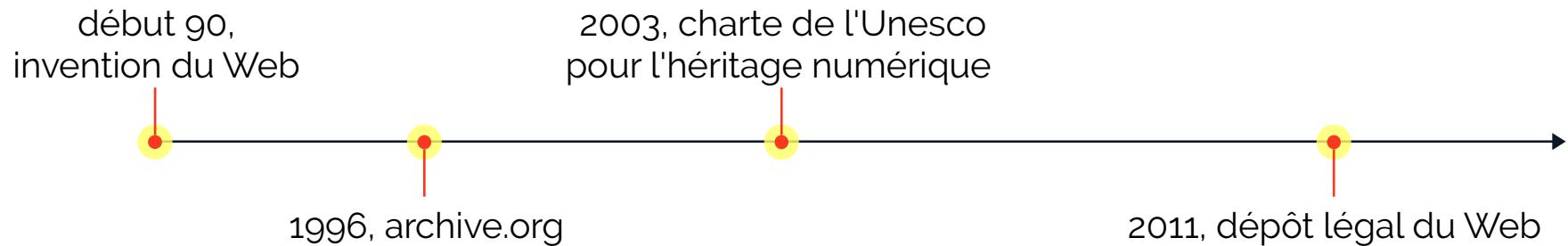


Mais les archives Web ne font l'objet que d'un **nombre confidentiel** d'études.

Pourquoi leur usage ne s'est-il **pas** encore **démocratisé** ?

Archiver le Web ? (3/3)

> Les archives Web sont aussi vieilles que le Web lui même



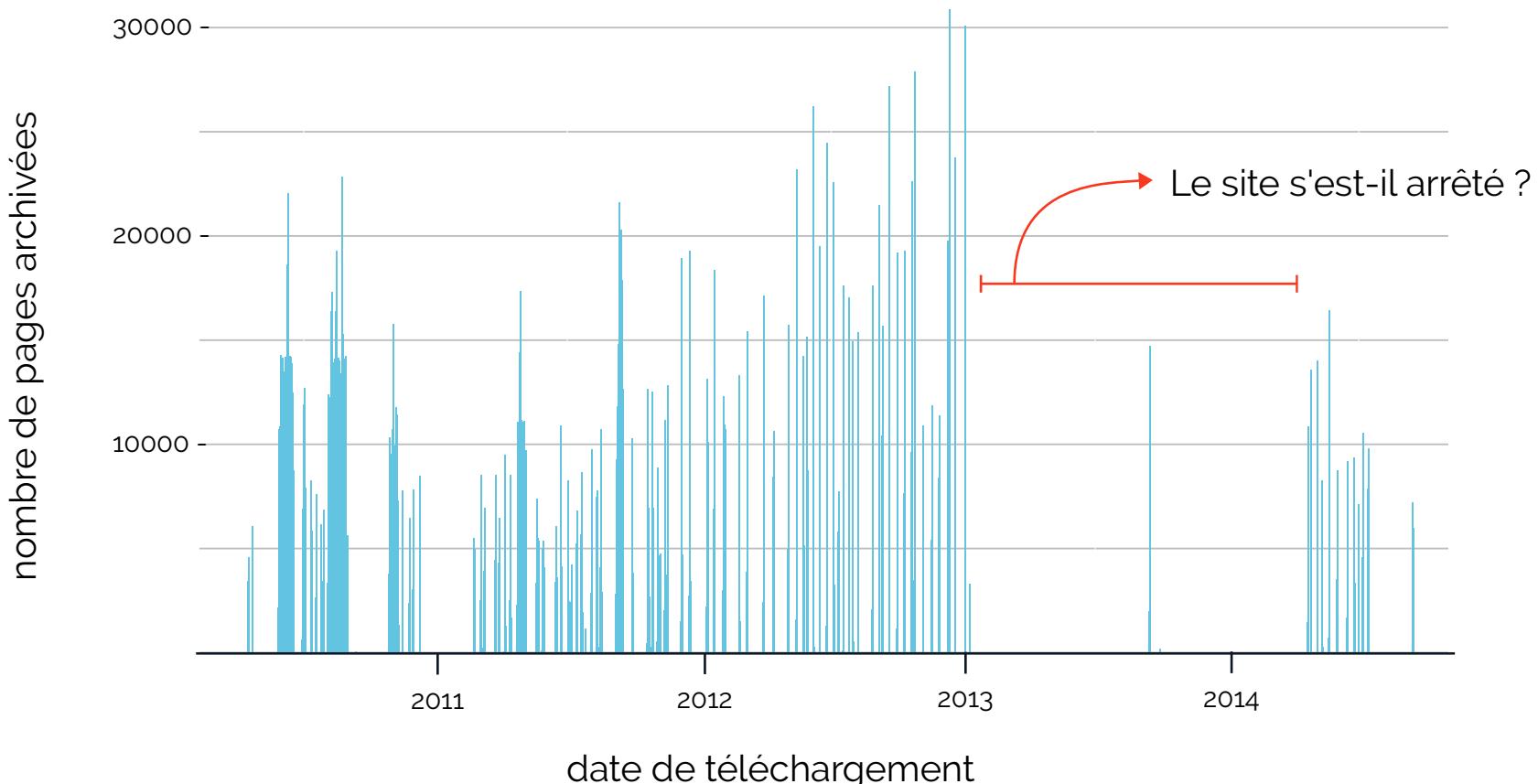
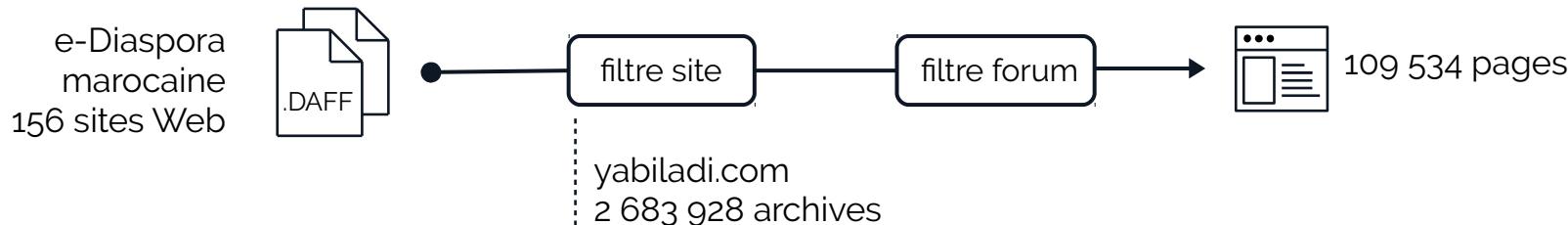
Mais les archives Web ne font l'objet que d'un **nombre confidentiel** d'études.

Pourquoi leur usage ne s'est-il **pas** encore **démocratisé** ?

Selon nous, le problème se situe au **coeur** même de la **structure** des corpus archivés.

Les archives Web ne sont pas des traces directes du Web (1/3)

> Soit l'expérience suivante



Les archives Web ne sont pas des traces directes du Web (2/3)

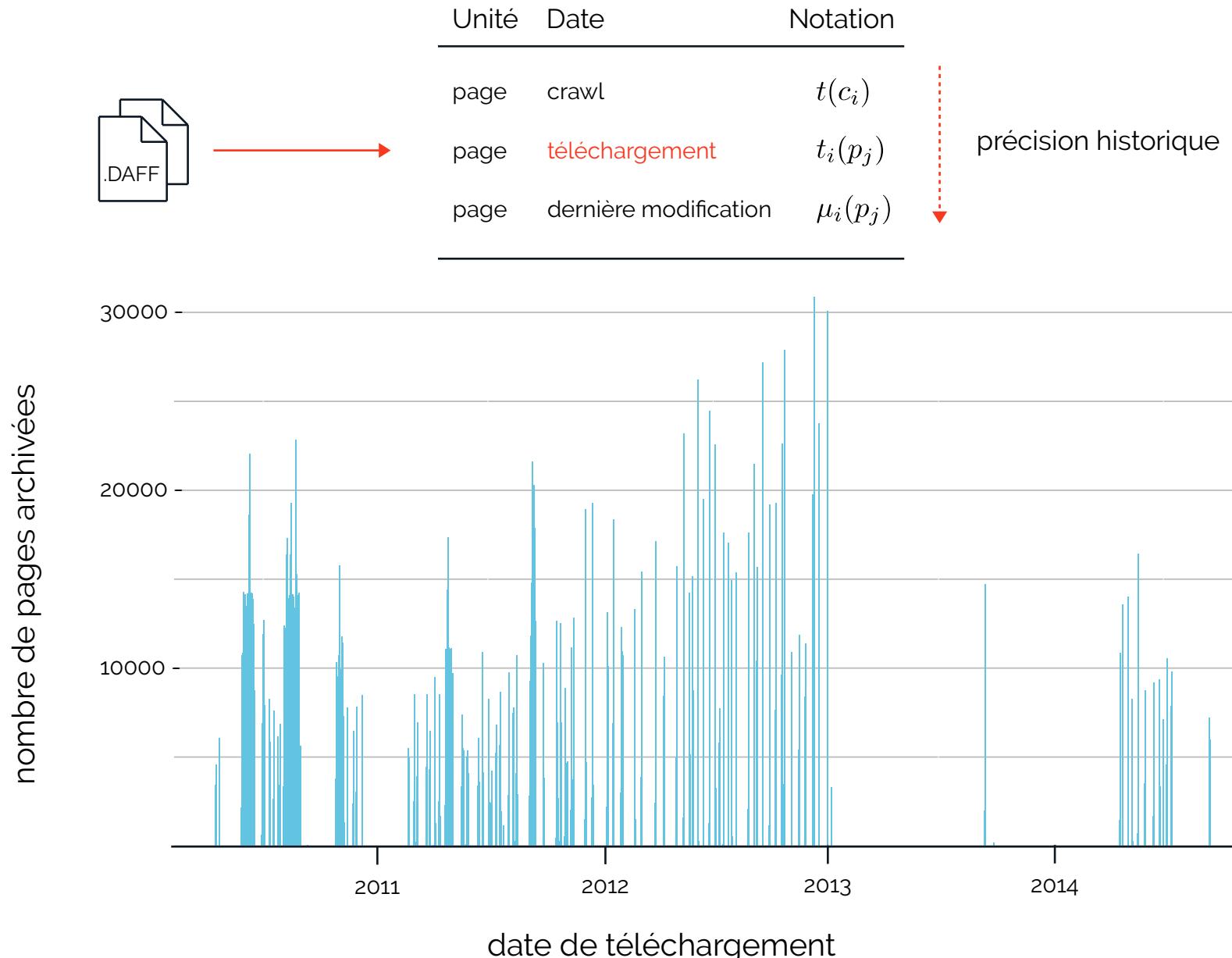
> Les archives Web sont les traces directes des outils de collecte



> « Boulevard du Temple », Louis Daguerre, 1838

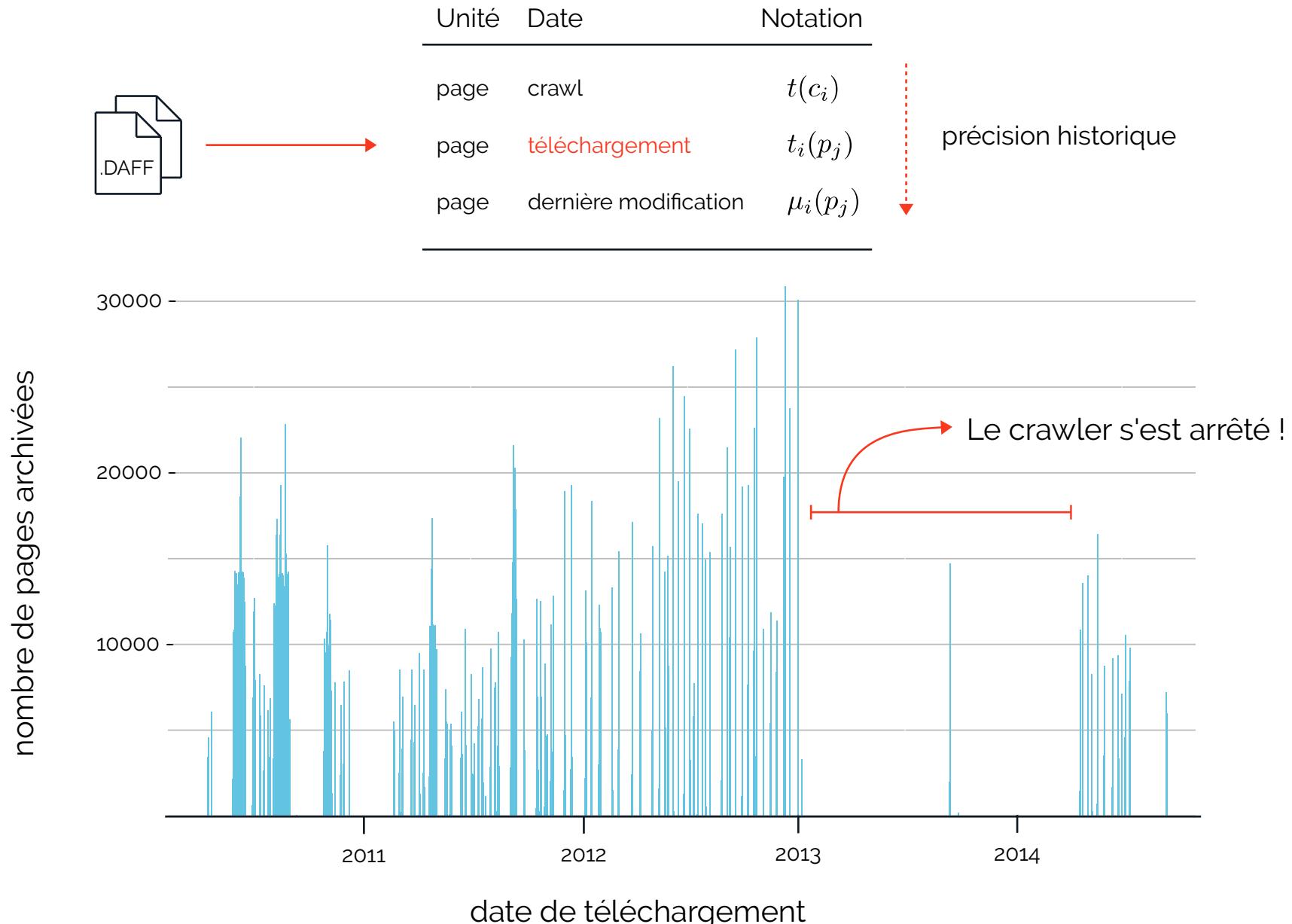
Les archives Web ne sont pas des traces directes du Web (3/3)

> Les archives Web sont arrachées à la temporalité du Web vivant



Les archives Web ne sont pas des traces directes du Web (3/3)

> Les archives Web sont arrachées à la temporalité du Web vivant



Les archives Web ne sont pas des traces directes du Web (3/3)

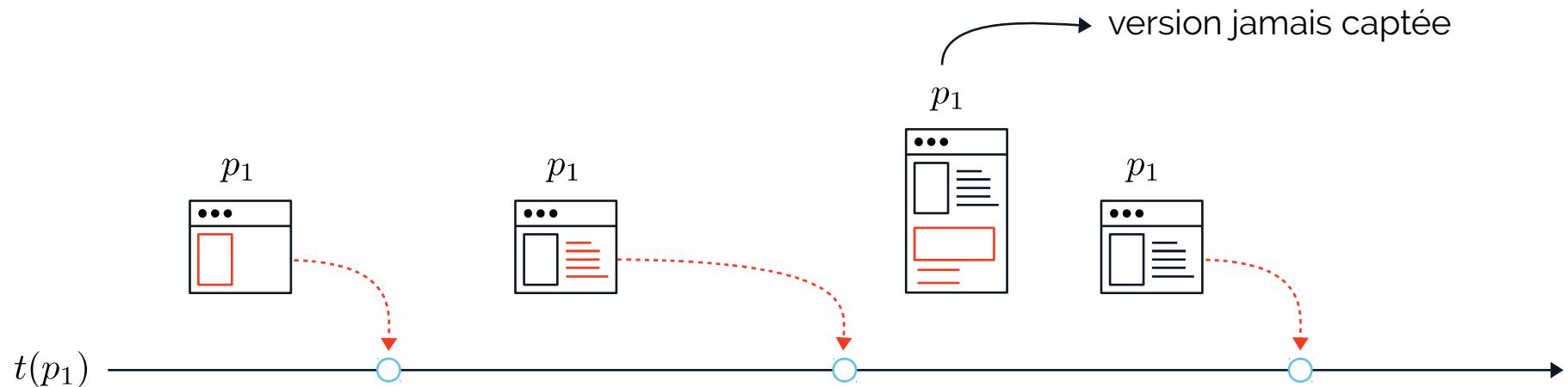
> Les archives Web sont arrachées à la temporalité du Web vivant

Unité	Date	Notation	
page	crawl	$t(c_i)$	
page	téléchargement	$t_i(p_j)$	précision historique
page	dernière modification	$\mu_i(p_j)$	

Par construction, les archives Web induisent la présence de multiples artefacts de crawl qui limitent toute exploration large du Web passé.

Comprendre les artefacts de crawl (1/4)

> Cécité de crawl

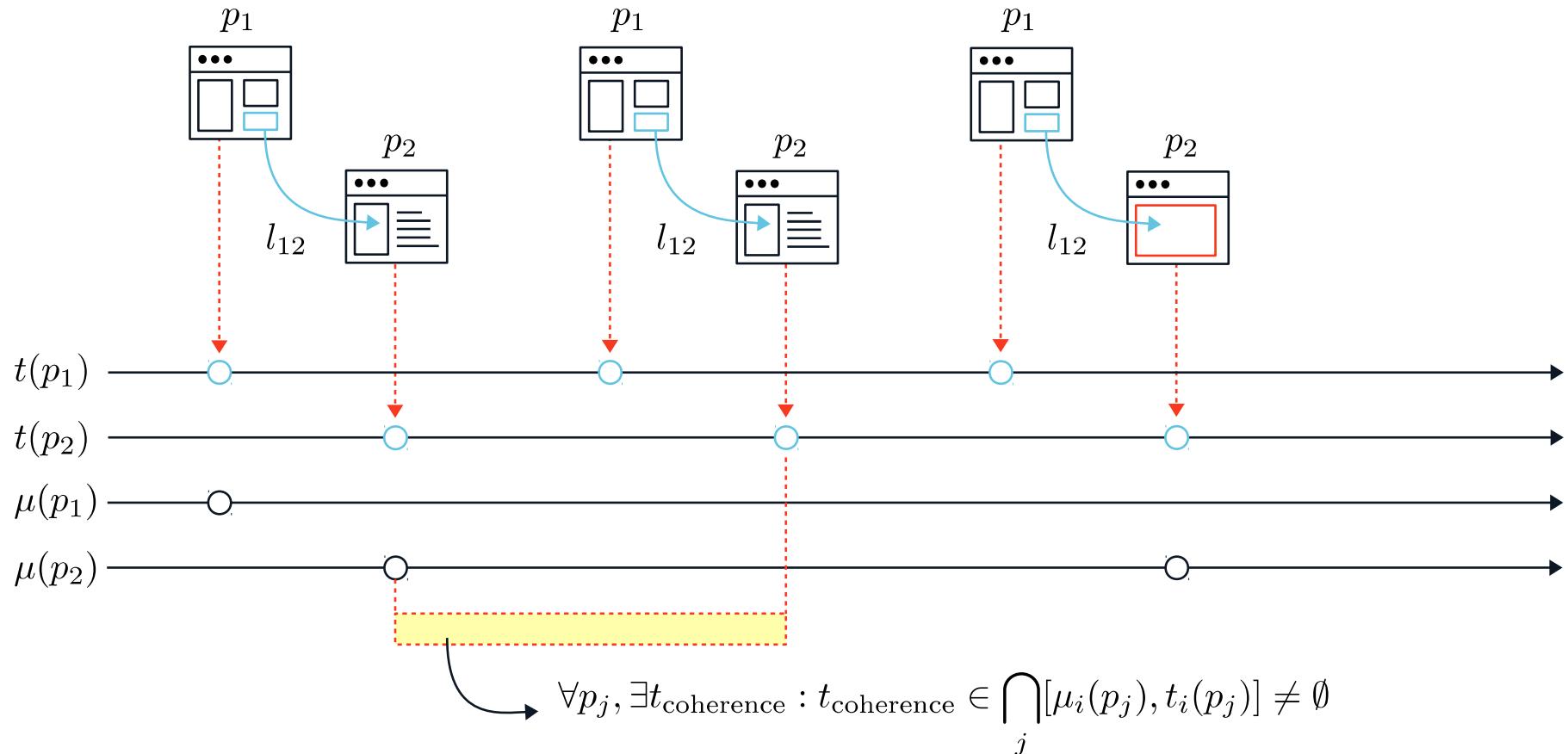


Certains changements subis par une page Web peuvent ne pas être captés par les crawlers

création
suppression
modification de contenu, de liens, etc.

Comprendre les artefacts de crawl (2/4)

> Cohérence entre pages archivées

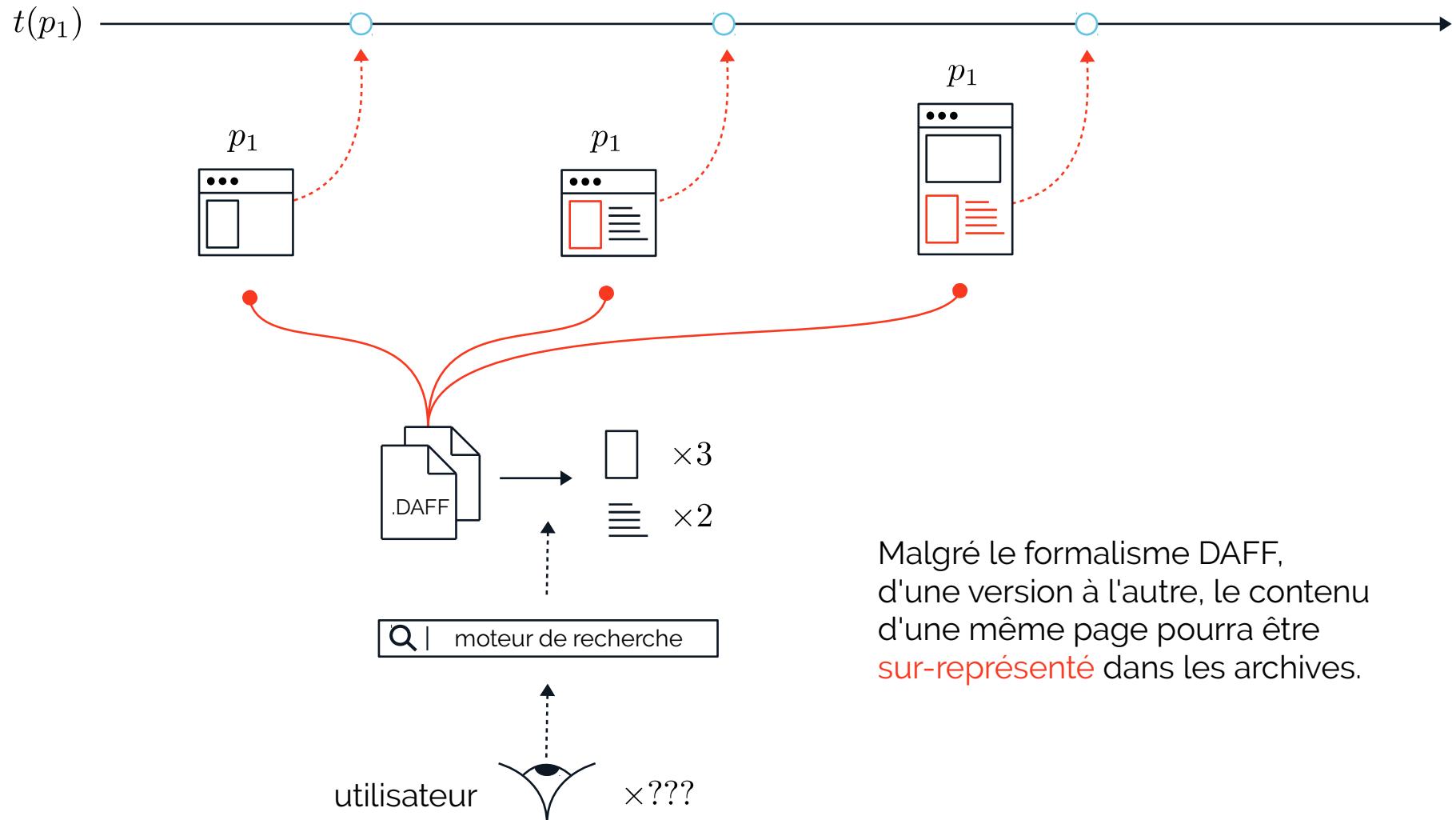


Le lien l_{12} a-t-il toujours la même signification ?

Ces pages restent-elles cohérentes ?

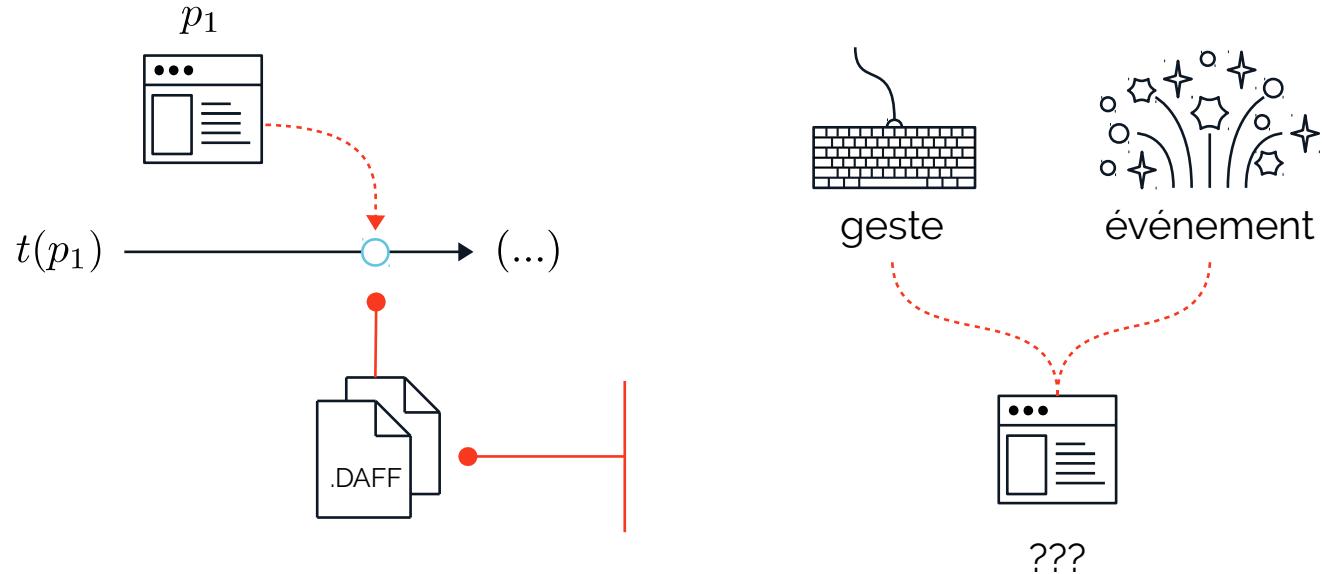
Comprendre les artefacts de crawl (3/4)

> Contenu dupliqué



Comprendre les artefacts de crawl (4/4)

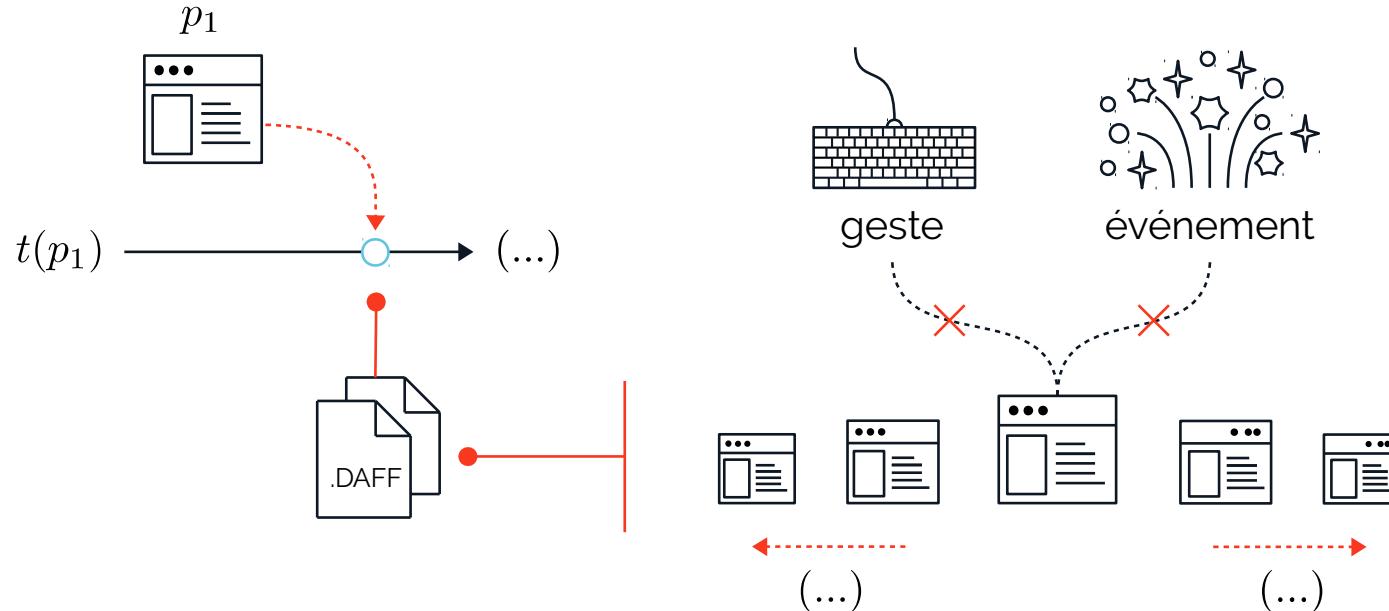
> Rupture de trace indiciaire



Toute archive est supposée être la **preuve** d'un geste, d'une activité, d'un événement, etc.
Elle en conserve la trace.

Comprendre les artefacts de crawl (4/4)

> Rupture de trace indiciaire



Toute archive est supposée être la **preuve** d'un geste, d'une activité, d'un événement, etc.
Elle en conserve la trace.

Or l'archivage du Web ne se définit que d'un point de vue technique, comme la complétion d'une collection de pages, **sans** établir de **lien organique** avec leurs contenus.

Plan de résolution

> Vers le fragment Web

1. Les représentations en ligne des diasporas
2. Archiver le Web
- 3. Introduction au fragment Web**
4. Les archives de la blogosphère marocaine
5. Les archives de yabiladi.com
6. Bilan et conclusion

Vers une exploration désagrégée (1/3)

> Face aux limites induites par les techniques d'archivages

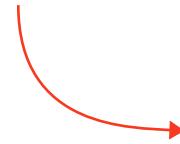
Il nous faut définir de nouveaux objectifs :

Vers une exploration désagrégée (1/3)

> Face aux limites induites par les techniques d'archivages

Il nous faut définir de nouveaux objectifs :

- dépasser la structure des archives Web



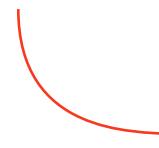
pour s'affranchir des crawlers

Vers une exploration désagrégée (1/3)

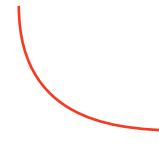
> Face aux limites induites par les techniques d'archivages

Il nous faut définir de nouveaux objectifs :

- dépasser la structure des archives Web

 pour s'affranchir des crawlers

- retrouver la temporalité du Web tel qu'il a été

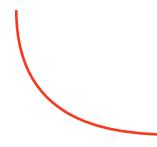
 pour suivre la trace des acteurs du Web passé

Vers une exploration désagrégée (1/3)

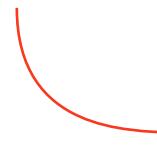
> Face aux limites induites par les techniques d'archivages

Il nous faut définir de nouveaux objectifs :

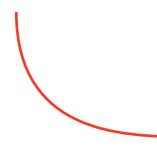
- dépasser la structure des archives Web

 pour s'affranchir des crawlers

- retrouver la temporalité du Web tel qu'il a été

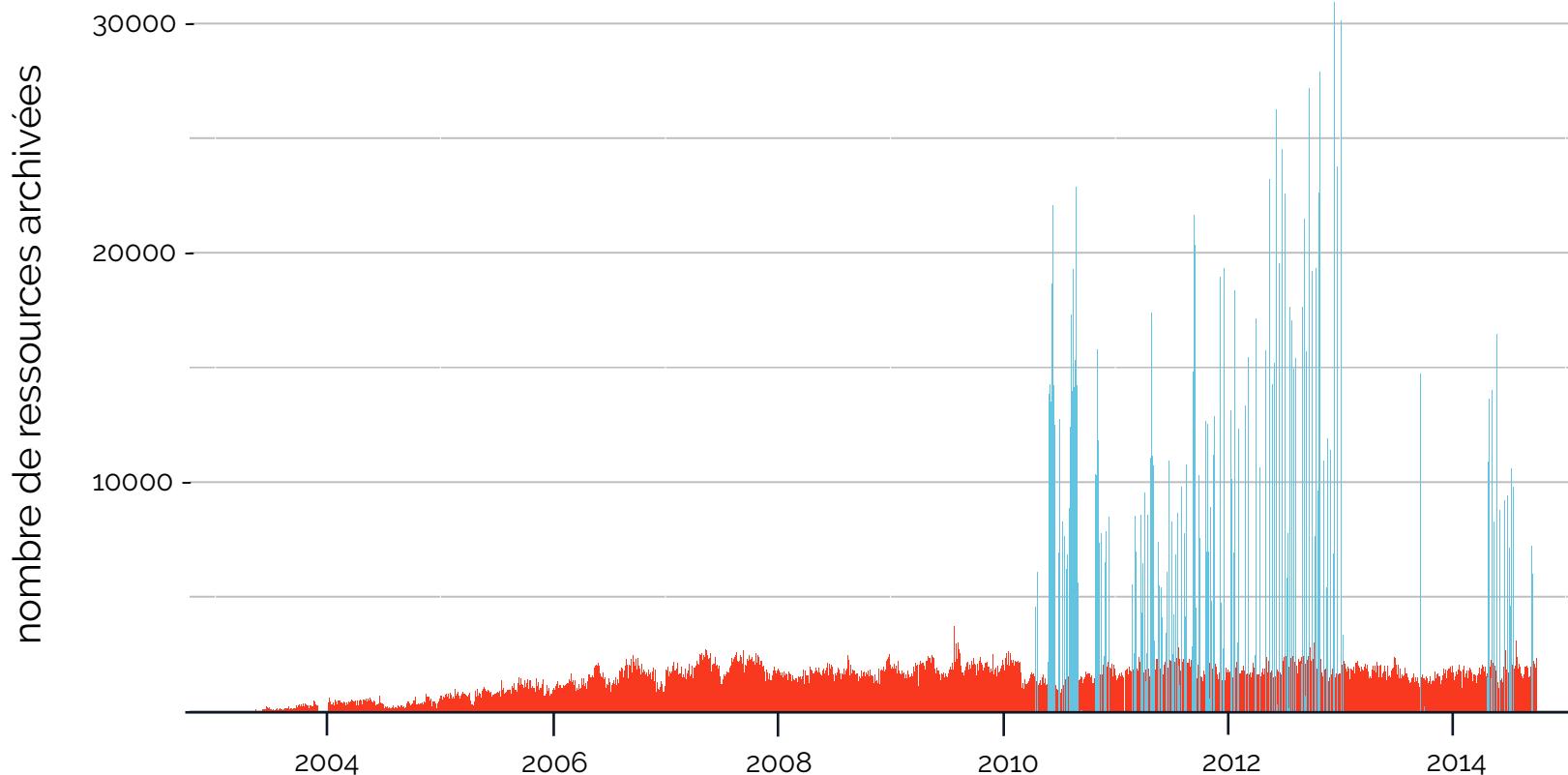
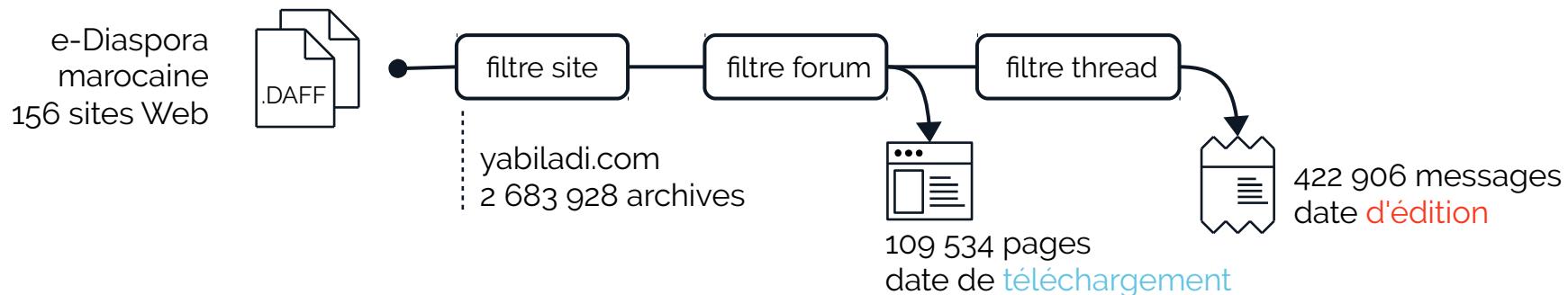
 pour suivre la trace des acteurs du Web passé

- retourner aux gestes, aux événements, aux activités

 pour faire s'asseoir les chercheurs à la table de montage

Vers une exploration désagrégée (2/3)

> Poursuivons notre expérience



Vers une exploration désagrégée (3/3)

> En dessous des pages archivées

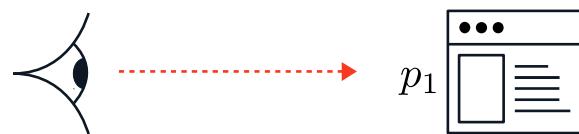
Pour mener à bien une exploration à grande échelle du Web passé.

Nous proposons d'introduire une **nouvelle unité d'exploration** des corpus d'archives Web.

Le fragment Web (1/3)

> Définition

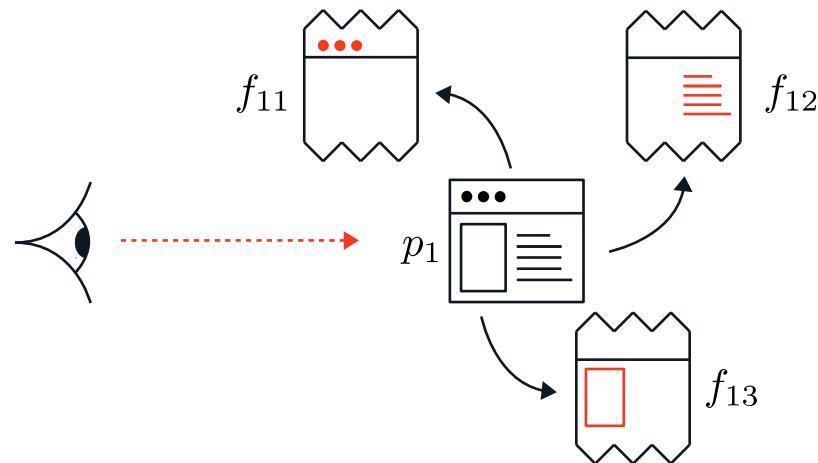
Considérant que du point de vue de la **perception humaine** une page Web est le résultat de l'agencement logique d'éléments distincts.



Le fragment Web (1/3)

> Définition

Considérant que du point de vue de la **perception humaine** une page Web est le résultat de l'agencement logique d'éléments distincts.

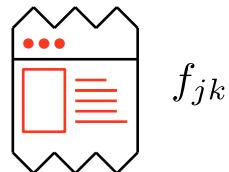


Nous appelons **fragment Web**, un sous-ensemble sémantique et syntaxique d'une page Web.

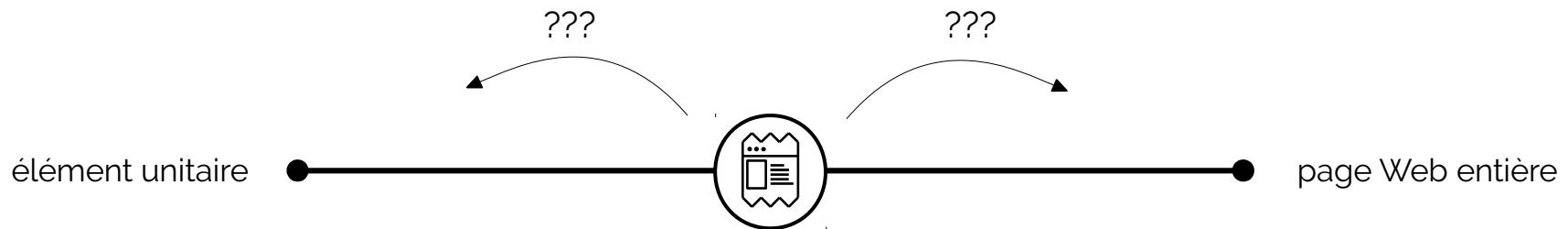
Le fragment Web (2/3)

> Définition

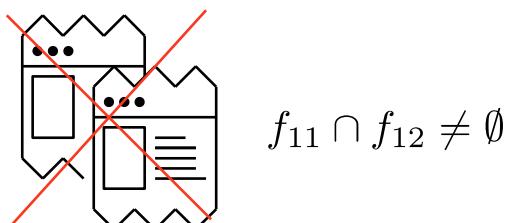
C'est un ensemble cohérent et autosuffisant d'élément textuels, visuels, sonores ou logiciels.



Il y a une relation d'échelle entre une page Web et ses fragments Web.



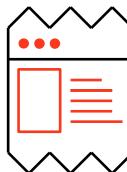
Au sein d'une même page Web, deux fragments Web ne peuvent pas se superposer.



Le fragment Web (3/3)

> Définition

Certains éléments d'un fragment Web peuvent faire l'objet d'une catégorisation.



f_{jk}

Y-a-t-il un titre ? Le nom d'un auteur ?
Ou une date d'édition ?

Unité	Date	Notation
page	crawl	$t(c_i)$
page	téléchargement	$t_i(p_j)$
page	dernière modification	$\mu_i(p_j)$
page	création	$\min_k \phi(f_{jk})$
fragment	édition	$\phi(f_{jk})$

precision historique

↓

Le fragment Web capture l'ensemble des dispositifs d'écriture et de partage de contenu.



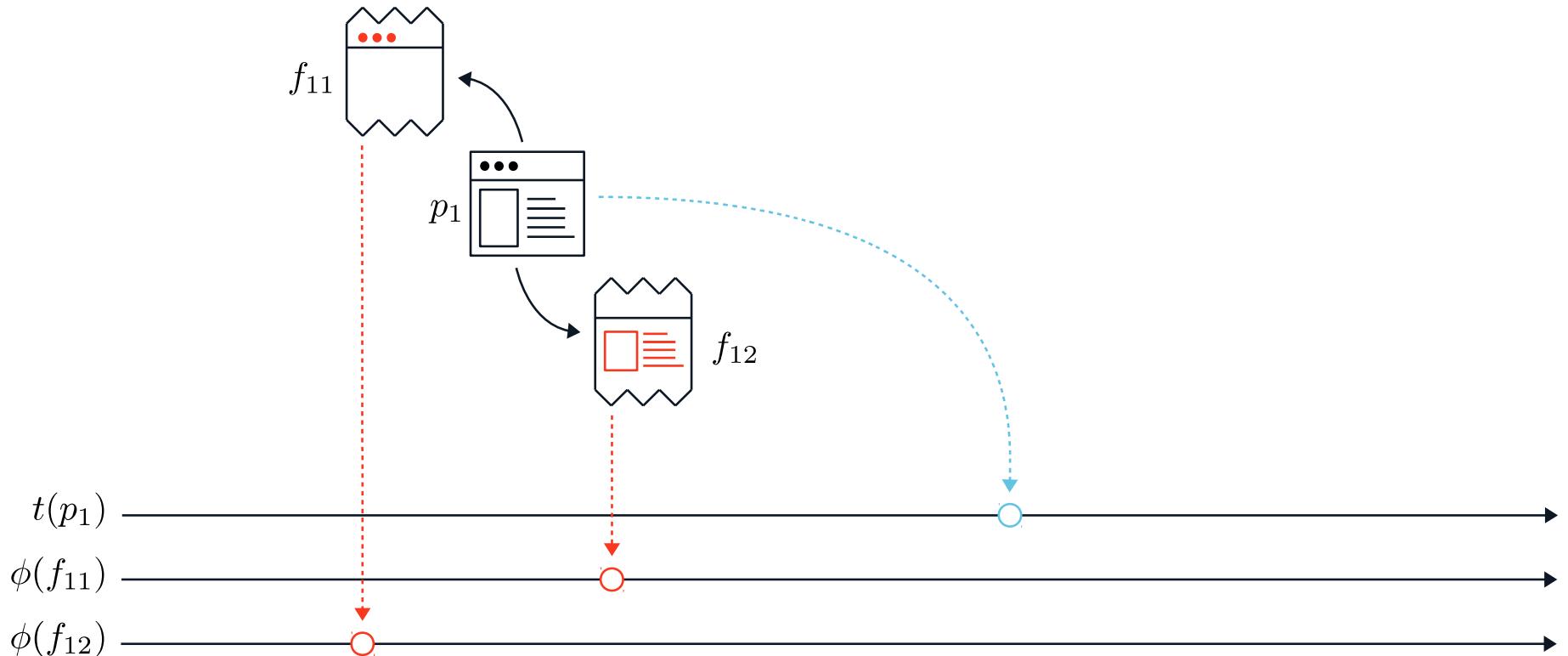
f_{jk}

Y-a-t-il un lien hypertexte ? Des widgets CMS ?
Un lien de syndication RSS ?

Penser une exploration désagrégée (1/3)

> Atténuer les cécités de crawl

En supposant que les **dates d'édition** sont fiables, alors $\forall p_j, f_{jk} \exists \phi(f_{jk}) : \phi(f_{jk}) \leq \mu_i(p_j) \leq t_i(p_j)$



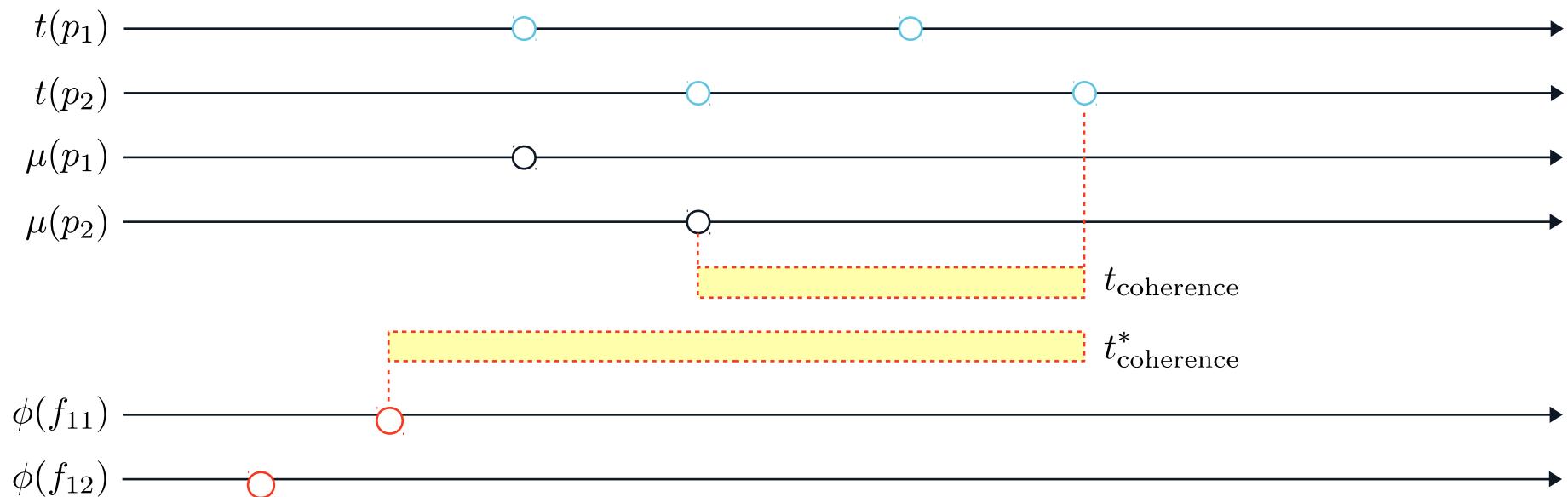
Pour yabiladi.com, les quartiles de $t_i(p_j) - \phi(f_{jk})$ en jours sont : (Q1) 256, (Q2) 777, (Q3) 1340 et il est possible de remonter ainsi jusqu'en 2003.

Penser une exploration désagrégée (2/3)

> Cohérence relative entre pages

Le chercheur peut **sélectionner** un ensemble de fragments Web d'intérêts tels que :

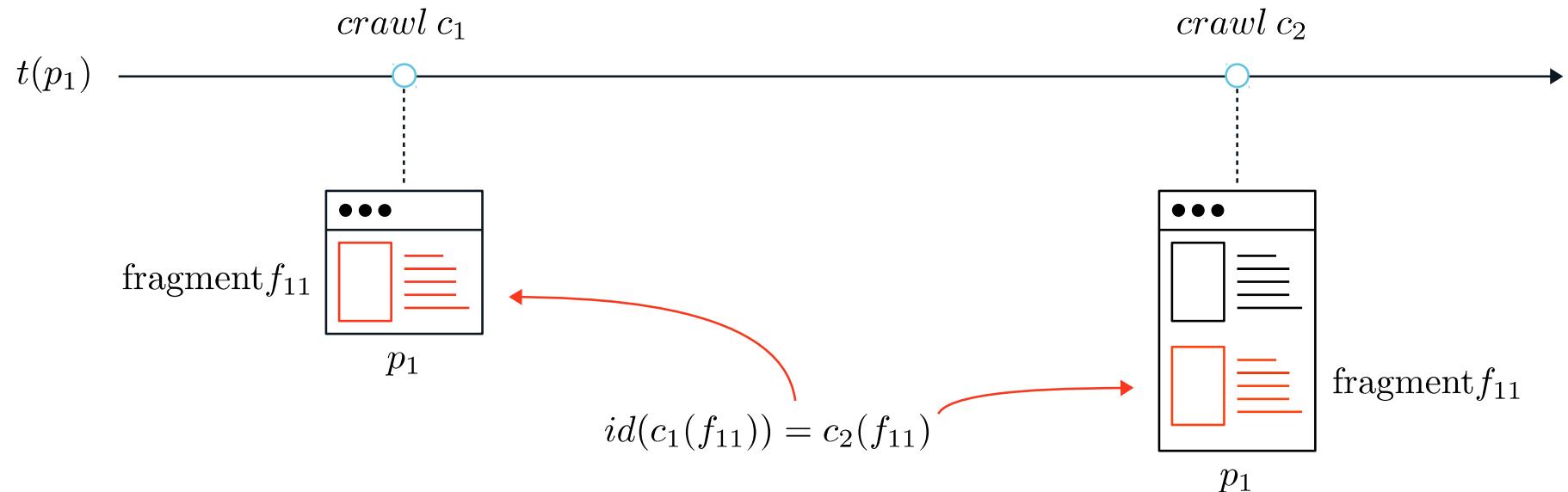
$$\forall p_j, \forall f_{jk}^* \in \{f_{j1}, \dots, f_{jm}\}, \exists t_{\text{coherence}}^* : t_{\text{coherence}}^* \in \bigcap_j [\phi(f_{jk}^*), t_i(p_j)] \neq \emptyset$$



Le fragment Web introduit une cohérence **plus souple** entre deux pages archivées en la définissant relativement à une question de recherche donnée

Penser une exploration désagrégée (3/3)

> Dé dupliquer les corpus

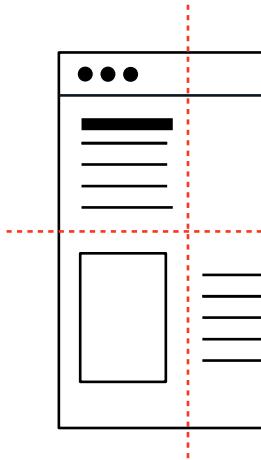


Il est possible de dé dupliquer des fragments en testant l'égalité d'une **clé de hachage sha256**.

Pour yabiladi.com, les quartiles du nombre de fragments dupliqués : (Q1) 1, (Q2) 1, (Q3) 2, (Max) 44

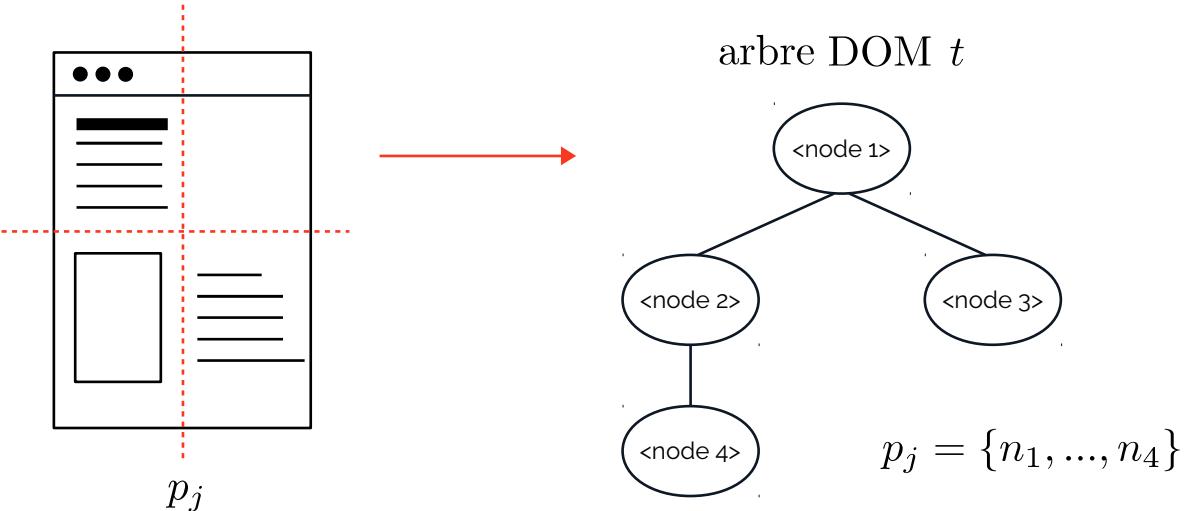
Trouver les fragment Web (3/3)

> Technique et méthodologie d'extraction



Trouver les fragment Web (3/3)

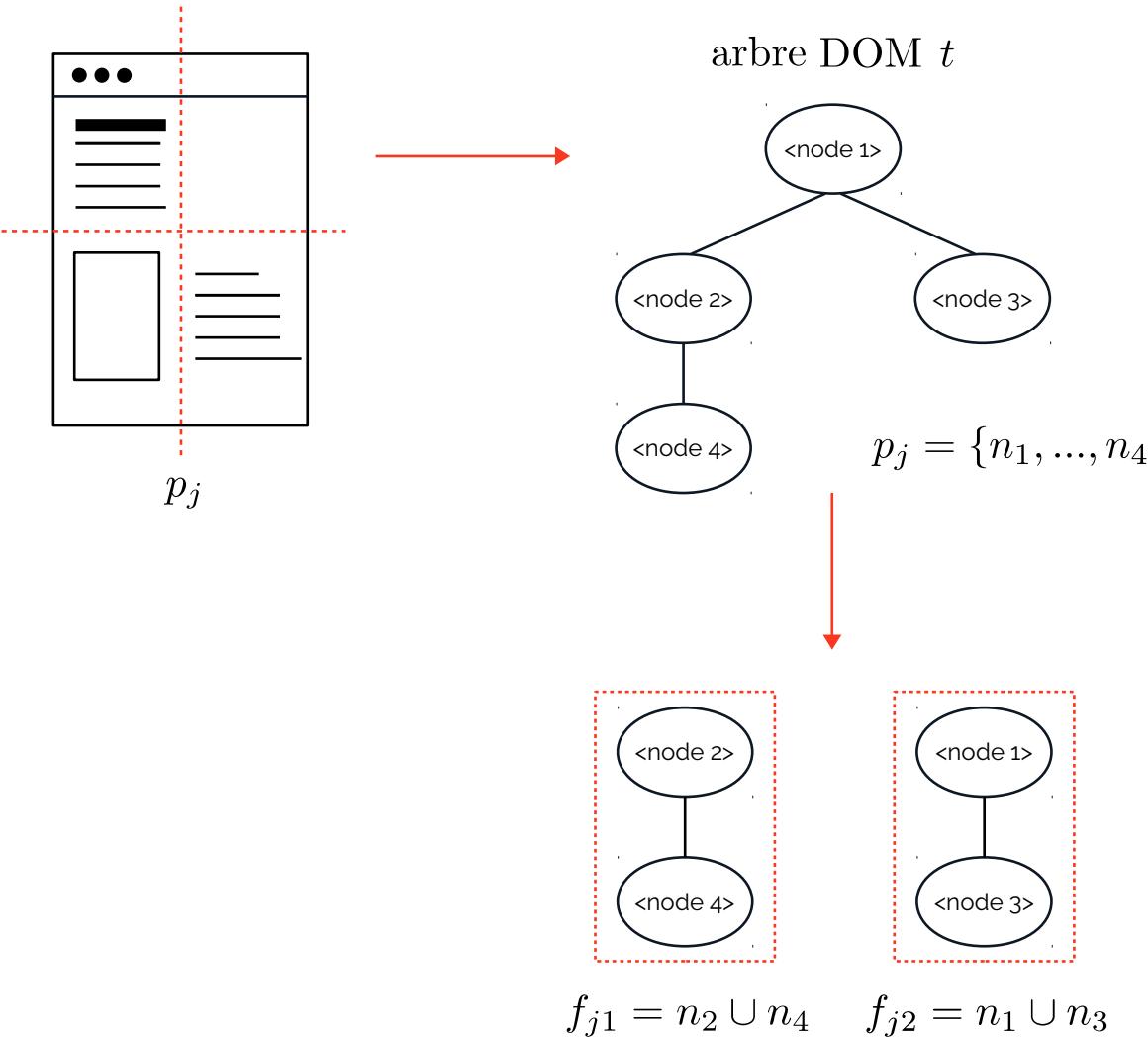
> Technique et méthodologie d'extraction



Trouver des **groupes de noeuds** HTML proches (clusters) en utilisant Readability et Fathom.

Trouver les fragments Web (3/3)

> Technique et méthodologie d'extraction

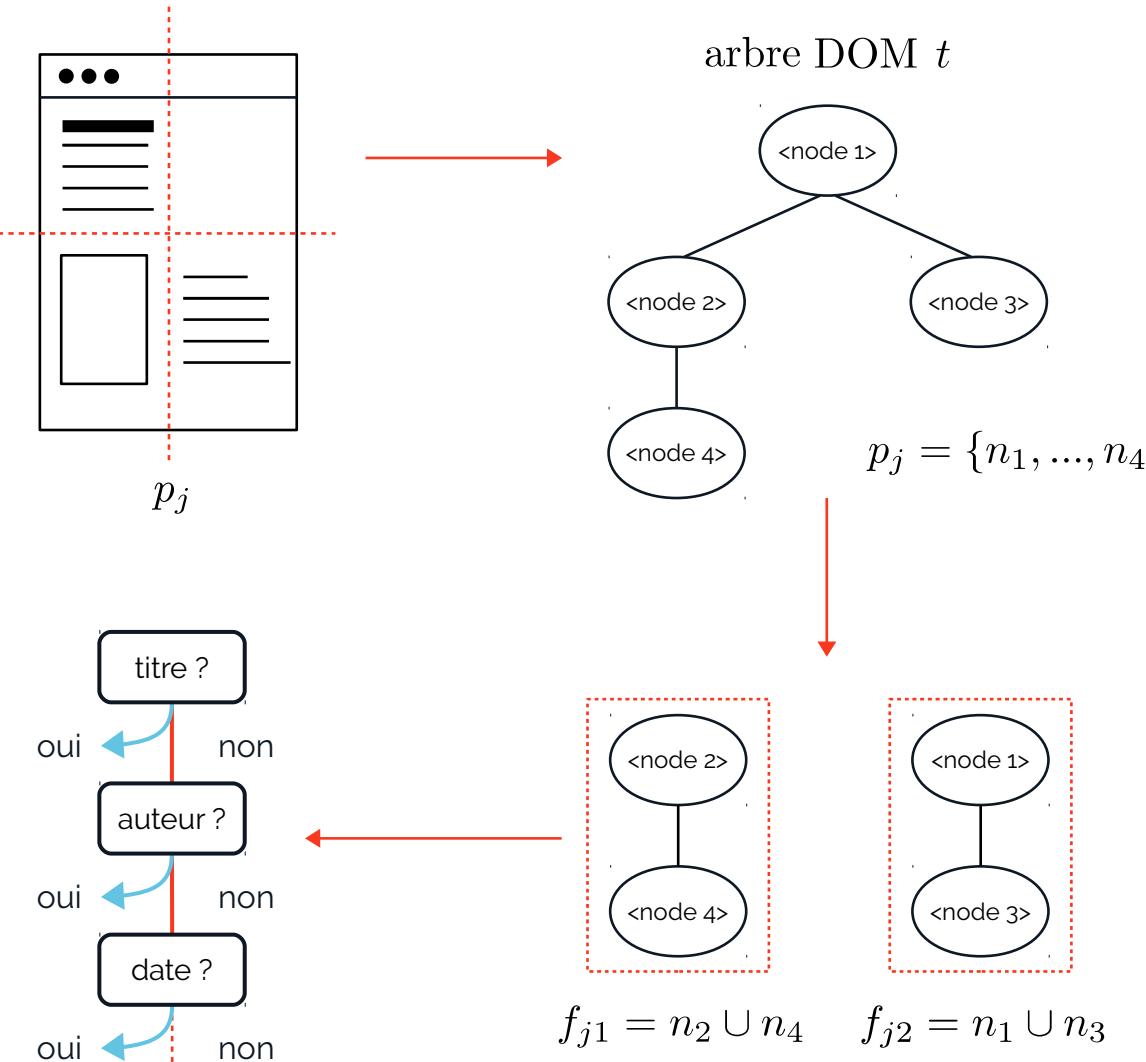


Trouver des **groupes de noeuds** HTML proches (clusters) en utilisant Readability et Fathom.

La **fonction de distance** est basée sur des malus visuels et sémantiques, ainsi que sur des heuristiques ad-hoc.

Trouver les fragment Web (3/3)

> Technique et méthodologie d'extraction

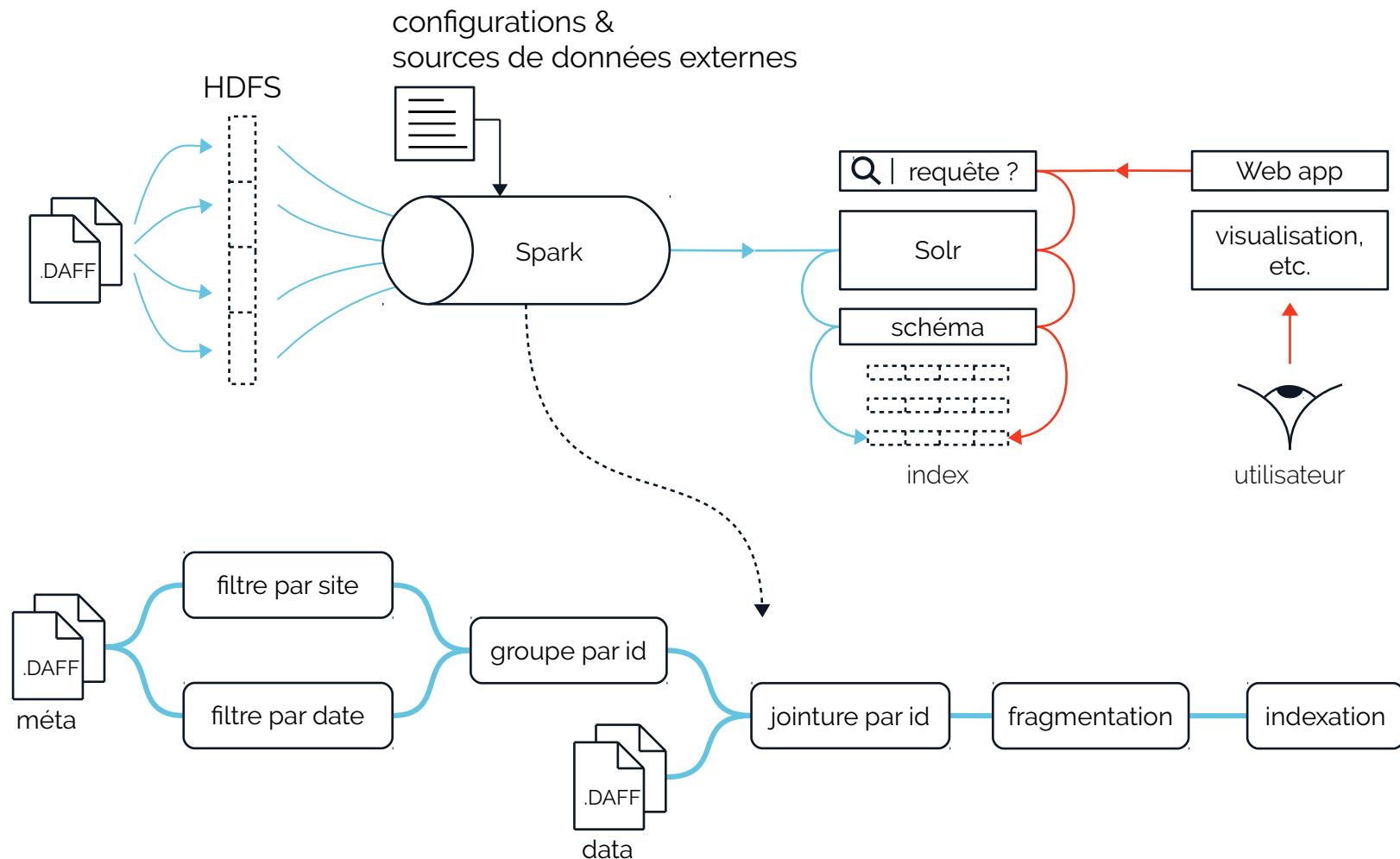


Trouver des **groupes de noeuds** HTML proches (clusters) en utilisant Readability et Fathom.

La **fonction de distance** est basée sur des malus visuels et sémantiques, ainsi que sur des heuristiques ad-hoc.

Construire un moteur d'exploration

> De l'archive à la visualisation



Plan de résolution

> Vers les archives de la blogosphère

1. Les représentations en ligne des diasporas
2. Archiver le Web
3. Introduction au fragment Web
4. **Les archives de la blogosphère marocaine**
5. Les archives de yabiladi.com
6. Bilan et conclusion

Retour à la blogosphère

> Formulons des hypothèses

En 2015, certains étudiants marocains nous indiquaient préférer les groupes Facebook aux blogs migrants précédemment cartographiés.



Retour à la blogosphère

> Formulons des hypothèses

En 2015, certains étudiants marocains nous indiquaient préférer les groupes Facebook aux blogs migrants précédemment cartographiés.



Facebook ? Twitter ? Etc.

1. des widgets

7didane ailleurs

Sur Gmail :
Sur twitter : [clique-ici](#)
Sur Facebook : [clique-ici](#)
etc...

```
<ul><li> <span>Sur twitter :</span>
    <span style="font-size: 11px; font-style: italic;">
        <a href="https://twitter.com/7didane">
            clique-ici</a></span><br></li>
<li><span>Sur Facebook :</span>
    <span style="font-size: 11px; font-style: italic;">
        <a href="https://www.facebook.com/7didane">
            clique-ici</a></span><br></li></ul>
```



```
<div class="textwidget">
<p style="text-align:center;">
<a class="twitter" href="http://twitter.com/AnasAlaoui"
    title="Suivez-moi sur Twitter - @AnasAlaoui"
    target="_blank">
```

2. des articles

larbi.org, 14 février 2011

« Non, le printemps arabe ne peut pas s'arrêter aux frontières du maroc - Revivez la journée de manifestations du 20 Février »
Morocco #Feb20 Maroc #20Fev lundi, février 14 2011
Morocco #Feb20 Maroc #20Fev
En direct de Tunisie 14/01/2011
En direct de Tunisie La révolution en Marche
En direct de Twitter Join the conversation

7didane.org, 16 juin 2009

Jun 16
L'épiphénomène **Twitter** (droit de réponse)
Exactement une semaine après mon billet sur **Twitter** (L'épiphénomène **Twitter**), et suite à une intervention de la maison blanche, le site me répond personnellement sur leur blog. Voici le texte ...

Retour à la blogosphère

> Formulons des hypothèses

En 2015, certains étudiants marocains nous indiquaient préférer les groupes Facebook aux blogs migrants précédemment cartographiés.



1. des widgets

A screenshot of a blog sidebar. At the top, it says "7didane ailleurs". Below that, there's a list of links: "Sur Gmail : clique-ici", "Sur twitter : clique-ici", "Sur Facebook : clique-ici", and "etc...".

```
<ul><li> <span>Sur twitter :</span>
    <span style="font-size: 11px; font-style: italic;">
        <a href="https://twitter.com/7didane">
            clique-ici</a></span><br></li>
<li><span>Sur Facebook :</span>
    <span style="font-size: 11px; font-style: italic;">
        <a href="https://www.facebook.com/7didane">
            clique-ici</a></span><br></li></ul>
```

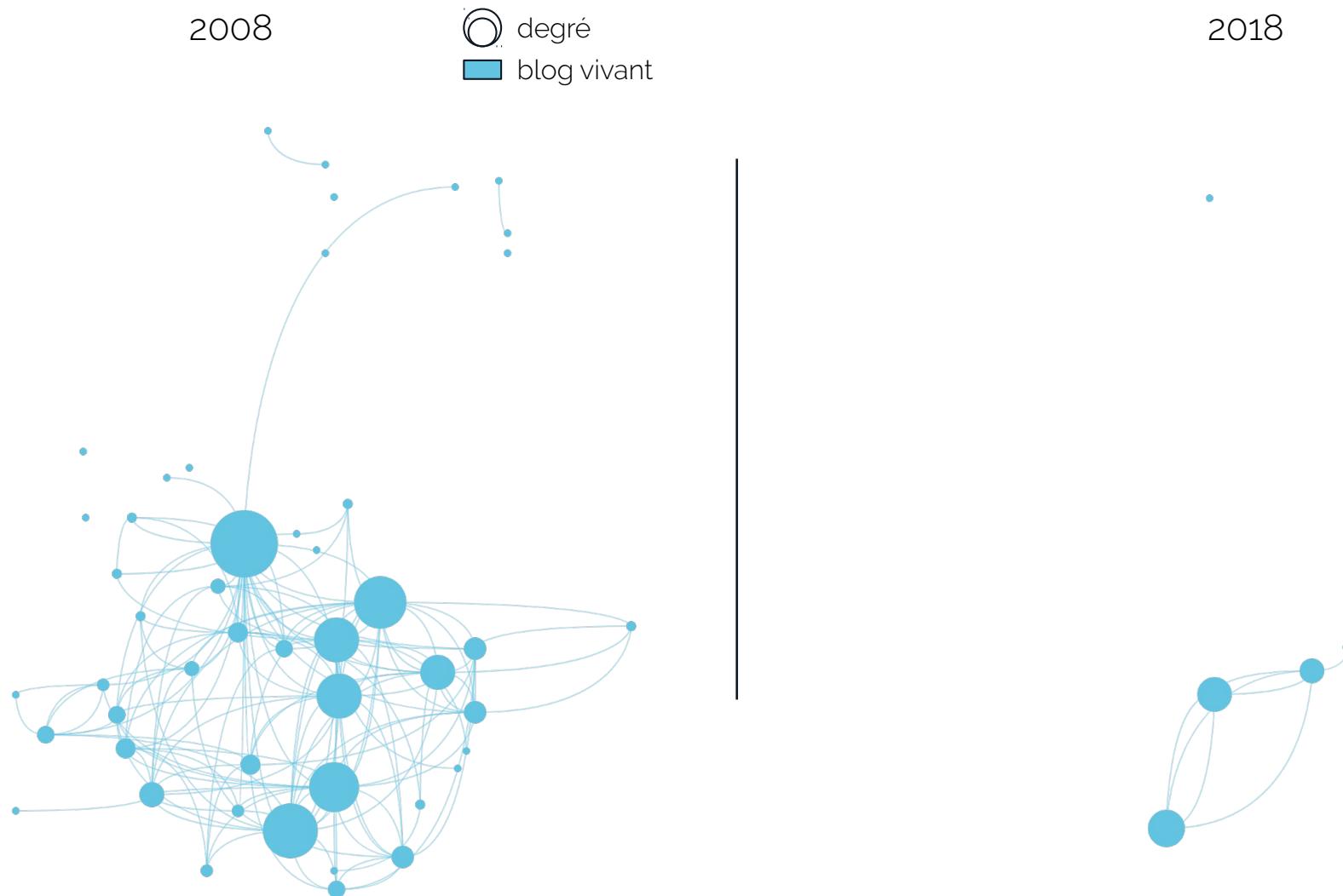


```
<div class="textwidget">
<p style="text-align:center;">
<a class="twitter" href="http://twitter.com/AnasAlaoui"
    title="Suivez-moi sur Twitter - @AnasAlaoui"
    target="_blank">
```

Des **avatars** ou des **pseudonymes** nous redirigent vers les divers comptes sociaux liés aux blogs migrants.

Les traces archivées d'une migration numérique (2/3)

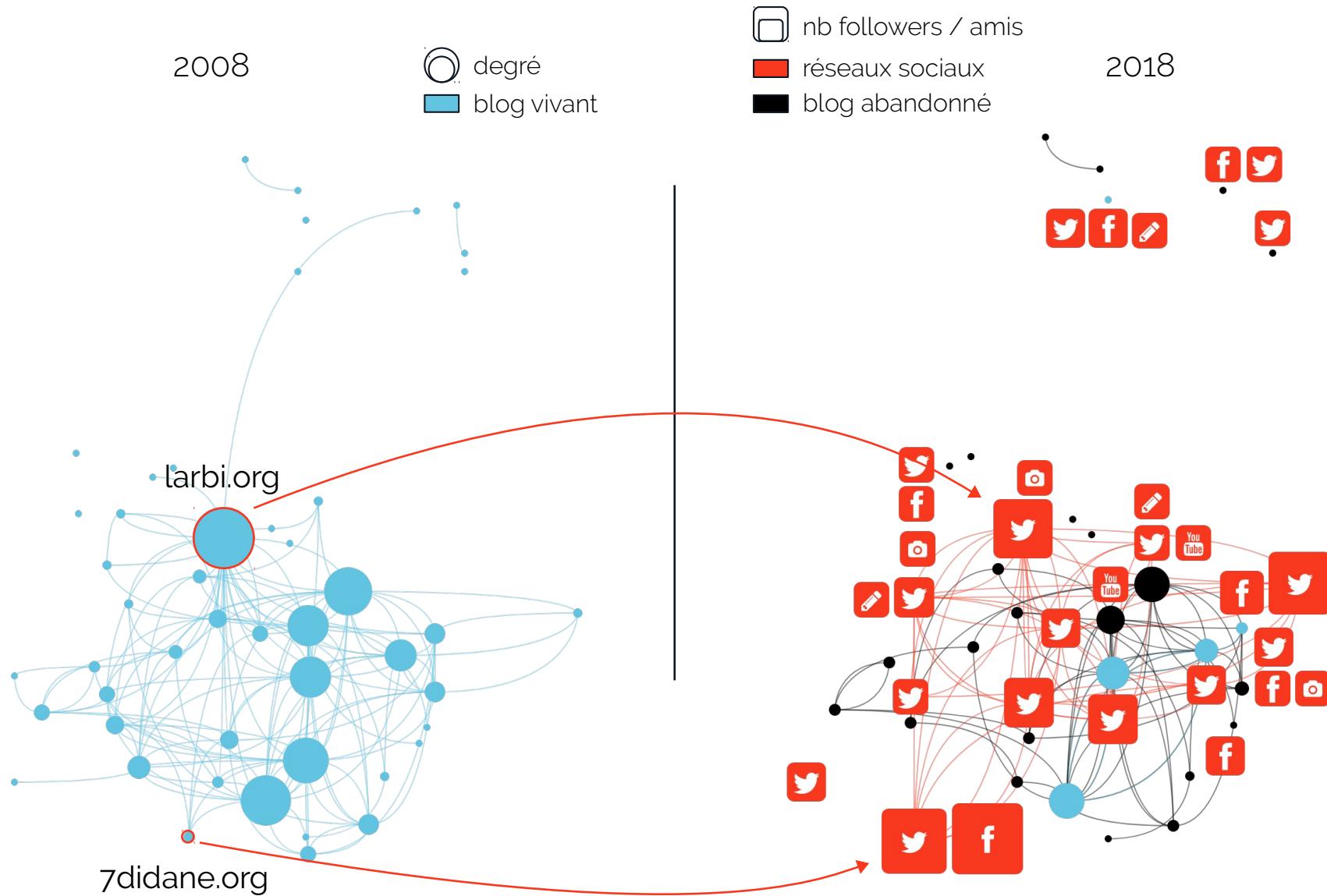
> Reconstruire un collectif éteint



> La blogosphère marocaine (focus)

Les traces archivées d'une migration numérique (2/3)

> Reconstruire un collectif éteint



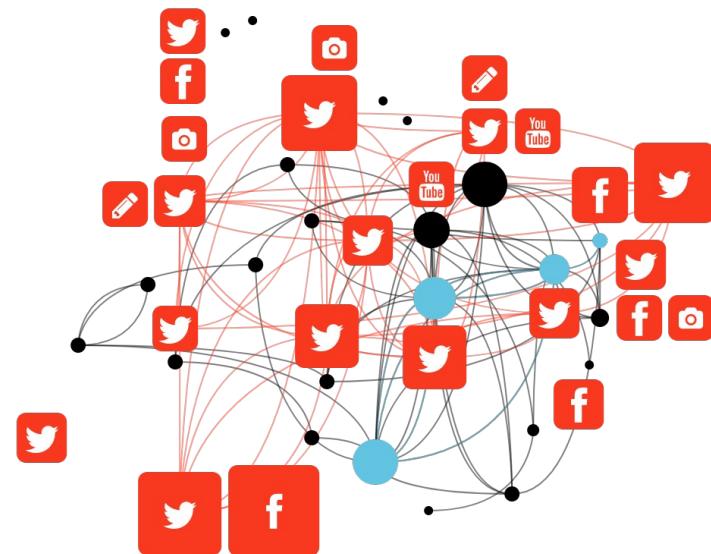
Les traces archivées d'une migration numérique (2/3)

> Reconstruire un collectif éteint



Les auteurs conservent leurs **pseudonymes** malgré la fermeture des blogs.

Les **moyens d'expression** en ligne sont multipliés et **spécialisés** par type de médium ou de communauté.



Les traces archivées d'une migration numérique (2/3)

> Reconstruire un collectif éteint

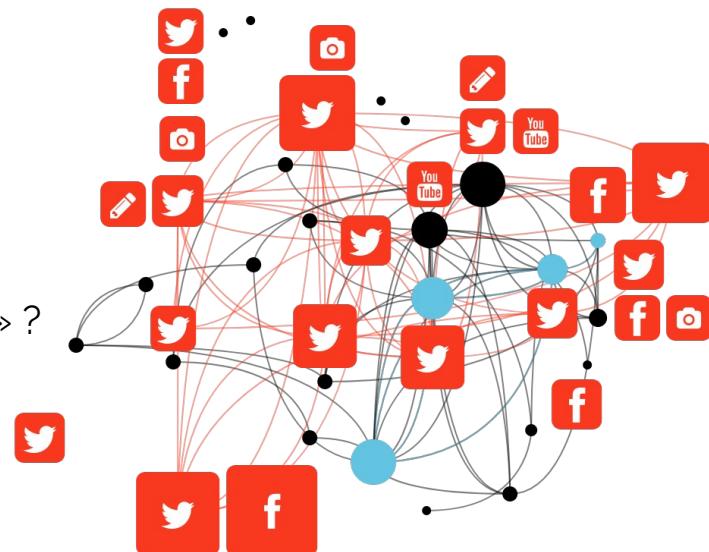


L'aspect **communauté** de l'ancienne blogosphère est conservé.

$$D = \frac{2 \times |E|}{|V| \times (|V| - 1)}$$

densité blogs (2008)
0,16
↓
0,24
densité twitter (2018)

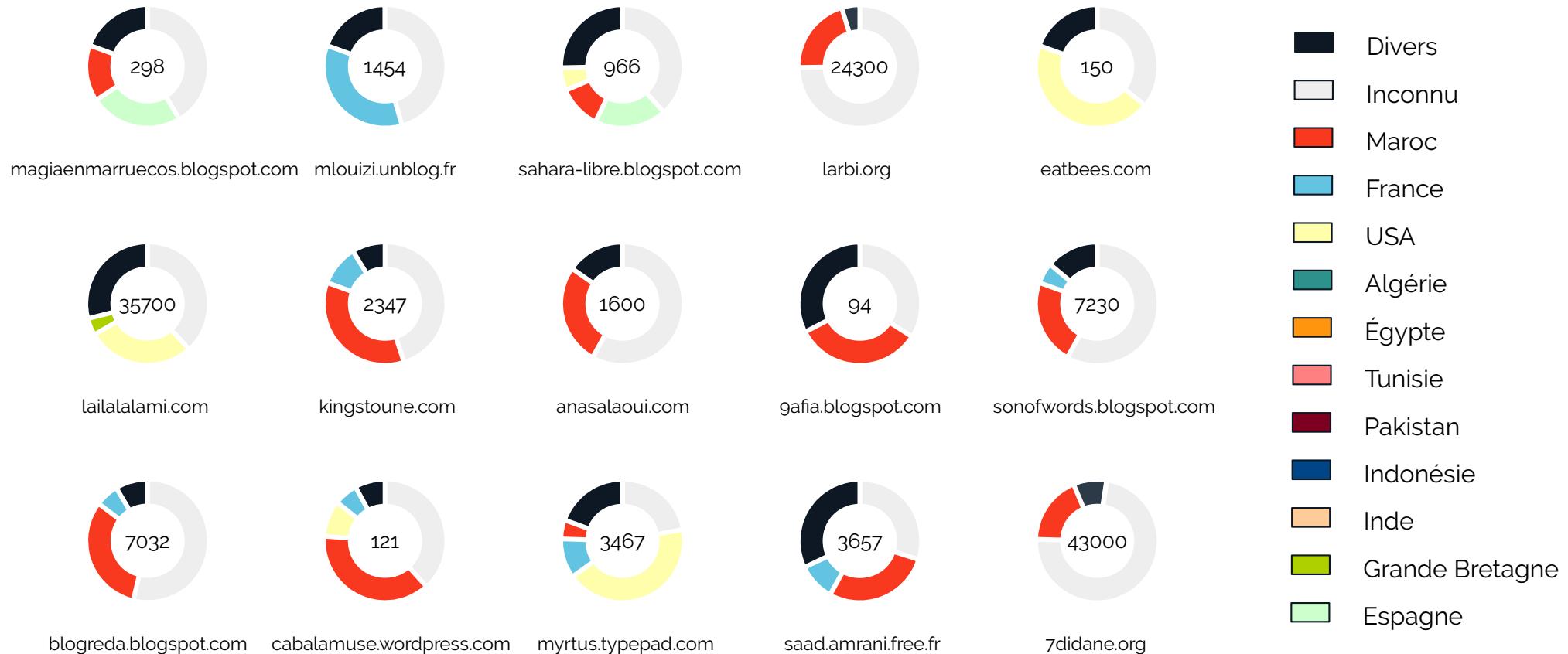
Mais la multiplication des liens est-elle le seul fait des auteurs ou un **effet lié** à l'environnement «réseau social» ?



Les traces archivées d'une migration numérique (3/3)

> Une communauté de lecteurs

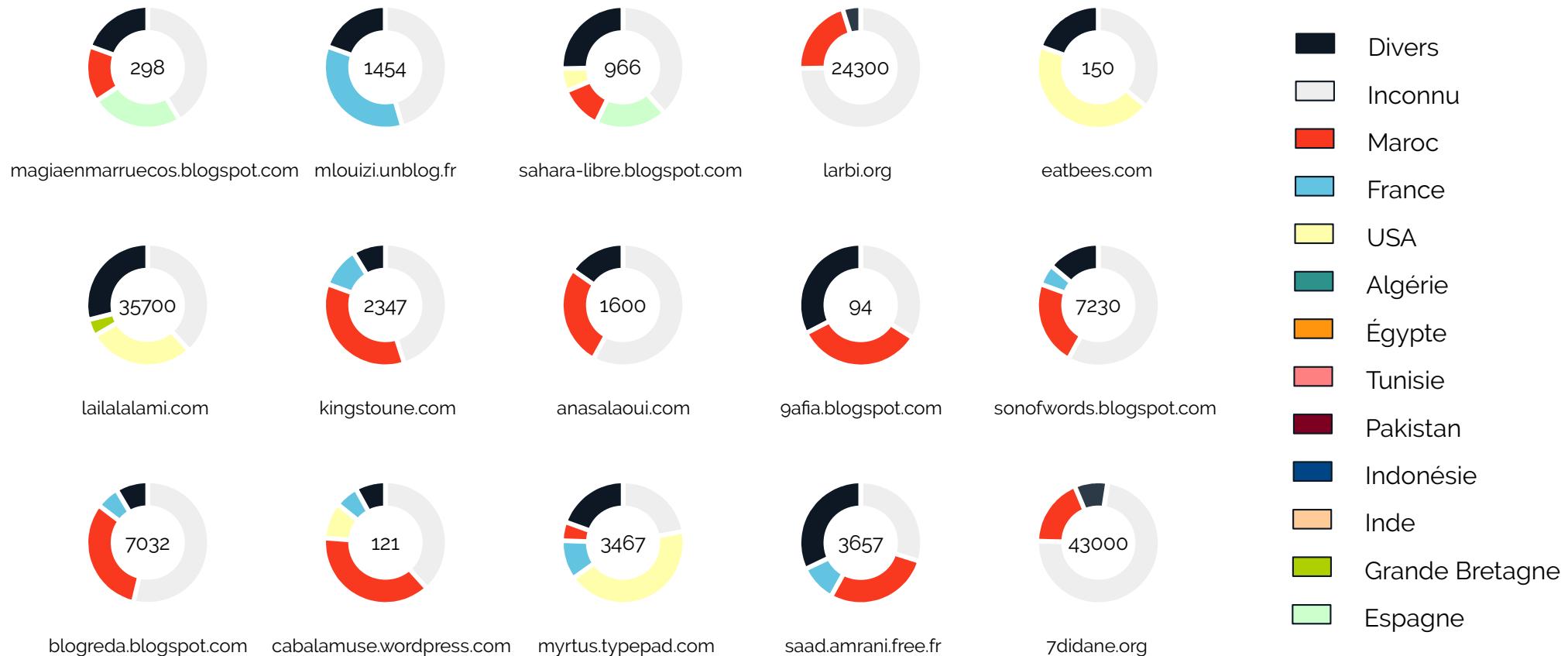
L'aspect **diasporique** de la blogosphère de 2008 est-il toujours présent en 2018 ?



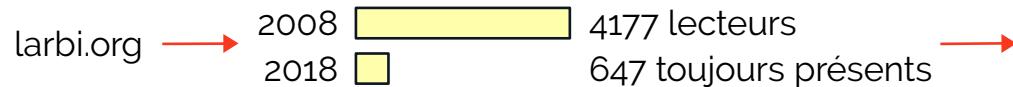
Les traces archivées d'une migration numérique (3/3)

> Une communauté de lecteurs

L'aspect **diasporique** de la blogosphère de 2008 est-il toujours présent en 2018 ?



Les lecteurs ont-ils **suivis** les auteurs dans leur migration ?



Borne inférieure avec hypothèse de conservation de pseudonymes

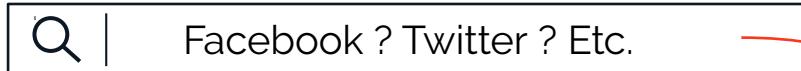
Plan de résolution

> Vers les archives de la yabiladi.com

1. Les représentations en ligne des diasporas
2. Archiver le Web
3. Introduction au fragment Web
4. Les archives de la blogosphère marocaine
- 5. Les archives de yabiladi.com**
6. Bilan et conclusion

Le Printemps arabe comme un moment clé ?

> Revenons sur un détail



des articles, peu nombreux
mais signifiants

larbi.org, 14 février 2011

« Non, le printemps arabe ne peut pas s'arrêter aux frontières du maroc - Revivez la journée de manifestations du 20 Février »
Morocco #Feb20 Maroc #20Fev lundi, février 14 2011
Morocco #Feb20 Maroc #20Fev
En direct de Tunisie 14/01/2011
En direct de Tunisie La révolution en Marche
En direct de Twitter Join the conversation

7didane.org, 16 juin 2009

Jun 16

L'épiphénomène Twitter (droit de réponse)

Exactement une semaine après mon billet sur Twitter (L'épiphénomène Twitter), et suite à une intervention de la maison blanche, le site me répond personnellement sur leur blog. Voici le texte :

Down Time Rescheduled (cliquez-ici)

A critical network upgrade must be performed to ensure continued operation of Twitter. In coordination with Twitter, our network host had planned this upgrade for tonight. However, our network partners at NTT America recognize the role Twitter is currently playing as an important communication tool in Iran. Tonight's planned maintenance has been rescheduled to tomorrow between 2-3p PST (1:30a in Iran).

Our partners are taking a huge risk not just for Twitter but also the other services they support worldwide—we commend them for being flexible in what is essentially an inflexible situation. We chose NTT America Enterprise Hosting Services early last year specifically because of their impeccable history of reliability and global perspective. Today's decision and actions continue to prove why NTT America is such a powerful partner for Twitter.

Hat off to the youth of iran !

Le Printemps arabe comme un moment clé ?

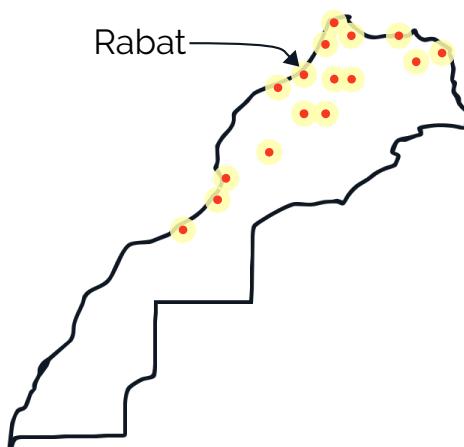
> Revenons sur un détail



des articles, peu nombreux mais signifiants

larbi.org, 14 février 2011

« Non, le printemps arabe ne peut pas s'arrêter aux frontières du maroc - Revivez la journée de manifestations du 20 Février »
Morocco #Feb20 Maroc #20Fev lundi, février 14 2011
Morocco #Feb20 Maroc #20Fev
En direct de Tunisie 14/01/2011
En direct de Tunisie La révolution en Marche
En direct de Twitter Join the conversation



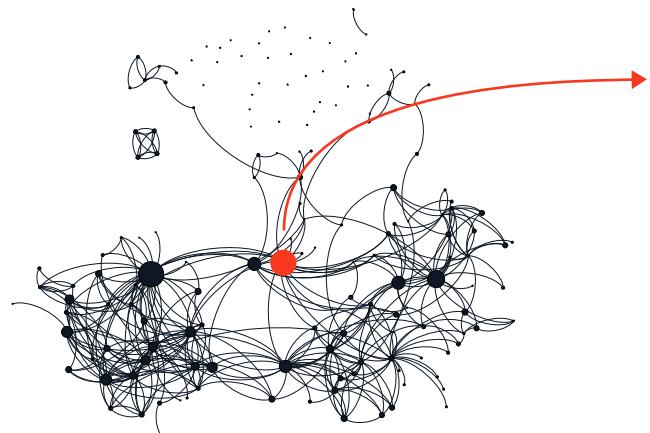
Le **20 février 2011**, un grand mouvement de protestation frappe le Maroc s'inscrivant dans la dynamique régionale du Printemps arabe.

Les manifestations ne sont pas spontanées mais **organisées en ligne** par un collectif : le **M2oF**

Cet événement a-t-il pu avoir un **impact** visible sur d'autres espaces de notre corpus ?

Yabiladi.com une porte d'entrée sur la diaspora

> Un site hybride et central



Un hub et un **pont** entre des constellations de sites qui ne se parlent pas.

Ya biladi signifie *Oh mon pays*.

Le site s'inscrit dans la tradition des **portails** diasporiques en ligne.



Yabiladi.com une porte d'entrée sur la diaspora

> Un site hybride et central

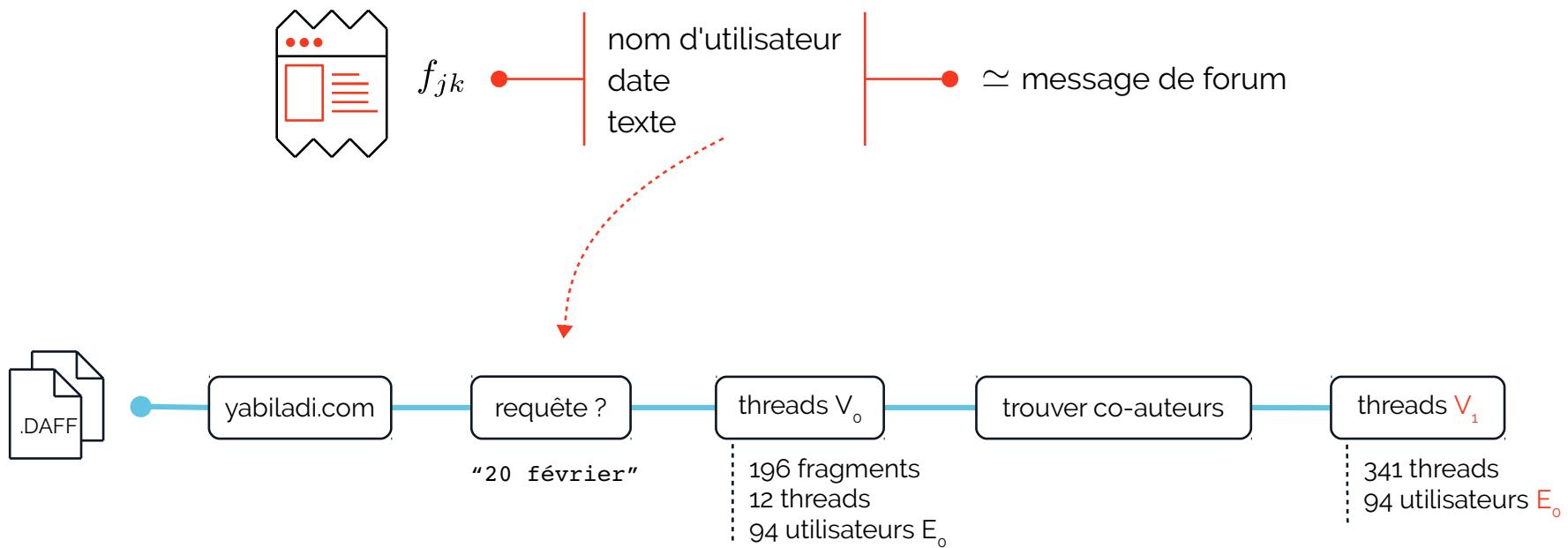
Peut-on extraire et voir le **réseau** des utilisateurs de yabiladi.com ayant parlé du 20 février 2011 ?



Yabiladi.com une porte d'entrée sur la diaspora

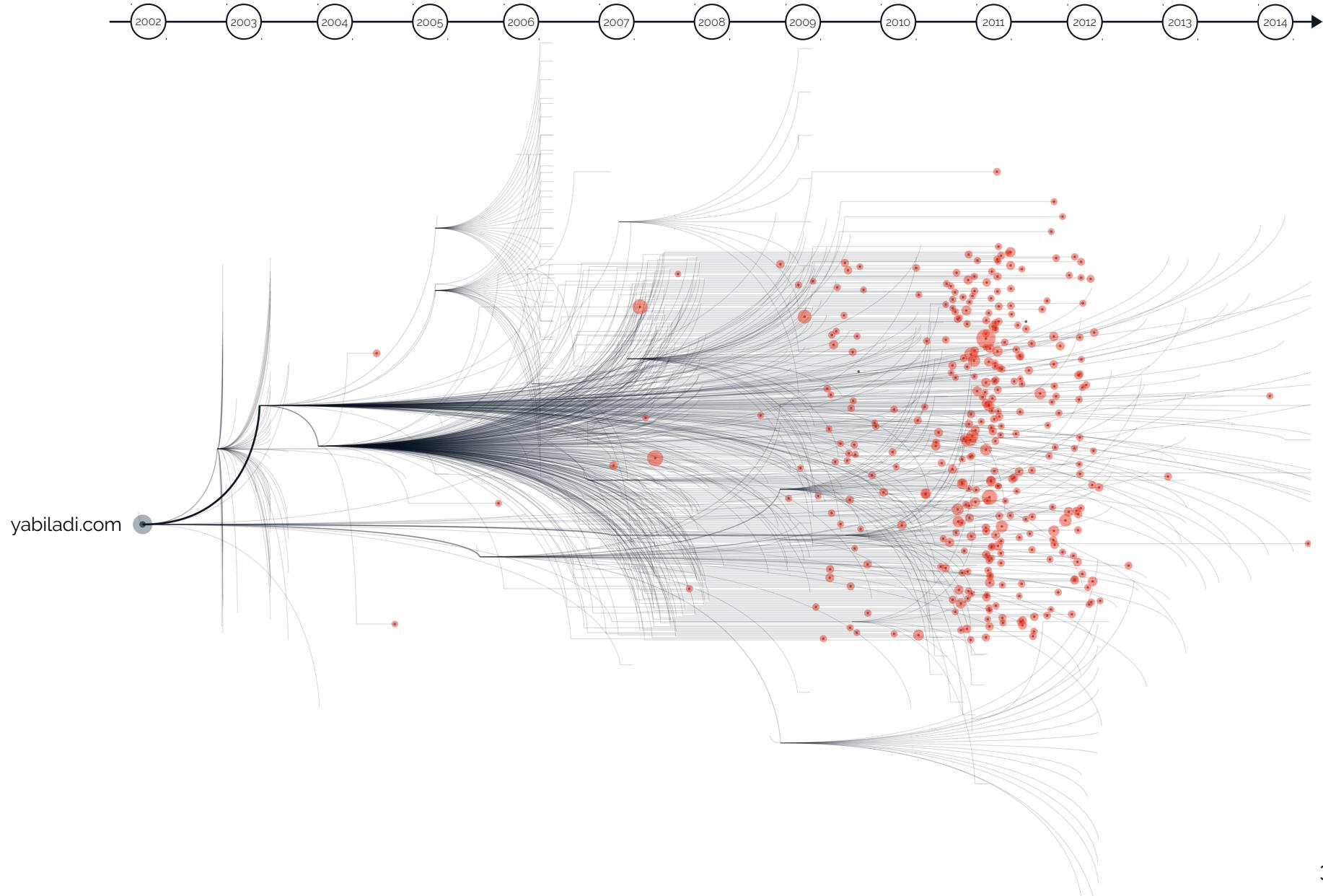
> Un site hybride et central

Peut-on extraire et voir le **réseau** des utilisateurs de yabiladi.com ayant parlé du 20 février 2011 ?



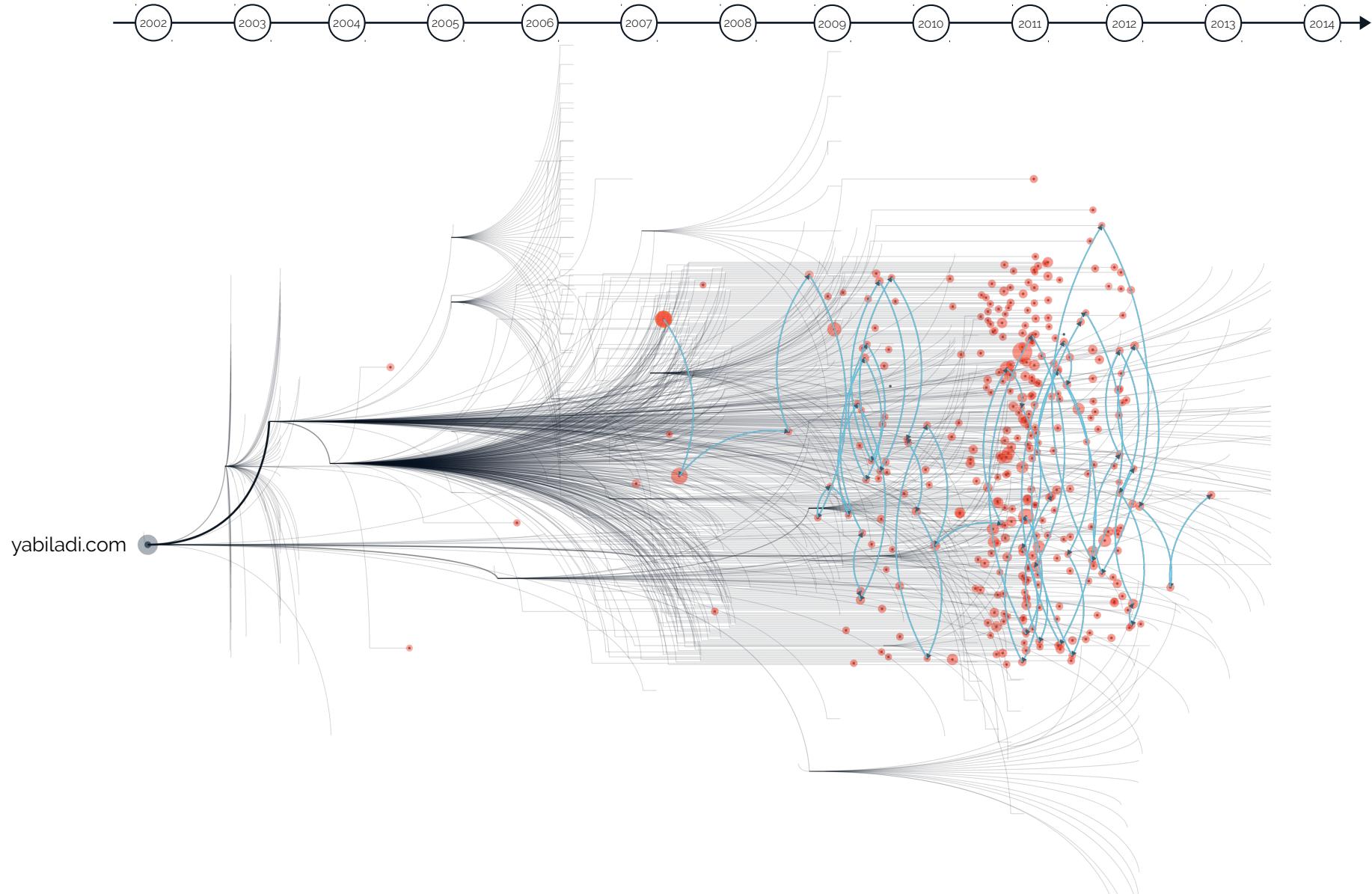
Un soulèvement en ligne éphémère (1/4)

> Les threads liés au 20 février 2011



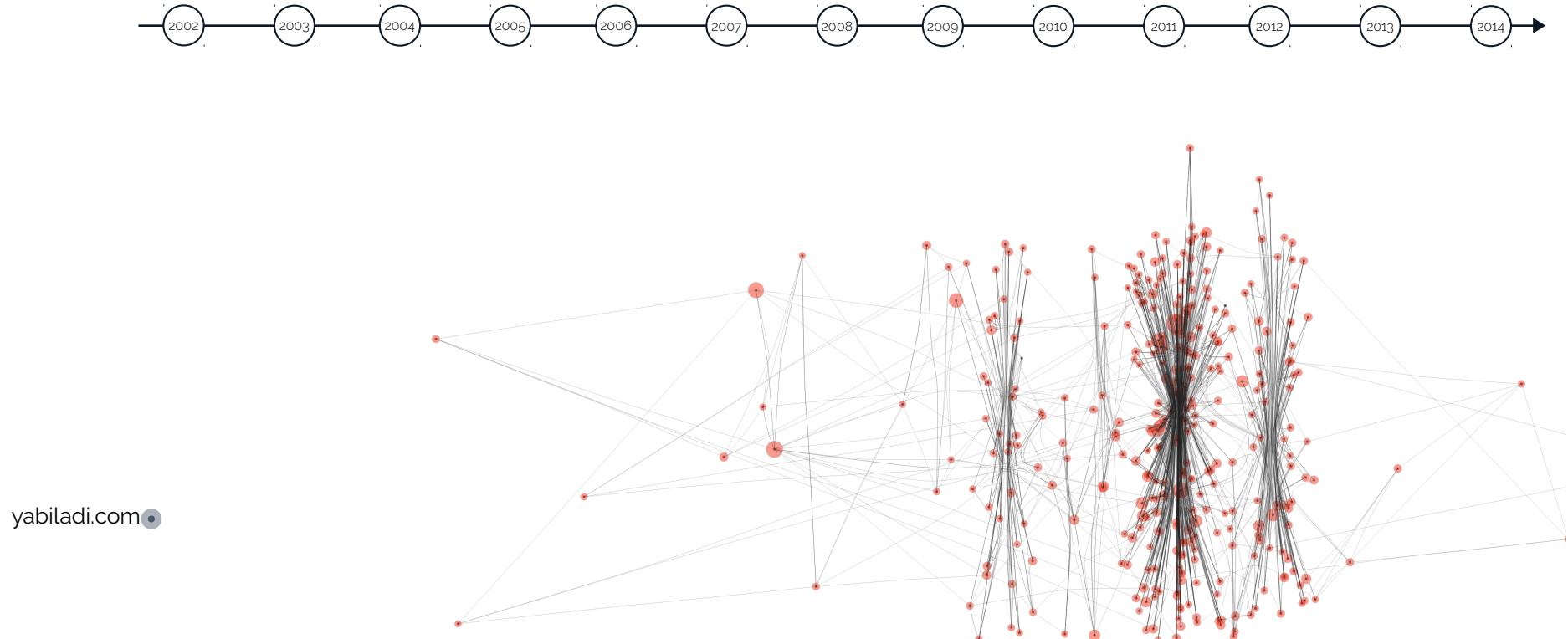
Un soulèvement en ligne éphémère (2/4)

> Suivre des trajets de contributeurs



Un soulèvement en ligne éphémère (3/4)

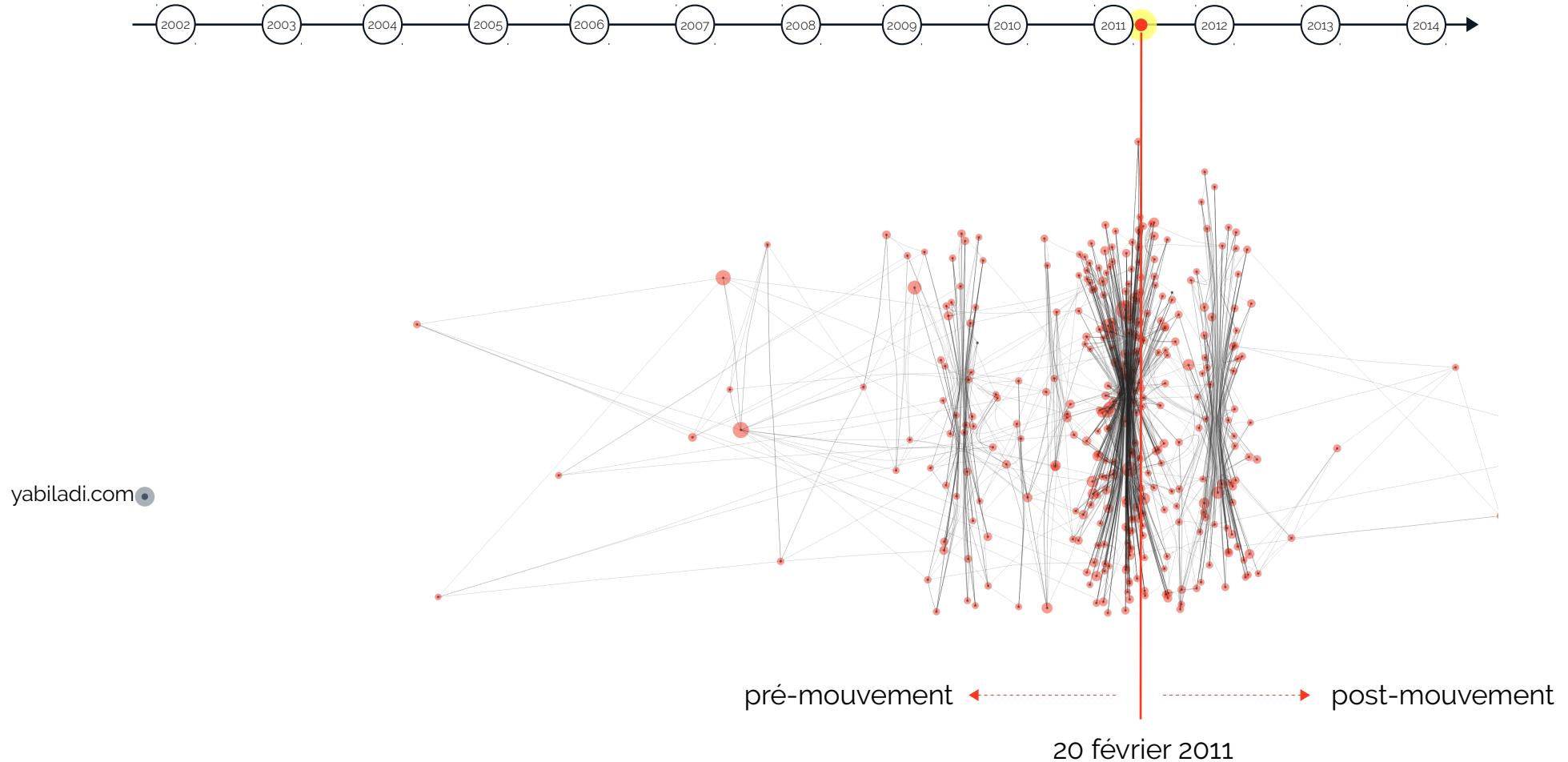
> D'anciens membres convergent, de nouveaux se joignent à eux



Tous les trajets sont agrégés pour former un réseau de membres contribuant collectivement à des threads.

Un soulèvement en ligne éphémère (3/4)

> D'anciens membres convergent, de nouveaux se joignent à eux



62 % des utilisateurs ont posté leur premier message avant le 20 février.

25 % des threads sont créés entre le 12/2010 et 03/2011

Un soulèvement en ligne éphémère (4/4)

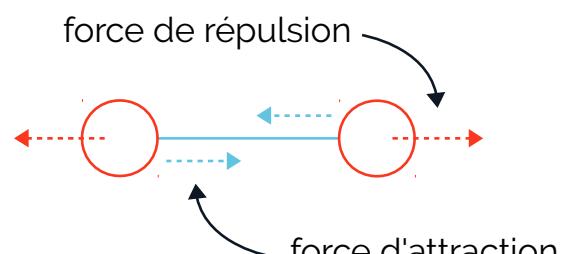
> Vers un embrasement soudain

Comment mieux saisir la **dynamique de convergence** des anciens et nouveaux membres ?

Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain

Comment mieux saisir la **dynamique de convergence** des anciens et nouveaux membres ?

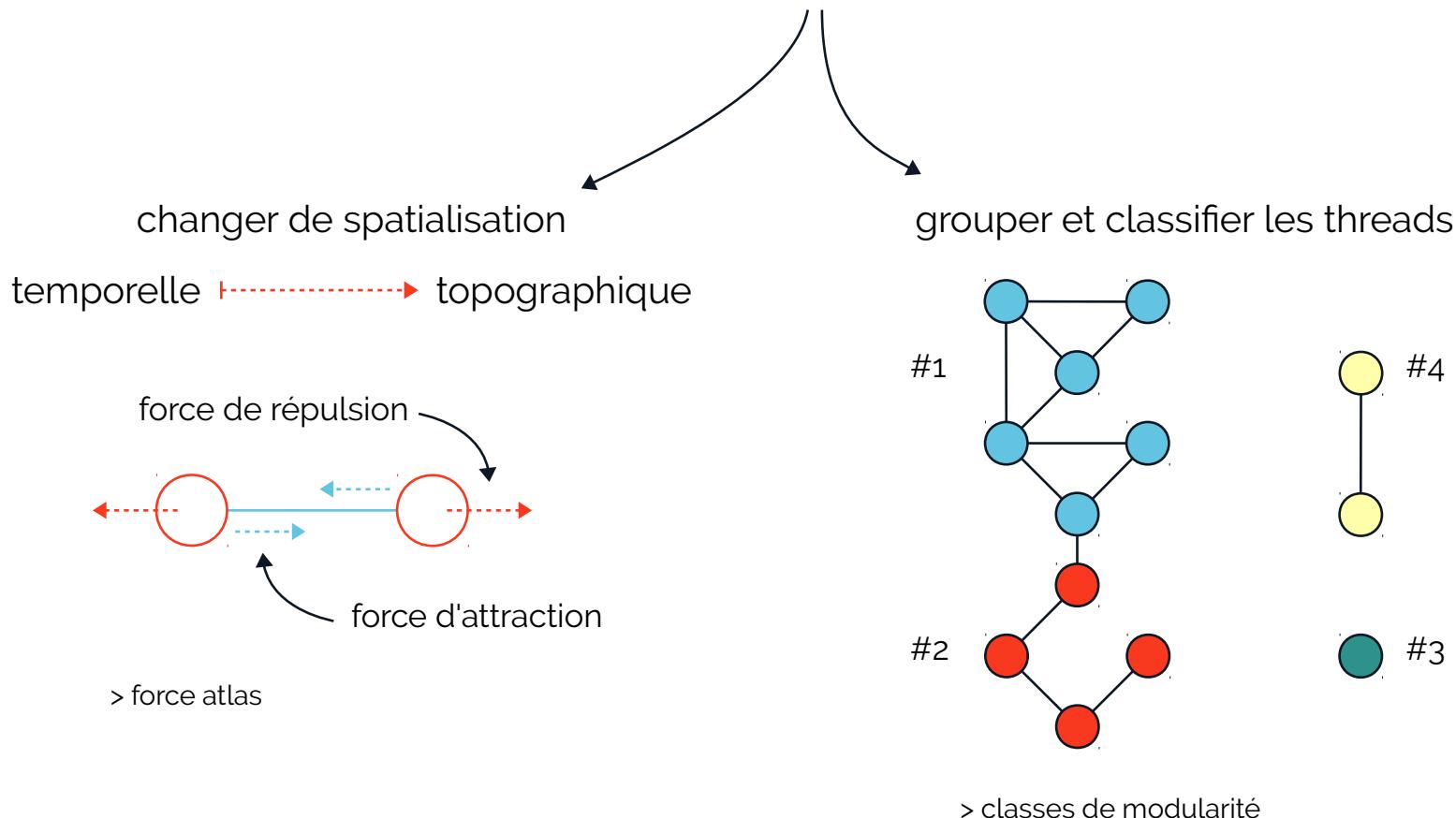


> force atlas

Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain

Comment mieux saisir la **dynamique de convergence** des anciens et nouveaux membres ?



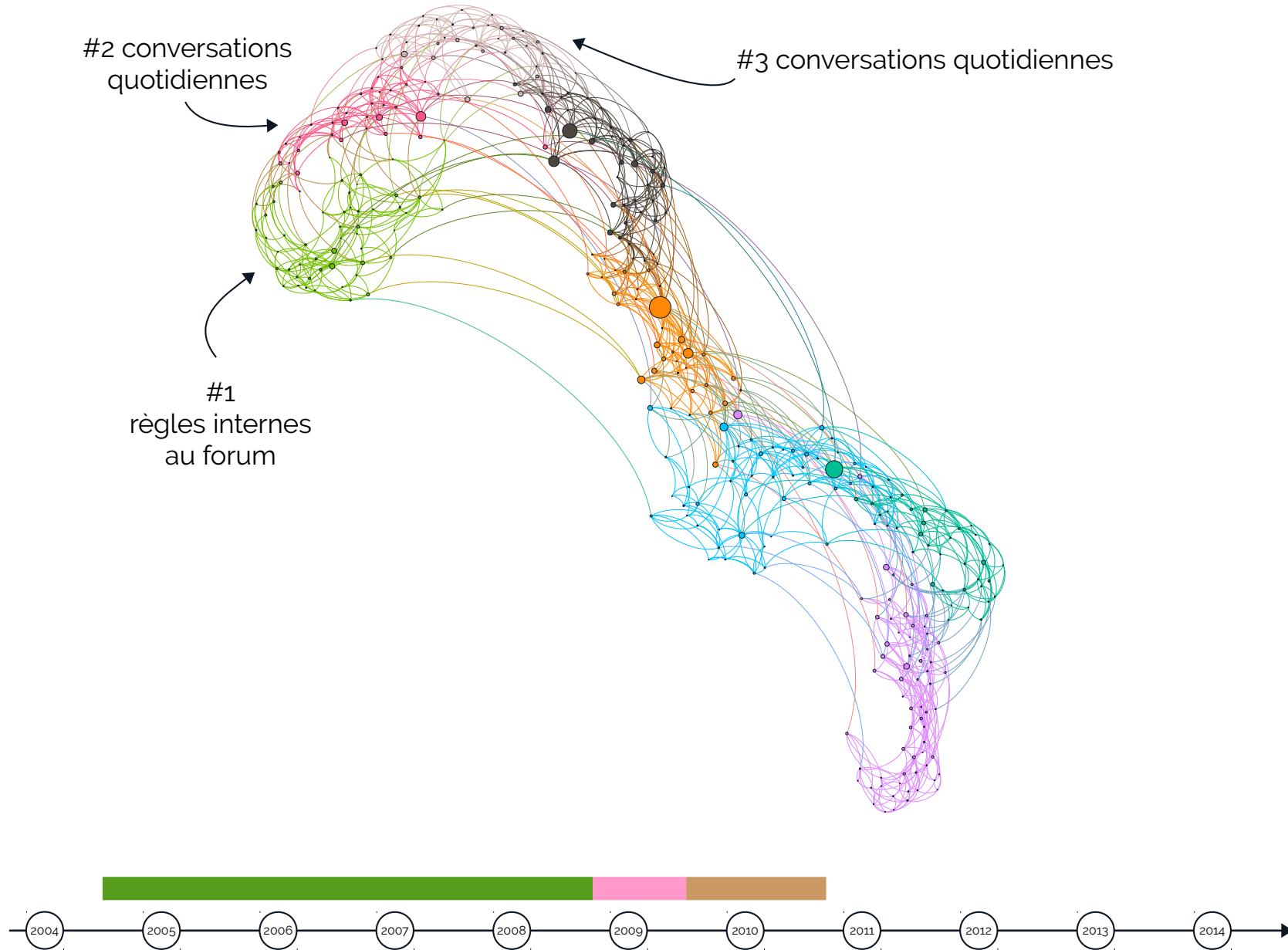
Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



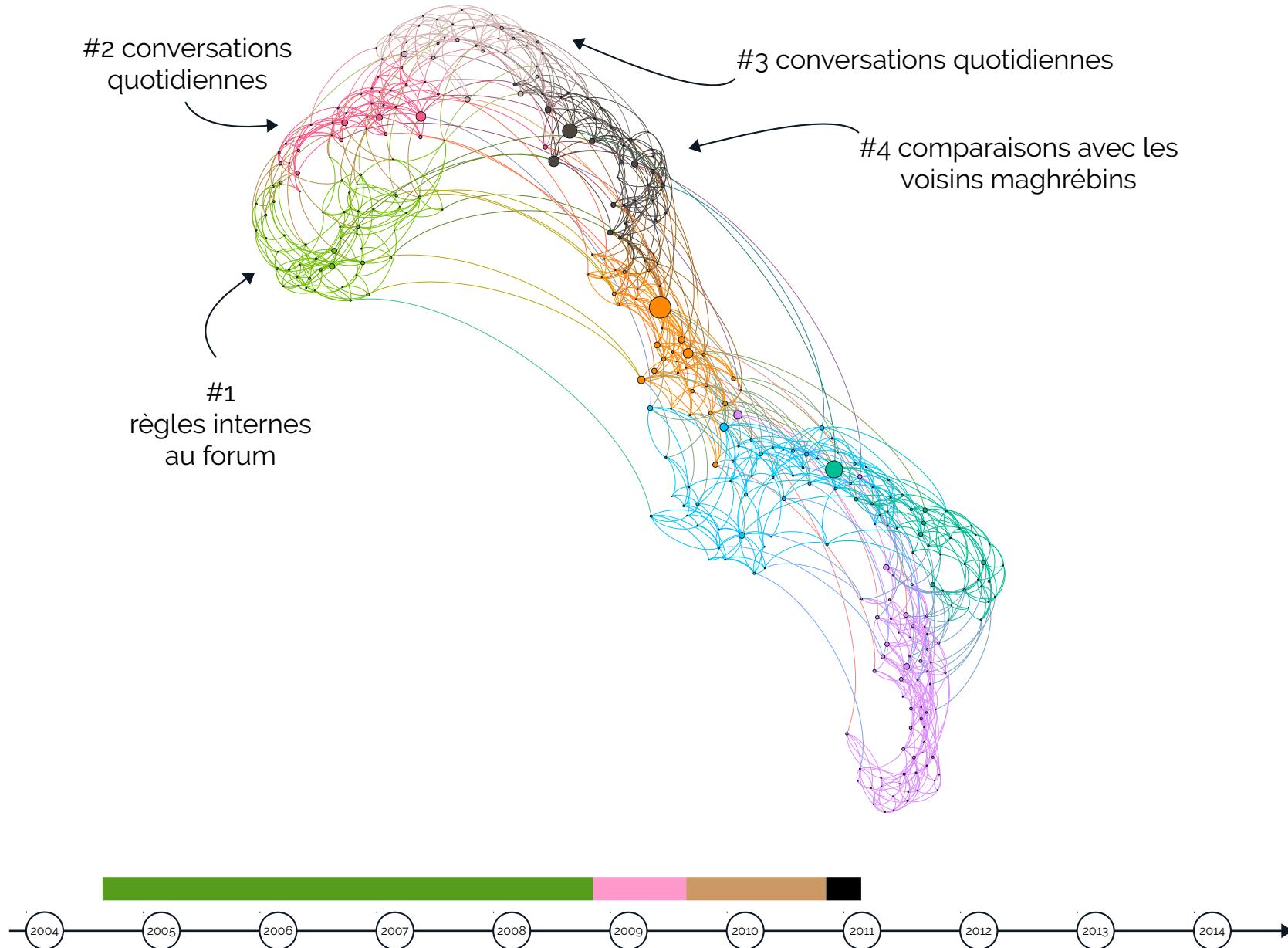
Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



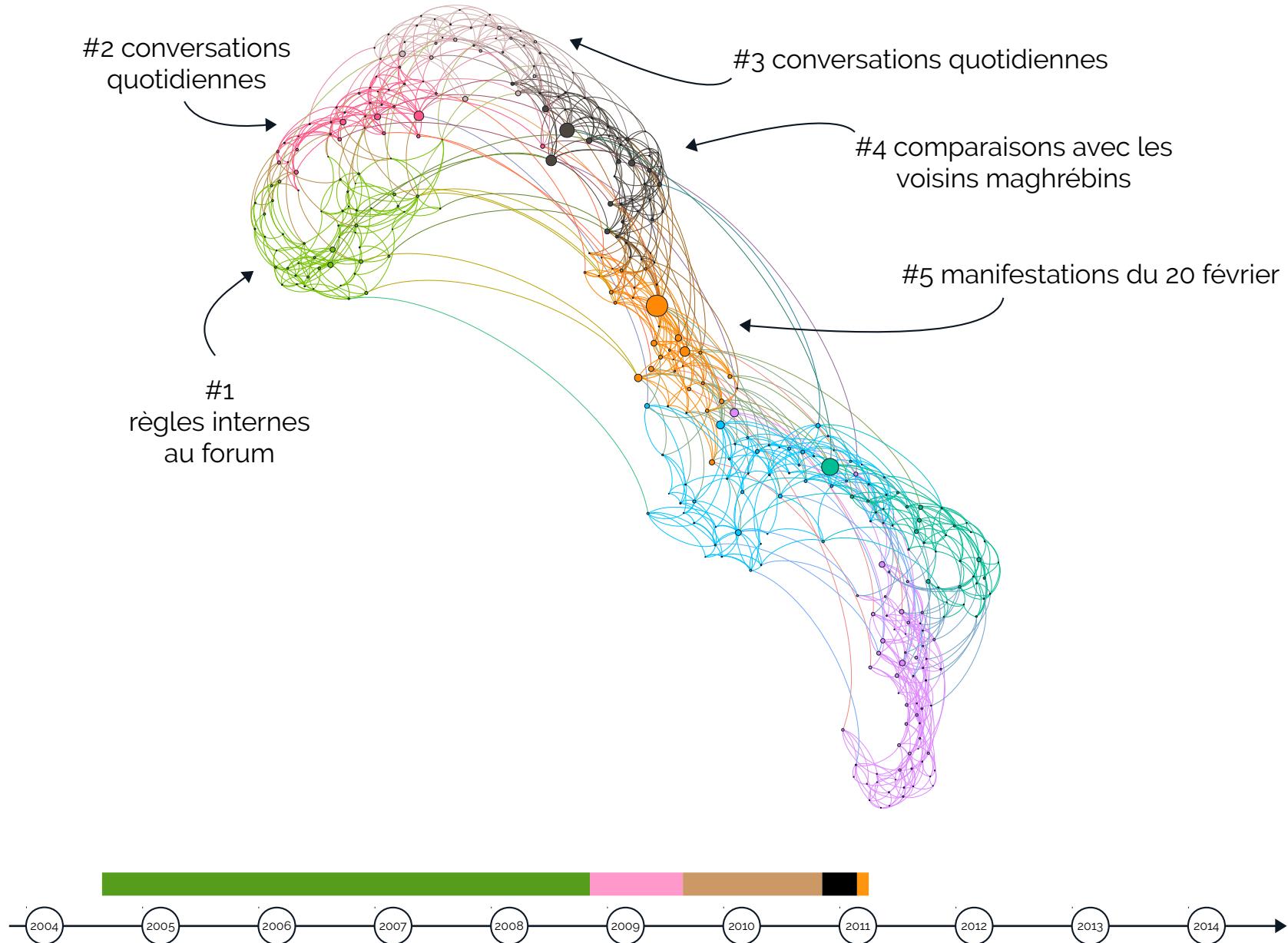
Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



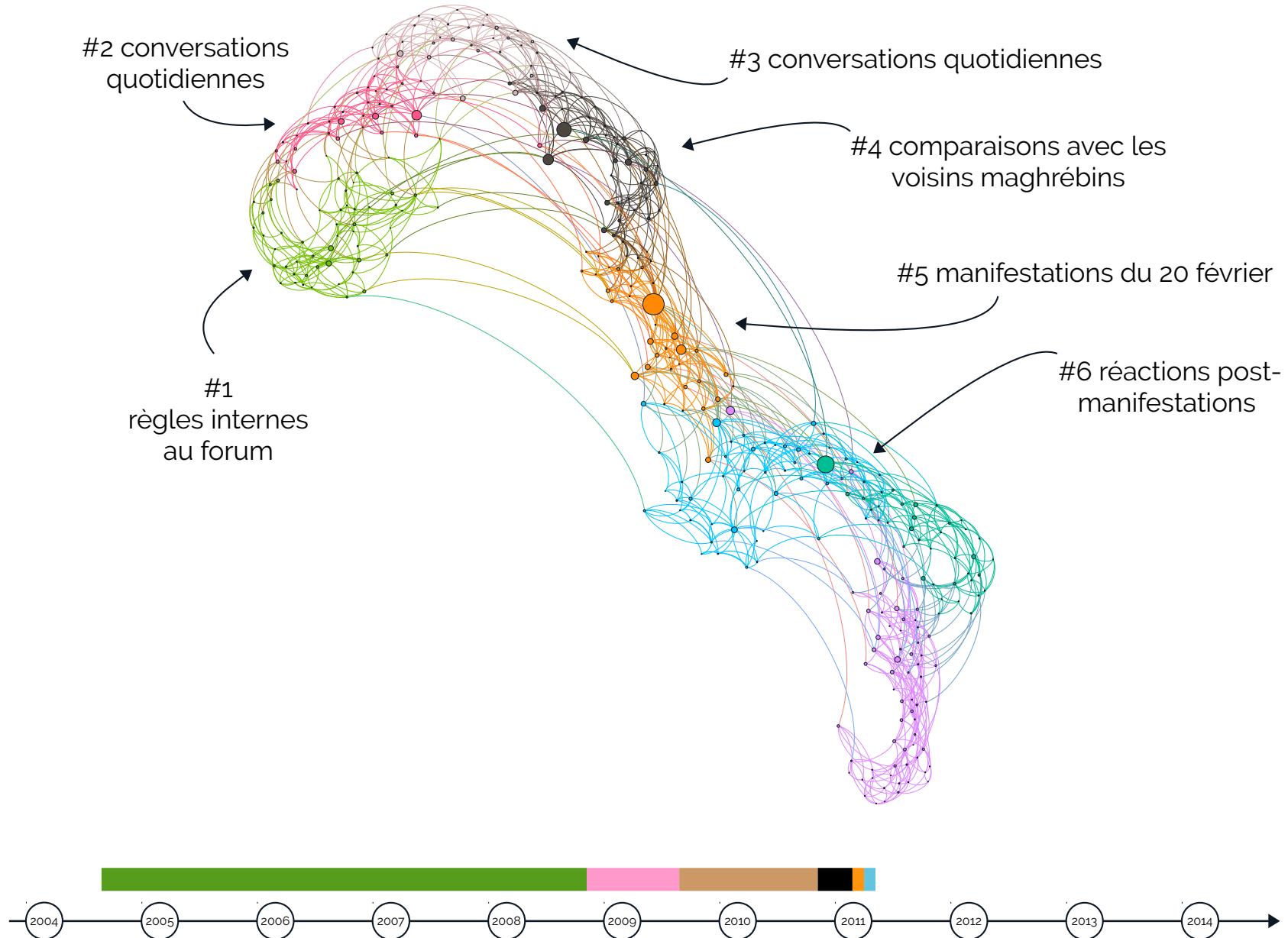
Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



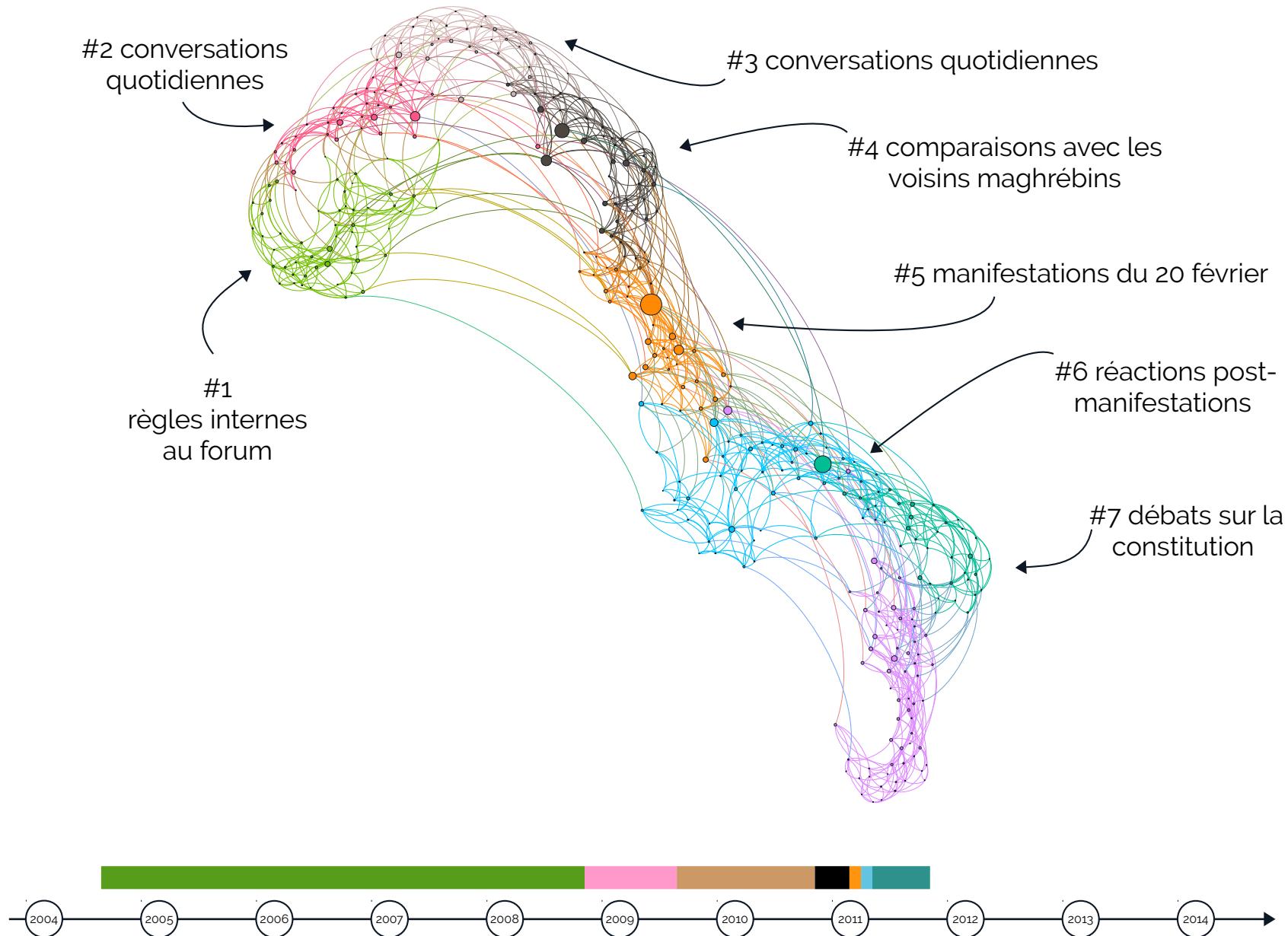
Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



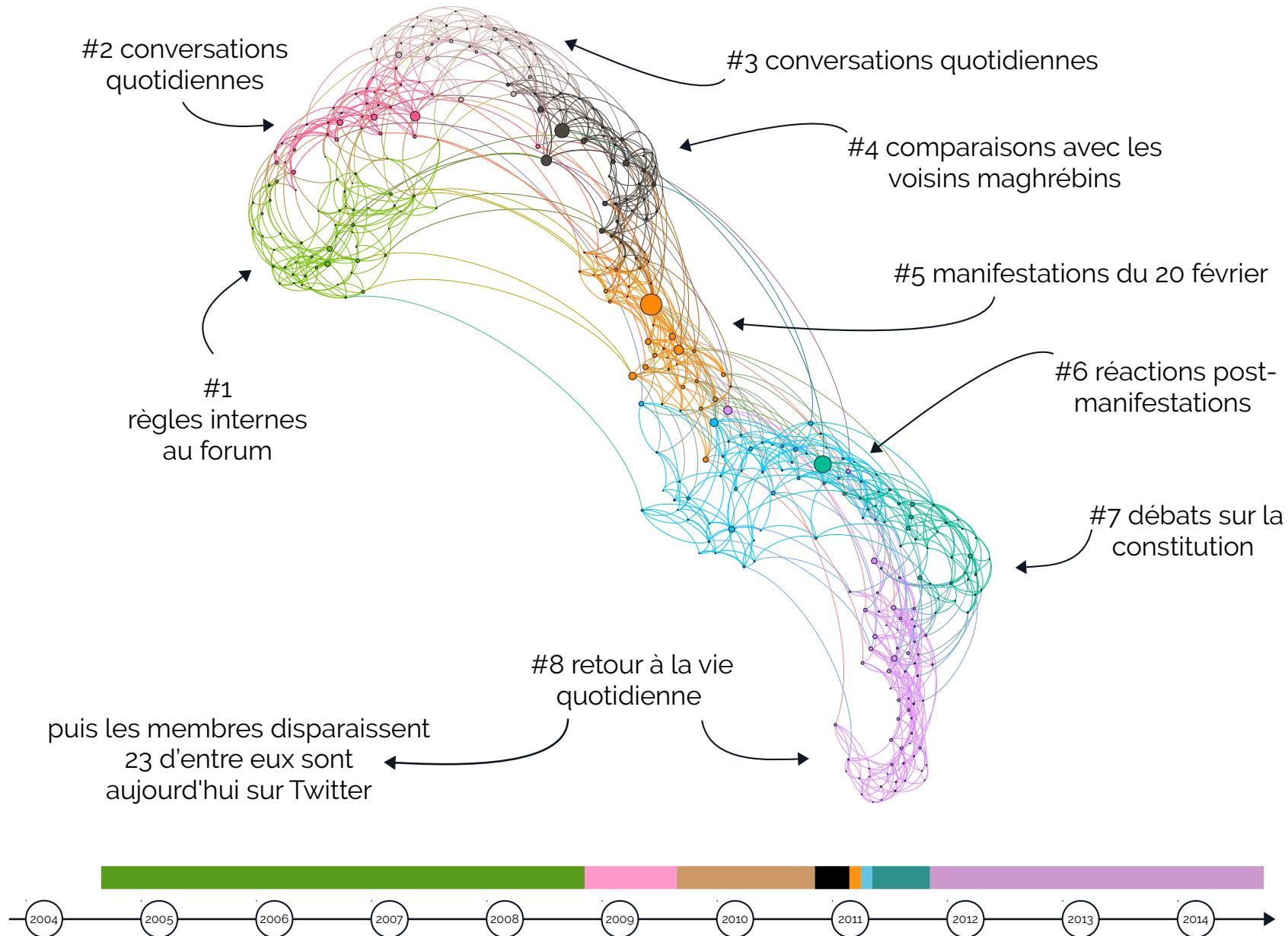
Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



Un soulèvement en ligne éphémère (4/4)

> Vers un embrasement soudain



Plan de résolution

> Vers une conclusion

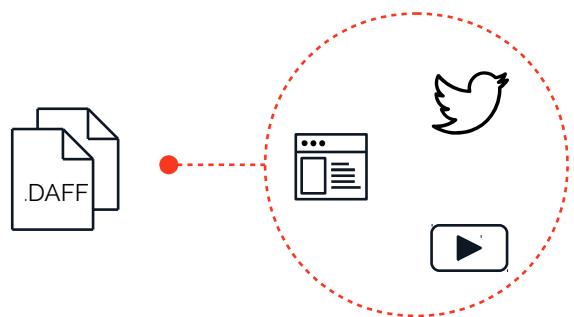
1. Les représentations en ligne des diasporas
2. Archiver le Web
3. Introduction au fragment Web
4. Les archives de la blogosphère marocaine
5. Les archives de yabiladi.com
- 6. Bilan et conclusion**

Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

Bien que le fragment Web nous ait permis de remonter à la racine des événements et des gestes du Web passé

Nous **ne pouvons aller plus loin** en nous basant sur les seules archives Web ...



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

Bien que le fragment Web nous ait permis de remonter à la racine des événements et des gestes du Web passé

Nous **ne pouvons aller plus loin** en nous basant sur les seules archives Web ...

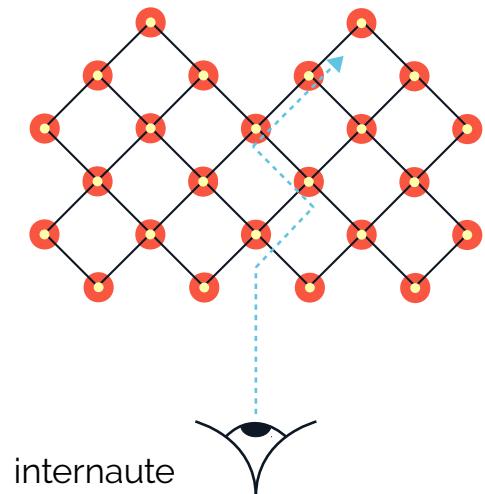


Les archives Web ont du mal à capturer le Web comme un **environnement** et de manière organique.

Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

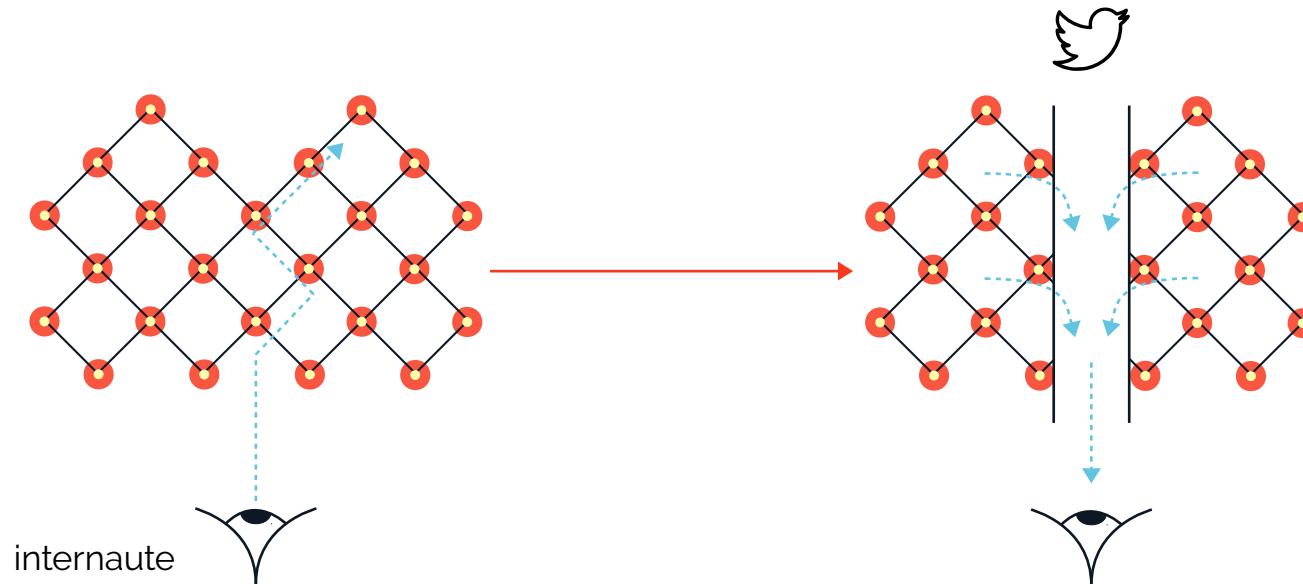
Alors que nous cherchions à étudier
les traces archivées des blogs et des forums ...



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

Alors que nous cherchions à étudier
les traces archivées des blogs et des forums ...



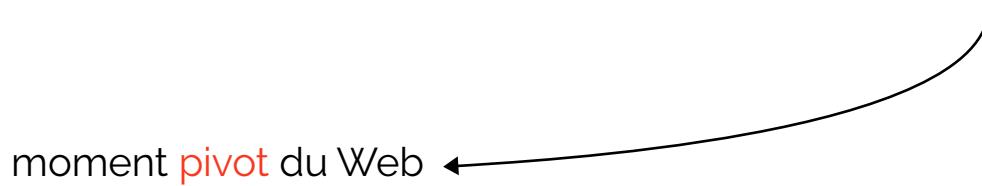
... les acteurs du Web avaient déjà
migré vers les réseaux sociaux

Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

Nos archives Web ne témoignent que du premier saut de cette **transition**

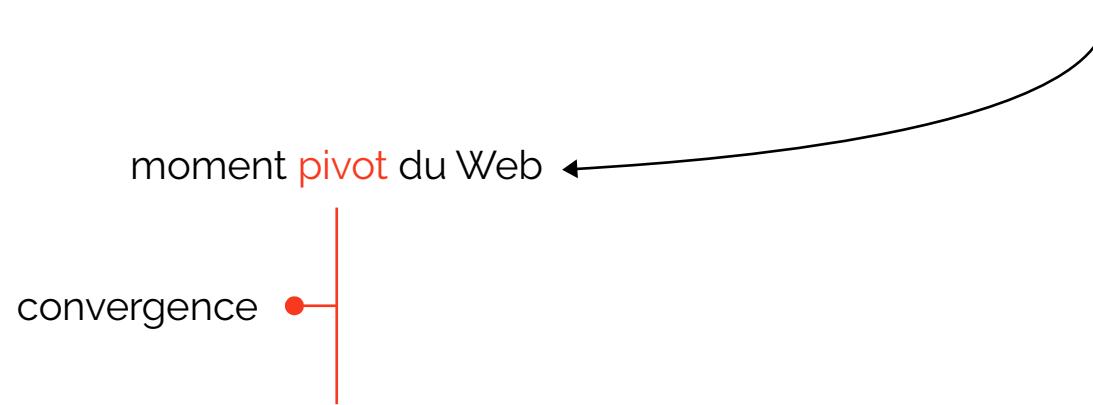
moment **pivot** du Web



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

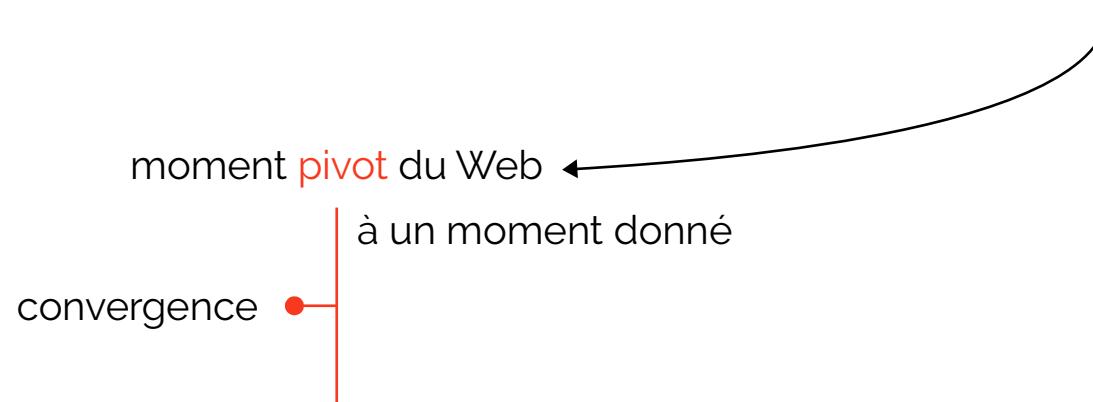
Nos archives Web ne témoignent que du premier saut de cette **transition**



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

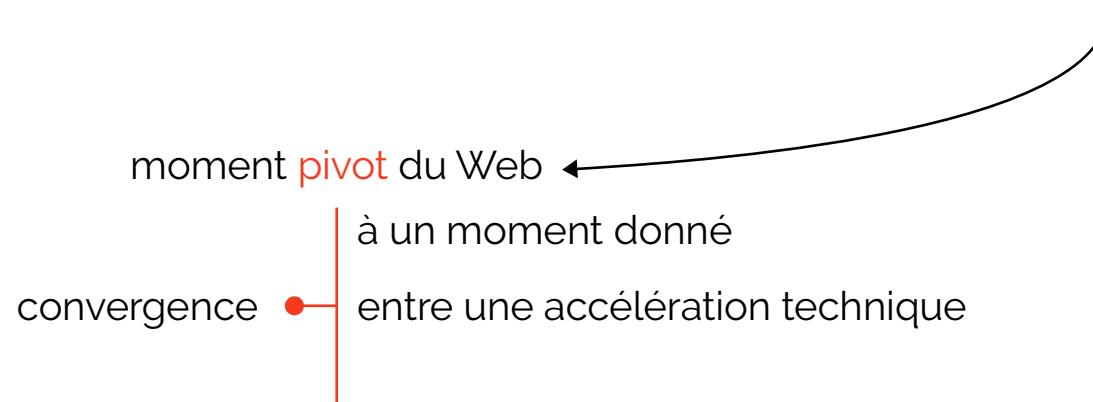
Nos archives Web ne témoignent que du premier saut de cette **transition**



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

Nos archives Web ne témoignent que du premier saut de cette **transition**



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

Nos archives Web ne témoignent que du premier saut de cette **transition**



Face aux **limites** de l'archivage du Web

> Réflexions sur les archives comme matière historique

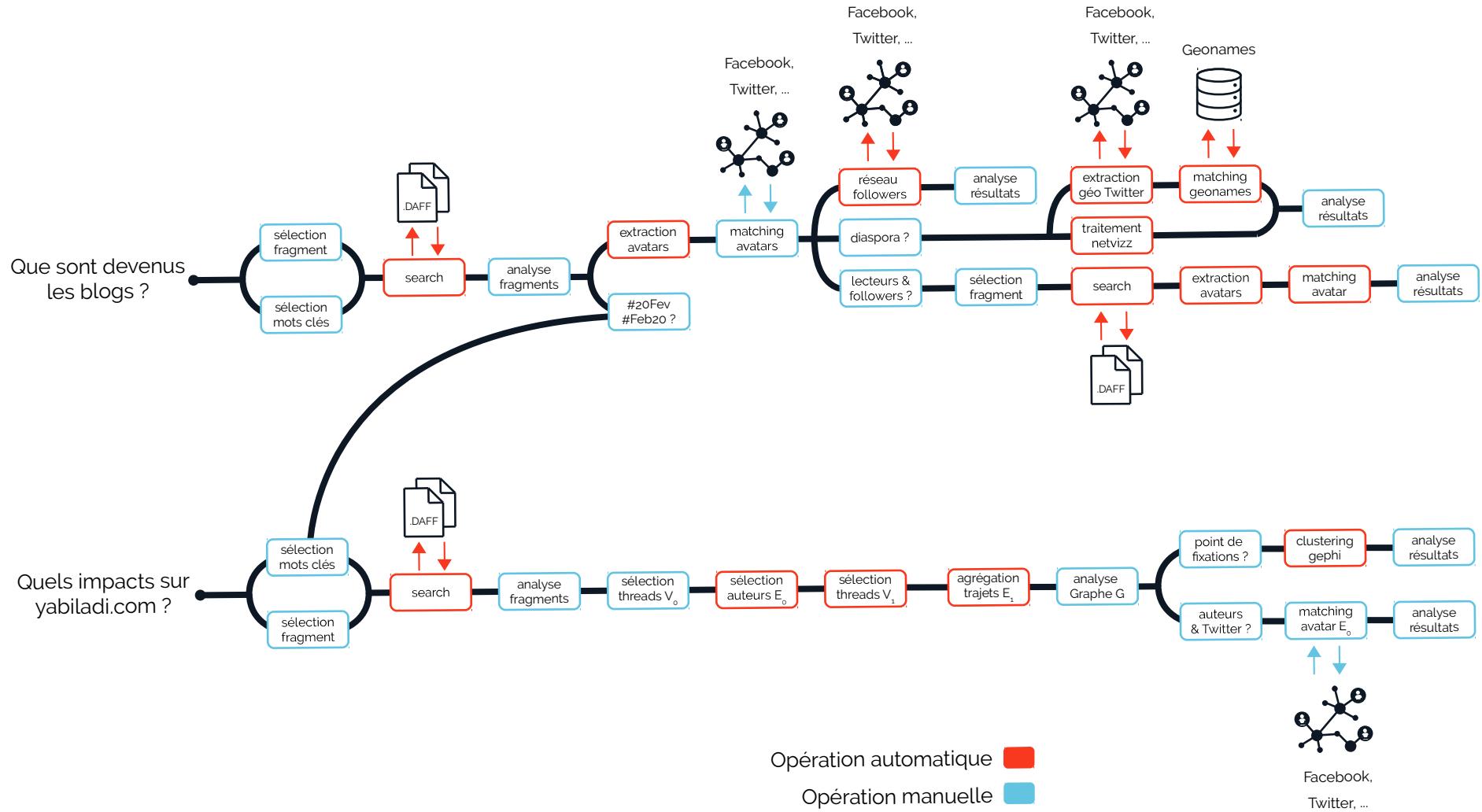
Nos archives Web ne témoignent que du premier saut de cette **transition**



Étudier les archives Web autour de ces moments pivots où le Web change de direction peut-être une **bonne stratégie** pour qui veut mener des recherches historiques.

Bilan de nos explorations

> Approche automatique, semi-automatique et manuelle



Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

1. Revenir au **terrain**

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

1. Revenir au **terrain**



2. Élargir nos travaux à **l'ensemble** des e-Diasporas

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

1. Revenir au **terrain**



2. Élargir nos travaux à **l'ensemble** des e-Diasporas



3. **Valider** et quantifier l'efficacité du fragment Web

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

1. Revenir au **terrain**
2. Élargir nos travaux à **l'ensemble** des e-Diasporas
3. **Valider** et quantifier l'efficacité du fragment Web
4. Multiplier les **outils** d'exploration

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

Les archives Web deviendraient objets d'**études historiques** à part entière.

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

Les archives Web deviendraient objets d'**études historiques** à part entière.

- ||| Comment étudier in situ le Web passé ? Reproduction ? Émulation ?
- ||| Peut-on étudier la circulation d'un contenu à travers les archives Web ?
- ||| Quels sont les sites qui perdurent et qui perdureront ?
- ||| Peut-on traquer la présence de fantômes dans les archives ?

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

Les archives Web deviendraient objets d'**études historiques** à part entière.

- ||| Comment étudier in situ le Web passé ? Reproduction ? Émulation ?
- ||| Peut-on étudier la circulation d'un contenu à travers les archives Web ?
- ||| Quels sont les sites qui perdurent et qui perdureront ?
- ||| Peut-on traquer la présence de fantômes dans les archives ?

Définissant alors les contours d'une future **archéologie** du Web !

Pour conclure

> De l'exploration d'archives Web à une discipline plus mature

Les archives Web deviendraient objets d'**études historiques** à part entière.

- ||| Comment étudier in situ le Web passé ? Reproduction ? Émulation ?
- ||| Peut-on étudier la circulation d'un contenu à travers les archives Web ?
- ||| Quels sont les sites qui perdurent et qui perdureront ?
- ||| Peut-on traquer la présence de fantômes dans les archives ?

Définissant alors les contours d'une future **archéologie** du Web !

Merci
pour votre attention !

Place
aux questions ...