

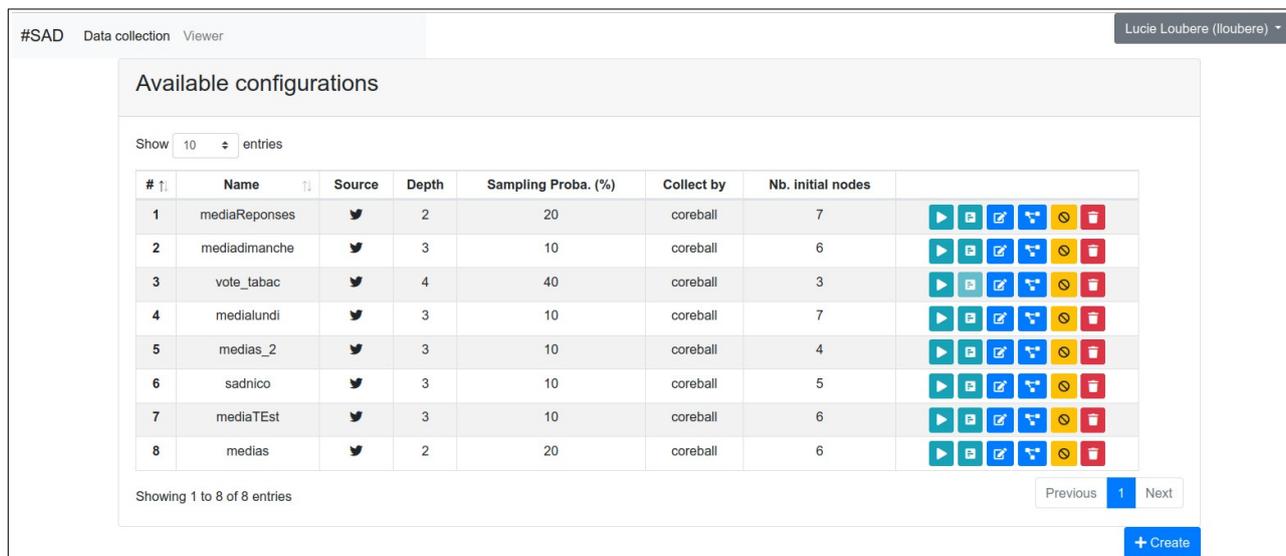
Documentation interface SAD.

Table des matières

1	Tableau de bord.....	2
2	Archive des collectes :.....	2
3	Paramètres.....	3
3.1	Paramètres de la collection.....	4
3.2	Paramètres du graphe.....	5
3.3	Paramètres Twitter.....	6
4	Gestions des comptes de départ.....	7
4.1	Recherche par hashtag.....	8
5	Blacklist.....	8
6	Visualisation.....	9
6.1	Informations accessibles.....	10
6.2	Bandeau général.....	10

1 Tableau de bord

L'entrée dans l'outil se fait par le tableau de bord. Accessible par l'onglet « Data Collection », il permet de créer un nouveau travail par le bouton .



The screenshot shows a dashboard titled '#SAD Data collection Viewer' with a user profile 'Lucie Loubere (loubere)'. Below the title is a section 'Available configurations' with a 'Show 10 entries' dropdown. A table lists 8 configurations:

#	Name	Source	Depth	Sampling Proba. (%)	Collect by	Nb. initial nodes	
1	mediaReponses	Twitter	2	20	coreball	7	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
2	mediadimanche	Twitter	3	10	coreball	6	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
3	vole_tabac	Twitter	4	40	coreball	3	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
4	medialundi	Twitter	3	10	coreball	7	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
5	medias_2	Twitter	3	10	coreball	4	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
6	sadnico	Twitter	3	10	coreball	5	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
7	mediaTEst	Twitter	3	10	coreball	6	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]
8	medias	Twitter	2	20	coreball	6	[Play] [Show run] [Update] [Manage nodes] [Black-list] [Delete]

At the bottom of the table, it says 'Showing 1 to 8 of 8 entries' and includes 'Previous', '1', and 'Next' navigation buttons, along with a '+ Create' button.

Il reprend également les différents travaux déjà opérés par l'utilisateur, avec pour chacun d'entre eux une description des principaux paramètres (Nom, source, nombre de strates, pourcentage de compte collecté par strate, nombre de comptes initiaux) ainsi que 6 modifications possibles :

	Play : cette commande permet de relancer la collecte et le nouveau graphe qui y correspond.
	Show run : permet d'afficher les collectes déjà opérées sur ce travail, ils s'afficheront sous le tableau de bord (voir paragraphe \$ref).
	Update collection parametres : permet d'accéder aux paramètres de la collecte et de la visualisation (voir paragraphe \$ref).
	Manage initials nodes : permet d'accéder à la sélection des comptes initiaux à partir desquels seront faits la collecte par boule de neige (voir paragraphe \$ref).
	Manage black-list : permet d'accéder à une liste de comptes à ignorer lors de la collecte par boule de neige (voir paragraphe \$ref).
	Delete : supprime le travail.

2 Archive des collectes :

Ce tableau est accessible par la touche  présente sur chaque travail du tableau de bord (voir \$ref).

Runs for collect «mediaReponses»

Show: 10 entries Search:

#	Status	Log	End date	
c157e205-f0f1-41e0-89c5-dbb87799ff50	completed		March 22, 2022, 5:54 p.m.	   
export avec 10% des comptes 26088d80-64bb-48b6-998b-c98d14e5f570	completed		Feb. 6, 2022, 12:56 p.m.	   
Graph has 90 nodes and 563 edges				
export avec 50% des comptes eba67e07-dd89-42f3-9839-6c6eecd27c3a	completed		Feb. 6, 2022, 12:50 p.m.	   

Showing 1 to 3 of 3 entries Previous 1 Next

Il affiche toutes les collectes déjà opérées pour le travail sélectionné. Lors du lancement d'une collecte, une ligne se crée automatiquement avec un nom codé aléatoirement. Les colonnes présentées sont :

- **#** : le nom, il est possible de le modifier avec la  touche ,
- **Status** : il reprend l'état d'avancement de la collecte, si elle est en cours, il affiche « running » ainsi que le % de l'a collecte,
- **Log** : contient un fichier texte reprenant les informations de la collecte,
- **End date** : la date à laquelle s'est achevée la collecte.

La dernière colonne donne accès à 3 options :

	Permet d'accéder à la visualisation
	Permet d'annoter cette collecte
	Supprime cette collecte

Des détails sur le nombre de nœud (comptes) et liens pour chaque collecte sont accessibles par la touche .

3 Paramètres

La touche  présente pour chaque travail du tableau de bord permet d'accéder aux paramètres de collecte. Séparés en 3 thématiques, ils reprennent les 3 éléments de travail de cette application (collecte, visualisation, gestion de l'api twitter). Ainsi, les différents blocs présents sur cette page permettent de personnaliser les exports sur :

- les paramètres de la boule de neige (nombre de strates, nombre de comptes récupérés par strates...),
- la construction du graphe (nombre minimum de liens entre de compte...)
- des paramètres spécifiques à Twitter (\$ref).

3.1 Paramètres de la collection

La collection se fait par « boule de neige », c'est-à-dire qu'à partir des comptes initiaux renseignés par le journaliste, l'application va sonder leurs derniers tweets et identifier les comptes mentionnés et/ou retweetés par ce groupe pour constituer une nouvelle liste à sonder pour former une nouvelle strate. Cette démarche permet de reconstituer un réseau de comptes, mais peut très vite amener à un nombre conséquent de comptes. De plus chaque étude est différente, afin de permettre à partir du même outil d'étudier une petite communauté au faible taux d'échange, mais aussi de grandes populations aux échanges multiples, l'interface propose de paramétrer plusieurs éléments structurant la collecte.



[Collection parameters](#)

Name*
mediaReponses

Name

Depth*
2

Number of exploration steps.

Sampling probability*
20

Percentage of nodes randomly selected at each layer

Collect by*
Coreball

Collect by

Max nodes per hop*
1000

Max number of new nodes at each hop

Les différents paramètres sont :

- **Name** : c'est le nom qui désignera ce travail dans le tableau de bord,

- **Depth** : il définit le nombre de strates à répéter durant la collecte, l'augmenter permet d'atteindre un graphe plus complexe, mais augmente le temps de traitement. Le diminuer, donne un graphe plus simple à lire et interpréter, mais limite la profondeur des échanges.
- **Sampling probability** : en fonction des réseaux étudiés, le nombre de comptes rencontré pour chaque strate peut devenir massif, afin de limiter les comptes et privilégier le nombre de strates (et donc la profondeur des réseaux), il est proposé ici de définir le pourcentage de compte à conserver pour chaque strate. Un taux de 100 % prendra tous les comptes rencontrés, un taux de 10 % ne prendra que 10 comptes tous les 100 rencontrés.
- **Collect by** : permet de choisir entre 2 algorithmes le Coreball par défaut ou le Fireball (encore en développement).
- **Max nodes per hop** : la récolte par boule de neige fait augmenter de façon exponentielle le nombre de comptes rencontrés à chaque strate. Ce paramètre permet de limiter l'effectif de comptes recueilli à chaque étape. En privilégiant ce paramètre au paramètre « sampling probability », cela permet de conserver une proportion élevée de comptes dans les premières strates (puisque l'on part de quelques comptes initiaux, les comptes rencontrés en première étape seront inférieurs à 1000), mais de limiter les strates suivantes.

3.2 Paramètres du graphe

Dans ce paragraphe, nous agissons après la collecte lors de la structuration du graphe. Le principe de cette étape est de placer sur un plan l'ensemble des comptes rencontrés, puis de les relier entre eux par leur nombre de connexions. C'est à partir de ce graphe structuré que sont calculées les communautés de comptes. Plus des comptes ont tendance à se retweeter et/ou se mentionner entre eux, plus ils structureront une zone du graphe dense et plus ils seront identifiés comme appartenant à une même communauté.

Les réseaux collectés par l'application pouvant contenir de nombreux comptes, afin qu'ils soient plus facilement lisibles, l'outil SAD permet de limiter les liens et comptes apparaissant selon les critères suivants :

- **Minimum weight** : c'est le nombre minimal de connexions entre deux comptes pour qu'ils soient reliés sur le graphe. Si on augmente cette valeur, on « efface » les liens ponctuels entre les comptes au profit d'échanges réguliers (forte connexion, ou conflit).
- **Minimum degree** : nombre minimum de voisins pour qu'un compte soit dans le graphe. En augmentant cette valeur, on « évince » les comptes reliés à peu de personnes, qui peuvent avoir été mentionné ou retweeté très occasionnellement.

- **Minimum community size** : nombre minimum de comptes pour qu'une communauté soit affichée. Si l'on augmente ce paramètre les petites communautés ne seront plus affichées, le graphe gagnera en clarté, mais perdra en précision.

Graph parameters

Minimum weight*

1

Minimum number of RT/mentions needed between 2 users to connect them.

Minimum degree*

2

Minimum number of neighbors a node must have to be included.

Minimum community size*

2

Communities of users are removed if they contain too few elements

3.3 Paramètres Twitter

L'outil SAD permet d'investiguer le réseau social Twitter, pour cela il commande l'API de cette plateforme, les tweets récupérés sont dotés de plusieurs données. Les paramètres suivants permettent de restreindre le nombre de tweets étudiés par utilisateur selon :

API version : permet d'utiliser l'API 1 ou l'API 2, cette dernière n'est rendue accessible par twitter uniquement aux chercheurs académiques.

- **Min mentions** : le nombre minimum de fois qu'un utilisateur est mentionné et/ou retweeté pour apparaître dans le graphe. L'augmentation de cette valeur permet d'effacer les comptes n'étant mentionnés que quelques fois.
- **Max days** : période maximale sur laquelle on collecte les tweets, elle se calcule à partir de la date d'export et est exprimée en jour. Réduire ce chiffre permet de limiter la temporalité de l'étude. En effet collecter les 100 derniers tweets sur un compte à l'activité réduite peut amener à une temporalité plus longue que prévu.
- **Max tweets** : nombre maximum de tweets collecté pour un usager. L'activité des comptes peut être très hétérogènes, de quelques tweets pas mois, à plusieurs dizaines par jours. Afin de limiter les comptes à grosse production, nous pouvons ici limiter le nombre de tweets.
- **Popular Tweets** : ce paramètre ne porte que sur le volet lexical de l'application, il permet de restreindre le nombre de tweets par utilisateur qui sera exploité pour constituer le nuage de mots (voir paragraphe \$ref).

[Twitter parameters](#)

Api version*

1

Twitter API version

Min mentions*

0

Minimum number of times a user must be mentioned by the already collected users to be included.

Max days*

45

Number of days max in the past when collecting a user timeline

Max tweets*

200

Maximal number of tweets collected from a user's timeline

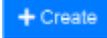
Popular tweets*

20

Number of popular tweets (for lex. analysis) for each user

4 Gestions des comptes de départ

La touche  accessible pour chaque travail présent sur le tableau de bord permet d'accéder au menu des nœuds initiaux. Dans cette interface, nous pouvons ajouter des comptes pour débiter la boule de neige. Les trois actions possibles sont :

- L'ajout d'un nouveau compte par le bouton 
- La modification d'un compte déjà enregistré 
- la suppression d'un compte enregistré 

Show entries

#	Name	Update / Delete
1	@fairmedia	 
2	@biofrontsau	 

Showing 1 to 2 of 2 entries

Previous Next

Hashtag search

Hashtag*



4.1 Recherche par hashtag

Dans ce menu, un module permet de rechercher les comptes par l'utilisation de hashtags. Pour cela, il faut renseigner un hashtag dans le menu déroulant et lancer la recherche. Les comptes et tweets utilisant ce hashtag s'affichent alors.

Hashtag search

Hashtag*

#LFI

Search

Add users

Users list

Save

Show 10 entries

#	User	Text	❤	↻
1	AQuatennens  	Aujourd'hui s'ouvre la séquence qui portera Jean-Luc #Melenchon au second tour de l'élection #présidentielle.... https://t.co/e9cggdZmYQ	1211	279
<small>143k 🗣️ 20k ❤️ Feb 2015 Sun 20 Mar 2022</small>				
2	AQuatennens  	La puissance de notre rassemblement est une réponse à #Macron. La présence de @JLMelenchon au second tour de la... https://t.co/6QHNP4Ho9r	426	160
<small>143k 🗣️ 20k ❤️ Feb 2015</small>				

orkpanda.xyz/collect/initial-nodes/32/search/présidentielle

Le bouton  permet d'ajouter le compte dans le champ « Add users », le bouton  permet d'enregistrer les comptes dans le menu des comptes initiaux.

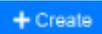
5 Blacklist

La touche  accessible pour chaque travail présent sur le tableau de bord permet d'accéder au menu des comptes à ignorer durant la collecte. Lors de l'étude, certains comptes très populaires (voire trop populaires) vont apparaître, c'est par exemple le cas de YouTube, ou autre plateforme. En fonction de ce qui est recherché, il se peut qu'ils produisent trop de bruit et/ou qu'ils fassent sortir le graphe de la thématique étudiée. Cela peut également être utile, si une personnalité extérieure au territoire étudié apparaît et produit par sa présence lors d'une strate une communauté autour d'elle. Le menu présenté ici permet de lister les comptes à ne plus prendre en compte lors de la collecte. Il faut les renseigner sans le @.

#	Name	Update / Delete
1	@S_Jacoby	 
2	@zentralplus	 
3	@ChTaubira	 
4	@AvecTaubira_CH	 

← Back + Create

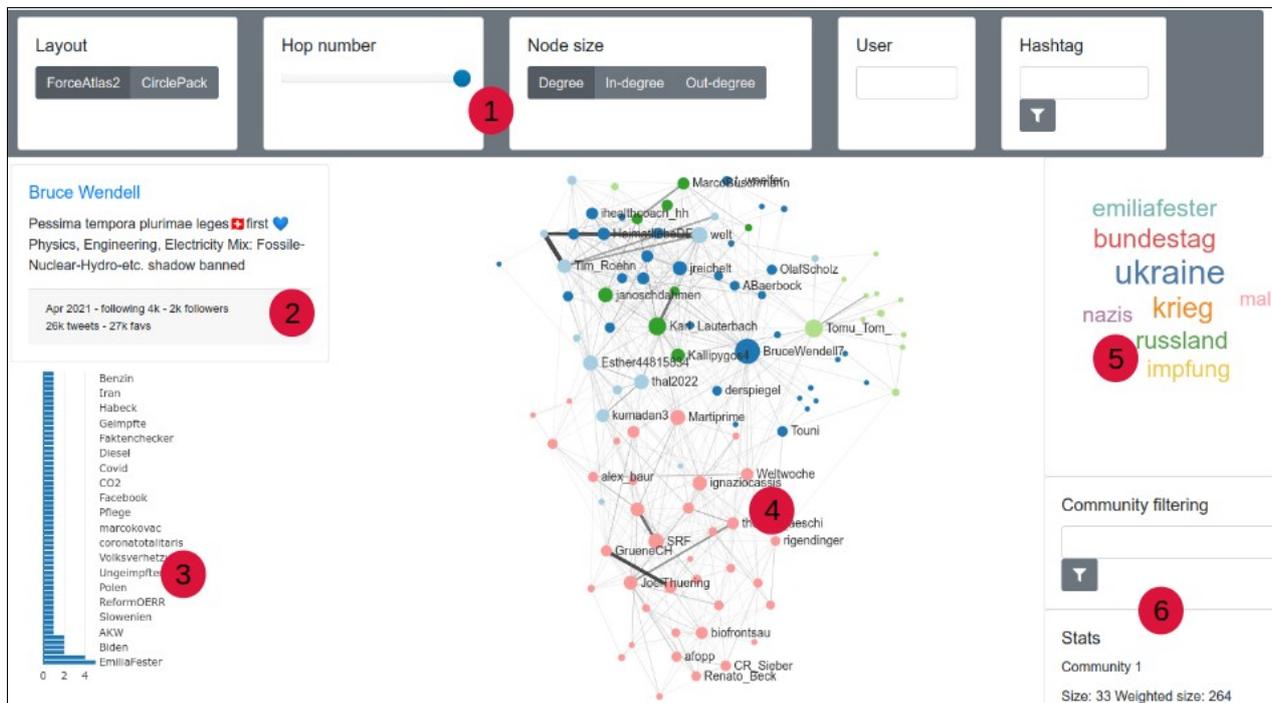
Les trois actions possibles sont :

- L'ajout d'un nouveau compte par le bouton 
- La modification d'un compte déjà enregistré 
- la suppression d'un compte enregistré 

6 Visualisation

La dernière étape du procédé consiste en la visualisation du graphe collecté, il permet de façon dynamique d'explorer les comptes reliés entre eux en apportant certaines informations sur eux.

6.1 Informations accessibles

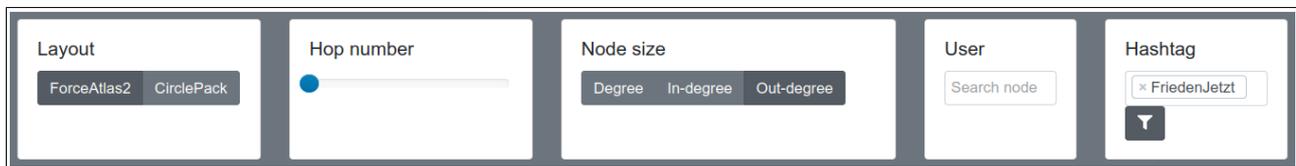


Nous trouverons comme différents blocs :

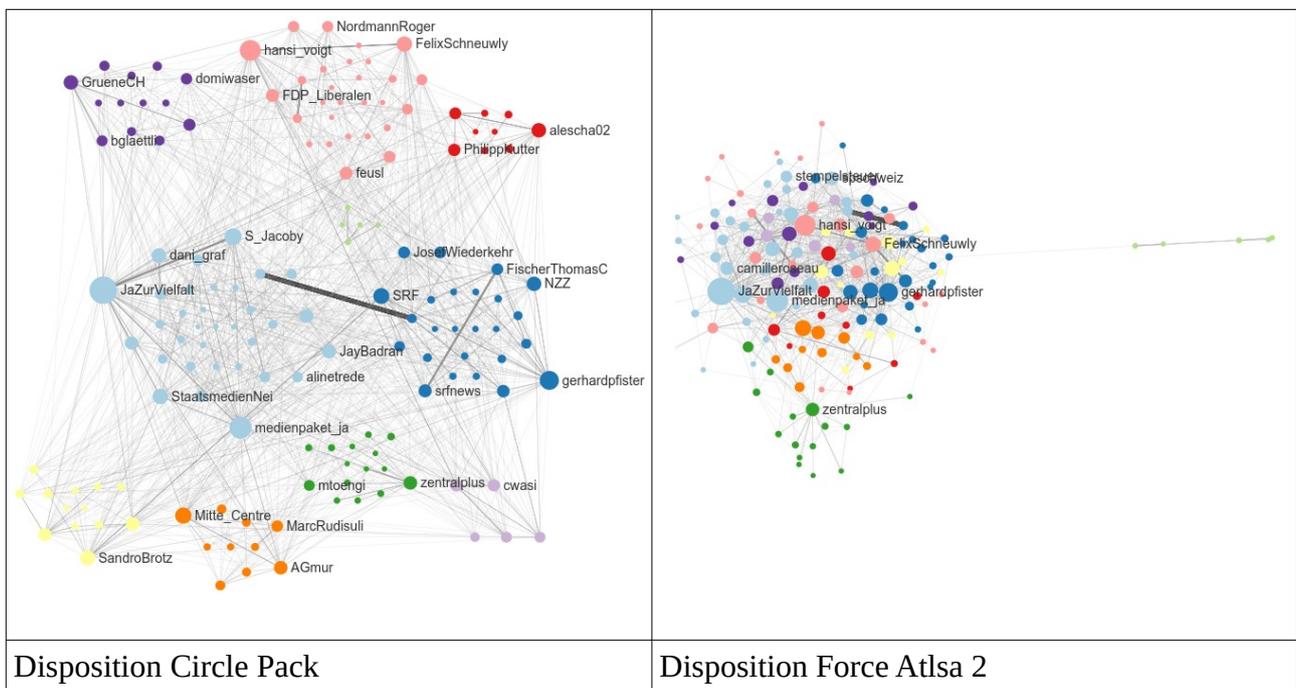
- 1 – bandeau de paramètre du graphique (voir paragraphe \$ref pour plus de détails)
- 2 – Description du compte twitter, ce sont les éléments présents sur la page twitter du compte,
- 3 – hashtag utilisé par le compte dans les tweets récoltés
- 4 – graphique de l’export,
- 5 – mots clés spécifiques de la communauté à laquelle appartient le compte,
- 6 – détails concernant la communauté du compte, le champ libre permet également de filtrer le graphique sur une ou des communautés.

6.2 Bandeau général

L’outil de visualisation permet plusieurs modulations du même graphe. Dans les paramètres qui suivront, le graphe ne change pas, c’est-à-dire que les comptes restent reliés entre eux par les mêmes liens, mais des choix existent sur leur disposition sur le plan ou sur la modification de leurs tailles selon plusieurs critères. Nous notons qu’il n’y a pas une visualisation qui sera meilleure qu’une autre, elles se complètent et permettent de mettre en valeur des critères différents.



- **Layout** : permet de choisir entre disposition des comptes, l’algorithme circlePack regroupe ces derniers par communautés, alors que Force Atlas 2 les disposent par interactions il fait donc ressortir les liens entre les communautés.



- **Hop number** : ce curseur permet de reprendre le graphe en n'affichant que les comptes et liens fait par strate.
- **Node size** : augmente la taille en fonction de 3 critères :
 - **Indegree** : nombre de fois que le compte a été mentionné ou retweeté, cela mettra en évidence les comptes les plus connus, ou les plus influents, mais réduira les comptes les plus actifs et dont les tweets ne sont pas beaucoup repris,
 - **Out-degree** : nombre de tweets produits par le compte mentionnant ou retweetant un autre compte, cela mettra en évidence les comptes les plus prolifiques, mais effacera les comptes influents, mais peu prolifiques.
 - **Degree** : c'est le cumul des deux derniers, soit le nombre de mentions du compte auquel est ajouté son nombre de tweets effectué.
- **User** : permet de rechercher un compte précis, il sera affiché en vert pomme.
- **Hashtag** : n'affiche que les comptes qui ont tweeté au moins une fois ce hashtag.