

# Qui est le leader d'opinion le plus influent sur Twitter en 2021 ?

## Usage de l'analyse des réseaux sociaux

*Florine COMLAN  
Jordy HOUNSINO  
Ramya HOUNTONDI  
Soulane Kabirou KANLANFEYI*

### Abstract

Considérée comme l'une des grandes thématiques de l'industrie 4.0, l'Intelligence Artificielle (IA) n'échappe pas aux discussions sur les réseaux sociaux qui sont aujourd'hui l'un des plus grands moyens de communication et d'influence. En se basant sur la fonction hashtag très en vogue sur ces derniers notamment sur Twitter, nous proposons une analyse de ce réseau social afin d'identifier le personnage le plus influent sur l'hashtag #IA durant une période donnée. Cette étude sera suivie d'une recommandation pour devenir influent sur Twitter en se basant sur des critères pertinents.

### INTRODUCTION

Les réseaux sociaux représentent aujourd'hui pour les entreprises une arme redoutable qui leur permet de s'assurer d'atteindre leur cible lors de campagnes de communication. Pouvoir identifier le bon canal pour faire passer leur message est primordial.

Si vous avez déjà cherché à identifier à quel compte twitter confier une campagne de communication sur l'intelligence artificielle ou encore quel compte suivre pour être sûr de ne rien rater, notre article qui propose une démarche d'étude permettant de mettre la main sur les tweetos à absolument suivre par rapport à ce sujet vous intéressera certainement.

#### I. CRITERES UTILISES POUR IDENTIFIER LES LEADERS D'OPINION SUR TWITTER

Afin de réaliser une analyse pertinente, nous avons défini les critères qui selon nous peuvent permettre d'identifier un leader d'opinion sur un réseau. Ces critères sont ceux qui, selon nous correspondent le plus à nos objectifs. Il s'agit de :

- **Eigen centrality** (centralité de vecteur propre) : En théorie des graphes, il permet de mesurer l'influence d'un nœud sur un réseau en fonction du nombre de liens qu'il possède avec d'autres nœuds du réseau. Il prend également en compte la qualité de la connexion d'un nœud et le nombre de lien de ses connexions à travers tout le réseau. Cet indicateur fonctionne de sorte à identifier les nœuds ayant une influence sur l'ensemble du réseau. Il s'agit donc d'un bon indicateur d'analyse.

- **Authority** : Permet de mesurer la pertinence de l'information contenue dans un compte associé à un acteur d'un réseau.
- **PageRank** : est une variante d'Eigen Centrality. La différence est que le PageRank prend également en compte le poids des liens. Il permet de découvrir les nœuds dont l'influence s'étend au-delà de leurs connexions directes dans un réseau plus large. Cet indicateur peut être utile pour comprendre l'autorité du nœud dans le réseau.
- **Weighted indegree** : est un indicateur qui nous fournit un classement des acteurs du réseaux avec le plus de mentions. Ce qui peut permettre d'observer le niveau d'écoute et de reconnaissance des acteurs du réseau.
- **Weighted degree** : mesure l'engagement d'un acteur dans un réseau indépendamment du nombre de voisins auxquels il est connecté .
- **Betweenness centrality** : mesure le nombre de fois où un nœud se trouve sur le chemin le plus court entre les autres nœuds. Il permet ainsi de faire voir quels nœuds sont des « ponts » entre les autres nœuds d'un réseau. Il est très utile pour analyser la dynamique de la communication dans un réseau.
- **Weighted outdegree** : classe les acteurs en fonction du nombre de tweets où ils ont mentionné d'autres comptes.

## II. PRESENTATION DU JEU DE DONNEES ET STATISTIQUES DESCRIPTIVES

Une fois les objectifs d'étude clairement définis, la première étape a consisté à collecter les données qui permettraient de les atteindre. Nous avons choisi de travailler avec les données fournies dans le challenge afin de bénéficier d'informations sur une période plus longue.

En effet une collecte a été faite avec python via l'API de Twitter. Malheureusement, cette API ne permet d'obtenir que 7 jours de données avec l'abonnement étudiant. Nous avons donc jugé qu'il était plus intéressant de se servir du jeu de données fourni qui lui nous offrait un champ d'observation sur environ 3 semaines (du 28 mars au 11 avril 2021).

Ces données une fois transformées présentaient les paramètres suivants : la date du tweet, le nom de la source (tweeter), le nom de la destination s'il y en a (retweeter), le contenu du tweet, le nombre d'abonnés du tweeter, sa localisation, son nombre d'abonnement, son nombre de tweets, la description de son profil et la relation entre la source et la destination.

Notre jeu de données est composé au total de **7767 observations** qui correspondent au nombre de liens (tweet et retweet) et de **4718 nœuds** (tweetos). Les tweets sont majoritairement localisés en France plus précisément en région parisienne comme le montre le graphique ci-dessous.

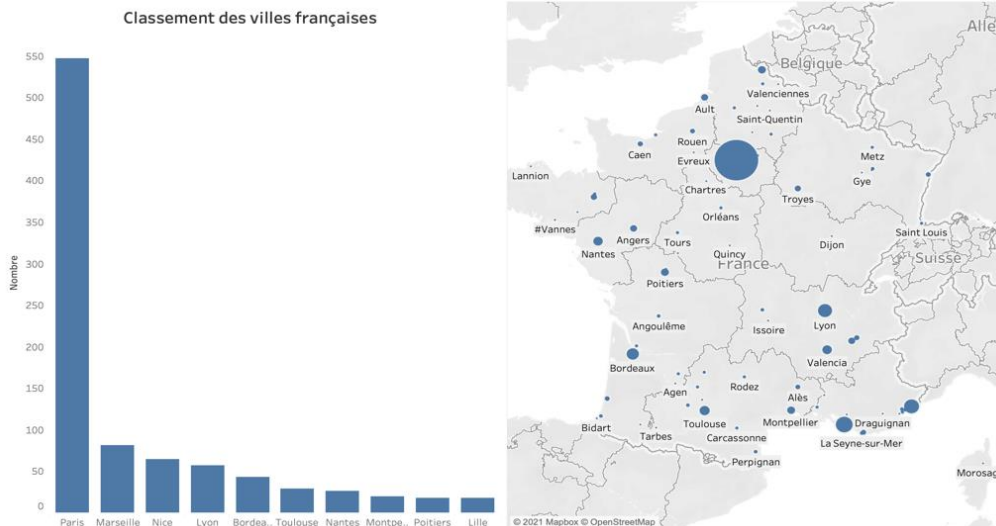


Fig 1 : Cartographie des sources de tweets en France

### III. METHODOLOGIE ET ETAPES SUIVIES POUR REALISER L'ANALYSE

- Création du jeu de données final

Trois fichiers Excel correspondant aux différentes semaines d'extraction ont été mis à notre disposition. La première étape de notre démarche fut l'agrégation de ces données afin d'obtenir un fichier unique sur lequel effectuer nos analyses. S'en est suivi le traitement du contenu de notre fichier notamment en nous servant des techniques de nettoyage et de dédoublonnage de données. Une fois terminé nous avons procédé à la fusion des différentes feuilles dudit fichier Excel grâce à l'outil Tableau ; ceci afin d'obtenir une feuille unique qui contient à la fois les nœuds et les liens de notre réseau.

- Cartographie grâce à l'outil Gephi

A cette étape, nous avons utilisé le logiciel Gephi pour une exploration complète de la structure du réseau. En y appliquant les différents critères d'analyse énumérés plus haut, nous avons généré différentes cartographies de notre réseau. Le facteur de spatialisation utilisé pour le déploiement est celui de Yifan Hu.

Par la suite, en nous basant sur les informations temporelles de notre jeu de données, nous avons visualisé l'évolution des interactions entre les différents nœuds de notre réseau sur les trois semaines considérées.

A la fin de l'exploration, l'outil Gephi nous permet d'obtenir des statistiques par rapport aux différents critères d'observation ; nous nous en sommes servis pour la suite de notre analyse.

- Classement des critères d'influences, scoring puis détection des nœuds influents du réseau

En se servant des statistiques générées par Gephi, nous avons créé une nouvelle liste contenant les 20 premiers nœuds les plus importants par critère d'analyse.

Afin de déterminer un classement unique en prenant en compte tous les critères d'influence à la fois, nous les avons rangés par ordre d'importance et leur avons affecté une pondération (fig 3). Cette pondération a été assignée à notre liste regroupant les nœuds selon leurs rangs pour que nous obtenions un premier classement d'influenceurs. Une dernière vérification est faite pour s'assurer de la justesse et de la pertinence du classement grâce à un traitement du contenu des pages associés.

- Analyse du contenu des tweets et création d'un Wordcloud

A partir du contenu des tweets et des hashtags présents, nous avons fait une analyse textuelle en python afin de déterminer les mots les plus fréquents. Lors de cette analyse ont été retenus uniquement les mots en français et en anglais. Certains tweets étaient en chinois et en espagnol mais nous avons choisi de ne pas les garder. S'en est suivi la représentation visuelle (fig 7). Les mots principaux qu'on peut y lire correspondent bien au sujet d'étude.

#### IV. RESULTATS

- Les cartographies des acteurs les plus influents en fonction des différents critères d'analyse et leur regroupement en communauté

Ce regroupement a été fait en se basant sur la classe de modularité avec une spatialisation dont la répartition selon les clusters est renseignée sur la figure suivante :

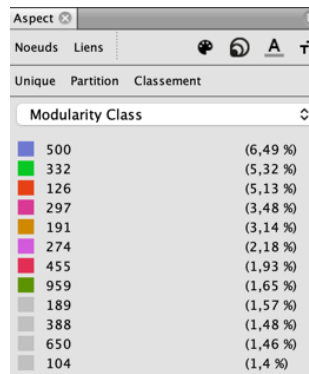
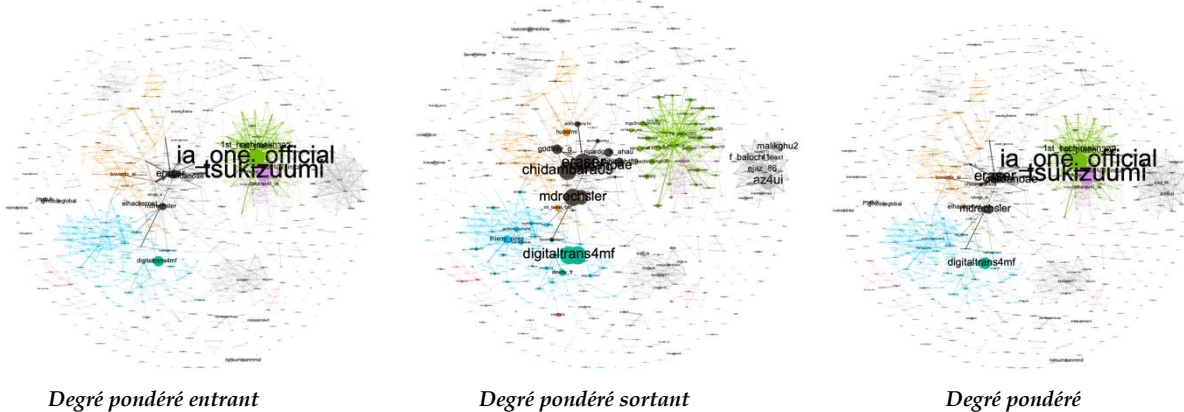
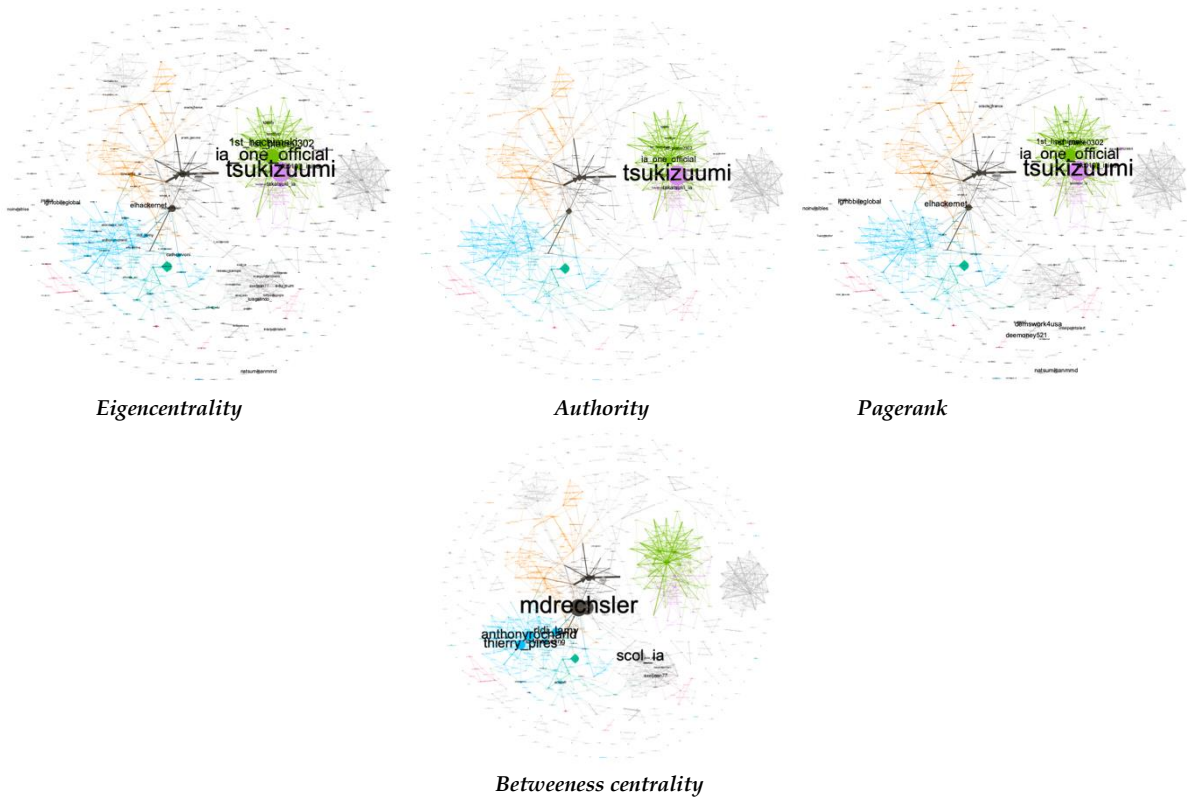
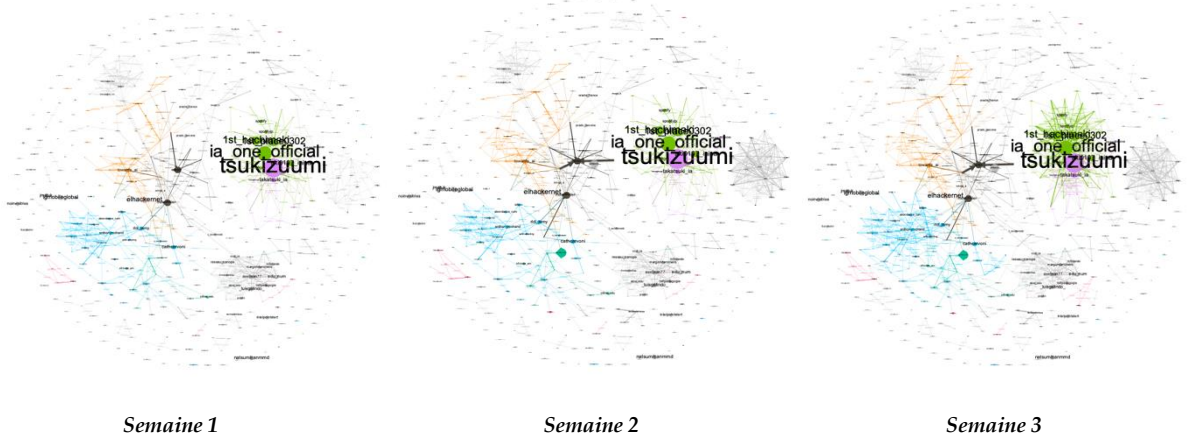


Fig 2 : Répartition des clusters par classe de modularité





○ La dynamique du réseau sur les 3 semaines qui font l'objet de notre analyse



L'interaction entre les membres des réseaux prépondérants reste plus ou moins stable tout au long des trois semaines ce qui justifie la force de leurs relations.

○ La pondération appliquée aux critères d'analyses suivant leur importance

	Eigencentrality	Authority	Pageranks	Weighted indegree	Weighted degree	Betweenness centrality	weighted outdegree
Classement	1er	2eme	3eme	4eme	5eme	6eme	7eme
Pondération	[20, 10]	[19, 9]	[18, 8]	[17, 7]	[16, 6]	[15, 5]	[12, 2]

Fig 3 : Tableau de pondération

○ Le top 15 des influenceurs après application de la pondération

Influenceurs	
ia_one_official	111,00
1st_place0302	100,00
tsukizuumi	97,50
1st_hachimaki	72,50
mdrechsler	63,00
eraser	61,50
cardanoae	59,50
aki0107_inica	58,00
elhackernet	46,00
takatsuki_ia	41,50
towards_ai	37,50
gfk_ee	36,50
digitaltrans4mf	35,50
lgmobileglobal	29,00
jmatuk	27,50

Fig 4 : Classement intermédiaire des influenceurs pour l'hashtag #IA

Après la phase d'analyse des profils des influenceurs de ce classement, nous avons remarqué que certains bien que contenant des hashtags cible de notre recherche ne correspondaient pas du tout à des profils du domaine spécifié. L'exemple type est le premier du classement actuel **ia\_one\_officiel** qui n'est en réalité que le compte officiel d'un vocaloïd (logiciel de synthèse vocale développé par Yamaha). Il était donc nécessaire de revoir le classement tout en excluant ces profils dont les contenus ne sont pas pertinents pour notre étude.

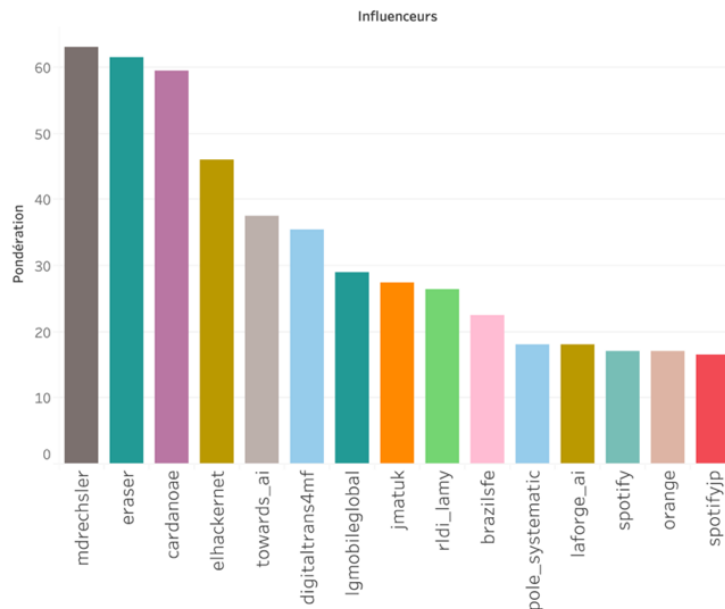


Fig 5 : Top 15 des influenceurs liés au #IA

Suite à nos résultats, nous avons pu dégager le podium et le top 15 des leaders d'opinion sur l'hashtag #IA sur la période considérée.

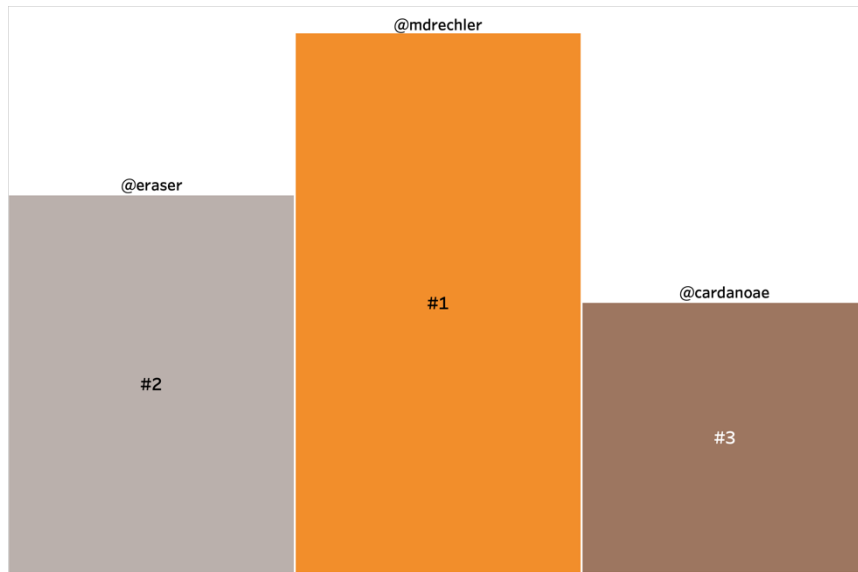


Fig 6 : Podium final des leaders d'opinion

Notre leader d'opinion identifié est @mdrechler. Michèle Drechsler (son vrai nom) détient un Doctorat en sciences de l'information et de la communication. Son parcours professionnel parsemé de nombreuses missions d'expertise, notamment celles menées au sein du Ministère de l'Education Nationale, justifierait son engagement et son leadership sur les sujets de l'IA.

### Les critères d'influences appliqués aux 15 tweetos les plus suivis en termes de connexion

Label	eigencentralit	Authority	pageranks	weighted indegree	weighted degree	betweenesscentrality	weighted outdegree	Eccentricity
mdrechler	0,157734	0	0,000709	75	135	0,000428	60	5
eraser	0,14416	0	0,000587	123	191	0,000179	68	2
cardanoae	0,072329	0	0,00066	85	143	0,000003	58	1
elhackernet	0,285583	0	0,003269	80	81	0	1	0
towards_ai	0,163932	0	0,000798	61	65	0	4	0
digitaltrans4mf	0,051576	0	0,000165	72	134	0	62	0
lgmobileglobal	0,232894	0	0,00267	68	68	0	0	0
jmatuk	0,176634	0	0,001443	67	68	0	1	1
ridi_lamy	0,172281	0	0,00076	31	43	0,00019	12	5
brazilife	0,010709	0	0,000071	27	52	0	25	0
pole_systematic	0,029882	0	0,000569	11	14	0,000001	3	1
laforge_ai	0,093566	0	0,000827	39	48	0	9	0
spotify	0,157532	0,102192	0,000762	28	28	0	0	0
orange	0,09683	0	0,00081	26	26	0	0	0
spotifyjp	0,157532	0,102192	0,000762	28	28	0	0	0

Fig 7: Tableau récapitulatif

- La visualisation du contenu des tweets



Fig 8 : Word cloud basé sur les mots contenus dans les tweets

## CONCLUSION

Même si nous avons utilisé des critères assez pertinents pour une identification précise de l'influenceur sur cette thématique, notre étude comporte néanmoins certaines limites qui sont à prendre à compte pour une analyse plus poussée.

### Limites de l'étude :

- ***Données non quantitatives***

Une étude sur durée plus large serait plus intéressante car elle nous permettra de déterminer si les leaders d'opinion identifiés le sont sur la durée ou s'il s'agit d'un effet de mode.

- ***Hashtag non exclusif***

Même si bon nombre des tweets présents dans notre jeu de données est lié à l'IA, nous avons néanmoins rencontré des incohérences sur le contenu de certains qui faisait référence à toute une autre thématique d'où la nécessité d'analyser le contenu de chaque tweet afin de s'assurer de leur pertinence.

- ***Élargissement du champ de recherche***

Nous pensons qu'il serait préférable de considérer d'autres hashtag supplémentaires liés à l'IA car même si #IA est assez représentatif, il n'englobe pas forcément tous les tweets liés à ce domaine donc certaines informations importantes pourraient nous passer sous le nez. Exemple d'hashtag à intégrer : #robot, #DataScience, #intelligenceartificielle, #data.

### Intérêt de l'analyse :



Cette analyse a un grand intérêt, surtout pour les entreprises souhaitant atteindre efficacement leur cible car elle leur permettra :

- D'identifier les acteurs clés ayant une influence sur une thématique donnée, ces derniers pouvant servir de canal à l'entreprise leur permettant ainsi :
  - D'avoir plus de visibilité sur les réseaux sociaux
  - De maximiser leurs chances d'atteindre leurs objectifs de campagnes de communication.
  - De cibler le bon public
- De suivre les tendances grâce aux interactions de la communauté du leader d'opinion qui peuvent être divergentes.

### **Préconisation :**

Afin de devenir un leader d'opinion sur Twitter, il faudrait dans un premier temps construire un réseau de qualité qui ne repose pas seulement sur la quantité de nos connexions, mais aussi sur l'influence de ces derniers sur le réseau. Il faudrait également créer un contenu assez riche et pertinent afin d'avoir une interaction fréquente pas qu'avec nos relations mais également avec d'autres membres. La capacité d'interagir avec les plusieurs groupes est aussi un atout très important pour un utilisateur en quête d'influence.

En somme, l'analyse de réseaux sociaux est très indispensable dans ce monde connecté pour les entreprises. Elle leur permet de mieux se connecter à leur cible. Une bonne utilisation de ces techniques d'analyse se basant sur différents critères apporte un bon coup de boost aux relations entreprises-cible/clients. Hormis le côté Marketing, cette analyse leur est également bénéfique pour faire passer leurs messages au bon public.

### **RÉFÉRENCES**

<https://management-datascience.org/challenges/identifier-le-leader-dopinion-le-plus-influent-sur-twitter/>  
<https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/>  
<https://datascientest.com/wordcloud-python>  
<https://seinecle.github.io/gephi-tutorials/generated-html/convertng-a-network-with-dates-into-dynamic.html>  
[http://helios.mi.parisdescartes.fr/~lerb/a\\_pedagogiques/Rapports/QRS\\_Rapport%202002/Florentine/Rapport2.html](http://helios.mi.parisdescartes.fr/~lerb/a_pedagogiques/Rapports/QRS_Rapport%202002/Florentine/Rapport2.html)  
<https://master-iesc-angers.com/utilisation-du-logiciel-gephi-pour-lanalyse-cartographique/>