

# Rapport de stage

## Rôle des systèmes de veille sanitaire dans la détection et diffusion des signaux d'émergence des maladies infectieuses animales et zoonotiques

BOUDOUA, El Bahdja

De Janvier à Juillet 2021

**Encadrement :** Valentin Sarah, Arsevska Elena, Hautefeuille Claire

**Maitre de stage :** Arsevska Elena

**Etablissement/ Formation :** Université Toulouse Paul Sabatier – Master en Gestion  
intégrée des maladies animales et tropicales - Parcours biologie santé – 2<sup>ème</sup> année

**Etablissement d'accueil :** Centre de coopération internationale en recherche  
agronomique pour le développement –Unité ASTRE

## Sommaire :

Introduction .....	9
Synthèse bibliographique .....	10
1.1 Concepts et définition.....	10
1.1.1 La veille sanitaire .....	10
1.1.2 La Surveillance basée sur les indicateurs .....	10
1.1.3 La surveillance basée sur les événements.....	10
2. Systèmes de veille sanitaire.....	11
3. Fonctionnement de PADI-web .....	12
3.1 Acquisition/ collecte des articles .....	12
3.2 Nettoyage et déduplication.....	13
3.3 Analyse et classification .....	13
3.4 Communication .....	13
4. Analyse de la diffusion de l'information au sein d'un système de veille sanitaire .....	13
4.1 L'analyse de réseau .....	13
4.2 Métriques utilisées dans l'étude des réseaux de communication .....	14
5. Objectifs de travail .....	15
Matériel et Méthode .....	16
1. Collecte des données .....	16
2. Création de la base de données.....	16
3. Construction du réseau .....	20
4. L'analyse du réseau .....	21
Résultats .....	23
1. Analyses descriptives .....	23
2. Analyse de réseau.....	23
2.1 Caractéristiques des chemins.....	24
2.2 Caractéristiques des sources .....	24
2.3 Mesures de centralité.....	25
a. Centralité de degré.....	25
b. Centralité d'intermédiarité.....	26
c. Centralité de vecteur propre.....	26
d. Densité du réseau.....	27
e. Réactivité du réseau.....	27
3. Visualisation du réseau.....	28

3.1 Détection des communautés .....	29
4. Aspects de complétude et d'exactitude .....	29
5. Résultats sur réseau pondéré .....	30
Discussion .....	32
Conclusion.....	35
Annexes .....	36
Bibliographie.....	39

## Liste des figures :

Figure 1 : Cadre conceptuel de la veille sanitaire (Eilstein et al., 2012) .....	11
Figure 2: Etapes de fonctionnement de PADI-web (Valentin et al., 2019).....	12
Figure 3 : Vue globale d'un réseau de veille sanitaire (adapté de (Weber & Monge, 2011)).....	14
Figure 4:Etapes successives mises en œuvre pour analyser le flux d'informations relatives à l'influenza aviaire détectées par PADI-web.....	16
Figure 5:Exemple d'un rapport final détecté par PADI-Web.....	17
Figure 6: Schéma simplifié de la base de données Microsoft Access. ....	19
Figure 7: (A) Extrait d'un rapport de Focus news et (B) sa représentation sous forme de graphe .....	20
Figure 8: Distribution géographique des signaux d'influenza aviaire observés durant la période d'étude .....	23
Figure 9: Longueur des chemins .....	24
Figure 10: réactivité des chemins.....	24
Figure 11:Types de sources primaires et secondaires impliquées dans le réseau.....	25
Figure 12: Distribution de degrés des sources.....	25
Figure 13: Réactivité de PADI-Web. ....	28
Figure 14: Réactivité observée chez les deux catégories d'hôtes : oiseaux domestiques et oiseaux sauvages .....	28
Figure 15 : Vue globale du réseau d'information sanitaire .....	29
Figure 16 : Communautés au sein du réseau sanitaire.....	29
Figure 17 : Proportion des informations de localisation (A), sous-type (B), dates d'observations du foyer (C) et nombre de cas (D).....	30
Figure 18: distribution du poids des liens au sein du réseau d'information sanitaire.....	30
Figure 19: Zoom sur les liens qui diffusent le plus grand nombre d'informations. ....	31

## Liste des tableaux

Tableau 1 : Les différents types de sources retrouvées dans le réseau d'informations sanitaires de PADI-web.....	18
Tableau 2: Extrait de la table signale .....	19
Tableau 3 : Aperçu de la table des chemins .....	20
Tableau 4 : Aperçu de la table Chemin-Signal.....	20

Tableau 5: Les mesures utilisées et leur interprétation dans le contexte de cette étude. Adapté de (Giuliani & Pietrobelli, 2011).....	22
Tableau 6: Classement des 5 premières sources ayant les valeurs de degrés la plus élevée .....	26
Tableau 7 : Classement des 5 premières sources ayant les valeurs de degré entrants et sortants les plus élevées .....	26
Tableau 8: Les 5 premières sources présentant les valeurs de centralité d’intermédierité les plus élevées .....	27
Tableau 9: Les 5 premières sources présentant les valeurs de centralité de vecteur propre les plus élevées .....	27
Tableau 10: Classement des 5 premiers liens ou diffuse le plus grand nombre de signaux dans le réseau.....	31

## Liste des acronymes :

**ADNS** : Animal Disease Notification System

**CIDRAP** : Center for Infectious Disease Research and Policy (Centre de recherche en maladies infectieuses)

**EMPRES-i** : Global Animal Disease Information System (Système de prévention et de réponse rapide contre les ravageurs et les maladies transfrontières des animaux et des plantes)

**ESA** : Plateforme Nationale d’épidémiosurveillance en Santé Animale (plateforme ESA)

**FAO** : Food and Agriculture Organisation – Organisation pour l’alimentation et l’agriculture

**IA** : Influenza aviaire

**OIE** : Organisation mondiale de la santé animale (Office internationale des épizooties)

**PADI-web**: Platform for Automated extraction of animal Disease Information from the Web

**PPA** : Peste porcine africaine

**USDA** : Département de l’agriculture des Etats-Unis

**VSI** : Veille Sanitaire Internationale

**WAHIS** : World Animal Health Informaton System (Système mondial d’information sanitaire)

**WHO** : World Health Organisation - Organisation Mondiale de la Santé (OMS)

## **Liste des annexes**

**Annexes A** : Comparaison de trois systèmes de veille sanitaire.

**Annexe B** : Indicateurs d'évaluation des systèmes de veille sanitaire.

**Annexe C** : vue globale de la base de données Microsoft Access.

## Remerciements

Tout d'abord je tiens à remercier Elena Arsevska pour sa confiance, sa bienveillance et pour tout ce qu'elle m'a appris, travailler aux côtés de son équipe durant ces derniers mois était une expérience extrêmement enrichissante, aussi bien sur le plan scientifique que sur le plan humain.

Merci également à Sarah Valentin pour sa gentillesse, sa disponibilité et tout ce qu'elle a apporté à ce travail.

Un grand Merci à Claire Hautefeuille pour sa patience, ses relectures et ses précieux conseils que je garderai bien en tête !

Et pour finir merci à toute l'équipe du MOOD café, nos réunions hebdomadaires étaient aussi rafraichissantes que motivantes.

## Résumé

Les maladies émergentes représentent un risque croissant pour la santé publique, la rapidité de détection de ces émergences et d'évaluation des risques qu'ils représentent est cruciale.

Pour faire face à ces nouvelles menaces, de nombreux pays adoptent une stratégie de veille sanitaire aussi appelée « intelligence épidémique ».

La veille sanitaire comprend toutes les activités liées à la détection précoce de dangers sanitaires en s'appuyant sur deux types d'informations : celles issues de sources « officielles », déclarées par les services de santé publique ; et celles issues de sources « non-officielles », qui relaient les informations trouvées dans les médias, les articles scientifiques, les données de laboratoires ou encore les réseaux sociaux.

Plusieurs systèmes de veille sanitaire (ProMed, HealthMap etc.) sont dédiés à la collecte et à la diffusion des données issues des sources non-officielles. En France la cellule de Veille Sanitaire Internationale (VSI) de la plateforme d'Epidémiosurveillance en Santé Animale (ESA) utilise depuis 2014 le système de veille sanitaire PADI-web (Platform for Automated extraction of Disease Information from the Web) pour compléter sa composante événementielle et assurer la détection, la vérification et communication des signaux de maladies infectieuses.

Le but de ce travail était de comprendre comment l'information sanitaire (signal) se propage d'une source primaire (émetteur) à une source finale (système de veille sanitaire) à travers des méthodes quantitatives et qualitatives d'analyses de réseau.

Pour se faire, les signaux d'influenza aviaire détectés par PADI-web entre 2018 et 2019 ont été analysés. Les résultats ont montré que l'OIE occupe une position centrale dans le réseau d'information sanitaire de PADI-web et que les médias en ligne suivis des agences de presse sont les principales sources secondaires (émettrices-réceptrices) du réseau. Il a également été constaté que la sensibilité de PADI-web par rapport à l'influenza aviaire avait augmenté de 10% en comparant aux résultats obtenus en 2016. Concernant la détection précoce, une différence a été observée entre les deux catégories d'hôtes étudiées. Chez les oiseaux sauvages, les délais entre la détection par PADI-web et la notification officielle peuvent être supérieurs à 3 mois. Chez les oiseaux domestiques, cela n'excédait pas les 33 jours.

En vue d'améliorer PADI-web, il serait intéressant de le comparer sur le même cas d'étude, à d'autres systèmes de veille sanitaire comme HealthMap ou PromedMail.

## Abstract

Emerging diseases represent an increasing public health risk, their fast detection and risk evaluation is crucial.

To face these new threats, numerous countries have adopted an « epidemic intelligence » strategy. This strategy encompasses all activities related to early detection by focusing on two types of information: public health « official sources » reports and « unofficial sources » which includes information found in media outlets, scientific publications, research laboratory data, and social media platforms.

Several event based surveillance (EBS) systems (ProMed, HealthMap, etc) are dedicated to the collection and dissemination of data from unofficial sources. In France, the Epidemic Intelligence System (VSI) of the Animal Health Epidemiological Surveillance platform (ESA) has been using PADI-web (Platform for Automated extraction of Disease Information from the Web) event based surveillance system since 2014 to complete its event component and ensure the detection verification and communication of infectious disease signals.

The goal of this work was to understand health information (signal) spread from a primary source (transmitter) to a final source (EBS system) through quantitative and qualitative network analysis methods.

To do this, avian influenza signals detected by PADI-web between 2018 and 2019 were analysed. The results showed that the OIE occupies a central position in the PADI-web information network and that online media followed by press agencies are the main secondary sources. It was also found that the sensitivity of PADI-web to avian influenza increased by 10% compared to the results obtained in 2016. In terms of early detection, a difference was observed between the two categories of hosts. In wild bird the delay between the detection by PADI-web and the official notification can be longer than 3 months while in domestic birds it does not exceed 33 days. To improve PADI-web, it would be interesting to compare the same study case to other event based surveillance systems like ProMedMail or HealthMap.

## Introduction

Dans un monde en rapide évolution des plateformes de partage d'information sanitaire sur le web (média, réseaux sociaux, etc.), les services nationaux chargés de la sécurité sanitaire ont de plus en plus recours aux nouvelles technologies pour identifier l'émergence de risques des maladies infectieuses (ré-) émergentes à travers de ces différentes plateformes.

La veille sanitaire, appelée aussi intelligence épidémiologique, fournit un cadre conceptuel qui permet aux pays d'adapter leurs systèmes de surveillance national pour répondre à ce nouveau défi d'accès à l'information sanitaire issue de multiples sources en ligne.

La veille sanitaire intègre à la fois une composante basée sur les indicateurs et une composante basée sur les événements (veille de médias et autres sources informelles) pour compléter les données issues de la surveillance basée sur les indicateurs et surmonter les limites de la surveillance traditionnelle, dans un souci de détection précoce et aussi pour reconnaître l'émergence de nouvelles menaces telles que le Syndrome respiratoire aigu sévère (SARS) et la grippe aviaire (Paquet et al., 2006).

En France, dans le cadre de Plateforme Nationale d'épidémiosurveillance en santé animale (Plateforme ESA), une cellule de « veille sanitaire internationale (VSI) » existe depuis 2013. La VSI vise à identifier, suivre et analyser les signaux des dangers sanitaires (en santé animale au sens large) menaçant le territoire français dans son ensemble, dans un but de production d'information sanitaire à des fins d'évaluation du risque d'une part (Anses), et de gestion du risque d'autre part (DGAL) (Arsevska et al., 2018).

PADI-web (Platform for Automated extraction of animal Disease Information from the Web) est un système de veille automatique d'articles médias publiés sur le Web, dédié à la surveillance des maladies animales. Il est utilisé depuis 2014 par la cellule de Veille sanitaire internationale (VSI) et a pour but de compléter sa composante événementielle (Arsevska et al. 2018).

Plusieurs études ont évalué les performances des systèmes de veilles sanitaires en termes de sensibilité, de réactivité par rapport à des maladies spécifiques, mais l'on sait peu de choses sur la manière dont l'information sanitaire circule sur le web avant d'être détectée par ces systèmes.

Le but de ce travail est de comprendre comment l'information sanitaire (signal) se propage d'une source primaire (émetteur) à une source finale (système de veille sanitaire) à travers des méthodes quantitatives et qualitatives d'analyses de réseau (topologie des sources, mesures de centralités etc.), de caractériser le réseau en terme de réactivité en explorant les délais entre la communication et la détection d'un signal et enfin d'analyser la véracité de l'information en la comparant avec les informations publiées officiellement par les organismes internationaux (OIE, FAO).

## Synthèse bibliographique

### 1.1 Concepts et définition

#### 1.1.1 La veille sanitaire

Dans le but de couvrir l'ensemble des risques sanitaires de nombreux pays adoptent une stratégie dite de veille sanitaire (Barboza, 2015). La veille sanitaire comprend toutes les activités liées à la détection précoce de dangers sanitaires susceptibles de présenter un risque pour la santé. Ces activités comprennent l'évaluation du risque, nécessaire à la mise en place de mesures de prévention et contrôle appropriées, elle repose sur deux composantes : la surveillance des indicateurs et la surveillance des événements (Paquet et al., 2006).

#### 1.1.2 La Surveillance basée sur les indicateurs

La surveillance basée sur les indicateurs est définie comme la collecte, le suivi, l'analyse et l'interprétation systématiques de données structurées c'est-à-dire d'indicateurs (nombre de cas ou foyers, la prévalence, le taux de mortalité, un seuil d'incidence franchi etc.) (WHO, 2014) et repose sur des définitions de cas établies.

Les organisations internationales sont les principaux acteurs de développement des systèmes de surveillances basés sur les indicateurs (Valentin, 2020), et il existe de nombreux systèmes de surveillance au niveau régional, national ou local. En santé animale, les notifications relatives aux dangers sanitaires sont principalement collectées par 3 systèmes (Cáceres et al., 2017).

- Le Système mondial d'information zoo sanitaire (WAHIS) de l'organisation mondiale de la santé (OIE)
- Le Système de prévention et de réponse rapide contre les ravageurs et les maladies transfrontières des animaux et des plantes (Empres-i) de l'organisation des nations unies pour l'alimentation et l'agriculture (FAO), et au niveau européen
- le système de communication des maladies animales (ADNS).

Ils sont souvent considérés comme gold-standard dans l'évaluation des systèmes de surveillance basés sur les événements car les informations recueillies par ces systèmes sont officiellement validées par les services sanitaires (Barboza, 2015).

#### 1.1.3 La surveillance basée sur les événements

La surveillance basée sur les événements est définie par l'OMS comme « la collecte organisée, le suivi et l'évaluation, et l'interprétation d'informations *ad-hoc*, principalement non structurées (sans format prédéfini) concernant des événements pouvant représenter un risque aigu pour la santé humaine (ou animale) (WHO, 2014). Les informations sanitaires recueillies proviennent de sources multiples ; officielles (OMS, OIE, FAO) ou non-officielles (médias en ligne, plateforme sociale etc.) (Barboza, 2015). La détection et la collecte de ces informations est assurée par des systèmes de veille sanitaire (Paquet et al., 2006).

Le dispositif de veille sanitaire (Figure 1) permet donc le recueil d'événements inhabituels de santé pouvant représenter une menace pour la santé publique et non pris en compte par les systèmes de surveillances traditionnels (Ilef et al., 2011).

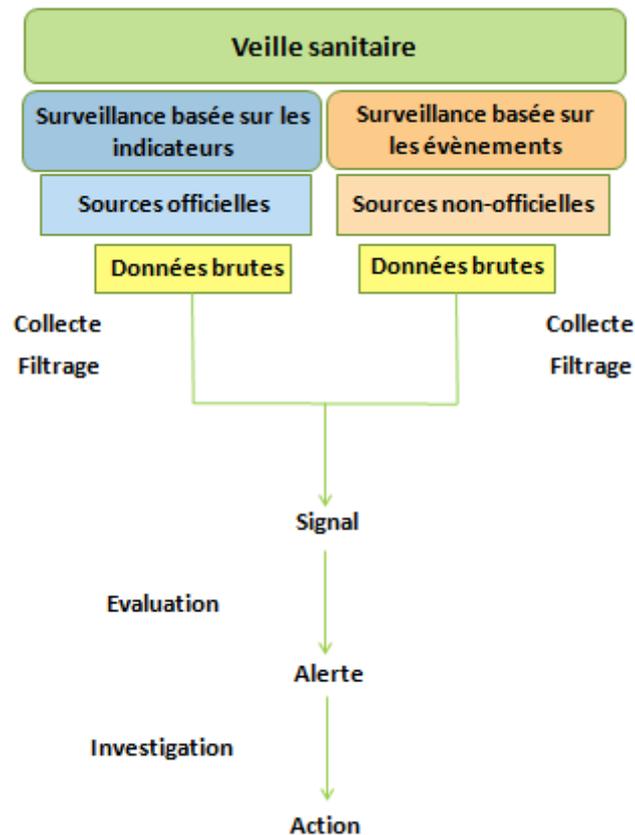


Figure 1 : Cadre conceptuel de la veille sanitaire (Eilstein et al., 2012)

## 2. Systèmes de veille sanitaire

Plusieurs systèmes de veille sanitaire (HealthMap, ProMed etc.) sont dédiés à l'acquisition et à la diffusion de données issues de sources non-officielles. Ces systèmes s'intéressent à un large éventail de risques sanitaires (maladies infectieuses humaines, animales ou végétales, risques environnementaux) (Valentin, 2020).

Certains systèmes sont entièrement automatisés, d'autres bénéficient d'une modération humaine. Néanmoins, ils présentent tous des caractéristiques similaires : ils collectent automatiquement, à partir d'Internet pour compléter les données issues de la surveillance basée sur les indicateurs (Paquet et al., 2006) et détecter de manière précoce de nouvelles émergences (Arsevska, 2017).

Annexe A comparaison de 3 systèmes de veille sanitaire

Les indicateurs utilisés par l'OMS pour évaluer les systèmes de surveillance des maladies infectieuses ont été adaptés au contexte de l'évaluation des systèmes de veille sanitaire (WHO, 2014) ainsi plusieurs études ont permis d'évaluer leurs performances par rapport à des maladies spécifiques, et leur capacité à détecter les menaces sanitaires précocement est reconnue. (Annexe B indicateurs d'évaluation).

En 2003 les rapports quotidiens envoyés par ProMedMail pour couvrir l'épidémie SARS ont contribué à avertir le personnel médical d'un hôpital de Toronto, ayant pris connaissance de cette maladie mortelle et

contagieuse, ils ont pu isoler les patients lorsque la maladie y est apparue pour la première fois (Yu & Madoff, 2004).

En 2009 durant l'épidémie de H1N1 le système HealthMap a donné des résultats impressionnants, avec des délais de 12 jours en moyenne entre la notification de cas suspects et de cas confirmés, cela peut considérablement modifier l'impact de la réponse sanitaire (Kman & Bachmann, 2012).

En Janvier 2014 ProMedMail était le premier système à communiquer des cas de peste porcine africaine chez les sangliers en Lituanie suite à des alertes émises par des chasseurs locaux (Carrion & Madoff, 2017).

En 2020 PromedMail fourni des informations utiles permettant d'identifier de nouveaux foyers d'épidémies pendant la pandémie de Covid-19 (Bonilla-Aldana et al., 2020).

En France la cellule de Veille Sanitaire Internationale (VSI) de la plateforme d'Epidémiosurveillance en Santé Animale (ESA) utilise depuis 2014 le système de veille sanitaire PADI-web (Platform for Automated extraction of Disease Information from the Web) pour compléter sa composante événementielle et assurer la détection, la vérification et communication des signaux de maladies infectieuses. a la différence des systèmes de veille sanitaire précédemment cités, PADI-web est axé sur les maladies animales et zoonotiques, et les maladies actuelles d'intérêt pour la VSI sont la peste porcine africaine (PPA), la fièvre aphteuse, la fièvre catarrhale du mouton, et l'influenza aviaire(Arsevaska et al., 2018).

### 3. Fonctionnement de PADI-web

Le fonctionnement de PADI-web repose sur 4 étapes successives (Figure2).

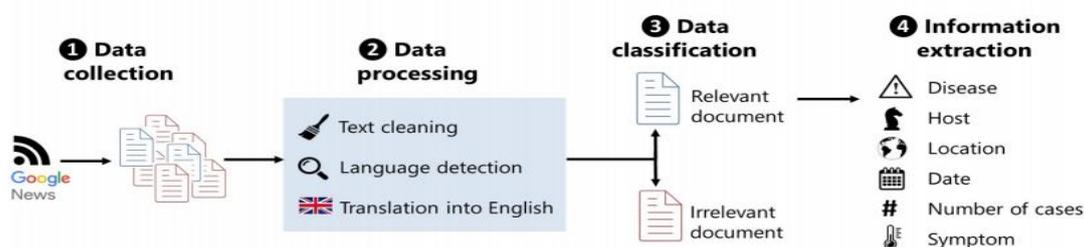


Figure 2: Etapes de fonctionnement de PADI-web (Valentin et al., 2019).

#### 3.1 Acquisition/ collecte des articles

La plupart des systèmes de surveillance basée sur l'événement collectent des données de façon automatique. La collecte est principalement réalisée à partir de sources non officielles sur le web, tels que les flux RSS (Freifeld et al., 2008).

Padi-web aspire des articles quotidiennement via l'agrégateur Google News grâce à deux types de requêtes :

- Les requêtes spécifiques incluent les noms de maladies connues (par exemple « Avian Flu » OR « Avian Influenza »)
- Les requêtes non spécifiques qui permettent de détecter des menaces non prédéfinies, consistent en une combinaison de symptômes et de noms d'hôtes (par exemple, « fever » AND « wild boar ») (Valentin et al., 2019)

### 3.2 Nettoyage et déduplication

Le contenu des articles est nettoyé et tout élément inutile est supprimé (images, publicités, etc.), et le contenu des articles est enregistré dans la base de données PADI-web accompagné du titre, date de publication et nom de la source (Valentin et al., 2019)

Le dédoublement (déduplication) consiste à identifier et à supprimer les documents identiques ou dont le contenu est identique (Barboza, 2015). PADI-web filtre les doublons en comparant les URL des articles nouvellement aspirés à ceux déjà présents dans la base de données (Valentin et al., 2019)

### 3.3 Analyse et classification

L'analyse des données collectées par les outils de veille sanitaire est délicate car, dans la plupart des cas, il s'agit de textes non structurés nécessitant d'être préalablement traités afin d'en extraire l'information désirée (Arsevska, 2017).

Cette étape permet de filtrer la quantité d'articles qui sera présentée à l'utilisateur en rejetant les articles impertinents (non liés à un danger sanitaire). Le classifieur (c'est-à-dire le programme informatique qui fait le tri) est issu d'un apprentissage automatique supervisé (le corpus d'apprentissage est un corpus annoté de 600 articles, 400 pertinents et 200 impertinents). Un échantillon de 100 articles aléatoirement sélectionné a été évalué manuellement par deux épidémiologistes et a obtenu une exactitude "*accuracy*" de 0.92 (Valentin et al., 2019).

### 3.4 Communication

Un des principaux moyens de diffuser l'information est sous la forme de page web comprenant une interface composée de cartes géographiques des lieux de foyers ou d'événements sanitaires (Arsevska, 2017). Les articles stockés dans la base de données PADI-Web sont consultables via une interface dédiée (Valentin et al., 2019).

## 4. Analyse de la diffusion de l'information au sein d'un système de veille sanitaire

### 4.1 L'analyse de réseau

L'analyse de réseau désigne un ensemble de méthodes, de notions, et de concepts fondés sur la théorie des graphes pour étudier un phénomène relationnel donné, en symbolisant les différents acteurs (individus, institutions etc.) par des points appelés « nœuds » et les relations entre ces acteurs par des lignes appelés « liens » (Beauguitte, 2016).

Depuis leur introduction en 1954, l'analyse des réseaux sociaux est un domaine de recherche très actif, bien que les premiers travaux ont principalement concerné les sciences sociales, leur capacité à modéliser un très grand nombre de phénomènes, a attiré d'autres communautés scientifiques, par exemple en épidémiologie pour étudier et visualiser la propagation d'une maladie infectieuse (Stattner & Vidot, 2011). Cependant très peu d'études utilisent les méthodes d'analyse de réseau pour comprendre comment l'information se propage dans un réseau de communication (Kloth et al., 2019), (Sabot et al., 2017).

Une étude comparative menée sur le réseau d'information Twitter a montré que même si un réseau est formé d'un grand nombre de sources, seules quelques-unes jouent le rôle d'acteurs clé dans la transmission de l'information (Yoon et al., 2020).

Une autre étude suggère que l'information sur le web circule selon le modèle « Source- Récepteur –émetteur- Récepteur ». Ce modèle montre que le réseau d'information est orienté (car l'information circule selon une direction donnée) et met en évidence 3 catégories d'acteurs différents selon leur rôle, les « Sources » sont les premières à émettre l'information, les « récepteurs-émetteurs » vont recevoir les informations, les filtrer et leur apporter un contenu supplémentaire avant de la transmettre aux « Récepteurs » (figure 3).

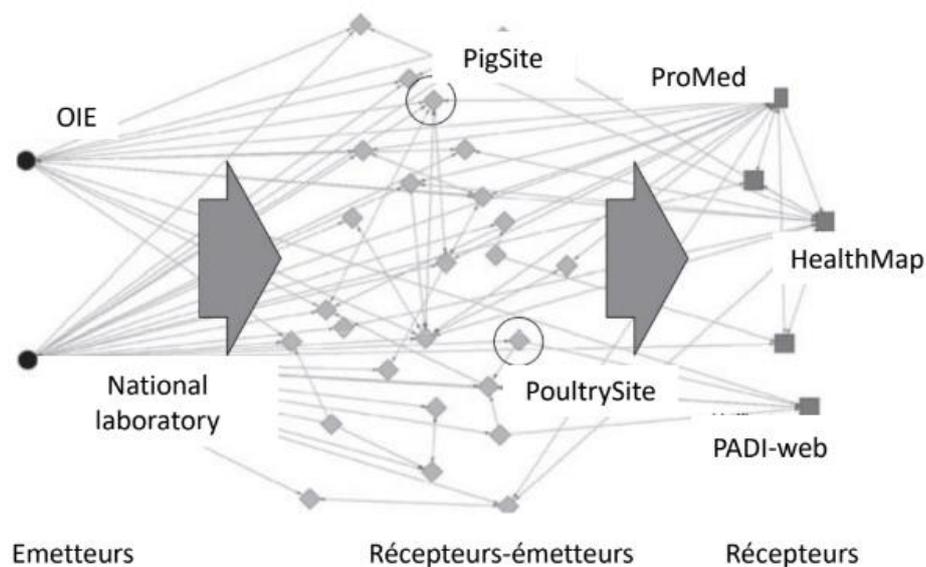


Figure 3 : Vue globale d'un réseau de veille sanitaire (adapté de (Weber & Monge, 2011)).

#### 4.2 Métriques utilisées dans l'étude des réseaux de communication

Selon la nature et les caractéristiques des liens, la position d'un acteur dans un réseau peut refléter son influence, son accès ou non aux informations (Giuliani & Pietrobelli, 2011). La mesure la plus simple est la centralité de degré, qui est le nombre de liens directement connectés à un nœud (Kirchhoff et al., 2008). Selon cette mesure, déterminer l'importance d'un nœud dans un graphe revient à calculer le nombre de ses sommets voisins, ou de manière équivalente, à calculer le nombre de liens qui lui sont incidents (Chikhi, 2010). Cette mesure est d'autant plus intéressantes lorsque le graphe est orienté car elle permet de représenter les liens entrant et sortants (Beauguitte, 2016), elle permet de savoir si un acteur a accès à un grand nombre d'information, si il communique l'information avec beaucoup de visibilité ou si il possède ces deux capacités combinées(Weber & Monge, 2011).

D'autres mesures de centralités servent à définir la situation d'un acteur du réseau via ses voisins directs (Gay et al., 2013). La centralité d'intermédierité par exemple, est particulièrement appropriée dans l'étude de réseaux de communication (Beauguitte, 2016), car elle permet de mettre en lumière les sources importantes dans la transmission de l'information. Les sources ayant une centralité intermédiaire élevée vont non

seulement filtrer et disséminer l'information mais aussi assurer une connexion entre les différents groupes du réseau (Yoon et al., 2020).

## 5. Objectifs de travail

En se basant sur une approche d'analyse de réseau, les informations détectées par PADI-web ont été étudiées avec pour objectifs de :

- Comprendre comment l'information circule au sein des systèmes de veille sanitaire en répondant aux questions de recherche suivantes :
  - Comment l'information sanitaire circule sur le web avant d'être détectée par PADI-web ?
  - Quel est le rôle des différentes sources dans la détection et dans la diffusion des informations ?
  - Quelles sont les premières sources à émettre l'information ?
- Caractériser le réseau en termes de réactivité et de qualité d'information :
  - Quelle est la réactivité du réseau (délai entre détection du signal et communication) ?
  - L'information transmise est-elle vraie et complète ?

Pour mener cette étude, l'influenza aviaire a été choisie comme maladie modèle car cette maladie virale représente un enjeu sanitaire et économique mondiale, aussi bien en santé animale qu'en santé humaine.

## Matériel et Méthode

Pour répondre à ces objectifs nous avons analysé le flux d'informations relatives à l'influenza aviaire détectées par PADI-web en suivant les quatre étapes décrites par la figure 4 pour construire le réseau de transmission de l'information sanitaire dans PADI-web.



Figure 4: Etapes successives mises en œuvre pour analyser le flux d'informations relatives à l'influenza aviaire détectées par PADI-web

### 1. Collecte des données

Tous les rapports relatifs à l'influenza aviaire publiés entre le 02 août 2018 et le 29 juin 2019 ont été extraits à partir de la base de données PADI-web, le choix de cette période est justifié par le fait que plusieurs événements d'influenza aviaire avaient été observés à ce moment-là dans le monde, notamment en Europe et Asie.

Dans cette étude, un rapport final est un contenu en ligne détecté par PADI-web décrivant un ou plusieurs foyers d'influenza aviaire.

Une seconde vérification (validation) a été faite afin de sélectionner uniquement les rapports pertinents et en extraire manuellement les informations épidémiologiques.

Un rapport pertinent a été défini comme étant un rapport final décrivant un foyer épidémiologique ou plus d'influenza aviaire chez les oiseaux domestiques et sauvages.

### 2. Création de la base de données

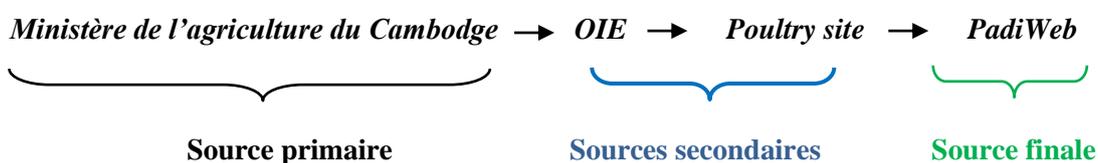
Dans un premier temps, tous les signaux décrits dans un rapport final ont été identifiés et classés en signaux officiels ou non officiels. Les signaux officiels sont des signaux trouvés dans le système mondial d'information sur les maladies animales de la FAO, Empres-i (considéré comme gold-standard). Les signaux non officiels sont des signaux déclarant un foyer d'influenza aviaire chez un hôte sensible (oiseaux domestiques ou sauvages) et absents de la base de données Empres-i.

Ensuite, les sources citées dans le rapport final ont été utilisées pour tracer le chemin de chaque signal. La figure 5 montre en exemple un rapport final détecté par PADI-web sur un foyer d'IA H5N6 au Cambodge.



Figure 5: Exemple d'un rapport final détecté par PADI-Web.

Le cheminement de l'information qui a permis d'obtenir ce rapport peut être illustré comme ceci :



- Source primaire : la source primaire est la première source à émettre le signal.
- Sources secondaires : Les sources secondaires sont toutes les sources impliquées dans la transmission de l'information et qui ne sont ni primaire ni finale.
- Source finale : La source finale est utilisée pour récupérer le rapport final. Dans notre étude, la source finale est toujours PADI-web (étant donné que les informations étudiées sont issues de sa base de données).

Chaque source était caractérisée par :

- **Spécialisation** : (spécialisée en santé vs générale). Certains types de sources sont spécialisés par défaut (par exemple, services vétérinaire). D'autres, comme les actualités en ligne, contiennent à la fois des sources spécialisées (par exemple PoultrySite) et générales (par exemple FocusNews).
- **Orientation géographique** (orientation nationale vs orientation générale). Certains types de sources ont une orientation géographique nationale ou locale par défaut (par exemple, les services vétérinaires). D'autres, comme les actualités en ligne, ont une portée nationale (par exemple, Times of India) ou internationale (par exemple, The Guardian).
- **Type de source** : au totale 11 types de sources ont été identifiés leurs définitions ainsi que des exemples sont indiqués dans le (tableau 1) ci-dessous.

<b>Type de sources</b>	<b>Définition</b>	<b>Exemple</b>
Organisation internationale	Organisation qui réunit plusieurs états représentés par leurs gouvernements.	OIE, FAO, OMS.
Autorités officielles nationales/locales	Toute source ayant le droit de prendre des mesures au niveau local ou national.	Ministères, douanes, police.
Services vétérinaires nationales/locales	Toute source ayant le droit de prendre des mesures sanitaires au niveau local ou national.	Autorités vétérinaires de Taiwan ou de Taipei.
Médias en ligne	Journaux utilisant internet comme support.	Gulf News, Focus News.
Organisations publiques/associations	Organismes publiques ou associations œuvrant dans le domaine de la santé et de l'environnement	Réseau d'information agricole (GAIN)
Laboratoires	Structure équipée dans le but de réaliser diverses analyses biologiques	Laboratoire vétérinaire central de Namibie
Personne	Toute personne citée comme source et qui n'entre dans aucune de ces catégories.	Un fermier, une source locale.
Compagnies privées	Organisme privé œuvrant dans le domaine de l'agroalimentaire	Compagnies de production de viande.
Agences de presse	Source qui collecte et communique les informations issues de différents journaux.	Xhinua, Reuters.
Radio/télévision	Chaîne de radio ou de télévision.	OdishaTV.
Plateforme sociale	Média en ligne permettant l'échange d'information entre les utilisateurs/abonnés	Avian Flu Diary

Tableau 1 : Les différents types de sources retrouvées dans le réseau d'informations sanitaires de PADI-web.

Une base de données a été créée sur Microsoft Access2019 pour stocker les données collectées, Cette base de données est formée d'un ensemble de tables reliées les unes aux autres comme le montre la figure 6. Chacune des tables comprend des données relatives aux informations collectées, ou aux composantes du réseau.

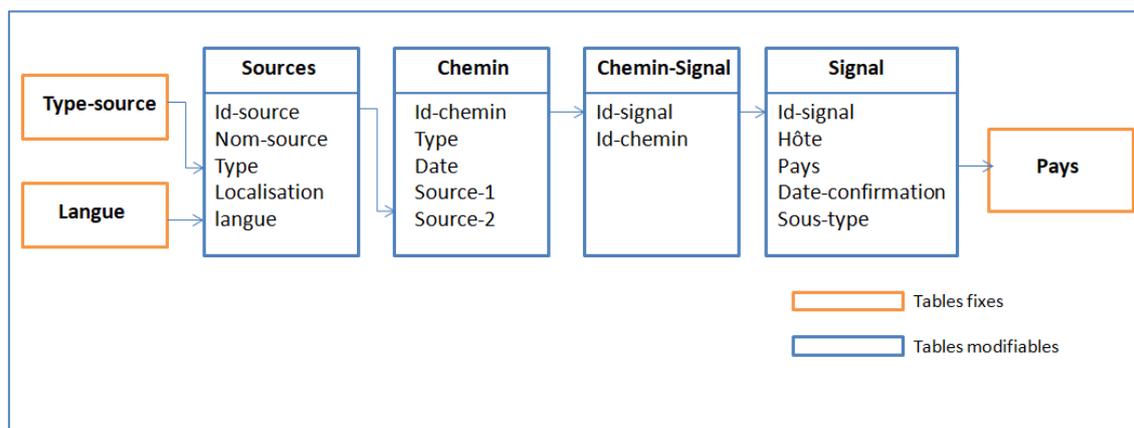


Figure 6: Schéma simplifié de la base de données Microsoft Access.

Le Schéma détaillé de la table Microsoft Access sont décrit en Annexe C

Les tables fixes servent à lister les pays et les types de sources.

Les tables modifiables servent à stocker les informations épidémiologiques extraites de chaque rapport.

**La table Signal** : Contient toutes les informations relatives aux signaux (événements épidémiologiques) ; le type de signal (officiel ou non officiel), l'hôte, le sous-type, la localisation, la référence Empres-i et la date de confirmation du signal du signal (présente sur Empres-i si le signal est officiel)

<b>Id signal</b>	<b>Type</b>	<b>Hôte</b>	<b>Sous-type</b>	<b>Localisation</b>	<b>Date de notification officielle</b>
1	Officiel	Oiseaux domestiques	H5N6	Locale (ville)	12/04/2019

Tableau 2: Extrait de la table signale

**La table Chemin** ; Contient les informations relatives au cheminement de l'information (nombre et noms des sources, date de transmission de l'information, nombre d'hôtes touchés, nombre de cas, sous-type, localisation).

Ainsi si on reprend l'exemple précédemment cité les informations seraient stockées comme l'indique la figure 8 ou chaque ligne représente un lien entre deux sources.

Id Chemin	Source 1 (réceptrice)	Source 2 (émettrice)	Date de publication
1	PADI-Web	The Poultry Site	19/04/2018
1	The Poultry Site	OIE	19/04/2018
1	OIE	Ministère de l'agriculture	19/04/2018

Tableau 3 : Aperçu de la table des chemins

**La table Source :** Contient toutes les informations relatives aux sources : le nom de la source, le type, la localisation.

**La table Chemin-Signal :** contient chaque signal relié au chemin qui lui correspond. Un signal peut être retrouvé dans différents chemins lorsque celui-ci est cité par différentes sources.

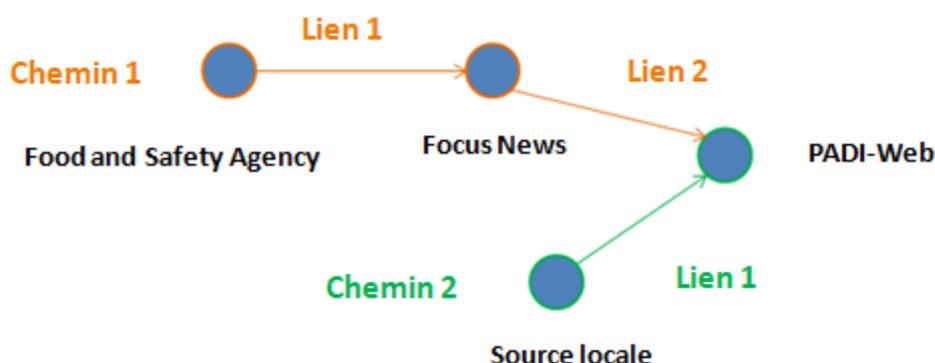
Id Chemin	Id Signal
1	1

Tableau 4 : Aperçu de la table Chemin-Signal

### 3. Construction du réseau

Un réseau dirigé a été construit à partir de la base de données. Selon la terminologie de la théorie des graphes, les nœuds représentent les sources. Les arêtes ou liens représentent le flux d'un signal entre deux sources (un émetteur, qui envoie le signal, et un récepteur, qui reçoit le signal). Toutes les arêtes consécutives représentaient un chemin.

**Focus News: Food and Safety Agency** reported a bird flu outbreak in the region of Haskovo, where more than 100,000 birds had to be culled. A **local source** confirmed the outbreak. (A)



(B)

Figure 7: (A) Extrait d'un rapport de Focus news et (B) sa représentation sous forme de graphe

## 4. L'analyse du réseau

Toutes les données ont été extraites manuellement et stockées dans une base de données Microsoft Access. Les analyses de réseau ont été faites en utilisant le package igraph disponible sur R version 3.6, la visualisation a été faite sur le logiciel Gephi 9.2.

Nous avons étudié les attributs qualitatifs des nœuds (types de sources, spécialisation, localisation) et quantitatifs (mesures de centralité et densité).

En se basant sur la bibliographie, les mesures et les concepts les mieux adaptées à l'analyse de réseaux de communication ont été utilisés :

- La centralité de degrés
- La centralité d'intermédiarité
- La centralité de vecteur propre
- La densité
- La détection de communautés

Leurs définitions et interprétations dans le contexte de cette étude sont présentées dans le tableau 5.

Ensuite, des analyses de réseau supplémentaires ont été faites sur réseau pondéré. La pondération permet de prendre compte le nombre de signaux transmis (un lien par lequel circulent 5 signaux a plus de poids qu'un lien qui transmet un seul signal).

Nous avons également exploré la réactivité et les aspects de complétude et d'exactitude du réseau.

La réactivité a été définie comme étant la différence en jours entre la détection d'un évènement épidémiologique (date de publication du rapport) par PADI-web et sa date de notification officielle (JO) qui est disponible sur la base de données Empres-i.

Une différence négative indique que PADI-web a permis une détection précoce du signal de foyer épidémique, et une différence que le signal a été détecté tardivement, une différence nulle indique que PADI-web a détecté le signal le jour même de sa notification officielle.

Pour évaluer les aspects de complétude (c'est-à-dire si l'information est complète) et d'exactitude (c'est-à-dire si l'information est vraie), les données collectées relatives à la localisation, sous-type, nombre de cas et date d'observation ont été comparées avec les données présentes sur Empres-i. (gold-standard).

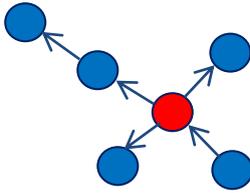
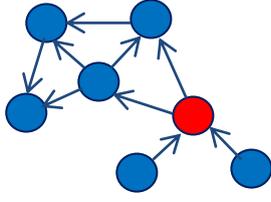
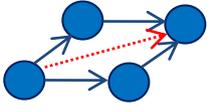
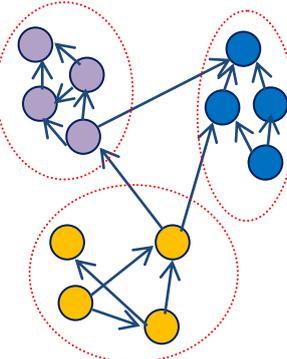
Concept d'analyse de réseau	Définition	Illustration	Interprétation
<b>Degrés entrants</b>	Nombre de liens ayant ce nœud pour extrémité		La source collecte l'information de plusieurs différentes sources
<b>Degrés sortants</b>	Nombre de liens ayant ce nœud pour origine		La source est souvent citée, communique l'information avec beaucoup de visibilité
<b>Degrés (entrants + sortants)</b>	Somme des degrés entrants et sortants		La source possède les deux capacités combinées
<b>Centralité d'intermédiarité</b>	Nombre de fois où le nœud est intermédiaire sur le plus court chemin entre deux autres nœuds		La source a un rôle important en tant qu'intermédiaire dans la transmission de l'information
<b>Centralité de vecteur propre</b>	La centralité du nœud est déterminée par la centralité des autres nœuds auxquels il est connecté		Cette source est connectée à d'autres sources fortement connectées dans le réseau
<b>Densité</b>	Ratio du nombre de liens existants sur le nombre de liens possible		Une densité élevée indique que de nombreuses sources sont connectées les unes aux autres
<b>Détection de communautés</b>	Les communautés sont des groupes qui ont des connexions denses à l'intérieur et peu de connexion avec le reste du graphique		Forte collaboration/communication entre les sources appartenant à la même communauté

Tableau 5: Les mesures utilisées et leur interprétation dans le contexte de cette étude. Adapté de (Giuliani & Pietrobelli, 2011).

## Résultats

### 1. Analyses descriptives

Entre le 02 Aout 2018 et le 29 Juin 2019 PADI-Web détecté 439 rapports finaux relatifs à l'influenza aviaire, parmi eux 340 rapports sont pertinents.

Le nombre total de signaux détectés était de **228** :

- 202 signaux **officiels** (confirmés par Empres-i) soit une proportion de 88,65%
- 26 signaux **non-officiels**, soit une proportion de 11,35%

64% (202/314) des signaux officiels présents dans Empres-i ont été détectés.

La sensibilité\* de PADI-web par rapport à l'influenza aviaire est de 39%, sur la figure 8 sont indiqués les signaux officiels détectés par PADI-web (en vert) et les signaux officiels présents dans la base de données Empres-i qui n'ont pas été détectés par PADI-web (en rouge).

Chaque rapport contenait minimum 1 signal, avec une moyenne de 2 signaux par rapport. Le nombre maximal de signaux (14) a été observé sur un rapport publié par un media en ligne spécialisé dans l'industrie avicole '*WattagNet*'.

78% (178/228) des signaux concernaient les oiseaux domestiques tandis que 22% (51/228) signaux concernaient les oiseaux sauvages.

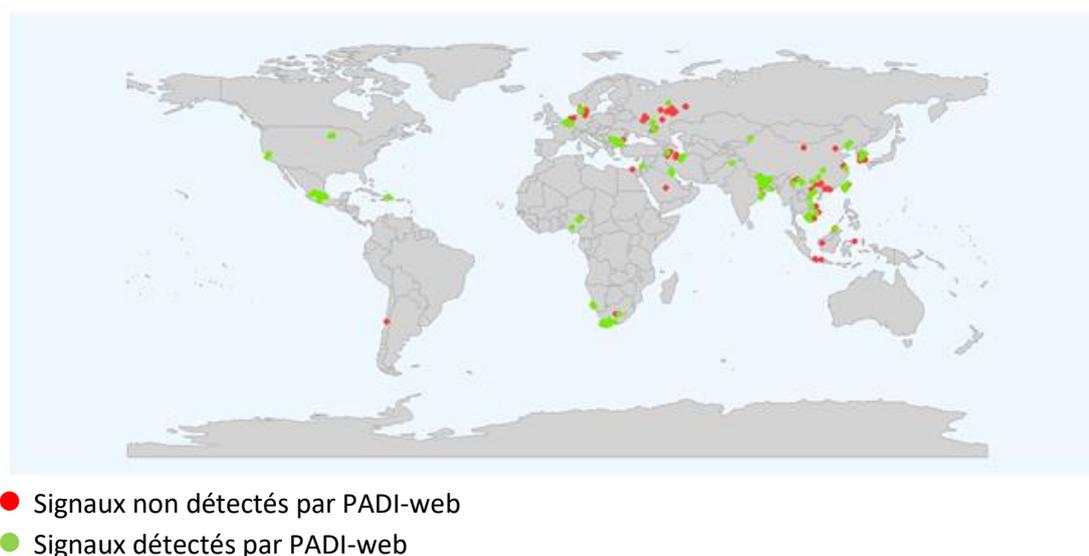


Figure 8: Distribution géographique des signaux d'influenza aviaire observés durant la période d'étude

### 2. Analyse de réseau

Ici les performances globales du réseau ont été analysées ; le réseau est composé de **254 sources** différentes, **445 liens** et **437 chemins**.

\* La sensibilité peut être définie comme la proportion d'événements sanitaires (confirmés) détecté par un outil de veille sanitaire = 
$$\frac{n \text{ événements détectés}}{(n \text{ événements détectés} + n \text{ événements non détectés})}$$

## 2.1 Caractéristiques des chemins

Sur 437 chemins, 90% des chemins (396/437) sont composés de 2 et 3 liens, 0.06% (27/437) des chemins sont composés d'un seul lien (ne citent aucune source), et 0.03% (14/437) des chemins sont constitués de 4 liens et plus (figure 9) 91% des chemins transmettaient l'information en moins d'une journée (figure 10).

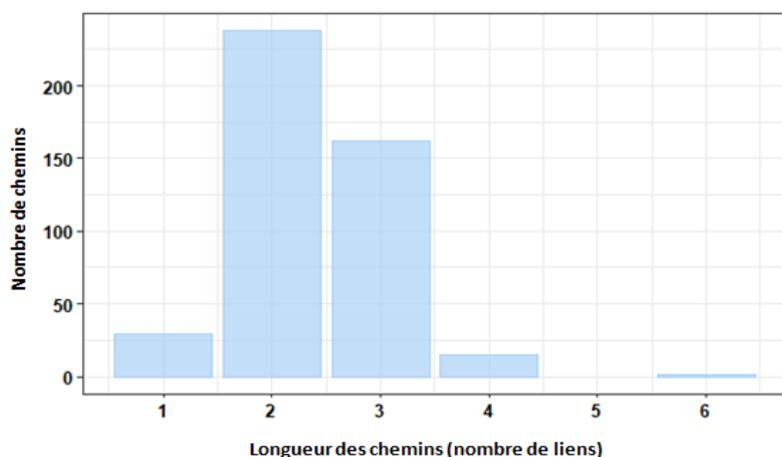


Figure 9: Longueur des chemins

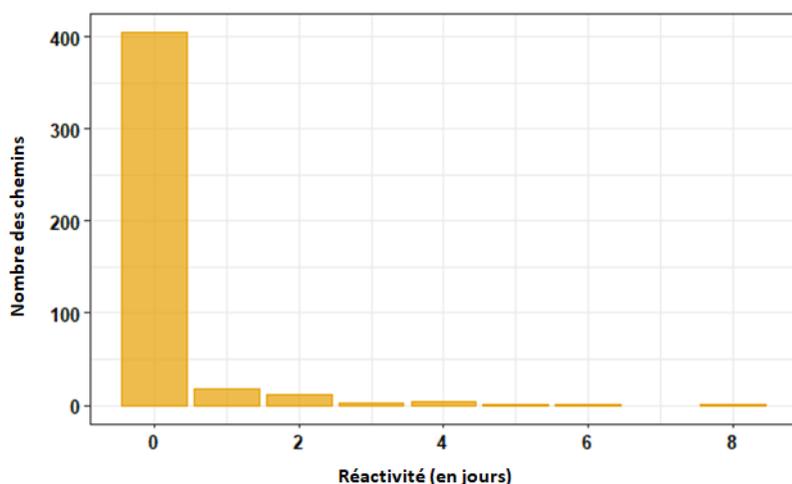


Figure 10: réactivité des chemins

## 2.2 Caractéristiques des sources

Les médias en ligne représentent 48% % (122/254) des sources impliquées dans le réseau suivis des autorités officielles 26%, des services sanitaires 9%, les agences de presse et les chaînes de radio/télévision représentent chacune 4% des sources impliquée dans le réseau. Les sources restantes qui sont : les organisations internationales et privées, les laboratoires, les réseaux sociaux et les sources locales ne représentent que 7% des du réseau.

Les autorités officielles (nationales et locales) et les services vétérinaires sont les sources primaires de 78% (354/437) des chemins, les services vétérinaires nationaux sont cités comme sources primaires dans tous les chemins ou ils sont impliqués. Les laboratoires, les compagnies privées et les personnes locales sont principalement citées comme sources primaires. Les organisations internationales (OIE et FAO), les médias en ligne et agences de presse sont principalement cités comme sources secondaires (figure 11).

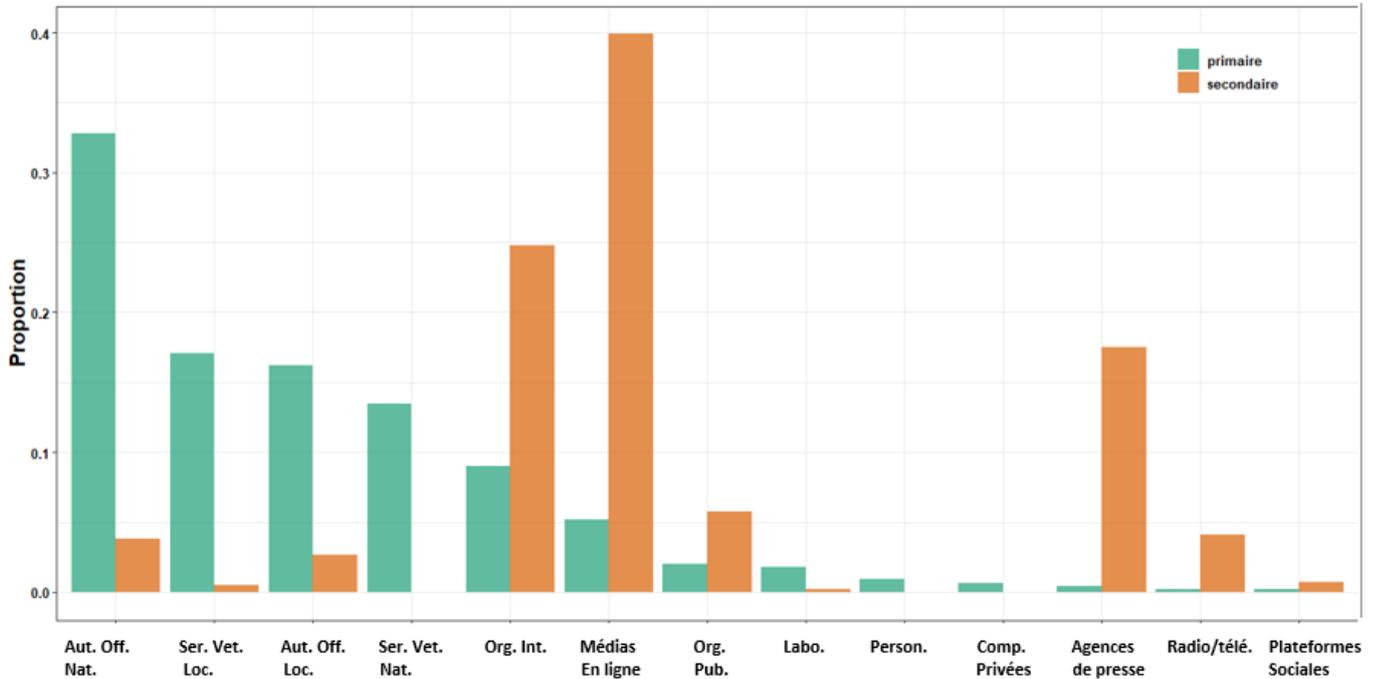


Figure 11: Types de sources primaires et secondaires impliquées dans le réseau.

## 2.3 Mesures de centralité

### a. Centralité de degré

La distribution de degrés (somme des liens entrants et sortants) montre un grand nombre de valeurs faibles car 25 % (64/254) des sources sont citées une seule fois (figure 12). Seulement 10 % (26/254) des sources du réseau sont caractérisées par un degré total supérieur à 5, l'OIE se démarque avec un degré total égal à 62. Le tableau 6 présente un classement des sources selon les valeurs de centralité de degré total (somme des degrés sortants et entrants). Sur le tableau 7, les sources sont classées séparément selon les valeurs de degrés sortants et degrés entrants.

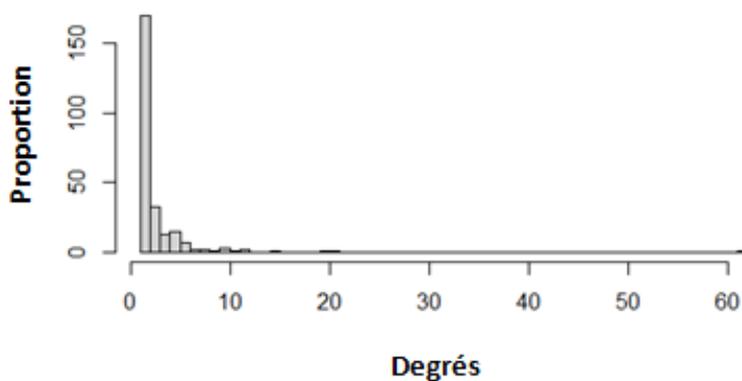


Figure 12: Distribution de degrés des sources

<b>Classement</b>	<b>Nom des sources</b>	<b>Degré total</b>
<b>1</b>	OIE	<b>62</b>
<b>2</b>	Reuters	<b>21</b>
<b>3</b>	Services vétérinaires en Inde	<b>20</b>
<b>4</b>	Bulgarian Food and Safety Agency	<b>15</b>
<b>5</b>	Xhinua/ services vétérinaires du Minnessota	<b>12</b>

Tableau 6: Classement des 5 premières sources ayant les valeurs de degrés la plus élevée

<b>Nom des sources</b>	<b>Degrés entrants</b>	<b>Noms des sources</b>	<b>Degrés sortants</b>
OIE	<b>38</b>	OIE	<b>24</b>
Xhinua	<b>10</b>	Services vétérinaires en Inde	<b>19</b>
PoultrySite	<b>9</b>	Reuters	<b>14</b>
WaTTagnet	<b>9</b>	Bulgarian Food and Safety Agency	<b>13</b>
CIDRAP/ Reuters	<b>7</b>	Services vétérinaires du Minnessota	<b>12</b>

Tableau 7 : Classement des 5 premières sources ayant les valeurs de degré entrants et sortants les plus élevées

### **b. Centralité d'intermédiarité**

La centralité d'intermédiarité renseigne sur l'importance d'une source en tant qu'intermédiaire dans la transmission de l'information, très peu de sources présentent des valeurs de centralités élevées. Sur le tableau 8, sont classés les 5 sources présentant les valeurs de centralité d'intermédiarité les plus élevées, l'OIE se distingue des autres sources avec un score de 0.035.

### **c. Centralité de vecteur propre**

La centralité de vecteur propre, prend en compte toutes les connexions présentes dans le réseau, selon cette mesure, une source connectée à une autre source fortement connectée vaut plus qu'une source connectée à une source faiblement connectée, les sources listées sur le tableau 9 sont celles qui ont les valeurs de centralité de vecteur propre les plus élevées.

Centralité intermédiaire	Nom	Type de source
0.035	OIE	Organisation internationale
0.006	Bulgarian Food and Safety Agency	Autorité officielle
0.006	Reuters	Agence de presse
0.005	Avian Flu Diary	Média en ligne
0.004	Ministère du climat et de l'environnement des Emirats Arabes Unis	Autorité officielle

Tableau 8: Les 5 premières sources présentant les valeurs de centralité d'intermédierité les plus élevées

Centralité de vecteur propre	Nom	Type de source
0.458	FAO	Organisation internationale
0.283	Poultry Site	Média en ligne
0.283	Gulf News	Média en ligne
0.283	Ministère du climat et de l'environnement des Emirats Arabes Unis	Autorité officielle
0.175	Arabianbusiness.com	Média en ligne

Tableau 9: Les 5 premières sources présentant les valeurs de centralité de vecteur propre les plus élevées

#### d. Densité du réseau

La densité du réseau est faible (0.006924777).

#### e. Réactivité du réseau

La réactivité varie de -112 jours et 96 jours, 22,7% des signaux ont été détectés précocement, 64% ont été détecté tardivement et 13,3% ont été détectés le jour de leur rapport officiel (figure 13).

La figure 14 illustre la différence de réactivité chez les deux catégories d'hôtes. 17,7% des signaux de foyer épidémiques d'influenza aviaire chez les oiseaux domestiques ont été détecté avant leur notification officielle. 68,5% des signaux ont été détectés tardivement, et 13,8% des signaux ont été détecté le même jour que leur notification officielle.

Chez les oiseaux sauvages 41% des signaux ont été détecté précocement (avec un record de 112 jours) 11% des signaux ont été détectés à J0 c'est-à-dire le jour de leur notification officielle et 47% des signaux ont été détectés tardivement.

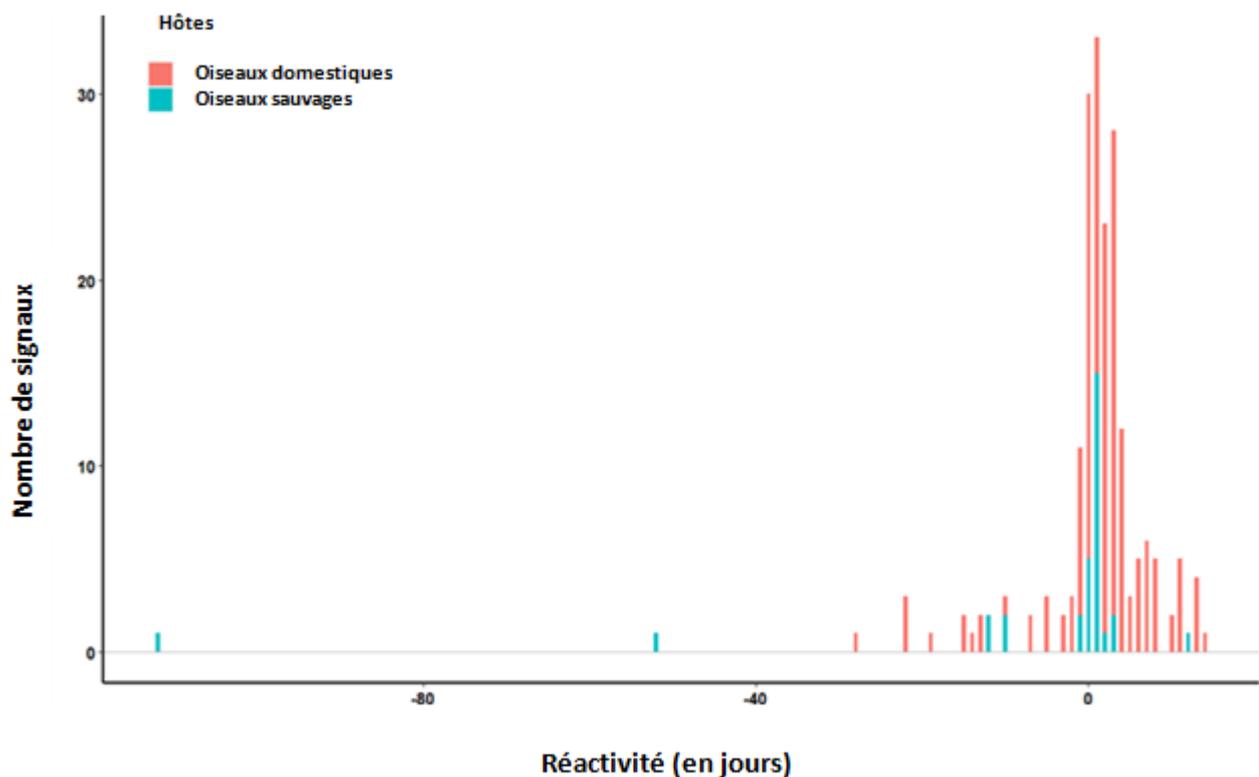


Figure 13: Réactivité de PADI-Web.

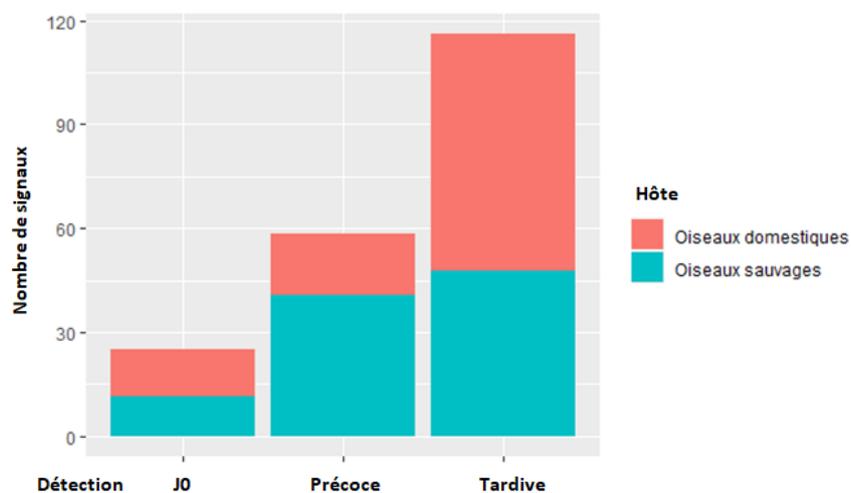


Figure 14: Réactivité observée chez les deux catégories d'hôtes : oiseaux domestiques et oiseaux sauvages

### 3. Visualisation du réseau

Sur les figures 15 et 16 la taille des nœuds est proportionnelle à la valeur de degrés (somme des liens entrants et sortants). La figure 15 montre que l'OIE occupe une position centrale dans le réseau d'information, les sources qui lui sont adjacentes sont les services vétérinaires et les autorités officielles, les sources directement connectées à PADI-Web sont les médias en ligne. Les agences de presse sont dispersées et semblent communiquer avec les sources de différents types.

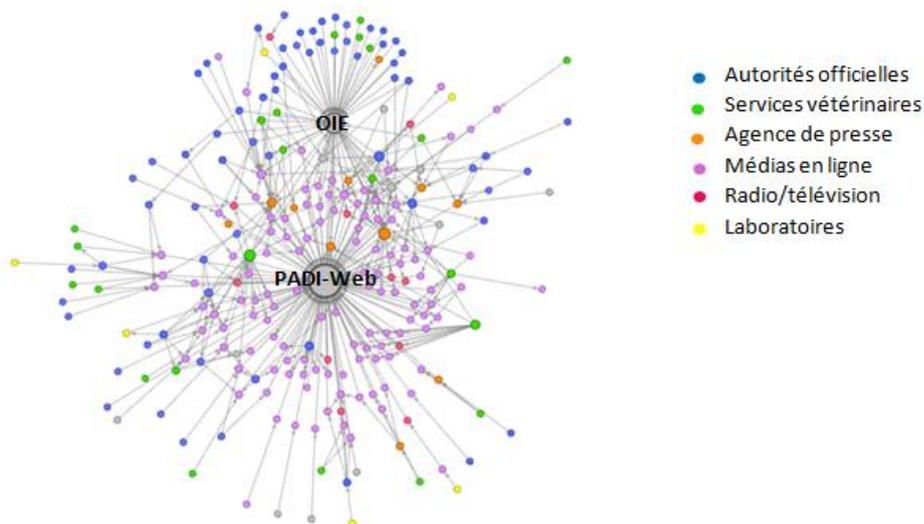


Figure 15 : Vue globale du réseau d'information sanitaire

### 3.1 Détection des communautés

Quatre communautés ont été obtenues en utilisant l'algorithme Louvain (Blondel et al., 2008). Les sources ayant la plus grande valeur de degrés dans chaque communauté sont mises en avant (figure 16).

La première (en termes de nombre de source) contient PADI-web et différents médias en ligne, dans la deuxième communauté sont retrouvées les autorités officielles et l'OIE, dans la troisième, la Bulgarian Food and Safety Agency ainsi que divers médias et agences de presse, et la quatrième renferme les médias en ligne et services sanitaire principalement localisé en Asie.

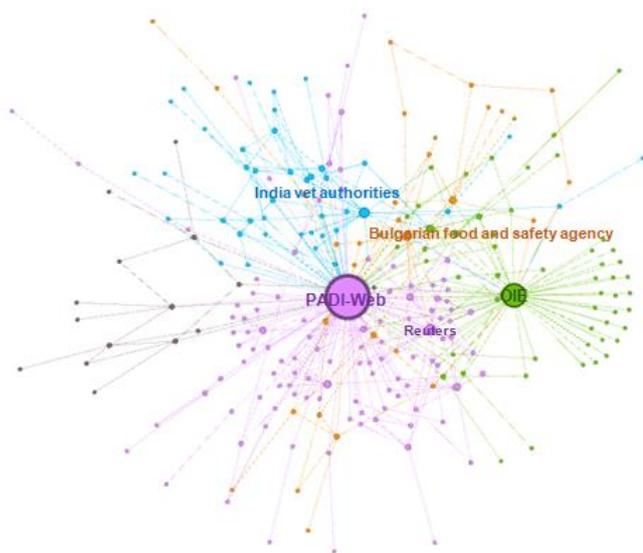


Figure 16 : Communautés au sein du réseau sanitaire.

## 4. Aspects de complétude et d'exactitude

Très peu d'informations incorrectes circulent dans le réseau. Lorsqu'elles existent, elles concernent les dates d'observation des foyers épidémiques, quant aux informations relatives à la localisation, au sous-type et au nombre de cas, elles peuvent être absentes ou manquer de précisions. Il arrive par exemple que le sous-type ne

soit cité que partiellement (H5, hautement pathogénique etc.), ou que la localisation soit donnée à l'échelle nationale seulement, ou que le nombre de cas soit approximatif.

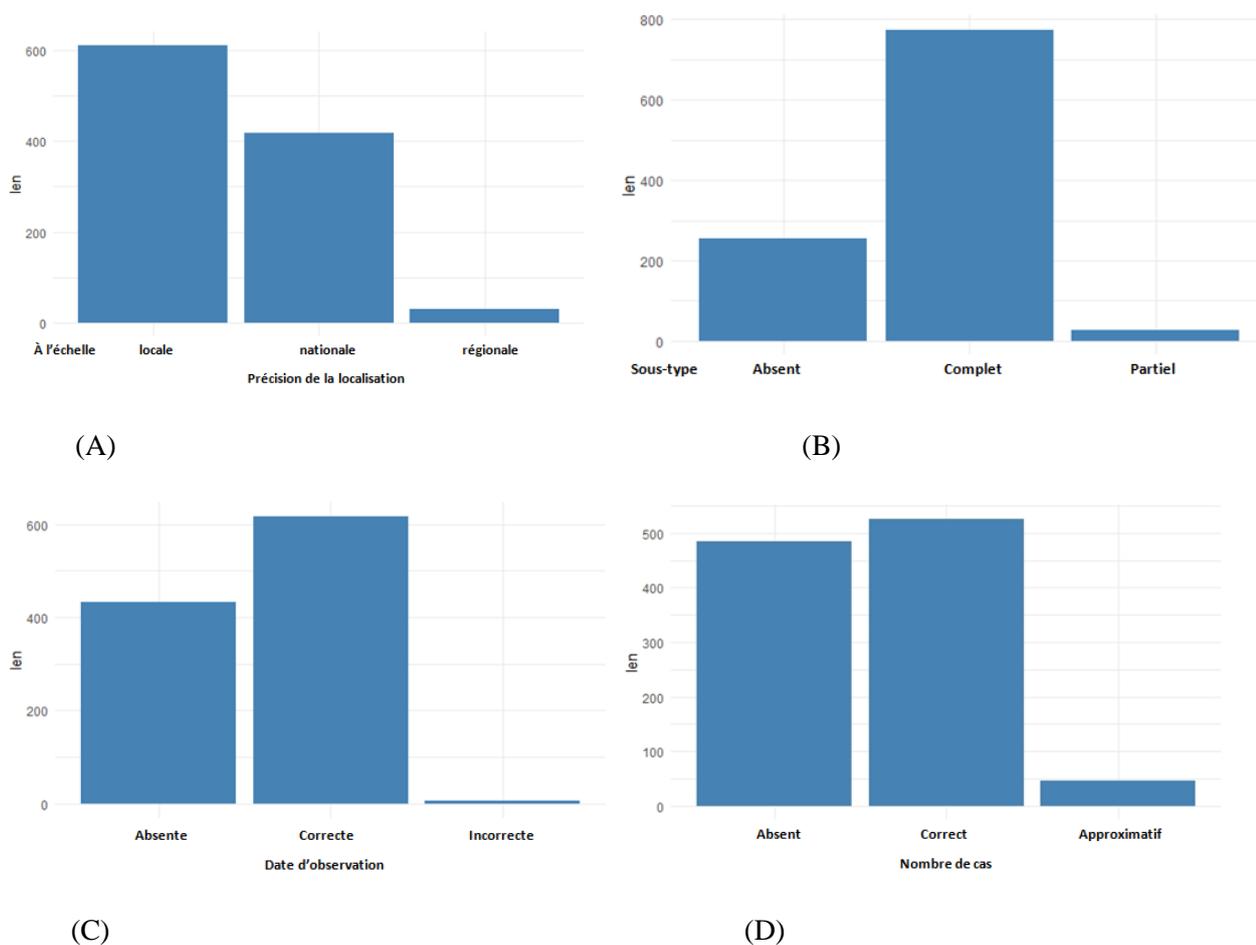


Figure 17 : Proportion des informations de localisation (A), sous-type (B), dates d'observations du foyer (C) et nombre de cas (D).

## 5. Résultats sur réseau pondéré

Dans chaque lien entre deux sources circule au minimum un signal, le nombre maximal de signaux circulant par le même lien est de 76 signaux. La figure 18 montre que la majorité des liens diffusent un seul signal

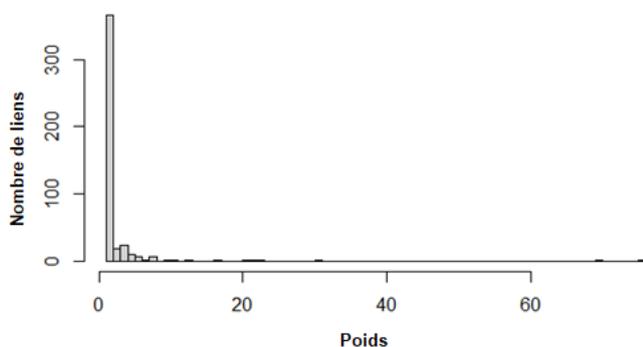


Figure 18: distribution du poids des liens au sein du réseau d'information sanitaire

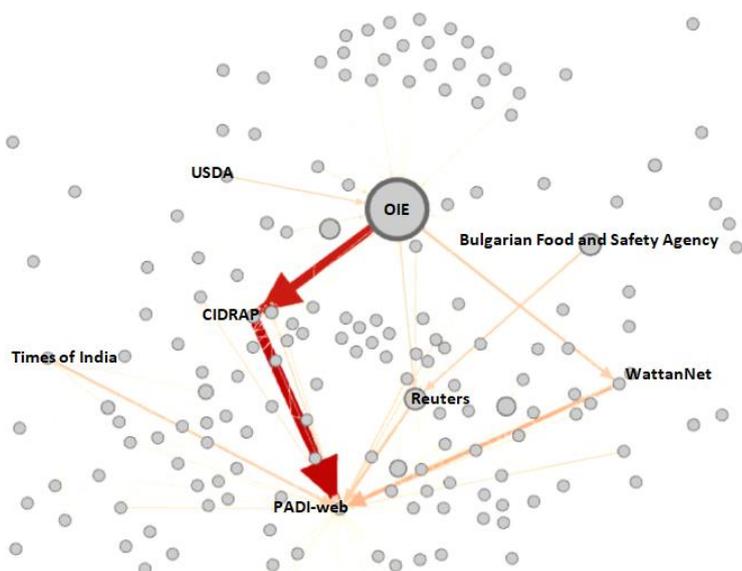


Figure 19: Zoom sur les liens qui diffusent le plus grand nombre d'informations.

La figure 19 représente le réseau d'information sanitaire de PADI-web. Ici la largeur des liens est proportionnelle à l'intensité du flux d'information, et la taille des nœuds proportionnelle aux valeurs de centralité intermédiaire.

, Le CIDRAP est le "Center for Infectious Disease Research and Policy" de l'université du Minnesota", Reuters est une agence de presse, WattagNet et Times of India sont des médias en ligne et l'USDA est le Département de l'agriculture des Etats-Unis

Le tableau 10 met en évidence les 5 liens du réseau PADI-web où circule le plus grand nombre d'informations.

Sur la colonne de gauche sont listées les sources émettrices et sur la colonne de droite sont listées les sources réceptrices.

Emettrice	Réceptrice	Poids du lien (nombre de signaux)
<b>CIDRAP</b>	<b>PADI-web</b>	76
<b>OIE</b>	<b>CIDRAP</b>	70
WattagNet	<b>PADI-web</b>	31
Reuters	<b>PADI-web</b>	23
<b>OIE</b>	WattagNet	22

Tableau 10: Classement des 5 premiers liens où diffuse le plus grand nombre de signaux dans le réseau.

## Discussion

Les systèmes de veille ont été décrits et évalués dans de nombreuses études, mais l'on sait peu de chose sur la manière dont l'information sanitaire circule avant qu'elle ne soit détectée. Récemment un travail de recherche de ce type, sur la peste porcine africaine (PPA) a été mené dans le cadre d'une thèse (Valentin 2020). Cependant à notre connaissance aucune autre étude n'a été réalisée pour comprendre comment se fait la diffusion des informations avant d'arriver aux systèmes de veille sanitaire.

Le but de ce travail était d'étudier comment l'information sanitaire circule sur le web avant d'être détectée par PADI-web. Pour cela, le travail a été mené sur un cas d'étude : la diffusion de l'information relative à l'influenza aviaire (IA). Le réseau d'informations sanitaires sur lequel repose ce système de veille a été étudié à travers des méthodes qualitatives (topologie des sources) et quantitatives (mesures de centralités de densité etc.) d'analyse de réseau. La réactivité de PADI-web et les aspects de complétude et d'exactitude des informations recueillies ont également été explorés.

Durant la période allant du 02/08/2018 au 29/06/2019 PADI-web a détecté 64% des signaux relatifs à l'influenza aviaire présents dans la base de données Empres-i, contre 51% lors d'une étude menée en 2016 (Arsevaska et al., 2018). Cette amélioration peut s'expliquer par l'intégration de nouveaux algorithmes qui permettent à PADI-web de traduire plus de langues étrangères en anglais, et par l'enrichissement du corpus d'annotation de son classifieur (Valentin et al., 2019).

- **PADI-Web permet de détecter les foyers d'IA précocement**

Cette étude a permis de montrer l'intérêt de l'utilisation de PADI-web pour la détection rapide d'informations concernant des foyers d'IA. En effet, pour ce cas d'étude, PADI-web détecté précocement 20.7% des rapports finaux validés par les gold-standard. L'information sanitaire dans le réseau d'information PADI-web se transmet rapidement d'une source primaire à une source secondaire et PADI-web détecte les rapports finaux automatiquement moins d'un jour après leur publication. La transmission rapide est justifiée par le fait que 90% des chemins soient courts et composés de 2 et 3 liens. Ces résultats ont été obtenus en analysant les informations relatives à l'IA. Des résultats similaires ont été observés lors de l'étude portant sur la PPA, en termes de caractéristiques générales du réseau (nombre de source, de liens, et de chemins) et de types de sources majoritairement impliquées (les médias en ligne et agences de presse sont des sources secondaires importantes dans les deux réseaux) (Valentin, 2020).

La vitesse de détection est un enjeu majeur dans la surveillance en santé animale. En effet, la détection précoce des foyers d'IA permet la mise en place rapide de mesures de contrôle afin de limiter les dégâts causés par la maladie. La détection précoce de foyers d'IA chez les oiseaux domestiques est primordiale d'une part à cause des pertes économiques qu'elle engendre (restrictions commerciales, perte du statut indemne des pays touchés) mais aussi pour le risque sanitaire qu'elle peut représenter pour l'homme.

Seulement, les oiseaux sauvages aussi peuvent participer à la dissémination de la maladie: le contact entre oiseaux sauvages et domestiques peut entraîner une contamination croisée, les oiseaux sauvages en particulier les oiseaux aquatiques (cygnes, canards etc.) peuvent contribuer à la propagation de la maladie par leur mouvements migratoires (Hill et al., 2019) . En 2019, par exemple, durant la période migratoire, plusieurs foyers d'IA au sein des fermes avicoles périphériques au lac Chilka en Inde ont été observés, et de nombreux foyers d'IA (aussi bien chez les oiseaux sauvages que domestiques) ont été détectés par PADI-web

En termes de détection précoce, une différence a été notée entre les deux catégories d'hôtes étudiées. Chez les oiseaux sauvages, les délais entre la détection par PADI-web et sa notification officielle peuvent être supérieurs à 3 mois tandis que chez les oiseaux domestiques cela n'excède pas les 33 jours. L'une des explications possibles serait que le processus de confirmation officielle d'IA chez les oiseaux sauvages prend plus de temps, notamment à cause des méthodes de capture et d'échantillonnage qui sont plus contraignantes chez les oiseaux sauvages que chez les oiseaux domestiques.

Il aurait été intéressant d'étudier les deux réseaux (oiseaux sauvage/ oiseaux domestiques) séparément pour voir les différents types de sources impliquées pour chacun. L'étude sur deux réseaux distincts n'a pas été faite car le corpus concernant les oiseaux sauvages comprenait seulement 26 signaux.

- **PADI-web permet d'obtenir des données sanitaires fiables**

Les sources primaires de PADI-web étaient majoritairement les autorités officielles et les services sanitaires. Les services vétérinaires représentent les premières sources émettrices de l'information dans tous les chemins où ils sont impliqués, Les organismes internationaux tels que l'OIE et la FAO sont principalement citées comme sources secondaires car elles vont recevoir l'information des autorités officielles et des services vétérinaires. Cela montre que les données circulantes dans le réseau du PADI-web proviennent majoritairement des services vétérinaires et des autorités officielles, les organismes internationaux ont un rôle majeur de centralisation des données en particulier l'OIE dont la valeur de centralité d'intermédiarité est la plus élevée.

Les médias en ligne suivi des agences de presse jouent majoritairement le rôle de sources secondaires dans le réseau d'information de PADI-web. Elles aussi reçoivent l'information des sources primaires telles que les services sanitaires et autorités locales ou nationales et contournent ainsi le processus de notification officielle à l'OIE, ce qui contribue vraisemblablement à la détection précoce de foyers épidémiques par PADI-Web. Ce schéma a également été observé lors d'une étude d'analyse similaire de PADI-web avec la PPA comme maladie modèle (Valentin, 2020).

La structure du réseau d'information sanitaire est semblable au modèle « Source - Réceptrice-émettrice – Réceptrice » décrit par une étude précédente (Weber & Monge, 2011) : Les sources secondaires (réceptrice-émettrice) peuvent apporter un contenu supplémentaire à l'information fournie par les sources primaires qui peut parfois être incorrect ou manquer de précision, avant de relayer l'information. Néanmoins très peu d'informations incorrectes ont circulé dans le réseau de transmission d'information de PADI-web. Cela peut s'expliquer par le fait que les acteurs clés soient des sources fiables (Yoon et al. 2020). L'analyse des rapports captés par PADI-web montre que les sources les plus fiables quant à la qualité de l'information sont les services vétérinaires, les laboratoires et les organismes privés. Ces deux derniers jouent aussi le rôle de sources primaires prioritaires bien qu'ils soient peu nombreux dans ce réseau.

Les valeurs de centralités obtenues dans cette étude indiquent que quelques sources jouent le rôle d'acteurs clés en diffusant l'information avec beaucoup de visibilité comme les agences de presse (Xinhua et Reuters), ou en agrégeant les informations de différentes sources comme les sites internet spécialisés. Ces résultats peuvent servir à définir des sources prioritaires à l'avenir (rapports à détecter en priorité par PADI-web) afin d'améliorer les performances de détection précoce de PADI-web. De plus, les résultats de cette étude ont montré que de nombreuses informations pertinentes ont été rapportées par des médias en ligne spécialisés en industrie avicole ou en élevage, par exemple "Wattagnet", "The PoultrySite" et "Avian Flu Diary". Cela a également été constaté dans l'étude du réseau d'information sanitaire PADI-web propre à la PPA avec des médias comme "PigSite" et "PigWorld" (Valentin, 2020)

Les résultats obtenus font également ressortir la ‘‘Bulgarian Food and Safety Agency’’ comme source clé dans ce réseau d’information, par sa position centrale dans le réseau (centralité d’intermédiation et de vecteur propre) et également par le nombre d’informations transmises. Ces résultats ont probablement été influencés par le choix de la maladie étudiée et aussi par la période d’étude. En effet pendant la période s’étalant d’Aout 2018 à Juin 2019, 22 foyers d’influenza aviaire ont été déclarés en Bulgarie (Empres-i 2019).

- **PADI-web permet de détecter des informations non-officielles**

La capacité des systèmes de veille sanitaire à détecter des événements épidémiologiques précocement a été évaluée et prouvée dans de nombreuses études (Barboza, 2015),(Carrion & Madoff, 2017). D’ordinaire, les systèmes de veille basée sur les indicateurs tels que l’OIE, la FAO, l’OMS servent de gold-standard aux études des systèmes de veille sanitaire, mais ces gold-standards sont tributaires de la volonté des pays concernés de communiquer ou non l’information (Barboza, 2015). Les données relatives aux signaux non-officiels détectés par PADI-web n’ont pas été exploitées. Une étude complémentaire serait nécessaire pour savoir si c’est signaux non-officiels sont de fausses alertes de foyers épidémiques ou s’il s’agit bel et bien d’évènements qui n’ont pas été notifiés aux organismes internationaux.

- **Perspectives d’étude pour approfondir l’évaluation de PADI-web**

A présent que cette étude a été menée pour une maladie animale émergente non zoonotique (PPA) et zoonotique (IA), il serait intéressant de mener à nouveau cette étude à travers le prisme d’une maladie émergente et nouvelle, par exemple en analysant rétrospectivement les rapports relatifs à la Covid-19 captés par PADI-web et suivre l’évolution du réseau sanitaire depuis le début de la pandémie (en termes de sources impliquées dans la diffusion de l’information, de termes utilisés pour décrire la maladie, de qualité de l’information etc.). Comprendre le fonctionnement de ce type de réseau d’information, pourrait apporter des éléments qui aideraient à améliorer la capacité de PADI-web à détecter les signaux d’une maladie nouvelle.

Les résultats obtenus sont représentatifs du réseau d’information sanitaire de PADI-web s’appuyant sur le cas d’étude de l’IA durant une période où de nombreux foyers ont été rapportés en Asie et en Europe. Il serait intéressant d’étudier les informations sanitaires collectées pour cette même maladie et cette même période via un autre système de veille et de faire une étude comparative des réseaux d’informations sanitaires, en termes de sources impliquées, de réactivité, et de qualité de l’information. Par exemple, en étudiant la base de données du système de veille : ProMED qui bénéficie d’une modération humaine assuré par des experts dans différents domaines (virologie, écologie etc.) (Yu & Madoff, 2004) ou la base de données de HealthMap. HealthMap serait un choix d’autant plus intéressant car il aspire ses rapports finaux des deux moteurs de recherche GoogleNews et Baidu (Freifeld et al., 2008), contrairement à PADI-web qui a recours à GoogleNews uniquement.. Aspirer les rapports de deux moteurs de recherche différent va certes enrichir la base de données du système de veille, mais on peut se poser la question quant à les redondances des informations collectés et à leur pertinence/fiabilité.

## Conclusion

La veille sanitaire est la stratégie adoptée par de nombreux pays pour faire face aux nouvelles menaces de maladies émergentes et zoonotiques.

PADI-web est un système de veille automatique du web axé en santé animale, qui collecte quotidiennement l'information sanitaire à travers les sources officielles (OMS, FAO) mais également à travers des sources non officielles (médias en ligne, agences de presse, plateformes sociales etc.). Ce travail a permis de montrer la plus-value de l'utilisation de PADI-web pour la détection précoce des foyers influenza aviaire, enjeu cruciale pour la gestion nationale et locale de cette maladie à répartition mondiale. Par ailleurs, cette étude a permis de démontrer la fiabilité des informations circulantes dans ce réseau. Elle a également permis de décrire avec précision le réseau de transmission de l'information sanitaire au sein de PADI-web et ainsi mieux comprendre le rôle des sources impliquées. Enfin, cette étude a également montré l'intérêt de l'utilisation de PADI-web dans l'obtention des données non-officielles, bien que dans ce cas, l'expertise d'épidémiologistes pour valider la pertinence des informations soit indispensable.

Les résultats obtenus pourraient également aider à définir des sources prioritaires afin d'améliorer les performances de détection précoce des systèmes de veille sanitaire. La méthodologie utilisée dans cette étude peut servir de modèle pour divers travaux d'analyses de réseaux de communication, avec d'autres systèmes de veille sanitaire et différents cas d'étude.

Dans l'ensemble, le travail mené au cours de ce stage a prouvé l'utilité de la veille sanitaire des articles publiés sur le web et a également montré qu'une approche multidisciplinaire informatique et épidémiologique permet d'améliorer les performances des systèmes de veille et d'enrichir leurs bases de données.

## Annexes

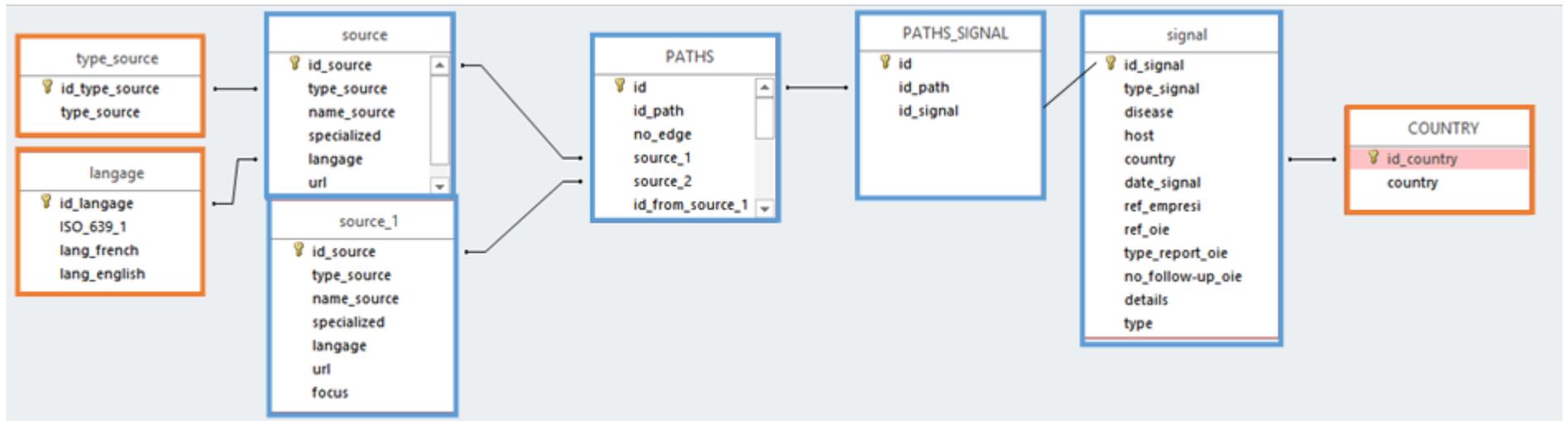
### Annexes A Comparaison de trois systèmes de veille sanitaire

	<b>ProMed</b>	<b>HealthMap</b>	<b>PADI-Web</b>
<b>Sources</b>	Formelles et informelles : Articles de presse, rapports officiels, observateurs locaux	Formelles et informelles : PromedMail GoogleNews Baidu	Formelles et informelles : GoogleNews
<b>Langues</b>	(5) : Anglais, Français, Espagnol, Russe, Portugais	(7) : Anglais, Français, Espagnol, Russe, Portugais, Arabe, Chinois	Multilingues
<b>Modération</b>	Modération humaine	Partiellement automatique	Automatique
<b>Maladies ciblées</b>	Humaines, Animales, Plantes, environnement	Humaines, Animales	Animales
<b>Couverture géographique</b>	Mondiale	Mondiale	Mondiale
<b>Référence</b>	(Carrion & Madoff, 2017) (Yu & Madoff, 2004)	(Freifeld et al., 2008)	(Arsevska et al., 2018; Valentin et al., 2019)

**Annexe B** Indicateurs d'évaluation des systèmes de veille sanitaire (Barboza, 2015), (WHO, 2014).

<b>Indicateurs</b>	<b>Définition</b>
Réactivité	Intervalle de temps entre la détection d'un événement par un outil de veille sanitaire et la notification de ce même événement par les autorités sanitaire en charge de prendre des mesures de contrôle
Sensibilité	<p>La sensibilité peut être définie comme la proportion d'événements sanitaires (confirmés) détecté par un outil de veille sanitaire</p> $\frac{n \text{ événements détectés}}{(n \text{ événements détectés} + n \text{ événements non détectés})}$
VPP (Valeur Prédictive positive)	<p>La probabilité qu'un signal détecté par un système corresponde à un évènement réel (proportion d'évènements parmi les signaux).</p> $\frac{n \text{ signaux correspondant aux évènements réels}}{(n \text{ signaux correspondant aux évènements réels}) + (n \text{ autres signaux})}$
Flexibilité	La capacité de s'adapter aux besoins des utilisateurs, ainsi que la mesure des capacités des outils de veille sanitaire à détecter de nouveaux risques sanitaires (maladies nouvelles)
Représentativité	<p>La capacité d'un système à décrire avec précision la survenue d'un événement sanitaire (temps, lieu, personne) elle peut être définie en termes de :</p> <p>a) couverture géographique (mondiale ou régionale)</p> <p>b) sujet : maladies infectieuses animales, zoonoses</p>

Annexe C vue global de la base de données Microsoft Access.



## Bibliographie

- Arsevka, E. (2017). Élaboration d'une méthode semi-automatique pour l'identification et le traitement des signaux d'émergence pour la veille internationale sur les maladies animales infectieuses. 251.
- Arsevka, E., Valentin, S., Rabatel, J., de Goër de Hervé, J., Falala, S., Lancelot, R., & Roche, M. (2018). Web monitoring of emerging animal infectious diseases integrated in the French Animal Health Epidemic Intelligence System. *PLoS ONE*, 13(8). <https://doi.org/10.1371/journal.pone.0199960>
- Barboza, P. (2015). Evaluation des systèmes d'intelligence épidémiologique appliqués à la détection précoce des maladies infectieuses au niveau mondial. 122.
- Beauguitte, L. (2016). L'analyse de réseaux en sciences sociales et en histoire. 18.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bonilla-Aldana, D. K., Holguin-Rivera, Y., Cortes-Bonilla, I., Cardona-Trujillo, M. C., García-Barco, A., Bedoya-Arias, H. A., Rabaan, A. A., Sah, R., & Rodriguez-Morales, A. J. (2020). Coronavirus infections reported by ProMED, February 2000–January 2020. *Travel Medicine and Infectious Disease*, 35, 101575. <https://doi.org/10.1016/j.tmaid.2020.101575>
- Cáceres, P., Awada, L., Barboza, P., Lopez-Gatell, H., & Tizzani, P. (2017). The World Organisation for Animal Health and the World Health Organization : Intergovernmental disease information and reporting systems and their role in early warning. *OIE Revue Scientifique et Technique*, 36, 539-548. <https://doi.org/10.20506/rst.36.2.2672>
- Carrion, M., & Madoff, L. C. (2017). ProMED-mail : 22 years of digital surveillance of emerging infectious diseases. *International Health*, 9(3), 177-183. <https://doi.org/10.1093/inthealth/ihx014>
- Chikhi, N. F. (2010). Calcul de centralité et identification de structures de communautés dans les graphes de documents. 182.

- Eilstein, D., Salines, G., & Desenclos, J.-C. (2012). Veille sanitaire : Outils, fonctions, processus. *Revue d'Épidémiologie et de Santé Publique*, 60(5), 401-411. <https://doi.org/10.1016/j.respe.2012.03.005>
- Freifeld, C. C., Mandl, K. D., Reis, B. Y., & Brownstein, J. S. (2008). HealthMap : Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports. *Journal of the American Medical Informatics Association : JAMIA*, 15(2), 150-157. <https://doi.org/10.1197/jamia.M2544>
- Gay, B., Dousset, B., & Wanassi, R. (2013). Les indicateurs structurels d'un graphe : Calculs, visualisation, interactivité. 5ème séminaire de Veille Stratégique Scientifique & Technologique (VSST 2013), 0. <https://hal.archives-ouvertes.fr/hal-02639447>
- Giuliani, E., & Pietrobelli, C. (2011). *Social Network Analysis Methodologies for the Evaluation of Cluster Development Programs*. 54.
- Hill, A., Gilling, S., Brouwer, A., & Breed, A. (2019). Quantifying the spatial risk of Avian Influenza introduction into British poultry by wild birds | *Scientific Reports*. <https://www.nature.com/articles/s41598-019-56165-9>
- Ilef, D., Derode, A., Morel, É., Rhône-Alpes, A., Perin, H., Champagne-Ardenne, A., Sarradet, A., Aoussou, F., Magne, P., Buisson, C., Caserio-Schönemann, C., Fournet, N., Gastellu-Etchegorry, M., & Lefranc, A. (2011). *La veille et l'alerte sanitaires en France*. 63.
- Kirchhoff, L., Stanoevska-Slabeva, K., Nicolai, T., & Fleck, M. (2008). Using social network analysis to enhance information retrieval systems. 21.
- Kloth, Y. M., Deutsch, K. M., Danielson, K. A., Strack, J., & Law, C. (2019). What Twitter teaches us about patient-provider communication on pain. *PLOS ONE*, 14(12), e0226321. <https://doi.org/10.1371/journal.pone.0226321>
- Kman, N. E., & Bachmann, D. J. (2012). *Biosurveillance : A Review and Update*. *Advances in Preventive Medicine*, 2012. <https://doi.org/10.1155/2012/301408>

- Paquet, C., Coulombier, D., Kaiser, R., & Ciotti, M. (2006). Epidemic intelligence : A new framework for strengthening disease surveillance in Europe. *Eurosurveillance*, 11(12), 5-6.  
<https://doi.org/10.2807/esm.11.12.00665-en>
- Sabot, K., Wickremasinghe, D., Blanchet, K., Avan, B., & Schellenberg, J. (2017). Use of social network analysis methods to study professional advice and performance among healthcare providers : A systematic review. *Systematic Reviews*, 6(1), 208. <https://doi.org/10.1186/s13643-017-0597-1>
- Stattner, E., & Vidot, N. (2011). Social network analysis in epidemiology : Current trends and perspectives. 1-11. <https://doi.org/10.1109/RCIS.2011.6006866>
- Valentin, S. (2020). Extraction et combinaison d'information à partir de sources informelles pour la veilles des maladies infectieuses.
- Valentin, S., Rabatel, J., Arsevska, E., Falala, S., De Goër, J., Mercier, A., Lancelot, R., & Roche, M. (2019). PADI-web : Un système automatique multilingue pour la veille sanitaire internationale en santé animale. 30èmes Journées Francophones d'Ingénierie des Connaissances (IC), 235-239.  
<https://hal.archives-ouvertes.fr/hal-02556918>
- Weber, M. S., & Monge, P. (2011). The Flow of Digital News in a Network of Sources, Authorities, and Hubs. *Journal of Communication*, 61(6), 1062-1081. <https://doi.org/10.1111/j.1460-2466.2011.01596.x>
- WHO. (2014). Early detection assessment and respons to accute public health event : Implementation of Early Warning and Response with a focus on Event-Based Surveillance.
- Yoon, S., Odlum, M., Broadwell, P., Davis, N., Cho, H., Deng, N., Patrao, M., Schauer, D., Bales, M. E., & Alcantara, C. (2020). Application of Social Network Analysis of COVID-19 related Tweets mentioning Cannabis and Opioids to Gain Insights for Drug Abuse Research. 6.
- Yu, V. L., & Madoff, L. C. (2004). ProMED-mail : An Early Warning System for Emerging Diseases. *Clinical Infectious Diseases*, 39(2), 227-232. <https://doi.org/10.1086/422003>



