

Analyse des communicants pivots à partir des données massives de communication

**Ou comment dire des
choses intelligentes
avec des données pas
trop structurées**

Maxime Bergeat, Anca Boboc

Journées de méthodologie statistique de l'Insee – 31 mars 2022



université
PARIS-SACLAY



Les communicants pivots ?

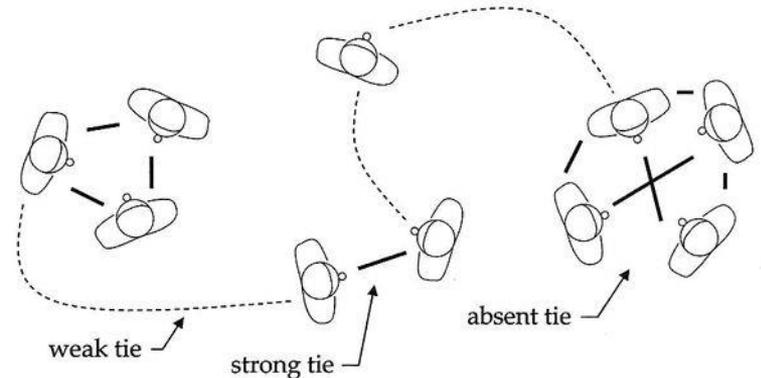
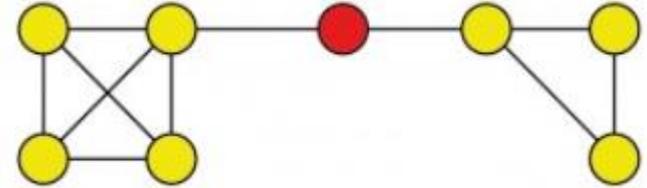
Données et méthodes

Typologie des profils de communication

Les communicants pivots ?

Deux théories canoniques sur l'analyse des réseaux sociaux

- **La force des trous structuraux (Burt, 1992)**
 - Une position de « trou structural » dans le réseau confère un *capital social* qui fournit un **avantage compétitif** au *broker*
→ accès plus rapide et enrichi à l'information, contrôle de la transmission de l'information
- **La force des liens faibles (Granovetter, 1977)**
 - Les liens faibles, un fort potentiel d'innovation
 - Une théorie très discutée par la suite, notamment dans le cadre de l'étude des processus de recrutement (Gee et al., 2017)



Objectifs

1. Détecter les communicants pivots

- À partir de traces sur les communications des salariés

2. Caractériser les communicants pivots

- Selon le métier exercé
- Selon les caractéristiques des salariés

⚠ Un communicant pivot

- Joue un rôle important dans la circulation de l'information (quantité et « qualité » des contacts noués dans l'entreprise)
- N'est pas nécessairement un influenceur (mais peut transformer l'info qu'il transmet, ou en produire en tant qu'expert de son domaine)

Objectif bonus

Donner une proposition de recette pour dire des choses intelligentes avec des données pas trop structurées

Données et méthodes

Méthodologie : comment caractériser les profils de communication ?

Cadrage

- Période d'analyse : septembre-novembre 2019
- Champ de salariés : Direction de l'Innovation d'une grande entreprise française du secteur des télécommunications (4500 salariés en France)

En quelques étapes

1. Définition d'indicateurs sur la communication par salarié
2. Analyse multivariée pour étudier les liens entre indicateurs
3. Regroupement des salariés ayant des comportements proches pour la communication
4. Étude de deux profils de communicants pivots



Comment dire des choses intelligentes avec des données pas trop structurées ?

1) Réduire la dimensionnalité du problème si c'est possible

→ 1 million de messages échangés par jour, c'est beaucoup (trop ?)

2) Ne pas avoir peur du vocabulaire des geeks



Données incluses dans l'analyse

- **Trois canaux de communication**

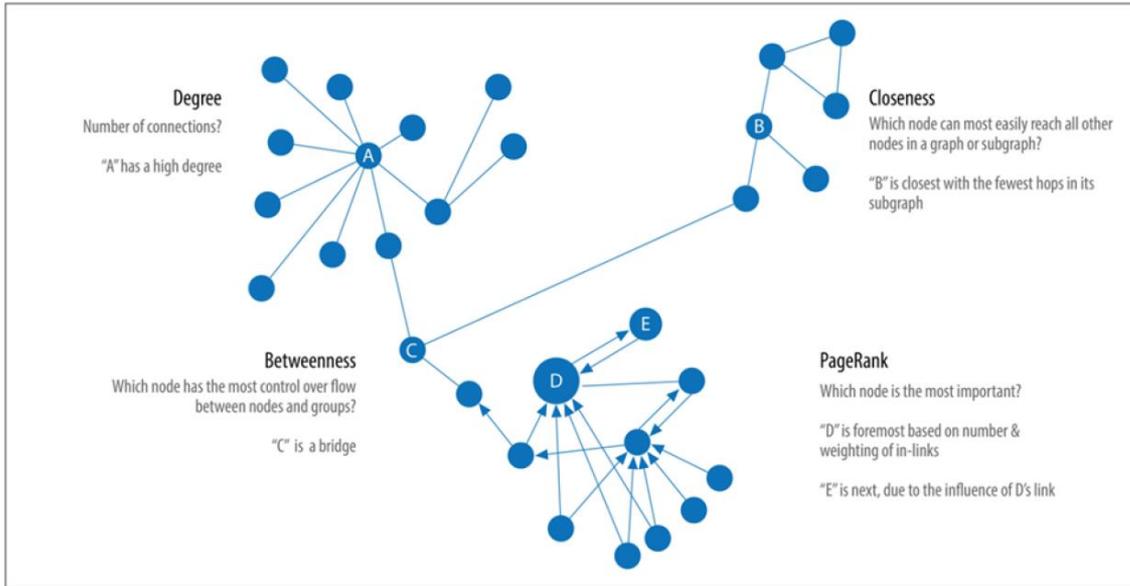
- E-mail
- Messagerie instantanée
- Réseau social interne d'entreprise



- **Différents indicateurs d'usage de ces outils**

- **Activité** : nombre de contacts à qui on écrit, messages envoyés, pages du réseau social consultées, actions réalisées sur le réseau social
- **Popularité** : nombre de contacts qui nous écrivent, messages reçus, nombre de vues et d'actions réalisées sur les pages gérées
- **Qualité du positionnement** : rapidité d'accès à l'information, contrôle de l'information, prestige
- **Transversalité** : intensité des échanges en dehors de la Direction de l'Innovation

L'analyse des réseaux sociaux apporte des indications sur l'accès à l'information !



$$\text{Poids lien} = \frac{\mathbb{1}_{\text{Envoi en } \dot{A}} + 0.5 \times \mathbb{1}_{\text{Envoi en CC ou CCI}}}{\# (\text{Destinataires en } \dot{A}) + 0.5 \times \# (\text{Destinataires en CC et CCI})}$$

→ Des variables fondées sur l'analyse de la distance minimale entre les nœuds du réseau

Comment dire des choses intelligentes avec des données pas trop structurées ?

3) Se ramener à des choses simples et interprétables

Unité statistique : trace (émetteur, multiples destinataires)

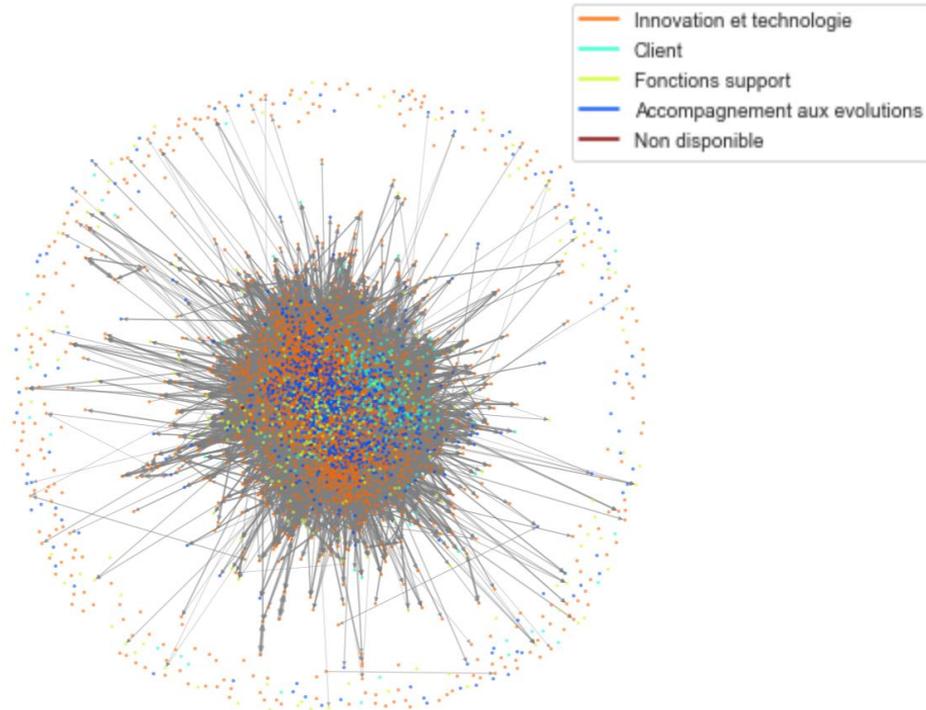
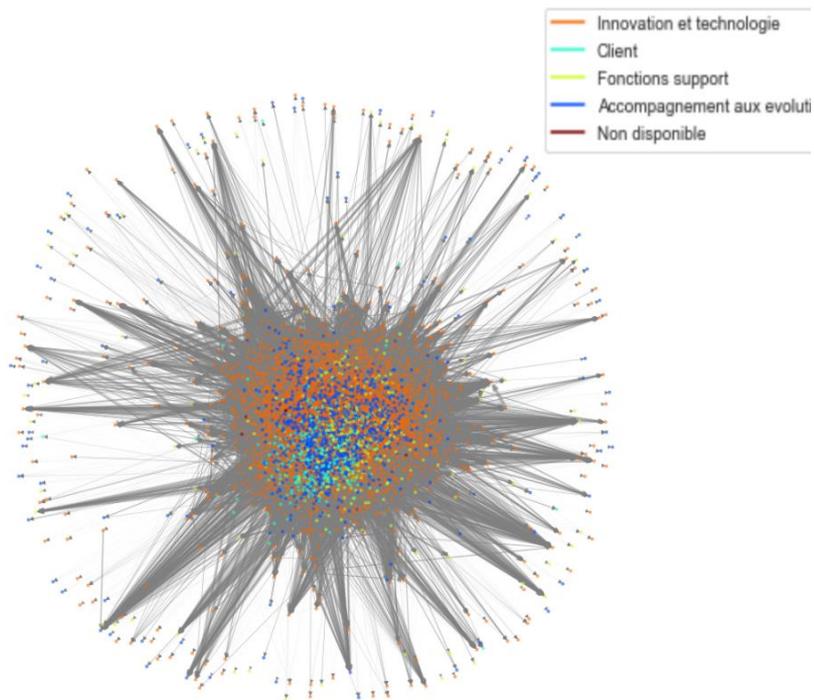


Unité statistique : lien pondéré (émetteur, destinataire)



Unité statistique : salarié

Des réseaux de communication peu denses et avec une forte réciprocity



Réciprocité : 67 %

Distance moyenne

entre deux nœuds : 2,3

E-MAIL

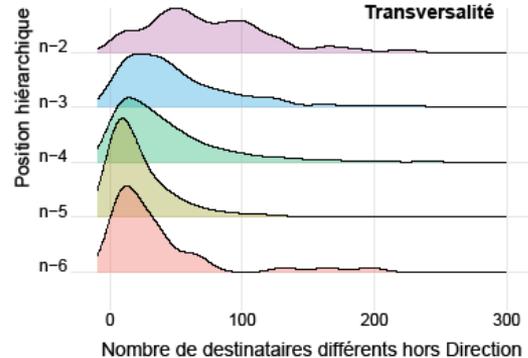
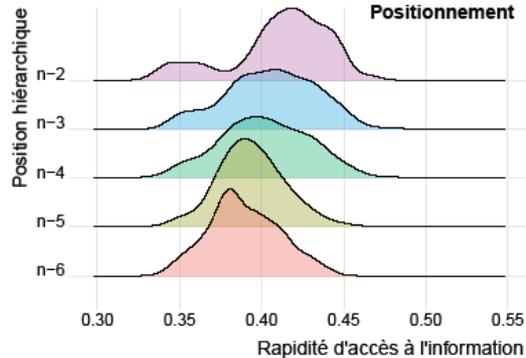
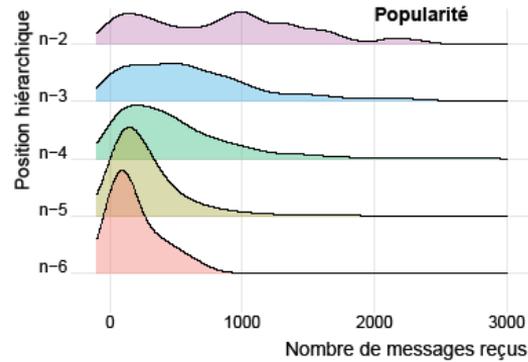
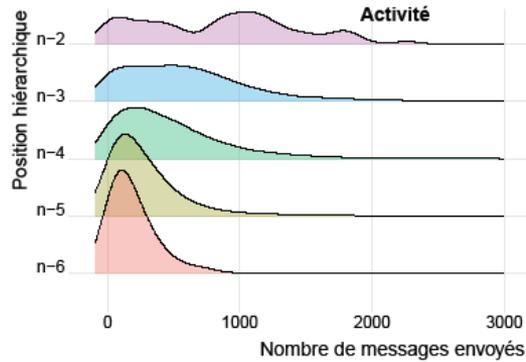


NetworkX
Network Analysis in Python

MESSAGERIE
INSTANTANÉE



Plus on progresse dans la hiérarchie de l'entreprise, plus on a une utilisation intensive du courrier électronique



! Une utilisation

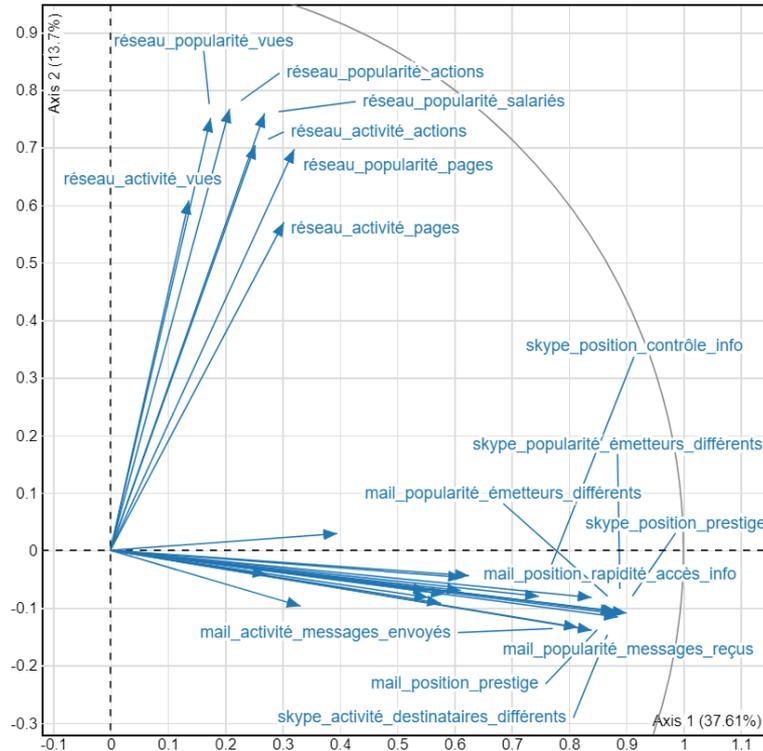
- Plus intensive d'un point de vue quantitative
- Avec un positionnement plus stratégique dans le réseau des échanges par mail
- Plus transversale

Comment dire des choses intelligentes avec des données pas trop structurées ?

4) Discuter de jolis graphiques

Typologie des profils de communication

Les salariés combinent souvent le mail et la messagerie instantanée



⚠ Méthodologie

→ Analyse en composantes principales

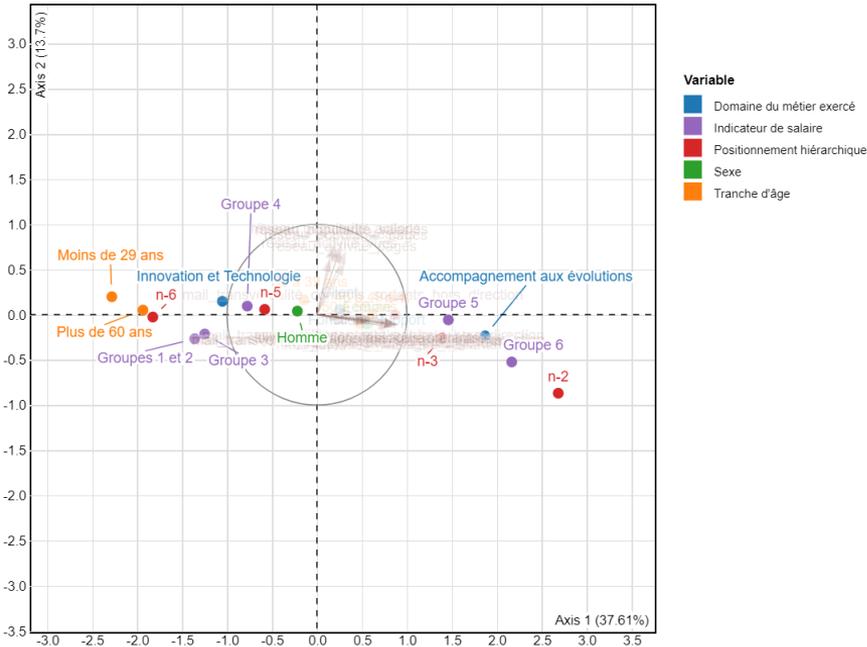
→ Variables actives : mesures d'utilisation des outils de communication

→ Variables supplémentaires : informations sur les salariés et leur métier

Axe horizontal : utilisation (quantitative et « qualitative ») du mail et de la messagerie instantanée

Axe vertical : activité et popularité sur le réseau social interne de l'entreprise

Les plus jeunes et les plus vieux sont plus silencieux que leurs collègues



⚠ Méthodologie

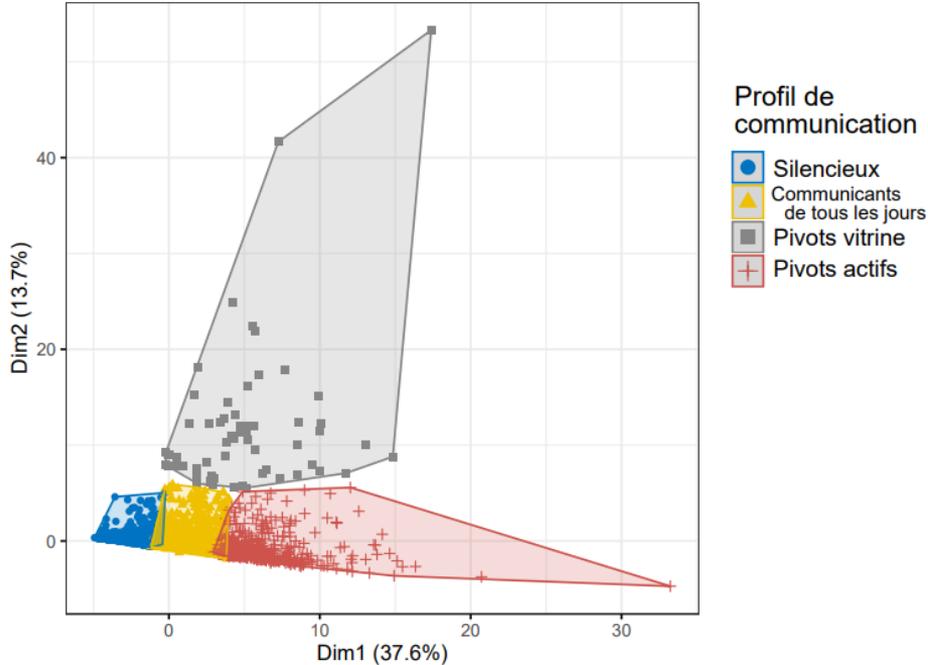
- Analyse en composantes principales
- Variables actives : mesures d'utilisation des outils de communication
- Variables supplémentaires : informations sur les salariés et leur métier

De forts effets liés au métier exercé : les *managers* communiquent plus

Les salariés de plus de 60 ans en **déprise** avec l'entreprise ?

Les plus jeunes et les plus vieux sont plus silencieux que leurs collègues

Classification des profils de communication
Direction de l'Innovation
Septembre – Novembre 2019



! **Méthodologie**

→ *Classification Ascendante Hiérarchique*

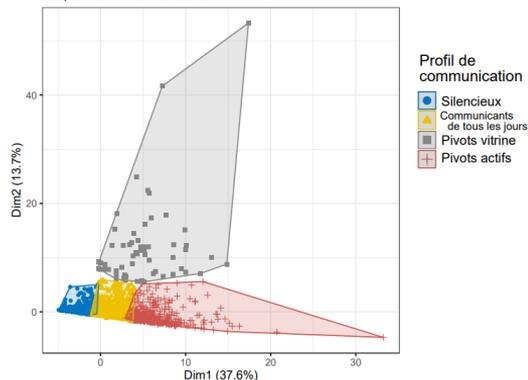
→ À partir des résultats de l'Analyse en Composantes Principales

Comment dire des choses intelligentes avec des données pas trop structurées ?

5) À la fin, utiliser des techniques statistiques « du monde d'avant » pour en tirer quelques résultats

Les pivots actifs nouent plutôt des liens forts

Classification des profils de communication
 Direction de l'Innovation
 Septembre - Novembre 2019



⚠ Méthodologie

→ *Classification Ascendante Hiérarchique*

→ **Statistiques descriptives** pour décrire les clusters

	Les silencieux	Les communicants de tous les jours	Les pivots actifs	Les pivots vitrine
Proportion	41 %	46 %	12 %	1 %
Part de femmes	24 %	29 %	40 %	32 %
Mails par contact en moyenne	8	12	15	12

Toutes choses égales par ailleurs, quelques résultats sur les différences entre femmes et hommes

1. Les femmes sont plus souvent des pivots actives que les hommes

- L'explication par les rôles de genre ? Les femmes garantes d'une « sociabilité plus enveloppante » ?

2. À métier, activité et popularité équivalents, les hommes bénéficient en moyenne d'un positionnement plus stratégique dans le réseau des communications par mail

- Les contacts des salariés femmes, moins pourvoyeurs d'information ?

❗ Difficile de confirmer les interprétations au vu des données à disposition

❗ Méthodologie

→ Classification Ascendante Hiérarchique

→ Régression logistique pour analyser les déterminants « toutes choses égales par ailleurs » des caractéristiques de la communication



Conclusion (1) – Comment dire des choses intelligentes avec des données pas trop structurées ?

1. Chercher à réduire la dimensionnalité du problème si c'est possible
2. Ne pas avoir peur du vocabulaire des *geeks*
3. Se ramener à des choses simples et interprétables
4. Discuter de jolis graphiques
5. À la fin, utiliser des techniques statistiques « du monde d'avant » pour en tirer quelques résultats



Conclusion (2) – Limites et perspectives

👍 Limites

- Généralisabilité des résultats obtenus ?
- On observe que des communications formalisées par un message instantané ou un mail (échanges oraux ? Documents consultés ou rédigés ?)
- On ne connaît pas le contenu des échanges

👍 Perspectives potentielles d'études

- Actualiser cette étude descriptive après la crise sanitaire
→ Dans le « nouveau monde » avec télétravail plus important qu'avant, des changements structurels dans les façons de communiquer ?
- Analyser le **contenu des échanges** → Une communication moins formelle va-t-elle de pair avec une communication plus horizontale ?
- Coupler les résultats quantitatifs **avec des analyses qualitatives** → Précision des mécanismes à l'œuvre

Merci 😊

Analyse des communicants pivots à partir des données massives de communication

Maxime Bergeat, Anca Boboc

Journées de méthodologie statistique de l'Insee – 31 mars 2022



université
PARIS-SACLAY

