

Challenge #2 : Zéro Email
Soumiya Razzouk, Hamidou Kane, Alexandre Kessler

Table des matières

I. Business Understanding.....	3
II. Data Understanding.....	5
III. Data Preparation.....	8
IV. Modeling.....	9
V. Evaluation.....	10

I. Business Understanding

L'analyse du "Business Understanding" est basée sur les concepts, résultats et discussions de l'article "Dépendance au courrier électronique : effets sur le technostress et la surcharge informationnelle et répercussions sur la performance" par Hajer KEFI, Michel KALIKA et Najma SAIDANI (2021).

a) Contexte et objectif du projet

Le courrier électronique est un moyen de communication omniprésent dans le monde professionnel, interrompant en moyenne les travailleurs 70 fois par jour (Jackson et al., 2003). Cette omniprésence pose des défis de gestion des communications par email, impactant la productivité et le bien-être des employés. Le projet vise à réduire la surcharge de mails au sein d'une entreprise et la dépendance des employés à leur messagerie électronique.

b) Qu'est-ce que la surcharge d'email et comment se caractérise-t-elle ?

La surcharge d'email est l'usage excessif des courriers électroniques au travail, se manifestant par une surcharge informationnelle et des formes de technostress (Soucek et Moser, 2010; J.-F. Stich et al., 2019). La techno-surcharge et la techno-invasion sont les principales sources de technostress. Nous définissons plus précisément ces concepts dans les paragraphes ci-après.

i) Surcharge informationnelle

La surcharge informationnelle se manifeste par :

- Déséquilibre informationnel : les besoins en information excèdent les capacités de traitement des utilisateurs (Galbraith, 1974).
- Non-criticité de l'information : la majorité des emails reçus ne sont pas critiques, submergeant l'utilisateur (Sumecki et al., 2011).
- Boîtes de réception comme outils d'archivage : la crainte de perdre des informations transforme les boîtes de réception en systèmes de rangement (Capra, Khanova, & Ramdeen, 2013).

ii) Techno-surcharge

La techno-surcharge se caractérise par :

- Augmentation de la pression de travail : la nécessité de vérifier et répondre aux emails fréquemment (Barley et al., 2011).
- Réduction des plages horaires pour tâches professionnelles : le temps consacré aux emails réduit celui pour les tâches spécifiques, augmentant la difficulté des tâches quotidiennes.

iii) Techno-invasion

La techno-invasion est causée par :

- Connectivité permanente : la possibilité d'accéder à son travail à tout moment via son téléphone par exemple brouille les frontières entre vie professionnelle et privée.
- Séparation vie professionnelle/privée : maintenir une séparation claire devient difficile avec l'ubiquité des outils de communication.

c) Mesures de la surcharge informationnelle, du techno-Surcharge et de la techno-invasion et impacts sur la performance individuelle des employés

L'évaluation de la surcharge d'email est subjective et non limitée au volume d'emails (KEFI et al., 2021). Cependant, comme nous ne disposons pas de données qualitatives, nous n'avons pas d'autre choix que de définir des indicateurs de surcharge d'email basés sur le volume d'emails échangés.

<u>Indicateurs de surcharge d'email</u>	<u>Items de mesure</u>	<u>Justifications</u>
Surcharge informationnelle	Nombre total d'emails envoyés sur la période par les employés.	Mesure de la charge d'information globale échangée dans l'entreprise par les collaborateurs.
Techno-surcharge	Fréquence d'email envoyés par employé par jour.	Mesure de la pression ressentie par les employés pour communiquer fréquemment.
Techno-invasion	Nombre d'emails envoyés ou reçus pendant les weekends et en dehors des heures de "bureau"	Signaler un empiètement de la vie professionnelle dans la vie privée et une attente constante de disponibilité de la part de des employés.

d) Traduction du problème métier en un problème data science et solution proposée

En tant que data scientists, notre approche consistera d'abord à analyser les échanges d'emails entre les employés, puis à les modéliser sous la forme d'un graphe.

Ce graphe sera pondéré et orienté, où :

- Nœud : représente un collaborateur et ses attributs (ex: grade).
- Arrête : représente un email envoyé d'un collaborateur à un autre.
- Poids de l'arête : représente le nombre d'emails envoyés d'un collaborateur à un autre.

Notre objectif sera de détecter, grâce à ces analyses, des schémas ou des pratiques de communication inefficaces au sein de l'entreprise. Nous proposerons des solutions ciblées qui, bien que ne mesurant pas directement les gains de productivité, permettront de réduire efficacement le volume des échanges par courriel. Cela contribuera indirectement à améliorer la productivité et le bien-être des salariés au sein de l'organisation.

La qualité de nos solutions sera mesurée en examinant le taux de réduction des indicateurs de surcharge d'emails après la mise en place des solutions recommandées.

e) Contraintes liées au projet

Plusieurs contraintes sont à souligner, car elles pourraient limiter la portée et l'efficacité de nos recommandations :

- Accès limité aux données : pas d'accès au contenu des emails.

- Période de données restreinte : données sur une période limitée, influencée par des événements spécifiques.
- Absence de données qualitatives : manque d'informations sur les perceptions personnelles des employés face à la surcharge d'emails.

II. Data Understanding

a) Collecte des données

Le dataset provient d'un cabinet de conseil et contient des échanges d'emails sur une période du 4 au 19 mars 2019. Le dataset comprend 1 174 928 lignes, chaque ligne représentant un email envoyé ou reçu par un collaborateur à une date et heure spécifiques.

b) Description des Données

Les variables initiales incluses dans le dataset sont les suivantes :

- GroupName1 : Poste/Titre du collaborateur
- RecipientName : Le collaborateur qui a envoyé ou reçu l'email
- MessSize : Taille du message en mégaoctets
- MessDate : Date et heure de l'email
- Subject : Le sujet de l'email
- Id_Direction : Indique si l'email a été envoyé (1) ou reçu (2).
- PartnerTypeName : Type d'interlocuteur (Interne, Internet, ou Adresse locale non identifiée)
- PartnerName : Nom de domaine si l'interlocuteur est externe
- Recipient_Display : L'interlocuteur (**il ne s'agit pas d'un identifiant unique et dépend du sujet de l'email**)
- Recipient_Name : Le collaborateur (**il ne s'agit pas d'un identifiant unique et dépend du sujet de l'email**)
- Id_Recipient : Identifiant (Id) unique du collaborateur
- Id_Regroup : Id unique de l'interlocuteur.

c) Qualité des données

L'analyse des valeurs manquantes dans les données révèle que la plupart des variables sont complètes, excepté deux :

- Grade_Interlocuteur : 161 110 valeurs manquantes.
- Subject : 74 637 valeurs manquantes.

Les valeurs des variables semblent également cohérentes avec les définitions données initialement.

d) Exploration des données

i) Nombre d'emails envoyés

Nous avons calculé le nombre moyen d'emails envoyés par collaborateur sur la période étudiée. Chaque collaborateur a envoyé en moyenne 268,73 emails, soit environ 16,80 emails par jour,

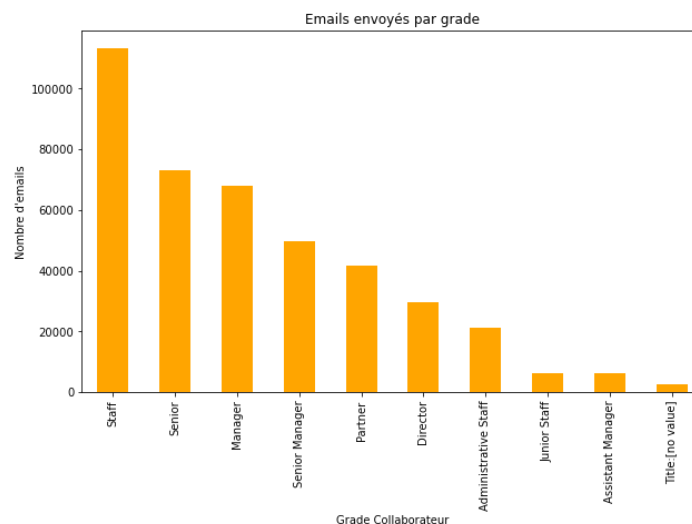
indiquant une charge d'e-mail significative pour chaque individu. Nous avons également observé que 4469 emails ont été envoyés à eux-mêmes, peut-être des raisons comme la prise de notes rapide, le transfert de fichiers entre appareils, les rappels de réunions, et la conservation d'emails.

ii) Nombre d'emails par discussion

L'analyse du nombre d'emails par discussion montre une forte variabilité, avec certaines discussions comportant jusqu'à 2024 emails, et d'autres seulement un seul. Le nombre moyen d'emails par discussion est de 17,83, suggérant des échanges prolongés entre collaborateurs, ce qui peut contribuer à la surcharge d'email.

iii) Répartition des emails selon les grades

La répartition des emails envoyés selon les grades montre que le grade "Staff" envoie le plus grand nombre d'emails (113 427), suivi des grades "Senior" (73 158) et "Manager" (68 108). Une catégorie "Title: [no value]" a également été identifiée avec 2444 emails, représentant les envois d'emails sans grade associé.

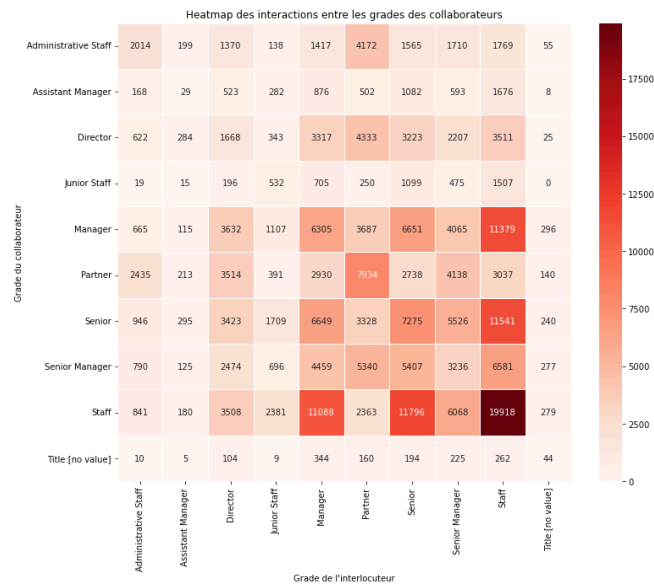


iv) Interlocuteurs Internes vs Externes

La majorité des emails (313 553) sont envoyés à des interlocuteurs internes, soit environ trois fois plus que ceux envoyés à des interlocuteurs externes. Cela indique que la communication par email est principalement intra-entreprise, suggérant que les efforts pour réduire la surcharge d'email devraient d'abord se concentrer sur les communications internes.

v) Interactions entre différents grades de collaborateurs

Le grade "Staff" est au centre des communications, envoyant et recevant le plus d'emails, notamment entre eux (19 918 emails). Ils ont des interactions élevées avec les "Managers" (11 379 emails) et les "Seniors" (11 541 emails). Les échanges significatifs incluent aussi les emails de "Staff" aux "Partners" (7 934 emails). En général, les grades supérieurs comme "Senior Manager", "Senior", et "Partner" interagissent fréquemment avec les "Staff", suggérant une communication régulière pour des directives ou collaborations. Ces observations indiquent que les "Staff" jouent un rôle central dans la communication au sein de l'organisation.



vi) Étude des sujets des mails

Nous avons observé 76 346 sujets uniques parmi les emails envoyés. Certains sujets comptent jusqu'à 2115 mails, tandis que d'autres ne sont associés qu'à un seul mail. Cette variété de sujets et la diversité dans le nombre de mails associés indiquent une large gamme de discussions au sein de l'entreprise.

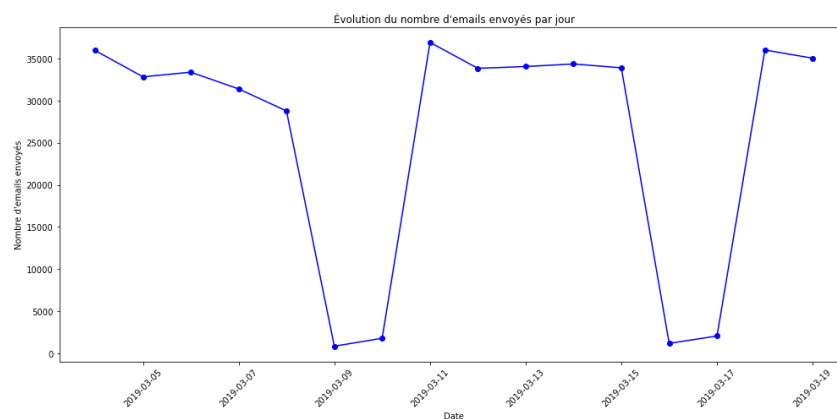
vii) Étude de la taille des mails

L'analyse de la taille des emails montre une gamme variée de tailles, allant de 0,0048 MB à 98,98 MB. Nous avons catégorisé les emails en trois groupes :

- Très légers / Légers : 0 à 1 MB
- Moyens : 1 à 10 MB
- Lourds / Très lourds : 10 à 100 MB

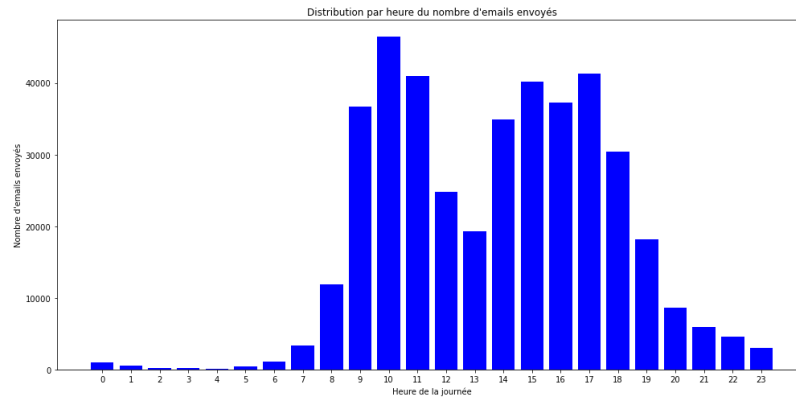
viii) Évolution Quotidienne du nombre d'emails envoyés

L'analyse de l'évolution quotidienne du nombre d'emails envoyés sur la période du 4 au 19 mars 2019 révèle des pics d'activité le lundi et des baisses significatives le week-end, ce qui est attendu dans un contexte professionnel.



ix) Distribution des emails envoyés par heure de la journée

La distribution des emails envoyés par heure montre un pic significatif entre 9h et 17h, avec une baisse entre 12h et 14h. Le nombre d'emails envoyés diminue fortement avant 8h et après 18h, reflétant une activité réduite en dehors des heures de bureau.



III. Data Preparation

a) Réindexation des colonnes

Nous avons ré-indexé les colonnes pour réorganiser l'ordre des colonnes pour les structurer de manière plus logique et faciliter l'analyse et la compréhension des données.

b) Renommage des colonnes

Nous avons renommé les colonnes pour des noms plus explicites et conformes aux besoins de l'analyse :

- Id_Regroup est renommé en Id_Interlocuteur.
- Id_Direction est renommé en Statut_Message (E=1 ou R=2).
- Id_Recipient est renommé en Id_Collaborateur.
- GroupName1 est renommé en Grade_Collobateur.
- Recipient_Name est renommé en Target if Statut_Message=2, Sender otherwise.
- PartnerTypeName est renommé en Statut_Interlocuteur.
- Partner_Name est renommé en Domaine_Name.
- Recipient_Display est renommé en Sender if Statut_Message=2, Target otherwise.

c) Association des grades aux collaborateurs

Nous avons créé un dictionnaire associant les Id_Collaborateur à leur Grade_Collobateur respectif.

d) Ajout du grade des interlocuteurs

Nous avons ajouté une nouvelle colonne "Grade_Interlocuteur" au DataFrame en utilisant le dictionnaire créé précédemment pour mapper les Id_Interlocuteur à leurs grades. Cela nous permet d'enrichir les données avec des informations supplémentaires sur les interlocuteurs des messages.

e) Filtrage des données

Nous avons filtré le data frame pour ne conserver que les messages envoyés (Statut_Message (E=1 ou R=2) == 1). Cela n'empêchera pas l'analyse du nombre d'emails reçus. En effet, nous modéliserons les emails envoyés sous la forme d'un graphe pondéré orienté, ce qui nous permettra d'analyser, pour chaque nœud, le nombre d'arcs entrants.

Pour plus de clarté, nous avons de nouveau renommé certaines colonnes :

- Target if Statut_Message=2, Sender otherwise est renommé en Sender.
- Sender if Statut_Message=2, Target otherwise est renommé en Target.

IV. Modeling

a) Centralité des degrés (Degree Centrality)

Objectif	
Permet d'identifier les collaborateurs qui envoient ou reçoivent le plus de mails. Un degré élevé peut indiquer une surcharge de communication.	
Avantages	Inconvénients
Facile à calculer et à interpréter.	Ne prend pas en compte l'importance des connexions.

b) Détection des communautés (Clustering)

Objectif	
Permet d'identifier des groupes de collaborateurs qui communiquent fréquemment entre eux, mais moins avec les autres. Utile pour détecter des silos.	
Avantages	Inconvénients
Facilite la détection de groupes de travail naturels.	Coût de calcul élevé pour les grands graphes. Peut être complexe à interpréter dans de grands graphes

V. Evaluation

a) **Réduction de la surcharge d'emails par la limitation des envois hors heures de bureau et par la taille des emails envoyés.**

Supprimer les emails envoyés en dehors des heures de bureau et le week-end pourrait réduire la surcharge d'emails. Les emails envoyés le week-end (5893 emails) et en dehors des heures de bureau en semaine pour un cabinet de conseil (8-19h) (20600 emails) représentent environ 8,5% du volume total d'emails échangés après nettoyage des données. La mise en œuvre d'une telle politique

encouragerait les employés à optimiser leurs communications pendant les heures de travail, en supposant que les employés ne reporteraient pas simplement ces envois aux heures de bureau, ce qui n'affecterait pas la surcharge, mais qu'ils intégreraient plutôt ces communications dans leurs échanges déjà planifiés pendant les heures autorisées.

De plus, la réduction du nombre d'emails envoyés pourrait être encore plus importante si l'on suppose que les emails de très grande taille, contenant probablement des pièces jointes, ne sont plus envoyés et que ces pièces jointes sont directement partagées dans un drive ou espace partagé accessible aux collaborateurs.

<u>Indicateurs de surcharge d'email</u>	<u>Items de mesure</u>	<u>Avant implémentation des solutions</u>	<u>Après implémentation des solutions</u>
Surcharge informationnelle	Nombre total d'emails envoyés sur la période par les employés.	309 088	282 595 (- 8,5 %)
Techno-surcharge	Fréquence d'email envoyés par employé par jour.	13.44 emails/jour	12.28 emails/jour
Techno-invasion	Nombre d'emails envoyés ou reçus pendant les weekends et en dehors des heures de "bureau"	20 600	0

b) Réduction de la surcharge d'emails par l'analyse de la centralité des degrés

Les résultats de la centralité par degré entrant montrent des nœuds ayant des très hauts degrés. Ces nœuds sont probablement des hubs dans le réseau, recevant des emails de nombreux autres nœuds. Ils jouent donc un rôle central dans le réseau, indiquant des points de convergence d'informations et donc potentiellement une surcharge d'informations. Ces individus peuvent avoir du mal à traiter et répondre efficacement à tous les messages. **Nous recommandons d'étudier ces nœuds pour comprendre leur rôle exact dans le réseau.**

Les résultats de la centralité par degré sortant montrent que certains nœuds, tels que 1620429, 1629, 1683, 1331, et 13215, ont un nombre de connexions sortantes extrêmement élevé, avec des degrés de 1468, 1239, 1120, 861 et 852 respectivement. Ces nœuds semblent donc jouer un rôle de "diffuseurs d'information" au sein de l'entreprise. En comparaison, le degré sortant moyen des nœuds ayant au moins une arête sortante est de 46,13. Pour approfondir cette analyse, nous avons identifié les nœuds ayant un degré sortant supérieur à 100, car cela représente une activité de communication particulièrement élevée par rapport aux autres. Ces nœuds ont envoyé un total de 109 070 emails sur la période. Parmi eux, nous avons sélectionné ceux ayant envoyé plus de 500 emails, bien au-dessus de la moyenne de 270 emails par collaborateur. L'hypothèse est que ces nœuds, en raison de leur nombre élevé de connexions et d'emails envoyés, participent activement à la communication interne de l'entreprise, en partageant à large échelle des informations comme les actualités de l'entreprise.

Nous avons ensuite calculé la somme des écarts entre la moyenne d'emails envoyés par l'ensemble des nœuds et le nombre d'emails envoyés par ces nœuds sélectionnés, obtenant une somme des écarts de 41 716. L'idée est de mesurer combien ces nœuds surpassent la moyenne d'envois d'emails et de supposer que cet écart représente les emails de communication interne à large échelle. En supposant que ces nœuds avaient limité leurs envois à 270 emails chacun et avaient utilisé une plateforme collaborative (ou d'autres outils) pour les communications internes, **nous aurions pu réduire le nombre total d'emails envoyés de 18 707, soit une réduction de 6,05% du nombre total d'emails envoyés**, après nettoyage des données. Des solutions alternatives de communication pourraient donc absorber une partie significative de cette activité, réduisant ainsi la charge d'emails.

<u>Indicateurs de surcharge d'email</u>	<u>Items de mesure</u>	<u>Avant implémentation des solutions</u>	<u>Après implémentation des solutions</u>
Surcharge informationnelle	Nombre total d'emails envoyés sur la période par les employés.	309 088	290 381 (- 6,05%)
Techno-surcharge	Fréquence d'email envoyés par employé par jour.	13,44 emails/jour	12,62 emails/jour
Techno-invasion	Nombre d'emails envoyés ou reçus pendant les weekends et en dehors des heures de "bureau"	20 600	20 600

c) Réduction de la surcharge d'emails par l'analyse des communications au sein de chaque communauté

L'utilisation de la méthode de Louvain pour l'analyse des réseaux sociaux a révélé l'existence de 18 communautés distinctes au sein de l'organisation. La plus grande de ces communautés regroupe 1087 membres, tandis que la plus petite en compte seulement 18, avec une taille moyenne de communauté s'élevant à 434 membres. Cette identification des communautés suggère une structuration complexe et variée des réseaux de collaboration et de communication au sein de l'organisation.

Dans une première phase d'analyse, nous nous sommes intéressés aux profils les plus impliqués dans les échanges de courriels. Il ressort que le personnel, avec 81,960 courriels envoyés et 59,754 reçus, occupe une position centrale, reflétant son rôle crucial dans les opérations quotidiennes. Ce groupe est suivi par les seniors et les managers qui démontrent également un volume élevé de communications, illustrant ainsi les dynamiques de communication inter-hiérarchiques.

En approfondissant l'analyse des communications du personnel administratif, nous observons que certains messages, tels que le mail identifié par sa taille (0.0727100372314453) et son sujet ('5495c59f42a8588c0e160569964c6b8a48ead180'), sont diffusés à un large éventail de destinataires de différents grades. Ce phénomène suggère que le contenu de ces messages pourraient être des annonces générales plutôt que des communications spécifiques à des projets. En effet, ce mail a été envoyé à 187 individus de grades variés, allant du personnel de bureau au management senior.

L'analyse des échanges intra et inter-grades révèle que 85,6% des communications sont inter-grades (verticales), tandis que 14,4% sont intra-grades (horizontales), soulignant la prédominance des échanges entre différents niveaux hiérarchiques au sein de l'organisation.

Pour optimiser la communication au sein de chaque communauté, il serait judicieux de créer des groupes dédiés sur des plateformes collaboratives comme Slack ou Microsoft Teams. Chaque groupe pourrait inclure des canaux spécifiques pour les annonces, les échanges professionnels et le partage de documents. Cette structuration permettrait de réduire la surcharge d'informations dans les espaces personnels et de faciliter la gestion des projets communs, améliorant ainsi l'efficacité et la clarté des communications internes.

Conclusion

Notre étude montre que l'analyse de la centralité des degrés et la détection des communautés sont des solutions potentielles pour réduire la surcharge d'emails et améliorer la communication interne en entreprise. Ces approches permettent d'identifier les points de surcharge et les dynamiques de groupe, proposant des solutions concrètes pour optimiser les échanges.

La centralité des degrés révèle des acteurs clés dans la gestion des informations, tandis que la détection des communautés aide à comprendre les structures de communication. Combinées à des mesures comme la restriction des emails hors des heures de bureau, ces analyses réduisent significativement la surcharge informationnelle.

L'introduction de plateformes collaboratives pourrait aussi transformer la communication interne en déplaçant les interactions vers des canaux plus interactifs et organisés, réduisant ainsi le volume d'emails et améliorant la réactivité et l'efficacité des échanges. Toutefois, il faut rester vigilant face aux risques de distractions et aux coûts. Une évaluation continue de l'efficacité des solutions est nécessaire.

En résumé, la combinaison de techniques analytiques et de technologies de communication adaptées est essentielle pour améliorer la gestion de la surcharge d'informations et optimiser la communication interne, soutenant ainsi l'efficacité et le bien-être des collaborateurs tout en favorisant une culture de travail collaborative et innovante.

Références

Barley, S. R., Meyerson, D. E., & Grodal, S. (2011). E-mail as a source and symbol of stress. *Organization Science*, 22(4), 887-906. DOI: 10.1287/orsc.1100.0573

Capra, R. G., Khanova, J., & Ramdeen, S. (2013). Work and personal e-mail use by university employees: PIM practices across domain boundaries. *Journal of the American Society for Information Science and Technology*, 64(5), 1029-1044. DOI: 10.1002/asi.22801

Galbraith, J. (1974). *Designing complex organizations*. Reading, MA: Addison-Wesley.

Hajer Kefi, Michel Kalika, & Najma Saidani (2021). Dépendance au courrier électronique : effets sur le technostress et la surcharge informationnelle et répercussions sur la performance. *Systèmes d'information & management*, 26(1), 45-83. DOI: 10.3917/sim.211.0045

Soucek, R., & Moser, K. (2010). Coping with information overload in email communication: Evaluation of a training intervention. *Computers in Human Behavior*, 26(6), 1458-1466. DOI: 10.1016/j.chb.2010.04.024

Stich, J.-F., Tarafdar, M., Cooper, C. L., & Stacey, P. (2019). Electronic communication in the workplace: Boon or bane? *Journal of Organizational Effectiveness: People and Performance*, 6(1), 42-64. DOI: 10.1108/JOEPP-07-2018-0043

Sumecki, D., Chipulu, M., & Ojiako, U. (2011). Email overload: Exploring the moderating role of the perception of email as a 'business critical' tool. *International Journal of Information Management*, 31(5), 407-414. DOI: 10.1016/j.ijinfomgt.2011.02.002

Jackson, T., Dawson, R., & Wilson, D. (2003). Reducing the effect of email interruptions on employees. *International Journal of Information Management*, 23(1), 55-65. [https://doi.org/10.1016/S0268-4012\(02\)00068-3](https://doi.org/10.1016/S0268-4012(02)00068-3)