UCLouvain

Rapport Web Mining

MLSMM2153 – Web Mining

Professeurs : Corentin Vande Kerckhove & Sylvain Courtain

Basé en partie sur les supports du Prof. P. Francq

Exploration du discours écologique dans les sites de festivals internationaux

Étudiants :
Diégo Charles-Pirlot 08112000
Valère Scheuren 72621800

Année académique 2025–2026

Table des matières

1 Introduction et Problématique

La durabilité est devenue un enjeu central dans l'organisation des festivals et événements culturels contemporains. Face à la pression croissante des publics, des autorités et des partenaires institutionnels, les organisateurs sont amenés à intégrer des considérations environnementales non seulement dans leurs pratiques opérationnelles, mais également dans leur **communication officielle**. Cette dernière joue un rôle stratégique : elle façonne la perception du public, légitime les choix organisationnels et contribue à la construction d'une image de marque responsable. Cependant, un décalage peut exister entre les actions concrètes mises en œuvre et le discours écologique affiché. Certains festivals mettent en avant des engagements précis et mesurables tels que la **mobilité durable**, la **gestion des déchets** ou l'**alimentation responsable** tandis que d'autres recourent davantage à des formulations vagues ou purement symboliques, dépourvues de détails opérationnels.

Objectifs et questions de recherche

Dans ce contexte, ce projet vise à analyser de manière systématique la manière dont les festivals communiquent sur la durabilité à travers leurs sites web officiels. En mobilisant des techniques de *Web Mining*, de *Text Mining* et de *Link Analysis*, l'objectif est d'identifier les thématiques dominantes du discours écologique et d'évaluer la cohérence de ces engagements. Notre étude s'articule autour de trois questions de recherche principales :

1. **Thématiques :** Quelles sont les thématiques écologiques les plus fréquemment mobilisées dans la communication numérique des festivals ?
2. **Profils :** Peut-on distinguer des profils de discours contrastés, notamment entre l'affichage d'engagements concrets et de promesses générales ?
3. **Réseaux :** Comment les thématiques écologiques s'organisent-elles et se connectent-elles entre elles au sein d'un réseau sémantique et relationnel ?

Cadre méthodologique et principes de reproductibilité

Le projet adopte une approche exploratoire, contrainte par l'hétérogénéité des architectures web et la diversité linguistique du secteur. Les choix méthodologiques (du filtrage de pages au **clustering K-means** et à l'**analyse de réseaux**) visent un compromis entre réduction du bruit sémantique, robustesse analytique et interprétabilité.

Dans une logique de **transparence et de réplicabilité**, l'ensemble du pipeline a été implémenté sous forme de notebooks *Jupyter*. Le code source complet est disponible dans un dépôt GitHub public [1], permettant la reproduction des résultats et leur extension à d'autres corpus.

1. https://github.com/ChilliAngel/Web-Mining-Festivals-Ecolo.git

2 Collecte de données et constitution du corpus

Notre processus de *Data Collection* repose sur un échantillon contrôlé de **30 sites web de festivals internationaux**. Afin de garantir une comparaison internationale pertinente, chaque source a été documentée avec son URL de départ, son pays d'origine et sa langue principale (principalement le français, l'anglais et le néerlandais).

2.1 Stratégie de Crawling et Exploration

À partir des URL racines, nous avons déployé une stratégie de *crawling* ciblée, paramétrée par des contraintes de faisabilité et de *politeness* (délais entre requêtes, profondeur maximale) afin de ne pas saturer les serveurs hôtes. Pour structurer l'exploration et limiter la dérive thématique, nous avons mobilisé un **thésaurus multilingue** [2] de mots-clés liés à l'écologie et à la durabilité

Cette approche méthodique nous a permis d'extraire initialement un volume élevé de pages et de segments textuels avant filtrage. Pour chaque document, le pipeline a récupéré automatiquement le titre, le texte principal (après suppression des scripts et éléments de navigation) ainsi que la structure des hyperliens internes, afin de permettre l'analyse relationnelle.

2.2 Filtrage et Purification du Corpus

Le corpus initial présentait un bruit structurel conséquent (pages de billetterie, boutiques, espaces presse). Pour obtenir une base de données exploitable et cohérente, nous avons appliqué trois niveaux de filtrage rigoureux :

- **Filtrage par longueur :** Seules les pages dépassant un seuil de **200 mots** ont été conservées pour éviter les contenus transactionnels ou pauvres en contenu sémantique.
- **Exclusion sémantique :** Une liste de motifs d'exclusion (*tickets, shop, login, partner, legal*) a permis d'écarter les pages structurellement hors-sujet.
- **Déduplication et normalisation :** Les URL ont été normalisées pour supprimer les doublons techniques et stabiliser le périmètre d'analyse.

Au final, le corpus exploité pour le Text Mining contient **55 pages**, tandis que les graphes Gephi agrègent l'information à un niveau plus haut (festivals et catégories thématiques), ce qui explique des tailles de réseaux différentes.

2. Un thésaurus est un vocabulaire contrôlé structurant les concepts d'un domaine et leurs relations sémantiques (synonymie, hiérarchie, association), couramment utilisé en sciences de l'information et en recherche d'information ; voir Aitchison, Gilchrist & Bawden (2000).

3 Prétraitement textuel et stratégie de sélection des contenus

Étant donné la masse d'informations, nous avons décidé de concentrer notre extraction sur les zones textuelles à forte densité sémantique. Notre protocole privilégie les segments où les mots les plus définissants sont localisés, souvent dans les premières sections de présentation des engagements.

3.1 Sélectivité des contenus et justification méthodologique

Ce choix repose sur les caractéristiques des documents identifiés lors de la phase de *Data Collection* :

- **Synthèse thématique :** Extraire le texte complet aurait généré trop de bruit lexical (mentions légales, menus). Les mots-clés essentiels se situent généralement dans les paragraphes introductifs des chartes écologiques.
- **Compensation par les liens :** En cas de perte mineure de sens lors de l'analyse textuelle, nous compensons cette limite par l'**analyse des liens** internes, préservant la cohérence globale du réseau thématique.

Après l'extraction, nous avons procédé à une **normalisation complète** et une **tokenisation word-level** adaptée au multilinguisme. Cette étape réduit l'espace nécessaire au stockage tout en préservant la significativité pour les calculs de similarité.

4 Vectorisation et mesures de similarité

Afin de permettre les regroupements thématiques et les analyses de proximité sémantique, le corpus a été transformé en une représentation vectorielle à l'aide de la pondération **TF–IDF** (*Term Frequency–Inverse Document Frequency*). Ce choix est adapté à la comparaison de documents web hétérogènes, en limitant l'influence des termes trop fréquents tout en valorisant les mots discriminants.

Chaque page est ainsi représentée par un vecteur TF–IDF, et les proximités entre documents sont évaluées à l'aide de la similarité cosinus, ce qui permet de comparer des contenus de longueurs et de structures différentes. Les paramètres ont été fixés afin de réduire le bruit *boilerplate* tout en préservant la capacité discriminante du vocabulaire : `min_df=3`, `max_df=0.25`, stopwords multilingues (FR/EN/NL) et stopwords experts liés à la navigation, à la billetterie, aux marques et à la temporalité événementielle.

4.1 Exemple de Similarité : Le Cas Tomorrowland

Le tableau ci-dessous présente les festivals ayant les scores de similarité les plus élevés par rapport à **Tomorrowland**, illustrant l'efficacité du regroupement thématique.

Document / Festival	Score de similarité
Dour Festival - Charte Environnement	0.5842
Esperanzah ! - Engagement Durable	0.5215
Hellfest - Green Team	0.4988
Montreux Jazz - Social Responsibility	0.4760
Burning Man - Net Zero Vision	0.4532

TABLE 1 – Scores de similarité sémantique par rapport à Tomorrowland.

5 Analyse des Résultats : Positionnements Discursifs

5.1 Matrice « Actions vs Promesses » : Typologie des Positionnements

Après le scoring sémantique, nous avons positionné chaque festival sur une matrice croisant la densité de formulations concrètes (Axe Y) et aspirationnelles (Axe X). La visualisation met en évidence des contrastes nets :

— **LEADERS (Haut-Droite) : Tomorrowland** s'impose ici, combinant un niveau très élevé de preuves techniques et une narration de valeurs puissante.
— **PRAGMATIQUES (Haut-Gauche) :** Des événements comme **Dour, Esperanzah ! et le Hellfest** privilégient une communication structurée autour d'éléments opérationnels vérifiables.
— **FAIBLE VISIBILITÉ (Bas-Gauche) :** Plusieurs festivals affichent un discours écologique marginal ou peu détectable par le lexique standard.

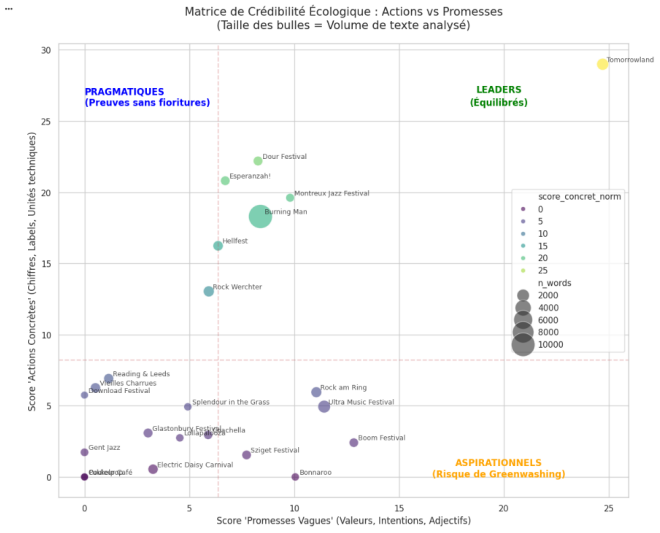

FIGURE 1 – Matrice de Crédibilité Écologique : Positionnement des festivals selon le ratio Actions concrètes vs. Promesses vagues (la taille des bulles indique le volume textuel analysé).

6 Clustering : Validation et Interprétation de la Dualité Discursive

Afin de synthétiser les thématiques transversales de la durabilité, nous utilisons une approche de clustering complétée par une analyse de modélisation textuelle. L'objectif est de vérifier si le corpus se segmente naturellement en profils discursifs distincts.

6.1 Clustering (K-means)

Nous appliquons K-means sur la matrice TF–IDF afin de regrouper automatiquement les pages selon leur profil lexical dominant. Le nombre de clusters est déterminé de manière exploratoire via la méthode du coude (inertie), et nous retenons $k = 2$ car il fournit une segmentation stable et interprétable du corpus en deux registres de communication (logistique terrain vs institutionnel/RSO).

(Seed fixé pour assurer la reproductibilité.)

6.2 Résultats : La Dualité Expérience vs Institution

L'analyse des termes dominants révèle que la structure éditoriale des sites influence fortement la segmentation. Nous identifions deux clusters principaux :

— **Cluster 0 : Le pôle « Expérience et Logistique Terrain »**
Mots-clés : city, park, access, camp, packages...
Représenté par des géants comme **Tomorrowland** ou **Rock Werchter**, ce groupe ancre l'écologie dans la réalité physique (gestion des campements, accès au site). La durabilité y est traitée comme un défi logistique lié à l'accueil de masse.

— **Cluster 1 : Le pôle « Institutionnel et Stratégie RSO »**
Mots-clés : billetterie, engagements, partenaires, RSO...
Porté par **Esperanzah !** ou le **Hellfest**, ce pôle adopte un ton formel. L'écologie est ici un argument de gouvernance, de valeurs et de partenariats officiels.

7 Analyse Visuelle et Validation des Résultats

Cette section mobilise des outils de visualisation avancés pour synthétiser les thématiques dominantes et valider statistiquement la structure de notre corpus.

7.1 Vision d'ensemble : Nuage de mots global

Le nuage de mots (*WordCloud*) présenté en Figure 2 synthétise les termes les plus fréquents du corpus après l'application de notre nettoyage sémantique expert. Cette représentation permet d'identifier immédiatement le "bruit de fond" thématique de la durabilité dans le secteur événementiel.

FIGURE 2 – Nuage de mots global du corpus : Vision synthétique de la durabilité.

La Figure 2 met en évidence les termes les plus fréquents après nettoyage. Cette visualisation est utilisée à des fins descriptives et exploratoires. La dominance de tokens génériques (p. ex. *people*, *event*, *year*) peut refléter un style éditorial général ou un discours marketing transversal, plutôt qu'un contenu environnemental substantiel. C'est précisément pour cette raison que le nuage de mots n'est jamais interprété isolément, mais systématiquement complété par l'analyse de réseau thématique et par des indicateurs ciblant la concrétisation des engagements (matrice Actions/Promesses, indice ICE).

7.2 Validation spatiale de la dualité discursive (PCA)

Afin de valider la séparation de nos clusters et de vérifier la cohérence de notre segmentation, nous appliquons une **Analyse en Composantes Principales (PCA)** sur la matrice TF-IDF. Cette méthode permet de projeter les documents, initialement dans un espace à haute dimension (2068 termes), dans un plan en deux dimensions.

La projection PCA (Figure 3) fournit une validation exploratoire de la séparation obtenue par

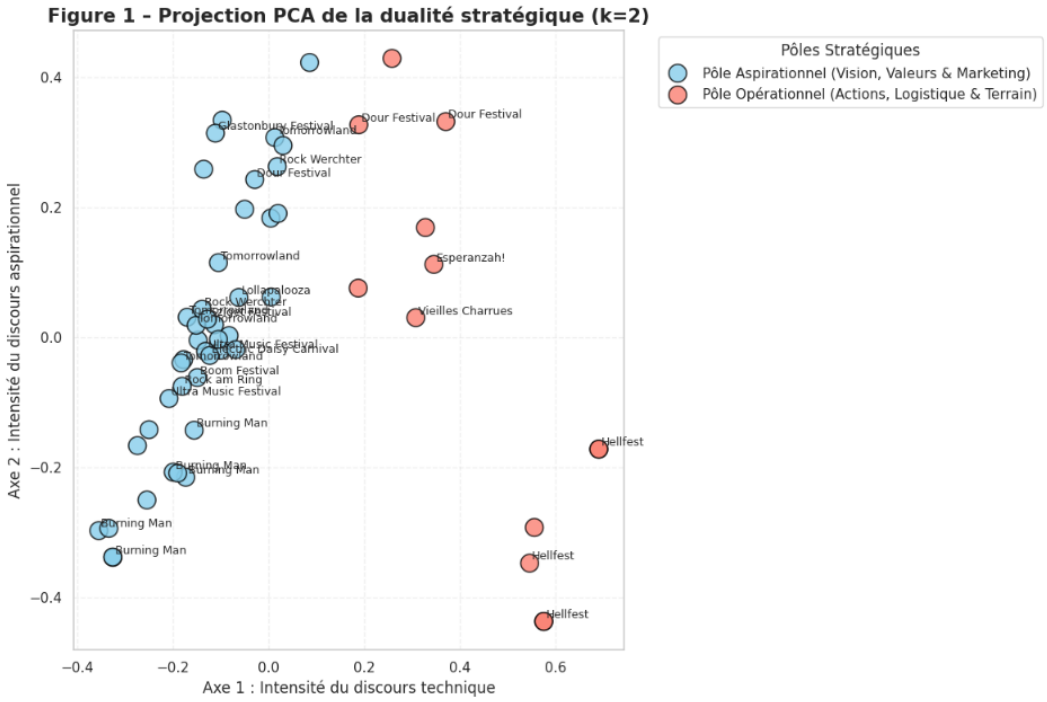

FIGURE 3 – Projection PCA de la dualité stratégique ($k = 2$).

K-means ($k = 2$). Elle met en évidence une structuration du corpus selon un gradient allant d'un registre davantage opérationnel/logistique à un registre davantage institutionnel/aspirationnel, ce qui renforce l'interprétation des deux profils discursifs.

8 Indicateurs Synthétiques et Limites

8.1 L'Indice ICE : Un Outil Exploratoire de Crédibilité

Nous avons construit l'indice **ICE** (*Indicator of Concrete Engagement*) afin de synthétiser, de manière exploratoire, la *crédibilité discursive* des communications environnementales observées. L'objectif n'est pas d'inférer une performance écologique réelle, mais de comparer des **styles de communication** : (i) un registre **concret/opérationnel** (actions, dispositifs, mesures, labels, indicateurs) versus (ii) un registre davantage **aspirationnel/valorisant**, appréhendé ici via un *Sentiment Score* utilisé comme **proxy lexical** d'un discours de valeurs et d'intentions.

Dans cette logique, l'ICE met en rapport la densité normalisée de formulations concrètes avec l'intensité du registre aspirationnel ; un plancher est appliqué au dénominateur afin d'éviter des explosions du ratio lorsque le score aspirationnel est proche de zéro :

$$ICE = \frac{\text{Score Concret Normalisé}}{\max(\text{Sentiment Score}, 0.01)} \tag{1}$$

Les résultats placent en tête des festivals tels qu'**Esperanzah !**, le **Hellfest** et **Burning Man**. Ces cas se caractérisent par une forte densité de marqueurs opérationnels (preuves, mesures, dispositifs) et une mobilisation plus contenue du lexique valorisant, ce qui correspond à un registre davantage orienté « discours-preuve » dans le corpus analysé.

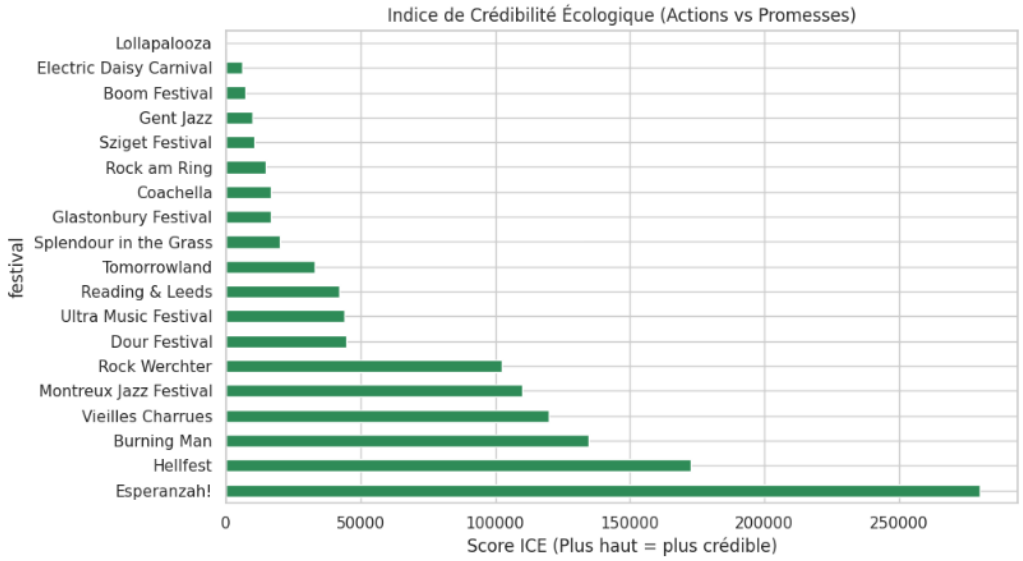

FIGURE 4 – Indice ICE : comparaison des festivals selon la densité d'actions concrètes et le registre aspirationnel.

8.2 Limites et Prudence d'Interprétation

Il est essentiel de souligner que l'ICE, au même titre que le clustering, constitue un **signal exploratoire**. Il mesure une **visibilité éditoriale** et un **style discursif**, et non une performance environnementale vérifiable sur le terrain. L'indice dépend fortement (i) des **lexiques** utilisés pour détecter le « concret » et l'« aspirationnel », (ii) des **seuils de prétraitement** (filtrage, stopwords), et (iii) des **volumes textuels** : malgré la normalisation, les documents courts peuvent produire des ratios instables. Enfin, le *Sentiment Score* ne doit pas être interprété comme une mesure psychométrique de « promesses vagues », mais comme un proxy linguistique du registre valorisant. Pour ces raisons, l'ICE doit être interprété comme un indicateur comparatif interne au corpus et idéalement complété par une validation qualitative (p. ex. concordances en contexte autour de termes-clés tels que *waste*, *transport*, *energy*).

9 Analyse de Réseaux (Link Analysis)

L'analyse textuelle a permis d'identifier les thématiques dominantes. Cependant, pour comprendre comment ces idées s'organisent et interagissent, il est nécessaire d'étudier la structure relationnelle du discours écologique. Cette section détaille les propriétés topologiques des réseaux formés entre les festivals et leurs engagements.

9.1 Extraction des liens et préparation des données

Pour construire notre réseau, nous avons importé les fichiers de nœuds (*nodes*) et d'arêtes (*edges*) générés lors de la phase de collecte dans le logiciel **Gephi**. Initialement, le volume de liens web est massif ; afin de rendre le graphe plus pertinent et visuel, nous avons appliqué des protocoles de filtrage rigoureux :

— **Filtrage par degré :** Exclusion des nœuds ayant moins de 10 connexions pour se concentrer sur les éléments significatifs.

— **Betweenness Centrality :** Sélection du top 20% des nœuds les plus interconnectés pour cartographier les piliers du discours.

Cette stratégie permet de focaliser l'analyse sur les connexions structurantes tout en préservant la lisibilité du réseau.

9.2 Graphe 1 : Interconnexion Festivals et Thématiques

Le premier réseau (Figure 5) illustre la manière dont les festivals gravitent autour des piliers environnementaux. La structure biparti révèle une modularité de 0,39, indiquant un réseau fortement interconnecté où les thématiques communiquent intensément entre elles.

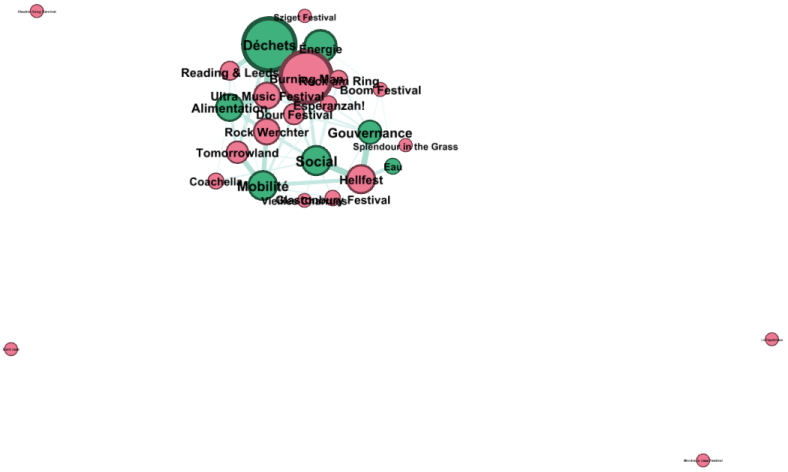

FIGURE 5 – Structure globale du réseau Festival ↔ Thématique.

9.3 Construction et caractéristiques du réseau Festival ↔ Thématique

Le graphe principal, de nature **bipartie**, modélise les co-présences discursives : un lien indique qu'un festival mobilise une catégorie environnementale identifiée (déchets, énergie, mobilité, etc.) dans sa communication officielle.

Caractéristique du graphe	Valeur
Nombre de nœuds	26
Nombre de liens	50
Degré moyen	3,846
Degré pondéré moyen	69,692
Diamètre du graphe	4

TABLE 2 – Caractéristiques principales du graphe Festival ↔ Thématique.

Le diamètre de 4 indique que deux nœuds quelconques restent reliés par un nombre limité d'étapes, témoignant d'une connectivité globale élevée au sein de l'espace discursif.

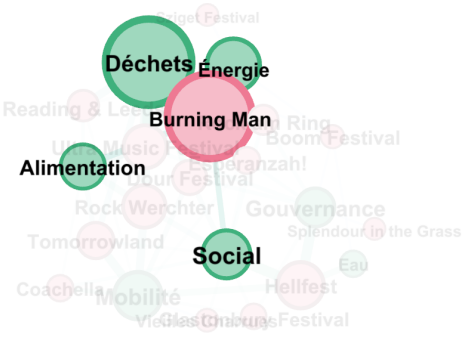

FIGURE 6 – Focus : Principaux enjeux liés au Burning Man.

9.4 Analyse des centralités et communautés

Nous analysons la centralité de degré pour repérer les thématiques les plus mobilisées, la betweenness centrality pour identifier les nœuds "ponts" reliant des sous-ensembles du réseau, et PageRank pour détecter les nœuds influents au-delà du simple degré. Enfin, la modularité (détection de communautés) met en évidence une organisation en sous-groupes, opposant des logiques de spécialisation thématique à des logiques de transversalité.

L'algorithme de **modularité** a été appliqué pour identifier des clusters. Les résultats suggèrent une dualité entre une logique de *spécialisation* (focus sur un seul axe comme les déchets) et une logique de *transversalité* (discours global et intégré).

9.5 Proximités discursives : Graphe Festival ↔ Festival

Un second réseau a été construit afin d'analyser les proximités discursives entre festivals. Ce graphe, non orienté, relie deux festivals lorsqu'ils partagent des profils thématiques similaires dans leur communication environnementale. Les caractéristiques globales de ce réseau sont les suivantes :

— **Connectivité :** Le réseau présente un degré moyen de 1,789 et un degré pondéré moyen de 1,529.
— **Étendue :** Le diamètre du graphe est de 7, indiquant une structure plus étirée que le réseau thématique.
— **Segmentation :** La modularité s'élève à 0,497, ce qui indique une segmentation claire entre des groupes de festivals partageant des styles discursifs proches.

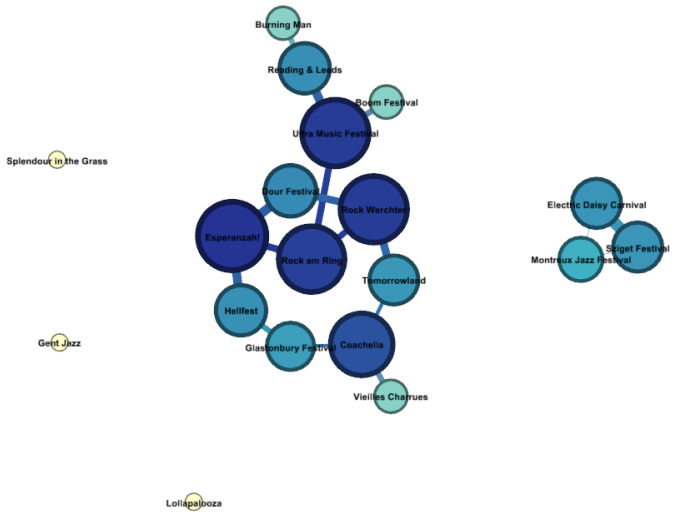

FIGURE 7 – Réseau de similarité Festival ↔ Festival.

Cette modularité relativement élevée traduit l'existence de noyaux fortement connectés face à d'autres festivals plus périphériques ou isolés, reflétant soit une communication singulière, soit une faible visibilité des enjeux écologiques.

9.6 Co-occurrences structurelles : Graphe Thématique ↔ Thématique

Un troisième graphe illustratif relie les thématiques apparaissant conjointement dans la communication des mêmes festivals. On observe des « paquets discursifs » récurrents :

— **Déchets et Énergie :** Association technique et logistique.
— **Social, Gouvernance et Eau :** Bloc lié à la responsabilité organisationnelle.
— **Mobilité et Alimentation :** Thématiques liées aux flux et à la consommation.

Ces associations structurelles montrent que les festivals articulent simultanément plusieurs dimensions de la durabilité pour légitimer leur discours global.

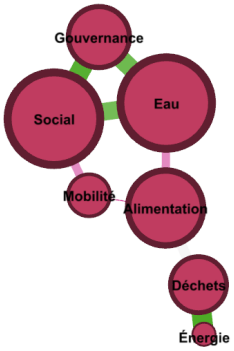

FIGURE 8 – Co-occurrences thématiques : réseau Thématique ↔ Thématique.

10 Discussion et Recommandations

En croisant les résultats de *Text Mining* et de *Link Analysis*, nous proposons une lecture critique de la manière dont les festivals rendent visibles (ou non) leurs engagements environnementaux sur leurs sites web.

10.1 Interprétation critique, limites et pistes d'amélioration

Les analyses mettent en évidence des logiques de communication contrastées. Une première famille de festivals mobilise un registre davantage **opérationnel**, structuré autour d'éléments concrets liés à l'organisation (déchets, énergie, mobilité, eau). Une seconde famille privilégie un registre plus **aspirationnel**, centré sur des valeurs et des intentions (*green, future, commitment*), sans détailler systématiquement des mécanismes vérifiables. La matrice « Actions vs Promesses » permet ainsi de situer des stratégies allant d'un *discours-preuve* à des formes de *valorisation symbolique* plus générales.

Il est toutefois essentiel de rappeler que ce projet mesure une **crédibilité discursive** c'est-à-dire la visibilité et la structuration d'un discours en ligne et non une performance environnementale réelle. Un festival peut mettre en œuvre des actions sans les documenter sur son site (faible visibilité), tandis qu'un autre peut valoriser des intentions sans fournir de détails opérationnels. Dans ce contexte, plusieurs limites structurantes doivent être prises en compte :

- **Biais d'architecture web :** certaines pages clés peuvent être sous-représentées lorsque les sites reposent sur des architectures dynamiques (*SPA / Next.js*), malgré l'injection manuelle réalisée pour quelques cas.
- **Biais de visibilité :** le volume textuel (et la densité de communication) influence mécaniquement la probabilité de détecter des marqueurs lexicaux, même lorsque des normalisations sont appliquées.
- **Limites des méthodes :** TF–IDF et K-means capturent parfois des régularités **éditoriales** (types de pages : infos pratiques vs chartes) avant de refléter des oppositions strictement conceptuelles sur la durabilité ; le choix de $k = 2$ reste néanmoins pertinent car il

isole une dualité stable et interprétable (logistique/expérience vs institutionnel/RSO). Pour une version ultérieure du pipeline, trois améliorations prioritaires peuvent renforcer la finesse thématique tout en réduisant l'effet « structure de site » : (i) l'élargissement du corpus (plus de pages pertinentes par festival), (ii) l'usage d'outils sémantiques multilingues (embeddings) afin de dépasser la correspondance exacte de tokens, et (iii) l'intégration d'une modélisation thématique (*BERTopic* ou *LDA*) permettant d'extraire des thèmes latents plus fins (p. ex. circularité, biodiversité, gouvernance, alimentation) et de compléter les analyses actuelles.

11 Conclusion

Ce projet a proposé une exploration du discours écologique de festivals internationaux à partir de leurs sites web officiels, en mobilisant un pipeline combinant *Web Mining*, *Text Mining* et analyse de réseaux. À partir de données textuelles et structurelles extraites automatiquement, l'objectif était d'analyser comment les enjeux environnementaux sont formulés, organisés et rendus visibles dans la communication numérique du secteur événementiel.

Les résultats mettent en évidence une forte hétérogénéité des stratégies observées : certains festivals structurent leur communication autour de marqueurs opérationnels explicites (déchets, mobilité, énergie), tandis que d'autres privilégient des registres plus institutionnels ou aspirationnels, davantage centrés sur des valeurs que sur des actions détaillées. Les outils mobilisés (matrice « Actions vs Promesses », similarité TF–IDF, clustering et réseaux thématiques) permettent ainsi de comparer des **styles de communication** et des logiques de visibilité, sans inférer une performance environnementale réelle.

Plus largement, ce travail montre que les méthodes de science des données appliquées au web permettent d'analyser non pas la durabilité intrinsèque des organisations, mais la manière dont celles-ci **construisent, hiérarchisent et légitiment** leurs engagements dans l'espace numérique. Une extension naturelle consisterait à confronter ces profils discursifs à des sources externes (labels, audits, indicateurs environnementaux) et à renforcer l'analyse sémantique multilingue, afin d'affiner l'interprétation et de réduire l'influence des structures éditoriales propres aux sites web.

12 Références

— Salton, G., & Buckley, C. (1988). *Term-weighting approaches in automatic text retrieval. Information Processing & Management*, 24(5), 513–523.
Référence fondatrice pour la pondération TF–IDF et la vectorisation des corpus textuels.

— Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.
Ouvrage de référence pour la similarité cosinus, la vectorisation de textes et les bases du Text Mining.

— MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations.* In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.*
Article fondateur de l'algorithme de clustering K-Means.

— Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking : Bringing order to the web.* Stanford InfoLab.
Travail fondateur sur les mesures de centralité et l'algorithme PageRank appliqué aux graphes web.

— Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi : An open source software for exploring and manipulating networks.* In *Proceedings of the International AAAI Conference on Weblogs and Social Media.*
Référence principale pour l'analyse et la visualisation de réseaux complexes avec Gephi.

— Vande Kerckhove, C., & Courtain, S. (2025–2026). *Web Mining – MLSMM2153.* Université catholique de Louvain.
Support de cours utilisé pour la construction du pipeline méthodologique, basé en partie sur les supports du Prof. P. Francq.

— OpenAI. *ChatGPT (version GPT-5.1).*
Disponible sur : https://chat.openai.com.
Utilisé exclusivement comme outil d'aide à la reformulation linguistique, sans génération de contenu analytique ou factuel original.