

**LA DIFFUSION DES MÉTHODES NLP DANS LA RECHERCHE EN MARKETING : UNE ANALYSE
SYSTÉMATIQUE**

CARON OLIVIER

Doctorant

Université Paris-Dauphine, Université PSL, CNRS, DRM, [M&R], 75016 PARIS, FRANCE

olivier.caron@dauphine.psl.eu

BENAVENT CHRISTOPHE

Professeur des universités

Université Paris-Dauphine, Université PSL, CNRS, DRM, [M&R], 75016 PARIS, FRANCE

christophe.benavent@dauphine.psl.eu

LA DIFFUSION DES MÉTHODES NLP DANS LA RECHERCHE EN MARKETING : UNE ANALYSE SYSTÉMATIQUE

Résumé : L'avènement des techniques d'analyse du langage naturel (NLP) révolutionne la recherche en marketing, marquée par une transition notable vers l'intelligence artificielle et les modèles de type "Transformers". Le NLP permet une analyse approfondie des données textuelles, transformant les chaînes de caractères en informations quantifiables pour décrypter le contenu sémantique, les styles expressifs et les intentions. Cette transformation est cruciale pour l'analyse des réseaux sociaux et des avis consommateurs, reflétant la corrélation entre la disponibilité des données et les avancées technologiques. Au-delà de l'aspect technique, l'adoption du NLP dans la recherche en marketing illustre un changement méthodologique majeur. Les chercheurs, en tant qu'agents sociaux, équilibrent les coûts d'apprentissage et de mise en œuvre avec les bénéfices tels que l'originalité et la légitimité des méthodes. Cet équilibre varie selon les domaines et l'expérience des chercheurs, entraînant des différences dans l'adoption et la diffusion du NLP. La formation de réseaux parmi les innovateurs et l'imitation stratégique par d'autres chercheurs définissent ainsi les contours d'une nouvelle ère dans la recherche marketing.

Mots clefs : diffusion ; NLP ; marketing ; méthodes ; revue systématique de littérature

THE DIFFUSION OF NLP METHODS IN MARKETING RESEARCH: A SYSTEMATIC ANALYSIS

Abstract: The advent of Natural Language Processing (NLP) techniques is revolutionizing marketing research, marked by a notable transition towards artificial intelligence and "Transformers" type models. NLP enables in-depth analysis of textual data, transforming character strings into quantifiable information to decipher semantic content, expressive styles and intentions. This transformation is crucial for the analysis of social networks and consumer reviews, reflecting the correlation between data availability and technological advances. Beyond the technical aspect, the adoption of NLP in marketing research illustrates a major methodological shift. Researchers, as social agents, balance the costs of learning and implementation with benefits such as the originality and legitimacy of methods. This balance varies from field to field and from researcher to researcher, leading to differences in the adoption and diffusion of NLP. The formation of networks among innovators and strategic imitation by other researchers thus define the contours of a new era in marketing research.

Keywords: diffusion; NLP; marketing; methods; systematic literature review

Introduction

L'intégration progressive des méthodes de traitement automatique du langage naturel (NLP) dans la recherche en marketing représente un changement méthodologique et pratique majeurs, répondant à la nécessité d'analyser des volumes croissants de données textuelles. Les interactions en ligne, les avis de consommateurs, et les plateformes de médias sociaux constituent en effet une masse d'informations précieuse. Dans ce contexte, l'utilisation des méthodes NLP devient nécessaire, permettant une nouvelle dimension d'analyse. En plus de gérer l'échelle et la complexité des données, elles offrent également des perspectives d'analyse inédites, enrichissant ainsi l'éventail méthodologique en complétant les approches qualitatives traditionnelles. Face à ces changements, cet article se propose d'explorer la diffusion des méthodes NLP dans la recherche en marketing grâce à une revue systématique de littérature.

Notre recherche se développe en s'appuyant sur les acquis des théories de l'adoption des innovations technologiques (TAM) que nous adaptons spécifiquement au champ académique.

L'évolution du système éditorial est caractérisée par une compétition renforcée et organisée par les classements de revues et des scores de citations. Cette dynamique modifie profondément les stratégies de publication des chercheurs, les incitant à adapter leurs travaux aux exigences de reconnaissance et d'impact au sein de la communauté académique.

Naviguant entre la pression à publier, la quête d'originalité et la maîtrise, les chercheurs doivent équilibrer leurs aspirations à l'innovation avec la nécessité d'assurer la légitimité de leurs méthodes et approches. Cette tension les amène à adopter des stratégies pour répondre aux attentes de la communauté académique tout en visant à se distinguer par la nouveauté de leurs contributions.

Parallèlement, l'adoption des méthodes NLP dans la recherche en marketing est stimulée par l'accès à de nouvelles bases de données et la capacité à aborder des nouvelles questions de recherche. Ces dynamiques s'entrelacent, reflétant comment la disponibilité de données inédites et l'applicabilité de nouvelles techniques façonnent les orientations de la recherche en marketing.

Cadre théorique

Depuis une vingtaine d'années, le cadre institutionnel est celui d'un régime de compétition où les classements de revues, les scores de citations des chercheurs, jouent un rôle central, et participe au moins en partie à la rémunération des chercheurs et à leur promotion professionnelle (Richard et al., 2015). Le passage d'un paradigme vocationnel de la recherche à un paradigme professionnel est concomitant à une transformation profonde du système éditorial qui s'est concentré autour de quelques grands bouquets tels que Elsevier ou Thomson, et a vu le nombre des revues croître considérablement ainsi que leur accès facilité par les interfaces de recherche bibliographiques (Lambiotte & Panzarasa, 2009 ; Evans et al., 2011). Au cours de ces dernières années, l'émergence des modèles en *open access* (MDPI, PLoS, Frontiers, par exemple) a accéléré le rythme de production grâce à un nouveau modèle d'affaires.

Ces transformations ont comme conséquence pour les chercheurs, au-delà du "*publish or perish*", de les conduire à raffiner leurs stratégies de publications (Kolesnikov et al., 2018). L'adoption étant pensée dans les termes du modèle TAM (Davis, 1989), les chercheurs en marketing ont tout intérêt à adopter les méthodes NLP s'ils en perçoivent clairement le bénéfice (a) et la facilité d'utilisation (b).

Le bénéfice se traduit par l'originalité et la pertinence : la méthode permet-elle de se distinguer de travaux analogues ? L'enjeu de l'originalité réside dans son coût, qui se manifeste sous deux formes principales : d'abord, en termes de légitimité, car une méthode innovante, encore peu testée, peut soulever des doutes chez les relecteurs peu familiers avec celle-ci. L'adoption d'une nouvelle méthode implique également un coût d'apprentissage, nécessitant souvent un investissement dans l'acquisition de nouvelles compétences ou connaissances. Ce coût d'apprentissage peut être influencé par l'adoption du NLP dans des communautés de pratique connexes (Hauser et al., 2006), telles que des publications dans des revues spécialisées, des participations à des conférences, des démonstrations pratiques, ou d'autres collaborations intellectuelles ne menant pas forcément à des publications (Evans et al., 2011).

Dans le cas qui nous intéresse, les bénéfices et les coûts pour les auteurs dépendent des comportements d'autres acteurs. En premier lieu, celui des revues qui acceptent ou refusent les propositions d'article en fonction de leur qualité et notamment de l'originalité et de la légitimité des techniques employées.

--- insérer figure 1 ---

Dans notre contexte, il est raisonnable d'y inclure les entreprises technologiques qui participent largement au développement et à la diffusion des outils dans ce domaine, ainsi que de multiples acteurs qui jouent un rôle de gourou au sens de Huczynski (1993). Cette diffusion des techniques NLP est bien sûr motivée par une myriade de possibilités

intrinsèques et nouvelles de traitement de données (Hartmann & Netzer, 2023). Comme (Peres et al., 2010) le soulignent, la diffusion, et ses modèles, doivent prendre en compte des forces sociales, et notamment des phénomènes de modes (Abrahamson, 1991) débordant l'impératif d'être à la pointe de la technologie, en permettant l'analyse de vastes volumes de données textuelles d'une manière qui, auparavant, se limitait principalement à des approches qualitatives.

Ainsi, l'adoption du NLP satisfait le désir d'individualité, d'innovation et d'originalité des chercheurs et les distingue ainsi de leurs collègues n'ayant pas adopté ces techniques. En contrepartie, elle maintient une certaine mesure de conformité en imitant les experts reconnus du domaine qui sont déjà parvenus à publier dans les revues les plus influentes en utilisant ce type de méthodes.

Méthodes

L'approche adoptée relève d'une forme de revue de littérature systématique (Stratton, 2019), au moins par la constitution du corpus. Elle vise cependant moins à répondre à une question spécifique, telle la relation entre deux concepts, qu'à dresser de manière systématique une carte de l'évolution de l'usage de méthodes d'analyse textuelles dans le champ du marketing et dans la communauté qui correspond.

En date du 18 août 2023, une recherche exhaustive sur la base de données Scopus a permis de récupérer un total de 369 articles. Ce corpus a été spécifiquement sélectionné pour inclure des articles publiés dans des revues dont les titres contiennent les termes "*marketing*" et "*consumer*". Les critères de recherche pour identifier ces articles comprenaient la présence des mots-clés suivants dans le titre, le résumé ou les mots-clés : "*nlp*", "*natural language processing*", "*text mining*", "*text-mining*", "*text analysis*", et "*text-analysis*". Un processus de filtrage manuel a ensuite été appliqué pour affiner davantage la sélection, en écartant les articles où les termes recherchés ne concernaient pas la recherche en marketing utilisant des méthodes d'analyse textuelle ("*NLP*" peut par exemple signifier "*neuro-linguistic programming*").

Pour comprendre de manière plus précise la nature de la diffusion, nous nous appuyons sur des analyses de réseaux d'auteur par période, en prenant comme point de départ l'article le plus ancien datant de 1996. Pour caractériser les collaborations, nous estimons que les auteurs d'un même article ont tous collaboré comme il est d'usage dans ce genre d'analyses (Barabási et al., 2002). Cette démarche vise à cartographier la structure sociale de la recherche dans ce domaine sur une période étendue, élément constituant de la diffusion des méthodes NLP. On cherchera tout particulièrement à identifier les communautés (au sens de l'algorithme de Louvain) et les auteurs isolés.

Dans un second temps nous étudions l'évolution des thèmes de recherche en utilisant une méthode appelée Structural Topic Modeling (STM) (Roberts et al., 2014). Cette approche permet de dégager une compréhension plus qualitative des différents domaines de recherche et de leur évolution en prenant en compte de manière explicite l'effet du temps sur la prévalence des topics.

Résultats

1 - Une mode académique ?

Cet effet de mode managériale est corroboré par une accélération récente et soudaine du nombre de publications, de conférences, de livres, etc. portant sur l'utilisation des méthodes NLP dans le champ du marketing comme le montre la figure 1 :

--- insérer figure 2 ---

Tandis que la production scientifique est relativement stable et faible dans un premier temps, elle croît considérablement et continuellement à partir de 2018 dont l'article de Humphreys et Wang fait un récapitulatif pertinent à la fois théorique et opérationnel (Humphreys & Wang, 2018). Auparavant, plusieurs articles pionniers puis largement cités ont été publiés (Ghose et al., 2012 ; Netzer et al., 2012 ; Ludwig et al., 2013) dans des revues majeures, envoyant un signal à l'ensemble des chercheurs en marketing et encourageant la diffusion des nouvelles méthodes NLP dans le champ du marketing. On remarque toutefois un délai notable entre ces publications et l'accélération du nombre de publications. Il peut être causé par un jeu stratégique des chercheurs dans cet environnement changeant. À l'instar des managers, les chercheurs en marketing font face au paradoxe décrit par Repenning et Sterman (2001) : l'accès aux nouvelles techniques, telles que le traitement automatique du langage naturel (NLP) s'est simplifié, cependant, leur intégration effective dans la pratique demeure un défi.

--- insérer figure 3 ---

2 - La structure des réseaux d'auteurs

L'analyse des réseaux de co-auteurs couvre une période de 1996 à 2023 que nous illustrons en se focalisant sur deux périodes distinctes.

--- insérer figure 4 ---

--- insérer figure 5 ---

Il est intéressant d'observer que le nombre d'auteurs a considérablement augmenté entre ces intervalles, passant de 116 auteurs pour la première période à 923 pour la seconde. Malgré cet accroissement, le nombre de collaborations n'a pas augmenté dans la même proportion, avec 123 collaborations enregistrées pour la période 1996-2015 et 1361 pour la période étendue jusqu'à 2023. Cela se traduit par une baisse de la densité du réseau (Melançon, 2006) qui est passée de 1.84% à 0.32%¹ respectivement pour les périodes 1996-2015 et 1996-2023. Les nouveaux venus dans le champ académique ne s'intègrent pas forcément aux structures majeures préexistantes, mais forment plutôt des communautés périphériques moins influentes

¹ Ce pourcentage est à comprendre comme le rapport entre le nombre de collaborations observées et le nombre maximal de collaborations possibles au sein du réseau. En d'autres termes, il représente la proportion des liens effectivement établis par rapport à l'ensemble des liens qui auraient pu théoriquement exister si tous les auteurs avaient collaboré les uns avec les autres.

et moins connectées (la dimension des nœuds sur les réseaux reflétant le nombre de citations recueillies par chaque chercheur).

Pour une compréhension plus fine, les concepts d'homophilie et de contrainte de focalisation peuvent être éclairants. L'homophilie dans la collaboration scientifique, où les chercheurs choisissent des partenaires de recherche similaires, peut contribuer au renforcement des macro-composants, tandis que les contraintes institutionnelles et géographiques peuvent favoriser la formation de communautés périphériques. En outre, la recherche a montré que la collaboration multi-auteurs est généralement plus citée, soulignant l'avantage des efforts collectifs par rapport aux efforts individuels (Wuchty et al., 2007) d'autant plus que ces collaborations ont lieu entre plusieurs universités dont au moins une est dite de premier plan (Jones et al., 2008). Cela est visible quand l'on s'intéresse à différentes métriques des réseaux.

--- insérer figure 6 ---

C'est également visible en analysant les principales communautés du réseau de 1996-2023. Les cinq plus grandes communautés (en nombre d'auteurs) regroupent 126 auteurs, soit 13.7% des auteurs du réseau et concentrent 46% du nombre de citations. Le diagramme de Pareto montre bien cette concentration des citations au sein d'une petite minorité de communautés, toutes manifestement expertes dans l'utilisation des méthodes NLP.

--- insérer figure 7 ---

3 - Une propagation concentrique

L'étude des thématiques menée par le modèle STM met en évidence un phénomène clair : le nombre de topics par période augmente de manière régulière sans que le nombre relatif d'articles par topic ne décroisse, procédant par accumulation plutôt que par substitution.

--- insérer figure 8 ---

--- insérer figure 9 ---

En calculant l'entropie par période, on s'aperçoit qu'en fin de période l'entropie est maximale (relativement au nombre de topic identifiés qui devrait croître dans le futur). La diffusion s'opère de manière concentrique à partir de sujets spécifiques.

--- insérer figure 10 ---

De manière plus qualitative, chaque poussée correspond à l'apparition dans le champ de la recherche de nouveaux sets de données associés avec de nouveaux problèmes. C'est le cas spectaculaire de la question du *e-wom* dès les années 2005 qui remet au goût du jour les phénomènes de bouche-à-oreille dans la communication marketing. Cette période correspond à l'émergence de nouveaux médias (les réseaux sociaux) et d'une nouvelle technique : l'analyse du sentiment. Un peu plus tard, le surgissement, avec les plateformes, de systèmes de notation et de collecte systématique d'avis de consommateurs, crée un nouveau champ d'études.

Discussion & Conclusion

L'usage du NLP n'est-il qu'une tendance passagère (*fads*), qu'une mode (*fashion*) ou une innovation destinée à s'installer durablement dans le champ ? Sans répondre de manière catégorique, on peut préciser certains processus à l'œuvre qui se révèlent dans les données.

Le premier est que le mouvement de diffusion est largement descendant, l'innovation s'enracine d'abord dans les publications de premier rang à la recherche d'originalité, et qui par conséquent la légitiment autant dans le public des chercheurs que dans les revues de second rang. On peut s'interroger sur la question des coûts d'apprentissage qui, dans le cadre précis du NLP, tendent à se réduire au cours de la période, par la constitution d'une *lingua franca*, la multiplication des tutoriels, scripts et autres moyens de disséminer les savoir-faire techniques. Ces derniers sont fréquemment ancrés au sein de communautés sociales internes aux organisations, comme le soulignent (Lave & Wenger, 1991). Pour intégrer efficacement de nouvelles pratiques, il est crucial pour les chercheurs d'accéder et de valoriser ces connaissances, souvent tacites et dispersées (Hauser et al., 2006). Ces communautés de pratique peuvent donc se retrouver au sein d'une même affiliation et diminuer les coûts d'apprentissage.

Le second élément à considérer est la localisation spécifique de l'adoption au sein de certaines aires thématiques du champ de recherche. Ce phénomène correspond à l'intersection entre l'émergence de nouvelles questions de recherche, l'accès à de nouvelles bases de données, et l'utilisation de techniques disponibles et appropriées (Lambiotte et Panzarasa, 2009). Bien que cet aspect soit de nature contingente, il joue un rôle déterminant dans le processus d'adoption et de diffusion des méthodes au sein de la communauté scientifique.

Notre analyse des données révèle que, bien que les techniques de “*topic analysis*” aient été introduites au début des années 2000, surtout avec la publication influente de Blei et al. (2003), leur usage régulier ne semble s'être généralisé qu'après 2015. De même, les embeddings, issus de l'article fondateur de Mikolov (Mikolov et al., 2013), ainsi que l'avènement des Transformers, initié par l'article de (Vaswani et al., 2017), n'ont à ce jour qu'une faible application observée dans notre corpus. La majorité de la littérature examinée révèle une prédominance de méthodes basées sur les dictionnaires pour des applications telles que l'analyse de sentiment et l'emploi du Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010). Toutefois, l'intégration des *Large Language Models* (LLM) dans la recherche en marketing est déjà en cours, rendue possible par un coût d'apprentissage réduit et les applications de messagerie, et promet d'excellentes performances en analyse du sentiment et en classification de texte (Krugmann & Hartmann, 2024).

Références

Abrahamson, E (1991) Managerial Fads and Fashions : The Diffusion and Rejection of Innovations. *The Academy of Management Review*, 16(3), 586.
<https://doi.org/10.2307/258919>

Blei, DM, Ng, AY, & Jordan, MI (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3, 993-1022.

Davis, FD (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319. <https://doi.org/10.2307/249008>

Evans, TS, Lambiotte, R, & Panzarasa, P (2011). Community structure and patterns of scientific collaboration in Business and Management. *Scientometrics*, 89(1), 381-396. <https://doi.org/10.1007/s11192-011-0439-1>

Ghose, A, Ipeirotis, PG, & Li, B (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3), 493-520. <https://doi.org/10.1287/mksc.1110.0700>

Hartmann, J, & Netzer, O (2023). Natural Language Processing in Marketing. In K. Sudhir & O. Toubia (Éds.), *Review of Marketing Research* (p. 191-215). Emerald Publishing Limited. <https://doi.org/10.1108/S1548-643520230000020011>

Hauser, J, Tellis, GJ, & Griffin, A(2006). Research on Innovation : A Review and Agenda for « Marketing Science ». *Marketing Science*, 25(6.), 687-717.

Huczynski, A. A. (1993). *Explaining the succession of management fads. The International Journal of Human Resource Management*, 4(2), 443-463. <https://doi.org/10.1080/095851993000000023>

Humphreys, A, & Wang, RJ-H. (2018). Automated Text Analysis for Consumer Research. *Journal of Consumer Research*, 44(6), 1274-1306. <https://doi.org/10.1093/jcr/ucx104>

Kolesnikov, S, Fukumoto, E, & Bozeman, B (2018). Researchers' risk-smoothing publication strategies : Is productivity the enemy of impact? *Scientometrics*, 116(3), 1995-2017. <https://doi.org/10.1007/s11192-018-2793-8>

Krugmann, J. O., & Hartmann, J. (2024). Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, 11(1), 1-19.

Lambiotte, R, & Panzarasa, P (2009). Communities, knowledge creation, and information diffusion. *Journal of Informetrics*, 3(3), 180-190. <https://doi.org/10.1016/j.joi.2009.03.007>

Lave, J, & Wenger, E (1991). *Situated Learning : Legitimate Peripheral Participation* (1^{re} éd.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>

Ludwig, S, de Ruyter, K, Friedman, M, Brüggén, EC, Wetzels, M., & Pfann, G (2013). More than Words : The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, 77(1), 87-103. <https://doi.org/10.1509/jm.11.0560>

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. 9.
- Netzer, O, Feldman, R, Goldenberg, J, & Fresko, M (2012). Mine Your Own Business : Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3), 521-543. <https://doi.org/10.1287/mksc.1120.0713>
- Peres, R., Muller, E., & Mahajan, V. (2010) Innovation diffusion and new product growth models : A critical review and research directions. *International Journal of Research in Marketing*, 27(2), 91-106. <https://doi.org/10.1016/j.ijresmar.2009.12.012>
- Repenning, NP, & Sterman, JD (2001). Nobody Ever Gets Credit for Fixing Problems that Never Happened: 43(4).
- Richard, J, Plimmer, G, Fam, K.-S., & Campbell, C (2015). Publishing success of marketing academics : Antecedents and outcomes. *European Journal of Marketing*, 49(1/2), 123-145. <https://doi.org/10.1108/EJM-06-2013-0311>
- Roberts, ME, Stewart, BM, Tingley, D, Lucas, C, Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064-1082. <https://doi.org/10.1111/ajps.12103>
- Stratton, SJ (2019). Literature Reviews : Methods and Applications. *Prehospital and Disaster Medicine*, 34(4), 347-349. <https://doi.org/10.1017/S1049023X19004588>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words : LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24-54. <https://doi.org/10.1177/0261927X09351676>
- Vaswani, A, Shazeer, N., Parmar, N, Uszkoreit, J, Jones, L, Gomez, AN, Kaiser, L, & Polosukhin, I (2017). Attention Is All You Need. <https://doi.org/10.48550/ARXIV.1706.03762>
- Wuchty, S, Jones, BF, & Uzzi, B (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827), 1036-1039. <https://doi.org/10.1126/science.1136099>

Figures et tableaux :

Figure 1 : Cadre conceptuel de l'adoption des techniques NLP dans la recherche en marketing

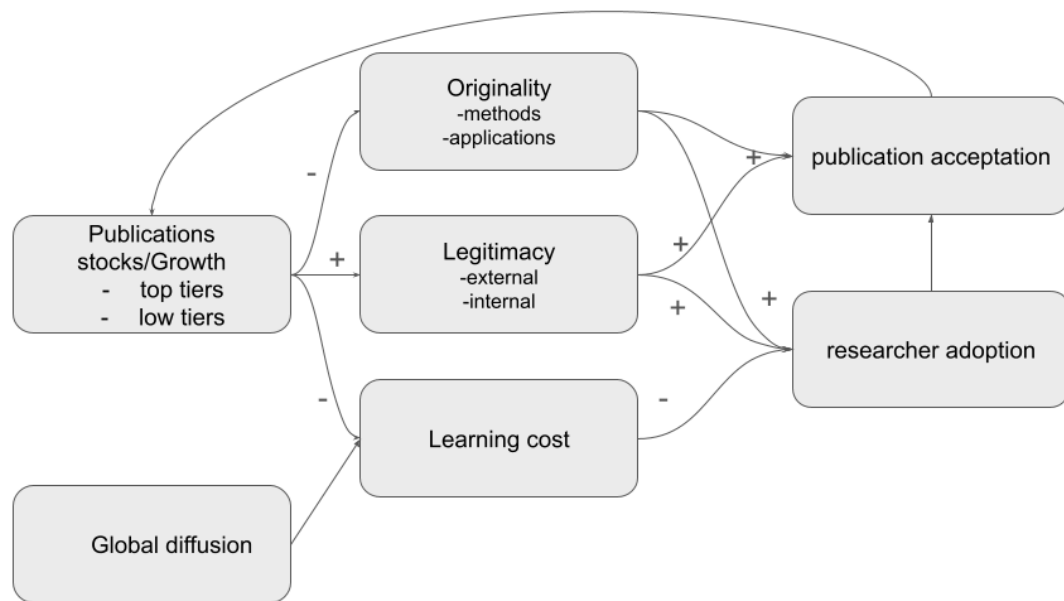
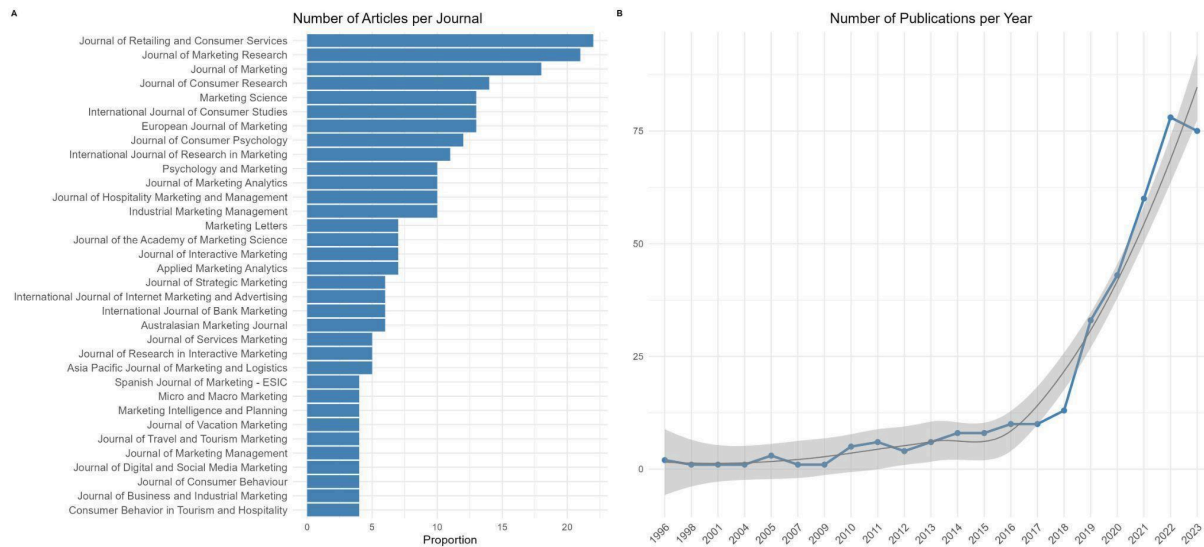
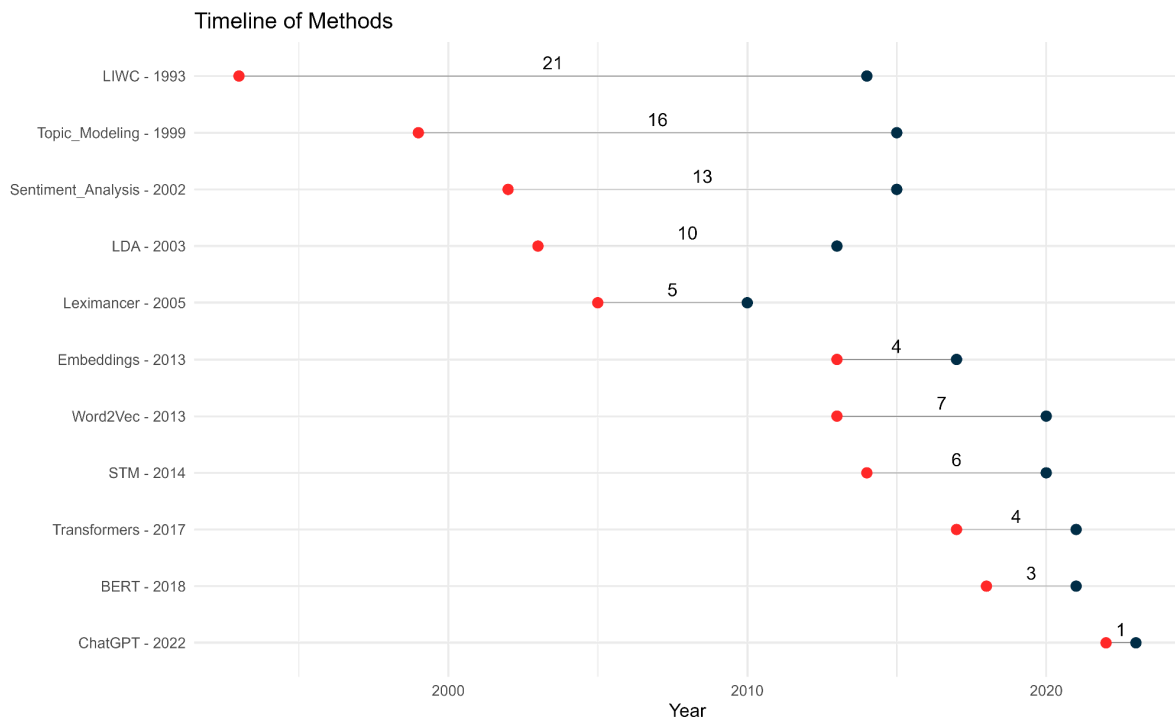


Figure 2 : L'essor des méthodes NLP dans la recherche en marketing : répartition des articles par revue et évolution annuelle des publications.



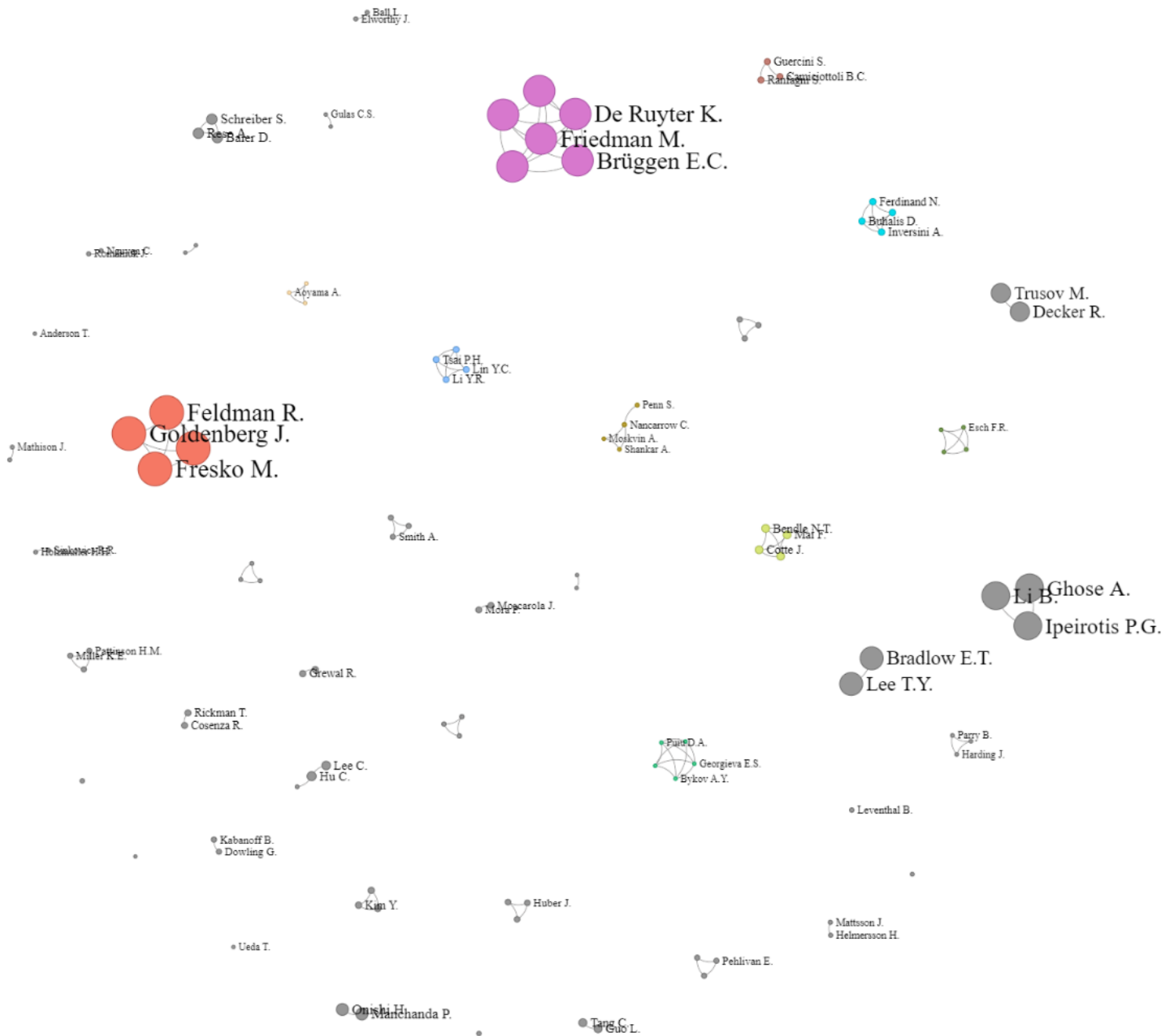
Cette visualisation graphique représente les résultats d'une recherche systématique menée sur la base de données Scopus le 18/08/2023. Les termes spécifiques utilisés pour la recherche dans le titre, l'abstract et les mots-clés sont "nlp", "natural language processing", "text-mining", "text mining", "text analysis" et "text-analysis". Les revues ciblées pour cette analyse sont celles contenant les mots 'marketing' ou 'consumer' dans leur titre. La partie A du graphique montre la répartition des articles par revue spécialisée. La partie B illustre l'évolution annuelle des publications, révélant une augmentation notable de l'emploi des méthodes NLP dans le domaine du marketing.

Figure 3 : Délais d'adoption des méthodes NLP dans la recherche en marketing



La visualisation graphique présente le délai en années entre la publication de l'article scientifique majeur présentant la méthode et son apparition dans notre corpus. Cette liste n'est pas exhaustive, mais elle présente les méthodes les plus utilisées avec quatre méthodes globales que sont Topic_Modeling, Sentiment_Analysis, Embeddings et Transformers.

Figure 4 : Le réseau des co-auteurs jusqu'en 2015



Ce graphe non dirigé des co-auteurs comprend l'ensemble des collaborations entre auteurs jusqu'en 2015. La taille des nœuds dépend du nombre de citations. L'épaisseur des arêtes dépend du nombre de collaborations entre deux auteurs. Le nombre de citations n'est pas disponible par année et est donc à considérer depuis l'année de parution de l'article jusqu'en août 2023. Les couleurs des nœuds sont attribuées aux dix communautés les plus nombreuses détectées grâce à l'algorithme de Louvain. Enfin, la disposition du graphe est gérée par l'algorithme de force ForceAtlas2, bien adapté aux dimensions des graphes.

Figure 5 : Le réseau des co-auteurs jusqu'en 2023

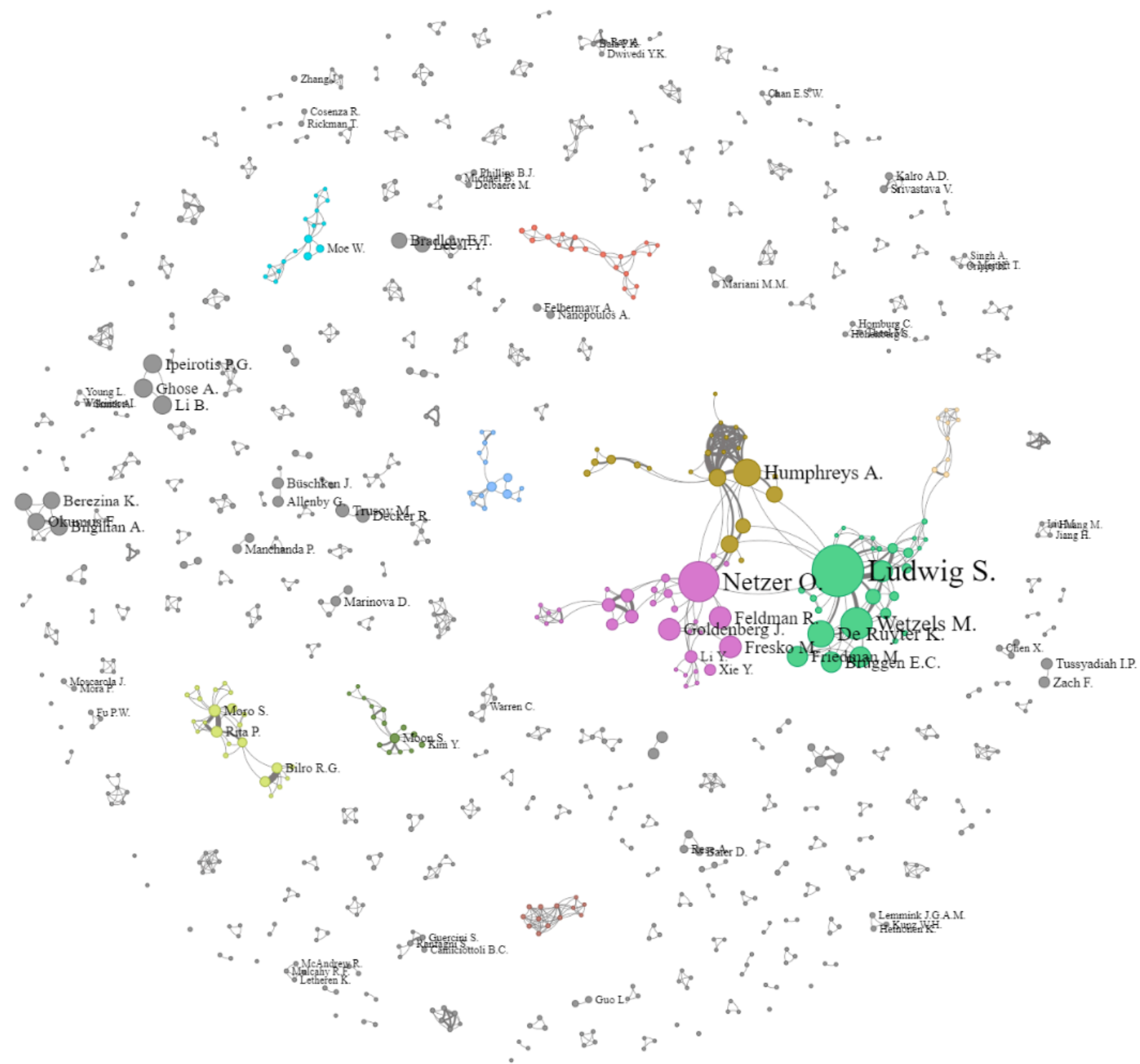


Figure 6 : Statistiques de structure des réseaux :

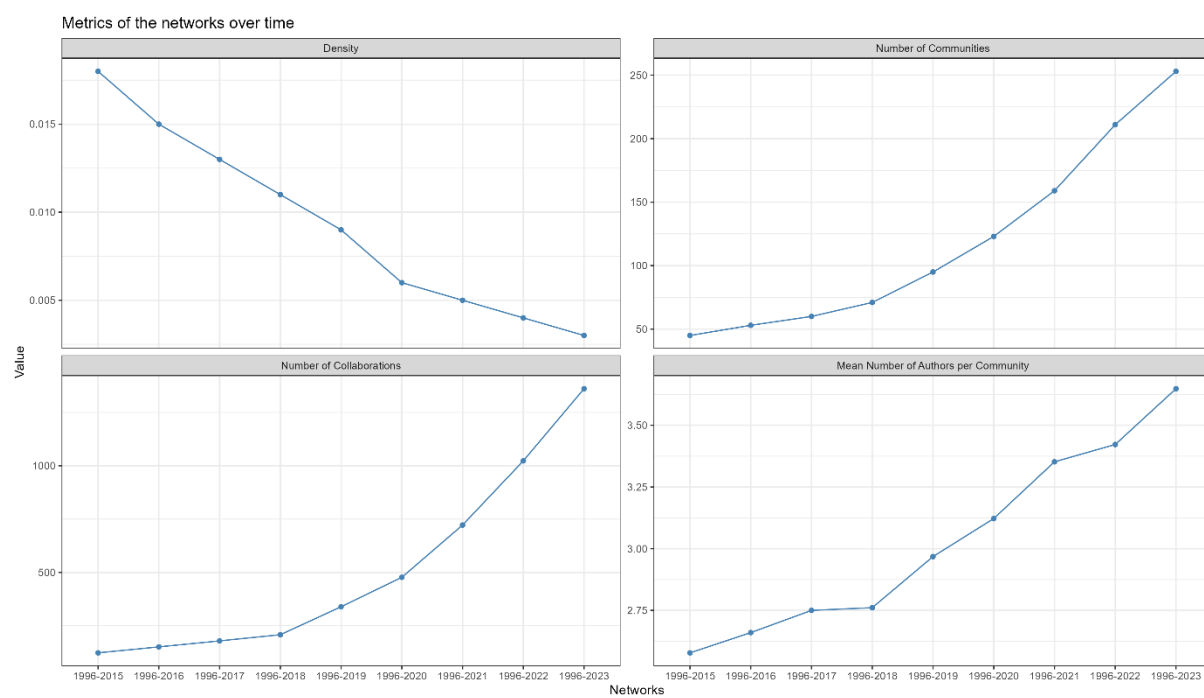


Figure 7 : Diagramme de Pareto des citations par communauté.

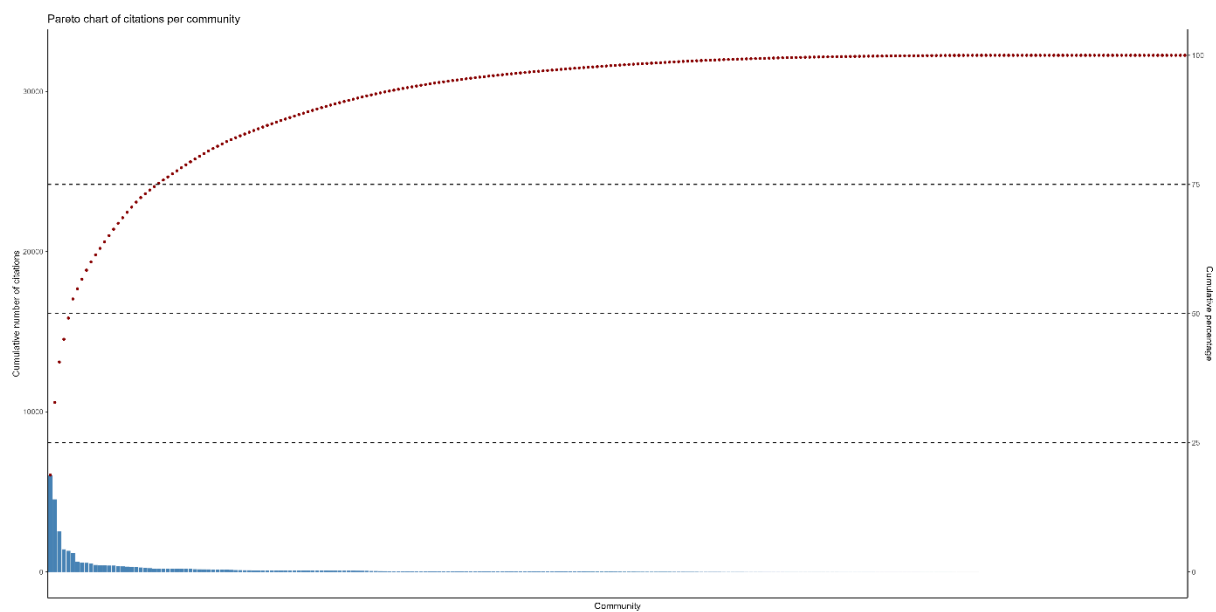
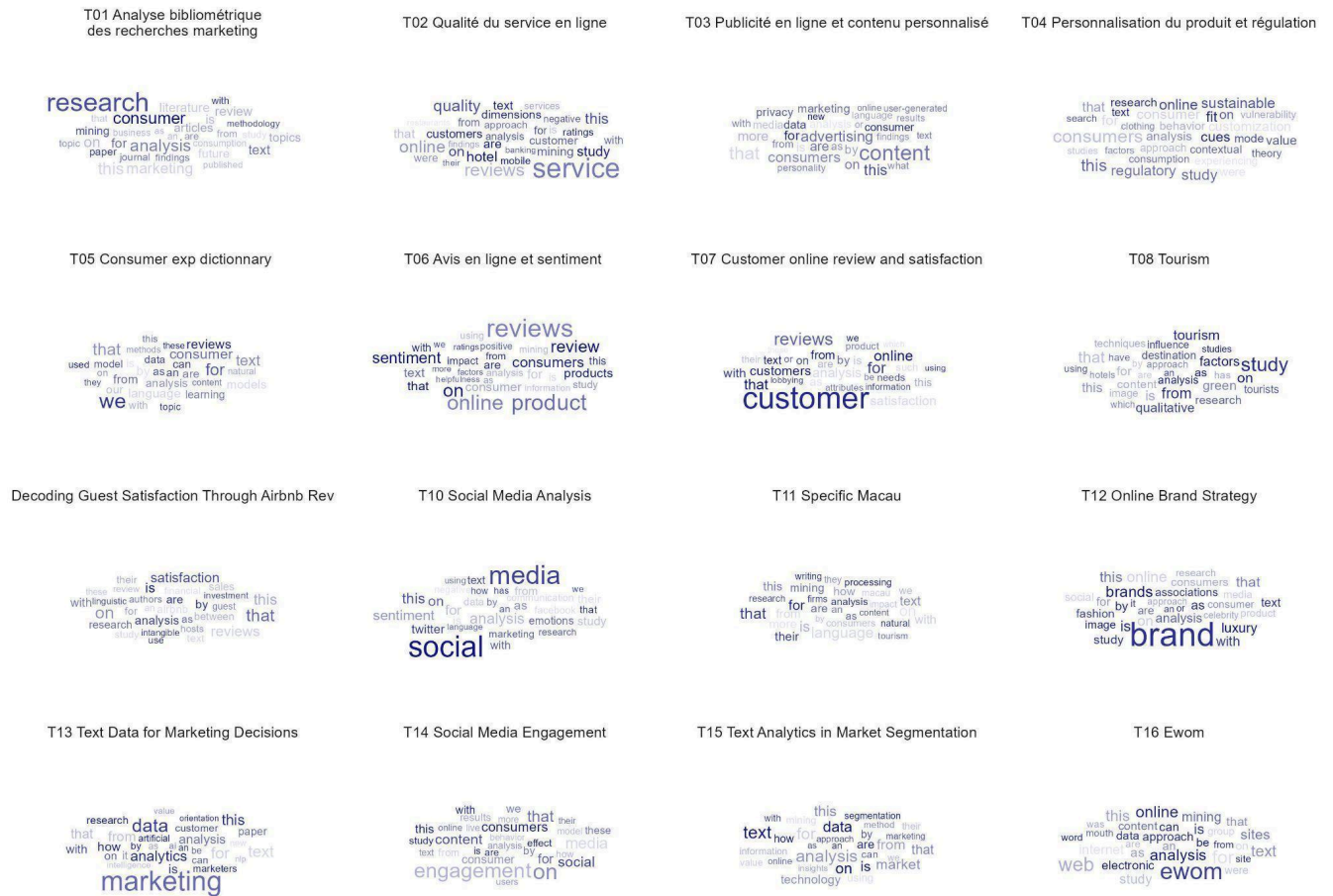
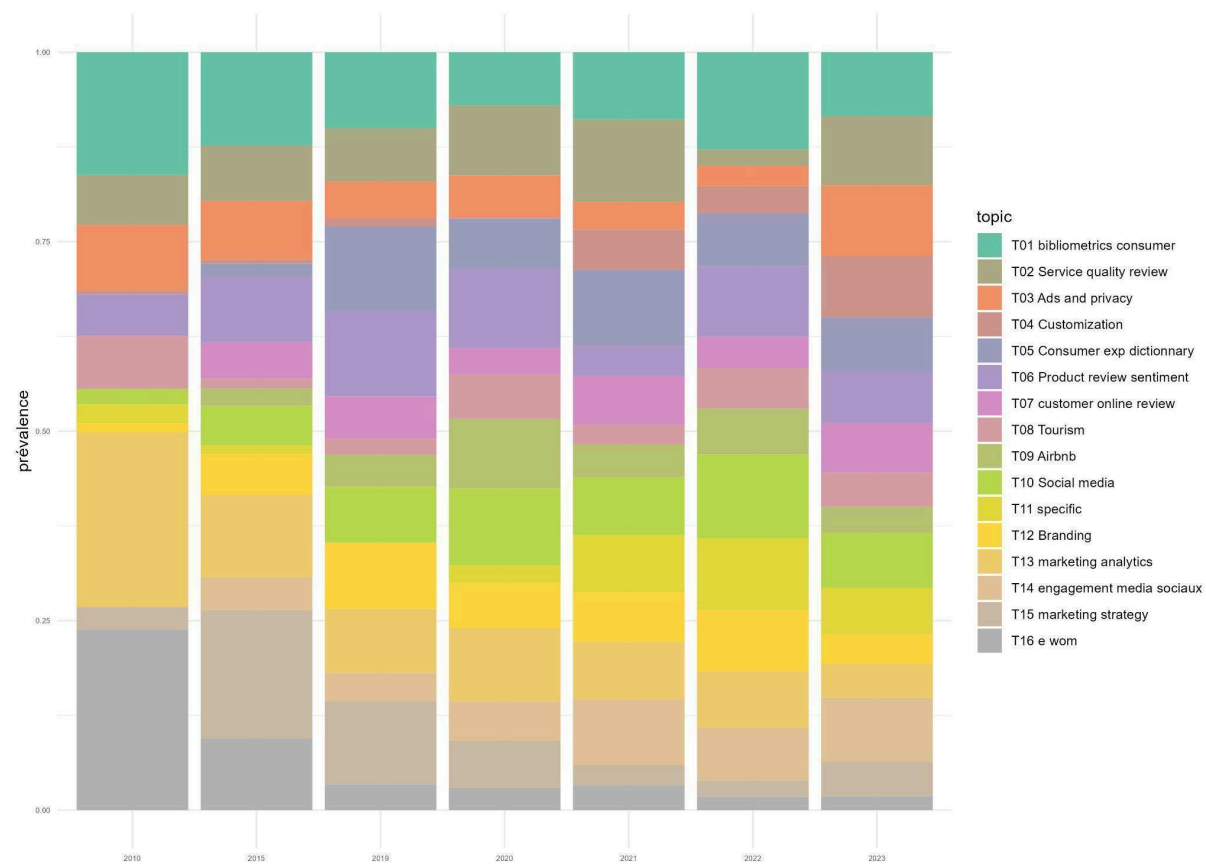


Figure 8 : Les différents thèmes (topics) détectés par l’algorithme STM



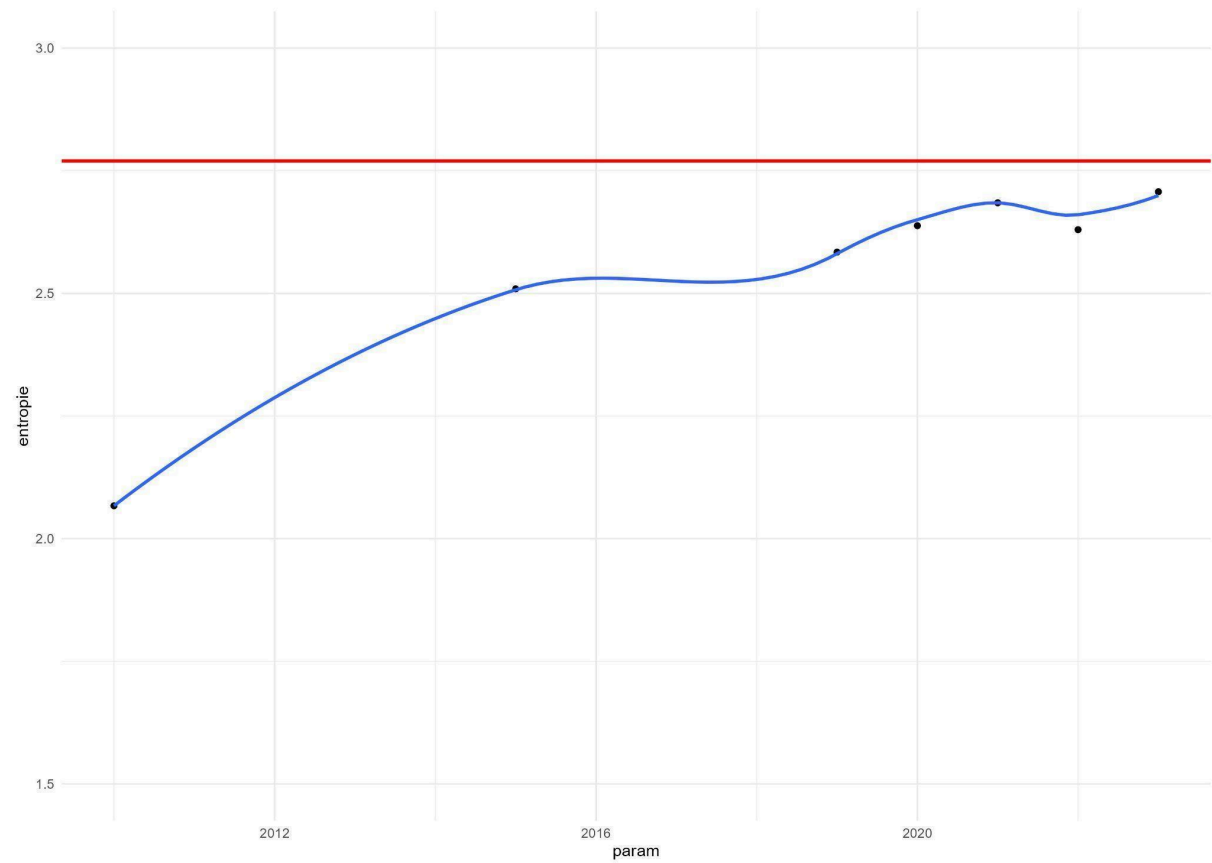
Description des topics sur la base de la probabilité des termes d’appartenir à un des topics (matrice β) indiqué par la taille des labels et par leur exclusivité (frex) indiquée par le degré de transparence du label.

Figure 9 : Répartition des topics par période de temps (algorithme STM)



Elle est calculée en prenant la moyenne de la prévalence des topics au travers des articles de la période. Compte tenu du faible nombre d'articles publié, les périodes 2010 concernent toutes les publications avant 2010 inclus, 2015, de 2011 à 2015 inclus, 2019, de 2016 à 2019 inclus, les périodes suivantes sont les années, 2023 est réduit aux neuf premiers mois de l'année.

Figure 10 : Évolution de l'entropie



L'entropie mesure la dispersion de la distribution de probabilité. Son maximum dans le cas discret correspond à l'équiprobabilité de la distribution des thèmes $E = - \sum_{i=1}^k p_i \log(p_i)$