UNIVERSITÀ DEGLI STUDI DI CASSINO E DEL LAZIO MERIDIONALE
CORSO DI DOTTORATO IN METODI, MODELLI E TECNOLOGIE PER L'INGEGNERIA
DIPARTIMENTO DI INGEGNERIA CIVILE E MECCANICA



Driving Open Innovation with Twitter: Strategies for Engaging Communities and Maximizing Impact

Antonio Giovanni Yury Di Russo
antoniogiovanniyury.dirusso1@unicas.it

In Partial Fulfillment of the Requirements for the Degree of
PHILOSOPHIAE DOCTOR in
Methods, Models and Technologies for Engineering

18/12/2024

TUTOR
Prof. Marco Greco
Ing. Amedeo Pata

COORDINATOR
Prof. Fabrizio Marignetti

XXXVII Cycle

“It is not the strongest of the species that survives,

nor the most intelligent that survives.

It is the one that is most adaptable to change.”

— Charles Darwin

Ringraziamenti

Questo traguardo non sarebbe stato possibile senza il sostegno, la guida e l'affetto di molte persone speciali che hanno illuminato il mio percorso di dottorato e, più in generale, il mio cammino di crescita personale e professionale. A loro va la mia più profonda gratitudine.

Un ringraziamento particolare va a *Leonardo S.p.A.*, nella persona di Amedeo Pata, per aver generosamente finanziato la borsa di studio che ha reso possibile questo percorso di ricerca. Il vostro investimento nella ricerca e nell'innovazione è un esempio tangibile di come l'industria possa contribuire al progresso scientifico e alla formazione delle future generazioni di ricercatori.

Al professor Marco Greco, mio tutor, vanno i miei più sinceri ringraziamenti. Le sue revisioni, puntuali e rigorose, che modificando e analizzando i miei paper mi ha spinto a migliorarmi costantemente e a raggiungere un livello di accuratezza e precisione che non credevo possibile. Anche se ci sono voluti fiumi di inchiostro virtuale e un numero incalcolabile di incontri, spesso accesi, il suo contributo è stato fondamentale per la qualità di questo lavoro. La sua dedizione, a titolo totalmente gratuito, è stata per me un esempio di integrità accademica e di passione per la ricerca. Nessuno più di Lei ha creduto in questo percorso.

Un grazie di cuore va ai miei colleghi, compagni di avventura in questo percorso a volte tortuoso. In particolare, Silvia e Valentina, con la vostra ironia e il vostro sostegno, avete reso più lievi anche i momenti più difficili.

Francesca. Non so davvero come avrei fatto senza di te. Sei stata la mia colonna portante, la mia instancabile compagna di sventure, colei che mi ha letteralmente trascinato per i capelli fino alla fine, soprattutto quando la motivazione vacillava e la stanchezza prendeva il sopravvento. La tua tenacia, la tua competenza e la tua amicizia sono state per me un dono inestimabile. Grazie. Grazie per il tuo click più veloce del West, grazie per essere stata così vicina a Miriam quando le mie parole non bastavano, grazie per esserci e grazie per essere.

Un pensiero speciale va a Miriam, amica di una vita e persona insostituibile. Grazie per esserci sempre stata, per aver ascoltato pazientemente i miei sfoghi e le mie preoccupazioni, per avermi offerto un rifugio sicuro nei momenti di sconforto e per aver sostenuto moralmente, con un affetto smisurato, il nostro Caro amico, Carne. Basta un cenno della testa o uno sguardo in una specifica direzione, che ti trasformi in un mostro da organizzazione, una segretaria, un consigliere, ma anche una confidente, una roccia (con tutta la spada conficcata) e nelle notti di luna piena, se ti gira, anche in un cavaliere. Grazie di abbracciarti ogni croce con lo stesso sorriso con cui si abbraccia un vecchio amico.

Francesca (mia sorella), di te non mi dimentico. Ti ringrazio per tutto quello che hai fatto per me, direttamente ed indirettamente. Non avrei conosciuto quest'università se tu non avessi spianato la strada, non avrei saputo come affrontare alcuni scogli senza la tua esperienza e non avrei saputo come scrivere questa tesi senza le nostre chiacchierate.

Ringrazio anche i nonni e i parenti, che hanno seguito questo percorso da lontano, ma sempre vicini per festeggiare le vittorie e patire le sconfitte.

Infine, il mio ringraziamento più grande e sentito va ai miei genitori e alla mia famiglia. Mia madre e mio padre, vi fidate di me ad occhi (soc)chiusi ogni giorno, dicendomi la vostra opinione ad ogni mia scelta, incoraggiandomi a inseguire un futuro. Siete stati il mio faro nei momenti di tempesta, la mia fonte inesauribile di forza e di amore. Avete sempre pensato, pure troppo, ma avete pensato sempre e solo al bene mio e di Mara. Mara, sorella mia, complice vicina e lontana, grazie per esserci sempre, per le risate, per l'espressione *sempre sorridente*, per i litigi e l'affetto incondizionato. Questo traguardo è dedicato a voi, alla vostra infinita pazienza e al vostro amore che mi ha permesso di arrivare fin qui. Senza di voi, nulla di tutto questo sarebbe stato possibile.

Grazie a tutti, di cuore.

Abstract

Open innovation (OI) is crucial for organizations in today's interconnected and rapidly changing technological landscape. Companies must leverage external knowledge and collaborations to stay competitive, and social media platforms have emerged as powerful tools for facilitating OI practices. However, despite the recognized importance of both OI and social media, there remains a significant gap in our understanding of how these two forces intersect, particularly on Twitter, a platform characterized by its real-time communication, global reach, and diverse user base. This dissertation addresses this gap by investigating the dynamics of OI communication on Twitter, employing a mixed-methods approach to analyse the sentiment, language, thematic structures, and network dynamics that shape discussions surrounding OI.

Traditional closed innovation models, with their reliance on internal R&D, have proven increasingly inadequate in today's fast-paced environment. OI, conceptualized by Chesbrough (2003), recognizes the value of external knowledge and collaboration, encouraging firms to actively seek outside partnerships and ideas. This shift requires effective communication strategies to facilitate knowledge exchange, build relationships, and foster a culture of collaboration. This study examines how these strategies are employed on Twitter, a platform offering unique opportunities and challenges for OI communication.

This dissertation employs a mixed-methods approach, combining quantitative analysis of a large Twitter dataset with qualitative case study insights and predictive modelling. This multifaceted approach provides a more comprehensive understanding of the OI discourse on Twitter, addressing a crucial gap in the literature by empirically investigating how OI is communicated and practiced on this platform.

The findings reveal several key insights. Sentiment analysis indicates a generally positive tone associated with OI communication, suggesting favourable

perceptions of this approach among Twitter users. Linguistic analysis uncovers distinct patterns in language use, including specialized vocabulary and an emphasis on collaboration. Topic modelling identifies eight key themes, ranging from funding and investment to emerging technologies and sustainability. Network analysis reveals a fragmented ecosystem, with clusters and influential actors shaping the conversation. Furthermore, a novel language-based cluster analysis identifies distinct communities characterized by unique communication styles, providing a more nuanced perspective than traditional hashtag analysis. A feed-forward neural network model successfully predicts tweet view count based on user characteristics, content, and timing, offering practical implications for optimizing communication strategies.

This research contributes to both theory and practice. It deepens our understanding of OI communication dynamics on Twitter, informing the literature on OI, social media, and organizational communication. The findings offer practical guidance for organizations seeking to leverage Twitter for OI, enabling them to tailor their communication to specific audiences, build relationships with key actors, and foster a culture of collaboration. The predictive model offers a valuable tool for maximizing message reach and impact.

The study acknowledges limitations, primarily its focus on a specific timeframe and platform. Future research could extend the analysis to other platforms, like LinkedIn, or conduct longitudinal studies to capture evolving trends. Further exploration of advanced machine learning techniques and integration of qualitative methods could provide even richer insights. This research serves as a first foundation for future investigations into the dynamic intersection of social media and open innovation.

Table of contents

Abstract	**Errore. Il segnalibro non è definito.**
Table of contents	iii
List of figures	iv
List of tables	v
Introduction	1
Chapter 1 The OI Paradigm	13
Chapter 2 Social Media as a Catalyst for Open Innovation	30
Chapter 3 Methods	44
Chapter 4 Sentiment Analysis and OI Keywords	62
Chapter 5 Thematic Exploration, Network Dynamics, and Language-Based Clustering	76
Chapter 6 Case Studies: OI Strategies on Twitter	97
Chapter 7 Impact and Engagement Measurement	112
Chapter 8 Predicting View Count: A Deep Learning Approach to Optimizing OI Communication	124
Chapter 9 Conclusions	134
References	141

List of figures

Figure 1: Time series of #openinnovation tweets 59

Figure 2: Time series of #innovation tweets 59

Figure 3: Sentiment Distribution for Dataset 2 64

Figure 4: Comparison of sentiment distrubutions in Dataset 2 65

Figure 5: Daily sentiment trend for dataset 2 66

Figure 6: Box plot of sentiment on a weekly basis 66

Figure 7: Mean relative frequency of OI and VC keywords 69

Figure 8: LDA visualization 79

Figure 9: Topic Distribution Comparison 80

Figure 10: Network Visualizations 85

Figure 11: PCA Visualization of Tweet Embeddings 89

Figure 12: A visual representation of the neural network shape 126

Figure 13: Training Loss history 129

List of tables

Table 1: T-test Results for Engagement Metrics 60

Table 2: Descriptive statistics for the compound sentiment score 64

Table 3: Top 10 OI and Value Capture Keywords in Dataset 2 68

Table 4: Topics Identified Through LDA 78

Table 5: Exemplary Tweets for Each Identified Topic 81

Table 6: Language-Based Clusters in the OI Discourse on Twitter 91

Table 7: Comparative Overview of OI Strategies on Twitter 108

Table 8: T-test Results for Impact and Engagement Metrics 114

Table 9:Permutation Importance of Input Variables 129

Introduction

For much of the 20th century, the dominant model of innovation was a closed, internally driven process. Companies relied primarily on their own research and development (R&D) departments, jealously guarding their intellectual property (IP) and striving for self-sufficiency. This "closed innovation" paradigm, as described by Chesbrough (2003), assumed that successful innovation stemmed from controlling the entire process, from ideation to commercialization, within the firm's boundaries. However, as the pace of technological advancements accelerated and knowledge became more widely distributed, the limitations of this closed approach became increasingly apparent (Chesbrough, 2003).

The Open Innovation Paradigm

In response to these limitations, a new paradigm for innovation emerged: open innovation. Pioneered by Henry Chesbrough (2003), open innovation (OI) recognizes the value of external knowledge and collaboration in driving innovation. It encourages companies to look beyond their internal walls, actively seeking out ideas, technologies, and partnerships from external sources. This approach allows firms to tap into a wider pool of knowledge, accelerate their innovation cycles, reduce R&D costs, and develop more customer-centric products and services (Chesbrough & Appleyard, 2007).

OI, however, is not a monolithic concept but rather a spectrum of approaches. Companies can engage in various forms of open innovation, including:

- **Inbound OI:** Actively acquiring and integrating external knowledge into the firm's innovation processes, such as through technology scouting, licensing agreements, or joint ventures (van de Vrande et al., 2009).
- **Outbound OI:** Leveraging the firm's internal knowledge externally, through mechanisms like licensing, spin-offs, or open-source

contributions, to generate new revenue streams or create new markets (Lichtenthaler, 2010).

- **Coupled OI:** Combining inbound and outbound practices, creating a synergistic flow of knowledge between the firm and external partners, often through joint development projects or strategic alliances (Mention, 2011).

OI in Practice

Numerous companies across various industries have successfully implemented OI strategies, demonstrating its transformative potential. Procter & Gamble, for instance, embraced OI through its "Connect + Develop" program, leveraging external partnerships to accelerate product development and access new technologies (Huston & Sakkab, 2006). Similarly, IBM has been a long-time advocate for OI, utilizing open-source software, collaborative platforms, and strategic partnerships to drive innovation in areas like artificial intelligence and cloud computing (Mention et al., 2019). These diverse examples highlight the flexibility and adaptability of OI, demonstrating its applicability across various industries and organizational contexts.

Social Media and the Transformation of Innovation

The emergence of social media has coincided with the rise of OI, creating a powerful synergy that is reshaping the innovation landscape. Social media platforms, with their unparalleled capacity to connect individuals and organizations across geographical boundaries, have become vital tools for facilitating OI practices.

Social Media as a Catalyst for OI

Social media platforms have fundamentally democratized the innovation process, empowering individuals and communities to actively participate in shaping the future of products and services (Kohler, 2015). They provide companies with unprecedented access to external knowledge, allowing them to tap into the collective intelligence of a global network of users. This access enables them to identify emerging trends, gather customer feedback, scout for promising technologies, and co-create solutions in a more collaborative and responsive manner (Mention et al., 2019; Saura et al., 2023).

Social media has impacted all stages of the innovation process, from the initial ideation phase to product development, marketing, and even post-launch feedback and improvement. By leveraging social media, companies can break down traditional silos, foster cross-functional collaboration, and accelerate the diffusion of innovations, reaching broader audiences and driving faster adoption (Jussila et al., 2012; Kaplan & Haenlein, 2010).

Twitter as a Real-Time Communication Platform

Among the diverse landscape of social media platforms, Twitter stands out as a particularly potent instrument for supporting OI. Its unique characteristics, namely its openness, speed, and reach, make it an ideal platform for monitoring real-time conversations, identifying emerging trends, and engaging with a diverse community of users.

Twitter's open and public nature fosters transparency and accessibility, allowing companies to observe and participate in discussions about innovation, gaining insights into user needs, industry trends, and competitor activities (Mention et al., 2019). Its real-time communication capabilities enable rapid dissemination of information and facilitate immediate feedback, accelerating the pace of innovation and enabling companies to respond quickly to changing market dynamics (Kaplan & Haenlein, 2010). Furthermore, Twitter's global reach provides access to a vast and diverse network of users, including individuals, companies, researchers, and investors, creating a dynamic ecosystem for OI.

The Evolving Role of Social Media in OI

The use of social media for OI has evolved significantly in recent years. Initial approaches often focused on simply broadcasting information about new products or initiatives, treating social media as a one-way communication channel. However, companies have increasingly recognized the value of engaging in two-way dialogue, actively listening to user feedback, and co-creating solutions with their online communities (Mention et al., 2019). This shift towards more collaborative and participatory approaches reflects a deeper understanding of the transformative potential of social media for OI.

Several companies have demonstrated innovative uses of social media for OI. For instance, LEGO Ideas, a platform where users submit designs for new LEGO sets, exemplifies the power of crowdsourcing through social media (Kohler, 2015). By tapping into the creativity of its user community, LEGO identifies innovative

ideas, validates their market potential, and engages customers in the product development process. Similarly, IBM actively uses Twitter to connect with experts, foster knowledge exchange, and gather insights from a global network of professionals (Mention et al., 2019). Through live Q&A sessions, discussions on emerging technologies, and crowdsourcing initiatives, IBM demonstrates its commitment to leveraging social media for OI.

Communication Patterns and Strategies

Language plays a pivotal role in shaping perceptions, framing debates, and influencing the adoption of new practices (West & Bogers, 2014). In the context of OI, understanding how this concept is communicated online is crucial for promoting its understanding, adoption, and ultimately, its success.

The Importance of Language

Language is not merely a neutral tool for conveying information; it actively constructs meaning, shapes perceptions, and influences behaviour (Fairclough, 1992). The words we choose, the phrases we use, and the overall tone of our communication all contribute to creating a specific image, message, and emotional impact. In the realm of OI, the language employed can significantly influence how this concept is understood, perceived, and ultimately, embraced by individuals and organizations.

A Linguistic Perspective of OI on Twitter

Analysing the specific language used in OI discussions on Twitter can provide valuable insights into the dynamics of this online community. The prevalence of certain terms, phrases, and communication styles can reveal the dominant themes, concerns, and values within this space, helping to identify trends and patterns that might not be apparent from a superficial analysis. For example, the frequency of terms like "co-creation," "crowdsourcing," or "open source" can indicate the degree of emphasis on collaboration, user involvement, and knowledge sharing within the OI discourse. Similarly, the use of inclusive language, such as "we" or "together", might suggest a community-oriented approach and a focus on shared goals.

This dissertation specifically leverages a predefined list of OI keywords derived from a comprehensive literature review by Toroslu et al. (2023). By analysing the frequency and distribution of these keywords across different datasets and user

groups, this research aims to uncover distinctive linguistic patterns associated with OI communication on Twitter.

Strategic Communication

Companies and organizations engaging in OI on Twitter can use language strategically to achieve their specific communication objectives. Understanding the different audiences, sentiments, and linguistic patterns present on the platform is crucial for crafting effective messages that resonate with the intended recipients.

For instance, a company seeking to build awareness about its OI initiatives might adopt a more general and enthusiastic tone, highlighting the positive impact of OI and using broad keywords like "innovation" or "collaboration." Conversely, a company looking for specific partners or technologies might use a more technical and targeted language, emphasizing its expertise and employing keywords related to its specific industry or area of interest.

Measuring the Impact of OI Communication

Measuring the success of OI initiatives, particularly those that leverage social media, presents unique challenges. Traditional metrics used to assess innovation success, such as patent counts, new product launches, or revenue generated, might not fully capture the complexities of OI, where value can be created and captured in diverse ways (Chesbrough & Appleyard, 2007). Furthermore, simply relying on social media metrics like likes or retweets provides a limited view of the impact of OI communication.

Challenges in Measuring OI Success

Measuring the success of OI requires a nuanced approach that considers both the quantitative and qualitative aspects of value creation (West & Bogers, 2014). Quantitative metrics, such as the number of collaborations initiated, the amount of external knowledge acquired, or the reduction in development time, can provide valuable insights into the efficiency and effectiveness of OI processes. However, these metrics often fail to capture the intangible benefits, such as enhanced brand reputation, increased customer loyalty, or the creation of a more innovative culture within the organization.

Qualitative methods, such as interviews, case studies, or surveys, can complement quantitative data by providing insights into the subjective

experiences of participants, the quality of the collaborations, and the long-term impact of OI initiatives.

Social Media Metrics

Social media metrics, such as view count, engagement rates, and sentiment scores, offer a valuable window into the dynamics of OI communication on Twitter. These metrics can provide insights into:

- **Reach and Visibility:** The extent to which OI messages are being disseminated and seen by the target audience.
- **Engagement and Interaction:** The level of interest and participation generated by OI content, as reflected in replies, retweets, likes, quotes, and bookmarks.
- **Sentiment and Perception:** The overall emotional tone and public opinion towards OI, as measured through sentiment analysis.

However, it's important to acknowledge the limitations of these metrics (Kaplan & Haenlein, 2010). Social media engagement does not necessarily translate into real-world impact, and metrics can be easily manipulated or influenced by factors unrelated to the content itself, such as bot activity, paid promotions, or the viral nature of certain types of content.

Connecting Online Interactions to Real-World Outcomes

To fully understand the impact of OI communication on Twitter, it's crucial to go beyond social media metrics and investigate the link between online interactions and offline outcomes. For example, do tweets about OI lead to the formation of new collaborations, the development of innovative products, or the adoption of new practices within organizations?

This dissertation aims to explore this connection by integrating the quantitative analysis of social media data with qualitative insights from case studies. By examining the communication strategies and experiences of companies actively engaged in OI on Twitter, this research sheds light on the mechanisms through which online interaction can translate into real-world impact.

Research Objectives and Questions

This dissertation aims to provide a comprehensive understanding of OI communication on Twitter, contributing to both theoretical and practical advancements in the field. The growing importance of OI as a driver of

organizational success, coupled with the pervasive influence of social media in shaping communication practices, necessitates a deeper exploration of how these two forces intersect. While prior research has established the value of OI (Chesbrough, 2003) and the transformative potential of social media for communication (Kaplan & Haenlein, 2010), less attention has been devoted to the specific dynamics of OI communication on Twitter. This study seeks to address this gap by investigating the sentiment, linguistic patterns, thematic structures, and network dynamics that characterize OI discourse on this influential platform. The first objective of this research is to analyse the sentiment and language surrounding OI on Twitter. Understanding the emotional tone and linguistic features associated with this concept is crucial for assessing its reception and for identifying potential barriers or facilitators to its adoption (West & Bogers, 2014). This analysis will compare the sentiment expressed in tweets related to OI with that of general innovation, unveiling unique linguistic patterns and communication styles that distinguish these discussions. This objective is operationalized through the following research questions:

1. What is the overall sentiment expressed towards OI on Twitter? How does it compare to the sentiment surrounding general innovation?
2. Are there distinct linguistic patterns associated with OI discussions on Twitter? How do these patterns differ between users who explicitly use the #openinnovation hashtag and those who do not?

The second objective of this research is to map the OI ecosystem on Twitter, uncovering the key actors, communities, and interaction patterns that shape this online space. This mapping will provide insights into the flow of information, the influence of different actors, and the formation of communities of practice around OI (Antcliff et al., 2007). By analysing the network structure and dynamics of the OI conversation on Twitter, this study aims to contribute to a better understanding of how online communities can support collaborative innovation. This objective is addressed through the following research questions:

3. How is the OI ecosystem on Twitter structured? Who are the key influencers, and what are the patterns of interaction between different actors?
4. Do different OI activities on Twitter (e.g., seeking feedback, promoting collaborations, announcing new products) lead to varying levels of engagement and impact?

The final objective of this research is to examine the relationship between online OI activities on Twitter and organizational performance. This exploration will investigate whether companies that actively engage in OI discussions on Twitter exhibit higher levels of innovation performance compared to those that are less active. This analysis will draw upon traditional performance indicators and qualitative case studies to provide a comprehensive understanding of the potential impact of online OI activities on real-world outcomes. This objective is explored through the following research question:

5. Can we develop a predictive model that accurately forecasts the impact of OI -related tweets based on factors like user characteristics, tweet content, and temporal variables?

By addressing these research questions, this dissertation aims to provide a significant contribution to the field of OI, offering both a theoretical understanding of the dynamics of OI communication on Twitter and practical insights for companies seeking to leverage this platform to enhance their innovation performance. The following chapter will outline the specific methodology employed in this study, detailing the data collection process, analytical techniques, and the research design used to address these objectives.

Methodology

This dissertation employs a mixed-methods approach to comprehensively address the research questions and objectives. The research design integrates quantitative analysis of Twitter data with qualitative case studies and predictive modelling. This multifaceted approach allows for the exploration of the linguistic and emotional characteristics of OI conversations on Twitter, their correlation with performance metrics, and the development of a tool to predict OI success on the platform.

Data Collection: Harvesting the Twittersphere

The research utilizes Twitter as its primary data source due to its unique attributes, as discussed in Chapter 2. The platform's open nature, rapid information dissemination, and global reach make it ideal for monitoring real-time conversations, identifying emerging trends, and gaining insights from a diverse community of users. The data collection process involved:

- **Search Strategy:** Tweets were collected using the hashtags #Innovation and #OpenInnovation, as these were identified as prominent markers of OI discussions in a previous study by Saura et al. (2023).
- **Data Acquisition:** Due to the limitations and cost of Twitter's API, a web scraper was employed to collect tweets over a six-month period.
    - **Dataset 1:** A random sample of 30 days was collected from January to March 2024.
    - **Dataset 2:** A complete dataset was collected for April, May, June, and July 2024.
- **Data Cleaning and Preprocessing:** The collected tweets underwent a rigorous cleaning process to remove irrelevant tweets, handle retweets and duplicates, and normalize the text for effective analysis.

Sentiment Analysis

Sentiment analysis was employed to assess the emotional tone and polarity of the collected tweets. VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon and rule-based sentiment analysis tool specifically designed for social media text, was used for this analysis (Hutto & Gilbert, 2014). VADER's ability to handle informal language, slang, and emoticons commonly used on Twitter makes it a suitable choice for this research. The analysis involved calculating the compound sentiment score for each tweet, ranging from -1 (most negative) to +1 (most positive), and comparing sentiment distributions across different hashtags and clusters.

Linguistic Analysis

To uncover distinctive language patterns in OI discussions, a keyword-based approach was employed, leveraging a predefined list of terms associated with OI. This list was extracted from a recent comprehensive literature review by Toroslu et al. (2023). The analysis involved calculating the frequency and relative frequency of these keywords within different datasets and user groups, comparing their usage across #innovation and #openinnovation tweets.

Thematic Analysis

To identify the prominent themes and topics discussed within the OI discourse on Twitter, LDA topic modelling was employed. This unsupervised machine learning technique automatically identifies underlying topics within a collection

of documents by analysing the distribution of words. The analysis involved pre-processing the tweet text, creating a dictionary and corpus, determining the optimal number of topics using the elbow method, and interpreting the extracted topics based on their associated keywords.

Network Analysis

To visualize and analyse the relationships between actors in the OI conversation on Twitter, network analysis was performed using NetworkX, a Python library for network analysis. Two distinct networks were created: one encompassing all tweets related to #innovation and #openinnovation, and another focused solely on #innovation. These networks were visualized using Gephi, and key network metrics, such as degree centrality, betweenness centrality, and clustering coefficient, were calculated to identify influential users and understand the overall structure of the network.

Language-Based Cluster Analysis

To uncover groups of users with similar communication patterns, regardless of their explicit use of hashtags, a language-based cluster analysis was performed. This involved generating tweet embeddings using a pre-trained BERT model for English, applying the k-means clustering algorithm to these embeddings using cosine similarity as the distance metric, and manually inspecting the resulting clusters to assign descriptive labels.

Case Study Analysis

To provide a more nuanced and contextualized understanding of how companies leverage Twitter for OI, three case studies were selected: WazokuCrowd, NineSigma, and PlugandPlayTC. These companies were chosen based on their active use of the #openinnovation hashtag and their diverse OI activities on Twitter. The analysis involved a detailed examination of their tweets, drawing on insights from the quantitative analyses (sentiment, linguistic, thematic, network, and cluster) to understand their communication strategies, engagement patterns, and overall approach to OI on Twitter.

Predictive Modelling

To explore the potential for predicting the impact of OI communication on Twitter, a neural network model was developed. The model incorporated the

following variables as predictors: *log(user.followersCount)*, *weekday*, *hour*, and *embeddings*. The model was trained using a supervised learning approach, with the viewCount variable as the target variable. The performance of the model was evaluated using the R-squared (R^2) metric, which measures the proportion of variance in the dependent variable that is predictable from the independent variables. The model achieved an impressive R^2 of 0.63, demonstrating its ability to predict tweet visibility with a high degree of accuracy.

Structure of the Dissertation

This dissertation unfolds through a series of chapters, each dedicated to exploring a specific facet of OI communication on Twitter. Following this introductory chapter, the research delves into the theoretical foundations of OI and the role of social media in Chapter 1. Chapter 2 provides a comprehensive overview of social media as enablers of OI, highlighting the unique characteristics and affordances of Twitter for supporting OI practices. Chapter 3 outlines the research methodology, detailing the data collection and analysis techniques employed throughout the study.

Chapters 4 through 8 present the findings of the individual analyses:

- **Chapter 4:** Sentiment Analysis
- **Chapter 5:** Thematic Exploration, Network Dynamics, and Language-Based Clustering
- **Chapter 6:** Case Studies
- **Chapter 7:** Impact and Engagement
- **Chapter 8:** Predicting View Count with a Neural Network

These chapters collectively address the research questions and objectives, providing a multifaceted understanding of the OI discourse on Twitter.

Conclusion

This dissertation has explored the dynamic and evolving landscape of OI communication on Twitter. The findings underscore the importance of social media as a platform for connecting diverse stakeholders, sharing knowledge, and driving innovation. By understanding the nuances of online communication, companies and organizations can effectively leverage Twitter to promote their OI initiatives, build relationships with key audiences, and contribute to a more collaborative and innovative future.

The insights gleaned from this research provide valuable guidance for practitioners, offering a roadmap for navigating the complexities of OI communication in the digital age. By embracing a data-driven and audience-centric approach, companies can transform Twitter into a powerful tool for fostering innovation, collaboration, and growth. Moreover, the study's focus on language, sentiment, and social dynamics underscores the importance of considering the human element in OI, recognizing that communication is not merely a means of transmitting information but a process of building relationships, shaping perceptions, and creating shared value.

As OI continues to gain momentum and as social media platforms evolve, future research should build upon the foundation laid by this dissertation, exploring new avenues for understanding the interplay between online communication and real-world OI practices. By embracing a multidisciplinary approach, integrating quantitative and qualitative methods, and leveraging the power of deep learning, future research can further enhance our understanding of this complex and dynamic field and contribute to a more open, collaborative, and innovative future.

Chapter 1 The OI Paradigm

1.1 Introduction

In today's rapidly evolving technological landscape, companies face unprecedented challenges to innovate effectively and stay ahead of the competition. The traditional model of closed innovation, where firms relied primarily on their internal resources and capabilities, has increasingly shown its limitations as the pace of technological change accelerates, knowledge becomes more widely distributed, and new business models emerge (Chesbrough, 2003). This closed paradigm struggles to harness the collective intelligence and distributed knowledge that characterize the modern innovation environment. The rise of social media platforms has fundamentally altered this landscape, offering new and powerful tools for organizations to embrace a more open and collaborative approach.

As a response to these challenges, a new paradigm for innovation has emerged: OI. This paradigm recognizes that valuable knowledge and ideas can reside outside the firm's boundaries and that companies can benefit from strategically engaging with external partners in a collaborative manner (Chesbrough & Appleyard, 2007). Social media platforms, in particular, have become increasingly important enablers of OI, facilitating communication, knowledge sharing, and collaboration among diverse stakeholders. Their ability to connect individuals and organizations across geographical and organizational boundaries creates opportunities for accessing external knowledge, accelerating innovation cycles, and developing more customer-centric products and services. This intersection of OI and social media will be explored in greater detail in Chapter 2.

This chapter delves into the concept of OI, exploring its principles, benefits, challenges, and implications for business strategy. We begin by tracing the evolution of the innovation concept, outlining the key drivers of the shift from closed to OI, and highlighting the changing role of knowledge and collaboration

in a globalized and interconnected world. Subsequently, we examine the principles and characteristics of OI, focusing on openness and collaboration, bidirectional knowledge flows, and the diversity of partners involved. We then discuss the potential benefits and challenges of adopting OI practices, analysing both the opportunities and risks associated with this strategic shift. Finally, we adopt the lens of the Resource-Based View to explore how OI can be integrated into business strategies and give rise to new business models.

1.2 The Evolution of the Innovation Concept

The concept of innovation has undergone a significant transformation in recent decades, moving from a closed, internally focused model to a more open and collaborative approach. This evolution has been driven by a multitude of factors, including rapid technological advancements, increasing globalization, and the emergence of new business models.

1.2.1 The Traditional Model of Closed Innovation

The traditional model of innovation, often referred to as "closed innovation," was characterized by a linear process in which companies relied primarily on their internal resources and capabilities to generate new ideas, develop products, and bring them to market (Chesbrough, 2003). This model was prevalent in the twentieth century, particularly in industries with high barriers to entry and relatively stable technological trajectories (Linton & Walsh, 2008). A key assumption of this paradigm was that companies could control the innovation process from start to finish, minimizing the risk of knowledge leakage and maintaining a competitive advantage by protecting their IP(Chesbrough, 2003).

1.2.2 The Rise of OI

The closed innovation model started to show its limitations in the late twentieth century, as the pace of technological change accelerated and the boundaries between industries became increasingly porous (Huizingh, 2011). Companies realized that they could no longer rely solely on their internal resources to innovate effectively and started to explore novel approaches to innovation, which led to the emergence of the OI paradigm (Chesbrough, 2003). This new paradigm recognizes that firms can and should use external as well as internal ideas, and internal and external paths to market, as they look to advance their technology

(Chesbrough & Appleyard, 2007). Several factors have been identified as key drivers of the shift from closed to OI.

One key driver has been the increasing availability and accessibility of external knowledge. Technological advancements in information and communication technology (ICT), such as the Internet, have created new channels for the transfer and sharing of knowledge, making it easier for companies to access external ideas and expertise (Dahlander & Gann, 2010). Moreover, the globalization of R&D has led to the emergence of new innovation hotspots around the world, providing companies with access to diverse pools of talent and knowledge (Haour & Jolly, 2014).

Another important driver has been the changing nature of competition. The traditional focus on building defensible positions against competitors through proprietary technologies and strong IP rights has given way to a more collaborative approach to innovation. Companies are increasingly forming strategic alliances, partnerships, and networks to share knowledge, develop technologies jointly, and bring new products to market faster (Chesbrough & Appleyard, 2007). This move toward collaboration is also a response to the challenges of innovation in complex technological fields, where a single company may lack the resources and expertise needed to innovate effectively (Pohl & Yarime, 2012).

This shift towards more collaborative approaches is particularly relevant to the use of social media in OI. Platforms like Twitter, with their ability to connect diverse stakeholders, can facilitate the very knowledge exchange and collaborative development that characterize this new paradigm (Mention et al., 2019).

1.2.3 The Role of Social Media in Enabling OI

The increasing prevalence of social media has presented a fertile ground for the advancement of OI principles and practices. Social media platforms, characterized by their inherent connectivity, interactivity, and user-generated content (UGC), offer companies a powerful set of tools to access external knowledge, collaborate with diverse stakeholders, and accelerate the diffusion of innovations (Kaplan & Haenlein, 2010; Kietzmann et al., 2011). These platforms have fundamentally democratized the innovation process, empowering individuals and communities to actively participate in shaping the future of products and services. This section explores how social media's affordances

contribute to enabling OI, and then delves into the literature examining the strategic use of social media for OI.

Connectivity, Interactivity and User-Generated Content

Connectivity is the bedrock upon which social media's influence on OI is built. These platforms enable individuals and organizations to connect, forming online communities and networks based on shared interests, professional affiliations, or expertise (Kaplan & Haenlein, 2010). This interconnectedness fosters knowledge exchange, where ideas, trends, and insights rapidly disseminate across a global network.

Interactivity, another defining feature, enhances social media's value for OI. Unlike traditional one-way communication, social media enables dialogical and participatory exchange (Saffer et al., 2013). Companies engage in direct conversations with customers, soliciting feedback, gathering ideas, and co-creating solutions (Kargaran et al., 2024). This fosters shared ownership and empowers individuals to contribute actively to innovation.

User-generated content (UGC) is vital for OI, providing rich insights into user needs, preferences, and experiences (Roma & Aloini, 2019). Companies leverage UGC to identify innovation areas, gather feedback, and explore new product development ideas. Platforms like LEGO Ideas exemplify this, demonstrating crowdsourcing's power through social media (Kohler, 2015). Similarly, IBM uses Twitter to connect with experts, foster knowledge exchange, and tap into collective intelligence (Mention et al., 2019).

What Social Media can do for OI

While the previous section explored how social media enables OI, it's crucial to examine the literature addressing the strategic use of social media for achieving specific OI objectives.

Mount and Martinez (2014) offer a comprehensive overview of social media as an OI tool, arguing that these platforms facilitate idea generation, knowledge sharing, and external collaboration. They also acknowledge the challenges, including managing information overload, protecting IP, and ensuring content quality, aligning with broader challenges discussed in Chapter 1 related to managing open innovation projects. However, their framework, while providing a valuable foundation, primarily focuses on broad applications of social media, lacking a deep dive into platform-specific strategies and neglecting the nuances

of real-time interaction and public discourse characteristic of platforms like Twitter.

Saura et al. (2023) explore OI boundaries using social media mining, demonstrating the potential of analysing this data to identify trends, understand needs, and uncover innovation opportunities. Their work, employing text mining and sentiment analysis on Twitter data, highlights the value of extracting insights for a real-time understanding of customer sentiment and emerging trends. However, while valuable for understanding the "what" of OI discussions, their focus on social media mining does not explicitly address the "how" – the strategic communication aspect of using these platforms for proactively driving OI.

Mention et al. (2019), using a qualitative case study approach, investigate how firms leverage social media to develop dynamic capabilities for innovation, focusing on sensing, seizing, and reconfiguring opportunities. Their research highlights the importance of integrating social media into a firm's dynamic capabilities framework, but their analysis lacks a specific focus on individual platforms like Twitter. While they acknowledge the potential of social media for OI, they do not delve into the specific communication strategies or metrics relevant to measuring the effectiveness of OI initiatives on these platforms. This leaves a gap in understanding how firms can operationalize the development of dynamic capabilities for OI through targeted social media strategies.

Hitchen et al. (2017), through a literature review and conceptual analysis, discuss social media's potential for facilitating open innovation in SMEs. Their work emphasizes the resource constraints faced by SMEs and how social media can help overcome these constraints by providing access to external knowledge and fostering collaboration. While valuable for understanding the specific context of SMEs, their study does not offer empirical evidence or specific guidance on leveraging individual social media platforms like Twitter for OI. Furthermore, their focus on SMEs limits the generalizability of their findings to larger organizations, which often have different resources and communication strategies.

This dissertation addresses these gaps by exploring the dynamics of the OI ecosystem on Twitter, specifically focusing on larger, established organizations. Building on the foundational work of Mount and Martinez (2014) on social media for OI, this research delves into the specific context of Twitter, leveraging its real-time nature and public discourse to analyze communication strategies and engagement patterns. While Saura et al. (2023) focus on mining social media data,

this research explores how organizations can actively use Twitter for OI, going beyond passive analysis to examine proactive communication strategies. Furthermore, while Mention et al. (2019) highlight the importance of dynamic capabilities, this dissertation provides a granular analysis of how these capabilities can be developed and leveraged through specific communication practices on Twitter. Finally, extending beyond the SME focus of Hitchen et al. (2017), this research examines how larger organizations, with their different resources and communication objectives, can leverage Twitter for OI. By combining quantitative analysis of Twitter data with qualitative case studies and the development of a predictive model, this dissertation provides a comprehensive mapping of the OI landscape on this platform, offering both theoretical insights and practical guidance for organizations seeking to maximize the impact of their OI initiatives. This research examines how companies use Twitter to engage communities and maximize the impact of their OI initiatives, developing a predictive model for understanding the factors that drive engagement with OI-related content on Twitter, offering companies a practical tool to optimize communication strategies. Through analysis of Twitter data, including tweet content, user characteristics, and engagement metrics, this work contributes to a more nuanced understanding of the strategic use of Twitter for OI within the context of larger organizations.

1.3 Principles and Characteristics of OI

OI is characterized by a set of principles that distinguish it from the traditional closed innovation model. These principles are rooted in the recognition that valuable knowledge and ideas can reside outside the boundaries of the firm, and that companies can benefit from engaging with external partners in a collaborative manner (Chesbrough, 2003).

1.3.1 Openness and Collaboration

One of the core principles of OI is the emphasis on openness and collaboration. This entails moving beyond the traditional, internally focused approach to innovation and embracing the idea that valuable knowledge and expertise can reside outside the company (Chesbrough & Appleyard, 2007). This openness extends to collaboration with a wide range of external partners, including customers, suppliers, universities, start-ups, and even competitors (Dahlander & Gann, 2010). The theoretical lens of Service-Dominant Logic (SDL) provides a

valuable framework for understanding this shift toward openness and collaboration in innovation (Wilden et al., 2017). SDL posits that value is co-created by firms and customers through the integration of resources and capabilities in a service ecosystem (Vargo & Lusch, 2004). In the context of OI, this implies that firms can create new value by engaging with customers and other stakeholders in a collaborative process of knowledge exchange and co-creation (Mention et al., 2019).

1.3.2 Bidirectional Knowledge Flows

OI involves bidirectional knowledge flows, which means that knowledge can flow both into and out of the firm (Dahlander & Gann, 2010). This contrasts with the traditional model of closed innovation, where knowledge flows were primarily unidirectional, from internal R&D to the market. Inbound OI refers to the process of acquiring external knowledge and integrating it into the firm's innovation activities (Chesbrough, 2003). Outbound OI involves leveraging the firm's internal knowledge externally, through licensing, spin-offs, or other mechanisms (Lichtenthaler, 2010). Coupled OI combines inbound and outbound practices, creating a synergistic process of knowledge exchange and co-creation with external partners (Van Lancker et al., 2016).

1.3.3 Breadth and Depth of OI

The diversity of partners involved in OI is a key characteristic that distinguishes it from the traditional model of closed innovation. This diversity, often referred to as the *breadth* of OI, encompasses a wide range of external actors, including customers, suppliers, universities, startups, and even competitors (Chesbrough & Appleyard, 2007). Companies are increasingly recognizing that each type of partner can bring unique knowledge, expertise, and perspectives to the innovation process.

However, the extent to which companies engage with these partners can vary significantly. This level of engagement, referred to as the *depth* of OI, represents the degree of integration and collaboration with external actors. The depth of OI can range from shallow interactions, such as simple licensing agreements or one-time consultations, to deep, long-term partnerships involving joint R&D, co-creation of new products, or the formation of strategic alliances (Garriga et al., 2013).

For example, a company engaging in inbound OI might acquire a technology license from a university (low depth) or engage in a joint research project with a research institution (high depth). Similarly, a company pursuing outbound OI might license its technology to another firm for a specific application (low depth) or create a spin-off company to commercialize its technology in a new market (high depth).

The choice of breadth and depth in OI strategies depends on a variety of factors, including the company's resources, capabilities, strategic goals, and the nature of the innovation challenge. Companies need to carefully consider the trade-offs between the benefits and challenges of different levels of engagement with external partners. Deeper collaboration can lead to greater knowledge sharing, faster innovation cycles, and the development of more radical innovations, but it also requires greater investment, coordination, and trust-building.

Measuring the success of the diverse approaches to OI, such as inbound, outbound, and coupled OI, requires a nuanced approach, especially when considering the dynamic environment of Twitter. Traditional metrics, such as patent counts or new product launches, might not fully capture the value created through OI, which can manifest in various forms, including enhanced knowledge, improved processes, and stronger relationships with external partners.

1.4 Benefits and Challenges of OI

The adoption of OI practices offers a multitude of potential benefits to firms, particularly in a rapidly changing and increasingly interconnected business environment. However, this strategic shift also presents a range of challenges that require careful consideration and effective management.

1.4.1 Benefits for Companies

OI offers a multitude of potential benefits to firms, particularly in a rapidly changing and increasingly interconnected business environment. By leveraging external knowledge and collaborative partnerships, companies can accelerate their innovation processes, reduce R&D costs, and gain access to new competencies and markets (van de Vrande et al., 2009).

Acceleration of Innovation

OI can significantly accelerate the innovation process by enabling companies to tap into a wider pool of ideas, expertise, and resources (Chesbrough, 2003).

Collaboration with external partners, such as universities, research institutes, startups, and even competitors, allows firms to leverage specialized knowledge, shorten development cycles, and bring new products and services to market faster (van de Vrande et al., 2009). This agility is particularly critical in industries characterized by rapid technological advancements and short product lifecycles, as exemplified by the electronics and software sectors (Wang et al., 2015).

For instance, Samsung, a leading electronics company, actively engages in OI by partnering with universities and research institutes to gain access to cutting-edge technologies and expertise in areas such as artificial intelligence, 5G, and advanced materials. These collaborations have allowed Samsung to accelerate its development of innovative products, such as foldable smartphones and AI-powered home appliances, maintaining its competitive edge in the fast-paced consumer electronics market. (Binti Mohd Haizar et al., 2020)

Cost Reduction

OI can significantly contribute to reducing R&D costs by enabling companies to share the financial burden of innovation with external partners. This is especially beneficial for SMEs, which often face resource constraints that limit their ability to invest heavily in internal R&D (Radziwon & Bogers, 2019). Collaborating with universities, research institutions, or larger companies allows SMEs to access specialized knowledge and expertise at a lower cost than developing it internally (Filieri, 2015).

A notable example is the case of Syngenta, an agribusiness company that established a strategic partnership with the University of Manchester to create a University Innovation Centre focused on developing new technologies for agriculture (Malik et al., 2011). By leveraging the university's research expertise and resources, Syngenta was able to accelerate the development of innovative solutions in areas such as agri-electronics and precision farming, while sharing the associated R&D costs. This collaboration has enabled Syngenta to gain a competitive advantage in the rapidly evolving agricultural technology market.

Access to New Competencies and Markets

Collaboration with external partners can provide firms with access to new technologies, skills, and market knowledge that they may not possess internally (van de Vrande et al., 2009). This access can enable them to expand their product portfolios, diversify their offerings, and enter new markets (Gkypali et al., 2018).

For companies seeking to overcome technological barriers or expand their global reach, OI can be a crucial strategic lever (Mention, 2011).

One example is the case of IBM, a global technology company that has embraced OI to access new technologies and markets. The company actively engages in strategic partnerships with startups, universities, and other technology leaders to develop innovative solutions in areas such as cloud computing, artificial intelligence, and blockchain. These collaborations have allowed IBM to expand its product portfolio, enter new market segments, and maintain its position among its competitors. (Yang et al., 2021)

1.4.2 Organisational and Managerial Challenges

Despite its potential benefits, OI also presents a range of organizational and managerial challenges that need to be addressed effectively (van de Vrande et al., 2009). One of the most significant hurdles lies in managing the inherent complexity of collaborating with external partners. This complexity stems from the need to coordinate activities, align objectives, and integrate knowledge across organizational boundaries, demanding new management skills and adaptable organizational structures (Thanasopon et al., 2016). Successfully navigating these complexities often necessitates developing robust collaborative frameworks and adopting agile methodologies to ensure effective communication, knowledge sharing, and conflict resolution between partners (Barbosa et al., 2020).

Another critical challenge is safeguarding IP within an OI environment. Sharing knowledge with external partners inherently introduces a risk of IP leakage and potential appropriation by competitors (Bogers, 2011). While selective revealing can be employed to control the flow of sensitive information (Henkel, 2006), robust contractual agreements are crucial to establish clear IP ownership rights and recourse in case of infringement. Cultivating strong relationships with partners based on trust, mutual respect, and shared values is also essential for mitigating the risk of opportunism and encouraging reciprocal knowledge sharing (Brockman et al., 2018).

Finally, shifting towards an OI model often requires a fundamental change in organizational culture (Chesbrough, 2003). Traditional organizational cultures rooted in secrecy, internal competition, and a "not-invented-here" (NIH) mentality can significantly impede the adoption of OI practices (Burcharth et al., 2014). To effectively embrace OI, firms need to foster a more collaborative and open culture that encourages knowledge sharing, external engagement, and a

willingness to learn from diverse sources. This involves promoting trust and transparency, rewarding employees for collaborating with external partners, and actively addressing any resistance to change (Burcharth et al., 2014).

1.4.3 Managing the Risks of OI

While OI offers numerous benefits, it also presents inherent risks that companies must actively manage to ensure the success and sustainability of their innovation efforts. These risks stem from the inherent openness and collaboration that define this paradigm, potentially exposing firms to IP leakage, challenges in managing external collaborations, and the need for cultural adaptation within the organization.

IP Leakage

Sharing knowledge with external partners, a cornerstone of OI, inherently increases the risk of IP leakage (Bogers, 2011). Companies must be vigilant in protecting their core technologies and valuable knowledge assets, especially when collaborating with competitors or other actors with potentially conflicting interests (Frishammar et al., 2015). Several strategies can be employed to mitigate this risk:

- Selective Revealing (Henkel, 2006): Companies can strategically choose what information to share with external partners, revealing only the necessary knowledge while protecting their core IP. This involves carefully evaluating the potential risks and benefits of sharing specific information and implementing mechanisms to control the flow of knowledge.
- Strong Contractual Agreements: Establishing clear contractual agreements with partners, outlining IP ownership rights, confidentiality clauses, and non-disclosure agreements, can help protect a firm's IP and establish legal recourse in case of infringement.
- Building Trust and Reputation (Brockman et al., 2018): Cultivating strong relationships with partners based on trust, mutual respect, and shared values can reduce the risk of opportunism and encourage reciprocal knowledge sharing. Companies can build trust through open communication, transparency, and a demonstrable commitment to collaborative goals.

Challenges in Managing External Collaborations

Collaboration across organizational boundaries can be complex, requiring effective management to ensure alignment of objectives, efficient coordination of activities, and smooth knowledge integration (van de Vrande et al., 2009). These challenges can be exacerbated by cultural differences, communication barriers, and potential conflicts of interest between partners.

- Partner Selection and Due Diligence: Carefully selecting partners based on their expertise, capabilities, and alignment with the firm's values and goals can reduce the risk of mismatches and conflicts. Thorough due diligence is crucial to assess the partner's track record, trustworthiness, and commitment to collaboration.
- Clearly Defined Roles and Responsibilities: Establishing clear roles, responsibilities, and decision-making processes for each partner can ensure smooth collaboration and minimize confusion or misunderstandings.
- Effective Communication and Knowledge Sharing: Facilitating open and transparent communication channels between partners is essential for effective knowledge transfer and coordination of activities. This may involve regular meetings, online collaboration tools, and knowledge management systems to enable the seamless flow of information.

Organizational Culture and Openness

The shift to OI often demands a fundamental change in organizational culture (Chesbrough, 2003). Traditional, internally focused cultures that emphasize secrecy and a "not-invented-here" mentality can hinder OI (Burcharth et al., 2014). Fostering a more collaborative and open culture is crucial to successfully embracing this paradigm.

- Leadership Commitment and Support: Leaders must champion OI, demonstrating their commitment through communication, resource allocation, and by actively participating in collaborative initiatives.
- Employee Training and Development: Training programs can help employees develop the skills and knowledge necessary to engage in OI effectively. This could include training on collaboration tools, knowledge sharing practices, and IP management.
- Reward and Recognition Systems: Recognizing and rewarding employees for their contributions to OI can reinforce a culture of

collaboration and encourage participation in external engagement activities.

By effectively addressing these risks, companies can unlock the full potential of OI, fostering a more dynamic and collaborative approach to innovation that enhances their competitiveness and drives sustainable growth.

1.5 OI and Business Strategies

The integration of OI into a firm's strategic management practices requires a shift in thinking from traditional, internally focused approaches to a more dynamic and externally oriented perspective. The resource-based view offers a valuable theoretical lens for understanding how firms can leverage OI to create and sustain a competitive advantage.

1.5.1 Integrating OI into Business Strategies

The Resource-Based View posits that a firm's competitive advantage stems from its unique and valuable resources and capabilities (Barney, 1991). These resources can be tangible, such as physical assets and financial capital, or intangible, such as knowledge, skills, and organizational processes (Wernerfelt, 1984). In the context of OI, the Resource-Based View suggests that firms can enhance their competitive advantage by strategically acquiring and integrating external knowledge and resources to complement their internal capabilities (Gkypali et al., 2018). This entails identifying and evaluating potential partners, building collaborative relationships, and developing mechanisms to effectively absorb and leverage external knowledge (Hutton et al., 2021).

Integrating OI into business strategies requires firms to move beyond a transactional approach to collaboration and embrace a more strategic and long-term perspective. Rather than simply seeking external knowledge to address immediate needs, firms should consider how OI can contribute to developing and strengthening their core competences and creating new strategic opportunities (Walsh & Linton, 2011). This entails aligning OI activities with the firm's overall business strategy, identifying potential synergies between internal and external knowledge, and developing dynamic capabilities to manage the complexities of collaboration.

1.5.2 Business Models based on OI

OI has profoundly changed how companies approach innovation, giving rise to novel business models that redefine value creation and capture (Chesbrough & Appleyard, 2007). These new models capitalize on the principles of openness, collaboration, and external knowledge integration to achieve competitive advantage and drive sustainable growth. Two prominent examples of such business models are platform-based models and crowdsourcing-based models.

Platform-Based Models:

Platform-based business models have emerged as a dominant force in the digital economy. These models leverage a central platform that provides a shared infrastructure and a set of rules for user interaction, facilitating transactions and fostering value co-creation among participants (Kohler, 2015). OI is crucial for platform-based models, enabling the flow of knowledge and ideas between participants, leading to the development of complementary products and services, and ultimately creating a thriving ecosystem (Gawer, 2014).

Companies like Uber, Airbnb, and Android exemplify the success of platform-based models:

- Uber: By creating an open platform connecting drivers and riders, Uber has built a global transportation network relying on independent contractors. This model fosters innovation through continuous service improvement based on user feedback, new feature development, and integration with other platforms (Ulloa et al., 2016).
- Airbnb: Airbnb's platform enables travellers to connect with hosts offering unique accommodations, forming a global marketplace built on the participation of individual homeowners. This model leverages OI to enhance the platform's features, integrate with travel-related services, and incorporate user feedback (Roma & Aloini, 2019).
- Android: Google's open-source mobile operating system exemplifies a platform-based model driven by OI. By making the Android source code freely available, Google has fostered a vast ecosystem of developers, hardware manufacturers, and app creators, resulting in rapid innovation and widespread adoption (Grönlund et al., 2010).

Crowdsourcing-Based Models:

Crowdsourcing-based models harness the collective intelligence and creativity of a large and diverse group of individuals to address challenges, generate ideas,

and develop new products and services (Kohler, 2015). These models leverage the power of distributed knowledge and open participation, often yielding innovative solutions that may not be achievable through traditional internal R&D approaches.

Companies like LEGO, InnoCentive, and Threadless have successfully employed crowdsourcing models to drive innovation:

- LEGO Ideas: LEGO Ideas exemplifies crowdsourced innovation by inviting users to submit designs for new Lego sets. This platform enables LEGO to tap into the creativity of its global fan base, identify potentially successful products, and engage customers in the development process (Dahlander et al., 2023).
- InnoCentive: This platform connects companies with a global network of problem-solvers to address complex scientific and technical challenges. By offering rewards for solutions, InnoCentive enables firms to access specialized expertise from a diverse pool of individuals and accelerate their innovation efforts (Schuhmacher et al., 2022).
- Threadless: This T-shirt design company relies entirely on crowdsourcing, hosting design contests where users submit designs and vote on their favourites. The winning designs are then produced and sold, demonstrating a customer-driven and highly innovative approach to product development (Burger-Helmchen & Siegel, 2020; Kohler, 2015).

OI is crucial for managing these crowdsourcing-based models by providing mechanisms for engaging with the crowd, incentivizing participation, ensuring quality, and fostering a collaborative environment. Companies need to establish clear contest guidelines, provide feedback to participants, protect intellectual property, and cultivate a sense of community to ensure the success and sustainability of their crowdsourcing initiatives.

1.6 Conclusions

This chapter has laid the theoretical groundwork for understanding the OI paradigm and its increasing importance in today's dynamic business environment. By tracing the evolution of the innovation concept from closed to open models, highlighting the shift from internal, secretive R&D to collaborative, externally-focused innovation practices, we have emphasized the critical role of external knowledge, collaboration, and, increasingly, social media in driving successful innovation. The limitations of traditional closed innovation models,

particularly in the face of rapid technological advancements and the increasingly distributed nature of knowledge, have been underscored, demonstrating why organizations need to embrace OI principles for sustained competitiveness.

The core tenets of OI – openness, collaboration, bidirectional knowledge flows, and engagement with a diverse range of partners – have been thoroughly examined, demonstrating the shift towards a more permeable and interconnected approach to innovation. Specifically, the discussion highlighted the importance of outside-in, inside-out, and coupled OI approaches, emphasizing how these strategies enable organizations to access and leverage external knowledge while also sharing their internal knowledge assets. Furthermore, the affordances of social media platforms, particularly their connectivity, interactivity, and capacity for generating user-generated content, have been discussed as key enablers of OI, facilitating the connection and collaboration necessary for successful open innovation practices. Critically examining the literature on the strategic use of social media for OI revealed a gap in understanding how organizations can effectively leverage platforms like Twitter to achieve specific OI objectives, particularly concerning maximizing impact and engagement with relevant communities. This dissertation seeks to address this gap by focusing on the dynamics of OI communication on Twitter within the context of larger organizations.

The potential benefits of OI, such as accelerated innovation cycles, reduced R&D costs, and access to new competencies and markets, have been contrasted with the inherent challenges, including the risk of IP leakage, the complexities of managing external collaborations, and the need to cultivate an open and collaborative organizational culture. The Resource-Based View has been introduced as a valuable framework for understanding how OI can be integrated into business strategies, creating a competitive advantage by strategically leveraging external knowledge and resources to complement internal capabilities. Furthermore, the discussion of emergent business models based on OI, including platform-based and crowdsourcing models exemplified by companies like Uber, Airbnb, and LEGO Ideas, has illustrated the transformative potential of this paradigm in reshaping value creation and capture.

This chapter has addressed the underlying theoretical context relevant to the research questions posed in the introduction. By examining the evolution of innovation, the core principles of OI, the role of social media in enabling OI practices, the benefits and challenges associated with OI adoption, the integration

of OI into business strategies through the lens of the Resource-Based View, and the emergence of novel business models based on OI, this chapter provides a robust foundation for understanding the empirical findings presented in the subsequent chapters. The next chapter will delve into the specifics of social media as a catalyst for open innovation, with a particular focus on the unique characteristics of Twitter that make it a valuable platform for studying OI communication. Subsequent chapters will analyze the empirical data collected from Twitter, exploring the sentiment, language, thematic structures, network dynamics, and impact of OI communication on this platform. This analysis will offer valuable insights into how organizations can strategically leverage Twitter to effectively communicate their OI initiatives, engage with relevant communities, and maximize the impact of their innovation efforts.

Chapter 2 Social Media as a Catalyst for OI

2.1 Introduction

The landscape of innovation is undergoing a dramatic transformation in the 21st century. Driven by rapid technological advancements, increasing globalization, and the rise of new business models, organizations are recognizing the limitations of traditional, closed innovation approaches (Chesbrough, 2003). The closed innovation paradigm, characterized by a reliance on internal resources and a protective stance towards intellectual property, struggles to keep pace with the accelerated flow of knowledge and the interconnected nature of modern innovation (Chesbrough & Appleyard, 2007). This shift necessitates a new approach, one that embraces external collaboration, knowledge sharing, and rapid iteration: OI (West & Bogers, 2014).

OI, at its core, recognizes the value of external knowledge and the potential for collaborative partnerships to drive breakthrough innovation. It encompasses a range of practices, from crowdsourcing and open-source development to strategic alliances and joint ventures, all aimed at leveraging the collective intelligence of diverse communities (Dahlander & Gann, 2010). However, the successful implementation of OI requires more than just a change in mindset; it demands effective mechanisms for connecting with external partners, facilitating communication, and managing the flow of knowledge.

In this context, social media has emerged as a powerful catalyst for OI (Mention et al., 2019). With their unprecedented reach, interactive nature, and capacity to connect diverse actors, social media platforms offer organizations new avenues for accessing external knowledge, fostering collaboration, and accelerating innovation processes (Kaplan & Haenlein, 2010). From identifying emerging trends and gathering user feedback to scouting for potential partners and co-

creating solutions, social media empowers organizations to tap into the collective intelligence of a global network of individuals and organizations.

This chapter delves into the transformative potential of social media as an enabler of OI. We begin by revisiting the core principles of OI and exploring how social media has reshaped these principles in the digital age. We then analyse the specific affordances of social media platforms that contribute to OI success, including connectivity, interactivity, user-generated content, and data availability. We introduce a unified theoretical framework, drawing on Social Capital Theory, Network Theory, Diffusion of Innovation Theory, and Social Influence Theory, to provide a comprehensive lens for understanding the dynamics of social media-enabled OI. Finally, we justify our selection of Twitter as the focus of our empirical analysis, highlighting its unique characteristics and outlining the specific research questions that will guide our investigation. This chapter lays the groundwork for understanding the role of social media in shaping the future of OI.

2.2 Defining and Conceptualizing OI in the Social Media Age

2.2.1 Core Tenets of OI Revisited

OI, as discussed in Chapter 1, marks a departure from the traditional reliance on internally generated knowledge and resources. Instead, it acknowledges that valuable insights and expertise can exist outside an organization's boundaries (Chesbrough, 2003). OI embraces a more permeable approach to innovation, advocating for the strategic integration of external sources to accelerate R&D, reduce costs, and develop more customer-centric solutions (van de Vrande et al., 2009). The core tenets of OI – outside-in, inside-out, and coupled – describe the various ways firms can leverage external knowledge. Outside-in OI involves actively acquiring and integrating external knowledge into a firm's own innovation activities (Chesbrough, 2003). Inside-out OI focuses on leveraging a firm's internal knowledge externally, such as through licensing or spin-offs, to generate new revenue streams or reach new markets (Lichtenthaler, 2010). Coupled OI, as described by Mention (2011), combines these approaches, fostering a synergistic flow of knowledge between the firm and its external partners. Social media, with its unparalleled capacity to connect diverse actors across geographical boundaries, has emerged as a significant catalyst for all three forms of OI.

2.2.2 Social Media's Transformative Impact on OI

The influence of social media extends beyond merely facilitating the existing OI processes; it has fundamentally transformed the ways organizations approach and engage in OI.

2.2.2.1 Distributed Knowledge and Collective Intelligence

Social media platforms have democratized access to knowledge, making it easier than ever before for firms to tap into the collective intelligence of a global network of individuals and organizations (Kohler, 2015). This distributed knowledge base offers a rich source of expertise, ideas, and perspectives that can be leveraged for OI. Platforms like LEGO Ideas exemplify this, enabling companies to identify innovative ideas from their user community and engage customers in the product development process (Kohler, 2015). Similarly, IBM actively uses Twitter to connect with experts, foster knowledge exchange, and gather insights from a global network of professionals (Mention et al., 2019).

2.2.2.2 Collaboration Beyond Boundaries

The ability of social media to connect individuals and organizations across geographical and organizational boundaries makes it an ideal tool for facilitating collaborative OI (Jussila et al., 2012). Companies can use these platforms to forge strategic alliances with external stakeholders, including universities, research institutions, startups, and even competitors. Such collaborations can take various forms, from joint development projects to online hackathons, enabling the sharing of resources, expertise, and risk (van de Vrande et al., 2009). Social media facilitates communication, coordination, and knowledge sharing among dispersed teams, fostering a more agile and responsive approach to innovation.

2.2.2.3 Democratization of Innovation

Social media has empowered individuals and communities to actively participate in shaping the future of products and services, blurring the traditional lines between "producers" and "consumers" (Kohler, 2015). This democratization of innovation aligns perfectly with the core principles of OI, allowing companies to tap into the collective creativity of a wider pool of individuals. Citizen science projects, for instance, have harnessed social media's power to recruit, retain, and train volunteers, expanding the reach and impact of scientific research (Oliveira et al., 2021). Platforms like Facebook provide space for knowledge sharing and awareness-building in citizen science initiatives, fostering a sense of community and collective learning (Liberatore et al., 2018). Beyond structured projects, social

media also enables "passive citizen science," where individuals share incidental observations of the natural environment, creating valuable datasets for ecological research (Edwards et al., 2021).

Similarly, user-led innovation communities, often focused on specific products or technologies, flourish on social media (Dong & Wu, 2015). These online communities leverage social capital built through interconnectivity, driving co-creation and influencing product development (Bugshan, 2015). Research has shown that the value derived from such online user innovation communities (OUICs) is tied to how effectively firms implement and act on user-generated ideas rather than simply the volume of ideas collected (Dong & Wu, 2015). Social media's capacity for fostering multi-faceted engagement within these communities, encompassing cognitive, integrative, and affective benefits, significantly stimulates user contributions and drives innovation (Ogink & Dong, 2019).

2.2.2.4 Real-Time Feedback and Accelerated Iteration

Social media's capacity for real-time communication enables companies to gather immediate feedback from users, accelerating the pace of innovation (Kaplan & Haenlein, 2010). This shortened feedback loop allows for rapid iteration and adaptation, fostering more customer-centric and market-responsive product development (Jussila et al., 2012). Companies can leverage social media for beta testing, gathering user input on prototypes, or soliciting ideas for new features. This direct engagement with users allows companies to identify and address potential issues early on, reducing development costs and increasing the likelihood of market success.

2.2.2.5 Managing the Challenges of Openness

The openness inherent in OI, particularly when facilitated by social media, introduces new challenges that organizations must navigate effectively. Protecting IP becomes crucial when collaborating with external partners. Companies need to carefully consider what information to share, implement robust contractual agreements, and build trust with their collaborators (Bogers, 2011). Another challenge is managing the deluge of information available on social media. The sheer volume of data can be overwhelming, potentially leading to information overload and hindering effective decision-making (Gouws & Tarp, 2016). Companies need to develop strategies for filtering, analysing, and extracting meaningful insights from social media data, distinguishing valuable

information from noise (Gomez Rodriguez et al., 2014). Additionally, organizations need to ensure that online discussions and activities translate into tangible innovation outcomes, bridging the gap between virtual engagement and real-world impact.

2.3 Social Media's Affordances for OI

Social media's impact on OI stems from a confluence of characteristics that distinguish it from traditional communication channels. These affordances make social media platforms particularly well-suited for supporting the open and collaborative nature of OI.

2.3.1 Connectivity and Network Effects

Building and Maintaining Networks

Social media platforms excel at facilitating connections between individuals and organizations. Through features like user profiles, groups, and following mechanisms, these platforms enable the formation and maintenance of vast networks based on shared interests, professional affiliations, or specific areas of expertise (Kaplan & Haenlein, 2010). For example, LinkedIn, with its focus on professional networking, provides a space for individuals and organizations to connect based on industry, skills, and experience. Specialized LinkedIn groups focused on OI, innovation management, or specific technologies facilitate the formation of communities where individuals can share knowledge, exchange ideas, and identify potential collaborators. Similarly, professional forums and online communities dedicated to innovation provide spaces for virtual networking and knowledge sharing.

Leveraging Network Effects

The value of a social media network for OI increases as more participants join. This phenomenon, known as the network effect, amplifies the potential for knowledge exchange, collaboration, and innovation. A larger network translates to a wider reach, allowing companies to access a more diverse pool of expertise and perspectives (Gkypali et al., 2018). Moreover, as more individuals and organizations connect, the network becomes more densely interconnected, facilitating the flow of information and ideas, and increasing the likelihood of serendipitous encounters that can lead to new collaborations or partnerships (Antcliff et al., 2007).

2.3.2 Interactivity and Engagement

Types of Interactions and Their Significance

Social media platforms foster interactivity and engagement through various mechanisms, each contributing to OI in distinct ways. Likes, shares, and comments provide immediate feedback, gauging user interest and sentiment towards ideas or initiatives. Retweets amplify the reach of messages, extending them beyond the initial audience and promoting broader awareness. Replies facilitate direct interaction and dialogue, fostering deeper engagement and collaborative discussions. Furthermore, polls and surveys can be utilized to gather user input and feedback, while online contests can encourage participation and co-creation.

Building Relationships and Fostering Trust

The interactive nature of social media communication fosters a more dialogical and participatory exchange between companies and their stakeholders, moving beyond traditional, one-way models of communication (Saffer et al., 2013). This shift towards two-way dialogue fosters a sense of shared ownership and empowers individuals to actively contribute to innovation processes. By engaging in conversations, responding to comments, and acknowledging user contributions, companies can build relationships based on trust, reciprocity, and mutual understanding (Brockman et al., 2018).

Engagement Strategies for OI

Effectively leveraging social media for OI requires strategic engagement practices. Companies need to develop a clear communication strategy, identifying their target audience and tailoring their messages accordingly. Actively participating in relevant conversations, responding to user inquiries, and fostering a sense of community can significantly enhance engagement levels. Furthermore, utilizing a variety of content formats, such as text, images, videos, and polls, can capture audience attention and encourage interaction (Kaplan & Haenlein, 2011). However, it's essential to avoid overly promotional or generic language, prioritizing authenticity and genuineness to build trust and credibility with the online community.

2.3.3 UGC and Co-creation

UGC as a Source of Innovation

Social media platforms are rich sources of UGC, providing valuable insights into user needs, preferences, and experiences (Roma & Aloini, 2019). This content can encompass a wide range of formats, including text, images, videos, reviews, and discussions, reflecting a multitude of perspectives and experiences with products and services. Companies can leverage this UGC to identify potential areas for innovation, gather feedback on existing products or services, and explore new ideas for product development.

Facilitating Co-creation through Social Media

Social media platforms provide tools and spaces for facilitating co-creation, enabling companies to engage with their online communities in collaborative problem-solving and product development (Kohler, 2015). Companies can leverage social media to host design contests, online hackathons, or virtual brainstorming sessions, inviting users to contribute their ideas and expertise. Platforms like LEGO Ideas and Threadless exemplify successful co-creation initiatives, demonstrating the power of harnessing the collective creativity of a user community to drive innovation.

Managing and Evaluating UGC for OI

While UGC offers significant potential for OI, it also requires careful management and evaluation. The sheer volume of UGC generated on social media platforms necessitates the use of tools and techniques for filtering, evaluating, and incorporating relevant content into the innovation process. As Meneghello et al. (2020) suggest, a robust knowledge management framework is necessary to unlock the full potential of UGC as a real-time data source for innovation. This includes automated methods for collecting and extracting data from embedded comments, reviews, and testimonials, as well as algorithms for analysing and interpreting this data to extract valuable insights. Companies also need to develop clear criteria for evaluating UGC, establish processes for integrating user feedback into product development, and ensure that they acknowledge and reward valuable contributions from their online communities.

2.3.4 Data and Analytics for OI (Overview)

Potential of Social Media Data

Social media platforms generate vast amounts of data that can be leveraged to understand user behaviours, trends, and opinions relevant for OI (Zachlod et al., 2022). Different types of analyses can be applied to this data, including sentiment analysis to understand user sentiment towards specific topics or brands, network analysis to map relationships and identify influencers within online communities, and content analysis to extract themes and patterns from user-generated text and multimedia content.

Measuring the Impact of OI Initiatives

Social media analytics tools provide valuable insights into the reach and effectiveness of OI campaigns (Mention et al., 2019). Companies can track metrics such as reach, engagement, and sentiment to understand how their OI messages are being received by their target audience. This data can be used to evaluate the success of specific campaigns, identify areas for improvement, and optimize communication strategies to maximize impact and engagement.

2.3.5 Challenges of Social Media for OI

Information Overload and Filtering

The sheer volume of data generated on social media platforms presents a significant challenge for organizations seeking to leverage this information for OI. Information overload can hinder effective decision-making, making it difficult for companies to discern relevant insights from the noise (Gouws & Tarp, 2017). Effective filtering strategies, utilizing keywords, hashtags, Boolean operators, and sophisticated social media analytics platforms, are crucial for managing this data deluge and extracting meaningful information.

Ensuring Data Quality and Addressing Misinformation

The open and participatory nature of social media makes it susceptible to the spread of misinformation, which can negatively impact the quality of insights derived from this data. The spread of misinformation has become a major threat to public interests, particularly in sensitive domains like disaster relief, healthcare, and politics (Muhammed T & Mathew, 2022). Companies need to be vigilant in verifying the accuracy of information, employing fact-checking techniques and cross-referencing information with reliable sources. Aïmeur et al. (2023) highlight the inherent difficulty in automatically detecting fake news, as its content is often

designed to closely resemble truth. They recommend combining AI-based techniques with information from trusted third parties for effective verification. Additionally, promoting media literacy and critical thinking skills within the online community can help mitigate the spread of misinformation and ensure that discussions are based on credible information.

This section has explored the various affordances of social media that make it a powerful tool for enabling and enhancing OI. These characteristics, when effectively leveraged, can facilitate the formation of valuable networks, foster meaningful engagement, generate innovative ideas through UGC, and provide insights through data analytics. However, it's crucial to acknowledge the potential challenges associated with social media, such as information overload, misinformation, and privacy concerns, and to implement appropriate strategies for mitigating these risks.

2.4 A Unified Theoretical Framework for Social Media and OI

To fully understand the complex interplay between social media and OI, we need a robust theoretical lens. This dissertation draws upon a multi-theoretical framework, integrating concepts from various disciplines to provide a comprehensive and nuanced understanding of how social media shapes OI dynamics.

2.4.1 Social Capital Theory (SCT)

SCT, originally proposed by Bourdieu (2011), emphasizes the value of social relationships and networks as a form of capital. It posits that individuals and organizations can leverage their social connections to access resources, information, and support that can enhance their ability to achieve their goals. In the context of OI, social media platforms enable the creation, maintenance, and leveraging of various forms of social capital.

Bridging Social Capital: Social media facilitates connections between individuals and organizations that might not otherwise interact, bridging structural holes and fostering collaboration across diverse communities. This bridging capital is particularly relevant for accessing external knowledge and forging partnerships for OI. For example, companies can leverage LinkedIn to connect with experts from different industries or research institutions, facilitating knowledge transfer and the identification of potential collaborators (Heger & Boman, 2015).

Bonding Social Capital: This form of capital refers to strong ties within tightly-knit communities. Social media can help foster bonding capital within OI communities by providing spaces for knowledge sharing, mutual support, and the development of shared identities (Antcliff et al., 2007). For example, online communities dedicated to specific technologies or areas of OI can provide a sense of belonging and facilitate knowledge exchange among members.

Linking Social Capital: Linking social capital refers to connections between individuals or organizations at different levels of power or hierarchy. Social media can bridge these hierarchical gaps, enabling individuals to connect with influential figures, industry leaders, or decision-makers, potentially facilitating access to resources or opportunities for OI (Oliveira-Dias et al., 2022).

2.4.2 Network Theory

Network theory provides a valuable framework for analysing the structure and dynamics of social relationships, highlighting the importance of connections and pathways within complex systems. When applied to social media and OI, network analysis can help visualize and understand the relationships between key actors (individuals, organizations, communities), mapping the flow of information, knowledge, and resources within the OI ecosystem (Park et al., 2015). Key network metrics, such as degree centrality (number of connections), betweenness centrality (bridging role), and clustering coefficient (density of connections within a group), provide insights into the influence of different actors, the pathways for knowledge diffusion, and the formation of communities of practice around OI.

2.4.3 Diffusion of Innovation Theory (DOI)

DOI, pioneered by Rogers (1995), explains how new ideas and practices spread through social systems. It highlights the importance of communication channels, opinion leaders, and the characteristics of the innovation itself in influencing the adoption process. Social media, with its capacity for rapid information dissemination and user-to-user communication, can significantly accelerate the diffusion of OI principles and practices. Companies can leverage social media to target different adopter categories (innovators, early adopters, laggards) with tailored messages, promoting awareness, and overcoming barriers to adoption (Kaplan & Haenlein, 2011). The interactive nature of social media also facilitates

peer-to-peer influence, leveraging social proof and testimonials to encourage adoption.

2.4.4 Social Influence Theory (SIT)

SIT, originally proposed by Deutsch & Gerard (1955), explains how individuals' thoughts, feelings, and behaviours are shaped by the influence of others. This theory is particularly relevant in the context of social media, where individuals are constantly exposed to information, opinions, and behaviours from their social network. SIT highlights three key mechanisms of social influence:

Informational Influence: Individuals are more likely to adopt a behaviour or belief if they perceive it to be supported by credible information or expert opinion. On social media, companies can leverage informational influence by sharing valuable insights, data, and case studies related to OI.

Normative Influence: Individuals are influenced by social norms and the desire to conform to the expectations of their peers. Companies can leverage normative influence by highlighting the growing adoption of OI practices within their industry or by showcasing successful examples of OI collaborations.

Comparative Influence: Individuals are influenced by observing the behaviours and outcomes of others, especially those they perceive as similar to themselves. Companies can leverage comparative influence by highlighting positive testimonials from other companies or individuals who have successfully implemented OI initiatives.

2.4.5 Institutional Theory

Institutional theory provides a valuable lens for understanding how social media contributes to the institutionalization of OI. This theory suggests that organizations adopt practices and structures considered legitimate and appropriate within their institutional environment, often driven by pressures to conform to industry norms, regulatory requirements, or societal expectations (Jiao et al., 2022). Social media, by providing a platform for observing and discussing industry trends and best practices, can amplify these institutional pressures. The visibility of successful OI implementations on social media can create a "mimetic pressure" (DiMaggio & Powell, 1983), encouraging other organizations to adopt similar practices to gain legitimacy and avoid being perceived as lagging behind. Furthermore, social media can serve as a channel for disseminating normative

pressure, highlighting the ethical or social responsibilities associated with OI and influencing organizational behaviour (Muller & Peres, 2019).

2.4.6 Integration of Perspectives

This multi-theoretical framework, integrating insights from SCT, Network Theory, DOI, SIT, and institutional theory, provides a holistic lens for understanding social media's role in shaping the OI landscape. By considering the various ways social media influences social capital, network structures, information diffusion, and social influence, we can gain a deeper understanding of how this technology shapes the adoption, practice, and impact of OI. This integrated perspective recognizes that social media is not simply a neutral communication channel, but a powerful force that influences individual and organizational behaviours, shaping the dynamics of the innovation ecosystem.
This section has outlined a robust theoretical framework for analysing the complex relationship between social media and OI. The subsequent sections will focus on the specific case of Twitter, justifying its selection as a platform for analysis and outlining the methodological approach used to investigate the OI discourse on this platform.

2.5 Twitter as a Case Study Platform: Justifying the Choice

While social media offers fertile ground for studying OI, this dissertation focuses specifically on Twitter. This choice is driven by several key features that make Twitter particularly well-suited for investigating OI communication and its impact.

2.5.1 Openness and Public Discourse

Twitter's defining characteristic is the predominantly public nature of its conversations. This openness provides researchers with unparalleled access to a vast and diverse dataset of real-time OI discourse. Unlike platforms like Facebook, where algorithmic feeds filter content and private groups restrict access, Twitter's open communication structure allows for a more comprehensive and less biased observation of how individuals and organizations discuss and engage with OI (Kaplan & Haenlein, 2010). This transparency enables the identification of emerging trends, the analysis of communication patterns, and the tracking of the diffusion of OI principles and practices at scale.

2.5.2 Real-Time Data and Dynamic Interactions

The real-time nature of communication on Twitter makes it ideal for capturing the dynamic and evolving nature of OI. Conversations unfold rapidly, allowing researchers to observe the formation of connections, the exchange of ideas, and the responses to emerging events or trends. This real-time data offers valuable insights into the speed of information diffusion, the influence of key actors, and the factors that drive engagement within the OI community. Twitter's concise message format and its focus on rapid information dissemination further enhance its suitability for studying these dynamic interactions.

2.5.3 Global Reach and Diverse Community

Twitter boasts a massive global user base, spanning diverse industries, demographics, and geographical locations (Smith, 2010). This broad reach provides a rich and heterogeneous dataset for investigating OI practices across different contexts. The platform's diverse community includes individuals, companies, researchers, policymakers, and non-profit organizations, offering a multifaceted perspective on the challenges and opportunities of OI. While acknowledging potential limitations regarding representativeness, Twitter's global reach provides a valuable window into the broader OI landscape.

2.5.4 Data Accessibility and Tools Available

Twitter provides access to its data through both its Application Programming Interface (API) and its publicly accessible search interface. The API offers a structured way to retrieve tweets and user information, while the search interface allows for more flexible keyword-based searches. Numerous tools and software packages, including Python libraries like Tweepy and NLTK, facilitate the collection, processing, and analysis of Twitter data. This accessibility of data, combined with the availability of powerful analytical tools, makes Twitter a valuable platform for conducting empirical research on OI communication.

2.6 Conclusion

This chapter has explored the transformative potential of social media as a catalyst for OI. By examining the core tenets of OI, social media's unique affordances, and a robust theoretical framework, we have laid the groundwork for understanding how this technology shapes the innovation landscape. The specific characteristics

of Twitter – its openness, real-time communication, global reach, and data accessibility – make it an ideal platform for investigating OI communication and its impact. The subsequent chapters will delve into the empirical analysis of the OI discourse on Twitter, using quantitative and qualitative methods to uncover the sentiments, linguistic patterns, thematic structures, network dynamics, and communication strategies that characterize this evolving online space.

Chapter 3 Methods

3.1 Introduction

This dissertation aims to comprehensively understand the communication of OI on Twitter, exploring its sentiment, linguistic patterns, thematic structures, network dynamics, and potential impact on organizational performance. The research seeks to unravel how individuals and organizations perceive, express, and engage with this evolving approach to innovation on this influential platform.

Twitter is a particularly relevant platform for this study due to its unique characteristics. The public nature of most Twitter conversations allows for the observation and analysis of OI discourse at a large scale, providing a rich dataset of real-time interactions. Moreover, the platform's global reach, encompassing a diverse range of users across various industries, demographics, and geographical locations, offers a unique opportunity to study the global dynamics of OI communication. Twitter's speed of information dissemination and real-time nature are ideal for capturing the dynamic evolution of conversations and identifying emerging trends in OI discourse.

Understanding how individuals and organizations communicate about OI on Twitter is crucial for both researchers and practitioners. The insights gleaned from this study can inform academics studying innovation and communication, providing a deeper understanding of how social media shapes the discourse around new approaches to innovation. Furthermore, the findings can inform practitioners seeking to leverage Twitter for OI initiatives, providing them with data-driven insights to optimize their communication strategies and engage with relevant communities.

To achieve these objectives, this study employs a mixed-methods approach, combining the strengths of quantitative and qualitative analysis. This approach is particularly valuable for studying complex social phenomena like OI, which

involves multiple actors, diverse perspectives, and intricate social dynamics. The quantitative analysis will leverage a large dataset of tweets related to OI, employing techniques such as sentiment analysis, topic modelling, and network analysis to uncover patterns and trends in the data. This quantitative foundation will then inform the qualitative investigation, using case studies to provide a deeper understanding of the nuances of language, context, and individual experiences. Finally, a predictive model will be developed to explore the potential for forecasting the impact of OI communication on Twitter. The integration of these diverse methods ensures a holistic and multi-faceted view of the OI discourse on Twitter, offering valuable insights for both academic understanding and practical application. The following sections of this chapter will detail the specific data collection process, sampling methodology, and analytical techniques employed in this study, ensuring transparency and replicability of the research.

3.2 Data Collection and Sampling

3.2.1 Selection of Twitter as the Data Source

This study focuses on Twitter as the primary data source for several reasons. The platform's predominantly public nature of conversations allows for the observation and analysis of OI discourse at a large scale, providing a rich dataset of real-time interactions. This openness distinguishes Twitter from platforms like LinkedIn or ResearchGate, which are primarily oriented towards professional networking and academic exchange, respectively.

LinkedIn, while valuable for connecting professionals and sharing industry insights, lacks the same level of open public discourse as Twitter. Its communication is often more restricted, with private messages, closed groups, and an algorithmic feed limiting access to broader conversations. Similarly, ResearchGate, while a valuable platform for sharing scientific publications and engaging in academic discussions, has a more specialized focus and limited real-time interaction compared to Twitter. The public nature of Twitter conversations enables the capture of a more diverse range of perspectives and opinions, making it more suitable for studying the broader dynamics of OI communication.

Furthermore, the real-time nature of Twitter makes it ideal for capturing the dynamic and evolving nature of OI discourse. The platform's fast-paced communication style and its focus on immediate sharing allow for the analysis of

trends as they emerge, the tracking of information diffusion, and the study of the formation and evolution of online communities.

Finally, the availability of data through Twitter's API and its publicly accessible search interface makes it a highly accessible platform for collecting large-scale data for research. Numerous tools and software packages, including Python libraries like Tweepy and NLTK, facilitate the collection, processing, and analysis of Twitter data. This accessibility, combined with Twitter's openness, real-time nature, and global reach, makes it the ideal platform for this study.

3.2.2 Data Collection Process and Challenges

The initial plan for this research was to collect data exclusively through the Twitter API. However, the limitations of the API, particularly its restricted access to historical data and its current prohibitive pricing structure, made this approach unfeasible. While the Twitter API once offered valuable functionalities for researchers, recent changes have significantly limited its utility. The current version restricts historical data access to a negligible amount, allowing the download of only a few thousand tweets per month. Furthermore, the pricing for accessing even this limited amount of data is prohibitively expensive for research purposes.

An additional limitation of the Twitter API is that the follower count provided is not the historical count at the time of the tweet's publication but rather the real-time count at the moment of data retrieval. This discrepancy introduces a potential bias as the follower count could have changed significantly since the tweet was posted. To avoid this bias and maintain the integrity of the data, the decision was made to focus on a more recent timeframe for data collection.

Due to these limitations, this research employed a web scraping approach using a custom-built software called twscrape. Twscrape utilizes a registry of Twitter user credentials created specifically for this research project to bypass the login wall that now restricts access to certain endpoints. The software directly queries Twitter's GraphQL endpoints, which were once publicly accessible but now require authentication, to retrieve the desired data. This approach allows for the collection of a larger and more comprehensive dataset than is currently possible through the Twitter API.

However, web scraping presents its own set of challenges. Twitter actively implements measures to prevent automated data collection, including rate limiting (restricting the number of requests per minute or hour) and IP blocking

(preventing access to the website from specific IP addresses). To address these challenges, twscrape was meticulously designed to mimic human user behaviour and to respect Twitter's terms of service.

The software incorporates random delays between requests, simulating the natural browsing behaviour of a human user. It also rotates IP addresses after a certain number of requests, making it more difficult for Twitter to detect and block the scraper. These measures ensured that the scraper operated within Twitter's acceptable usage guidelines and minimized the risk of being blocked.

Despite these precautions, the dynamic nature of Twitter's website structure, with frequent updates and changes, required ongoing adaptation and maintenance of twscrape to ensure its continued functionality.

Two distinct datasets were collected for this research, reflecting the evolving data collection process:

- Dataset 1: This dataset was collected between January 1st, 2024, and March 31st, 2024, using a random sample of tweets. This approach was initially adopted due to the need to manage the volume of data collected during the initial stages of the research project, as the scraper was still under development.
- Dataset 2: This dataset encompasses a complete capture of all relevant tweets from April 1st, 2024, to July 31st, 2024. This comprehensive collection was possible after refining twscrape and ensuring its consistent and reliable performance.

3.2.3 Sampling Methodology

Dataset 1 (Random Sample): To ensure a representative sample for the initial three months, a simple random sampling method was employed. This involved randomly selecting 30 days from the calendar within the period of January 1st, 2024, to March 31st, 2024. This sampling ensures that each day within the three-month period has an equal chance of being selected, minimizing potential bias in the sample. All tweets posted on those selected days were then retrieved. This approach, supported by the work of Kim et al. (2018), ensures that the sample reflects the overall patterns and trends in the Twitter data for that period. Kim et al. (2018) demonstrate that a ratio of 1 day out of every 3 days is sufficient to obtain a representative sample on this platform.

Dataset 2 (Complete Capture): For the subsequent four-month period (April 1st, 2024, to July 31st, 2024), all relevant tweets were collected. This complete capture

approach provides a more detailed and comprehensive perspective on the OI discourse during this specific timeframe but limits the generalizability of the findings to this specific period.

The use of these two distinct datasets has implications for the generalizability of the findings. While Dataset 1, with its random sampling, offers a broader representation of the overall OI discourse during the first three months of 2024, its limited size might not capture the full diversity of perspectives and trends within this conversation. Dataset 2, on the other hand, provides a more detailed and nuanced view of the OI discourse during a specific four-month period, but its findings might not be generalizable to other periods, as the nature and dynamics of online conversations can shift over time. Therefore, it's important to acknowledge that these datasets represent snapshots of a dynamic and evolving online conversation, and the conclusions drawn from their analysis should be interpreted within the context of their specific limitations.

3.2.4 Inclusion and Exclusion Criteria

This study focuses specifically on discussions related to OI on Twitter. To achieve this focus, only tweets containing the hashtags #Innovation and #OpenInnovation were included in the analysis. These hashtags were chosen due to their established prominence as markers of OI discussions on Twitter (Saura et al., 2023). However, it's important to acknowledge that relying solely on these hashtags might introduce a selection bias, as users could engage in discussions about OI using alternative terms or phrases. For example, users might use terms like "co-creation," "crowdsourcing," "open source," or "collaborative innovation" to discuss concepts related to OI without explicitly using the chosen hashtags. Future research could explore the use of a wider range of keywords and phrases, or alternative methods for identifying OI-related tweets, to address this potential limitation and capture a more comprehensive picture of the OI discourse on Twitter.3.3 Data Analysis Methods

This section delves into the specific analytical techniques employed in this study to examine the collected Twitter data. The research utilizes a multi-faceted approach, combining various methods to provide a comprehensive and nuanced understanding of the OI discourse on Twitter.

3.2.5 Pre-processing and Data Cleaning

Before conducting any analysis, the raw tweet data underwent several pre-processing steps to ensure data quality and prepare the text for analysis:

- Removing Irrelevant Elements: URLs, mentions (@mentions), hashtags, and emojis were removed from the tweet text. This step ensures that the analysis focuses solely on the meaningful textual content of the tweets, preventing these elements from introducing bias into subsequent analyses. For example, removing hashtags ensures that the topic modelling algorithm isn't unduly influenced by the explicit topic labels users have chosen.
- Removing Retweet Markers: The "RT" marker, indicating a retweet, was removed to ensure that the analysis focuses on original content and to prevent retweets from disproportionately influencing the data. This step helps to distinguish between original thoughts and ideas and those that are simply re-shared by other users.
- Removing Punctuation: All punctuation marks were removed to prepare the text for analysis. Removing punctuation helps to standardize the text and prevent these symbols from interfering with text processing algorithms.
- Converting to Lowercase: All text was converted to lowercase to ensure case-insensitive matching during analysis. This step ensures that words with different capitalization (e.g., "Innovation" and "innovation") are treated as the same word, preventing variations in capitalization from affecting the results.

These pre-processing steps are essential for cleaning and standardizing the text data, ensuring that the subsequent analyses focus on the meaningful content of the tweets and are not influenced by irrelevant elements or variations in formatting (Garg & Sharma, 2022).

3.2.6 Sentiment Analysis

Sentiment analysis was employed to assess the emotional tone and polarity of the collected tweets, providing insights into the overall sentiment associated with OI and innovation on Twitter. The analysis specifically targeted the sentiment expressed within the communication itself (i.e., the tweet content) and did not consider the sentiment of replies or reactions to these tweets.

This study utilized VADER, a lexicon and rule-based sentiment analysis tool specifically designed for analysing social media text (Hutto & Gilbert, 2014). VADER is particularly well-suited for this task as it effectively handles informal language, slang, and emoticons commonly used in social media communication. Its effectiveness in capturing sentiment in social media contexts has been demonstrated in previous research (Hromic & Hayes, 2019).

VADER assigns a "compound sentiment score" to each tweet, ranging from -1 (most negative) to +1 (most positive), representing the overall sentiment expressed in the tweet. This score reflects the intensity and balance of positive and negative expressions within the text.

While sentiment analysis provides a valuable overview of the emotional tone surrounding OI on Twitter, it's important to acknowledge its limitations. The informal nature of social media language, including the frequent use of sarcasm, irony, and humor, can introduce biases and affect the accuracy of sentiment classification. To mitigate these limitations, the findings of the sentiment analysis will be interpreted in conjunction with other analytical methods, including qualitative analysis of tweet content and the exploration of contextual factors.

3.2.7 Topic Modelling with LDA

Topic modelling was employed to identify the prominent themes and topics discussed within the OI discourse on Twitter. This study utilized Latent Dirichlet Allocation (LDA), a generative probabilistic model for automatically discovering underlying topics within a collection of documents (Blei et al., 2002). LDA assumes that documents are mixtures of topics, and topics are probability distributions over words. The model aims to identify the latent topics that best explain the observed word distributions within the corpus of documents.

LDA is particularly well-suited for analysing Twitter data due to its robustness to data sparsity and its ability to capture the semantic meaning of words within short and noisy text. Unlike other topic modelling methods, such as Non-negative Matrix Factorization (NMF), which often perform better with longer and more structured documents, LDA has been shown to be effective in identifying meaningful topics from the short and informal text typical of tweets.

Several parameters influence the performance of LDA, including:

- Number of Topics (k): This parameter determines the number of latent topics to be extracted from the corpus. The optimal number of topics was determined using the "elbow method," which involves plotting the coherence

score of the model against different values of k. The elbow point on the plot, where the coherence score begins to plateau, indicates the optimal number of topics. The final k value chosen for this study was 8, based on the analysis of coherence scores.

- Alpha and Beta: These parameters control the distribution of topics within documents and the distribution of words within topics, respectively. Alpha influences the sparsity of topic assignments – a lower alpha value results in documents being assigned to fewer topics, while a higher alpha value allows for more diverse topic mixtures. Beta influences the concentration of words within topics – a lower beta value results in topics being defined by a smaller set of highly probable words, while a higher beta value allows for a more even distribution of words within topics. The specific alpha and beta values used in this study were determined empirically, based on the characteristics of the dataset and the desired level of topic granularity.

The quality of the LDA model was assessed using coherence scores. Coherence measures the semantic similarity between the top words associated with each topic. A higher coherence score indicates that the words within a topic are more semantically related, resulting in a more interpretable and meaningful topic.

The interpretation of the extracted topics involved analysing the top words associated with each topic and identifying a representative label that captures the essence of the theme. This process was guided by a qualitative examination of tweets belonging to each topic to ensure the accuracy and relevance of the assigned labels.

While LDA offers a valuable approach to identifying the key themes within the OI discourse on Twitter, it's essential to acknowledge its limitations. The interpretation of topics can be subjective, and different researchers might arrive at slightly different interpretations of the same topic based on their understanding of the domain and the context of the data. To mitigate this subjectivity, the topic interpretations were validated through discussions with experts in the field of OI and by cross-referencing the findings with existing literature.

3.2.8 Network Analysis

Network analysis was employed to visualize and analyse the relationships between actors engaged in the OI discourse on Twitter. This approach provides valuable insights into the structure, dynamics, and key players within the OI ecosystem on the platform.

The network was constructed using user mentions as indicators of relationships. Each Twitter user who either posted a tweet or was mentioned in a tweet was represented as a node in the network. A directed edge was created from user A to user B if user A mentioned user B in a tweet. This approach captures the flow of communication and attention within the OI community on Twitter.

Two separate networks were created for this analysis:

- Combined OI Network: This network includes all tweets from Dataset 2 that contain either #innovation or #openinnovation. This network provides a broad overview of the interactions among users engaging in any aspect of innovation discourse, regardless of their explicit focus on OI.
- Innovation Network: This network includes only tweets from Dataset 2 that contain #innovation (excluding those that also contain #openinnovation). This network enables a more focused analysis of communication patterns among users who are not explicitly engaging in OI discussions, allowing for comparisons between the two groups.

To enhance clarity and interpretability and to focus on the most significant connections, the networks were simplified by:

- Removing Self-Loops: Edges connecting a node to itself, which do not contribute to understanding interactions between different users, were removed.
- Removing Single Nodes: Nodes with no connections, representing users isolated from the larger network, were removed.

After simplification, key network metrics were calculated using the NetworkX library in Python:

- Degree Centrality: This metric measures the number of connections a node has, reflecting its prominence within the network. Nodes with high degree centrality are frequently mentioned by others and have a broader reach.
- Betweenness Centrality: This metric measures how often a node lies on the shortest path between other nodes. Nodes with high betweenness centrality act as bridges or connectors between different parts of the network, facilitating the flow of information and potentially influencing communication across diverse communities.
- Clustering Coefficient: This metric measures the degree to which a node's neighbours are connected to each other. High clustering coefficients indicate tightly knit groups or communities, where users are more likely to interact with each other and share information within their group.

When plotting the graphs, to make the graph easier to plot and visually interpret, it was further simplified by:

- Removing Low-Weight Edges: Edges with a weight less than 10, indicating infrequent interactions, were removed. This threshold was chosen based on the distribution of edge weights to significantly reduce network density without eliminating crucial connections.
- Retaining Top 1000 Nodes for the Innovation Network: Due to its larger size and complexity, only the top 1000 nodes based on degree centrality (explained below) were retained in the Innovation Network. This simplification reduces visual clutter while preserving the core structure and highlighting the most influential actors.

The networks were visualized using Gephi, an open-source software for network analysis and visualization, to facilitate exploration of network structures, identify key actors, and analyse community formations.

3.2.9 Language-Based Cluster Analysis

This study goes beyond analysing explicit mentions of OI by examining language patterns to identify communities of users with similar communication styles, regardless of their use of specific hashtags. This language-based cluster analysis was conducted using Bidirectional Encoder Representations from Transformers (BERT) embeddings and the k-means clustering algorithm.

BERT is a powerful language model pre-trained on a massive dataset of text and code, enabling it to understand the context and meaning of words in a sentence (Devlin et al., 2019). For each tweet, BERT generates a 768-dimensional embedding vector that represents the tweet's semantic content.

K-means clustering is an unsupervised machine learning algorithm that groups data points into k clusters based on their similarity. In this case, the data points are the BERT embeddings representing the tweets. The algorithm iteratively assigns data points to clusters, aiming to minimize the distance between points within the same cluster and maximize the distance between points in different clusters.

Cosine similarity was used as the distance metric for comparing tweet embeddings. This metric measures the angle between two vectors in a multi-dimensional space, ignoring their magnitude. Cosine similarity is well-suited for comparing text embeddings as it focuses on the semantic relationships between

words rather than their absolute values, which can vary depending on the length of the text.

The elbow method was used to determine the optimal number of clusters (k) for the k-means algorithm. The elbow method involves plotting the within-cluster sum of squares against different values of k and identifying the point where the plot bends, resembling an elbow. This point suggests a balance between minimizing the within-cluster variance and avoiding an overly complex model with too many clusters.

After the clusters were formed, they were manually inspected by examining a sample of tweets within each cluster to identify common themes, language patterns, and user types. Based on this analysis, each cluster was assigned a descriptive label that captures its essence. This approach reveals how language use and communication styles can connect users with shared interests in OI, even if they do not explicitly use the same hashtags.

3.2.10 Predictive Modelling

To understand the factors influencing the reach of OI-related communication on Twitter, a predictive model was developed to forecast the viewCount of tweets. viewCount is a metric readily available in Twitter data that represents the number of times a tweet has been viewed, serving as a proxy for the tweet's overall visibility and potential impact.

This research employs a feedforward neural network for predictive modelling. This architecture, well-suited for modelling complex non-linear relationships, processes information in a hierarchical manner, extracting increasingly abstract features through multiple layers of interconnected nodes (neurons). The model takes a combination of numerical and categorical variables as input and outputs a single value representing the predicted viewCount of a tweet.

The neural network model was designed with five dense layers, each with a decreasing number of neurons (512, 256, 128, 64, 32), followed by a single output layer. The input layer receives four distinct input features:

- *log(user.followersCount)*: The logarithm of the user's follower count was used as a proxy for reach and influence. The logarithmic transformation mitigates the impact of outliers and normalizes the distribution of follower counts, which is typically heavily skewed.
- *weekday*: This categorical variable represents the day of the week when the tweet was published. It is one-hot encoded, transforming the categorical

information into a binary vector representation, allowing the network to process this information effectively.

- *hour*: Similar to weekday, this categorical variable represents the hour of the day when the tweet was published and is also one-hot encoded.
- *embeddings*: This variable represents the 768-dimensional embedding vector generated by BERT for each tweet. The embedding captures the semantic content of the tweet, allowing the network to learn from the language and context of the message.

The choice of these features was based on their potential relevance for predicting viewCount, informed by previous research and descriptive analyses of the dataset.

The model was trained using a supervised learning approach, where the algorithm learns from labeled examples to map the relationship between input features and the target variable (viewCount). The dataset was randomly split into a training set (80% of the data) and a test set (20% of the data). The model's parameters were adjusted during training to minimize the difference between the predicted and actual viewCount values in the training set.

Several evaluation metrics were used to assess the model's performance:

- R-squared (R^2): This metric measures the proportion of variance in the target variable that is predictable from the independent variables. A higher R^2 value indicates a better fit.
- Mean Squared Error (MSE): This metric measures the average squared difference between the predicted and actual values. A lower MSE indicates better accuracy.
- Mean Absolute Error (MAE): This metric measures the average absolute difference between the predicted and actual values. A lower MAE indicates better accuracy.

The model achieved an impressive R^2 value of 0.60 on the test set, indicating its ability to explain 60% of the variance in viewCount. This performance demonstrates the potential of using deep learning techniques to predict the reach and impact of OI communication on Twitter.

3.3 Analysis Tools

This study employed a variety of software tools and libraries, carefully selected for their specific functionalities and suitability in addressing the research questions. Data collection was conducted using twscrape, an open-source web

scraping software developed specifically for this project. twscrape bypasses Twitter's current limitations on historical data access by utilizing a registry of user credentials to access previously public endpoints, now restricted by a login wall. The software directly queries Twitter's GraphQL API, retrieving the data in a structured format. The accuracy of twscrape was validated through manual inspection of a subset of the downloaded tweets. While a small number of irrelevant tweets were found, this discrepancy is attributed to changes in Twitter's search algorithm, which now includes promotional content in search results.

After data collection, the raw data was cleaned and pre-processed using Python, a versatile programming language widely used in data analysis. Python's flexibility and its extensive libraries make it an ideal choice for manipulating and analysing large datasets. Pandas, a powerful data manipulation library in Python, was employed for data cleaning, filtering, transformation, and organization, enabling efficient handling of the large Twitter dataset.

Sentiment analysis was conducted using VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon and rule-based sentiment analysis tool specifically tailored for social media text (Hutto & Gilbert, 2014). VADER's strength lies in its ability to effectively handle informal language, slang, and emoticons, common in Twitter communication, ensuring a more accurate assessment of sentiment compared to tools designed for formal text (Hromic & Hayes, 2019). Its lexicon incorporates sentiment-related words and phrases commonly used in social media, and its rule-based system accounts for grammatical and syntactical conventions of online language, such as capitalization, punctuation, and emojis. This makes VADER particularly well-suited for analyzing tweets related to open innovation, where the language can be informal and expressive.

For topic modelling, the Gensim library in Python was employed to implement Latent Dirichlet Allocation (LDA). Gensim's selection was driven by its efficiency in handling large text corpora and its comprehensive suite of functionalities for topic model evaluation, including coherence scores, which provide a measure of topic interpretability. LDA was chosen for its ability to uncover underlying thematic structures within the tweet dataset, revealing the key topics and trends discussed within the OI discourse. Other topic modelling methods, such as Non-negative Matrix Factorization (NMF), were considered but deemed less suitable for the short and often noisy text characteristic of tweets. The number of topics was determined using the coherence score, balancing model complexity with interpretability.

Social Network Analysis (SNA) was performed using NetworkX, a Python library specifically designed for creating, manipulating, and studying the structure, dynamics, and functions of complex networks. NetworkX facilitates the calculation of key network metrics, such as degree centrality, betweenness centrality, and clustering coefficient, which are essential for understanding the relationships and influence of different actors within the OI ecosystem on Twitter. Gephi, an open-source network visualization software, was used to generate interactive visualizations of the networks, providing a visual representation of the connections and communities within the OI discourse. The choice of user mentions as the basis for network construction was motivated by their explicit representation of interactions and relationships between users.

Language-based cluster analysis was conducted using BERT (Bidirectional Encoder Representations from Transformers) embeddings, generated using the transformers library from *Hugging Face*, and the k-means clustering algorithm from the scikit-learn library. BERT embeddings, pre-trained on a massive text and code dataset, capture the semantic meaning of words within their context, offering a richer representation of tweet content compared to traditional word embeddings (Devlin et al., 2019). The use of BERT was motivated by its demonstrated effectiveness in capturing nuanced meanings and contextual information in short texts, making it well-suited for analyzing tweets. K-means clustering, a widely used unsupervised learning algorithm, was selected for its ability to group similar tweets based on their BERT embeddings. Cosine similarity was chosen as the distance metric for comparing embeddings due to its focus on the angular difference between vectors, capturing semantic relationships more effectively than Euclidean distance in high-dimensional spaces. The optimal number of clusters was determined using the elbow method, balancing cluster homogeneity with model complexity. This method provides insights into different communication patterns and styles within the OI discourse beyond the limitations of hashtag-based analyses. While other clustering methods were considered, k-means was deemed appropriate due to its efficiency and interpretability.

Finally, a feedforward neural network model was developed using Keras, a high-level deep learning API running on top of TensorFlow, to predict the viewCount of tweets. Keras was chosen for its user-friendly interface, ease of customization, and integration with TensorFlow, which provides the underlying computational framework for deep learning. A neural network was selected for its ability to

model complex, non-linear relationships between variables, which was deemed necessary to capture the intricate interplay of factors influencing tweet visibility. The model architecture, including the number of layers, neurons per layer, activation functions, and input features (log(user.followersCount), weekday, hour, and BERT embeddings) was determined through experimentation and evaluation based on performance metrics such as R-squared, Mean Squared Error (MSE), and Mean Absolute Error (MAE). While other machine learning models, such as linear regression, were considered, they were found to be less effective in capturing the nuances of the data and achieving satisfactory predictive accuracy. The choice of input features was guided by their theoretical relevance and potential influence on tweet visibility, informed by prior research and exploratory data analysis. The omission of image-based features was primarily due to the focus of this research on textual content and the complexities associated with processing and integrating visual data into the model.

3.4 Sample Description and Statistical Analyses

3.4.1 Dataset Overview

This section provides a descriptive overview of the collected Twitter data, highlighting key characteristics of the two datasets. Dataset 1, a random sample of tweets collected from January 1st, 2024, to March 31st, 2024, contains 57,266 tweets. Of these, 57,121 tweets were associated with the #innovation hashtag and 145 with the #openinnovation hashtag. Dataset 2, encompassing all relevant tweets from April 1st, 2024, to July 31st, 2024, totals 264,784 tweets, with 264,135 associated with #innovation and 649 with #openinnovation. This significant difference in volume across both datasets suggests that discussions surrounding "innovation" are considerably more prevalent on Twitter compared to those explicitly labeled as "open innovation," potentially indicating a broader usage and awareness of the general term "innovation."

Table 3.2 provides insights into the reach and potential influence of users participating in these discussions by presenting descriptive statistics for the number of followers per user in each dataset.

Table 3.2: User Follower Count Statistics

Dataset	*Hashtag*	*Mean*	*Median*	*Standard Deviation*
Dataset 1	#innovation	5,995	283	103,266
Dataset 1	#openinnovation	9,526	1,098	37,747
Dataset 2	#innovation	5,995	285	103,266

Dataset 2	#openinnovation	9,526	1,084	37,747

The table shows that users engaging in discussions related to #openinnovation tend to have a larger following compared to those discussing #innovation, suggesting that the OI audience might be comprised of more influential users, such as industry experts, thought leaders, or organizations with a larger online presence.

Analysis of the temporal distribution of tweets reveals a consistent weekly cyclical pattern in the volume of tweets related to both #innovation and #openinnovation.

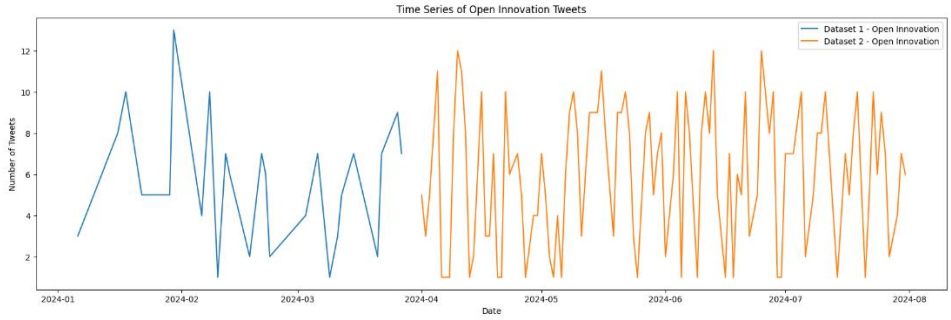

Figure 1: Time series of #openinnovation tweets

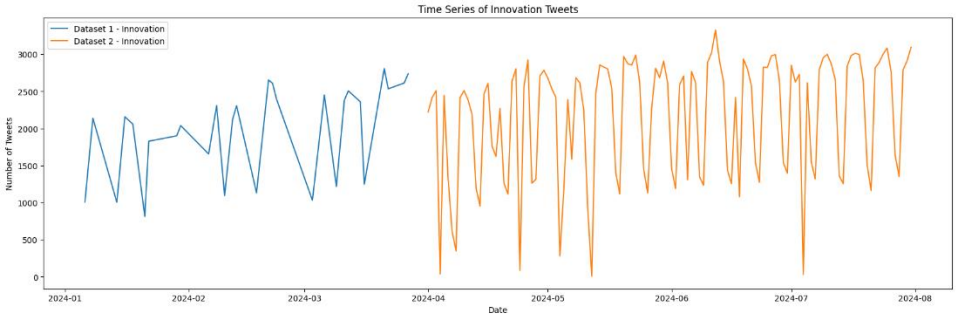

Figure 2: Time series of #innovation tweets

The visualization, as shown above, illustrates a clear increase in activity during weekdays, peaking around mid-week, followed by a decline in volume during weekends. This pattern suggests that conversations surrounding OI and innovation on Twitter are influenced by work schedules and online activity patterns, with higher engagement during weekdays when users are more likely to be actively working and participating in online discussions.

3.4.2 Statistical Comparisons

To investigate whether there were significant differences in user engagement between tweets discussing general innovation and those specifically related to OI, a t-test was conducted on the five engagement metrics: *replyCount*, *retweetCount*, *likeCount*, *quoteCount*, and *bookmarkedCount*. The results, presented in Table 1, confirm a statistically significant difference in engagement between the two groups, with higher engagement levels for tweets containing the #openinnovation hashtag across all five metrics.

Table 1: T-test Results for Engagement Metrics

Metric	*T-statistic*	*P-value*	*Degrees of Freedom*
replyCount	2.69	0.007	647
retweetCount	2.46	0.014	647
likeCount	2.39	0.017	647
quoteCount	2.12	0.034	647
bookmarkedCount	2.58	0.010	647

While the t-test reveals a statistically significant difference, it's important to consider the effect size to understand the practical significance of this difference. In this case, Cohen's d, a commonly used measure of effect size, was calculated for each metric. The effect sizes ranged from small to moderate, indicating that while the differences in engagement are statistically significant, the magnitude of these differences might not be substantial in practical terms.

3.5 Conclusion

This chapter has provided a thorough account of the methodology employed in this study, detailing the rationale for each methodological choice, the specific tools and techniques utilized, and the challenges encountered during the data collection and analysis process. The combination of quantitative and qualitative methods, along with a predictive modelling approach, allows for a comprehensive exploration of the OI discourse on Twitter, encompassing both the broad patterns and the nuanced details of online communication.

The meticulous data collection process, using a web scraper (twscrape) to overcome limitations of the Twitter API, ensured the retrieval of a rich and

comprehensive dataset. The implementation of two distinct sampling methodologies – a random sample for Dataset 1 and a complete capture for Dataset 2 – provides both a representative overview and a detailed perspective on the OI discourse. The selection of specific hashtags, while acknowledging potential limitations, provides a focused lens for analysing relevant conversations.

The rigorous application of various analytical techniques, including sentiment analysis, topic modelling, network analysis, language-based cluster analysis, and predictive modelling, allows for a multifaceted understanding of the OI discourse on Twitter. These analyses will be explored in detail in the following chapters, uncovering the sentiments, language, thematic structures, network dynamics, and potential impact of OI communication on this influential platform.

Chapter 4 Sentiment Analysis and OI Keywords

This chapter delves into the heart of OI communication on Twitter, examining the sentiment and linguistic patterns that characterize discussions surrounding #openinnovation. Understanding how individuals and organizations perceive and communicate about OI is crucial for its successful implementation and diffusion (West & Bogers, 2014). While prior research has extensively explored the theoretical frameworks and practical implications of OI (Chesbrough & Appleyard, 2007; Helfat, 2006), less attention has been devoted to the specific language and emotional tone employed in online discussions.

This chapter aims to bridge this gap by investigating the sentiment and linguistic features that characterize tweets related to OI. Sentiment analysis will provide insights into the overall emotional valence associated with #openinnovation, revealing whether discussions tend towards positive, negative, or neutral perspectives. This analysis will also explore temporal trends in sentiment, identifying any shifts in emotional tone over time.

Complementing sentiment analysis, linguistic analysis will uncover distinctive language patterns employed in #openinnovation discussions. Drawing on established keywords related to OI and value capture (Toroslu et al., 2023), this analysis will identify the prevalence of key terms and phrases, providing insights into the specific themes, concepts, and practices that dominate the discourse.

By integrating these two analytical lenses, this chapter aims to provide a deeper understanding of how OI is perceived and communicated on Twitter. This is particularly important considering the increasing role of social media platforms in shaping public opinion and influencing the adoption of new practices (Hitchen et al., 2017). The findings will contribute to a more nuanced understanding of the

social dynamics surrounding OI, offering valuable insights for both researchers and practitioners.

The chapter is structured as follows: Section 4.2 presents the methodology and results of the sentiment analysis, exploring the emotional landscape of #openinnovation. Section 4.3 delves into the linguistic analysis, uncovering distinctive language patterns associated with the hashtag. Finally, Section 4.4 discusses the integrated findings, highlighting key insights and implications for future research and practice.

4.1 Sentiment Analysis

This section delves into the findings of the sentiment analysis conducted on Dataset 2, aiming to uncover the emotional tone associated with discussions surrounding #innovation and #openinnovation. Understanding the sentiment expressed towards these concepts can provide valuable insights into how individuals and organizations perceive these phenomena and engage in related conversations online.

Dataset 2 was chosen because it contains tweets for each day in the interval. This allows us to more easily discern trends in the data.

4.1.1 Methodology

Sentiment analysis is widely employed in social media research to gauge the emotional tone and polarity of text data (Culotta & Cutler, 2016). This analysis relies on natural language processing (NLP) algorithms and lexicons to classify text as expressing positive, negative, or neutral sentiments.

For this research, VADER was chosen as the sentiment analysis tool. VADER is a lexicon and rule-based sentiment analysis tool specifically designed for social media text (Hutto & Gilbert, 2014). It is particularly effective in handling the informal language, slang, and emoticons often used on Twitter. VADER has been shown to be effective in capturing sentiment in social media contexts (Hromic & Hayes, 2019), making it a suitable choice for analysing tweets related to innovation.

VADER assigns a "compound sentiment score" to each tweet, ranging from -1 (most negative) to +1 (most positive). This score reflects the overall sentiment of the tweet, taking into account the intensity and balance of positive and negative expressions.

4.1.2 Sentiment Distribution

The sentiment analysis revealed a predominantly positive sentiment score associated with both #innovation and #openinnovation across Dataset 2. This suggests that these concepts are generally communicated with a positive tone in tweets.

Table 2: Descriptive statistics for the compound sentiment score

Metric	*#innovation*	*#openinnovation*
Count	264135	649
Mean	0.659868	0.629782
Standard Dev.	0.251321	0.287291
Min	-0.9794	-0.6705
25%	0.5106	0.3818
50%	0.7184	0.7269
75%	0.8519	0.8481
Max	0.9986	0.9740

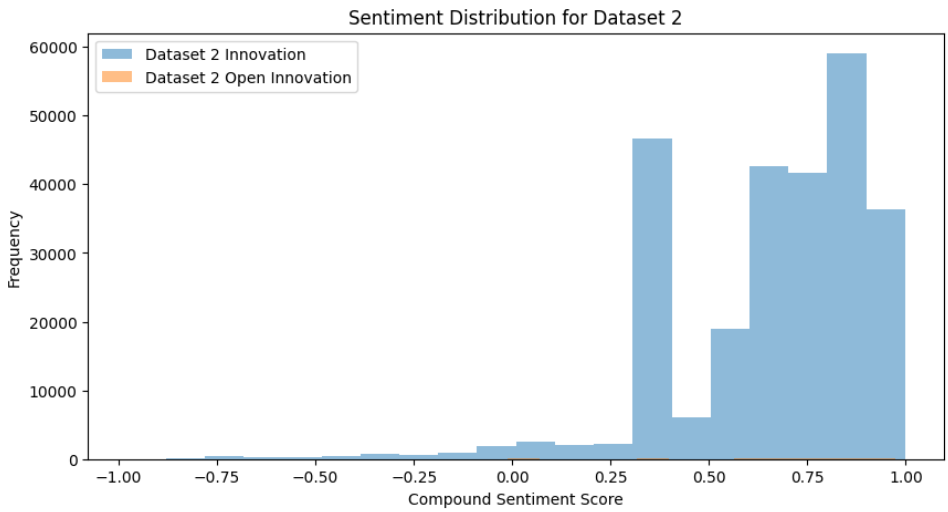

Figure 3: Distribution of compound sentiment scores for tweets containing #innovation (blue) and #openinnovation (orange) in Dataset 2. The compound sentiment score ranges from -1 (most negative) to +1 (most positive).

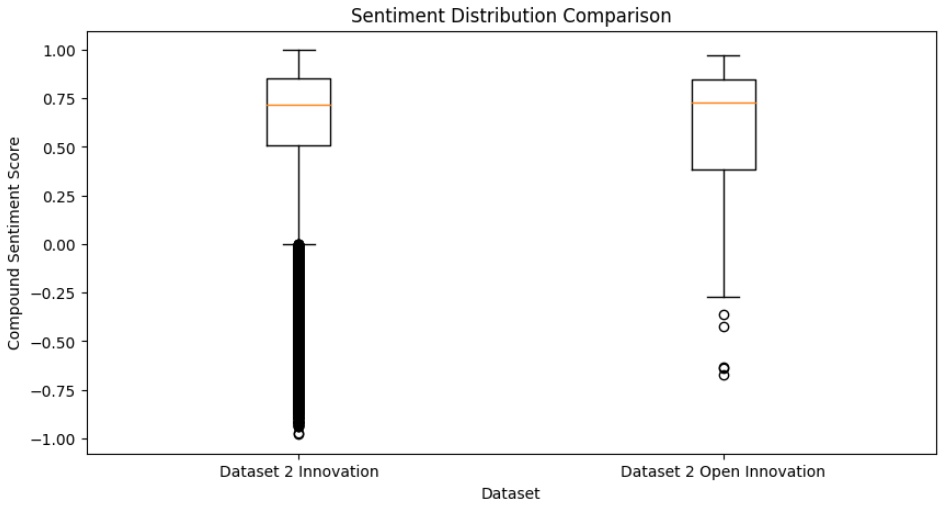

Figure 4: Comparison of compound sentiment score distributions for tweets containing #innovation and #openinnovation in Dataset 2. The box plot shows a slightly higher median and a narrower interquartile range for #innovation.

It's important to clarify that this sentiment analysis focuses on the sentiment expressed within the OI/innovation communication itself (i.e., the tweet content), and not on the sentiment expressed in replies or reactions to those tweets.

Discussion

As demonstrated by the table and visualizations, the mean compound sentiment scores for both hashtags are positive, exceeding 0.6 for both. This suggests that innovation and OI are generally communicated with a positive tone in tweets. However, it's important to acknowledge that the nature of social media, particularly its tendency to amplify positive messages and successes, might contribute to this positive bias. Future research could explore the relationship between the sentiment expressed in tweets and the actual sentiment of the broader online community towards OI.

It is worth noting that the standard deviations for both hashtags are relatively high, indicating a range of sentiment expressions within each category. A direct comparison using the box plot confirms that while both hashtags are associated with positive sentiment, #innovation exhibits a slightly higher median sentiment score and a narrower interquartile range, suggesting a more pronounced and consistent positive sentiment. #OpenInnovation, while still positive overall, shows more variability in sentiment expressions. This greater variability in sentiment for #openinnovation compared to #innovation might be attributed to

the inherent complexities of this approach, which often involves navigating challenges related to intellectual property, collaboration dynamics, and the integration of external knowledge (Chesbrough & Appleyard, 2007). These complexities could lead to a wider range of opinions and experiences, resulting in a more diverse range of tones used when communicating these topics.

4.1.3 Temporal Trends in Sentiment

Analysing the sentiment expressed towards #innovation and #openinnovation over time reveals insightful patterns.

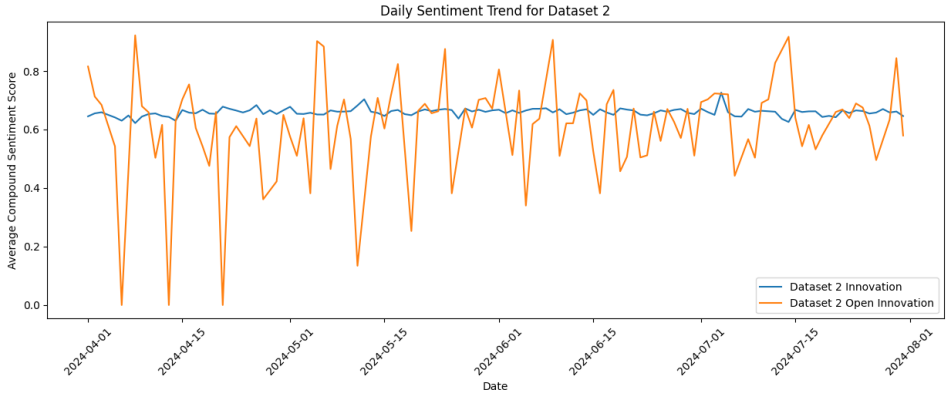

Figure 5: Daily sentiment trend for dataset 2

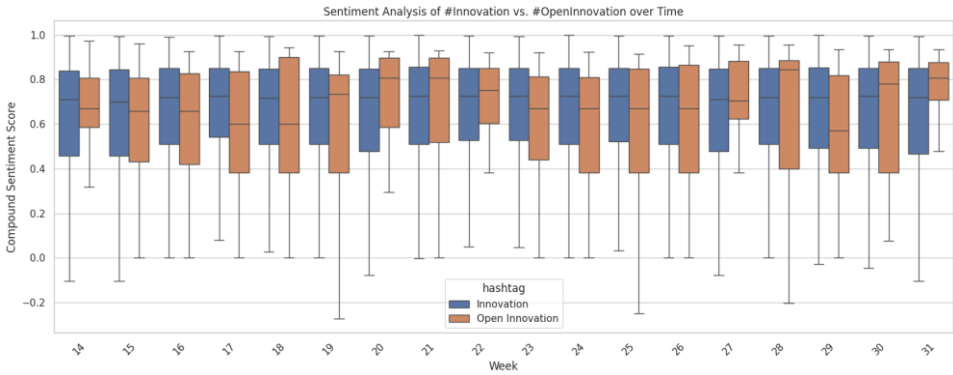

Figure 6: Box plot of sentiment on a weekly basis

Discussion

The temporal analysis highlights the dynamic nature of sentiment on Twitter. The sentiment towards #innovation shows more noticeable fluctuations over time, with some days exhibiting spikes in positive sentiment. These fluctuations are

likely driven by specific events or news related to innovation that generated excitement or enthusiasm within the Twitter community. For instance, a peak in positive sentiment for #innovation on April 12th coincided with the release of a highly anticipated tech product, as evidenced by numerous tweets celebrating the launch.

Conversely, the sentiment towards #openinnovation, while still generally positive, demonstrates greater stability over time, with fewer pronounced spikes or dips. This finding suggests a more consistent and potentially community-driven conversation, less influenced by external factors or daily events. This aligns with research that highlights the enduring nature of OI principles and the focus on long-term collaborative practices within dedicated communities (Bogers et al., 2017).

4.2 Analysis of OI Keywords

This section delves into the linguistic features that characterize tweets related to #innovation and #openinnovation. By examining the prevalence of specific keywords derived from a robust theoretical framework, this analysis aims to unveil distinctive patterns in language use and shed light on the thematic focus and conceptual underpinnings of these discussions on Twitter.

4.2.1 Methodology

The linguistic analysis employed a keyword-based approach, leveraging a predefined list of terms associated with OI. The keyword list was extracted from a recent literature review by Toroslu et al. (2023), ensuring a strong theoretical foundation and alignment with established academic discourse. This specific list was chosen because it provides a comprehensive and nuanced overview of the diverse terminology used in the field of OI.

To ensure consistency and accuracy, the following steps were taken:

- Data Preparation: The datasets were pre-processed by converting all text to lowercase to ensure case-insensitive matching of keywords.
- Keyword Counting: The frequency of each keyword was counted in the text of every tweet within each dataset.
- Relative Frequency Calculation: The relative frequency of each keyword was calculated by dividing its count by the total number of words in the tweet. This normalization accounts for variations in tweet length and provides a comparable measure across tweets.

4.2.2 Keyword Frequencies

Table 3: Top 10 OI and Value Capture Keywords in Dataset 2

Keyword	*#openinnovation (MRF)*	*#innovation (MRF)*
open innovat	0.003488	6.583e-06
co-creat	0.000297	1.241e-05
crowdsourcing	0.000341	3.988e-06
hackathon	0.000256	0.000123
cooperat	0.000191	5.046e-05
open source	8.006e-05	2.525e-05
joint venture	-	6.130e-06
industry collaborat	-	4.509e-06
innovation partnership	-	2.949e-06
co-innovat	-	2.207e-06
value	0.000539	0.000516
destruct	-	8.270e-06
appropriat	-	5.386e-06
leakage	-	1.679e-06
misappropriat	-	1.967e-07
spillover	-	1.352e-07

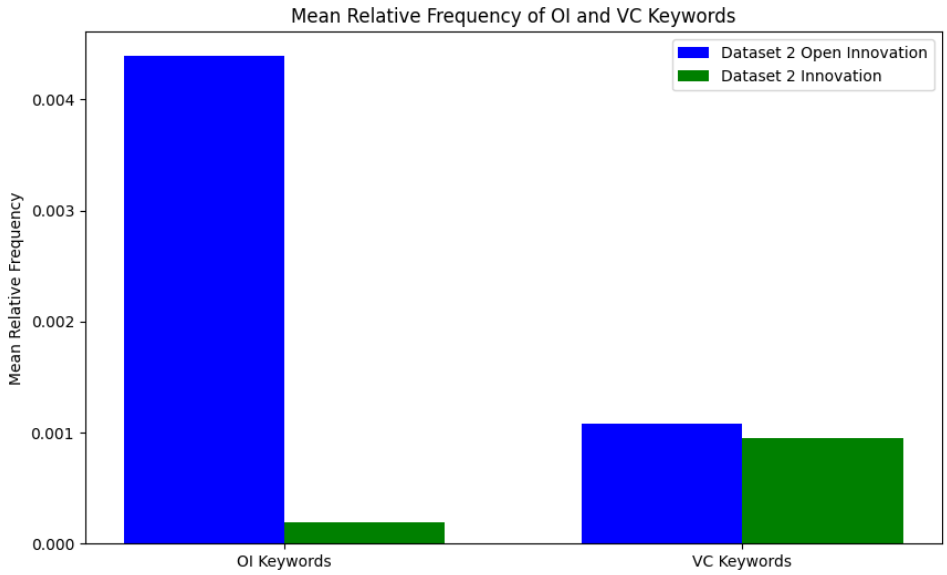

Figure 7: Mean relative frequency of OI and VC keywords

Discussion

The linguistic analysis reveals clear differences in the use of OI keywords between tweets containing #innovation (excluding #openinnovation) and those with #openinnovation. The latter exhibits a significantly higher relative frequency of OI keywords, indicating that users employing this hashtag are more likely to engage in discussions directly related to OI concepts and practices.

Certain keywords are particularly distinctive of #openinnovation tweets, including "industry collaborat," "open innovat," "co-creat," and "crowdsourcing." This suggests that discussions surrounding #openinnovation tend to emphasize the collaborative and participatory nature of this approacH. These findings align with research emphasizing the key role of collaboration, external partnerships, and user involvement in OI (Mention, 2011; West & Bogers, 2014)

In contrast, #innovation tweets exhibit a more diverse range of keywords, reflecting the broader scope of this term. However, some terms, such as "hackathon," appear more frequently in #innovation discussions, potentially indicating a greater focus on technology-driven events or competitions.

4.3 Integrating Sentiment and Language

This section synthesizes and interprets the findings from the sentiment and linguistic analyses, aiming to provide a deeper understanding of how OI is

perceived and discussed on Twitter. By examining the convergences and divergences between the two analyses, this discussion aims to uncover the nuances of OI communication and highlight practical implications for companies and organizations.

4.3.1 Synthesis of Findings

The sentiment analysis revealed a predominantly positive sentiment associated with both #innovation and #openinnovation on Twitter, suggesting a generally favourable view of these concepts. However, #openinnovation exhibited greater variability in sentiment scores compared to #innovation, indicating a more diverse range of opinions and potentially a more nuanced discourse. The sentiment towards #innovation also showed greater sensitivity to daily fluctuations, potentially reflecting a greater responsiveness to external events or news cycles, while sentiment towards #openinnovation remained more stable over time.

The analysis of OI keywords further highlighted the distinction between #innovation and #openinnovation on Twitter. Tweets using #openinnovation consistently demonstrated a higher frequency of keywords directly related to OI principles and practices, such as "industry collaborat," "open innovat," and "crowdsourcing." This finding suggests that #openinnovation is used by a more specialized community focused specifically on the collaborative and participatory aspects of innovation.

4.3.2 Convergences and Divergences

Both the sentiment analysis and the keyword analysis converge in revealing a positive sentiment associated with innovation in general. This suggests that communication about innovation tends to employ a positive frame, aligning with the widespread recognition of innovation as a driver of progress and growth (Chesbrough, 2003). This positive framing might be amplified on social media platforms like Twitter, where users are more likely to share successes and positive experiences, potentially contributing to a bias towards positive communication.

The two analyses also converge in highlighting the more nuanced and multifaceted nature of discussions surrounding #openinnovation compared to #innovation. The wider distribution of sentiment scores and the higher frequency of specialized keywords suggest that conversations about OI encompass a broader spectrum of perspectives, reflecting the inherent complexities and

challenges of this approach. As Chesbrough and Appleyard (2007) note, OI requires careful management of intellectual property, collaboration dynamics, and the integration of external knowledge, all of which can lead to a more diverse range of opinions and experiences.

Despite these convergences, the two analyses also reveal key divergences. While both hashtags are associated with positive sentiment, the sentiment analysis highlighted a statistically significant difference in the mean sentiment score between the two, with #innovation exhibiting a more pronounced positive bias. This difference could reflect the different audiences and purposes associated with each hashtag. #Innovation, as a broader term, might attract a wider audience and be used in more celebratory or promotional contexts, leading to a stronger positive sentiment. #OpenInnovation, on the other hand, might be used by a more specialized community engaged in more critical or nuanced discussions about the specific challenges and opportunities of this approach.

Furthermore, the temporal analysis revealed a distinction in the dynamics of sentiment expression. The sentiment towards #innovation fluctuated more noticeably over time, potentially reflecting a greater responsiveness to external events, news cycles, or trending topics. In contrast, the sentiment towards #openinnovation remained relatively stable, suggesting a more consistent and community-driven discourse, less influenced by external factors. This aligns with research by Bogers et al. (2017) that emphasizes the enduring nature of OI principles and the importance of long-term collaborative practices within dedicated communities.

4.3.3 Two Distinct Discourses

The findings of the sentiment and linguistic analyses raise an important question: do #innovation and #openinnovation represent two facets of the same phenomenon or two distinct discourses on Twitter? The evidence suggests that, while conceptually related, these two hashtags function as distinct markers of different communities and conversations on the platform.

The consistent difference in sentiment, both in terms of average scores and temporal dynamics, points to different underlying dynamics shaping the discussions around each hashtag. #Innovation appears to be associated with a more promotional and celebratory tone, potentially reflecting a broader audience and a greater focus on general announcements or news related to innovation. #OpenInnovation, on the other hand, is characterized by a more nuanced and

varied sentiment, suggesting a more critical and engaged community focused on the specific challenges and opportunities of this approach.

This distinction is further reinforced by the analysis of OI keywords. The significantly higher prevalence of these keywords in #openinnovation tweets compared to #innovation tweets demonstrates that users employing this hashtag are more actively engaging with the specific concepts, practices, and challenges associated with OI. The use of specific keywords like "industry collaborat" or "co-creat" in #openinnovation tweets illustrates this focus on collaborative and participatory approaches, which are less pronounced in #innovation tweets.

These findings suggest that #innovation and #openinnovation function as distinct markers of different communities and conversations on Twitter. While both hashtags relate to the broader concept of innovation, #openinnovation signals a more specialized and engaged community focused on the specific principles and practices of this approach, while #innovation captures a wider range of discussions about innovation in general.

4.3.4 Practical Implications

The insights from the sentiment and linguistic analyses provide valuable guidance for companies and organizations seeking to communicate about OI effectively on Twitter. Understanding the different audiences, sentiments, and language patterns associated with #innovation and #openinnovation can help tailor communication strategies to specific goals and audiences.

- **For Building Brand Awareness:**
    - Leverage the generally positive sentiment associated with innovation by using enthusiastic and positive language to highlight the company's commitment to innovation and its potential benefits.
    - Share success stories and case studies that showcase the positive impact of OI on the company's products, services, or overall strategy.
    - Utilize the broader reach of the #innovation hashtag to increase visibility among a wider audience.
- **For Fostering Engagement and Collaboration:**
    - Adopt a more inclusive and collaborative tone, using language that encourages participation and dialogue, such as "Let's collaborate!" or "Share your ideas!"

    - Focus on specific OI themes and challenges, using relevant keywords and the #openinnovation hashtag to connect with a more specialized and engaged community.
    - Actively participate in relevant conversations and respond to tweets and comments, demonstrating a commitment to dialogue and interaction.
- **For Seeking Partners:**
    - Showcase the company's expertise and specific capabilities related to OI, using a more technical and professional tone.
    - Promote OI challenges or calls for proposals, clearly outlining the company's needs and inviting external collaboration.
    - Utilize relevant hashtags and keywords to reach potential partners with specific expertise or interests.

By tailoring their communication strategies to the specific audiences, sentiments, and language patterns associated with #innovation and #openinnovation, companies and organizations can effectively communicate about their OI initiatives, build relationships with key stakeholders, and achieve their desired outcomes on Twitter.

4.4 Conclusions

This chapter has delved into the world of OI discourse on Twitter, employing sentiment analysis and linguistic analysis to uncover the nuances of how this concept is perceived and communicated online. The findings provide a deeper understanding of the social dynamics surrounding OI, offering valuable insights for both researchers and practitioners.

The analysis revealed that OI is generally viewed positively on Twitter, as evidenced by the predominantly positive sentiment associated with both #innovation and #openinnovation. However, #openinnovation discussions exhibited greater variability in sentiment, suggesting that this approach – while promising – is also perceived as more complex and challenging. This finding has implications for organizations implementing OI strategies, highlighting the need to address potential concerns and complexities through clear and transparent communication.

Furthermore, the analysis of OI keywords revealed that #openinnovation is used by a more specialized community that actively engages with the specific principles and practices of this approach. The prevalence of terms like "industry

collaborat," "co-creat," and "crowdsourcing" in #openinnovation tweets underscores the importance of collaboration, external partnerships, and user involvement in this discourse. This finding suggests that companies seeking to engage with the OI community on Twitter should adopt a more participatory and collaborative communication style, using language that resonates with this specific audience.

The temporal analysis of sentiment highlighted the dynamic nature of the conversation surrounding innovation on Twitter. While the sentiment towards #innovation was more volatile, fluctuating in response to external events or news cycles, the sentiment towards #openinnovation remained relatively stable, suggesting a more enduring and community-driven discourse. This stability might indicate that the #openinnovation community is less influenced by hype or short-term trends, and more focused on the long-term principles and practices of OI.

This chapter has demonstrated the value of combining sentiment analysis and linguistic analysis for a more nuanced understanding of OI discourse on Twitter. By examining both the emotional tone and the specific language used, we can gain insights into how different communities perceive and engage with this concept. This integrated approach offers a powerful tool for researchers and practitioners seeking to understand the social dynamics of OI and to develop effective communication strategies.

4.4.1 Future Research

Several promising avenues for future research emerge from this study:

- Longitudinal Analysis: Analysing the evolution of sentiment and language over a more extended period could reveal how the OI discourse on Twitter changes over time and how it responds to major events or shifts in the field.
- Comparative Analysis: Comparing the sentiment and language used for OI on Twitter with other social media platforms, such as LinkedIn or specialized OI platforms, could provide insights into the platform-specific dynamics of this discourse.
- Network Analysis of Interactions: Expanding the network analysis to include replies and threads could reveal more nuanced patterns of interaction and collaboration within the OI community, providing a

deeper understanding of how knowledge is shared and ideas are developed on the platform.

- Qualitative Analysis of Tweet Content: Conducting a qualitative analysis of a subset of tweets could provide richer insights into the specific topics, challenges, and motivations driving the OI discourse on Twitter. This qualitative analysis could complement the quantitative findings and provide a deeper understanding of the context and meaning behind the language used.
- Impact of Communication Strategies: Investigating the impact of different communication strategies on sentiment and engagement with OI messages on Twitter could provide valuable guidance for companies and organizations seeking to optimize their communication efforts. This could involve analysing the effects of different language styles, content types, hashtag usage, and interaction patterns.

By pursuing these research directions, we can further enhance our understanding of the OI discourse on Twitter and its implications for the future of innovation in the digital age.

Chapter 5 Thematic Exploration, Network Dynamics, and Language-Based Clustering

5.1 Introduction

OI has emerged as a critical paradigm for organizations seeking to accelerate their innovation processes and tap into the collective intelligence of diverse communities. As companies increasingly recognize the value of external knowledge and collaboration, social media platforms have become essential tools for facilitating OI initiatives. Twitter, with its public nature, global reach, and real-time communication capabilities, offers a unique environment for observing and analysing the dynamics of OI discourse.

This chapter delves into the vibrant landscape of OI on Twitter, employing a multi-faceted approach to uncover the key themes, actors, and communication patterns that shape this dynamic online space. By combining topic modelling, network analysis, and language-based clustering, we aim to provide a holistic and nuanced understanding of the OI ecosystem on this influential platform. These three complementary methods offer distinct yet interconnected perspectives on the OI discourse. Topic modelling reveals the key areas of focus within OI conversations, highlighting the diverse interests and concerns of the community. Network analysis visualizes and quantifies the connections between individuals and organizations, identifying influential actors, communities, and pathways for knowledge diffusion. Language-based clustering goes beyond explicit hashtag use to identify groups of users with similar communication styles, revealing nuanced communities of practice within the broader OI discourse.

By integrating the insights from these three analyses, this chapter aims to create a rich tapestry of the OI ecosystem on Twitter, illuminating the interplay between

themes, actors, and communication styles. This understanding is crucial for both researchers seeking to deepen their knowledge of OI in the social media age and practitioners aiming to leverage Twitter effectively for OI initiatives. The following sections will detail the methodology and findings of each analysis, providing a comprehensive and insightful exploration of the OI landscape on Twitter.

5.2 Topic Modelling with LDA

This study utilizes topic modelling to identify the key themes and trends that define OI discussions on Twitter. Specifically, LDA, a probabilistic model for automatically discovering underlying topics within a collection of documents, is employed to analyse the corpus of tweets (Blei et al., 2002). LDA operates under the assumption that each document, in this case, a tweet, is a mixture of various topics, each characterized by a distribution of words. By analysing the co-occurrence patterns of words within the tweets, LDA uncovers the latent topics that best explain the observed word distributions.

5.2.1 Methodology

Before applying LDA, the tweet data underwent a series of pre-processing steps to ensure data quality and focus the analysis on meaningful content. These steps included removing irrelevant elements such as URLs, mentions, hashtags, and emojis, as well as converting all text to lowercase and removing punctuation. These actions ensure that the analysis is based solely on the substantive textual content and not influenced by user-specific formatting or non-textual elements.

Following pre-processing, the text was tokenized, separating it into individual words. Stop words, common words like "the," "a," and "is," that carry little semantic meaning, were removed to further refine the dataset. Finally, the remaining words were lemmatized, reducing them to their base or dictionary form to consolidate different forms of the same word, enhancing the accuracy of the topic modelling process. These pre-processing steps are commonly used in text analysis, particularly when dealing with social media data, to ensure that the analysis captures the essential content and is not skewed by irrelevant elements or variations in language use (Clark & Araki, 2011).

The optimal number of topics (k) for the LDA model was determined using the elbow method, a commonly used technique for identifying the point of diminishing returns when increasing the number of clusters in a dataset. This

method involves plotting the coherence score, a measure of the semantic similarity between the top words within a topic, against different values of *k*. The "elbow" on the plot, where the coherence score begins to plateau, suggesting minimal improvement with additional topics, was identified as the optimal number of topics. Based on this analysis, k = 8 was chosen as the optimal number of topics for this study, providing a balance between capturing the diversity of themes and ensuring the coherence and interpretability of each topic.

The LDA model was implemented using the Gensim library in Python, a powerful tool for topic modelling that offers efficient algorithms and functionalities for evaluating model performance. The alpha and beta parameters, which control the distribution of topics within documents and the distribution of words within topics, respectively, were set to their default values within Gensim. These default values are commonly used in LDA implementations and provide a good starting point for analysis. The quality of the model was assessed using coherence scores, ensuring that the extracted topics were semantically meaningful and interpretable.

5.2.2 Results

The LDA analysis identified eight distinct topics that capture the prominent themes within the OI discourse on Twitter. Table 5.1 presents these topics along with their associated keywords and interpretive labels, summarizing the essence of each theme.

Table 4: Topics Identified Through LDA

Topic	*Salient Terms*	*Interpretive Label*
0	*project, invest, financ, consult, fund, busi, onpass, market, partner, growth*	Funding & Investment in Innovation
1	*blockchain, futur, crypto, excit, game, web, develop, metavers, decentral, nft*	Blockchain, Cryptocurrencies & Gaming
2	*like, technew, news, follow, appl, ai, read, latest, know, articl*	Technology News & Updates
3	*busi, leadership, success, new, growth, world, transform, improv, digit, learn*	Business, Leadership & Growth
4	*design, tech, robot, futur, creativ, build, experienc, think, space, world*	Design, Robotics & Creativity

5	*join, day, excit, event, research, 1, opportun, workshop, team, learn*	Innovation Events & Conferences
6	*tech, futur, read, latest, insight, articl, trend, learn, share, develop*	Education & Future Skills in Technology
7	*effici, sustain, solut, product, construct, build, improv, design, world, work*	Efficiency, Sustainability & Innovative Solutions

Figures below provide a visual comparison of the topic distributions for tweets containing #innovation and those containing #openinnovation, highlighting the differences in the prevalence of certain topics between the two groups.

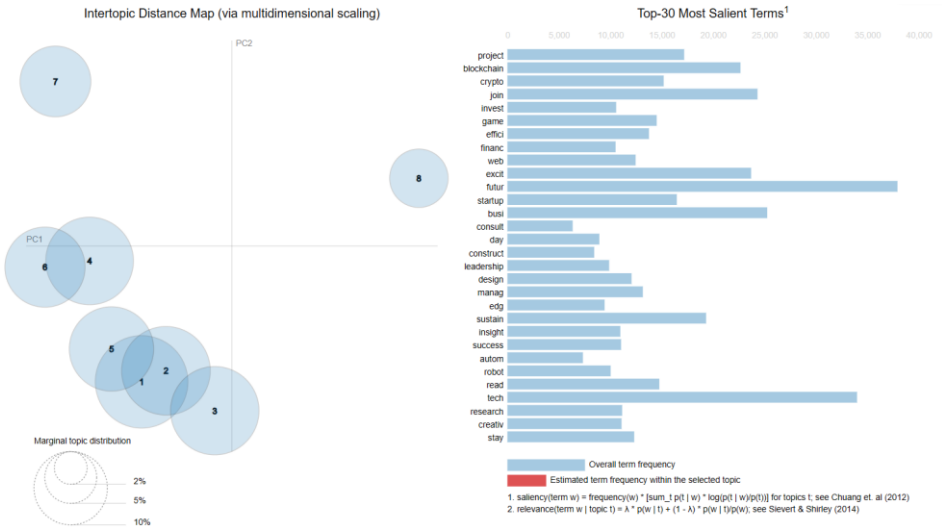

Figure 8: LDA visualization

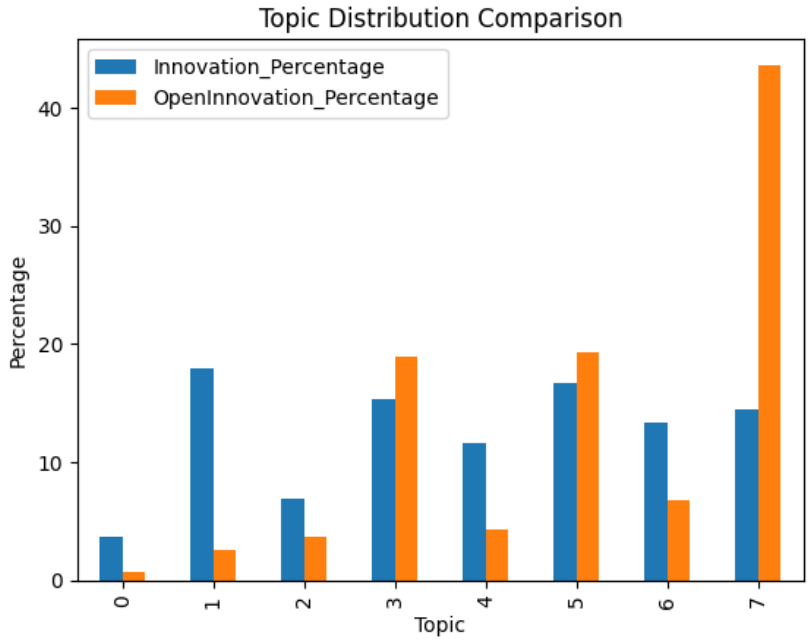

Figure 9: Topic Distribution Comparison

For example, discussions related to "Funding & Investment in Innovation" (Topic 0) are significantly more prevalent in tweets associated with #openinnovation (almost 18%) compared to #innovation (less than 4%). This suggests that those engaging in conversations specifically about OI are more likely to focus on the financial and investment aspects of this approach. Similarly, "Business, Leadership & Growth" (Topic 3) is more prominent in the #openinnovation tweets, potentially indicating that these discussions tend to centre around the practical application and implementation of OI within business contexts.

Conversely, topics like "Technology News & Updates" (Topic 2) and "Design, Robotics & Creativity" (Topic 4) are more prevalent in tweets containing #innovation. This could suggest that the broader conversation surrounding innovation on Twitter encompasses a wider range of technological advancements and creative applications, whereas those focused on #openinnovation prioritize the strategic and organizational aspects of collaborative innovation. Topic 7, "Efficiency, Sustainability & Innovative Solutions," is the most prevalent topic overall, indicating a shared focus on these themes across both #innovation and #openinnovation discussions.

To further illustrate the content and scope of each identified topic, Table 5.2 provides representative examples of tweets from each topic category.

Table 5: Exemplary Tweets for Each Identified Topic

Topic ID	*Example Tweet*
0	*Big Bang ! OKX Ventures is investing in Jambo with a shared mission of enhancing access to Web3 technologies and driving adoption. @okx @okxweb3 @okx_vietnam @JamboTechnology Read more: https://t.co/ldlFMniPzu #OKXVentures #JAMBO #PHONE #Blockchain #Innovation https://t.co/6cDyBiLMAT*
1	*#OpenInnovation @ethereum is the verifiable internet. Rollup is the verifiable web server. AVS is verifiable SaaS. @eigenlayer is the verifiable cloud. Welcome to the verifiable digital commons!*
2	*🚀 Exciting Upcoming Changes in OxygenOS 15/ColorOS 15! 🚀 #OxygenOS15 #ColorOS15 #TechUpdates #Smartphones #Innovation #TechNews #TechibeeReview Retweet for more updates! 🔄 https://t.co/DWXR1UHzCo*
3	*#Innovation and #quality are the keys to driving the global appeal of Chinese goods, said Wang Lingjun, vice minister of China's General Administration of Customs, debunking the "#overcapacity" discourse trap. #PressroomHighlights https://t.co/2E4KRW6aep*
4	*Very useful! This device stitches you up without the need for stitches Credit: Zipstitcht #techforgood #innovation #technology #healthtech https://t.co/frPmHfQvka*
5	*🚀 Already over 100 companies on our waitlist! Join today and don't miss out! 🌟 #JoinUs #Innovation #Tech #AI #Recruiting https://t.co/p4waIMmyu1 https://t.co/fW22wqp7jn*
6	*🎓 DE Do you want to #study in #Germany? ❗💡 Here's everything you need to know about: 🤯 Universities for innovation 🦾 Courses about robotics and bionics 🤲 Future-oriented programs #StudyInGermany #HigherEducation #Innovation #Future #technology https://t.co/tQk5ywSddP*
7	*🌍📰 @ReSkills_ is booming with 1M online students, expanding from Asia and Africa into North and Latin America! @vincemedia1 Watch the full interview here ⬇️⬇️⬇️ https://t.co/4JSB9n0kYO #ReSkills #Innovation #growth #enterpreneur #ecommerce #digitalmarketing #careergrowth #business #newtothestreet*

These examples provide valuable insights into the diverse conversations happening within the OI community on Twitter, showcasing the range of topics, perspectives, and communication styles. The prominence of Topic 0 ("Funding & Investment in Innovation") in the #openinnovation tweets aligns with the

emphasis on collaboration and openness central to the OI paradigm. This suggests that Twitter serves as a platform where stakeholders discuss and engage with the practical implications of OI, particularly its financial and strategic aspects.

5.2.3 Interpretation of Results

The LDA topic modelling analysis reveals a rich tapestry of themes within the OI discourse on Twitter, reflecting the multifaceted nature of this evolving field. The identified topics highlight the various dimensions of OI, encompassing financial aspects, technological advancements, business strategies, collaborative partnerships, future skills, and sustainability solutions.

The comparison of topic distributions between #innovation and #openinnovation suggests a distinction between the broader conversation surrounding innovation and the more specialized focus on OI. For instance, the significant difference in the prevalence of "Funding & Investment in Innovation" (Topic 0) points to the importance of financial resources and collaborative partnerships in driving OI. While visual inspection of the topic distribution suggests differences in topic prevalence between #innovation and #openinnovation, further statistical analysis, such as a chi-squared test and the calculation of Cramer's V for effect size, is needed to confirm the statistical and practical significance of these differences.

The prominence of topics related to funding, business, and leadership in the #openinnovation tweets suggests a focus on the practical implementation and strategic implications of this approach. Conversely, the higher prevalence of topics related to technology news, design, and creativity in the #innovation tweets indicates a broader interest in technological advancements and their applications within the general innovation discourse. Topic 7, "Efficiency, Sustainability & Innovative Solutions," is the most prevalent topic overall, indicating a shared focus on these themes across both #innovation and #openinnovation discussions. This shared concern for efficiency and sustainability aligns with the growing recognition of these issues as crucial drivers of innovation across various industries.

The example tweets provide further insights into the content and scope of each topic, showcasing the diversity of perspectives and communication styles within the OI community on Twitter. These findings lay the groundwork for a deeper exploration of the OI ecosystem on Twitter, which will be further investigated through network analysis and language-based clustering in the subsequent sections.

5.3 Network Analysis

This study employs network analysis to delve into the structure and dynamics of the OI ecosystem on Twitter. By visualizing and quantifying the relationships between actors (individuals and organizations) engaged in OI discussions, network analysis provides valuable insights into the flow of information, the influence of key players, and the formation of communities around shared interests. This approach allows us to explore the interconnectedness of the OI community on Twitter and understand how different actors contribute to the conversation.

5.3.1 Methodology

The network for this analysis was constructed using user mentions as indicators of relationships. Each unique Twitter user in the dataset, whether they posted a tweet or were mentioned in a tweet, was represented as a node in the network. A directed edge was created from user A to user B if user A mentioned user B in a tweet. This approach captures the flow of communication and attention within the OI community on Twitter, highlighting who is engaging with whom and how information is being disseminated.

To gain a broader perspective on the OI discourse, two distinct networks were constructed:

- **Combined OI Network:** This network encompasses all tweets from Dataset 2 (April-July 2024) that contain either #innovation or #openinnovation. This inclusive approach provides a comprehensive view of the interactions among users engaged in any aspect of innovation discourse on Twitter, regardless of their explicit focus on OI.
- **Innovation Network:** This network includes only tweets from Dataset 2 containing #innovation (excluding those also containing #openinnovation). By focusing solely on #innovation, this network provides a more targeted analysis of communication patterns among users who are not explicitly engaging in discussions about OI, allowing for a comparison between the two groups.

To enhance the clarity and interpretability of the network visualizations and to focus the analysis on the most significant connections, both networks were simplified. Initially, a threshold of 100 was used to remove low-weight edges, aiming to reduce the number of nodes and make the network visualizations

manageable. However, this initial threshold resulted in an overly simplified network, potentially obscuring important connections and relationships. Therefore, the threshold was adjusted to 10, retaining a larger number of edges while still ensuring visual clarity. This adjustment revealed a more nuanced view of the network structure, allowing for a more in-depth analysis.

The simplification process also included removing self-loops (edges connecting a node to itself) and single nodes (nodes with no connections), as these do not provide information about interactions between different users. Due to its significantly larger size and complexity, the Innovation Network was further simplified by retaining only the top 1000 nodes with the highest degree centrality. This strategy further reduces visual clutter while preserving the core structure of the network and highlighting the most influential actors.

After simplification, key network metrics were calculated for both networks using the NetworkX library in Python. These metrics, grounded in the principles of Network Theory, provide insights into the structure, influence, and community formation within the networks (Park et al., 2015):

- **Degree Centrality:** This metric measures the number of connections a node has, representing how well-connected a user is within the network. Users with high degree centrality are often considered influential, as they are frequently mentioned by others and have a wider reach within the community.
- **Betweenness Centrality:** This metric captures a node's ability to act as a bridge or connector between different parts of the network. Users with high betweenness centrality often play a critical role in facilitating the flow of information and connecting diverse communities within the network.
- **Clustering Coefficient:** This metric quantifies the degree to which a node's neighbours are connected. High clustering coefficients indicate tightly-knit communities, where users are more likely to interact and share information amongst themselves.

To visualize the network structures and facilitate their exploration, both networks were visualized using Gephi, an open-source network analysis and visualization software. The visualizations employed force-directed layouts to illustrate the overall network structure and community detection algorithms to highlight clusters of interconnected users. Due to the complexity of the networks, even after

simplification, the visualizations were limited to the top 1000 nodes and edges with a weight greater than 10 to ensure visual clarity and interpretability.

5.3.2 Results

The network visualizations generated using Gephi reveal distinct characteristics of the OI discourse on Twitter.

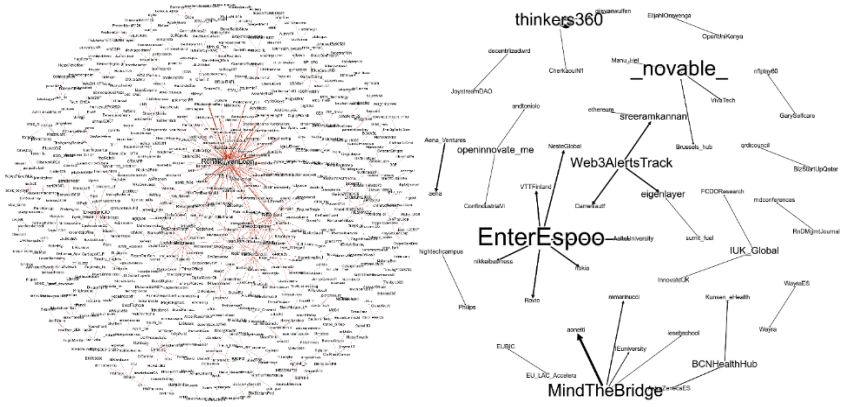

Figure 10: Network Visualizations

The Combined OI Network, encompassing tweets related to both #innovation and #openinnovation, exhibits a sparse and fragmented structure. This suggests that the broader conversation surrounding innovation on Twitter is relatively decentralized, with limited direct interaction and interconnection among users. However, despite its overall sparseness, the network does reveal the presence of clusters, indicating the formation of communities around specific topics, individuals, or organizations.

The Innovation Network, focusing solely on tweets containing #innovation, reveals a denser and more complex structure compared to the Combined OI Network. This suggests that the conversation surrounding general innovation on Twitter is more interconnected and involves a wider range of actors.

Analysis of the degree centrality metric reveals a statistically significant difference between the #innovation and #openinnovation networks (T-statistic: -237.850, P-value: < 0.001). This finding indicates that users discussing #openinnovation tend to have a lower degree centrality than those discussing #innovation. This difference could indicate a more centralized and interconnected conversation within the broader #innovation community, while the #openinnovation community might be more diffuse and less reliant on a few central actors.

While initial analysis suggested a statistically significant difference in clustering coefficient between the two networks, further analysis with a refined threshold for low-weight edges revealed no statistically significant difference (T-statistic: -1.886, P-value: 0.059). This suggests that while there might be a tendency for slightly higher clustering within the #openinnovation network (p = 5.9%, borderline significance), the difference is not strong enough to be considered statistically significant.

Calculation of betweenness centrality, a metric that could further illuminate the bridging role of users between different communities, proved computationally challenging due to the size and complexity of the networks. The analysis was repeatedly halted due to memory constraints, suggesting a need for more powerful computational resources or alternative strategies for analysing betweenness centrality in large networks.

5.3.3 Interpretation of Results

The network analysis paints a revealing picture of the OI ecosystem on Twitter. While the broader innovation discourse appears more interconnected and potentially reliant on a few central actors, as indicated by the higher degree centrality in the #innovation network, the community specifically focused on #openinnovation exhibits a more diffuse structure with a tendency towards slightly higher clustering, suggesting a possible emphasis on community formation within this niche.

These findings align with the theoretical framework presented in Chapter 2. The higher degree centrality in the #innovation network could suggest a stronger influence of traditional, hierarchical structures within the broader innovation discourse, whereas the #openinnovation community, with its lower degree centrality and tendency towards higher clustering, might be more aligned with the principles of decentralization and collaborative networks that are characteristic of OI (Chesbrough, 2003).

The limitations of Twitter as a platform for fostering large-scale collaborative OI projects are also revealed through this analysis. The overall sparseness of the Combined OI Network suggests that while Twitter can facilitate knowledge sharing and awareness building, sustaining complex, long-term collaborations might be more challenging on this platform. However, the presence of clusters within both networks indicates that Twitter can support the formation of smaller,

more focused communities of practice, enabling collaboration and knowledge exchange within those specific niches.

The findings from this section provide a valuable foundation for understanding the structural dynamics of the OI ecosystem on Twitter. The next section will delve into the nuances of communication styles within this ecosystem, using language-based cluster analysis to uncover hidden communities and explore how language shapes the OI discourse.

5.4 Language-Based Cluster Analysis

This study utilizes language-based cluster analysis to identify groups of users with similar communication patterns, regardless of their explicit use of specific hashtags. This approach complements the hashtag-based analysis by revealing nuanced differences in language use and underlying themes, providing a deeper understanding of the diversity of voices and perspectives within the OI discourse on Twitter. By examining the semantic similarities between tweets, we aim to uncover hidden communities and explore how language shapes the OI discourse.

5.4.1 Methodology

The language-based cluster analysis was conducted using BERT embeddings and the k-means clustering algorithm. BERT, a powerful language model pre-trained on a massive dataset of text and code, excels at understanding the context and meaning of words in a sentence (Devlin et al., 2019). For each tweet in the dataset, BERT generates a 768-dimensional embedding vector that represents the tweet's semantic content. These embeddings capture the nuanced meaning of words within their context, providing a more accurate representation of tweet content compared to traditional word embeddings, which often fail to capture the subtleties of language use in social media.

Specifically, BERT's ability to consider the relationships between words bidirectionally – both preceding and following a given word – allows it to grasp the nuances of meaning that are often lost in traditional word embeddings, which typically analyse words in isolation or only consider one direction of context. Furthermore, BERT's training on a massive and diverse dataset allows it to develop a rich understanding of language, including slang, idiomatic expressions, and other linguistic features common in social media communication. Several studies have demonstrated BERT's superiority in capturing semantic similarity in social media text, particularly when compared to traditional word embedding

methods like Word2Vec or GloVe. For example, Konstantinov, Moshkin, and Yarushkina (2021) found that BERT outperformed Word2Vec in sentiment analysis tasks using social media data, while Başarslan and Kayaalp (2021) demonstrated BERT's effectiveness in sentiment analysis of social media reviews. In the context of disaster prediction using Twitter data, Chanda (2021) highlighted the efficacy of BERT embeddings in accurately capturing the semantic meaning of tweets related to disasters.

K-means clustering is an unsupervised machine learning algorithm that groups data points into *k* clusters based on their similarity. In this case, the data points are the BERT embeddings representing the tweets. The algorithm iteratively assigns data points to clusters, aiming to minimize the distance between points within the same cluster and maximize the distance between points in different clusters.

Cosine similarity was chosen as the distance metric for comparing tweet embeddings. This metric measures the angle between two vectors in a multi-dimensional space, ignoring their magnitude. Cosine similarity is particularly well-suited for comparing text embeddings as it focuses on the semantic relationships between words rather than their absolute values, which can vary depending on the length of the text.

The elbow method was used to determine the optimal number of clusters (*k*) for the k-means algorithm. The elbow method involves plotting the within-cluster sum of squares against different values of *k* and identifying the point where the plot bends, resembling an elbow. This point suggests a balance between minimizing the within-cluster variance and avoiding an overly complex model with too many clusters. The analysis indicated that 4 clusters was the optimal number for this dataset.

Figure 5.2 presents a visualization of the tweet embeddings using Principal Component Analysis (PCA), a dimensionality reduction technique that projects the high-dimensional embeddings onto a two-dimensional plane. This visualization provides a visual representation of the distribution of tweets based on their semantic similarity.

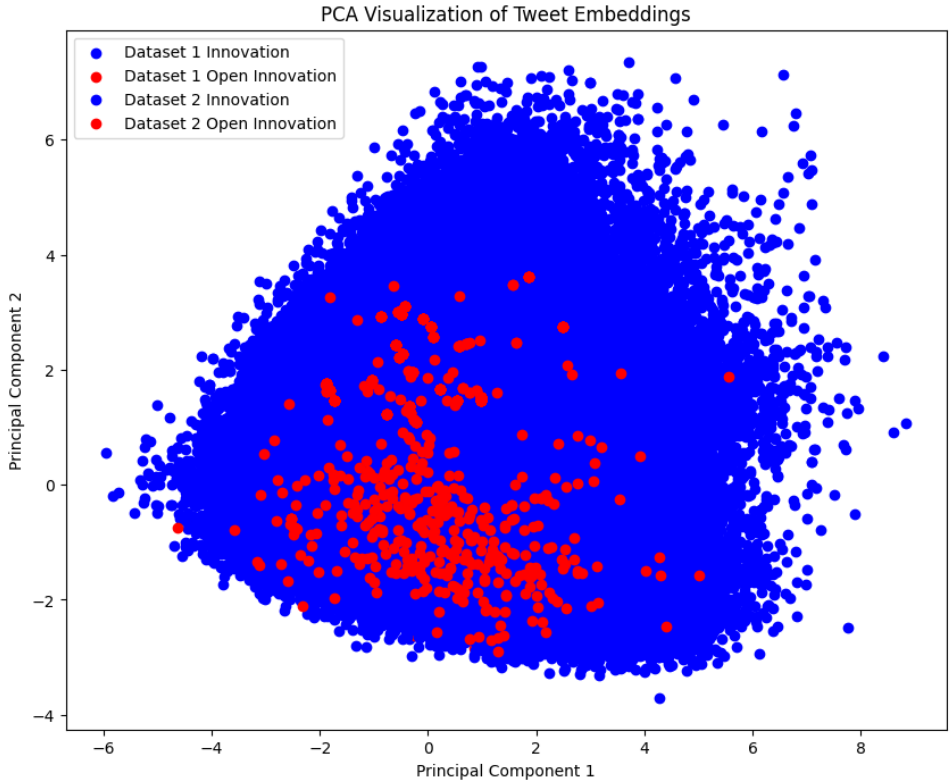

Figure 11: PCA Visualization of Tweet Embeddings. Each point represents a tweet, projected into a two-dimensional space based on its semantic similarity (calculated using BERT embeddings). Colours indicate the hashtag used (#innovation or #openinnovation).

After the clusters were formed, they were manually inspected by examining a sample of tweets within each cluster. This process involved identifying common themes, language patterns, and user types to gain a better understanding of the characteristics that define each cluster. Based on this analysis, each cluster was assigned a descriptive label that captures its essence.

5.4.2 Results

The language-based cluster analysis revealed four distinct groups of users engaged in discussions related to OI and innovation on Twitter. These clusters transcend the limitations of hashtag-based analysis, highlighting the diversity of voices and communication styles within the broader conversation. Table 5.3 provides a summary of each cluster, including its interpretive label, key themes, and an example tweet.

Table 6: Language-Based Clusters in the OI Discourse on Twitter

Cluster Label	*Key Themes*	*Example Tweet*
Tech Enthusiasts & Community Engagement	Emerging technologies, blockchain, AI, decentralized solutions, future of tech, enthusiastic tone, community-oriented language	"gained attention for its innovative sharding technology and unique approach to scalability. Exciting times ahead for NEAR Protocol! #decentralized #blockchain #innovation"
Corporate OI & Startup Promotion	Startup acceleration, corporate innovation programs, investment opportunities, partnerships, ecosystem building, promoting events and initiatives	"🎉 We've recently reached 20,000 followers on Instagram! 🎉\nHere's to many more milestones together! 🤗❤️\nFollow us! https://t.co/G3yFnCBpMv \n#ThankYou #CorporateInnovation #OpenInnovation #StartupInvestor #StartupAcceleration #CorporateVentureCapital #OnePNPTC https://t.co/wfKiCATXAI"
Research & Development Focus	Scientific advancements, technical innovations, research findings, industry-specific challenges, solutions and technologies, data-driven insights	"Colgate-Palmolive is on the lookout for groundbreaking technologies that can revolutionize tooth whitening!\n\nSubmit your ideas here: 🔗 https://t.co/E5Uz9ELSai \n\n#Innovation #ToothWhitening #ColgatePalmolive #DentalCare #TechInnovation #OralHealth https://t.co/H7IlCGN5fb"

Strategic & Collaborative OI	Best practices, frameworks, partnerships, impact of OI, connecting corporations with startups, fostering collaboration, solution-oriented language	"🔑Inocrowd is the key to unlocking groundbreaking solutions and driving your business forward. We help companies and organizations find innovative solutions to their toughest challenges.\n\nFor more information, visit 👉 https://t.co/KBrlFSAGiY\n\n#Innovation #OpenInnovation https://t.co/7U0rUvENu9"

5.4.3 Interpretation of Results

The language-based cluster analysis reveals the presence of distinct communities within the OI discourse on Twitter, each characterized by a unique communication style and thematic focus. These findings align with the concept of *communities of practice*, as described in Social Capital Theory, where individuals with shared interests and goals come together to exchange knowledge, learn from each other, and build a collective identity (Wenger, 1998).

The variation in communication styles across clusters suggests that the dissemination of OI principles and practices on Twitter is not a monolithic process. Different communities engage with OI in distinct ways, reflecting their diverse backgrounds, motivations, and goals. For instance, the "Tech Enthusiasts & Community Engagement" cluster, characterized by an informal and enthusiastic tone, often employs language that fosters community interaction and highlights the transformative potential of emerging technologies. This cluster's focus on blockchain, AI, and decentralized solutions suggests a strong interest in disruptive innovations and a belief in the power of technology to drive positive change. On the other hand, the "Corporate OI & Startup Promotion" cluster adopts a more formal and promotional tone, frequently highlighting investment opportunities, partnerships, and the success stories of startups. This cluster's language reflects a focus on building a vibrant and thriving OI ecosystem, connecting startups with corporations and investors to accelerate innovation.

The "Research & Development Focus" cluster stands out for its emphasis on scientific advancements, data-driven insights, and industry-specific challenges. This community's communication style is characterized by a more technical and research-oriented tone, often referencing scientific publications, research

findings, and innovative solutions to complex problems. Finally, the "Strategic & Collaborative OI" cluster emphasizes best practices, frameworks, and partnerships, reflecting a more strategic and solution-oriented approach to OI. This cluster's language often centers around the benefits of collaboration, knowledge sharing, and the impact of OI on organizational success.

The identification of these language-based clusters provides valuable insights for organizations seeking to leverage Twitter for OI initiatives. By recognizing the distinct communication styles and interests of different communities, companies can tailor their messages and engagement strategies to resonate with specific audiences. For example, organizations targeting the "Tech Enthusiasts & Community Engagement" cluster should adopt a more informal and enthusiastic tone, highlighting the transformative potential of their innovations and emphasizing community involvement. Conversely, those seeking to engage with the "Strategic & Collaborative OI" community should adopt a more formal and professional approach, focusing on the strategic benefits of OI and showcasing successful collaborations.

Future research could expand on this analysis by exploring the temporal evolution of these clusters, investigating whether their composition and communication styles change over time. Furthermore, a deeper qualitative analysis of the tweets within each cluster could provide a richer understanding of the specific motivations, challenges, and perspectives that shape their discourse.

5.5 Discussion

This section synthesizes the findings from the topic modelling, network analysis, and language-based cluster analysis, highlighting the interconnected nature of the OI ecosystem on Twitter. By examining the relationships between the identified topics, actors, and communication styles, we aim to provide a holistic understanding of how OI is discussed and practiced on this platform.

5.5.1 Relationships and Intersections

The three analyses conducted in this chapter reveal a complex and multifaceted OI landscape on Twitter. While each analysis provides a unique lens for understanding this landscape, they also complement and reinforce each other, creating a richer and more nuanced picture when considered together.

For instance, the topic modelling analysis revealed a higher prevalence of topics related to funding, business, and leadership in the #openinnovation tweets. This

finding aligns with the network analysis, which showed that the #openinnovation community exhibits a higher degree centrality and clustering coefficient. This suggests that these users are more interconnected and potentially engaged in more focused discussions related to the practical application and strategic implications of OI. Conversely, the prominence of topics related to technology news, design, and creativity in the #innovation tweets is reflected in the denser and more complex structure of the Innovation Network, indicating a broader and more diverse conversation surrounding general innovation.

Furthermore, the language-based cluster analysis provides additional insights into the characteristics of these communities. For example, the "Tech Enthusiasts & Community Engagement" cluster, with its informal language and focus on emerging technologies, aligns with the prominence of Topic 1 ("Blockchain, Cryptocurrencies & Gaming") and Topic 4 ("Design, Robotics & Creativity") in the #innovation tweets. This suggests that this community is actively engaged in exploring and promoting novel technological advancements. Conversely, the "Strategic & Collaborative OI" cluster, with its more formal and professional tone, aligns with the higher prevalence of Topic 0 ("Funding & Investment in Innovation") and Topic 3 ("Business, Leadership & Growth") in the #openinnovation tweets, reflecting this cluster's interest in the strategic and practical aspects of OI.

5.5.2 Implications for OI Communication on Twitter

The findings of this chapter have significant implications for organizations seeking to leverage Twitter for OI initiatives. By understanding the thematic landscape, the network structure, and the distinct communication styles of different communities, organizations can tailor their messages and engagement strategies to resonate with specific audiences.

For example, companies seeking to attract investment or build partnerships could focus on content related to "Funding & Investment in Innovation" (Topic 0) and engage with users within the #openinnovation community, which exhibits a higher concentration of individuals interested in these topics. Conversely, organizations seeking to promote new technologies or foster a broader conversation about innovation could target a wider audience using the #innovation hashtag and focus on content related to "Technology News & Updates" (Topic 2) or "Design, Robotics & Creativity" (Topic 4).

The language-based cluster analysis provides further insights for tailoring communication strategies. Companies targeting the "Tech Enthusiasts & Community Engagement" cluster should adopt a more informal and enthusiastic tone, highlighting the transformative potential of their innovations and encouraging community involvement. Those seeking to engage with the "Strategic & Collaborative OI" community should adopt a more formal and professional approach, focusing on the strategic benefits of OI and showcasing successful collaborations. By understanding the nuances of communication within these distinct communities, organizations can increase the effectiveness of their messaging and build stronger relationships with potential collaborators and partners.

5.6 Limitations and Future Research Directions

While this study provides insights into the OI discourse on Twitter, it's important to acknowledge its limitations.

Firstly, the generalizability of the findings to other social media platforms or contexts beyond Twitter is limited. Twitter, with its unique characteristics, might not accurately represent the broader landscape of OI communication. Future research could explore OI discussions on other platforms like LinkedIn or ResearchGate, comparing and contrasting the findings across different social media environments.

Secondly, the reliance on specific hashtags to identify relevant tweets introduces a potential selection bias. As discussed in Section 3.2.4, users might engage in OI-related discussions without explicitly using the #innovation or #openinnovation hashtags. Future research could employ a wider range of keywords, phrases, or topic modelling techniques to capture a more comprehensive view of the OI discourse.

Thirdly, the cross-sectional nature of the data limits the ability to capture the evolution of the OI conversation over time. Longitudinal studies, tracking the changes in themes, actors, and communication styles over an extended period, could provide a richer understanding of the dynamics of OI on Twitter.

Finally, the computational challenges encountered in calculating betweenness centrality for the full networks highlight a limitation in terms of computational resources. Future research could explore alternative network analysis techniques or leverage more powerful computing resources to overcome this limitation.

5.7 Conclusion

This chapter has explored the OI ecosystem on Twitter through a multifaceted lens, combining topic modelling, network analysis, and language-based cluster analysis. The findings highlight the diversity of themes, actors, and communication styles that shape this dynamic online space, offering valuable insights for both researchers and practitioners.

The topic modelling analysis revealed eight distinct topics that capture the key areas of focus within the OI discourse. The network analysis uncovered differences in network structure and engagement levels between the broader innovation community and the more specialized OI community. The language-based cluster analysis identified four distinct communities with unique communication styles and thematic foci. These analyses, when considered together, provide a comprehensive view of the OI landscape on Twitter, illuminating the interconnected nature of themes, actors, and communication patterns.

The findings have practical implications for organizations seeking to leverage Twitter for OI initiatives. By understanding the thematic landscape, the network structure, and the distinct communication styles of different communities, organizations can tailor their messages and engagement strategies to effectively reach their target audience and foster a more vibrant and collaborative OI ecosystem.

Future research can build upon these findings by exploring the OI discourse on other social media platforms, adopting a longitudinal approach to capture the evolution of the conversation, and addressing the computational challenges associated with analysing large and complex networks. By continuing to explore the dynamics of OI in the digital age, we can gain a deeper understanding of how social media can be effectively leveraged to promote collaboration, accelerate innovation, and create a more open and interconnected world.

Chapter 6 Case Studies: OI Strategies on Twitter

6.1 Introduction

While quantitative analyses provide a valuable overview of trends and patterns, case studies offer a deeper understanding of how OI principles are implemented and communicated in real-world settings. This chapter shifts from broad quantitative analysis to a more focused qualitative exploration of how specific companies leverage Twitter to promote and enact their OI strategies.

Case studies are particularly valuable in the context of OI research because they allow us to:

- Examine the complexities of real-world OI practices: Quantitative data can reveal general trends, but case studies shed light on the nuances of how companies actually implement OI, including the challenges they face and the strategies they employ.
- Gain insights into the decision-making processes: Case studies provide a glimpse into the factors that influence companies' decisions to engage in OI, the choices they make in terms of partners and methods, and the ways they measure success.
- Develop best practices and recommendations: By analysing successful (and unsuccessful) OI initiatives, case studies can help identify best practices and develop practical recommendations for other companies or organizations.

This chapter focuses on three companies: WazokuCrowd, NineSigma, and PlugandPlayTC. These companies were chosen from a list of the 20 most active users with the #openinnovation hashtag, ensuring they actively engage with the OI community on Twitter. It is important to note that these companies represent

diverse approaches to OI and operate within different segments of the ecosystem. WazokuCrowd operates a crowdsourcing platform, connecting organizations with individual problem solvers globally. NineSigma, in contrast, acts as an OI consulting firm, focusing on connecting large corporations with external sources of innovation, such as research institutions and specialized technology providers. PlugandPlayTC functions as a global innovation platform and startup accelerator, bridging the gap between startups, corporations, and investors. These fundamental differences in their business models, target audiences, and the scope of their OI activities are reflected in their respective Twitter communication strategies. By analyzing these diverse approaches, this chapter aims to provide a richer and more nuanced understanding of how Twitter can be leveraged to support various OI objectives. The following sections will delve into the individual case studies, examining each company's Twitter presence, thematic focus, sentiment and language use, network position, and linguistic cluster alignment.

6.2 Case study 1: WazokuCrowd

6.2.1 Company Profile and Twitter Presence

WazokuCrowd operates a leading crowdsourcing platform that connects organizations with a global network of "solvers" to tackle challenges and generate innovative ideas. Their mission is to democratize innovation by enabling anyone, regardless of their background or expertise, to contribute to solving real-world challenges. Their target audience includes businesses of all sizes, government agencies, and non-profit organizations seeking external expertise and creative solutions.

WazokuCrowd maintains an active presence on Twitter, regularly tweeting about their platform, promoting ongoing challenges, and showcasing success stories from their clients. Their communication style is generally positive and action-oriented, encouraging participation in crowdsourcing initiatives and emphasizing the benefits of this approach. They frequently use hashtags related to OI (#openinnovation), crowdsourcing (#crowdsourcing), challenges (#challenge), and solutions (#solution), reflecting their core business and target audience.

6.2.2 Thematic Focus

Analysis of WazokuCrowd's tweets through the LDA topic modelling reveals a clear focus on three key thematic areas:

- Business, Leadership & Growth: A significant portion of their tweets emphasize the benefits of OI for businesses, aligning with Topic 3 ("Business, Leadership & Growth") identified in the LDA analysis. They promote challenges related to business growth, leadership development, and process improvement, demonstrating how crowdsourcing can help organizations achieve their strategic goals.
- Efficiency, Sustainability & Innovative Solutions: WazokuCrowd frequently promotes challenges related to sustainability, environmental issues, and social impact, reflecting a commitment to using crowdsourcing for positive change and addressing global challenges. This thematic focus aligns with Topic 7 ("Efficiency, Sustainability & Innovative Solutions").
- Design, Robotics & Creativity: WazokuCrowd also highlights their focus on fostering creativity and innovation through their challenges. They often showcase innovative solutions from their community, emphasizing the diversity of expertise within their network and their commitment to pushing the boundaries of what's possible. This theme aligns with Topic 4 ("Design, Robotics & Creativity").

6.2.3 Sentiment and OI Keywords

The sentiment analysis of WazokuCrowd's tweets reveals a consistently positive tone. Their language is often enthusiastic and action-oriented, encouraging participation and collaboration. They frequently use terms like "open innovation," "challenge," "solution," "crowdsourcing," and "make a difference," reflecting their core business and mission of empowering problem-solvers. This positive and action-oriented language likely resonates with their target audience of individuals and organizations seeking innovative solutions and a platform to contribute their ideas.

6.2.4 Network Position and Role

WazokuCrowd occupies a relatively peripheral position in the Combined OI Network, indicating limited direct interaction with other users. This suggests that their communication strategy on Twitter focuses on broadcasting their own

messages and promoting their platform rather than engaging in extensive dialogue or building relationships with other actors. This limited interaction with other actors aligns with a focus on broadcasting their own message and building *bonding social capital* within their own user base, rather than actively engaging in the broader OI network and developing *bridging social capital*. This approach could be a deliberate choice, reflecting their focus on driving traffic to their platform and engaging users within their own ecosystem. However, this strategy could also limit their opportunities to form strategic partnerships and engage with the broader OI community on Twitter.

6.2.5 Linguistic Cluster

WazokuCrowd's tweets predominantly align with the "Tech Enthusiasts & Community Engagement" and "Business & Strategy" clusters. This blend of clusters reflects their dual focus on promoting innovative solutions and highlighting the business benefits of crowdsourcing. Their communication style resonates with both technology enthusiasts and business leaders seeking innovative solutions to drive growth.

6.2.6 SWOT Analysis

Strengths

- **Strong Brand and Platform:** WazokuCrowd benefits from a well-established brand and a user-friendly platform, making it easy for organizations to launch and manage crowdsourcing challenges.
- **Global Reach and Diverse Solver Network:** Their platform connects organizations with a global pool of talent, providing access to diverse perspectives and expertise.
- **Proven Track Record:** They have a successful track record of facilitating impactful crowdsourcing projects, evidenced by numerous case studies and client testimonials.

Weaknesses

- **Reliance on a Single Platform:** Their heavy reliance on their own platform might limit their flexibility to adapt to evolving trends and the emergence of new technologies.

- **Potential for Limited Engagement:** Their focus on promoting their platform could overshadow opportunities to engage in broader OI discussions and build relationships with key influencers.

Opportunities

- **Expanding Services and Partnerships:** Exploring new services, such as OI consulting or workshops, could expand their offerings and attract a broader audience.
- **Leveraging Social Media for Deeper Engagement:** Experimenting with diverse content formats and interactive initiatives, such as Twitter chats or Q&A sessions with experts, could foster a more engaged community.

Threats

- **Competition from Other Crowdsourcing Platforms:** The crowdsourcing market is increasingly competitive, with new platforms emerging and existing platforms expanding their services.
- **Evolving User Preferences:** Changes in user behaviour and preferences on Twitter could impact the effectiveness of their current communication strategy.

6.3 Case Study 2: NineSigma

6.3.1 Company Profile and Twitter Presence

NineSigma is a well-established OI consulting firm that helps corporations and organizations solve their toughest challenges by connecting them with external sources of innovation. They have a long track record of facilitating OI initiatives for clients across various industries, including healthcare, consumer goods, energy, and advanced materials. Their services include technology scouting, OI challenge design and management, partner search, and IP licensing.

On Twitter, NineSigma maintains a professional and informative tone, reflecting their expertise in the field of OI. They often share insights about OI best practices, highlight client success stories, and promote upcoming events or webinars. Their activity level is moderate, and their tweets are generally concise and focused on key aspects of OI.

6.3.2 Thematic Focus

NineSigma's tweets primarily focus on two key themes:

- **Industry-Specific Challenges:** A significant portion of their tweets highlight their involvement in projects and initiatives related to specific industry challenges. They showcase their ability to connect clients with solutions and technologies from around the world, emphasizing their expertise in areas like advanced materials, biotechnology, and sustainable solutions. This thematic focus aligns with the high prevalence of Topic 7 ("Efficiency, Sustainability & Innovative Solutions") in their tweets, as identified through the LDA analysis.
- **Global Collaboration:** NineSigma frequently emphasizes the importance of global collaboration for innovation. Their tweets often promote events and initiatives that connect corporations with international experts, startups, and research institutions. They showcase their network and expertise in facilitating cross-border collaboration and knowledge sharing, aligning with the core principles of OI.

6.3.3 Sentiment and OI Keywords

The sentiment analysis of NineSigma's tweets reveals a generally neutral sentiment. This pragmatic tone reflects their role as consultants, focusing on providing practical solutions to their clients' challenges rather than expressing strong emotions or engaging in hype. Their language is generally informative and solution-oriented, using terms like "innovation," "technology," "solutions," and "challenges." They frequently use the hashtag #InnovationChallenge, highlighting their focus on facilitating challenges as a key mechanism for driving OI.

6.3.4 Network Position and Role

NineSigma occupies a bridging position in the Combined OI Network, connecting different clusters and facilitating knowledge exchange. While their degree centrality is relatively low, indicating limited direct interaction with other users, their strategic connections to key actors within different industries and technology domains suggest a bridging role. They often mention specific organizations or research institutions, highlighting their ability to connect diverse stakeholders within the OI ecosystem. This bridging function aligns with the concept of *bridging social capital* (Bourdieu, 2011), highlighting their ability to connect otherwise disconnected actors and facilitate knowledge transfer across different parts of the OI ecosystem.

However, focusing exclusively on large corporations could be a limitation for NineSigma. While they excel in connecting these corporations with external sources of innovation, this specialization might limit their ability to engage with individual innovators, startups, or smaller companies that could also contribute valuable ideas and solutions. They might be missing opportunities to contribute to a more inclusive and democratized innovation landscape.

6.3.5 Linguistic Cluster

The majority of NineSigma's tweets align with the "Business & Strategy" cluster, reflecting their focus on providing consulting services to large corporations. Their communication style emphasizes the strategic value of OI, showcasing their expertise in navigating complex challenges, and promoting solutions that drive business growth. This alignment with a specific cluster suggests that they are effectively targeting a distinct audience with their communication, focusing on the needs and interests of large corporations seeking to leverage OI for strategic advantage.

6.3.6 SWOT Analysis

Strengths

- **Established Expertise and Network:** NineSigma has a long track record and a well-established network of experts, partners, and clients within the OI space, giving them a strong competitive advantage.
- **Focus on Industry-Specific Challenges:** Their specialization in addressing industry-specific challenges allows them to provide tailored solutions and expertise to their clients.
- **Global Reach:** Their global network and experience in facilitating cross-border collaborations enable them to connect clients with a wide range of potential solutions.

Weaknesses

- **Limited Engagement with Individuals and Startups:** Their focus on large corporations might limit their reach and engagement with individual innovators or smaller companies, potentially missing out on valuable contributions.

- **Relatively Low Visibility on Twitter:** Their relatively low degree centrality suggests that they could benefit from increasing their visibility and engagement within the broader OI conversation on Twitter.

Opportunities

- **Expand Engagement with a Wider Audience:** Developing strategies to engage with individual innovators and startups could broaden their reach and tap into a wider pool of talent and ideas.
- **Increase Visibility and Thought Leadership:** Actively participating in relevant discussions and sharing insights on emerging trends in OI could enhance their visibility and establish them as thought leaders in the field.

Threats

- **Competition from Other Consulting Firms:** The OI consulting market is becoming increasingly competitive, with new players emerging and existing players expanding their services.
- **Shifting Trends in OI:** The evolving nature of OI and the emergence of new technologies and platforms require constant adaptation and innovation to stay relevant and competitive.

6.4 Case Study 3: PlugandPlayTC

6.4.1 Company Profile and Twitter Presence

PlugandPlayTC is a global innovation platform that connects startups, corporations, and investors to foster collaboration and accelerate innovation. They operate a network of industry-specific accelerator programs, providing startups with mentorship, resources, and connections to potential partners and investors. Their target audience is diverse, encompassing startups, corporations, and investors seeking to drive innovation and create new business opportunities. PlugandPlayTC's Twitter presence is characterized by a dynamic and engaging tone, reflecting their active role in the global startup ecosystem. They frequently promote their accelerator programs, showcase the achievements of their portfolio companies, share insights about innovation trends, and highlight their events and partnerships. Their tweets often include visuals, such as images and videos, to capture attention and make their content more engaging. They maintain a high level of activity, frequently posting and interacting with other users.

6.4.2 Thematic Focus

Thematic analysis of PlugandPlayTC's tweets reveals a consistent focus on three key areas:

- **Startup Acceleration:** They actively promote their accelerator programs, showcasing their expertise in supporting startup growth and connecting startups with valuable resources. They frequently tweet about the application process, the benefits of participating in their programs, and the successes of their alumni companies. This focus aligns with Topic 5 ("Innovation Events & Conferences"), highlighting their commitment to providing opportunities for startups to connect, learn, and grow.
- **Corporate Innovation:** PlugandPlayTC also promotes their corporate innovation programs, helping established companies embrace OI principles and build relationships with startups. They often showcase collaborations between startups and corporations, emphasizing the mutual benefits of these partnerships. Their tweets often highlight the role of OI in driving corporate transformation and staying ahead of the competition, aligning with Topic 3 ("Business, Leadership & Growth").
- **Ecosystem Building:** A significant portion of their tweets is dedicated to promoting events, summits, and initiatives that bring together startups, corporations, investors, and other stakeholders in the innovation ecosystem. They emphasize the importance of collaboration, knowledge sharing, and building a vibrant and supportive community. This focus on ecosystem building aligns with the principles of Network Theory, recognizing the value of interconnectedness and the role of social capital in driving innovation.

6.4.3 Sentiment and OI Keywords

The sentiment analysis of PlugandPlayTC's tweets reveals a consistently positive sentiment, reflecting their enthusiasm and support for the startup ecosystem and the transformative potential of innovation. Their language is often collaborative and optimistic, using terms like "open innovation," "startup," "corporate innovation," "collaboration," "ecosystem," "growth," and "future" to highlight their commitment to building a thriving OI ecosystem.

6.4.4 Network Position and Role

PlugandPlayTC occupies a central position in the Combined OI Network, with high degree centrality, indicating a high level of interaction with other users. Their connections span a diverse range of actors, including startups, corporations, investors, and media outlets, reflecting their role as a connector and facilitator within the OI ecosystem. They actively engage in conversations, respond to tweets, and promote collaborations, fostering a dynamic and interactive community around their brand. Their high level of activity and diverse connections suggest a strong capacity for developing both *bridging and bonding social capital*. They not only connect different actors within the OI ecosystem, but also foster a strong sense of community and shared identity among their followers.

However, building and managing a global innovation ecosystem is a complex and challenging endeavor. PlugandPlayTC might face challenges in maintaining the quality and cohesion of this ecosystem, ensuring that the connections they facilitate translate into tangible outcomes and successful collaborations.

6.4.5 Linguistic Cluster

PlugandPlayTC's tweets primarily align with the "Tech Enthusiasts & Community Engagement" and "Business & Strategy" clusters. This blend of communication styles reflects their dual focus on nurturing a vibrant startup community and fostering corporate innovation. They effectively bridge the gap between technology enthusiasts and business leaders, creating a space where both audiences can connect and engage with their shared interest in innovation.

6.4.6 SWOT Analysis

Strengths

- **Global Network and Reach:** PlugandPlayTC has a vast global network of startups, corporations, and investors, providing them with access to a diverse range of resources, expertise, and opportunities.
- **Strong Brand and Reputation:** They have established a strong brand and reputation as a leading innovation platform, attracting top talent and attracting interest from major corporations.

- **Industry-Specific Expertise:** Their network of industry-specific accelerator programs allows them to provide tailored support and guidance to startups in various sectors.

Weaknesses

- **Reliance on Physical Events:** Their strategy relies heavily on physical events and summits, which can be disrupted by unforeseen circumstances (e.g., pandemics, travel restrictions).
- **Competition from Other Accelerators:** The startup accelerator market is highly competitive, with many players vying for the attention of startups, corporations, and investors.

Opportunities

- **Leveraging Digital Platforms:** Expanding their digital presence and developing online programs could broaden their reach and impact.
- **Strategic Partnerships:** Forging strategic partnerships with key players in the innovation ecosystem could further enhance their reach and resources.

Threats

- **Economic Downturn:** A global economic downturn could impact investment in startups and corporate innovation initiatives, potentially affecting their growth.
- **Emerging Technologies:** The rapid pace of technological advancements requires constant adaptation and innovation to stay ahead of the curve.

6.5 Comparing the Case Studies

The three case studies – WazokuCrowd, NineSigma, and PlugandPlayTC – provide a valuable snapshot of the diverse approaches to OI communication on Twitter. While each company has a unique business model and target audience, their strategies share common threads: leveraging Twitter to promote OI, connect with key stakeholders, and build a community around their brand. Their approaches differ in terms of thematic focus, communication style, network position, and linguistic cluster alignment, reflecting their specific goals and target audiences.

6.5.1 A Tapestry of Strategies

WazokuCrowd focuses on democratizing innovation through crowdsourcing, targeting a broad audience of businesses and individuals seeking innovative solutions. Their communication style is action-oriented and emphasizes the positive impact of crowdsourcing, aligning with the "Tech Enthusiasts & Community Engagement" and "Business & Strategy" clusters. Their network position is relatively peripheral, reflecting their focus on promoting their platform and engaging users within their own ecosystem.

NineSigma, as an OI consulting firm, targets large corporations and organizations, emphasizing their expertise in connecting clients with external sources of innovation. Their communication style is more formal and informative, reflecting their professional approach and alignment with the "Business & Strategy" cluster. They occupy a bridging position in the network, facilitating connections between different communities.

PlugandPlayTC, a global innovation platform and startup accelerator, targets a diverse audience of startups, corporations, and investors. Their communication style is dynamic and engaging, reflecting their active involvement in the startup ecosystem. They occupy a central position in the network, serving as a hub for connecting diverse actors and fostering a vibrant OI community.

6.5.2 Shared Objectives

Despite their diverse approaches, the three companies share common objectives in their use of Twitter for OI communication. All three companies actively promote the benefits of OI, showcase successful examples of OI initiatives, and seek to build a community around their brand. They recognize the power of social media to connect with stakeholders, disseminate information, and foster a collaborative environment for innovation.

6.5.3 Comparative Overview

Table 6.1 provides a comparative overview of the three companies, highlighting their key characteristics and strategies:

Table 7: Comparative Overview of OI Strategies on Twitter

Company	*Business Model*	*Target Audience*	*Thematic Focus*	*Sentiment*	*Network*	*Linguistic Cluster*

					Position	
WazokuCrowd	Crowdsourcing Platform	Businesses, Individual Innovators	OI Challenges, Sustainability, Technology	Positive	Peripheral	Tech Enthusiasts & Community Engagement, Business & Strategy
NineSigma	OI Consulting	Large Corporations, Organizations	Industry-Specific Challenges, Global Collaboration	Neutral	Bridging	Business & Strategy
PlugandPlayTC	Startup Accelerator	Startups, Corporations, Investors	Startup Acceleration, Corporate Innovation, Ecosystem Building	Positive	Central Hub	Tech Enthusiasts & Community Engagement, Business & Strategy

6.6 Discussion

The analysis of these three case studies reveals valuable insights and best practices for organizations seeking to leverage Twitter for effective OI communication.

6.6.1 Key Insights from the Case Studies

- **Clarity and Conciseness:** Given the limitations of Twitter's character count, communicating messages in a clear, concise, and easily digestible manner is crucial. NineSigma exemplifies this practice, using precise

language to highlight their expertise and convey key messages effectively.

- **Strategic Hashtag Use:** Hashtags play a vital role in increasing tweet discoverability and reaching the target audience. WazokuCrowd demonstrates a strategic approach, using a combination of general and specific hashtags to reach both a broad audience and a more specialized community.
- **Compelling Content:** Providing valuable content that resonates with the target audience is essential for attracting attention and driving engagement. PlugandPlayTC effectively shares articles, insights, and success stories that highlight the benefits and impact of OI, attracting interest from startups, corporations, and investors.
- **Community Building:** Fostering a sense of community around the brand is crucial for building a loyal following and encouraging interaction. PlugandPlayTC actively engages with their followers, responds to tweets, and promotes collaborations, creating a sense of belonging within their online community.
- **Consistency and Authenticity:** Maintaining a consistent brand voice and tone is crucial for building trust and credibility. Authenticity is paramount, and companies should avoid overly promotional or generic language. WazokuCrowd consistently uses an inclusive and action-oriented tone, reflecting their commitment to "democratizing innovation."

6.6.2 Adapting Strategies to Different Contexts

While these best practices provide a valuable framework for OI communication on Twitter, it's essential for organizations to tailor their strategies to their specific goals, target audience, and resources. For example, a startup with limited resources might focus on building a strong online community and engaging in relevant Twitter chats to increase their visibility and connect with potential partners. A large corporation, on the other hand, might leverage its resources to host Twitter events, sponsor challenges, or collaborate with influencers to promote its OI initiatives and attract a wider audience.

6.6.3 Limitations of the Case Study Approach

This case study analysis provides valuable insights into the real-world application of OI communication on Twitter. However, it's crucial to acknowledge the limitations of this approach. The small sample size (three companies) limits the generalizability of the findings. Additionally, the focus on Twitter might not capture the full spectrum of OI communication, as other platforms, like LinkedIn, play a significant role in professional networking and knowledge exchange. Moreover, the companies selected for this study represent only a subset of the linguistic clusters and thematic topics identified in the previous chapters. Future research could expand this analysis to include a more diverse range of companies, representing different communication styles and areas of focus within the OI discourse on Twitter.

6.7 6.7 Conclusion

This chapter has provided a nuanced and insightful exploration of OI strategies on Twitter, examining the diverse approaches taken by three companies: WazokuCrowd, NineSigma, and PlugandPlayTC. By combining quantitative findings with qualitative analysis, this chapter highlighted how these companies leverage Twitter to achieve their unique OI goals.

The analysis revealed the importance of tailoring communication strategies to specific objectives and audiences. WazokuCrowd, with its focus on democratizing innovation, adopts a more inclusive and action-oriented approach. NineSigma, as a consulting firm, emphasizes its expertise and connects clients with external resources. PlugandPlayTC, as a global accelerator, focuses on building a vibrant ecosystem that connects startups, corporations, and investors.

Despite their diverse approaches, the three companies demonstrate several best practices for OI communication on Twitter, including clarity and conciseness, strategic hashtag use, compelling content, community building, and consistency and authenticity. By adhering to these best practices and learning from the successes and challenges of these companies, other organizations can effectively leverage Twitter to advance their own OI goals, engage with relevant communities, and contribute to a more collaborative and innovative future.

Chapter 7 Impact and Engagement Measurement

7.1 Introduction

This chapter delves into the quantitative dimension of OI communication on Twitter, focusing on the impact and engagement of tweets related to innovation. While previous chapters explored the thematic content and network structures of this discourse, this chapter seeks to understand the reach and resonance of these messages, providing empirical evidence for the effectiveness of different communication strategies.

Specifically, this chapter investigates:

1. Are there differences in impact and engagement between tweets explicitly using the #openinnovation hashtag and those using only #innovation?
2. Do the linguistic clusters, identified in Chapter 5, exhibit variations in impact and engagement?
3. Are specific topics associated with higher levels of impact and engagement?
4. What is the relationship between various tweet characteristics and user engagement, considering factors beyond just hashtag use?

Through rigorous quantitative analysis, we will uncover the multifaceted nature of effective communication on Twitter, considering the interplay between reach, interaction, and content. The findings of this chapter hold valuable insights for practitioners seeking to optimize their OI communication strategies on the platform.

7.2 Impact and Engagement Metrics

This study utilizes a combination of impact and engagement metrics to quantify the reach and resonance of OI-related tweets on Twitter. These metrics are not only crucial for assessing the effectiveness of general communication strategies, as highlighted by Töllinen and Karjaluoto (2011), but they are particularly relevant in the context of OI, where success often hinges on the ability to engage a diverse community of stakeholders, foster collaboration, and disseminate knowledge widely. Impact, in this context, reflects the visibility and potential reach of a message, signifying its potential to influence a broader audience of potential collaborators, partners, and contributors to the OI process. Engagement, on the other hand, highlights the level of interaction and response generated, showcasing how actively the audience participates in the discourse and contributes to the collaborative nature of OI.

The following metrics, readily available in the Twitter data, are employed in this analysis:

- Impact:
    - *viewCount*: This metric captures the number of times a tweet has been viewed, providing a measure of its overall visibility and potential reach. A high view count suggests that the message has reached a wide audience, increasing its potential for impact.
- Engagement:
    - *replyCount*: The number of direct replies a tweet has received, indicating active conversation and engagement with the tweet's content. Replies demonstrate a willingness to participate in the conversation and engage directly with the author.
    - *retweetCount*: The number of times a tweet has been retweeted. Retweets significantly expand a tweet's reach, as they appear on the retweeter's followers' timelines, amplifying the message and increasing its visibility within the network.
    - *likeCount*: The number of times a tweet has been liked. Likes serve as a simple yet powerful indicator of approval and engagement with the tweet's content. They suggest that users found the message relevant or engaging.
    - *quoteCount*: The number of times a tweet has been quoted. Quotes allow users to share a tweet with their own commentary, adding to the conversation or providing their own perspective on the

topic. A high quote count suggests that the tweet has sparked further discussion and debate.

- *bookmarkedCount*: The number of times a tweet has been bookmarked. Bookmarks indicate that users found the tweet valuable enough to save for future reference, suggesting a deeper level of engagement and interest in the content.

By analysing these diverse metrics, this study aims to gain a comprehensive understanding of how users are interacting with OI content on Twitter and to identify the factors that contribute to both the visibility and the engagement of these messages. It is important to acknowledge that while impact and engagement are interconnected, a high view count does not always translate to high engagement, and vice versa. A tweet might be seen by many users but fail to resonate with them or prompt any interaction. Conversely, a tweet with a lower view count might spark a lively discussion within a smaller, more engaged community. Therefore, analysing both impact and engagement is crucial for a holistic understanding of communication effectiveness.

7.3 Comparing #innovation and #openinnovation

This section directly addresses the question of whether explicitly using the #openinnovation hashtag leads to a measurable difference in the impact and engagement of tweets compared to those using only the broader #innovation hashtag. To answer this question, a series of t-tests were conducted to compare the means of the six key metrics (viewCount, replyCount, retweetCount, likeCount, quoteCount, and bookmarkedCount) between the two groups of tweets.

Table 8 presents the results of these t-tests.

Table 8: T-test Results for Impact and Engagement Metrics

Metric	*T-statistic*	*P-value*	*Interpretation*
viewCount	0.298	0.765	Not Significant
replyCount	0.763	0.445	Not Significant
retweetCount	0.097	0.922	Not Significant
likeCount	0.057	0.954	Not Significant
quoteCount	0.119	0.904	Not Significant

bookmarkedCount	0.019	0.984	Not Significant

The t-tests reveal no statistically significant differences in any of the measured metrics. This finding indicates that, within this dataset, the explicit use of the #openinnovation hashtag does not appear to be a statistically significant driver of higher levels of impact or engagement. However, this analysis focuses solely on statistical significance. The absence of significant differences does not necessarily negate the importance of other factors, such as the content of the tweets or the characteristics of the users engaging with them. The following sections will delve deeper into these nuances, exploring potential trends and patterns in impact and engagement across different linguistic clusters and tweet characteristics.

7.4 Exploring the Impact of Tweet Characteristics on Engagement

While the t-tests presented in Section 7.3 provided a basic comparison of impact and engagement based on hashtag use, this section delves deeper into the factors influencing user engagement with OI-related tweets by employing regression analysis. This approach allows us to examine the relationship between various tweet characteristics and user engagement metrics, providing a more nuanced understanding of the drivers of engagement on Twitter.

7.4.1 Regression Model Overview

Ordinary Least Squares (OLS) regression, a statistical technique for modelling the relationship between a dependent variable and one or more independent variables, was employed to analyse the impact of various tweet characteristics on user engagement. The analysis focused on six engagement metrics as dependent variables: view Count, reply Count, retweet Count, like Count, quote Count, and bookmarked Count. Each metric represents a different dimension of user engagement, providing a comprehensive view of how users interact with tweets.

7.4.2 Independent Variables and Model Specification

The following independent variables were included in the regression models:

- **cluster:** A categorical variable representing the four distinct tweet clusters (0-3) identified through the language-based cluster analysis in Chapter 5. This variable captures the potential influence of communication style and thematic focus on engagement.

- **topic:** A categorical variable representing the eight different tweet topics (0-7) identified through the LDA topic modelling analysis in Chapter 5. This variable accounts for the potential impact of topic relevance on user engagement.
- **main_type:** A categorical variable representing the type of innovation hashtag used in the tweet: 0 for "#innovation" and 1 for "#openinnovation." This variable directly addresses the research question about the impact of hashtag choice on engagement.
- **user.followersCount:** A continuous variable representing the number of followers a user has. This variable captures the potential reach and influence of the user.
- **weekday:** A categorical variable representing the day of the week the tweet was posted (0-6, Monday-Sunday). This variable accounts for potential variations in user behaviour and engagement patterns throughout the week.
- **hour:** A categorical variable representing the hour of the day the tweet was posted (0-23). This variable allows for the exploration of potential time-of-day effects on engagement.
- **has_link:** A boolean variable indicating the presence of a link within the tweet. This variable captures the potential impact of including external links on engagement.
- **has_image:** A boolean variable indicating the presence of an image within the tweet. This variable accounts for the potential influence of visual content on engagement.

While using two separate dummy variables for #innovation and #openinnovation could be considered, the current model specification uses a single variable (main_type) to represent the hashtag type. The model focuses on overall effects and interactions with other variables, and preliminary analyses indicated multicollinearity issues when including both hashtags as separate variables.

To address this issue, subsequent model iterations explored specific interaction models and focused on analysing the main effects of individual variables, particularly those related to the research questions of this chapter.

7.4.3 Results

Despite the initial multicollinearity issues, the regression analyses revealed valuable insights into the factors influencing user engagement with OI-related tweets.

Firstly, across various engagement metrics, the interaction effects of main_type (hashtag type) with cluster and topic were not statistically significant. This suggests that the choice of #innovation or #openinnovation did not significantly alter the relationship between these variables and engagement. However, descriptive analysis revealed subtle differences in average engagement based on the hashtag used. For example, tweets with #openinnovation tended to exhibit slightly higher average view counts compared to those with #innovation.

Secondly, the number of followers a user has (user.followersCount) consistently emerged as a highly significant predictor of engagement across all metrics. This finding underscores the importance of audience size in determining the reach and impact of tweets. A larger audience translates to greater potential visibility and a higher likelihood of generating interactions.

Thirdly, the inclusion of links in tweets (has_link) was found to have a significant negative impact on engagement. For example, the coefficient for has_link in the model predicting viewCount was -418.45 (p-value < 0.001), indicating that tweets containing links received, on average, 418 fewer views compared to those without links. This finding suggests that users on Twitter might be more likely to engage with self-contained content and less inclined to click on external links.

Finally, the presence of images in tweets (has_image) was associated with a positive effect on specific engagement metrics, particularly replyCount and likeCount. This finding suggests that incorporating visual content can enhance engagement, potentially by making tweets more visually appealing or by sparking conversation and discussion.

7.4.4 Discussion of Findings

The regression analysis provides valuable insights into the factors driving user engagement with OI-related tweets. The consistent influence of user follower count across all engagement metrics underscores the importance of building a large and engaged audience on Twitter. As Xu and Saxton (2019) emphasize, a strong online social capital, built through active engagement and relationship building, is crucial for achieving mission-related outcomes on social media platforms. In the context of OI, this implies that companies and organizations

should prioritize cultivating a strong online presence and actively engaging with their followers to amplify their message and encourage participation in OI initiatives.

The negative impact of links on engagement suggests that users on Twitter might prefer engaging with self-contained content. This finding aligns with the observation that users on social media platforms often have limited attention spans and are more likely to engage with content that is easily digestible and readily available within the platform. Companies should prioritize crafting tweets that provide value directly within Twitter, rather than relying heavily on external links. However, strategic use of links to relevant resources or calls to action could still be effective if the tweet itself provides sufficient context and value to capture the user's attention.

The positive effect of images on reply and like counts suggests that incorporating visuals can enhance engagement, potentially by making tweets more appealing and encouraging interaction. This finding is consistent with research on the role of visual communication in online environments (Valentini et al., 2018). Images can capture attention, convey information quickly, and evoke emotional responses, making them a powerful tool for increasing engagement on social media.

The analysis of engagement across linguistic clusters revealed nuanced differences in communication styles and preferences. This aligns with the principles of audience segmentation and message tailoring, emphasizing the importance of adapting communication strategies to resonate with different target audiences (Rimal & Adkins, 2003). Understanding the unique characteristics, interests, and communication styles of each cluster can help organizations craft more targeted and effective messages that are more likely to generate engagement and achieve their communication goals.

7.5 Discussion

This section synthesizes the findings from the quantitative analyses presented in this chapter, exploring the apparent paradox between the importance of OI and the lack of statistically significant differences in impact and engagement based solely on hashtag use. It also delves into the practical implications of the results for organizations seeking to leverage Twitter for effective OI communication.

7.5.1 The Impact-Engagement Paradox

The findings presented in this chapter reveal an interesting paradox. While OI has gained significant traction as a critical approach to innovation, and Chapter 4 demonstrated an overall positive sentiment towards OI on Twitter, the quantitative analysis of impact and engagement metrics did not reveal a statistically significant advantage for tweets explicitly using the #openinnovation hashtag.

This seeming contradiction raises questions about the role of Twitter in facilitating and promoting OI. Could it be that the general conversation about innovation on Twitter has already incorporated the principles and practices of OI, making the explicit use of the #openinnovation hashtag less relevant?

Several factors could contribute to this phenomenon. Firstly, the widespread adoption of OI principles and language across various industries might have blurred the lines between traditional innovation and OI, leading to the integration of OI concepts into broader discussions about innovation.

Secondly, the reliance on specific hashtags as the sole indicator of OI discussions might be limiting. Users could be engaging in conversations about collaboration, knowledge sharing, and external partnerships – core tenets of OI – without explicitly using the #openinnovation hashtag.

Finally, the dynamics of social media communication, with its emphasis on speed, brevity, and virality, might favour content that is easily digestible and shareable, potentially overshadowing nuanced discussions about specific approaches to innovation like OI.

7.5.2 Strategic Insights for Practitioners

Despite the lack of statistically significant differences between #innovation and #openinnovation in terms of overall impact and engagement, the findings of this chapter offer valuable insights for organizations seeking to leverage Twitter for OI.

Firstly, the regression analysis highlights the importance of audience size. The consistent influence of user follower count across all engagement metrics underscores the need for building a large and engaged following on Twitter. Companies seeking to maximize the reach and impact of their OI communication should prioritize strategies for growing their follower base and fostering a community around their brand.

Secondly, the content of the tweet matters. While the type of innovation hashtag used did not significantly alter the relationship between other variables and engagement, the choice of topic and the inclusion of visual elements can influence user interaction. Companies should carefully consider the relevance of their content to their target audience and experiment with different formats, such as images and videos, to enhance engagement.

Thirdly, the negative impact of links on engagement suggests that users on Twitter might prefer engaging with self-contained content. Companies should prioritize crafting tweets that provide value directly within the platform, rather than relying heavily on external links. However, strategic use of links to relevant resources or calls to action could still be effective if the tweet itself provides sufficient context and value to capture the user's attention.

Finally, the analysis of engagement across linguistic clusters revealed nuanced differences in communication styles and preferences. Organizations should tailor their messages and engagement strategies to resonate with the specific characteristics of each cluster. For example, when targeting the "Tech Enthusiasts & Community Engagement" cluster, a more informal and enthusiastic tone, emphasizing the transformative potential of emerging technologies, might be more effective. Conversely, when communicating with the "Strategic & Collaborative OI" cluster, a more formal and professional approach, focusing on the strategic benefits of OI and showcasing successful collaborations, would likely be more appropriate.

By understanding the interplay between reach, interaction, and content, and by adapting their communication strategies to the specific dynamics of the OI discourse on Twitter, organizations can enhance the effectiveness of their messages, foster meaningful engagement, and contribute to a more vibrant and collaborative OI ecosystem.

7.6 Limitations and Future Research Directions

While this chapter provides valuable insights into the impact and engagement of OI communication on Twitter, it is essential to acknowledge its limitations. Understanding these limitations is crucial for interpreting the findings within a proper context and for identifying avenues for future research.

7.6.1 Addressing the Boundaries of the Study

- **Generalizability:** The findings of this study are based on a specific dataset of tweets collected within a limited timeframe. It is important to recognize that the dynamics of online communication can vary significantly across platforms and over time. Therefore, the results might not be generalizable to other social media platforms or to the broader landscape of OI discourse beyond Twitter.
- **Hashtag Reliance:** The study relies solely on the presence of the hashtags #innovation and #openinnovation to identify relevant tweets. As discussed in previous chapters, this approach might have excluded relevant conversations that use alternative terms or phrases to discuss OI. This limitation could have contributed to the lack of statistically significant differences observed between the two hashtags.
- **Cross-Sectional Data:** This research utilizes a cross-sectional dataset, capturing a snapshot of the OI discourse on Twitter during a specific period. This approach limits the ability to analyse trends and changes in communication patterns over time. A longitudinal study, tracking the evolution of OI discourse over an extended period, would provide a richer and more dynamic understanding of this phenomenon.
- **Multicollinearity:** The regression analysis faced challenges related to multicollinearity, potentially stemming from high correlations between certain independent variables. While efforts were made to address this issue through model simplification and analysis of interaction effects, it is important to acknowledge that multicollinearity might have influenced the interpretation of the regression coefficients.

7.6.2 Avenues for Future Exploration

The limitations of this study point to several promising avenues for future research.

- **Comparative Analyses:** Future studies could explore the OI discourse on other social media platforms, such as LinkedIn, Facebook, or specialized OI platforms like NineSigma or Wazoku. Comparative analyses across platforms could reveal platform-specific dynamics, communication styles, and factors influencing impact and engagement.
- **Expanding Keyword Scope:** Future research could expand the scope of data collection by incorporating a wider range of keywords and phrases

related to OI. This could involve using a combination of keyword searches, topic modelling techniques, or natural language processing algorithms to identify relevant tweets.

- **Longitudinal Studies:** A longitudinal study, tracking the OI discourse on Twitter over an extended period, would allow for the analysis of trends, changes in communication patterns, and the impact of specific events or interventions. This approach would provide a more dynamic and nuanced understanding of the evolution of OI communication.
- **Advanced Statistical Techniques:** Future research could employ more advanced statistical techniques to address the challenges of multicollinearity and to explore complex interactions between variables. Techniques like principal component analysis or ridge regression could help mitigate multicollinearity, while interaction terms and non-linear models could capture more nuanced relationships between tweet characteristics and engagement.
- **Qualitative Insights:** Complementing the quantitative analysis with qualitative methods, such as interviews with key actors in the OI ecosystem or content analysis of highly engaging tweets, could provide a deeper understanding of the motivations, strategies, and challenges involved in communicating OI on Twitter.

7.7 Conclusion

This chapter has provided a quantitative exploration of the impact and engagement of OI communication on Twitter, utilizing a combination of t-tests and regression analysis to examine the relationship between hashtag use, tweet characteristics, and user interaction metrics.

While the analysis did not reveal statistically significant differences between #innovation and #openinnovation in terms of overall impact and engagement, the findings offer valuable insights into the factors that influence engagement with OI-related content on Twitter. The regression analysis highlighted the importance of audience size, the use of visual content, the impact of including links, and the nuances of communication styles across different linguistic clusters.

Despite the limitations of relying solely on hashtags and the cross-sectional nature of the data, this study provides a valuable contribution to the growing body of research on OI communication in the digital age. The findings have practical implications for organizations seeking to leverage Twitter for OI initiatives,

providing data-driven insights to guide their communication strategies and maximize their reach and engagement.

Future research should address the limitations of this study by exploring OI communication on other platforms, utilizing a broader range of keywords and phrases, adopting a longitudinal approach, and employing more advanced statistical techniques. By continuing to investigate the complex interplay between social media and OI, we can foster a deeper understanding of how this technology can be leveraged to promote collaboration, accelerate innovation, and create a more open and interconnected world.

Chapter 8 Predicting View Count: A Deep Learning Approach to Optimizing OI Communication

8.1 Introduction

This chapter presents a novel approach to understanding and predicting the reach of OI communication on Twitter, leveraging the power of deep learning to model the complex relationships between user characteristics, tweet content, and temporal factors. Previous chapters have explored various facets of OI discourse on Twitter using traditional statistical methods, such as sentiment analysis, thematic analysis, and linear regression. While these methods have provided valuable insights, they have also revealed limitations in their ability to fully capture the intricate interplay of factors influencing tweet visibility, as evidenced by the consistently low R-squared values obtained in the regression models.

Recognizing these limitations, this chapter explores the potential of neural networks, a powerful class of machine learning algorithms capable of learning complex, non-linear relationships in data. Neural networks have gained prominence in recent years, particularly in the field of natural language processing, due to their ability to extract meaningful patterns from unstructured text data and to make accurate predictions (Devlin et al., 2019).

The primary objective of this chapter is to present the architecture, training process, and results of a neural network model specifically designed to predict the viewCount of OI -related tweets. By incorporating user-level features, temporal variables, and tweet embeddings generated by BERT, the model aims to provide a more accurate and nuanced understanding of the factors driving tweet visibility.

The practical implications of this model are significant. By predicting the viewCount of tweets, companies and organizations can optimize their communication strategies on Twitter, maximizing the reach and impact of their messages about OI. The model can assist in identifying the optimal timing for publishing tweets, tailoring content to specific audiences, and understanding the potential impact of different communication strategies.

8.2 Model Architecture

The neural network model developed for this research utilizes a feedforward architecture, consisting of multiple layers of interconnected nodes (neurons) that process information in a hierarchical manner. The model takes as input a combination of numerical and categorical variables, extracts meaningful features through hidden layers, and outputs a single value representing the predicted viewCount of a tweet.

The model consists of five dense layers, each with a decreasing number of neurons, followed by a single output layer. The input layer receives four distinct input features:

- log(user.followersCount): This numerical variable represents the logarithm of the user's follower count, a crucial indicator of the user's potential reach and influence on the platform. The logarithmic transformation is applied to mitigate the impact of outliers and to normalize the distribution of this variable.
- weekday: This categorical variable represents the day of the week when the tweet was published. It is encoded using one-hot encoding, a technique that transforms categorical variables into a binary vector representation, allowing the network to process this information effectively.
- hour: Similar to weekday, this categorical variable represents the hour of the day when the tweet was published and is also encoded using one-hot encoding.
- embeddings: This variable represents the 768-dimensional embedding vector generated by BERT for each tweet. The embedding captures the semantic content of the tweet, allowing the network to learn from the language and context of the message.

The first hidden layer consists of 512 neurons and employs the Leaky ReLU activation function, a variation of the ReLU (Rectified Linear Unit) activation

function that allows for small, non-zero gradients for negative input values, potentially improving the network's ability to learn complex patterns. The remaining four hidden layers consist of 256, 128, 64, and 32 neurons, respectively, also using the Leaky ReLU activation function. These layers progressively extract more abstract and high-level features from the input data, allowing the network to learn complex relationships between the variables.

The final output layer consists of a single neuron with a linear activation function, producing the predicted viewCount value. The linear activation function is appropriate for this task, as the output is a continuous numerical value.

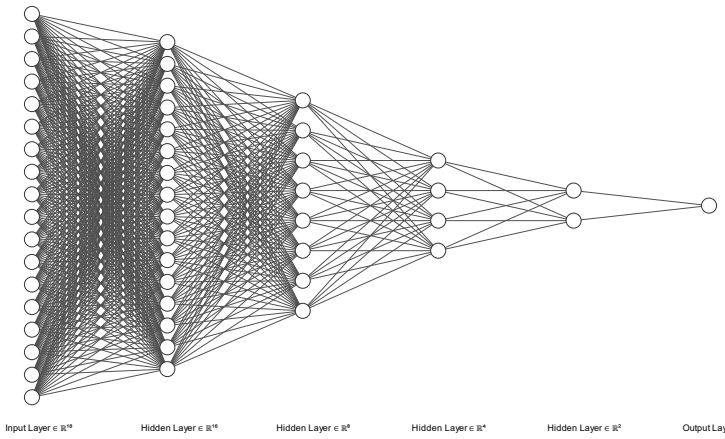

Figure 12: A visual representation of the neural network shape

8.3 Input Data and Pre-processing

Before feeding the data into the neural network, several pre-processing steps are crucial to ensure that the model can effectively learn from the input variables and make accurate predictions. The selection of input variables was guided by a combination of theoretical considerations and practical insights gained from the previous analyses. It's important to note that the presence of specific hashtags (#innovation and #openinnovation) was not included as a separate binary variable in the model. This is because the BERT embeddings, used to represent the semantic content of the tweets, already capture the presence and influence of these hashtags within the context of the tweet.

The model utilizes the following input features:

- **log(user.followersCount):** The logarithm of the user's follower count was used as a proxy for the user's reach and influence on Twitter. Users with

a larger following have a wider audience and are more likely to have their tweets seen by a larger number of people. The logarithmic transformation mitigates the impact of outliers and normalizes the distribution of follower counts, which is typically heavily skewed. This transformation is also consistent with the theoretical understanding that the marginal impact of an additional follower decreases as the total follower count increases.

- **weekday:** This categorical variable represents the day of the week when the tweet was published. It is one-hot encoded to create a binary vector representation, allowing the network to effectively process this information.
- **hour:** Similar to weekday, this categorical variable represents the hour of the day when the tweet was published and is also one-hot encoded.
- **embeddings:** This variable represents the 768-dimensional embedding vector generated by BERT for each tweet. The embedding captures the semantic content of the tweet, allowing the network to learn from the language and context of the message.

The choice of these variables was based on their potential relevance for predicting viewCount, as supported by previous research and descriptive analyses. Pre-processing these variables is crucial to ensure that the neural network can effectively learn from the data. Applying a logarithmic transformation to user.followersCount mitigates the impact of outliers and creates a more normal distribution, which improves the model's performance. One-hot encoding of weekday and hour transforms these categorical variables into numerical representations, which are more suitable for neural network processing. Finally, the BERT embeddings are normalized to ensure that all values fall within a similar range, preventing certain features from dominating the learning process and improving the model's convergence speed.

8.4 Training Process

The neural network model is trained using a supervised learning approach, where the model learns to predict the viewCount of tweets based on labelled examples from the dataset. The dataset is divided into two sets: a training set, used to train the model, and a test set, used to evaluate the model's performance on unseen data.

The training process involves feeding the pre-processed input data into the neural network and adjusting the model's parameters (weights and biases) to minimize the difference between the predicted viewCount values and the actual viewCount values in the training set. This process is repeated iteratively, with the model learning from its errors and gradually improving its predictions.

The following key parameters and configurations were used for training the neural network:

- Batch Size: 32. This parameter determines the number of training examples processed in each iteration of the training process.
- Number of Epochs: 4000. This parameter determines the number of times the entire training set is passed through the network.
- Optimizer: Adam (Adaptive Moment Estimation). This is a popular optimization algorithm that adaptively adjusts the learning rate based on the gradients of the loss function, allowing for efficient convergence.
- Loss Function: Mean Squared Error (MSE). This function measures the average squared difference between the predicted viewCount values and the actual viewCount values.
- Evaluation Metrics: R-squared (R^2), Mean Squared Error (MSE), and Mean Absolute Error (MAE). These metrics provide different perspectives on the model's predictive accuracy.

The model's performance is evaluated on both the training set and the test set, monitoring the loss function and evaluation metrics to track the progress of the training process.

Remarkably, the model did not exhibit any signs of overfitting, even after training for 4000 epochs. Overfitting occurs when the model learns the training data too well, memorizing specific patterns rather than generalizing to unseen data. The absence of overfitting in this case suggests that the model has learned the underlying relationships between the variables and is likely to generalize well to new, unseen tweets.

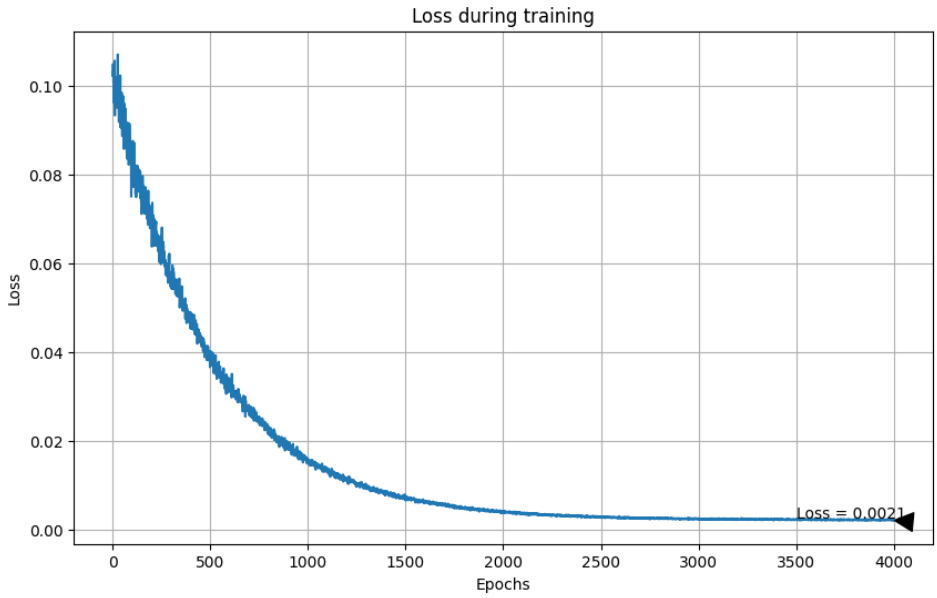

Figure 13: Training Loss history

8.5 Model Results and Evaluation

The neural network model achieved impressive results, demonstrating its ability to predict the viewCount of OI -related tweets with a high level of accuracy. The model achieved an R^2 value of 0.63 on the test set, indicating that it can explain 63% of the variance in viewCount based on the input variables. To further understand the influence of each variable on the model's predictions, a permutation importance analysis was conducted.

Table 8.1 presents the results of the permutation importance analysis. This analysis involves shuffling the values of each input variable and measuring the resulting decrease in the model's performance (in this case, a decrease in R^2). A larger decrease in R^2 indicates a greater importance of the variable in the model's predictions.

Table 9:Permutation Importance of Input Variables

Variable	*Permutation Importance (Decrease in R^2)*
log_followers	0.0566
embeddings	0.0096
weekday	1.44e-05

hour	3.32e-05

As shown in Table 8.1, the logarithm of the user's follower count (log_followers) emerged as the most influential variable, with a permutation importance of 0.0566. This indicates that shuffling the values of log_followers causes the largest decrease in the model's R^2, suggesting that the model heavily relies on this variable to make accurate predictions.

The BERT embeddings (embeddings) also exhibit a notable influence, with a permutation importance of 0.0096, indicating that the semantic content of the tweet plays a significant role in predicting viewCount.

Conversely, the day of the week (weekday) and the hour of the day (hour) have a negligible impact on the model's performance, with permutation importance values close to zero. This suggests that the timing of the tweet might not be as crucial a factor in determining visibility as the content itself and the reach of the user.

As explained in Section 8.3, the presence of the specific hashtags #innovation and #openinnovation was not included as a separate binary variable in the model. This is because the BERT embeddings, which capture the semantic context of the tweets, already account for the influence of these hashtags on the tweet's meaning and potential for engagement.

The dominance of log_followers as a predictor highlights the importance of audience size in determining the visibility of tweets on Twitter. However, the significant influence of the BERT embeddings indicates that the model also considers the content of the tweet, suggesting that crafting engaging and relevant content can contribute to higher view counts, even for users with a smaller following.

The model's strong performance and the insights gained from the permutation importance analysis demonstrate the potential of deep learning for understanding and predicting the dynamics of online communication.

8.6 Discussion

The successful implementation of this neural network model has profound implications for understanding and optimizing OI communication on Twitter.

8.6.1 Model Accuracy

The R^2 of 0.63 achieved by the neural network model, while impressive, also highlights the inherent complexity of OI discourse on Twitter. Despite incorporating several key variables, the model can only explain 63% of the variance in viewCount. This suggests that other factors, not captured in the current model, also play a role in determining tweet visibility. These factors might include:

- Content Quality: The specific wording, tone, and style of the tweet, which are not fully captured by the BERT embeddings.
- Visual Elements: The presence of images or videos, which can significantly enhance engagement and visibility.
- Network Effects: The dynamics of retweets, replies, and mentions, which can amplify the reach of a tweet beyond the initial audience.
- Trending Topics: The alignment of a tweet with current trending topics or events, which can significantly increase its visibility.

8.6.2 Deep Learning for Social Media Analysis

The success of the neural network model in predicting viewCount underscores the power of deep learning for analysing social media data. Traditional statistical methods, such as linear regression, are often limited in their ability to capture the complex, non-linear relationships present in social media data. Neural networks, with their ability to learn from vast amounts of data and to model intricate patterns, offer a more powerful and flexible approach to social media analysis.

This study demonstrates that deep learning can be effectively applied to understand and predict the dynamics of OI communication on Twitter. The findings pave the way for future research to explore other applications of deep learning in this domain, such as:

- Sentiment Analysis: Developing deep learning models that can more accurately capture the nuanced sentiment expressed in tweets, considering factors like irony, sarcasm, and context.
- Topic Modelling: Using deep learning techniques to identify more fine-grained and context-aware topics within OI discussions.
- Network Analysis: Applying deep learning to predict the formation of connections and the flow of information within the OI network.

8.6.3 Optimizing Tweet Timing

Beyond the primary goal of predicting viewCount, this research also explores the potential for the developed neural network model to inform tweet scheduling strategies. While the model's current performance suggests limitations in directly optimizing timing based solely on hour and day, this section investigates, in a preliminary fashion, how the model could be used to explore the relationship between tweet timing and visibility. This is presented as an exploratory analysis to demonstrate the potential applications of the model, rather than a definitive guide for tweet scheduling.

Instead of manually adjusting the day and hour while keeping other variables constant, a more efficient and effective approach is to utilize a grid search technique. This involves fixing the followersCount and using a script to evaluate a range of potential tweets with varying content characteristics, such as different wording, hashtags, and emojis.

Because neural networks operate in batches, testing all possible hours of the week in a single batch for each combination of follower count and embeddings is computationally efficient. This allows for the rapid evaluation of a large number of potential tweets and the identification of the optimal timing for each tweet.

This automated grid search approach provides several advantages:

- **Efficiency:** Evaluating a large number of tweet variations in parallel significantly reduces the time and effort required for optimization.
- **Comprehensiveness:** Testing all possible hours of the week ensures that the optimal timing is identified, accounting for daily and weekly fluctuations in user activity.
- **Data-Driven Decision Making:** The model's predictions provide data-driven insights to guide tweet scheduling, maximizing the potential for reach and engagement.

To address the concern of changing user behaviour and the potential for the model's performance to degrade over time, online learning techniques can be employed. This involves continuously updating the model with new data as it becomes available, allowing it to adapt to shifts in user behaviour and maintain its predictive accuracy. By leveraging the Twitter API's enterprise features or alternative data collection methods, new tweets can be collected and integrated into the model's training data, ensuring that it remains relevant and effective in the dynamic landscape of social media.

It is important to acknowledge that the current model, as highlighted in the permutation importance analysis in Section 8.5, does not identify weekday and hour as significant predictors of viewCount. This suggests that optimizing tweet timing based solely on these two variables might not yield substantial improvements in visibility. This is likely because the optimal timing is influenced by a multitude of complex factors, including the specific topic of the tweet, the activity patterns of the target audience, and trending topics on Twitter, which are not fully captured by the current model. Further research, incorporating a wider range of variables and potentially more sophisticated model architectures, could provide more definitive insights into tweet timing optimization. Furthermore, it should be noted that the primary purpose of the grid search analysis in this research was to evaluate the potential impact of different tweet characteristics, rather than to provide actionable guidelines for tweet scheduling. Therefore, the findings regarding timing should be interpreted as exploratory and preliminary.

8.7 Conclusion

This chapter has demonstrated the power of deep learning for analysing and understanding OI communication on Twitter. The neural network model, with its ability to accurately predict viewCount, provides a valuable tool for companies and organizations seeking to optimize their communication strategies and maximize the reach of their messages.

The success of the model also highlights the increasing importance of data-driven approaches to communication in the digital age. By leveraging advanced analytics and machine learning techniques, organizations can gain deeper insights into the dynamics of online conversations, tailor their messages to specific audiences, and measure the effectiveness of their communication efforts.

Future research should build upon this foundation, exploring other applications of deep learning for analysing OI communication and developing more sophisticated models that incorporate additional variables and data sources. By embracing the power of data and AI, we can unlock new insights and strategies for fostering innovation and collaboration in the digital world.

Chapter 9 Conclusions

9.1 Introduction

This dissertation set out to explore the multifaceted world of OI communication on Twitter. By employing a mixed-methods approach, combining quantitative analysis of Twitter data with qualitative insights from case studies and the development of a predictive model, this research has sought to provide a comprehensive understanding of how individuals and organizations perceive, communicate, and engage with OI on this influential platform. This chapter synthesizes the key findings of this journey, highlighting the theoretical and practical implications, acknowledging the limitations of the study, and proposing promising avenues for future research.

9.2 Summary of Key Findings

This section synthesizes the key findings of this dissertation, directly addressing each research question posed in the introduction and connecting them to the theoretical frameworks established in Chapters 1 and 2.

Research Question 1: What is the overall sentiment expressed towards OI on Twitter? How does it compare to the sentiment surrounding general innovation?

Chapter 4's sentiment analysis, using VADER, revealed predominantly positive sentiment for both #openinnovation and #innovation, suggesting a generally favorable public perception of these concepts on the platform. However, #openinnovation exhibited greater variability in sentiment, indicating a more nuanced and potentially more critical discourse surrounding this approach. This aligns with the complexities and challenges inherent in implementing OI, such as managing intellectual property, navigating collaborations, and fostering an open

organizational culture, as discussed in Chapter 1. The more consistently positive sentiment towards #innovation may be attributed to its broader scope, encompassing a wider range of initiatives and attracting more promotional content. This result provides a novel perspective on public perception of OI in online discussions, emphasizing the need to consider the diverse range of opinions and the potential for critical engagement.

Research Question 2: Are there distinct linguistic patterns associated with OI discussions on Twitter? How do these patterns differ between users who explicitly use the #openinnovation hashtag and those who do not?

The linguistic analysis in Chapter 4, based on OI keywords identified by Toroslu et al. (2023), revealed distinct linguistic patterns. Tweets using #openinnovation featured a significantly higher frequency of specific OI terms (e.g., "co-creation," "crowdsourcing"), signifying a focused and specialized discourse within that community. In contrast, #innovation tweets displayed broader vocabulary, reflecting the more general nature of the term. This difference supports Chapter 2's discussion on communities of practice and the role of shared language in defining these communities. It suggests that the #openinnovation hashtag serves as a marker for a community deeply engaged with the core principles and practices of OI. This finding contributes significantly to our understanding of how specialized communities communicate about OI on Twitter, offering valuable insights for tailoring communication strategies to reach and engage these specific audiences.

Research Question 3: How is the OI ecosystem on Twitter structured? Who are the key influencers, and what are the patterns of interaction between different actors?

Chapter 5's network analysis, using user mentions to map connections, revealed a fragmented OI ecosystem on Twitter. While the combined network for both #openinnovation and #innovation was sparse, indicating limited direct user interaction, the #openinnovation community exhibited higher degree centrality and clustering coefficients. This suggests a more interconnected and collaborative structure, aligning with the network theory principles of strong ties and knowledge sharing discussed in Chapter 2. The difficulty in pinpointing specific key influencers due to computational constraints highlights the need for further research with more robust methods. Despite these limitations, this analysis

provides a valuable first mapping of the OI ecosystem on Twitter, contributing to our understanding of how OI-related discussions are structured and disseminated on social media.

Research Question 4: Do different OI activities on Twitter (e.g., seeking feedback, promoting collaborations, announcing new products) lead to varying levels of engagement and impact?

Chapters 6 and 7 addressed this question by combining case study analysis and quantitative investigation of impact and engagement metrics. The case studies revealed diverse communication strategies aligned with different OI objectives. WazokuCrowd focuses on community engagement through challenges, NineSigma highlights expertise by connecting clients with resources, and PlugandPlayTC fosters a broader ecosystem. However, the quantitative analysis in Chapter 7 found no statistically significant differences in engagement based on specific OI activities, suggesting that factors beyond the type of activity, such as audience size and content quality (explored through regression analysis), play a more significant role. This aligns with the Social Capital Theory discussed in Chapter 2, which emphasizes the importance of building a large and engaged audience (bridging social capital) for maximizing the reach and impact of communication on social media. This integrated approach provides a nuanced understanding of how different OI activities are communicated on Twitter and their relative effectiveness in engaging users.

Research Question 5: Can we develop a predictive model that accurately forecasts the impact of open innovation-related tweets based on factors like user characteristics, tweet content, and temporal variables?

Chapter 8 explored the potential of deep learning for predicting tweet view count as a proxy for impact. The developed neural network model, incorporating follower count, content (BERT embeddings), and temporal variables, achieved an R^2 of 0.60. This demonstrates the applicability of deep learning to analyze social media data for OI research, as discussed in Chapter 3, and offers a practical tool for optimizing communication strategies. While the model's performance is promising, limitations like the strong reliance on follower count and the exclusion of visual elements suggest avenues for future research, such as incorporating image analysis, exploring alternative network architectures, or considering other features related to content quality and network effects. This study contributes to

the nascent field of predictive analytics for OI communication, providing a valuable foundation for future research and development of more sophisticated models.

9.3 Contributions of the Research

This dissertation makes several significant contributions to the growing body of research on OI, social media, and communication strategies. By employing a multi-method approach, this research provides a more comprehensive and nuanced understanding of the OI conversation on Twitter than studies relying on single-method approaches. The identification of distinct linguistic clusters and communication styles within the OI community offers a new perspective on the diversity of actors and their interactions. The development of a neural network model for predicting tweet view count demonstrates the potential of deep learning for understanding and optimizing social media communication. Finally, the identification of best practices and the discussion of practical implications provide valuable guidance for companies and organizations seeking to leverage Twitter for successful OI.

9.4 Practical Implications

This dissertation offers valuable practical implications for companies and organizations seeking to leverage Twitter for OI. By understanding the nuances of online communication revealed through this research, companies can develop more effective strategies for engaging with key stakeholders and promoting their OI initiatives.

- **Target Specific Communities:** The identification of distinct linguistic clusters and thematic communities underscores the need to tailor communication strategies to specific audiences. Companies should identify the communities most relevant to their innovation goals and adapt their language, content, and engagement tactics accordingly. For example, a company targeting the "Tech Enthusiasts & Community Engagement" cluster should use an informal and enthusiastic tone, while a company targeting the "Business & Strategy" cluster should focus on a more formal and professional approach.
- **Build Relationships with Influencers:** The network analysis revealed the influence of key actors in the OI ecosystem. Companies should proactively identify and engage with these influencers to amplify their

message, build credibility, and foster collaborations. This could involve retweeting their content, responding to their tweets, or inviting them to participate in online discussions or events.

- **Craft Compelling Content:** The analysis of tweet characteristics and engagement metrics emphasizes the importance of content quality. Companies should create informative, engaging, and visually appealing content that resonates with their target audience. The insights from the predictive model can be used to optimize the timing and content of tweets to maximize their potential reach, ensuring that messages are delivered when the target audience is most active on the platform.
- **Foster a Culture of Collaboration:** The overall positive sentiment towards OI on Twitter suggests a receptive audience for collaborative initiatives. Companies should leverage this opportunity to foster a culture of collaboration online, inviting feedback, engaging in open discussions, and co-creating solutions with their online community.
- **Measure and Monitor Performance:** Companies should track relevant metrics, such as view count, engagement rates, and sentiment, to assess the effectiveness of their communication strategies and identify areas for improvement. The insights from the various analyses presented in this dissertation can provide valuable guidance for measuring and monitoring the performance of OI initiatives on Twitter, allowing companies to refine their strategies and maximize their return on investment.

By integrating these practical recommendations into their social media strategies, companies can leverage Twitter as a powerful tool for driving OI and achieving their innovation goals.

9.5 Limitations and Future Research Directions

While this dissertation has provided a valuable contribution to the field of OI, it is essential to acknowledge its limitations and suggest avenues for future research.

The analysis is based on a dataset of tweets collected during a specific period. Future research could adopt a longitudinal approach, tracking the evolution of OI communication on Twitter over an extended period, to understand how conversations and trends change over time. Furthermore, the study focuses exclusively on Twitter. Future research could explore OI communication on other

social media platforms, such as LinkedIn or Facebook, to compare and contrast platform-specific dynamics and identify cross-platform trends.

The network analysis was limited by computational constraints, particularly in calculating betweenness centrality for the full networks. Future research could leverage more powerful computational resources or explore alternative network analysis techniques to overcome these limitations.

This research relied on a pre-defined list of OI keywords and the use of specific hashtags to identify relevant tweets. Future research could explore the use of alternative methods, such as natural language processing or machine learning, to identify and analyse a broader range of terms and phrases related to OI, potentially uncovering hidden conversations and capturing a more complete picture of the OI discourse.

Finally, this study primarily employed quantitative methods. Future research could integrate qualitative methods, such as interviews or focus groups, to gain a deeper understanding of the motivations, challenges, and perspectives of individuals and organizations engaged in OI on Twitter. Combining quantitative and qualitative approaches would provide a richer and more nuanced understanding of the complex interplay between social media and OI.

9.6 Conclusion

This dissertation has explored the dynamic and evolving landscape of OI communication on Twitter. The findings underscore the importance of social media as a platform for connecting diverse stakeholders, sharing knowledge, and driving innovation. By understanding the nuances of online communication, companies and organizations can effectively leverage Twitter to promote their OI initiatives, build relationships with key audiences, and contribute to a more collaborative and innovative future.

This research has provided valuable insights for practitioners, offering a roadmap for navigating the complexities of OI communication in the digital age. By embracing a data-driven and audience-centric approach, companies can transform Twitter into a powerful tool for fostering innovation, collaboration, and growth. Moreover, this study's focus on language, sentiment, and social dynamics highlights the importance of the human element in OI, recognizing that communication is not merely a means of transmitting information, but a process of building relationships, shaping perceptions, and creating shared value.

As OI continues to gain momentum and social media platforms evolve, future research should build upon this foundation, exploring new avenues for understanding the interplay between online communication and real-world OI practices. By embracing a multidisciplinary approach, integrating quantitative and qualitative methods, and leveraging the power of deep learning, future research can further enhance our understanding of this complex and dynamic field and contribute to a more open, collaborative, and innovative future.

References

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, *13*(1), 30. https://doi.org/10.1007/s13278-023-01028-5

Antcliff, V., Saundry, R., & Stuart, M. (2007). Networks and social capital in the UK television industry: The weakness of weak ties. *Human Relations*, *60*(2), 371–393. https://doi.org/10.1177/0018726707075880

Barbosa, A. P. P. L., Salerno, M. S., Brasil, V. C., & Nascimento, P. T. de S. (2020). Coordination Approaches to Foster Open Innovation R&D Projects Performance. *Journal of Engineering and Technology Management*, *58*, 101603. https://doi.org/10.1016/j.jengtecman.2020.101603

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Başarslan, M. S., & Kayaalp, F. (2021). Sentiment Analysis on Social Media Reviews Datasets with Deep Learning Approach. *Sakarya University Journal of Computer and Information Sciences*, *4*(1), 35–49. https://doi.org/10.35377/saucis.04.01.833026

Binti Mohd Haizar, N. F., Kee, D. M. H., Chong, L. M., & Chong, J. H. (2020). The Impact of Innovation Strategy on Organizational Success: A Study of Samsung. *Asia Pacific Journal of Management and Education*, *3*(2), 93–104. https://doi.org/10.32535/apjme.v3i2.849

Blei, D. M., Ng, A. Y., & Jordan, M. T. (2002). Latent dirichlet allocation. *Advances in Neural Information Processing Systems*, *3*, 993–1022. https://doi.org/10.5555/944919.944937

Bogers, M. (2011). The open innovation paradox: Knowledge sharing and protection in R&D collaborations. *European Journal of Innovation Management*, *14*(1), 93–117. https://doi.org/10.1108/14601061111104715

Bogers, M., Zobel, A.-K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., Frederiksen, L., Gawer, A., Gruber, M., Haefliger, S., Hagedoorn, J., Hilgers,

D., Laursen, K., Magnusson, M. G., Majchrzak, A., McCarthy, I. P., Moeslein, K. M., Nambisan, S., Piller, F. T., … Ter Wal, A. L. J. (2017). The open innovation research landscape: established perspectives and emerging themes across different levels of analysis. *Industry and Innovation*, *24*(1), 8–40. https://doi.org/10.1080/13662716.2016.1240068

Bourdieu, P. (2011). The forms of capital.(1986). *Cultural Theory: An Anthology*, *1*(81–93), 949.

Brockman, P., Khurana, I. K., & Zhong, R. (Irene). (2018). Societal trust and open innovation. *Research Policy*, *47*(10), 2048–2065. https://doi.org/10.1016/j.respol.2018.07.010

Bugshan, H. (2015). Co-innovation: the role of online communities. *Journal of Strategic Marketing*, *23*(2), 175–186. https://doi.org/10.1080/0965254X.2014.920905

Burcharth, A. L. de A., Knudsen, M. P., & Søndergaard, H. A. (2014). Neither invented nor shared here: The impact and management of attitudes for the adoption of open innovation practices. *Technovation*, *34*(3), 149–161. https://doi.org/10.1016/j.technovation.2013.11.007

Burger-Helmchen, T., & Siegel, E. J. (2020). *From Open Design to Open Strategy: Managing the Evolution of Crowdsourcing-Based Business Models* (pp. 73–107). https://doi.org/10.1142/9789811219238_0004

Chanda, A. K. (2021). Efficacy of BERT embeddings on predicting disaster from Twitter data. In *Archive â•Ž21* (Vol. 1, Issue 1). Association for Computing Machinery. http://arxiv.org/abs/2108.10698

Chesbrough, H. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Chesbrough, H., & Appleyard, M. M. (2007). Open Innovation and Strategy. *California Management Review*, *50*(1), 57–76. https://doi.org/10.2307/41166416

Clark, E., & Araki, K. (2011). Text Normalization in Social Media: Progress, Problems and Applications for a Pre-Processing System of Casual English. *Procedia - Social and Behavioral Sciences*, *27*, 2–11. https://doi.org/10.1016/j.sbspro.2011.10.577

Culotta, A., & Cutler, J. (2016). Mining Brand Perceptions from Twitter Social Networks. *Marketing Science*, *35*(3), 343–362. https://doi.org/10.1287/mksc.2015.0968

Dahlander, L., Beretta, M., Thomas, A., Kazemi, S., Fenger, M. H. J., & Frederiksen, L. (2023). Weeding out or picking winners in open innovation? Factors driving multi-stage crowd selection on LEGO ideas. *Research Policy*, *52*(10),

104875. https://doi.org/10.1016/j.respol.2023.104875

Dahlander, L., & Gann, D. M. (2010). How open is innovation? *Research Policy*, *39*(6), 699–709. https://doi.org/10.1016/j.respol.2010.01.013

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2), 147. https://doi.org/10.2307/2095101

Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *The Journal of Strategic Information Systems*, *24*(2), 113–127. https://doi.org/10.1016/j.jsis.2015.04.003

Edwards, T., Jones, C. B., Perkins, S. E., & Corcoran, P. (2021). Passive citizen science: The role of social media in wildlife observations. *PLOS ONE*, *16*(8), e0255416. https://doi.org/10.1371/journal.pone.0255416

Fairclough, N. (1992). *Discourse and Social Change*. Polity Press.

Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, *68*(6), 1261–1270. https://doi.org/10.1016/j.jbusres.2014.11.006

Frishammar, J., Ericsson, K., & Patel, P. C. (2015). The dark side of knowledge transfer: Exploring knowledge leakage in joint R&D projects. *Technovation*, *41*, 75–88. https://doi.org/10.1016/j.technovation.2015.01.001

Garg, N., & Sharma, K. (2022). Text pre-processing of multilingual for sentiment analysis based on social network data. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(1), 776. https://doi.org/10.11591/ijece.v12i1.pp776-784

Garriga, H., von Krogh, G., & Spaeth, S. (2013). How constraints and knowledge impact open innovation. *Strategic Management Journal*, *34*(9), 1134–1144. https://doi.org/10.1002/smj.2049

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239–1249.

https://doi.org/10.1016/j.respol.2014.03.006

Gkypali, A., Arvanitis, S., & Tsekouras, K. (2018). Absorptive capacity, exporting activities, innovation openness and innovation performance: A SEM approach towards a unifying framework. *Technological Forecasting and Social Change*, *132*, 143–155. https://doi.org/10.1016/j.techfore.2018.01.025

Gomez Rodriguez, M., Gummadi, K., & Schoelkopf, B. (2014). Quantifying Information Overload in Social Media and Its Impact on Social Contagions. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 170–179. https://doi.org/10.1609/icwsm.v8i1.14549

Gouws, R. H., & Tarp, S. (2016). Information overload and data overload in lexicography. *International Journal of Lexicography*, ecw030. https://doi.org/10.1093/ijl/ecw030

Grönlund, J., Sjödin, D. R., & Frishammar, J. (2010). Open Innovation and the Stage-Gate Process: A Revised Model for New Product Development. *California Management Review*, *52*(3), 106–131. https://doi.org/10.1525/cmr.2010.52.3.106

Haour, G., & Jolly, D. (2014). China: the next innovation hot spot for the world. *Journal of Business Strategy*, *35*(1), 2–8. https://doi.org/10.1108/JBS-05-2013-0037

Heger, T., & Boman, M. (2015). Networked foresight—The case of EIT ICT Labs. *Technological Forecasting and Social Change*, *101*, 147–164. https://doi.org/10.1016/j.techfore.2014.02.002

Helfat, C. E. (2006). Open Innovation: The New Imperative for Creating and Profiting from Technology. *Academy of Management Perspectives*, *20*(2), 86–88. https://doi.org/10.5465/AMP.2006.20591014

Henkel, J. (2006). Selective revealing in open innovation processes: The case of embedded Linux. *Research Policy*, *35*(7), 953–969. https://doi.org/10.1016/j.respol.2006.04.010

Hitchen, E. L., Nylund, P. A., Ferràs, X., & Mussons, S. (2017). Social media: open innovation in SMEs finds new support. *Journal of Business Strategy*, *38*(3), 21–29. https://doi.org/10.1108/JBS-02-2016-0015

Hromic, H., & Hayes, C. (2019). Characterising and evaluating dynamic online communities from live microblogging user interactions. *Social Network Analysis and Mining*, *9*(1), 30. https://doi.org/10.1007/s13278-019-0576-8

Huizingh, E. K. R. E. (2011). Open innovation: State of the art and future perspectives. *Technovation*, *31*(1), 2–9. https://doi.org/10.1016/j.technovation.2010.10.002

Huston, L., & Sakkab, N. (2006). Connect and develop: Inside Procter & Gamble's new model for innovation. *Harvard Business Review*, *84*(3), 58–66. http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=19707485&site=bsi-live

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Hutton, S., Demir, R., & Eldridge, S. (2021). How does open innovation contribute to the firm's dynamic capabilities? *Technovation*, *106*, 102288. https://doi.org/10.1016/j.technovation.2021.102288

Jiao, H., Yang, J., & Cui, Y. (2022). Institutional pressure and open innovation: the moderating effect of digital knowledge and experience-based knowledge. *Journal of Knowledge Management*, *26*(10), 2499–2527. https://doi.org/10.1108/JKM-01-2021-0046

Jussila, J. J., Kärkkäinen, H., & Leino, M. (2012). Social media's opportunities in business-to-business customer interaction in innovation process. *International Journal of Technology Marketing*, *7*(2), 191. https://doi.org/10.1504/IJTMKT.2012.046905

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, *54*(2), 105–113. https://doi.org/10.1016/j.bushor.2010.09.004

Kargaran, S., Shahri, M. H., Ghorbani, Z., Saberi, A., Jamali, S. M., & Aleebrahim, N. (2024). Patterns of publications in social media-based co-creation: a bibliometric analysis. *VINE Journal of Information and Knowledge Management Systems*, *54*(3), 578–596. https://doi.org/10.1108/VJIKMS-09-2021-0222

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. https://doi.org/10.1016/j.bushor.2011.01.005

Kim, H., Jang, S. M., Kim, S.-H., & Wan, A. (2018). Evaluating Sampling Methods for Content Analysis of Twitter Data. *Social Media + Society*, *4*(2), 205630511877283. https://doi.org/10.1177/2056305118772836

Kohler, T. (2015). Crowdsourcing-Based Business Models: How to Create and

Capture Value. *California Management Review*, *57*(4), 63–84. https://doi.org/10.1525/cmr.2015.57.4.63

Konstantinov, A., Moshkin, V., & Yarushkina, N. (2021). *Approach to the Use of Language Models BERT and Word2vec in Sentiment Analysis of Social Network Texts* (pp. 462–473). https://doi.org/10.1007/978-3-030-65283-8_38

Liberatore, A., Bowkett, E., MacLeod, C. J., Spurr, E., & Longnecker, N. (2018). Social Media as a Platform for a Citizen Science Community of Practice. *Citizen Science: Theory and Practice*, *3*(1), 3. https://doi.org/10.5334/cstp.108

Lichtenthaler, U. (2010). Technology exploitation in the context of open innovation: Finding the right 'job'' for your technology.' *Technovation*, *30*(7–8), 429–435. https://doi.org/10.1016/j.technovation.2010.04.001

Linton, J. D., & Walsh, S. T. (2008). A theory of innovation for process-based innovations such as nanotechnology. *Technological Forecasting and Social Change*, *75*(5), 583–594. https://doi.org/10.1016/j.techfore.2007.02.005

Malik, K., Georghiou, L., & Grieve, B. (2011). Developing New Technology Platforms for New Business Models: Syngenta's Partnership with the University of Manchester. *Research-Technology Management*, *54*(1), 24–31. https://doi.org/10.5437/08953608X540124

Meneghello, J., Thompson, N., Lee, K., Wong, K. W., & Abu-Salih, B. (2020). Unlocking Social Media and User Generated Content as a Data Source for Knowledge Management. *International Journal of Knowledge Management*, *16*(1), 101–122. https://doi.org/10.4018/IJKM.2020010105

Mention, A.-L. (2011). Co-operation and co-opetition as open innovation practices in the service sector: Which influence on innovation novelty? *Technovation*, *31*(1), 44–53. https://doi.org/10.1016/j.technovation.2010.08.002

Mention, A.-L., Barlatier, P.-J., & Josserand, E. (2019). Using social media to leverage and develop dynamic capabilities for innovation. *Technological Forecasting and Social Change*, *144*, 242–250. https://doi.org/10.1016/j.techfore.2019.03.003

Mount, M., & Martinez, M. G. (2014). Social Media: A Tool for Open Innovation. *California Management Review*, *56*(4), 124–143. https://doi.org/10.1525/cmr.2014.56.4.124

Muhammed T, S., & Mathew, S. K. (2022). The disaster of misinformation: a review of research in social media. *International Journal of Data Science and Analytics*, *13*(4), 271–285. https://doi.org/10.1007/s41060-022-00311-6

Muller, E., & Peres, R. (2019). The effect of social networks structure on innovation performance: A review and directions for research. *International Journal of*

Research in Marketing, *36*(1), 3–19. https://doi.org/10.1016/j.ijresmar.2018.05.003

Ogink, T., & Dong, J. Q. (2019). Stimulating innovation by user feedback on social media: The case of an online user innovation community. *Technological Forecasting and Social Change*, *144*, 295–302. https://doi.org/10.1016/j.techfore.2017.07.029

Oliveira-Dias, C., Morais, S., & Costa, A. R. (2022). Sex differences in the association between social capital and healthcare use—Results from the Survey of Health, Ageing and Retirement in Europe (SHARE). *Health & Social Care in the Community*, *30*(6). https://doi.org/10.1111/hsc.13889

Oliveira, S. S., Barros, B., Pereira, J. L., Santos, P. T., & Pereira, R. (2021). Social Media Use by Citizen Science Projects: Characterization and Recommendations. *Frontiers in Environmental Science*, *9*. https://doi.org/10.3389/fenvs.2021.715319

Park, S. J., Lim, Y. S., & Park, H. W. (2015). Comparing Twitter and YouTube networks in information diffusion: The case of the "Occupy Wall Street" movement. *Technological Forecasting and Social Change*, *95*, 208–217. https://doi.org/10.1016/j.techfore.2015.02.003

Pohl, H., & Yarime, M. (2012). Integrating innovation system and management concepts: The development of electric and hybrid electric vehicles in Japan. *Technological Forecasting and Social Change*, *79*(8), 1431–1446. https://doi.org/10.1016/j.techfore.2012.04.012

Radziwon, A., & Bogers, M. (2019). Open innovation in SMEs: Exploring inter-organizational relationships in an ecosystem. *Technological Forecasting and Social Change*, *146*, 573–587. https://doi.org/10.1016/j.techfore.2018.04.021

Rimal, R. N., & Adkins, A. D. (2003). Using computers to narrowcast health messages: The role of audience segmentation, targeting, and tailoring in health promotion. In *The Routledge Handbook of Health Communication* (pp. 511–528). Routledge.

Rogers, E. M. (1995). Diffusion of Innovations: modifications of a model for telecommunications. *Die Diffusion von Innovationen in Der Telekommunikation*, 25–38.

Roma, P., & Aloini, D. (2019). How does brand-related user-generated content differ across social media? Evidence reloaded. *Journal of Business Research*, *96*, 322–339. https://doi.org/10.1016/j.jbusres.2018.11.055

Saffer, A. J., Sommerfeldt, E. J., & Taylor, M. (2013). The effects of organizational Twitter interactivity on organization–public relationships. *Public Relations*

Review, *39*(3), 213–215. https://doi.org/10.1016/j.pubrev.2013.02.005

Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Exploring the boundaries of open innovation: Evidence from social media mining. *Technovation*, *119*, 102447. https://doi.org/10.1016/j.technovation.2021.102447

Schuhmacher, A., Gassmann, O., Bieniok, D., Hinder, M., & Hartl, D. (2022). Open innovation: A paradigm shift in pharma R&D? *Drug Discovery Today*, *27*(9), 2395–2405. https://doi.org/10.1016/j.drudis.2022.05.018

Smith, B. G. (2010). Socially distributing public relations: Twitter, Haiti, and interactivity in social media. *Public Relations Review*, *36*(4), 329–335. https://doi.org/10.1016/j.pubrev.2010.08.005

Thanasopon, B., Papadopoulos, T., & Vidgen, R. (2016). The role of openness in the fuzzy front-end of service innovation. *Technovation*, *47*, 32–46. https://doi.org/10.1016/j.technovation.2015.11.007

Töllinen, A., & Karjaluoto, H. (2011). Marketing communication metrics for social media. *International Journal of Technology Marketing*, *6*(4), 316. https://doi.org/10.1504/IJTMKT.2011.045911

Toroslu, A., Schemmann, B., Chappin, M. M. H., Castaldi, C., & Herrmann, A. M. (2023). Value capture in open innovation: A literature review and a research agenda. *Industrial Marketing Management*, *114*, 297–312. https://doi.org/10.1016/j.indmarman.2023.08.012

Ulloa, D., Saleiro, P., Rossetti, R. J. F., & Silva, E. R. (2016). Mining social media for open innovation in transportation systems. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 169–174. https://doi.org/10.1109/ITSC.2016.7795549

Valentini, C., Romenti, S., Murtarelli, G., & Pizzetti, M. (2018). Digital visual engagement: influencing purchase intentions on Instagram. *Journal of Communication Management*, *22*(4), 362–381. https://doi.org/10.1108/JCOM-01-2018-0005

van de Vrande, V., de Jong, J. P. J., Vanhaverbeke, W., de Rochemont, M., Vrande, V. Van De, & Jong, J. P. J. De. (2009). Open innovation in SMEs: Trends, motives and management challenges. *Technovation*, *29*(6–7), 423–437. https://doi.org/10.1016/j.technovation.2008.10.001

Van Lancker, J., Mondelaers, K., Wauters, E., & Van Huylenbroeck, G. (2016). The Organizational Innovation System: A systemic framework for radical innovation at the organizational level. *Technovation*, *52–53*, 40–50. https://doi.org/10.1016/j.technovation.2015.11.008

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a New Dominant Logic for

Marketing. *Journal of Marketing*, *68*(1), 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036

Walsh, S. T., & Linton, J. D. (2011). The Strategy-Technology Firm Fit Audit: A guide to opportunity assessment and selection. *Technological Forecasting and Social Change*, *78*(2), 199–216. https://doi.org/10.1016/j.techfore.2010.06.023

Wang, C.-H., Chang, C.-H., & Shen, G. C. (2015). The effect of inbound open innovation on firm performance: Evidence from high-tech industry. *Technological Forecasting and Social Change*, *99*, 222–230. https://doi.org/10.1016/j.techfore.2015.07.006

Wenger, E. (1998). The Systems Thinker – Communities of Practice: Learning as a Social System - The Systems Thinker. *Systems Thinker*, *9*(5), 2–3. https://thesystemsthinker.com/communities-of-practice-learning-as-a-social-system/

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180. https://doi.org/10.1002/smj.4250050207

West, J., & Bogers, M. (2014). Leveraging external sources of innovation: A review of research on open innovation. *Journal of Product Innovation Management*, *31*(4), 814–831. https://doi.org/10.1111/jpim.12125

Wilden, R., Akaka, M. A., Karpen, I. O., & Hohberger, J. (2017). The Evolution and Prospects of Service-Dominant Logic. *Journal of Service Research*, *20*(4), 345–361. https://doi.org/10.1177/1094670517715121

Xu, W. (Wayne), & Saxton, G. D. (2019). Does Stakeholder Engagement Pay Off on Social Media? A Social Capital Perspective. *Nonprofit and Voluntary Sector Quarterly*, *48*(1), 28–49. https://doi.org/10.1177/0899764018791267

Yang, J., Chesbrough, H., & Hurmelinna-Laukkanen, P. (2021). How to Appropriate Value from General-Purpose Technology by Applying Open Innovation. *California Management Review*, 000812562110417. https://doi.org/10.1177/00081256211041787

Zachlod, C., Samuel, O., Ochsner, A., & Werthmüller, S. (2022). Analytics of social media data – State of characteristics and application. *Journal of Business Research*, *144*, 1064–1076. https://doi.org/10.1016/j.jbusres.2022.02.016