UNIVERSITY OF TURIN

Ph.D. in Modeling and Data Science

XXXV CYCLE

Final dissertation

The role of images in online knowledge consumption: studying readers' behavior on Wikipedia

Supervisor: Prof. Rossano SCHIFANELLA

Candidate: Daniele RAMA

ACADEMIC YEAR 2022/2023

Abstract

Over the past decades, the quantity of available multimedia content on the Web, especially images, has exponentially increased, thanks to their power of easily capturing our attention and facilitating communication. Experimental psychologists has demonstrated their fundamental cognitive functions in learning and educational contexts, while computational scientists have underlined their significance in effective online information search and navigation. Investigating the impact of images on online knowledge consumption is crucial for understanding how images shape our Web experience and for integrating visual content into online knowledge platforms.

This dissertation explores the role of images in engaging with and navigating knowledge on the English Wikipedia, the largest online encyclopedia. Despite its prominence as a platform for open knowledge, little is known about how images influence readers' interactions and navigation through its concepts.

The dissertation comprises two main contributions. In the first part, we use passively collected digital traces to uncover patterns of engagement with images on the English Wikipedia. Our findings demonstrate that images significantly drive interactions, and we provide a comprehensive overview of the visual and contextual factors associated with image engagement. Furthermore, our research suggests evidence for images' function of fulfilling the need for contextual information while reading articles.

In the second part, we investigate the influence of images on knowledge navigation. Analyzing previously collected logs from Wikipedia, we observe that illustrated article sections receive more interactions compared to non-illustrated ones, providing evidence for the attentive function of visual content. Then, through longitudinal analysis, we assess the beneficial impact of adding illustrations to previously non-illustrated sections. Finally, we design a human-navigation experimental study, revealing that the presence of images makes people more efficient at finding the information they need.

This dissertation aims to advance our comprehension of the role of multimedia content in online knowledge consumption. While focusing on the context of Wikipedia, our findings have broader implications for the importance of visual content in consuming and navigating knowledge on the broader Web.

Contents

1 Introduction **1**
1.1 Motivations . . . 1
1.2 Contributions and thesis outline . . . 2

2 Background and related work **5**
2.1 Images in learning and online platforms . . . 5
2.1.1 Cognitive theories of multimedia learning . . . 5
2.1.2 Image interestingness . . . 6
2.2 Information-seeking behaviors on the Web . . . 7
2.2.1 Web search and navigation through visual cues . . . 7
2.2.2 Navigation on Wikipedia . . . 9
2.3 Research on Wikipedia's visual content . . . 11

3 Visual content on Wikipedia **15**
3.1 Wikipedia . . . 15
3.2 Images on Wikipedia . . . 17

4 Quantifying interactions with images on Wikipedia **21**
4.1 Scope . . . 21
4.2 Data collection and methods . . . 22
4.2.1 Collecting article and image counts data . . . 23
4.2.2 Collecting article and image traffic data . . . 23
4.2.3 Mining image content and context . . . 25
Contextual factors . . . 25
Visual factors . . . 28
4.2.4 Engagement metrics . . . 29
4.3 RQ1: To what extent are readers interacting with images in Wikipedia? 30
4.3.1 Overall engagement with images: the global click-through rate 31
4.3.2 Average engagement with individual images: image-specific click-through rate . . . 31
4.4 RQ2: What drives reader's engagement with images when reading Wikipedia articles? . . . 32
4.4.1 Exploratory analysis . . . 32
4.4.2 Regression analysis . . . 33
4.4.3 Identifying prototypical image groups . . . 35

4.4.4 Are faces engaging on Wikipedia? 37
4.5 RQ3: Do images support reader's need for additional information when navigating Wikipedia? 39
4.6 Conclusions 40

5 The role of images in navigating Wikipedia 43
5.1 Scope 43
5.2 RQ1: Do readers engage more with illustrated knowledge? 44
5.2.1 Materials and Methods 45
Section data 45
Pageloads 45
Section click-through rate 46
5.2.2 The effect of images 46
5.2.3 The effect of topics 48
5.2.4 The effect of the section offset 49
5.2.5 Does engagement with sections change when adding images? . 50
5.3 RQ2: Do images support navigation on Wikipedia? 52
5.3.1 Materials and Methods 53
Wikispeedia 53
Data 54
Metrics 54
5.3.2 Results 55
Task duration 55
Path length 56
Dwell time 56
Are images relevant to their context? 56
5.4 Conclusions 57

6 Conclusions 59
6.1 Summary of findings 59
6.2 Limitations and opportunities 61
6.3 Implications and broader perspectives 62

Bibliography 65

List of Figures

3.1 Example of an article in the English Wikipedia. 16
3.2 Examples of modalities through which readers can interact with images on a Wikipedia article: directly clicking on them (the images are then visualized in a different page, called the *Media Viewer* or indirectly through article *previews*, a tooltip showing a short summary—eventually illustrated—of linked articles (only via desktop browsers). . 17
3.3 Examples of the three types of locations (*infobox*, *inline*, and *gallery*) where images are placed within a Wikipedia article. 19
3.4 Wikipedia size as a function of (a) the percentage of illustrated articles and (b) the average number of images per article. 20

4.1 Cumulative distributions of (A) number of images per article and (B) the number of articles per image. 23
4.2 Cumulative distributions of number of sessions by (A) imageviews, (B) pageviews, and (C) previews partitioned by desktop (in blue) and mobile device (in orange). Page previews are available only on desktop devices. 24
4.3 (A) Fraction of images by topic (in blue) and fraction of images with faces (in orange). (B) Image-specific CTR by article topic. 26
4.4 Feature distributions. Spearman's rank correlation coefficient ρ between the numerical features and the iCTR at the top of each panel ($p < 0.001$ for each feature). 27
4.5 Distributions of (A) values of iCTR by page popularity partitioned by device and (B) number of images per page popularity. Spearman's rank correlation coefficient ρ_{iCTR} between iCTR and pageviews in the inset ($p < 0.001$). The axes are in log scale. 31
4.6 Examples of *high* and *low* image-specific CTR images by page popularity (left) and image quality (right). We ranked images by iCTR, popularity and quality, and picked examples from the top-100 ("high") and bottom-100 ("low") for each dimension. 32
4.7 Association of the features with the image iCTR expressed as coefficients of the logistic regressions. (A) Coefficients of the model trained with topics of the article as predictors. (B) Coefficients of the model trained with the other variables of the image. Error bars represent 95% confidence intervals. 33

4.8 Visual representation of the clustering. The radar plots show for a group centroid the intensity of each feature on a three classes scale. We summarize in green the topics that cover at least 85% of the images categories in a cluster. 37

4.9 Comparison of the iCTR for images with faces (orange) and without faces (blue) as function of the popularity (*pageviews*). Error bands represent bootstrapped 95% CIs. 38

4.10 Comparison of the conversion rate for preview tooltip with an image (purple) and without image (green) as function of the page popularity (*pageviews*). Error bands represent bootstrapped 95% CIs. 40

5.1 Effect of the section's covariates on the probability of having an image. The effects are estimated with logistic regression (area under the ROC curve: 0.78). 95% confidence intervals are too narrow to be appreciated. 48

5.2 Comparison of median section click-through rate between matched pairs of sections with (yellow) and without (green) images, stratified by topic. Significant differences (Mann-Whitney U test with $p < 0.01$) are marked with an asterisk (*). 49

5.3 Relation between the section click-through rate and the section offset, stratified by quartiles of article length (with bootstrapped 95% confidence intervals). 50

5.4 Difference-in-difference treatment effect between matched pairs of sections, stratified by month of image addition (with 95% confidence intervals). 53

5.5 Mean task duration (a) and effective path length (b), stratified by shortest path length (with 95% bootstrapped confidence intervals). Average time spent on articles (dwell time) (c), stratified by quartiles of article length. Error bands are 95% bootstrapped confidence intervals. 55

Chapter 1

Introduction

1.1 Motivations

In the digital age, we are witnessing an unprecedented surge in the volume of multimedia content available online. Platforms like Instagram and TikTok, which primarily revolve around visual content, have gained immense popularity, capturing the attention of millions of users worldwide. Few years later, it was the turn of Flicker and Tumblr to capitalize on the allure of visual imagery. This pervasive use of images in social media and other online platforms begs the question: *what motivates this extraordinary growth and why are images such powerful tools of communication?*

There is no single reason behind this success. Images have a unique ability to evoke emotions, often surpassing the limitations of words alone. They can capture moments and convey stories, all in a single frame. Images transcend language barriers, enabling communication and understanding across diverse cultures and linguistic backgrounds. This linguistic universality of visuals makes them essential in today's interconnected and multicultural world. Images simplify complex concepts. Complex ideas often require intricate explanations, but a well-designed image can distill these complexities into a concise and easily comprehensible form. By presenting information visually, images facilitate efficient knowledge transfer and promote better understanding among diverse audiences. Furthermore, images have an innate ability to capture and retain our attention, making them powerful tools for engagement and learning.

Extensive research has been conducted to explore the role of images, particularly in learning and educational contexts. These studies have shed light on the cognitive processes underlying image perception and comprehension, opening avenues for further exploration and application in educational settings. With the advent of the Web, various computational approaches have also been employed to analyze the impact of images in different domains. Several works have studied why images are interesting to online readers, and how visual cues enhance knowledge search and navigation in online information networks.

However, despite the growing recognition of the significance of images, studying their role in online contexts presents several challenges. Traditional approaches, such as surveys or in-lab studies, often rely on small-scale samples and narrow audiences, thus lacking generalizability. On the other hand, novel approaches involving the use of digital traces are still hindered by the scarcity of large-scale datasets on user interactions including visual content. Overcoming these limitations is crucial for a comprehensive understanding of how images shape knowledge dissemination and online interactions in the digital era.

1.2 Contributions and thesis outline

This dissertation aims to advance our comprehension of the role of images in online information consumption, specifically in the context of encyclopedic knowledge. The research focuses on Wikipedia, the largest online free source of knowledge, which has democratized access to information providing vast amounts of textual and visual content to billions of readers every month. Investigating how visual content impacts knowledge consumption on Wikipedia is crucial for understanding the role of images in shaping the way we access and interact with knowledge.

By leveraging the rich network of concepts and resources available on Wikipedia, this work employs large-scale observational and experimental studies to offer a novel, comprehensive perspective on the impact of visual content on knowledge consumption. The findings of this research hold implications that extend beyond the scope of Wikipedia, benefiting various audiences.

This work is based on two major research studies conducted during the Ph.D. program. Namely, Chapter 4 is adapted from a published research contribution (Rama et al., 2022), and Chapter 5 is adapted from another work recently submitted to a top-tier international conference. The structure of the dissertation, methodologies employed, and main results are summarized in the following.

Chapter 2 provides an overview of the relevant literature that contextualizes the research. First, it covers fundamental works in cognitive psychology related to multimedia learning and computational studies on image interestingness. Then, it explores the literature on how images impact Web search and navigation, and presents relevant studies on the information needs and navigational patterns of Wikipedia readers. Finally, it reviews the existing research on the visual aspects of Wikipedia.

Chapter 3 delves into the visual ecosystem of Wikipedia. It provides a detailed overview of Wikipedia's structure and explores the landscape of its visual content.

Chapter 4 presents a large-scale characterization of readers' engagement patterns with images on Wikipedia. Utilizing billions of requests collected from Wikipedia's logs, the study measures how readers read articles and interact with images. The results confirm that overall engagement with images surpasses interactions with other types of article content. Computer vision algorithms are employed to extract visual characteristics from the images. Additionally, methods from causal inference are employed to comprehensively examine factors associated with image engagement. The findings reveal that interactions occur more frequently with detail-rich and unfamiliar illustrations. Furthermore, evidence suggests that images support readers' need for additional contextual information when navigating Wikipedia.

Chapter 5 investigates the role of images in the navigation and consumption of knowledge on Wikipedia. Using previously collected logs from Wikipedia, the study reveals that illustrated article sections receive more interactions compared to non-illustrated sections. Furthermore, the impact of adding illustrations to previously non-illustrated sections is evaluated, indicating that sections are more likely to receive interactions when image additions are recent. Additionally, a crowdsourced experimental study is designed, where users are asked to navigate a network of Wikipedia articles by following hyperlinks. The findings demonstrate that information wayfinding is more efficient when articles are illustrated, particularly when images are relevant to the surrounding textual context.

Chapter 6 summarizes the main findings of the research and discusses the limitations arising from the methodology employed. Potential areas for improvement in future work are suggested. Moreover, the practical and theoretical implications of the findings are explored, providing broader perspectives on how the results can be utilized within the context of Wikipedia and the broader web.

Chapter 2

Background and related work

2.1 Images in learning and online platforms

In our daily lives, we are used to read and learn from text that is often illustrated to ehnance our ability to learn and comprehend the topic. This section offers a comprehensive overview of the existing literature on the importance of images, exploring their role from both psychological and computational perspectives.

2.1.1 Cognitive theories of multimedia learning

The recognition of the importance of illustrations dates back to the 19th century, with the pioneering work of scholars such as John Amos Comenius (Wojewoda, 2018). Comenius, in his book *The Great Didactic* (Comenius, 1907), emphasized the educational role of images, particularly as an introduction to knowledge acquisition during early stages of life. It was only in the second half of the 19th century that researchers began hypothesizing that learning could be more effective when multiple representations (i.e., visual and verbal) were used to construct and coordinate information (Gyselinck et al., 1999; W. Howard Levie et al., 1982b; Mayer, 2020).

The most influential work in this field is Mayer's theory of *Multimedia Learning* (Mayer, 2020). Basing his theories on several experimental observations, Mayer formulated his *multimedia* principles, suggesting that learning materials incorporating illustrated texts (i.e., multimedia materials) can enhance learning and promote deeper understanding compared to non-illustrated materials. The theoretical foundations of his principles stem from Atkinson and Shiffrin's multi-store model (Atkinson et al., 1968), Baddeley's working memory model (Baddeley, 1992), and Paivio's dual-coding theory (Paivio, 1990). In particular, Paivio's dual-coding theory proposes that pictures have advantages over words in terms of coding and retrieval of stored memory, as pictures are easier to code and can be retrieved from symbolic mode (also referred to as the *picture superiority effect*). Consistent with the dual-coding theory, the theory of multimedia learning suggests that presenting information through both visual and verbal channels enhances memory retention (Mayer, 2020). Complementary, W. Howard Levie et al. (W. Howard Levie et al., 1982b) suggested that illustrations serve multiple functions: (i) the attentional function, attracting attention to textual material and then directing attention within it, (ii) the affective function, enhancing emotions and attitudes in the reader, (iii) the cognitive function, pictures can facilitate learning information in the text by improving comprehension and improving the retention of it, and (iv) the compensatory function, assisting poor readers.

Despite their positive effects, caution must be taken when adding images. The *signaling principle* (Gog, 2014; Alpizar et al., 2020; Schneider et al., 2018) posits that pictures can be used to guide the learner's attention to essential part of the text. Nonetheless, not all signaling attempts are helpful; decorative images or incorrectly positioned visuals can act as distractors, leading to additional cognitive processes that do not contribute to goal-oriented learning (Mautone et al., 2001).

Over the past fifty years, the validity of multimedia principles has been supported by numerous studies in education (J. Li et al., 2019; Almarabeh et al., 2015; Rudolph, 2017), cognitive psychology (Frick et al., 2023; C. I. Johnson et al., 2012; Ginns, 2005) , and linguistics (Alobaid, 2020). However, only recently advancements in information technologies and the internet have fostered research in online settings by computer scientists (Tempelman-Kluit, 2017; Khamparia et al., 2017; Guo et al., 2020). For instance, studies have shown that computer-based media offer advantages in terms of affordability and interactivity over traditional book-based media in instructional settings (Shih, 2007). Online courses and resources make knowledge more accessible to underserved populations (Clark, 2014). Videos facilitate the transfer of knowledge and the structural knowledge acquisition (Ibrahim et al., 2012; Yoon et al., 2021). The adoption of virtual reality tools as a pedagogical method in education has even challenged the conceptual definition of learning environment (Hamilton et al., 2021; Mystakidis et al., 2022).

It is important to note that the majority of these works are often based on experimental analyses conducted on small samples or limited segments of the population, such as primary school students, and may focus on specific topics (Shih, 2007; Clark, 2014). As a result, the significance and generalizability of the conclusions may be limited.

2.1.2 Image interestingness

With the proliferation of online multimedia content, there has been a growing interest in the comprehension and prediction of user perceptions of visual content. The concept of *interestingness* typically refers to the ability to evoke curiosity, capture attention, and generate interest (Berlyne, 1960). This notion has been intensively investigated for years in psychology and vision research (Chen et al., 2001; Elazary et al., 2008). Although interestingness is inherently subjective and depends on personal preferences and experiences, a considerable consensus exists among users regarding which images are more captivating than others. This consensus has motivated researchers in the field of computer vision to delve deeper into this topic.

Researchers in computer vision have traditionally described interestingness in two ways. Visual interestingness is the extent to which an image can hold or catch the viewer's attention due to its intrinsic visual qualities (Constantin, Ştefan, et al., 2021). Studies have shown that images tend to be more interesting when they possess greater aesthetic appeal, visual complexity, or unfamiliar (Gygli et al., 2013). On the other hand, social interestingness, often called popularity, corresponds to the extent to which an image is liked by a large number of people in a community (Constantin, Ştefan, et al., 2021; Berson et al., 2019; Amengual et al., 2015). Social interestingness depends on the social dynamics of the platforms where images are shared, the pictures visual content (Khosla et al., 2014a; Bakhshi et al., 2014; Ding et al., 2019), and the text associated with them (Zhang et al., 2018). Most of these previous works focus on predicting image popularity in photo sharing platforms such as Flickr (Khosla et al., 2014b; Zhang et al., 2018), or Instagram (Bakhshi et al., 2014), specifically designed to increase social interestingness in images. However, the degree to which images are interesting to readers on online instructional and information-sharing platforms, like Wikipedia, has received limited attention so far.

While testing the principles of the multimedia learning theory and related research is beyond the scope of this dissertation, the present work grounds some ideas in this field to design experiments, observational studies, and analyses to study how visual content influences the reader's experience in the context of online encyclopedic knowledge. Moreover, in Chapter 4, to model the complex interplay between encyclopedic pictorial representations and user engagement, visual features explicitly inspired by the literature on image interestingness are extracted.

2.2 Information-seeking behaviors on the Web

Given the Internet diffusion, a significant portion of the information we consume comes from the Web. This section provides an overview of research describing our information-seeking behavior on the Web and how this is influenced by the presence of visual cues. In addition, it reviews the literature on how readers consume and navigate information on Wikipedia.

2.2.1 Web search and navigation through visual cues

The advent of the World Wide Web has opened the doors for the digital revolution, making vast amounts of information available anywhere in the world. With this abundance of resources, navigating and efficiently searching through all the digital information has become increasingly challenging. Studying users' navigational traces can enhance our understanding of the most effective strategies and enable us to make better-informed decisions. In the last decades, researchers have devoted significant efforts to studying how readers navigate the Web. They have concluded that there are primarily two strategies for exploring and discovering information spaces (Levene, 2011): *navigation*, where users explore interconnected pieces of information by following hyperlinks, and *search*, where users formulate a query with a specific target in mind.

Over the years, the study of information-seeking behaviors has received attention from several disciplines, from sociology to cognitive psychologists, and, thanks to the recent availability of digital trace data, also from computer scientists. In their search for information, people are driven by information needs. While this concept remains unclear to define, starting from the 1960's scholars have tried to observe and model the strategies we use to find piece of information (WILSON, 1981; Taylor, 1962). An interesting point of view coming from cognitive psychology is that humans are informavores and seek information in analogy with how animals search food (Machlup, 1983). This idea inspired the formulation of the theory of information foraging (Pirolli et al., 1999), which describes humans as behaving akin to predators in the information space. Like predators relying on scents to find the paths for food, humans search for *information scents* (Chi et al., 2003) to maximize the chances of reaching the piece of information they need. The strength of information scent is determined by textual and visual cues, shaping users' movements toward subsequent information patches. Recently, researchers have applied these theories to investigate how users navigate the Web. A common finding is that people tend to revisit similar contents and pages multiple times (Tauscher et al., 1997; Anderson et al., 2014). Many works describes navigation trajectories and predict next navigation steps via Markhov chains (Deshpande et al., 2004) and hybrid models (Awad et al., 2007). Some works have shown that the Web page layout can have an impact on navigation. For instance, link at the top of a page are more likely to attract attention (Craswell et al., 2008). In addition, models based on the information foraging theory have been developed to predict user behaviour during the information seeking process and also to predict the information scent for specific interfaces and web pages (Blackmon et al., 2002; Chi et al., 2003). However, most of these works focus on the presence of textual cues, while visual cues have largely been ignored. Among these few works, Oostendorp et al. (Oostendorp et al., 2012a) proposed a cognitive model of Web navigation that takes into account the semantic information from graphical elements present on a Web page to compute the information scent value of the hyperlinks. In follow-up work, Karanam et al. (Karanam et al., 2012) demonstrated that the model can predict the hyperlinks on the shortest path more frequently, and also with greater information scent, compared to earlier cognitive models of Web navigation that relied only on textual information.

Complementary to this line of research, recent works in the field of information retrieval have investigated web user behavior in Search Engine Result Pages (SERPs) and image search engines. For instance, Loumakis et al. (Loumakis et al., 2011a) discovered that the addition of images to text snippets on Search Engine Results Pages (SERPs) increase participants' confidence in finding answers, irrespective of image quality. Researchers have found that, in general, the most popular queries in image search engines are about people, celebrities, and entertainment (Tsikrika et al., 2014a; Jansen et al., 2004; Huang et al., 2009). By comparing image search behavior with text search behavior, several studies have found that image search sessions are heavier in interaction and exploration than the more focused textual sessions (Jansen et al., 2004), although, in a later study, Park et al. (Park et al., 2015) found that web image search behavior depends on the query type.

2.2.2 Navigation on Wikipedia

Twenty years after its birth, Wikipedia has become the largest free source of encyclopedic knowledge on the Web. Its content is produced by the remarkable effort of thousands of editors and is consumed by billions of readers worldwide every day. Given its central role, in recent years, researchers have extensively studied how readers consume and navigate information when reading Wikipedia.

Motivations, consumption patterns and content engagement in the Wikipedia readership In a seminal work, Singer, Lemmerich, et al. (Singer, Lemmerich, et al., 2017a) observed that users visit Wikipedia with different information needs, motivated by a variety of different factors such as current events, media coverage of a topic, personal curiosity, work or school assignments, or boredom. Following works have observed how these motivations lead to different use cases and reading activities, which could be classified into four main patterns: exploration, focus, trending, and passing (Lehmann et al., 2014a). The way in which users read and consume information in Wikipedia also varies across cultural and demographical dimensions and is often accompanied by biases. For instance, some works showed that readers exhibit specific consumption patterns based on their country, such as more frequent in-depth reading in the Global South and countries with low Human Development Index (Lemmerich et al., 2019a; TeBlunthuis et al., 2019). Waller (Waller, n.d.) observed difference among the various lifestyle segments in the use of Wikipedia among the Australian population, with high-income segments more likely to use it. In addition, gender differences exists among the Wikipedia readership (I. Johnson et al., 2021), with women being underrepresented among readers and visiting fewer pages per reding session than men. Finally, researchers investigated the effect of external factors, like an Academy Award nomination (Ratkiewicz et al., 2010) or the promulgation of COVID-19 mobility restrictions (Ribeiro et al., 2021), and internal design changes, like the introduction of the preview feature (Chelsy Xie et al., 2019).

Another line of research has characterized how readers engage with articles' content on Wikipedia. Recent work shows that readers tend to engage more with article references about recent events or people's experiences (Piccardi, Redi, et al., 2020a), medical content and external link in the infoboxes (Piccardi, Gerlach, and West, 2022). Regarding the topical preferences of the readers, research findings indicate that people tend to share a common interest and curiosity for entertainment, e.g. movies, music, sports independently of their language (Miz et al., 2020; Lehmann et al., 2014a; Waller, n.d.; Spoerri, 2007), but also that men and women exhibit specific topical preferences (I. Johnson et al., 2021).

Modeling and analysis of navigation patterns In previous years, several studies have reported on the navigation behaviors of the readers within Wikipedia, following different approaches. Server logs have been used by researchers to study the dynamics of content navigation with observational studies. Recently, Piccardi, Gerlach, Arora, et al. (Piccardi, Gerlach, Arora, et al., 2023b) leveraged a massive dataset of 6.5B pageloads to characterize reading behaviors, finding that most navigation paths are shallow, comprising a single pageloads, but that longer reading sessions largely varies with topics, device, and time of day. Similar works have shown that reader navigation exhibit daily and weekly (Reinoso, Gonzalez-Barahona, et al., 2009; Reinoso, Muñoz-Mansilla, et al., 2012; Piccardi, Gerlach, and West, 2023a) and intra-session (Halfaker, Keyes, et al., 2015) temporal regularities. Also, some readers tend to fall into "rabbit holes", i.e., long exploratory navigation sessions that often starts from the same topics (e.g., entertainment, sports, politics and history) (Piccardi, Gerlach, and West, 2022).

In alternative to raw server logs, which are usually not accessible to the public due to privacy reasons, the Wikimedia Foundation releases the public clickstream[1], which contains monthly page-to-page transitions for all the Wikipedia articles in multiple language editions. Recently, Arora et al. (Arora et al., 2022) proved that it approximates the real navigation patterns with a good level of accuracy. Researchers using this dataset showed that some topics tend to relay more traffic than others (Dimitrov et al., 2018a) and how readers' navigation paths tend to start general and become incrementally more semantically focused at every step (Rodi et al., 2017). Some works have found that the article network structure has significant impact on the navigation (Gildersleve et al., 2018), sometimes biasing readers' navigation towards specific topics (Menghini et al., 2019). Lamprecht et al. (Lamprecht et al., 2017) also showed that the article layout has an impact on the navigation, showing that readers tend to click more on links at the top of the page.

Finally, another approach to analyze user navigation relies on digital traces collected through Wikipedia navigation games, such as Wikispeedia (West, n.d.) and TheWikiGame[2]. These online games are designed to collect human-generated data

[1] https://meta.wikimedia.org/wiki/Research:Wikipedia_clickstream
[2] https://www.thewikigame.com/

in a gamified way. Specifically, players are asked to reach a target article starting from a random one following as few internal links as possible. In contrast to server logs or clickstream data, these paths carry an unambiguous definition of success, which allows researchers to study how users drift away from the best path and when they abandon their search (Koopmann et al., 2019). Several works have shown that users are able to employ efficient search strategies wherein they make progress towards the destination in the first part of the exploration by jumping towards high degree nodes, and then traverse the semantic space with smaller step sizes (West and Leskovec, 2012a; Helic, 2012). These targeted navigation traces have also been used to test models of human navigation, showing high predictive power (Trattner et al., 2012; Helic et al., 2013; Singer, Helic, et al., 2014; Koopmann et al., 2019)

This dissertation builds upon existing research on information-seeking behaviors on the Web and extends the current literature on readership patterns in Wikipedia, with a specific focus on images. Chapter 4 incorporates relevant insights from previous studies on Web image search, while Chapter 5 delves into the significance of images in facilitating reader navigation within the encyclopedia.

2.3 Research on Wikipedia's visual content

Despite their importance on the Wikipedia platform and for the entire Web community, imagery on Wikipedia has largely been ignored in previous research. Only in recent years, researchers have started to investigate the characteristics and the importance of the Wikipedia's visual knowledge space.

The editors community and image content A prominent line of works has focused on image-related editing behaviors and on image characteristics and content. A seminal work by Fernanda B. Viegas studied for the first time the characteristics and the motivations of the community of visual collaborators. Basing their analysis on a survey conducted with image contributors, they found that this community is one of the fastest growing in the site showing a high degree of motivation. Their findings also reveal key differences in collaborating around images as opposed to text, suggesting that, even though image editing is a more isolated activity, the sense of community is an important motivation for image contributors. More recently, few works have focused on image characteristics and content. Measuring the visual diversity across 25 Wikipedia language editions, He et al. found that image diversity often exceeds that found in text, and that a large fraction of images are unique to different language editions. Beytía et al. investigated the visual gender bias in biographical content. They observed that written and visual biases are dissimilar, with men being more visually represented than women across languages, but women having images of higher quality. In addition, a couple of works (Navarrete et al., 2020; Bertacchini et al., 2022) explored the dynamics of production, fruition and reuse of images of archaeological heritage and paintings present on Wikimedia Commons, finding that most of these works are extensively used to illustrate also non-art-related topics. The extensive range of visual content necessitates the utilization of tools for its classification. Recently, Vieira Bernat developed a deep learning model that classifies images into a predetermined set of topics. Moreover, images are often inaccessible to blind and low-vision users. In a recent work, Kreiss et al. delve into this matter and characterized image accessibility across languages through the presence of image's alt descriptions. Unfortunately, only 10% of the images, on average, are covered by alt descriptions, showing a universal accessibility issue.

Visio-linguistic datasets and multimodal learning Thanks to the availability of free, large-scale Wikipedia content data[3], a parallel line of research has worked on building massive multimodal datasets. DBpedia Commons (DBC) (Vaidya et al., 2015) was the first dataset offering a large media collection of images and their metadata from Wikimedia Commons. Two years later, Ferrada et al. presented IMGpedia which enhances the DBC dataset with visual descriptors and similiratiry relations between images. Recently, to accompany the milestone improvements in the field of multimodal learning, Srinivasan et al. have introduced the Wikipedia-based Image Text (WIT) dataset, composed of around 37.5M entity rich image-text pairs across 108 Wikipedia languages. In a follow-up work, Burns et al. presented the Wikipedia Webpage 2M (WikiWeb2M) suite, the first to retain the full set of images, text and structure data available in each Wikipedia page, obtained scraping all the articles in the WIT dataset. Similarly, the Authoring Tools for Multimedia Content (AToMiC) dataset (Yang et al., 2023) is designed on top of the WIT dataset for the purpose of assisting authors of Web pages in multimedia content creation. Recently, these datasets have been used to achieve mileston improvements in many fields of multilingual, multimodal learning, such as contextualized image captioning (Nguyen et al., 2022), image-text multimodal retrieval and search (Yang et al., 2023), Web page understanding (Burns et al., 2023a), image-text alignment (Singh et al., 2022), and visual question answering (Lerner et al., 2023).

The economic and social value of images beyond Wikipedia Finally, a couple of works have studied the economic and social value of the visual side of Wikipedia beyond the platform. Heald et al. (Heald et al., 2015) first developed a methodology to estimate the monetary value added by Wikipedia's public domain images in terms of costs saved to Web page developers and increased traffic to Web pages. Drawing on the same methods, Erickson et al. (Erickson et al., 2018b) estimated a potential contribution of around $29B from donwstream use of Wikimedia Commons images over the lifetime of the project (until 2018).

[3] https://dumps.wikimedia.org/

Wikipedia readers Despite the importance of images on Wikipedia is recognized in several research areas, their role and impact on the Wikipedia readership is barely been investigated. In a small survey with 15 participants, Lucassen et al. (Lucassen et al., 2010) showed that images are among the most important features that readers use to estimate the trustworthiness of Wikipedia articles. Studying art-related pictures, Navarrete et al. found that images of paintings receive a much larger online audience than the physical artworks do in the museums. During the years, important limitations have held down the development of this line of research: surveys and lab-based experiments, tipically involving small samples consisting of biased populations (e.g., university students), are often not representative and hard to generalize (Olteanu et al., 2019); also, studies based on aggregated and filtered public versions of the traces generated by Wikipedia readers (e.g., pageviews and page-to-page transitions), although capturing local, page-level behaviors accurately (Arora et al., 2022), may lack relevant reader and image-specific preferences.

This dissertation aims to fill this gap with a comprehensive overview of the importance of the visual content for the Wikipedia readership.

Chapter 3

Visual content on Wikipedia

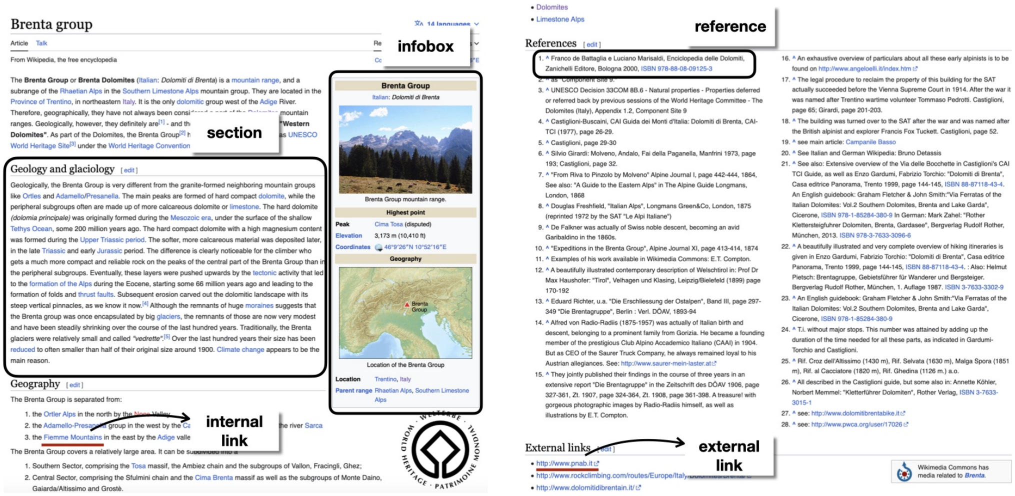
3.1 Wikipedia

Twenty years after its birth, Wikipedia has become the largest online source of free encyclopedic content and one of the most visited sites on the Web. Its purpose is "to benefit readers by containing information on all branches of knowledge."[1]. Wikipedia is hosted by the Wikimedia Foundation (WMF)[2], a non-profitable organization, together with many other sister project, such as Wikidata[3], Wikinews[4], and Wikimedia Commons[5], the largest online media file repository.

Wikipedia content is added collaboratively by hundreds of thousands of volunteers from all over the world[6]. Anyone can become a Wikipedia editor, adding text, references, and images to already existing content or newly created articles. What is written must conform with Wikipedia's policies[7]: edits must be done in good faith, verifiable by sources[8], and kept neutral with respect to editors' opinions and personal beliefs[9]. To contribute, editors write content in a markup language called *Wikitext*[10], also known as *wiki markup* or *wikicode*. The Wikitext consists of a variety of keywords and a dedicated syntax that the *MediaWiki*[11] PHP engine converts into HTML format for the browser.

Wikipedia reflects the diversity of its contributors and readers, and is therefore available in multiple languages. As of May 2023, there exists 320 active Wikipedia language editions[12], containing more than 60M articles. The number of articles is not

[1] https://en.wikipedia.org/wiki/Wikipedia:About
[2] https://wikimediafoundation.org/
[3] https://www.wikidata.org/
[4] https://en.wikinews.org/
[5] https://commons.wikimedia.org/
[6] https://en.wikipedia.org/wiki/Wikipedia:Wikipedians
[7] https://en.wikipedia.org/wiki/Wikipedia:Editing_policy
[8] https://en.wikipedia.org/wiki/Wikipedia:Verifiable
[9] https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view
[10] https://en.wikipedia.org/wiki/Help:Wikitext
[11] https://www.mediawiki.org/wiki/Manual:What_is_MediaWiki?
[12] https://meta.wikimedia.org/wiki/List_of_Wikipedias



Brenta group

14 languages

Article Talk

From Wikipedia, the free encyclopedia

The **Brenta Group** or **Brenta Dolomites** (Italian: *Dolomiti di Brenta*) is a mountain range, and a subrange of the Rhaetian Alps in the Southern Limestone Alps mountain group. They are located in the Province of Trentino, in northeastern Italy. It is the only dolomitic group west of the Adige River. Therefore, geographically, they have not always been consi mountain ranges. Geologically, however, they definitely are[1] - and th **"Western Dolomites"**. As part of the Dolomites, the Brenta Group[2] h as UNESCO World Heritage Site[3] under the World Heritage Convention

Geology and glaciology [edit]

Geologically, the Brenta Group is very different from the granite-formed neighboring mountain groups like Ortles and Adamello/Presanella. The main peaks are formed of hard compact dolomite, while the peripheral subgroups often are made up of more calcareous dolomite or limestone. The hard dolomite *(dolomia principale)* was originally formed during the Mesozoic era, under the surface of the shallow Tethys Ocean, some 200 million years ago. The hard compact dolomite with a high magnesium content was formed during the Upper Triassic period. The softer, more calcareous material was deposited later, in the late Triassic and early Jurassic period. The difference is clearly noticeable for the climber who gets a much more compact and reliable rock on the peaks of the central part of the Brenta Group than in the peripheral subgroups. Eventually, these layers were pushed upwards by the tectonic activity that led to the formation of the Alps during the Eocene, starting some 66 million years ago and leading to the formation of folds and thrust faults. Subsequent erosion carved out the dolomitic landscape with its steep vertical pinnacles, as we know it now.[4] Although the remnants of huge moraines suggests that the Brenta group was once encapsulated by big glaciers, the remnants of those are now very modest and have been steadily shrinking over the course of the last hundred years. Traditionally, the Brenta glaciers were relatively small and called *"vedrette"*.[5] Over the last hundred years their size has been reduced to often smaller than half of their original size around 1900. Climate change appears to be the main reason.

Geography [edit]

The Brenta Group is separated from:

1. the Ortler Alps in the north by the Noce Valley
2. the Adamello-Presanella group in the west by the Ca the river Sarca
3. the Fiemme Mountains in the east by the Adige valle

The Brenta Group covers a relatively large area. It can be subdivided into a

1. Southern Sector, comprising the Tosa massif, the Ambiez chain and the subgroups of Vallon, Fracingli, Ghez;
2. Central Sector, comprising the Sfulmini chain and the Cima Brenta massif as well as the subgroups of Monte Daino, Gaiarda/Altissimo and Grostè.

Brenta Group

Italian: *Dolomiti di Brenta*

Brenta Group mountain range.

Highest point

Peak	Cima Tosa (disputed)
Elevation	3,173 m (10,410 ft)
Coordinates	46°9′26″N 10°52′16″E

Geography

Location of the Brenta Group

Location	Trentino, Italy
Parent range	Rhaetian Alps, Southern Limestone Alps

- Dolomites
- Limestone Alps

References [edit]

1. ^ Franco de Battaglia e Luciano Marisaldi, Enciclopedia delle Dolomiti, Zanichelli Editore, Bologna 2000, ISBN 978-88-08-09125-3
2. ^ as "Component Site 9."
3. ^ UNESCO Decision 33COM 8B.6 - Natural properties - Properties deferred or referred back by previous sessions of the World Heritage Committee - The Dolomites (Italy), Appendix 1.2, Component Site 9
4. ^ Castiglioni-Buscaini, CAI Guida dei Monti d'Italia: Dolomiti di Brenta, CAI-TCI (1977), page 26-29.
5. ^ Castiglioni, page 29-30
6. ^ Silvio Girardi: Molveno, Andalo, Fai della Paganella, Manfrini 1973, page 193; Castiglioni, page 32.
7. ^ "From Riva to Pinzolo by Molveno" Alpine Journal I, page 442-444, 1864, See also: "A Guide to the Eastern Alps" in The Alpine Guide Longmans, London, 1868
8. ^ Douglas Freshfield, "Italian Alps", Longmans Green&Co, London, 1875 (reprinted 1972 by the SAT "Le Alpi Italiane")
9. ^ De Falkner was actually of Swiss noble descent, becoming an avid Garibaldino in the 1860s.
10. ^ "Expeditions in the Brenta Group", Alpine Journal XI, page 413-414, 1874
11. ^ Examples of his work available in Wikimedia Commons: E.T. Compton.
12. ^ A beautifully illustrated contemporary description of Welschtirol in: Prof Dr Max Haushofer: "Tirol", Velhagen und Klasing, Leipzig/Bielefeld (1899) page 170-192
13. ^ Eduard Richter, u.a. "Die Erschliessung der Ostalpen", Band III, page 297-349 "Die Brentagruppe", Berlin : Verl. DÖAV, 1893-94
14. ^ Alfred von Radio-Radiis (1875-1957) was actually of Italian birth and descent, belonging to a prominent family from Gorizia. He became a founding member of the prestigious Club Alpino Accademico Italiano (CAAI) in 1904. But as CEO of the Saurer Truck Company, he always remained loyal to his Austrian allegiances. See: http://www.saurer-mein-laster.at
15. ^ They jointly published their findings in the course of three years in an extensive report "Die Brentagruppe" in the Zeitschrift des DÖAV 1906, page 327-361, Zt. 1907, page 324-364, Zt. 1908, page 361-398. A treasure! with gorgeous photographic images by Radio-Radiis himself, as well as illustrations by E.T. Compton.
16. ^ An exhaustive overview of particulars about all these early alpinists is to be found on http://www.angeloelli.it/index.htm
17. ^ The legal procedure to reclaim the property of this building for the SAT actually succeeded before the Vienna Supreme Court in 1914. After the war it was named after Trentino wartime volunteer Tommaso Pedrotti. Castiglioni, page 65; Girardi, page 201-203.
18. ^ The building was turned over to the SAT after the war and was named after the British alpinist and explorer Francis Fox Tuckett. Castiglioni, page 52.
19. ^ see main article: Campanile Basso
20. ^ See Italian and German Wikipedia: Bruno Detassis
21. ^ See also: Extensive overview of the Via delle Bocchette in Castiglioni's CAI TCI Guide, as well as Enzo Gardumi, Fabrizio Torchio: "Dolomiti di Brenta", Casa editrice Panorama, Trento 1999, page 144-145, ISBN 88-87118-43-4. An English guidebook: Graham Fletcher & John Smith:"Via Ferratas of the Italian Dolomites: Vol.2 Southern Dolomites, Brenta and Lake Garda", Cicerone, ISBN 978-1-85284-380-9 In German: Mark Zahel: "Rother Klettersteigfuhrer Dolomiten, Brenta, Gardasee", Bergverlag Rudolf Rother, München, 2013. ISBN 978-3-7633-3096-6
22. ^ A beautifully illustrated and very complete overview of hiking itineraries is given in Enzo Gardumi, Fabrizio Torchio: "Dolomiti di Brenta", Casa editrice Panorama, Trento 1999, page 144-145, ISBN 88-87118-43-4. : Also: Helmut Pietsch: Brentagruppe, Gebietsführer für Wanderer und Bergsteiger. Bergverlag Rudolf Rother, München, 1. Auflage 1987. ISBN 3-7633-3302-9
23. ^ An English guidebook: Graham Fletcher & John Smith:"Via Ferratas of the Italian Dolomites: Vol.2 Southern Dolomites, Brenta and Lake Garda", Cicerone, ISBN 978-1-85284-380-9
24. ^ T.i. without major stops. This number was attained by adding up the duration of the time needed for all these parts, as indicated in Gardumi-Torchio and Castiglioni.
25. ^ Rif. Croz dell'Altissimo (1430 m), Rif. Selvata (1630 m), Malga Spora (1851 m), Rif. al Cacciatore (1820 m), Rif. Ghedina (1126 m.) a.o.
26. ^ All described in the Castiglioni guide, but some also in: Annette Köhler, Norbert Memmel: "Kletterführer Dolomiten", Rother Verlag, ISBN 3-7633-3015-1
27. ^ see: http://www.dolomitibrentabike.it
28. ^ see: http://www.pwca.org/user/17026

External links [edit]

- http://www.pnab.it
- http://www.rockclimbing.com/routes/Europe/Italy
- http://www.dolomitidibrentain.it/

Wikimedia Commons has media related to *Brenta*.

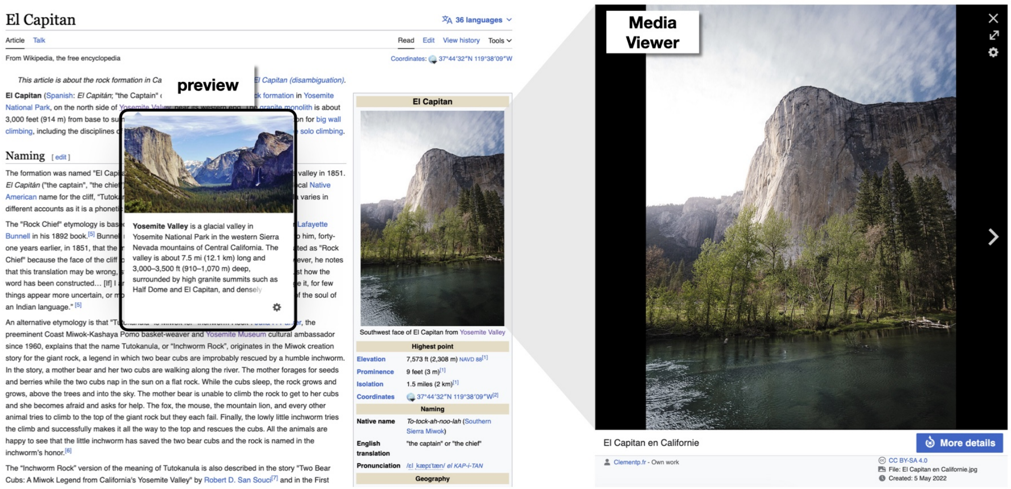
FIGURE 3.1: Example of an article in the English Wikipedia.

evenly distributed: the English Wikipedia is the largest one, with around 6.7M articles, while over 50 languages have fewer than 10K articles. A similar trend is observed in the distribution of contributors, with the English Wikipedia having around 100K active users, six times more than the second one, French.

Wikipedia's encyclopedic content is conveyed through articles[13] (Figure 13.1). Articles mostly contain written information, but also images, links (both to external websites and other Wikipedia pages) and references. Articles can be divided into smaller paragraphs, called *sections*[14], to better organize their content into a coherent structure. Within an article, readers can interact with its content in several ways. Links can be clicked to reach external Web pages (*external* links), navigate another article within the same Wikipedia (*wikilinks* or internal links), or jump to a page of another Wikimedia project website (*interwiki* links). References are usually marked in the text as superscript footnote numbers. Hovering these numbers display a reference tooltip, while clicking on them brings the reader to the *References* section[15]. Recently, the *page preview* feature has been introducted on desktop devices[16]. Page previews are small tooltips that are displayed whenever a reader hovers over a link to another article (Figure 3.2, right side). This tooltip contains a short summary of the subject of the article linked and an image, if available. Finally, also images can be clicked to visualize their content in further detail. When clicked, images are displayed in a visualization tool called Media Viewer[17] (Figure 3.2, left side). The Media Viewer overlays the article and displays the image in a larger size, with additional metadata below.

Since its birth, Wikipedia has maintained an open access policy to increase global

[13] https://en.wikipedia.org/wiki/Wikipedia:What_is_an_article?
[14] https://en.wikipedia.org/wiki/Help:Section
[15] https://en.wikipedia.org/wiki/Wikipedia:Citing_sources
[16] https://www.mediawiki.org/wiki/Page_Previews
[17] https://en.wikipedia.org/wiki/Wikipedia:Media_Viewer



El Capitan

From Wikipedia, the free encyclopedia

Coordinates: 37°44′32″N 119°38′09″W

This article is about the rock formation in Ca … *El Capitan (disambiguation).*

El Capitan (Spanish: *El Capitán*; "the Captain" … ck formation in Yosemite National Park, on the north side of Yosemite Valley, near its western end. The granite monolith is about 3,000 feet (914 m) from base to su… on for big wall climbing, including the disciplines o… solo climbing.

Naming [edit]

The formation was named "El Capi… valley in 1851. *El Capitán* ("the captain", "the chief… cal Native American name for the cliff, "Tutok… varies in different accounts as it is a phoneti…

The "Rock Chief" etymology is bas… Lafayette Bunnell in his 1892 book.[5] Bunnel… him, forty-one years earlier, in 1851, that the… ted as "Rock Chief" because the face of the cliff… ever, he notes that this translation may be wrong,… st how the word has been constructed… [If] I a… e it, for few things appear more uncertain, or m… f the soul of an Indian language." [5]

An alternative etymology is that "Tu… er, the preeminent Coast Miwok-Kashaya Pomo basket-weaver and Yosemite Museum cultural ambassador since 1960, explains that the name Tutokanula, or "Inchworm Rock", originates in the Miwok creation story for the giant rock, a legend in which two bear cubs are improbably rescued by a humble inchworm. In the story, a mother bear and her two cubs are walking along the river. The mother forages for seeds and berries while the two cubs nap in the sun on a flat rock. While the cubs sleep, the rock grows and grows, above the trees and into the sky. The mother bear is unable to climb the rock to get to her cubs and she becomes afraid and asks for help. The fox, the mouse, the mountain lion, and every other animal tries to climb to the top of the giant rock but they each fail. Finally, the lowly little inchworm tries the climb and successfully makes it all the way to the top and rescues the cubs. All the animals are happy to see that the little inchworm has saved the two bear cubs and the rock is named in the inchworm's honor.[6]

The "Inchworm Rock" version of the meaning of Tutokanula is also described in the story "Two Bear Cubs: A Miwok Legend from California's Yosemite Valley" by Robert D. San Souci[7] and in the First

Yosemite Valley is a glacial valley in Yosemite National Park in the western Sierra Nevada mountains of Central California. The valley is about 7.5 mi (12.1 km) long and 3,000–3,500 ft (910–1,070 m) deep, surrounded by high granite summits such as Half Dome and El Capitan, and densely

El Capitan	
Southwest face of El Capitan from Yosemite Valley	
Highest point	
Elevation	7,573 ft (2,308 m) NAVD 88[1]
Prominence	9 feet (3 m)[1]
Isolation	1.5 miles (2 km)[1]
Coordinates	37°44′32″N 119°38′09″W[2]
Naming	
Native name	*To-tock-ah-noo-lah* (Southern Sierra Miwok)
English translation	"the captain" or "the chief"
Pronunciation	/ɛl ˌkæpɪˈtæn/ *el KAP-i-TAN*
Geography	

El Capitan en Californie

Clementp.fr - Own work

CC BY-SA 4.0
File: El Capitan en Californie.jpg
Created: 5 May 2022

FIGURE 3.2: Examples of modalities through which readers can interact with images on a Wikipedia article: directly clicking on them (the images are then visualized in a different page, called the *Media Viewer* or indirectly through article *previews*, a tooltip showing a short summary—eventually illustrated—of linked articles (only via desktop browsers).

access to knowledge[18]. For this reason, every month the WMF releases the content of Wikipedia, and of all its wiki projects, in a public online database[19]. The WMF also collects a vast set of user interactions which are stored as server logs by the analytics instrumentations[20]. These logs contist of all the requests to the Wikimedia's servers, including pageviews, image clicks, page previews, and page-to-page transitions via internal links. These datasets, containing sensitive information, such as IPs and geolocations, are deleted every 90 days for privacy reasons, and accessed only with priviledges. To foster research, the WMF also releases pageview and page-to-page transaction data aggregated at the page level with hourly, monthly, and country granularity[21].

3.2 Images on Wikipedia

Images are a crucial aspect of the reader's experience on Wikipedia as they can help break down complex concepts and ideas, making them easier to grasp and retain. The English Wikipedia guidelines state that "Wikipedia uses a variety of multimedia files to enhance content and explain concepts that are difficult to convey via text alone"[22].

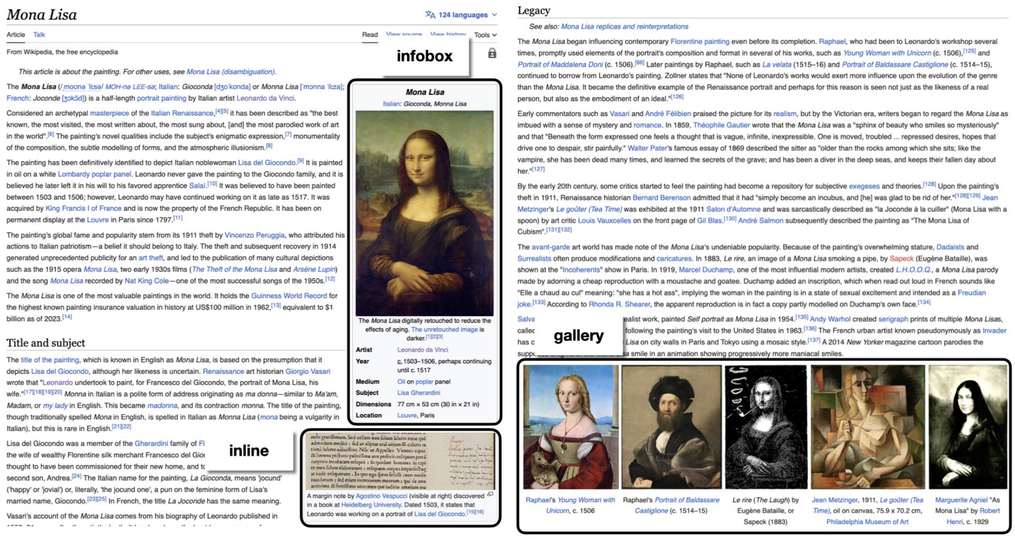
[18] https://foundation.wikimedia.org/wiki/Policy:Open_access_policy
[19] https://dumps.wikimedia.org/
[20] https://wikitech.wikimedia.org/wiki/Analytics/Data_Lake/Traffic/Webrequest
[21] https://dumps.wikimedia.org/other/analytics/
[22] https://en.wikipedia.org/wiki/Help:Viewing_media

Image contributors Images are added to Wikipedia by its large community of editors, following a Manual of Style maintained by the Wikipedia community[23]. In essence, images added to Wikipedia articles have to be relevant to the article's content and of high photographic quality. However, they should be significant illustrative aids to understand concepts, hence not primarily decorative. Images can be uploaded locally to a specific Wikipedia language edition, but the majority of images in the encyclopedia are hosted in Wikimedia Commons[24], the largest free visual knowledge repository. Wikimedia Commons was launched in 2004 as a media file repository, with the purpose of making public domain and freely licensed educational media content publicly available to everyone. From its start, Commons has constantly grown, reaching a total of 90M media files (around 90% of which are images). Images on Commons must then be either of the public domain or licensed under a free license allowing anyone to reuse the material for any purpose, including commercial purposes. All the files uploaded to Commons can be used like locally uploaded files on all other Wikimedia projects, including the different Wikipedia sites, facilitating the reuse of visual content across languages.

Types of images and characteristics Wikimedia Commons contains huge amounts of images with a variety of formats, typologies, and characteristics. Photographic images are mostly uploaded in JPEG format (84%)[25], but other popular formats, like PNG and SVG, are supported as well. Mathematical formulas can also be created and presented as images through an editing tool that converts a subset of TEX markup into PNG files. To make it easy to users to find the images they are looking for, images are often tagged. Once tagged, images are automatically assigned to categories. Like Wikipedia articles, images can have multiple categories that are often organized into a tree structure. Popular categories are *Nature*, *Society*, *Culture*, *Science*, and *Engineering*, each of them with several subcategories. Tags are manually assigned by volunteers, making this process laborious and time consuming. Images are also categorized according to their quality and value. *Featured* images represent what the community has chosen to be of the highest quality among all the images on the site. *Quality* images are required to be generally well-composed and executed, and must be produced by Commons contributors. *Value* images are considered particularly valuable by the Commons community for use in other Wikimedia sites or other online contexts. They do not need to be of extraordinary technical quality, but must be of extreme value in terms of diversity and usability, and are meant to encourage other editors' to contribute to subjects that are difficult to illustrate or obtain.

On the Wikipedia articles, images can be placed in different ways (Figure 3.3). Image can be found in the *infobox*, a table summarizing the main facts about the article's

[23] https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Images
[24] https://commons.wikimedia.org/wiki/Main_Page
[25] https://commons.wikimedia.org/wiki/Special:MediaStatistics



Mona Lisa

From Wikipedia, the free encyclopedia

This article is about the painting. For other uses, see Mona Lisa (disambiguation).

The **Mona Lisa** (/ˌmoʊnə ˈliːsə/ *MOH-nə LEE-sə*; Italian: *Gioconda* [dʒoˈkonda] or *Monna Lisa* [ˈmɔnna ˈliːza]; French: *Joconde* [ʒɔkɔ̃d]) is a half-length portrait painting by Italian artist Leonardo da Vinci.

Considered an archetypal masterpiece of the Italian Renaissance,[4][5] it has been described as "the best known, the most visited, the most written about, the most sung about, [and] the most parodied work of art in the world".[6] The painting's novel qualities include the subject's enigmatic expression,[7] monumentality of the composition, the subtle modelling of forms, and the atmospheric illusionism.[8]

The painting has been definitively identified to depict Italian noblewoman Lisa del Giocondo.[9] It is painted in oil on a white Lombardy poplar panel. Leonardo never gave the painting to the Giocondo family, and it is believed he later left it in his will to his favored apprentice Salaì.[10] It was believed to have been painted between 1503 and 1506; however, Leonardo may have continued working on it as late as 1517. It was acquired by King Francis I of France and is now the property of the French Republic. It has been on permanent display at the Louvre in Paris since 1797.[11]

The painting's global fame and popularity stem from its 1911 theft by Vincenzo Peruggia, who attributed his actions to Italian patriotism—a belief it should belong to Italy. The theft and subsequent recovery in 1914 generated unprecedented publicity for an art theft, and led to the publication of many cultural depictions such as the 1915 opera *Mona Lisa*, two early 1930s films (*The Theft of the Mona Lisa* and *Arsène Lupin*) and the song *Mona Lisa* recorded by Nat King Cole—one of the most successful songs of the 1950s.[12]

The *Mona Lisa* is one of the most valuable paintings in the world. It holds the Guinness World Record for the highest known painting insurance valuation in history at US$100 million in 1962,[13] equivalent to $1 billion as of 2023.[14]

Title and subject

The title of the painting, which is known in English as *Mona Lisa*, is based on the presumption that it depicts Lisa del Giocondo, although her likeness is uncertain. Renaissance art historian Giorgio Vasari wrote that "Leonardo undertook to paint, for Francesco del Giocondo, the portrait of Mona Lisa, his wife."[17][18][19][20] *Monna* in Italian is a polite form of address originating as *ma donna*—similar to *Ma'am*, *Madam*, or *my lady* in English. This became *madonna*, and its contraction *monna*. The title of the painting, though traditionally spelled *Mona* in English, is spelled in Italian as *Monna Lisa* (*mona* being a vulgarity in Italian), but this is rare in English.[21][22]

Lisa del Giocondo was a member of the Gherardini family of Fl the wife of wealthy Florentine silk merchant Francesco del Gioc thought to have been commissioned for their new home, and tc second son, Andrea.[24] The Italian name for the painting, *La Gioconda*, means 'jocund' ('happy' or 'jovial') or, literally, 'the jocund one', a pun on the feminine form of Lisa's married name, Giocondo.[23][25] In French, the title *La Joconde* has the same meaning.

Vasari's account of the *Mona Lisa* comes from his biography of Leonardo published in

Mona Lisa
Italian: *Gioconda, Monna Lisa*

The *Mona Lisa* digitally retouched to reduce the effects of aging. The unretouched image is darker.[1][2][3]

Artist	Leonardo da Vinci
Year	c. 1503–1506, perhaps continuing until c. 1517
Medium	Oil on poplar panel
Subject	Lisa Gherardini
Dimensions	77 cm × 53 cm (30 in × 21 in)
Location	Louvre, Paris

A margin note by Agostino Vespucci (visible at right) discovered in a book at Heidelberg University. Dated 1503, it states that Leonardo was working on a portrait of Lisa del Giocondo.[15][16]

Legacy

See also: Mona Lisa replicas and reinterpretations

The *Mona Lisa* began influencing contemporary Florentine painting even before its completion. Raphael, who had been to Leonardo's workshop several times, promptly used elements of the portrait's composition and format in several of his works, such as *Young Woman with Unicorn* (c. 1506),[125] and *Portrait of Maddalena Doni* (c. 1506).[66] Later paintings by Raphael, such as *La velata* (1515–16) and *Portrait of Baldassare Castiglione* (c. 1514–15), continued to borrow from Leonardo's painting. Zollner states that "None of Leonardo's works would exert more influence upon the evolution of the genre than the *Mona Lisa*. It became the definitive example of the Renaissance portrait and perhaps for this reason is seen not just as the likeness of a real person, but also as the embodiment of an ideal."[126]

Early commentators such as Vasari and André Félibien praised the picture for its realism, but by the Victorian era, writers began to regard the *Mona Lisa* as imbued with a sense of mystery and romance. In 1859, Théophile Gautier wrote that the *Mona Lisa* was a "sphinx of beauty who smiles so mysteriously" and that "Beneath the form expressed one feels a thought that is vague, infinite, inexpressible. One is moved, troubled ... repressed desires, hopes that drive one to despair, stir painfully." Walter Pater's famous essay of 1869 described the sitter as "older than the rocks among which she sits; like the vampire, she has been dead many times, and learned the secrets of the grave; and has been a diver in the deep seas, and keeps their fallen day about her."[127]

By the early 20th century, some critics started to feel the painting had become a repository for subjective exegeses and theories.[128] Upon the painting's theft in 1911, Renaissance historian Bernard Berenson admitted that it had "simply become an incubus, and [he] was glad to be rid of her."[128][129] Jean Metzinger's *Le goûter (Tea Time)* was exhibited at the 1911 Salon d'Automne and was sarcastically described as "la Joconde à la cuiller" (Mona Lisa with a spoon) by art critic Louis Vauxcelles on the front page of Gil Blas.[130] André Salmon subsequently described the painting as "The Mona Lisa of Cubism".[131][132]

The avant-garde art world has made note of the *Mona Lisa*'s undeniable popularity. Because of the painting's overwhelming stature, Dadaists and Surrealists often produce modifications and caricatures. In 1883, *Le rire*, an image of a *Mona Lisa* smoking a pipe, by Sapeck (Eugène Bataille), was shown at the "Incoherents" show in Paris. In 1919, Marcel Duchamp, one of the most influential modern artists, created *L.H.O.O.Q.*, a *Mona Lisa* parody made by adorning a cheap reproduction with a moustache and goatee. Duchamp added an inscription, which when read out loud in French sounds like "Elle a chaud au cul" meaning: "she has a hot ass", implying the woman in the painting is in a state of sexual excitement and intended as a Freudian joke.[133] According to Rhonda R. Shearer, the apparent reproduction is in fact a copy partly modelled on Duchamp's own face.[134]

Salva ealist work, painted *Self portrait as Mona Lisa* in 1954.[135] Andy Warhol created serigraph prints of multiple *Mona Lisa*s, callec following the painting's visit to the United States in 1963.[136] The French urban artist known pseudonymously as Invader has c *Lisa* on city walls in Paris and Tokyo using a mosaic style.[137] A 2014 *New Yorker* magazine cartoon parodies the suppo a smile in an animation showing progressively more maniacal smiles.

Raphael's *Young Woman with Unicorn*, c. 1506

Raphael's *Portrait of Baldassare Castiglione* (c. 1514–15)

Le rire (*The Laugh*) by Eugène Bataille, or Sapeck (1883)

Jean Metzinger, 1911, *Le goûter (Tea Time)*, oil on canvas, 75.9 x 70.2 cm, Philadelphia Museum of Art

Marguerite Agniel "As Mona Lisa" by Robert Henri, c. 1929

FIGURE 3.3: Examples of the three types of locations (*infobox*, *inline*, and *gallery*) where images are placed within a Wikipedia article.

subject, usually placed in the top-right corner of the page on desktop browsers or at the top of the screen on mobile browsers. Images can also be added *inline*, namely individually near the relevant text in the article body. Finally, when images are too many to be placed within the text body, they can be collected into *galleries*, collections of images generally added at the bottom of articles. On the English Wikipedia (as of March 2021), the majority of the images are placed inline (48%)—often at the top of articles—, 36% are added to infoboxes, and only 16% can be found in galleries. A more detailed characterization of the visual content on the English Wikipedia will be provided in Chapter 4.

Visual content across languages The space of visual content on Wikipedia is huge and rapidly growing. As of May 2023, there are 54M images across 320 Wikipedia language editions. The distribution of illustrations largely varies across editions (Figure 3.4A). Among the languages with at least 100 articles, English Wikipedia contains the largest amount of images, with 6.5M images in 6.7M articles, while Nigerian Pidgin Wikipedia only contains 188 images in 955 articles. Across languages, on average 50% of articles are illustrated. In the English Wikipedia, around 50% of the articles are illustrated. In proportion, Navajo Wikipedia is the most illustrated one, with 93% of articles that contain images, while Pali Wikipedia is the least illustrated (4.6%). The average number of images in illustrated articles is more uniformly distributed (Figure 3.4B). On average, there are 2.6 illustrations per article across languages. The majority of the language editions have between 2 and 4 images per article. English Wikipedia has 3 images per article on average, while Moksha Wikipedia is the edition with the highest average number of images per article (10).

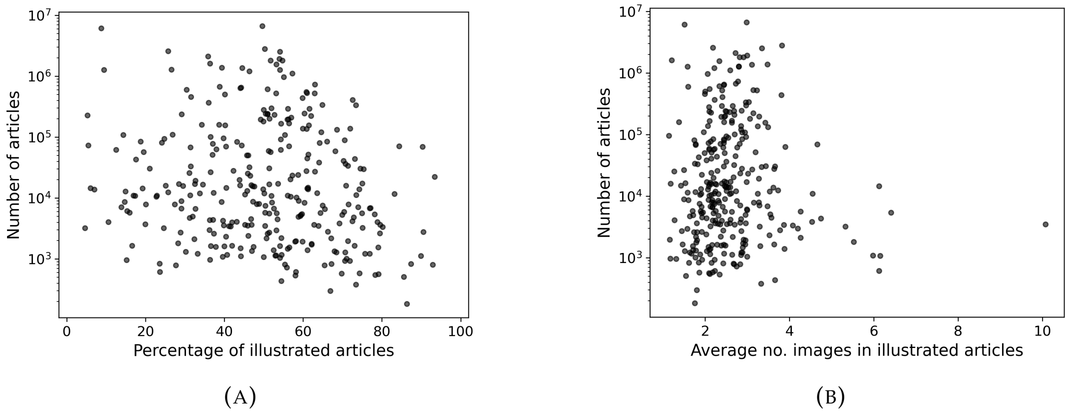


FIGURE 3.4: Wikipedia size as a function of (a) the percentage of illustrated articles and (b) the average number of images per article.

Initiatives to enhance the visual side of Wikipedia One reason for the large volume of missing images is the effort needed to illustrate articles. Wikipedia editors need to find the right image match for an article by searching through millions of images in Wikimedia Commons. When relevant images are not available in Commons, editors have two options. On one hand, editors can create and publish new visual content on their own. However, the creation of content of good quality often requires adequate technical skills, special equipment, and time. Contributors without these possibilities have to search through online sources. While less comples, this process may still be demanding. First, the right pictures for an article need to exist somewhere on the Web otherwise someone—Wikipedia editors, photographers, GLAM institutions, or other users—must create or retrieve them. Second, images have to be free to reuse. If images are not free-licensed, editors' and authors' efforts will be needed to make them publicly available. Only then, images can be hosted in Wikimedia Commons and finally added to Wikipedia articles. To help with these efforts, the Wikimedia movement organizes several initiatives. Among the most notable there are Wiki Loves Monuments[26], Wiki Loves Earth[27], and Wiki Loves Folklore[28] which are photographic contests trying to encourage participants to contribute with images of cultural and natural heritage from all around the world. Moreover, the Wikipedia Pages Wanting Photoscampaign[29] is designed to help editors add images to unillustrated Wikipedia articles, also promoting the use of digital media files collected from various Wikimedia photography contests. These initiatives have achieved significant results over the years, and proved to be powerful tools to engage users and improve the visual content on Wikipedia. For instance, the WPWP campaign contributed to illustrate around 100K articles on the English Wikipedia alone over the last three years[30].

[26] https://www.wikilovesmonuments.org/

[27] https://wikilovesearth.org/

[28] https://wikilovesfolklore.org/

[29] https://meta.wikimedia.org/wiki/Wikipedia_Pages_Wanting_Photos

[30] https://meta.wikimedia.org/wiki/Wikipedia_Pages_Wanting_Photos/Results

Chapter 4

Quantifying interactions with images on Wikipedia

4.1 Scope

A vast body of literature in experimental psychology has shown the impact of images for learning and engaging with knowledge. Images positively affect comprehension and increase attention on the textual material (W Howard Levie et al., 1982a). In the context of online open-knowledge, Wikipedia, accessed by millions of readers every month, has emerged as a prominent platform. Given its crucial role as a central hub for knowledge sharing and learning, understanding how images are used on Wikipedia is crucial. However, while other aspects of Wikipedia have been widely studied (Yasseri et al., 2012; Lemmerich et al., 2019a; Halfaker and Geiger, 2020), little is known about visual content and its usage, with only a few studies looking at cross-language image diversity (He et al., 2018a), and the communities of Wikipedia "image" editors (Fernanda B Viegas, 2007a).

In this Chapter, we fill this gap by providing for the first time a comprehensive overview of how readers interact with images in the English Wikipedia. We quantify and characterize reader engagement with images when browsing the encyclopedia using traffic data and we explore the role played by images in the exploration of free knowledge. To operationalize reader engagement, we adopt the most widely-used metrics in web user studies (Bakhshi et al., 2014; Park et al., 2015): we compute click-through rate on images, and conversion rate on illustrated and unillustrated page previews. While only partially representing the complex, multifaceted notion of interest (Constantin, Redi, et al., 2019), these implicit signals do reflect an expression of engagement with visual content and they provide a solid baseline for an initial overview of readers' interactions with images. More specifically, we address three major research questions:

RQ1 **To what extent are readers interacting with images on Wikipedia?** And what is the relation with engagement values on other types of content?

RQ2 What drives reader's engagement with images when reading Wikipedia articles? What are the visual and contextual factors that influence image interactions?

RQ3 Do images support reader's need for additional information when navigating Wikipedia? Are images helpful to delve into contextual information provided by the article?

In addressing these questions, we make the following contributions:

- We collect a large dataset of reader interactions with images in English Wikipedia over one month and characterize the landscape of Wikipedia images with several features inspired by experimental psychology and web user studies (Section 4.2.3). We quantify reader engagement with images and find that, on average, readers click with images 1 in every 29 pageviews on English Wikipedia, ten times more often than with references (RQ1, Section 4.3).
- To visualize the factors impacting reader engagement with Wikipedia images, we perform a set of multivariate analyses on the image features extracted and find that readers interact more often with images of monuments, maps, vehicles, and unfamiliar faces (RQ2, Section 4.4).
- To understand whether images support readers' need for additional contextual information when navigating Wikipedia, we design a matched observational study based on page previews, i.e., the short article summaries that are displayed when users hover on links to other Wikipedia pages (RQ3, Section 4.5). We find a negative effect of the presence of images on the proportion of articles' page previews that convert into a visualization of the full article page.

We thus conclude that the visual preferences of Wikipedia readers are radically different compared to web users in photo-sharing platforms or image search engines, where images of people and celebrities largely predominate. We also find that images on Wikipedia appear to fulfill part of the *cognitive* function typical of illustrations in instructional settings supporting readers' information need. Finally, we discuss theoretical implication of this research and its important repercussions on how the Wikipedia communities organize and prioritize the inclusion of visual content and how the broader web and content creators could contribute to the web with free visual knowledge.

4.2 Data collection and methods

To answer our research questions, we first need to estimate the volume of Wikipedia articles and their images, collect data about reader interactions with those, and characterize them through feature extraction.

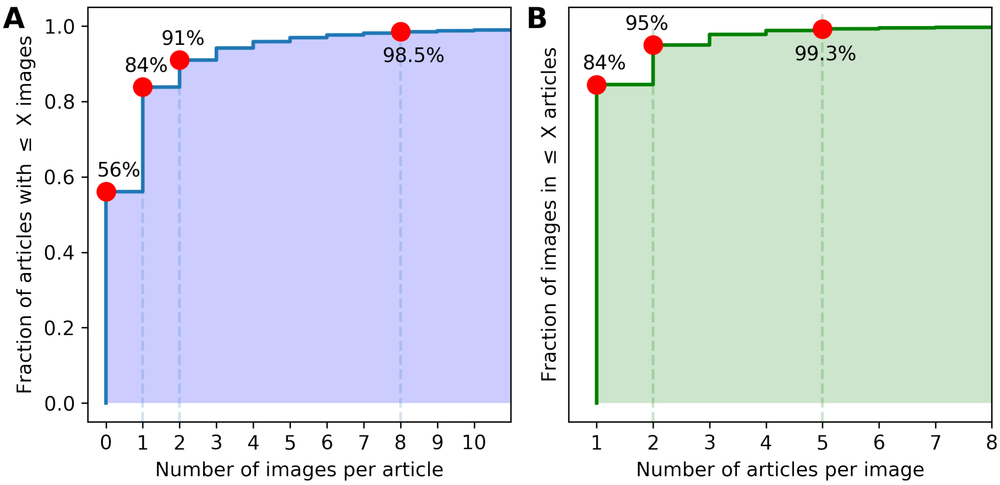


FIGURE 4.1: Cumulative distributions of (A) number of images per article and (B) the number of articles per image.

4.2.1 Collecting article and image counts data

To measure the number of articles and images, we use the HTML version of English Wikipedia at the end of March 2021. We collect 6.2M documents, and we parse them to extract the images' URLs, caption text, resolution, and position on the page. Using the CSS class in the HTML code, we exclude all images that appear as icons (for example, portals or Wikiprojects). Additionally, for each page, we also record the article length as the number of characters.

Out of the 6.2M articles, 2.7M (44%) contain at least one image, for a total of 5M unique images across all English Wikipedia articles. The vast majority of the articles (91%) contain two images or less, while only 1.5% has more than eight images (Figure 4.1A). On average, there are 2.3 images per illustrated article. Around 84% of images is unique to the article where it appears, while 16% of the images appear in more than one article (Figure 4.1B).

4.2.2 Collecting article and image traffic data

We collect the reader interactions with images for desktop and mobile browsers by processing the server access logs[1] for the period from 1st to 28th of March 2021. We restrict our analysis to only human interactions by ignoring traffic from bots thanks to a set of heuristics developed by Wikimedia's Analytics team[2]. For privacy reasons, we use an anonymized version without sensitive information. Since the logs do not contain any explicit identifier for the user, before the anonymization, we assign a random id based on IP and user-agent similar to previous work (Lemmerich et al., 2019b). In addition, we discard all the events coming from logged-in users, the events of any user that edited a page, and the events originated from countries where

[1] https://wikitech.wikimedia.org/wiki/Analytics/Data_Lake/Traffic/Webrequest

[2] https://techblog.wikimedia.org/2020/10/05/bot-or-not-identifying-fake-traffic-on-wikipedia/

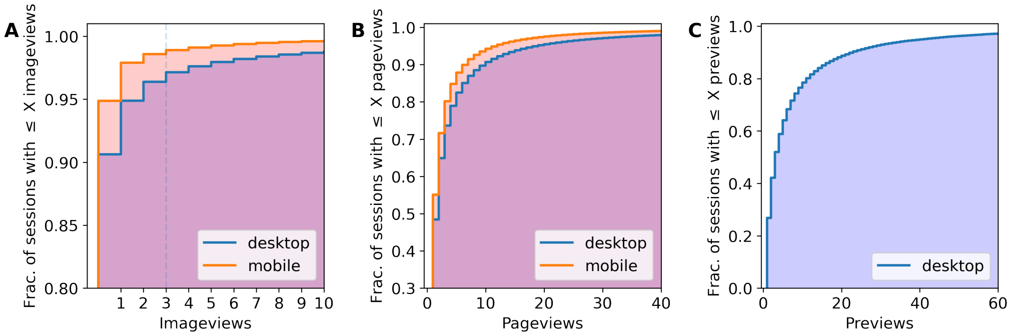


FIGURE 4.2: Cumulative distributions of number of sessions by (A) imageviews, (B) pageviews, and (C) previews partitioned by desktop (in blue) and mobile device (in orange). Page previews are available only on desktop devices.

not all days have more than 500 pageviews consistently. This filtering ensures more privacy for the Wikipedia readers by dropping around 3% of the data.

Over the selected period, we extract from the web logs all requests that reflect three types of actions:

- **Imageviews:** these requests correspond to image visualizations in the Media Viewer after a user clicks on an article image (Figure 3.2, left side);
- **Pageviews:** these are requests logged every time a user visits a Wikipedia page. For the scope of this study, we select only pageviews of articles with at least one image;
- **Page previews:** these requests are logged whenever a user hovers over a link to an article (Figure 3.2, right side). To remove the effect of casually generated page previews, we only keep those previews that are shown for at least one second. Note that page previews are generated only on desktop devices.

We aggregate these image-related events at the user level by using the previously assigned identifier to obtain sorted sequences of actions from the same user, which we refer to as *sessions*.

Finally, to exclude the potential impact of exogenous time-dependent events' and the consequent traffic directed by external image search engines to Wikipedia, we filter out all the incoming traffic generated from Google Image Search, which represents by far the most used image retrieval engine from which people access Wikipedia's visual content. Nevertheless, the pageviews originating from Google Image search account for 0.006% of the total, making their impact negligible.

In our data collection, we extract interactions for 1.5B sessions. In Figure 4.2 we report the distributions of sessions by the number of imageviews, pageviews, and previews. The distributions are heavily skewed, with 91% and 94% of sessions having less than 10 pageviews on desktop and mobile devices, respectively, and 99% of sessions having less than 10 imageviews both on desktop and mobile devices during our data collection period. Similarly, 79% of sessions have generated less than 10 previews. Users with extensive sessions (i.e. "power users"), that may be over-represented, are therefore limited in our analysis. Over one month, 100% of the illustrated articles have been loaded at least once, accounting for a total of 7.1B pageviews, 461M imageviews, and 49M previews events in our dataset. We find that most pageviews are generated from mobile devices (59% from the mobile site), while most imageviews are generated from desktop (58% from desktop).

4.2.3 Mining image content and context

To investigate the factors that make images engaging when reading Wikipedia, we characterize the pictures in our dataset with several features related to the visual context and content. Our choice of features is largely inspired by the literature around the cognitive perception of images in instructional or web environments.

Contextual factors

Images on Wikipedia are not isolated items. Instead, they exist in *context*, providing epistemic support to the article they are illustrating. To extract features from the image context, we resort to previous literature on Wikipedia reader behavior and experimental psychology studies on the role of images in instructional settings. Note that 16% of the images appear in multiple articles. Since the same image may appear in very different articles, thus belonging to very different contexts, we treat such images as distinct.

Page topic In a previous study, Piccardi et al. (Piccardi, Redi, et al., 2020b) found that Wikipedia reader engagement with references varies with the article topic. To test whether the reader's need for visual support similarly varies across subject matters, we extract, for each page in our dataset, a *topic vector*, using the Wikidata topic model[3]. The classifier takes as input the Wikidata item of a Wikipedia page, and it returns a 64-dimensional vector containing the probability that the article belongs to the topics of the Wikiproject hierarchy[4]. To reduce the dimensionality of the topic vector, we consider the second level of the topic taxonomy accounting for 31 topics. We then rearranged some of the topics into coarse-grained topics, namely media, internet culture, and performing arts into *entertainment*, chemistry and biology into *biology*, computing and libraries & information into *computer science*, mathematics and physics into *maths & physics*. Figure 4.3A shows the distribution of images by article topic. Geographic articles are the most illustrated, containing 1/4 of the images in our dataset. Biographies, making up 30% of the articles on Wikipedia, also contain around 15% of the images. Topics such as entertainment (movies, plays, books),

[3] Wikidata topic model. https://github.com/geohci/wikidata-topic-model. Accessed March 2021.
[4] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Directory

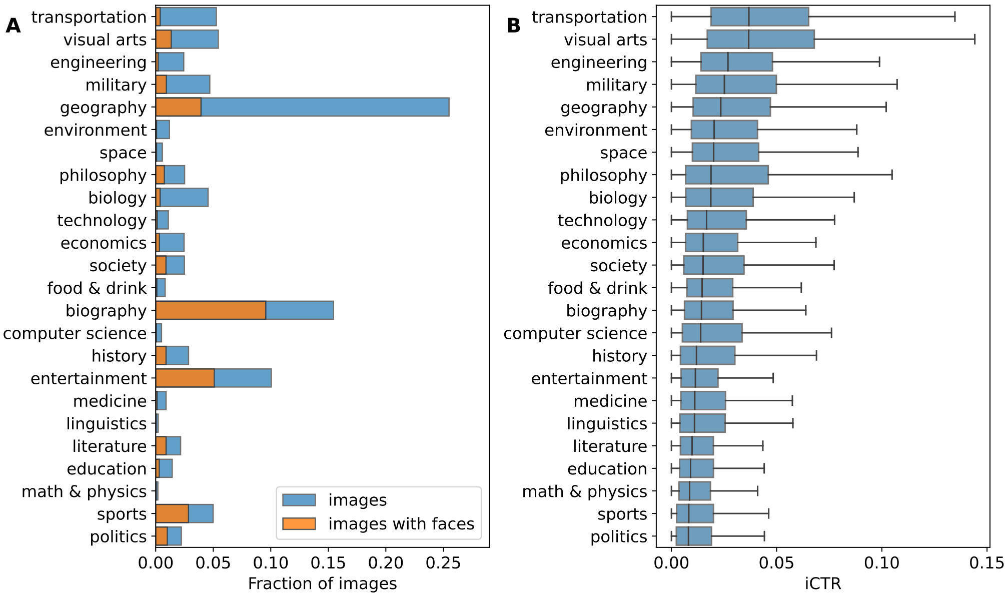


FIGURE 4.3: (A) Fraction of images by topic (in blue) and fraction of images with faces (in orange). (B) Image-specific CTR by article topic.

visual arts, transportation, military, biology, and sports follow, covering together another third of the images in English Wikipedia. A summary of the numerical values can be found in the Supplementary Material (Supplementary Table).

Page length One of the possible functions of text illustrations in learning contexts is to enrich and complement the textual content with additional material (W Howard Levie et al., 1982a). To investigate to which extent images are used to complement the lack of textual information, we measure the textual richness as the *length* of each article in characters. The distribution of the number of images by text length as shown in Figure 4.4 is log-normal, with most images in English Wikipedia being found in articles between 1K and 100K characters long.

Page popularity Previous work analyzing reader behavior with respect to Wikipedia citations (Piccardi, Redi, et al., 2020b) found that there is an inverse relation between article popularity and reference click-through rate. To test whether this relation is valid also in the case of interactions with images, we compute a page *popularity* feature for each image by computing the total monthly pageviews for the page where an image appears. As in Piccardi et al. (Piccardi, Redi, et al., 2020b), the page popularity follows a power-law distribution (Figure 4.5), with 70% images having average monthly pageviews in the range between 50 and 10K.

Readability Although not completely verified, another function of images in textual knowledge is to facilitate text comprehension, especially in the case of reading difficulties (W Howard Levie et al., 1982a). To take into account this function in our study, we quantify the reading ease by computing the Flesch *readbility* score,

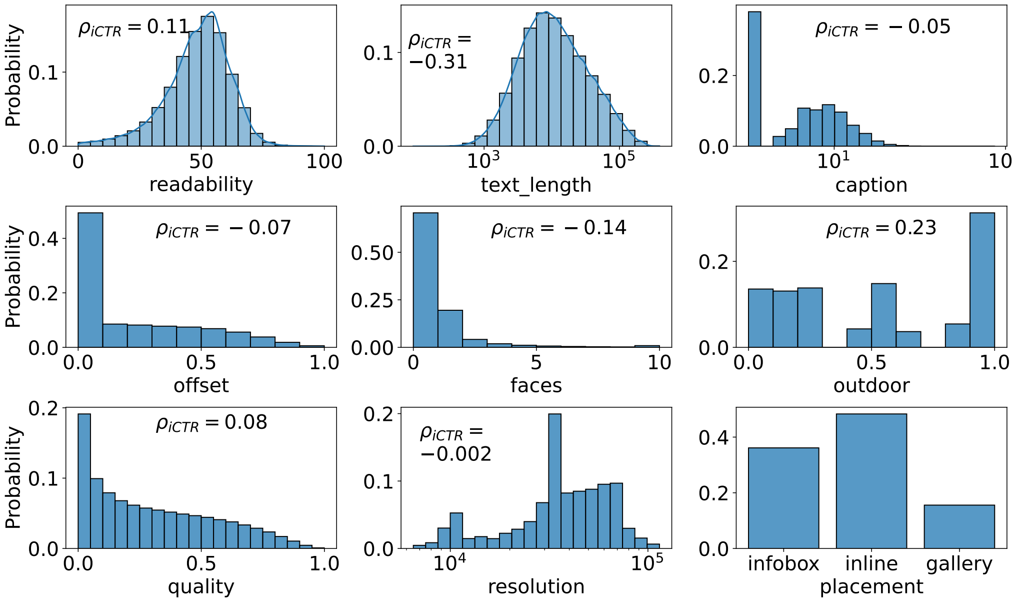


FIGURE 4.4: Feature distributions. Spearman's rank correlation coefficient ρ between the numerical features and the iCTR at the top of each panel ($p < 0.001$ for each feature).

reflecting the "comprehension difficulty of written material" (Flesch, 1948), on the text of each article in our dataset. We compute the readability score for all the pages containing an image, and plot the resulting distribution in Figure 4.4: most of the images on Wikipedia are in articles detected as "Fairly difficult to read" (score 50-60), or "Difficult to read" (score 30-50).

Length of the image caption Studies in educational technologies have found that the usage of captions marginally enhances the usefulness of text illustrations (Bernard, 1990). To operationalize the presence of captions as a contextual feature of the images in our dataset, we store the average number of words used to caption each image when appearing in a Wikipedia article. We can see from Figure 4.4 the caption length following a Tweedie distribution with a large fraction of the images without a description and the majority of existing captions centered around ten words.

Image placement How images are placed in the text can play a crucial role in the knowledge exploration experience (Peeck, 1993), and researchers investigating Wikipedia reader behavior showed that people tend to engage more with content (in this case, internal hyperlinks) which lies at the top of the article (Paranjape et al., 2016). At the same time, Wikipedia editors follow specific placement guidelines when illustrating an article. To investigate the role of image placement on Wikipedia article consumption, we extract the image's *text offset*, i.e., the relative position of the image with respect to the length of the article, as well as the image *position*, a categorical feature which can take the values $\{infobox, inline, gallery\}$, depending on the template used to add the image to the article. From the plots in Figure 4.4 we can see that only 36% of the images in our dataset is generally placed in infoboxes, while only 16% can be found in galleries, and that the majority of inline images are generally placed at the top of the article (see *offset*). A summary of the numerical values can be found in the Supplementary Material (Supplementary Table).

Image resolution In addition to their position, the viewer's attention may also be driven by the size of an image. According to the Wikipedia's Image Size guidelines[5], editors should choose the appropriate image size in proportion of its level of details. However, readers may still tend to click on small images that are inherently difficult to observe. To investigate the role of the image size, we compute the image *resolution* in pixels for each image. As shown in Figure 4.4, image resolutions vary across different scales, mostly ranging from 10K to 100K pixels.

Visual factors

The content of pictures plays a key role in driving readers' attention to both the images (Gygli et al., 2013) and the text on the page (W Howard Levie et al., 1982a). To understand the type of visuals that elicit higher levels of interactions with Wikipedia images, we run a set of computer vision-based classifiers. Since training a classifier to detect every concept in Wikipedia's visual knowledge would be practically infeasible, we instead focus on three main indicators, based on extensive literature from visual and social interestingness prediction.

Image quality Visual aesthetics, or image quality, is one of the top visual factors driving the viewer's attention to an image (Gygli et al., 2013). At the same time, researchers have found that not all images which receive much attention from web communities are actually of high quality (Schifanella et al., 2015), and that a lot of socially uninteresting pictures are very beautiful. We investigate here whether the quality of an image plays an important role in eliciting Wikipedia reader attention. To do so, we design a *Wikipedia Image Quality* classifier, as follows.

- We collect a training set of images annotated with a binary (high/low) image quality score. To annotate images, we resort to the highly curated categories that Wikimedia Commons editors assign to images. We download 141,984 images from the *Quality images* category from Commons[6]: these are high-quality images that have to meet Commons' quality guidelines[7] before being voted and promoted as Quality images by the community through a highly selective process. Only a few images make it to the "image quality" category: there is, therefore, a large consensus on the quality of the images in that category. To collect low-quality images, we simply randomly sample an approximately equal number of pictures (169,310) from the large pool of Commons images.

[5] https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Images#Size
[6] https://commons.wikimedia.org/wiki/Commons:Quality_images
[7] https://commons.wikimedia.org/wiki/Commons:Image_guidelines

These are very likely to be low quality, as images randomly drawn from Commons tend to have a small resolution, and they are rarely used to illustrate Wikipedia articles (Erickson et al., 2018a).

- We next train a deep neural network using transfer learning: we fine-tune a pre-trained model, originally designed to classify image objects, using the image quality data collected. We use the Inception-v3 (Szegedy et al., 2016) deep network pre-trained on the 1000-classes ImageNet dataset (Deng et al., 2009), as it was proved to be a good starting dataset for transfer learning tasks (Huh et al., 2016). We use 90% of the data for training and the rest for validation, and we train the last layer of the network over 10,000 iterations with the data collected. The fine-tuned classifier achieves 77% accuracy on a balanced test set.

The resulting image quality classifier, given any image, outputs a *quality score* in the range $[0, 1]$ which corresponds to the probability that the image belongs to the "High Quality" class. As shown in Figure 4.4, most images in our dataset have a very low-quality score.

Presence of faces In line with several studies showing the importance of faces for web users' positive reactions and engagement with images (Bakhshi et al., 2014; Pappas et al., 2016), we also extract information about the presence of faces of people in the image. We use MTCNN (Wen et al., 2016) to detect faces and their bounding box in an image. For a given image, we then output a binary feature indicating whether it contains at least one face or not. We find that around 1/3 of the images on Wikipedia have at least one face (see Figure 4.4), and most of those are in articles about biographies, entertainment, and sports (see Figure 4.3A).

Outdoor setting Literature around image interestingness and aesthetics (Gygli et al., 2013) has shown that outdoor images tend to elicit the viewer's interest more than indoor images do. To extract the information about the image scene setting, we use a Wide Residual Network (Zagoruyko et al., 2016) trained on MIT's Places (Zhou et al., 2017), an image dataset with 10M images annotated with 365 scene types, and indoor/outdoor labels. This classifier, given an image, outputs an *outdoor* score which reflects the probability of the image being an outdoor scenery. When the feature is $\leq$ 0.5, the image is likely to be an indoor scenery. In our dataset, indoor and outdoor images are almost equally distributed, with a slight prevalence of outdoor pictures.

4.2.4 Engagement metrics

To quantify the volume of readers' interactions with visual content, we introduce the following metrics:

Global click-through rate. The global click-through rate (gCTR) measures the overall reader engagement with images. It is defined as the fraction of reading sessions with at least one interaction with an image. Formally, for each session s, let $C(s, p)$ be the indicator function that is 1 if at least one image was clicked on page p by the respective reader, and 0 otherwise. Moreover, let $N(p)$ be the number of distinct reading sessions during which page p was loaded. We define the global click-through rate as

$$gCTR = \frac{\sum_s \sum_p C(s, p)}{\sum_p N(p)} \tag{4.1}$$

where p ranges over the set of pages that contain at least one image.

Image-specific click-through rate. The image-specific click-through rate (iCTR) measures how much engagement a Wikipedia image elicits. It is defined as the ratio of clicks to impressions. Formally, let $N(i)$ be the number of distinct sessions with clicks on image i and $N(p_i, i)$ the number of distinct sessions that viewed page p_i where the image is placed, the image-specific click-through rate is

$$iCTR(i) = \frac{N(i)}{\sum_{p_i \in P_i} N(p_i, i)} \tag{4.2}$$

where p_i ranges over the set P_i of pages containing i.

Conversion rate. The conversion rate (CR) quantifies the probability of clicking on an article link after its preview is shown in another article. Formally, for each page p and session s, we denote by $C(s, p)$ the indicator function that is one if session s has clicked on a link to page p after seeing its preview. Moreover, we denote by $N(p)$ the total number of distinct sessions that loaded a preview of p. The conversion rate for page p can be written as:

$$CR(p) = \frac{\sum_s C(s, p)}{N(p)} \tag{4.3}$$

In the following sections, we restrict our analyses to images visualized by at least 50 readers during the period of our data collection in order to reduce the effect of rarely viewed articles and obtain a reliable estimate of the quantities above. This results in a set of 3.2M unique images displayed in 2.7M articles.

4.3 RQ1: To what extent are readers interacting with images in Wikipedia?

The first step of our analysis is to quantify the volume of readers' interactions towards visual content when reading Wikipedia. To this aim, we compute the global

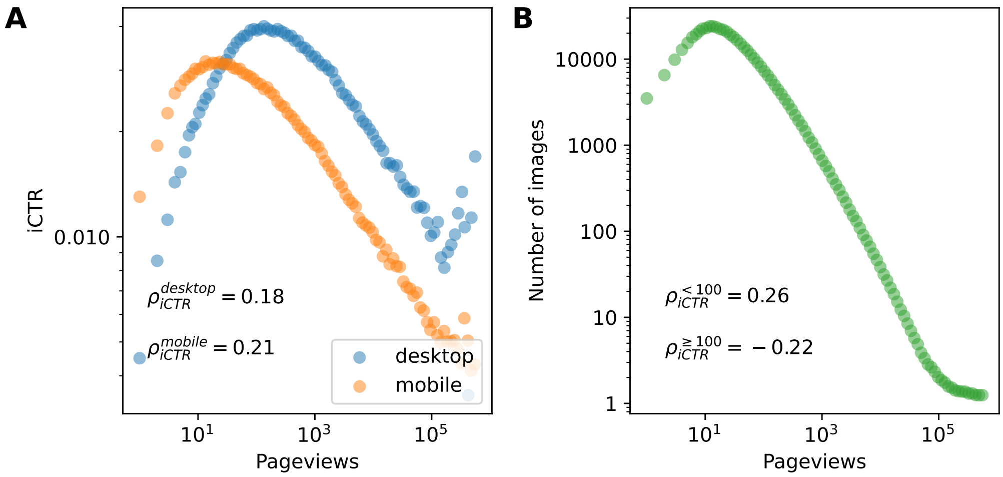


FIGURE 4.5: Distributions of (A) values of iCTR by page popularity partitioned by device and (B) number of images per page popularity. Spearman's rank correlation coefficient ρ_{iCTR} between iCTR and pageviews in the inset ($p < 0.001$). The axes are in log scale.

click-through rate and image-specific click-through rate on our data and find the following.

4.3.1 Overall engagement with images: the global click-through rate

We find that the gCTR across all pages in English Wikipedia with at least one image is 3.5%, meaning that around 3.5 out of 100 times readers visit a page, they also click on an image. This metric is higher for desktop (5.0%) and lower for mobile web users (2.6%), probably due to differences in the way readers navigate Wikipedia on the two devices and the better Media Viewer experience on desktop. Over time, the behavior also changes depending on the device used. For example, on desktop, readers tend to click more often on images during weekdays (Monday to Friday), with an increase of 5.5% over weekends. However, on mobile, there is no significant difference between week and weekends. To understand whether these values represent a high or low level of engagement, we can compare them with engagement metrics on another type of article content, namely article's references. According to Piccardi et al. (Piccardi, Redi, et al., 2020b), the gCTR on citations in English Wikipedia is 0.29%, thus around ten times lower than for images. This observation suggests that images tend to elicit a different level of engagement than those on references for English Wikipedia.

4.3.2 Average engagement with individual images: image-specific click-through rate

On average, an image in a Wikipedia article gets clicked 2.6 times every 100 impressions. Again, the iCTR is higher (3.2%) for desktop than for mobile users (2.2%). In Figure 4.6 we report examples of highly engaging and less engaging images. By

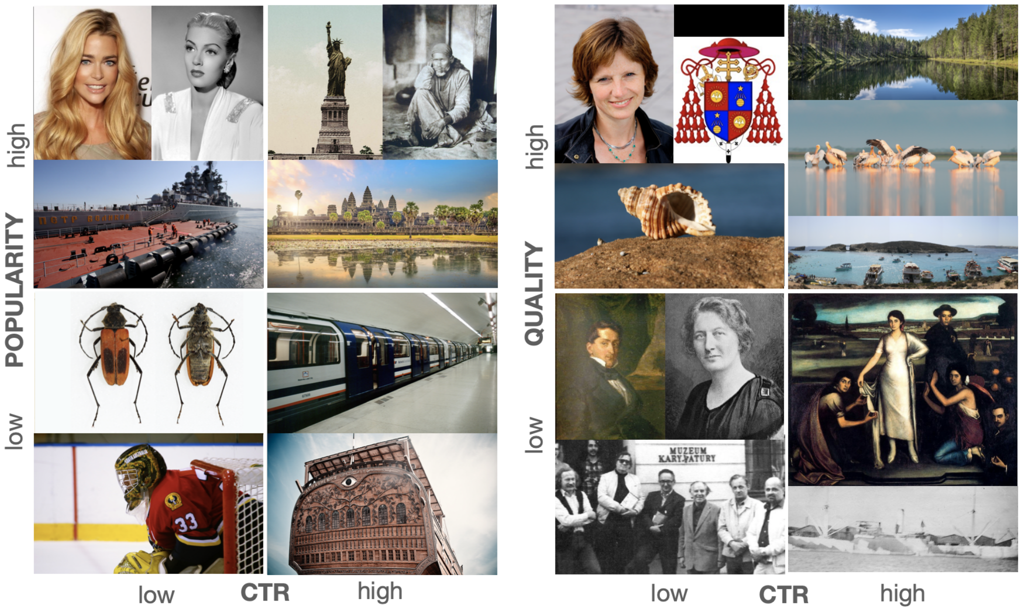


FIGURE 4.6: Examples of *high* and *low* image-specific CTR images by page popularity (left) and image quality (right). We ranked images by iCTR, popularity and quality, and picked examples from the top-100 ("high") and bottom-100 ("low") for each dimension.

visually inspecting these results, we can see some visual trends: highly engaging images seem to depict outdoor environments. In contrast, among the images with low levels of iCTR, we can find human faces.

4.4 RQ2: What drives reader's engagement with images when reading Wikipedia articles?

To address RQ2, we now model reader interaction with images on Wikipedia using the factors listed in Section 4.2.3.

4.4.1 Exploratory analysis

We start our analysis by seeking a relationship between our target metric, the iCTR, and each of the contextual and visual factors in Section 4.2.3. We report the Spearman's rank correlation coefficients ρ_{ctr} between the iCTR and the scalar predictors in Figure 4.4 and 4.5. Considering the contextual factors, we observe a negative correlation with article length and popularity ($\rho = -0.31$ and $\rho = -0.21$, respectively). When further investigating the relationship with article popularity (Figure 4.5), we find that it seems non-linear: engagement with images is low for highly unpopular pages. It becomes higher for pages in a mid-level bucket of popularity and drops again for highly viewed pages. Regarding the image size, despide images are displayed in different resolutions, this does not have a clear relation with the iCTR ($\rho = -0.002$). When considering the position in the page instead, the median iCTR is higher for images in galleries (median iCTR=0.024) than for images in the infobox

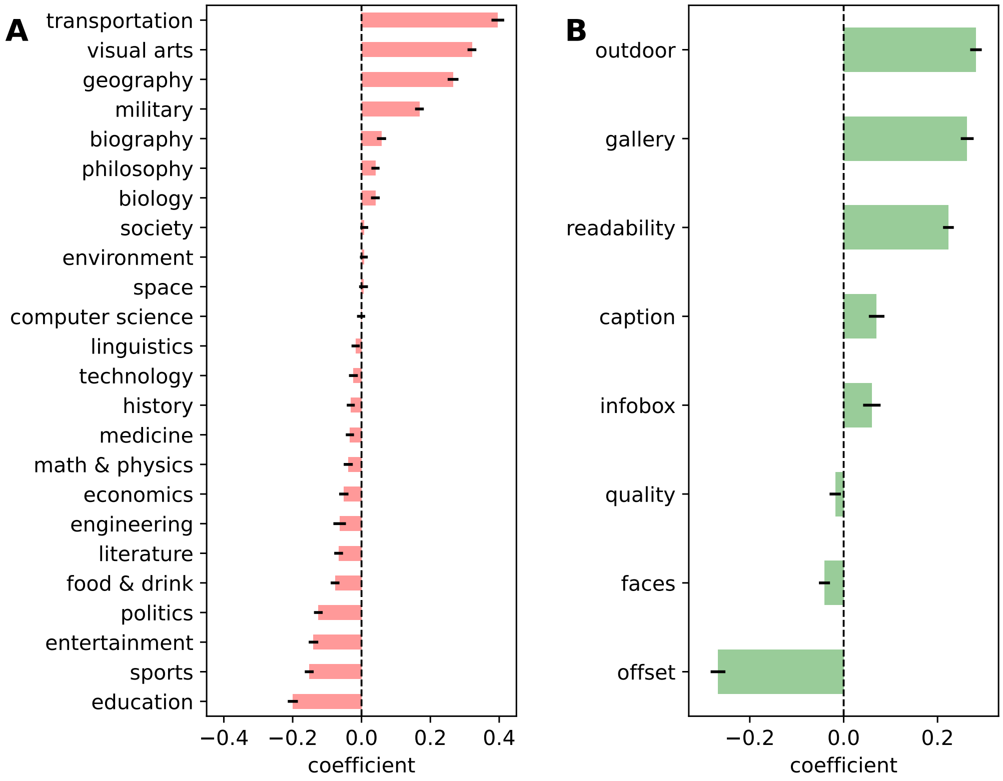


FIGURE 4.7: Association of the features with the image iCTR expressed as coefficients of the logistic regressions. (A) Coefficients of the model trained with topics of the article as predictors. (B) Coefficients of the model trained with the other variables of the image. Error bars represent 95% confidence intervals.

(median iCTR=0.019) and inline (median iCTR=0.016). Moreover, we see signals of reader visual preferences in terms of article topics (Figure 4.3B): the topics with the highest median value are transportation (0.037) and visual arts (0.037), while politics and sports show the lowest level of interaction with a median iCTR of 0.008. Finally, the correlation analysis of the visual factors confirms our initial intuitions from the visual analysis. There is a positive correlation between the iCTR and outdoor scenery ($\rho = 0.23$) and a negative relation between the presence of faces and readers' engagement ($\rho = -0.14$). A complete summary of the numerical values discussed can be found in the Supplementary Material (Supplementary Table).

4.4.2 Regression analysis

Next, we aim to understand how much these features are predictive of reader engagement with images. To do so, we perform a logistic regression analysis that classifies images according to their iCTR.

Study design We build the training set as follows. We take the median value of iCTR and label the images in our dataset with two classes of *high* and *low* iCTR according to whether their iCTR is above or below the median[8]. We use the contextual and visual factors described in Section 4.2.3 as predictors and the binary iCTR as the target variable. Moreover, we split the predictors into two sets of features and train two separate logistic regression models. The first set of features consists of the topic vectors, while the second consists of the remaining other factors. In the second set of features, we log-transform variables that span over different scales, such as page popularity, text length, caption length, and the number of faces. Moreover, to reduce the amount of multicollinearity among the predictors, we manually inspect the correlation table and compute the Variance Inflation Factor (Kutner et al., 2005) for each variable. We decide to exclude the *inline* variable, as it shows strong collinearity with *gallery* and *infobox*. Finally, we standardize each predictor in the two sets of features.

Impact of image resolution We found that images on Wikipedia are displayed in different resolutions. Before running the regression analysis, we test the hypothesis that the image size could be decisive in attracting clicks, i. e. readers may tend to click on smaller images as it may be harder to see the details. In Section 4.4.1 we found the correlation coefficient to be -0.002 (with $p < 0.001$), indicating no clear relationship between the two variables. Moreover, we observe that image resolution is highly related to its position within the page: the median resolution is about 46, 36, and 11 megapixels respectively for images in the infobox, inline, and in galleries. Also, image resolution is highly correlated with some topics, e. g. it has large positive correlation with biography and entertainment, and large negative correlation with geography and visual arts. Since the image resolution does not seem to be directly related to the iCTR, while it seems to be influenced by some other independent variables, and thus may act as a confounder, we decide not to take it into account in the subsequent analyses.

Controlling for page length and popularity Similar to what was described in previous work on engagement with Wikipedia content (Piccardi, Redi, et al., 2020b), we found that the page popularity and the text length have strong negative correlations with the target. Since page popularity and text length show large variations across the other predictors, especially across topics, we remove the effect of these two confounding variables with a matched study. We build a bipartite graph with images of low and high iCTR as nodes of the two halves. We split the log-transformed page popularity and text length ranges into 100 bins of equal size each, and assign the nodes to these bins, linking two nodes of opposite iCTR when falling into the same bins of popularity and length. Finally, we use min-weight matching on the bipartite

[8] We repeat the logistic regression analysis with different thresholds splitting the two classes, namely we focus on the highest vs. the lowest percentiles of the images according to their iCTR. We find no significant differences on the resulting regression coefficients. Therefore, we choose the median as the cutoff to maximize the presence of images in the analysis.

graph to find pairs of high/low iCTR samples that minimize the Euclidean distance between all pairs. This procedure succesfully balanced the dataset, with the standardized mean difference of text length and pape popularity across the two classes dropping from −0.54 and −0.51 to −0.010 and −0.007, respectively.

Results The resulting regression models have an area under the ROC curve (AUC) of 0.67 and 0.62 for the model trained on the topics and the model trained on the other variables, respectively. Figure 4.7 shows the resulting models coefficients. In Figure 4.7A, we observe that clicks on images are more often related to topics such as transportation, visual arts, geography, and military. On the contrary, clicks on images are less likely in education, sports, and entertainment articles. In Figure 4.7B, we observe that the most important negative predictor is the text offset, i.e. the relative position of the image with respect to the length of the article, meaning that images are more clicked if placed in the upper part of an article. Regarding the visual content, we observe a strong positive effect of outdoor settings, consistently with the positive coefficients of transportation and geography, topics in which a large portion of images display outdoor scenes. Regarding the image position on the page, we find that images in galleries show a high level of engagement, as well as images in the infobox, even though with a moderate effect. Noteworthy, the presence of faces has negative impact in predicting a high level of interactions with images, contrary to what we would expect from the literature (Bakhshi et al., 2014). In the remainder of this section, we further investigate this inconsistency in depth, by performing a clustering experiment and an observational study on the images in our dataset.

4.4.3 Identifying prototypical image groups

To dive deeper into the results emerging from the regression analysis, we provide in this section a non-linear multivariate analysis of our data.

Study design Our goal is to draw a complementary picture of the complex interplay between reader engagement and image features, identifying prototypical groups of Wikipedia images with homogeneous characteristics. To this extent, we perform a density-based clustering using HDBSCAN (Campello et al., 2013), which seeks partitions with high density areas of points separated by low density areas, possibly containing noise objects. The advantage of using HDBSCAN is threefold: first, its density-based structure allows to better identify areas of continuous, non-globular points compared to other clustering algorithms that rely on the assumptions of spherical shape clusters, e.g., k-means (Lloyd, 1982). Second, by labeling the sparse background points as noise, it aggregates data into coherent clusters rather than partitions. Finally, it extends DBSCAN (Ester et al., 1996) by implementing a hierarchical clustering approach that allows to extract the optimal flat grouping based on the stability of the clusters, allowing to find groups with non homogeneous density in contrast to a global density threshold adopted by DBSCAN.

We run HDBSCAN[9] on the features set described in Section 4.4.2 including the binary iCTR variable and limiting the analysis to the eight most popular topics (geography, biography, entertainment, visual arts, transportation, sports, military, and biology) that account for 92% of the images in our corpus. HDBSCAN has two main hyper-parameters that have significant practical effect on the clustering: *min_cluster_size* which refers to the minimum number of grouped items to consider as a cluster, and *min_samples* which provides a measure of how conservative the clustering would be defining the level at which points are considered noise. The larger the value, the more conservative the clustering, that implies more points will be declared as noise, and clusters will be restricted to progressively more dense areas. We explore the hyper-parameter space with a grid search approach to find the best configuration that maximizes the Density-Based Clustering Validation (DBCV) index (Moulavi et al., 2014). Due to computational constraints, we perform the clustering on a random sample of 50K images, we repeat the procedure 5 times to assess the stability of the tuning phase. We achieve the best configuration with $min\_cluster\_size = 600$ and $min\_samples = 5$ in the majority of the runs. With these settings, we identify 23 clusters, with a number of images ranging between 600 and 5000.

Results We summarize in Figure 4.8 the characteristics of the centroids of the 12 most populated clusters, where each facet represents the mean value of that feature across the examples in that cluster. For ease of visualization, we discretize continuous variables in three classes: *low*, *medium*, or *high*, according to whether the value falls, respectively, in the first, second, or third quantile of the feature distribution. To provide a more clear visual representation of the clusters, we labeled them with descriptive names. We also manually inspected the images in each cluster and chose two to four representative images among the most popular ones.

In the rest of this section, we explore more in depth image quality and its interplay with images containing faces. Even though quality appears, on aggregate, to be moderately positively associated with the tendency to click on images, the underlying phenomenology is more nuanced. On one hand, high-quality images within the geography, transportation, visual arts, military, and biology categories (clusters 2, 3, 5, 6, 7, 8, and 9) show high iCTR across a wide range of contextual factors. A large portion of these images depicts outdoor sceneries that is coherent with the positive coefficient of the *outdoor* feature in the regression in Section 4.4.2. On the other hand, low quality images are often associated with the presence of faces, especially in topics such as biography, entertainment, and sports, wich overall tend to have a lower click-through rate. Focusing on the interplay between biographies and iCTR reveals significant differences across page popularity and topics worth studying. Images within unpopular biographies, predominantly inline and with a

[9] To run the algorithm, we use the *hdbscan* Python library (McInnes et al., 2017): https://hdbscan.readthedocs.io.

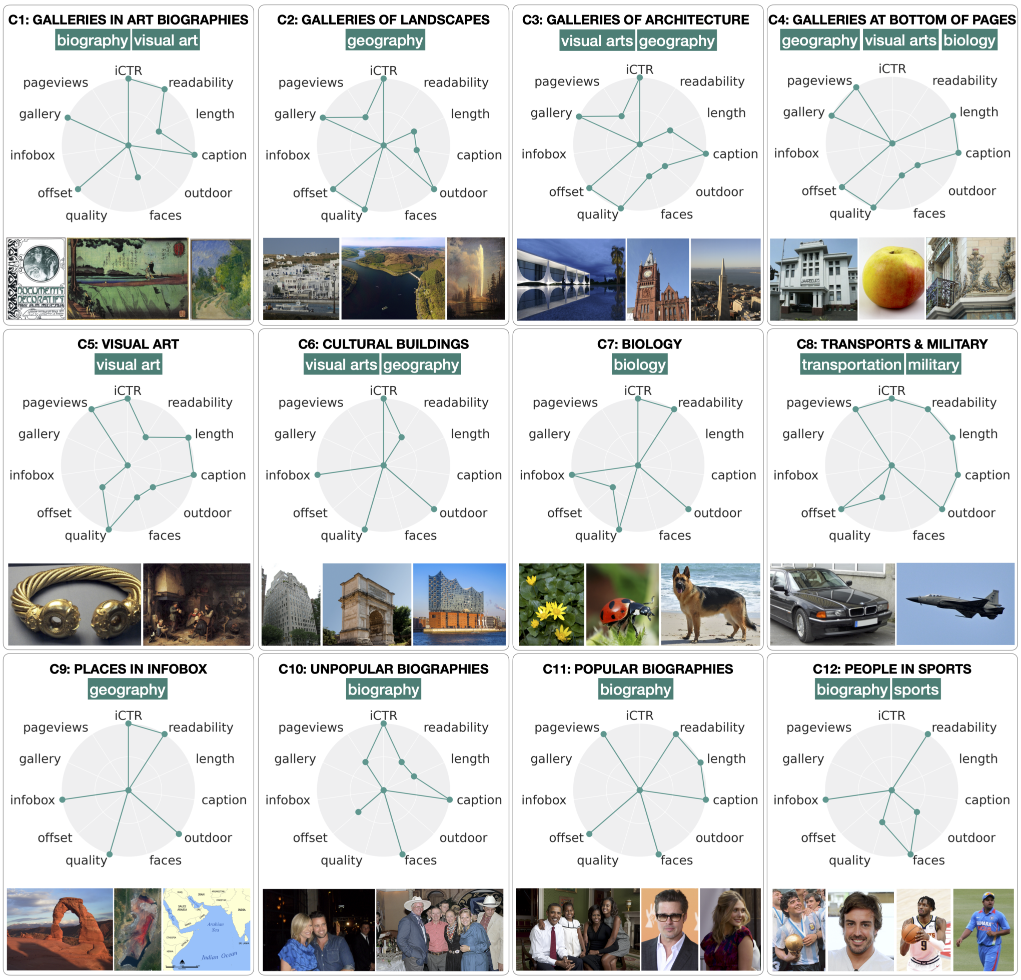


FIGURE 4.8: Visual representation of the clustering. The radar plots show for a group centroid the intensity of each feature on a three classes scale. We summarize in green the topics that cover at least 85% of the images categories in a cluster.

curated textual description, show high iCTR (cluster 10), as well as images placed in galleries in biographies of unpopular artists (cluster 1). On the contrary, popular biographies (cluster 11) or pages that present popular athletes (cluster 12), experience a low iCTR. A possible explanation for this behavior is that users may tend to click on an image in a biography if they do not recognize immediately the subject depicted, while for prominent celebrities, especially if the image is accessible in the infobox, the information need is fulfilled without the need of a click and the interaction with the Media Viewer.

4.4.4 Are faces engaging on Wikipedia?

As pointed out in Section 4.2.3, images with faces generally elicit high social engagement. In Section 4.4.2, we found that the number of faces has negative weight with respect to the iCTR, while in Section 4.4.3 we observed that Wikipedia readers are more likely to click on images with faces only when placed in less popular biographies. To further investigate this aspect, we design a matched observational study in which we compare the iCTR between images with and without faces. To reduce the effects of confounding factors, we perform a pairwise comparison of images with similar covariates using propensity score matching.

Propensity score matching Propensity score matching (Abadie et al., 2006) is a statistical technique to evaluate the efficacy of a treatment against a control group, while taking into account the effect of confounding factors. The propensity score is defined as the probability of a sample being treated as a function of the covariates, and it is obtained by training a logistic regression with the covariates as predictors, and the treatment/control variable as target. As a result, observations with the same propensity scores have the same distribution across the observed covariates.

In our experiment, we define images with at least one face as receiving the treatment, images without a face as the control group, and the variables used in the logistic regression (except for the topics and the page popularity) as the covariates.

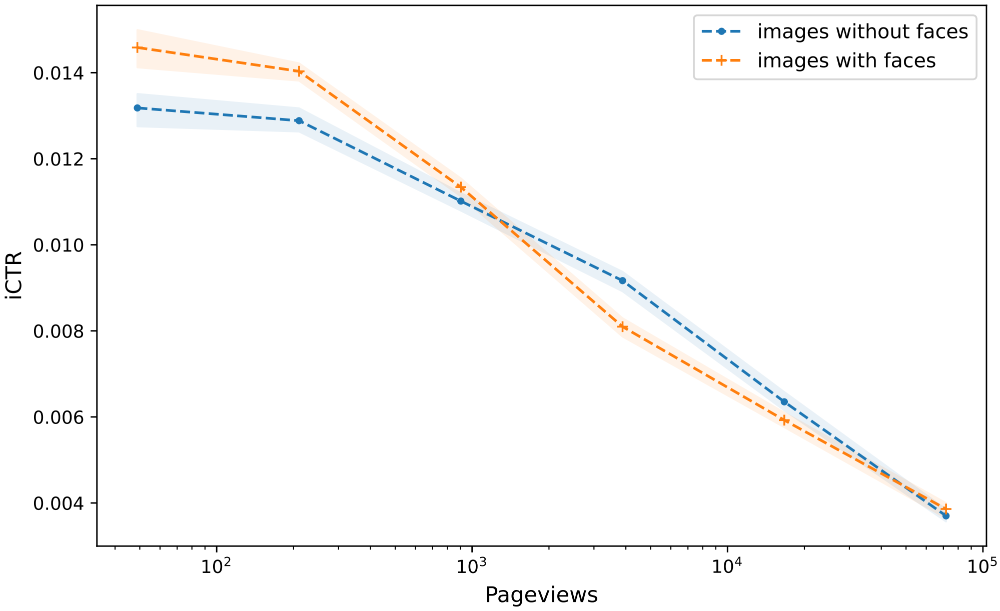


FIGURE 4.9: Comparison of the iCTR for images with faces (orange) and without faces (blue) as function of the popularity (*pageviews*). Error bands represent bootstrapped 95% CIs.

Results We consider images in articles about biography, entertainment, and sports, accounting for 90% of all images with at least one face. We find pairs of images minimizing the propensity score within pairs of articles. Figure 4.9 shows the iCTR as a function of the page popularity, for images with (in orange) and without (in blue) faces. According to a Mann-Whitney U test, the difference between the two distributions is statistically significant, with $p < 0.001$. The tendency to click on images with faces varies depending on page popularity. On pages with less that 1,000 monthly pageviews, the presence of faces induces higher level of interactions, with a difference of 0.1%, whereas, after 1,000 pageviews, we observe the opposite behavior, with a difference of 0.06%. This also confirms the findings of the clustering analysis.

To ascertain that our findings remain valid also for non-biographical articles, we replicate the same study by including all the topics in the matching procedure. In this case, we observe a different behavior. Images with faces are less likely to be clicked than others, across all the popularity range. This may explain the overall negative coefficient of the faces feature in the regression analysis, and highlight the role that faces play in increasing engagement on biographical articles.

4.5 RQ3: Do images support reader's need for additional information when navigating Wikipedia?

We found that readers show a signal of interest in images when reading Wikipedia articles. But are images useful to fulfill part of the reader's information need when navigating the website? To address this question, we design an additional study that attempts to estimate whether the presence of an image in an article preview can complement the textual information and support in-depth reading.

Matching articles. To check the difference in terms of conversion rate between articles having and not having an image, we first need to reduce the impact of exogenous factors that may potentially drive reader attention on articles, other than the presence of an image. For example, events localized in time can have the effect of sporadically increasing the interest towards specific articles, and therefore on the number of edits (Georgescu et al., 2013). Similarly, the probability of clicking on an article may also depend on its centrality in the article network, i.e. on its *in-degree*, which is the number of page links pointing to that article. Ideally, we would like to find pairs of articles—one with, the other without image in the preview—that are similar in such factors. To control for these factors, we resort again to propensity score matching. In this experiment, articles with an image in the preview are the treatment group, articles without images are the control, and we use text length, number of edits, and in degree as variables for the matching procedure.

Results We find pairs of articles by minimizing the propensity score within pairs of articles. Figure 4.10 shows the conversion rate as a function of article popularity (total number of page views), for articles with (in blue) and without (in yellow) an image in the preview. We find that, according to a Mann-Whitney U test, the difference is statistically significant ($p < 0.001$), across all the popularity spectrum, with a difference of 2% in the conversion rate. We rank all pages by conversion rate, and manually inspect the top and bottom articles, with and without images. We find that most of the illustrated articles with higher conversion rate tend to be

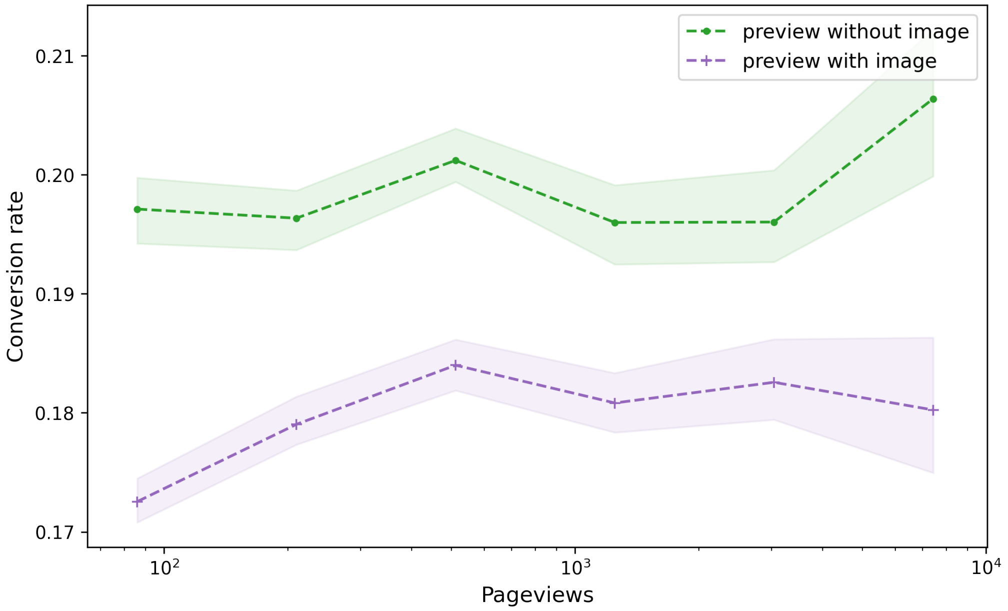


FIGURE 4.10: Comparison of the conversion rate for preview tooltip with an image (purple) and without image (green) as function of the page popularity (*pageviews*). Error bands represent bootstrapped 95% CIs.

long lists of aggregated pieces of content related to the same topic, e,g., achievements / publications (movies, books, articles) from notable people or shows. Highly clicked illustrated page previews are often also historical events, or elections, namely information-dense articles where the lead image is only partially useful to grasp the entire article content and its complexity. Conversely, illustrated pages with low conversion rate are articles talking about a specific place (e.g., "Old Fortress, Corfu"), or a specific person, object or spieces (e.g., "Microvelia Macgregori"), namely articles where an illustration can satisfy most of the information need.

Unillustrated page previews with high conversion rate are much more diverse, they go from individual objects or people, e.g. ("Fanny Sidney"), where more textual information is needed to understand the subject in absence an image, to lists and events. Unillustrated articles with lower conversion rate instead tend to be about subjects where a visual explanation is not necessarily needed in order to fully understand the information: for example, generic concept such as "Authority". "Minis-eries", or "Bachelor of Science", where images could actually be misleading or give a biased perception of the abstract piece of knowledge.

4.6 Conclusions

In this work, we provided a comprehensive overview over Wikipedia's visual world and how readers interact with it. We analyzed reader interactions with visual encyclopedic knowledge and found that images attract more attention than other interactive parts of the article: on average, click-through rate on images is 3.5%, while, for example, reference clicks happen only for 1 in 300 pageviews (Piccardi, Redi, et al., 2020b). Our insights can be summarized as follows:

- *Images serve a cognitive purpose.* We found a negative relation between article length and iCTR. This suggests that, similar to references (Piccardi, Redi, et al., 2020b), images might be used by readers to complement missing information in the article, fulfilling part of their *cognitive* function of providing knowledge complementary to the text (W Howard Levie et al., 1982a). Through a matched observational study, we also found that readers tend to click more often on unillustrated Wikipedia page previews to expand their content. On the contrary, conversion rate on illustrated page previews is consistently much lower across popularity buckets, thus suggesting that readers' need for contextual information is often fulfilled by the presence of an image on the preview popup. In this work, we also tested the relation betwen readers' interactions with images and article readability: our hypothesis was that images provide a *compensatory* function for articles that are difficult to read. However, we found evidence of the opposite trend: more readable articles tend to elicit higher engagement with images. While this is a preliminary result, further investigation is needed to understand how images support learning in low readability contexts.
- *We engage more with images illustrating the world and complex objects.* Our different layers of analysis consistently expose that Wikipedia readers are attracted by images about geographic locations, especially monuments and maps, and illustrations about biological sciences. Moreover, while we did not explicitly encode the notion of image *complexity* into our models, we found that Wikipedia readers tend to interact more often with images of complex objects, such as the ones in articles about visual arts, transportation, and military topics. A similar relation between the complexity of the image and its visual interestingness, i.e., the extent to which an image catches the viewer attention, has been widely explored and verified in experimental psychology and computer vision literature (Constantin, Redi, et al., 2019). While this relation can be influenced by different visual factors, such as the image size and its content, our results seem to support similar hypothesis, and provide a starting point for further investigation on the relation between image complexity and reader engagement.
- *Faces engage us, but only if unfamiliar.* Consistently, research works from different fields suggest that people and web users engage more with faces (Morton et al., 1991) and face pictures (Bakhshi et al., 2014), especially celebrities (Tsikrika et al., 2014b), than with other objects or subject, both in online platforms and in the real world. In this work, we found an opposite trend: for Wikipedia readers, images with faces seem to be much less engaging than, for example, more "encyclopedic" images about monuments or transportation. However,

we also found that readers do interact with face images when they are placed in unpopular articles, i.e. when those faces represent less well-known people or are *unfamiliar*. This positive relation between unfamiliarity and engagement again confirms findings from previous research linking the interestingness of a visual object with its familiarity to the observer (Constantin, Redi, et al., 2019).

Chapter 5

The role of images in navigating Wikipedia

5.1 Scope

In the digital age, Wikipedia has become the largest online source of free encyclopedic knowledge, serving billions of readers worldwide (Lemmerich et al., 2019b). Understanding how readers navigate and consume information on the platform is of crucial importance for enhancing user experience and meeting their information needs. Extensive research has been conducted to delve into the information-seeking behavior of Wikipedia readers, e.g., exploring navigation patterns (Piccardi, Gerlach, Arora, et al., 2023a), temporal regularities (Piccardi, Gerlach, and West, 2023b), curiosity needs (Lydon-Staley et al., 2021), and topical preferences (Dimitrov et al., 2018b). According to information foraging theories, users—when searching for information—actively look for patches that emit the strongest scent similar to animals in search of food. The strength of the scent is influenced by textual and visual cues from the environment, which translates to textual and visual content in the information environment analogy. Along this line, researchers have explored readers' interactions with various aspects of Wikipedia's content, ranging from references (Piccardi, Redi, et al., 2020b) to external links (Piccardi, Redi, et al., 2021).

While primarily text-based, Wikipedia also includes vast amounts of visual content. The key role of images as information scent and visual cues has been highlighted by computational and social sciences. Computational models have integrated the presence of images to better describe how people seek information online (Loumakis et al., 2011b; Oostendorp et al., 2012b). Meanwhile, experimental psychologists have underscored the multifaceted cognitive functions served by illustrations (W Howard Levie et al., 1982a), emphasizing the role of images as potent cues in online instructional contexts (Khamparia et al., 2018; Rudolph, 2017; Tempelman-Kluit, 2006). Despite such evidence, the role of images in information consumption and navigation on Wikipedia remains a relatively understudied area, with only a few works exploring visual encyclopedic knowledge from the perspective of the editors' community (Capra et al., 2013; He et al., 2018a; Beytıa et al., 2022).

This Chapter investigates the role of images in navigation and knowledge consumption on the English Wikipedia. Specifically, to understand whether images enchance the surrounding contents' information scent, we ask the following research questions:

RQ1 Are readers more likely to engage with knowledge in the presence of images?

RQ2 Do images support navigation on Wikipedia?

To address RQ1 (Sec. 5.2), we design an observational study on a large-scale dataset of reader interactions with sections from English Wikipedia. Through a matched cross-sectional study design, we compare engagement between pairs of illustrated and non-illustrated sections to estimate the causal effect of the presence of images. Our findings indicate that sections accompanied by images exhibit higher overall engagement, with links in illustrated sections being 8% more likely to be clicked compared to their unillustrated counterparts. We also observe that images tend to elicit increased interactions with geographical content, and content located further down the page, particularly in short articles. Then, through a longitudinal study design, we measure the impact of adding images to previously unillustrated sections. We observe that the effect of adding an image contributes to increased engagement if the addition is recent, but the effect vanishes with time. We dig deeper into these findings by addressing RQ2 (Sec. 5.3). We analyze navigation traces obtained from an online crowdsourced experiment where people play a game of finding the path from two given articles following links. The experiment has two setups: one in which all the articles are illustrated, and the other where all the images are removed. We observe that information-seeking is more efficient in the presence of images. Participants who completed tasks with illustrations take 19% less time to find the solution, with shorter paths.

We conclude that the impact of text illustrations on the way readers navigate and consume information on Wikipedia is significant. From a broader perspective, our work contributes to a growing body of research on multimedia learning and information consumption. In addition to advancing the theoretical understanding, our study has practical implications for both educators and platform designers seeking to enhance the user experience beyond Wikipedia, as it sheds light on the potential benefits of incorporating visual content into educational materials and online platforms.

5.2 RQ1: Do readers engage more with illustrated knowledge?

Wikipedia articles serve as a repository of encyclopedic information. Articles are typically divided into smaller sections for the purpose of organizing their content and supporting readers to comprehend and retain information. Sections convey information mainly in the textual form, but the use of visual aids, such as images and graphs, can often enhance the reader's understanding of the subject matter. To answer our first research question, we collect a large-scale dataset of reader interactions with sections, and perform a cross-sectional and longitudinal observational study to characterize its dynamic.

5.2.1 Materials and Methods

Section data

We download the *wikitext*[1] of the public Wikipedia snapshots of the English version released at the end of March 2021 and November 2022. The wikitext, also called *wiki markup* or *wikicode*, specifies the syntax and the keywords adopted by the MediaWiki software to format an article. It is released every month on a public online repository[2] for each Wikimedia project. From the wikitext, we extract first-level sections for each article. Each section is then characterized by a set of features such as the section length (in characters), the section offset (i.e., its relative position in characters from the beginning of the article), the number of links, and the presence of images. Also, we label each article with a topic extracted from the ORES topic classifier[3], and we assign to each section the topic of its parent article. Note that, within an article, the same link could be repeated in multiple sections. In these cases, from the server logs it is impossible to detect which one was actually clicked. As our analysis is based on link clicks, we discard all the sections that include such links to avoid overestimating user activity. In total, we removed 30% of sections. In the end, our dataset consists of 16.7 million sections, 20% of which are illustrated.

Pageloads

To quantify engagement with sections, we base our analysis on the Wikipedia server logs[4] for the English language edition. This data contains a log of the user activities on any Wikimedia project, retained for analytic purposes and deleted after 90 days for privacy reasons. We filter out the requests coming from bots and we collect the pageloads for 6.2 million articles in the main namespace (i.e., whose MediaWiki namespace is 0) over two periods of four consecutive weeks: between March 1 and 28, 2021 for the cross-sectional analysis, and between October 31 and November 27, 2022, for the longitudinal analysis. To preserve the privacy of the users, we take the following precautions: we discard pageloads from users that were logged in or made any edit in the period of the data collection; we removed pageloads from countries with at least one day of fewer than 300 pageloads; we assign a random identifier to each user based on its user-agent and IP address (Paranjape et al., 2016)

[1] https://en.wikipedia.org/wiki/Help:Wikitext

[2] https://dumps.wikimedia.org/

[3] https://www.mediawiki.org/wiki/ORES

[4] https://wikitech.wikimedia.org/wiki/Analytics/Data_Lake/Traffic/Webrequest

and drop any geographical information. Moreover, we drop the pageloads referring to the *Main Page*, as they do not represent any article in particular. In summary, our dataset contains 6.5 billion pageloads from 1.5 billion user identifiers.

Section click-through rate

We quantify reader engagement with links at the section level using the click-through rate. Our choice is driven by the vast literature on user engagement with web content where click-through rate is widely adopted to measure image relevance, user search satisfaction, engagement with illustrated ads (Lehmann et al., 2014b; Richardson et al., 2007; Edizel et al., 2017), and reader interactions with content on Wikipedia (Piccardi, Redi, et al., 2020b; Rama et al., 2021). In this work, we define the *section click-through rate (CTR)*, for a given section, as the ratio between clicks and views. Clicks are counted as the number of reading sessions where at least one link was clicked, whereas views are counted as the number of reading sessions upon which the section was displayed. We use article pageloads as a proxy to count section views. Moreover, it frequently happens that the same user views the same article multiple times in the same session (e.g., because reloading the page or clicking the back button). To avoid overcounting such multiple views, we consider each reading session visiting the same page once. Formally, the section CTR is defined as:

$$CTR = \frac{\sum_i C_{i,s}}{\sum_i V_{i,s}} \tag{5.1}$$

where $\sum_i C_{i,s}$ is the number of unique reading sessions i that clicked on at least one link in section s, and $\sum_i V_{i,s}$ the number of unique reading session upon which section s was displayed. Intuitively, the section CTR can be interpreted as the probability of observing a click on any link in the section, during a reading session in which that section is viewed.

5.2.2 The effect of images

As a first step, we aim to study the effect of text illustrations on readers' engagement. To perform our analysis, we conduct a cross-sectional study on the section data collected in March 2021. We divide the sections into two groups: the sections that contain at least one image (the *treatment* group), and the sections without any images (the *control* group). Our aim is to compare engagement, as measured using the section CTR, between these two groups.

Propensity score matching The first approach would be to simply compare the average section CTR between the two groups. However, such a comparison could potentially be impacted by confounders and selection bias. For example, it may be the case that images are more likely to be included in sections of highly popular articles, or in the first sections rather than at the bottom of a page. To reduce the effect of these confounders, we adopt a *matched* pairs design. In this setup, the goal is to find pairs of sections–one that is illustrated, and one that is not–that share similar characteristics. Our aim is to balance potential confounding variables within pairs so that we are able to observe if the presence of an image is associated with a change in engagement. To find such pairs, we resort to propensity score matching (Rosenbaum et al., 1983). Propensity score matching is a quasi-experimental technique that attempts to estimate the effect of a treatment by accounting for the covariates that predict receiving the treatment. In practice, we simulate a randomized treatment experiment by assigning the propensity of receiving the treatment to each section. Matching pairs of treated and control sections with similar propensity scores results in balanced covariates distributions between groups that only differ for the assignment of the treatment and gives us an indication of the causal impact of the presence of images on the CTR.

In our case, we define the presence of images as the treatment and we estimate the propensity scores by fitting a logistic regression on a set of covariates that capture the section characteristics. These covariates concern the structure of the section (the section length and the number of links in the section) and the context of the section (the length, the popularity, the topic of the article, and the section offset). The model accurately predicts the probability of receiving the treatment (area under the ROC curve: 0.78). We examine the model coefficients in Fig. 5.1. We observe that the section offset (i.e., the position of the section in the article) is the largest coefficient in absolute value and is negative. This means that images tend to be added in the first sections of an article. Sections in the two groups are then matched ensuring that two potential matches have similar propensity scores within a caliper of $std(propensity\_score) \cdot 0.2$ (Austin, 2011). Moreover, to improve section similarity, we require an exact match in two cases: we match sections if they are the lead sections (i.e., the first section of an article, often including the infobox), and sections of the same topic[5]. Finally, we extract matched pairs from the candidates that have the closest propensity scores by performing k-nearest neighbors matching.

Our approach yields 310,000 pairs of matched sections, with indistinguishable characteristics between the two groups. We measure the standardized mean difference (SMD) to ensure the balance in the covariate distributions: we observe the matching procedure greatly reduces the variance between groups from an average of 0.50 to 0.073, with all the SMD below 0.2 for each variable (Austin, 2009).

Results At this point, we are able to estimate the causal effect of the presence of images on engagement. We find that the section CTR is 1.70% and 1.58% for sections with and without images, respectively. The difference is statistically significant according to a Wilcoxon signed-rank test with $p << 0.001$. This means that even

[5] We only consider the main level of the topical structure. In total, at this level, there are four topic categories: *Culture*, *STEM*, *History and society*, and *Geography*

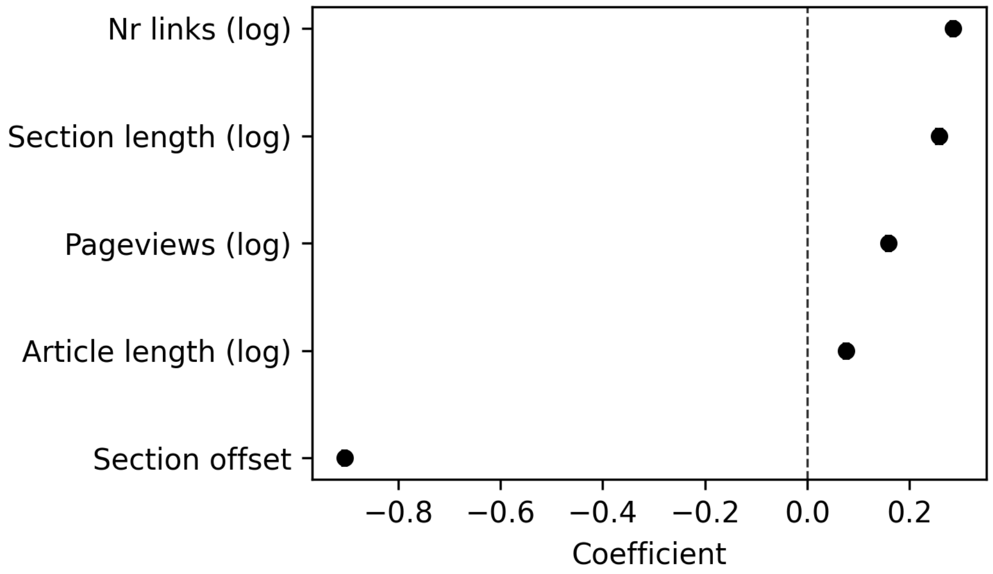


FIGURE 5.1: Effect of the section's covariates on the probability of having an image. The effects are estimated with logistic regression (area under the ROC curve: 0.78). 95% confidence intervals are too narrow to be appreciated.

accounting for possible confounders between the two groups, sections that are illustrated tend to elicit a higher level of interactions with links with respect to their counterpart. Quantitatively, the difference between the two is 0.12%, meaning that, on average, a random reader will be 8% more likely to engage with a section, by clicking on its links, in the presence of an image.

5.2.3 The effect of topics

We further deepen our investigation by performing a topical analysis. We repeat the previous procedure, this time stratifying by topic. We use the second level of the topical taxonomy as a reference. As a result, each set of matched pairs consists of sections belonging to the same topic. Fig. 5.2 shows the median section CTR for the two groups. The topics are ordered in decreasing order with respect to the difference in section CTR. We observe that *Geography.Geographical* is the topic with the largest median section CTR difference, meaning that it is much more likely to click links on illustrated sections rather than their counterpart. This is in accordance with previous findings (Rama et al., 2021) that show how geographical articles are among the ones where images are more likely to be interacted with. We note that *Culture.Food and drink* and *Culture.Sports* also have a large positive section CTR difference despite the corresponding images being less likely to be clicked (Rama et al., 2021). We explain this behavior by observing how the presence of images could drive interest in the text, without eliciting user clicks on them. For *STEM.Mathematics*, *STEM.Physics*, and *Culture.Internet culture* the median section CTR difference is not statistically significant according to a Mann-Whitney U test with $p < 0.01$. This could be due to

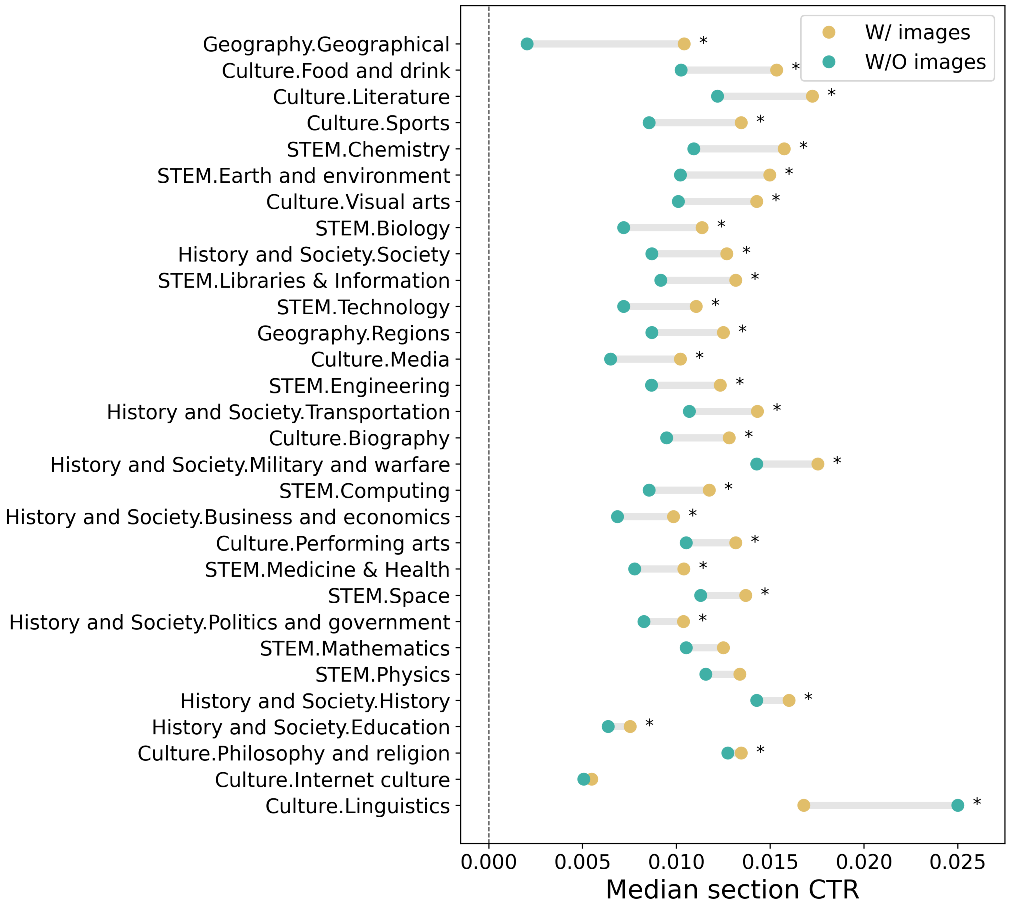


FIGURE 5.2: Comparison of median section click-through rate between matched pairs of sections with (yellow) and without (green) images, stratified by topic. Significant differences (Mann-Whitney U test with $p < 0.01$) are marked with an asterisk (*).

both the limited presence of images on these topics, and the readers' informational needs driven by textual rather than visual cues. The only topic showing a negative difference (i.e., links are more likely to be clicked if a section does not have images) is *Culture.Linguistics*. It is worth noting how linguistics shows the smallest presence of images, and, upon manual inspection, we observe how the vast majority consists of alphabetic symbols and icons.

5.2.4 The effect of the section offset

Previous work has shown that engagement is lower for longer articles (Piccardi, Redi, et al., 2020b), and it tends to decrease with the length within the same page (Rama et al., 2021). Moreover, in Fig. 5.1, we observe that images are less likely to be added at the bottom of a page. Since images tend to elicit higher levels of engagement on average, we ask: at the same offset, is engagement higher for illustrated or non-illustrated sections? To estimate the effect of images in sections at different positions within a page, we rely on the matched dataset from the previous analysis. To make a fair comparison, we stratify the analysis by article length. We divide articles into

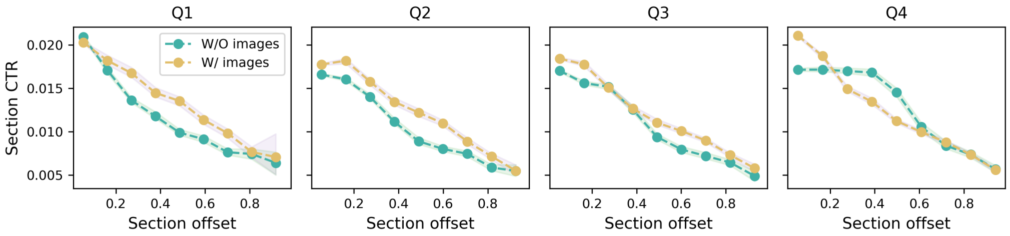


FIGURE 5.3: Relation between the section click-through rate and the section offset, stratified by quartiles of article length (with bootstrapped 95% confidence intervals).

quartiles of article length, and group each section accordingly. Fig. 5.3 shows the section CTR as a function of the section offset, colored by the presence of an image. Each subplot corresponds to a quartile of the article length distribution. The lowest value of the section offset range (i.e., 0) corresponds to the lead section, while the highest value (i.e., 1) to sections at the bottom of an article. First, we observe that the section CTR decreases with the offset, which confirms the previous findings: engagement tends to decrease as we progress towards the end of an article. Second, we note that illustrated sections are characterized by higher CTR compared to their counterpart, across strata of the article length, except for highly popular articles. This has two effects: on one hand, given two illustrated and non-illustrated sections at the same position in two different articles, links in the section with images will elicit more engagement on average. On the other hand, within the same article, images help sections to receive attention further down the page. Third, while at the bottom of an article engagement is similar, at the top (i.e., on the lead section and the infobox) the section CTR is higher for illustrated sections, except for short articles.

5.2.5 Does engagement with sections change when adding images?

The findings above show that images tend to elicit a higher level of interaction with the surrounding content. Next, we ask whether images are able to increase engagement over time. What happens to the engagement of the sections that become illustrated? Will links be more likely to be clicked?

Propensity score matching. To investigate this aspect, we conduct a longitudinal study comparing sections data between March 2021 and December 2022. The time span between the two time periods is chosen to allow a comparison between consecutive time snapshots since we are interested in the causal effect of the addition of an image to a previously unillustrated section. We divide the sections into two groups: the *treatment* group made of sections that have no images in the first period (*pre-treatment*) and that have at least one image in the second (*post-treatment*), and the *control* group made of sections that remained without images. We aim at comparing the section CTR between the two groups and understanding whether and how much adding an image impacts engagement. The first approach to answering this question would be to compute the difference in section CTR between the post and the pre-treatment periods and compare the average differences between the two groups. The problems with this setup are two. First, each section and its context may have changed significantly between the two periods: editing activities could have modified its length, the length of its article, and therefore its position relative to the length of the article (i.e., its offset). Also, the article's popularity might have changed due to seasonality and/or specific events localized in time. We aim to reduce these biases as follows: we assign each section to a percentile concerning its page length, popularity, section length, and section offset distributions for the pre and post-treatment periods. We discard sections whose respective percentiles have changed between the two periods. Moreover, even after accounting for the temporal variations of the section characteristics, the comparison between the control and the treatment groups could be subject to other confoundings and selection bias. For example, it might be the case that images are more likely to be added to sections that are already more popular, or to sections placed at the top of an article. As previously, we turn to propensity score matching to disentangle these effects. We aim to compare pairs of sections from the two groups with similar pre-treatment characteristics and be able to observe how the addition of an image is associated with a consequent change in engagement over time. To model the propensity score, we use the same covariates as in the previous analysis, and train a random forest (area under the ROC curve: 0.87). We extract matched pairs from the candidates, repeating the same procedure as before, but performing maximum weight matching on the weighted bipartite graph, where nodes are sections, and the weights are similarity scores based on the Mahalanobis distance in covariates. Our approach yields 45,500 pairs of matched sections. The SMD changes from 0.31 before matching to 0.02 after, with all the SMD below 0.2 for each variable.

Difference-in-differences analysis. To estimate the causal effect of adding images to sections, we conduct a difference-in-differences (DiD) analysis (Lechner et al., 2011). The idea is to compute the differences in the outcome variable (i.e., the section CTR) for each section of the treatment group between the pre and post-treatment periods. To eliminate the effect of time-varying factors, DiD also computes the same difference for the control group. Then, it computes the difference in treatment effect between each section in each matched pair. Finally, it calculates the overall effect of the treatment by averaging these differences across all pairs of matched sections.

In practice, the DiD estimate is computed using a linear regression model in the following form:

$$y_{i,t} = \alpha + \beta \cdot treatment_i + \gamma \cdot period_t + \delta \cdot treatment_i \cdot period_t + \epsilon_{i,t} \tag{5.2}$$

where $y_{i,t}$ is the CTR of section i in the period t. t is an indicator variables that takes values of 0 and 1 when capturing the pre and post-treatment period, respectively. The independent variable $treatment_i$ indicates whether section i has a new image or not (1 if it has been added, 0 otherwise), and $period_t$ indicates whether section i is in the pre or post-treatment period. The coefficient δ of the interaction term is the DiD estimation of the treatment effect, i.e., the effect of adding an image to a previously unillustrated section to its CTR.

We fit the linear regression to the set of matched sections to compute the DiD estimates. We do not observe a positive effect on the CTR due to the addition of an image, with $\delta = 0.0002$ (97.5% CI $[-0.001, 0.002]$). A reason for this behavior could be due to how we measure the outcome variable. In our analysis, we do not keep track of how engagement varies over time. However, in the treatment group, images could have been added throughout the full period, while we only consider the condition of the section at the beginning and at the end of the period. In this setting, changes in the outcome variable between sections where images have been added at different times could be hard to compare. For example, for sections where the image was added more recently, we could not witness a change in CTR yet; on the other hand, for sections where the image was added two years ago, the effect could already have vanished.

Stratifying by the time of image addition. To overcome this limitation, we measure the difference in engagement pre and post-treatment stratifying by the time when the image was added. We extract the section data bimonthly from March 2021 until November 2022, for all the matched section pairs found in the previous analysis. For each time period, we extract the sections where images have been added and we perform the DiD estimation on those. Fig. 5.4 displays the DiD estimate as a function of addition time. Results show evidence of a time-dependent relationship: we see that the coefficient is around zero for sections that became illustrated earlier, while is positive for sections recently illustrated. We suppose that newly added images could represent a novelty, thus bringing an immediate attention shift towards the content of the section (Schomaker et al., 2015). Also, this effect seems to decrease over time.

5.3 RQ2: Do images support navigation on Wikipedia?

Previously, we observed that illustrated sections tend to receive more clicks than non-illustrated ones. While we provided quantifiable evidence about this phenomenon, the reason *why* we observe this is still unclear: on one hand, images could drive attention by generating curiosity towards the sections they illustrate; on the other hand, they could help readers find the information they need among all the concepts of an article. Since it is impossible to test these hypotheses using just observational data, we resolve to an online crowdsourcing experiment to shed light on the role of images in information wayfinding.

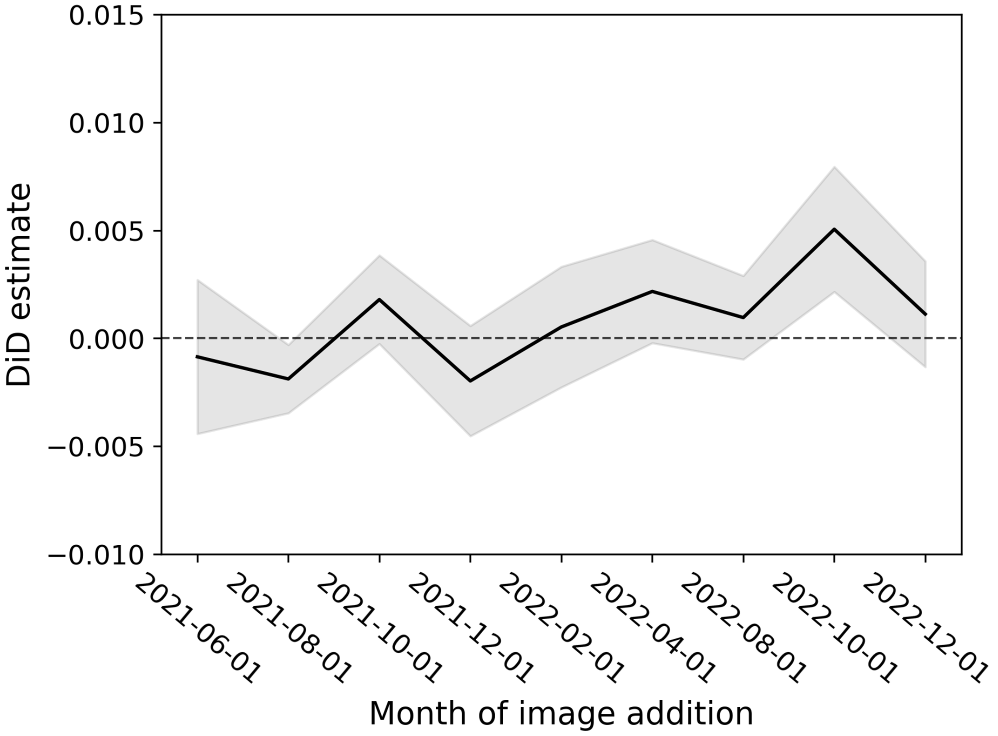


FIGURE 5.4: Difference-in-difference treatment effect between matched pairs of sections, stratified by month of image addition (with 95% confidence intervals).

Condition	No. tasks	Mean task duration	Mean effective path length	Mean complete path length	Shortest path length
Illustrated	105	117 [110; 125]	5.1 [4.9; 5.4]	5.9 [5.6; 6.2]	2.7 [2.7; 2.8]
W/O images	105	145 [135; 158]	6.3 [5.9; 7.3]	7.4 [6.9; 8.3]	2.8 [2.7; 2.8]
Images shuffled	102	150 [140; 163]	5.1 [4.9; 5.3]	6.0 [5.8; 6.4]	2.7 [2.5; 2.7]
White spaces	102	143 [134; 152]	5.9 [5.6; 6.2]	6.5 [6.2; 6.8]	2.6 [2.6; 2.7]

TABLE 5.1: Number of tasks, mean task duration, mean effective/complete path length, and mean shortest path length across the four conditions in which the tasks are presented to participants (with 95% bootstrapped confidence intervals).

5.3.1 Materials and Methods

Wikispeedia

We design an online experiment in which we simulate readers' behavior when trying to find information in the Wikipedia articles' network. As a tool, we use Wikispeedia (West and Leskovec, 2012b), an online human-computation game specifically designed with the purpose of collecting data about how humans navigate online information networks and what strategies they use. During each task, people are given two random Wikipedia articles and asked to solve the task of navigating from one to the other by clicking hyperlinks in the shortest time possible. While playing the game, people have no knowledge of the articles' network structure, and need to rely solely on the information they find on each page—consisting of text, images, and links—and their previous knowledge about how the concepts they navigate could be connected to each other.

We base our tool on a Wikipedia version designed for schools in 2007[6] with around 4,000 articles and 120,000 links. We decide to keep the Wikipedia graph the same as in the original Wikispeedia game for the following reasons: to have a baseline to compare our pilot studies against; articles are written in English, which makes readership comparable to that of the sections above and crowdsourced data available from a larger audience; the network and topic structure is similar to that of the English version, with a skewed degree distribution and few high-degree hubs connecting a vast portion of the network.

Data

We advertise our tool on Twitter and academic-related mailing lists, and we collect source-target search paths from voluntary participants. Each data point consists of a pair of source-target articles. Tasks are randomly selected and stratified by task difficulty. We consider the optimal solution (i.e., the shortest path length (SPL) between the two nodes) as a proxy for the task difficulty. We randomly selected 1,000 source-target pairs for SPL values from 2 to 4. To test our hypothesis, we decline the experiment in two main variants: one in which participants are exposed to the original *illustrated* articles[7], and the other where all the images have been removed. To eliminate the effect of possible confounders due to the heterogeneity of the article characteristics, we require each task (i.e., each source-target pair) to be completed in each of the two conditions by two distinct participants. We collect data for 105 source-target pairs for both conditions, resulting in a total of 210 tasks. Before analyzing the data, we identify suspiciously long or short paths, discarding paths whose duration and path length differ more than 4 standard deviations from their mean value.

Metrics

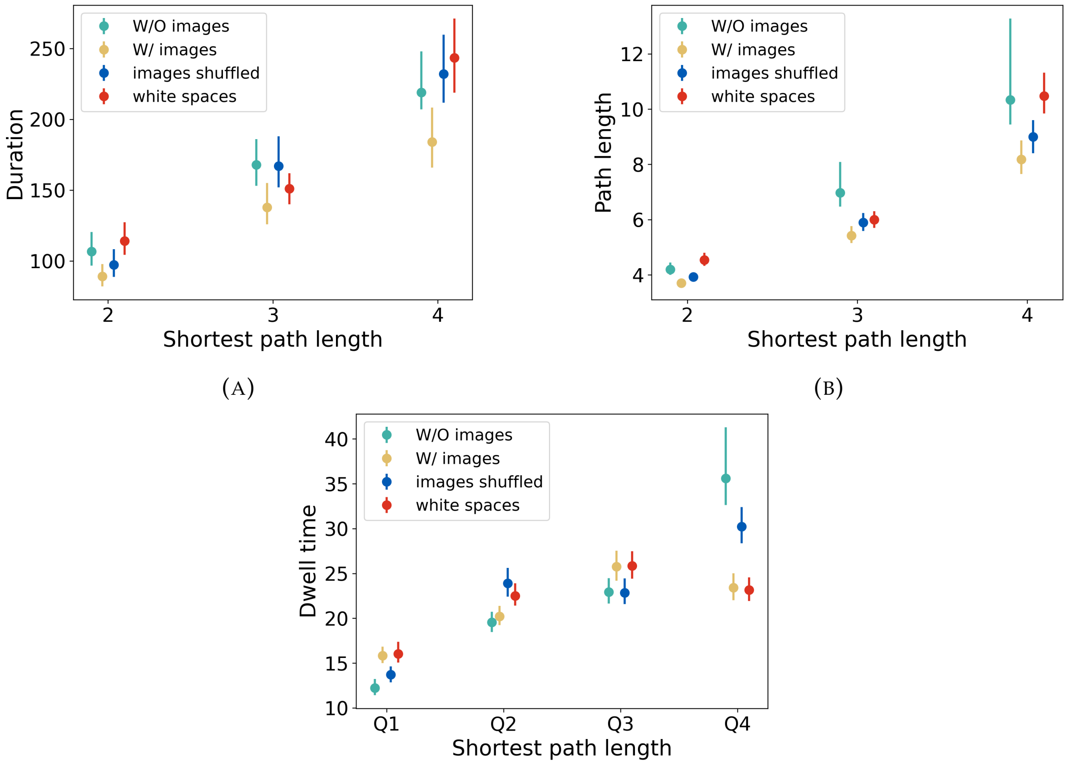
The key metrics in our analysis are the following.

Path length Previous research has shown that humans are efficient at navigating information networks (West and Leskovec, 2012b). To measure the efficacy in finding short paths, quantitatively, we computed for each task the *full* path length, as the number of links visited from source to target, and the *effective* path length, which excludes undone and back clicks from the complete path.

Dwell time We define the *dwell time* as the time (in seconds) the participant spends on each page before clicking on the next one. The dwell time is assumed to be related, beyond the mental semantic association task, also to the visual search and visiospatial attention abilities involved in our task. Indeed participants, in order to

[6] https://web.archive.org/web/20071006054112/http://schools-wikipedia.org/

[7] We make sure to retain only those articles that are illustrated from the networks, accounting for around 90% of the total.



(A)

(B)

(C)

FIGURE 5.5: Mean task duration (a) and effective path length (b), stratified by shortest path length (with 95% bootstrapped confidence intervals). Average time spent on articles (dwell time) (c), stratified by quartiles of article length. Error bands are 95% bootstrapped confidence intervals.

complete the task, have to visually explore the article to find possible relevant links that could be semantically associated with the final target.

Task duration Search paths can consist of long chains of articles visited spending a short time in each, short chains where articles have been read in-depth, or a combination of the two. We then measure the task *duration* to evaluate the trade-off between path length and dwell time.

5.3.2 Results

Task duration

We find that the median task duration is less for tasks containing images, i.e., 117 seconds against the 145 seconds when images are removed. The difference is statistically significant with $p < 0.1$ according to an independent samples T-test. This means that participants spent less time on average to complete the same tasks in the presence of illustrations. Moreover, Fig. 5.5a shows how duration significantly varies according to task difficulty. Non-illustrated tasks consistently show longer duration sessions.

Path length

We observe a statistically significant difference between the effective path length in the case of image presence, 5.1 steps, and image removal, 6.4 steps (meaning 20% reduction). We find the same effect for the complete path length, i.e., 5.9 steps versus 7.4 steps, respectively. In Fig. 5.5b the effect of the difficulty is shown, confirming that tasks with increasing difficulty consist of, as expected, longer chains exhibiting a wider difference between the illustrated and unillustrated cases.

Dwell time

Finally, we focus on the time spent on each article of the sequence. Since participants naturally tend to spend more time on longer articles, we stratify the analysis into quartiles of article length. Fig. 5.5c shows the mean dwell time across task conditions stratified by quartile of article length. We observe that there are significant differences between the two conditions when articles are short and long. On short pages (Q1), on average participants spend less time on articles without images. We hypothesize that when articles contain a small amount of text and they are, therefore, less demanding to parse, it is easier for the reader to internalize the content and find the next concept of the chain. In this case, images could even play the role of distractors. On the other hand, for large-sized articles, the average dwell time is consistently shorter in the presence of images confirming their positive role in enabling information finding.

Are images relevant to their context?

The previous analysis in Sec. 5.3.2 reveals that participants are more efficient in finding the information they seek when images are present. This leads us to the question: why do images strengthen information scent? One possible explanation comes from previous studies on online information foraging, suggesting that the relevance of an image to its surrounding context is crucial (Oostendorp et al., 2012b; Capra et al., 2013; Loumakis et al., 2011b). To delve deeper into this aspect, we investigate whether our results can be attributed to the effect of image relevancy. On Wikipedia, images are carefully added by editors to be "significant and relevant in the topic's context, not primarily decorative."[8] We thus compare the original version of an article with two additional alternative settings. In the first one, we modify the original image positions by shuffling them within the same article. In the second one, we replace all the images with white spaces while maintaining the original structure of the article. This enables us to evaluate the impact of both the effect of the image content and its placement within the article structure. We collect 204 tasks, evenly distributed across the two conditions while keeping the same source-target pairs used in the previous settings. Tab. 5.1 presents the results. We observe that the average task duration is 150 and 143 seconds for the shuffled images and white spaces

[8] https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Images

variants, respectively. In both cases, the average durations are significantly higher compared to the illustrated case with $p < 0.1$ according to an independent samples T-test. Furthermore, the durations are comparable to those of the tasks where images are completely removed. These findings suggest that, on one hand, users actively examine the content of the images during information search, and that illustrations enhance efficacy when relevant. On the other hand, even when the visual content is replaced with white spaces but the article structure remains unchanged, information wayfinding becomes less efficient. Overall, our results confirm prior research on the effectiveness of relevant images.

5.4 Conclusions

In this work, we provided insights regarding the role of text illustrations on knowledge navigation in the context of English Wikipedia. By combining large-scale observational analyses and a crowdsourced experimental design, we conclude that illustrations have a causal effect on the way readers navigate and consume information in the context of encyclopedic information. Our findings can be summarized as follows:

- **Readers engage more with textual content in the presence of images.** We compared engagement between sections that are illustrated to those that are not after removing potential confounders in a quasi-experimental design. We observed that readers are more likely to click links in sections that contain images. We also note that the same trend repeats across almost any topic, with geographical content being the one for which images make the largest difference. This suggests that images might partially fulfill the *attentional function* of directing the casual reader to specific portions of the text where the visuals are present (W Howard Levie et al., 1982a).

- **Textual content is more engaging around images, even further down the article.** Previous works have shown that engagement is lower for longer articles (Piccardi, Redi, et al., 2020b) and decreases with length within the same article (Rama et al., 2021). Our findings show that images help readers reach textual content further down the page, especially in short articles.

- **Engagement increases around newly added images, but only when additions are recent.** We estimated the causal effect of adding images in a longitudinal analysis. We found that, overall, the effect is negligible. However, when stratifying the analysis by month of image addition, we observed that images have a positive effect when the addition is recent.

- **The presence of images facilitates information wayfinding.** Finally, we investigated the influence of images simulating the process of human information wayfinding. We leveraged Wikispeedia, a human-computation game where participants are asked to link concepts in the Wikipedia article network. We found that, overall, participants are more efficient at finding information in the presence of images. They take less time to reach their target and with shorter path lengths. This is particularly evident when images are relevant to the surrounding text. As a matter of fact, when we modified the original position of the images on the page and shuffle them around, we noticed that participants mostly ignore the content of the images, and the effectiveness levels are similar to the navigation patterns of articles without images. This is in line with previous studies (Oostendorp et al., 2012b; Capra et al., 2013; Loumakis et al., 2011b), showing that paragraphs or search snippets with relevant pictures increase the (perceived) value of information scent for hyperlinks that are relevant to the information need or task, hence increasing the probability of selecting the “correct” hyperlink in a browsing or search session.

Chapter 6

Conclusions

This dissertation provides, for the first time, an overview of the impact of images on how users interact and navigate Web encyclopedic content. In the scope of this work, Wikipedia represents the ideal candidate, being the largest online free source of knowledge with billions of views every month. By analyzing large-scale datasets of article textual and visual content and user interaction logs collected from the English edition of the encyclopedia, this work takes a step in the direction of understanding the role and importance of images in shaping our experience on the Web.

This dissertation contributes to the existing literature with two major contributions. The first contribution explores the characteristics of the visual space and sheds light on the engagement patterns with the visual content on Wikipedia. The second contribution uncovers the role of images in the dynamics that guides people in navigating information on the encyclopedia. The following sections summarize the findings and discuss the implications and limitations of the present work.

6.1 Summary of findings

Quantifying interactions with images on Wikipedia Chapter 4 provides a comprehensive overview on the visual space of the English Wikipedia and how its readers engage with its visual content. By analizing a large-scale dataset of reader interactions with images, we discovered that images drive a significant amount of readers' attention when browsing Wikipedia. We quantified the click-through rate on images to be 3.5%, meaning that, on average, clicks on images happen at least once every 29 pageviews. This is higher compared to to other types of interactive content within articles, such as references (Piccardi, Redi, et al., 2020a), for which the click-through rate is ten times smaller, and external links (Piccardi, Redi, et al., 2021). Similar to previous research, we found that readers are attracted by images about geographical locations, especially monuments and maps, and illustrations about biological sciences. Moreover, readers tend to interact more often with images in articles about visual art, transportation, and military topics. While not explicitly factoring the notion of image complexity into our analysis, these results seem to support previous research (Constantin, Redi, et al., 2019) on the relation between the image complexity and its interestingness, thus providing a starting point for further investigation on this aspect in the context of encyclopedic knowledge. On the contrary, we noticed a peculiar behavior with respect to the presence of faces. While previous research has shown that people engage more with faces, especially of celebrities, than with other objects or subjects, both in online platforms and in the real world (Morton et al., 1991; Bakhshi et al., 2014; Tsikrika et al., 2014a), we found that, on Wikipedia, the presence of faces usually elicits less interactions than other visual characteristics. In addition, readers tend to interact with face images when placed in unpopular articles, e.g., of less well-known people or when unfamiliar. Finally, we found that readers tend to click more often on non-illustrated Wikipedia page previews to expand their content, suggesting that the the need for additional information is often fulfilled by the presence of an image on the preview tooltip. These findings are aligned with previous research in experimental psychology anc confirm that images serve strong cognitive purposes of providing knowledge complementary to the text (Mayer, 2020).

The role of images in navigating Wikipedia Chapter 5 contributes to the field of online knowledge consumption by exploring the role of images on the dynamics of navigation on the Engligh Wikipedia. To understand if readers engage more with illustrated parts of articles while browsing Wikipedia, we divided articles into sections and compared engagement between sections that are illustrated to those that are not. We performed matched cross-sectional study to reduce the effect of potential confounders and discovered that links in illustrated sections are 8% more likely to be clicked compared to those in the non-illustrated ones. As in Chapter 4, this result supports evidence of the attentional function of images, directing readers to specific portions of the text where visuals are present (W. Howard Levie et al., 1982b). We also observed that the level of interactions with links decreases with article length, similarly to previous research (Piccardi, Redi, et al., 2020a; Rama et al., 2022), but images help readers reach textual content further down the page, expecially in short articles. Moreover, we estimated the causal effect of adding an image to a previously non-illustrated section. Our longitudinal analysis reveals that engagement increases around newly added images, but only when the addition are recent. Finally, to simulate the process of information wayfinding, we analyzed the navigation traces collected from a human-computation game where participants were asked to link concepts in the Wikipedia article network. We found that, overall, people are more efficient at searching for information in the presence of images, in particular when images are relevant to the surrounding text. This aligns with previous studies of foraging theories (Capra et al., 2013; Loumakis et al., 2011a; Oostendorp et al., 2012a), in which images are described as powerful visual cues, able to increase value of information scent for hyperlinks relevant for the information need, hence increasing the probability of selecting the correct path towards the target of a navigation session.

6.2 Limitations and opportunities

We acknowledge some shortcomings and limitations in the presented research. First, one main limitation of this work is that it focuses on the English Wikipedia only. While it represents by far the largest version in terms of content and popularity, it still provides a limited view of the broader Wikipedia project, available in more than 300 languages and accessed all over the world. Other works have shown that both readers' characteristics and visual content can vary significantly across Wikipedias (Lemmerich et al., 2019b; Beytıa et al., 2022). Therefore, in future work, we aim to replicate this analysis on a heterogeneous set of languages. Moreover, Wikipedia readers come from different parts of the world and have different information needs (Lemmerich et al., 2019a). In addition, multimedia research has shown that different language communities (Pappas et al., 2016) and geographies (You et al., 2017) perceive and produce visual content in different ways. However, we did not take into accout into our analyses the characteristics of *readers*, such as gender, age, educational attainment, geographic location, internet connection availability, and native language. Preliminary findings—not included in this work—on a global scale show that the way in which readers interact with images on Wikipedia tend to differ across geographic locations, mainly due to broadband availability, access modality, and availability of content in the native language. Future research could extend the analyses presented in this dissertation to explore the behavior of diverse groups of readers with visual encyclopedic content and the impact of exogenous events on image viewership. In addition, while the use of large-scale logs gives us a broad overview of the user behavior on Wikipedia and offers important advantages over more traditional methods when interested in quantitative measurements of population-scale phenomena [salganikBitByBit], it still carries some limitations. Previous research shows that large-scale data are often prone to biases introduced by data collection problems, preprocessing and measurement errors [wagner,lazer2014], and algorithmic dynamics [wagner,lazer2021]. Moreover, passively collected logs do not reveal readers motivations and information needs during their navigation sessions. Future work should extend what previous research has already discovered about readers motivation on Wikipedia (Singer, Lemmerich, et al., 2017b; Lemmerich et al., 2019b) with surveys and user studies to learn why readers look at images and how they use visual content on Wikipedia.

From a methodological perspective, both parts of this thesis make estensive use of causal inference techniques to estimate the causal impact of illustrations. We employed quasi-experimental study designs to match treatment-control pairs of observational data, to reduce the effect of confounders. We cannot exclude the possibility that other potential counfounding factors may still be present in our analysis. For example, external events or editors' activity may lead to an increase in attention to certain articles or topics. Future analysis should take into consideration statistical techniques, such as sensitivity analysis [rosenbaum2005sensitivity], to mitigate the effect of unnoticed confounders. Furthermore, in Chapter 4, we utilized visual features that rely on the outputs of commonly used machine learning models, including ORES (Halfaker and Geiger, 2020), MTCNN (Wen et al., 2016), and the proposed Wikimedia Image Quality classifier. Although these models have proven to be effective for our task, their fairness and inclusivity have not been thoroughly examined. To enhance the current research, further work could incorporate models that are as unbiased as possible and can be readily applied to images and articles from diverse global sources.

6.3 Implications and broader perspectives

This dissertation takes a first step in the direction of understanding the role of images for free knowledge ecosystems. While specific to the context of Wikipedia, our findings speak to the importance of visual content when consuming and navigating knowledge in the broader Web, with several implications.

Implications for theoretical research Our work offers an unprecedented large-scale analysis of user behaviors with visual content on the largest free knowledge platform online. While not exempt from limitations, the use big data offers advantages and new perspectives over traditional data collection methods. We believe that the approaches used in this work can serve as an analytical framework for future research based on user-generated logs and can easily be adapted to other online contexts. A large proportion of readers use Wikipedia for intrinsic learning, for school, or work-related tasks (Lemmerich et al., 2019b). Wikipedia is used in instructional settings by students and teachers (Knight et al., 2012), often with successful outcomes (Wannemacher et al., 2010). Our findings show the feasibility of large scale studies to understand the role of images in instructional settings using a multimodal, computational approach. While explaining the cognitive role of images in learning is outside the scope of this work, we still draw inspiration and ground our analyses on the large body of literature pertaining to experimental psychology and multimedia learning. Our findings can foster exciting directions for future work in these fields to dig deeper in the role of text illustrations for education. To this end, experiments could be designed along the same lines of this research, analyzing data coming from, for example, online learning platforms or Massive Open Online Courses (MOOCs).

In addition, these findings have significant implications for future research in modeling user navigation. Currently, there is limited research examining the role of visual aids in online information seeking. This work can inspire future research to inspire future investigations to gain deeper insights into the factors that capture readers' attention while they follow specific trails, including visual cues. Enhancing our understanding of how humans traverse information networks will contribute to the development of more robust theories.

Furthermore, our study focused on readers' interest and usage of images using implicit, large-scale signals, namely the click-through rate and the conversion rate. We selected these metrics based on the extensive literature on user interactions with web content. Click-through rates and conversion rates are commonly employed to measure image relevance, search satisfaction, user interest in illustrated ads, and reader engagement with content on platforms like Wikipedia. While these metrics provide big picture of readers' behavior, a more detailed examination of user interactions could offer complementary insights. Future research should explore a broader range of metrics, such as hovers, dwell-time, and eye tracking movements, to expand the analysis conducted in this study. This comprehensive approach will enrich our understanding of how people consume and navigate information, thereby informing the development of theories and models in this field.

Improving Wikipedia and webpage content From our analyses, we found strong evidence that images receive significant level of interactions and elicit interest in the sourrounding text, especially enhancing the information scent of the surrounding links. Highlighting links that are relevant to readers' information needs is especially crucial when we think of Wikipedia as a platform for learning. For instance, readers who use Wikipedia for learning assignments tend to spend more time on the encyclopedia, with longer, topic-specific sessions(Singer, Lemmerich, et al., 2017b). From this perspective, it becomes essential to enhance the visibility of article contents that facilitate readers' navigation and aid them in achieving their objectives. Editors who are committed to enhancing the educational function of the encyclopedia can play a crucial role by giving priority to the addition of informative images in close proximity to links that follow a prerequisite sequence for a learning objective. To support these efforts, researchers can contribute by developing models that provide guidance on optimal image placement to maximize engagement and learning on Wikipedia. Insights from our study can be leveraged by future recommender systems to prioritize the inclusion of missing content for editors. These models can incorporate signals of readers' interest in visual content. For instance, tools designed to automatically predict reader engagement with images can be integrated into services and models designed to identify and prioritize suitable images for Wikipedia articles. Considering the limited information available to editors regarding how readers interact and learn from Wikipedia content, gaining visibility into the potential usefulness of an image in an article would greatly improve editor workflows.

Additionally, these findings can easily extend beyond Wikipedia, especially to instructional settings and e-learning settings. Researchers and designers could benefit from the methodologies outlined in this study to test not only the role of images but also the impact of different content arrangement setups, on navigational behavior and information foraging on other learning platforms or websites. Moreover, as conversational AI interfaces are shaping new ways in which users interact with information on the Web, we envision future work focusing on understanding the interplay between visual content and these evolving machine-mediated paradigms to navigate online information spaces.

Impact on the Wikimedia ecosystem and beyond Our findings indicate that areas of Wikipedia content are more *visible* and engaging when images are present. Improving the visibility of Wikipedia content is especially relevant to the existing efforts around addressing knowledge gaps(Redi et al., 2020), or existing inequalities in Wikimedia projects. One of the most evident knowledge gaps in the encyclopedia is the gender gap, with around 80% of biographies being around men, across different language editions(Beytıa et al., 2022). Efforts to make content about underrepresented groups more "visible" exist across the Wikimedia movement, with initiatives supporting the creation and improvement of new and existing articles about women (e.g., Wikiproject Women in Red[1]), specifically in STEM professions (Vitulli, 2017), or the curation of LGBT+ content (Ribé et al., 2021). A few campaigns such as "Visible Wiki Women"[2], or "Wiki Unseen"[3] also encourage the inclusion of images of people from under-represented communities into Wikipedia articles. Our results show that when curating content about marginalized groups, the placement of images in the article should also be taken into account, as it can be key to improving the visibility of such knowledge. More in general, our work further underlines the importance of the many initiatives across the Wikimedia movement aimed at improving the image coverage on the Wikimedia projects, such as Wikipedia Pages Wanting Photos (WPWP)[4] and Wiki Loves Monuments[5]. While these initiatives focus on improving the visual coverage of the encyclopedia, adding high-quality images to Wikipedia is a gift to the whole Web ecosystem. Wikipedia content is widely re-used across different parts of the web, with its images being extremely visible and highly ranked at the top of image and text search results, and often used to train large computer vision models. Fostering and motivating these initiatives is therefore crucial not only for the encyclopedia but for the whole of human knowledge.

[1] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red

[2] Whose Knowledge? https://whoseknowledge.org/

[3] https://wikimediafoundation.org/participate/unseen/

[4] Wikipedia Pages Wanting Photos https://meta.wikimedia.org/wiki/Wikipedia_Pages_Wanting_Photos_2022

[5] https://www.wikilovesmonuments.org/

Bibliography

Abadie, Alberto and Guido W Imbens (2006). “Large sample properties of matching estimators for average treatment effects”. In: *econometrica* 74.1, pp. 235–267.

Almarabeh, Hilal, Ehab Amer, and Amjad Sulieman (Dec. 22, 2015). “The Effectiveness of Multimedia Learning Tools in Education”. In: *International Journal of Advanced Research in Computer Science and Software Engineering* 5, p. 761.

Alobaid, Azzam (Sept. 15, 2020). “Smart multimedia learning of ICT: role and impact on language learners’ writing fluency— YouTube online English learning resources as an example”. In: *Smart Learning Environments* 7.1, p. 24. ISSN: 2196-7091. DOI: 10.1186/s40561-020-00134-7. URL: https://doi.org/10.1186/s40561-020-00134-7 (visited on 06/21/2023).

Alpizar, David, Olusola O. Adesope, and Rachel M. Wong (Oct. 1, 2020). “A meta-analysis of signaling principle in multimedia learning environments”. In: *Educational Technology Research and Development* 68.5, pp. 2095–2119. ISSN: 1556-6501. DOI: 10.1007/s11423-020-09748-7. URL: https://doi.org/10.1007/s11423-020-09748-7 (visited on 06/21/2023).

Amengual, Xesca, Anna Bosch, and Josep Lluís de la Rosa (2015). “Review of Methods to Predict Social Image Interestingness and Memorability”. In: *Computer Analysis of Images and Patterns*. Ed. by George Azzopardi and Nicolai Petkov. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 64–76. ISBN: 978-3-319-23192-1. DOI: 10.1007/978-3-319-23192-1_6.

Anderson, Ashton et al. (2014). “The dynamics of repeat consumption”. In: *Proceedings of the 23rd international conference on World wide web*, pp. 419–430.

Arora, Akhil et al. (2022). “Wikipedia reader navigation: When synthetic data is enough”. In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 16–26.

Atkinson, R. C. and R. M. Shiffrin (Jan. 1, 1968). “Human Memory: A Proposed System and its Control Processes11This research was supported by the National Aeronautics and Space Administration, Grant No. NGR-05-020-036. The authors are indebted to W. K. Estes and G. H. Bower who provided many valuable suggestions and comments at various stages of the work. Special credit is due J. W. Brelsford who was instrumental in carrying out the research discussed in Section IV and whose overall contributions are too numerous to report in detail. We should also like to thank those co-workers who carried out a number of the experiments discussed in the latter half of the paper; rather than list them here, each will be acknowledged at the appropriate place.” In: *Psychology of Learning and Motivation*.

Ed. by Kenneth W. Spence and Janet Taylor Spence. Vol. 2. Academic Press, pp. 89–195. DOI: 10.1016/S0079-7421(08)60422-3. URL: https://www.sciencedirect.com/science/article/pii/S0079742108604223 (visited on 06/20/2023).

Austin, Peter C (2009). "Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples". In: *Statistics in medicine* 28.25, pp. 3083–3107.

– (2011). "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies". In: *Pharmaceutical statistics* 10.2, pp. 150–161.

Awad, Mamoun A and Latifur R Khan (2007). "Web navigation prediction using multiple evidence combination and domain knowledge". In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37.6, pp. 1054–1062.

Baddeley, A. (Jan. 31, 1992). "Working memory". In: *Science (New York, N.Y.)* 255.5044, pp. 556–559. ISSN: 0036-8075. DOI: 10.1126/science.1736359.

Bakhshi, Saeideh, David A Shamma, and Eric Gilbert (2014). "Faces engage us: Photos with faces attract more likes and comments on instagram". In: *Proc. Conference on human factors in computing systems (SIGCHI)*.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. Conflict, arousal, and curiosity. Pages: xii, 350. New York, NY, US: McGraw-Hill Book Company. xii, 350. DOI: 10.1037/11164-000.

Bernard, Robert M (1990). "Using extended captions to improve learning from instructional illustrations". In: *British Journal of Educational Technology* 21.3, pp. 215–225.

Berson, Eloise, Ngoc Q.K. Duong, and Claire-Helene Demarty (Sept. 2019). "Collecting, Analyzing and Predicting Socially-Driven Image Interestingness". In: *2019 27th European Signal Processing Conference (EUSIPCO)*. 2019 27th European Signal Processing Conference (EUSIPCO). A Coruna, Spain: IEEE, pp. 1–5. ISBN: 978-90-827970-3-9. DOI: 10.23919/EUSIPCO.2019.8902803. URL: https://ieeexplore.ieee.org/document/8902803/ (visited on 06/21/2023).

Bertacchini, Enrico, Enrico Ferraris, and Alice Fontana (Dec. 2022). "La fruizione delle collezioni digitali di beni archeologici: un'esplorazione delle immagini su Wikimedia Commons". In: *DigitCult* 7.2, pp. 99–116. ISSN: 2531-5994. DOI: 10.36158/97888929562237. URL: https://doi.org/10.36158/97888929562237 (visited on 06/15/2023).

Beytía, Pablo et al. (May 31, 2022). "Visual Gender Biases in Wikipedia: A Systematic Evaluation across the Ten Most Spoken Languages". In: *Proceedings of the International AAAI Conference on Web and Social Media* 16, pp. 43–54. ISSN: 2334-0770. DOI: 10.1609/icwsm.v16i1.19271. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/19271 (visited on 06/15/2023).

Beytıa, Pablo et al. (2022). "Visual gender biases in wikipedia: A systematic evaluation across the ten most spoken languages". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 16, pp. 43–54.

Blackmon, Marilyn Hughes et al. (Apr. 20, 2002). “Cognitive walkthrough for the web”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’02. New York, NY, USA: Association for Computing Machinery, pp. 463–470. ISBN: 978-1-58113-453-7. DOI: 10.1145/503376.503459. URL: https://doi.org/10.1145/503376.503459 (visited on 06/22/2023).

Burns, Andrea et al. (May 5, 2023a). *A Suite of Generative Tasks for Multi-Level Multimodal Webpage Understanding*. arXiv:2305.03668. type: article. arXiv. arXiv: 2305.03668[cs]. URL: http://arxiv.org/abs/2305.03668 (visited on 06/15/2023).

– (May 9, 2023b). *WikiWeb2M: A Page-Level Multimodal Wikipedia Dataset*. arXiv:2305.05432. type: article. arXiv. arXiv: 2305.05432[cs]. URL: http://arxiv.org/abs/2305.05432 (visited on 06/15/2023).

Campello, Ricardo JGB, Davoud Moulavi, and Jörg Sander (2013). “Density-based clustering based on hierarchical density estimates”. In: *Proc. Pacific-Asia conference on knowledge discovery and data mining (PAKDD)*.

Capra, Robert, Jaime Arguello, and Falk Scholer (2013). “Augmenting web search surrogates with images”. In: *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pp. 399–408.

Chelsy Xie, Xiaoxi, Isaac Johnson, and Anne Gomez (May 13, 2019). “Detecting and Gauging Impact on Wikipedia Page Views”. In: *Companion Proceedings of The 2019 World Wide Web Conference*. WWW ’19. New York, NY, USA: Association for Computing Machinery, pp. 1254–1261. ISBN: 978-1-4503-6675-5. DOI: 10.1145/3308560.3316751. URL: https://doi.org/10.1145/3308560.3316751 (visited on 06/18/2023).

Chen, A., P. W. Darst, and R. P. Pangrazi (Sept. 2001). “An examination of situational interest and its sources”. In: *The British Journal of Educational Psychology* 71 (Pt 3), pp. 383–400. ISSN: 0007-0998. DOI: 10.1348/000709901158578.

Chi, Ed H. et al. (Apr. 5, 2003). “The bloodhound project: automating discovery of web usability issues using the InfoScent simulator”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’03. New York, NY, USA: Association for Computing Machinery, pp. 505–512. ISBN: 978-1-58113-630-2. DOI: 10.1145/642611.642699. URL: https://doi.org/10.1145/642611.642699 (visited on 06/22/2023).

Clark, Ruth Colvin (2014). “Multimedia learning in e-courses”. In: *The Cambridge handbook of multimedia learning, 2nd ed*. Cambridge handbooks in psychology. New York, NY, US: Cambridge University Press, pp. 842–881. ISBN: 978-1-107-61031-6 978-1-107-03520-1 978-1-139-99016-5. DOI: 10.1017/CBO9781139547369.040.

Comenius, Johann Amos (1907). *The great didactic of John Amos Comenius*. Open Library ID: OL24155794M. London: A. and C. Black. 316 pp.

Constantin, Mihai Gabriel, Miriam Redi, et al. (2019). “Computational understanding of visual interestingness beyond semantics: literature survey and analysis of covariates”. In: *ACM Computing Surveys (CSUR)* 52.2, pp. 1–37.

Constantin, Mihai Gabriel, Liviu-Daniel Ştefan, et al. (May 1, 2021). "Visual Interestingness Prediction: A Benchmark Framework and Literature Review". In: *International Journal of Computer Vision* 129.5, pp. 1526–1550. ISSN: 1573-1405. DOI: 10.1007/s11263-021-01443-1. URL: https://doi.org/10.1007/s11263-021-01443-1 (visited on 06/21/2023).

Craswell, Nick et al. (Feb. 11, 2008). "An experimental comparison of click position-bias models". In: *Proceedings of the 2008 International Conference on Web Search and Data Mining*. WSDM '08. New York, NY, USA: Association for Computing Machinery, pp. 87–94. ISBN: 978-1-59593-927-2. DOI: 10.1145/1341531.1341545. URL: https://doi.org/10.1145/1341531.1341545 (visited on 06/22/2023).

Deng, Jia et al. (2009). "Imagenet: A large-scale hierarchical image database". In: *Proc. Conference on computer vision and pattern recognition (CVPR)*.

Deshpande, Mukund and George Karypis (2004). "Selective markov models for predicting web page accesses". In: *ACM transactions on internet technology (TOIT)* 4.2, pp. 163–184.

Dimitrov, Dimitar et al. (May 15, 2018a). "Query for Architecture, Click through Military: Comparing the Roles of Search and Navigation on Wikipedia". In: *Proceedings of the 10th ACM Conference on Web Science*. WebSci '18. New York, NY, USA: Association for Computing Machinery, pp. 371–380. ISBN: 978-1-4503-5563-6. DOI: 10.1145/3201064.3201092. URL: https://doi.org/10.1145/3201064.3201092 (visited on 06/18/2023).

– (2018b). "Query for architecture, click through military: Comparing the roles of search and navigation on wikipedia". In: *Proceedings of the 10th ACM conference on web science*, pp. 371–380.

Ding, Keyan, Kede Ma, and Shiqi Wang (2019). "Intrinsic Image Popularity Assessment". In: *Proc. International Conference on Multimedia (MM)*.

Edizel, Bora, Amin Mantrach, and Xiao Bai (2017). "Deep character-level click-through rate prediction for sponsored search". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 305–314.

Elazary, Lior and Laurent Itti (Mar. 7, 2008). "Interesting objects are visually salient". In: *Journal of Vision* 8.3, pp. 3.1–15. ISSN: 1534-7362. DOI: 10.1167/8.3.3.

Erickson, Kristofer, Felix Rodriguez Perez, and Jesus Rodriguez Perez (2018a). "What is the Commons Worth? Estimating the Value of Wikimedia Imagery by Observing Downstream Use". In: *Proc. International Symposium on Open Collaboration (OpenSym)*.

– (Aug. 22, 2018b). "What is the Commons Worth?: Estimating the Value of Wikimedia Imagery by Observing Downstream Use". In: *Proceedings of the 14th International Symposium on Open Collaboration*. OpenSym '18: The 14th International Symposium on Open Collaboration. Paris France: ACM, pp. 1–6. ISBN: 978-1-4503-5936-8. DOI: 10.1145/3233391.3233533. URL: https://dl.acm.org/doi/10.1145/3233391.3233533 (visited on 06/15/2023).

Ester, Martin et al. (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *International Conference on Knowledge Discovery and Data Mining (KDD)*.

Ferrada, Sebastián, Benjamin Bustos, and Aidan Hogan (2017). “IMGpedia: A Linked Dataset with Content-Based Analysis of Wikimedia Images”. In: *The Semantic Web – ISWC 2017*. Ed. by Claudia d’Amato et al. Vol. 10588. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 84–93. ISBN: 978-3-319-68203-7 978-3-319-68204-4. DOI: 10.1007/978-3-319-68204-4_8. URL: https://link.springer.com/10.1007/978-3-319-68204-4_8 (visited on 06/15/2023).

Flesch, Rudolph (1948). “A new readability yardstick.” In: *Journal of applied psychology* 32.3, p. 221.

Frick, Pauline and Anne Schüler (Oct. 1, 2023). “Extending the theoretical foundations of multimedia learning: Activation, integration, and validation occur when processing illustrated texts”. In: *Learning and Instruction* 87, p. 101800. ISSN: 0959-4752. DOI: 10.1016/j.learninstruc.2023.101800. URL: https://www.sciencedirect.com/science/article/pii/S0959475223000695 (visited on 06/21/2023).

Georgescu, Mihai et al. (2013). “Extracting event-related information from article updates in wikipedia”. In: *Proc. European Conference on Information Retrieval (ECIR)*.

Gildersleve, Patrick and Taha Yasseri (2018). “Inspiration, Captivation, and Misdirection: Emergent Properties in Networks of Online Navigation”. In: *Complex Networks IX*. Ed. by Sean Cornelius et al. Springer Proceedings in Complexity. Cham: Springer International Publishing, pp. 271–282. ISBN: 978-3-319-73198-8. DOI: 10.1007/978-3-319-73198-8_23.

Ginns, Paul (Aug. 1, 2005). “Meta-analysis of the modality effect”. In: *Learning and Instruction* 15.4, pp. 313–331. ISSN: 0959-4752. DOI: 10.1016/j.learninstruc.2005.07.001. URL: https://www.sciencedirect.com/science/article/pii/S0959475205000459 (visited on 06/21/2023).

Gog, Tamara van (Jan. 1, 2014). “The signaling (or cueing) principle in multimedia learning”. In: DOI: 10.1017/CBO9781139547369.014. URL: https://repub.eur.nl/pub/90919/ (visited on 06/21/2023).

Guo, Daibao et al. (Sept. 2020). “The Impact of Visual Displays on Learning Across the Disciplines: A Systematic Review”. In: *Educational Psychology Review* 32.3, pp. 627–656. ISSN: 1040-726X, 1573-336X. DOI: 10.1007/s10648-020-09523-3. URL: http://link.springer.com/10.1007/s10648-020-09523-3 (visited on 06/21/2023).

Gygli, Michael et al. (2013). “The interestingness of images”. In: *Proc. International Conference on Computer Vision (ICCV)*.

Gyselinck, Valérie and Hubert Tardieu (1999). “The role of illustrations in text comprehension: What, when, for whom, and why?” In: *The construction of mental representations during reading*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, pp. 195–218. ISBN: 978-0-8058-2428-5 978-0-8058-2429-2.

Halfaker, Aaron and R Stuart Geiger (2020). “Ores: Lowering barriers with participatory machine learning in wikipedia”. In: *Proc. Human-Computer Interaction (HCI)*.

Halfaker, Aaron, Os Keyes, et al. (May 18, 2015). “User Session Identification Based on Strong Regularities in Inter-activity Time”. In: *Proceedings of the 24th International Conference on World Wide Web*. WWW ’15. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, pp. 410–418. ISBN: 978-1-4503-3469-3. DOI: 10.1145/2736277.2741117. URL: https://doi.org/10.1145/2736277.2741117 (visited on 06/18/2023).

Hamilton, D. et al. (Mar. 1, 2021). “Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design”. In: *Journal of Computers in Education* 8.1, pp. 1–32. ISSN: 2197-9995. DOI: 10.1007/s40692-020-00169-2. URL: https://doi.org/10.1007/s40692-020-00169-2 (visited on 06/21/2023).

He, Shiqing et al. (2018a). “The_tower_of_babel. jpg: diversity of visual encyclopedic knowledge across Wikipedia language editions”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 12. 1.

– (June 15, 2018b). “The_Tower_of_Babel.jpg: Diversity of Visual Encyclopedic Knowledge Across Wikipedia Language Editions”. In: *Proceedings of the International AAAI Conference on Web and Social Media* 12.1. Number: 1. ISSN: 2334-0770. DOI: 10.1609/icwsm.v12i1.15037. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/15037 (visited on 06/15/2023).

Heald, Paul, Kristofer Erickson, and Martin Kretschmer (2015). “The Valuation of Unprotected Works: A Case Study of Public Domain Images on Wikipedia”. In: *Harvard Journal of Law & Technology* 29, p. 1. URL: https://heinonline.org/HOL/Page?handle=hein.journals/hjlt29&id=7&div=&collection=.

Helic, Denis (May 2012). “Analyzing user click paths in a Wikipedia navigation game”. In: *2012 Proceedings of the 35th International Convention MIPRO*. 2012 Proceedings of the 35th International Convention MIPRO, pp. 374–379.

Helic, Denis et al. (May 1, 2013). “Models of human navigation in information networks based on decentralized search”. In: *Proceedings of the 24th ACM Conference on Hypertext and Social Media*. HT ’13. New York, NY, USA: Association for Computing Machinery, pp. 89–98. ISBN: 978-1-4503-1967-6. DOI: 10.1145/2481492.2481502. URL: https://doi.org/10.1145/2481492.2481502 (visited on 06/18/2023).

Huang, Jeff and Efthimis N Efthimiadis (2009). “Analyzing and evaluating query reformulation strategies in web search logs”. In: *Proc. Conference on Information and Knowledge Management (CIKM)*.

Huh, Minyoung, Pulkit Agrawal, and Alexei A Efros (2016). “What makes ImageNet good for transfer learning?” In: *arXiv preprint arXiv:1608.08614*.

Ibrahim, Mohamed et al. (Sept. 1, 2012). “Effects of segmenting, signalling, and weeding on learning from educational video”. In: *Learning, Media and Technology* 37.3. Publisher: Routledge _eprint: https://doi.org/10.1080/17439884.2011.585993, pp. 220–235. ISSN: 1743-9884. DOI: 10.1080/17439884.2011.585993. URL: https://doi.org/10.1080/17439884.2011.585993 (visited on 06/21/2023).

Jansen, Bernard J., Amanda Spink, and Jan Pedersen (Aug. 31, 2004). "THE EFFECT OF SPECIALIZED MULTIMEDIA COLLECTIONS ON WEB SEARCHING". In: *Journal of Web Engineering*, pp. 182–199. ISSN: 1544-5976. URL: https://journals.riverpublishers.com/index.php/JWE/ (visited on 06/19/2023).

Johnson, Cheryl I. and Richard E. Mayer (2012). "An eye movement analysis of the spatial contiguity effect in multimedia learning". In: *Journal of Experimental Psychology: Applied* 18. Place: US Publisher: American Psychological Association, pp. 178–191. ISSN: 1939-2192. DOI: 10.1037/a0026923.

Johnson, Isaac et al. (May 22, 2021). "Global Gender Differences in Wikipedia Readership". In: *Proceedings of the International AAAI Conference on Web and Social Media* 15, pp. 254–265. ISSN: 2334-0770. DOI: 10.1609/icwsm.v15i1.18058. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/18058 (visited on 06/18/2023).

Karanam, Saraschandra, Herre van Oostendorp, and Bipin Indurkhya (Jan. 1, 2012). "Evaluating CoLiDeS + Pic: the role of relevance of pictures in user navigation behaviour". In: *Behaviour & Information Technology* 31.1. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0144929X.2011.606335, pp. 31–40. ISSN: 0144-929X. DOI: 10.1080/0144929X.2011.606335. URL: https://doi.org/10.1080/0144929X.2011.606335 (visited on 06/19/2023).

Khamparia, Aditya and Babita Pandey (2017). "Impact of Interactive Multimedia in E-Learning Technologies: Role of Multimedia in E-Learning". In: *Enhancing Academic Research With Knowledge Management Principles*. IGI Global, pp. 171–199. ISBN: 978-1-5225-2489-2. DOI: 10.4018/978-1-5225-2489-2.ch007. URL: https://www.igi-global.com/chapter/impact-of-interactive-multimedia-in-e-learning-technologies/www.igi-global.com/chapter/impact-of-interactive-multimedia-in-e-learning-technologies/179191 (visited on 06/21/2023).

– (2018). "Impact of interactive multimedia in E-learning technologies: Role of multimedia in E-learning". In: *Digital Multimedia: Concepts, Methodologies, Tools, and Applications*, pp. 1087–1110.

Khosla, Aditya, Atish Das Sarma, and Raffay Hamid (Apr. 7, 2014a). "What makes an image popular?" In: *Proceedings of the 23rd international conference on World wide web*. WWW '14. New York, NY, USA: Association for Computing Machinery, pp. 867–876. ISBN: 978-1-4503-2744-2. DOI: 10.1145/2566486.2567996. URL: https://doi.org/10.1145/2566486.2567996 (visited on 06/21/2023).

– (2014b). "What makes an image popular?" In: *Proc. International World Wide Web Conference (WWW)*.

Knight, Charles and Sam Pryke (2012). "Wikipedia and the University, a case study". In: *Teaching in higher education* 17.6, pp. 649–659.

Koopmann, Tobias et al. (Sept. 12, 2019). "On the right track! Analysing and Predicting Navigation Success in Wikipedia". In: *Proceedings of the 30th ACM Conference on Hypertext and Social Media*. HT '19. New York, NY, USA: Association for Computing Machinery, pp. 143–152. ISBN: 978-1-4503-6885-8. DOI: 10.1145/3342220.3343650. URL: https://doi.org/10.1145/3342220.3343650 (visited on 06/18/2023).

Kreiss, Elisa et al. (May 15, 2023). *Characterizing Image Accessibility on Wikipedia across Languages*. arXiv:2305.09038. type: article. arXiv. arXiv: 2305.09038[cs]. URL: http://arxiv.org/abs/2305.09038 (visited on 06/15/2023).

Kutner, Michael H et al. (2005). *Applied linear statistical models*. Vol. 5.

Lamprecht, Daniel et al. (Jan. 2, 2017). "How the structure of Wikipedia articles influences user navigation". In: *New Review of Hypermedia and Multimedia* 23.1. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13614568.2016.1179798, pp. 29–50. ISSN: 1361-4568. DOI: 10.1080/13614568.2016.1179798. URL: https://doi.org/10.1080/13614568.2016.1179798 (visited on 06/18/2023).

Lechner, Michael et al. (2011). "The estimation of causal effects by difference-in-difference methods". In: *Foundations and Trends® in Econometrics* 4.3, pp. 165–224.

Lehmann, Janette et al. (Sept. 1, 2014a). "Reader preferences and behavior on Wikipedia". In: *Proceedings of the 25th ACM conference on Hypertext and social media*. HT '14. New York, NY, USA: Association for Computing Machinery, pp. 88–97. ISBN: 978-1-4503-2954-5. DOI: 10.1145/2631775.2631805. URL: https://doi.org/10.1145/2631775.2631805 (visited on 06/18/2023).

– (2014b). "Reader preferences and behavior on Wikipedia". In: *Proceedings of the 25th ACM conference on Hypertext and social media*, pp. 88–97.

Lemmerich, Florian et al. (Jan. 30, 2019a). "Why the World Reads Wikipedia: Beyond English Speakers". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. WSDM '19: The Twelfth ACM International Conference on Web Search and Data Mining. Melbourne VIC Australia: ACM, pp. 618–626. ISBN: 978-1-4503-5940-5. DOI: 10.1145/3289600.3291021. URL: https://dl.acm.org/doi/10.1145/3289600.3291021 (visited on 06/18/2023).

– (2019b). "Why the world reads Wikipedia: Beyond English speakers". In: *Proceedings of the twelfth ACM international conference on web search and data mining*, pp. 618–626.

Lerner, Paul, Olivier Ferret, and Camille Guinaudeau (Jan. 11, 2023). *Multimodal Inverse Cloze Task for Knowledge-based Visual Question Answering*. arXiv: 2301.04366[cs]. URL: http://arxiv.org/abs/2301.04366 (visited on 06/18/2023).

Levene, Mark (Jan. 14, 2011). *An Introduction to Search Engines and Web Navigation*. John Wiley & Sons. 500 pp. ISBN: 978-1-118-06034-6.

Levie, W Howard and Richard Lentz (1982a). "Effects of text illustrations: A review of research". In: *Ectj* 30.4, pp. 195–232.

– (1982b). "Effects of text illustrations: A review of research". In: *Educational Communication & Technology Journal* 30. Place: US Publisher: Association for Educational Communications & Technology, pp. 195–232. ISSN: 0001-2890.

Li, Jingwei, Pavlo D. Antonenko, and Jiahui Wang (Nov. 1, 2019). "Trends and issues in multimedia learning research in 1996–2016: A bibliometric analysis". In: *Educational Research Review* 28, p. 100282. ISSN: 1747-938X. DOI: 10.1016/j.edurev.2019.100282. URL: https://www.sciencedirect.com/science/article/pii/S1747938X18305736 (visited on 06/21/2023).

Lloyd, S. (1982). “Least squares quantization in PCM”. In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137.

Loumakis, Faidon, Simone Stumpf, and David Grayson (Oct. 24, 2011a). “This image smells good: effects of image information scent in search engine results pages”. In: *Proceedings of the 20th ACM international conference on Information and knowledge management*. CIKM ’11. New York, NY, USA: Association for Computing Machinery, pp. 475–484. ISBN: 978-1-4503-0717-8. DOI: 10.1145/2063576.2063649. URL: https://doi.org/10.1145/2063576.2063649 (visited on 06/19/2023).

– (2011b). “This image smells good: effects of image information scent in search engine results pages”. In: *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 475–484.

Lucassen, Teun and Jan Maarten Schraagen (Apr. 27, 2010). “Trust in wikipedia: how users trust information from an unknown source”. In: *Proceedings of the 4th workshop on Information credibility*. WICOW ’10. New York, NY, USA: Association for Computing Machinery, pp. 19–26. ISBN: 978-1-60558-940-4. DOI: 10.1145/1772938.1772944. URL: https://doi.org/10.1145/1772938.1772944 (visited on 06/15/2023).

Lydon-Staley, David M et al. (2021). “Hunters, busybodies and the knowledge network building associated with deprivation curiosity”. In: *Nature Human Behaviour* 5.3, pp. 327–336.

Machlup, Fritz (1983). *The Study of Information: Interdisciplinary Messages*. Wiley.

Mautone, Patricia D. and Richard E. Mayer (2001). “Signaling as a cognitive guide in multimedia learning”. In: *Journal of Educational Psychology* 93. Place: US Publisher: American Psychological Association, pp. 377–389. ISSN: 1939-2176. DOI: 10.1037/0022-0663.93.2.377.

Mayer, Richard E. (July 9, 2020). *Multimedia Learning*. Higher Education from Cambridge University Press. ISBN: 9781316941355 Publisher: Cambridge University Press. DOI: 10.1017/9781316941355. URL: https://www.cambridge.org/highereducation/books/multimedia-learning/FB7E79A165D24D47CEACEB4D2C426ECD (visited on 06/20/2023).

McInnes, Leland, John Healy, and Steve Astels (Mar. 2017). “hdbscan: Hierarchical density based clustering”. In: *The Journal of Open Source Software* 2.11.

Menghini, Cristina, Aris Anagnostopoulos, and Eli Upfal (Dec. 2019). “Wikipedia Polarization and Its Effects on Navigation Paths”. In: *2019 IEEE International Conference on Big Data (Big Data)*. 2019 IEEE International Conference on Big Data (Big Data), pp. 6154–6156. DOI: 10.1109/BigData47090.2019.9005566.

Miz, Volodymyr et al. (Apr. 20, 2020). “What is Trending on Wikipedia? Capturing Trends and Language Biases Across Wikipedia Editions”. In: *Companion Proceedings of the Web Conference 2020*. WWW ’20. New York, NY, USA: Association for Computing Machinery, pp. 794–801. ISBN: 978-1-4503-7024-0. DOI: 10.1145/3366424.3383567. URL: https://doi.org/10.1145/3366424.3383567 (visited on 06/18/2023).

Morton, John and Mark H Johnson (1991). "CONSPEC and CONLERN: a two-process theory of infant face recognition." In: *Psychological review* 98.2, p. 164.

Moulavi, Davoud et al. (2014). "Density-based clustering validation". In: *Proc. SIAM international conference on data mining (SDM)*.

Mystakidis, Stylianos, Athanasios Christopoulos, and Nikolaos Pellas (Mar. 1, 2022). "A systematic mapping review of augmented reality applications to support STEM learning in higher education". In: *Education and Information Technologies* 27.2, pp. 1883–1927. ISSN: 1573-7608. DOI: 10.1007/s10639-021-10682-1. URL: https://doi.org/10.1007/s10639-021-10682-1 (visited on 06/21/2023).

Navarrete, Trilce and Elena Villaespesa (Jan. 1, 2020). "Image-based information: paintings in Wikipedia". In: *Journal of Documentation* 77.2. Publisher: Emerald Publishing Limited, pp. 359–380. ISSN: 0022-0418. DOI: 10.1108/JD-03-2020-0044. URL: https://doi.org/10.1108/JD-03-2020-0044 (visited on 06/15/2023).

Nguyen, Khanh et al. (Sept. 21, 2022). *Show, Interpret and Tell: Entity-aware Contextualised Image Captioning in Wikipedia*. arXiv:2209.10474. type: article. arXiv. arXiv: 2209.10474[cs]. URL: http://arxiv.org/abs/2209.10474 (visited on 06/15/2023).

Olteanu, Alexandra et al. (2019). "Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries". In: *Frontiers in Big Data* 2. ISSN: 2624-909X. URL: https://www.frontiersin.org/articles/10.3389/fdata.2019.00013 (visited on 06/18/2023).

Oostendorp, Herre van, Saraschandra Karanam, and Bipin Indurkhya (Jan. 1, 2012a). "CoLiDeS+ Pic: a cognitive model of web-navigation based on semantic information from pictures". In: *Behaviour & Information Technology* 31.1. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0144929X.2011.603358, pp. 17–30. ISSN: 0144-929X. DOI: 10.1080/0144929X.2011.603358. URL: https://doi.org/10.1080/0144929X.2011.603358 (visited on 06/22/2023).

– (2012b). "CoLiDeS+ Pic: a cognitive model of web-navigation based on semantic information from pictures". In: *Behaviour & Information Technology* 31.1, pp. 17–30.

Paivio, Allan (Sept. 13, 1990). "Dual Coding Theory". In: *Mental Representations: A dual coding approach*. Ed. by Allan Paivio. Oxford University Press, p. 0. ISBN: 978-0-19-506666-1. DOI: 10.1093/acprof:oso/9780195066661.003.0004. URL: https://doi.org/10.1093/acprof:oso/9780195066661.003.0004 (visited on 06/20/2023).

Pappas, Nikolaos et al. (2016). "Multilingual visual sentiment concept matching". In: *Proc. International Conference on Multimedia Retrieval (ICMR)*.

Paranjape, Ashwin et al. (2016). "Improving website hyperlink structure using server logs". In: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 615–624.

Park, Jaimie Y et al. (2015). "A large-scale study of user image search behavior on the web". In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 985–994.

Peeck, Joan (1993). "Increasing picture effects in learning from illustrated text". In: *Learning and instruction* 3.3, pp. 227–238.

Piccardi, Tiziano, Martin Gerlach, Akhil Arora, et al. (Jan. 2023a). “A Large-Scale Characterization of How Readers Browse Wikipedia”. In: *ACM Trans. Web.* ISSN: 1559-1131. DOI: 10.1145/3580318. URL: https://doi.org/10.1145/3580318.

– (Apr. 3, 2023b). “A Large-Scale Characterization of How Readers Browse Wikipedia”. In: *ACM Transactions on the Web* 17.2, 11:1–11:22. ISSN: 1559-1131. DOI: 10.1145/3580318. URL: https://doi.org/10.1145/3580318 (visited on 06/18/2023).

Piccardi, Tiziano, Martin Gerlach, and Robert West (Aug. 16, 2022). “Going Down the Rabbit Hole: Characterizing the Long Tail of Wikipedia Reading Sessions”. In: *Companion Proceedings of the Web Conference 2022*. WWW ’22. New York, NY, USA: Association for Computing Machinery, pp. 1324–1330. ISBN: 978-1-4503-9130-6. DOI: 10.1145/3487553.3524930. URL: https://doi.org/10.1145/3487553.3524930 (visited on 06/18/2023).

– (May 16, 2023a). *Curious Rhythms: Temporal Regularities of Wikipedia Consumption*. DOI: 10.48550/arXiv.2305.09497. arXiv: 2305.09497[cs]. URL: http://arxiv.org/abs/2305.09497 (visited on 06/18/2023).

– (2023b). “Curious Rhythms: Temporal Regularities of Wikipedia Consumption”. In: *arXiv preprint arXiv:2305.09497*.

Piccardi, Tiziano, Miriam Redi, et al. (Apr. 20, 2020a). “Quantifying Engagement with Citations on Wikipedia”. In: *Proceedings of The Web Conference 2020*. WWW ’20. New York, NY, USA: Association for Computing Machinery, pp. 2365–2376. ISBN: 978-1-4503-7023-3. DOI: 10.1145/3366423.3380300. URL: https://doi.org/10.1145/3366423.3380300 (visited on 06/18/2023).

– (2020b). “Quantifying engagement with citations on Wikipedia”. In: *Proc. International World Wide Web Conference (WWW)*.

– (2021). “On the Value of Wikipedia as a Gateway to the Web”. In: *Proc. International World Wide Web Conference (WWW)*.

Pirolli, Peter and Stuart Card (1999). “Information foraging”. In: *Psychological Review* 106. Place: US Publisher: American Psychological Association, pp. 643–675. ISSN: 1939-1471. DOI: 10.1037/0033-295X.106.4.643.

Rama, Daniele et al. (2021). “A Large Scale Study of Reader Interactions with Images on Wikipedia”. In: *EPJ Data Science*.

– (Jan. 3, 2022). “A large scale study of reader interactions with images on Wikipedia”. In: *EPJ Data Science* 11.1, p. 1. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-021-00312-8. URL: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00312-8 (visited on 06/15/2023).

Ratkiewicz, Jacob et al. (Oct. 8, 2010). “Characterizing and Modeling the Dynamics of Online Popularity”. In: *Physical Review Letters* 105.15, p. 158701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.105.158701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.105.158701 (visited on 06/18/2023).

Redi, Miriam et al. (2020). “A taxonomy of knowledge gaps for wikimedia projects (second draft)”. In: *arXiv preprint arXiv:2008.12314*.

Reinoso, Antonio J., Jesus M. Gonzalez-Barahona, et al. (July 2009). "A quantitative approach to the use of the Wikipedia". In: *2009 IEEE Symposium on Computers and Communications*. 2009 IEEE Symposium on Computers and Communications. ISSN: 1530-1346, pp. 56–61. DOI: 10.1109/ISCC.2009.5202401.

Reinoso, Antonio J., Rocio Muñoz-Mansilla, et al. (May 2012). "Characterization of the Wikipedia Traffic". In: *Proceedings of The Seventh International Conference on Internet and Web Applications and Services | The Seventh International Conference on Internet and Web Applications and Services | 27/05/2012 - 01/06/2012 | Stuttgart, Alemania*. The Seventh International Conference on Internet and Web Applications and Services. Num Pages: 7. Stuttgart, Alemania: E.T.S.I. Caminos, Canales y Puertos (UPM). ISBN: 978-1-61208-200-4. URL: http://www.thinkmind.org/index.php?view=article&articleid=iciw_2012_5_50_20194 (visited on 06/18/2023).

Ribé, Marc Miquel, Andreas Kaltenbrunner, and Jeffrey M Keefer (2021). "Bridging LGBT+ Content Gaps Across Wikipedia Language Editions". In: *The International Journal of Information, Diversity, & Inclusion* 5.4, pp. 90–131.

Ribeiro, Manoel Horta et al. (May 22, 2021). "Sudden Attention Shifts on Wikipedia During the COVID-19 Crisis". In: *Proceedings of the International AAAI Conference on Web and Social Media* 15, pp. 208–219. ISSN: 2334-0770. DOI: 10.1609/icwsm.v15i1.18054. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/18054 (visited on 06/18/2023).

Richardson, Matthew, Ewa Dominowska, and Robert Ragno (2007). "Predicting clicks: estimating the click-through rate for new ads". In: *Proceedings of the 16th international conference on World Wide Web*, pp. 521–530.

Rodi, Giovanna Chiara, Vittorio Loreto, and Francesca Tria (Feb. 2, 2017). "Search strategies of Wikipedia readers". In: *PLOS ONE* 12.2. Publisher: Public Library of Science, e0170746. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0170746. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170746 (visited on 06/18/2023).

Rosenbaum, Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1, pp. 41–55.

Rudolph, Michelle (2017). "Cognitive theory of multimedia learning". In: *Journal of Online Higher Education* 1.2, pp. 1–10.

Schifanella, Rossano, Miriam Redi, and Luca Maria Aiello (2015). "An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures". In: *International Conference on Web and Social Media (ICWSM)*.

Schneider, Sascha et al. (Feb. 1, 2018). "A meta-analysis of how signaling affects learning with media". In: *Educational Research Review* 23, pp. 1–24. ISSN: 1747-938X. DOI: 10.1016/j.edurev.2017.11.001. URL: https://www.sciencedirect.com/science/article/pii/S1747938X17300581 (visited on 06/21/2023).

Schomaker, Judith and Martijn Meeter (2015). "Short-and long-lasting consequences of novelty, deviance and surprise on brain and cognition". In: *Neuroscience & Biobehavioral Reviews* 55, pp. 268–279.

Shih, Yuhsun Edward (June 15, 2007). “Setting the New Standard with Mobile Computing in Online Learning”. In: *The International Review of Research in Open and Distributed Learning* 8.2. ISSN: 1492-3831. DOI: 10.19173/irrodl.v8i2.361. URL: https://www.irrodl.org/index.php/irrodl/article/view/361 (visited on 06/21/2023).

Singer, Philipp, Denis Helic, et al. (July 11, 2014). “Detecting Memory and Structure in Human Navigation Patterns Using Markov Chain Models of Varying Order”. In: *PLoS ONE* 9.7. Ed. by Zhongxue Chen, e102070. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0102070. URL: https://dx.plos.org/10.1371/journal.pone.0102070 (visited on 06/18/2023).

Singer, Philipp, Florian Lemmerich, et al. (Apr. 3, 2017a). “Why We Read Wikipedia”. In: *Proceedings of the 26th International Conference on World Wide Web*. WWW ’17. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, pp. 1591–1600. ISBN: 978-1-4503-4913-0. DOI: 10.1145/3038912.3052716. URL: https://doi.org/10.1145/3038912.3052716 (visited on 06/18/2023).

– (2017b). “Why we read Wikipedia”. In: *WebConf’17*, pp. 1591–1600.

Singh, Amanpreet et al. (June 2022). “FLAVA: A Foundational Language And Vision Alignment Model”. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, pp. 15617–15629. ISBN: 978-1-66546-946-3. DOI: 10.1109/CVPR52688.2022.01519. URL: https://ieeexplore.ieee.org/document/9880206/ (visited on 06/18/2023).

Spoerri, Anselm (Apr. 2, 2007). “What is popular on Wikipedia and why?” In: *First Monday*. ISSN: 1396-0466. DOI: 10.5210/fm.v12i4.1765. URL: https://firstmonday.org/ojs/index.php/fm/article/view/1765 (visited on 06/18/2023).

Srinivasan, Krishna et al. (July 11, 2021). “WIT: Wikipedia-based Image Text Dataset for Multimodal Multilingual Machine Learning”. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’21. New York, NY, USA: Association for Computing Machinery, pp. 2443–2449. ISBN: 978-1-4503-8037-9. DOI: 10.1145/3404835.3463257. URL: https://dl.acm.org/doi/10.1145/3404835.3463257 (visited on 06/15/2023).

Szegedy, Christian et al. (2016). “Rethinking the inception architecture for computer vision”. In: *Proc. Conference on computer vision and pattern recognition (CVPR)*.

Tauscher, Linda and Saul Greenberg (Mar. 27, 1997). “Revisitation patterns in World Wide Web navigation”. In: *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. CHI ’97. New York, NY, USA: Association for Computing Machinery, pp. 399–406. ISBN: 978-0-89791-802-2. DOI: 10.1145/258549.258816. URL: https://dl.acm.org/doi/10.1145/258549.258816 (visited on 06/22/2023).

Taylor, Robert S. (1962). “The process of asking questions”. In: *American Documentation* 13.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.5090130405, pp. 391–396. ISSN: 1936-6108. DOI: 10.1002/asi.5090130405. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090130405 (visited on 06/22/2023).

TeBlunthuis, Nathan, Tilman Bayer, and Olga Vasileva (Aug. 20, 2019). "Dwelling on Wikipedia: investigating time spent by global encyclopedia readers". In: *Proceedings of the 15th International Symposium on Open Collaboration*. OpenSym '19: The 15th International Symposium on Open Collaboration. Skövde Sweden: ACM, pp. 1–14. ISBN: 978-1-4503-6319-8. DOI: 10.1145/3306446.3340829. URL: https://dl.acm.org/doi/10.1145/3306446.3340829 (visited on 06/18/2023).

Tempelman-Kluit, Nadaleen (2006). "Multimedia learning theories and online instruction". In: *College & Research Libraries* 67.4, pp. 364–369.

– (Apr. 25, 2017). "Multimedia Learning Theories and Online Instruction | Tempelman-Kluit | College & Research Libraries". In: DOI: https://doi.org/10.5860/crl.67.4.364. URL: https://crl.acrl.org/index.php/crl/article/view/15811 (visited on 06/21/2023).

Trattner, Christoph et al. (Sept. 5, 2012). "Exploring the differences and similarities between hierarchical decentralized search and human navigation in information networks". In: *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*. i-KNOW '12. New York, NY, USA: Association for Computing Machinery, pp. 1–8. ISBN: 978-1-4503-1242-4. DOI: 10.1145/2362456.2362474. URL: https://doi.org/10.1145/2362456.2362474 (visited on 06/18/2023).

Tsikrika, Theodora and Christos Diou (Apr. 13, 2014a). "Multi-evidence User Group Discovery in Professional Image Search". In: *Proceedings of the 36th European Conference on IR Research on Advances in Information Retrieval - Volume 8416*. ECIR 2014. Berlin, Heidelberg: Springer-Verlag, pp. 693–699. ISBN: 978-3-319-06027-9. (Visited on 06/19/2023).

– (2014b). "Multi-evidence user group discovery in professional image search". In: *Proc. European Conference on Information Retrieval (ECIR)*.

Vaidya, Gaurav et al. (2015). "DBpedia Commons: Structured Multimedia Metadata from the Wikimedia Commons". In: *The Semantic Web - ISWC 2015*. Ed. by Marcelo Arenas et al. Vol. 9367. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 281–289. ISBN: 978-3-319-25009-0 978-3-319-25010-6. DOI: 10.1007/978-3-319-25010-6_17. URL: http://link.springer.com/10.1007/978-3-319-25010-6_17 (visited on 06/15/2023).

Viegas, Fernanda B (2007a). "The visual side of wikipedia". In: *Proc. Hawaii International Conference on System Sciences (HICSS)*.

– (Jan. 2007b). "The Visual Side of Wikipedia". In: *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07). ISSN: 1530-1605, pp. 85–85. DOI: 10.1109/HICSS.2007.559.

Vieira Bernat, Matheus (2023). *Topical Classification of Images in Wikipedia : Development of topical classification models followed by a study of the visual content of Wikipedia*. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-191161 (visited on 06/15/2023).

Vitulli, Marie A (2017). *Writing women in mathematics into Wikipedia*.

Waller, Vivienne (n.d.). “The search queries that took Australian Internet users to Wikipedia”. In: ().

Wannemacher, Klaus and Frank Schulenburg (2010). “Wikipedia in academic studies: corrupting or improving the quality of teaching and learning?” In: *Looking toward the future of technology-enhanced education: Ubiquitous learning and the digital native*. IGI Global, pp. 295–311.

Wen, Yandong et al. (2016). “A discriminative feature learning approach for deep face recognition”. In: *Proc. European conference on computer vision (ECCV)*.

West, Robert (n.d.). “Wikispeedia: An Online Game for Inferring Semantic Distances between Concepts”. In: ().

West, Robert and Jure Leskovec (Apr. 16, 2012a). “Human wayfinding in information networks”. In: *Proceedings of the 21st international conference on World Wide Web*. WWW ’12. New York, NY, USA: Association for Computing Machinery, pp. 619–628. ISBN: 978-1-4503-1229-5. DOI: 10.1145/2187836.2187920. URL: https://doi.org/10.1145/2187836.2187920 (visited on 06/18/2023).

– (2012b). “Human wayfinding in information networks”. In: *Proceedings of the 21st international conference on World Wide Web*, pp. 619–628.

WILSON, T.D. (Jan. 1, 1981). “ON USER STUDIES AND INFORMATION NEEDS”. In: *Journal of Documentation* 37.1. Publisher: MCB UP Ltd, pp. 3–15. ISSN: 0022-0418. DOI: 10.1108/eb026702. URL: https://doi.org/10.1108/eb026702 (visited on 06/19/2023).

Wojewoda, Mariusz (2018). “The Concept of Image According to John Amos Comenius and New Media”. In: *Philosophy and Canon Law* 4. Publisher: Wydawnictwo Uniwersytetu Slaskiego, pp. 85–99. ISSN: 2450-4955, 2451-2141. URL: https://www.ceeol.com/search/article-detail?id=892627 (visited on 06/20/2023).

Yang, Jheng-Hong et al. (Apr. 4, 2023). *AToMiC: An Image/Text Retrieval Test Collection to Support Multimedia Content Creation*. arXiv:2304.01961. type: article. arXiv. arXiv: 2304.01961[cs]. URL: http://arxiv.org/abs/2304.01961 (visited on 06/15/2023).

Yasseri, Taha et al. (2012). “Dynamics of conflicts in Wikipedia”. In: *PloS one* 7.6, e38869.

Yoon, Meehyun, Jungeun Lee, and Il-Hyun Jo (June 1, 2021). “Video learning analytics: Investigating behavioral patterns and learner clusters in video-based online learning”. In: *The Internet and Higher Education* 50, p. 100806. ISSN: 1096-7516. DOI: 10.1016/j.iheduc.2021.100806. URL: https://www.sciencedirect.com/science/article/pii/S1096751621000154 (visited on 06/21/2023).

You, Quanzeng et al. (2017). “Cultural diffusion and trends in facebook photographs”. In: *Proc. International Conference on Web and Social Media (ICWSM)*.

Zagoruyko, Sergey and Nikos Komodakis (2016). “Wide residual networks”. In: *arXiv preprint arXiv:1605.07146*.

Zhang, Wei et al. (2018). “User-guided hierarchical attention network for multi-modal social image popularity prediction”. In: *Proc. International World Wide Web Conference (WWW)*.

Zhou, Bolei et al. (2017). “Places: A 10 million image database for scene recognition”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.6, pp. 1452–1464.