A DECADE OF SCIENTIFIC COMMUNICATION ON TED TALKS : A BIBLIOMETRIC ANALYSIS

by

Hita Swamy

Submitted in partial fulfilment of the requirements

for the degree of Master of Electronic Commerce

at

Dalhousie University

Halifax, Nova Scotia

July, 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Researchers involved in communicating their works with the general public and other scientists for various reasons, including but not limited to citations, publications, or funding. Robert K Merton identified four norms of science: universalism, communalism, organized skepticism and disinterestedness. These norms of science point out that all researchers with the same credentials must receive equal recognition. However, there is bias in gender or race and affects various aspects of a researcher's career. Social media is gaining popularity in the scientific realm, where the recent happenings can reach a wider community than earlier. TED talks are one platform that invites various subject matter experts to share their knowledge and discoveries with the audience and make this available online. This study aims to understand the socio-demographic characteristics and scientific impact of researchers who appear on TED talk compared to a larger population. For this purpose, the dataset was taken from TED talks and the Web of Science database. And various algorithms were adopted for author disambiguation purposes. Also, the socio-demographic variables (country of affiliation, organization, gender) were considered independent variables. A binomial logistic regression model was used to predict the chances of appearing on TED talk and the correlation of the variables. From the study, it was seen that the Mathew effect, as defined by Robert K Merton, is seen to take place on the TED talk platform as the majority of researchers who appear on the TED platform were male researchers affiliated to top universities in the United States. Additionally, the prediction analysis also showed that the chances of appearing on TED talks decreases when a researcher is from other continents apart from North America, Female researchers, universities that do not belong to the top 100 universities globally. Thus, by showing that the researchers who are at the peak of their career get more recognition when compared to the rest. Further, this acts as a base for various platforms such as TED to understand and widen the category of guests to participate on their platform.

Keywords: Bibliometrics, Science Communication, Scholarly Communications, TED.

ACKNOWLEDGEMENT

This thesis is dedicated to anyone who wishes to deviate from the path to see where life takes us.

I would take this opportunity to express my gratitude to my supervisor, Dr Philippe Mongeon for guiding me throughout this process and making this journey highly knowledgeable and interesting. Dr Colin Conrad for being an amazing mentor and support throughout my master's degree. Faculty of computer science, Dalhousie university for this incredible journey where I have met the most amazing people ever.

Words can’t express the gratitude to my family and friends who have been a constant support without which this would have never been possible.

CHAPTER 1 INTRODUCTION

1.1 Background

"A scientific discovery is only as good as its communication; the key is to accommodate the multiple communication paths from that discovery" (Abraham, 2020, pp.3-4). The idea of science communication has always been a component of the scientific process, and scientists have always been keen on sharing their ideas and discoveries with others (Fleming, 2009). Furthermore, scientific communication takes place between scientists and between the scientific community and laypeople. It is vital for laypersons to grasp scientific ideas and issues; according to Chang et al. (2018), the scientific literacy model proposed by Miller (1998) says that educating the public eradicates misperceptions regarding the discoveries, which creates greater support for science (Chang et al., 2018). Also, scientific communication increases public involvement in policymaking (Weber & Schell Word, 2001). Therefore, one can say that science communication plays a role beyond sharing scientific advances (Chang et al., 2018).

There are many ways to communicate scientific findings to the public and other researchers. Communication is generally categorized as either verbal or written and can be either formal and traditional (e.g., peer-reviewed articles within scientific journals) or informal and modern (i.e., sharing one's scientific findings through social media applications like Twitter and Instagram). Sharing results through informal venues such as blogs and social media is becoming increasingly more common. Science has become more available with the internet and enabled users to publish their content on to the internet via blogs, user-generated tweets, videos, and others (Abraham, 2020). As such, nearly 60% of the general public is believed to rely on social media for scientific information (Tsubokura et al., 2018).

Other prevalent sources of scientific information for laypeople are websites that aim to make science more accessible to the general public. TED website is an example of such a website, and this also includes its YouTube channel and TED podcast. Technology, Entertainment, and Design is abbreviated as TED. While TED was founded as a tech conference in 1984, it has since expanded to include talks on a wide variety of topics, with its conference presentations being made available online. The TED website hosts recorded videos of talks given by various experts, celebrities, and prominent individuals (Tsou et al., 2014). TED is funded by advertising and members, and its presentations are available to view online for free. TED talks have been deemed a science communication phenomenon, with talks in over 100 languages and millions of views on its videos (Tsou et al., 2014). As various speakers narrate their ideas during the speech, we can say that TED talks are an example of narrative communication, a form of communication that influences how individuals perceive and respond to the world (Dahlstrom, 2010). Narratives are powerful enough to introduce new information, defamiliarize existing information, help individuals understand information, and provide new behavioural models (Dahlstrom, 2010).

Due to its prevalence and popularity, the TED talk organization and community should be better studied (Tsou et al., 2014). For instance, studies like the one conducted by Sugimoto et al. (2013a) and Tsou et al. (2014) show that only 27% of these presentations are by female presenters, indicating gender bias. Findings such as these suggest that biases in who gets invited to give a tend talk exist, and these biases can impact how laypersons view the scientific community.

While communicating one's results is a crucial component of the scientific process, not all scientists are afforded the same opportunity to share their discoveries with the scientific community and the broader public. Becoming a member of the scientific community has many intrinsic and extrinsic rewards, including peer recognition and the opportunity to make breakthrough discoveries, either individually or while working as part of a scientific team**.** However, factors such as being affiliated with a prestigious university or early career launch can give some scientists an advantage over others. In addition, scientists that are already prominent within their discipline are often provided with a disproportionate amount of credit for their contributions, either when working as part of a collaborative team or when they are only one of several scientists who have independently made the same scientific discovery. This phenomenon is commonly known as the Mathew Effect for scientific achievement (Merton, 1968). Scientists who gain this disproportionate credit or recognition due to the Matthew Effect receive additional visibility and prestige through increased citations and funding. In contrast, productive yet less visible, scientists remain in the background. The Matthew Effect can be exacerbated by and lead to inequalities in scientific recognition based on demographic variables such as scientists' gender, race, and socio-economic status.

1.2 Objectives of the study

While TED invited individuals from academia, industry backgrounds, and celebrities and other public figures to give talks, the current study will focus exclusively on TED speakers who work within academia, as the general public is said to have more trust in academic researchers compared to their industry counterparts (Sugimoto et al., 2013). The key objective of this study is to explore various characteristics of the academic TED speakers to help us understand the reward system of science its biases. The specific aims of this study are listed below:

1. Exploring socio-demographic and professional characteristics of academic TED talk presenters and non-TED talk researchers in the same field of research. The characteristics are:
    a. Gender
    b. Academic age (Number of years the researcher is in academia)
    c. Country of affiliation
    d. University/Organization
    e. University ranking (Leiden University Ranking)
2. Bibliometric Analysis.

    A citation analysis will be conducted to understand and compare the scientific impact of researchers who have given a TED talk. Those researchers have not given a TED talk, using different indicators such as publications and citations.
3. Predictive Analysis

    A logistic regression model will be used to predict whether or not an academic researcher will be invited to give a TED talk according to their socio-demographic characteristics.

Moving further, chapter 2 discusses the literature review relevant to the study. The methodology used to conduct the study is given in chapter 3. And, chapter 4 covers the results, while chapter 5 provides the analysis of the experiment. Finally, chapter 6 concludes the experiment.

CHAPTER 2 LITERATURE REVIEW

2.1 Goal, norms, and reward system of science

Every type of industry has its unique goals and norms. Likewise, as an institution, science has its norms, which can be referred to as scientific norms.

2.1.1 Scientific norms

According to Merton (1938), despite different institutional constraints, the persistent development of science occurs only in societies of a certain order. Active participation of capable persons in scientific pursuits is needed for science to advance; therefore, it is vital to understand the factors that might prevent otherwise skilled and productive scientists from achieving recognition and encouragement for their work (Merton, 1938). Furthermore, Merton (1938) mentions that science has characteristics such as honesty, organized skepticism, disinterestedness and impersonality, and its traditional autonomy and ethos, often challenged by an external authority. Thus, Merton identified four norms to ensure the functioning of science (Merton, 1973, 1942):

1. Universalism: Scientists are to be rewarded solely based on their work and results. In other words, scientific works are to be judged on their merits alone, and not in regards to characteristics of the scientist or researcher, such as class, race, and religion.

2. Communism: Knowledge is said to be a product of collective and cumulative efforts by the scientific community. Also, scientists gain recognition when they make their work public and available to others. Therefore, it can be said that

results and discoveries belong to the community as a whole and not merely a property of an individual researcher.

3. Disinterestedness: Scientists enter into their professors for the primary goal of advancing scientific knowledge rather than to achieve goals such as gaining personal recognition.

4. Organized Skepticism: Researchers must subject all findings to a critical appraisal until the necessary proof has been obtained, including evidence to support their own hypotheses.

These norms form a foundation for how science, ideally, grows as an industry. Further, when examining the career or research prospects of the persons involved in science and the broader society's reception of science, Merton discusses the aforementioned Matthew Effect, elaborated upon in the other sections. In addition to understanding scientific norms and the Matthew Effect, it is also vital to understand the reward system of science and its possible effects.

2.1.2 Reward System of Science

There are several different rewards of engaging in science. Scientists who have best fulfilled their roles are given recognition and esteem, and this is a reward system that the institution of science has developed (Merton, 1957). Rewards can be intrinsic (e.g., the satisfaction that comes from knowing that you have advanced human knowledge, pure enjoyment when engaging in scientific endeavours), or they can be extrinsic. Cole and Cole (1967) mention a few extrinsic rewards, such as "recognition" in the form of being granted honorific awards and memberships in honorific societies. Recognition for scientific work also comes in the form of being granted positions within top-ranked departments (Cole & Cole, 1967). While it is important that productive researchers be recognized for their hard work and contributions, the current reward system also tends to prioritize research that is deemed "relevant" and novel over research that is creative and exploratory (Cole & Cole, 1967).

The first person to document the priority rule in the reward system of science and document it was Robert Merton (1957). The following are a few priority rules that he noted (Merton, 1957):

1. While rewards are given to the first team or scientist who makes the discovery, second runners get very little or nothing.

2. Prestige (i.e., peer recognition) is the scientists' primary reward rather than money.

3. Prizes such as Field Medal, Nobel prize, promotions and eponymy are few forms of prestige that researchers consider.

Related to the concept of prestige is the concept of symbolic capital. One aspect of the reward system of science can be seen as symbolic capital. One reason why researchers engage in science communication is to acquire symbolic capital. So the question that follows is, how do scientists acquire symbolic capital?. Attention, including from others within the scientific community, is one way to acquire symbolic capital (Franck, 1999). Citations indicate the amount of attention that a researcher receives from the other community members, and researchers use citations to show support to other researchers in the community who have similar goals.

2.2 Bias in Scientific Recognition

As covered thus far, scientists have many motivations for engaging in science and receive many different rewards. Ideally, according to the scientific norms, a scientist conducts research objectively to advance human knowledge. However, the reality is that rewards such as obtaining prestige and symbolic capital also motivate scientists. The desire for prestige and symbolic capital may not always hamper scientific advancement. Still, scientific advancement can become limited when productive scientists do not receive the prestige and symbolic capital that they might otherwise receive if it were not for biases in who receives recognition for their contributions. "Universalism asserts that contribution to science and that scientific merit should remain independent of race, nationality, culture or gender"(Hogan & Sweeney, 2013). However, inequalities seem to exist within science. According to (Rossiter, 1993), the question of inequality within science is often related to how well-known scientists should be and how specific and widespread a scientific reputation should be. Furthermore, she also proposes having a scale or measure to understand the unrecognized yet deserving scientists (Rossiter, 1993). To understand the reasons behind formulating such questions, one can look into different factors contributing to inequalities faced in science.

2.2.1 Mathew Effect

While the Matthew Effect has been applied more broadly to matters of prestige and status, it is most commonly applied to the institution of science and how scientists who are already well-recognized continue to gain recognition (Merton, 1968). Many scientific advancements are a result of collaboration, so questions arise regarding how credit for scientific discoveries should be divided among members of a team. The answer is that the already well-recognized collaborator generally gets in the greater share of recognition in the form of prizes, book deals, and more, while the others might barely get mentioned in the footnote of the biography or during their more well-known colleagues' Nobel prize speeches (Merton, 1968). A related finding is that that scientists at major universities gain more recognition compared to scientists from lesser-known universities, despite being equally productive (Merton, 1968). Furthermore, scientists who receive recognition for their research early in their careers are more productive than the ones who do not (Bol et al., 2018; Merton, 1968). Additionally, Merton links the unequal division of recognition to "accumulation of advantage", in which scientists who have charisma, existing recognition within their field, acquiring-ranking positions in their institutions are the ones who gain more fame (Merton, 1968; Rossiter, 1993).

In short, the Matthew Effect, as relating to scientific achievement, is when scientists who are already successful are more likely to advance further in their careers than scientists who are not as successful (Bol et al., 2018; Merton, 1968). Scientific achievement may entail involvement in further discoveries, receiving, citations or being given awards or funding. For example, receiving funds earlier in one's career has been linked to the likelihood of getting funding later, and those who were unsuccessful in receiving funds are less likely to participate in future funding opportunities (Bol et al., 2018), in large part due to the fact that grant reviewers are more likely to highly score grants submitted by individuals who have already been awarded grants in the past.

2.2.2 Gender Bias

In addition to the type of institution a researcher is affiliated with and metrics such as the number of citations and past grant funding, researcher gender has been studied in regard to the Matthew Effect. According to O'Brien et al. (2019), "the principle of gender equality is widely embraced but not clearly defined". An increasing number of women are rising to various leadership roles and gaining prominence within the scientific community, and there has also been an increased interest in examining gender inequality within science. According to O'Brien et al. (2019), various factors and interactions at the individual, family, and societal levels can lead to inequalities within an industry. While the principle of equality is touted within science, women scientists receive less recognition and fewer opportunities compared to their male colleagues (O'Brien et al., 2019). One contributing factor, which O'Brien et al. (2019) refer to as the "mother of all conflicts", is the work-home conflict faced by many female workers, which encompasses unconscious biases regarding the role of women in society and the fact that many women take on more caregiving roles and do more domestic chores than their male significant others, leaving less time and energy to devote to their careers and limiting their opportunities for advancement.

In order for science to live up to its ideal of being a meritocracy, then equal achievements should receive similar recognition (Rossiter, 1993). However, female scientists repeatedly receive less recognition than their male collaborators (Rossiter,1993). There are exceptions, such as Marie Currie, who, unlike many wife-collaborators, was not Pierre Curie's silent partner or invisible co-worker but was the sole author of many important papers (Rossiter, 1993) and who received widespread acclaim during her lifetime. However, many female scientists have not only been under-recognized during their lifetimes but have been lost or nearly lost to history, perhaps in part due to biases on the part of historians (Rossiter, 1993). An example provided by Rossiter (1993) is Trota of Salerno, an eleventh-century physician in Salerno, Italy, who was, due to her accomplishments, assumed to have been male by historians for many years afterwards. Another example of gender inequality would be the experiences of Lise Meitner, who worked for decades alongside Otto Hahn in Germany. Despite Meitner's contributions, only her male collaborator was awarded the Nobel prize for Nuclear Fission in 1940 (Rossiter, 1993).

Throughout history, there have been many examples of women not receiving the same degree of recognition as their male counterparts, and this is particularly true for female scientists who are married and involve in collaborative research with their husbands (Rossiter, 1993). Another example of minimizing the presence of women in science can be seen in the naming of various journals, such as *American Men of Science* (Rossiter, 1993). Additionally, in the past, women were included in data collection as subjects but were then omitted from the final publications of the research. A study of American Chemists by Anslem Strauss and Lee Rainwater in 1962 is an example of this scenario (Rossiter, 1993).

The likelihood of receiving citations, authorship, grants, and awards is also subject to gender bias. In a global and cross-disciplinary bibliometric analysis of the relationship between gender and research output, Larivière et al. (2013) find that the extent of collaboration, and scientific impact of articles published between 2008 to 2012, found that women in dominant author positions receive fewer citations than men in the same positions, at least in countries that have a high research output. Additionally, papers with sole female authorship, first-authorship, and last-authorship attracted fewer citations when compared to papers with sole male authorship, first-authorship, and last-authorship

(Andersen et al., 2019; Larivière et al., 2013). Because citation-based indicators play a central role in the evaluation processes of researchers, for example, in determining whether or not a researcher should be awarded tenure, women are at a disadvantage compared to their male colleagues (Ghiasi et al., 2018; Larivière et al., 2013).

In regard to receiving funding, the research on gender biases has yielded mixed results. For instance, an analysis of health services and policy research funding in Canada found that, overall, women and men were equally likely to receive National Science Foundation (NSF) funding, except for female researchers under 45 years of age, who had lower success rates in terms of being awarded funding. Additionally, this study also found that fewer than expected women were submitting grant proposals in all fields except engineering (Rissler et al., 2020; Tricco et al., 2017). Receiving grants significantly impacts career advancement in academia. Yet, the extent of gender bias in awarding grant funding remains unclear (Rissler et al., 2020; Tricco et al., 2017).

2.2.3 Racial Bias

While many studies have focused on gender as a socio-demographic factor of bias, racial bias can also lead to inequality within science. An early example can be found in reference to psychologist Anne Roe's book, Making of a Scientist (1953), where scientists were eliminated from participating in the study and being included in the subsequent book on the basis of being "foreign-born" (Rossiter, 1993). In Nazi Germany, many were excluded or eliminated from universities and scientific institutes by virtue of not having "Aryan" ancestry (Merton, 1938), including many eminent scientists.

Racial bias is seen in reference to funding, citations, and other types of scientific activity. This bias can be seen across many disciplines, including the sciences (Ray, 2018).

Because citations are a measure of one's intellectual influence, racial bias in the number of citations has a significant impact on a scientist's ability to receive tenure and advance in other ways within his or her career.

In regards to funding, one study showed that African-American/Black scientists are less likely to receive a National Institute of Health (NIH) R01 grant compared to White scientists, despite the fact that the NIH promotes a diverse biomedical workforce by having various control factors such as educational background, country of origin, training, previous research awards and employer characteristics (Hoppe et al., 2019). Again, because receiving funding is linked to the likelihood of receive future funding, and because funding influences career advancement, such as receiving tenure, these biases can put non-White researchers at a significant disadvantage compared to their White colleagues.

2.3 Science-society interactions/ social impact of research/ science communication

2.3.1 Aims of scientific communication and public involvement in science.

Communication is vital not only for the creation of knowledge but also for its diffusion (Chawla & Singh, 2012). Burns et al. (2013) mention five aims of science communication: increased awareness, enjoyment, interest, opinion-forming, and understanding. Further, the National Academies of Sciences (2016) identified the five goals of science communication as:

1. To share the recent discoveries and excitement surrounding science.

2. Science communication can be used to understand and navigate the modern world by increasing appreciation for science.

3. Increasing knowledge and understanding of science addressing a particular matter that requires a decision.

4. Influencing the opinions, policy preferences or behaviour of people when the available evidence points out the concerns for public safety and societal concern.

5. To get perspectives of a diverse group to address concerning societal problems.

Finally, Kappel & Holmen (2019) posit that scientific information can be disseminated in two ways, one-way transmission of information or public participation paradigm with public involvement.

The involvement of the public in science leads to scientific institutions gaining more trust from the broader community, leading to greater acceptance of findings that might otherwise not be easily accepted (Kappel & Holmen, 2019). Science communication aims to achieve social acceptance, which, in turn, entails creating a positive attitude in the larger population regarding the importance of funding science, governance, and application of science (Kappel & Holmen, 2019). Researchers involved in science communication should receive public input regarding acceptable research aims and applications of science and allow the public to provide their interpretation of results so that scientists do not fall victim to researcher biases or narrow interpretations of their work (Kappel & Holmen, 2019). Furthermore, science communication aims at generating political support for the discoveries (Kappel & Holmen, 2019).

The public can be kept informed of scientific advancements in different ways, including public hearings, science shops, and mass media (Kappel & Holmen, 2019). In addition, through the internet, different platforms and media applications have gained prominence (Kappel & Holmen, 2019). Further and colleagues (2016) found a positive impact on the perceived moral trustworthiness of scientists who use blogs to disseminate their research when they include a discussion on the possible ethical implications of scientific findings in their blog post.

Furthermore, opportunities in terms of funding, such as the ones offered by the National Academies of Sciences, Engineering and Medicine, the Swiss National Science Foundation, Alfred P Sloan Foundation, VolkswagenStiftung Foundation, and MIF's Science Communications Funders Network, have launched in response to a growing interest in science communication (Ngumbi, 2020). Therefore, communicating scientific results to the broader public, and giving laypeople a chance to engage with research, can have profound implications for scientific advancement.

2.3.2 Impact of science communication on researchers.

Research in science communication sheds light on scientists' motivations for communicating their findings to the public, such as Ngumbi (2020), where the results mention that science communication increases the chances of funding and citations. Also, Shugrat and Racaniello (2015) mention growing evidence to support the statement that outreach can positively impact scientists' careers. Results show that scientists who engaged with the public actively published more papers and were more frequently cited than their peers who did not engage with the public.

Developing a personal brand and gaining visibility can be accomplished through increased scientific communication. According to Ngumbi (2020), benefits such as increased attention from grant funding agencies, increased citations and new pathways to influence science policy can be achieved through greater public visibility. Scientific communication correlates positively with increased citations in ecology and conservation research, an example of the previously mentioned scenario (Ngumbi, 2020). Also, tweeting can disseminate research results widely, further elevating the scientist's visibility (Ngumbi, 2020).

Finally, engaging in science communication increases communication with potential collaborators and is linked to future collaboration and a greater likelihood of engaging in interdisciplinary research. These benefits of engaging in scientific communication can advance a scientists' career (Ngumbi, 2020).

Because of the benefits of widely disseminating research findings and engaging with the general public, many funding agencies now explicitly state the importance of scientific communication. For instance, proposals submitted to NSF have been advised to include research products or broader impact, such as those generated by scientific communication. Additionally, scientists are allowed by academic institutions to make use of metrics provided by blog sites or social media platforms where they share their research to understand the impact scientists' output has (Ngumbi, 2020).

Further, a study conducted by Liang et al. (2014) shows that social media platforms like Twitter can advance a scientist's career by promoting their research. This can be in different ways like funding, citations and collaborations. Likewise, platforms such as

TED talks allow scientists to disseminate their research to a large, general public and experts within their own fields. Due to the increasing popularity of TED talks, it is helpful to understand how TED talks advance scientific communication and understand which types of scientists are more likely to get the opportunity to use this method of sharing their research.

2.4 TED Communication

Nowadays, an increasing number of individuals gather information using online courses (Sugimoto et al., 2013). Therefore, there are numerous ways in which the internet can serve the purpose of popularizing science. However, according to Sugimoto et al. (2013), online videos are an interactive and novel platform for popularizing science, and one particularly popular and successful online platform that has done so is the TED (Technology, Entertainment and Design) conference and its corresponding website, YouTube channel, and podcast.

2.4.1 Introduction and History of TED

TED started in 1984 as a conference hosted in Monterey, California, focusing on technology, entertainment, and design; today, the TED conference covers a broader range of topics, ranging from business to global issues in over 100 languages (*Our organization, n.d.*).TED is a nonprofit organization devoted to spreading ideas, most typically using short talks developed by topic experts. In 2001, TED was acquired by Chris Anderson's nonprofit sapling foundation (*Our organization, n.d.*). As a result, the presenters at TED became more diverse in their area of expertise and included individuals working in technology and philanthropists, scientists, and musicians.

Three significant additions were made to the TED organization between 2001-2006, those being:

1. TEDGlobal, a sister conference held in various locations of the world.
2. TED Prize, granting its winners a wish to change the world.
3. Podcasts of TED talks, which allowed the best TED content to be made more accessible online.

Additionally, there are various subtypes of TED conferences dedicated to specific topics of interest groups, such as TEDWomen and the aforementioned (*Conferences*, n.d.). As TED continues to expand its scope and focus, and as it continues to make its content more accessible via online outlets, it also attracts larger audiences.

2.4.2 Speaking at TED

TED presenters are usually the most elitist individuals known for creating ground-breaking media, invent world-changing devices and run the most admired companies (*Speaking at TED, n.d*). Speakers who inform and inspire, surprise and delight are often searched by TED year-round (*Speaking at TED, n.d*). "TED also seeks out emerging artists, scientists and thinkers, introducing them to TED community well before they hit the mainstream" (*Speaking at TED, n.d*).

Further, it is mentioned on the TED website that, collectively, TED speakers have won every major prize awarded for excellence, including the Nobel, Pritzker, Fields, Pulitzer, Oscar, Grammy, Emmy, Tony and MacArthur "genius" grant. Such award-winning speakers include Jane Goodall, Bill gates, Simone Giertz (*Speaking at TED, n.d*).

They also provide the speaker nomination form that people can use to nominate someone or themselves (*Speaking at TED, n.d*).

2.4.3 Impact of TED talks

In the digital age, one of the most successful outreach initiatives is the TED website, which hosts videos of presentations given at TED conferences by academics, industry figures, artists, and other prominent individuals (Tsou et al., 2014). With the internet providing easy access to a wide variety of information for all age groups and target audiences, TED recorded videos are easily accessible to anyone with an internet connection. However, when we talk of the scope and impact of TED talks, it seems that TED talks primarily impact the public sphere rather than the academic sphere (Sugimoto & Thelwall, 2013). This impact on the public sphere might be primarily because these TED talks are designed and marketed towards a broad audience that includes laypeople, rather than an audience comprised exclusively of other industry experts (di Carlo, 2014). Also, TED talks, beyond merely disseminating knowledge, encourage the audience to "make a change" and are often related to real-life concerns, goals, and interests (di Carlo, 2014). Yet another aspect of TED talks that makes them accessible to a general audience is that these talks focus more on the relationship of the experts with the subject matter than on the specific identity or the reputation of the expert (di Carlo, 2014). Additionally, it is mentioned on TED partnerships site that TED is one of the most trusted brands (Whatever your brand CHALLENGE,TED PARTNERSHIPS has a solution, 2021). While online platforms such as TED might recruit or invite experts to submit material, the relationship between overall online presence and career advancement within academia remains unclear. It is a new and growing research topic. For example, a study conducted by (McClain & Neeley, 2015) shows no correlation between scientific careers and an active social media platform. Because the relationship between online presence and academic success is a new and growing research topic, it would be beneficial to have a deeper understanding of how characteristics of academics influence their likelihood of getting invited to share their knowledge through popular organizations that make their content available online.

2.4.3 Previous TED talk Studies

There have been some studies conducted examining TED talks. For example, a study by Sugimoto et al. (2013) aimed at understanding the different characteristics of academic TED talk presenters and the impact of giving a TED talk on the number of citations. This study adopted various methodologies, such as coding the gender of the presenters upon examining both the videos and pronouns provided on the TED website and whether presenters were in academia or not(Sugimoto et al., 2013). Furthermore, Sugimoto et al. (2013) also obtained university rankings from Times Higher Education (THE) world university rankings 2011-2012 to rank the universities that TED academic presenters were affiliated with (Sugimoto et al., 2013). Finally, the publication data retrieved from Web of Science; given that all possible matches were retrieved, the best possible match was retained after comparing various attributes present on their CV like research institute (Sugimoto et al., 2013a).

Sugimoto et al. (2013) found that 21% of the TED presenters were academics, 27% were female, and no statistical differences in the distribution of gender by academic status. Female presenters were younger than their male counterparts (Sugimoto et al., 2013). Additionally, the majority of the presenters, 73%, held the title of professor, followed by associate professor, assistant professors, and other positions, such as adjunct, lecturer, and research scientist, in that order (Sugimoto et al., 2013). More than a quarter of the presenters were affiliated with California-based universities (Sugimoto et al., 2013). Also, most institutions were based in the United States, and most presenters were affiliated with institutions in the United States (Sugimoto et al., 2013).

Furthermore, there was no correlation between the popularity of their TED talk and the overall number of citations received by the presenter (Sugimoto et al., 2013). Finally, to understand the impact of TED talks on academic presenters, citations received by academics were analyzed before and after giving the TED talk. Perhaps surprisingly, giving a TED talk was not associated with increased citations (Sugimoto et al., 2013). Therefore, appearing on TED talks did not affect scientists' or presenters' research impact.

Another study regarding TED talks was conducted by Sugimoto and Thelwall (2013). Along with identifying various aspects of TED, the study aimed to understand the reaction academic presenters received for scientific talks compared to their non-academic counterparts (Sugimoto & Thelwall, 2013). The methodology of this study included different metrics, including TED metrics, YouTube metrics, Google Scholar citations, Web of Science citations, Google Books results, and information from syllabi and other institutional documents (Sugimoto & Thelwall, 2013). As a result, it is found that academics' videos attracted more citations or and mentions, as well as a higher proportion of likes on social media platforms when compared to non-academic presenters of scientific talks (Sugimoto & Thelwall, 2013). Therefore, the study suggests that the general public pays more attention to academic researchers' discoveries than non-academic ones.

CHAPTER 3 METHODOLOGY

3.1 General methodological approach

The current study is a quantitative descriptive analysis of a decade of scientific communication on TED talks. Exploratory analysis of the data will be conducted using various statistical functions such as grouping and measures of central tendency and bibliometric methodologies such as citation analysis. Further, the predictive analysis model is adopted to understand the impact of various characteristics of speakers that would affect their appearance on TED talk. Binomial logistic regression will be used to identify whether or not a researcher would appear on TED talk based on characteristics such as gender, academic age, university, and race.

Bibliometric analysis is a type of quantitative analytics that investigates external characteristics of scientific literature. Before exploring the dataset used to conduct the study, it is beneficial to look into the meaning of "bibliometrics". The first appearance of the term bibliometrics was seen in 1969 in the Journal of Documentation (Broadus, 1987). Since then, there have been various definitions of bibliometrics, including: "bibliometrics is the quantitative study of written communication through its physical realization" (Broadus, 1987, pp.373-379), and bibliometrics being is the usage of statistical analyses for studying publication patterns (McBurney & Novak, 2002).

Bibliometrics offers many tools and methods to study the structure and process of scholarly communication, such as citation analysis (Borgman & Furner, 2002). With the assumption that the articles are somehow related to one another, citation analysis looks at the relationship between a paper and the papers it has cited (McBurney & Novak, 2002). Unfortunately, fields, institutions, countries, and authors are few variables recorded and analyzed using citation analysis (McBurney & Novak, 2002). Journal impact factor is another tool used in bibliometrics. It is calculated by dividing the number of citations a journal receives for articles published in the two previous years by the number of articles published in that journal in those same years (McBurney & Novak, 2002). As the name suggests, the journal impact factor measures how much impact a journal has on the scientific community.

The derived metrics of bibliometrics are ideally counts of frequencies of events or the probability of occurrence (Borgman & Furner, 2002). Bibliometric distributions are formed based on the probability distribution and form the basis for bibliometric laws. Examples of bibliometric distributions are the Bradford, Lotka, and Zipf distributions of journals, authors, and words, respectively (Borgman & Furner, 2002). The following are few examples of bibliometric indicators of science that will be used in the current study:

1. The number of papers

Scientific output is reflected by the number of papers published by an academic (Okubo, 1997). The term paper refers to various scientific texts and includes books, journal articles, newspapers, reviews, reports, and other articles (Okubo, 1997). An approximate and simplified measure of scientists, laboratories, schools, teams, or countries' quantity of work is provided by paper counts (Okubo, 1997). The number of papers provides a more significant measure of scientific impact when compared to other metrics and is useful in monitoring trends in scientific impact (Okubo, 1997). This indicator would be used to understand the differences between the two broad categories (i.e., TED and Non-TED researchers) regarding scientific output.

2. The number of citations

The impact of the articles cited and their utility and timeliness can be measured by citations (Okubo, 1997). Authors cite the works of others when the earlier work acts as a foundation or to highlight the innovation in the article; the author acknowledges the earlier work (Okubo, 1997). Methods such as the ones mentioned above help determine the scientific output and the impact of researchers. In the current study, researchers will also be grouped based on whether or not they have given a TED talk.

3.2 Dataset

The current study is based on data related to TED talks and data related to publications and citations.

3.2.1 TED talks data

TED talks data is taken from Kaggle.com (Jr, 2020), which contains various TED speakers. The dataset contains information regarding 4,000 speakers and for 19 attributes and is available in multiple languages. The table below lists the various attributes that are present in the dataset, along with their definitions.

Table 1: TED Talk data attributes

No.	Attribute	Description
1	Talk_id	The identity number of the particular talk. This is unique to every TED talk.
2	Title	The title of the talk.
3	Speaker_1	The name of the TED talk presenter.
4	Speakers	Names of all the presenters if there are more than one.
5	Occupations	Occupation of the presenters
6	About_speakers	A brief description of the presenter.
7	Views	The number of times the talk has been viewed.
8	Recorded_date	The date when the talk was recorded.
9	Published_date	The date when the talk was published.
10	Event	TED includes specific events, such as TEDx and TEDMED. This attribute indicates the event this talk belongs to.

No.	Attribute	Description
11	Native_lang	The native language in which the TED talk took place.
12	Available_lang	All the languages that the TED talk episode is available in.
13	Comments	Comments on the episode of the TED talk
14	Duration	The duration of the talk.
15	Topics	Topics that are covered in the talk show.
16	Related_talks	TED talk episodes that are related to the current talk.
17	URL	Link to view the particular talk.
18	Description	Description of the TED talk
19	Transcript	Transcript of the TED talk.

The dataset consists of various speakers, some of whom are academics, CEOs, and entrepreneurs. Among 4000 speakers, 3300 speakers' profiles were considered (speakers who appeared on TEDed were excluded). The speakers were divided into two categories, academic and non-academic, and only academic speakers were considered. To achieve this, the profiles of every speaker on the TED.com website were thoroughly researched and compared with profiles on Google Scholar and other social networking websites, such as LinkedIn. Speakers who held academic positions when the talk was given, and speakers who were Doctor of Philosophy students, adjunct professors, and adjunct professors, were also classified as academic speakers. Speakers who did not fall into these groups were therefore categorized as non-academic speakers. Two hundred fifty-three (253) speakers were classified as academics and were considered for our study.

3.2.2 Citation and publications data

To answer questions regarding the number of citations received by a presenter, papers and articles that the presenter authors must be retrieved. The current study derives the citation and publication data from the Web of Science. Web of Science is a publisher-independent global citation database that contains every article and cited reference from all journals (Trusted publisher-independent citation database, 2021). These articles have been indexed, creating a comprehensive and complete citation network that aids in assessment and discovery for future studies (Trusted publisher-independent citation database, 2021) .

3.2.3 Author disambiguation

An anagram program was developed to retrieve all the papers that matched the speakers' first initial and last name. Various authors have identical names, and, at certain times, individual authors publish under multiple names, causing author ambiguity. To resolve author ambiguity, we follow the process to find all papers that match the TED speaker's name anagram, specifically, the last name and initial of the first name. This also included collecting co-authored papers (i.e., having more than one author) for which one of the authors matched the TED speaker anagram. Upon retrieval of the papers, the metadata such as first name, journal, and title were considered to identify the valid articles (i.e., authored by the TED speaker).

Two approaches were adopted to aid the process of matching TED talk presenters with Web of Science author publications. First, Google Scholar was used for identifying the profiles of the TED speakers. Then, the Google Scholar API was used to retrieve all publications from the TED speaker's Google Scholar profile. The next step was to match the retrieved Google Scholar publications with the Web of Science publications to identify relevant publication clusters containing the speaker's publication. Moving further, to remove the false positives, the list of publications is manually removed. Google Scholar accounts of some TED talk speakers could not be identified. In this scenario, accounts of all authors that matched the presenter's last name and first initial were identified, and all the publications in that account were retrieved. Furthermore, the publications were manually matched to those authored by the TED presenters. Google Scholar publications that matched the web of science publications were tagged as true positives (i.e., authored by the same TED speaker).

The second approach is to use an algorithm written by Caron and van Eck (2014). The algorithm creates clusters of related publications that are likely to belong to the same researcher. The algorithm consists of three parts: pre-processing, rule-based scoring and clustering, and post-processing. The following image depicts the workflow, and descriptions of each step are provided below.

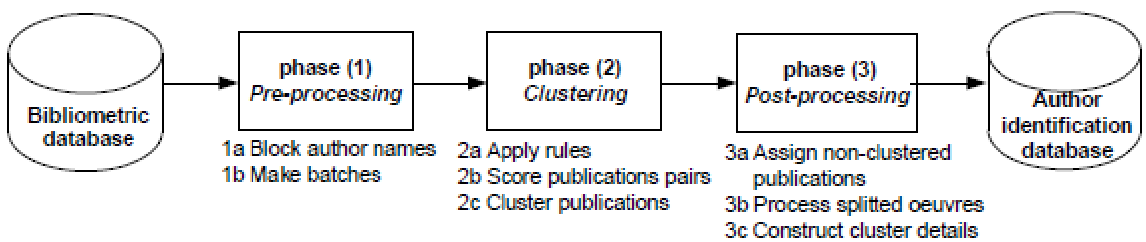

Figure 1:Author name disambiguation process (Caron & van Eck, 2014, pp 80)

Phase 1: Data Pre-Processing

During this phase, author name blocks are constructed based on last name and first name initial and by removing the non-alphabetic characters (Caron & van Eck, 2014). For example, the author name "Caron, E." would be placed with block Carone (Caron & van Eck, 2014). The author name blocks are further divided into classes of block size 1-6, depending on the number of publications within a block (Caron & van Eck, 2014).

Phase 2: Rule-based scoring and clustering

In this phase, publications within blocks are detected that are likely to be written by the same author based on a set of scoring rules (Caron & van Eck, 2014). The scoring rules are based on the idea that the higher the number of bibliographic elements between two publications, the more evidence they are written by the same author (Caron & van Eck, 2014). The categories used in this scoring are author, article, source and citation (Caron & van Eck, 2014). Table 2 gives the list of scores.

Table 2: Rules, scores and Threshold for block size 2 (Caron & van Eck, 2014, pp 81)

Category	Rule	Field	Criterion	Score
Author	1	email		100
	2a	all initials, more than one	two initials	5
	2b		more than two initials	10
	2c		Conflicting initials	-10
	3a	first name	General name	3
	3b		Non-general name	6
	4a	address (linked to author)	country, city	4
	4b		Country, city, org	7
	4c		Country, city, org, dep.	10
Article	5a	shared co-authors	one	4
	5b		two	7
	5c		More than two	10
	6	grant number		10
	7a	address (not linked to author)	country, city	2
	7b		country, city, org	5
	7c		country, city, org, dep.	8
	8a	subject category		3
	8b	journal		6
	9	self-citation		10
	10a	bib coupling	one	2
	10b		two	4

Category	Rule	Field	Criterion	Score
Citation	10c		three	6
	10d		four	8
	10e		more than four	10
	11a	co-citation	one	2
	11b		two	3
	11c		three	4
	11d		four	5
	11e		more than four	6
			Threshold	11

A publication pair can be defined as two publications that have scored on at least one of the scoring rules mentioned above. Once the publication pairs are scored, the pairs with a total score above the threshold are considered. This threshold depends on the block size (i.e., author name blocks created before). Finally, all matched publication pairs above a certain threshold are clustered using single-linkage clustering. The final cluster represents a part of an author's work (Caron & van Eck, 2014).

Phase 3: Post-processing

This step concerns the publications that are not clustered. These publications are labelled as separate clusters (Caron & van Eck, 2014). These are handled by using a correction procedure over generated clusters by matching email addresses between clusters (Caron & van Eck, 2014). In the end, matching publications are re-assigned to the larger cluster (Caron & van Eck, 2014). The above process will help identify the papers or publications written by individual authors and form a cluster. This process also helps in dealing with complexities such as popular names and hyper authorship. The above algorithm was run on the publications that were retrieved from the Web of Science. The next step would be to match the TED talk presenters with the authors of the publications.

The process mentioned above dealt with assigning the papers to TED speakers. Furthermore, to analyze the distribution of TED talks across different research fields, publication and citation data retrieved from Web of Science and stored in a relational database hosted by the Center for Science and Technology Studies (CWTS) at Leiden University was used. This was mainly used to classify different research areas and assign papers to them.

Classification systems within science are vital tools in bibliometric and scientometric studies, which often allots individual publications or journals to research areas (Waltman & Eck, 2012). The CWTS classification system refers to the classification system of science proposed by Waltman and Eck (2012). In this method, publications are clustered based on citation relations, where each publication is assigned to a single research area (Waltman & Eck, 2012). In a hierarchical structure, research areas begin broad and contain subfields, allowing for the classification of many publications (Waltman & Eck, 2012). Further, indicators need to be field normalized to make meaningful comparisons between researchers from different fields.

Non- TED talks researchers.

Upon retrieving papers belonging to researchers who have given a TED talk, the socio-demographic information and publications of the remaining researchers *who belong to the same cluster* in the CWTS WOS database were retrieved. These subjects comprised the non-TED talk speaker group. For selecting the authors for the non-TED category from the clusters, a threshold of 5 was used, and researchers with less than five paper publications were excluded. Certain bibliometric studies, such as the one conducted by (Nielsen & Andersen, 2021), mentions that the author-disambiguation process or algorithm is more reliable above this threshold value. And hence, we excluded researchers who had less than five paper publications.

3.2.4 Field delineation

The publications are matched to the five main fields that are present on web of science. These fields are Biomedical and health sciences, life and earth sciences, mathematics and computer science, physical sciences and engineering and social sciences and humanities. Each publication of a university belongs to one, or more main fields and when a publication belongs to more than one main field, the publication is assigned fractionally to each of the main fields (Waltman & Eck, 2012) Web of science assigns publications into main fields algorithmically (Waltman & Eck, 2012) . The following are steps that are used:

1. 4140 micro-fields of science are constructed algorithmically and each web of science publication is assigned to one 4140 fields (Waltman & Eck, 2012). This assignment is based on large scale analysis of hundreds of millions of citation relations between publications (Waltman & Eck, 2012).

2. Further, 254 journal subject categories defined in web of science that these 4140 micro-fields overlap are determined (Waltman & Eck, 2012).

3. Each subject category in web of science is linked to one of the five main fields (Waltman & Eck, 2012) . Each of the 4140 micro- fields are assigned to one or more of the five main fields based on the link between subject categories and the main fields (Waltman & Eck, 2012). If atleast 25% of the publications in the micro-level field

belong to subject category linked in the main field then the micro-field is linked to the main field (Waltman & Eck, 2012).

Further, each publication in web of science has an assignment to a micro-field level, and each micro-field in turn has an assignment to one or more main field (Waltman & Eck, 2012) . With this each publication, is assigned to one or more main fields (Waltman & Eck, 2012) . Depending on which main field the publication isassigned to, the author of the publication is considered to belong to that main field for this study (Waltman & Eck, 2012) . The image below shows the micro-fields and the size of the circle determines the number of publications in that field. The color determines the main field to which the micro-field is assigned to. Micro-fields assigned to biomedical and health sciences are in green, life and earth sciences are in yellow, mathematics and computer science are in purple, physical sciences and engineering are in blue and social sciences and humanities are in red.

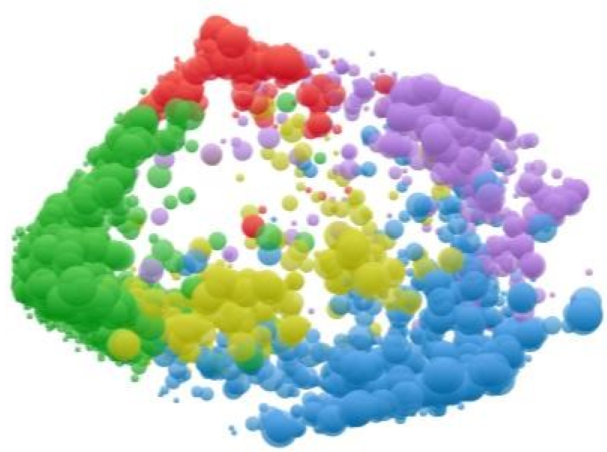

Figure 2: Micro-fields (Centre for Science and Technology Studies (CWTS),n.d)

The below table gives the total number of researchers across the main fields in both TED and non-Ted categories. It can be seen that when the number of researchers in each field are added exceeds the total number of researchers. This is because one researcher might belong to more than one research field and add to the count in both fields.

Table 3: Number of researchers by field and overall

	TED	**Non-TED**
Physical Sciences and Engineering	78	5347
Life and earth sciences	28	8084
Mathematics and computer science	39	11653
Social sciences and Humanities	26	23912
Biomedical and health sciences	82	32489
Total number of researchers	245	81485

3.3 Analysis

3.3.1 Descriptive Analysis

This study aims to understand different characteristics, such as gender, ethnicity, and academic age of academic researchers, of academic presenters who either have given a TED talk or have not given a TED talk but are matched for the research area. The following are the methodologies adopted to explore the presenter characteristics.

Gender

Subjects' profiles on different social media websites like LinkedIn, TED speaker profiles and simple google search were primarily used to gather data regarding the gender of the speakers, and gender was coded as M (male), F (Female) and U (Unisex). Additionally, the GenderChecker (Genderchecker,n.d.) dataset was used to record the gender based on the individuals' first names. Further, the gender data will be visually depicted in the form of a bar chart.

Organization

Presenters' organizational affiliations were also recorded. For analyzing this variable, the total number of presenters for various organizations was calculated. This data will be depicted using a bar graph. Organizations that had less than five presenters were grouped into one category.

Academic age

To analyze the academic age of different presenters, the year when presenters published their first and last articles were considered. The difference between these two years was calculated, and academic age was categorized based on multiples of ten to know the number of presenters who fall in each category.

Country of affiliation

Certain presenters are affiliated with more than one organization, which might be present in different countries. These organizations were termed alternate organizations. The country where the alternate organizations were present were considered for analyzing the country of affiliation. The total number of affiliations country-wise was plotted.

University Ranking

The organizations or universities with which the presenters are associated are compared with the ranking made available from Leiden Ranking. In addition, the recent ranking of the university was taken to categorize the number of presenters that belonged to top-ranked universities, which were again divided by multiples of five.

Research output and impact

Table 4: Research output and indicator description

Indicator	Abbreviation	Description
Citation Score	CS	Number of citations received by a publication
Normalized Citation Score	NCS	Number of citations field average for the same publication year
X% top-cited publication in the field	P(top 10%)	Dichotomous variable indicating that a paper is among
Number of publications	N_pubs	Number of publications by the researcher

Note: Among different types of indicators made available by the Center for Science and Technology Studies (CWTS) at Leiden University, scientific impact indicators normalized by micro fields are used to measure the impact made by researchers who gave a TED talk compared to those who did not. The table above shows the list of indicators that will be used in this study.

For this study, the researchers in both groups (i.e., academic presenters who have given a TED talk and who have not given a TED talk) are further grouped based on the research area of the presenter. The average of every indicator mentioned above is calculated for all researchers, as the dataset contains one or more publications for all researchers. Further, this data is used to compare the two groups. Questions of interest include: (1) What is the average number of researchers' publications in each research area for both groups? (2) What is the average normalized citation score for each research cluster for both groups?

3.3.2 Predictive Analysis

A binomial logistic regression model will be used to predict whether or not a researcher might appear on TED talk, depending on their demographic characteristics, such as gender, race, and academic age. “A binomial logistic regression attempts to predict the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical” (*Laerd Statistics*, n.d). The binomial logistic regression model will be conducted using SPSS. The requirements for completing a binomial logistic regression are as follows (*Laerd Statistics*, n.d):

1. One dependent variable: Binomial logistic regression models require one dichotomous dependent variable (i.e., a nominal variable with two outcomes). For this study, appearing to TED talks is the dependent variable and has two outcomes: “Yes” or “No”.
2. One or more independent variables: Binomial logistic regression models require independent variables to be measured on either a continuous or nominal scale. Different academic presenters, such as gender, race, academic age, research area, and country of origin, are independent variables for this study, as they might predict the dependent variable.

“Binomial logistic regression is a part of a larger statistical group of tests called Generalized Linear Models (GzLM)” (*Laerd Statistics*, n.d). It allows a relationship to be modelled between multiple independent variables and predicts a single dichotomous variable (*Laerd Statistics*, n.d). Additionally, a transformation is applied to predict the logit of the dependent variable instead of the category of the binomial logistic regression directly (*Laerd Statistics*, n.d).

For example, if we consider four independent variables to be "X1" through "X4", and the dependent variable to be "Y", the formula of the binomial logistic regression model is (*Laerd Statistics*, n.d):

$$Logit(Y)=\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \varepsilon^5$$

Where β_0 is the intercept or constant, β_1 is the slope parameter or slope coefficient for X1, and so forth, and ε represents the errors.

Upon deciding the dependent and independent variables, the binomial logistic regression model is run using the data. Further, the results are used to identify outliers. This table highlights the cases with standardized residuals (i.e., ZResid greater than +/- 2 standard deviations) (*Laerd Statistics*, n.d). Cases with residual values greater than 2.5 would be inspected further to determine whether to remove them from the analysis if need be.

The probability of an event occurring is estimated by binomial logistic regression. If the likelihood of the event occurring is more significant than 0.5, the event is classified as occurring, whereas events with a probability less than 0.5 are classified as not occurring. The observed and predicted classifications are present in the classification table. From the classification table, we also determine the values of percentage accuracy in classification (PAC), sensitivity, specificity, positive predictive value, and negative predictive values, which are described below (*Laerd Statistics*, n.d):

1. PAC: To improve the overall prediction of cases into their observed categories of the dependent variable, independent variables are added. This measure is referred to as percentage accuracy in classification (PAC).

2. Sensitivity: The percentage of cases that had the observed characteristic that the model correctly predicted.
3. Specificity: The percentage of cases that did not have the observed characteristic and were correctly classified by the model as not having the characteristic.
4. Positive predictive value: The percentage of correctly predicted cases with the observed characteristic compared to the total number of cases predicted as having the characteristic.
5. Negative predictive value: The percentage of correctly predictive cases without the observed characteristic compared to the total number of cases predicted as not having the characteristic.

These measures help assess the ability of the binomial regression model to classify cases correctly (*Laerd Statistics*, n.d). The classification table mentions the cut-off point used to classify the cases with the characteristic of interest (*Laerd Statistics*, n.d). A higher cut-off point will increase specificity but lower sensitivity (*Laerd Statistics*, n.d). Instead of having one cut-off point, it is possible to include many cut-off points. When multiple cut-off points are used, a visual representation is generated, referred to as a receiver operating characteristic (ROC) curve, a plot of sensitivity versus specificity (*Laerd Statistics*, n.d).

Interpreting the ROC curve

The value of the area under the ROC curve is described as the *area* under the curve table. The area under the ROC concordance probability is the most common measure of the ability of a generalized linear model to discriminate (*Laerd Statistics*, n.d).

The area can range from 0.5 to 1.0, with higher values representing better discrimination (*Laerd Statistics*, n.d). The rules of thumb for the area under the curve are as follows:

Table 5:Values for interpreting the area under the curve (Web of Science, n.d)

Area under the curve (AUC)	**Classification**
0.5	NO discrimination
0.5<AUC<0.7	Poor discrimination
0.7<= AUC<0.8	Acceptable discrimination
0.8<=AUC<0.9	Excellent discrimination
AUC >=0.9	Outstanding discrimination

To see the contribution of each independent variable to the model and their statistical significance, we will use the variables in the equation table. Below are descriptions of the tests and statistics referenced in the study (*Laerd Statistics*, n.d).

1. Wald test: The Wald test is used to determine statistical significance for each of the independent variables.
2. Sig (p-value): Indicates the statistical significance of the test.
3. *B* coefficients: "These are used to predict the probability of an event occurring, but not in an intuitive manner" (*Laerd Statistics*, n.d).

The coefficients show the change in the log odds for a one-unit change in an independent variable when all other independent variables are kept constant.

CHAPTER 4 RESULTS

This chapter presents the results of the analysis. The results are divided into three sections based on the three main research questions. The first section describes the various socio-demographic characteristics of TED presenters compared to the population of researchers active in the same research areas. The second section focuses on the research impact of both categories of researchers. The third section presents the logistic regression model aimed at identifying the predictors of TED appearances.

4.1 Descriptive Analysis

The following are the description and visualization of various socio-demographic characteristics of the researchers from both groups (i.e., TED and Non-Ted researchers), divided into five broad research fields. The socio-demographic considered are Gender, University ranking (according to Leiden Ranking) etc.

4.1.1 Gender

Figure 2 presents the distribution of gender by gender for TED and non-TED researchers. The results show that, in all research fields, men are typically more likely than women to present at TED. The data also show that Male researchers are over-represented in TED talks, and there are no significant differences in the distribution of gender by fields. Nearly 50% of researchers in both categories across all fields are male. The results also show that in biomedical and health sciences and life and earth sciences the proportion of women in both TED and non-TED categories are almost the same. This could be due a greater involvement of women in these areas (*The stem gap: Women and girls in science, technology, engineering and math,* 2020). Further, in mathematics and computer science field, the percentage of men and women in TED category is almost same. It's interesting to see that the representation of women in TED category is more compared to non-TED which might indicate that women in mathematics and computer science field which can indicate that in one way women in this area are being more encouraged.

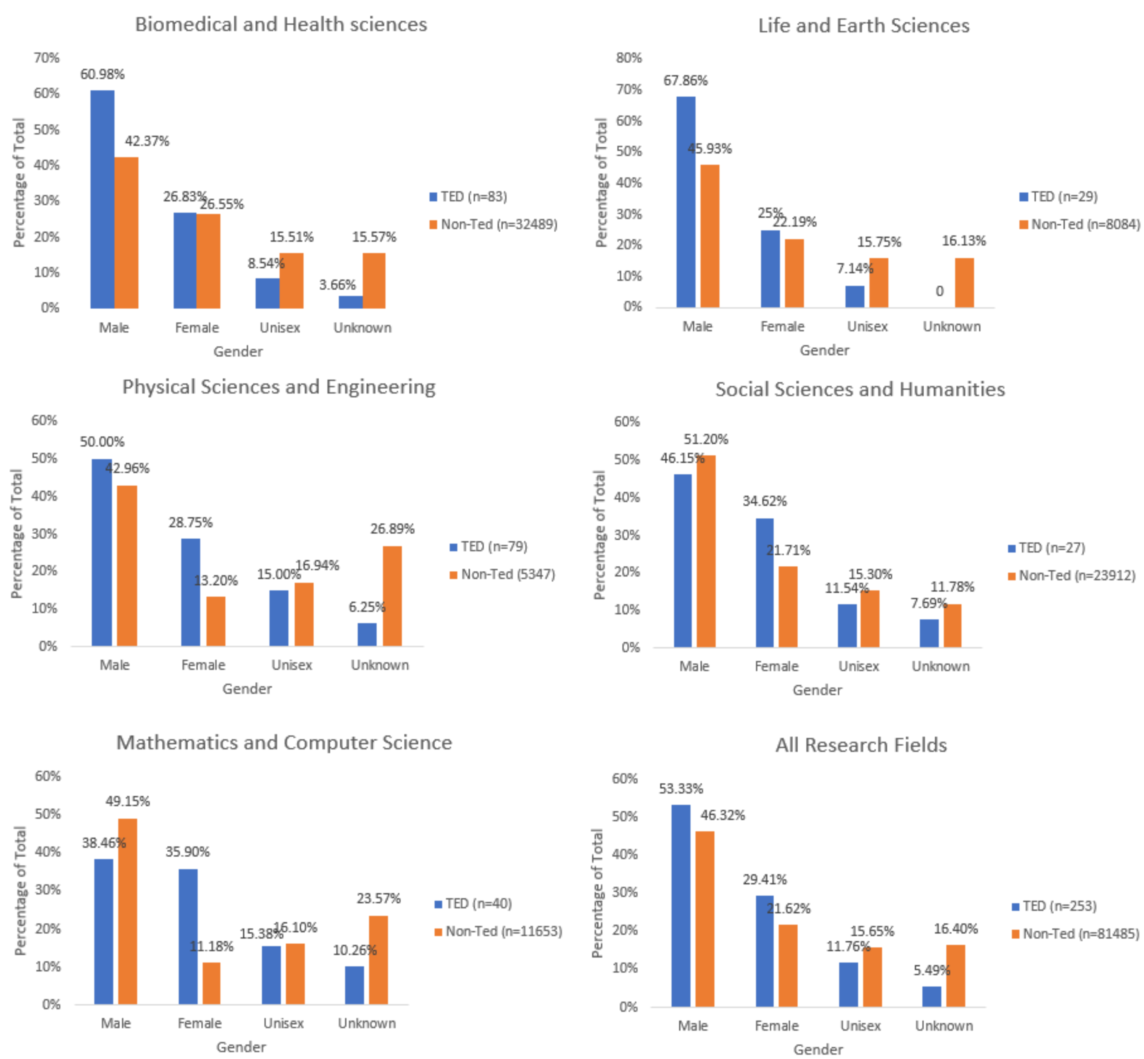

Figure 3: Total percentage of each gender in TED and Non-TED researchers category

4.1.2 Academic Age

Figure 3 displays the academic age distribution for researchers of both groups. Nearly 30 to 60% of TED researchers fall into category 0-9 years of experience. And also, 35 to

50% of Non-Ted researchers fall into this category. We find that the minority of researchers fall into the category of 30+ years of experience for all the research fields for TED and Non-TED researchers. Additionally, the percentage of researcher distribution based on academic age for both TED and Non-TED researchers categories across all research fields is identical.

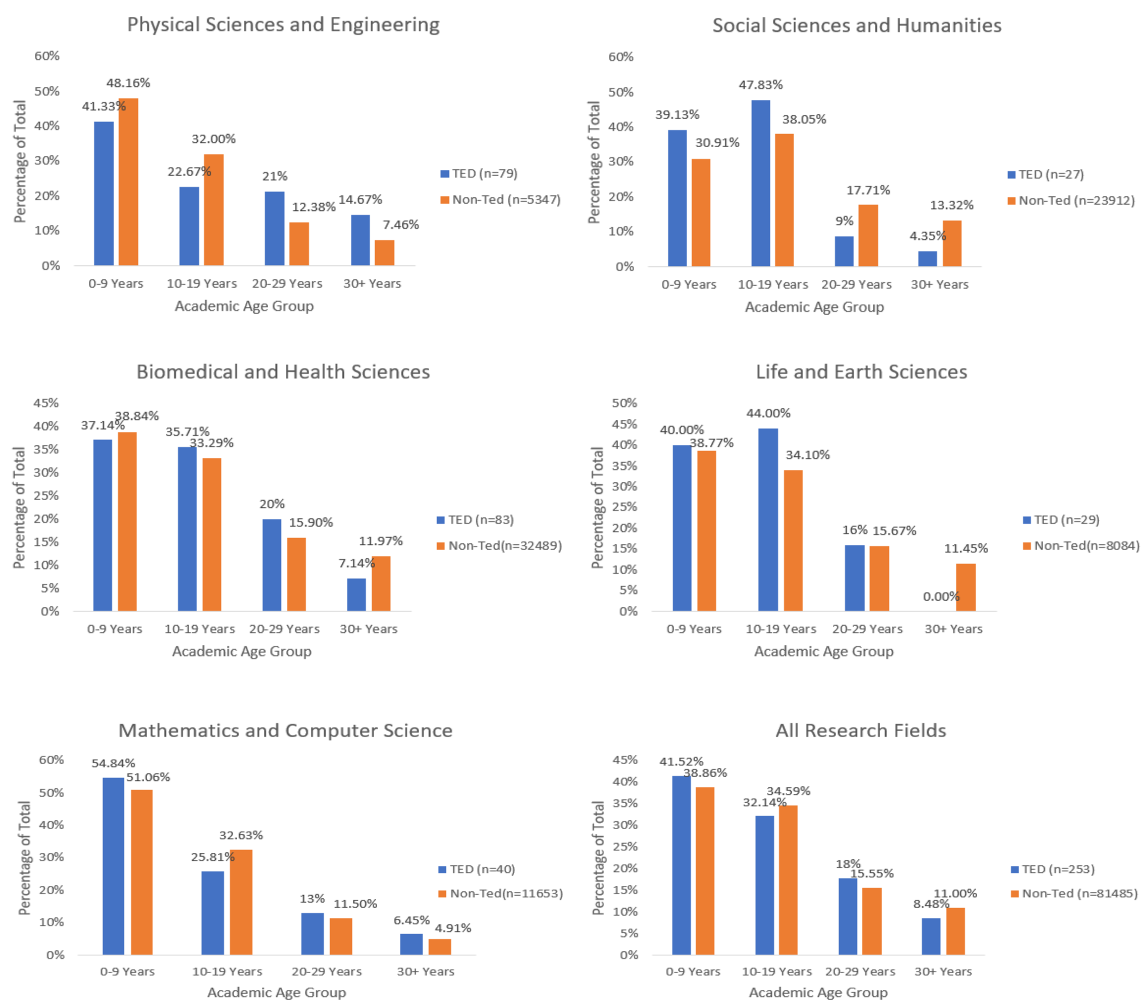

Figure 4:Total percentage of each academic age group in TED and Non-Ted researchers category

The exception being Life and earth sciences, where there are more TED speakers in the 10-19 years category (10-19 years - 44%, 0-9 years- 40%) compared to their Non-Ted counterparts (10-19 years – 34.10%, 0-9 years- 38.77%). Also, there are no TED researchers in 30+ years in the Life and Earth field, whereas 11.45% of Non-TED researchers. Furthermore, there seems to be an inclination in the 10-19 years academic age group for both TED and Non-TED researchers categories in Social Sciences and Humanities research field. An exception to this is TED speakers of the Social Science and Humanities, where 53.84% belong to the 10-19 years category, and 46.15% belong to the 0-9 years category.

4.1.3 University Ranking

The bar graphs below show the picturization of organizations to which researchers are affiliated.

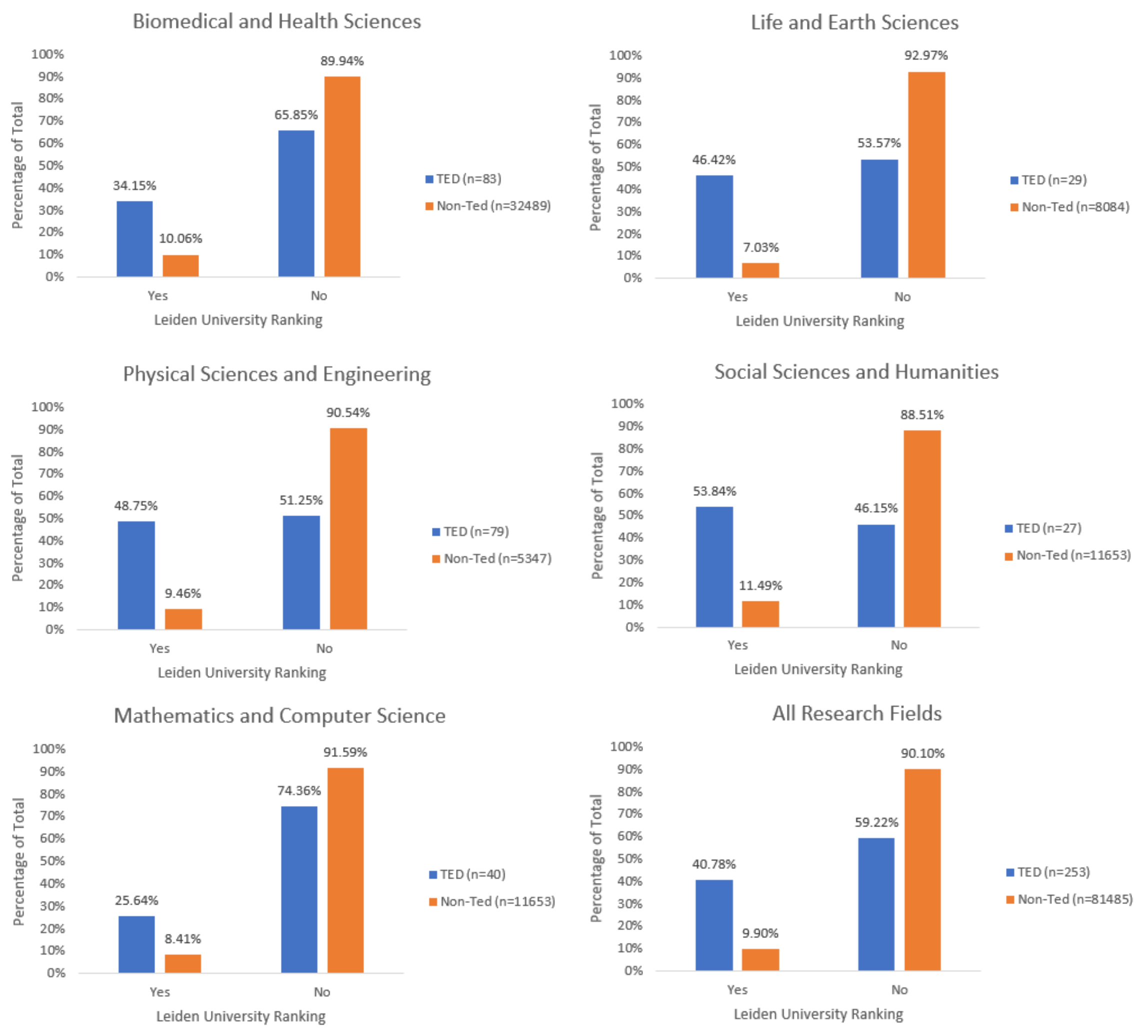

Figure 5:Total percentage of Universities appearing in top 100 rankings in Leiden University ranking for both TED and Non-Ted researchers category

According to Leiden University rankings, these organizations are classified based on whether or not they appear in the top 100 universities. Based on the visualizations above, researchers from the top 100 universities are over-represented in TED talks: they range from 25% to 55% for all the research fields. On the other hand, only 10% of the researchers belong to the Top 100 universities in the Non-Ted category. Also, more researchers belong to universities that are not in the top 100 universities for both TED and Non-TED researchers for all research fields.

4.1.4 Country

By analyzing the data of country where researcher's universities are located, i.e., country of affiliation, most researchers are from the United States than other countries for both Ted and non-ted researchers categories. Therefore, the top ten countries of each research field for both TED and Non-TED categories are shown above, except for some TED research fields due to the number of academic researchers in those fields. Also, more than 60% of TED researchers are from the United States and are over-represented compared to other countries.

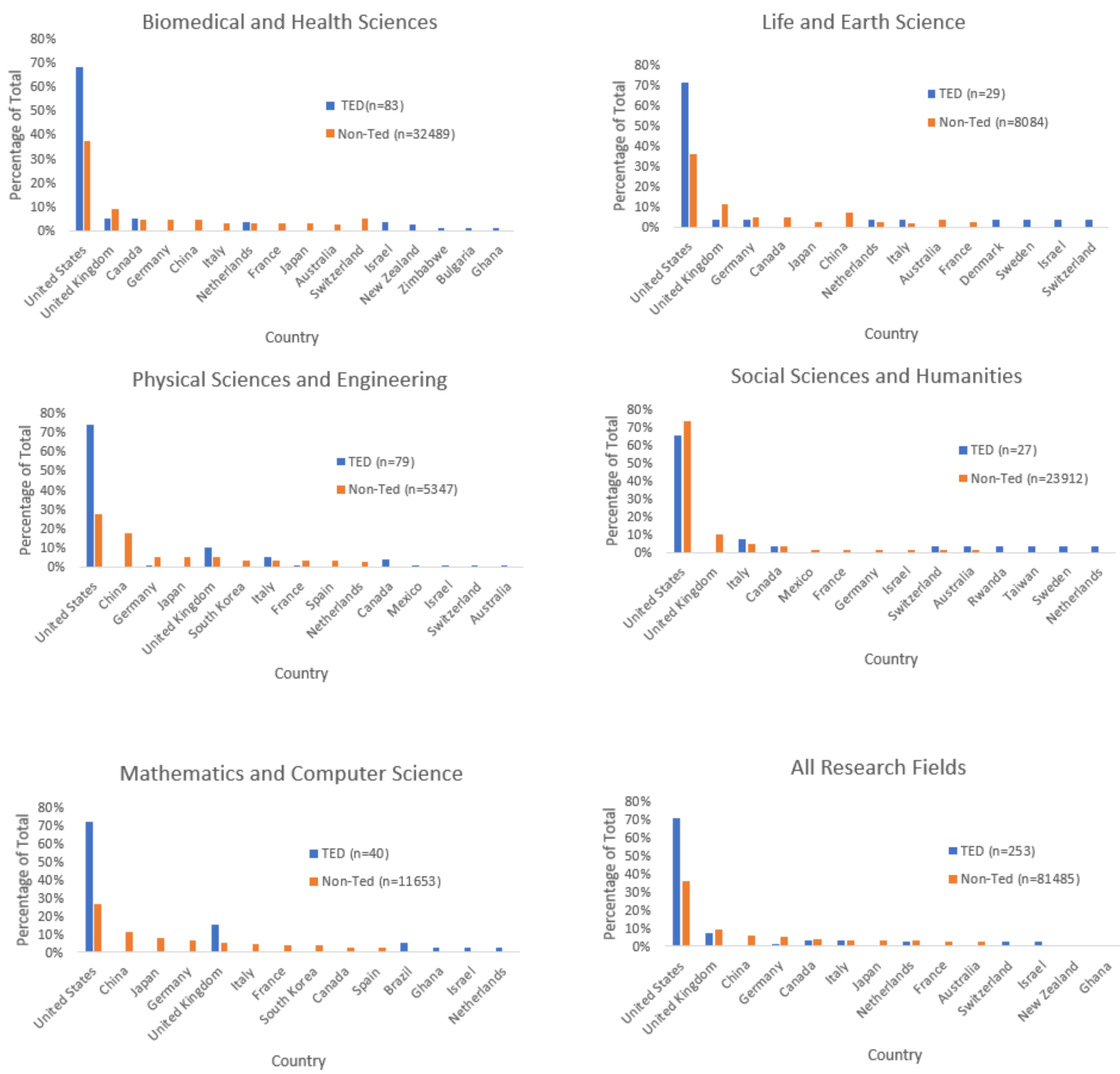

Figure 6:Total percentage of top 10 countries of affiliation of TED and Non-TED researchers.

4.1.5 Organization

The bar graphs show the top 10 organizations based on the percentage of researchers from various organizations. It can be seen that the list does not vary much between the research fields or the categories of researchers. Also, it can be observed that more researchers appear from universities (e.g., Harvard, MIT, Yale) for almost all research fields. On the other hand, we can see that most of the top 10 universities for the TED and

Non-TED categories are different in each research field. However, few organizations were common for both TED and Non-TED categories except for Social Sciences and Humanities research field.

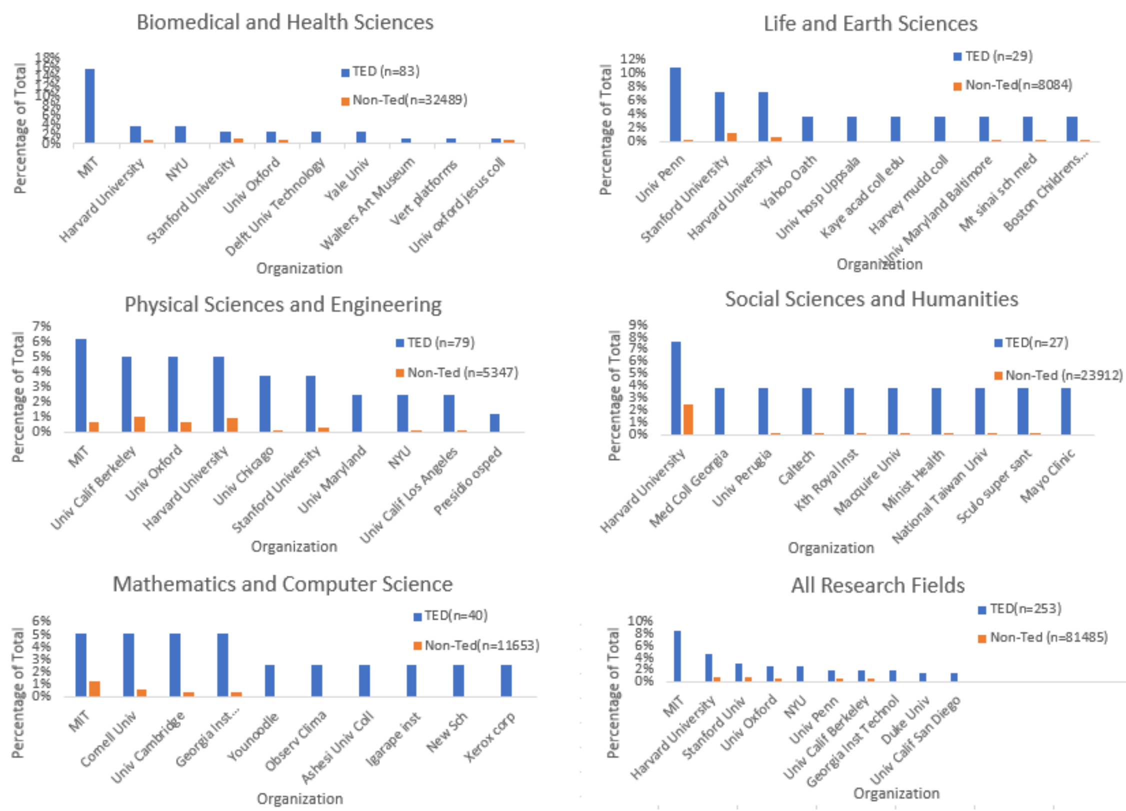

Figure 7:Top 10 organizations of TED researchers and presence of top 10 Non-TED researchers' organization in it.

4.2 Research output and impact indicators

4.2.1 Research Output

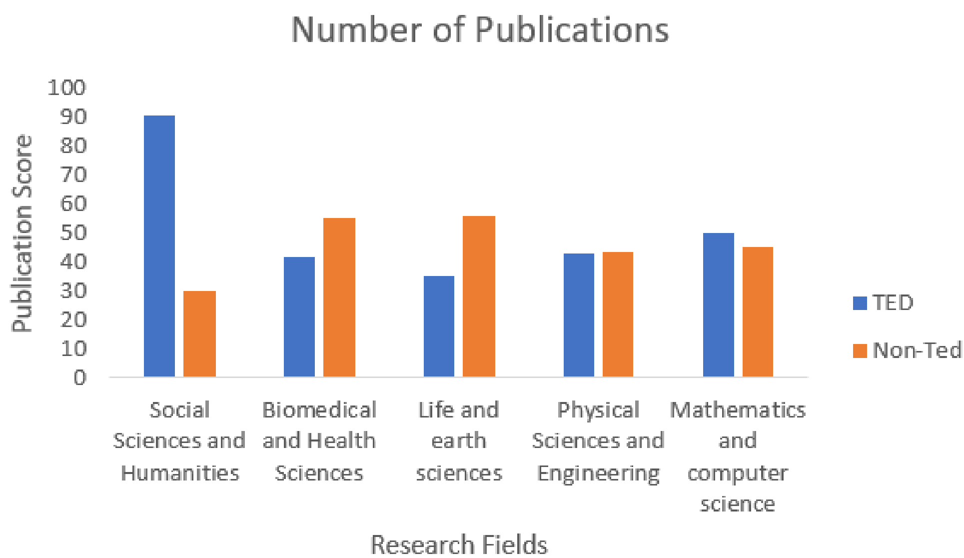

Figure 8:Average number of publications of TED and Non-Ted researchers across different research fields

We use the number of publications to measure the research output of researchers. As shown in Figure 7, TED presenters are, on average, published more papers than their non-TED colleagues in Social sciences and Humanities (90 vs 29), Physical Sciences and Engineering (43 vs 43) and Mathematics and Computer Science (50 vs 45). However, an exception to this is Biomedical and health sciences and Life and earth sciences, where the number of publications of TED researchers are 41 and 35, respectively. On the other hand, the number of publications for non-TED researchers of the same fields is 55 and 56, respectively.

4.2.2 Impact indicators

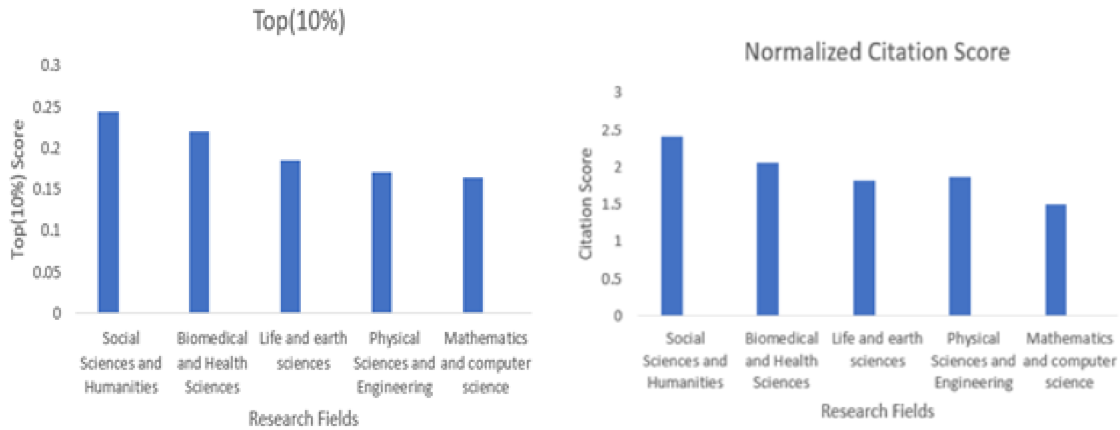

Figure 9:Normalized citation score and Top 10% most cited publication of TED researchers across different research fields.

As proxies for research impact, we use the citation score, the normalized citation score, and the share of publications in the top 10% most cited in the field. The above graphs show normalized citation scores and Top 10 % of most-cited publications in each field for TED researchers. From the graphs, it can be seen that the citation score and Top 10% values are higher for Social Science and Humanities and Biomedical and Health Sciences research fields when compared to Life and earth sciences, Physical Sciences and Engineering and Mathematics and Computer Science research fields. However, we can conclude that, on average, researchers who appeared in TED talks had more citations, publications, and papers that appear under the top(10%) in that particular research field.

4.3 Logistic regression

4.3.1 Category Prediction

We used Binomial logistic regression to determine whether TED participation can be correctly predicted using our set of socio-demographic variables. Laerd SPSS statistic tool was referred to for the interpretation of the results generated by the logistic regression (*Laerd Statistics, n.d.*). Table 5 shows the effectiveness of the predicted classification against the actual classification. The probability of an event occurring is estimated by Binomial logistic regression. If the estimated probability of the event occurring is greater than or equal to 0.5, then the event is classified as occurring (i.e., appearance on TED Talks). Conversely, if the probability is less than 0.5, then the event is classified as not occurring (i.e., no appearance on TED talk).

There are fewer TED researchers than Non-TED researchers in our dataset, which leads to an unbalanced dataset. Hence, the dataset was oversampled by using SMOTE library in python, which uses synthetic methods to increase the number of cases for the minority dependent variable category to make it a balanced dataset.

Table 6:Classification Table

	Observed	**TED Appearance**		**Percentage Correct**
		No	**Yes**	
TED Appearance	No	51588	16463	75.8
	Yes	17514	50537	74.3
Overall Percentage				75.0

Note: The cut value is .500

It must be noted that the cut-off value for this test is 0.500, which means that if the probability of a case being classified into Yes is more than 0.500, then that particular case is classified into the "Yes" category. Otherwise, it's categorized into the "No" category. The overall percentage measure is called percentage accuracy in classification (PAC), 75% in this case. The table also gives the information about the specificity, i.e., the percentage that did not have the observed value, in this case, 75,8%. Sensitivity is the percentage of cases that had the observed value (74.3%). Positive predictive value (100(50537/16463+50537)= 75.42%) and Negative predictive value i(100(51588/51588+17514)= 74.65%). 75.8% of the researchers who did not appear on TED talk were predicted correctly by the model as “No appearance on TED talk.” Further, the table also provides details about sensitivity which gives the cases of those who appeared on TED talk. In this case, 74.3% who appeared on TED talks were correctly categorized by the model as appeared on TED talk.

The contribution of each independent variable to the model and its statistical significance is given by the variables in the equation table. The Wald test column from the table is used to determine the independent variables’ statistical significance. The significance is given in the sig column. It is seen that the significance value is for all the variables is <=0.05. Wald test is another name for the Chi-square test. Therefore, P <= 0.05 is considered to be of statistical significance. In this case, all the variables are statistically significant. The variables in the equation table generated by SPSS provide both B coefficients and the odds ratio. The odds ratio shows the increase in odds for every unit increase for all the variables. A value is below 1 indicates decreased odds for an increase in one unit of the independent variable.

From the table, it can be seen that Biomedical and health sciences are taken as the base for research fields, and for every unit change, the odds are given in the table in the odds ratio column. The highest odds of appearing on TED talks is seen for Physical sciences and engineering, 5.977 times more likely than researchers from Health Sciences.

Likewise, the chances of appearing in TED talk reduces if the researcher is from Social Sciences and Humanities (odds ratio = 0.183).

When it comes to gender, it can be seen that the odds of appearing on TED talks for Female and Unisex name researchers reduces by 0.843 and 0.328, respectively. Also, the organization's presence under Top 100 in Leiden University ranking increases the odds by 3.652. Further, it is also seen that the odds of TED appearance decreases by 0.308, 0.610 and 0.583 respectively for academic age groups 10-19 years, 20-29 years and 30+ years compared to 0-9 years academic age group. Speaking of Continents which gives the location of the researchers, it is seen that the odds of appearing on TED talk reduces for all the continents except for North America. Researchers with the lowest odds of appearing on TED talks are from Asia (odds ratio of 0.135) and Australia (odds ratio of 0.224).

Table 7:Variables in the equation

95% C.I. for Odds Ratio								
	B	**S.E**	**Wald**	**df**	**Sig**	**Odds Ratio**	**Lower**	**Upper**
Biomedical and health sciences			16368.531	4	.000			
Life and earth science	.324	.022	219.899	1	<.001	1.383	1.325	1.443
Mathematics and cs	.385	.020	386.747	1	<.001	1.469	1.414	1.527
Physical sciences and engineering	1.788	.022	6572.668	1	.000	5.977	5.724	6.241
Social sciences and Humanities	-1.696	.022	6031.620	1	.000	.183	.176	.191

95% C.I. for Odds Ratio								
	B	**S.E**	**Wald**	**df**	**Sig**	**Odds Ratio**	**Lower**	**Upper**
Male			2756.457	2	.000			
Female	-.171	.016	114.513	1	<.001	.843	.817	.870
Unisex	-1.114	.021	2746.517	1	.000	.328	.315	.342
Leiden Score(1)	1.295	.018	4981.444	1	.000	3.652	3.523	3.785
0-9 years			4578.802	3	.000			
10-19 years	-1.178	.017	4556.397	1	.000	.308	.297	.319
20-29 years	-.494	.020	620.723	1	<.001	.610	.587	.634
30+ years	-.540	.023	568.398	1	<.001	.583	.558	.609
North America			11007.430	5	.000			
Europe	-1.435	.016	8017.774	1	.000	.238	.231	.246
Asia	-2.003	.030	4481.511	1	.000	.135	.127	.143
Aus-NZ	-1.494	.051	869.161	1	<.001	.224	.203	.248
South America	-1.086	.056	373.201	1	<.001	.338	.302	.377
Africa	-.478	.054	78.630	1	<.001	.620	.558	.689
Constant	.847	.017	2607.471	1	.000	2.332		

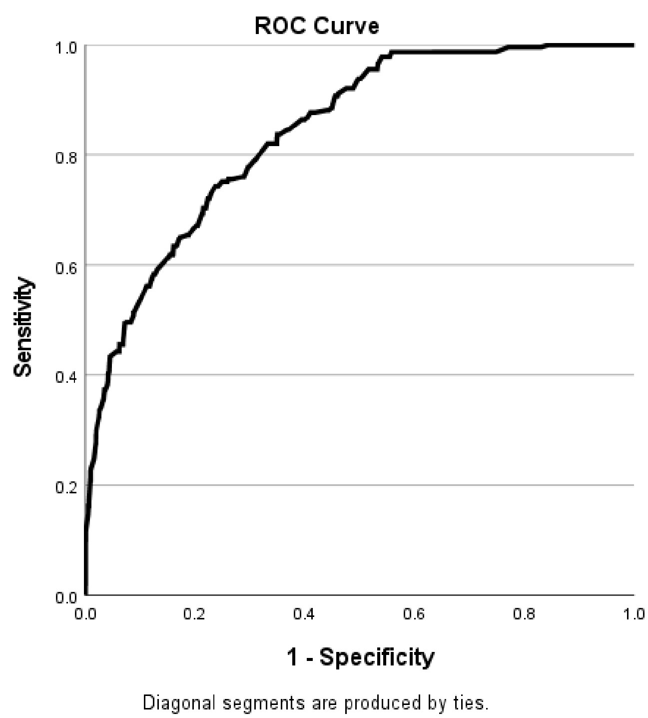

Figure 10:ROC Curve

The area under the curve is 0.842. Hosmer et al. (2013) say that with values between 0.8 and 0.9, the model is considered to have excellent discrimination.

Table 8:Model Summary

Step 1	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	132072.615[a]	.340	.454

[a] Estimation terminated at iteration number 5 because parameter estimates changed by less than 0.001

Table 9:Hosmer and Lemeshow Test

Chi-square	**df**	**Sig.**
3126.973	8	.000

Table 9 shows that the model is poorly fit ($p < 0.001$).

CHAPTER 5 DISCUSSION

5.1 Summary of findings

Looking into the socio-demographic attributes of both TED and Non-Ted categories across all the research fields that were taken into consideration, it was seen that there are more male researchers than their female counterparts. These results fall in line with O'Brien et al. (2019) findings that despite the number of years of women involved in research and science, they receive lesser recognition and opportunities. Furthermore, it can be seen that male researchers are over-represented in TED talks compared to female researchers. An interesting aspect to remember is that there are various other aspects such as Mother of Conflicts- work home balance affecting women's involvement in science fields such as research and leadership. With this, it becomes difficult to identify whether lesser female representation in TED talks is due to under-representation because of male researchers getting more prominence or fewer female researchers present in those fields.

Speaking of organizations and country of affiliation of researchers, relating to Mathew's effect, Merton (1968) mentions that scientists from major universities received more recognition when compared to scientists from lesser-known universities. Also, the Mathew effect mainly refers to over recognition of the scientists at the peak of their profession(Merton,1968). Relating the above to our findings surrounding organization and country of affiliation of researchers, we can see that most researchers for both categories across all research fields are from the United States and followed by the United Kingdom for some fields. Also, most researchers were from universities or organizations such as MIT, Harvard, Stanford, especially TED researchers. Additionally, most of these universities are located in North America. Thereby, we can say that the Ted talk gives more platform for researchers from top universities, especially the ones in North America. Another reason for the over-representation of researchers from the United States might be the origin of the TED talk show in the United States. The above explanation holds good even for University ranking (Leiden Ranking) as the proportion of TED talk researchers from the top 100 universities in the world is more than Non-Ted research counterparts in each field.

Bol et al. (2018) say that receiving funds at an early stage in a career also acts as an advantage for researchers. This might also be one reason most researchers in both categories are from 0-9 years of academic age. However, this can also suggest that only the most motivated researchers might have continued in the field, leading to a decrease in the percentage of researchers with an increase in years of experience.

Further from seeing the values of bibliometric indicators used in the study, it was seen that, on average, TED researchers had more citation and publications. compared to non-Ted researchers. Therefore, scientists whose research was predominantly better known to the scientific community and the general public were invited to TED talks. This can be related to the Mathew effect (Merton, 1968) scientists. At the peak of their career receive over recognition. Also, the accumulation of advantages were scientists who had a previous reputation (i.e., citations, publications), charisma, well-placed positions had more chances of accumulating more fame (Merton, 1968). Also, the appearance on the TED talk platform would further enhance the chances of receiving grants as the reviewer would also consider past achievements of applicants as criteria (Bol et al., 2018).

Researchers are involved with the scientific community by informing them about various research activities in the form of publications. They are also engaging with the public on various social media platforms such as Twitter. According to Shugrat & Racaniello (2015), researchers' careers positively impact researchers' careers by engaging with the general public. This could be one of the reasons why researchers would like to appear on the TED talk platform. From our study, on average, researchers who appeared on TED talk had more citations and publications than those who did not.

Further, scientists use such platforms to increase visibility which additionally helps them receive more funds or citations. Another example would be funding organizations like NSF, which has begun to include scientific impact by science communication for grant approval (Ngumbi,2020). Hence, we can say that the general public is kept informed of the recent happenings in research, and scientists benefit variously through scientific Communication. Ted is an abbreviation for Technology, Education and Design, where experts provide insights into their work. This would assist in promoting scientific impact in the case of researchers (Liang et al., 2014). Therefore, attracting more and more researchers to participate and appear on the TED talks platform.

When it comes to our study's predictive analysis, which considered various independent variables and predicted the appearance of academic researchers on TED talks, our results once again bring us to a similar conclusion as we found in our descriptive analysis section. As mentioned before, a binomial logistic regression model was being used to predict the appearance of researchers in TED talks by using Academic age, gender, Organization, Leiden University ranking and continent as the independent variables. Our results are mentioned in how the chances of appearing on TED talks increase or decrease with a change in our nominal independent variable. It is seen that the chances of appearing on TED talk reduce if the researcher is from any other continent apart from the

North American continent, and according to our descriptive analysis, the United States is over-represented in TED talks. Likewise, it is seen that our model predicts that the chances of appearing on TED talk increase with the researcher's background where they are from top tier universities, presence of their university in top 100 lists in the world, being a male professor and also the research stream to some extent. Thus, directing us towards the Mathew effect in science and scientific communication.

5.2 Limitations

The study involves researchers whose gender was not determined, this is one of the limitations as we don't know these unknowns are more or less equaly divided between men and women. Our results could be affected if this is not the case.. Also, academic age was being calculated by taking into consideration today's year and the year the researcher first published. This was performed to achieve consistency while calculating the academic age of TED and non-Ted researchers. Whereas, for TED researchers the academic age should have been calculated by using the year they appeared on TED talk and the year of their first publication.

The study makes use of certain bibliometric indicators such as citation score. There are limitations for using citation score as an performance indicator. An example to these limitations would be the coverage of database that's used and the reference pattern used (Aksnes, 2019). And in fields like social sciences and humanities, publishing in books is more common than international journals (Aksnes, 2019).

5.3 Future Research

From the works of Ray (2018), it is known that there is racial bias in citations in certain disciplines such as law and philosophy. However, the study does not include the race of researchers. It would be interesting to see the description of the race of researchers across different research fields and to know how this variable relates to the likelihood that researchers appear on TED talks. Additionally, this study mainly focused on comparing the TED and Non-TED researchers and understanding independent variables' role. However, future research can aim at understanding the impact of appearing on TED talk platforms on researchers, for example, determining if there have been any significant changes in the citation score and publication score for researchers after appearing on the TED platform.

Furthermore, along with the TED talk platform, other social media platforms such as Twitter and YouTube. can be taken into consideration. For example, a study can be designed to see whether researchers got more followers and comments on their Twitter posts upon appearing for TED talks if the researchers' YouTube videos had more views. In other words, to see if there is an impact on the social media presence of the researchers upon appearing on the TED talk platform.

CHAPTER 6 CONCLUSION

When understood an institution, science must abide by certain norms and identified Universalism, Communalism, Disinterestedness and organized skepticism as to the four norms of science. Further, individuals at the peak of their career receive more recognition and accumulate an advantage over the others. Science communication is now a vital part of science and takes place between the general public and scientists and researchers. Our study aims to understand various demographics of academic researchers appearing on the TED talk platform and compare them to a broader population of the same research cluster.

The study aims to address three questions. 1. Visualizing various socio-demographic attributes of both TED and Non-Ted Researchers. 2. Understanding the scientific and research impact of both researchers' categories by using various bibliometric indicators. 3. Understanding the odds of appearance of researchers on the TED talk platform given their socio-demographic attributes by using a prediction algorithm.

For the study, TED talks from 2009 to 2019 were considered and filtered out by academic researchers. Various socio-demographic features of these researchers were determined. Further, the researchers who belong to the same clusters were selected from the CWTS Web of Science database, and the socio-demographic attributes were retrieved. These researchers were categorized as Non-TED. Further, to understand the researchers' scientific impact, indicators like the number of citations and publications were retrieved from the Web of Science database. Finally, they were matched with the researchers whose description best suited the author demographics of that paper. This author disambiguation process was done by using the algorithm written by Caron and van Eck

(2014) and considering the google scholar account of the authors. And, the researchers were divided into five research fields: Biomedical Research and Health Sciences; Social Sciences and Humanities; Physical Sciences and Engineering; Mathematics and Computer Science; and Life and Earth Sciences.

The results of this study were presented with the help of various graphs and tables. Socio-demographic attributes such as gender, organization, university ranking, country of affiliation, academic age of both categories of researchers were visualized using bar graphs. It was seen that researchers from top organizations and organizations whose ranking is in the top 100 in the world are over-represented in TED talks. Also, more male researchers and researchers in academic age of 0-9 years in almost all the research fields compared to the Non-Ted category of researchers.

Coming to the scientific impact of the researchers, bibliometric indicators like citation score, normalized citation score, number of publications and top 10% pp was taken into consideration. Our study showed that, on average, the indicators' values were higher for TED researchers than Non-TED researchers. Further, for the prediction analysis, logistic binomial regression was being used. This model was generated with the help of the SPSS tool. Finally, the socio-demographic attributes of the researchers were taken into consideration as independent variables, and the dependent variable is TED appearance. From the results of the model, it was determined that the chances of appearing on Ted talk was more for a researcher who was from North American Continent, whose gender is male, belongs to the top organization whose ranking is within the top 100 and academic age of 0-9 years when compared to all other possibilities.

Thus, we can say that the Mathew effect has impacted scientific Communication as of today. This effect, described by Merton (1957), occurs when researchers who are already recognized receive more credit for their work than their lesser-known colleagues. Also, male researchers are more promoted than female researchers, and top universities get more limelight.

Further, TED talk, a popular social media platform, allows researchers to communicate scientific knowledge to a broader community. Hence, it becomes interesting to understand the demographics of speakers. This study would help narrative talk shows, including TED talks, broaden their speakers base and include a wider variety of researchers and not only the ones who seem to be doing better in terms of being well-known to the scientific community and the general public. Additionally, it provides a base for future predictive analysis where demographics of guests appearing on similar platforms such as TED can be understood better.

REFERENCES

Abraham, G. (2020). The Importance of Science Communication. *Metallography, Microstructure, and Analysis*, *9*(1), 3–4. https://doi.org/10.1007/s13632-020-00613-w

Andersen, J. P., Schneider, J. W., Jagsi, R., & Nielsen, M. W. (2019). Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *ELife*, *8*, e45374. https://doi.org/10.7554/eLife.45374

Bol, T., Vaan, M. de, & Rijt, A. van de. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, *115*(19), 4887–4890. https://doi.org/10.1073/pnas.1719557115

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, *36*(1), 2–72. https://doi.org/10.1002/aris.1440360102

Broadus, R. N. (1987). Toward a definition of "bibliometrics." *Scientometrics*, *12*(5–6), 373–379. https://doi.org/10.1007/BF02016680

Caron, E., & van Eck, N. J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. *Proceedings of the 19th International Conference on Science and Technology Indicators. CWTS-Leiden University*, 79–86.

Centre for Science and Technology Studies (CWTS). (n.d.). Fields. Retrieved July 27, 2021, from https://www.leidenranking.com/information/fields

Chang, J.-H., Kim, S.-H., Kang, M.-H., Shim, J. C., & Ma, D. H. (2018). The gap in scientific knowledge and role of science communication in South Korea. *Public Understanding of Science*, *27*(5), 578–593. https://doi.org/10.1177/0963662516685487

Cole, S., & Cole, J. R. (1967). Scientific Output and Recognition: A Study in the Operation of the Reward System in Science. *American Sociological Review*, *32*(3), 377–390. https://doi.org/10.2307/2091085

Conferences. (n.d.). Retrieved February 12, 2021, from https://www.ted.com/about/conferences

Dag W. Aksnes, L. (2019). Citations, citation indicators, and Research Quality: An overview of basic concepts and theories - dag W. Aksnes, Liv langfeldt, Paul WOUTERS, 2019. Retrieved July 28, 2021, from https://journals.sagepub.com/doi/full/10.1177/2158244019829575

Dahlstrom, M. F. (2010). The Role of Causality in Information Acceptance in Narratives: An Example From Science Communication. *Communication Research*, *37*(6), 857–875. https://doi.org/10.1177/0093650210362683

di Carlo, G. S. (2014). The role of proximity in online popularizations: The case of TED talks. *DISCOURSE STUDIES*, *16*(5), 591–606. https://doi.org/10.1177/1461445614538565

Fleming, J. S. (2009). *Talking with Barmaids: The importance of Science Communication in Today's Changing World*. *1*. https://www-researchgate-net.ezproxy.library.dal.ca/profile/Jean-

Fleming/publication/233720350_Talking_with_Barmaids_The_Importance_of_Science_Communication_in_Today%27s_Changing_World/links/02e7e5191556470911000000/Talking-with-Barmaids-The-Importance-of-Science-Communication-in-Todays-Changing-World.pdf

Franck, G. (1999). Scientific Communication—A Vanity Fair? *Science*, *286*(5437), 53–55. https://doi.org/10.1126/science.286.5437.53

Genderchecker. (n.d.). Retrieved July 27, 2021, from https://genderchecker.com/

Ghiasi, G., Mongeon, P., Sugimoto, C., & Larivière, V. (2018). Gender homophily in citations. In *STI 2018 Conference Proceedings* (pp. 1519–1525). Centre for Science and Technology Studies (CWTS). https://hdl.handle.net/1887/65291

Hogan, N. M., & Sweeney, K. J. (2013). Social networking and scientific communication: A paradoxical return to Mertonian roots? *Journal of the American Society for Information Science and Technology*, *64*(3), 644–646. https://doi.org/10.1002/asi.22842

Hoppe, T. A., Litovitz, A., Willis, K. A., Meseroll, R. A., Perkins, M. J., Hutchins, B. I., Davis, A. F., Lauer, M. S., Valantine, H. A., Anderson, J. M., & Santangelo, G. M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Science Advances*, *5*(10), eaaw7238. https://doi.org/10.1126/sciadv.aaw7238

Jr, M. (2020). *TED – Ultimate Dataset*. https://kaggle.com/miguelcorraljr/ted-ultimate-dataset

Kappel, K., & Holmen, S. J. (2019). Why Science Communication, and Does It Work? A Taxonomy of Science Communication Aims and a Survey of the Empirical

Evidence. *Frontiers in Communication*, *4*. https://doi.org/10.3389/fcomm.2019.00055

Laerd Statistics. (n.d.). Retrieved July 28, 2021, from https://statistics.laerd.com/premium/spss/blr/binomial-logistic-regression-in-spss.php

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, *504*(7479), 211–213. https://doi.org/10.1038/504211a

Liang, X., Su, L. Y.-F., Yeo, S. K., Scheufele, D. A., Brossard, D., Xenos, M., Nealey, P., & Corley, E. A. (2014). Building Buzz: (Scientists) Communicating Science in New Media Environments. *Journalism & Mass Communication Quarterly*, *91*(4), 772–791. https://doi.org/10.1177/1077699014550092

McBurney, M. K., & Novak, P. L. (2002). What is bibliometrics and why should you care? *Proceedings. IEEE International Professional Communication Conference*, 108–114. https://doi.org/10.1109/IPCC.2002.1049094

McClain, C., & Neeley, L. (2015). A critical evaluation of science outreach via social media: Its role and impact on scientists. *F1000Research*, *3*, 300. https://doi.org/10.12688/f1000research.5918.2

Merton, R. K. (1938). Science and the social order. *Philosophy of Science*. http://www.jstor.org/stable/184838

Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*(3810), 56–63. https://doi.org/10.2307/1723414

National Academies of Sciences, E. (2016). *Communicating Science Effectively: A Research Agenda*. https://doi.org/10.17226/23674

Ngumbi, E. (2020, February 12). *Is science communication making me a better scientist?* Hindawi. https://www.hindawi.com/post/science-communication-making-me-better-scientist/

Nielsen, M. W., & Andersen, J. P. (2021). Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences*, *118*(7). https://doi.org/10.1073/pnas.2012208118

O'Brien, K. R., Holmgren, M., Fitzsimmons, T., Crane, M. E., Maxwell, P., & Head, B. (2019). What Is Gender Equality in Science? *Trends in Ecology & Evolution*, *34*(5), 395–399. https://doi.org/10.1016/j.tree.2019.02.009

Okubo, Y. (1997). *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples* (OECD Science, Technology and Industry Working Papers No. 1997/01; OECD Science, Technology and Industry Working Papers, Vol. 1997/01). https://doi.org/10.1787/208277770603

Ray, V. (2018, April 27). *The racial exclusions in scholarly citations (opinion) | Inside Higher Ed*. https://www.insidehighered.com/advice/2018/04/27/racial-exclusions-scholarly-citations-opinion

RissleR, L. J., Hale, K. L., Joffe, N. R., & Caruso, N. M. (2020). Gender Differences in Grant Submissions across Science and Engineering Fields at the NSF | BioScience | Oxford Academic. *BioScience*, *70*(9), 814–820. https://doi-org.ezproxy.library.dal.ca/10.1093/biosci/biaa072

Rossiter, M. W. (1993). The Matthew Matilda Effect in Science. *Social Studies of Science*, *23*(2), 325–341.

Speaking at TED. (n.d.). Retrieved July 28, 2021, from https://www.ted.com/about/conferences/speaking-at-ted

Sugimoto, C. R., & Thelwall, M. (2013). Scholars on soap boxes: Science communication and dissemination in TED videos. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*, *64*(4), 663–674. https://doi.org/10.1002/asi.22764

Sugimoto, C. R., Thelwall, M., Larivière, V., Tsou, A., Mongeon, P., & Macaluso, B. (2013a). Scientists Popularizing Science: Characteristics and Impact of TED Talk Presenters. *PLoS ONE*, *8*(4), e62403. https://doi.org/10.1371/journal.pone.0062403

Sugimoto, C. R., Thelwall, M., Larivière, V., Tsou, A., Mongeon, P., & Macaluso, B. (2013b). Scientists Popularizing Science: Characteristics and Impact of TED Talk Presenters. *PLOS ONE*, *8*(4), e62403. https://doi.org/10.1371/journal.pone.0062403

The stem gap: Women and girls in science, technology, engineering and math. (2020, October 05). Retrieved July 27, 2021, from https://www.aauw.org/resources/research/the-stem-gap/

Tricco, A. C., Thomas, S. M., Antony, J., Rios, P., Robson, R., Pattani, R., Ghassemi, M., Sullivan, S., Selvaratnam, I., Tannenbaum, C., & Straus, S. E. (2017). Strategies to Prevent or Reduce Gender Bias in Peer Review of Research Grants: A Rapid Scoping Review. *PLoS One*, *12*(1), e0169718. http://dx.doi.org.ezproxy.library.dal.ca/10.1371/journal.pone.0169718

Trusted publisher-independent citation database. (2021, July 07). Retrieved July 26, 2021, from https://clarivate.com/webofsciencegroup/solutions/web-of-science/

Tsou, A., Thelwall, M., Mongeon, P., & Sugimoto, C. R. (2014). A Community of Curious Souls: An Analysis of Commenting Behavior on TED Talks Videos. *PLoS ONE*, *9*(4), e93609. https://doi.org/10.1371/journal.pone.0093609

Waltman, L., & Eck, N. J. van. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

Weber, J. R., & Schell Word, C. (2001). The Communication Process as Evaluative Context: What Do Nonscientists Hear When Scientists Speak?: Scientists and nonscientists benefit by recognizing that attempts at mutual influence, multiple frames of reference, and “objective” information in science communication are not neutral but evaluated with other social influences. *BioScience*, *51*(6), 487–495. https://doi.org/10.1641/0006-3568(2001)051[0487:TCPAEC]2.0.CO;2

Whatever your brand CHALLENGE,TED PARTNERSHIPS has a solution. (2021, March 10). Retrieved July 29, 2021, from https://tedpartnerships.com/

Yeager, K. (n.d.). *LibGuides: SPSS Tutorials: Pearson Correlation*. Retrieved June 24, 2021, from https://libguides.library.kent.edu/SPSS/PearsonCorr