

Factors Affecting Wikipedia Notability of Filipino Athletes: A Study on Education and Geographic Inequality

Sylvan Dan M. Moldes
Keys School Manila

Kevin S. Cuevas
University of the Philippines Diliman

Abstract

As the Philippines celebrates its 100th year of participation in international sporting events, this paper looked into the Wikipedia pages of Filipino athletes and investigated how the presence or absence of quality education and geographic inequalities affect the notability of Filipino sports personalities. The authors hypothesized that certain terms co-occur frequently in the descriptions and infoboxes of notable Filipino athletes, and that terms related to popular sports, urbanized areas, and prestigious educational institutions have higher centrality in the network. Information from the Wikipedia pages were web scraped using Python, while the article title, infobox, and description of the entries were harvested using BeautifulSoup and Wikipedia API. Through the pre-processed, extracted noun phrases, a co-occurrence network was built and imported to Gephi, and was visualized via Force Atlas 2 algorithm. The results suggest that there are three major network clusters from the data obtained: (1) Olympics, regional, and local sports; (2) mainstream, collegiate, and professional sports, and (3) boxing and combat sports. Through this research, the authors hope to fill the gap in local literature on sports, education and geographic inequalities in the country. The findings of this research offer an innovative approach by making use of data analysis in suggesting policy recommendations to stakeholders.

Keywords: Wikipedia notability, Co-occurrence Network, Filipino athletes, UN Sustainable Development Goals (SDG), sports education, geographic inequality

BACKGROUND

The year 2024 marks the centennial anniversary of the Philippines sending delegates to international sporting competitions. Back in 1924, the Philippines sent its lone delegate to represent the country in the Summer Olympics held in Paris, France. Filipino runner David Nepomuceno participated in the 100- and 200-meter races, although unable to secure an Olympic medal. Since then, the country has been sending athletes to various international sporting events such as the Olympics, the Asian Games, and the Southeast Asian Games. In the 2024 Summer Olympics, 22 Filipino athletes participated in different sporting events (Dioquino, 2024).

When world leaders met in 2015, a universal agenda for sustainable development was adopted, and it was called *The 2030 Agenda for Sustainable Development*. This policy agenda, accompanied by 17 Sustainable Development Goals (SDGs) and 169 targets, aims to call to action on a global scale for the next 15 years. This paper focuses its lens on two of the 17 SDGs, more specifically, Goal 4 and Goal 10. Goal 4 sets its sights on providing quality education, ensuring inclusive and equitable quality education, and promoting lifelong opportunities for all, while Goal 10 concentrates on reducing inequality within and among countries. In the context of this paper, Goal 4 of the UN SDGs is examined using Target 4.5, which focuses on eliminating gender disparities in education

and ensuring equal access to all levels of education. Goal 10 is examined based on Target 10.2, which empowers and promotes social, economic, and political inclusion for all, as seen from the Philippine sporting landscape.

Wikipedia Notability

A handful of articles pertaining to Wikipedia shed light on the aspect of notability of people across different disciplines. Tripodi (2021) looked into Wikipedia biographies nominated for deletion and found that “women’s biographies are more frequently miscategorized as non-notable than men’s” adding to the knowledge that not only is there a gender gap, but also Wikipedians have to “work nearly twice as hard to prove women’s notability” (p. 10). A similar study conducted by Martini (2023) analyzed the content of Wikipedia articles for deletion and quantitatively investigated gender differences in discussions on deletion. Among its significant findings, the paper discussed that “women’s biographies are questioned more often but not deleted more often compared to men’s biographies” (p. 13). Meanwhile, Hinnoosaar (2019) suggests that the provision of information to editors about gender inequality may help reduce the gender gap in Wikipedia contributions.

Gauthier and Sawchuk (2017) identified a number of Wikipedia entries that act as sites of contestation where academic scholars and Wikipedians debate on aspects of notability, verifiability, and tone. Their study found that there is an inconsistency in the application of notability guidelines, which are particularly biased against women, cultural minorities, sexualities, and genders. Due to the nature of Wikipedia being “predicated on secondary sources rather than primary sources”, visibility in secondary sources therefore builds notability (pp. 397-398). To add to this, several other studies noted that Wikipedia contributors tend to use online or Internet-based references (Huvila, 2010; Kaffee & Elsahar, 2021; Luyt & Tan, 2010).

Known Inequalities on Wikipedia

Literature has also shown that inequalities exist within Wikipedia. Among many other discussions on Wikipedia inequalities, two specific categories stand out: gender and ethnicity.

A content analysis from Ferran-Ferrer et al. (2023) identified three areas of research that focused on gender gaps on Wikipedia, which include gaps in content, editing and participation, and readership.

Adams et al. (2019) discussed both gender and racial gaps on Wikipedia. Their research showed that male sociologists are more than twice as likely to have a Wikipedia page as their female counterparts. Similarly, on issues of race/ethnicity, sociologists who are white are twice as likely to have a page as non-white/minority sociologists. Lemieux et al. (2023) looked into Wikipedia articles for deletion, which were tagged as *WP:Search Engine Test* and *WP:Too Soon*, and found that the application of these two criteria was not applied consistently “across race and gender for biographies of academics” (p. 9). As a result, academics who are white women and Black, indigenous, and people of color (BIPOC) are at a disadvantageous position when it comes to keeping their Wikipedia pages. Furthermore, the study noted that women in Wikipedia are more notable than men with the use of “internal and external proxy measures for notability” (p. 19).

Despite extensive coverage on Wikipedia on inequality in aspects of gender, ethnicity, and how the Wikipedia guidelines on notability are interpreted in favor of or against these two among different disciplines, there has been minimal coverage on how other demographic aspects affect the notability of athletes. This study will attempt to explore other factors affecting the notability of Filipino athletes. In particular, inequalities involving gender, education and geographic location will be discussed. In order for inequalities on these aspects to surface, the researchers rely on the degree centrality of terms within the text network and interpret the data accordingly. The following research questions and hypotheses guide this inquiry:

Research Questions

1. What are the most frequently co-occurring terms in the descriptions and infoboxes of notable Filipino athletes?
2. What is the centrality of terms related to high-profile sports, geographic locations, and educational institutions within the text network?

Hypotheses

1. Certain terms (e.g., names of cities, schools, sports) frequently co-occur in the descriptions and infoboxes of notable Filipino athletes, indicating common paths to notability.
2. Terms related to high-profile sports (e.g., basketball, boxing), metropolitan areas, and prestigious educational institutions have

higher centrality in the network, suggesting these factors are prominently featured in the profiles of notable athletes.

Data and Information Science in Policymaking

Data science is a field concerned with the responsible extraction of data to gather information and insights (Martens, 2022). Recently, there has been a growing interest in the academe on the relationship between data science and evidence-based decision-making, more specifically, how data science contributes to the development of more effective policies (Kim, 2024). For example, Buttow (2024) analyzed policy documents produced by an intergovernmental organization through the lenses of Critical Data Studies (CDS) and regulation theory. The use of data-driven technologies in all the stages of the policy cycle framework has been found to contribute positively to evidence-based policymaking in the government, bring policy stages together, and transform the government's response from reactive to proactive. Data collection and data analytics play a central role in all the stages of the policy and regulatory cycle. Idzalika et al. (2019) looked into the evolving role of big data analytics in the development of evidence-based policymaking in population dynamics and social development research. Big data from various sources (social media, mobile phone calls, financial records, crowdsourcing, transportation records) have been used as information tools to fill gaps in policy-making and traditional datasets.

On the other hand, information science deals with the creation, management, and use of information. As policymakers often refer to various sources of information (scientific, practical, and opinion-based), information science professionals can contribute to the provision of curated resources that aid in the creation of legislation and policies across all stages of policymaking. Albeit the differences in the use of terminologies, Cairney (2016) and Buttow (2024) identified stages in the policy-making cycle where information professionals may be of help. For instance, during the stage of agenda setting or defining problems at hand, information professionals may provide policymakers with relevant data to determine which problem requires the most attention. Notably, librarians are "expert searchers" (Crosby, 2001) who can access a variety of sources to fill gaps in knowledge and information. In the consultation phase, librarians and other information professionals may be invited as resource speakers, sharing their insights on a range of public issues as

experts.

The authors, a data scientist and a librarian, will formulate policy recommendations informed by the relationships found in web-scraped data through co-occurrence network analysis, text mining, and natural language processing. These policy recommendations are mainly directed to the Philippine Sports Commission (PSC), which is the policy-making body of sports development in the Philippines.

METHODS

Data was obtained from existing Wikipedia pages about Filipino athletes. This was done by manually going through the Wikipedia biographies of Filipino athletes from dedicated Wikipedia pages of international sporting events where the Philippines had sent a delegation (i.e., Olympics, Paralympics, Winter Olympics, Asian Games, Southeast Asian Games) in a span of 100 years, from 1924 to 2024. The Wikipedia page on Filipino sportspeople was also included in the data gathering. From these, 990 Wikipedia pages on athletes were harvested through web scraping technique using Python. From the list of article links of athletes, a combination of scripts using the packages BeautifulSoup and the Wikipedia API were used to collect the article's title, infobox, and description.

The pre-processing was done in three phases:

1. **Standardizing Article Names:** This step entailed the elimination of unnecessary parentheses and descriptive elements within the titles to establish the athletes' names with article names as the basis. Regular expressions were employed to facilitate the cleaning and standardization of these names.
2. **Extracting Noun Phrases:** Two distinct models were utilized for the extraction of noun phrases from the descriptions provided in the article. Hugging Face's RoBERTa model, configured with default hyperparameters, was leveraged to identify noun phrases present within the text (Liu et al., 2019). These noun phrases were subsequently refined to ensure their relevance and accurate representation of key terms associated with the athletes.
3. **Refining Noun Phrases:** OpenAI's turbo 3.5 algorithm was employed for the further refinement of noun phrases. The temperature was set at 0 to guarantee deterministic results. The following prompt was utilized to guide the

model in this process:

Here are the extracted noun phrases from the Wikipedia biography of an athlete. Please refine the list by: Eliminating overly generic terms that do not add specific value. Ensuring each noun phrase accurately reflects important aspects of the athletes' careers and achievements.

Next, a co-occurrence network was constructed using the extracted noun phrases. Pairs of noun phrases that co-occurred within the same text were identified, and their frequencies of co-occurrence were recorded. This resulted in a list of pairs of co-occurring terms along with the frequency of their co-occurrence. This data was then imported into Gephi for further analysis. Within Gephi, the network was visualized using the Force Atlas 2 algorithm, which is commonly used for spatializing graphs based on the repulsion and attraction forces between nodes (Jacomy et al., 2014). The Force Atlas 2 algorithm in Gephi was configured with the following settings:

- Threads number: 6
- Tolerance (speed): 1.0
- Approximate Repulsion: Checked
- Approximation: 1.2
- Scaling: 3.0
- Stronger Gravity: Unchecked
- Gravity: 1.5
- Dissuade Hubs: Unchecked
- LinLog mode: Unchecked
- Prevent Overlap: Unchecked
- Edge Weight Influence: 1.0
- Normalize edge weights: Unchecked
- Inverted edge weights: Unchecked

The Force Atlas 2 algorithm utilized within the Gephi software facilitates the visual clustering of nodes grounded in their co-occurrence. This process enhances the identification of communities or groupings of related terms. Gephi employs the Louvain method for community detection, an iterative process that optimizes modularity to identify clusters within the network (Blondel et al., 2008). The resolution parameter in community detection was deliberately set to 3.0. This elevated resolution setting enabled more refined identification of communities

for smaller, more specialized clusters within the network's structure. This methodology combines text mining, natural language processing, network analysis, and visualization techniques to explore and understand the complex relationships within the data.

FINDINGS

Based on the results informed by the web-scraped data, a total of 13,414 nodes and 630,038 edges were produced. The hypotheses of the researchers were confirmed: (1) certain terms frequently co-occur in the descriptions and infoboxes of notable Filipino athletes, indicating common paths to Wikipedia notability, and (2) terms related to high-profile sports (e.g., basketball, boxing), metropolitan areas, and prestigious educational institutions have higher centrality in the network, suggesting these factors are prominently featured in the profiles of notable athletes.

Major Network Clusters

1. **Olympics, Regional, and Local Sports (70.43%):** This cluster includes athletes and sports that are region-specific or less mainstream but have significant local or international representation. Examples include niche sports and events where Filipino athletes have historically participated, such as the Southeast Asian Games and various niche sports competitions. These sports often highlight the diversity and range of athletic talent in the Philippines.

The presence of regional and niche sports in this cluster emphasizes the breadth of sports culture in the Philippines. These sports, while not always in the international spotlight, play a crucial role in the development of athletic talent and community engagement. The Olympic sports included here show the long-standing commitment of Filipino athletes to compete on the global stage, despite challenges and limited resources. This cluster is notable for its rich historical context and the emotional connection that fans have with athletes representing their heritage and nation in prestigious events like the Olympics.

2. **Mainstream, Collegiate and Professional Sports (25.00%):** This cluster focuses on sports that have significant commercial backing and a structured collegiate system. It includes popular sports like basketball and volleyball, which have

robust leagues and are widely followed in the Philippines. The collegiate sports system, particularly in universities, serves as a breeding ground for future professional athletes.

Commercial and collegiate sports are critical for the sustainability and growth of sports in the Philippines. The commercial aspect brings in sponsorships, media coverage, and fan engagement, which are essential for the financial viability of sports leagues. Collegiate sports ensure a steady pipeline of talent, providing young athletes with the necessary training and exposure to excel professionally. This cluster is highly notable and liked due to its broad appeal, extensive media coverage, and the strong fan base that supports teams and individual athletes. The frequent co-occurrence of terms in this cluster highlights the interconnected nature of commercial success and athletic development in popular sports.

3. **Boxing and Combat Sports (4.56%):** This cluster is distinguished by the inclusion of combat sports, which have historically been areas where Filipino athletes excel. Boxing, in particular, has brought international acclaim to the Philippines with legends like Manny Pacquiao. Mixed Martial Arts (MMA) and other combat sports have also seen a rise in popularity, with many athletes competing internationally.

Combat sports hold a special place in the Philippine sports landscape due to their storied history and the global success of Filipino athletes. Boxing has been a source of national pride, and the growing popularity of MMA indicates a shift towards embracing a variety of combat sports. This cluster is considered highly notable and liked because of the dramatic, high-stakes nature of combat sports, the compelling personal stories of athletes, and the significant achievements on the world stage. The frequent co-occurrence of terms in this cluster reflects the strong public interest and media attention that combat sports receive.

Tertiary Education and Inequality in Philippine Sports: Case in Point

Data revealed that noun phrases associated with universities and colleges found in Metropolitan Manila have higher centrality in the network. Higher education institutions that are members of athletic

associations (e.g. UAAP, NCAA) tend to have higher centrality in the network, as shown in Table 1.

Table 1 *Philippine Universities and Colleges with the 10 Highest Degree Centralities in the Network*

Institution	Degree Centrality
University of Santo Tomas	989
Far Eastern University	737
Ateneo de Manila University	735
San Beda College	695
University of the East	659
De La Salle University	643
Adamson University	524
University of the Philippines	519
National University	514
University of the Cordilleras	213

From the table above, it can be noted that nine out of the ten universities and colleges with the highest degree centralities in the network are located in Metro Manila. Eight of those are institutions that are members of the University Athletic Association of the Philippines (UAAP), and one is a member of the National Collegiate Athletic Association (NCAA). Meanwhile, the lone university located outside of Metro Manila, the University of the Cordilleras, is a member of the Baguio-Benguet Educational Athletic League (BBEAL). Altogether, these institutions are in mainland Luzon. Institutions located in Metro Cebu and Metro Davao have degree centralities that are lower than their Metro Manila counterparts.

Given this data, athletes who have affiliations with universities and colleges that have the highest degree centralities in the network are connected to a higher number of nodes within the network. This suggests that having connections or ties with these institutions open up a path to notability. Conversely and unfortunately, athletes who are affiliated with institutions that do not have a high degree of centrality are less likely to be considered notable. With most of the institutions with high degree centrality being geographically located in Luzon, the likelihood of notability for athletes who are living in Visayas and Mindanao is slim.

Key Sports Personalities: Side by Side

Moving from the discussion on the geographic inequality concerning tertiary education and degree centralities in the network, this section now shifts its

focus to select Philippine sports personalities and corresponding direct connections. While not exhaustive, the examples below show the disparity between personalities who have low and high degree centralities in the network.

Table 2 *Select Sports Personalities and Corresponding Degree Centralities in the Network*

Name	Degree Centrality
Nesthy Petecio	31
Carlos Yulo	49
Hidilyn Diaz	98
Onyok Velasco	310
Alyssa Valdez	145
Lydia de Vega	242
Manny Pacquiao	255
James Yap	264

Table 2 provides a snapshot of key sports personalities in the Philippines. The first four personalities are Filipino Olympians who secured at least one (1) Olympic medal: Nesthy Petecio (1 Silver, 1 Bronze for Boxing), Carlos Yulo (2 Gold for Artistic Gymnastics), Hidilyn Diaz (1 Gold, 1 Silver for Weightlifting), and Onyok Velasco (1 Silver for Boxing). The latter four, meanwhile, are notable sports personalities who, albeit not being Olympic medalists, have high degree centralities in the network. Alyssa Valdez is a multi-awarded volleyball player, Lydia de Vega is a track and field sprinter, once considered Asia's Fastest Woman, Manny Pacquiao is a professional boxer known for being the only eight-division world champion boxer, and James Yap is a professional basketball player from the Philippine Basketball Association (PBA).

From the table presented, it may be noted that sports personalities with high degree centralities fall under the three major clusters identified in this paper. Alyssa Valdez and James Yap play sports that are considered popular in the Philippines. As previously noted, these popular sports have extensive media coverage, fan engagement, and appeal, which contribute toward having a high degree of centrality in the network. Lydia de Vega may be considered as an athlete who played for regional sporting events such as the Asian Games and the Southeast Asian Games. Manny Pacquiao and Onyok Velasco, who are boxers, also had high degree centralities, which may be attributed to the high-stakes nature of contact sports like boxing.

While these data may be interpreted in multiple ways, it is also worth noting how the inclusion of sports personalities in the three major clusters, taking into

consideration the percentages each of those clusters occupy in the pie, does not automatically equate to having a high degree of centrality in the network. To illustrate, Nesthy Petecio, Carlos Yulo, and Hidilyn Diaz, who are Olympic medalists belonging to the Olympics, Regional, and Local Sports cluster, have relatively low degree centrality, indicating fewer nodes connected to them as compared to their counterparts who belong to either Mainstream, Collegiate and Professional Sports cluster or Boxing and Combat Sports cluster. This may be explained by different factors, including but not limited to: age, gender, university or college these athletes attended, general appeal of the sport they represent, and media coverage of their respective sports.

Interestingly, male sports personalities tend to have higher degree centralities in the network than females. For example, male athletes who have high degree centralities include Carlos Loyzaga (436), Brandon Vera (413), Alvin Patrimonio (396), LA Tenorio (389), and Efren "Bata" Reyes (350). The highest degree centrality for a female Filipino athlete is Lydia de Vega (242), followed by Bong Coe (234), Lita dela Rosa (224), Akiko Thomson (203), and Alyssa Valdez (145). By looking only at the difference between the degree centralities of male and female sports icons, it may be inferred that male athletes are more notable than female athletes.

Additionally, male sports personalities who are involved in team sports are likely to be more notable than those involved in individual sports: Carlos Loyzaga, Alvin Patrimonio, and LA Tenorio are professional basketball players, while Brandon Vera and Efren "Bata" Reyes are into combat and individual sports, respectively. Meanwhile, the opposite may be observed for female athletes: those who play individual sports had a higher degree centrality in the network, in Lydia de Vega (track and field), Bong Coe and Lita dela Rosa (bowling), and Akiko Thomson (swimming), than athletes who play team sports, in Alyssa Valdez (volleyball).

CONCLUSION AND RECOMMENDATIONS

This study contributes to the existing discourse on Wikipedia notability and the body of literature on education and gender inequality by offering an innovative approach through the use of text mining, natural language processing, network analysis, and visualization techniques on Wikipedia pages of Filipino sports personalities. Three major clusters have been observed to exist within the dataset in this study: (1) Olympics, Regional, and Local Sports

(70.43%), (2) Mainstream, Collegiate and Professional Sports (25.00%), and (3) Boxing and Combat Sports (4.56%).

From the discussion of findings, it was revealed that certain inequalities as identified by published literature (Gauthier & Sawchuk, 2017; Tripodi, 2021; Wagner et al., 2015) are also evident in our data, particularly that of gender inequality on Wikipedia notability. Meanwhile, the geographical inequality in terms of access to tertiary education is apparent in the discussion of higher education institutions (HEIs) with the highest degree centralities—all are found in Luzon island, leaving Visayas- and Mindanao-based HEIs with lower degree centralities, therefore being less notable.

Challenges and Future Directions

The researchers hope that the findings from this study will help advance the discussion on inequality in sports, specifically on aspects of gender, education, and geographical location in the Philippine context. As the Philippines celebrates its 100th year of sending out delegates to international sporting competitions in 2024, this milestone also provides an opportunity to look back and look further ahead—reflect on challenges and issues that the nation faced in past sporting events and take concrete steps to address inconveniences—in order to provide an equitable landscape of sports in the country.

Looking at Wikipedia articles of Filipino sports personalities, this research endeavor offered a unique way of exploring factors that contribute towards notability through the lenses of data science and information science. The research questions and hypothesis of this study were answered and confirmed, but more significantly, analysis of data resulted in uncovering existing inequalities faced by sports personalities on aspects of gender, education, and geographical location. Moving forward, this study hopes to rally fellow scholars and interested stakeholders to take an active stance in multiple fronts: action research for those who are in the academe and policy recommendation for those involved in policymaking.

Policy Recommendations

Given the potential of the results of this study to impact the Philippine sports landscape, the researchers have also formulated suggestions and/or amendments to policies in order to address the inequalities in education and sports in the country.

By integrating the results of this study in the different stages of policy-making within the PSC, future policies will hopefully address inequalities in education and sports (as suggested by high centralities and low centralities in the network) in the Philippines.

Provide Equitable Media Coverage to Filipino Athletes

To ensure visibility of athletes, the researchers suggest providing equitable media coverage to Filipino athletes across sports, across genders, and across regions. To do this, there should be sufficient media coverage for sporting events which are less popular in the country (e.g. athletics, taekwondo, weightlifting, etc.). These events should be given enough media coverage similar to sports which are already mainstream in the country (e.g. basketball, volleyball, etc.). Filipina athletes who experience inequity in sports should also receive a fair amount of media coverage similar to their male counterparts. Media coverage of athletes from various regions in the country should also be promoted, if not highlighted, especially those coming from areas (or HEIs) which do not belong to geographical locations with high degree centrality. That way, not only will the geographical inequality be addressed, but also the capacity of educational institutions in providing quality education to student-athletes outside Luzon will be recognized.

Provide Quality Access to Education for Filipinos Who Are Into Sports

The institution of the Philippine High School for Sports (PHSS) is a step towards providing quality access to education for Filipinos who are into sports. As an educational institution and facility that will be attached to the Department of Education and PSC, its vision of providing scholarship and holistic training for students who excel in sports will provide a solid foundation for quality access to sports education in the country (Reganit, 2020).

Policies may be implemented in a local and national context. For example, the Philippine government, through the PSC, may adopt policies that expand on existing scholarship programs that are tailored towards student-athletes who are into a variety of sports. Whenever applicable, regional and/or local governments may establish branches of the PHSS and the National Academy of Sports (NAS) such that these branches may be geographically accessible to people outside metropolitan areas. Local Government Units (LGUs) may invest in the construction of sports facilities that will serve as training grounds and

competition venues.

Strengthen Partnerships with Media and Educational Institutions

In order to provide comprehensive coverage of all sports and athletes, the PSC should engage in active collaboration with media organizations and educational institutions. This collaborative effort could encompass joint initiatives, scholarships, media training for athletes, and the promotion of sporting events that highlight the wide range of athletic achievements. To effectively administer these collaborations and maintain consistent efforts, the PSC should establish a dedicated division focused on media and educational partnerships. By creating a specialized division, the PSC can streamline efforts to enhance the visibility and recognition of Filipino athletes, ensuring that their achievements are celebrated and supported at both national and international levels.

Enhance Secondary Source Visibility and Training for Wikipedia Editors

To establish the notability of Filipino athletes on Wikipedia, it is crucial for sports organizations, educational institutions, and media outlets to publish articles, reports, and studies showcasing their achievements. This will create a robust repository of secondary sources that can be used for this purpose, as Wikipedia notability is often based on visibility in such sources (as noted by Huvila, 2010; Kaffee & Elsahar, 2021; Luyt & Tan, 2010).

Additionally, training programs for Wikipedia editors should be implemented to address biases in notability and coverage. These programs, led by members of the National Innovation Council, should include information on gender and geographic inequalities, as well as the importance of diverse sources and perspectives.

Furthermore, the training should emphasize the significance of Wikipedia edits as a measure of the Global Innovation Index (GII). The GI considers the quality and quantity of Wikipedia contributions as an indicator of a country's innovation and knowledge dissemination. Therefore, improving the quality and inclusiveness of Wikipedia content not only enhances the representation of Filipino athletes but also contributes to the country's overall innovation ranking.

With these policy recommendations being taken into consideration, Target 4.5 of UN SDG 4 and Target

10.2 of UN SDG 10 are hopefully addressed to a certain extent, reducing gender disparity and providing more avenues for inclusion, regardless of age, sex, ethnicity, etc.

DECLARATION OF CONFLICTING INTEREST

The author/s declared no potential conflicts of interest with respect to research, authorship, and/or publication of this article.

DECLARATION ON SOURCE OF FUNDING

The author/s received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

- Adams, J., Brückner, H., & Naslund, C. (2019). Who counts as a notable sociologist on Wikipedia? Gender, race, and the “Professor Test.” *Socius*, 5, 237802311882394. <https://doi.org/10.1177/2378023118823946>
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Buttow, C. V. (2024). Data-Driven Policy Making and its Impacts on Regulation: A study of the OECD Vision in the light of Data Critical Studies. *European Journal of Risk Regulation*, 1–19. <https://doi.org/10.1017/err.2024.73>
- Cairney, P. (2016). The politics of Evidence-Based Policy making. In *Palgrave Macmillan UK eBooks*. <https://doi.org/10.1057/978-1-137-51781-4>
- Crosby, O. (2001). Librarians: information experts in the Information Age. *Occupational Outlook Quarterly*, 44(4), 2–15. <https://www.bls.gov/careeroutlook/2000/Winter/art01.pdf>
- Dioquino, D. (2024, July 26). 2024 Paris Olympics: Meet Team Philippines. *RAPPLER*. <https://www.rappler.com/sports/things-to-know-members-team-philippines-paris-olympics-2024/>
- Ferran-Ferrer, N., Boté-Vericad, J.-J., & Minguillón, J. (2023). Wikipedia gender gap: A scoping review. *El Profesional de La Información*. <https://doi.org/10.3145/epi.2023.nov.17>
- Gauthier, M., & Sawchuk, K. (2017). Not notable enough: feminism and expertise in Wikipedia. *Communication and Critical/Cultural Studies*, 14(4), 385–402. <https://doi.org/10.1080/14791420.2017.1386321>
- Hinnosaar, M. (2019). Gender inequality in new media: Evidence from Wikipedia. *Journal of*

- Economic Behavior & Organization*, 163, 262–276. <https://doi.org/10.1016/j.jebo.2019.04.020>
- Huvila, I. (2010). Where does the information come from? Information Source Use Patterns of Wikipedia. *Information Research*, 15(3). <https://informationr.net/ir/15-3/paper433.html>
- Idzalika, R., Pramestri, Z., Amin, I., Riyadi, Y., & Hodge, G. (2019). Big data for population and social policies. *United Nations Global Pulse*. <https://www.unglobalpulse.org/wp-content/uploads/2023/12/2019-01-24-08-58-31.pdf>
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9(6), e98679. <https://doi.org/10.1371/journal.pone.0098679>
- Kaffee, L., & Elsahar, H. (2021). References in Wikipedia: The Editors' perspective. *Companion Proceedings of the Web Conference 2021*. <https://doi.org/10.1145/3442442.3452337>
- Kim, E. (2024). Can data science achieve the ideal of evidence-based decision-making in environmental regulation? *Technology in Society*, 78, 1–9. <https://doi.org/10.1016/j.techsoc.2024.102615>
- Lambiotte, R., Delvenne, J. C., & Barahona, M. (2009). Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*. <https://arxiv.org/abs/0812.1770>
- Lemieux, M. E., Zhang, R., & Tripodi, F. (2023). “Too Soon” to count? How gender and race cloud notability considerations on Wikipedia. *Big Data & Society*, 10(1). <https://doi.org/10.1177/20539517231165490>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*. <https://arxiv.org/abs/1907.11692>
- Luyt, B., & Tan, D. (2010). Improving Wikipedia's credibility: References and citations in a sample of history articles. *Journal of the American Society for Information Science and Technology*, 61(4), 715–722. <https://doi.org/10.1002/asi.21304>
- Martens, D. (2022). Data Science Ethics: Concepts, techniques and cautionary tales. In Oxford University Press eBooks. *Oxford University Press*. <https://doi.org/10.1093/oso/9780192847263.001.0001>
- Martini, F. (2023). Notable enough? The questioning of women's biographies on Wikipedia. *Feminist Media Studies*, 1–17. <https://doi.org/10.1080/14680777.2023.2266585>
- Reganit, J. C. (2020, February 26). Senate OKs bill creating Philippine High School for Sports. *Philippine News Agency*. <https://www.pna.gov.ph/articles/1094940>
- Tripodi, F. (2021). Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, 25(7), 1687–1707. <https://doi.org/10.1177/14614448211023772>
- Wagner, C., Garcia, D., Jadidi, M., & Strohmaier, M. (2015). It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, 9(1), 454–463. <https://doi.org/10.1609/icwsm.v9i1.14628>

AUTHOR BIOGRAPHIES

Sylvan Dan M. Moldes, RL, graduated magna cum laude in 2017 and valedictorian for Class Iridescent at the University of the Philippines School of Library and Information Studies (UP SLIS). After graduation, he took the Librarian Licensure Examination in September 2017 and emerged Top 2 out of 904 takers. At present, he is the Head Librarian of Middle School and College Preparatory in Keys School Manila. He also teaches the Extended Essay in Grades 11 and 12 while pursuing Master of Science in Library and Information Science (MSLIS) at UP SLIS.

Email: smmoldes@up.edu.ph

Kevin S. Cuevas is a licensed agriculturist, freelance journalist, and data scientist. He earned his bachelor's degree in agribusiness management from the University of the Philippines in Los Baños. He also pursued economics at Brown University before finishing a master's degree in data science from the Technological Institute of the Philippines. His interests include advanced analytics, marketing, strategic marketing, customer segmentation, sales incentives, growth strategies, time series analysis, monitoring and evaluation, and corporate reporting. At present, he is taking the Master of Legal Studies and Juris Doctor (MLS-JD) program at Adamson University College of Law.

Email: kscuevas@up.edu.ph



This work is licensed under a Creative Commons Attribution 4.0 International License.



The PhJLIS is published by the School of Library and Information Studies, University of the Philippines Diliman. ISSN 2719-0471 (Online)