Inequality at Scale: Unpacking the causes and consequences of Instagram's Network Distribution

Introduction

In this paper, I explore the power law distributions on Instagram, a phenomenon where a select group of users, referred to as influencers in this essay, hold disproportionate connectivity. I unpack the causes and consequences of this phenomenon through the introduction of 'privilege-based attachment'—a concept that interrogates the systemic and often invisible advantages that allow certain nodes within the network to gain and exert greater influence.

This paper will first establish the presence and characteristics of power law distributions within Instagram. It then explores the theoretical implications of these findings through the lens of attachment, focusing on both preferential and fitness-based attachment mechanisms, before introducing an alternative: 'privilege-based attachment'. This alternative sheds light on the inherent biases within digital social structures, emphasising how these platforms are not merely tools for social interaction but pivotal spaces for the creation and perpetuation of societal norms and hierarchies where inequalities are essentially scaled. This is followed by a discussion of the epistemic consequences of power-law distributions before a brief outline of potential solutions and areas for future work.

1. Understanding Instagram's Network Distribution

Past Findings

Power law distributions are a key concept in network theory, particularly in the context of infinite or 'scale-free' social media platforms such as Instagram (Boy and Uitermark, 2017). Over the past decade, multiple studies focusing on online and digital platforms have highlighted the skewed nature of online networks, with the popularity of websites, social media posts and blogs consistently having skewed power law distributions (Johnson et al., 2014; Cameron, 2022; Artico et al., 2020). These distributions are characterised by the principle that a small number of nodes (or users) accumulate a vast majority of connections (or followers), while the majority of nodes have significantly fewer connections.

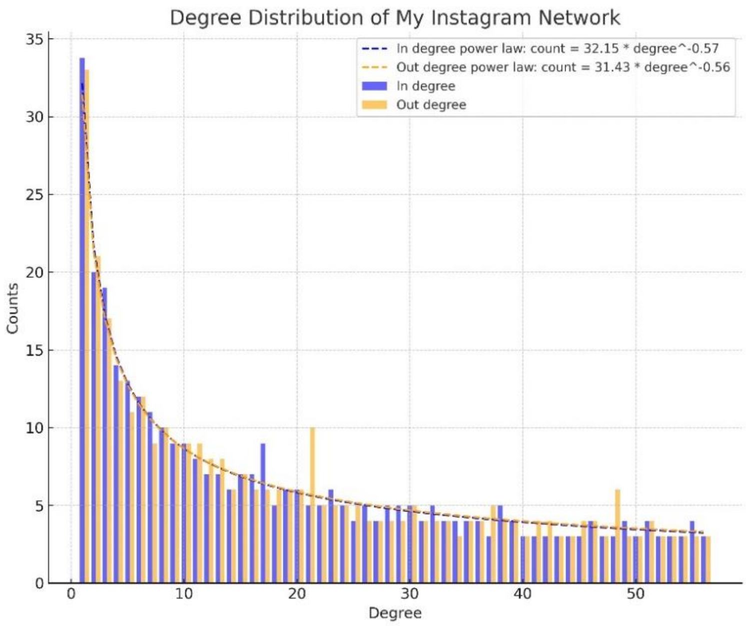
In Instagram's case, only a small subset of users have millions of followers, whereas the vast majority struggle to reach even a thousand (Christin and Lu, 2023). Cameron's (2022) analysis of the network's top 5000 users yielded an adjusted R-squared value of 0.996, suggesting a near-perfect power-law distribution. The term 'near-perfect' in describing (2022) findings on Instagram's power-law distribution is intentional as definitively distinguishing a pure power-law distribution from a network with an exponential cut-off can be difficult, especially given that while Instagram can be described as an infinite social network (Cameron, 2022), the data myself and other researchers have been able to mine from the platform is finite. The degree distribution – or the way connections are spread among the nodes in a network – can seem to follow a power-law until it reaches this distant cut-off point. Beyond this point, the frequency of nodes with a high number of connections drops off more sharply than what a power-law distribution would predict (Artico et al., 2020).

With this in mind, Artico et al. (2020) develop and employ a tail-testing method which involves setting a cut-off in the degree distribution to ensure the test can detect deviations from a power law. Testing 4482 networks, they find that almost 65% had a power-law tail with at least 80% power, suggesting that the fact that networks are finite does not necessarily make them rarer than previously assumed (Artico et al., 2020). The determined frequency of power laws leads me to assume that what Cameron (2022) finds in their analysis of Instagram is a power-law distribution for the sake of the analysis. It is worth noting that even if this assumption is proven to be false, pattern changes typically occur far enough along the distribution curve that the initial pattern would still hold valuable insights into the causes behind the network's structure, which is what we are interested in.

A Starting Point

To begin exploring the structure of Instagram networks with the limited data available, I engaged with a tutorial by Piessen (2019) to extract and visualise my own Instagram network data to gain an initial understanding. Although specifics are not disclosed to respect my network's privacy and other ethical considerations, Figure 1.0 illustrates the degree distribution of my network—a key metric for understanding network topology as it reveals the pattern of connections per node (Artico et al., 2020).

Figure 1.0: Visualising the Degree Distribution of my Instagram Network



Source: Instagram (2024)

Analysing the graph, we observe the in-degree (number of followers) and out-degree (number of accounts followed). Both distributions reveal that most profiles in my network follow and are followed by a small number of other profiles, with a power-law distribution indicating a few nodes with significantly higher connectivity. Following this quantitative analysis, I conducted a mixed-methods tail analysis of my network. By reviewing profiles and their content, and considering various theoretical frameworks, I sought to understand the reasons behind the varying number of connections and inequalities among users. This methodology was then applied to the tail of Cameron's (2022) dataset, focusing on a network revolving around the top 200 Instagram influencers (results to be discussed in section 3).

The subsequent section delves into the applicability of academic theories in interpreting the emergence of power-law distributions to these networks. Section 3 expands on the empirical evidence and theoretical interpretations, setting the stage for a discussion on the implications of these findings and forward-looking recommendations.

2. Attachment Type

In this section, I explore underlying dynamics that shape the formation of scale-free networks, such as Instagram. I investigate why and how certain nodes within these networks become more connected than others. Specifically, I ask why do certain nodes within the network attract more attachments than others?

Preferential Attachment

Central to this discussion is the concept of preferential attachment as outlined by the Barabási-Albert model (1999). It outlines two critical conditions for preferential attachment: the continual addition of new nodes to the system and the capacity of these new nodes to discern and connect with the more popular existing nodes (Barabási and Albert 1999). This model explains the emergence of power law distributions in complex networks, by determining that new nodes are more likely to form connections with already popular ones. Such a tendency echoes what sociologist Robert K. Merton (1968) identified as the "Matthew effect," a socio-economic principle where the "rich get richer and the poor get poorer", originally observed within academic circles where established scientists often receive disproportionate recognition for their contributions compared to lesser-known researchers, even if their work is of similar quality.

On Instagram, this appears to manifest in the dynamics of follower growth and engagement: those who already have substantial followings are more likely to attract new followers due to the model outlined above. This effect is also observed in the application's algorithm, with these users being more likely to be featured on Explore pages or suggested to other users, thus becoming more visible and likely to engage with an even broader audience (Christin and Lu, 2023). This amplification through algorithmic promotion can create a reinforcing loop: as users gain more followers, their content is more frequently highlighted, leading to further follower growth. In contrast, users with smaller followings may find it challenging to achieve the same level of exposure and may continue to be overshadowed within the network. This phenomenon of algorithmic preferential attachment on Instagram could itself be considered a distinct topic of analysis. Given the opaque nature of Instagram's algorithms and the challenges in assessing them, this analysis treats the algorithmic influence as a reinforcing mechanism—essentially a tool that amplifies existing network dynamics driven by user behaviour and pre-existing connections.

My analysis, supported by more recent studies, suggests that this is only part of the story. While the static snapshots of networks often used in studies can mask the temporal aspects of growth, Kunegis et al. (2013) present a dynamic view, finding that networks adhere to a nonlinear preferential attachment model, challenging the notion that networks grow strictly according to a linear preferential attachment model (Johnson et al., 2014). This suggests that online community dynamics are shaped by a variety of social mechanisms, not just a single, universal principle like preferential attachment. I suggest that attributes, such as the intrinsic qualities of a node, also significantly influence the rate at which connections are made. In the case of Instagram, these could include the quality of content, the social identity of the user, or economic factors.

Fitness based Attachment

Following the discussion of preferential attachment, often considered as a given when discussing power law distributions and defined by the 'rich get richer' mechanism, I consider an alternative model offered by Caldarelli et al. (2002), who present a "good-get-richer" mechanism. This theory asserts that the 'fitness' or inherent quality of nodes dictates their level of connectivity in a network, irrespective of the number of connections they already have or the network's age. According to Caldarelli et al. (2002), when the distribution of nodes' fitness possesses a distribution akin to a power law, the network is likely to display a power law degree distribution. Analogous to the concept of utility in economics, a node's fitness in a network might be equated to its appeal or competitiveness (Bell et al., 2017). This 'fitness' could encapsulate a variety of factors that make a node desirable to others in the network—fame, content quality, or strategic positioning.

Privilege Based Attachment

In the landscape of Instagram's network, where connections are currency, I put forth a “privileged get richer" mechanism. This concept, borrowing from Bell et al.’s (2017) reading of the good get-richer mechanism, also views fitness as synonymous with utility—a measure of a node's inherent appeal or desirability. Here, I argue that nodes with inherent advantages are predisposed to attract more connections, enhancing their influence within the network. However, this attraction is not distributed evenly across all nodes, what contributes to one's 'fitness' is often intrinsically linked to systemic inequality (O’Connor, 2023; Christin and Lu, 2023; Mills, 2007). This will be looked at in more detail in the subsequent section.

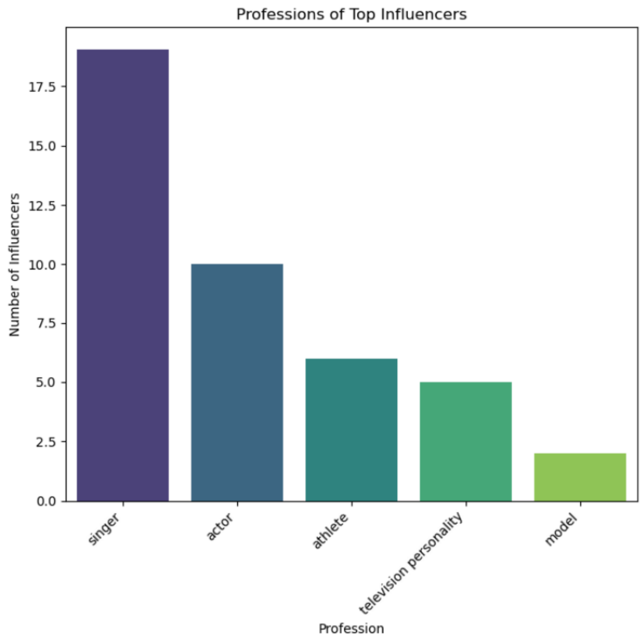
Users do not all start from the same position of privilege. Attributes that contribute to an individual's fitness on Instagram, much like in broader society, are often predetermined and beyond personal control. These can range from socioeconomic status to cultural capital, which can bolster one's visibility or, conversely, render one invisible. This observation echoes the findings of Boy and Uitermark’s (2017) analysis of Instagram networks, who note that inequalities extend beyond material resources such as money to include recognition and visibility. Socially marginalised groups and spaces—like the suburbs, the disabled, the elderly, immigrants, and the homeless—are often relegated to the periphery, while more privileged spaces and individuals are highlighted and celebrated (Boy and Uitermark, 2017). This mechanism combines and adds to the previous mechanisms discussed above, as richness (preferential attachment) and goodness as a utility (fitness-based attachment) can be seen to privilege those who possess them.

3. Barriers to Fitness

The following section will leverage findings from my analysis to support and build on my privilege-based attachment model. Following this framework, my analysis identifies three key attributes that act as forms of privilege for some but pose barriers to fitness for others within the network. These attributes, discovered through the examination of the tail end of Cameron’s (2022) (the top 200 followed), will be outlined as they significantly influence the distribution and flow of connections within Instagram's social web. In the context of our discussion, influencers are defined as hubs within Instagram's network, significant for their vast number of followers rather than their specific occupations.

Class Barriers

Figure 2.0: Main Professions of Top Influencers



Source: Social Blade (2024)

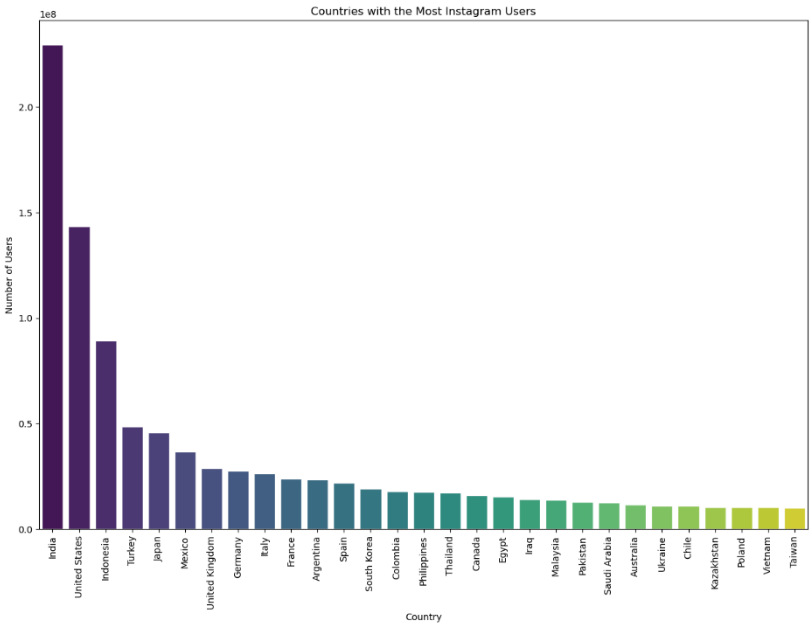
The class dynamics among Instagram's leading figures are revealed by analysing the economic standings of top influencers, which includes taking their professions into consideration. Figure 2.0 indicates a disproportionate representation of individuals from high-earning professions, such as singers and actors, suggesting that financial and social capital are significant precursors to online influence. This pre-established visibility translates into rapid network growth on Instagram, as seen with public figures like Jennifer Aniston whose entry into Instagram was marked by a large influx of followers, setting records for the fastest time to reach 1 million (Sorto, 2019). Aniston's existing popularity allowed her network to essentially migrate with her, highlighting a form of privilege-based attachment- a process where pre-existing societal privilege influence and recognition act as catalysts for exponential growth within and beyond the network, beyond the typical metrics of node connectivity.

As Instagram paves the way for a new creative industry, those hailing from high-profile careers possess a clear advantage (Christin and Lu, 2023). However, this advantage extends beyond the realm of the already famous. Even less renowned individuals benefit from class privileges, such as existing connections to influential actors and resources, allowing them to leverage both preferential attachment and fitness within the platform's dynamics. Time, as much as money, becomes a currency on Instagram. Those not constrained by extensive work hours can devote significant efforts to content creation, curating personas that embody a blend of luxury and relatability, thus attracting a diverse follower base (Giles and Edwards, 2018; Christin and Lu, 2023). Similarly, Marwick’s (2015) analysis of fast-growing Instagram users finds that accessibility to 'Instafame' is skewed towards those who either possess wealth or can effectively simulate an affluent lifestyle, noting that younger users with wealthy parents can amass large followings, with many billionaires’ and famous peoples’ children becoming renowned content creators on the platform. This phenomenon often obscures the class inequalities within the network as these creators are often found to hide the origins and realities of their wealth and privilege (Marwick, 2015). The convergence of class, wealth, and time thus helps shape the network's power structure, aligning with the proposed "privileged get richer" mechanism.

Geographic Barriers

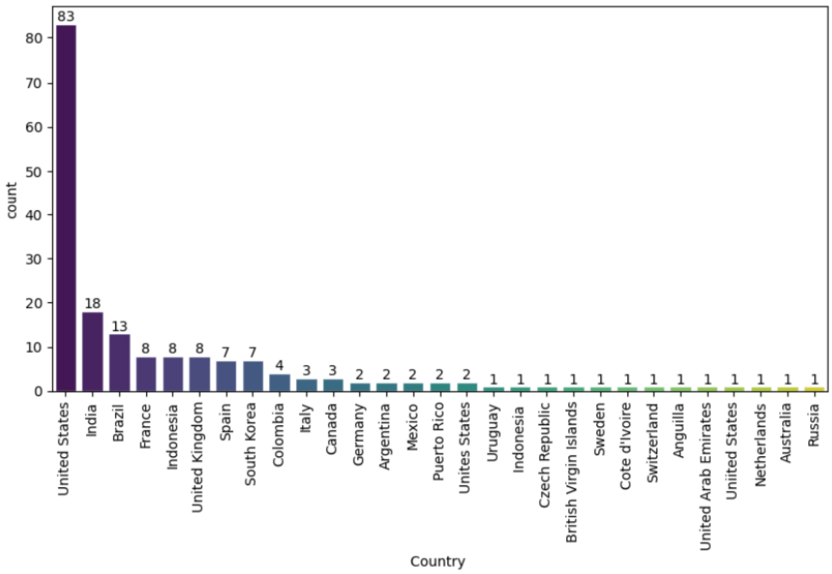
This subsection focuses on the relationship between a user’s geographic location and their network size on Instagram, establishing that high user activity in a given region does not equate to influence. Figures 3.0 and 4.0, representing the number of users and top influencers by country respectively, suggest a disconnect between the widespread availability of the platform and the concentration of influencer status.

Figure 3.0- Number of Instagram users Per country



Source: World Population Review (2024)

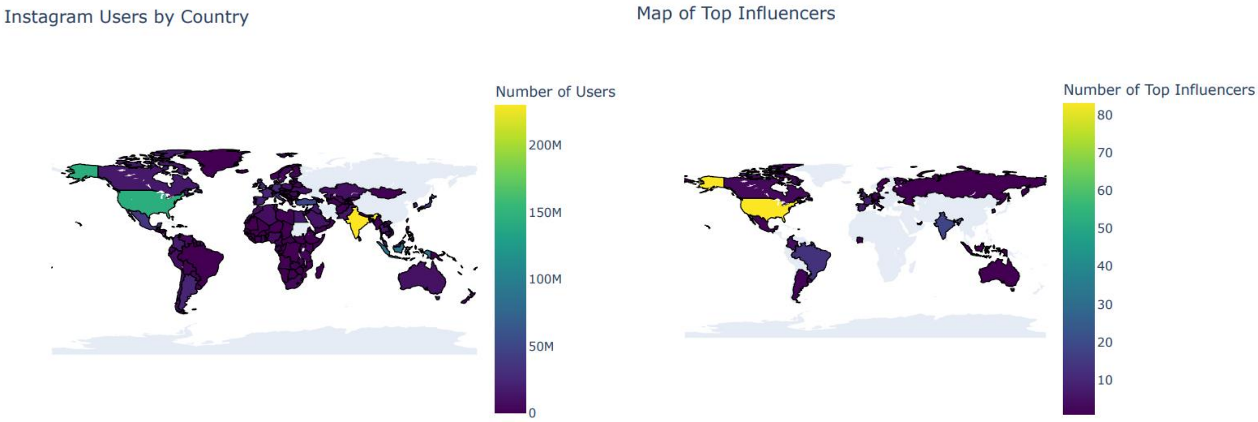
Figure 4.0: Number of Top Influencers per Country



Source: Social Blade (2024)

Moyo's (2017) decolonial framework provides a lens through which to interpret these findings. It posits that access to digital platforms is not solely about the technology; cultural and systemic biases play a crucial role in who becomes visible and who remains in the shadows. This framework can be used to analyse the observed geographic distributions: countries in the Global South may show high numbers of Instagram users (Figure 3.0), yet the map of top influencers (Figure 4.0) reveals a concentration in the Global North, suggesting that factors beyond access influence online prominence.

Figure 5.0: Side-by-side maps of users and top influencers by Country



Sources: Social Blade (2024) and World Population Review (2024)

Figure 5.0, which shows maps of users and top influencers, further highlights this disparity. Countries with significant user bases do not necessarily translate into proportional influencer representation. The cultural dominance of Euro-American content, as discussed by Dutton (2003), contributes to this imbalance by favouring certain languages and aesthetics over others. Such dominance can deter the visibility of local content creators from the Global South, affecting their ability to amass followers and engagement, crucial for influence on the platform.

This skewed distribution can be perceived as a manifestation of a 'privileged get richer' phenomenon within the context of global digital engagement. While the digital divide has traditionally been framed around access (Moyo, 2017), these findings suggest that the divide extends into the realms of recognition and visibility. Even in areas with high Instagram access and usage, like those depicted in the warmer tones in Figure 5, systemic biases embedded within the platform and its user base may suppress the influence of users from these regions.

Thus, through the lens of Moyo's theoretical framework, the figures reveal that the digital divide on Instagram is not only a matter of who is online but also who is seen and heard. The platform's algorithms and content moderation practices, which are not neutral but steeped in cultural biases, are likely to amplify this divide, enabling certain groups to amass influence while others remain marginalised. This suggests a digital landscape where geographic accessibility has improved on the surface, yet the barriers to becoming an influential presence on Instagram are still insurmountably high for many in certain regions.

Racial Barriers

During my ‘tail analysis’, a significant racial imbalance was uncovered—57.14% of the top influencers are White. This finding underscores a systemic issue given that, particularly in the United States where many influencers are concentrated, White Instagram users do not constitute the majority; instead, the platform is predominantly utilised by Black and Hispanic individuals (O’Connor, 2023). The limited data on the exact racial makeup of Instagram's global user base complicates a full understanding of these dynamics but does not obscure the visibility gap evident among influencers. The study by Christin and Lu (2023) supports the existence of this gap. Their research, involving over 500 influencers who remained anonymous, provides a window into the racial composition of influential social media figures: 44% identified as White, 29% as Black, 15% as Asian, and 2% as Hispanic. These statistics reveal a skew towards White influencers that does not mirror the diverse nature of Instagram's user demographics.

O'Connor (2023) delves into the contested nature of online spaces for people of colour, who encounter the same structural barriers on Instagram that position them as less desirable in societal hierarchies. On such a visually-driven platform, the prevailing beauty standards have substantial implications. Instagram's network, largely based on appearance, magnifies the impact of Western European and American beauty ideals, which have been globalised to the point where they dominate the digital sphere (Dimitroc and Kroumpouzos, 2023). These norms, as O'Connor (2023) argues, are intertwined with 'disciplined whiteness,' rewarding those who align with them and marginalising those who do not. Consequently, the platform inadvertently promotes a form of digital colonialism that privileges Western aesthetics, impacting not only which profiles gain influence but also the financial and engagement opportunities available to influencers of colour (Christin and Lu, 2023).

This visual network thus reflects and reinforces racial hierarchies, placing barriers to fitness for influencers who do not fit the Western-centric archetype of beauty and desirability. It can be argued that on Instagram, racial privilege interplays with the "privileged get richer" mechanism, where systemic biases in favour of White influencers catalyse their growth in visibility and influence, further perpetuating the racial divide within the digital landscape.

Implications and Solutions

In unpacking the implications of my findings we observe a pattern of reinforcing inequities and exclusion. Instagram's network, while presenting new opportunities, is skewed by systemic biases that favour certain profiles—typically those aligning with wealth, residing in the global north and whiteness—and, as Christin and Lu (2023) suggest, may even institutionalise a form of automated racism. Influencers from the marginalised groups mentioned in the previous section are more often subject to uncompensated labour and face disparities in brand collaboration and compensation, widening class, geographic and racial divides (O'Connor, 2023; Giles and Edwards, 2018; Christin and Lu, 2023). These systemic biases not only exacerbate existing social and economic divides but also influence who is perceived as credible or desirable on the platform, with significant epistemic consequences for how knowledge and cultural capital are distributed and recognised. This section focuses on the epistemic implications of Instagram's network distribution.

Epistemic Inequalities

As previously mentioned the 'rich get richer' mechanism, foundational to many of the perspectives mentioned including my own, originally emerged from the credit economy of academia, to indicate the unequal nature and distribution of epistemic networks. Instagram, given its substantial influence on public discourse (Christin and Lu, 2023), constitutes a digital epistemic network, where information is circulated or suppressed and connections determine what is seen and to what degree (Cinelli et al., 2021).

This digital landscape, characterised by power law distributions of influence, is vulnerable to Charles Mills' concept of "white ignorance" (Mills, 2007). Mills articulates that this form of ignorance is a constructed lack of knowledge that upholds the racial status quo by selectively omitting certain narratives and emphasising others. The distribution of knowledge on Instagram, thus, not only mirrors but also reinforces the power law distribution of its network. This structure allows a few highly connected, often privileged, nodes to disproportionately shape public knowledge, perpetuating existing geographic, racial, and class hierarchies.

Instagram's epistemic network structure marginalises non-Western and minority cultures, significantly shaping the global information ecosystem (Moyo, 2017). This dominance restricts the diversity of narrative and knowledge, echoing Marwick's (2015) observations about how social and economic divisions influence narrative control. This is because influential figures within the network can manipulate consensus and dictate the prevailing narrative (Holman and Bruner, 2015), embedding systemic biases deeply within the network's structure as we have seen before that privilege determines influence. This is a problem as epistemic networks have been found to benefit from greater epistemic diversity (Wu, 2022; Fazelpour and Steel, 2022) because diversity can help us overcome the Zollman effect, where tightly knit networks may prematurely converge on suboptimal or incorrect beliefs due to homogenous or misleading information (Zollman, 2013).

One proposed solution is the creation of anonymous social networks, such as Whisper or 4chan, which are theoretically egalitarian spaces, free from the influence of existing privileges (Mondal et al., 2020). Drawing on Mills, I argue that anonymity does not address the underlying structural biases that perpetuate 'white ignorance' and inequality—merely 'veiling' the identifiers of race or privilege does not dismantle the entrenched systems that govern knowledge distribution and have the power to change the status quo. The following subsection outlines alternative solutions.

Recommendations and areas for future work

To mitigate the epistemic inequalities identified in social media networks, particularly on platforms like Instagram, a multi-faceted approach involving platform responsibilities, regulatory oversight, and ethical AI development is essential. Platforms can adjust their content recommendation algorithms to balance promoting popular content with discovering and elevating emerging creators. This approach, inspired by Bell et al. (2017), involves diversifying user feeds to prevent echo chambers and introducing a wider range of content, thereby facilitating a more equitable distribution of knowledge and influence. Regulators could develop guidelines that ensure algorithmic transparency and promote diversity, with the goal of preventing the monopolisation of influence by a few nodes. These regulations can include mandatory audits of platform practices and opaque algorithms to ensure compliance with set standards. Lastly, as AI increasingly influences user recommendations, ethical AI development becomes essential. This includes working with diverse teams when designing algorithms to ensure they do not amplify biases or contribute to social divides, ensuring a fair and inclusive online environment.

Conclusion

To conclude, this paper has explored the skewed power law distributions on Instagram, emphasising how a select group of users, dominate connectivity within the network. By introducing the concept of 'privilege-based attachment,' I have highlighted how systemic advantages significantly amplify the influence of certain users over others. The findings underscore that Instagram is not just a platform for social interaction but a space where societal norms and hierarchies are reflected and perpetuated. These distributions are shaped by structural forces that privilege certain nodes, leading to significant disparities in visibility and influence.

As scale-free networks like Instagram play a critical role in shaping public discourse and knowledge dissemination, understanding these dynamics is crucial for policymakers and developers seeking to mitigate digital inequalities and foster a more inclusive digital ecosystem. This study prompts further research into the systemic biases of digital spaces, advocating for equitable visibility algorithms to balance the digital landscape.

References

Artico, I., et al. 2020 “How Rare Are Power-Law Networks Really?” *Proceedings: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2241, pp. 1–17. *JSTOR*, https://www.jstor.org/stable/27097269. Accessed 5 Apr. 2024.

Barabási, A.L. and Albert, R., 1999. Emergence of scaling in random networks. *science*, *286*(5439), pp.509-512.
Bell, M., Perera, S., Piraveenan, M., Bliemer, M., Latty, T. and Reid, C., 2017. Network growth models: A behavioural basis for attachment proportional to fitness. Scientific reports, 7(1), pp.1-11.
Boy, J.D. and Uitermark, J., 2017. Reassembling the city through Instagram. Transactions of the Institute of British Geographers, 42(4), pp.612-624.
Bruner, Justin and Cailin O'Connor, 2017, "Power, Bargaining, and Collaboration," in *Scientific Collaboration and Collective Knowledge*, Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg, eds., Oxford: Oxford University Press. doi:10.1093/oso/9780190680534.003.0007
Caldarelli, G., Capocci, A., Rios, P. and Munoz, M.A., 2002. Scale-free networks without growth or preferential attachment: Good get richer. arXiv preprint cond-mat/0207366.
Cameron, M., 2022. Zipf's Law across social media. University of Waikato.
Chicheportiche R, Bouchaud J-P. 2012 Weighted Kolmogorov-Smirnov test: accounting for the tails. Phys. Rev. E 86, 041115. (doi:10.1103/PhysRevE.86.041115)
Christin, A. and Lu, Y., 2023. The influencer pay gap: Platform labor meets racial capitalism. new media & society, p.14614448231164995.
Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M., 2021, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, 118(9): e2023301118.
Dimitrov, D. and Kroumpouzos, G., 2023. Beauty perception: A historical and contemporary review. Clinics in Dermatology, 41(1), pp.33-40.
Dutton, W.H. (2003). "Information Society." In Encyclopedia of international media and communications (Vol. 2). London: Academic Press, pp. 407–418. Edited by Downing Johnston
Fazelpour, S., & Steel, D. (2022). The Zollman effect and the epistemic value of diversity in scientific communities. Philosophical Studies, 1-22.
Giles, D.C. and Edwards, L., 2018. Instagram and the rise of the social media 'influencer'. In Twenty-first century celebrity: Fame in digital culture (pp. 155-173). Emerald Publishing Limited.

Holman, Bennett & Bruner, Justin P., 2015, “The Problem of Intransigently Biased Agents,” *Philosophy of Science*, 82(5): 956–968. doi:10.1086/683344

Instagram (2024)*. (@hayakassem) • Instagram photos and videos* (no date). Available at: https://www.instagram.com/hayakassem (Accessed: 3 May 2024).

Johnson, S.L., Faraj, S. and Kudaravalli, S., 2014. Emergence of power laws in online communities. *Mis Quarterly*, *38*(3), pp.795-A13.

Kunegis, J., Blattner, M. and Moser, C., 2013, May. Preferential attachment in online networks: Measurement and explanations. In Proceedings of the 5th annual ACM web science conference (pp. 205-214).

Marwick, A.E., 2015. Instafame: Luxury selfies in the attention economy. Public culture, 27(1 (75)), pp.137-160.

Mills, Charles, 2007, “White Ignorance,” in Shannon Sullivan & Nancy Tuana, eds., *Race and Epistemologies of Ignorance*, Albany: State Univ of New York Press, 11–38.

Mondal, M., Correa, D. and Benevenuto, F., 2020, July. Anonymity effects: A large-scale dataset from an anonymous social media platform. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (pp. 69-74).

Moyo, L., 2017. Rethinking the information society: A decolonial and border gnosis of the digital divide in Africa and the Global South. In *Theorizing digital divides* (pp. 133-145). Routledge.

O'Connor, S., 2023. The racial economy of Instagram. Transactions of the Institute of British Geographers.

Social Blade (2024) *Top 250 Instagram Business Accounts sorted by Followers - Socialblade Instagram Stats* | *Instagram Statistics*. Available at: https://socialblade.com/instagram/top/250 (Accessed: 2 May 2024).

Sorto, G. (2019) *Jennifer Aniston broke Instagram with her debut*, *CNN*. Available at: https://www.cnn.com/2019/10/16/entertainment/jennifer-aniston-broke-instagram-trnd/index.html (Accessed: 20 April 2024).

Weatherall, James Owen, Cailin O’Connor, and Justin P. Bruner, 2020, “How to Beat Science and Influence People: Policymakers and Propaganda in Epistemic Networks,” *The British Journal for the Philosophy of Science*, 71(4): 1157–1186.

World Population Review (2024) *Instagram Users by Country 2024*. Available at: https://worldpopulationreview.com/country-rankings/instagram-users-by-country (Accessed: 2 May 2024).

Wu, J. (2022). The Zollman effect and the epistemic value of demographic diversity. Philosophical Studies, 1-20.

Zollman, Kevin J. S., 2007, “The Communication Structure of Epistemic Communities,” *Philosophy of Science*, 74(5): 574–587. Reprinted in Goldman & Whitcomb, eds., 2011, 338–350. doi:10.1086/525605

——, 2010, “The Epistemic Benefit of Transient Diversity,” *Erkenntnis*, 72(1): 17–35. doi:10.1007/s10670-009-9194-6

Appendix

Code used inspired by: Ceviz, F. *Top200 Influencers Data Analysis* (2022). Available at: https://kaggle.com/code/fethullahceviz/top200-influencers-data-analysis (Accessed: 2 May 2024).

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import re
import matplotlib.pyplot as plt
import missingno as msno
import seaborn as sns
import plotly.graph_objects as go
import plotly.express as px
from plotly.offline import init_notebook_mode, iplot
init_notebook_mode(connected=True)
import os
df=pd.read_csv(r'C:\Users\hayak\OneDrive\Documents\Scaling
Data\instagram analysis file.csv')
df['channel_info'] = df['channel_info'].str.replace('\n','')
number_liste=[ 'followers', 'avg_likes', 'posts','new_post_avg_like','total_likes']
for i in number_liste:
```

```
    lst =  df[i]
    tbl = {'k':1000, 'm': 1000000, 'b': 1000000000,}
    df[i] = [int(i) for i in (re.sub(r'([\d\.]+)(k|m|b)', lambda v:
str(int(float(v.groups()[0]) * tbl[v.groups()[1]])), i) for i in lst)]
df1 = df.groupby('Country ').size().reset_index(name='Size')
import pandas as pd
import plotly.graph_objects as go

# Load the datasets
df_users = pd.read_csv(r'C:\Users\hayak\OneDrive\Documents\Scaling
Data\instagram-users-by-country-2024.csv')
df_top_followers =
pd.read_csv(r'C:\Users\hayak\OneDrive\Documents\Scaling Data\instagram
analysis file.csv')

df_users_grouped = df_users.groupby('Country ').agg({'Instagram Users':
'sum'}).reset_index()
df_top_followers_grouped = df_top_followers.groupby('Country
').size().reset_index(name='Followers Count')

fig_users = go.Figure(data=go.Choropleth(
    locations=df_users_grouped['Country '],
    z=df_users_grouped['Instagram Users'],
    locationmode='country names',
    colorscale='Viridis',
    colorbar_title="Number of Users",
    marker_line_color='black'
))
fig_users.update_layout(
    title_text='Instagram Users by Country',
    geo=dict(
        showframe=False,
        showcoastlines=False
    )
```

```
)
fig_users.show()
fig_followers = go.Figure(data=go.Choropleth(
    locations=df_top_followers_grouped['Country '],
    z=df_top_followers_grouped['Followers Count'],
    locationmode='country names',
    colorscale='Viridis',
    colorbar_title="Number of Top Influencers",
    marker_line_color='black'
))
fig_followers.update_layout(
    title_text='Map of Top Influencers',
    geo=dict(
        showframe=False,
        showcoastlines=False
    )
)
fig_followers.show()
df_top_29 = df_users.sort_values('Instagram Users',
ascending=False).head(29)
# Function to create the count plot without bar labels
def count_plot_no_labels(data, x, y, title, x_label, y_label):
    plt.figure(figsize=(x, y))
    ax = sns.barplot(x='Country ', y='Instagram Users', data=data,
palette="viridis", edgecolor='white', linewidth=1, order=data['Country '])
    plt.xticks(rotation=90)
    plt.title(title)
    plt.xlabel(x_label)
    plt.ylabel(y_label)
    # Removed the line that adds labels to each bar
    plt.show()

# Call the function to create the plot without labels
```

count_plot_no_labels(df_top_29, 15, 10, 'Countries with the Most Instagram Users', 'Country', 'Number of Users')

Link to Zip file of tutorial I used, adjustments I made to some of the scripts are included before: https://medium.com/@maximpiessen/how-i-visualised-my-instagram-network-and-what-i-learned-from-it-d7cc125ef297#:~:text=We%20see%20that%20the%20distributions,for%20in%20and%20out%20degree.

```
import random
import time
import re
import collections
import sys
class Bot:
    def __init__(self):
        options = webdriver.ChromeOptions()
        options.add_argument('--no-sandbox')
        options.add_argument('--disable-gpu')
        options.add_argument("--lang=en")
        self.times_restarted = 0
        self.driver = webdriver.Chrome(options=options)
        self.driver.implicitly_wait(20)

    def tear_down(self):
        self.driver.quit()

    def go_to_page(self, url):
        try:
            self.driver.get(url)
        except NoSuchElementException as ex:
            self.fail(ex.msg)



    def login(self, username, password):
```

```
        self.driver.get("https://www.instagram.com")
        # Wait for the cookie popup to be clickable and dismiss it
        WebDriverWait(self.driver, 20).until(
            EC.element_to_be_clickable((By.XPATH, "//button[text()='Allow all
cookies']"))
        ).click()

        # Fill in the username and password
        username_field = WebDriverWait(self.driver, 20).until(
            EC.presence_of_element_located((By.NAME, "username"))
        )
        username_field.send_keys(username)

        password_field = WebDriverWait(self.driver, 20).until(
            EC.presence_of_element_located((By.NAME, "password"))
        )
        password_field.send_keys(password)

        # Submit the login form
        submit_button = WebDriverWait(self.driver, 20).until(
            EC.element_to_be_clickable((By.XPATH, "//button[@type='submit']"))
        )
        submit_button.click()

        # Wait for the 'Save Your Login Info?' prompt and dismiss if present
        WebDriverWait(self.driver, 20).until(
            EC.element_to_be_clickable((By.XPATH, "//button[text()='Not now']"))
        ).click()

        # Add a wait here to give time for any other potential popups
        time.sleep(5)

    def get_my_followers(self, username):
        self.go_to_page("https://instagram.com/" + username + "/")
```

```
        time.sleep(5)
        my_followers_set = set()
        followers = self.driver.find_elements_by_class_name("-nal3")
        followers[1].click()
        time.sleep(2)
        initialise_vars = 'elem = document.getElementsByClassName("isgrP")[0];
followers =
parseInt(document.getElementsByClassName("g47SY")[1].innerText); times =
parseInt(followers * 0.14); followersInView1 =
document.getElementsByClassName("FPmhX").length'
        initial_scroll = 'elem.scrollTop += 500'
        next_scroll = 'elem.scrollTop += 1500'

        with open('./jquery-3.3.1.min.js', 'r') as jquery_js:
            # 3) Read the jquery from a file
            jquery = jquery_js.read()
            # 4) Load jquery lib
            self.driver.execute_script(jquery)
            # scroll down the page
            self.driver.execute_script(initialise_vars)
            #self.driver.execute_script(scroll_followers)
            self.driver.execute_script(initial_scroll)
            time.sleep(3)

            next = True
            while(next):
                n_li_1 = len(self.driver.find_elements_by_class_name("FPmhX"))
                self.driver.execute_script(next_scroll)
                time.sleep(1.5)
                n_li_2 = len(self.driver.find_elements_by_class_name("FPmhX"))
                if(n_li_1 != n_li_2):
                    following = self.driver.find_elements_by_xpath("//*[contains(text(),
'Following')]")
                    for follow in following:
```

```
                el = follow.find_element_by_xpath('../..')
                el = el.find_element_by_tag_name('a')
                profile = el.get_attribute('href')
                my_followers_set.add(profile)
            else:
                next = False

        return list(my_followers_set)

    def get_followers(self, my_followers_arr, start_profile, relations_file):
        n_my_followers = len(my_followers_arr)
        count_my_followers = start_profile - 1

        for current_profile in my_followers_arr[start_profile - 1 : -1] +
[my_followers_arr[-1]]:
            print("Start scraping " + current_profile)
            self.go_to_page(current_profile)
            time.sleep(random.randint(5, 20))
            last_5_following = collections.deque([1, 2, 3, 4, 5])  # queue to keep
track of Instagram blocking scroll requests
            count_my_followers += 1

            with open('start_profile.txt', 'w+') as outfile: # keep track of last profile
checked
                outfile.write(str(count_my_followers))

            followers = self.driver.find_elements_by_class_name("-nal3")
            followers[2].click()
            time.sleep(2)
            initialise_vars = 'elem =
document.getElementsByClassName("isgrP")[0]; followers =
parseInt(document.getElementsByClassName("g47SY")[1].innerText); times =
parseInt(followers * 0.14); followersInView1 =
document.getElementsByClassName("FPmhX").length'
```

```
        initial_scroll = 'elem.scrollTop += 500'
        next_scroll = 'elem.scrollTop += 2000'

        with open('./jquery-3.3.1.min.js', 'r') as jquery_js:
            # 3) Read the jquery from a file
            jquery = jquery_js.read()
            # 4) Load jquery lib
            self.driver.execute_script(jquery)
            # scroll down the page
            self.driver.execute_script(initialise_vars)
            # self.driver.execute_script(scroll_followers)
            self.driver.execute_script(initial_scroll)
            time.sleep(random.randint(2, 5))

            next = True
            follow_set = set()
            # check how many people this person follows
            nr_following =
int(re.sub(",","",self.driver.find_elements_by_class_name("g47SY")[2].text))

            n_li = 1
            while next:
                print(str(count_my_followers) + "/" + str(n_my_followers) + " " +
str(n_li) + "/" + str(nr_following))
                time.sleep(random.randint(7, 12) / 10.0)
                self.driver.execute_script(next_scroll)
                time.sleep(random.randint(7, 12) / 10.0)
                if not (n_li < nr_following - 11):
                    next = False

                n_li = len(self.driver.find_elements_by_class_name("FPmhX"))
                last_5_following.appendleft(n_li)
                last_5_following.pop()
```

```
                # if instagram starts blocking requests, reload page and start
again
                if len(set(last_5_following)) == 1:
                    print("Instagram seems to keep on loading. Refreshing page in
7 seconds")
                    self.times_restarted += 1
                    if self.times_restarted == 4:
                        print("Instagram keeps on blocking your request.
Terminating program. Start it again later.")
                        sys.exit()
                    time.sleep(7)
                    self.get_followers(my_followers_arr, count_my_followers,
relations_file)

            self.times_restarted = 0

            following = self.driver.find_elements_by_class_name("FPmhX")
            for follow in following:
                profile = follow.get_attribute('href')
                if profile in my_followers_arr:
                    follow_set.add((current_profile, profile))

            with open(relations_file, "a") as outfile:
                for relation in follow_set:
                    outfile.write(relation[0] + " " + relation[1] + "\n")

            print("This person follows " + str(len(follow_set)) + " of your
connections. \n")

        sys.exit()
```