Beyond Access: Contextualizing the Benefits of Broadband through Contributor Dynamics on Wikipedia

Yaxuan Yin
University of Wisconsin–Madison
Madison, Wisconsin, USA
yaxuan.yin@wisc.edu

Jacob Thebault-Spieker
University of Wisconsin–Madison
Wisconsin, USA
jacob.thebaultspieker@wisc.edu

Abstract

Broadband infrastructure is often assumed to reduce informational disparities by expanding access to digital platforms. Yet less is understood about how broadband shapes participation in peer production communities, where knowledge is collectively created and maintained. Using spatial regression models, we examine how broadband coverage influences who contributes and how participation patterns shift in geo-tagged Wikipedia edits across U.S. counties. We find that broadband expansion is strongly associated with increased contributions from local casual and regular editors while reducing reliance on bot-driven activity. However, contributions remain highly concentrated, as prolific editors continue to dominate production. Moreover, we uncover spatial spillover effects, where broadband gains in one county decrease participation in neighboring areas, revealing competitive dynamics in peer production. These findings challenge the assumption that access alone fosters equity, showing that broadband reshapes but does not evenly redistribute editorial influence, with implications for infrastructure policy, platform design, and sustaining inclusive peer production.

CCS Concepts

• **Human-centered computing → Empirical studies in collaborative and social computing**; **Empirical studies in HCI**.

Keywords

Wikipedia, Broadband Access, Infrastructure, Participation Inequality, Peer Production

ACM Reference Format:
Yaxuan Yin and Jacob Thebault-Spieker. 2026. Beyond Access: Contextualizing the Benefits of Broadband through Contributor Dynamics on Wikipedia. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3772318.3790903

1 Introduction

Broadband Internet access has become a critical component of modern infrastructure, shaping opportunities for education, economic development, civic engagement, healthcare, and social participation [58, 70]. In the United States, federal initiatives such as the 2021 Infrastructure Investment and Jobs Act and the Broadband Equity, Access, and Deployment (BEAD) program have allocated billions of dollars to expand high-speed Internet access nationwide. These investments reflect a growing recognition that broadband is not just a technical resource but a driver of equity, influencing who can learn, work, and participate in public life. Yet, while broadband's impact has been examined in areas such as health informatics [22], local economic development [69], and education [11], its influence on peer production systems remains poorly understood. Peer production systems are collaborative online platforms in which users create and maintain shared knowledge. Platforms such as Wikipedia, play a central role in shaping whose knowledge is produced, whose voices are represented, and how communities are made visible online. Understanding how broadband access shapes participation in these systems is essential for contextualizing the broader impact of digital infrastructure on knowledge equity and for informing policies and platform practices that aim to support more inclusive forms of contribution.

The Wikimedia Foundation has likewise noted that internet access is a key determinant of who can participate in Wikipedia and whose knowledge becomes represented on the platform [59]. Wikipedia, the largest peer production project, serves as a global reference infrastructure that organizes and disseminates knowledge across domains and geographies. Its geo-tagged articles, in particular, provide a rich repository of place-based information, capturing local histories, landmarks, and community identities [28, 29]. This knowledge does not merely serve human readers, it also forms a foundational input for computational systems. For example, large language models (LLMs) such as GPT-4 and Gemini ingest Wikipedia content to build general-purpose knowledge about people, places, and events [55]. As a result, geographic disparities in Wikipedia's coverage can propagate into downstream systems, shaping what knowledge AI models represent—and fail to represent—about different regions [2]. Understanding how infrastructural conditions such as broadband access shape participation in Wikipedia therefore has consequences not only for improving digital knowledge equity but also for ensuring fairness and representational balance in AI-driven systems.

Despite Wikipedia's open and collaborative design, which is often described as a platform where "anyone can edit almost every page" [72], content production remains highly uneven across geographic contexts. Prior research has documented stark disparities in the volume and quality of Wikipedia content across urban and rural areas, affluent and underserved communities, and Global North and Global South regions [7, 28, 37, 64]. Within the United States, geotagged Wikipedia content is disproportionately concentrated in densely populated, well-connected counties, while rural and socioeconomically disadvantaged regions often remain underrepresented

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3790903

[44, 49, 63]. In addition to these geographic disparities, peer production systems such as Wikipedia and OpenStreetMap also exhibit strong contribution inequalities, where a small number of highly active contributors perform most of the editorial labor [64, 76]. Such concentration of effort can introduce risks for data quality and for the long-term sustainability of contributor communities. Taken together, these patterns raise concerns about whether peer production models adequately serve marginalized communities, whose perspectives often remain underrepresented, and about how much space exists for locally grounded or less-experienced contributors to influence content [35]. However, most prior work has explained geographic disparities by using population density as a proxy for socioeconomic resources and for urban–rural divides. Much less attention has been given to the role of broadband infrastructure, a foundational condition for digital participation that may shape both where contributions occur and who is able to participate.

Our work addresses this gap by examining how broadband access relates to patterns of participation in geo-tagged Wikipedia content, guided by two central research questions:

- RQ1: How does broadband coverage inform participation dynamics?
- RQ2: How do these patterns reflect contributor participation in regions with different broadband coverage?

Focusing on counties in the United States, we combine geo-tagged Wikipedia edit data with county-level broadband coverage to investigate how digital infrastructure relates to participation. We account for population density as a proxy for urban–rural and economic divides, and we employ spatial regression models to capture geographic dependencies in contribution patterns. This design enables us to estimate both the direct effects of broadband access within a county and the spillover effects that occur when nearby counties expand their broadband coverage. Our study makes the following contributions:

- We show that higher broadband access is associated with increased participation in Wikipedia, including more contributors and more geo-tagged content. However, increased access does not reduce the underlying concentration of contribution effort. A small group of prolific contributors continues to account for most editing activity, reflecting the persistence of power-law participation patterns.
- We find that broadband expansion is linked to more diverse participation, particularly through increased involvement from locally oriented and casual contributors. These contributors broaden the range of place-based perspectives represented in Wikipedia, even though they contribute a relatively small share of total edits compared to highly prolific users.
- We contextualize the benefits of broadband by highlighting the structural dynamics that shape participation in peer production. While broadband access is a necessary condition for engagement, it is not sufficient for ensuring equitable knowledge creation. Our findings underscore the need for complementary strategies in platform design, governance, and policy to support more inclusive participation.

2 Related Work

Our study builds on three strands of related work: (1) Infrastructure and Inequity in Digital Participation, which frames broadband access as essential for inclusion but emphasizes that access alone does not ensure meaningful engagement. (2) Geographic Disparities in Peer Production, which document how knowledge creation on platforms like Wikipedia remains unevenly distributed across regions. (3) Contribution Inequality in Peer Production, which shows that a small fraction of highly active contributors produce the bulk of content, reinforcing long-standing inequalities in visibility, influence, and whose perspectives shape the knowledge commons.

2.1 Infrastructure and Inequity in Digital Participation

Access to broadband internet has become foundational to civic life, education, and participation in digital knowledge production. Recognizing its importance, governments and international organizations have invested heavily in expanding connectivity. In the United States, federal initiatives such as the FCC's Rural Digital Opportunity Fund and the Broadband Equity, Access, and Deployment (BEAD) program target underserved areas [52]. At the global level, programs like the EU's Gigabit Society strategy [14] and the UN's Broadband Commission frame broadband as essential infrastructure for achieving the Sustainable Development Goals (SDGs) [16, 74]. Despite these efforts, broadband access remains uneven: nearly 24 million Americans still lack coverage—particularly in rural, Tribal, and low-income communities [16]—and global access is even more unequal, with just 27% of people in low-income countries connected, compared to 92% in high-income countries [42].

However, broadband infrastructure alone does not ensure that individuals can or will participate equitably in digital knowledge production. Even with improved technical access, longstanding structural, economic, and design-related barriers continue to shape who is able to contribute and under what conditions. Prior work shows that when broadband is unreliable, intermittent, or prohibitively expensive, participation often becomes selective, strategic, and shared across households or communities rather than continuous and individual [12]. For example, residents in Havana created StreetNet (SNET), a grassroots mesh network that enabled local information exchange outside of costly and state-controlled internet infrastructure. Likewise, in Bangladesh and other low-bandwidth regions, people navigate infrastructural limitations through offline-first practices, shared-access devices, and informal distribution networks for digital content. These behaviors represent more than pragmatic improvisations; they reflect collective forms of care, resource sharing, and alternative participation models developed in response to structural constraints [60].

Scholars have also called attention to how dominant narratives of digital participation often render these practices invisible. Postcolonial and feminist perspectives within HCI argue that the needs and realities of marginalized communities—particularly in the Global South—are often excluded from platform design and infrastructural planning [66]. Even when access is technically available, design norms, onboarding systems, and sociotechnical expectations may discourage or prevent meaningful contribution. As Graham et al.

[28] and others have noted, broadband expansion alone does not dismantle the structural inequalities that determine whose knowledge is produced and recognized.

Taken together, this literature positions broadband access as a sociotechnical condition: it is a foundational enabler of digital participation, but not a sufficient one. The benefits of connectivity remain unequally distributed, and without inclusive design and community-centered practices, infrastructure may reinforce rather than reduce disparities in who contributes and who is represented in open knowledge systems.

2.2 Geographic Disparities in Peer Production

Prior research has consistently demonstrated substantial geographic disparities across peer production platforms, particularly Wikipedia and OpenStreetMap. These disparities often align with socioeconomic and urban–rural divides, resulting in uneven representation across different regions of the world [31, 37, 44, 64, 76, 77]. Such gaps are not merely reflective of contributor preferences, but stem from structural inequalities in digital infrastructure, contributor concentration, and regional visibility.

Globally, Wikipedia has been shown to disproportionately feature people, places, and events associated with the Global North and Western countries. Beytía [4] quantified this imbalance across multilingual editions, showing that articles linked to the Global South are significantly underrepresented. Hickman et al. [40] found that similar patterns can manifest across more and less well-resourced languages even within a Global South regions as well. Similarly, Graham et al. [28] revealed that articles from more geographically central and well-connected regions tend to attract more edits and visibility, amplifying already dominant knowledge geographies.

At the subnational level, the divide between urban and rural areas is particularly striking. Johnson et al. [44] found that articles about rural U.S. towns are often minimal and rely heavily on bot-generated templates rather than community-authored content. This pattern reflects a lack of active contributors in these areas and mirrors similar dynamics in OpenStreetMap, where low-density, low-income regions are often sparsely mapped or entirely absent [53]. Prior work has attributed these gaps in part to "self-focus bias", where contributors are more likely to add information about locations they are physically or emotionally connected to [36].

Infrastructure and contributor availability further shape the geography of participation. Warncke-Wang et al. [68] show that mismatches between where contributors are active and where high-quality content is most needed often result in underdeveloped articles, particularly in less populated or underrepresented regions. In areas with limited broadband access or fewer digitally literate contributors, knowledge gaps can emerge and persist despite Wikipedia's open-access model. Graham et al. [28] similarly argue that participation on Wikipedia is deeply uneven, noting that some African countries have fewer geo-tagged articles than individual cities in Europe or North America. Even targeted efforts to encourage global participation, such as edit-a-thons [46, 73], struggle to counteract the concentration of editorial activity in already well-represented regions [19, 46].

2.3 Contribution Inequality in Peer Production

Wikipedia, like many peer production systems, relies on contributions from both humans and automated agents. Given the scale and complexity of maintaining a global encyclopedia, bots have become central to sustaining large-scale knowledge production. By 2009, bots were responsible for approximately 28.49% of all edits to the English-language Wikipedia, making them a major force in shaping the platform's content and structure [23, 67]. Bots are not a monolithic group, they vary in purpose, functionality, and affiliation [9, 25, 33]. For example, Geiger and Ribes [26] categorize bots into three main types: (1) content bots that make direct edits (e.g., fixing typos or enforcing formatting standards), (2) task bots that support editorial workflows (e.g., tagging uncategorized articles or detecting vandalism), and (3) community bots that engage with contributors (e.g., welcoming newcomers or enforcing community guidelines). Bots have been credited with improving Wikipedia's efficiency and content quality, particularly by handling high-volume, repetitive tasks. However, bots can also introduce challenges to community dynamics. The centralization of bot activity, where a small number of operators control the majority of automated contributions, raises concerns about transparency and participatory equity in what is intended to be an open and decentralized system [25]. Additionally, bot interventions can unintentionally undermine creativity, collaboration, and the sense of community belonging, particularly when newcomers interpret automated messages or reverts as impersonal or overly bureaucratic [25, 41].

Turning to human contributors, editing trajectories follow recognizable patterns: many editors make only brief or sporadic contributions, while a small "core" group sustains long-term, intensive engagement [32, 56]. At the system level, Wikipedia and other peer production platforms such as OpenStreetMap exhibit strong power-law contribution dynamics [36, 63, 76, 78], in which a small fraction of highly active users produces a disproportionately large share of all content [15]. Contributors participate for varied reasons—including altruism, ideological commitments, domain interest, enjoyment of editing, and a sense of belonging [1, 21, 54]—yet these individual motivations aggregate into strongly skewed patterns of labor. Social and organizational structures further reinforce these dynamics. Contributors with elevated roles or institutional status (e.g., administrators) tend to remain highly active, and governance norms often reflect their practices [1]. As a result, a small group of prolific editors exerts outsized influence over editorial conventions, content standards, and the interpretive frames through which knowledge is documented [24, 32]. These entrenched norms can pose challenges for newcomers, whose contributions may be misaligned with expectations shaped by a narrow contributor cohort [63]. Moreover, prior work shows that power-law contribution dynamics deepen geographic and topical imbalances [30] and make it difficult to sustain diverse contributor communities. Taken together, these forms of contribution inequality can constrain whose perspectives ultimately shape the knowledge commons and which communities remain underrepresented.

3 Methodology

In this study, we investigate how broadband access influences geographically anchored Wikipedia editing activity across U.S. counties.

A central methodological challenge is spatial dependence: neighboring counties often share infrastructural conditions, demographic characteristics, and contributor behaviors [65]. To address this, we employ a Spatial Durbin Model (SDM), which explicitly models spatial autocorrelation in both the predictors and the outcome. The subsections that follow describe our modeling specifications (Section 3.1), variable selection (Sections 3.3 and 3.4), robustness checks (Section 3.5), and additional methodological details.

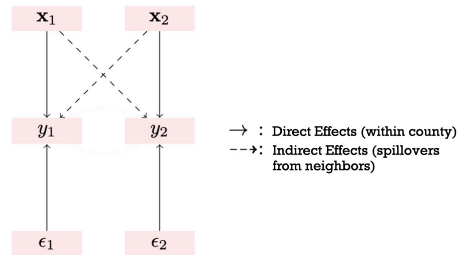


Figure 1: Conceptual illustration of the Spatial Durbin Model (SDM). Solid arrows represent direct effects, which capture within-county relationships between broadband availability and Wikipedia participation. Dashed arrows represent indirect effects, which capture spatial spillovers through which broadband conditions in neighboring counties influence participation in the focal county.

3.1 Analytical Approach: Spatial Durbin Model

Participation on Wikipedia does not emerge independently across counties. Instead, it is shaped by shared regional infrastructures, demographic conditions, and knowledge-production dynamics [65]. When such spatial processes are present, traditional regression models become inappropriate because their assumption of independent errors is violated.

To assess whether this issue applies to our data, we first estimated a non-spatial OLS model for each outcome and examined the residuals. As shown in Appendix A, the residual maps reveal clear geographic clustering: counties in similar regions tend to have systematically positive or negative residuals. This pattern indicates that OLS leaves meaningful spatial structure unexplained. We then formally tested for spatial dependence using Global Moran's I with a Queen contiguity spatial-weights matrix. Across all participation measures included in our analysis (see Appendix A), the residuals exhibit strong and statistically significant spatial autocorrelation. Taken together, these diagnostics show that a standard OLS specification is misspecified and motivate the use of a spatial regression framework that explicitly accounts for spatial dependence.

To address this, we apply the Spatial Durbin Model (SDM) [51]. As illustrated in Figure 1, the SDM captures two key pathways: (1) direct effects, which represent the relationship between broadband availability and Wikipedia participation within a county, and (2) indirect effects, which represent spatial spillovers through which broadband conditions in neighboring counties influence participation in the focal county. This structure provides a more realistic representation of geographically patterned participation than models that assume counties behave independently. The SDM is widely used in spatial econometrics and has been applied in geo-HCI research on digital participation [8, 65]. We define spatial relationships using a Queen contiguity matrix, in which counties that share either a boundary or a vertex are considered neighbors. As illustrated in Appendix Figure 7, this specification is well suited for county-level analysis, because diffusion can occur through both shared borders and corner adjacency. As a robustness check, we also estimated results using a Rook contiguity matrix and found that the substantive findings remained consistent.

For each dependent variable, we estimate a separate SDM that includes broadband coverage and log-transformed population density, along with spatially lagged terms that represent the influence of neighboring counties. Following standard practice in prior work [8, 65], we report *direct* and *indirect* effects rather than raw model coefficients to ensure the interpretability of our results. *Direct* effects estimate how broadband coverage relates to participation within a focal county, while *indirect* effects capture spillover relationships, how broadband levels in surrounding counties are associated with outcomes in the focal county.

3.2 Spatial Unit of Analysis

In this study, we analyze 3,133 of the 3,143 U.S. counties, excluding those without residential broadband coverage data to ensure consistency across models. Counties provide both a policy-relevant and analytically robust unit of analysis. Federal broadband initiatives rely on county-level reporting for funding allocation and infrastructure assessment [17, 52, 70], so aligning our study with this scale enhances the practical relevance of our findings. Counties also represent one of the smallest geographic units for which broadband, demographic, and user-generated content data can be reliably aggregated. This makes them particularly well suited for integrating diverse datasets and for detecting spatial disparities in participation. Moreover, county-level analysis is well established in research on digital inequality and socio-technical systems [39, 44, 53, 65], providing a strong methodological foundation for our approach.

3.3 Independent and Control Variables

Our primary independent variable of interest is county-level broadband coverage, obtained from the Federal Communications Commission (FCC) Broadband Data Collection (BDC) dataset released in December 2024. We specifically focus on residential fixed wired service coverage, defined as the percentage of locations within a county with access to connections offering at least 100 Mbps download and 20 Mbps upload speeds (100/20). This benchmark reflects the federal standard defined by the Infrastructure Investment and Jobs Act (IIJA) and adopted by the National Telecommunications and Information Administration (NTIA) [52], and it serves as a meaningful threshold for enabling modern digital participation.

To account for structural conditions that may independently shape participation [44, 76], we include population density as a control variable. Population density is calculated as residents per square mile using 2023 U.S. Census estimates. Because the distribution is highly skewed (skewness = 26.50), we apply a base-2 logarithmic transformation, which reduces skewness to 0.01 and enables more meaningful comparisons across counties. Controlling for population density is essential because densely populated counties tend to have greater internet infrastructure investments, higher concentrations of active contributors, and more coverage in user-generated platforms [63]. Accounting for this variation allows us to isolate the association between broadband availability and Wikipedia participation from underlying urban–rural differences that might otherwise confound the relationship.

3.4 Dependent Variables

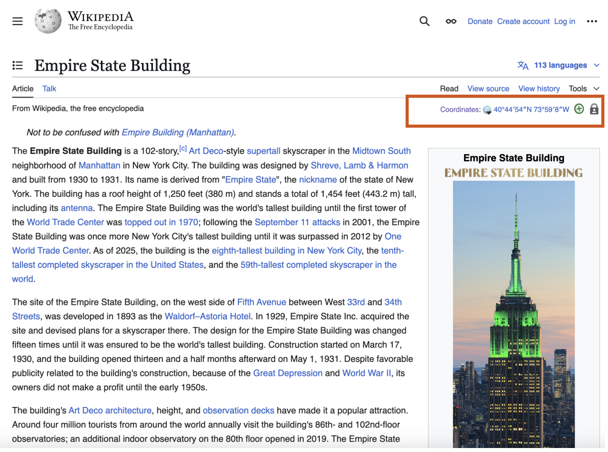


Figure 2: An example of a geo-tagged English Wikipedia page. The article on the Empire State Building includes latitude and longitude metadata, which enables spatial assignment to a U.S. county.

We began by extracting English Wikipedia articles and selecting those that contain valid geographic coordinates from an English Wikipedia dump file dated December 10, 2024. This date aligns with the temporal coverage of our broadband data. Using each article's latitude and longitude metadata (Figure 2), we assigned articles to U.S. counties by performing a spatial overlay with TIGER/Line county boundaries from the U.S. Census Bureau. We then retrieved complete edit histories for these geo-tagged articles from Wikipedia database dumps, enabling us to identify contributors and construct the following set of participation measures.

3.4.1 Participation Volume and Concentration. For RQ1, which examines overall participation and contribution dynamics, we aggregate all participation metrics at the county level (see Table 1). These measures serve as the dependent variables in our Spatial Durbin Models. We estimated seven SDMs in total, two of which use log-transformed outcomes—ln(Number of Pages) and ln(Number of Unique Contributors)—to normalize skewed distributions and facilitate interpretation in terms of percent change.

- Number of Wikipedia Pages: We counted the total number of geo-tagged Wikipedia articles assigned to each county. Each article was matched to a county using a spatial overlay, and the count reflects the total number of pages associated with a given geographic area. We also applied a natural logarithmic transformation (ln) to normalize the distribution and facilitate proportional interpretation.
- Number of Unique Contributors: We measure the number of distinct user accounts that edited at least one geo-tagged article within each county. Contributors are identified using user IDs or usernames extracted from Wikipedia edit history logs. To address skewness in the distribution and to facilitate proportional interpretation of model coefficients, we include the natural logarithm (ln) of this measure in our analysis.
- Average Total Edits per Page: For each county, we calculated the total number of edits across all geo-tagged articles and divided this by the number of articles to obtain an average. This metric reflects contributor effort, capturing the average volume of editorial activity per article.

While the participation metrics above measure how much editing occurs within each county, they do not capture how that editing labor is distributed among contributors or across pages. To assess disparities in the distribution of work, we measure inequality of participation using the Gini coefficient. The Gini ranges from 0 (perfect equality) to 1 (extreme inequality), providing a principled way to quantify the concentration of editorial labor. We operationalize participation inequality using two complementary Gini-based measures:

- Contribution Concentration per Page (Gini of Edits per Page): We computed the Gini coefficient for the number of edits per contributor on each page, then averaged these across all pages in a county. This captures how evenly contributors share editing responsibility. Values near 0 indicate balanced contributions, values near 1 suggest that a small number of contributors did most of the editing.
- Labor Concentration (Gini of Contributors Across Pages): We computed the Gini coefficient for the number of contributors per page across all articles in a county. This captures how evenly editorial effort is distributed. Values near 0 reflect broad participation across many pages, values near 1 indicate that contributors focus on a small subset of pages.

3.4.2 Participation Composition. For RQ2, which examines the composition of participation and the formation of contribution dynamics, we constructed metrics that capture who contributes across U.S. counties. These metrics include both bot and human contributors and characterize human contributors by their geographic proximity and level of editorial engagement (see Table 1).

To identify bot contributors, we used two complementary approaches. First, we applied username heuristics commonly used in Wikipedia research, such as detecting the presence of terms like "bot" or "script" in account names [71]. Second, we incorporated platform-provided metadata that flag accounts as automated, following established practices in peer production studies [25]. For each county, we also computed both the percentage and the absolute number of bot contributors to examine the role of automation in shaping local patterns of peer production.

For human contributors, rather than relying on noisy point-based IP geolocation, we operationalize localness by examining the spatial footprint of each editor's geo-tagged contributions. Prior work has shown that IP-based geolocation is an unreliable indicator of contributor location. For instance, it is available only for anonymous

Table 1: Descriptive statistics for dependent and independent variables at the county level (N = 3,133).

Variable	Mean	SD	Min	Max
Independent variables				
Broadband Coverage (pp)	75.66	21.93	0.00	100.00
Population Density (people/mi^2)	104.40	697.28	0.01	27819.80
Dependent variables				
Participation volume and concentration (RQ1)				
Number of Wikipedia Pages	198.44	377.92	4.00	8145.00
Number of Unique Contributors	9437.37	24219.50	270.00	642451.00
Average Total Edits per Page	124.66	46.80	21.74	398.00
Contribution Concentration (Gini of Edits per Page)	0.41	0.07	0.17	0.74
Labor Concentration (Gini of Contributors Across Pages)	0.49	0.05	0.28	0.69
Participation composition (RQ2)				
Bot Contributors (#)	107.04	37.61	35.00	467.00
Bot Contributors (%)	6.35	3.11	0.29	20.42
Local Prolific Contributors (#)	46.59	211.30	0.00	6938.00
Local Prolific Contributors (%)	0.87	0.95	0.00	11.91
Local Regular Contributors (#)	181.12	607.56	0.00	18999.00
Local Regular Contributors (%)	3.73	2.28	0.00	24.40
Local Casual Contributors (#)	813.59	2826.98	4.00	88985.00
Local Casual Contributors (%)	16.55	8.62	0.70	55.94
Non-local Prolific Contributors (#)	1662.74	1887.10	149.00	26991.00
Non-local Prolific Contributors (%)	66.08	10.48	16.96	89.56
Non-local Regular Contributors (#)	212.66	574.94	0.00	13677.00
Non-local Regular Contributors (%)	5.06	3.08	0.00	54.86
Non-local Casual Contributors (#)	57.23	179.67	0.00	4639.00
Non-local Casual Contributors (%)	1.32	1.07	0.00	10.16
Aggregated contributor composition (for interpretability)				
Local Contributors (#)	1041.30	3630.88	4.00	114064.00
Non-local Contributors (#)	1932.64	2557.90	167.00	44540.00
Local Contributors (%)	21.16	10.83	0.92	71.71
Non-local Contributors (%)	72.48	8.82	28.00	97.31
Prolific Contributors (#)	1709.33	2050.25	151.00	33071.00
Prolific Contributors (%)	66.97	10.17	20.79	89.66
Regular Contributors (#)	393.79	1152.99	3.00	32676.00
Regular Contributors (%)	8.80	4.27	0.51	55.12
Casual Contributors (#)	870.82	2896.12	4.00	92857.00
Casual Contributors (%)	17.88	9.05	0.74	58.37

Note: All (%) variables use a 0-100 percentage-point scale.

editors, often resolves to ISP hubs rather than user locations, and can be highly inaccurate for mobile or institutional networks [38]. In contrast, geo-tagged editing behavior provides a more stable signal of the places where contributors choose to focus their activity. Wikipedia editors' geographic contributions have been shown to cluster around a primary region of engagement [38, 44] and follow strong distance-decay patterns [34].

Guided by this evidence, we define a contributor as local to a county when the majority of their geo-tagged edits concern articles in that county, following established peer-production practice [56, 64]. Prior work shows that such contributors often live near the location they edit about: a substantial share of edits to geographic Wikipedia articles come from people who are located—and likely reside—near the subject of the article [34, 36]. Others have found that contributors may also be "local" to a place even if they do not currently live there because of personal ties, familiarity, or contextual knowledge of a place, which results in more accurate and diverse contributions [13, 44, 80]. Thus, in our study, we use "local" to describe where contributors concentrate their editing activity, representing contributors that have sufficient expertise to regularly, effectively, contribute about these places. Our data does not allow us to confirm where contributors currently reside, nor where they lived at the time of editing. However, empirically, contributors in our dataset made an average of 86% of their geo-tagged edits within a single county, suggesting that editors typically concentrate their geographic editing in one primary area, echoing findings in prior work as well [64].

Building on these localness classifications, we also characterize contributors by their level of activity. Following established practices in peer-production research, we define prolificness based on

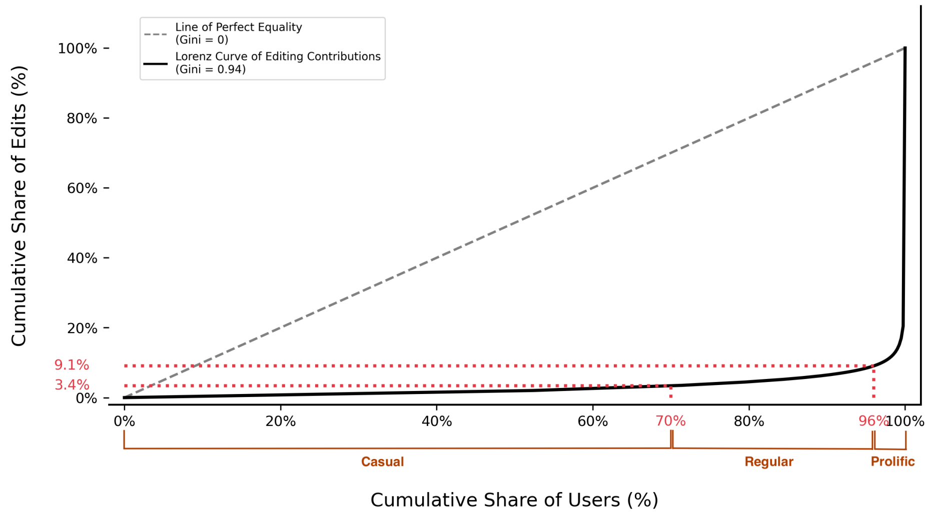


Figure 3: Lorenz curve of contributor editing activity. The figure shows the cumulative share of edits contributed by the cumulative share of users, illustrating the highly unequal distribution of labor in geo-tagged Wikipedia editing (Gini = 0.94). Horizontal dotted lines denote the edit shares contributed by casual, regular, and prolific users.

each contributor's total number of edits across all geo-tagged articles [56, 64]. As shown in Figure 3, the most active 4% of contributors collectively account for 90.9% of all geo-tagged edits, and we classify these individuals as prolific contributors. The next 25% of contributors, who together produce 5.7% of edits, are categorized as regular contributors. The remaining 70% of contributors—responsible for only 3.4% of edits—are classified as casual contributors.

Combining localness (local vs. non-local) with activity level (prolific, regular, casual), we constructed six mutually exclusive contributor subgroups: Local Prolific, Local Regular, Local Casual, Non-local Prolific, Non-local Regular, and Non-local Casual. For each subgroup, we calculated two county-level metrics: (1) the percentage of all contributors in that category and (2) the absolute number of contributors in that category.

3.5 Robustness Checks

We conducted several robustness checks to assess the stability of our Spatial Durbin Model (SDM) estimates. First, because spatial regression results can be sensitive to the specification of the spatial-weights matrix, we re-estimated all SDM models using alternative definitions of spatial proximity. While a "Queen contiguity" matrix is theoretically well suited for county-level analyses, we also estimated models using first-order "Rook contiguity" (which includes only shared borders and excludes corner adjacency) and k-nearest-neighbor (k = 4) matrices (a distance-based alternative). Broadband's direct and indirect effects remained consistent in direction, magnitude, and significance of the statistical trends across these specifications, indicating that our findings are not dependent on a particular spatial-weights choice (see Appendix D for full robustness tables).

Second, we addressed the multiple-testing problem arising from estimating direct and indirect effects across multiple dependent variables. Following best practices in spatial econometrics and quantitative HCI research, we applied the Benjamini–Hochberg False Discovery Rate (FDR) correction [3] to all p-values within each research question. Treating RQ1 and RQ2 as distinct families of related hypotheses avoids the excessive conservatism of Bonferroni-type corrections while appropriately controlling false discoveries. All significance levels reported in our section 4 reflect FDR-adjusted p-values using standard thresholds (*** $p < .001$, ** $p < .01$, * $p < .05$).

3.6 Methodological Limitations

While our findings provide new insights into the relationship between broadband coverage and local information production, some methodological decisions limit the generalizability of our results.

First, our analysis relies on cross-sectional county-level data, which limits our ability to make strong causal claims. The observed relationships between broadband availability and Wikipedia participation should therefore be interpreted as *associations* rather than direct causal effects, partly due to the limited availability and potential inaccuracies of longitudinal broadband data [5, 45, 48]. Future research could incorporate panel datasets or quasi-experimental designs to better identify potential causal mechanisms and examine how changes in broadband access over time influence local information production, as prior work has demonstrated that temporal variations can be significant [64].

Second, it is important to acknowledge the Modifiable Areal Unit Problem (MAUP) in geography, which suggests that the spatial units selected for analysis may influence the observed trends. Our analysis operates at the county level, which provides a balance between data availability and interpretability, but this choice may smooth over important variation within counties or across other geographic scales. Future studies could extend this work by examining broader state-level patterns or finer-grained spatial units, such

Table 2: Effects of broadband coverage and population density on participation volume and concentration.

	Broadband (pp)		Pop. Density ($\log_2$)	
Dependent Variables	**Direct Effect**	**Indirect Effect**	**Direct Effect**	**Indirect Effect**
Number of Pages	2.58***	0.64	-0.20	23.56***
ln (Number of Pages)	0.01***	-0.01	0.00***	0.08***
Number of Unique Contributors	47.32***	16.64	-11.99	403.47***
ln (Number of Unique Contributors)	0.02***	-0.01***	0.01	0.07***
Average Total Edits per Page	0.22***	0.07	-0.25	-5.06***
Contributions Concentration per Page	0.00***	-0.00***	-0.00	-0.01***
Labor Concentration	0.00***	-0.00***	0.00***	0.00

Note. Broadband (pp) indicates coefficients represent the effect of a one–percentage-point increase in county broadband coverage. Asterisks denote significance after Benjamini–Hochberg FDR correction (* $p < .05$, ** $p < .01$, *** $p < .001$).

as census tracts, to provide a more nuanced understanding of how broadband access shapes participation dynamics.

Third, one key modeling decision concerns our choice of control variables. We include population density but exclude broader socioeconomic status (SES) measures such as median income or the urban–rural divide, for both conceptual and methodological reasons. SES variables are strongly correlated with broadband availability ($r = 0.71$ correlation between median household income and broadband coverage), which risks multicollinearity in our modelling. As a result, some of the observed association between broadband and participation in our findings may also reflect underlying economic differences across counties. We included population density as one form of statistical control, knowing that this approach cannot fully capture the role of local policies, community initiatives, and sociocultural factors, such as county subsidies, municipal broadband programs, or digital literacy campaigns, that may produce unexpectedly high participation in lower-density or lower-income areas. Future work should further investigate the relationship between other control variables and broadband availability, and integrate these qualitative dimensions to better understand how local contexts mediate the relationship between broadband availability and online knowledge production.

4 Results

We present the results by organizing them around recurring patterns observed across the participation measures. In particular, we address RQ1 (Section 4.1) and RQ2 (Sections 4.2, 4.3, and 4.4) through the relevant findings discussed below. In Sections 4.3 and 4.4, we highlight the key broadband effects in the main result tables for ease of interpretation, while the complete spatial regression outputs for all models are provided in the Appendix C.

For RQ1, we find that broadband access is associated with more contributors and more content production on Wikipedia, but the overall concentration of contributions across users and pages changes very little. In other words, increased access expands participation without reducing existing inequalities. This motivates RQ2, where we examine who is contributing and how productivity is distributed. We focus on two dimensions: *localness*, which captures whether broadband enables more contributions from editors whose activity is tied to a place, and *prolificness*, which indicates whether participation growth is driven by a small set of highly active users or a broader base of contributors.

To better contextualize these findings, we draw on baseline expectations for broadband expansion in the United States. Counties undergoing federally supported broadband development typically experience an annual increase of approximately 10% in 100/20 Mbps coverage [18]. For example, in Midland County, Michigan, the proportion of underserved residences decreased from 15% to 7% within a single year—an 8% improvement attributed to federally supported broadband initiatives [43]. Therefore, we adopt this 10% annual change as a reference point to aid interpretation of effect sizes reported in subsequent sections.

4.1 Broadband Facilitates Participation but Not Contribution Equity

In RQ1, we found that broadband coverage strongly predicts the scale of content production and participation. When comparing two otherwise similar counties, a county with 10% higher broadband coverage is associated with the approximately 26 additional geo-tagged Wikipedia pages, corresponding to a 10.5% increase in page production based on the log-transformed model. Similarly, compared to a comparable county, one with 10% higher broadband coverage is associated with approximately 473 more Wikipedia contributors, a 22% increase relative to the baseline level. Beyond volume, broadband also shapes contributor engagement. We see that between two otherwise similar counties, the one with 10% higher broadband coverage is associated with an increase of approximately 2 more edits per page, indicating that infrastructure not only influences the volume of contributions, but may also shape the intensity of contributor engagement.

Turning to the indirect (spatial spillover) effects of broadband coverage on participation, we find no significant spillover effects on the total number of Wikipedia pages—meaning that broadband expansion in neighboring counties does not significantly affect a county's overall page production. However, we observe a significant spatial spillover effect on the log-transformed number of contributors: if neighboring counties experience a 10% increase in broadband coverage, the number of unique contributors within a given county is predicted to decrease by approximately 9.5% relative to its original contributor count. This effect appears on the log

Table 3: Effects of broadband coverage and population density on bot contributor participation.

Category	Metric	Broadband (pp)		Pop. Density ($\log_2$)	
		Direct Effect	Indirect Effect	Direct Effect	Indirect Effect
Bot	Bot Contributors (%)	-0.045***	0.017*	-0.023	-0.433***
	Bot Contributors (#)	0.488***	-0.238**	-0.250	2.252**

Note. Broadband (pp) indicates coefficients represent the effect of a one–percentage-point increase in county broadband coverage. Asterisks denote significance after Benjamini–Hochberg FDR correction (* $p < .05$, ** $p < .01$, *** $p < .001$).

scale but not in raw counts, suggesting that broadband expansion in nearby regions primarily influences the proportional distribution of contributors rather than driving consistent, large absolute changes. This pattern points to potential contributor displacement, where editors may be drawn toward better-connected regions, thereby reshaping geographic participation patterns.

To examine broader contribution dynamics, we assess two forms of concentration using the Gini coefficient: contribution concentration per page, which captures how edits are distributed across contributors, and labor concentration, which captures how contributor activity is distributed across pages. Although the direct and indirect effects of broadband coverage are statistically significant, their magnitudes are extremely small—near zero in both cases. In short, increased broadband access relates to higher participation and greater content production, but it does not meaningfully alter how contributions are distributed across contributors or how contributor labor is distributed across pages. This suggests that while broadband expands participation, it does not substantially change the underlying inequality in contribution patterns.

4.2 Broadband Coverage Is Associated With More Human Than Automated Participation

We next examine how broadband relates to the composition of contributors, beginning with bot activity. As shown in Table 3, when comparing two otherwise similar counties, a county with 10% higher broadband coverage is associated with approximately 4.8 additional active bot accounts. Despite this increase in absolute numbers, the proportion of bot contributors declines by roughly 0.45%. Given that counties in our dataset average 107.04 bot contributors, representing 6.35% of all contributors, this change indicates that bots account for a smaller share of overall activity in better-connected counties. In other words, although more bots operate in counties with higher broadband coverage, their relative prominence declines because human participation grows at a substantially faster rate overall and locally. This pattern suggests a rebalancing of participation in which human-led contributions become more central as broadband infrastructure improves.

We also find evidence of significant spatial spillover effects of broadband coverage on bot participation. When comparing two otherwise similar counties, if the neighboring counties of one have 10% higher broadband coverage, the proportion of bot contributors in that focal county is predicted to increase by about 0.17%, even though the absolute number of active bots decreases slightly (2.4 fewer bots on average).

Taken together, these patterns show that in the context of local knowledge production, such as geo-tagged Wikipedia pages, counties with higher broadband access tend to exhibit contributor mixes in which human participants account for a slightly larger share relative to bots. Moreover, when neighboring counties within the same region have higher broadband coverage, the focal county is predicted to show a slightly higher proportion of bot contributors, even as the absolute number of bots is somewhat lower. In practice, this means that counties with greater broadband access display contributor populations dominated even more strongly by human editors, while adjacent counties without similar broadband expansion show contributor mixes in which bots make up a modestly larger share of activity, despite humans remaining the dominant group across all settings. These dynamics suggest that automation and human labor do not operate as fixed layers of the contributor ecosystem but instead appear in changing proportion across regional infrastructural contexts. We elaborate on these editorial dynamics, and their implications for how different forms of contribution surface in local knowledge production in the Section 5.2.

4.3 Broadband Access Aligns with Growth in Grassroots and Local Participation

As shown in Table 4, when examining overall local and non-local participation, our results suggest that, holding other factors constant, a 10% increase in broadband coverage is associated with approximately 209 more local contributors and 261 more non-local contributors. However, despite these absolute increases, the proportional composition of the contributor base shifts in a seemingly paradoxical way: the share of local contributors rises by about 1.63%, while the share of non-local contributors declines by roughly 0.99%. This discrepancy highlights a key dynamic in platform participation—while broadband expansion stimulates growth in contributor headcounts across both groups, the baseline size of the non-local contributor population (averaging 1932.63 contributors) is nearly double that of the local contributor base (averaging 1041 contributors). As a result, even smaller absolute increases in local contributors yield proportionally larger effects, whereas relatively larger increases among non-local contributors yield modest or even negative shifts in their share.

This finding reveals a broader implication: while infrastructure improvements like broadband expansion can increase participation among both local and non-local contributors, they also shift the relative balance between these groups. Specifically, broadband access appears to boost local engagement, helping to address geographic information gaps, but does not fundamentally resolve deeper structural disparities in who contributes. Rather, these gains may signal

Table 4: Estimated broadband effects on human contributor participation, disaggregated by contributor type and geography. The final column reports the sum of Local and Non-local Direct effects†, excluding Indirect effects. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. † Sum (Direct) = Local Direct + Non-local Direct.

Contributor Type	Local		Non-local		Sum (Direct)
	Direct	Indirect	Direct	Indirect	
Number of Contributors (#)					
Prolific (#)	0.961***	0.444	21.454***	-0.680	**22.415**
Regular (#)	3.692***	2.305	3.691***	1.749	**7.383**
Casual (#)	16.264***	11.904*	0.996***	0.640	**17.260**
Total Count (Direct)	**20.917**	–	**26.141**	–	–
Share of Contributors (%)					
Prolific (%)	0.006***	-0.004**	-0.122***	-0.092***	**-0.116**
Regular (%)	0.029***	-0.006	0.018***	-0.013*	**0.047**
Casual (%)	0.128***	-0.018	0.005***	-0.003*	**0.133**
Total Share (Direct)	**0.163**	–	**-0.099**	–	–

Note. The coefficients represent the effect of a one–percentage-point increase in county broadband coverage. Asterisks denote significance after Benjamini–Hochberg FDR correction (* $p < .05$, ** $p < .01$, *** $p < .001$).

a redistribution within an already unequal system: local contributors, though growing in number, still represent a smaller and more fragile portion of the overall contributor base. As a result, improvements in access alone may be insufficient to close longstanding representational gaps without additional, targeted strategies that promote sustained and equitable local participation.

Turning to spillover effects, we observe consistent negative associations between neighboring broadband expansion and the share of non-local contributors in a given county. Specifically, a 10% increase in broadband coverage in adjacent counties is associated with a statistically significant decline in the proportion of non-local contributors to the focal county: 0.92% for prolific contributors, 0.13% for regular, and 0.03% for casual contributors. However, we do not observe significant changes in the absolute number of non-local contributors across these categories.

This pattern suggests that broadband improvements in neighboring areas do not directly displace non-local contributors from a county in raw terms. Instead, they appear to alter the relative composition of participants. One plausible explanation is that regional broadband improvements encourage greater local participation within the focal county, thereby reducing the proportional presence of non-local contributors. These findings point to a compositional shift rather than an overall decline in non-local engagement, and they align with broader theories of geographically bounded participation and infrastructural crowding effects [10, 61].

4.4 Broadband Expansion Is Associated with Shifting Participation Dynamics Across Contributor Types

As shown in the rightmost column of Table 4, broadband coverage is associated with distinct patterns of participation growth across contributor types. When disaggregating contributors by their level of prolificness, we find that higher broadband availability correlates with differential changes in both the number and relative composition of contributors. Specifically, comparing two otherwise similar counties, a 10% increase in broadband coverage corresponds to an estimated 224 additional prolific contributors in absolute terms, even though their relative proportion within the overall contributor base declines by about 1.16%. In contrast, the same broadband increase is associated with 73 additional regular contributors (a 0.47% increase in share) and 173 additional casual contributors (a 1.33% increase in share). By contrast, the same 10% increase in broadband coverage is associated with around 73 additional regular contributors, representing an increase of roughly 0.47% , and approximately 173 additional casual contributors, an increase of about 1.33%.

This pattern suggests that broadband expansion is linked to broader participation by attracting more casual and regular contributors—those typically constituting the long tail of engagement—while the relative influence of prolific editors declines. These dynamics may help contextualize the rising participation inequality observed in RQ1, as reflected in increased Gini coefficients. In other words, while broadband may expand the contributor base, it also appears to amplify disparities in engagement levels, highlighting structural imbalances in who contributes and how often.

5 Discussion

Taken holistically, our findings align with and advance the Wikimedia Foundation's broader agenda of understanding Interaction Gaps, which refer to the structural and experiential barriers that shape individuals' ability to access and engage with Wikimedia sites [59]. By examining broadband availability through the lens of geo-tagged Wikipedia production, our results highlight both the benefits and the limitations of broadband expansion for supporting equitable knowledge contribution. These patterns reveal practical trade-offs for broadband initiatives and policy efforts to consider, and they also point toward important directions for future research and design. We elaborate on these implications below.

5.1 Mitigating Information Gaps, But Maintaining Contribution Dynamics

Controlling for population density to account for rural–urban disparities, our results show that improved broadband access is associated with substantial growth in geo-tagged Wikipedia content and contributor participation. Specifically, a county with 10% higher broadband coverage is expected to have approximately 470 additional contributors and 25.8 more geo-tagged Wikipedia pages. This relationship illustrates how infrastructural access not only expands connectivity [28], but also enables new forms of collaborative engagement, lowering sociotechnical barriers to entry and broadening the landscape of participation.

Yet, this growth does not translate into a more equitable distribution of contribution effort. As shown in Section 4.4, more individuals participate when broadband expands, but this growth is concentrated among non-prolific contributors who make only a small fraction of total edits. As a result, the core labor of knowledge production remains carried out by a small subset of highly active users. This pattern mirrors observations in other digital participation initiatives, including humanitarian mapping efforts [76, 78], where efforts to broaden engagement successfully expand the contributor base but can unintentionally intensify the concentration of labor among a small core group. Together, these findings suggest that expanding infrastructural access brings more people into the ecosystem, but does not meaningfully redistribute who performs the majority of the work, or whose perspectives ultimately shape the resulting knowledge.

This persistent concentration of labor is not merely a descriptive pattern in data production, but has important implications for how peer production communities evolve over time. Contribution inequality affects both the quality of content and the long-term development of the community. When a small group of highly prolific contributors performs most of the editorial labor, their practices and perspectives disproportionately shape emerging norms and governance structures—potentially narrowing the interpretive frames through which knowledge is created [24, 32]. Over time, these concentrated patterns of activity can create subtle barriers for newcomers, as highly entrenched norms, routines, and expectations reflect the habits of a narrow contributor group rather than the broader population of potential participants [63].

In geographic knowledge production, like geo-tagged Wikipedia editing, these dynamics carry additional risks. Prior work shows that uneven contributor distributions can reinforce spatial biases and reproduce gaps in how places are represented, particularly in rural or underserved regions [28–30]. When prolific contributors dominate, their perspectives—not necessarily the perspectives of local residents—are more likely to shape how communities, landmarks, and histories are documented. This risks amplifying representational inequalities and limiting opportunities for less-experienced or locally embedded contributors to meaningfully influence what becomes part of the public record.

Therefore, addressing these kinds of imbalances requires complementary strategies that go beyond improving access. These may include inclusive platform design features, targeted community support, and localized editorial campaigns to encourage sustained and distributed contributions. For instance, in Humanitarian OpenStreetMap, Yin et al. [76] found that when third party organizations help facilitate participation, it can promote more equitable and long-term knowledge production in regions that suffer from data scarcity. Likewise, initiatives like Growing Local Language Content [50], organize localized edit-a-thons [46, 73], or scaffold newcomer participation through structured pathways [57] have been shown to broaden the contributor base and strengthen place-based knowledge practices. These efforts highlight that infrastructural access is necessary but insufficient: equitable participation requires sociotechnical ecosystems that actively support contributors whose knowledge has historically been underrepresented.

Designing for equitable participation, then, requires not only enabling access but also reshaping the social and technical conditions that determine how—and by whom—contributions are made. Our findings point toward the need for novel, locally oriented modalities of information production that better support situated contributors and amplify community-grounded perspectives (e.g., [57]). Pursuing this direction will require substantive, community-driven needs assessments to understand the design space for such tools and the goals of the communities who might use them. We see this as an essential avenue for future work, particularly in efforts to reduce geographic disparities in public knowledge infrastructures.

5.2 Infrastructure Access and the Centrality of Human-Led Participation Over Automation

Prior work has found that automated editing tends to become more prominent where human participation is limited [20, 24, 28, 33, 44]. Bots often fill structural gaps by performing large-scale maintenance tasks—reverting vandalism, enforcing formatting rules, or ensuring consistency across articles—and do so at speeds and scales that far exceed human capacity. However, these strengths coexist with limitations: bots cannot easily incorporate contextual nuance, local cultural knowledge, or situated histories that human editors contribute [20]. These differences manifest geographically: Johnson et al. [44] show that bots account for more than 23% of contributions to rural articles, compared to only 4.5% in urban ones, illustrating how automation fills gaps where human activity is sparse.

Our findings add nuance to this line of work by showing that automation's visibility also varies with regional differences in broadband access. In counties with higher broadband coverage, human contributors constitute a larger share of activity, and bots appear less central in relative terms, even as their absolute numbers increase. These patterns point to a shifting balance in which human and automated work coexist within infrastructural conditions that shape not only whether people can contribute, but how labor is distributed across different forms of participation. Bots remain part of the editorial landscape, yet their prominence depends on the volume and distribution of surrounding human activity.

Taken together, these observations highlight how infrastructural conditions shape the mix of human and automated labor across localities. Prior work suggested that regions with stronger human participation may generate more contextually informed or locally grounded content, whereas areas with slower participation growth may continue to rely more heavily on automated processes. These findings raise opportunities for future work to examine how automation adapts to increasingly localized contributions, how the roles, functions, and geographic reach of bots shift alongside changing participation patterns, and how these evolving dynamics shape equitable and contextually rich knowledge production at scale.

5.3 Grassroots and Local Participation Under Supportive Infrastructural Conditions

Our findings show that broadband access is associated with meaningful shifts in *who* participates in producing geo-tagged Wikipedia content. Counties with higher broadband coverage see increases across all contributor types, but the most substantial growth emerges among local contributors. At the same time, the relative share of non-local contributors declines, even though their absolute numbers remain stable. This pattern suggests that supportive infrastructural conditions strengthen community-grounded participation by enabling people who are more closely connected to a place—whether through residence, familiarity, or interest—to play a larger role in characterizing their local environments online.

Our results here extend prior research showing that technological access is foundational for inclusive participation in distributed knowledge systems [28, 75]. Our analysis moves beyond documenting access disparities to identify which contributors are most likely to engage when infrastructure improves. The largest increases occur among less-prolific local contributors—individuals who edit intermittently but whose situated, place-based knowledge cannot easily be substituted by non-local contributors or automated systems. Their increased presence broadens the range of grounded perspectives reflected in the Wikipedia articles about that area.

At the same time, broadband access seems to decentralize who participates even as contribution volume remains unequal. The influx of casual local contributors extends the long tail of participation, but their relatively small edit volumes mean that highly prolific editors continue to dominate total contributions. This highlights an important duality: infrastructural improvements diversify the perspectives represented, but they do not automatically reduce longstanding inequalities in contribution volume. Moreover, the expansion of participation that broadband enables does not necessarily guarantee sustained or equitable involvement. Research on digital inclusion and community engagement shows that durable participation—particularly among newer or less prolific editors—depends on sociotechnical scaffolding such as mentorship networks, local editing campaigns, discoverability mechanisms, and tools that support novice contributors [58]. Similar supports may be necessary to ensure that contributors activated through improved broadband access continue to participate and that their on-the-ground perspectives remain visible amid established editorial workflows.

These findings also have implications for AI systems trained on geo-tagged Wikipedia. For instance, GPT-3 used the entirety of the Wikipedia corpus nearly four-times over [6], and recent evidence suggests that Wikipedia comprises 14% of training data in LLMs [47]. Because AI models can internalize the perspective biases of their underlying datasets [36], shifts in who contributes to Wikipedia may influence the kinds of geographic knowledge that downstream systems—like generative AI tools—learn. Increased participation from locally grounded editors expands the presence of situated, context-specific perspectives in the training data for these tools. As a result, AI systems trained on these sources may encode representations that reflect a different mix of place-based experiences—potentially altering how local environments are interpreted computationally. While future work should interrogate this empirically, if indeed this bears out, local representation in online communities like Wikipedia may be a fundamentally necessary component of being accurately represented and characterized in next-generation information systems like Large Language Models.

Taken together, these findings suggest that broadband infrastructure does more than increase participation: it reshapes the contributor ecology and the epistemic character of the data produced. While improved infrastructure may attract more locally grounded contributors, whether their contributions persist, gain influence, or are reused downstream depends on sociotechnical conditions that support novice editors and maintain the visibility of their work within established editorial workflows. This dynamic points toward future research on platform design, community structures, and governance approaches that support equitable, geographically grounded public knowledge production, and improve the representational quality of datasets that increasingly inform both human understanding and AI systems.

5.4 Towards Understanding the Spatial Redistribution of Contributor Activity

Our results on indirect, cross-county effects point to an important dimension of digital participation: contributions are shaped not only by a county's own infrastructural conditions but also by those of its surrounding regions. We observe that when neighboring counties have higher broadband coverage, the focal county shows a decline in the proportion of both local prolific and non-local contributors, even though their absolute numbers remain essentially unchanged. Rather than indicating a loss of contributors, this pattern suggests a form of spatial rebalancing in which participation becomes redistributed across nearby areas as infrastructural conditions vary.

This observation aligns with work in spatial diffusion research, which has long shown that technological improvements rarely unfold as isolated or uniform changes. In many domains, new infrastructural openings lead to regional rearrangements in activity rather than parallel growth across all locations [79]. In our context, contributors, particularly those who are highly prolific, may adjust their geographic focus as new areas become more viable to edit, gradually diffusing their attention across a wider set of places. Related scholarship on innovation diffusion shows that initial gains often emerge in more connected centers before spreading to peripheral regions [62]. Our findings reflect a similar pattern. As broadband improves in some places, contributor activity becomes spread across a broader spatial field, which reduces the relative concentration of participation within any single county.

These dynamics highlight the importance of moving from a narrow, county-level perspective to a regional network perspective. Contributor pools operate within geographically interdependent ecosystems. Changes in activity in one area may coincide with shifts in another, and participation patterns may reflect broader regional connectivity rather than isolated local conditions. This interdependence raises several directions for future work. One important question concerns the circumstances under which broadband access aligns with locally anchored participation as opposed to a redistribution of contributor attention across counties. Another concerns how uneven infrastructural improvements shape knowledge equity, visibility, and representation across adjacent regions. A further opportunity is to examine what kinds of platform mechanisms, collaborative programs, or regional partnerships can support locally meaningful contributions even as participation becomes more spatially diffuse.

Together, these directions show that broadband access is not only associated with participation within specific boundaries. It also appears to shape a spatially interconnected knowledge ecosystem. Understanding these dynamics is important for developing policy and sociotechnical systems that support equitable, sustainable, and geographically balanced forms of public knowledge production.

6 Ethical Considerations

We believe this observational, aggregated by county, statistical work presents no major ethical concerns. All datasets we used are publicly available and contain no personally identifiable information, including: geo-tagged Wikipedia revision histories from a dump file, county-level broadband coverage from the FCC, and U.S. Census demographics. While we do not anticipate any negative societal impact or misuse, we caution against interpreting our metrics as value judgments about specific regions. Our findings are intended to inform equity-aware digital policy and platform design.

7 Conclusion

Our work examined how broadband access shapes patterns of participation in geo-tagged Wikipedia contributions across U.S. counties. While broadband expansion is strongly associated with increased participation, it does not substantially reduce the underlying concentration of contributions. By analyzing contributors through the lenses of localness and prolificness, we show that broadband access is most strongly associated with growth in place-based local participation and in contributions from casual editors. However, these contributors add relatively few edits compared to highly prolific users, illustrating the continued dominance of a small core group in peer production. Our analysis also reveals spatial spillover effects, suggesting that improvements in internet access in one region may redistribute, rather than expand, participation in neighboring areas. Taken together, these findings contextualize the benefits and limitations of broadband infrastructure in enabling inclusive digital knowledge production. While expanding access is a necessary step, it is not sufficient to address deeper structural patterns in who contributes, how much they contribute, and the influence they hold.

Acknowledgments

Support for this research was provided by the University of Wisconsin–Madison Division of Extension and Office of the Vice Chancellor for Research.

References

[1] Ofer Arazy, Hila Lifshitz-Assaf, and Adam Balila. 2018. Neither a Bazaar nor a cathedral: The interplay between structure and agency in Wikipedia's role system. *Journal of the Association for Information Science and Technology* 70 (2018). https://doi.org/10.1002/asi.24076

[2] Jack Bandy. 2022. Problematic machine behavior: A systematic literature review of algorithm audits. *Comput. Surveys* 55, 3 (2022), 1–38. doi:10.1145/3485128

[3] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300.

[4] Pablo Beytía. 2020. The Positioning Matters: Estimating Geographical Bias in the Multilingual Record of Biographies on Wikipedia. In *Companion Proceedings of the Web Conference 2020 (WWW '20)*. ACM, New York, NY, USA, 806–810. doi:10.1145/3366424.3383534

[5] Emily Birnbaum. 2019. *Critics dismiss new FCC report showing increased broadband access.* https://thehill.com/policy/technology/445985-fcc-releases-broadband-deployment-report-amid-dissent The Hill.

[6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* 33 (2020), 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html?utm_source=transaction&utm_medium=email&utm_campaign=linkedin_newsletter

[7] Ewa S. Callahan and Susan C. Herring. 2011. Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology* 62, 10 (2011), 1899–1915. doi:10.1002/asi.21577

[8] Ashley Colley, Jacob Thebault-Spieker, Allen Yilun Lin, Donald Degraen, Benjamin Fischman, Jonna Häkkilä, Kate Kuehl, Valentina Nisi, Nuno Jardim Nunes, Nina Wenig, Dirk Wenig, Brent J. Hecht, and Johannes Schöning. 2017. The Geography of Pokémon GO: Beneficial and Problematic Effects on Places and Movement. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017).

[9] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*. ACM, New York, NY, USA, 32–41. doi:10.1145/1216295.1216309

[10] Justin Cranshaw, Raz Schwartz, Jason I. Hong, and Norman Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 6. AAAI, 58–65. doi:10.1609/icwsm.v6i1.14278

[11] Lisa J. Dettling, Sarena Goodman, and Jonathan Smith. 2018. Every Little Bit Counts: The Impact of High-Speed Internet on the Transition to College. *Review of Economics and Statistics* 100, 2 (May 2018), 260–273. doi:10.1162/REST_a_00673

[12] Michaelanne Dye, Annie Antón, and Amy S. Bruckman. 2016. Early Adopters of the Internet and Social Media in Cuba. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*. ACM, 1295–1309. doi:10.1145/2818048.2819954

[13] Melanie Eckle and João Porto de Albuquerque. 2015. Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. In *Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2015)*. Kristiansand, Norway.

[14] European Commission. 2021. Europe's Digital Decade: 2030 Digital Targets. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en Accessed May 2025.

[15] Frank A. Farris. 2010. The Gini Index and Measures of Inequality. *The American Mathematical Monthly* 117, 10 (2010), pp. 851–864. https://www.jstor.org/stable/10.4169/000298910x523344

[16] Federal Communications Commission. 2021. Broadband Deployment Report. https://www.fcc.gov/sites/default/files/bdac-low-income-communities-approved-rec-12172020.pdf Accessed May 2025.

[17] Federal Communications Commission. 2023. Fixed Broadband Deployment Data from FCC Form 477. https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477. Accessed: 2025-02-03.

[18] Federal Communications Commission. 2024. 2024 Communications Marketplace Report. https://www.fcc.gov/reports-research/reports/consolidated-communications-marketplace-reports/CMR-2024 Accessed: 2025-05-15.

[19] Heather Ford, Shilad Sen, David R. Musicant, and Nathaniel Miller. 2013. Getting to the source: Where does Wikipedia get its information from?. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW)*. ACM, 734–741. doi:10.1145/2441776.2441861

[20] Heather Ford and Judy Wajcman. 2017. 'Anyone Can Edit', Not Everyone Does: Wikipedia's Infrastructure and the Gender Gap. *Social Studies of Science* 47, 4 (2017), 511–527. doi:10.1177/0306312717692172

[21] Andrea Forte and Amy Bruckman. 2005. Why do people write for Wikipedia? Incentives to contribute to open-content publishing. In *Proceedings of GROUP 2005 Workshop: Sustaining Community.*

[22] Vikas Gawai and Steven Deller. 2025. Is There a Link Between Access to Broadband and Health Outcomes? *American Journal of Economics and Sociology* 84, 2 (March 2025), 361–386. doi:10.1111/ajes.12608 First published December 7, 2024.

[23] R. Stuart Geiger. 2009. The Social Roles of Bots and Assisted Editing Programs. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09)*. ACM, New York, NY, USA, 30:1–30:2.

[24] R. Stuart Geiger. 2014. Bots, Bespoke Code, and the Materiality of Software Platforms. *Information, Communication & Society* 17, 3 (2014), 342–356. doi:10.1080/1369118X.2013.873069

[25] R. Stuart Geiger and Aaron Halfaker. 2013. When the levee breaks: without bots, what happens to Wikipedia's quality control processes?. In *Proceedings of the 9th International Symposium on Open Collaboration* (Hong Kong, China) *(WikiSym '13)*. Association for Computing Machinery, New York, NY, USA, Article 6, 6 pages. doi:10.1145/2491055.2491061

[26] R. Stuart Geiger and David Ribes. 2010. The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 117–126. doi:10.1145/1718918.1718941

[27] Chris Gentry. 2024. Spatial Regression Analysis — Contiguity Neighbors. https://chrismgentry.github.io/Spatial-Regression/#111_Contiguity. Accessed: 2025-07-25.

[28] Mark Graham, Bernie Hogan, R. K. Straumann, and Ahmed Medhat. 2014. Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers* 104, 4 (2014), 746–764. doi:10.1080/00045608.2014.910087

[29] Mark Graham, Ralph Straumann, and Bernie Hogan. 2015. Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers* 105, 6 (Sept. 2015), 1158–1178. doi:10.1080/00045608.2015.1072791

[30] Muki Haklay. 2016. Why is Participation Inequality Important? In *European Handbook of Crowdsourced Geographic Information*, Cristina Capineri, Muki Haklay, Hai-Ying Huang, Vyron Antoniou, Juhani Kettunen, Frank Ostermann, and Ross Purves (Eds.). Ubiquity Press, 35–44. doi:10.5334/bax.c

[31] Mordechai (Muki) Haklay, Sofia Basiouka, Vyron Antoniou, and Aamer Ather. 2010. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal* 47 (2010), 315 – 322.

[32] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist* 57, 5 (2013), 664–688. doi:10.1177/0002764212469365

[33] Aaron Halfaker and John Riedl. 2012. Bots and Cyborgs: Wikipedia's Immune System. *Computer* 45, 3 (March 2012), 79–82. doi:10.1109/MC.2012.82

[34] Darren Hardy, James Frew, and Michael F. Goodchild. 2012. Volunteered geographic information production as a spatial process. *International Journal of Geographical Information Science* 26, 7 (2012), 1191–1212. doi:10.1080/13658816.2011.629618

[35] Jean Hardy. 2019. How the Design of Social Technology Fails Rural America. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion (DIS '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 189–193. doi:10.1145/3301019.3323906

[36] Brent Hecht and Darren Gergle. 2009. Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the Fourth International Conference on Communities and Technologies* (University Park, PA, USA) *(C&T '09)*. Association for Computing Machinery, New York, NY, USA, 11–20. doi:10.1145/1556460.1556463

[37] Brent Hecht and Darren Gergle. 2010. The Tower of Babel Meets Web 2.0: User-Generated Content and Its Applications in a Multilingual Context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Atlanta, GA, USA, 291–300. doi:10.1145/1753326.1753370

[38] Brent J. Hecht and Darren Gergle. 2010. On the "localness" of user-generated content. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW)*. 229–232. doi:10.1145/1718918.1718962

[39] Brent J. Hecht, Johannes Schöning, Muki Haklay, Licia Capra, Afra Jahanbakhsh Mashhadi, Loren G. Terveen, and Mei po Kwan. 2013. Geographic human-computer interaction. *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (2013).

[40] Molly G. Hickman, Viral Pasad, Harsh Kamalesh Sanghavi, Jacob Thebault-Spieker, and Sang Won Lee. 2021. Understanding Wikipedia Practices Through Hindi, Urdu, and English Takes on an Evolving Regional Conflict. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 34:1–34:31. doi:10.1145/3449108

[41] Benjamin Mako Hill and Aaron Shaw. 2014. The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. In *Proceedings of the 2014 ACM Conference on Computer Supported Cooperative Work*. ACM, 507–512. doi:10.1145/2531602.2531656

[42] International Telecommunication Union. 2022. Facts and Figures 2022: Internet Use by Country Income Level. https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/ Accessed May 2025.

[43] Isabelle Pasciolla. 2024. *Midland County Internet Connectivity Committee presents to county board on projects, new GIS map.* https://www.ourmidland.com/news/article/midland-county-broadband-work-continues-new-gis-20163107.php Accessed: 2025-05-13.

[44] Isaac L. Johnson, Yilun Lin, Toby Jia-Jun Li, Andrew Hall, Aaron Halfaker, Johannes Schöning, and Brent Hecht. 2016. Not at Home on the Range: Peer Production and the Urban/Rural Divide. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 13–25. doi:10.1145/2858036.2858123

[45] Ryan Johnston. 2019. *FCC's annual broadband report criticized for 'inconsistent' methodology.* https://statescoop.com/fccs-annual-broadband-report-criticized-for-inconsistent-methodology/ StateScoop.

[46] Shyong (Tony) K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl. 2011. WP: Clubhouse? An exploration of Wikipedia's gender imbalance. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (2011), 1–10. doi:10.1145/2038558.2038560

[47] Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, Xinyi Wu, Enrico Shippole, Kurt Bollacker, Tongshuang Wu, Luis Villa, Sandy Pentland, and Sara Hooker. 2023. The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI. http://arxiv.org/abs/2310.16787 arXiv:2310.16787 [cs].

[48] David Major, Ross Teixeira, and Jonathan Mayer. 2020. No WAN's Land: Mapping U.S. Broadband Coverage with Millions of Address Queries to ISPs. In *Proceedings of the ACM Internet Measurement Conference (IMC '20)* (Virtual Event, USA). Association for Computing Machinery, New York, NY, USA, 393–419. doi:10.1145/3419394.3423652

[49] Connor McMahon, Isaac Johnson, and Brent Hecht. 2017. The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies. In *Proceedings of the 11th International Conference on Web and Social Media (ICWSM 2017)*. AAAI Press, 142–151. doi:10.1609/icwsm.v11i1.14883

[50] Meta-Wiki. 2025. Supporting Indian Language Wikipedias Program — Meta-Wiki, discussion about Wikimedia projects. https://meta.wikimedia.org/w/index.php?title=Supporting_Indian_Language_Wikipedias_Program&oldid=28164257 [Online; accessed 1-December-2025].

[51] Jesús Mur and Ana Angulo. 2006. The Spatial Durbin Model and the Common Factor Tests. *Spatial Economic Analysis* 1, 2 (2006), 207–226. doi:10.1080/17421770601009841

[52] National Telecommunications and Information Administration. 2023. Broadband Equity, Access, and Deployment (BEAD) Program. https://www.internetforall.gov. Accessed: February 3, 2025.

[53] Pascal Neis, Dennis Zielstra, and Alexander Zipf. 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet* 5 (2013), 282–300.

[54] Oded Nov. 2007. What motivates Wikipedians? *Commun. ACM* 50, 11 (2007), 60–64. doi:10.1145/1297797.1297798

[55] OpenAI. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* (2023). arXiv:2303.08774 [cs.CL] https://arxiv.org/abs/2303.08774

[56] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: A study of power editors on Wikipedia. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. ACM, 51–60. doi:10.1145/1531674.1531682

[57] Katherine Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. 2010. Lurking? cyclopaths?. In *the 28th international conference*. ACM Press, New York, New York, USA, 1917. doi:10.1145/1753326.1753615

[58] Massimo Ragnedda and Glenn W. Muschert (Eds.). 2015. *The Digital Divide: The Internet and Social Inequality in International Perspective*. Routledge, New York.

[59] Miriam Redi, Martin Gerlach, Isaac Johnson, Jonathan Morgan, and Leila Zia. 2020. A Taxonomy of Knowledge Gaps for Wikimedia Projects. *arXiv preprint arXiv:2008.12314* (2020). https://arxiv.org/abs/2008.12314

[60] Anasuya Sengupta, Siko Bouterse, and Kira Allmann. 2018. Build an Internet for, and from, us all. *Nature* 563, 7730 (2018), S147–S148. doi:10.1038/d41586-018-07506-7 Part of Nature Outlook: Digital revolution.

[61] Susan Leigh Star and Karen Ruhleder. 1994. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (Chapel Hill, North Carolina, USA) *(CSCW '94)*. Association for Computing Machinery, New York, NY, USA, 253–264. doi:10.1145/192844.193021

[62] M. Straccamore, V. Loreto, and P. Gravino. 2023. The geography of technological innovation dynamics. *Scientific Reports* 13 (2023), 21043. doi:10.1038/s41598-023-48342-8

[63] Jacob Thebault-Spieker, Aaron Halfaker, Loren G. Terveen, and Brent Hecht. 2018. Distance and Attraction: Gravity Models for Geographic Content Production. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. doi:10.1145/3173574.3173722

[64] Jacob Thebault-Spieker, Brent J. Hecht, and Loren G. Terveen. 2018. Geographic Biases are 'Born, not Made': Exploring Contributors' Spatiotemporal Behavior in OpenStreetMap. *Proceedings of the 2018 ACM International Conference on Supporting Group Work* (2018).

[65] Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. Toward a Geographic Understanding of the Sharing Economy: Systemic Biases in UberX and TaskRabbit. *ACM Trans. Comput.-Hum. Interact.* 24, 3, Article 21 (April 2017), 40 pages. doi:10.1145/3058499

[66] Michaelanne Thomas. 2025. Relational Infrastructures of Survival: How Social Dependencies Enable Alternative Internets in Cuba. *Cuban Studies* 54 (2025), 126–146. doi:10.1353/cub.2025.a967310

[67] Denny Vrandečić. 2012. Wikidata: A New Platform for Collaborative Data Collection. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12 Companion)*. ACM, New York, NY, USA, 1063–1064. doi:10.1145/2187980.2188242

[68] Morten Warncke-Wang, Vivek Ranjan, Loren Terveen, and Brent Hecht. 2015. Misalignment between supply and demand of quality content in peer production communities. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. 493–502.

[69] Brian Whitacre. 2024. The fibre broadband housing premium across three US States. *Regional Studies, Regional Science* 11, 1 (February 2024), 38–62. doi:10.1080/21681376.2024.2305951 Received 4 August 2023; Accepted 4 January 2024; Published online 8 February 2024.

[70] Brian Whitacre, Roberto Gallardo, and Sharon Strover. 2014. Broadband's contribution to economic growth in rural areas: Moving towards a causal relationship. *Telecommunications Policy* 38, 11 (2014), 1011–1023. doi:10.1016/j.telpol.2014.05.005

[71] Wikimedia Contributors. 2023. Research:Identifying bot accounts. https://meta.wikimedia.org/wiki/Research:Identifying_bot_accounts. Accessed: May 14, 2025.

[72] Wikipedia contributors. 2015. Wikipedia:Introduction. https://en.wikipedia.org/wiki/Wikipedia:Introduction. Accessed: 2024-12-15.

[73] Wikipedia contributors. 2025. Edit-a-thon — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Edit-a-thon&oldid=1285562637. [Online; accessed 14-May-2025].

[74] World Bank. 2023. Digital Transformation Overview: Development news, research, data. https://www.worldbank.org/en/topic/digital/overview Accessed May 2025.

[75] Susan Wyche and Jennifer Olson. 2018. Kenyan women's rural realities, mobile Internet access, and "Africa Rising". *Information Technologies and International Development* 14 (01 2018).

[76] Yaxuan Yin, Longjie Guo, and Jacob Thebault-Spieker. 2024. Productivity or Equity? Tradeoffs in Volunteer Microtasking in Humanitarian OpenStreetMap. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 113 (April 2024), 34 pages. doi:10.1145/3637390

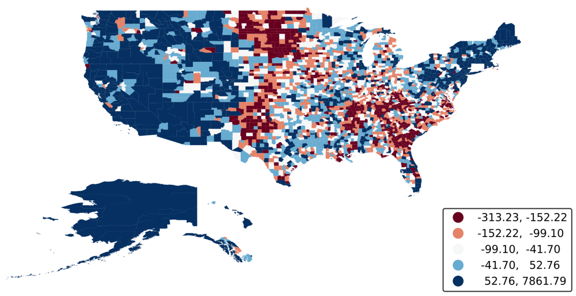
[77] Yaxuan Yin and Jacob Thebault-Spieker. 2025. Beyond 'Geo' HCI: Exploring Cultural Dimensions of Disparity in OpenStreetMap Road Safety Metadata. 9, 7, Article CSCW275 (Oct. 2025), 31 pages. doi:10.1145/3757456

[78] Yaxuan Yin and Jacob Thebault-Spieker. 2025. The Effect of Population Density on Remote Humanitarian Mapping Activities: A Triple-Difference Analysis. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW018 (May 2025), 21 pages. doi:10.1145/3710916

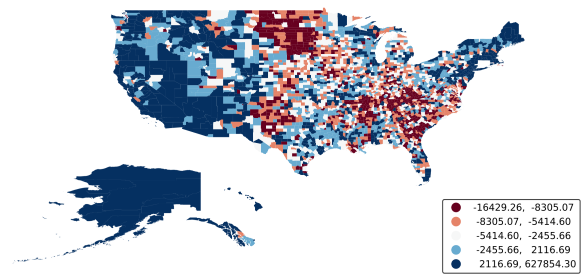
[79] H. Zhang, K. Zhang, T. Yan, et al. 2025. The impact of digital infrastructure on regional green innovation efficiency through industrial agglomeration and diversification. *Humanities and Social Sciences Communications* 12 (2025), 220. doi:10.1057/s41599-025-04512-9

[80] Dennis Zielstra, Hartwig H. Hochmair, Pascal Neis, and Francesco Tonini. 2014. Areal Delineation of Home Regions from Contribution and Editing Patterns in OpenStreetMap. *ISPRS International Journal of Geo-Information* 3, 4 (2014), 1211–1233. doi:10.3390/ijgi3041211

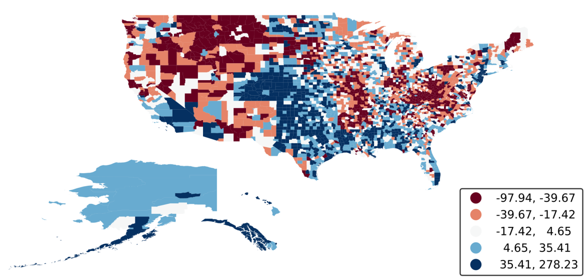
A Appendix: OLS Residual Diagnostic Maps



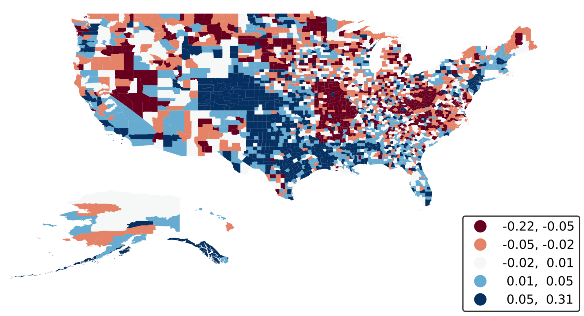
(a) Wikipedia Pages. Moran's I = 0.339, p < .001



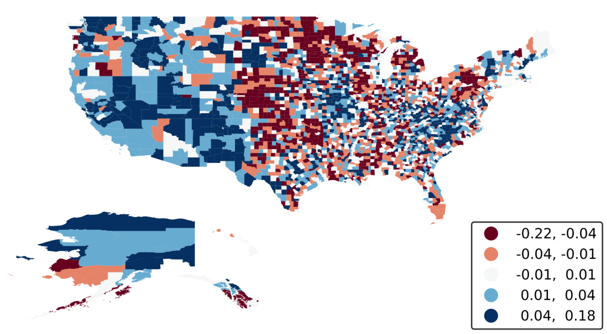
(b) Unique Contributors. Moran's I = 0.299, p < .001



(c) Edits per Page. Moran's I = 0.543, p < .001

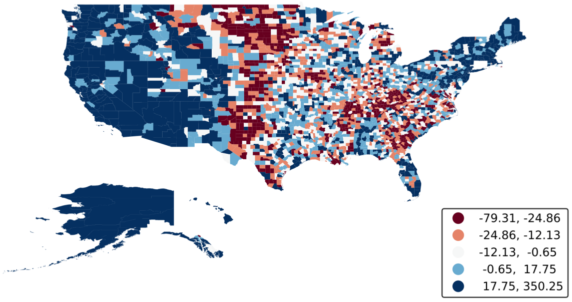


(d) Edit Concentration (Gini). Moran's I = 0.596, p < .001

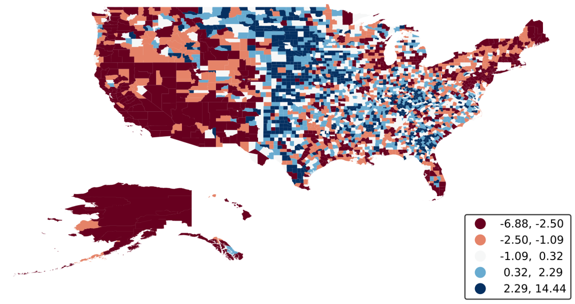


(e) Contributor Concentration (Gini). Moran's I = 0.350, p < .001

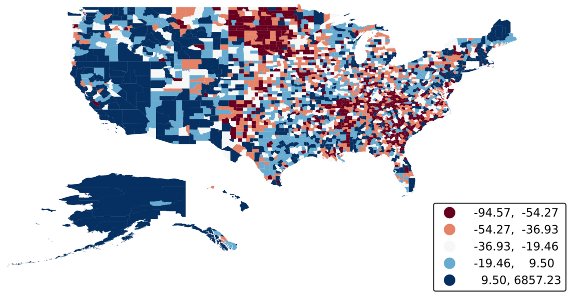
Figure 4: OLS residual diagnostic maps for page-level participation outcomes (Part 1). Each panel displays the spatial distribution of residuals from the non-spatial OLS model for a core page-level editing metric. Substantial spatial clustering, as indicated by the Moran's I values, shows that the OLS specification leaves significant geographic structure unexplained, motivating the use of spatial regression models.



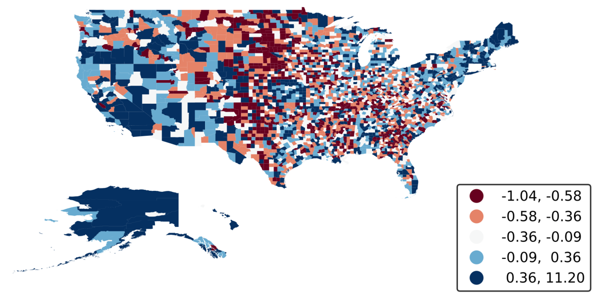
(a) Bot Contributors (#). Moran's $I = 0.497$, $p < .001$



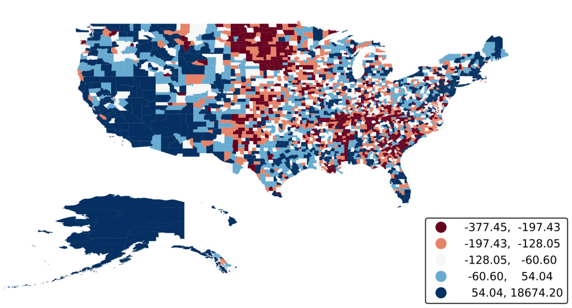
(b) Bot Contributors (%). Moran's $I = 0.441$, $p < .001$



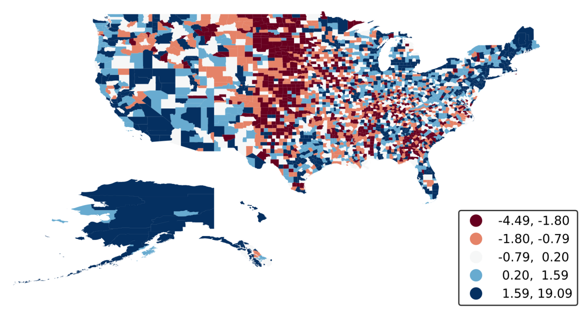
(c) Local Prolific (#). Moran's $I = 0.216$, $p < .001$



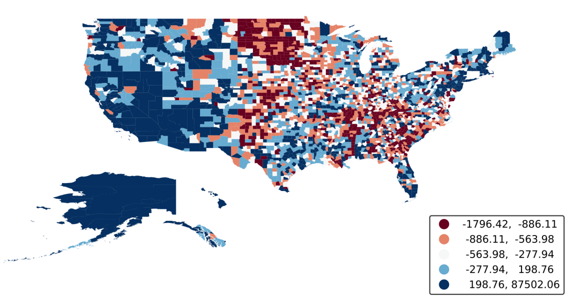
(d) Local Prolific (%). Moran's $I = 0.107$, $p < .001$



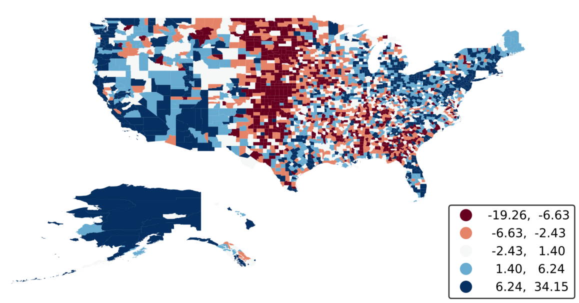
(e) Local Regular (#). Moran's $I = 0.292$, $p < .001$



(f) Local Regular (%). Moran's $I = 0.356$, $p < .001$

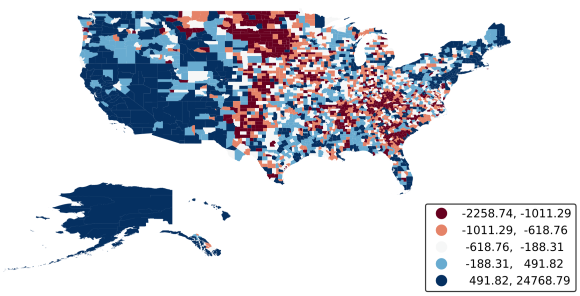


(g) Local Casual Contributors (#). Moran's $I = 0.296$, $p < .001$

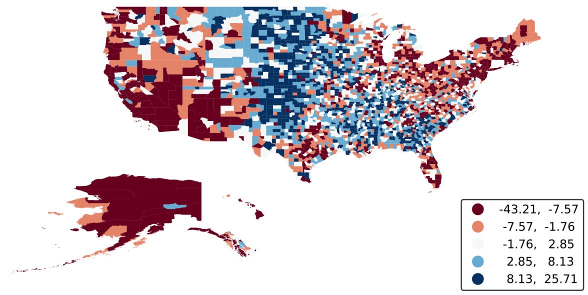


(h) Local Casual Contributors (%). Moran's $I = 0.508$, $p < .001$

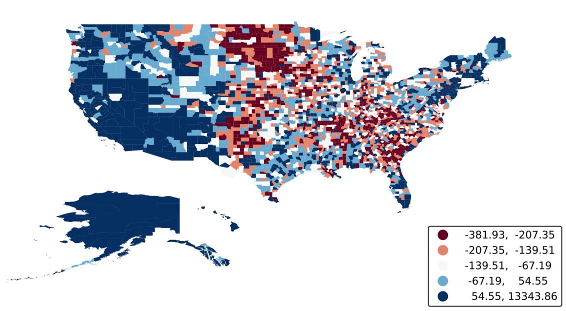
Figure 5: OLS residual diagnostic maps for automated activity and local contributor roles (Part 2). Residual maps for bot activity and for prolific, regular, and casual *local* contributors continue to exhibit strong spatial autocorrelation. The persistence of these geographic patterns further indicates that participation dynamics are spatially structured and cannot be adequately captured by non-spatial OLS models.



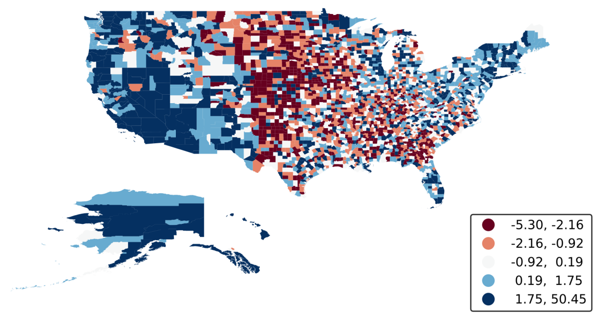
(a) Non-local Prolific (#). Moran's $I = 0.439$, $p < .001$



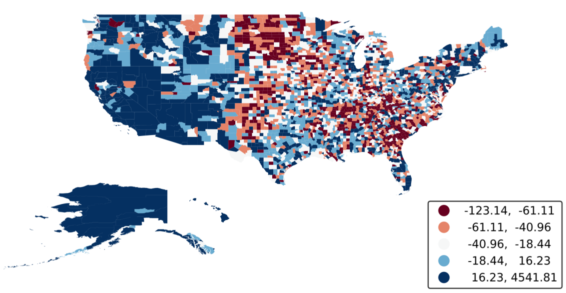
(b) Non-local Prolific (%). Moran's $I = 0.499$, $p < .001$



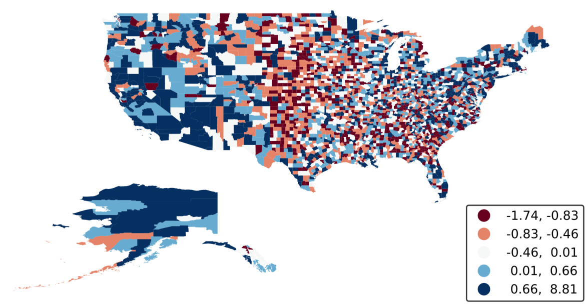
(c) Non-local Regular (#). Moran's $I = 0.307$, $p < .001$



(d) Non-local Regular (%). Moran's $I = 0.157$, $p < .001$



(e) Non-local Casual (#). Moran's $I = 0.288$, $p < .001$



(f) Non-local Casual (%). Moran's $I = 0.081$, $p < .001$

Figure 6: OLS residual diagnostic maps for non-local contributor roles (Part 3). The same diagnostic patterns emerge for prolific, regular, and casual *non-local* contributors, with significant residual clustering across counties. Together, Parts 1–3 demonstrate that spatial dependence is a systematic feature of Wikipedia participation across all contributor categories, reinforcing the need for the Spatial Durbin Model employed in the main analysis.

B Appendix: Spatial Weights and Contiguity Illustration

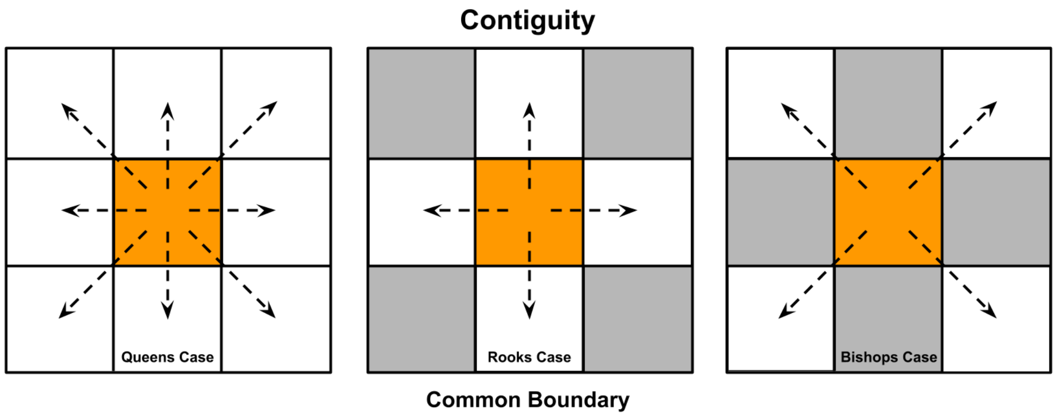


Figure 7: Illustration of contiguity relationships (Queen, Rook, Bishop) used to define neighbor structure in spatial models. Adapted from Chris Gentry's spatial regression tutorial [27].

C Appendix: Full Results

Table 5: Full Spatial Durbin Model (SDM) results for all participation outcomes using first-order Queen contiguity.

Metric	Broadband (pp)		Pop. Density ($\log_2$)	
	Direct Effect	Indirect Effect	Direct Effect	Indirect Effect
Number of Pages	2.58***	0.64	-0.20	23.56***
ln(Number of Pages)	0.01***	-0.01	0.00***	0.08***
Number of Unique Contributors	47.32***	16.64	-11.99	403.47***
ln(Number of Unique Contributors)	0.02***	-0.01***	0.01	0.07***
Average Total Edits	0.22***	0.07	-0.25	-5.06***
Contribution Concentration per Page	0.00***	-0.00***	-0.00	-0.01***
Labor Concentration	0.00***	-0.00***	0.00***	0.00
Bot Contributors (%)	-0.045***	0.017*	-0.023	-0.433***
Bot Contributors (#)	0.488***	-0.238**	-0.250	2.252**
Local Prolific Contributors (%)	0.006***	-0.004**	0.003	0.028*
Local Prolific Contributors (#)	0.961***	0.444	-0.076	8.918**
Local Regular Contributors (%)	0.029***	-0.006	0.056***	0.281***
Local Regular Contributors (#)	3.692***	2.305	-0.178	37.294***
Local Casual Contributors (%)	0.128***	-0.018	0.158***	1.506***
Local Casual Contributors (#)	16.264***	11.904*	-2.149	182.016***
Nonlocal Prolific Contributors (%)	-0.122***	-0.092***	-0.150*	-1.812***
Nonlocal Prolific Contributors (#)	21.454***	-0.680	-6.992	133.265**
Nonlocal Regular Contributors (%)	0.018***	-0.013**	0.016	0.145**
Nonlocal Regular Contributors (#)	3.691***	1.749	-2.545	31.886**
Nonlocal Casual Contributors (%)	0.005***	-0.003*	0.028***	0.082***
Nonlocal Casual Contributors (#)	0.996***	0.640	0.653	11.538***

Note. Broadband (pp) indicates coefficients represent the effect of a one–percentage-point increase in county broadband coverage. Asterisks denote significance after Benjamini–Hochberg FDR correction (* $p < .05$, ** $p < .01$, *** $p < .001$).

D Appendix: Robustness to Alternative Spatial-Weights Matrices

Table 6: Robustness check using first-order Rook contiguity. Spatial Durbin Model (SDM) results.

Metric	Broadband (pp)		Pop. Density ($\log_2$)	
	Direct Effect	Indirect Effect	Direct Effect	Indirect Effect
Number of Pages	2.41***	0.51	-0.18	21.87***
ln(Number of Pages)	0.010***	-0.008	0.000***	0.073***
Number of Unique Contributors	44.83***	14.92	-10.44	381.02***
ln(Number of Unique Contributors)	0.017***	-0.009**	0.008	0.064***
Average Total Edits	0.204***	0.061	-0.223	-4.82***
Contribution Concentration per Page	0.000***	-0.000***	-0.000	-0.010***
Labor Concentration	0.000***	-0.000***	0.000***	0.000
Bot Contributors (%)	-0.041***	0.014*	-0.022	-0.406***
Bot Contributors (#)	0.451***	-0.221**	-0.232	2.094**
Local Prolific Contributors (%)	0.005***	-0.003**	0.003	0.025*
Local Prolific Contributors (#)	0.885***	0.401	-0.072	8.432**
Local Regular Contributors (%)	0.027***	-0.005	0.053***	0.263***
Local Regular Contributors (#)	3.523***	2.114	-0.166	35.612***
Local Casual Contributors (%)	0.121***	-0.016	0.148***	1.431***
Local Casual Contributors (#)	15.713***	10.923*	-1.951	174.883***
Nonlocal Prolific Contributors (%)	-0.113***	-0.085***	-0.142*	-1.692***
Nonlocal Prolific Contributors (#)	20.532***	-0.601	-6.574	126.91**
Nonlocal Regular Contributors (%)	0.016***	-0.011**	0.015	0.134**
Nonlocal Regular Contributors (#)	3.421***	1.632	-2.414	29.774**
Nonlocal Casual Contributors (%)	0.004***	-0.003*	0.026***	0.077***
Nonlocal Casual Contributors (#)	0.927***	0.593	0.614	10.801***

Note. Broadband (pp) indicates coefficients represent the effect of a one–percentage-point increase in county broadband coverage. Asterisks denote significance after Benjamini–Hochberg FDR correction (* $p < .05$, ** $p < .01$, *** $p < .001$).

Table 7: Robustness check using 4-nearest-neighbor ($k = 4$) spatial weights. Spatial Durbin Model (SDM) results.

Metric	**Broadband (pp)**		**Pop. Density ($\log_2$)**	
	Direct Effect	**Indirect Effect**	**Direct Effect**	**Indirect Effect**
Number of Pages	2.66***	0.71*	-0.21	24.84***
ln(Number of Pages)	0.011***	-0.012	0.000***	0.087***
Number of Unique Contributors	49.65***	18.03*	-12.61	419.32***
ln(Number of Unique Contributors)	0.023***	-0.012**	0.01	0.074***
Average Total Edits	0.236***	0.084*	-0.268	-5.34***
Contribution Concentration per Page	0.000***	-0.000***	-0.000	-0.012***
Labor Concentration	0.000***	-0.000***	0.000***	0.000
Bot Contributors (%)	-0.048***	0.020*	-0.024	-0.462***
Bot Contributors (#)	0.513***	-0.257**	-0.268	2.331**
Local Prolific Contributors (%)	0.006***	-0.004**	0.003	0.030*
Local Prolific Contributors (#)	1.012***	0.491	-0.083	9.523**
Local Regular Contributors (%)	0.031***	-0.007	0.058***	0.298***
Local Regular Contributors (#)	3.834***	2.512	-0.193	39.442***
Local Casual Contributors (%)	0.134***	-0.021	0.165***	1.587***
Local Casual Contributors (#)	17.033***	12.503*	-2.389	189.334***
Nonlocal Prolific Contributors (%)	-0.129***	-0.099***	-0.158*	-1.923***
Nonlocal Prolific Contributors (#)	22.618***	-0.745	-7.213	140.871**
Nonlocal Regular Contributors (%)	0.019***	-0.014**	0.017	0.152**
Nonlocal Regular Contributors (#)	3.953***	1.884	-2.671	33.402**
Nonlocal Casual Contributors (%)	0.006***	-0.003*	0.029***	0.089***
Nonlocal Casual Contributors (#)	1.074***	0.702	0.689	12.207***

Note. Broadband (pp) indicates coefficients represent the effect of a one–percentage-point increase in county broadband coverage. Asterisks denote significance after Benjamini–Hochberg FDR correction (* $p < .05$, ** $p < .01$, *** $p < .001$).