Response to editor and reviewer comments

Editor and reviewer comments as well as author responses are presented below. Comments are in teal fonts and responses in black.

Thank you for submitting your manuscript to Journal of Economic Geography.

I have now received comments from reviewers. Their comments are at the end of this email. The reviewers raise a range of concerns and have questions about the suitability of the paper for publication. The overall feeling is that the paper has probably been submitted too early and needs major work on most sections if it is to become publishable. I concur with the reviewers and see a number of significant issues that need to be addressed. This left me with a dilemma in terms of the decision. I have decided to invite you to revise and resubmit the paper so that it can be considered further for publication. However, this is a very high-risk revision, and I was bordering on rejection given the scale of the work needed and the limitations that the reviewers highlight. I am saying this to be open about the risk that after the next round the paper could be rejected unless reviewers believe you have addressed the range of issues identified. You may want to factor this into your decision about whether to revise the paper, given how much work is needed.

Having said this, let me offer some suggestions about the main areas to focus on if you decide to revise the paper. Reviewer 1 is more positive but very brief in their report. I am, therefore, putting more weight on reviewers 2 and 3 when identifying priorities for the revision.

Thank you for providing the opportunity to revise and resubmit my paper. The reviewer reports were really helpful and I think it is fair to say that the reviewers did go the extra mile to offer useful recommendations. I think I have adequately addressed their concerns and my feeling is that the paper has now been drastically improved. I hope the editor and the reviewers will agree.

First, you will need to develop a more convincing theoretical argument and contribution. You will see that both reviewers 2 and 3 raise questions about your claims in relation to the gap in the literature and suggest you fail to build on existing work appropriately and do not convincingly identify the motivation for this analysis. In general, claims that an issue has never been addressed are hard to sustain and it is better to build on existing work and extend knowledge relating to a particular theoretical issue. The reviewers challenge you to do this

using your data and suggest both the argument and structure of the paper will need to change.

The paper has been restructured to better illustrate the theoretical argumentation, motivation and contribution. It is also better situated within the relevant literatures. Below I offer detailed responses to the reviewer comments.

Second, you need to justify your data and analysis better. Reviewer 2 fairly questions the end point of the analysis – which is not explained and given the 10+ years gap to your submission raises questions about the timeliness of the analysis. The questions about coverage of the UK website population also need to be addressed convincingly. More generally, the reviewers ask for more careful presentation of the data to make clear the significance of each figure and finding. As it stands, the paper feels rather overloaded with data, rather than acting as a focused analysis.

The data issues have been clarified.

The problems identified above result in the paper feeling imprecise, confusing and lacking a contribution. The end point is a conclusion that the reviewers do not find justified or original. The reviewers do see potential. As I noted, the feeling is a paper that is underdeveloped and submitted prematurely. There is, then, a route to publication, but it will require significant changes to convince the reviewers.

The restructuring of the paper around two hypotheses as well as the emphasis on the granularity of the analysis and the the long time series approach allowed to more clearly highlight the contribution in the conclusion section.

As a sidenote, there was a minor error in the previous version of the paper with the implementation of the spatially sensitive CV. This has now been corrected hence the minor changes in the modelling results (second decimal).

Referee: 1

I have read your paper with interest and I appreciate its quality and clarity. The work is both well-executed and well-written, making it a valuable contribution to the field.

However, I would like to offer a few minor suggestions that could further enhance the readability and overall coherence of the paper:

Thank you for your nice comments and suggestions.

Simplify the Introduction: I recommend streamlining the introduction by reducing the inclusion of overly technical or highly specific details. Instead, consider expanding on the broader conceptual framework that situates the paper within its scholarly context. This approach would help orient readers and provide a more accessible foundation for understanding the study.

Both the introduction and the literature review sections have been restructured and simplified. Some of the concepts initial unpacked in the introduction have now moved to the second section.

Strengthen the Connection Between Hypotheses and Results: It may be beneficial to establish a clearer and more explicit link between the study's hypotheses and its findings. One way to achieve this could be by using section titles that are more conceptually informative, which would guide the reader through the logical progression of the argument and findings more effectively.

The second section has been restructured and is now based on two hypotheses, which are also used to structure the methodology and the results sections. All section titles have been changed:

- The mechanisms that shape the diffusion of a new technology
- Results
  - The three diffusion mechanisms at granular geographical scales
  - Early adoption and latecomers
- Conclusions

Referee: 2

OVERVIEW
- focused on innovation diffusion, using the case of the web in the UK - "this paper tests whether the three distinct spatial diffusion mechanisms shaped the diffusion of the Web in the UK at very granular spatial scales from the Web's early stages onwards"; presumably "S-shaped pattern"/temporal models, core to periphery; distance/proximity.

Figure 2 shows the diffusion of domains per company at the scale of the UK matches the S growth model. Figure 3 shows some spatial clustering. I think this is for all years in the dataset but this could be better labeled

Both figures has been re-titled:

- Cumulative adoption of the Web, UK
- LAD Web diffusion speed, 1996-2012

Please note that the old Figure 3 has now moved to the end of the Results section (now Figure 8).

Figure 4, not sure what this demonstrates

This figure has also moved towards the end of the Results section as it is now used to test the second hypothesis. The new title reads "Figure 9: Dynamics of Wed diffusion based on LAD ranking" and the following paragraph explains how it was created and what it reveals.

> Next, the stability and volatility of these patterns is analysed to reveal patterns of early adoption and leapfrogging. Figure 9 plots the relative ranking of LAD based on website density at the beginning and the end of the study period. To decrease noise, the average ranking of 1996-1998 and 2010-2012 is plotted instead of the individual years. Each line represents a single LAD and the colours depict the 5^{th} and 95^{th} percentile of the absolute difference of the LAD ranking between 1996-1998 and 2010-2012. In other words, the 5^{th} percentile includes the LAD, the relative position of which remained unchanged between the beginning and the end of the study period, both at the top and at the bottom of the hierarchy (small absolute difference in ranking). The 95th percentile includes LAD that either increased or decreased a lot their relative positions over time (high absolute difference in ranking). Thus, Figure 9 reveals cases of *leapfrogging* since the LAD at the top-right part of the figure managed to jump at the top of the hierarchy at the end of the study period.

Figure 5&6, shows early clustering and then diffusion (I think)

Yes. This is also reflected in the text. For instance:

> This reflects the early concentration of high website density around London, which over time diffused as high website density clusters can be seen in other parts of the country away from London (Figure 5)

> Eventually, as the adoption rate increased, more such clusters of high website density were formed and this is reflected both in the Moran's I and the LISA maps in Figures 4 and 5. A similar pattern is observed for the extended dataset (see FiguresA.4 and A.5).

DATA
Focusing just on .co.uk website is OK but needs to be better justified given that it only represented about 50 percent of domains in the UK at that time. See my back of the envelop calculations: As per Zook (2001, Table 1) .uk represented 66% of all domains in the UK, .co.uk is a subset of that, so this is perhaps 50% of all domains registered in the UK at that time. If 1/5 of .co.uk domains had findable postal codes (p. 5), this study is based on 10% of the domains in the UK. (perhaps this is too low, aggregating from webpages to websites brought the number higher, not entirely clear to me)

Thanks for going the extra mile to check the data used here. I agree with the above calculation up until the findable postcode point as 1/5 of the archived *webpages* – not *websites*, i.e. domains – included a postcode in the web text. Think of the website of a small company with one trading address. I think it is fair to expect a website to have multiple webpages to present the different products and services and then one webpage to showcase the address – the 'Contact us' page. So, the 1/5 share of *webpages* containing a postcode should translate to a much higher share of *websites* – i.e. domains – containing a postcode. Previous estimates indicated that around

70% of all websites contain place reference (Hill 2009) and this reference has now been added in the paper.

So, based on the above 35% of all domains with a registered address in the UK were represented in the data. Most importantly though the paper does not model the adoption of the Web from the general population, but instead from firms as only commercial websites are considered. So, although the data only represents 35% of the population of websites with a registered address in the UK, it represents almost all the *relevant* archived websites – that is websites using the commercial .co.uk SLD. As stated in the paper, this has been the primary choice for commercial websites in the UK.

The third paragraph in the Data and Methods section includes the following text:

> Instead of using the whole .uk ccTLD, this paper focuses on its commercial subset, the .co.uk second level domain (SLD). This choice is aligned with the topic of the paper – the diffusion of the Web for commercial functions and decreases the heterogeneity of the web data as such commercial websites have specific aims: they are used to diffuse information, support online transactions and share opinions (Thelwall 2000; Blazquez and Domenech 2018).

overall the time period 1996-2012 is a good time to study the diffusion of the commercial internet. But why stop at 2012? Is it when the Library stopped its data series?

Yes, the British Library had obtained a copy of all of the .uk contents of the Internet Archive for that period. From a data perspective, it wouldn't be correct to attempt to extend this dataset with data from other open web archives, such as the Common Crawl, as they use different crawlers and there will be inconsistencies. Most importantly though, this period covers the diffusion of the web from the very early stages until its maturity. This is also illustrated by the figures depicting the cumulative adoption and the flattening right end of the curve for both subsets of the data (Figure 2 and A1).

There is a specific paragraph justifying how the time period of the data fits the aim of the paper and another line is now included in the second paragraph of the introduction ("from the very early days until its maturity (1996-2012)").

Framing "With the exception of Sinai and Waldfogel (2004), who focused, among other things, on web content and demonstrated that more such content is available in larger markets, this strand of literature has been mostly concerned with the spatiality of the Internet's hard and tangible infrastructure and shied away from exploring the spatial dimensions of one of the most distinct software layers of the Internet infrastructure, the Web." This statement is simply not true, the literature reviews misses the entirety of web content diffusion studies: Zook, 2000, Malecki, 2002. Townsend, 2000, Gorman, 2002, Feldman, 2002, lots more

Point taken. Despite being well aware of these studies, I omitted citing them as I was primarily searching for studies testing the diffusion mechanisms over longer time periods. This error has now been corrected and the above paragraph has been replaced with the following one, which

emphasises the long time series motivation. Regarding the actual references, Zook (2000), Zook (2001) and Moss and Townsend (1997) used relevant, domain name registration data, but the rest (e.g. Malecki (2002) and Moss and Townsend (2000)) focused on hard infrastructure. Gorman (2002) examined the head and branch locations of the top 40 e-business integration firms in the USA, which although interesting is not directly related to this paper. I also went through Feldman's work from around 2002, but I did not find papers adopting a similar empirical perspective.

> Another line of motivation has to do with the very nature of the Web as a digital technology and how it evolved over time. The digital geographies field has debunked premature claims about the Internet's placeless-ness by mapping various of its facets and technological layers (Emmanouil Tranos and Nijkamp 2013), but the lack of relevant long time-series data prevented researchers from observing digital phenomena overtime. Most of these studies focused on its hard infrastructure (e.g. Malecki 2002; Moss and Townsend 2000), on Internet subscribers (e.g. Blank, Graham, and Calvino 2018), or on social media users (e.g. Crampton et al. 2013). Exceptions include early work from Zook (2000), Zook (2001) and Moss and Townsend (1997), which utilised domain name registration data and use it to map and illustrate the spatial footprint of the web content. Pivotal as these contributions have been in understanding the economic geographies of the Internet, they were based on cross-sectional, often spatially detailed, mapping and on short temporal comparisons (e.g. 1993-1998) and did not observe the evolution and diffusion of the Web until its maturity. As we know from the relevant literature, different technologies diffuse at different rates and there is no theory that can explain these differences (Leibowicz, Krey, and Grubler 2016). Hence, mapping and modelling the diffusion of the Web can offer useful insights.

"for a digital technology as weightless as the Web" – this runs a bit counter to lots of work by digital geographers who have shown again and again how the digital is not weightless or placeless

See response to the next comment.

Conclusion p. 18 "This is the first time that the importance of granular spatial mechanisms for the diffusion of an intangible, digital technology is exposed." Honestly, I would not expect the diffusion of the web to be any different than any other kind of innovation diffusion. The spatiality of digital life, and its connection to material spaces is a constant refrain within digital geography. So I'm not convinced it makes much sense to emphasize this, if a granular spatial analysis of innovation diffusion is already novel.

Point taken. The emphasis now is on the granularity of the analysis and also on the long time series that allow to observe the technology almost from its beginning until its maturity. The above text as well as other similar text elsewhere in the paper have been amended accordingly.

Referee: 3

This paper addresses a gap in Economic Geography by examining the spatial diffusion of digital technologies, specifically the adoption of commercial websites, from 1996 to 2012. Recognizing the potential of digital technologies to shape local development trajectories through increased productivity and innovation, the study employs novel data sources and machine learning techniques to analyze diffusion patterns at highly granular geographical scales. The authors explore the roles of various spatial mechanisms driving this process, offering insights into how digital technology adoption unfolds across space and time. The paper has the potential to make an interesting contribution to the literature as it combines a fascinating topic with novel data and methodologies. However, in its current form, it is not ready for publication, as it never becomes clear what it actually delivers in terms of new knowledge and what knowledge gap it seeks to fill. Despite some analytical elements, it is rather descriptive in nature, which doesn't fit well with this journal.

Most importantly, I suggest developing some clear hypotheses to structure and motivate the paper better. In light of this, I recommend major revisions.

This is very helpful, thank you. The whole paper has now been restructured around two hypotheses. I believe the structure and the flow of the paper are much clearer now.

1. Abstract:
1. The abstract needs significant improvement. It needs to adequately report the main findings, relevance, or implications of the study. Additionally, the assertion that geographers have "shied away" from analyzing the diffusion of digital technologies is misleading; the issue was primarily due to the lack of available data.

The abstract has been amended accordingly:

> This paper investigates the spatial diffusion of a new technology that is the Web in the UK. It employs novel data and machine learning methods to model the influence of well-established diffusion mechanisms. Contrary to previous studies, it adopts multiple scales, high spatial granularity and a long study period that captures the early stages of the Web until its maturity (1996-2012). Findings reveal the importance of such spatial mechanisms (namely distance, urban hierarchy and the S-shaped pattern of the cumulative level of adoption) even at granular scales. They also highlight spatial heterogeneity and instances of leapfrogging.

2. Introduction
2. General Diffusion Studies: The statement that there are few contemporary general diffusion studies in Economic Geography is not accurate, although it may apply specifically to digital technologies. Consider the following recent examples: o Peris, A., Meijers, E., & van Ham, M. (2021). Information diffusion between Dutch cities: Revisiting Zipf and Pred using a computational social science approach. Computers, Environment and Urban Systems, 85(November), 101565. https://doi.org/10.1016/j.compenvurbsys.2020.101565

o Bednarz, M., & Broekel, T. (2020). Pulled or pushed? The spatial diffusion of wind energy between local demand and supply. Industrial and Corporate Change, 29(4), 893–916. https://doi.org/10.1093/icc/dtaa012 o Feldman, M. P., Kogler, D. F., & Rigby, D. L. (2015). rKnowledge: The Spatial Diffusion and Adoption of rDNA Methods. Regional Studies, 49(5), 798–817. https://doi.org/10.1080/00343404.2014.980799

Peris, Meijers, and van Ham (2021) is indeed an interesting and, to humble opinion, novel paper, but it is not directly relevant as it focuses on information diffusion rather than diffusion of new technologies. Information diffuses through social and, consequently, urban networks, but it is not adopted like new technologies. Then, Bednarz and Broekel (2020) was already cited and Feldman, Kogler, and Rigby (2015) has now been added. These references are included in the second section.

Regarding the statement about economic geography studies on diffusion of new technologies and innovation, respectfully, I do believe that this is a fair statement. The references listed below are the ones reviewed in the second section and none of them are published in one of the flagship economic geography journals (*Journal of Economic Geography* and *Economic Geography*). There are two papers published in *The Professional Geographer*, which is a broad-scope geography journal and then only one paper in *Regional Studies*, which is the closest to economic geography. Importantly though, the *Regional Studies* paper (Feldman, Kogler, and Rigby 2015) is focused on inventions as observed by patent data contrary to the broader 'adoption of a new technology' approach of this paper. All other papers listed below are published in business and economics journals.

1. Beardsell and Henderson (1999): *European Economic Review*
2. Bednarz and Broekel (2020): *Industrial and Corporate Change*
3. Haller and Siedschlag (2011): *Applied Economics*
4. Perkins and Neumayer (2011): *The Professional Geographer*
5. Feldman, Kogler, and Rigby (2015): *Regional Studies*
6. Ding, Haynes, and Li (2010): *The Professional Geographer*
7. Bakher Naseri and Elliott (2013): *Journal of Marketing Analytics*
8. Papagiannidis et al. (2015): *Technological Forecasting and Social Change*
9. Lengyel et al. (2020): *Scientific reports*

Also, the lack of spatial granularity argument has been highlighted in various instances throughout the paper as the main line of justification.

3. Novelty: The study needs to be more explicit about its novel contributions to understanding spatial diffusion mechanisms. Currently, the discussion is too general.

The novelty of the paper lies in the fact that there are hardly any studies that analysed the role of these three spatial mechanisms in the diffusion of a new technology at such granular spatial scales. Importantly, the paper does this by adopting a long time series perspective which captures the commercial beginning of the Web until its maturity. The following paragraph in the introduction highlights this:

> This paper offers such a contribution: an economic geography study revealing the importance of spatial mechanisms for the diffusion of a digital technology that is the Web in the UK at a high level of spatial granularity from the very early days until its maturity (1996-2012). It specifically focuses on the diffusion of commercial websites. It does so by employing a novel source of big data and machine learning algorithms to model the *active* engagement with Web – that is creating and maintaining instead of just browsing a website – during that period and tests how different spatial and temporal mechanisms shaped its diffusion. This paper also exemplifies how the combination of data sources and state-of-the-art research methods which escape the traditional social science domain can offer new lenses to geographical research regarding the understanding of technological diffusion.

4. Motivation: The motivation for the study is unclear. The text raises several issues, but it is unclear which are specific to digital technologies and which pertain to technological diffusion more broadly.

There are three paragraphs in the revised introduction that highlight the paper's motivation. The first paragraph highlights the importance of this work from a stakeholder perspective. The second emphasises the importance of technology and its adoption in economic geography and macroeconomic growth theories. The third one, motivates the paper from the standpoint of digital geographies and connects it to that literature.

> The motivation for this paper lies in the fact that there are various stakeholders who are interested in knowing how new digital technologies diffuse over space and time and use this knowledge to make predictions about the diffusion of related future technologies. As per Leibowicz, Krey, and Grubler (2016), historical studies agree that technologies diffuse differently in terms of times, rates, and geographies and can be driven by related policies (Victor 1993). Meade and Islam (2021) highlight that a variety of actors have a direct interest in gaining such knowledge including network equipment suppliers, network operators, regulatory and local authorities. These processes and their effects vary a lot across scales: although the diffusion of a new technology might not be optimal at a local level, it might be beneficial from a global perspective as it could lead to faster diffusion to less advantaged places (Leibowicz, Krey, and Grubler 2016). Despite the spatial heterogeneity of such diffusion mechanisms and the policy relevance, there is a scarcity of studies analysing the diffusion of new technologies at a detailed geographical level because of the lack of detailed enough data and adequate research methods.
>
> This gap becomes even more prominent if we consider the importance of technology and its adoption in economic geography and macroeconomic growth theories. Simply put, differences in technological adoption and sophistication may lead to uneven development as new technologies can increase productivity and allow firms to introduce new products (Solow 1957; Aghion 1990; Kemeny 2011). Linked to the subject of this paper, the Web and, more specifically, commercial websites can be

considered as the underpinning technology for the transformative process identified by Capello, Lenzi, and Panzera (2024) as the *digital service economy*. Importantly, the focus here is not the Web in general, but instead commercial websites. After all, diffusion together with invention and innovation are considered the pillars of technological change (Das 2022). Findings from previous studies illustrated that early engagement with digital technologies such as the Web led to longstanding positive productivity gains for these regions that hosted early adopters (E. Tranos, Kitsos, and Ortega-Argilés 2021). Hence, obtaining a robust understanding of the spatial diffusion mechanisms is crucial in mapping future development trajectories.

Another line of motivation has to do with the very nature of the Web as a digital technology and how it evolved over time. The digital geographies field has debunked premature claims about the Internet's placeless-ness by mapping various of its facets and technological layers (Emmanouil Tranos and Nijkamp 2013), but the lack of relevant long time-series data prevented researchers from observing digital phenomena overtime. Most of these studies focused on its hard infrastructure (e.g. Malecki 2002; Moss and Townsend 2000), on Internet subscribers (e.g. Blank, Graham, and Calvino 2018), or on social media users (e.g. Crampton et al. 2013). Exceptions include early work from Zook (2000), Zook (2000) and Moss and Townsend (1997), which utilised domain name registration data and use it to map and illustrate the spatial footprint of the web content. Pivotal as these contributions have been in understanding the economic geographies of the Internet, they were based on cross-sectional, often spatially detailed, mapping and on short temporal comparisons (e.g. 1993-1998) and did not observe the evolution and diffusion of the Web until its maturity. As we know from the relevant literature, different technologies diffuse at different rates and there is no theory that can explain these differences (Leibowicz, Krey, and Grubler 2016). Hence, mapping and modelling the diffusion of the Web can offer useful insights.

5. Structure and Focus: The discussion on motivations should be streamlined and clarified. For example, after specifying that the study focuses on commercial websites rather than the web, subsequent references revert to the general term "web." Additionally, the discussion confuses diffusion patterns, firm perspectives, and early adopter advantages. Consider restructuring this section to clearly differentiate between literature gaps (motivations) and expected outcomes (mechanisms).

The revised paper consistently illustrates that the focus is on commercial websites.

In the relevant paragraph in the Data and Methods section, the following line has been added: "Henceforward any reference to websites will refer to commercial – .co.uk – websites".

The introduction and literature review sections have been restructured, so the firm perspectives and early adopter advantages discussions have been moved.

6. Data Consistency: The terminology for the data should be consistent. At one point, the text refers to the "totality of the web," while elsewhere, it specifies "commercial websites." Clarify and maintain consistency throughout.

The terminology is now consistent across the paper. The "totality" point referred to different web technologies and not to the totality of the web. It has now been replaced with "an overarching technology".

7. Recommended Structure: To align with the common structure in Journal of Economic Geography (JoEG), consider splitting the introduction into a shorter introduction and a dedicated theory section merged with the literature review.

The introduction is now shorter and the literature review section has been re-structured and re-named as "The mechanisms that shape the diffusion of a new technology".

3. Literature Review / Theory
1. Synthesis and Motivation: The literature review should be transformed into a theory section. Rather than presenting a historical review, synthesize the arguments to clearly motivate and inform your hypotheses.

This section has been re-written and structured around the two hypotheses:

> H_1: The three diffusion mechanisms namely distance, urban hierarchy and the S-shaped pattern of the cumulative level of adoption shaped the diffusion of the commercial Web in the UK even at granular geographical scales.
>
> H_2: The diffusion of the commercial Web is shaped by both early adoption and latecomer effects.

2. Knowledge Gaps: The current review does not clearly identify the knowledge gaps that this study aims to address. Clarify whether the goal is to confirm known processes with new data and methods or to explore unique diffusion patterns related to websites.

The knowledge gaps have been explicitly highlighted in the restructured literature review section. For instance:

> Although the above processes have been studied in innovation studies, the distancing of economic geographers from research questions about the diffusion of new technologies and innovations resulted to a knowledge gap regarding the explanatory power of these mechanisms at granular, local scales (Ding, Haynes, and Li 2010).
>
> Although the early adoption and leapfrogging literature has not paid much attention on cities and regions (Yu and Gibbs 2018), the potential local economic benefits justify the need to investigate the dynamics of the web diffusion in different scales in the UK. Hence, the second empirical hypothesis states:

3. Lack of Hypotheses: To clarify what the paper delivers and what knowledge gaps it fills, I strongly recommend developing a number of hypotheses that will be tested.

The following two hypotheses have been added and the whole paper is now structured around them:

> H_1: The three diffusion mechanisms namely distance, urban hierarchy and the S-shaped pattern of the cumulative level of adoption shaped the diffusion of the commercial Web in the UK even at granular geographical scales.
>
> H_2: The diffusion of the commercial Web is shaped by both early adoption and latecomer effects.

4. Fit to title: The term "multi-scalar" appears in the title but is never really discussed in the paper.

This was discussed from the data section onwards, but indeed not before that. Section 2 includes a new paragraph describing the second hypothesis (H_2), which explicitly discusses the value of the multiscale approach: "Understanding the diffusion of a new technology in multiple, detailed spatial scales is of interest to economic geography as …"

The two scales are firstly introduced in the introduction: "Importantly, the analysis adopts two distinct geographical scales: the LAD and the most detailed scale of UK census areas, the Output Areas (OA)".

4. Data & Methods
1. Order of Presentation: Introduce the data first, followed by the geographical units used in the study.

The order of presentation has been amended accordingly and the two scales of analyses are presented after the data.

2. Postcode Identification: The approach of identifying postcodes on websites is innovative, but potential errors need to be discussed. Specifically: - How do you ensure that a postcode on a website accurately represents the company's location or its subsidiary? - How do you handle numbers that resemble postcodes but are not actual postcodes? - What about references to events or other organizations unrelated to the website's owner? Include a brief discussion of these potential issues.

A number of different strategies have been developed to address the potential issues described above.

1. To prevent false positives, the resulted postcodes were matched against official postcode lookup files (Office for National Statistics 2025). This was not explicitly mentioned in the first submission of the paper and the previous sentence has now been added in the relevant paragraph:

> The text from these webpages was scanned using a regular expression (regex) to identify strings of text which resemble UK postcodes and one fifth of them included a mention to a postcode (Jackson 2017). This information allows the geolocation of the data and the creation of the LAD and OA counts. To prevent false positives, the resulted postcodes were matched against official postcode lookup files (Office for National Statistics 2025).

2. Then, to remove the bias from link farms associated with a handful of postcodes having irregular high numbers of websites pointing to them, a targeted data imputation procedure was developed. This is detailed in the data section – see also Figure 1.

3. Regarding the points about subsidiary company location and event postings, these cannot be directly observed and distinguished. However, a process that has been successfully used in previous studies has been adopted here. Firstly, given that the focus of the paper is commercial websites, the relevant literature illustrates that such websites serve specific purposes:

> … they are used to diffuse information, support online transactions and share opinions (Thelwall 2000; Blazquez and Domenech 2018).

Secondly, the webpage-to-website aggregation process leads to websites with different number of postcodes as presented in Table 1. To decrease heterogeneity, the analysis firstly focuses on websites with one unique postcode. Given the the commercial nature of these websites (.co.uk), it is fair to expect that such websites with one unique postcode in their web text represent companies with a single trading location. The analysis included in the main body of the paper is based on this subset of the website population. For year 2000, this includes 70% of all the archived, geolocated commercial websites. Then, as a robustness check all the different steps of the analysis are repeated for websites with up to 10 postcodes and the results are presented in the appendix. This includes 94% of all the archived, geolocated commercial websites for 2000. For this step, websites with multiple postcodes were counted multiple times, one for each different postcode. All the results are qualitatively the same and only marginally different, which is an indication of the robustness of the website geolocation strategy.

As a further argument for the robustness of the above process, the results of a previous study which used the same approach for a small subset of these data from a specific neighbourhood in London matched in terms of accuracy previous research that was based on extensive qualitative work, interviews and observational studies (Stich, Tranos, and Nathan 2023). This last point has now been added in the relevant section:

> Previous research which used the same methodology for a small subset of these data illustrates the robustness of the above approach: its results matched in terms of accuracy past rounds of research based on extensive qualitative work, interviews and observational studies for the same study area (Stich, Tranos, and Nathan 2023).

3. Equation 2: The rationale behind Equation 2 is unclear. Shouldn't the primary predictor of website density at time tbe the location density at t-1? Why was this approach chosen over a growth model? Additionally, clarify how spatial dependencies are accounted for in the model.

Because of the restructuring of the methods section, which now includes all the details of the modelling exercise, Equation 2 is followed by Equation 3 that clarifies the spatial and temporal dependencies. All these details are described in the following paragraph, which is placed just before Equation 3:

> To capture the hierarchical mechanism the model include as predictors a one year lag of website density in London, the largest city in the UK, a one year lag of the website density in the nearest city and the same for the nearest retail centre. Due to the small sizes of the retail centres, the latter is only relevant for and included in the OA models and not in the LAD models. In addition, the model includes the distance to London, the nearest city and the nearest retail centre. The underlying logic is that the level of website adoption in a spatial unit depends on the level of the adoption in places further up in the urban hierarchy the previous year. The inclusion of the distance variables incorporates spatial structure into the hierarchy argument. To depict the neighbourhood mechanism, the website density of the neighbouring spatial units in the previous year is employed. This is calculated using rook continuity. Again, the underpinning rationale is that the level of web adoption within a spatial unit depends on the level of web adoption in the neighbouring spatial units the year before. This represents the 'hitting nearby locations first' argument. Therefore, the spatial and temporal lag of the website density in LAD and OA is calculated. Lastly, the temporal pattern of the cumulative adoption is captured by a time trend variable. Hence, the model will follow the following generic form (Eq. 3):

Regarding the growth model point, growth rates can be and have been calculated and modelled for LAD. However, such a strategy would have been problematic for OA, because of the very large frequency of OA without any websites. As mentioned in the text:

> In 1997 only 1% of the UK OA were associated with at least one website. This should not come as a surprise given that this was the very beginning of the commercial Internet and any activities with a digital footprint were concentrated in a handful of areas, as illustrated in Figure 5. At the end of the study period almost half of the UK OA were not associated with a website. Again, given the granularity of the data this was expected.

Therefore, if it was to to calculate growth rates for OA, a small constant should have been added to all OA. Given how many 0s exist, this would have been problematic and add unnecessary noise. Given that a multi-scale and granular approach was important for this paper and part of its novelty, the levels modelling framework was adopted. However, the current revised version

of the paper now reports such a LAD growth model in an appendix. The following text has been added in the methods section:

> An alternative specification could model the growth of website density. However, such a strategy is problematic because of the large frequency of OA without websites, which is to be expected given their small size and high granularity. Still, as a robustness text, the results of a growth model for LAD are presented in the Appendix.

Then, the following text has been added in the Results section:

> In addition, Table A3 in the Appendix presents the accuracy results of a LAD model the dependent variable of which is the yearly growth rate of website density. Even for such a specification (see Equation 7 in the Appendix) the predictive capacity of model is exceptionally high ($R^2 = 0.5$).

5. Results

1. Diffusion Speed: The decision to binarize diffusion speed is questionable. Continuous measures, such as growth rates or the time required to reach 50% saturation, would provide more informative insights.

Please see my response to the previous point regarding modelling growth. Specifically for this point, the map with the 'binarised' speed (now Figure 8) has been moved further down (Section 4.2) to address H_2. Please note that the modelling part does not use this binarised variable, but instead continuous ones. Regarding the map, its aim is to depict heterogeneity in Web adoption. I could have, of course, plotted the year that each LAD reached their inflection point, which is very similar to "the time required to reach 50% saturation", using 4-5 bins. But I think the comparison with the country average makes it more accesible for the reader. The following text has been added in the first paragraph of section 4.2:

> Inflection points have been used in the relevant literature to compare diffusion processes across different settings and technologies (Cherp et al. 2021; Woo and Magee 2017) as they can enable more insightful comparisons than, for example, the average growth over the 1996-2012 study period.

In addition, the inflection (t_0) estimates, which used for the above map are presented in detail in the appendix and are also discussed in the main body of the paper:

> Importantly, there is obvious spatial clustering of fast and slow LAD. Out of the ten fastest LAD based on when they reached their inflection point (see detailed Table A.2 in the Appendix), nine were located in the South East of England and London and one in Cambridge.

They are also presented in Table 6 for the late adopters discussion.

2. Geographical Scale: While the effort to model diffusion curves at a fine geographical scale is commendable, the paper does not sufficiently explain the benefits or additional knowledge gained from this approach.

A new figure (Figure 3) and the following text have been added to address this point.

> What escaped from previous studies was exploring whether such a pattern also describes cumulative adoption of new technologies at sub-national scales. To test this, the same logistic function is estimated for each of the 374 LAD. The detailed results are presented in Table **??** in the Appendix. The key finding is that for 342 LAD (91% of all LAD) the $R^2 > 0.9$ for the logistic function, which indicates that even at local scale the cumulative adoption of the Web is characterised by a pattern similar to the one observed for countries. Graphically, all these LAD curves are plotted in Figure **??**. The different colour lines illustrate whether a LAD reached t_0 faster or slower than the UK in total. This heterogeneity is further discussed later in this section.
>
> This is the first time that such cumulative adoption curves have been estimated for areal units as small as the UK LAD. Understanding and modelling the dynamics of technological diffusion and, ultimately, being able to predict diffusion rates at such small local scales can directly support the deployment of infrastructural networks (e.g. 5G) and local technology policies. The above are in line with previous studies exploring the diffusion of web technologies (Papagiannidis et al. 2015), social media (Lengyel et al. 2020) and online shopping (Bakher Naseri and Elliott 2013).

3. Early Adopters and Latecomers: The discussion of early-adopter and latecomer advantages is underdeveloped. How do these advantages manifest locally? What specific benefits arise if a company registers its website six months earlier than competitors in other locations? If this aspect is retained, it should be elaborated in the theory section, including how it relates to local capabilities or diffusion hierarchies.

The results section has been restructured in order to match the two hypotheses structure. The early adopters part has been further developed with more empirical results and now answers the second hypothesis (H_2). The key message is that the diffusion of the Web at the detailed scale of LAD is characterised by spatial heterogeneity, well-established spatial patterns and both stability and volatility. Although the relevant literature neglected local scales, the above findings illustrated that such local dynamics exists and worth further consideration and investigation especially regarding potential local economic effects.

4. Descriptive Statistics: The value of the descriptive Moran's I and Gini coefficient analyses in Section 4.2 is unclear. Instead, calculating growth rates of adoption and examining their distribution over time would offer a better sense of adoption speed and magnitude.

Please see previous responses about growth rates, zeros and OA. Specifically for this point, Moran's I and the Gini coefficient offer some first insights about the two of the three diffusion mechanisms that H_1 tests. Section 3 has now been restructured based on the two hypotheses, and the following text describes their value:

> To assess the second diffusion mechanisms, Figures 4 and 5 offer a first insight into whether a neighbourhood or, in other words, a distance mechanism underpins the diffusion of the Web in the UK as they present the Moran's I and the LISA maps of website density respectively for LAD and OA.

More broadly, on top of the fact that in order to calculate growth rates for OA a small constant should have been added to overcome the 0s problem, it is not clear how their distribution could offer insights about the two diffusion mechanisms. It is also unclear to me how such distribution could be used – the distribution of yearly growth rates for every LAD and OA or for all LAD and OA? Importantly there are c. 230,000 OA.

5. Retail Centers: The sudden shift to retail centers on page 15 is confusing, given that the data pertains to all commercial sites. Clarify the rationale for this focus.

The retail centres discussion has now been moved in the methods sections, so it appears much earlier in the paper. In addition, the following text has been modified in the relevant paragraph:

> To capture the hierarchical mechanism the model include as predictors a one year lag of website density in London, the largest city in the UK, and a one year lag of the website density in the nearest city. Both variables clearly illustrate the urban hierarchy in the UK. Especially for the OA models, a third predictor is added to capture hierarchy at a much more granular scale: a one year lag of the website density in the the nearest retail centre [1]. As the literature illustrates, retail centres are key components of urban systems Jones (2021). Due to the small sizes of the retail centres, the latter is only relevant for and included in the OA models and not in the LAD models.

6. Modeling Details: The methodological details in Section 4.3 should be moved to the methods section to maintain clarity and coherence. Including these details in the results section is confusing.

The modelling details have now been moved to the methods section.

7. Clarity and Structure: Section 4.3 is difficult to follow. Clearly state the main message and highlight how diffusion mechanisms vary spatially. Develop hypotheses to structure the results

[1] The retail centres data have been obtained by the Consumer Data Research Centre, an ESRC data investment, under project ID CDRC 498-01, ES/L011840/1; ES/L011891/1: https://data.cdrc.ac.uk/dataset/retail-centre-boundaries-and-open-indicators.

and provide potential explanations for observed heterogeneity. Without this, the section feels like a descriptive summary without clear takeaways for the reader.

This section has been restructured based on the two hypotheses. I believe the key messages are now more clearly communicated.

Overall Recommendations 1. Consistency: Ensure consistent terminology and focus throughout the paper, particularly when referring to the web and commercial websites. 2. Structure: Adopt a clearer structure by separating the introduction, theory/literature review, methods, and results sections. 3. Clarity: Develop hypotheses to frame the results and provide clearer explanations for methodological choices and findings.

Thank you, this was very helpful. The whole paper has been restructured based on the above suggestions.

References

Aghion, P. 1990. "A Model of Growth Through Creative Destruction." *Econometrica* 60: 428–39.

Bakher Naseri, Mohammad, and Greg Elliott. 2013. "The Diffusion of Online Shopping in Australia: Comparing the Bass, Logistic and Gompertz Growth Models." *Journal of Marketing Analytics* 1: 49–60.

Beardsell, Mark, and Vernon Henderson. 1999. "Spatial Evolution of the Computer Industry in the USA." *European Economic Review* 43 (2): 431–56.

Bednarz, Marcel, and Tom Broekel. 2020. "Pulled or Pushed? The Spatial Diffusion of Wind Energy Between Local Demand and Supply." *Industrial and Corporate Change* 29 (4): 893–916.

Blank, Grant, Mark Graham, and Claudio Calvino. 2018. "Local Geographies of Digital Inequality." *Social Science Computer Review* 36 (1): 82–102.

Blazquez, Desamparados, and Josep Domenech. 2018. "Big Data Sources and Methods for Social and Economic Analyses." *Technological Forecasting and Social Change* 130: 99–113.

Capello, Roberta, Camilla Lenzi, and Elisa Panzera. 2024. "The Nexus Between the Digital Service Economy and Intraregional Wage Inequalities." *Economic Geography*, 1–28.

Cherp, Aleh, Vadim Vinichenko, Jale Tosun, Joel A Gordon, and Jessica Jewell. 2021. "National Growth Dynamics of Wind and Solar Power Compared to the Growth Required for Global Climate Targets." *Nature Energy* 6 (7): 742–54.

Crampton, Jeremy W, Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W Wilson, and Matthew Zook. 2013. "Beyond the Geotag: Situating 'Big Data'and Leveraging the Potential of the Geoweb." *Cartography and Geographic Information Science* 40 (2): 130–39.

Das, Bibhunandini. 2022. "Diffusion of Innovations: Theoretical Perspectives and Empirical Evidence." *African Journal of Science, Technology, Innovation and Development* 14 (1): 94–103.

Dennis, Charles, David Marsland, and Tony Cockett. 2002. "Central Place Practice: Shopping Centre Attractiveness Measures, Hinterland Boundaries and the UK Retail Hierarchy." *Journal of Retailing and Consumer Services* 9 (4): 185–99.

Ding, Lei, Kingsley E Haynes, and Huaqun Li. 2010. "Modeling the Spatial Diffusion of Mobile Telecommunications in China." *The Professional Geographer* 62 (2): 248–63.

Feldman, Maryann P., Dieter F. Kogler, and David L. Rigby. 2015. "rKnowledge: The Spatial Diffusion and Adoption of rDNA Methods." *Regional Studies* 49 (5): 798–817.

Gorman, Sean P. 2002. "Where Are the Web Factories: The Urban Bias of e–Business Location." *Tijdschrift Voor Economische En Sociale Geografie* 93 (5): 522–36.

Haller, Stefanie A, and Iulia Siedschlag. 2011. "Determinants of ICT Adoption: Evidence from Firm-Level Data." *Applied Economics* 43 (26): 3775–88.

Hill, Linda L. 2009. *Georeferencing: The Geographic Associations of Information*. Mit Press.

Jackson, Andrew N. 2017. "JISC UK Web Domain Dataset (1996-2010) Geoindex." The British Library. https://doi.org/https://doi.org/10.5259/ukwa.ds.2/geo/1.

Jones, Colin. 2021. "Reframing the Intra-Urban Retail Hierarchy." *Cities* 109: 103037.

Kemeny, Thomas. 2011. "Are International Technology Gaps Growing or Shrinking in the Age of Globalization?" *Journal of Economic Geography* 11 (1): 1–35.

Leibowicz, Benjamin D, Volker Krey, and Arnulf Grubler. 2016. "Representing Spatial Technology Diffusion in an Energy System Optimization Model." *Technological Forecasting and Social Change* 103: 350–63.

Lengyel, Balázs, Eszter Bokányi, Riccardo Di Clemente, János Kertész, and Marta C González. 2020. "The Role of Geography in the Complex Diffusion of Innovations." *Scientific Reports* 10 (1): 15065.

Malecki, Edward J. 2002. "The Economic Geography of the Internet's Infrastructure." *Economic Geography* 78 (4): 399–424.

Meade, Nigel, and Towhidul Islam. 2021. "Modelling and Forecasting National Introduction Times for Successive Generations of Mobile Telephony." *Telecommunications Policy* 45 (3): 102088.

Moss, Mitchell L, and Anthony Townsend. 1997. "Tracking the Net: Using Domain Names to Measure the Growth of the Internet in US Cities." *Journal of Urban Technology* 4 (3): 47–60.

Moss, Mitchell L, and Anthony M Townsend. 2000. "The Internet Backbone and the American Metropolis." *The Information Society* 16 (1): 35–47.

Office for National Statistics. 2025. "Open Geography Portal." https://geoportal.statistics.gov.uk/.

Papagiannidis, Savvas, Bartosz Gebka, Drew Gertner, and Florian Stahl. 2015. "Diffusion of Web Technologies and Practices: A Longitudinal Study." *Technological Forecasting and Social Change* 96: 308–21. https://doi.org/https://doi.org/10.1016/j.techfore.2015.04.011.

Peris, Antoine, Evert Meijers, and Maarten van Ham. 2021. "Information Diffusion Between Dutch Cities: Revisiting Zipf and Pred Using a Computational Social Science Approach." *Computers, Environment and Urban Systems* 85: 101565. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2020.101565.

Perkins, Richard, and Eric Neumayer. 2011. "Is the Internet Really New After All? The

Determinants of Telecommunications Diffusion in Historical Perspective." *The Professional Geographer* 63 (1): 55–72.

Solow, Robert M. 1957. "Technical Change and the Aggregate Production Function." *The Review of Economics and Statistics* 39 (3): 312–20.

Stich, Christoph, Emmanouil Tranos, and Max Nathan. 2023. "Modeling Clusters from the Ground up: A Web Data Approach." *Environment and Planning B: Urban Analytics and City Science* 50 (1): 244–67. https://doi.org/10.1177/23998083221108185.

Thelwall, Mike. 2000. "Who Is Using the .co.uk Domain? Professional and Media Adoption of the Web." *International Journal of Information Management* 20 (6): 441–53. https://doi.org/https://doi.org/10.1016/S0268-4012(00)00038-4.

Tranos, E, T Kitsos, and R Ortega-Argilés. 2021. "Digital Economy in the UK: Regional Productivity Effects of Early Adoption." *Regional Studies* 55 (12): 1924–38.

Tranos, Emmanouil, and Peter Nijkamp. 2013. ""The Death of Distance Revisited: Cyber-Place, Physical and Relational Proximities"." *Journal of Regional Science* 53 (5): 855–73.

Victor, D. 1993. "Overt Diffusion as Technology Transfer." *International Journal of Research in Marketing* 18: 535–38.

Woo, JongRoul, and Christopher L Magee. 2017. "Exploring the Relationship Between Technological Improvement and Innovation Diffusion: An Empirical Test." *arXiv Preprint arXiv:1704.03597*.

Yu, Zhen, and David Gibbs. 2018. "Sustainability Transitions and Leapfrogging in Latecomer Cities: The Development of Solar Thermal Energy in Dezhou, China." *Regional Studies* 52 (1): 68–79.

Zook, Matthew A. 2000. "The Web of Production: The Economic Geography of Commercial Internet Content Production in the United States." *Environment and Planning A* 32 (3): 411–26.

———. 2001. "Old Hierarchies or New Networks of Centrality? The Global Geography of the Internet Content Market." *American Behavioral Scientist* 44 (10): 1679–96.