Dear Prof. Bian,

Thank you for providing the opportunity to revise and resubmit our paper. We found the reviewer comments very useful and constructive and we believe they enabled us to improve the paper. We were also very pleased to see that the reviewers found our paper . . .

blah blah

Kind Regards,

[removed for anonymity]

Reviewer: 1

> The paper reports research of detecting clusters of network access speed, especially under the pandemic for resilience. Such a topic is of great importance as we are all in the pandemic and we rely on the network cyberinfrastructure for work as stated, and also for online education for all levels of students. Therefore, it worth supports with a few major revisions needed:

Thank you for your very supportive comments.

> 1. The VGI approach of obtaining network speed maybe bias given that this probably is the digital guru group that will respond, maybe a discussion on this and picture the steps of selecting/soliciting and how representative this is.

We have added the following sentence to the first paragraph in section 3.2, explaining why we believe any bias balances out: "Indeed, those who seek to test their broadband are most likely to do so because they are experiencing slower speeds than expected, although any skew towards slower speeds is balanced by the likelihood that those who test their broadband are also more 'tech-savvy' and / or have purchased higher speed packages that are not delivering the promised level of service."

As we explain, our intentions were to analyse a dataset of experienced internet speeds. Those who use the speedcheck service are not solicited in any way, and no personally identifying data is included other than the geolocation of the test.

> 2. The results for network speed for the different regions are not very different with 1-2 out of 10Mbps difference and probably can not reflect the social disparities. Not sure if this is the facts in U.K. If so, then it's not a problem?

Firstly, mean speeds are not likely to be that different between regions of the UK (e.g. North or South England). The greater differences are between urban, built-up areas, and more rural areas within regions, although there are some local initiatives to address this in some areas. However, we assume that your comment refers to Table 1 and the differences in mean network speed between clusters? Whilst these differences are not large, the range for upload speeds is smaller than for download speeds, and a difference of 1-2Mbps slower may well be noticeable at the wrong moment. Importantly, we only consider the test which took place during working hours, so when speed drops during, let's say, a work Zoom call or when uploading large files, it will be noticeable. This is why we created the clusters by how the time series of speeds changes by day of the week and hour of the day, rather than focusing our analysis on mean speeds.

> 3. The relevance to pandemic is not very straightforward. This may be improved by adding some studies, e.g., how the clusters impact the effectiveness of work from home through survey. Or relates to the employment rate.

Regrettably, we did not have the opportunity to run a survey. As we say in the introduction, prior to the pandemic, the demand for telecommuting would not have put a strain on internet connection speeds. We have clarified this with the addition of the following sentence: "High levels of demand are one of the main causes of reduced reliability and slower connection speeds, as network bandwidth becomes congested". We also have added a variable to the auxiliary regression tracking the share of the population that was put on furlough. This was the scheme in the UK which, at least initially, prevented much change in the employment rate, but did show who could work and who could not at the height of the first wave of the pandemic.

4. The time series may be enhanced by putting a time range according to different events of the pandemic or other factors that relates to the pandemic.

We chose the time range to reflect the period of first lockdown in the UK, as we say in the first paragraph of 3.2: "We are particularly interested in upload speeds and the frequency of speed tests over the period from March to May 2020, as government statements indicate this encompasses the period when UK workers were first told to work from home if at all possible (GOV.UK 2020)".

We have also added the following text: "Schools and various retail, leisure and hospitality businesses were closed from late March, and restrictions were gradually eased from late May".

5. Would suggest a review of the spatiotemporal challenges of the covid-19.

We have now added the following paragraph in the 'Digital divides and economic resilience' section.

"Currently there is a hotly debated discussion in the literature regarding these exact changes in transportation and the structure of cities. We now have enough hard data to observe the drastic change in the space-time geography of cities around the world (e.g. Google 2020; Shibayama et al. 2021). Although there is a broad agreement that these changes during the pandemic have played a pivotal role in stopping the spread of the virus (Jia et al. 2020; Yang et al. 2020; Mu, Yeh, and Zhang 2020), we still have no evidence to what extend these increased levels of working from home, and the consequent decrease in commuting flows and the structure of cities, will remain post-pandemic".

6. A review of cyberinfrastructure and how network varies in different region would be useful.

In section 2.2 of the literature review, we have added additional detail specific to the UK:

"This finding broadly applies in the UK, where studies that also analysed broadband speed checks concluded that average speeds are lower in rural areas, something that has not been improved by policy measures to increase competition (Riddlesden and Singleton 2014; Nardotto, Valletti, and Verboven 2015). Riddlesden and Singleton (2014) found that levels of deprivation did not correlate with first level digital divides".

and

"Dense urban areas were shown to suffer more from slow down during peak hours, although these services were more likely to be improved by increased competition between providers, such as between new entrants and Virgin Media cable connections (Riddlesden and Singleton 2014; Nardotto, Valletti, and Verboven 2015). The latter were historically available to only 45% of premises in the UK (OfCom 2016), where the more lucrative and competitive market originally attracted the cable TV provider".

The latter bit about Virgin Media is moved from the results section, in response to reviewer 2, q7.

7. The first part introduction can be shorted to reflect the importance of network bandwidth to resilience of pandemic impacted life.

As mentioned in response to comment 3, we have added a sentence of clarification to highlight the link between speed and bandwidth. We have also restructured the introduction, so this becomes the second paragraph, the 3rd paragraph is slightly shorter, and we introduce our case study approach in the 4th paragraph. We also moved the paragraph describing the three levels of digital divides to the literature review.

Reviewer: 2

This article presents interesting and timely research in digital geography, which is directly relevant to digital divide associated with the COVID-19 pandemic. The pandemic provides a once-in-a-lifetime opportunity to examine the barriers of broadband usage for telecommuting. The first two sections were very well written. I enjoyed reading them. However, I have a few major concerns on the methodology and a lack of discussion.

Thank you for your very supportive comments.

1. The paper used k-means, a popular unsupervised classification in this study. The optimal number of clusters was determined as 13 based on cluster validation indices (CVIs). While

this number could be well justified, I feel it is too high to properly interpret the results. The interpretation and knowledge discovery from clustering results with a high k tend to be tedious. Being affected by the high k, the Results section appears to be very difficult to read and follow. As there are many different CVIs, I feel a lower k will benefit the interpretation of results.

Please see our response to the next point.

2. Another major comment is on the multinomial logit regression (MLR). This is partly related to my first comment on a larger k. While there is no hard limit on the number of categories in MLR, the interpretation of results will be difficult with a high k as well. In addition, if the dependence among independent variables (collinearity) occurs, the estimation of the multinomial logit model parameters becomes inaccurate. According to Table 2, it is very likely some of the variables could present collinearity (e.g., distance to nearest metro and tech jobs).

Following your suggestion, we have now opted to present the cluster analysis and the subsequent multinomial regression using $k = 13$. We indeed believe that the results are more clear now. we have added the following justification in the Methods and Data section:

"We initially run the algorithm for $k \in \mathbb{N} \bigcap [5, 15]$, calculated the cluster validity indices (CVIs) and then run the subsequent multinomial regression – see the end of this section for more details for the different k. Following Sardá-Espinosa (2019), to identify the optimal k we used the majority vote for the following CVIs: Silhouette (max), Score function (max), Calinski-Harabasz (max), Davies-Bouldin (min), Modified Davies-Bouldin (DB*, min), Dunn (max), COP (min). Nevertheless, we opted against using the optimal $k = 13$ solution as it was too large to allow for communicable LAD clusters. Instead we opted for a smaller $k = 9$, which led to a rather similar spatial pattern and, importantly, to a much higher R-squared in the subsequent explanatory regression (0.44 instead of 0.34)."

Regarding multicolinearity, the highest correlations were observed between $NVQ4+$ and $professional jobs$ (0.78) and $NVQ4+$ and earnings (0.69). Hence, we decided to exclude $NVQ4+$ from the regression.

3. Although Fig. 1 and 2 exhibit different speed tests patterns in 2019 and 2020, the data mining outcome lacks a baseline (prior to 2020) to compare with. It is very likely that the clusters in 2019 would be different from that in 2020. It will make the results more robust. Of course, I will not push this given the page limit. But some insights may be helpful in the Discussion (however, a Discussion section is missing).

ET: after we finalise k, we can rerun k-means for 2019 and decide - could also show mean speeds on a map for 2019 - this would be part of the response to reviewer 1 q2 and q6. **Let's discuss this**

4. No discussion was included in this manuscript. I expect some insights into various important issues, such as problem causes, any broadband policies and socioeconomic solutions, equitable development, and research uncertainties.

We have tightened up the results section, and begun this discussion in the last paragraph of the results section and then expanded the conclusion section to include a discussion of the results.

5. There are three levels of digital divide. According to the last paragraph in Section 2, only the intersection of level-1 and level-3 was included in this research. I am wondering how level-2 may fit in the general research framework. Could education and demographic variables be used to represent level-2 digital divide?

We clarify that the occupations included in the regression are more about the level-2 digital divide, whereas the average earnings and, particularly during the pandemic the % furloughed, are indications of level-3 digital divides. We did have education in the previous submission (NVQ4+), but have removed it due to multicollinearity – see response for point 2, which give more indication of not only education, but digital skills.

6. A few suggestions on the figures. Fig 1 and 2 have crowded tick labels for x-axis month and weekday. Fig 3 and 4 may be placed side by side with the same y-axis for easy comparison. Actually, I feel "Figure 3 shows that the high standard deviation" may be inaccurate as Fig 4 appears to exhibit higher variability. Fig 5 maps may include the locations of metropolitan or major cities to help interpretation. The warmer-cooler color scheme may match the population size instead of the ordinal cluster numbers. The current color scheme does not help interpretation at all.

x-axis for Figures 1 and 2 have been amended accordingly.

The temporal profiles of the upload clusters are now included in Figure 3.

Figure 5 now labelled Figure 4: the map has now been redesigned using a different colour palette to better reflect the discrete nature of the underpinning data – the time-series clusters. Points and labels of the main UK cities were also added.

7. A few minor things. In Sec 2.2, the statement "These returns have been even greater . . . " appears to be controversial. LAD acronym appears before it was defined. Virgin Media as a major broadband provider should be briefly introduced in the manuscript.

We have toned down the statement on returns to say: "In some aspects, these returns may have increased during the Covid-19 crisis. . . "

Apologies on LAD acronym, now corrected.

We also moved and integrated introduction of Virgin Media to the literature review s2.2. See also response to reviewer 1, q6.

References

Google. 2020. "See How Your Community Is Moving Around Differently Due to Covid-19. Google Mobility Reports." 2020. https://www.google.com/covid19/mobility/.

GOV.UK. 2020. "News and Communications." 2020. https://www.gov.uk/search/news-and-communications?level_one_taxon=5b7b9532-a775-4bd2-a3aa-6ce380184b6c&order=updated-oldest.

Jia, Jayson S, Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A Christakis. 2020. "Population Flow Drives Spatio-Temporal Distribution of Covid-19 in China." *Nature* 582 (7812): 389–94.

Mu, Xiaoyan, Anthony Gar-On Yeh, and Xiaohu Zhang. 2020. "The Interplay of Spatial Spread of Covid-19 and Human Mobility in the Urban System of China During the Chinese New Year." *Environment and Planning B: Urban Analytics and City Science*, 2399808320954211.

Nardotto, Mattia, Tommaso Valletti, and Frank Verboven. 2015. "Unbundling the Incumbent: Evidence from Uk Broadband." *Journal of the European Economic Association* 13 (2): 330–62.

OfCom. 2016. "Connected Nations Report 2016." https://www.ofcom.org.uk/__data/assets/pdf_file/0035/95876/CN-Report-2016.pdf.

Riddlesden, Dean, and Alex D Singleton. 2014. "Broadband Speed Equity: A New Digital Divide?" *Applied Geography* 52: 25–33.

Sardá-Espinosa, Alexis. 2019. "Time-Series Clustering in R Using the dtwclust Package." *The R Journal* 11 (1): 22–43. https://doi.org/10.32614/RJ-2019-023.

Shibayama, Takeru, Fabian Sandholzer, Barbara Laa, and Tadej Brezina. 2021. "Impact of Covid-19 Lockdown on Commuting: A Multi-Country Perspective." *European Journal of Transport and Infrastructure Research* 21 (1): 70–93.

Yang, Chaowei, Dexuan Sha, Qian Liu, Yun Li, Hai Lan, Weihe Wendy Guan, Tao Hu, et al. 2020. "Taking the Pulse of Covid-19: A Spatiotemporal Perspective." *International Journal of Digital Earth* 13 (10): 1186–1211.