1. Guest Editor Comments
   1. While clearly heading in the right direction, this paper will benefit considerably from further expositional clarity and additional content relating to the underlying themes of the Special Issue to which it is intended to contribute.  
      The need for expositional clarity is raised by both reviewers, but more particularly by reviewer 1. In this vein, our view is that it is not clear the included math is adding much if anything at all – the underlying concepts seem rather intuitively obvious; is the math actually needed rather than just a slightly more considered verbal approach?

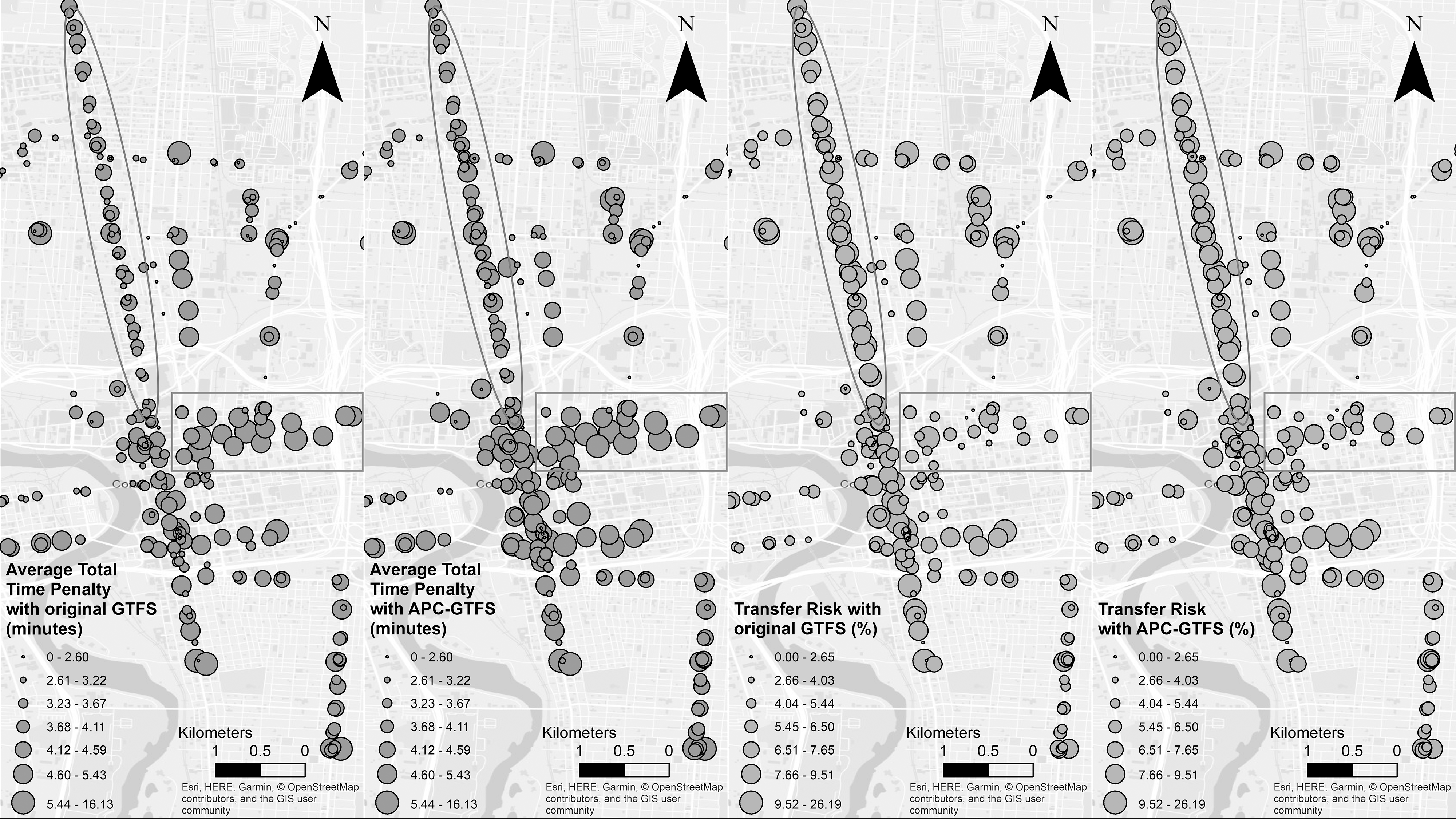
**Response**: Thank you for pointing this out. We incorporated formula 1 and 4 (last draft)[[1]](#footnote-1) into the paragraph to avoid confusions. Per reviewer 1’s comments, we simplified the notations for formula 1 and formula 2 (all current draft).

* 1. In terms of diagrams, Figure 2 is confusing – again, is it actually needed?

**Response**: It is a very good point. As the guest editor and reviewer pointed out, the definitions of three types of transfer are intuitive and obvious. Therefore, we decided to incorporate the Figure 2 (last draft) into the paragraph and remove the figure.

* 1. Figure 3 is, as reviewer 2 notes, currently hard to read – and will not replicate well in grayscale. In overall terms, more attention to the quality and suitability of the graphics will greatly enhance the approachability of the paper.

**Response**: It is also a good point. We changed the symbology of the map from graduated colors (circles with same size and different colors) to graduated symbols (circles with different sizes and same color). We believe current version will replicate better in grayscale: Picture 1 is a grayscale version of Figure 1 (current draft). For more clarification, please see our responses to reviewer 2’s comment below.



Picture 1: Grayscale version of Figure 1 (current draft).

* 1. The views of reviewer 2, on a need for the paper to speak more directly and substantively to an Urban Studies audience (and specifically to the special issue themes: Big Data as a tool for advancing our understanding of the urban, and to enhance urban well-being) also merit further emphasis. Attention to this matter has the potential to move the paper from a somewhat specialised contribution that will appeal to a limited cadre of transport experts, into something of much more general value and significance.

**Response**: We added many clarifications so that the paper is more suitable for the special issue theme. Generally, we focused on three parts in terms of emphasizing on the SI theme: the introduction, the literature review, and the conclusion.

In the introduction part, we emphasize the introduction of big data and its implication, both opportunities and challenges, on the transfer studies. To leverage and overcome the complicated nature of big data, we thereby introduce the measures and their advantages over other traditional measures.

In the literature review part, we start by introducing the history of transfer studies’ data sources. Moreover, we adopt the suggestion of reviewer 2; we changed the “deliberate versus byproduct data” to “small versus big data” to avoid unclear definitions and classifications, such as the review 3.3 raised by reviewer 2. In the second part, we focus on the development of transfer measures from non-real-time to real-time. We highlight the real-time characteristic of the two measures we are about to introduce and the challenge and academic gaps. By demonstrating the pros and cons of big data and the development of real-time measures, we also demonstrate the core contribution of the new measures: high-resolution, real-time, and computationally intensive measures of transfers with big data.

In the conclusion part, we adopt the reviewer 2’s comment about re-engaging the SI theme.

1. Reviewer 1

The paper makes a valuable contribution in public transport research by developing intuitive quality of transfer measures based on scheduled and real-time data. The paper is well written for the most part, however there are some unclear sentences. Some minor comments are listed below:

* 1. Fig 1: x-axis label further away than the legend headers, makes figure a bit hard to read.

**Response**: We adjusted the graph according to the comment (see Figure 1). We indicated x and y to the corresponding label to avoid confusion; we also moved the x-axis label to the middle to keep it consistent with the y-axis label.

* 1. The prime symbol used in the mathematical notation for “actual departure time” is quite difficult to notice. Can you please highlight that you use this symbol in text or alternatively use different symbols for clarity?

**Response**: We simplified the notations in the formula 1 (3.3, transfer time penalties section) and 2 (same section) in the current version. We now use uppercase T for the actual departure time and lowercase t for the scheduled departure time, which is much more obvious than prime symbols. Moreover, since all the time mentioned in the current formulas are now departure time, we removed the “departure” notations in the subscripts.

* 1. Fig 2: this is not giving much additional information. It could be changed to depict all the scenarios on page 14. There should also be a scheduled arrival time for the figure to make more sense. Furthermore, the figure text could highlight that the blue line is the chosen option.

**Response**: As the reviewer 1 and editor pointed out, the figure 2 in the last draft was confusing and did not give much additional information. Therefore, we removed figure 2 and added some further explanation to the definition of each transfer type.

* 1. Explanations for nonobvious abbreviations could be repeated in analysis section for convenience.

**Response**: We added several explanations in the analysis section.

* 1. Are there large differences in frequency between weekdays and weekends? What are the impacts on the measures? (similar effects as on the time of day comparison?)

**Response**: We added frequency in Figure 4 and Figure 5: the points are very few (3 for weekdays and 19 for hours), so it may be trivial to add another scatter point plot. Meanwhile, it is obvious to see the correlation from current line/bar plots. We also added some description and explanation in 4.2 – hourly pattern and weekday pattern.

Weekdays and impacts – Yes, there are. COTA system have three schedules: weekdays, Saturdays, and Sundays. We can observe that frequency and the ATTP/TR measures have a positive correlation.

From Equation (2), we can see TTP

This can be explained by the overall delays: weekdays have more delays (), therefore, compared with ,

* 1. Page 23, row 50: I think it is fair to assume zero delays for demonstrative purposes, however, it would be good to also highlight that it is a highly speculative assumption that there would be no delays on a BRT line. The only conclusion that can be drawn from this simulation is that improving punctuality even on one route will reduce ATTP.

**Response**: Yes, this is a good point. We added some explanation in 4.3, paragraph 2 to highlight the assumption is hypothetical. We also revised the conclusion from “DBL is an effective solution for reducing ATTP” to “improving punctuality even on one route will reduce ATTP”.

* 1. Page 17, row 45: unclear sentence: “To investigate the spatial pattern of transfer risk, the first thing is spatial aggregation, since trip patterns (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns.”

**Response**: Changed to: “To investigate the spatial pattern of transfer risk, the first thing is to aggregate trips based on their stops, since trip patterns (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns.”

* 1. Page 19, row 47: unclear sentence: “However, for the APC-GTFS dataset, we observe ATTP on Sundays is second lowest compared to Fridays, which is the lowest for original GTFS dataset.”

**Response**: Changed to: “However, we observe Sundays have the lowest ATTP for the APC-GTFS dataset while Saturdays have the lowest ATTP for the original GTFS dataset.”

1. Reviewer 2

The paper presents two measures of transport network performance. It is clearly written, mostly easy to follow and quite technical in nature. The editor may want to take a view to which extent the paper fits the remit of this special issue and the Urban Studies audience more generally. In its current form, the paper seems to be more suited to a transportation-focused journal.

Some suggestions.

* 1. The authors could engage more thoroughly with the themes highlighted in the call of the SI. I imagine that an Urban Studies readership would be interested in a discussion of how big data can offer new understanding of urban transport systems. Of course, there are already numerous reviews on this question, but a more focused discussion in view of the SI and the particular specialism of the authors might add to the literature. This would also help readers appreciate the specific contribution of this paper.

**Response:**

* 1. For the benefit of an Urban Studies audience, I would suggest that the authors dedicate more space to a fuller discussion of the specialist literature, and highlight the different objectives of those studies they cite. What are the pros and cons of existing measures of evaluating transfer effectiveness? In which ways are the measures proposed by authors superior to existing ones?

**Response**: Thanks for pointing this out. We adjusted the literature review so that it fits better with the paper topic and the SI theme. We start by

TR and TTP are the first measures that provide attainable solution to quantify the real-time performance of public transit transfers.

* 1. How does the commonly made distinction between ‘deliberate data’ and ‘byproduct data’ apply to the datasets used by the authors? I certainly agree that smartcard data are ‘byproduct data’, but datasets such as APC are deliberately collected for the purpose of passenger counts. Similarly, GTFS data are more than just byproduct; they are purposively structured, standardised and documented. The authors should clarify the ways in which their work relates to the SI’s theme of ‘Big Data in the City’, and depending on their focus, offer a fuller discussion of ‘byproduct data’ potentially extending it to issues of bias and computational cost. Alternatively, perhaps an emphasis of ‘small data’ versus ‘Big data’ may be more appropriate for this particular paper.

**Response**: This is a very nice proposal. To avoid similar confusions about how to classify “deliberate data” and “byproduct data” and keep the theme of the paper connected to the SI theme, we decided to change the two categories to “small data” versus “big data” as the reviewer 2 proposed. In this way, we can clearly define GTFS and APC as typical examples of “big data in the city” and ATTP and TR as new measures to manipulate big data.

Correspondingly, we know that

* 1. There is a tension between the technical measurement and passengers’ experience, which would warrant further discussion. To which extent are the components of transfer time penalties actually experienced by passengers? On high frequency services, ATP may actually not matter that much. And, if I understand correctly, transfer risk (which would be better called ‘risk of transfer loss’ or something) will increase for high frequency services. If a ‘receiving’ service runs every 5 minutes, the risk is inherently higher than if a service runs every 20 minutes, and yet the impact on the passengers’ experience is higher in the latter case. Of course, this dimension is picked up by the other measures, but I do wonder how meaningful ‘risk of transfer loss’ is without considering frequency.

Response: This is a good observation. The best and perhaps the only way to analyze passengers’ experience is to conduct surveys or interviews on the passengers per se. Since we did not do survey on passengers, we cannot have enough solid evidence to support a definite conclusion. This is also what we are conceptualizing in the conclusion part about the future research: more detailed and more abundant big data sources beyond real-time vehicle data. Although we do not have survey results, we can still discuss the questions by following aspects:

1. To which extent are the components of transfer time penalties actually experienced by passengers?

We will argue that the user will not experience the components at all, since the user will not see or perceive (the delay of actual receiving bus (the second bus in the transfer)) and (additional time penalty, the sum of receiving bus headways). For example, the nature of ATP determined that it sometimes cannot be physically perceived by the users at all: ATP can be negative. Instead, the user may experience TTP, since it is the “net time loss” for the users’ transfer trip: user can simply calculate or perceive it by comparing with the current time when the receiving bus moves () and the time that the trip planner promised (). However, since we do not have the data on user experience, we cannot conclude the exact threshold of the passengers’ perceived transfer penalty.

1. On high frequency services, ATP may actually not matter that much.

Yes, but this is partially right. On high frequency services, ATP, which is the sum of missed headways of receiving buses, is assumed to have lower value if the frequency increases. However, if the delay of generating bus (the first bus in the transfer) does not change, the ATP will not change even if the headway is smaller, since it will simply miss more bus. But smaller headway will generally make it easier for transfers to synchronize, thus incur smaller time penalty.

1. Transfer risk (which would be better called ‘risk of transfer loss’ or something) will increase for high frequency services. If a ‘receiving’ service runs every 5 minutes, the risk is inherently higher than if a service runs every 20 minutes, and yet the impact on the passengers’ experience is higher in the latter case. Of course, this dimension is picked up by the other measures, but I do wonder how meaningful ‘risk of transfer loss’ is without considering frequency.

We added the correlation analyses between TR/ATTP and frequency shown in Figure 5 (current draft). We can see that for both datasets, ATTP have significant negative correlation with frequency while TR does not have significant correlation with frequency. Therefore, we can conclude from the data that frequency/headway does not have a significant impact on transfer risk. This is also intuitive: if the frequency of both generating and receiving buses simultaneously increases, the synchronization result is likely to stay the same. Since frequency does not have significant impact on TR, we can conclude TR can be a good measure to gauge the desynchronization.

Meanwhile, as reviewer 2 commented, TR is only one dimension, which focuses on the performance of system side; for users’ experience, ATTP is more appropriate since all that user cares is time penalty instead of a percentage. The future research can measure the perceived ATTP and compare it with actual ATTP.

* 1. Similarly, I couldn’t follow why ‘ugly, pre-emptive’ transfers would be experienced negatively. ‘Pre-emptive’ suggests that passengers pre-emptively transfer to avoid risk elsewhere; but in this particular context, they simply get on an earlier vehicle without necessarily being aware of this. Except for potential crowding, they may not be any different from ‘good’ transfers.

**Response**: Yes, it is a very good point. The reason why we called it “ugly” is not because preemptive transfers will be experienced negatively and we added corresponding clarification in the paper about this in section 3.3 (current draft), subsection “Transfers: the good, the bad, and the ugly”, paragraph “The ugly”.

Just like the comment says, compared to the schedule, a preemptive transfer may result in positive, or zero, or negative TTP. This is exactly the reason why we define transfer risk as the proportion of missed transfers alone, instead of missed transfer and preemptive transfers together. We acknowledged that a preemptive transfer can be as “good” as or even better than a normal transfer. However, we have to point out that preemptive transfers are different from the normal transfers. With the metaphor of “ugly”, we intended to describe its random and chaotic nature: although it may be as good, but it is not intended. Even if it may achieve a random better performance, it is not sustainable.

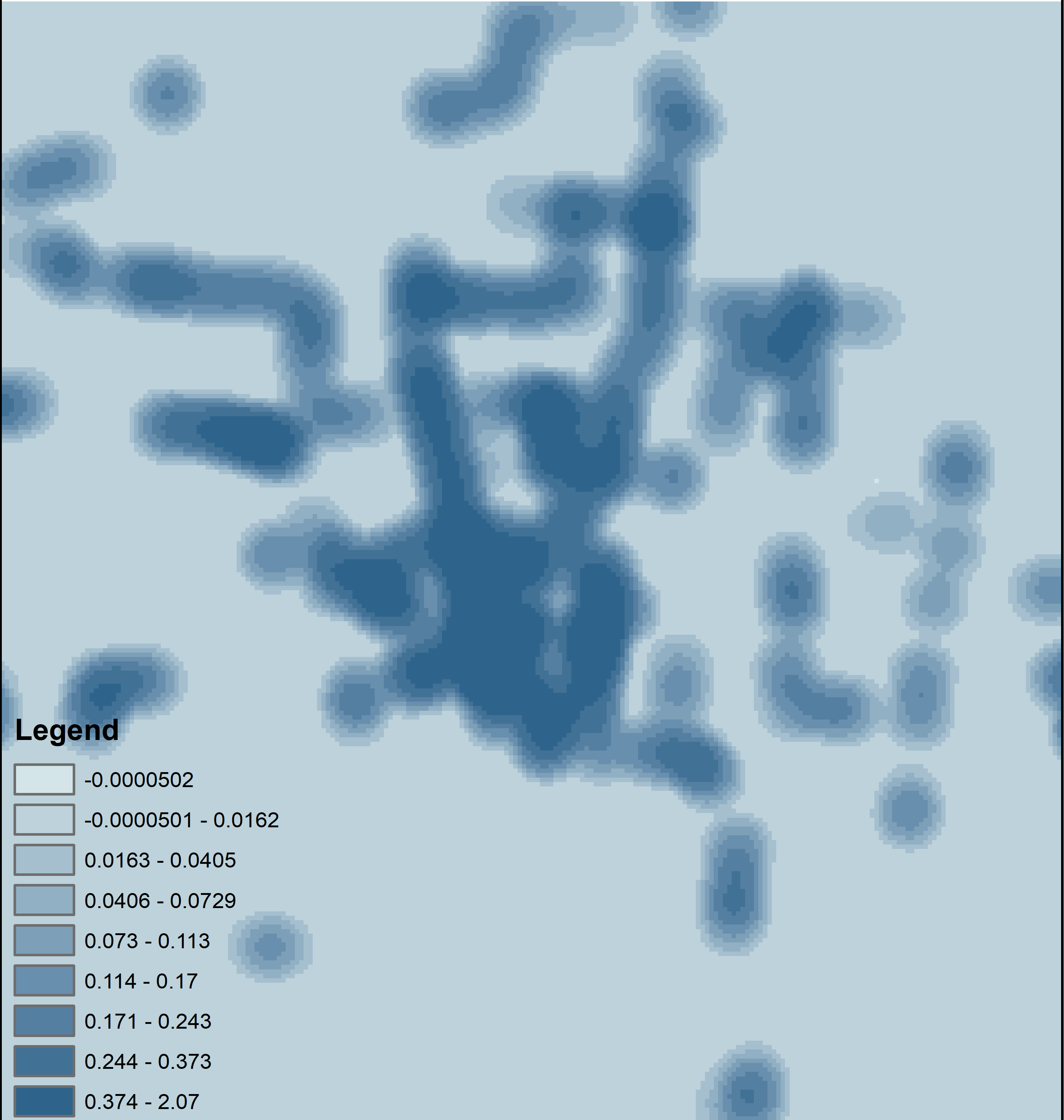
* 1. P.16, last sentence, “Only those combinations with … unique routes are selected”. I couldn’t follow what unique routes means in this context. Does this mean that origin-destination pairs with multiple transfer possibilities have been excluded? More explanation would be helpful.

**Response**: We added some clarification in the text. The idea is that: for different generating stop (the stop the user gets off from the first bus) - receiving stop (the stop the user take the second bus) pairs (short for *stops pair*), if the generating stop is the same and the transfer routes pair is also the same, there is really no point for them to coexist. For example, we have *stops pairs* A->A and A->B. Stop A has bus route No. 1 and No. 2 and Stop B also has bus route No. 2, there is barely a possibility that a user will walk from A to B to catch a No.2 Bus, since she/he can always directly transfer at stop A. By doing this, we can reduce some redundancy and remove some transfers that will never be used by any users.

* 1. Figure 3 is hard to read. The authors may want to explore alternative types of visualisations, e.g. heat maps or contour maps.

**Response**: This is a very good point. We adjusted the figure in following several aspects:

1. We changed the extent of the map to the center of the city of Columbus to keep consistent with the geographic pattern described in 4.1. Spatial patterns. The center of city is the area with most ridership in the Central Ohio Transit Authority (COTA) bus system. By doing this, we can not only cover most interesting patterns, but also the symbols are much clearer to see.
2. We did not use heat map or contour map and there are several reasons. First, the two visualization types requires interpolation, which means the visualized symbols does not necessarily represent the actual values. Therefore, it will cause some information loss and homogenize the heterogeneous patterns. We produced an example shown in Picture 2 in quantile classification and we can see heat map cannot show the heterogeneous and microscope patterns between each stops. This is also because the unique feature of bus stops: each stop is extremely close to each other while their difference can be very large.



Picture 2: A heat map version of Figure 2 (current draft).

1. Instead, we changed the visualization type from graduated colors (circles with same size and different colors) to graduated symbols (circles with same color and different sizes). The figure also replicates better in grayscale in terms of the editor’s concern, as shown in Picture 1.

Combined with closer zoom of the map extent, we believe it is easier for readers to distinguish the major patterns in Figure 2 (current draft) on the High street (marked by red circles) and the downtown area (marked by green rectangle).

* 1. P.21 first paragraph, I wonder if statistical tests should be added to assess differences on days with rain or a football match or in the DBL scenario. By the look of the values, I don’t see how they indicate ‘considerable impact’ as the authors conclude. The authors could also focus on connections with larger differences or on those with higher passenger counts.

**Response**: Thanks for pointing this out, we apologize for our neglect. We added corresponding statistical tests to each sections and the results are:

* Football. We did the Kolmogorov–Smirnov (KS) test to see whether the two distributions are the same. For ATTP (average total time penalty), the p-value for original GTFS is 0.036 and the p-value for APC-GTFS is 0.53; for TR (transfer risk), the p-value for original GTFS is smaller than and for APC-GTFS is 0.65. Therefore, for original GTFS, the difference between football game days and ordinary days is significant and for APC-GTFS it is not.
* Precipitation. The results of KS test are all not significant. For ATTP, the p-value for original GTFS is 0.46 and the p-value for APC-GTFS is 0.40; for TR, the p-value for original GTFS is 0.27 and for APC-GTFS is 0.88.

Even if the football days are significantly different from ordinary days for original GTFS data, the results show that the two results in the last draft are not statistically significant under most scenarios. Therefore, as reviewer 2 pointed out, we also would like to put more attention to geographic and temporal patterns; the causality between the introduced measures and other social and environmental factors is not the major research question of this paper. Meanwhile, also because of the strict words limit, we decided to delete it to save space for other more important parts.

* Dedicated bus lane (DBL). The KS test shows very strong significant results for the difference between two scenarios. For ATTP, the p-value for original GTFS is 0.006 and the p-value for APC-GTFS is 0.005. We also added the analysis results to section 4.3 (current draft), second paragraph.
  1. In view again of the theme of the SI, I would expect authors to re-engage with the theme of Big Data in the conclusions. This may include a discussion of what new insights we may gain with regard to urban dwellers’ experience of their cities.

**Response**: Will do.

1. To avoid confusion, in this response, we will use brackets + last/current draft to indicate whether the reference is from the last draft or the current draft. [↑](#footnote-ref-1)