Dear Editor, Professor Yuan,

Many thanks for your help to organize the review process. We appreciate your efforts for collecting and summarizing the feedback from the three reviewers. These comments, critiques and suggestions are very helpful to revise the previous submission. In this resubmission, we went through the suggested content and made substantial changes accordingly. Please see the detailed response we have provided in this letter (in color). We hope this revision has addressed most of, if not all, the raised issues. Thanks again.

To all the reviewers, we are grateful for your detailed feedback for improving the manuscript. In particular, we had added related work to justify research questions we were addressing, and a discussion section to discuss several aspects of the results and findings in this paper, the implications, and the current limitations. In particular, we added a gravity model to discuss the effects of geographical distance in our approach in linking human mobility research with the delineation of urban boundaries based on Twitter user spatial interactions.

Best Regards

Reviewers' Comments to Author:  
Reviewer: 1  
  
<b>Comments to the Author</b>  
This paper introduces analysis from Twitter with the goal of inferring non-administrative urban boundaries based on Twitter user’s spatial interaction. The delineation of urban boundaries based on analyzing spatial interactions and the proposed methodological approach, are addressing a relevant subject worthwhile being published.

Dear reviewer,

Many thanks for your very detailed comments and suggestions. We tried to incorporate your comments in the revised manuscript. In particular, regarding the comments concerning the usage of geo-located Twitter data in this study, we had carefully discussed the limitation and suitability in this revised version. Also, please our respective response below. Thanks again.

Please find my respective comments below:  
  
[General]  
Generally I am missing research questions which would significantly help to strengthen the results of the paper as a whole. Refer to other studies: what is the same or/and what is different to the state of the art studies dealing with human mobility analysis and the inference of underlying urban structures? What are recommendations for future research? What new analyses and study set-ups could contribute to the same or new questions? These questions and the study limitations need to be later addressed within the discussion section of the paper (currently the discussion section e.g. regarding study limitations etc. is rather short).

Thanks for the comments. In this revised version, we had added content to clarify our research questions, the implications of the results and findings, as well as discussion regarding the limitation of the current work. Please see the revised version and we are looking forward to hearing your thoughts. Below, we provided our response to communicate with you, some of which were incorporated in the revised manuscript and some were not.

Please discuss and consider the following issues and their bias on the applied methods and the conducted case study:  
  
•       Location uncertainty  
- Inaccuracy of GPS receivers (intrinsic effects: mobile device characteristics, extrinsic effects: built environment & GPS dilution of precision). In particular since you conduct your study across London (Figure 7). How is your result influenced by the location inaccuracy because of the "banding and splotching effect" e.g. in London Twitter data? ([https://www.mapbox.com/blog/twitter-map-every-tweet/](https://webmail.illinois.edu/owa/redir.aspx?C=kVoktyTWktS39UkDFN4XfZRmk4cMPNPAq-8Bd3RfM4Wwu5MfE7vTCA..&URL=https%3a%2f%2fwww.mapbox.com%2fblog%2ftwitter-map-every-tweet%2f))

The reviewer pointed to a very critical aspect of the geo-located Twitter data in terms of the accuracy of the recorded information embedded in the geo-located Tweets. First of all, because

Regarding the “banding and splotching” effect, we went back to the dataset and did not encounter such an effect in our dataset, we were not sure how it happened in the case from the provided link.

- Privacy: users can individually choose to add a precise geographic location to a tweet or just a general location (city or neighborhood). -> [https://support.twitter.com/articles/78525#](https://webmail.illinois.edu/owa/redir.aspx?C=yMtf6-eR2N0aI4Bo7WIpTOJo5KVCsyE6bhM08wEMgHawu5MfE7vTCA..&URL=https%3a%2f%2fsupport.twitter.com%2farticles%2f78525%23)  
Both appear as a proper WGS 84 coordinate within a tweets location. How do you deal with this? Does this bias your results?

The reviewer is

thanks for your very detailed comments. We found that

- Sampling Biases  
Heterogeneous spatial and spatiotemporal distribution of tweets (sparseness). Within your study you are more or less relying on single heavy Twitter user’ which generate multiple geotagged

tweets across the UK within the selected timeframe.

Many thanks for your very detailed comments. We found that

- Mismatch between overall population and sampling frame: exclusion or under/over-representation of certain population groups. This factor heavily influence your study results since you delineate geographical regions with the assumption of a more or less heterogeneous population distribution (and demographics), which is especially in the case of Twitter not true. How do varying geo-demographics of Twitter user influence your spatial interaction model?

Many thanks for your very detailed comments. We found that

- Generation of unrepresentative subsets and different sample sizes from the whole amount of tweets depending on the analysis (only extracting georeferenced tweets means 99% of non-geotagged tweets are neglected , since only "about 1% of publically available Tweets are geotagged (Longley et al., 2014)). Spatial interactions are inferred based on geo-tagged tweets. What about users having spatial interactions but generating geotagged and non-geotagged tweets?

Thanks for the comment. Regarding the 1% Twitter data policy, we would like to clarify that it does

What can we learn from delineating boundaries that makes the method valuable and worth reproducing? The authors do not justify adequately what can be gained from such an analysis, since the study is in line and more or less confirms results from existing studies. How can we benefit from this analysis, when the inference of collective mobility patterns based on spatial interactions from chapter (3.1) and the inferred tail distributions have already been confirmed by other studies with much more reliable data e.g. using mobile phone records:  
  
Zheng, Z. (2015): Two-regime Pattern in Human Mobility: Evidence from GPS Taxi Trajectory Data.[http://onlinelibrary.wiley.com/doi/10.1111/gean.12087/pdf](https://webmail.illinois.edu/owa/redir.aspx?C=lSjcf018IS8LGSfgOx_SCgGWbn70wE3Gc_CLKvbxLD-wu5MfE7vTCA..&URL=http%3a%2f%2fonlinelibrary.wiley.com%2fdoi%2f10.1111%2fgean.12087%2fpdf)  
  
Sobolevsky, S. (2015). Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries.[http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014248](https://webmail.illinois.edu/owa/redir.aspx?C=lRjQteoUW9iLCRPiTJWZlWpSdLLLVkgkqr50yUClcOSwu5MfE7vTCA..&URL=http%3a%2f%2fjournals.plos.org%2fplosone%2farticle%3fid%3d10.1371%2fjournal.pone.0014248)  
  
Ratti, C. (2010): Redrawing the Map of Great Britain from a Network of Human Interactions.[http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081707](https://webmail.illinois.edu/owa/redir.aspx?C=XWPNobEiFT66GAaDCWabtRBjF8kf3FqkWGAWQVoAixywu5MfE7vTCA..&URL=http%3a%2f%2fjournals.plos.org%2fplosone%2farticle%3fid%3d10.1371%2fjournal.pone.0081707)  
  
Thus, a correlation of study results from GPS trajectory datasets, mobile phone data records and social media feeds to find similar/dissimilar overlapping patterns would make a real significant contribution. Why should we use highly uncertain social media feeds if we already have observed these patterns from other datasets? How can social media add or further contribute new geographic information and knowledge we would not be able to extract e.g. by solely analyzing mobile phone data?

First of all, thanks for the

[Abstract]  
“[…] many unexpected and interesting boundaries were identified”. Can you please specify these? I would suggest to briefly highlight and point out the most important empirical results from the case study.  
  
P.1 L28: “we delineated the urban boundaries of Great Britain using a spatial Twitter user interaction network“.  
This statement is a arguable: A user interaction network cannot directly reflect the human interaction with the physical space. It reflects the interaction between users, and perhaps this is can be a proxy for the inference of mobility between real world geographic regions.  
  
[1]  
P.2 L6: “Urban boundaries that respect the human interaction space can be very import to city  
planning, traffic management and resource allocation.”  
This has been mentioned several times by the authors. However, it is not further explained why and the importance of these associations. When conducting empirical research, this cannot be purely data-driven. I am missing a substantial underlying theoretical body that further elaborates these associations.   
  
P. 2 “than 69 million Twitter messages from June 1st to December 31st”  
Can you give reasons why this time period in particular? (Please consider/discuss here seasonal trends and effects of human activities and changing spatial interaction patterns which might be neglected)

This is very good point. There is no specific reason why we only collected the dataset during this time span. In fact, we were interested to collect this the Great Britain to carry out comparative studies in other countries. To collect data from June to December, we thought half of a year collection would provide more stable measurements.

P.2 L32 "In addition, Twitter data are not as sensitive to user privacy issues and do not exhibit spatial granularity that is limited to the postal code level Thiemann et al. “  
Could the authors further explain this statement? When analyzing geo-tagged Twitter data on an individual user level, private user locations and trajectories are potentially disclosed. This raises even more privacy issues. Without using individual geographic locations from Twitter user, this study and the estimation of spatial interactions would not be feasible at all.

This is very good point. First of all, we had mentioned that the privacy issues in using other mobility dataset, such as mobile phone call records,  
  
P.2 L50“Our study provides a first step in connecting human mobility research with the definition of non-administrative anthropographic urban boundaries based on Twitter user spatial interaction.”

P.2 L33 „It provides a new understanding of the interactions between human activity and urban structures.”   
What are these new insights? The mentioned state of the art provides a good methodological overview. However, the authors claim to provide “novel analysis results” regarding the application, but a lot of key literature with similar and related research questions regarding the connection of human mobility and underlying urban structures is missing:  
  
Gao, S. (2014): Detecting Origin-Destination Mobility Flows From Geotagged Tweets in Greater Los Angeles Area.   
  
Huang et al. (2016): Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us?[http://www.tandfonline.com/doi/full/10.1080/13658816.2016.1145225](https://webmail.illinois.edu/owa/redir.aspx?C=GO08zA9H1mEVYAZ-B9CgFV-HyARE4fnxgJpE-p2sJLywu5MfE7vTCA..&URL=http%3a%2f%2fwww.tandfonline.com%2fdoi%2ffull%2f10.1080%2f13658816.2016.1145225)  
  
Liu et al. (2014): Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data.[http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0086026](https://webmail.illinois.edu/owa/redir.aspx?C=ur5FEOsTjtw_m9Geha_d3WleFkPMpdNrXhT_hy623jewu5MfE7vTCA..&URL=http%3a%2f%2fjournals.plos.org%2fplosone%2farticle%3fid%3d10.1371%2fjournal.pone.0086026)  
  
Lovelace, R. (2014): Geotagged tweets to inform a spatial interaction model: a case study of museums.[http://arxiv.org/abs/1403.5118](https://webmail.illinois.edu/owa/redir.aspx?C=7W2d0veBmRRljlhyrZ7NSlUJgs12mN5M7_zpj2EhciGwu5MfE7vTCA..&URL=http%3a%2f%2farxiv.org%2fabs%2f1403.5118)  
  
Luo, Feixiong (2015): Explore Spatiotemporal and Demographic Characteristics of Human Mobility via Twitter: A Case Study of Chicago. [http://arxiv.org/pdf/1508.00188v1.pdf](https://webmail.illinois.edu/owa/redir.aspx?C=iJsSe6KIhzVCWWEBHmYCExpgY-T0QT__cBBKXOaz39mwu5MfE7vTCA..&URL=http%3a%2f%2farxiv.org%2fpdf%2f1508.00188v1.pdf)  
  
Steiger, E. et al. (2015): Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data.[http://dx.doi.org/10.1016/j.compenvurbsys.2015.09.007](https://webmail.illinois.edu/owa/redir.aspx?C=SpnIRObU6c3VDj7A9lgxBxtBNzavX2992JshdiGWqWSwu5MfE7vTCA..&URL=http%3a%2f%2fdx.doi.org%2f10.1016%2fj.compenvurbsys.2015.09.007)  
  
Thanks for the suggesting the related literature.

P. 5 L48 “non-human users” -> this is more or less an assumption, since the application of heuristics does not result in the absolute distinction between human and non-human behavior. Also note the fact that user can manually “geotag” their tweets (as mentioned above).

The reviewer is right that this heuristic does not result in absolute distinction between human and non-human behavior. While recognizing this limitation, this heuristic is derived from the existing literature

P.9 “One clear observation […] is that most of the geographic divisions are centered around big urban cores with relatively high populations”. This statement provides only little new insights. As mentioned earlier, this might be due to population and sampling bias.

Thanks for the comment. The reviewer is right about this statement in terms of providing new insights. However, this statement was not intended to be a major finding in this paper, it was an observation that   
  
[Methods]  
Geographic scale effects (MAUP) -> You address this within your paper when referring to other studies. The authors point out that administrative boundaries are a classic top-down approach, only merely reflecting real world geographic regions. I completely agree. However, when you “[…] chose 10 km as the cell size to quantify the spatial coverage of the majority of Twitter users in Great Britain” P.6 L3., and aggregate point based observations (tweets) into an artificially created construct of fishnet cells, this does also not reflect real world geographic boundaries and the presented analysis results are also prone to various geographical scale effects. Spatial complexities within tweets are unknown, but parameter estimations like distance measures (in your case 10km thresholds) are nevertheless performed beforehand. Can you further elaborate the strength of your fishnet model with respect to overcome MAUP in comparison to existing studies?  
  
P.5 L26 “[…] corresponding weights are assigned by the accumulated volume of Twitter”  
Within your created geographic boundaries you aggregate geo-tagged tweets. This requires a normalization to filter out (averaging out) “heavy Twitterer” within these polygon areas due to varying tweet frequencies in order to avoid an over- or underrepresentation of particular Twitter users over all Twitter user, biasing the spatial interaction model. Otherwise your results are not valid, only reflecting geographic boundaries of very active Twitter user and biased by the general population distribution.

Thanks for the comments.

P.5 Equation 1: The radius of gyration has two problems at this point :  
a) Outlier Locations move the centroid disproportionate.  
b) Outliers affect relatively strong the radius itself (and tweets are generally “noisy” with a high amount of outliers)  
Would it not be better instead of the centroid to use the geometric median and then to construct an ellipse which contains 95 % of the points around the median?

Thanks for the comments. However, in this

The definition of radius of gyration considers the

In this study,   
  
[Figures & Tables]  
The images and captions are good and self-explanatory.  
Figure 2 & 5 please add the geographic scale  
Figure 4 What does the green area symbolize? Please label!  
Figure 5 is barely readable/interpretable. Please provide a better resolution of the image and maybe reduce colors. The display of major cities in black does not lead to more insights and confuses the reader, since its almost impossible to distinguish between raster results delineated from tweets and NUTS.  
  
[Minor]  
The authors ubiquitously use the term “urban areas”, thus focusing on urban areas. However the study is conducted across “all areas” (e.g. also rural communities). Therefore please specify/adapt the term.   
  
P. 1 L57 “schliephake” in capitals  
  
P.2 L6: “import” -> important  
  
P.2 L12: Lancichinetti and Fortunato in brackets  
  
P.6 L41: This only describes one aspect of MAUP, MAUP also here includes issues with the positioning and the shape of the cells.

Reviewer: 2  
  
<b>Comments to the Author</b>  
This paper presented a method to redraw urban boundaries based upon human interaction with the physical space by using Twitter users’ mobility data. The idea for the paper is interesting.  However, there are several concerns and issues should be addressed first.   
  
1.      Identify and separate Twitter users. Right now, the authors use all twitter users that have posted tweets within the boundary of Great Britain. However, many users, especially these tweeting in the major cities, such as London, could be tourists instead of residents within a community. The spatial mobility of such users can not reflect people’s daily interactions with the urban space, and therefore should not be incorporated for delineating the urban boundaries.   
  
2.      A comparison and discussion should be performed when trying other fishnet cell sizes. For the national scale boundary delineation, while the authors choose 10 km as the cell size by only examining the probability distributions of the radius of gyration, it is highly possible that different fishnet cell sizes would yield totally different community boundaries. Similarly, the authors only tested a fishnet of 1 km cells for the city level boundary delineation.   
  
3.      Literature review. Relevant work on using crowd-sourced data to determine the spatial extent of places should be reviewed and compared. Some examples include:  
  
Vasardani, M., S. Winter, and K.-F. Richter. 2013. "Locating place names from place descriptions." International Journal of Geographical Information Science 27 (12):2509-32.  
Stefanidis, A., A. Cotnoir, A. Croitoru, A. Crooks, M. Rice, and J. Radzikowski. 2013. "Demarcating new boundaries: mapping virtual polycentric communities through social media content." Cartography and Geographic Information Science 40 (2):116-29.  
Hollenstein, L., and R. Purves. 2015. "Exploring place through user-generated content: Using Flickr tags to describe city cores." Journal of Spatial Information Science (1):21-48.  
  
4.      The conclusions are brief, does not provide a critical review and vision on using social media data for boundary delineation.  
  
Minor issues:  
  
1.      For the comparison purpose, the authors could add legends and labels for the derived boundaries in the Figure 6 and 7.   
2.      Most of citations are not in a correct form.   
  
  
Reviewer: 3  
  
<b>Comments to the Author</b>  
In this manuscript the authors attempt to draw urban boundaries based on a network derived from geotagged Twitter messages. Roughly speaking, Twitter users connect locations in this network based on traveling from point A to B. The aggregated network based on these interactions is then used to detect communities (or clusters of strongly connected nodes/locations). Mapped back onto geographic space, this yields regions in a similar fashion to more conventional methods of regionalization in spatial analysis.  
  
The manuscript is well-written, clearly argued and the methodological section is very solid. The authors rightfully steer away from the more common modularity optimization approaches to community detection (as they have, for example, a resolution problem) and use the Infomap framework instead. This is a worthwhile addition and the merits of different community detection approaches deserve more attention/discussion within the field. I would suggest a few amendments and clarifications to increase readability of the manuscript for the larger IJGIS audience:  
  
-       The authors discuss how urban regions (or grid cells) are connected when a Twitter users movement begins in one region and ends in another. What remains unclear is how one determines where a movement starts and ends. E.g. if I tweet from location A, then tweet from B, and finally tweet from C, do you draw a connection between A and C only, or also include the intermediate B? Furthermore, if I tweet from A on day 1 and from B on day 2, does that still constitute a connection? In other words, please clarify the exact operationalization of these connections.

-       The authors state they use a directed network. Have they considered the effect and meaning of direction in this Twitter network? What happens if I tweet from A, then from B, and I travel back to A but don’t tweet from there again? Does the direction have an effect on the community detection results? I suggest testing this by ignoring direction in the Infomap algorithm

-       The manuscript correctly discusses MAUP, the effect of grid cell size and the effect of the length of ‘displacement’. It also discusses the issues that arise with using modularity optimization approaches to community detection with regards to resolution. One of the key advantages of the Infomap framework is that allows the researcher to influence the resolution of the communities by setting the Markov time parameter. Greater Markov time should theoretically yield larger urban regions and vice versa. However, there’s no discussion of this aspect of the algorithm despite it being in line with the aforementioned discussion.

-       I can’t seem to find the actual description length L(M) for the community divisions shown in Fig 5. This might help in further understanding the effects of the different ‘displacement lengths’.

-       Double-check the reference software used. Citations are currently interjected oddly (e.g. without parenthesis), which hampers readability.