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 revised version of the manuscript and 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. Many thanks for your help.

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, suggestions, advices. We tried to incorporate your feedback 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, as well as its suitability in this revised version. Also, please our respective response below. Thanks again for your help.

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 discussions regarding the limitations 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 comment 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. Yes, in dense built urban environment, the GPS receiver will tend to have lower accuracy and it would affect the calculation of distance between two consecutive locations, which we mentioned in the manuscript.

On the other hand, the mentioned concern of inaccuracy of GPS receiver spreads across all the existing mobility datasets, such as taxi trip records, location based social media data from GPS enabled smartphones, and trajectory data collected by hand-held GPS devices. In this paper, we discussed the accuracy of mobility data from mobile phone records, census migration records and geo-located location based media data. Actually, the geo-located Twitter data is considered to be high-accuracy mobility data in comparison with mobile phone records by cell-tower triangulation (Jurdak et al., 2015).

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D. 2015. Understanding Human Mobility from Twitter. PLoS ONE 10, e0131469. doi:10.1371/journal.pone.0131469

Regarding the mentioned “banding and splotching” effect, we went back to the dataset and did not encounter such an effect in our dataset, which is visualized in Figure 1. We were not sure how it happened in the case mentioned in the provided link.

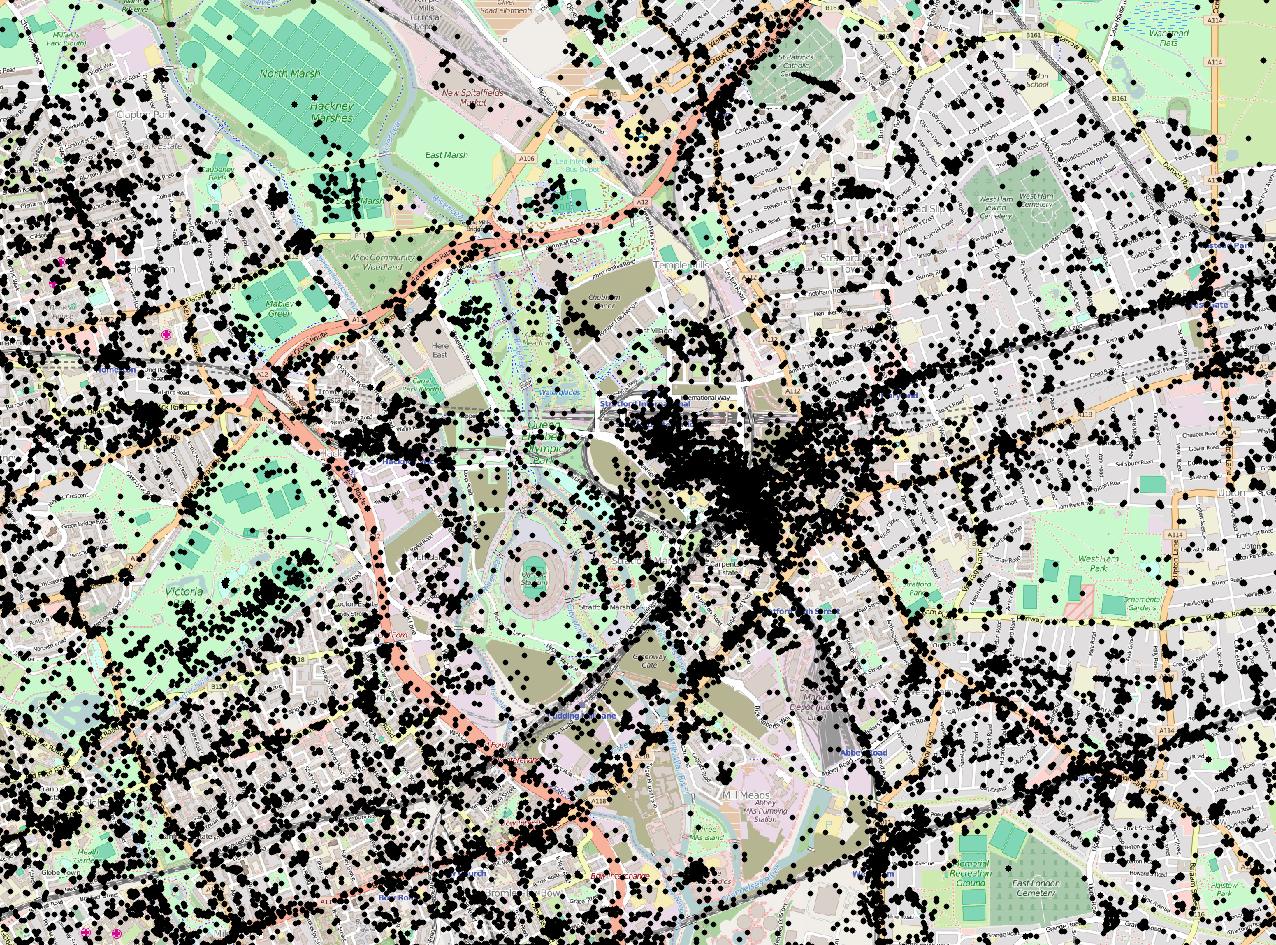


Figure 1: Visualization of the geo-located Tweets around at the Prime Meridian in London

- 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#  
Both appear as a proper WGS 84 coordinate within a tweets location. How do you deal with this? Does this bias your results?

Thanks for the comment. Yes, the mentioned situation could happen. We mentioned in our data processing part that in this study we first filtered the raw data and kept those tweets with the “geo” attribute set to true, which means it is derived from a GPS receiver, rather than from reverse geo-coding process, which was visualized in Figure 1 in the paper.

On this note, we had noticed that the capability for adding general location (city or neighborhood) started from April 2015, the release of the new Twitter mobile app. Before that, if the precise location is not available, the address is found and labelled as a polygon and the “geo” attribute would be false. And, yes, in those cases it would result in reduce the number of Tweets with precise geo-location and interfere the process to utilize Twitter user movements as connections of geographical spaces.

- 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.

Thanks for the comment. Regarding the concern of heterogeneous spatial and spatiotemporal distributions of tweets, we had incorporated the comment in the revised manuscript for discussions, as well as support illustrated in the supplement materials regarding the spatial coverage and temporal stability of the measurement of radius of gyration.

Yes, some Twitter users generated far more number of tweets than others, which were also evident from Figure 2. As we have discussed that such behaviors were observed in many natural systems, such as the distribution of emails people wrote. The point is that although there seems to exist vast heterogeneity individually, the interaction strength in the mobility network is measured collectively and we do not rely on single heavy Twitter user’s heavy tweeting behavior but rather collectively. In other words, we are studying the Twitter population in the Great Britain rather than sampled individuals. This is the case for using the probability distribution of the collective radius of gyrations of Twitter users. For example, some heavy tweeting users may generate multiple geo-tagging tweets, but their radius of gyration is small (say, they only tweet at home), therefore, their movements would not contribute in linking different geographical regions at all.

- 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?

Thanks for the excellent comment. Yes, mismatch between the overall population (in particular, the demographics) is one of the limitation in this current work, which we have taken your advice and carefully discussed in the revised manuscript and tuned down some of the statements made in the previous version.

Regarding the demographics in Twitter data, although current literature presents some preliminary studies (the methods used were arguable), it should be certainly incorporated in future studies. However, the focus of this paper is to present the methodology to use a mobility network to delineate the urban boundaries and how geographical distances in the mobility pattern affecting the outcome of the delineation.

- 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 means the Twitter streaming API would provide 1% of the whole Tweets (at that moment) available at twitter.com for downloading. As the selected area is around British Isle, and the Twitter population in the Great Britain only accounts for small percentage of the whole population around the world, our crawler did not receive any warning regarding exceeding the 1% limit, which means we have collected all the geo-located Tweets in the Great Britain. The refereed statement was not correct in (Longley et al., 2014).

On the other note, yes, not everyone posts geo-tagged Tweets, some may generate non-tagged tweets and geo-tagged tweets during the time span, which is the limitation we have discussed in the paper regarding use geo-located Twitter data for human mobility studies and it is also why we referred it as Twitter user displacements.

We would also like to mention that to continuously monitor people’s movements may only be possible through equipping GPS units at all time so there would be no gap, in which the subject’s movements are missing, such as taxi trip records. In this paper, we have discussed this aspect and referred to some of the recent work using geo-located Twitter data as valuable proxies for studying human mobility patterns. In short, the spatial interaction model will not be able to capture the in-between Twitter user movements that were missing.

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?

Thanks for the comments and directing to the related work. In terms of the rational of this study, its implications and its value worth reproducing, it is amended and revised in this revision.

Regarding inferencing the collective mobility patterns, the statistical analysis was based on the latest development in the literature, such as (Jurdak et al., 2015). The provided reference (Zheng, 2015) assumes that human mobility is power-law based and applied that theory in studying taxi data, which is somehow contradicted to the study from (Liang, 2012). The reference from (Sobolevsky, et al., 2010) and (Ratti, et al., 2010) were discussed and referred in this paper, where we pointed that the result of delineated boundaries were instances of the spatial proximity effect and the distance decay effects in human mobility patterns are linked and contributed to the geographical distance in affecting the interaction strength between different urban regions. Nevertheless, although the mobility studies are major part of this paper, our intention was not trying to find new insights in itself but seeking explanation for the urban boundary delineation based on physical human (Twitter user) movements.

The following content is for the sake of communication with the reviewer, which is not entirely embedded in the current paper. In terms of carrying out comparative human mobility studies relying on the existing findings from “reliable” mobility source, we discussed in the paper that (1) the accessibility issue, the mobile phone call records are almost impossible to get (2) from personal observation, even there are some existing studies that uses mobile phone data, I have never come across detailed mobile phone data in developed countries, such as USA or UK. (3) Each mobility dataset has its own limitation for human mobility study, even it is assumed that the demographics are heterogeneous, no existing studies were able to confirm as the privacy issues prevent such attempts. (4) There are also risks in not knowing the transitional human movements in-between to recorded locations. (5) As we have seen from the reference 1, applying the mobility pattern from mobile phone data is not necessarily consistent with the patterns found from taxi data.

In short, depending on specific angle and research question, geo-located Twitter data can be a valuable source to assist. Having said all these, such data current presents some limitations and we are encouraged that we are discussing these issues rather than using them directly without caution.

Liang, X., Zheng, X., Lv, W., Zhu, T. and Xu, K., 2012. The scaling of human mobility by taxis is exponential. Physica A: Statistical Mechanics and its Applications, 391(5), pp.2135-2144.

[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.

We have revised the abstract. As it is placed in the abstract, this statement was talking about the findings in delineation the urban boundaries in general. Delineation of urban boundaries is one thing and explaining the reason is another, that is why we kept it as a general description to draw reader’s attention.

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.

Thanks for the comment. We would like to clarify that it was a misusage of the term “spatial Twitter user interaction network”. In the revision, we referred it as a mobility network of Twitter user spatial interactions. The fishnet cells are the nodes of the network and edges between nodes are built form the Twitter user movements. Also, the spatial interaction refers a user’s re-allocation (i.e., user displacement) in space moving from one cell to another. It is not referred as a user connection network.  
  
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.

Thanks for your comments. In this revision, we have added related work and explained the motivation of this study. We hope the new version address your concern.   
  
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 particular reason why we only collected the dataset during this time span. In fact, we were interested in collecting this data in the Great Britain to carry out comparative studies with other countries (with data already in collection in our lab). To collect data from June to December, we thought half of a year collection would provide more stable measurements, for which we have discussed in this paper and also showed in the supplement material. Like mobile phone data, extend the data collection time span would not help in improving the representation of overall population (Zhao, et. al, 2016).

Zhao, Z., Shaw, S.L., Xu, Y., Lu, F., Chen, J. and Yin, L., 2016. Understanding the bias of call detail records in human mobility research. International Journal of Geographical Information Science, 30(9), pp.1738-1762.

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, is sensitive and almost inaccessible for most researchers, nor does it provide high-resolution location information, and due to privacy concerns, the analysis sometimes are aggregated in postal code level. On the other hand, while Twitter data is openly accessible, it does not mean users’ privacy should not be protected. In this study, when we processing the data, we do see the original latitude/longitude of user locations. However, we only measured the displacement (i.e., the distance between two consecutive locations) and radius of gyration (also distance), and finally the interactions are modelled on the fishnet level, in a sense, aggregated in the cell level. As we have discussed in the paper, this approach is different from the spatial clustering method, where a cluster of locations in a user’s trajectory can potentially lead to discover a user’s work or home location.   
  
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. We have taken some of the literature in the related work and discussed the difference between our approach and these existing work. We also want to highlight that our study is not solely on seeking mobility pattern or performing community detections, but to connect these two aspects by explaining the effects of distance decay effects, which are identified as a gap in the existing 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).

You are 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 as the current strategy (Jurdak, 2015; Hawelka, 2014).

In addition, as suggested by another reviewer, to remove the effects of tourists, we had added a new constraint that a user has be observed in the dataset more than 30 days between its first and last appearances. In other words, a user who has been identified to stay in the selected time period more than 30 days is considered as a resident.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo- located Twitter as proxy for global mobility patterns. Cartography and Geographic Information Science, 41, pp. 260-271. doi:10.1080/15230406.2014.890072

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 regarding providing new insights. However, this statement was not intended to be a major finding in this paper but an observation. In the revision, it was used as a setup to illustrate the effects of spatial proximity effects explained by a gravity model, which was not clearly stated in the previous version.  
  
[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?

Thanks you for the insightful comment. In the revised version, we also added an experiment by arbitrarily set the fishnet cell-size to 5 km (supplement material). We also discussed the reason for performing statistical analysis of the radius of gyration and chose 10 km as cell-size. We also demonstrated a partition scheme using the ward division of the Great Britain (which is the smallest administrative boundary in UK) and its problem related to its underlying population.

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. However, with respect to this particular point, we want to communicate with the reviewer that we have some different way of thinking. First, as we have discussed above, there do exist heavy tweeting users in the dataset but it is not abnormal from a twitter user population point of view. In particular, averaging out “heavy Twitter users” is problematic, because it is not clear what average means. Also, a heavy twitter user does not mean his/her geographical coverage is larger than others (in terms of connecting different urban regions) or his/her displacement is large than others. We do agree that geo-located Twitter data can be potentially over or under represent the overall population and therefore the results are limited. In this revision, we tried to discuss this issue. More importantly, we also focus on explaining why the interaction model would yield such results.

On the other hand, we studied the collective mobility patterns to differentiate different user groups, for example, people with spatial coverage within 30 meters, [50m 10km], and [10km, 100km] and select displacements at different ranges.

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. The suggested way to use standard deviation ellipse or weighted standard deviation ellipse (WSDE) may be a good choice for studying the trajectory pattern of moving objects with the consideration of directional changes, it may not be a good choice for studying collective mobility patterns in terms of deviation in coverage, i.e., radius.

Outlier is a relative term with respect to a defined model.

In definition, the radius of gyration considers the all the historical points in a user’s trajectory, in this sense, there are no outlier locations. Let’s say intuitively a person would have two major locations in his trajectory, i.e., home and work, a new location in shopping area or at the airport is should not be considered as an outlier, especially it indicates this user move from a previous location (one urban region) to the airport (another urban region).

As we have discussed above, we are handling the Twitter population with heterogeneity in term of recorded number of locations in their trajectory, the statistical significant level of 95% is not fair to be applied for the users.   
  
[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.

Thanks for your advices, please see the revised figures in the revision.

[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.

Thanks you for pointing those out. We have gone through the manuscript and revised them accordingly.

Regarding the MAUP with the positioning issues, in this study, we first used the national boundary of the Great Britain and then imposed the fishnet on top of it, additional cell-size (e.g., 5 km cell-size) was also used and discussed (please refer the discussion section and supplement material).

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.

Dear reviewer,

Many thanks for your efforts and help for providing the feedback. Regarding the concern of related work and discussion, we have added a related work section for comparisons with different approaches to delineating the spatial extent of urban boundaries. We also added a discussion session to analyze the results, comparisons using different settings and the implications for future studies. Please see our detailed response below. Thanks again for your help.

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.

Thanks for providing this advice. We did not consider such effects. To remove the effects of tourists, we had added a new constraint that requires a user has to be observed in the dataset more than 30 days between its first and last appearances. In other words, a user who has been identified to stay in the selected time period more than 30 days is considered as a resident.

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.

Thanks for the suggestion. In this revision, we have added discussion for using different fishnet-cell size with an experiment with cell-size as 5 km (supplement material). We discussed the reason of using 10 km at the national level and 1 km at the greater London regions.

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.

Thanks for providing the related work, we tried to discuss and compare those studies carefully in the related work section of the revision.  
  
4.      The conclusions are brief, does not provide a critical review and vision on using social media data for boundary delineation.

Thanks for pointing out this issue. We have tried to take your advice with a separate discussion section to address the review of this study. Please see the revised manuscript.  
  
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. 

Thanks for pointing out these issues. We have incorporated your advices and corrected them in the revision.

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:

Dear Reviewer,

Many thanks for your comments and suggestions. We have tried to incorporate your feedback to revise the manuscript accordingly. In particular, it helps to clarify the effect of using different setting of the network (i.e., directed or undirected network), how the network is constructed. Please see our detailed response below. Thanks again for your help, appreciated.

-       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.

Thanks for the comment. In this study, we first constructed a trajectory for each Twitter user in the dataset by appending all the recorded locations (with the same user\_ID) in the chronological order (based on the timestamps). Then, in the case moving from A to B and then C, we used the term “displacement” to represent such movement and the connection in the network is represented as A to B and B to C (A to C is not directly connected in this case, which will leave to InfoMap to figure out the connectivity.). Regarding the second case that the time interval between two consecutive recorded locations exceeds more than 1 day, we still treat it as a valid displacement and a connection is built. We have tried to clarify in the revised version.

-       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

Thanks for the advice. In this revised version, we had discussed the case of using un-directed graph to represent the network, however, it would not generate meaningful delineation (visualized in the supplement materials). This suggestion clarifies why a directed and weighted graph was used.

-       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.

Thanks for the comments and advices. Regarding the parameter of Greater Markov time in Infomap, we only used the default value set to 1. You are right that Greater Markov time should theoretically yield larger urban regions and vice versa. We intended to explore this aspect in the revised version, however, as we have discussed in the paper that the fishnet with different cell-size could lead to similar effects as mentioned in tuning the Markov time parameter. And for the revised version, one of our focuses is to explain the spatial proximity effects in forming the geographically cohesive urban areas. Therefore, in this study we did not explore this parameter.

-       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’.

Thanks for pointing to this aspect. We have added the L(M) accordingly in the revised figure and also in the discussion section regarding the relationship between the value and the delineation.

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

Thanks for the advice. Our apologies for not correctly using the citation style in Latex, which are now fixed throughout in this revision.