THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS A

Manifold cities: social variables of urban areas in the UK

Edmund Barter and Thilo Gross

Article citation details
Proc. R. Soc. A **474**: 20180615.
http://dx.doi.org/10.1098/rspa.2018.0615

Review timeline
Original submission: 7 September 2018
Revised submission: 15 November 2018
Final acceptance: 29 November 2018

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPA-2018-0615.R0 (Original submission)

Review form: Referee 1 (Danny Dorling)

Is the manuscript an original and important contribution to its field?
Yes

Is the paper of sufficient general interest?
Yes

Is the overall quality of the paper suitable?
No

Quality of the paper
A good paper worth publishing in Proceedings.

Can the paper be shortened without overall detriment to the main message?
No

Do you think some of the material would be more appropriate as an electronic appendix?
Yes

For papers with colour figures – is colour essential?
Yes

If there is supplementary material, is this adequate and clear?
Yes

Are there details of how to obtain materials and data, including any restrictions that may apply?
Yes

Do you have any ethical concerns with this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
You suggest very early on that cities have greater problems with traffic and pollution. Fewer people tend to be injured and killed by cars in cities as compared to the countryside. Pollution per person is also generally lower in cities as greater use can be made of public transport and there is less need to heat homes in an urban heat island.

Later you say that "Selecting a specific set of variables can introduce bias to a model". However, the census variables are very carefully selected and are themselves representative of bias. For instance, I have tried for many decades to have income added to the UK census, in the same as it is included in the USA census. Its omission effects how we see cities in the UK. We don't like to ask about income here – that itself is one of a myriad of biases that will influence what you find and that are largely out of your control but still very important.

You have identified Fettes College, a large private boarding school. Please could you check whether you also successfully identify prisons and similar establishments? One of the first Borstals was located near Bristol and it may still be used for similar purposes. Output areas with many people in them who are in communal establishments stand out in census data.

You mention Knowle West, which is also associated with the artist Tricky, but you do not mention him. Nor the (probably an urban myth) suggestion that he once met Samantha Cameron (when she was a student at UWE), wife of David, a former prime minister a rival for power of Tony Blair, a Fettes boy. Fettes, like Eton (David's school) has close connections to the military. The reason for mentioning all this is that, although your method may identify things for you to look at, how you and I would interpret those things would very much reflect our differing ways of seeing the world and the connections we are likely to make. Whether one sees Fettes as a good thing, or Bedminster as improving (or simply gentrifying and becoming unaffordable and unfriendly), depends on from where one is looking. I have not been to Bedminster for 20 years. When I was last there, people complained that there was so many babies their baby alarms kept on picking up the crying of other infants! But it was a very friendly place. Is it still so friendly?

The point of me writing the paragraph above is to illustrate the problem of subjectivity. The social sciences have been an attempt to deal with that problem.

What worries me is I cannot see that you have discovered anything that was not already known. In theory your method should make that possible but, of course, you would then have a problem convincing people of what you have found if it not match their predilections for explanation.

I very much enjoyed your paper, but I would like to know if places like Fettes, or Wills Hall, are similar to young offenders residential institutions as far as census data is concerned and if not, why not. What do you find that is shocking or surprising?

I hope the paper is published and that these comments are useful in improving it first.

All best wishes,

Danny Dorling

Review form: Referee 2 (Ernesto Estrada)

Is the manuscript an original and important contribution to its field?
Yes

Is the paper of sufficient general interest?
Yes

Is the overall quality of the paper suitable?
Yes

Quality of the paper
An excellent paper making an important contribution to the field: should be published.

Can the paper be shortened without overall detriment to the main message?
No

Do you think some of the material would be more appropriate as an electronic appendix?
No

For papers with colour figures – is colour essential?
Yes

If there is supplementary material, is this adequate and clear?
Not applicable

Are there details of how to obtain materials and data, including any restrictions that may apply?
Yes

Do you have any ethical concerns with this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
This is an excellent paper, which is very sounded and well written. In it the authors used an approach developed by Coifman in which diffusion maps are used to generate a geometry of

graphs. My only concern is that the manuscript does not mention other similar approaches that are also suitable for the type of analysis done in this work. In fact, it could be argued that they are more appropriate as diffusion maps, but it is a matter of discussion. My suggestion is to include the following (or some of the following) references:
M. Pereda, Estrada, E. Visualization and machine learning analysis of complex networks in hyperspherical space. Pattern Recognition 86 (2019) 320-331.
M. Akbarzadeh, Estrada, E., Communicability geometry captures traffic flows in cities. Nature Human Behaviour, 2, 2018, 645-652.
E. Estrada, N. Hatano, Communicability angle and the spatial efficiency of networks. SIAM Review, 58, 2016, 692-715 .
E. Estrada, M. G. Sanchez-Lirola, J. A. de la Pena, Hyperspherical Embedding of Graphs and Networks in Communicability Spaces. Discrete Applied Mathematics 176 2014, 53-77.
Estrada, E., The communicability distance in graphs. Linear Algebra and its Applications, 436 2012, 4317-4328.

Review form: Referee 3

Is the manuscript an original and important contribution to its field?
Yes

Is the paper of sufficient general interest?
Yes

Is the overall quality of the paper suitable?
Yes

Quality of the paper
A good paper worth publishing in Proceedings.

Can the paper be shortened without overall detriment to the main message?
Yes

Do you think some of the material would be more appropriate as an electronic appendix?
No

For papers with colour figures – is colour essential?
Yes

If there is supplementary material, is this adequate and clear?
Not applicable

Are there details of how to obtain materials and data, including any restrictions that may apply?
Yes

Do you have any ethical concerns with this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
This is an interesting paper looking at a novel methodology to find the most relevant variables describing a system, going beyond PCA. The method is applied to urban systems, and in particular to census data, taking Bristol as a case study.

My first comment, which is already discussed by the authors, is that the step of thresholding seems somehow arbitrary, and there's scope for future work to find a more rigorous methodology. In particular, the thresholding does not correspond to the values but to the number of links constructed between OAs. I would argue that for an initial approximation, considering only 10 neighbours gives good enough results, but a rigorous method in which the distribution of weights is considered to decide on an uneven set of neighbours could be explored. This is comment for future work.

With respect to the paper in its state for publication, Here's a list of questions, comments:
1) validation: the method seems to be able to pick up places that have related characteristics, such as the universities and student halls. Were the authors able to verify that ALL relevant student places are contained in the OAs picked up by the first eigenvector?
2) Difference between census and selected variables: there is indeed a whole problem of how the census differs between England and Wales, and Scotland. It would be insightful to say how the 231 variables were selected.
3) Given that there's no 1-1 map between OAs in Bristol and any other selected case study, more clarity is needed on how the values are assigned from Bristol to other cities following the distance in the similarity space, in particular, when the destination is a city whose variation is much larger than Bristol.
4) With respect to the meaning of the eigenvectors, this also needs clarification. Given the large set of variables, how were these probed in order to understand from "local knowledge" that the first eigenvector corresponded to student populations? The same comment holds for the second eigenvector. Is there a second process of correlating the results at OA level with the variables? Is there a geographic weighted regression carried on, or what sort of correlation? This needs to be explicit in the text.

Overall, I found the paper interesting and inspiring for future applications and generalisations. Although as it stands, it does not provide a rigorous answer to the variables that need to be considered in a specific urban context, since it seems that previous knowledge needs to intervene, it does give a first step in this direction, in particular since it goes beyond linear models. For this reason, I strongly recommend publication.

Decision letter (RSPA-2018-0615.R0)

08-Nov-2018

Dear Dr Barter,

On behalf of the Editor, I am pleased to inform you that your Manuscript RSPA-2018-0615 entitled "Manifold Cities: Social variables of urban areas in the UK" has been accepted for publication subject to minor revisions in Proceedings A. Please find the referees' comments below.

The reviewer(s) have recommended publication, but also suggest some minor revisions to your

manuscript. Therefore, I invite you to respond to the reviewer(s)' comments and revise your manuscript. Please note that we have a strict upper limit of 28 pages for each paper. Please endeavour to incorporate any revisions while keeping the paper within journal limits. Please note that page charges are made on all papers longer than 20 pages. If you cannot pay these charges you must reduce your paper to 20 pages before submitting your revision. Your paper has been ESTIMATED to be 12 pages. We cannot proceed with typesetting your paper without your agreement to meet page charges in full should the paper exceed 20 pages when typeset. If you have any questions, please do get in touch.

It is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let me know in advance of the due date.

To revise your manuscript, log into https://mc.manuscriptcentral.com/prsa and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you will be able to respond to the comments made by the referee(s) and upload a file "Response to Referees" in "Section 6 - File Upload". You can use this to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the referee(s).

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission process.

In addition to addressing all of the reviewers' and editor's comments, your revised manuscript MUST contain the following sections before the reference list (for any heading that does not apply to your work, please include a comment to this effect):

- Data accessibility
- Competing interests
- Authors' contributions
- Acknowledgements
- Funding statement
- Ethics statement

See https://royalsociety.org/journals/authors/author-guidelines/ for further details.

When uploading your revised files, please make sure that you include the following as we cannot proceed without these:

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tif, eps or print-quality pdf preferred). The format should be produced directly from original creation package, or original software format.

3) Electronic Supplementary Material (ESM): all supplementary materials accompanying an accepted article will be treated as in their final form. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details where possible (authors, article title, journal name). Supplementary files will be published alongside the paper on the journal website and posted on the online figshare repository (https://figshare.com). The heading and legend provided for each supplementary file during the submission process will be used to create the figshare page, so please ensure these are accurate and informative so that your files can be found in searches. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Alternatively you may upload a zip folder containing all source files for your manuscript as described above with a PDF as your "Main Document". This should be the full paper as it appears when compiled from the individual files supplied in the zip folder.

Article Funder

Please ensure you fill in the Article Funder question on page 2 to ensure the correct data is collected for FundRef (http://www.crossref.org/fundref/).

Media summary

Please ensure you include a short non-technical summary (up to 100 words) of the key findings/importance of your paper. This will be used for to promote your work and marketing purposes (e.g. press releases). The summary should be prepared using the following guidelines:

*Write simple English: this is intended for the general public. Please explain any essential technical terms in a short and simple manner.
*Describe (a) the study (b) its key findings and (c) its implications.
*State why this work is newsworthy, be concise and do not overstate (true 'breakthroughs' are a rarity).
*Ensure that you include valid contact details for the lead author (institutional address, email address, telephone number).

Cover images

We welcome submissions of images for possible use on the cover of Proceedings A. Images should be square in dimension and please ensure that you obtain all relevant copyright permissions before submitting the image to us. If you would like to submit an image for consideration please send your image to proceedingsa@royalsociety.org

Once again, thank you for submitting your manuscript to Proceedings A and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes

Alice Power
Publishing Editor
Proceedings A
proceedingsa@royalsociety.org

on behalf of
Dr Ernesto Estrada
Board Member
Proceedings A

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
You suggest very early on that cities have greater problems with traffic and pollution. Fewer people tend to be injured and killed by cars in cities as compared to the countryside. Pollution per person is also generally lower in cities as greater use can be made of public transport and there is less need to heat homes in an urban heat island.

Later you say that "Selecting a specific set of variables can introduce bias to a model". However, the census variables are very carefully selected and are themselves representative of bias. For instance, I have tried for many decades to have income added to the UK census, in the same as it is included in the USA census. Its omission effects how we see cities in the UK. We don't like to ask about income here – that itself is one of a myriad of biases that will influence what you find and that are largely out of your control but still very important.

You have identified Fettes College, a large private boarding school. Please could you check whether you also successfully identify prisons and similar establishments? One of the first Borstals was located near Bristol and it may still be used for similar purposes. Output areas with many people in them who are in communal establishments stand out in census data.

You mention Knowle West, which is also associated with the artist Tricky, but you do not mention him. Nor the (probably an urban myth) suggestion that he once met Samantha Cameron (when she was a student at UWE), wife of David, a former prime minister a rival for power of Tony Blair, a Fettes boy. Fettes, like Eton (David's school) has close connections to the military. The reason for mentioning all this is that, although your method may identify things for you to look at, how you and I would interpret those things would very much reflect our differing ways of seeing the world and the connections we are likely to make. Whether one sees Fettes as a good thing, or Bedminster as improving (or simply gentrifying and becoming unaffordable and unfriendly), depends on from where one is looking. I have not been to Bedminster for 20 years. When I was last there, people complained that there was so many babies their baby alarms kept on picking up the crying of other infants! But it was a very friendly place. Is it still so friendly?

The point of me writing the paragraph above is to illustrate the problem of subjectivity. The social sciences have been an attempt to deal with that problem.

What worries me is I cannot see that you have discovered anything that was not already known. In theory your method should make that possible but, of course, you would then have a problem convincing people of what you have found if it not match their predilections for explanation.

I very much enjoyed your paper, but I would like to know if places like Fettes, or Wills Hall, are similar to young offenders residential institutions as far as census data is concerned and if not, why not. What do you find that is shocking or surprising?

I hope the paper is published and that these comments are useful in improving it first.

All best wishes,

Danny Dorling

Referee: 2

Comments to the Author(s)
This is an excellent paper, which is very sounded and well written. In it the authors used an approach developed by Coifman in which diffusion maps are used to generate a geometry of graphs. My only concern is that the manuscript does not mention other similar approaches that are also suitable for the type of analysis done in this work. In fact, it could be argued that they are more appropriate as diffusion maps, but it is a matter of discussion. My suggestion is to include the following (or some of the following) references:
M. Pereda, Estrada, E. Visualization and machine learning analysis of complex networks in hyperspherical space. Pattern Recognition 86 (2019) 320-331.
M. Akbarzadeh, Estrada, E., Communicability geometry captures traffic flows in cities. Nature Human Behaviour, 2, 2018, 645-652.
E. Estrada, N. Hatano, Communicability angle and the spatial efficiency of networks. SIAM Review, 58, 2016, 692-715 .
E. Estrada, M. G. Sanchez-Lirola, J. A. de la Pena, Hyperspherical Embedding of Graphs and Networks in Communicability Spaces. Discrete Applied Mathematics 176 2014, 53-77.
Estrada, E., The communicability distance in graphs. Linear Algebra and its Applications, 436 2012, 4317-4328.

Referee: 3

Comments to the Author(s)
This is an interesting paper looking at a novel methodology to find the most relevant variables describing a system, going beyond PCA. The method is applied to urban systems, and in particular to census data, taking Bristol as a case study.

My first comment, which is already discussed by the authors, is that the step of thresholding seems somehow arbitrary, and there's scope for future work to find a more rigorous methodology. In particular, the thresholding does not correspond to the values but to the number of links constructed between OAs. I would argue that for an initial approximation, considering only 10 neighbours gives good enough results, but a rigorous method in which the distribution of weights is considered to decide on an uneven set of neighbours could be explored. This is comment for future work.

With respect to the paper in its state for publication, Here's a list of questions, comments:
1) validation: the method seems to be able to pick up places that have related characteristics, such as the universities and student halls. Were the authors able to verify that ALL relevant student places are contained in the OAs picked up by the first eigenvector?
2) Difference between census and selected variables: there is indeed a whole problem of how the census differs between England and Wales, and Scotland. It would be insightful to say how the 231 variables were selected.
3) Given that there's no 1-1 map between OAs in Bristol and any other selected case study, more clarity is needed on how the values are assigned from Bristol to other cities following the distance

in the similarity space, in particular, when the destination is a city whose variation is much larger than Bristol.
4) With respect to the meaning of the eigenvectors, this also needs clarification. Given the large set of variables, how were these probed in order to understand from “local knowledge” that the first eigenvector corresponded to student populations? The same comment holds for the second eigenvector. Is there a second process of correlating the results at OA level with the variables? Is there a geographic weighted regression carried on, or what sort of correlation? This needs to be explicit in the text.

Overall, I found the paper interesting and inspiring for future applications and generalisations. Although as it stands, it does not provide a rigorous answer to the variables that need to be considered in a specific urban context, since it seems that previous knowledge needs to intervene, it does give a first step in this direction, in particular since it goes beyond linear models. For this reason, I strongly recommend publication.

Board Member Comments to Author(s):

Although diffusion maps are useful tools for the embedding of graphs, other alternatives not mentioned in the paper are also important. For instance, communicability geometry has emerged as an important embedding to analyse cities, as shown in:
https://www.nature.com/articles/s41562-018-0407-3. Such approaches should be discussed in the paper.

Author's Response to Decision Letter for (RSPA-2018-0615.R0)

See Appendix A.

Decision letter (RSPA-2018-0615.R1)

29-Nov-2018

Dear Dr Barter

I am pleased to inform you that your manuscript entitled "Manifold Cities: Social variables of urban areas in the UK" has been accepted in its final form for publication in Proceedings A.

Our Production Office will be in contact with you in due course. You can expect to receive a proof of your article soon. Please contact the office to let us know if you are likely to be away from e-mail in the near future.

Open access
You are invited to opt for open access, our author pays publishing model. Payment of open access fees will enable your article to be made freely available via the Royal Society website as soon as it is ready for publication. For more information about open access please visit

http://royalsocietypublishing.org/site/authors/open_access.xhtml. The open access fee for this journal is £1700/$2380/€2040 per article. VAT will be charged where applicable.

Note that if you have opted for open access then payment will be required before the article is published – payment instructions will follow shortly.

If you wish to opt for open access then please inform the editorial office (proceedingsa@royalsociety.org) as soon as possible.

Your article has been estimated as being 13 pages long. Our Production Office will inform you of the exact length at the proof stage.

Proceedings A levies charges for articles which exceed 20 printed pages. (based upon approximately 540 words or 2 figures per page). Articles exceeding this limit will incur page charges of £150 per page or part page, plus VAT (where applicable).

We are keen to promote all published material in the journal. If you wish us to highlight the publication of your paper to appropriate colleagues, please send me by return email the names and email addresses of up to 5 people and we will ensure that they are notified once the paper goes online.

Under the terms of our licence to publish you may post the author generated postprint (ie. your accepted version not the final typeset version) of your manuscript at any time and this can be made freely available. Postprints can be deposited on a personal or institutional website, or a recognised server/repository. Please note however, that the reporting of postprints is subject to a media embargo, and that the status the manuscript should be made clear. Upon publication of the definitive version on the publisher's site, full details and a link should be added.

You can cite the article in advance of publication using its DOI. The DOI will take the form: 10.1098/rspa.XXXX.YYYY, where XXXX and YYYY are the last 8 digits of your manuscript number (eg. if your manuscript number is RSPA-2017-1234 the DOI would be 10.1098/rspa.2017.1234).

For tips on promoting your accepted paper see our blog post: https://blogs.royalsociety.org/publishing/promoting-your-latest-paper-and-tracking-your-results/

On behalf of the Editor of Proceedings A, we look forward to your continued contributions to the Journal.

Sincerely,

Raminder Shergill
proceedingsa@royalsociety.org

on behalf of
Dr Ernesto Estrada
Board Member
Proceedings A

Appendix A

We appreciate the reviewers comments, and have produced changes aligned with them which have improved the manuscript. Below are our detailed responses to each reviewer. The reviewers comments are reproduced in full and typeset in italics for clarity.

Reviewer 1

You suggest very early on that cities have greater problems with traffic and pollution. Fewer people tend to be injured and killed by cars in cities as compared to the countryside. Pollution per person is also generally lower in cities as greater use can be made of public transport and there is less need to heat homes in an urban heat island.

We agree. Our intention was not to draw a comparison between city and countryside life but rather to mention some aspects that people criticize in cities. While countryside living causes more pollution to the environment than city living the overall air quality tends to be a much more hotly debated topic in cities. We have rephrased this to make it clearer.

Later you say that "Selecting a specific set of variables can introduce bias to a model". However, the census variables are very carefully selected and are themselves representative of bias. For instance, I have tried for many decades to have income added to the UK census, in the same as it is included in the USA census. Its omission effects how we see cities in the UK. We don't like to ask about income here – that itself is one of a myriad of biases that will influence what you find and that are largely out of your control but still very important.

Yes this is an important point. We meant there is no additional bias in the analysis after the data has been gathered. Clearly the design of the census questions introduce some bias. However, we judge this bias to be much smaller than the one that is introduced by restricting the variables to a much smaller set and directly interpreting them in a predetermined way, which is the current approach taken by the ONS. To make it clearer that there is still residual bias from the questions we use the formulations "no bias, except for the bias already in the source data" and "no additional bias".

You have identified Fettes College, a large private boarding school. Please could you check whether you also successfully identify prisons and similar establishments? One of the first Borstals was located near Bristol and it may still be used for similar purposes. Output areas with many people in them who are in communal establishments stand out in census data.

We checked Horfield prison, which is the main prison in Bristol and houses some young offenders. It is not picked up in either the deprivation or the Univeristy eigenvector. We beleive that is due to the demographics of inmates to be sufficiently dissimilar from these groups to distinguish them. One dedicated YOI

(modern day Borstal) exists in the analyzed area. It also is not picked up by the deprivation or the Univeristy eigenvector. This is likely becuase it is small, 24 residents, and so makes only a small impact on the OA statistics. Further investigation found that YOI or prison populations don't stand out in any of the leading eigenvectors. Our interpretation is that these facilities are too rare to be picked out based on an analysis of just the Bristol area. However, we expect that they would be discovered when the method is applied to larger regions.

You mention Knowle West, which is also associated with the artist Tricky, but you do not mention him. Nor the (probably an urban myth) suggestion that he once met Samantha Cameron (when she was a student at UWE), wife of David, a former prime minister a rival for power of Tony Blair, a Fettes boy. Fettes, like Eton (David's school) has close connections to the military. The reason for mentioning all this is that, although your method may identify things for you to look at, how you and I would interpret those things would very much reflect our differing ways of seeing the world and the connections we are likely to make. Whether one sees Fettes as a good thing, or Bedminster as improving (or simply gentrifying and becoming unaffordable and unfriendly), depends on from where one is looking. I have not been to Bedminster for 20 years. When I was last there, people complained that there was so many babies their baby alarms kept on picking up the crying of other infants! But it was a very friendly place. Is it still so friendly?

The point of me writing the paragraph above is to illustrate the problem of subjectivity. The social sciences have been an attempt to deal with that problem.

Yes, this is of course a deep fundamental problem that we cannot solve. However, we believe that we have made some progress by showing that there are patterns in the city that provide explanatory variables for the census responses. Independently, of how one views university students, we show that University campuses induce a coherent response in the census in their proximity. This leads to a testable hypothesis that this coherent response is due to the student population and the pattern itself gives us insights in how this population spreads through the city. We have now emphasized these possibilities in the conclusions.

What worries me is I cannot see that you have discovered anything that was not already known. In theory your method should make that possible but, of course, you would then have a problem convincing people of what you have found if it not match their predilections for explanation.

Admittedly our analysis of just 2 eigenvectors of the diffusion map only scratched the surface, and our main intention was to bring a new methodology to bear on these questions. However, returning to the point of bias discussed above, our method recovers insights that where so far only known from more subjective analysis. While we still need human interpretation to attach the label of "deprivation" to one of our eigenvectors, the pattern that this eigenvector highlights

is discovered by the algorithm without any human pointing it in this direction. The method thus reveals the existence of an explanatory variable that is predictive of census responses.

I very much enjoyed your paper, but I would like to know if places like Fettes, or Wills Hall, are similar to young offenders residential institutions as far as census data is concerned and if not, why not. What do you find that is shocking or surprising?

I hope the paper is published and that these comments are useful in improving it first.

All best wishes,
Danny Dorling

We thank the reviewer very much for his very constructive and insightful comments, while we regret that we cannot answer the question on young offenders just yet, it is reassuring that the methodology is sensitive enough to distinguish them from university students. We agree that is actually an interesting question whether we can pick out this population on a larger scale and we will likely follow up this question in future works. We feel that the reviewers perspective has helped us very much by making us see our own paper in a slightly different light. We realize now that the points about bias and the novelty of discovering structures without looking specifically for poverty or students could have been much more clearly made. We have revised the conclusions accordingly.

Reviewer: 2

This is an excellent paper, which is very sounded and well written. In it the authors used an approach developed by Coifman in which diffusion maps are used to generate a geometry of graphs. My only concern is that the manuscript does not mention other similar approaches that are also suitable for the type of analysis done in this work. In fact, it could be argued that they are more appropriate as diffusion maps, but it is a matter of discussion. My suggestion is to include the following (or some of the following) references: M. Pereda, Estrada, E. Visualization and machine learning analysis of complex networks in hyperspherical space. Pattern Recognition 86 (2019) 320-331. M. Akbarzadeh, Estrada, E., Communicability geometry captures traffic flows in cities. Nature Human Behaviour, 2, 2018, 645-652. E. Estrada, N. Hatano, Communicability angle and the spatial efficiency of networks. SIAM Review, 58, 2016, 692-715 . E. Estrada, M. G. Sanchez-Lirola, J. A. de la Pena, Hyperspherical Embedding of Graphs and Networks in Communicability Spaces. Discrete Applied Mathematics 176 2014, 53-77. Estrada, E., The communicability distance in graphs. Linear Algebra and its Applications, 436 2012, 4317-4328.

We thank the reviewer for the nice words. It is indeed embarrassing that we overlooked the link to the work of Estrada at al. We think about diffusion maps more in the context of harmonic analysis and less in the context of graph embedding. But we agree with the reviewer that the link is clearly there and

now discuss it in the revised version, including several of the suggested refernces.

Reviewer: 3

This is an interesting paper looking at a novel methodology to find the most relevant variables describing a system, going beyond PCA. The method is applied to urban systems, and in particular to census data, taking Bristol as a case study.

My first comment, which is already discussed by the authors, is that the step of thresholding seems somehow arbitrary, and there's scope for future work to find a more rigorous methodology. In particular, the thresholding does not correspond to the values but to the number of links constructed between OAs. I would argue that for an initial approximation, considering only 10 neighbours gives good enough results, but a rigorous method in which the distribution of weights is considered to decide on an uneven set of neighbours could be explored. This is comment for future work.

We thank the reviewer and fully agree that the thresholding step merits more attention. We chose the 10 neighbour rule based on the basic intuition of thresholding strongly but avoiding fragmentation. We also verified that (within reason) other choices do not significantly impact the result, but this will indeed be an important target for future work.

With respect to the paper in its state for publication, Here's a list of questions, comments: 1) validation: the method seems to be able to pick up places that have related characteristics, such as the universities and student halls. Were the authors able to verify that ALL relevant student places are contained in the OAs picked up by the first eigenvector?

Yes, in the area under consideration all university accommodation is highlighted by the first eigenvector. We don't have data on the students living in private accommodation, but the distribution suggested by the eigenvector is very plausible. We now point this out in the manuscript.

2) Difference between census and selected variables: there is indeed a whole problem of how the census differs between England and Wales, and Scotland. It would be insightful to say how the 231 variables were selected.

The variables selected were those for which the identifying name (column header) in the England and Wales, and Scotland could be easily matched using an unsophistcated automatic method. We have described this now in some more detail in the text.

3) Given that there's no 1-1 map between OAs in Bristol and any other selected case study, more clarity is needed on how the values are assigned from Bristol to other cities following the distance in the similarity space, in particular, when the destination is a city whose variation is much larger than Bristol.

If the destination city would have a much greater variation, this type of transfer would fail. However our results show that Bristol is diverse enough to approximate the manifolds for other British cities, including London. Regarding the question of how the most similar area is selected: We pick the closest Bristol OA to each Edinburgh OA by Euclidean distance in the 231 diemnsional space. Subsequently not every Bristol OA must have a match in Edinburgh and multiple Edinburgh OAs can be assigned the value from a single Bristol OA. We have added this explanation to the text.

4) With respect to the meaning of the eigenvectors, this also needs clarification. Given the large set of variables, how were these probed in order to understand from "local knowledge" that the first eigenvector corresponded to student populations? The same comment holds for the second eigenvector. Is there a second process of correlating the results at OA level with the variables? Is there a geographic weighted regression carried on, or what sort of correlation? This needs to be explicit in the text.

The eigenvectors identify coherent sets of responses to the census. The process of identifying the cause of that coherence for the first two eigenvectors was done by visual inspection of the patterns produced on maps. Measuring the correlation of eigenvectors with each individual column of the census data did not significant correlation in any situation due to the nonlinearity of the underlying manifold. In the results section we have emphasised the identification is by inspection, and returned to this in more detail in the conclusions

Overall, I found the paper interesting and inspiring for future applications and generalisations. Although as it stands, it does not provide a rigorous answer to the variables that need to be considered in a specific urban context, since it seems that previous knowledge needs to intervene, it does give a first step in this direction, in particular since it goes beyond linear models. For this reason, I strongly recommend publication.

We appreciate the reviewers comments and thank them for their constructive input that we believe have helped us improve the manuscript. We also appreciate his suggestion for considering the role of the threshold in future work. We agree that a rigorous method would improve the continued application and represent a step forward in method's development. At present the method rigorously identifies the most informative variables, and with the help of some human interpretation provides at least testable hypothesis regarding the meaning of these variables.

Board Member Comments to Author(s):

Although diffusion maps are useful tools for the embedding of graphs, other alternatives not mentioned in the paper are also important. For instance, communicability geometry has emerged as an important embedding to analyse cities,

as shown in: https://www.nature.com/articles/s41562-018-0407-3. Such approaches should be discussed in the paper.

We agree that other methods for the embedding of graphs, especially low dimensional embeddings, are also important. We have added reference to these in our conclusions, including the likely role they play in the future analysis of cities.