Answers to reviewers' comments:

First of all, we would like to thank the reviewers for the positive and constructive comments.

Reviewer #1:

This paper analyzes a relevant social issue, which is migration under humanitarian crisis. It is a valuable contribution from the computational social science perspective, and shows that the usage of data from social media can help to real-time track migration events. The paper is clearly written and the methodology is well described, so I think that it can be a valuable contribution for PLOS ONE.

However, I have a couple of minor issues and doubts which I suggest that the authors address:

1) In Fig. 1 and Fig. 5, I think that the correlation values in term of the R^2 do not provide a clear highlight when the distribution of values seems to obey a power law. This R^2 is probably mainly dominated by the first 2 or 3 countries. I suggest to remove this indications, or either justify their usage, or either compute them in a logarithmic scale.

The log-log scale in the representation is only a matter of convenience since the data points have several orders of magnitude. Both figures are showing an identity relation, as we are comparing expected versus estimated flows. The important question is whether the points fall over the diagonal and, consequently, the analysis on the "goodness" of fit must be done linearly to have sense. We cannot expect any scaling law out of linearity because that would imply that the model is not reproducing well the numbers observed in official data and that a systematic bias has been introduced. We added in the captions of Fig. 1 and 5 a note to clarify this point.

2) It would be interesting to point out that this methodology does not fit well for some countries in the South Cone which are further from Venezuela, like Chile (~290k "official" migrants vs ~800k? estimated) and Argentina (~130k "official" vs ~600k? estimated). This might be related to higher Twitter penetration in those countries, which promotes its usage by migrants; or to the correlation between distance travelled and socio-economic status, or to other factors that make the upscaling work incorrectly for those countries. I think that this issue deserves to be briefly discussed.

We thank the reviewer for noticing this distortion. This is indeed an important observation to make. We think that the factors acting here may be two: on one side, as the reviewer states, the upscaling factor here is probably affected by the fact that the penetration rate in these countries is different. From a social perspective, migrants moving to other countries may conform to the local culture in the way the social platforms are used. However, this change of habits take time and, in some cases, it can be even generations. On the other hand, the difference could be this high because of a miss-representation of the Venezuelan migrants from the official statistics. Note what happened in Brazil, where the official numbers from the UN organizations are quite below the ones from the Federal Police and ours (Table 2). This is why it is so important to get information from extra sources. As we see from Figure 2, the routes of migration from Venezuela extend up to Santiago and Buenos Aires. We added a paragraph mentioning this issue in the Discussion Section.

3) I did not understand the following phrase in the Discussion: "We could have used a stricter criterion and request two or more tweets abroad but this does not affect the average flows (the upscaling factors absorb it), although it notably enhances the statistical fluctuations". As far as I understand, the scaling factor is computed as the ratio between Venezuela population and the amount of residents (TUV's). The criterion for detecting a migration situation does not affect any of the previous quantities. In this sense, if we apply a stricter "migration criterion" then the upscaled migration amount will be affected as well.

As the reviewer observed it is true that "the scaling factor is computed as the ratio between Venezuela population and the amount of residents (TUV's)". The sentence "We could have used a stricter criterion and request two or more tweets abroad but this does not affect the average flows (the upscaling factors absorb it), although it notably enhances the statistical fluctuations" is intended to say that if we want to use a stricter criterion to classify people abroad as migrants, we should be consistent on the residents classification as well. In this sense, if we want to check for two consecutive tweets abroad in a specific country, in order not to lose

the consistency, we should ask for residents to tweet at least twice in that year to consider them as active. On the other hand, by doing this, one notably reduces the sampling of the data and narrows the statistics, hence we discarded this option. In order to make the text clearer, we added this reflection to the above sentence.

Reviewer #2:

The authors propose a novel method for assessing and studying the phenomenon of migration using twitter geo-located data. The authors apply the method to the Venezuelan Migration crisis showing they are able to estimate the amounts of migrants in certain years. The estimates are compatible with those found by international organizations. Moreover, they provide a way to study in detail the geographic distribution of routes of migration. I find the idea of using Twitter data for migrations quite appealing, despite the limitations this kind of data might have (even though the authors provide a discussion of such limitations in the conclusions). Researchers interested in migration patterns do not have always access to private data from mobile phone companies, and surveys made by international organizations might not have the desired level of detail for certain studies. Hence, I would recommend the article for publication with minor revisions I am certain the authors will be able to address easily:

- in page 9 the authors propose a way to estimate the fluxes crossing the border of the Venezuelan country. However, I was not able to understand precisely how this is done (maybe due to my limited comprehension ability). Is it estimated by counting the number of vectors crossing the line? Are the authors able of following the trajectory of an individual and hence assess whether he crosses the border? Please rephrase it better in the text.

We thank the reviewer for pointing out this issue on the understanding of what are the numbers estimated in our methods. In the previous part of the manuscript, the method used to define a previously classified Venezuelan resident as a migrant was to detect at least one tweet from them in a second country as a proof of border crossing (Fig.1 and Table 2). We added this explanation in Validation of external flows subsection, lines 246 and following. On the other hand, from line 291, we introduce a different method, which requires a stricter sampling. We now want to assess whether migrants moving on the ground crossed a specific line along their trajectory. The results of this new measure are depicted in Figure 3 and Table 3. In order to be sure that they crossed the dashed line, we have to take at least one tweet on one side and one tweet on the other side of the dashed line. The vectorial depiction is a way to characterize the general direction of movement but it is not used to count the number of crossings. We added a clarification regarding this measure in lines 293 and following.

- The authors at page 9 makes distinction between migration patterns on land and by airplane. Of course the second ones belong to less disadvantaged individuals but still the flux might be relevant for migration studies. Do the authors think it would be possible to identify this kind of migrations as well?

There is possibility to detect air trips by having tweets of the same user in two faraway places and with a time interval compatible with a flight speed (between 300 and 900 km/h). We found a few cases in our data but they are not enough to do proper statistics. One can always use more relaxed criteria, like assuming that tweets happening between distant locations are footprints of air displacements regardless of the time between them. However, this can lead to false positives, like people who traveled on the ground by car/bus and never tweeted along the route.

- In the discussion section the authors state that the work has proven to be helpful for humanitarian agencies. Do they mean it has already been applied by these agencies for some of their studies? In this case, I would add a reference if available. Otherwise, I would state that the method "could be useful" or "have potential use for" these agencies.

It must be noticed that part of the authors belongs to UNICEF, specifically they are based in Brasilia and New York. Some of them are operatives and our results and data were discussed during the decision-making process regarding the Venezuelan crisis in Brazil and other nearby countries. This statement was included by them during the writing process and the rest of authors has no reason to consider it as false.

In particular, the insights from this research helped UNICEF to keep a broader vision of the scale of the migration problem beyond the border with Venezuela, which was the case before. Based on this, the UNICEF team moved into looking at (i) how to integrate the humanitarian response into our regular program of cooperation, especially our Municipal Seal of Approval; and (ii) expanding the reach of an AI-inspired project on xenophobia beyond the State of Roraima, close to the border.

- In general the work is interesting due to the fact the data used is publicly available. I would discuss a little bit about possible comparisons with private data in order to further validate the method.

We have added a paragraph in the Discussion section commenting on the possible comparisons that one could make with other data sources like private data.

After having addressed this minor comments, in my opinion the work will be ready for publication.