Responses to the comments of Reviewer #2

1. A large challenge that articles of this type face is how to demonstrate the effectiveness of a visual and/or interactive tool. It is not usually convincing for the authors of the tool to simply claim that the tool performs well. Formal experiments with human subjects might provide convincing evidence, but they are difficult and costly to conduct. In this case, the fact that no comparable tool exists perhaps justifies the work in itself. However, in this particular case the interactive tool results could at least be compared with the results of applying algorithmic tools (does the human do better or worse?).

   Response: We now include a simple illustration, that the human eye may indeed be better than algorithms in some cases. That is, we added a small network in the introduction that algorithmically is clustered multiple ways, but with gravicom the true structure is easily seen by the human eye. To formally compare gravicom to algorithmic approaches, we would need human subjects which is outside the scope of this introductory (rather that evaluatory) paper. Additionally, as mentioned in the paper, "accuracy" of community detection isn't necessarily the goal here, but rather finding community structure in a novel way.

2. the file format for data upload was not clear to me (page 6)

   Response: We have revised to now mention an appendix which contains description of the file structure necessary for user data.

3. [page 7] "extend to which" should be "extent to which"

   Response: We fixed the typos "of if plausibility or extend to which" to read "of plausibility or extent to which"

4. once a selection of groups of nodes has been made, how can that selection be "exported" (and in what form), so that it can be used in further work (e.g., to provide initial conditions for algorithmic tools).

   Response: We now explain the form of data export on page 7 and how communities are stored in the `gml` file.

5. [page 12] "is a fundamentally" should be "is fundamentally"

   Response: We changed the wording "a fundamentally" to "fundamentally".

6. [page 13] VERY nice description and portrayal of the interactive use of the system!

   Response: Thank you!

7. it is not clear what the second example is adding; could it be used to demonstrate the upload of a new data set into the system?

   Response: In the second example, we have now added an image showing the process of uploading user data into gravicom, specifically the political books example. This small example is adding another data set that has a clear visual structure. We have added a few sentences to the section hopefully making that clear.

8. [page 17] it is not clear how this is "dynamic" or why this is "temporal". Is this just a matter of comparing pairs of graphs based on a categorical variable ? (i.e., facetting?)

   Response: You're right, this example isn't dynamic because that capability isn't available in gravicom currently. However, this example illustrates the potential of what dynamic community detection could contribute. We have rearranged the section and added some content making it more clear that it would be nice to be able to look at this dynamically, allowing for a temporal structure, however currently this is not possible within gravicom. Secondly, this serves to illustrate temporal (shapsnot) data because there is an inherent sequential ordering of the graphs.

Responses to the comments of Reviewer #3

1. Page 5, line 17: "is comprised of" should be changed to "comprises." (Comprises is equivalent to includes. It's a pet peeve of mine.)

   Response: We have changed "is comprised of" to read "comprises".

2. Page 6, the paragraph describing the Navbar doesn't agree with Figure 1. The final button is the link to GitHub (from what I can tell), and the third button, which appears to be an envelope, is the link to the author's website.

   Response: We have switched the order in navbar description paragraph to correspond to the application.

3. Page7, line 4: Change "idea of if the plausibility of extend" to "idea of the the plausibility of extent".

   Response: We fixed typos "of if plausibility or extend to which" to read "of plausibility or extent to which".

4. Now the quibble. The authors state in the paragraph beginning around line 27 on page 2 that community detection has been mainly the realm of computer science and physics with the statistics community only lately coming to the game. However, I would like to point out that social network analysis predates much of the literature cited in the references. Holland and Leinhardt published their paper on exponential families for directed graphs in 1981 in JASA. Faust and Wasserman's book "Social Network Analysis" was published in 1994. I point this out, not as a claim for precedence, but to point out that there are many different criteria for determining communities beyond cliques and near-cliques, which the authors may wish to explore in the future.

   Response: We have added text acknowledging the earlier notions in the literature about defining and modeling structure in network data through graph topological features regarding node connectedness (e.g., counts of triangles, degree counts) which are related to the underlying aspects of community structure.

5. Which brings me to my questions. It's my experience that the datasets analyzed in the paper are almost at the limits of what the human is capable of clustering by eye. If one were to double the number of Division 1 football teams the task illustrated in Section 3.1 would become more difficult, but how much more? How does the task of visual clustering scale compared to algorithmic clustering? Have the authors considered an approach where they might first do an incomplete, suboptimal algorithmic approach to clustering nodes (a low-hanging fruit pass) followed by a visual clustering for the harder cases? How exactly do they envision combining the human and the machine for the clustering task?

   Response: This is an interesting thought. We have added a discussion in section 4.1 to address the possibility of using gravicom as a post-processing step in regards to an algorithmic approach for very large or complex graphs. In order to actually answer the question of how things scale we would need a human subjects experiment, which is outside the scope of this paper.

6. (As an aside, since finding networks with known structure can be difficult, I would like to share with the authors the paper "Benchmark graphs for testing community detection algorithms," http://arxiv.org/pdf/0805.4770v4.pdf in case they are unaware of it.)

   Response: Thank you very much for sharing this resource. We have added this reference in the introduction.