Dear Editor-in-chief and referees,

We thank the reviewers for the time and expertise they have invested in their reviews. We sincerely appreciate the comments and suggestions from the referees and the editor. The suggestions offer a valuable help for the improvement of our research. We have made a careful revision to adequate our manuscript to the format required by the journal and to comply with the reviewer's requests. The following is the revision statement corresponding to each comment. We look forward to complying with all of your suggestions.

Best regards,
I. P. Matsuno, R. G. Rossi, R. M. Marcacini and S. O. Rezende

Reviewer B

Comment: *A more rigorous formalization of the approach should be provided. An illustration of the approach should be presented in section 3.2 following the example presented in section 3.1*

Response: Thank you for your suggestion. We emphasize that the novelty and contributions of our article are two-fold: (1) the ASBA problem representation through a heterogeneous networks and (2) a robust evaluation semi-supervised learning algorithms for heterogeneous networks. Section 3.1 presents in detail our proposed approach to the ASBA problem representation through a heterogeneous networks (bipartite networks), including a rigorous formalization of the approach and illustrative examples. Section 3.2 formalizes the semi-supervised learning algorithms for heterogeneous networks, in particular the regularization functions used in our bipartite networks.

We agree with the Reviewer B that Section 3.2 lacks an illustrative example to facilitate understanding of the problem formalization. We added the Figures 5 and 6 to illustrate the iterations of semi-supervised learning algorithm. In these figures we show the initial state of the bipartite heterogeneous network, the label propagation and the final state of the bipartite network (for both aspect identification and aspect sentiment classification steps). We believe that this general explanation can help the reader to better understand the problem formalization and label propagations algorithms.

Comment: *A more complete analysis of the experimental results should be provided:, e.g. discussion concerning precision and recall separately; why is F-measure lower for sentiment analysis?; why does restaurant dataset have worst results then laptop dataset?*

Response: In the revised version, we improve Section 4.1. to summarize the characteristics of the datasets used in the experimental evaluation (number of aspects and sentiment class distributions). In particular, we calculate the S-Index for both restaurants and laptops datasets. S-Index intend to measure the overlap among the classes. When S-Index value is close to 1 means that the classes are well separated, i.e, there is no overlap among the classes. Otherwise, the value is close to 0. Therefore, S-Index can be used to understand the classification complexity for each dataset in controlled scenarios. In addition, S-Index measure will facilitate the reader to answer questions such as *"why does restaurant dataset have worst results then laptop dataset?"*.

Regarding the comment "*discussion concerning precision and recall separately*", we unpretentiously believe that the Micro-Averaging and Macro-Averaging are appropriate measures for a general experimental comparison of the algorithms. Moreover, we fully agree with the reviewer that the use of multiple evaluation metrics is an important issue that deserves attention. Thus, due to space constraints, we publish a detailed experimental results with other evaluation measures such as precision and recall in the webpage of our proposal (http://gepic.ufms.br/asphn2016/), including the datasets used in the experimental evaluation and ASPHN source code.

Finally, to improve the quality of the experimental analysis and to make the results even more convincing, we added a statistical significance analysis of the experimental results (Non-parametric Friedman Test with Nemenyi post-hoc test – 95% of confidence level). In this analysis, we compared our proposed approach with to the traditional supervised learning approach. The statistical analysis is summarized in Fig. 8 and Fig 10 by a critical difference diagram.

Comment: *About the quality of the text and its presentation. Can the paper be published as it is?: No*

Response: We carefully review the article and try to improve the text by ourselves. Then we send the article to be reviewed by English experts. We think we have improved the article writing.

Comment: *Along with this evaluation I sent the pdf with comments in it.*

Response: Thank you for your comments. We greatly appreciate the reviewer's efforts to carefully review the article and the valuable suggestions offered. With respect to the concerns raised by the reviewer, we have taken into account almost all points and modified our article as described below:

1. All typos were fixed;
2. We changed the text of the first paragraph of page 2, as the reviewer's suggestion. In this paragraph, we have added references that the reviewer asked about the use of NLP and ML for aspect-based sentiment analysis;
3. Regarding the comment "*If the aspect candidates can be nouns, verbs, adjectives and adverbs, why only those 4 terms are considered?*", we correct our example. In addition, we have added a footnote to make clear that some words are not used in the representation because they are considered stopwords;
4. The symbols and notations of the formulas have been corrected to improve our problem formalization;
5. Regarding the comment about graph regularization assumptions, we improved our text according to the definition in Zhu, X. and Goldberg, A. B. Introduction to semi-supervised learning. Morgan and Claypool Publishers, 2009. We also rewrote the graph regularization functions to reflect each assumption: "In each regularization function the first term is related with the first assumption and, analogously, the second term describes the second assumption.";
6. We provide more details on the parameters of the algorithms as the reviewer's suggestions (PDF annotated comments);

7. Regarding the comment "*It is not clear that the approach allows aspect identification. In the experiments, for example, the aspect came from annotated data. In section 3.1, it does not seem that the 4 aspects were automatically identified*", note that our ASPHN proposed approach allows to model both the identification of aspects as the sentiment classification by using heteregeneous networks (in a unified representation). The first step (aspect identification) uses linguistic features and a small sample of labeled aspects to learn classifier model. This classifier identifies new aspects from unlabeled sentences of the domain. For example, the algorithm TCBHN achieved an identification rate of new aspects (Macro-F1 and Micro-F1) between 0.77 and 0.81 using only 10% of labeled aspects, for laptop and restaurants domains respectively. To improve the quality of the experimental analysis and to make the results even more convincing, we added a statistical significance analysis of the experimental results (Non-parametric Friedman Test with Nemenyi post-hoc test – 95% of confidence level). In this analysis, we compared our proposed approach with to the traditional supervised learning approach. The statistical analysis is summarized in Fig. 8 and Fig 10 by a critical difference diagram; and
8. Regarding the commend "*A detailed discussion bringing reasons for the low value on F1 would improve the paper. Moreover, it is interesting to know if the approach is better on precision or recall.*", we improve Section 4.1. to summarize the characteristics of the datasets used in the experimental evaluation (number of aspects and sentiment class distributions). We also added a measure to understand the classification complexity (class overlapping) for each dataset in controlled scenarios (S-Index). Due to space constraints, we publish a detailed experimental results with other evaluation measures such as precision and recall in the webpage of our proposal (http://gepic.ufms.br/asphn2016/), including the datasets used in the experimental evaluation and ASPHN source code.

Reviewer C

Comment:
Does the paper have enough contribution that justifies its publication at JIDM?: Yes.
About the quality of the text and its presentation. Can the paper be published as it is?: Yes

Response: Thanks for your appreciation of our paper.

Reviewer D

Comment: *Does the new material in this paper corresponds to 30% of the paper length? This paper corresponds to an extension of a previously published work (Matsuno et. al 2015) and it is clear that the authors present an improved version.*

Response: We much appreciate the reviewer's careful review.

Comment: *This work proposed an approach to perform Aspect-based Sentiment Analysis through semi-supervised learning techniques. The main motivation is to perform ABSA using a less amount of labeled data when compared to supervised learning scenario. In general, the paper is well-written and organized. In addition, the experiments are reasonably well described and the results are promising. This paper corresponds to an extension of a previously published work (Matsuno et. al 2015) and it is clear that the authors present an improved version, including in the experimental evaluation more semi-supervised learning algorithms and more supervised learning algorithms. In addition the authors provided a more complete literature review.*

Response: We appreciate the positive feedback from the reviewer.

Comment: *In Section 2.2 the authors mention that machine learning approaches considering supervised learning require a huge amount of labeled data to obtain an accurate classifier. It is not always true. In addition, the authors highlight that labeling data is time consuming and makes the use of supervised learning approaches unfeasible in pratical situations. Again, as there are several works that apply supervised learning approaches to solve real problems, I consider that this is a very strong statement.*

Response: Thank you for your comments. We have taken into account almost all points and modified our manuscript accordingly.

We improved our text to emphasize that labeling data is time consuming in the context of aspect-based sentiment analysis. In this context, first we need to label which textual expressions are aspects for a particular application domain. Secondly, we need to label the phrases in which the aspects occurred to define the sentiment (positive, negative or neutral).

In fact, this well-known difficulty to obtain labeled data for sentiment analysis is the great motivation of several studies that explore semi-supervised learning algorithms, as we have cited in the last paragraph of section 2.2.

Comment: *The authors affirm that aspect based sentiment analysis with semi-supervised learning is underexplored. So, is there any paper dealing with aspect based sentiment analysis using semi-supervised learning?*

Response: Note that our ASPHN proposed approach allows to model both the identification of aspects as the sentiment classification by using heteregeneous networks (in a unified representation). A similar study investigates only the aspects identification step by semi-supervised learning, but still rely on a significant amount of labeled data for sentiment classification. On the other hand, there are studies investigating semi-supervised learning algorithms for sentiment classification only for document-level or sentence-level sentiment analysis. Thus, we believe that the our ASPHN is the first work for semi-supervised learning aspect-based sentiment analysis in a unified view.

We emphasize that most machine learning literature methods for sentiment analysis exploit supervised learning algorithms. Thus, to achieve a fair and robust evaluation, we chose the traditional supervised learning algorithms and the main semi-supervised learning algorithms for heteregenous networks.

Comment: *The words used in the first column of Table II do not correspond to the given example.*

Response: Thank you. We correct the Table II.

Comment: *It was very useful the exampless adopted in Section 3.1 to explain the two bipartite heterogeneous networks used by proposed approach. However, for sake of clarity I recommend the authors to use the same example to generate Tables I and II.*

Response: Thanks for valuable comments. We kept separate examples to illustrate to the reader (in an intuitive way) the two main steps: aspects identification and sentiment classification. The use of a single example for both steps requires more phrases and thus (1) generate a heterogeneous network with many nodes, (2) require more iterations for convergence. We believe that separate examples dedicated to each step is a simpler way to understand the approach.

Comment: *While Section 3.1 is well explained and includes a helpful example, Section 3.2 lacks of clarity. Why not also using an illustrative example in this section? It should be important the authors provide the relationships between each set/vector (e.g., C, T, B, F(O), y, w) and the itens of the problem in question.*

Response: Thank for you providing these ideas. We added two new figures (Fig 5 and 6) to illustrate the algorithms presented in Section 3.2. We use the Fig. 5 and Fig. 6 to illustrate the relationship between the each set (e.g., C, T, B, F(O), y, w) and the problem formulation.

Comment: *In addition, concerning algorithms adopted to perform graph regularization, the authors have presented the regularization function used by them. Nevertheless, they didn't provide explanation about these functions. If the authors consider that it is important to present the regularization functions, they should be clearly described.*

Response: As we have mentioned in the problem formalization on Semi-Supervised Learning based on Bipartite Heterogeneous Networks, "*The graph regularization has to satisfy two assumptions: (i) the class information of neighbors must be close and (ii) the class information assigned during the classification process must be close to the real class information.*" (see Section 3.2).

Different regularization functions have been proposed in recent years and we reformulate these functions using the same variable definitions of our proposed problem formulation. We provide references to the original works. A detailed explanation of the regularization functions for sentiment analysis is beyond the scope of this work.

Comment: *I would like to see more details about the datasets, i.e., number os aspects and distribution of aspect polarities.*

Response: Thanks for the suggestion. We reformulated the sentence about the datasets and include more details about number of aspects e class distributions. Please see the revised manuscript.

Comment: *For all algorithms used in the experiments the authors evaluated different parameter settings. However, they didn't provide the parameters used by algorithms to generate the results presented in Figures 5 and 6. Based on results plotted in Figures 5 and 6 the authors presented two comparative evaluations: semi-supervised learning algorithms versus supervised learning algorithms and algorithms based on networks versus algorithms based on vector space model. Nevertheless, all conclusions were obtained without statistical tests. Thus, statistical tests could be used in order to obtain a more precise analysis of results.*

Response: We appreciate the reviewer's insightful comment. Due to space constraints, we publish a detailed experimental results with other evaluation measures such as precision and recall in the webpage of our proposal (http://gepic.ufms.br/asphn2016/), including the datasets used in the experimental evaluation and ASPHN source code.

Finally, to improve the quality of the experimental analysis and to make the results even more convincing, we added a statistical significance analysis of the experimental results (Non-parametric Friedman Test with Nemenyi post-hoc test – 95% of confidence level). In this analysis, we compared our proposed approach with to the traditional supervised learning approach. The statistical analysis is summarized in Fig. 8 and Fig 10 by a critical difference diagram.

Comment: *I suggest you rewrite the last sentence of Section 4.3. It is quite confuse.*

Response: Thanks for the suggestion. The paragraph has been revised.

Comment: *The link provided by athours (http://gepic.ufms.br/asphn2016/) was broken.*

Response: Thank you. We correct this error and update the webpage contents.

Comment: *Some minor errors:*

...despite the many step that can...
...such approaches requires a huge...
...to perform ABSA, that allow learning a model...
...linguist features and aspect candidates...
...we present and extension about...
...bipartite network composed the aspects...
...aspect identification and sentiment analysis though semi-supervised learning...
...Figure 6(a) classification performance...
...restaurant dataset an Macro-F1...
...we considered two dataset....
...to classify web objects connect to social....
...in this experimental we....
...Eech review may contain more none, one or more than...
Response: Done. Thank you.