Dear chair and reviewers:

Thanks to the reviewers' insightful comments which have significantly helped us to prepare the camera-ready version of the paper. In the following, we present our response to the comments from each reviewer.

In response to reviewer 1:

Reviewer 1's 1st main concern:

"As mentioned above, the evaluation experiments in the paper needs much improvements. First of all, the approach seems like having a strong tendency toward picking more popular entities as a map. As an example, I am not sure at all how the system matches a not-so-popular Wei Wang (who perhaps has published a paper in the same conference that the famous Wei Wang has published in) to the correct entity in DBLP. even the entity object model seems to be unable of finding the correct Wei Wang in this case. On the other hand the text data set in the experiment section has been generated from a search engine which of course give you the text for the most popular persons. Thus the high accuracy reported in the paper might be a little biased."

Answer: Our proposed model is a combination of the entity popularity model and the entity object model. From the experimental results shown in Table 5, we can see that $\text{SHINE}_{\text{all-eom}}$ that just leverages the entity object model alone achieves very high accuracy (i.e., 0.931), which demonstrates that the entity object model is the key component for entity linking. Combining the two models, $\text{SHINE}_{\text{all}}$ obtains a little higher accuracy (i.e., 0.942) than $\text{SHINE}_{\text{all-eom}}$, which also shows that the entity popularity model is helpful for the entity linking task. Only when the entity object model gives the same score for the candidates, we prefer to link with the most popular one. In the case you mentioned, even if the not-so-popular Wei Wang published a paper in the same conference that the famous Wei Wang has published in, our entity object model could leverage the coauthor information, the title term information and the year information to link it with the correct entity.

To generate the test document collection, we try to find the Web documents for different candidate entities (both popular and not-so-popular ones) with respect to each name mention via adding some domain representative phrases (such as "database", "data mining", "graphics", etc.) in the search queries. The baseline POP method that always links a name mention to its most popular candidate entity achieves very low accuracy (i.e., 0.487) shown in Table 5, which demonstrates that our constructed data set is not biased.

Reviewer 1's 2nd main concern:

"Second, a larger data set should be considered for scalability analyses. Even for other parts of the experiments that would be very beneficial. Also, the current results indicates a linear growth in time while the data set size is linearly increased. This actually does not show the approach is scalable at all."

Answer: To the best of our knowledge, there is no publicly available benchmark data set for the

task of entity linking with a HIN. Since the annotation task for entity linking with a HIN consists of many steps, the annotation task is difficult and very time consuming. Therefore, due to the limited time, we just created this data set for evaluation and made it online available for future research. Evaluation over a larger data set is left for future work.

In the paragraph under Algorithm 1, we state that the running time of this algorithm is linear to the number of entity mentions in *M*. When the number of entity mentions in *M* is enormous, our learning algorithm becomes a little expensive. At that time, we could use the stochastic gradient descent method which samples a subset of entity mentions at each iteration and updates the parameters wp's on the basis of these sampled entity mentions only [1] and has been shown very effective for large-scale learning problem,. Then, the running time of our learning algorithm is linear to the number of sampled entity mentions.

Reviewer 1's 3rd main concern:

"Third, the weight case study should have been expanded since it is one of the main contributions of the paper. currently you are showing that the weight improves the accuracy with respect to only one other case (SHINE4). You need to show it works better for most of possible combinations."

Answer: In Section 5.5, we show the effectiveness of our learned weights to the performance of SHINE not $SHINE_4$. SHINE is the complete model we propose for entity linking with a HIN in this paper so we tested the impact of the learned weights to this model. Other models shown in Table 5 are either baseline methods or truncated models we propose to demonstrate the performance of different sub-models (i.e., entity popularity model and entity object model). So in this paper, due to limited space, we just show the effectiveness of our learned weights for the complete SHINE model. Additionally, from the experimental results for other models, we could obtain similar results.

Reviewer 1's 4th main concern:

"As for the other issue mentioned above, the paper leaves the pattern selection for the future work, but I think the authors should have at least studied the effect of pattern selections on the final results as well as the baselines. An interesting idea here is to harvest some of possible such meta-paths as explained in section 3.2 (for a smaller portion of data perhaps) and use your weight estimation approach to select the most effective ones. This will give you a more complete approach that you can then claim it would work for other HINs."

Answer: We dealt with the pattern selection problem in Section 4. In this section, we proposed a learning algorithm that can automatically learn the weights for meta-paths. A larger weight indicates a higher importance for the meta-path with respect to the entity linking task. The extreme case of weight = 0 means this pattern embedded in the meta-path is totally irrelevant to the entity linking process. In Section 5.5, we studied the effectiveness of our learned weights to the performance of SHINE.

Reviewer 1's 5th main concern:

"Some minor issues:

1) it would be nice to discuss more applications on different HINs."

Answer: We discussed more applications on other HINs in our paper. We added the IMDb network as another example for the task of entity linking with a HIN and discussed it throughout the paper. Specifically, we added the IMDb network schema in Figure 2, and illustrated the IMDb network above Definition 2. We also introduced four different meta-paths in the IMDb network along with their semantic meanings denoted by each meta-path above Section 2.2. In the last paragraph of Section 4, we discussed how to use our proposed model to link actor name mentions in Web text with the IMDb network. In addition, in the same paragraph, we also discussed how to link author name mentions with a new bibliographic network that has a new object type (i.e., organizations (ORG)) and a new relation type (i.e., isAffiliatedWith).

Reviewer 1's 6th main concern:
"2) the paper should make clear which part of the task is performed as preprocessing and which part will be repeated for each new text documents. Otherwise it's really hard to understand the time performance evaluations."

Answer: The operations introduced in Section 5.1 are performed as preprocessing including network construction, test document generation, candidate entity generation, and object recognition in documents. The time consumed in these operations is not counted in the running time evaluation in Figure 4(a). To make this clear, we added a sentence at the end of Section 5.1.

Reviewer 1's 7th main concern:
"3) seems like the results in table 2 are for the page rank (pr(e)) not the popularity (P(e))."

Answer: The results in Table 2 are the popularity scores *P(e)* for candidate entities in Example 1 calculated using Formula 7.

Reviewer 1's 8th main concern:
"4) you may want to study the impact of using many more meta-paths to prove the scalability of the approaches."

Answer: In our paper, we demonstrated the scalability of our approach via using different sizes of entity mentions which need to be linked. As the number of meta-paths created for some specific entity linking task is usually a small constant, we did not study the scalability performance of using more meta-paths, which could be left for future work due to limited space.

In response to reviewer 2:

Reviewer 2's 1st main concern:

"W1. The problem setting assumes all of the mentions occur in the information network. This is often not true in practice. Ideally given a "Wei Wang" that does not exist in DBLP yet, your model should decide there is no match. Only so it allows the growth of the information network."

Answer: In our paper, for simplicity, we assume that the heterogeneous information network contains all the mapping entities for all the named entity mentions. We think the method for predicting entity mentions that do not have their corresponding entity records in the network could be regarded as a meaningful extension of our current work.

Reviewer 2's 2nd main concern:

"W2. The models requires (nearly) exact string match and sets P(m|e) as a constant. Ideally it should compute P(m|e) from string similarity and I assume this extension should not be hard."

Answer: For simplicity, we assume that the probability *P(m|e)* of observing the name of surface form *m* given each candidate entity *e* for mention *m* is the same and defined as a constant, because in the candidate entity generation step introduced in Section 5.1, we leverage a method based on strict name string comparison. All author entities in the DBLP network that satisfy the predefined strict rules are extracted as the candidate entities, so these candidate entities are very likely to be the correct candidates for this mention. Therefore, we think it is reasonable to consider it as a constant in our problem setting. Of course, the idea to compute *P(m|e)* from string similarity is a very interesting extension we could consider.

Reviewer 2's 3rd main concern:

"W3. P_g(v) is not well explained; there are only 2-3 sentences about it and they are not very clear. Fig 5 shows that P_g(v) plays a dominating role in the model but P_e(v), which is explained in details, is not that important. What is P_g(v) after all and why is it so powerful?"

Answer: *P_g(v)* is actually a smoothing function to deal with the data sparse problem. It is independent of each entity and is the same with respect to different entities. If we set *theta* to 0 in Formula 9 (i.e., just leverage the generic object model *P_g(v)* in the entity object model), our model reduces to the POP baseline method which performs badly in the experiments shown in Table 5. This means, the entity specific object model *P_e(v)* which we explained in details is the key component and plays a dominating role for entity linking. Therefore, the parameter *theta* in Formula 9 does not indicate the relative importance of the two parts, and we revised the description under Formula 9 and in Section 5.4 accordingly.

Reviewer 2's 4th main concern:

"When I readn Eq (1), I can hardly understand what P(e) means, P(m|e) means, and P(d|e) means.

Problem 2 should be defined right at the beginning so the readers won't get lost by Eq (1) . Problem 1 can be merged to Sec 3.2."

Answer: We moved the original Problem 2 (i.e., Inference problem) to the beginning of Section 3 and merged the original Problem 1 (i.e., Estimation problem) to the end of Section 3.2.

Reviewer 2's 5th main concern:
"Table 3. Did you get this set by enumerating all paths?"

Answer: We obtained this set according to our knowledge. The creation of this meta-path set is very easy and won't consume much time.

Reviewer 2's 6th main concern:
"4th paragraph of Sec 5.1. Why each Web document has a single mention? Is that what happens for this data set, or do you just choose one mention?"

Answer: For each document, we just chose one mention to be linked. For other mentions in the same document, their mapping entities may not be disambiguated in our partially disambiguated DBLP network so that we cannot link them with the network.

Reviewer 2's 7th main concern:
"Table 4-5: What is #?"

Answer: # means the number of correctly linked author mentions. In these two tables, we changed it to "# correctly linked" to make it clear.

Reviewer 2's 8th main concern:
"Table 6: It will be easier to read if you rank the meta-paths by the weight."

Answer: We rearranged the content of Table 6 to demonstrate the meta-paths by their weights in decreasing order.

In response to reviewer 3:

Reviewer 3's 1st main concern:

"*D1. It is not clear if the proposed model can be used for other applications. Throughout the paper, the authors focused on a very specific example, linking the name mention "Wei Wang" to the correct name entity "Wei Wang" in the DBLP data set. It is not clear if it will work for other types of entity linking, or for other types of heterogeneous information networks. More discussion or case studies are needed.*

D4. The authors only conducted the experiments with one data set and only focused on the author names linking. I suggest them to test the proposed model on multiple different data sets, or give analysis of more examples."

Answer: Our probabilistic model can be applied to the task of entity linking with arbitrary heterogeneous information networks directly. To illustrate this claim more clearly, we gave analysis of more examples in our paper. We added the IMDb network as another example for the task of entity linking with a HIN and discussed it throughout the paper. Specifically, we added the IMDb network schema in Figure 2, and illustrated the IMDb network above Definition 2. We also introduced four different meta-paths in the IMDb network along with their semantic meanings denoted by each meta-path above Section 2.2. In the last paragraph of Section 4, we discussed how to use our proposed model to link actor name mentions in Web text with the IMDb network. In addition, in the same paragraph, we also discussed how to link author name mentions with a new bibliographic network that has a new object type (i.e., organizations (ORG)) and a new relation type (i.e., isAffiliatedWith).

To the best of our knowledge, there is no publicly available benchmark data set for the task of entity linking with a HIN. Since the annotation task for entity linking with a HIN consists of many steps, the annotation task is difficult and very time consuming. Therefore, due to the limited time and space, we just created one data set for evaluation and made it online available for future research. Evaluation over more data sets is left for future work.

Reviewer 3's 2nd main concern:

"D2. In the entity popularity model in Section 3.1, it is not clear what a link refers to. Are all the links between objects bi-directional? Is a citation relation a link?"

Answer: A link in the network refers to a relation between objects. The structure of the networks and the semantic meanings of nodes and edges in the networks are introduced in Section 2.1. The network schemas for the DBLP network and the IMDb network are shown in Figure 2. The edges between objects in the networks are bi-directional. For example, between authors and papers in the DBLP network, there are two types of relations (i.e., *write* and $write^{-1}$). Additionally, the arrows in Figure 2 are all bi-directional, which also demonstrates that the links between objects are bi-directional. In the DBLP network we adopted in the experiments, there is no citation relation.

Reviewer 3's 3rd main concern:

"D3. Since multiple objects may refer to the same entity, and later the authors evaluated their model on a partially disambiguated DBLP network, it is not clear how to incorporate this disambiguation information into the entity popularity model and entity-object model."

Answer: The disambiguation information is mainly captured in the entity object model. In the experiments, the partially disambiguated DBLP network was constructed by combining the disambiguation results from DBLP and a publicly available data set used in [29], which contains 110 author names and their gold standard disambiguation results. Here, disambiguation means we determine which author names in the publication records refer to the same author entity. In this partially disambiguated DBLP network, for each disambiguated author entity, all her publication information is linked with it. So we use meta-path constrained random walks starting from each disambiguated entity to generate the object distribution for this entity. This generated object distribution captures their disambiguation information since for different entities the networks associated with them are different and are dependent on their disambiguation results. Figure 3 shows the examples of the generated object distributions for different author entities with the same name "Wei Wang". Then we use these different object distributions to calculate the probability of generating the context with respect to different entities in the entity object model.

Reviewer 3's 4th main concern:

"*D5. Although the authors compared their model with some baseline methods, more comparisons are needed. For example, since the first and second baselines are orthogonal, what is the performance to combine them? Such a comparison is reasonable since the proposed model is also a combination of the entity-popularity mode and the entity-object model. Also, compared with the proposed probabilistic model, what is the performance for a simple probabilistic topic model without considering the meta-path?"*

Answer: We have evaluated the performance of the method that combines the first and second baselines via multiplying the entity popularity score by the cosine similarity score, the same way as our proposed model. The combined baseline method obtains the accuracy of 72.5% which is better than the performance of the first baseline but worse than the performance of the second baseline shown in Table 5. Thus, we did not demonstrate its result in our paper.