

## Checklist of Datasets and Benchmarks Contribution

### CompanyKG: A Large-Scale Heterogeneous Graph for Company Similarity Quantification

1. For all authors...
  - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? [Yes] Section 1
  - (b) Did you describe the limitations of your work? [Yes] Appendix 5
  - (c) Did you discuss any potential negative societal impacts of your work? [Yes] Appendix 5
  - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
  - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
  - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
  - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Appendix A
  - (b) Did you specify all the training details (e.g., data splits, hyper-parameters, how they were chosen)? [Yes] Appendix D
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Table 2 in main paper and Table 7 in Appendix
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Appendix D
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
  - (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 3 and Appendix F.2
  - (b) Did you mention the license of the assets? [Yes] Appendix A
  - (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
  - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? [Yes] See Section 3, Appendix F.2
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See question ⑬ in Appendix F.2
5. If you used crowdsourcing or conducted research with human subjects...
  - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [Yes] Appendix F.4.③
  - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [Yes] Appendix F.3.⑥
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [Yes] Appendix F.3.④