

Topical Word Importance for Fast Keyphrase Extraction

Lucas Sterckx, Thomas Demeester, Johannes Deleu, Chris Develder
Internet Based Communication Networks and Services, Ghent University - iMinds

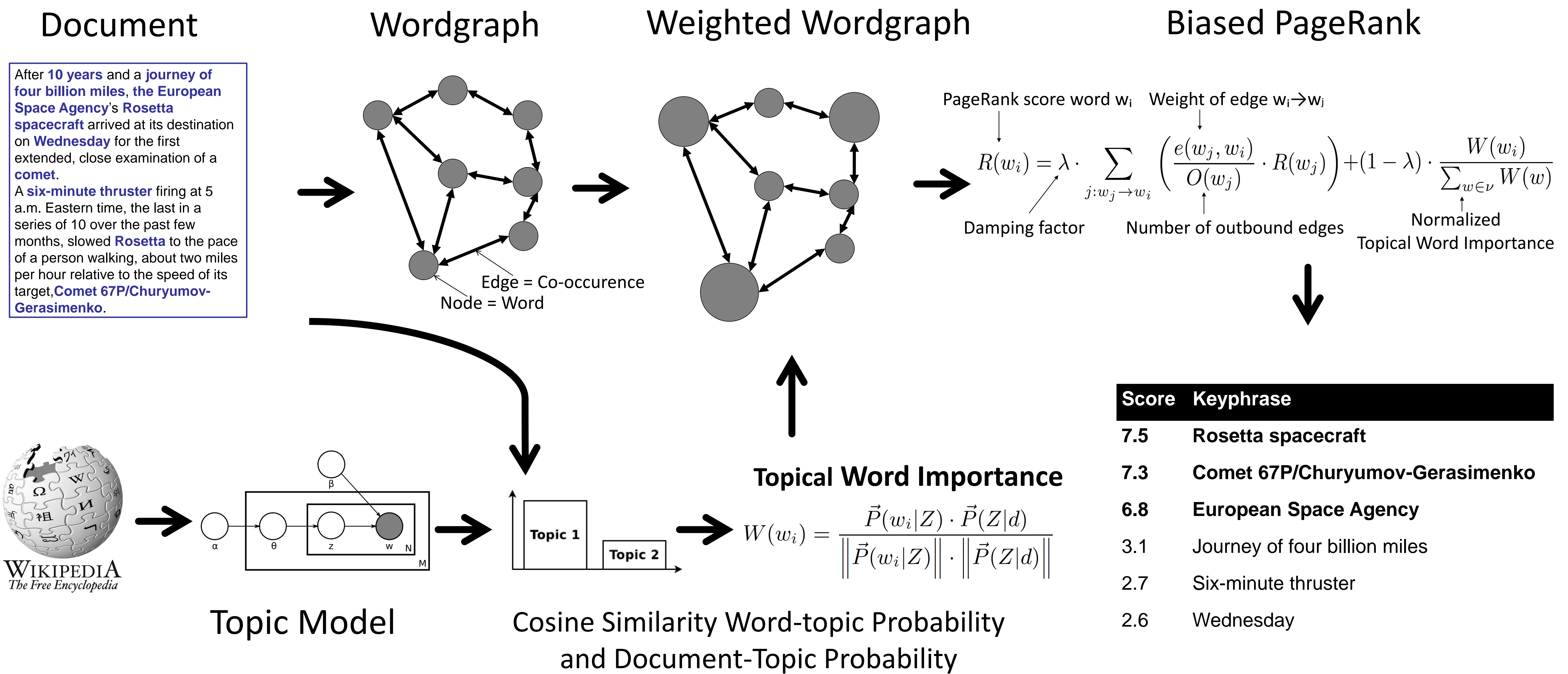
Topical Keyphrase Extraction

- Automatic selection of important and topical phrases
- Useful for Summarization, Contextual Web Advertisement, classification, ...
- **We introduce a fast and scalable algorithm for the incorporation of information from topic models in an unsupervised graph-based algorithm for keyphrase extraction**

Topical Pagerank

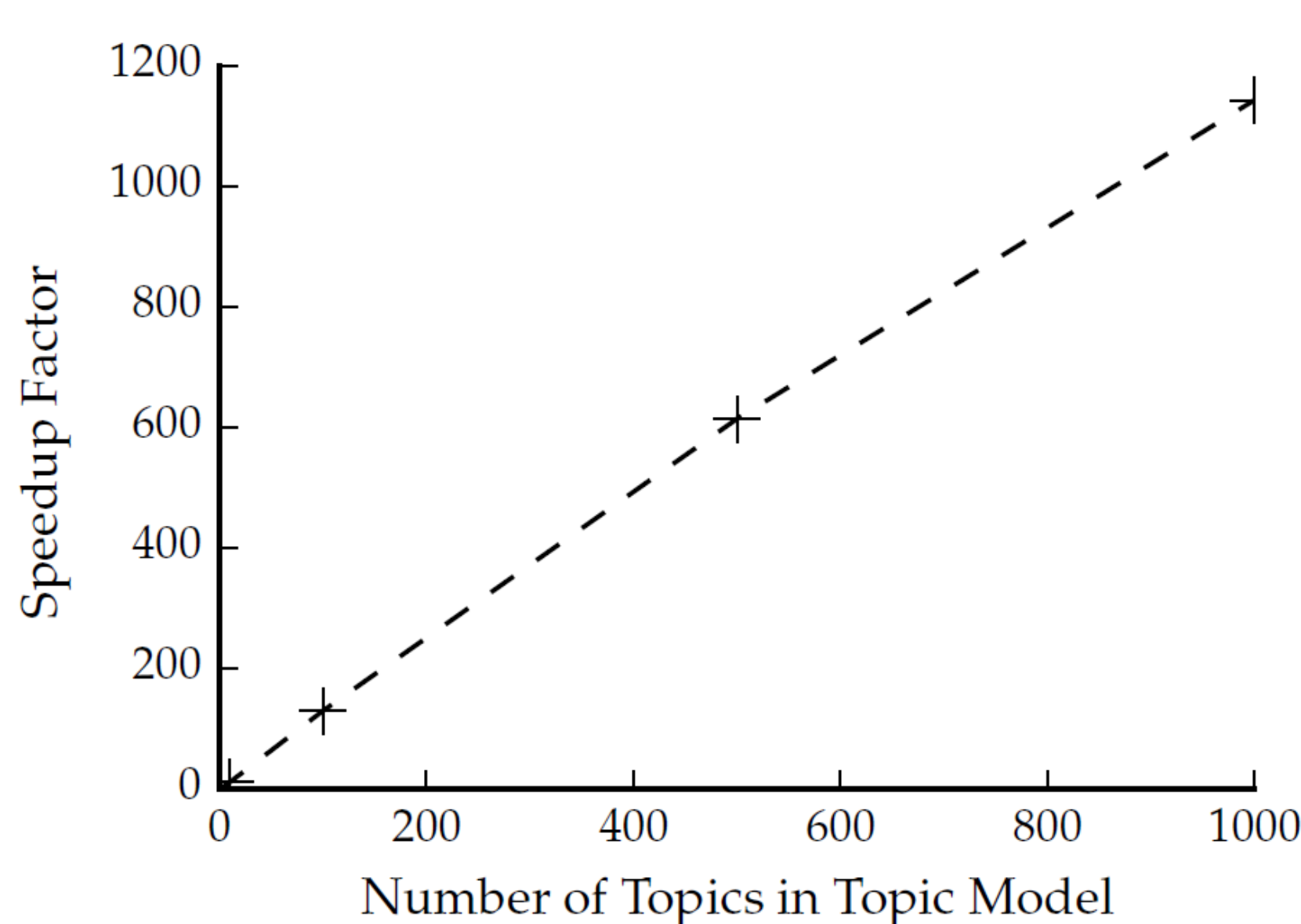
- Variation on the TextRank-algorithm
- Incorporates topical information by increasing the weight of important topical words in the word graph based on the topic-document and word-topic distributions generated by a topic model
- **Requires a random walk for each topic in the topic model!**

Single-PageRank Topical Keyphrase Extraction

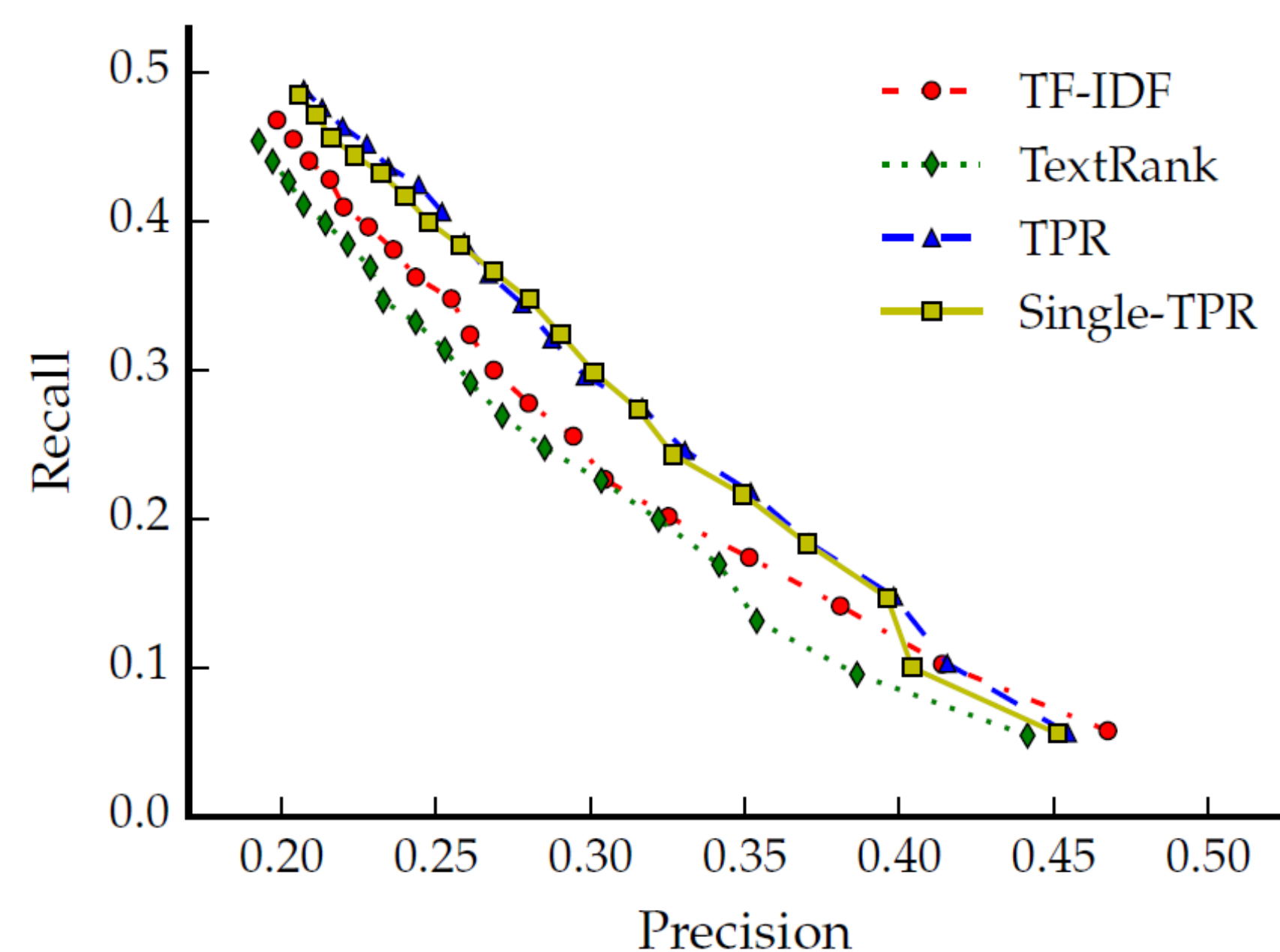


Experiments

News articles from the 2001 Document Understanding Conference



Speed-up Increases with Topic Model size



Performance Matches Original TPR

Conclusion

- **Single value for topical word importance** in a PageRank algorithm based on the cosine similarity between the vector of word-topic probabilities and the document-topic probabilities of the document
- **Constant computation time, independent of the topic model**
- **Does not alter the effectiveness** while reducing the computation time by a large margin