	doc_1		doc_2		decision	id
		Zhelezniak, V.Savkov, A.	authors	 Vitalii Zhelezniak Aleksandar Savkov April Shen Nils Y. Hammerla 		
	authors title	• Shen, A.	title	Correlation Coefficients and Semantic Textual Similarity		
cases		,	publication_dat	lication_date 2019-05-19 18:23:14+00:00		
			source	SupportedSources.ARXIV		
		Correlation Coefficients and Semantic Textual Similarity	journal	None		
	nublication date	e 2019-01-01 00:00:00	volume			
	source SupportedSources.CROSSREF		doi		<u> </u>	1
	journal		urls	• http://arxiv.org/pdf/1905.07790v1		S 341
	volume			 http://arxiv.org/abs/1905.07790v1 http://arxiv.org/pdf/1905.07790v1 		
	doi	10.18653/v1/n19-1100				
	urls	• http://dx.doi.org/10.18653/v1/n19- 1100	id	id3435085937423583027		
			abstract	A large body of research into semantic textual similarity has focused on constructing state-of-the-art embeddings using sophisticated modelling, careful choice of learning signals and many clever tricks. By contrast, little attention has been devoted to similarity measures between these embeddings, with cosine similarity being used		
	id	id4011132769841544343		unquestionably in the majority of cases. In this work, we illustrate that for all common word vectors, cosine similarity is essentially equivalent to the Pearson correlation coefficient, which provides some justification for its use. We thoroughly characterise cases where Pearson correlation (and thus cosine similarity) is unfit as similarity measure. Importantly, we show that Pearson correlation is appropriate for some word vectors but not others. When it is not appropriate, we illustrate how common non-parametric rank correlation coefficients can be used instead to significantly improve performance. We support our analysis with a series of evaluations on word-level and		
	abstract					
	versions					
				sentence-level semantic textual similarity benchmarks. On the latter, we show that even the simplest averaged word vectors compared by rank correlation easily rival the strongest deep representations compared by cosine similarity.		
			versions			