Evaluation: document similarities

How do we test that *similar documents* have *similar vectors*?

- ArXiv triplets: paper A, similar paper B, dissimilar paper C
- Measure the accuracy of guessing the dissimilar paper

http://arxiv.org/pdf/1206.5743	http://arxiv.org/pdf/cond-mat/0403258	http://arxiv.org/pdf/1408.0189
http://arxiv.org/pdf/1209.0268	http://arxiv.org/pdf/1307.7598	http://arxiv.org/pdf/math/0504051
http://arxiv.org/pdf/hep-ph/9908436	http://arxiv.org/pdf/nucl-th/9707019	http://arxiv.org/pdf/1112.3014
http://arxiv.org/pdf/1111.2905	http://arxiv.org/pdf/1303.2538	http://arxiv.org/pdf/1109.1922
http://arxiv.org/pdf/nucl-ex/0112013	http://arxiv.org/pdf/physics/9704013	http://arxiv.org/pdf/1408.4595
http://arxiv.org/pdf/0709.3419	http://arxiv.org/pdf/quant-ph/0611134	http://arxiv.org/pdf/0902.0616
http://arxiv.org/pdf/hep-th/9609148	http://arxiv.org/pdf/solv-int/9710009	http://arxiv.org/pdf/astro-ph/0508060

Andrew Dai, Cristopher Olah, Quoc Le. Document Embedding with Paragraph Vectors, CoRR, 2015.