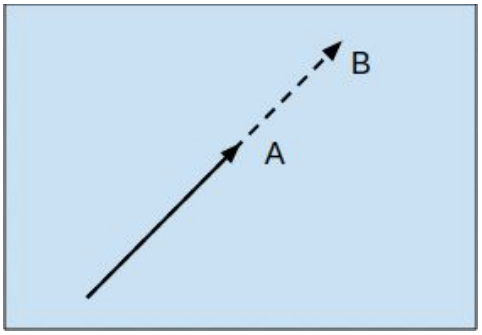
Paper Critique:

Interactive Analysis of Word Vector Embeddings

F. Heimerl,M. Gleicher

This paper brings out some interesting problems associated with word embeddings visualizations, especially the large scale paper review (111 ACL publications related to word embedding vectors). But a close inspection of the paper reveals that many of the visualization methods have some problems.



The **single neighborhood** task uses cosine distance which has been stated as the most common choice in the literature. Cosine distance though might be a good metric for comparing vector distance like [man-woman] and [king-queen] but word-embeddings are coordinates in the R_n dimensional space (scalar). If we consider them as vectors that would mean [word-origin] vector. Consider the example shown in the figure on left, both vectors have cosine similarity of 1 but they are not close in the vector space.

The coloring scheme used in the **multiple-neighborhood** task is quite confusing. A simpler design would have used lines when hovering over the words (even they had to resort to drawing a line on ‘cousin’ between the uncle and aunt axes to explain their coloring scheme).

It’s not clear how the paper computes **reconstructed co-occurrence** values of word pairs. The **embedding stability** analysis is concluded using banal arguments like ‘ stochastic nature of training embeddings can result in unexpected and unwanted local results’, the last visualization does not lead to a conclusion which essentially means that it is ineffective in meeting its goal (analyzing the stability of embeddings).