

Na Figura 1 é mostrado um exemplo de uma matriz de similaridade onde a intensidade do ponto (i, j) representa similaridade entre as sentenças i e j e linha diagonal representa a similaridade entre as mesmas sentenças. Observa-se que a matriz é simétrica e revela quadrados ao longo da diagonal que indicam os as regiões com maior coesão léxica.

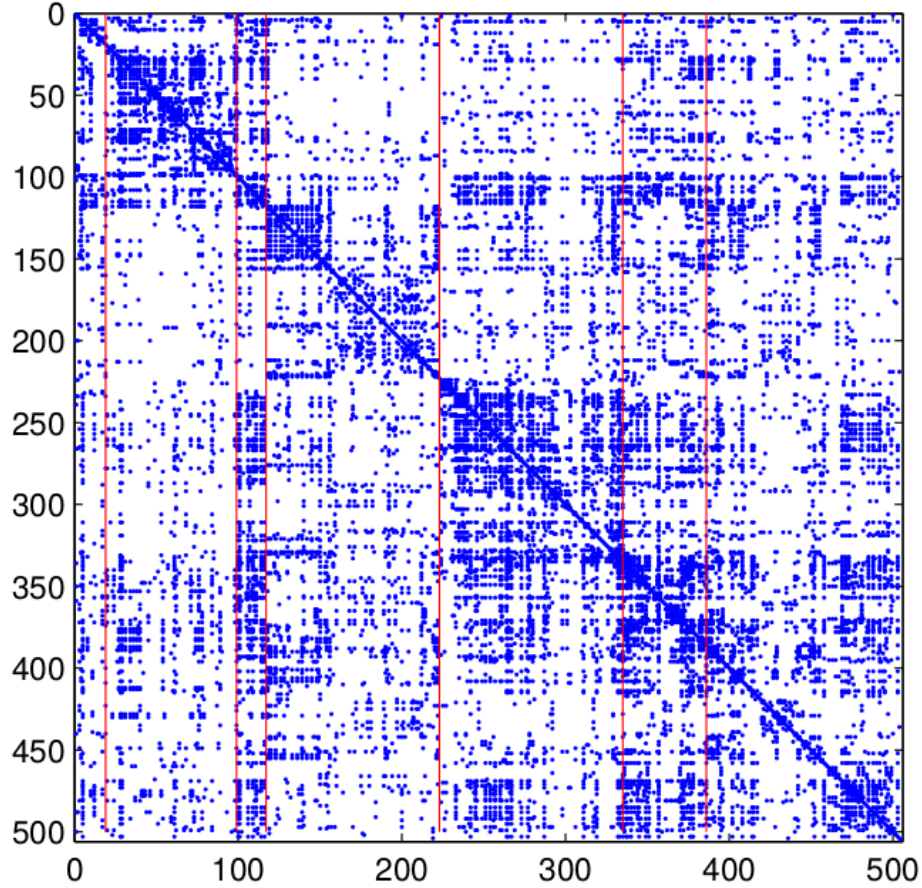


Figure 1: *DotPlot* da similaridade entre sentenças onde as linha verticais representam segmentos reais.

O processo de identificação dos limites é baseado no método DotPlotting [?]. Um segmento é representado por duas sentenças i e j que representam uma região quadrada ao longo da diagonal da matriz de similaridades.

Seja $s_{i,j}$ a somatória dos *rankings* do segmento que inicia na sentença i e termina na sentença j e $a_{i,j}$ sua área interior. Seja $B = \{b_1, \dots, b_m\}$ a lista de m segmentos.

A medida de densidade D é dada por

$$D = \quad (1)$$

onde s_k a_k são a somatória dos valores dos rankings e a área de um segmento k em B .