Na Figura 1 é mostrado um exemplo de uma matriz de similaridade onde a intensidade do ponto(i, j) representa a similaridade entre as sentenças i e j. Observa-se que a matriz é simétrica, assim cada ponto na linha diagonal representa a similaridade quanto $i = j$ (ou seja, com a mesma sentença) e revela quadrados com maior concentração de pontos ao longo da diagonal. Essas regiões indicam porções de texto com maior coesão léxica.

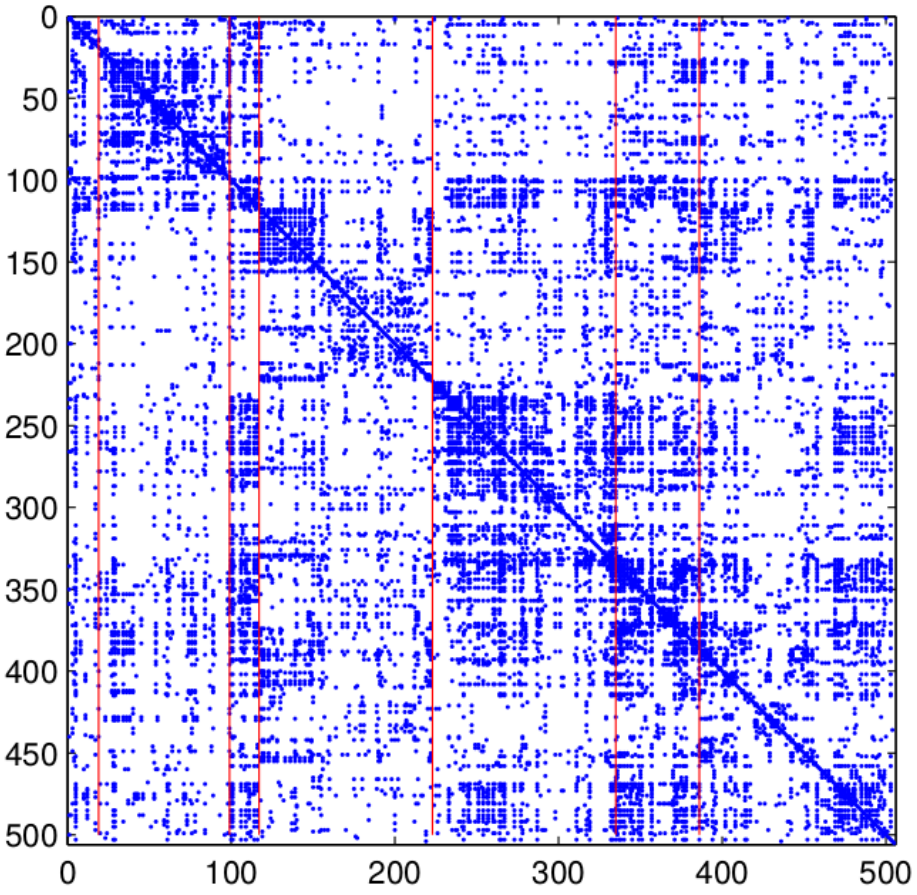

Figure 1: *DotPlot* da similaridade entre sentenças onde as linha verticais representam segmentos reais.

O processo de intentificação dos limites é baseado no método DotPloting [**?**] que usa regiões com maior densidade em uma matriz de similaridades para determinar como os segmentos são distribuídos. Um segmento é definido por duas sentenças i e j que representam uma região quadrada ao longo da diagonal da matriz. Calcula-se a densidade dessa região como mostrado na Equação 1. Seja $s_{i,j}$ a somatória dos *rakings* de um segmento e $a_{i,j}$ sua área interior. Seja $B = \{b1, ..., b_m\}$ a lista de m segmentos es_k a_k são a somatória dos valores dos rankings e a área de um segmento k em B. Então, a densidade é computada por:

$$D = \frac{\sum_{k=1}^{m} s_k}{\sum_{k=1}^{m} a_k} \tag{1}$$

O processo incia com um único segmento formado por todas as sentenças do documento e o divide recursivamente em m segmentos. Cada passo divide um dos segmentos em B no ponto ij que maximiza D (Equação 1). O processo ser repte até atingir o número de segmento desejados ou um limiar de similaridade.