**创新点改进方案：**

在 SEMMA 的 4.3 节文本关系图环节，替换固定阈值为数据集自适应阈值：

语义密度计算：对每个 KG 的关系文本嵌入，计算所有关系对的 cosine 相似度分布，用 “K-means 聚类拐点法” 找阈值，

根据如下公式：



k表示簇的个数，表示聚类中心

总结：根据各簇类中心对应的cosine相似度求平均得到数据集的动态阈值

实现：代码量少，无需额外模型；

阈值确定后，仍沿用原有的 “过滤边→构建文本关系图流程，仅替换阈值参数。

改动优势

改动极小：仅修改文本关系图构建中的 “阈值参数” 计算方式，不影响后续 模型的处理；

代码具体实现要求：

1. 配置中添加参数dynamic-threshold bool类型
2. 如果配置中dynamic-threshold为真，则按照上述的方案根据不同数据集的分布选择不同的阈值
3. 代码要求简洁，可根据参数去掉或者加上，不影响其他代码，只根据当前计算关系图的数据集进行计算阈值。
4. 上述方案若有不妥，请指出并询问是否按照修改后的方案修改代码