文本分析术语：

1. 节点（顶点）：在文本分析中，一个节点代表一个人物。
2. 度：节点的度定义为与该节点相连的边的数目。
3. 社交网络:由许多节点构成的一种社会结构。节点通常是指个人或组织，而社交网络代表着各种社会关系。
4. 网络平均度：网络平均度反应了网络的疏密程度，而通过度分布则可以刻画不同节点的重要性。
5. 度中心性：度中心度是指与该节点直接相连的节点的数量。度中心性（Degree Centrality）是在网络分析中刻画节点中心性（Centrality）的最直接度量指标。一个节点的节点度越大就意味着这个节点的度中心性越高，该节点在网络中就越重要。
6. 入度：通常指有向图中某点作为图中边的终点的次数之和。
7. 出度：对于有向图来说，顶点的出边条数称为该顶点的**出度。**
8. **邻居：节点连接到的节点。**
9. **特征向量中心性：**特征向量中心性理论认为一个节点的重要程度与其相 连其他节点的重要程度息息相关，即对于一个节点来说， 如果该节点与很多本身具有较高中心度的点相连接的话， 那么该点就具有高的重要程度。
10. **聚类系数（Clustering Coefficient）：**用于描述网络中与同一节点相连的节点间也互为相邻节点的程度。其用于刻画社交网络中一个人朋友们之间也互相是朋友的概率，反应了社交网络中的聚集性。
11. **介数（Betweeness）：**为图中某节点承载整个图所有最短路径的数量，通常用来评价节点的重要程度，比如在连接不同社群之间的中介节点的介数相对于其他节点来说会非常大，也体现了其在社交网络信息传递中的重要程度。
12. 词云：由词汇组成类似云的彩色图形。“词云”就是对网络文本中出现频率较高的“关键词”予以视觉上的突出，形成“关键词云层”或“关键词渲染”，从而过滤掉大量的文本信息，使浏览网页者只要一眼扫过文本就可以领略文本的主旨。
13. 气泡线（Bubblelines）：语料库中的每个文档均以一条水平线表示，并分为等长的段（默认为50段）。每个选定的单词都表示为气泡，气泡的大小指示相应文本段中单词的频率。气泡越大，单词出现的频率就越高。
14. 气泡（Bubbles）: 气泡是按文档形式对字词频率的生动呈现。高频字词按文档顺序读取，当前字词显示在右上角。下面的字词列表显示了文档中该点的累积频率等级。第一次遇到字词时，会创建一个气泡并将其添加到画布的主要部分。该字词在读取时以黄色闪烁。气泡的相对大小表示相对于文档中当前点的其他字词而言的频率。
15. 语料库搭配（Corpus Collocates）:
16. 停用词（Stopwords）：
17. 词频（TF）
18. **逆向文件频率（IDF）：逆向文件频率**（inverse document frequency，IDF）
19. 词频-反转文件频率（**term frequency–inverse document frequency，**TF-IDF）
20. 字典树：
21. 后缀树和后缀字典树：
22. 自由度
23. 凝固度