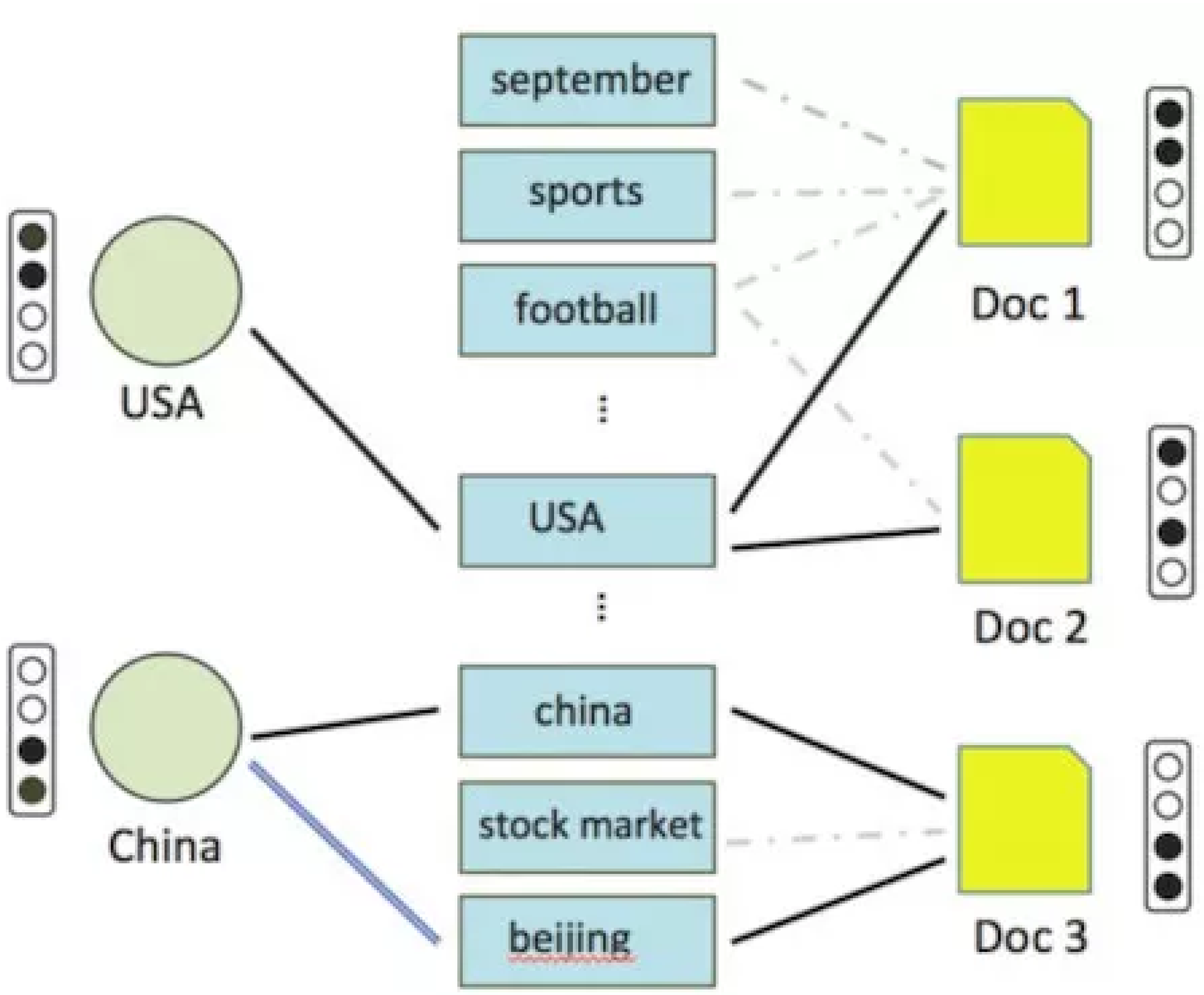
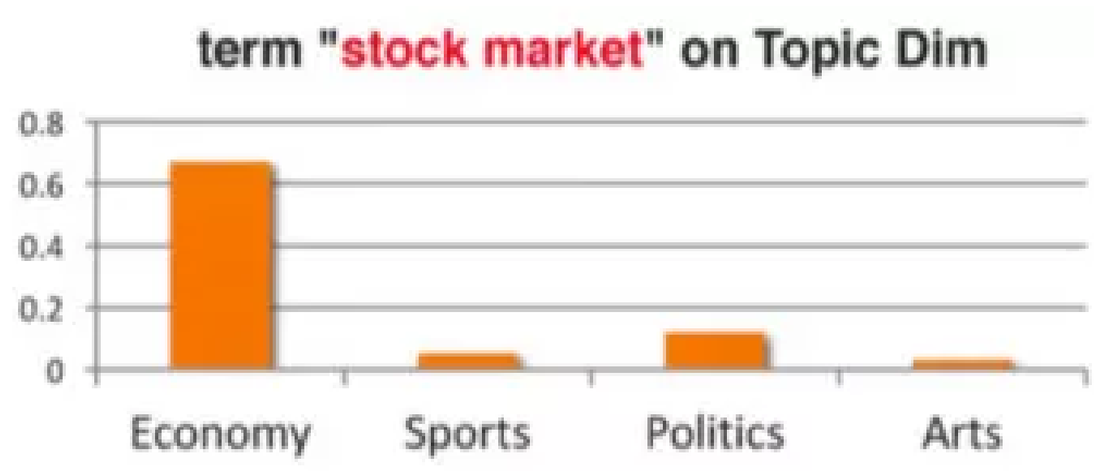
一个现实的问题就是，假如给你 100 万个 Documents，而只有少量几个标签（例如上述 Location、Topic 的标签），那么你能否自动地生成成百上千的标签，并将文本正确地放入到这些标签构建的多维 Text Cube 中呢？

首先去做的当然是 Embedding，但是已知的标签太少了。所以韩家炜他们建了一个 L-T-D（Label-Term-Document）图，其中的 Term 是从文本中抽取出来的。



我们查看每个 Term 在每个已知 Label 中的分布情况。



例如「stock market」，它在每个 Location 维度中分布的概率基本一致，这说明「stock market」这个 term 不属于 Location 这个维度；而另一方面，它在 Topic 维度的分布则有很强的差别性。根据一个称为 Dimension-Focal Score 的标准可以判别出它是属于 economy 标签下的。

依据上面的方法以及该 term 在这个标签下的普遍程度（如果大于某个值），则可以判断出这个 Term（例如「stock market」）属于相应标签维度下的一个标签。藉此，我们可以自动地生成大量的标签，并同时将文本放入到这些标签构建的多维度 Text Cube 当中。

Dimension	Label	1st Expansion	2nd Expansion	3rd Expansion
Topic	*Movies*	films	director	hollywood
	Baseball	inning	hits	pitch
	Tennis	wimbledon	french open	grand slam
	Business	company	chief executive	industry
	Law Enforcement	litigation	law	county courthouse
Location	*Brazil*	brazilian	sao paulo	confederations cup
	Australia	sydney	australian	melbourne
	Spain	madrid	barcelona	la liga
	China	chinese	shanghai	beijing

构建出这样的 Text Cube 之后，再去进行数据挖掘就会方便很多。