1)Алгоритм Ароры - отдельная статья. Основная идея такая: выделяем якорные слова, с помощью них по формуле Байеса инициализируем матрицу Phi слова-темы. Основное предположение - в матрице Phi есть диагональная подматрица, то есть в каждой теме есть ровно одно якорное слово - слово, которое встречается только в этой теме и ни в какой другой.

2)Случайная инициализация - либо полностью случайная, либо случайно-разреженная (может быть даже не с равномерным распределением). Эти все случаи несильно отличаются друг от друга.

3)Кластеризация слов - кластеризуем слова, найденные центроиды объявляем темами. Нормализуем столбцы матриц

4)SVD инициализация - применяем к исходной матрице SVD разложение - F=USV. Используем столбцы U и строки V, соответствующие наибольшим сингулярным числам, для заполнения матриц Phi, Theta (при этом используем их положительную или (отрицательную проекцию\*-1)). Отнормируем столбцы матриц

5)Пересчитаем матрицу слова документы в tf-idf, и запустим кластеризацию как в пункте 3