

# Ensemble Topic Modeling via Matrix Factorization

Mark Belford, Brian Mac Namee, Derek Greene

Insight Centre for Data Analytics, University College Dublin, Ireland  
mark.belford@insight-centre.org, brian.macnamee@ucd.ie,  
derek.greene@ucd.ie

**Abstract.** Topic models can provide us with an insight into the underlying latent structure of a large corpus of documents, facilitating knowledge discovery and information summarization. A range of methods have been proposed in the literature, including probabilistic topic models and techniques based on matrix factorization. However, these methods tend to have stochastic elements in their initialization, which can lead to their output being unstable. That is, if a topic modeling algorithm is applied to the same data multiple times, the output will not necessarily always be the same. With this idea of stability in mind we ask the question – how can we produce a definitive topic model that is both stable and accurate? To address this, we propose a new ensemble topic modeling method, based on Non-negative Matrix Factorization (NMF), which combines a collection of unstable topic models to produce a definitive output. We evaluate this method on an annotated tweet corpus, where we show that this new approach is more accurate and stable than traditional NMF.

## 1 Introduction

Topic models aim to discover the latent semantic structure or topics within a corpus of documents, which can be derived from co-occurrences of words across the documents. Popular approaches for topic modeling have involved the application of probabilistic algorithms such as Latent Dirichlet Allocation (LDA) [2, 15], and also, more recently, matrix factorization algorithms [19]. In both cases, these algorithms include stochastic elements in their initialization phase, prior to the optimization phase. This random component can affect the final composition of the topics and the rankings of the terms that describe those topics. This is problematic when seeking to capture a definitive topic modeling solution for a given corpus. Such issues represent a fundamental *instability* in these algorithms – different runs of the same algorithm on the same data can produce different outcomes [8]. Most authors do not address this issue and instead simply utilize a single random initialization and present the results of the topic model as being definitive. Another challenge in topic modeling is the identification of coherent topics using noisy texts, such as tweets [1]. The noisy and sparse nature of this data makes topic modeling more difficult when compared to analyzing longer, cleaner texts such as political speeches or news articles.

Here we consider the idea of ensemble learning, the rationale for which is that the combined judgment of a group of algorithms will often be superior to that of an individual [4]. Such techniques have been well-established for both supervised classification tasks [13] and also for unsupervised cluster analysis tasks [17]. In the case of the latter, the goal is to produce a more accurate or useful clustering of the data, which also avoids the issue of instability which is inherent in algorithms such as  $k$ -means. The application of unsupervised ensembles generally involves two distinct stages: 1) the generation of a collection of different clusterings of the data; 2) the integration of these clusterings to yield a single more accurate, informative clustering of the data. A variety of different strategies for both generation and integration have been proposed in the literature [7].

In this paper we propose an ensemble method for topic modeling, based on the generation and integration of the results produced by multiple runs of Non-negative Matrix Factorization (NMF) [11] on the same corpus. The integration aspect of the algorithm builds on previous work involving the combination of topics from different time periods with NMF [10]. To evaluate this method, we make use of a new Twitter corpus, the *20-topics* dataset, which provides partial ground truth annotations for user accounts. The results on this data indicate that the combination of many diverse models into a single ensemble topic model produces a more definitive and stable solution, when compared with randomly initialized NMF.

The paper is structured as follows. In Section 2 we explore related work in the areas of topic modeling and ensemble clustering. In Section 3 we describe how the two step process of our ensemble method works, before evaluating this new method in comparison to randomly initialized NMF in Section 4. Finally in Section 5 we conclude the paper with ideas for future work.

## 2 Related Work

In this section we will examine related work regarding topic modeling and the popular algorithms that are employed frequently in the field. We will also look briefly at ensemble clustering and the two main phases involved as outlined by previous literature.

### 2.1 Topic Modeling

Topic models attempt to discover the underlying thematic structure within a text corpus without relying on any form of training data. These models date back to the early work on latent semantic indexing by [5], which proposed the decomposition of term-document matrices for this purpose using Singular Value Decomposition [3]. A topic model typically consists of  $k$  topics, each represented by a ranked list of strongly-associated terms (often referred to as a “topic descriptor”). Each document in the corpus can also be associated with one or more topics. Considerable research on topic modeling has focused on the use of probabilistic methods, where a topic is viewed as a probability distribution over

words, with documents being mixtures of topics, thus permitting a topic model to be considered a generative model for documents [15]. The most widely-applied probabilistic topic modeling approach is Latent Dirichlet Allocation (LDA) [2].

Alternative algorithms, such as Non-negative Matrix Factorization (NMF) [11], have also been effective in discovering the underlying topics in text corpora [8, 19]. NMF is an unsupervised approach for reducing the dimensionality of non-negative matrices. When working with a document-term matrix  $\mathbf{A}$ , the goal of NMF is to approximate this matrix as the product of two non-negative approximate factors  $\mathbf{W}$  and  $\mathbf{H}$ , each with  $k$  dimensions. These dimensions can be interpreted as  $k$  topics. Like LDA, the number of topics  $k$  to generate is chosen beforehand. The values in  $\mathbf{H}$  provide term weights which can be used to generate topic descriptions, while the values in  $\mathbf{W}$  provide topic memberships for documents. One of the advantages of NMF methods over existing LDA methods is that there are fewer parameter choices involved in the modeling process. Typically NMF is initialized by populating  $\mathbf{W}$  and  $\mathbf{H}$  with random values before applying the optimization process. As noted previously, this can lead to different solutions of the two factors when applied to the same input matrix  $\mathbf{A}$ .

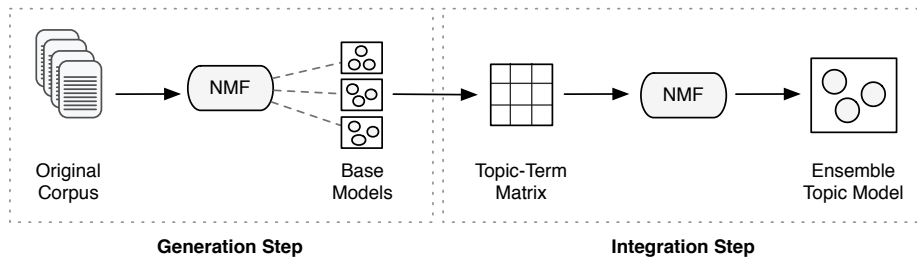
## 2.2 Ensemble Clustering

In the machine learning literature, it has been shown that combining the strengths of a diverse set of clusterings can often yield more accurate and stable solutions [16]. Such ensemble clustering approaches typically involve two phases: a *generation* phase where a collection of “base” clusterings are produced, and an *integration* phase where an aggregation function is applied to the ensemble members to produce a consensus solution. Generation often involves repeatedly applying a “base” algorithm with a stochastic element to different samples selected at random from a larger dataset. The most frequently employed integration strategy has been to use the information provided by an ensemble to determine the level of association between pairs of objects in a dataset [16, 6]. The fundamental assumption underlying this strategy is that pairs belonging to the same natural class will frequently be co-assigned during repeated executions of the base clustering algorithm. Other strategies have involved matching together similar clusters from different runs of the base algorithm.

While most of this work has focused on producing disjoint clusterings (*i.e.* each item in the dataset can only belong to a single cluster), researchers have considered combining probabilistic clusterings [14] and factorizations produced via NMF [9]. In the latter case, the approach was applied to identify hierarchical structures in biological network data.

## 3 Methods

In this section we will give a brief overview of how our proposed two step ensemble approach operates before delving deeper into how each of these steps work in greater detail.



**Fig. 1.** Illustration of the two steps involved in the ensemble topic modeling algorithm: generation and integration.

### 3.1 Overview

In this section we propose a new method for topic modeling, which involves applying ensemble learning in the form of two layers of NMF, in order to produce a stable and accurate final set of topics. This method builds on previous work on dynamic topic modeling involving the combination of topics from different time periods [10]. Fig. 1 shows an overview of the method, which can naturally be divided into two steps, following previous ensemble approaches:

1. *Ensemble generation:* Create a set of *base topic models* by executing multiple runs of NMF applied to the same document-term matrix  $\mathbf{A}$ .
2. *Ensemble integration:* Transform the base topic models to a suitable intermediate representation, and apply a further run of NMF to produce a single *ensemble topic model*, which represents the final output of the method.

We now discuss each of these steps in more detail.

### 3.2 Ensemble Generation

Unsupervised ensemble procedures typically seek to encourage diversity with a view to improving the quality of the information available in the integration phase [18]. Therefore, in the first step of our approach, we create a diverse set of  $r$  base topic models – *i.e.* the topic term descriptors and document assignments will differ from one base model to another. Here we encourage diversity by relying on the inherent instability of NMF with random initialization – we generate each base model by populating the factors  $\mathbf{W}$  and  $\mathbf{H}$  with values based on a different random seed, and then applying NMF to  $\mathbf{A}$ . In each case we use a fixed pre-specified value for the number of topics  $k$ . After each run, the  $\mathbf{H}$  factor from the base topic model (*i.e.* the topic-term weight matrix) is stored for later use. Note that in our experiments we use the fast alternating least squares implementation of NMF introduced by Lin [12].

### 3.3 Ensemble Integration

Once we have generated a collection of  $r$  factorizations, in the second step we create a new representation of our corpus in the form of a topic-term matrix  $\mathbf{M}$ . The matrix is created by stacking the transpose of each  $\mathbf{H}$  factor generated in the first step. It is important to note that this process of combining the factors is order independent. This results in a matrix where each row corresponds to a topic from one of the base topic models, and each column is a term from the original corpus. Each entry  $M_{ij}$  holds the weight of association for term  $i$  in relation to a single topic from a base model. To standardize the range of the values, we apply L2 normalization to the columns of  $\mathbf{M}$ .

Once we have created  $\mathbf{M}$ , we apply the second layer of NMF to this matrix to produce the final ensemble topic model. The reasoning behind applying NMF a second time to these topic descriptors is that they explicitly capture the variance between the base topic models. To improve the quality of the resulting topics, we generate initial factors using the popular Non-negative Double Singular Value Decomposition (NNDSVD) initialisation approach of [3]. As an input parameter to NMF, we specify a final number of  $k'$  topics. While this value can be set to be the same as the number of topics  $k$  in the base models, in practice we observe that an appropriate value of  $k'$  may be larger than this due to the ensemble approach being able to capture topics that only appear intermittently among a diverse set of base topic models. The resulting  $\mathbf{H}$  factor provides weights for the terms for each of the  $k'$  ensemble topics – the top-ranked terms in each column can be used as descriptors for a topic. To produce weights for the original documents in our corpus, we can “fold” the documents into the ensemble model by applying a projection to the document-term matrix  $\mathbf{A}$ :

$$\mathbf{D} = \mathbf{A} \cdot \mathbf{H}^T$$

Each row of  $\mathbf{D}$  now corresponds to a document, with columns corresponding to the  $k'$  ensemble topics. An entry  $D_{ij}$  indicates the strength of association of document  $i$  in ensemble topic  $j$ .

## 4 Experimental Evaluation

In this section we will give a brief summary of the dataset collected for this paper, the experimental setup, and finally an evaluation of our ensemble approach in comparison to randomly initialized NMF with respect to accuracy and stability of the topic models produced.

### 4.1 Data

One current area of interest for topic modeling is in the analysis of Twitter data [1]. However, annotated ground truth text corpora are rarely available for this platform, due to the scale of data involved. To evaluate our proposed method in the context of social media data, we collected a new corpus, the *20-topics*

**Table 1.** Number of tweets, unique user accounts, and user documents for each topic in the *20-topics* dataset.

Category	Tweets	Users	User Documents
Aviation	186,641	57	2,440
Basketball	245,359	61	1,467
Business	223,148	70	1,876
Energy	125,130	40	1,621
Fashion	159,819	40	1,227
Food	159,615	45	1,775
Football	359,393	89	1,524
Formula One	143,197	42	1,757
Health	209,941	60	2,542
Irish Politics	170,000	50	2,318
Movies	139,337	38	1,395
Music	208,838	56	1,539
NFL	255,554	80	1,388
Rugby	265,123	76	2,264
Space	127,280	51	2,157
Tech	250,486	66	1,947
Tennis	139,067	41	1,427
UK Politics	245,651	77	3,182
US Politics	332,766	103	4,503
Weather	224,037	65	2,373

dataset, which consists of tweets from 1,200 user accounts corresponding to 20 different distinct ground-truth categories, as can be seen in Table 1. These different categories were manually identified by leveraging community-maintained lists of high-profile users who predominantly tweet about a single topic, such as fashion or music. Therefore, each user is assigned to a single category. Using the Twitter REST API we collected 4,170,382 tweets for these 1,200 “core” users over the period March 2015 to February 2016. In addition, to make the topic modeling task more challenging, we identified a second set of 4,000 users who were randomly selected from among the friends of the core users. These users are not annotated with a ground truth category label, and their content does not necessarily pertain to any of the categories. We collected 16,429,510 tweets for these “friend” users. We randomly divide this second set into blocks of 1,000 users, which allow us to vary the level of noise in our dataset when evaluating topic model solutions.

The full set of tweets was processed as follows. Firstly, all links and user mentions were stripped from the tweet text. Hashtags were kept, but the # prefix was removed. At this point, the tweets for each user for a given week were concatenated into a single “weekly user document”. The justification for this is that individual tweets are short and often contain little textual content that is useful from the perspective of topic modeling. However, by combining multiple tweets from the same user into a single, longer document, we can perform topic modeling more effectively.

After creating these weekly user documents, we apply standard text pre-processing steps:

1. Find all individual tokens in each document, through conversion to lower-case and string tokenization. These tokens include both ordinary words and hashtags.
2. Remove single character tokens, emoticons, and tokens corresponding to generic stop words (*e.g.* “are”, “the”) and Twitter-specific stop words (*e.g.* “rt”, “mt”).
3. Remove documents containing less than 3 tokens.
4. Construct a document-term matrix based on the remaining tokens and documents. Apply TF-IDF term weighting and document length normalization.

The resulting dataset consisted of a total of 40,722 weekly documents for core users and an additional 155,758 documents for friend users.

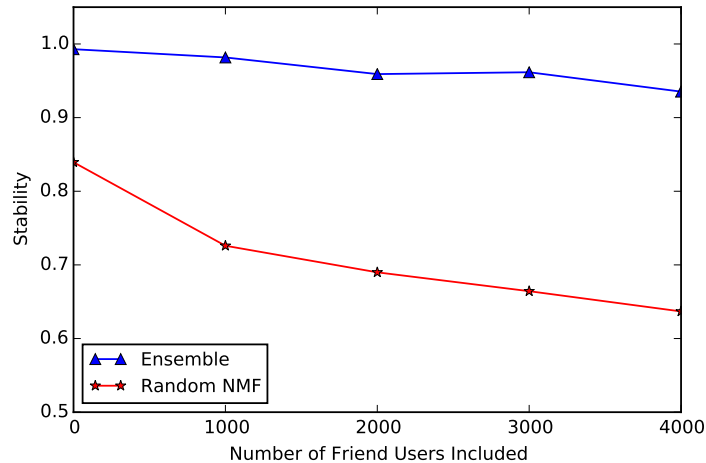
## 4.2 Experimental Setup

To evaluate the proposed method, we generated  $r = 100$  base topic models using NMF with random initialization and combine them as described in Section 3.3. In each case we set the number of base topics ( $k = 20$ ) and the number of ensemble topics ( $k' = 20$ ) to correspond to the number of ground truth categories. We ran this process on the initial set of 1,200 core users, and then repeated the process after including (1000, 2000, 3000, 4000) additional friend users, up to the case where all  $\approx 195k$  weekly documents were included. These friend users were added to evaluate the accuracy and stability of randomly initialized NMF and our ensemble approach with respect to varying levels of noise.

## 4.3 Evaluation of Stability

The goal of our first experiment was to quantify the extent to which instability is a problem with randomly-initialized NMF, and whether an ensemble approach can mitigate this instability. Firstly, we examined the stability between 100 base runs of randomly-initialized NMF to evaluate whether topics become less coherent with varying levels of noise. To do this, we assign each weekly user document to a single topic for which it has the highest weight according to the factor  $H$ , and then measure the agreement between the document assignments for different runs. As a measure of agreement, we use Normalized Mutual Information (NMI), which has previously been used in the evaluation of ensemble clusterings [16]. A pair of topic models that are identical will achieve a NMI score of 1.0 (*i.e.* high stability), while a pair with little agreement will achieve a lower score (*i.e.* low stability). We compute an overall stability score by calculating the NMI between all pairs of models for a given number of friend users and calculating the mean of these values.

We calculated the NMI score for each unique pair of topic model outputs. To evaluate the stability of randomly-initialized NMF with respect to varying



**Fig. 2.** Comparison of stability for randomly-initialized NMF and ensemble topic modeling, based on mean pairwise NMI agreement, for increasing numbers of friends users.

levels of noise, this was repeated while adding weekly summary documents from the friend user set. To vary the level of noise added these were added in blocks of 1,000 at a time, up to 4,000 friend users. Fig. 2 shows the stability scores for randomly-initialized NMF for each case. It is clear that as the level of background noise increases, we see a greater variation in the outputs produced by NMF, as it becomes more challenging to identify a definitive solution.

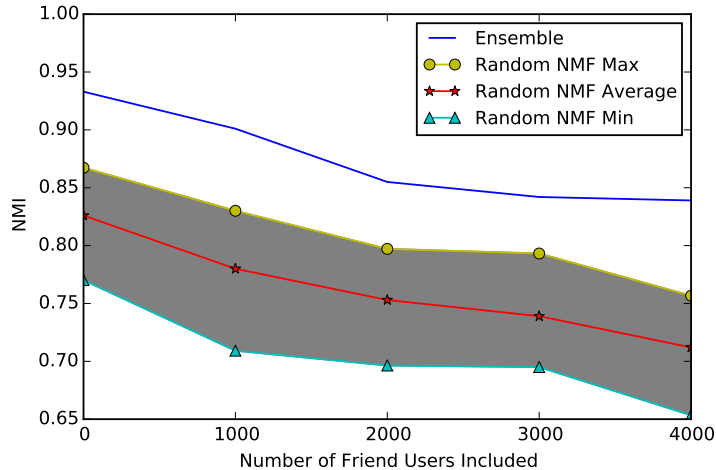
To provide some context as to what this instability means in practice, Table 2 shows an example of descriptors for a topic relating to UK politics, as they appear in five different runs of NMF. While each case does appear to be related to politics, we see variation in the composition and ordering of the top-ranked terms, with terms such as “Cameron” and “tax” appearing intermittently.

To determine whether our proposed approach can address this problem, we generated 10 ensemble topic models, each comprised of 100 different base topic models initialized with different random seeds. Again we compute the mean

**Table 2.** Example of instability between 5 different runs of randomly-initialised NMF, for topics relating to UK politics.

Run	Top 10 Terms
1	labour, tories, tory, nhs, people, cameron, uk, party, mp, support
2	labour, people, ge16, tories, vote, support, tory, party, government, nhs
3	labour, tories, uk, tory, nhs, people, cameron, tax, mp, party
4	labour, people, ge16, tories, vote, tory, support, party, government, nhs
5	labour, people, ge16, tories, uk, government, vote, support, govt, tory





**Fig. 3.** Comparison of NMI accuracy for randomly-initialized NMF and ensemble topic modeling, for increasing numbers of friends users.

pairwise agreement between the document assignments for all runs. We see from Fig. 2 that outputs from the ensemble method produces a much more stable solution, even when increasing the level of noise in the data. The stability scores for the ensemble approach have quite a small variation, ranging from 0.9929 to 0.9353 while the scores for randomly initialized NMF vary much more, ranging from 0.8394 to 0.6368. Our ensemble approach manages to produce a definitive topic modeling solution which crucially can be replicated across different runs.

#### 4.4 Evaluation of Accuracy

While stability is an important requirement, we also need to ensure that we can produce a topic model which accurately summarizes the contents of the corpus. Specifically, we now focus on whether combining a base set of unstable topic models using our ensemble method produces an accurate result relative to the ground truth annotations in the *20-topics* corpus. Firstly, we can manually inspect the topic descriptors generated by applying ensemble topic modeling. Table 3 shows the descriptors for the case where ensemble topic modeling is applied to the set of 1,200 users, along with a manually selected label corresponding to the most similar ground truth category. We see that 18 out of 20 ground truth categories are clearly identified, with two categories (‘Irish politics’ and ‘football’) replaced by two extra topics relating to ‘energy’ and ‘technology’. In general we observed that, across all experiments on this corpus, the ‘Irish politics’ topic consistently overlapped with the ‘UK politics’ topic, while the ‘football’ topic frequently overlapped with the ‘NFL’ topic. This is perhaps unsurprising given the partially shared vocabulary in both cases.

**Table 3.** Topic descriptors for 20 topics generated by applying ensemble topic modeling to the *20-topics* corpus, using tweets from 1,200 core users. The most similar ground truth category for each topic is also listed.

Category	Top 10 Terms
Energy 1	fracking, shale, gas, energy, natgas, natural, naturalgas, pa, epa, emissions
US Politics	gopdebate, president, gop, obama, senate, clinton, bill, hillary, trump, congress
Rugby	rugby, rwc2015, england, cup, wales, ireland, world, try, match, rbs6nations
NFL	game, nfl, season, win, patriots, team, league, football, tonight, goal
Tech 1	apple, watch, applewatch, google, app, music, tv, ios, facebook, macbook
UK Politics	labour, ge16, people, tories, vote, tory, party, government, nhs, support
Basketball	bulls, rose, butler, hoiberg, gasol, nba, game, noah, jimmy, pau
Weather	rain, snow, weather, forecast, showers, storm, tornado, severe, dry, winds
Business	china, stocks, market, fed, markets, stock, growth, tech, uk, ftse
Health	health, cancer, study, risk, patients, care, diabetes, zika, drug, disease
Music	album, music, video, listen, song, track, remix, tour, premiere, check
Aviation	avgeek, aviation, boeing, flight, airlines, air, aircraft, airbus, airport, paxex
Tech 2	iphone, ios, ipad, mac, apple, app, os, apps, beta, plus
Fashion	fashion, daily, nyfw, stories, style, collection, dress, wear, beauty, show
Food	recipes, recipe, food, chicken, best, dinner, delicious, chocolate, chef, restaurant
Formula One	f1, race, ferrari, hamilton, mclaren, mercedes, renault, rosberg, gp, bull
Movies	film, review, movie, trailer, star, wars, movies, films, awakens, oscars
Tennis	tennis, atp, murray, djokovic, federer, serena, nadal, wimbledon, ausopen, wta
Space	space, yearinspace, pluto, earth, nasa, mars, mission, launch, journeytomars, science
Energy 2	oil, energy, gas, crude, prices, opec, offshore, production, exports, oilandgas

To quantitatively evaluate accuracy, we can use NMI to measure the degree to which document assignments from a topic model agree with the ground truth categories listed in Table 1. Again we consider the case where increasing numbers of noisy documents from friend users are added to the data. Note that, while we add friend users we only consider the document-topic assignments for our set of ‘core’ users when calculating the NMI score.

Based on 100 runs of randomly-initialized NMF, Fig. 3 shows the mean, minimum, and maximum NMI scores. We can make two observations based on these

results. Firstly, the mean accuracy of the topic models decreases considerably as more friend users are added. Secondly, there is considerable variation in accuracy across the 100 runs, due to random initialization. In contrast, Fig. 3 shows that ensemble topic modeling achieves a level of accuracy above the accuracy maximum for the ensemble members from which it was compromised – in this case the ensemble topic model is “greater than the sum of its parts”. Taking the result in conjunction with the results from Section 4.3, this suggests that the combination of many unstable and diverse base topic models can produce a more accurate topic model. From Fig. 3, we also observe that the decline in NMI as more friend users are added is less pronounced, suggesting that the ensemble method is more robust to noise.

## 5 Conclusions

In this paper we have proposed a new ensemble topic modeling method, based on the combination of multiple matrix factorizations to produce a single ensemble model. We compared its performance to standard NMF on a tweet corpus, in terms of both stability and accuracy. We have observed that the proposed method not only yields a more accurate topic model with respect to document-topic assignments, it also produces a far more stable output, with little variation across multiple runs.

There are a number of future avenues of research which we would like to explore. Firstly, we intend to evaluate the proposed method on a range of other datasets, which consist of not only tweets but other sources of text such as news articles. We would also like to investigate alternative ensemble generation strategies, such as random subsampling of documents and terms, to evaluate if promoting further diversity improves the quality of the ensemble results. We would also like to investigate the number of base topic models required in the ensemble generation phase to generate an accurate and stable solution. Finally, we would be interested in generalizing our ensemble approach to work with other topic modeling algorithms, such as LDA, where instability is also an issue.

**Acknowledgement.** This research was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

## References

1. Aiello, L.M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., Jaimes, A.: Sensing trending topics in Twitter. *IEEE Transactions on Multimedia* 15(6), 1268–1282 (2013)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022 (2003)
3. Boutsidis, C., Gallopoulos, E.: SVD based initialization: A head start for non-negative matrix factorization. *Pattern Recognition* (2008)
4. Breiman, L.: Bagging predictors. *Machine Learning* 24(2), 123–140 (1996)

5. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. *Journal of the American Society of Information Science* 41(6), 391–407 (1990)
6. Fred, A.: Finding consistent clusters in data partitions. In: *Proc. 2nd International Workshop on Multiple Classifier Systems (MCS'01)*. vol. 2096, pp. 309–318. Springer (January 2001)
7. Ghaemi, R., Sulaiman, M., Ibrahim, H., Mustapha, N.: A Survey: Clustering Ensembles Techniques. In: *Proceedings of World Academy of Science, Engineering AND Technology*. vol. 38, pp. 2070–3740 (2009)
8. Greene, D., O'Callaghan, D., Cunningham, P.: How Many Topics? Stability Analysis for Topic Models. In: *Proc. European Conference on Machine Learning (ECML'14)*. pp. 498–513. Springer (2014)
9. Greene, D., Cagney, G., Krogan, N., Cunningham, P.: Ensemble Non-negative Matrix Factorization Methods for Clustering Protein-Protein Interactions. *Bioinformatics* 24(15), 1722–1728 (2008)
10. Greene, D., Cross, J.P.: Exploring the political agenda of the european parliament using a dynamic topic modelling approach. In: *5th Annual General Conference of the European Political Science Association (EPSA'15)* (2015)
11. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–91 (1999)
12. Lin, C.: Projected gradient methods for non-negative matrix factorization. *Neural Computation* 19(10), 2756–2779 (2007)
13. Opitz, D.W., Shavlik, J.W.: Generating accurate and diverse members of a neural-network ensemble. *Neural Information Processing Systems* 8, 535–541 (1996)
14. Punera, K., Ghosh, J.: *Soft Cluster Ensembles*. In: *Advances in Fuzzy Clustering and Its Applications*. Wiley (2007)
15. Steyvers, M., Griffiths, T.: *Latent Semantic Analysis: A Road to Meaning*, chap. Probabilistic topic models. Laurence Erlbaum (2007)
16. Strehl, A., Ghosh, J.: Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3, 583–617 (December 2002)
17. Strehl, A., Ghosh, J.: Cluster ensembles - a knowledge reuse framework for combining partitionings. In: *Proc. Conference on Artificial Intelligence (AAAI'02)*. pp. 93–98. AAAI/MIT Press (July 2002)
18. Topchy, A., Jain, A., Punch, W.: Clustering ensembles: Models of consensus and weak partitions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(12), 1866–1881 (December 2005)
19. Wang, Q., Cao, Z., Xu, J., Li, H.: Group matrix factorization for scalable topic modeling. In: *Proc. 35th SIGIR Conf. on Research and Development in Information Retrieval*. pp. 375–384. ACM (2012)