Heinz 95-845: Predicting False News on Twitter Applied Analytics: the Machine Learning Pipeline

Nicolas Banholzer NBANHOLZ@ANDREW.CMU.EDU
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Jasper Meijering JMEIJERI@ANDREW.CMU.EDU
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Abstract

The spread of false news on social media platforms continues to concern many. Various methods have been applied to classify veracity of twitter content. Assessing the veracity of tweets based on measures of information diffusion using a hybrid neural network is a novel approach to distinguish false and true news. The model has been trained and tested on 42,081 tweet cascades. The results show that it is possible to predict veracity based on measures of information diffusion to some extent. Yet, further research is advised to increase the model's performance.

1. Introduction

Following the 2016 US presidential election, many have expressed concern about the effects of false stories, circulated largely through social media (Allcott and Gentzkow, 2017). Within the domain of social media, there are three major areas where credibility assessment is important: "(1) the detection of opinion spam in review sites, (2) the detection of fake news and spam in microblogging, and (3) the credibility assessment of online health information" (Viviani and Pasi, 2017). Lazer et al. (2018) propose two ways to counteract the second point. First, empowering individuals in critical evaluation of content. Second, platform-based automated evaluation of the quality of content through algorithms, followed by false news detection and possibly prevention. While the first proposition probably requires long-term improvements in the educational system, the second proposition might be in closer reach, particularly when considering the data mining capabilities of Google, Facebook, Twitter and the like to predict user-relevant content (Lazer et al., 2018). Also, primarily social media platforms have strong incentives to check the veracity of content as they are especially susceptible to false news because they allow easy and cheap distribution of news (Shu et al., 2017) and because internal information on social media diffuses much faster than information coming from external sources (Yoo et al., 2016).

Multiple scholars including Alrubaian et al. (2017) developed methods to predict veracity of twitter content based on reputation-based judgments, credibility classifier engines, user experience components, and models that analyze the text shared in the tweets. To the best of our knowledge, assessing the veracity of tweets based on measures of information

diffusion using a neural network is a novel approach to distinguish false and true news. The research question of this study is:

Can veracity of news shared on Twitter be predicted based on measures of information diffusion?

2. Background

Shu et al. (2017) give an overview of features and data mining techniques that can help analyze news and predicting their veracity. They distinguish between news content and social context features. News content features describe the linguistic or visual style of the content and can for example inform on how false news are written. Indeed, Vosoughi et al. (2018) show for a smaller sample of their data that false news exhibit greater novelty. However, the main focus of their analysis is on how false news diffuse compared to true news. Since their analysis is mainly descriptive, we build on their work by predicting the veracity of news based on features measuring information diffusion and controlling for features of social context.

3. Dataset

The data by Vosoughi et al. (2018) consists of 126,000 tweet cascades in 4.5 million tweets by 3 million people between 2006 and 2017. A cascade can be thought of as a tree. At the root is the initial tweet which spreads through the network when being retweeted by the followers. Cascades can be about the same story.The veracity of the stories is independently checked by 6 fact-checking organizations and classified as either false, mixed or true.
Since the dataset covers both static and dynamic measures, a hybrid modeling approach is used in this study to predict the veracity of news.

4. A Hybrid Neural Network for Grouped Dynamic and Static Data

Figure 1 illustrates the model architecture chosen for this study. We follow a hybrid approach that essentially consists of two steps. In the first step, we model the dynamic part of the data with two Long Short-Term Memory (LSTM) neural networks. In the second step, we combine the predictions of the two LSTMs with the static part of the data into an overall neural network.

The modeling approach is motivated by the structure the dataset by Vosoughi et al. (2018). It consists of both dynamic and static data, which are displayed in Table 1. The reader interested in the specifics of the dataset is referred to Vosoughi et al. (2018). We highlight the two points relevant to our model architecture.
First, note that information diffusion can be measured over the whole cascade (static) or over time (dynamic). In a simple model, one might only attempt to predict the veracity of news based on the static measures. However, we believe that the way the measures change over time might contain valuable information for prediction. Second, attempting to model the dynamic and static data with a single large model is challenging as both the sequence

Figure 1: The model consists of two LSTM networks applied to the dynamic measures grouped by sequence length. The output of the LSTMs is used as input for the overall model, which concatenates one input layer for continuous and one input layer for categorical variables. For each cascade, the combined model outputs a vector containing the softmax probabilities.

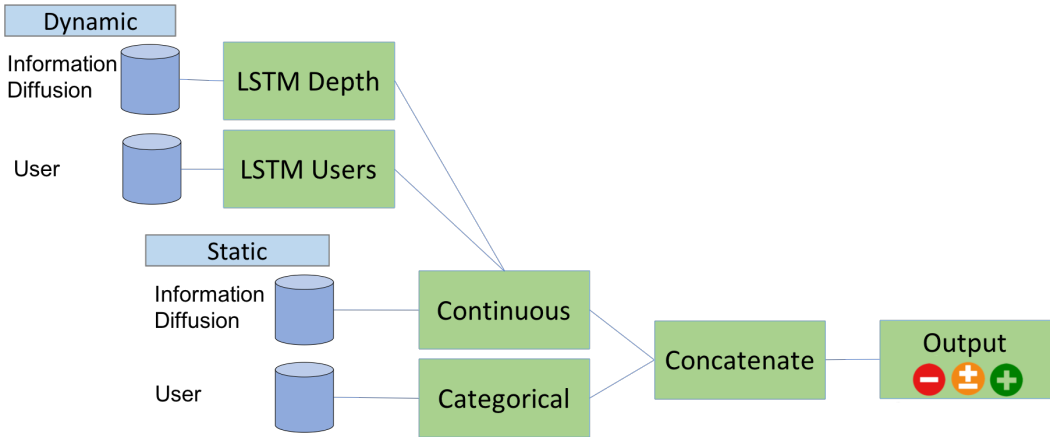


Table 1: The dataset by Vosoughi et al. (2018) consists of both dynamic and static measures that can be grouped by user-specific measures and measures of information diffusion. This table shows all measures that were included in our predictive study.

Static		Dynamic	
User	**Information Diff.**	**User**	**Information Diff.**
No. of followers	Breadth	Unique users to time	Depth to time
No. of followees	Depth		Depth to unique users
Engagement	Size		Depth to breadth
Verification	Virality		
Rumor Category			

length within dynamic and between dynamic measures varies.

Therefore, in order to exploit the full set of information about each cascade while maintaining a decent level of model complexity, we choose a hybrid approach that is explained in more detail in the following section.

5. Experimental Setup

Not every news spreads. A large fraction of the dataset consists of cascades of length zero, which means that the initial tweet is not retweeted at all. From our perspective, a rumor without audience is not really a rumor. For that reason we only select only tweets cascades

of length greater than zero. This already alters the outcome distribution as shown in 2. Apparently, relatively more true or mixed news are not retweeted at all compared to false news.

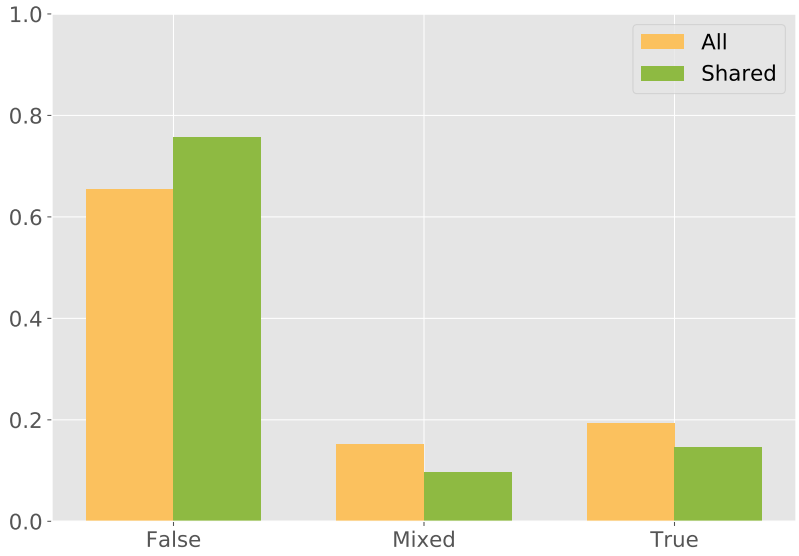


Figure 2: Distribution of fake, mixed and true news before (All) and after sub-setting the dataset (Shared).

The experimental setup consists of the following steps:

1. **Generate the static data:** After sub-setting the data, there are in total 42,081 cascades. 75.7% of the cascades spread false news, 9.7 % spread mixed and 14.6 % spread true news. Number of followers, number of followees and verification respectively have 136 missing values that are each conditionally imputed with a separate model using the so called MICE (Multivariate Imputation by Chained Equations) package.[1] Table 5 presents descriptive statics for the static measures of the dataset.

2. **Generate the dynamic depth data:** Variables depth to time, depth to unique users and depth to breadth have the same sequence length for each cascade and form the first set of dynamic measures. However, the length of the sequence varies from cascade to cascade. To put the LSTM at work, we therefore perform a so-called bucketing approach. Thereby, the depth data is grouped by sequence length (each group is referred to as a "bucket"). Figure 3 illustrates the bucketing approach and

1. Regarding the intuition of conditional imputation in this setting, assume that a user with a higher number of followers is probably more likely to spread news deeper than a user with fewer followers. Hence, number of followers is possibly imputed conditional on the information diffusion measures.

Figure 4 shows the bucket size by sequence length for the dynamic depth data. For each bucket the data is split 50/50 into a training and test set. The LSTM model is then fitted on the train set and predictions are obtained for both training and test set.[2]

3. **Generate the dynamic user data:** Analogously to the dynamic depth data, the bucketing approach is applied to the set of dynamic user measures, which only consists of the measure unique users to time. However, the train/test split is defined by the dynamic depth dataset. Therefore, contrary to the depth data, for the user data, the data is first split and then bucketed.
   Furthermore, two issues are addressed within the bucketing approach. First issue, the sequence length in the dynamic user data varies considerably so that buckets could be of very small size. Second issue, the outcome distribution is highly unbalanced, i.e. there are generally more cascades of false news than mixed or true news. For better training performance, we therefore apply an iterative approach where a new group is only created if it consists of more than 100 cascades and the training set must at least have a ratio of 20/80 of mixed and true to false news. Otherwise, the last n time steps of the sequence are pruned.
   Figure 5 shows the bucket size by sequence length for the training and test set of the user data.

4. **LSTM model predictions:** For both dynamic depth and user data, we use a neural network with a single LSTM layer and one output layer at the end of the sequence. The predicted softmax probabilities for the three outcomes in the training and test set are then appended to the static dataset.

5. **Combined model:** For the combined model, we concatenate two input layers, one for the continuous and one for the categorical variables. We added two dense layers with a ReLu activation function and a dropout layer with a learning rate of 0.2.

   [3] `More information about the architecture`

6. Results

To predict veracity of news we trained and tested three versions of the model on a dataset of 42,081 cascades, of which 75.7 % spread false false news, i.e. the baseline for evaluating model accuracy.

1. **Static Model:** Only considers the static measures.

2. **Static + LSTM Depth:** In addition to the static measures, also considers the predictions of the LSTM using the dynamic depth measures.

2. In addition, note that to run the LSTM, each group has to be a multiple of the batch size. Therefore, each group was padded by randomly sampling observations from the group.
3. Note that for both LSTM and the combined model, all continuous input variables are first standardized before running the model.

Figure 3: For each set of dynamic measures the cascades are grouped by sequence length. Within each bucket the data is split into a training and test set and the same model is fitted on the training data and predictions are obtained for both training and test data.

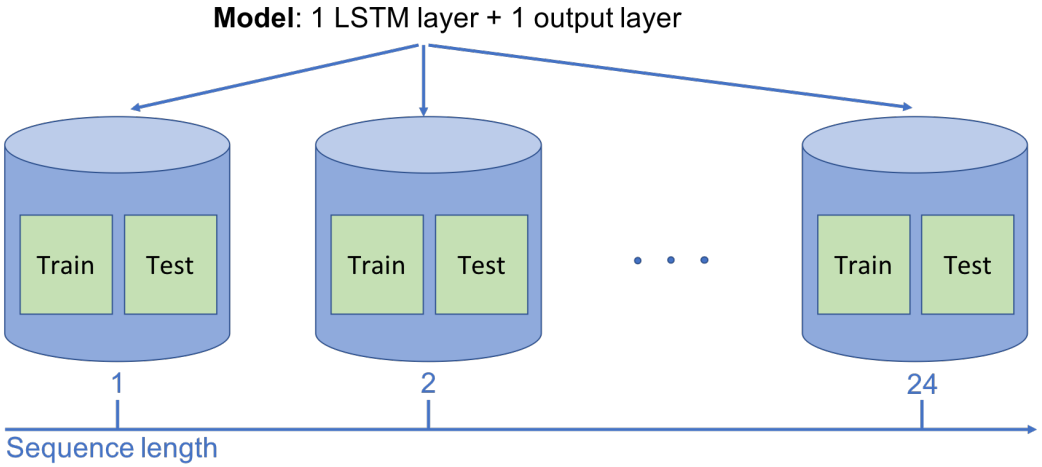


Figure 4: The dynamic depth data is split into six buckets, whereby the bucket size is inversely related to the sequence length.

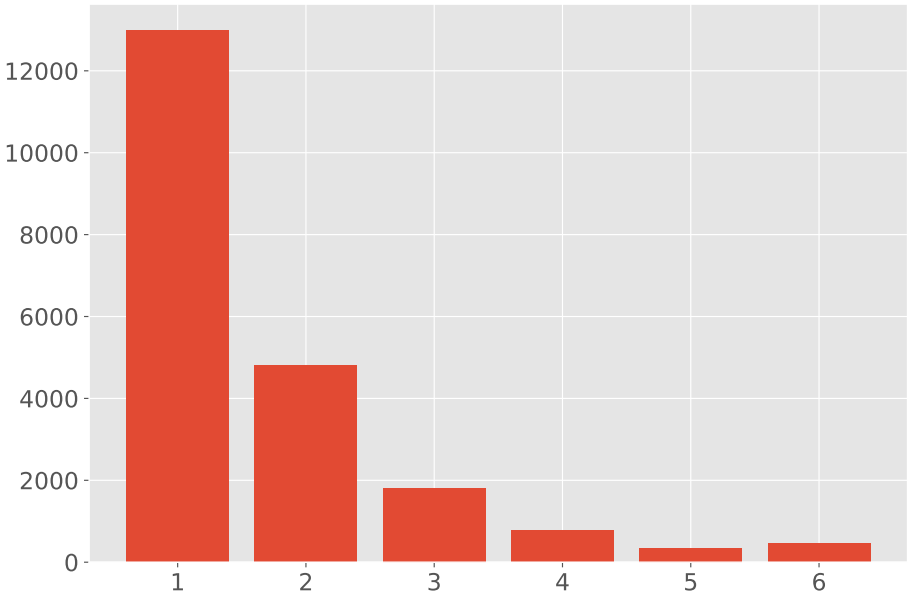
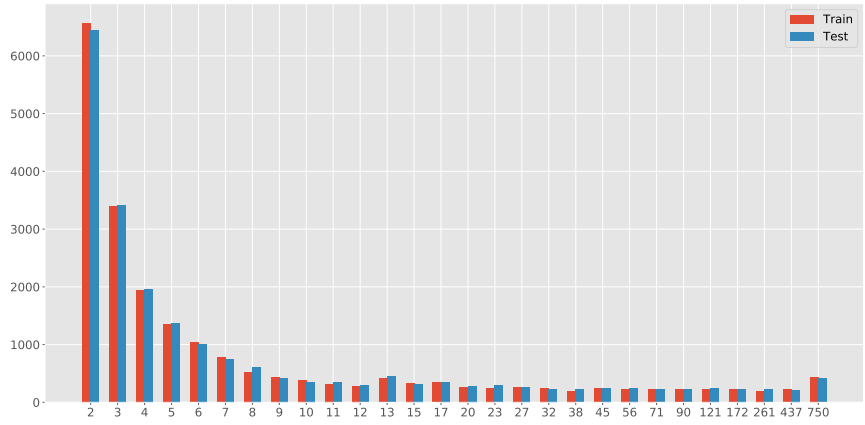


Figure 5: The dynamic user data is split into 29 buckets, whereby the bucket size inversely related to the sequence length.



3. **Static + LSTM Depth + LSTM User:** In addition to the static and dynamic depth measures, also considers the predictions of the LSTM using the dynamic user measure.

All three models yield comparable results with accuracy ranging between 78% and 79%, with the static model performing slightly better than the models incorporating the dynamic measures. That the LSTM predictions do not improve predictive performance could have several reasons. First, the length of sequences are mostly very short. When evaluating the model fit, the loss is lower for buckets with greater sequence lengths, i.e. more information. Second, the static model already incorporates information diffusion measures. So possibly, the aggregation of the information diffusion measures already captures all relevant informa-

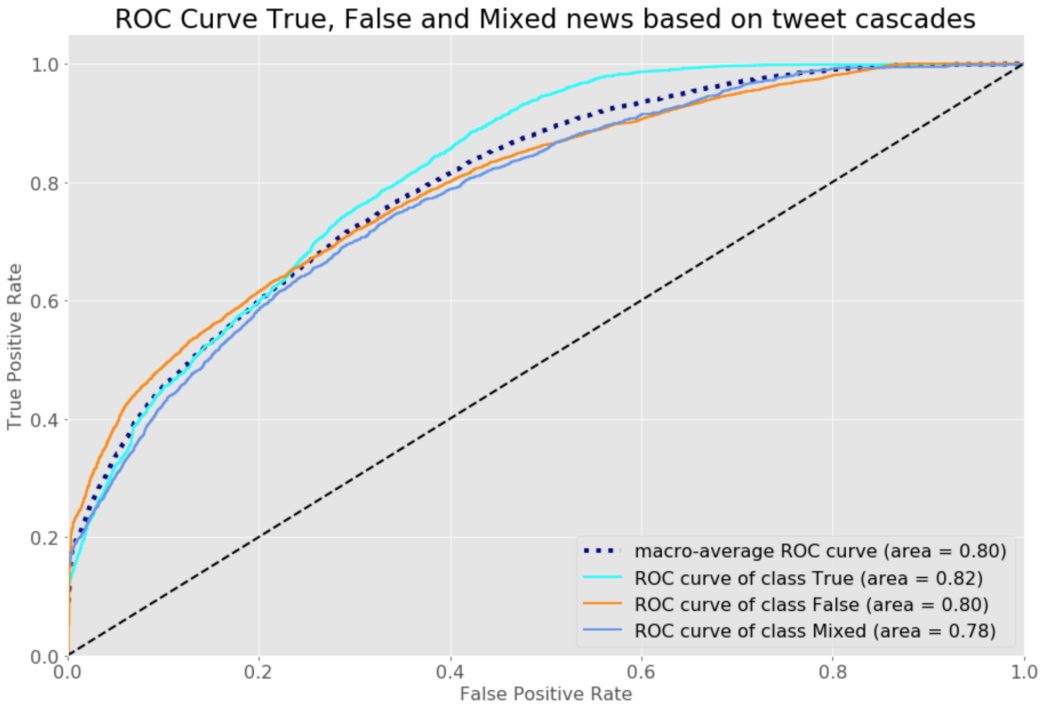
Table 2: Descriptives for the static measures. Regarding the categorical data, note that most users are not verified and that the dominant rumor category is Politics.

	breadth	**depth**	**engagement**	**nfollowees**	**nfollowers**	**size**	**virality**
mean	59.4	1.7	77.6	7287.9	59055.7	93.9	1.6
std	527.5	1.3	121.2	20441.7	597136.1	950.7	0.7
min	2	1	0	0	-61193	2	1
25%	2	1	6.1	470	676	2	1
50%	3	1	21.6	1660	2846	4	1.5
75%	8	2	80.3	7397	19417	9	2
max	29527	24	1248.3	1303465	55776569	46895	10.2

tion for predicting false news.

Figure 6 shows the ROC curve of the static model. The average area under the ROC curve (AUC) is 0.80. A comparison of the AUCs indicates that the model is generally better at distinguishing true and false news compared to mixed news. Overall however, there is only a slight improvement over the baseline of about 4%.

Figure 6: The ROC curve of the model. The area under the curve for predicting true news is higher than the AUC for false and mixed news.



7. Conclusion

Summary

Vosoughi et al. (2018) find that false news spread faster, farther and deeper than true news. Based on their finding, we show that measures of information diffusion in combination with user characteristics can predict the veracity of news to some extent. However, our results show that the dynamic nature of information diffusion does not seem to provide additional predictive power.

Limitations and further research

There are limitations to this study which we leave to further research.

- We eliminate tweets that do not spread. However, these tweets may still capture a small audience, i.e. the immediate followers of the initiator of the cascade. Also from

a practical perspective, a user that sees a new tweet might already be interested in its veracity. Our model could only provide a probability for news that spread.

- Although our hybrid modeling approach incorporates information contained in the dynamic measures, most dynamic measures of a cascade have only a very short sequence length. This limits the benefit and applicability of the LSTM models. Further research might therefore consider to engineer additional sequential features. For example, the raw data of Vosoughi et al. (2018) also provides the time stamp of each tweet. This could be used to compute the reaction time to each parent tweet in a cascade.

- The main contribution of this study is to provide a model architecture to predict false news on Twitter based on user characteristics and information diffusion measures. Further research should consider parameter tuning, cross-validation and robustness checks to enhance and validate the modeling approach. In addition, rather than using the full cascade, the model could be refitted on a pruned cascade as part of a sensitivity analysis. From a practical perspective, it is also interesting to find out at which point in the cascade the veracity of the news can be predicted very reliably.

- Vosoughi et al. (2018) describes the spread of false news on Twitter and we attempt to predict false news on Twitter based on their descriptive findings. Yet, a strong predictor of whether news is false or true, is the content of the news itself. Therefore, any prediction might see significant improvement from

Practical Implications

When predicting the veracity of news, most attention is paid to the content and sentiment of news. Vosoughi et al. (2018) find that false news spread faster, farther and deeper than true news. Based on their finding, we show that measures of information diffusion in combination with user characteristics predict the veracity of news to some extent. Social networks that like to prevent the spread of false news or inform their users about it, may use these findings jointly with findings on content and sentiment to develop an algorithm that assesses the veracity of news. Eventually, the algorithm might result in a tool that can enable social media users to automatically check the veracity of the news appearing in their social network.

References

Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, May 2017. doi: 10.1257/jep.31.2.211. URL http://www.aeaweb.org/articles?id=10.1257/jep.31.2.211.

M. Alrubaian, M. Al-Qurishi, M. Hassan, and A. Alamri. A credibility analysis system for assessing information on twitter. *IEEE Transactions on Dependable and Secure Computing*, pages 1–1, 2017. ISSN 1545-5971. doi: 10.1109/TDSC.2016.2602338.

David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018. ISSN 0036-8075. doi: 10.1126/science.aao2998. URL http://science.sciencemag.org/content/359/6380/1094.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.*, 19(1):22–36, September 2017. ISSN 1931-0145. doi: 10.1145/3137597.3137600. URL http://doi.acm.org/10.1145/3137597.3137600.

Marco Viviani and Gabriella Pasi. Credibility in social media: Opinions, news, and health information – a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5):e1209, 2017. doi: 10.1002/widm.1209. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1209.

Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018. ISSN 0036-8075. doi: 10.1126/science.aap9559. URL http://science.sciencemag.org/content/359/6380/1146.

Eunae Yoo, William Rand, Mahyar Eftekhar, and Elliot Rabinovich. Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. *Journal of Operations Management*, 45:123 – 133, 2016. ISSN 0272-6963. doi: https://doi.org/10.1016/j.jom.2016.05.007. URL http://www.sciencedirect.com/science/article/pii/S0272696316300419. Special Issue on Humanitarian Operations Management.