

Network Motifs May Improve Quality Assessment of Text Documents

Thomas Arnold and Karsten Weihe

Research Training Group AIPHES / Algorithmic Group

Department of Computer Science, Technische Universität Darmstadt

Hochschulstrasse 10, 64289 Darmstadt, Germany

arnold@aiphes.tu-darmstadt.de, weihe@cs.tu-darmstadt.de

Abstract

Motif analysis counts the number of small building blocks (the *motifs*) in a network and relates these statistical numbers to the inherent semantics of the network. In the realm of natural language processing, the networks are induced by texts. We demonstrate that motif analysis may help assess the quality of a document. More specifically, we consider the German Wikipedia and use the label “featured” as the (binary) quality criterion. The length (number of words) of an article is a comparatively good predictor for this label. We show that a well-designed combination of this criterion and motif statistics yields a significant improvement. We also found that a deeper look into the most relevant motifs may improve our understanding of quality.

1 Introduction

Recent work has shown that *motif analysis* is quite promising for natural language processing (Biemann et al., 2012; Mesgar and Strube, 2015). Roughly speaking, the occurrences of small subgraphs like those in Figure 1 are counted, and relations between this *motif signature* and the semantics of the network are analyzed. We adapt this technique to the assessment of document quality. Our computational study is based on the German Wikipedia. The label “featured” indicates high quality. A certain combination with the length of an article, which is a very good predictor for an article’s quality in this specific sense (Blumenstock, 2008), yields a significant improvement over the length alone.

This paper is organized as follows. First we briefly sketch our contribution in Section 1.1. Then, in Section 2, we discuss the state of the art and explain the concept of motif analysis. We illustrate the composition of our corpus in Section 3 and details of our approach in Section 4. Quantitative and qualitative results are presented in Section 5 and 6, respectively. Finally, we summarize our work and outline next steps in Section 7.

1.1 Our Contribution

The research questions In this paper, we address three general questions:

1. *Quantitative*: does motif analysis as a stand-alone tool help us assess the quality of text documents statistically?
2. *Quantitative*: does it help us in conjunction with other quality measures?
3. *Qualitative*: does it help us understand the nature of quality any better?

The German Wikipedia is our basis. The Wikipedia allows the community to assign the label “featured” to an article via an extensive communication and revision process, based on a collection of stylistic and content-based quality criteria. We use the distinction featured / non-featured as a (binary) quality criterion. Our corpus comprises all featured articles and a purely random selection of non-featured articles such that 7% of all articles in our corpus are featured.

In summary, we address the three general research questions in the following specific form: does motif

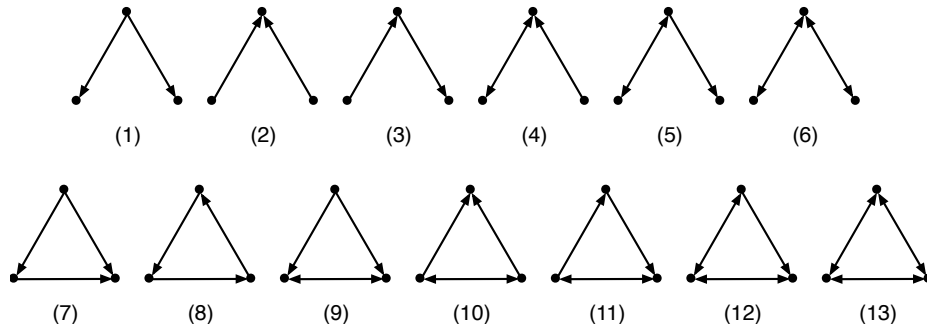


Figure 1: the directed motifs on three nodes. A double arrow indicates the presence of two mutually opposite arcs.

analysis help us—alone or in conjunction with another criterion—to distinguish between featured and non-featured articles, and if so, does it yield a deeper understanding of the nature of featured articles?

The bare length of an article is a surprisingly good predictor for whether or not an article is featured. So, for the second research question, we will combine the article length with our motif analysis.

Methodological contribution The revision history of a Wikipedia article allows us to analyze the development of the induced network and its motif signature over time. So, for each article we analyzed a series of “snapshots” and the temporal tendency of the motif signature.

2 Related Work and Background

2.1 Related Work

Network analysis of text documents There is a large scientific body of methods and applications of network analysis (Aggarwal and Wang, 2010; Aggarwal, 2011). Graph mining—the art of detecting and analyzing patterns and structures in graphs—is the specific focus of the surveys (Cook and Holder, 2006; Fortunato, 2010).

It seems reasonable to classify network analysis techniques by the level of granularity they address. Elementary statistical measures such as the node degree distribution operate on the level of single nodes and edges. In the opposite extreme case, on the global level, the structure of a network is captured in a single (scalar) numerical value. Examples for global measures are the average shortest path length, the diameter, as well as simple characteristics such as node and edge count. See the above-mentioned surveys for a systematic discussion.

Motif analysis Motifs such as those in Figure 1 capture local structure and are thus, in a sense, on an intermediate level between measures on single nodes and edges on one hand and global measures on the other hand.

Motif analysis has first been investigated in computational biology (Shen-Orr et al., 2002) and has since been applied to a variety of network types in biology and biochemistry (Schreiber and Schwöbbermeyer, 2010). The underlying insight is that biological and biochemical dynamics are statistically related to the occurrence of small functional blocks, which have specific structures. This insight is well captured by motif signatures, and in fact, many computational studies reveal significant relations. Due to this success, it did not take long time until this technique has been applied to networks from other domains. For example, (Milo et al., 2002; Milo et al., 2004) compare networks from biology, electrical engineering, natural language and computer science and find that the motif signatures from different domains are so different that they may serve as a “fingerprint” of the respective domain. Krumov et al. (2011) use motif analysis on co-authorship networks to find relations between particular motifs and citation frequency. They reveal one particular motif that implies high average citation frequency, and provide explanations based on social processes that are covered by the network.

Tran et al. (2015) explored differences in directed and undirected networks of various disciplines, including ecology, biology and social science. They conclude that motifs in undirected networks are very similar. However, motif analysis of directed networks was able to distinguish networks from different fields. Furthermore, larger motifs captured more

information about individual differences than small motifs.

Quite recently, motif analysis has been applied to text processing. In (Biemann et al., 2012), human-written texts and artificial texts with quite similar characteristics were compared by means of the motif signatures of certain induced networks. For several natural languages, the motif signatures were so different that they alone were sufficient to distinguish the human-written from the artificial texts.

Mesgar and Strube (2015) applied motif analysis to Wall Street Journal news articles. These texts were represented as a combined entity and discourse relation graph. They identified several motifs that are highly correlated with manual readability ratings.

Quality of documents Defining and measuring the quality of a document in a formalized way is an intrinsically difficult task. Various mathematical measures have been proposed for individual aspects of quality, like correct grammar (Tetreault and Chodorow, 2008) or spelling (Brill and Moore, 2000). Another part of quality, information ordering, has been evaluated with rank correlation metrics (Lapata, 2006). Louis and Nenkova (2013) investigated text quality at a much broader level that incorporates features of emotions and surprise.

A network based approach on quality assessment of school essays was presented by (Antiqueira et al., 2007). This research used global statistical network features of text-induced graphs, including mean clustering coefficient and average shortest path length. The authors presented correlations between these metrics and manually annotated quality scores.

2.2 Background: Networks and Motifs

Graphs vs. networks All graphs in this paper are directed. Basically, we use the terms *graph* and *network* synonymously. However, we will speak of *networks* only if the nodes and edges have a meaning in the application context.

More specifically, in our work, the nodes of a network are the sentences of a document, and two nodes are connected by an edge if, and only if, the two represented sentences have at least one noun in common and are separated by at most two other sen-

tences. The direction of the edge reflects the relative order of the sentences in the document, from the earlier to the later sentence. Such a definition of directed edges, which combines content with locality, has turned out to be quite promising.

Motif analysis A *motif* is simply a small connected graph—typically, but not exclusively, with three or four nodes. Figure 1 shows all directed motifs on three nodes.¹

For a *motif analysis* of a set of networks, a set of pairwise non-isomorphic motifs is to be firmly selected a priori. To analyze a network N , the number of occurrences of each motif in N is counted. In that, an *occurrence* of a motif M in N is a set X of nodes of N such that the subgraph of N induced by X is isomorphic to M .² Since no two motifs are isomorphic, each set of nodes of N can be the occurrence of at most one motif. And in fact, even if a motif bears some symmetries such as (1), (2), (6), (8), (9), (10), and (13) in Figure 1, the underlying node set is counted as *one* occurrence only.

The *motif signature* of a network N with respect to a selection of motifs comprises the number of occurrences of every selected motif in N , scaled to a sum of 1.

Generally speaking, *motif analysis* relates the structure of a network to the investigated semantic properties of the network, which are induced by the application context. There is a quantitative and a qualitative side to this. The *quantitative* analysis identifies statistical relations between the motif signature and the semantic properties. On the other hand, the *qualitative* analysis is based on the idea that the motifs – if well selected – may be interpreted as building blocks of the network, and that an unexpectedly high or low frequency of some motifs may yield a deeper understanding of these semantic properties.

¹Unfortunately, the terminology is not consistent throughout the scientific literature: in many articles, only the selected subgraphs with unexpectedly high frequency in a network are called *motifs* of the networks', and the subgraphs with unexpectedly low frequency are then called *anti-motifs*.

²Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two directed graphs such that $|V_1| \leq |V_2|$, and let $V'_2 \subseteq V_2$ such that $|V'_2| = |V_1|$. Then G_1 is the *subgraph of G_2 induced by V'_2* if there is a bijection $\varphi : V_1 \rightarrow V'_2$ such that for all $x, y \in V_1$, it is $(x, y) \in E_1$ if and only if $(\varphi(x), \varphi(y)) \in E_2$.

3 Data

Our corpus is a subset of the German Wikipedia. We utilize the quality label “featured” as an indicator for high quality articles. Therefore, we include all 2,338 featured articles of a complete snapshot of the German Wikipedia from June 2015. The proportion of featured articles in this snapshot is 0.13%. Adding all non-featured articles would result in a dataset of very large proportions and with an extremely skewed distribution of the relevant “featured” label. This is a huge problem for most machine learning techniques. We deal with this problem by under-sampling the overrepresented class (Drummond et al., 2003). For this reason, we restricted the set of all non-featured articles to a purely random – and thus representative – sample of 33,295 articles, which increases the share of featured articles in our corpus to 7.02%.

For each article in our corpus, we selected 10 distinct article versions from the article’s revision history. A new version of an article is created every time submitted revisions or additions to this article are approved by the community. For every featured article, we split all of its article versions into a set of versions with the featured label and a set of version without the featured label. On average, this divides the versions of a featured article into about 57% non-featured versions and 43% featured versions. For every featured article, we select five random versions from each part.

We split the versions of non-featured articles similarly into “early” and “late” parts of the revision history with the same proportions as the featured articles’ featured and non-featured parts. This way, we pick five random versions from the earliest 57% versions of a non-featured article, and five random versions from the latest 43% versions.

For every article version, we create a network according to our graph representation, search the network for motifs and compute the corresponding motif signatures, as explained below.

4 Our Approach

Graph representation We use a sentence level graph representation based on shared nouns.

For every selected version of a selected Wikipedia article, we construct a network $G = (V, E)$ with

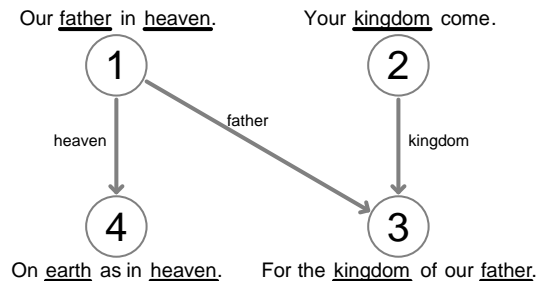


Figure 2: Exemplary graph representation with four consecutive sentences.

a set of nodes V and a set of directed edges E as follows:

The nodes of this network are the sentences of the article version. Two nodes are connected by an edge if and only if these two conditions are fulfilled:

1. There exists at least one noun token that appears in both corresponding sentences.
2. The two sentences are separated by at most two other sentences in the document.

Edges are directed and point from the sentence earlier in the text to the latter one. Figure 2 shows an example of this representation.

Motif analysis To each network created from a Wikipedia article version, we apply a motif analysis. In our case, we search for subgraphs of three or four connected nodes. Furthermore, we only search for motifs of three or four directly consecutive sentences. With this constraint, we can only discover discourse connections of sentences that follow right after each other. Because of their close proximity, we can be pretty sure that these sentences have a strong discourse connection and are likely to share a common topic to some extent.

The resulting motifs are quite rare. If we relaxed this constraint and searched for all connected subgraphs of three or four nodes, the number of occurrences of motifs increases significantly. However, we found that the motif analysis yields worse results.

The way our networks are constructed limits the number of possible subgraphs considerably. The directions of the edges have to follow the order of appearance of the corresponding sentences, which

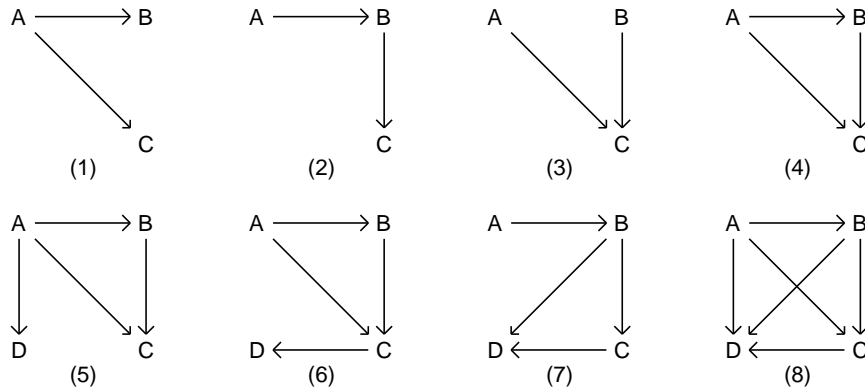


Figure 3: All possible directed motifs on three nodes, and four selected motifs on four nodes.

rules out every form of loops. All four possible connected subgraphs on three nodes and four subgraphs (out of 42) on four nodes are shown in Figure 3.

The node order together with the adjacency condition allows for very efficient searching for these motifs with a sliding window. The occurrences of all three-node motifs and all four-node motifs are scaled to a sum of 1, respectively. The results build the motif signatures, as defined in Section 2.2.

Machine learning setup We use the values of the motif signatures as 46 numeric features for our machine learning experiments. In addition, we include the word count of the article version as an additional baseline feature according to (Blumenstock, 2008), for comparison and combination. The experiments were performed with J48, a Java implementation of the C4.5 tree learning algorithm, included in Weka (Witten et al., 1999; Quinlan, 2014). The tree structures allow us to interpret the model and analyze the most determining features. We use default parameters with the exception of “minNumObj”, the minimum number of instances per leaf. The default value of this parameter is 2. We set it to 100 to reduce overfitting effect, and will present results for both configurations. The evaluation is performed with 10-fold cross validation over 10 experiment iterations.

5 Quantitative Results

Baseline Our corpus contains 7% featured article versions. Therefore, consistently predicting the majority class “non-featured” produces a lower bound baseline accuracy of 0.93. The baseline we want to

compare with is created by a J48 experiment with the word count feature only, which achieves 0.95 accuracy.

Stand-alone We evaluate the predictive power of our motifs with experiments that use only 3-node motifs, only 4-node motifs or both. The results for these experiments with default number of 2 instances per leaf are presented in Table 1. Experiments with all 3-node motifs without word count could not reach the baseline, but using all 4-node motifs without word count performed much better at 0.9745 accuracy. This includes significant overfitting effects, as the corresponding tree model is very large and consists of over 3,000 nodes and leaves.

Used Features	Accuracy	Tree size
Majority class baseline	0.93	
3N + 4N	0.9748	3605
4N	0.9745	3598
W (baseline)	0.95	6
3N	0.9488	1417

Table 1: J48 results for motifs alone, parameter minNumObj = 2, W = word count, 3N = all 3-node motifs, 4N = all 4-node motifs

Table 2 shows the results with 100 minimum instances per leaf (up from 2), which reduces the model size and the overfitting effects. Lower bound and baseline accuracy are the same as in the first setup, at 0.93 and 0.95, respectively. In this setup, using only 3-node motifs yields an accuracy of 0.943. 4-node motifs alone or in conjunction with 3-node motifs do not outperform the word count base-

Used Features	Accuracy	Tree size
Majority class baseline	0.93	
3N + 4N	0.9508	217
4N	0.9504	188
W (baseline)	0.95	3
3N	0.943	140

Table 2: J48 results for motifs alone, parameter minNumObj = 100, W = word count, 3N = all 3-node motifs, 4N = all 4-node motifs

line considerably, either. We conclude that our motif analysis as a stand-alone tool did not lead to notable statistical improvements.

Combination In our experiments, we also combined the baseline feature word count with motif features. The results for these combinations are shown in Table 3 with default number of 2 instances per leaf and Table 4 with 100 instances per leaf. At default setup, a combination of word count with 3-node and 4-node motifs shows excellent performance at 0.978 accuracy. Again, very large decision tree models were created, which indicates overfitting. Limiting the tree size reduces the decision tree to very moderate size and still results in 0.96 accuracy. Motifs together with word count close the gap from the baseline’s 5% miscategorized examples to 4%.

Used Features	Accuracy	Tree size
Majority class baseline	0.93	
W + 3N + 4N	0.9780	3179
W + 4N	0.9779	3254
W + 3N	0.9567	847
W (baseline)	0.95	6

Table 3: J48 results for combination, parameter minNumObj = 2, W = word count, 3N = all 3-node motifs, 4N = all 4-node motifs

We confirmed this improvement and its statistical significance on reduced subsets of our data, and also in a balanced setting. We created the reduced subsets by a purely random selection of 10% featured and non-featured article versions. Compared to the full dataset, these subsets have reduced size, but the same ratio of featured vs. non-featured instances. To create balanced subsets, we combined all featured article versions and an equal amount of ran-

Used Features	Accuracy	Tree size
Majority class baseline	0.93	
W + 3N + 4N	0.96	179
W + 4N	0.9577	200
W + 3N	0.9542	59
W (baseline)	0.95	3

Table 4: J48 results for combination, parameter minNumObj = 100, W = word count, 3N = all 3-node motifs, 4N = all 4-node motifs

domly selected, non-featured article versions. We constructed 20 subsets for both settings to measure variance and calculate statistical significance. See Table 5 for the results of the reduced subsets, and Table 6 for the results of the balanced scenario.

Used Features	Mean acc.	Standard dev.
W + 3N + 4N	95.116	0.073
W + 4N	95.093	0.071
W + 3N	95.079	0.085
W (baseline)	94.741	0.096

Table 5: Mean accuracy and standard deviation for 20 reduced datasets, J48 parameter minNumObj = 100, W = word count, 3N = all 3-node motifs, 4N = all 4-node motifs

Used Features	Mean acc.	Standard dev.
W + 3N + 4N	94.963	0.111
W + 4N	94.941	0.124
W + 3N	94.852	0.107
W (baseline)	94.44	0.109

Table 6: Mean accuracy and standard deviation for 20 balanced datasets, J48 parameter minNumObj = 100, W = word count, 3N = all 3-node motifs, 4N = all 4-node motifs

Due to the reduced amount of data, the mean accuracy was lower compared to previous experiments with same features and parameters, but the order remained constant. The accuracy is very stable with respect to data composition, with standard deviations between only 0.073 and 0.124. The mean accuracy of every combination setup surpassed the baseline by at least three standard deviations. Computation of p-values confirmed that all results are statistically highly significant at $p < 0.001$.

Motif	Accuracy
(4)	95.1713845
(8)	95.1328903
(3)	95.0983271
(1)	95.080967
(7)	94.9560177
(2)	94.9386417

Table 7: Accuracy of J48 experiments with single motifs in combination with word count

Most dominant motifs The motifs with highest impact on quality are obtained with two different methods.

To determine the most effective motifs in our machine learning setup, we performed additional experiments with single motifs in combination with word count. Table 7 displays the results with highest accuracy, which indicates a connection between these motifs and quality. The motif labels correspond to Figure 3.

As a second approach, we directly evaluated the correlation of the motif signatures to the quality label of the corresponding text. Since our variable for quality is dichotomous, we use the point biserial correlation coefficient. The value distribution of our motif signature entries imposes potential problems for this computation. An example of this distribution is shown in Figure 4. We see large proportions of the extreme values 0 and 1 in our motif signature distributions. A value of 1 in the signature can only happen if the respective motif was the only motif to be found, which only happens in very small texts. A 0 entry also hints to small texts, as large texts tend to contain at least a fraction of every motif type. Small articles in Wikipedia are rarely featured, so both extreme values of the distribution largely correspond to non-featured articles, which distorts the point biserial correlation coefficient. To eliminate the effects of article length in the correlations, we create sub-collections of featured and non-featured article versions with similar amounts of motifs, measured by mean average values and standard deviation.

We found two motifs with a distinctively negative correlation to the featured quality label, and three motifs with distinctively positive correlation coefficients. See Table 8 for these results. All these motifs have also been identified by our machine learn-

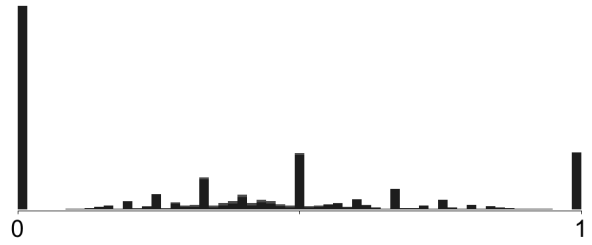


Figure 4: Distribution of values for motif (4) over all motif signatures.

Motif	Correlation
(1)	0.2015307459
(3)	0.1847741318
(2)	0.1803760107
(8)	-0.2691052665
(4)	-0.27433768

Table 8: All motifs with correlation coefficient of absolute value > 0.15

ing approach, which confirms our findings. We concentrated our qualitative interpretation on the motifs with most dominant results, which are motifs (1), (3), (4) and (8) of Fig 3.

6 Qualitative Results

The two motifs that are most positively correlated with the featured quality label are motifs (1) and (3). The huge amount of data made an exhaustive investigation of all motif occurrences infeasible. Therefore, we examined random examples of these motifs in our data set.

Many examples of motif (1) showed this sentence structure: The first sentence introduces two connected entities; the second sentence offers details about the first entity; the third sentence explains the other entity. One example from the German Wikipedia:

- (a) Von den drei Hartstrahlen der Rückenflosse wurde der erste zur Angel (Illicium) mit anhängendem Köder (Esca) umgebildet.
- (b) Das Illicium hat oft eine Streifenzeichnung.
- (c) Die Esca hat bei den einzelnen Anglerfischarten eine unterschiedliche Form und ist ein wichtiges Unterscheidungsmerkmal zwischen den Arten.

Analyzing motif (3) revealed a similar, but decisively different pattern: The first sentence introduces one entity, the second sentence introduces a second one. The last sentence combines the two mentions and draws a connection.

- (a) Die Haupthenne bleibt mit dem Hahn oft über mehrere Jahre zusammen.
- (b) Bei den Nebenhennen handelt es sich meistens um recht junge Weibchen.
- (c) Der Hahn paart sich zunächst mit der Haupthenne, sodann mit den Nebenhennen.

These two ways to introduce and connect entities are an indication of good writing style. The reader can see explanations for the two mentioned entities and their connection in direct vicinity. This makes it easy to understand and to follow the argument structure.

Motifs (4) and (8) are highly negatively correlated with the Wikipedia article quality. They share a very similar structure: Motif (4) is a maximally connected 3-node subgraph, motif (8) is a maximally connected 4-node subgraph. Many text examples of these motifs share a pattern of repetition. One noun is used three or four times in a row in very close proximity.

- (a) Die Nahrungssuche erfolgt in der Regel einzeln, seltener als Paar.
- (b) Bei Beobachtungen waren die Tiere in 92 % aller Fälle allein auf Nahrungssuche.
- (c) Beutejagd und Nahrungssuche bestehen vor allem darin, dass die Füchse ihre Beute zumeist zwischen und unter Steinen suchen und gelegentlich auch graben.

Repetition is a strong stylistic device that can enhance learning effects, but it can also make the text tedious and reduce the attention of a reader (Grill-Spector et al., 2006; Tannen, 2007). Too much repetition is a sign of bad writing style and is certainly avoided in good articles, as our findings demonstrate.

7 Conclusion and Outlook

We have seen that motif analysis can improve the assessment – and our understanding – of the quality of a document. For that, we explored one particular setting and presented the results in this paper. We formulated the following research questions:

1. *Quantitative*: does motif analysis as a stand-alone tool help us assess the quality of text documents statistically?
2. *Quantitative*: does it help us in conjunction with other quality measures?
3. *Qualitative*: does it help us understand the nature of quality any better?

Our empirical findings In our corpus, only 7% of all articles are featured. Hence, categorizing *all* articles as non-featured gives quite a high base line. If the threshold is well chosen, the word count of an article miscategorizes 5%.

Our motif analysis alone is not better than that, so the answer to the first research question is not strictly positive. However, we showed that our combination of both criteria reduces the share of miscategorized articles from 5% down to 4%.

For our third research question, we identified a subset of motifs with high positive or negative correlation to the featured label. Two motifs occur outstandingly frequently in the featured articles, and two other motifs occur outstandingly frequently in the non-featured articles. All four motifs are indeed indicators of text quality as desired: the two former ones are frequently induced by two concepts of good writing style, whereas the two latter ones are frequently induced by two cases of repetitive style.

Future work In the future, we will extend our work to other concepts of quality. However, we anticipate that other quality criteria will require completely different approaches with different networks and, possibly, even different sets of motifs.

Acknowledgments

This work has been supported by the German Research Foundation as part of the Research Training Group Adaptive Preparation of Information from

Heterogeneous Sources (AIPHES) under grant No. GRK 1994/1.”

References

- Charu C. Aggarwal and Haixun Wang. 2010. Managing and mining graph data. *Database*, 40:487–513.
- Charu C Aggarwal. 2011. *Social Network Data Analytics*. Kluwer.
- Lucas Antiquiera, M das Gracas V Nunes, ON Oliveira Jr, and L da F Costa. 2007. Strong correlations between text quality and complex networks features. *Physica A: Statistical Mechanics and its Applications*, 373:811–820.
- Chris Biemann, Stefanie Roos, and Karsten Weihe. 2012. Quantifying semantics using complex network analysis. In *COLING 2012*.
- Joshua E Blumenstock. 2008. Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web*, pages 1095–1096. ACM.
- Eric Brill and Robert C Moore. 2000. An improved error model for noisy channel spelling correction. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 286–293. Association for Computational Linguistics.
- D. J. Cook and L. B. Holder. 2006. *MLing Graph Data*. Wiley.
- Chris Drummond, Robert C Holte, et al. 2003. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on learning from imbalanced datasets II*, volume 11.
- Santo Fortunato. 2010. Community detection in graphs. *Physics Reports*, 486:75–174.
- Kalanit Grill-Spector, Richard Henson, and Alex Martin. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, 10(1):14–23.
- Lachezar Krumov, Christoph Fretter, Matthias Müller-Hannemann, Karsten Weihe, and Marc-Thorsten Hütt. 2011. Motifs in co-authorship networks and their relation to the impact of scientific publications. *European Physical Journal B*, 84(4):535–540.
- Mirella Lapata. 2006. Automatic evaluation of information ordering: Kendall’s tau. *Computational Linguistics*, 32(4):471–484.
- Annie Louis and Ani Nenkova. 2013. What makes writing great? first experiments on article quality prediction in the science journalism domain. *Transactions of the Association for Computational Linguistics*, 1:341–352.
- Mohsen Mesgar and Michael Strube. 2015. Graph-based coherence modeling for assessing readability. *Lexical and Computational Semantics (* SEM 2015)*, page 309.
- R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, October.
- R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. 2004. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, March.
- J Ross Quinlan. 2014. *C4. 5: programs for machine learning*. Elsevier.
- Falk Schreiber and Henning Schwöbbermeyer, 2010. *Statistical and Evolutionary Analysis of Biological Network Data*, chapter Motifs in biological networks, pages 45–64. Imperial College Press/World Scientific.
- S.S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature genetics*, 31(1):64–68.
- Deborah Tannen. 2007. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*, volume 26. Cambridge University Press.
- Joel R. Tetreault and Martin Chodorow. 2008. The ups and downs of preposition error detection in esl writing. In *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1*, COLING ’08, pages 865–872, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ngoc Tam L Tran, Luke DeLuccia, Aidan F McDonald, and Chun-Hsi Huang. 2015. Cross-disciplinary detection and analysis of network motifs. *Bioinformatics and Biology insights*, 9:49.
- Ian H Witten, Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, and Sally Jo Cunningham. 1999. *Weka: Practical machine learning tools and techniques with java implementations*.