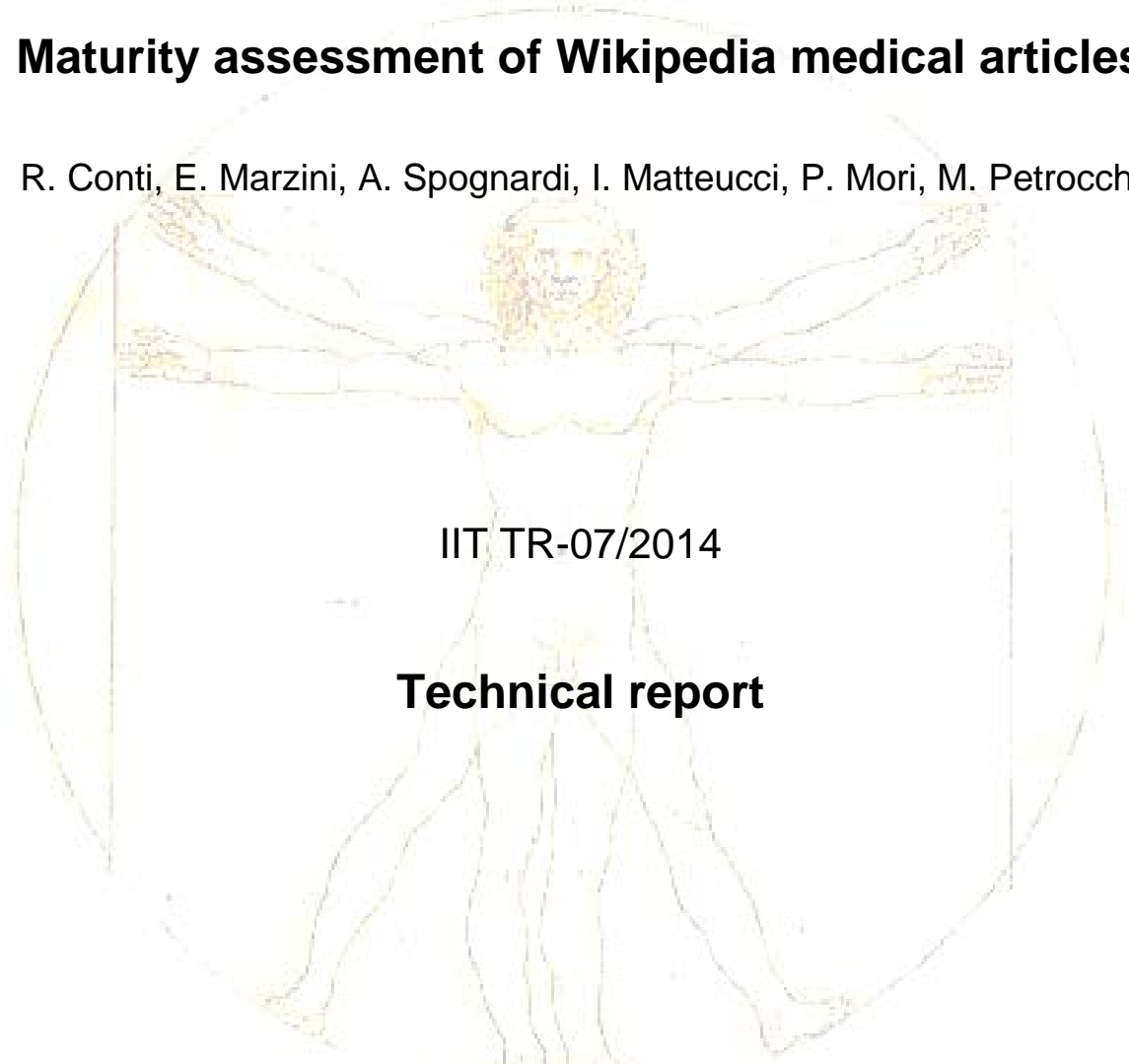




Consiglio Nazionale delle Ricerche

Maturity assessment of Wikipedia medical articles

R. Conti, E. Marzini, A. Spognardi, I. Matteucci, P. Mori, M. Petrocchi



IIT TR-07/2014

Technical report

Marzo 2014



Istituto di Informatica e Telematica

Maturity assessment of Wikipedia medical articles

Riccardo Conti, Emanuel Marzini, Angelo Spognardi, Ilaria Matteucci, Paolo Mori, Marinella Petrocchi
Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy
email: {name.surname@iit.cnr.it}

Abstract—Recent studies report that Internet users are growingly looking for health information through the Wikipedia Medicine Portal, a collaboratively edited multitude of articles with contents often comparable with professionally edited material. Automatic quality assessment of the Wikipedia medical articles has not received much attention by Academia and it presents open distinctive challenges. In this paper, we propose to tag the medical articles on the Wikipedia Medicine Portal, clearly stating their maturity degree, intended as a summarizing measure of several article properties. For this purpose, we adopt the Analytic Hierarchy Process, a well known methodology for decision making, to evaluate the maturity degree of more than 24000 Wikipedia medical articles, based on several features.

I. INTRODUCTION

Nowadays, Internet is one of the simplest and most immediate means to retrieve a *plethora* of information. However, not all of them are equally accurate and/or reliable. This lack of accuracy and reliability is a big concern, in particular if one deals with dissemination of data related to, *e.g.*, instruction, law, culture, medicine, and so on. In particular, many e-health websites are growingly capturing the interests of citizens [1], to such an extent that their use has been proposed to promote global public health [2], even if trust towards such websites is still an open challenge [3]. The Wikipedia Medicine Portal is one of the leading sources of medical information, targeting patients and health care professionals [4], [5]. The portal has a large number of documents and a well structured quality rating scheme, handled by the WikiProject Medicine¹. WikiProject Medicine represents a community of people “interested in medical and health content on Wikipedia”, with the aim of “giving the general public and health care professionals a text they can all read, appreciate, and respect, free of charge”. In pursuing their goal, they manually assess the quality level of the published articles, to aid the recognition of excellent contributions and identify topics that instead need further work. They adopt a well structured scheme that assigns to each article a *quality degree* and an *importance level*, placing them in a fixed place inside the structure of the article (the *banner*). The quality classes are ordered as follows *stub*, *start*, *class C*, *class B*, *Good article*, *Featured article* (class A is not used). The Featured and Good article grades are the highest possible assessments, and they require a community consensus and an official review, while all the others can be achieved with a simple review. The importance level represents the degree of relevance of the article as a *medical article*. As an example, an article about Alexander Fleming (who discovered the antibiotic substance penicillin) has a lower importance level than an article about penicillin itself. A case study reports that the power

of collaborative editing about medical articles produces content comparable, in accuracy and depth – a bit less in readability, with professionally edited material [6]. WikiProject provides details about the characteristics that an article should exhibit to achieve a given quality level. However, such characteristics are mainly qualitative and they heavily rely on the expertise of the editors. Primarily, Featured Articles have to be well-written, comprehensive, well-researched, neutral and stable.

In this paper, we propose an alternative way to tag the medical articles on the Wikipedia Medicine Portal. We intend to attach to each article a piece of information clearly stating, in a quantitative way, its *maturity degree*, intended as a measure of content level, stability, and structural properties. We believe that the use of this metric could improve the consciousness of the users about the type of information they are accessing, going side-by-side with an improved quality offered by the portal. In few words, the maturity degree will summarize in a numerical value the work behind each medical article. We proceed as follows: i) we collect and analyze articles from the Wikipedia Medicine Portal in order to understand if and how the WikiProject guidelines for quality stage assessment are applied; ii) we survey and evaluate several criteria proposed by the academic literature for quality assessment of Wikipedia articles, and then iii) we define a metric to assess the *maturity degree* of a Wikipedia medical article and we compare the new metric with the quality level proposed by WikiProject. The *maturity degree* is a measure to synthesize the features of a medical article, highlighting its relevance to all the quality classes proposed by WikiProject.

The evaluation of the maturity degree is calculated using the Analytic Hierarchy Process (AHP) [7], [8], a well-known methodology for multi-criteria decision making. The choice of AHP is driven by several aspects. It allows to automatically solve a decision problem, where several different *alternatives* can be chosen to reach a *goal*. AHP returns the *most relevant alternative* with respect to a set of previously established *criteria*. It is modular, *i.e.*, able to be expanded and tuned with new criteria, hierarchical, *i.e.*, able to group together different criteria and to assign different weights to each, and it is sound, since it produces consistent results. As final step, we discuss our results to draw some critical points about the automatic assessment of Wikipedia pages.

The paper is organized as follows: next section discusses related work in the area of automatic quality assessment of electronic documents. Section III recalls the AHP approach and presents our instantiation for the evaluation of the maturity degree of medical articles. In Section IV, we discuss our results and in Section V we draw some conclusions and future works.

This research has been partially supported by the Tuscany region projects *MyChoice* and *Pics*, and by the IIT-CNR project *MobiCare*

¹http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine

II. RELATED WORK

WikiProject provides a set of criteria helpful to recognize the quality of an article. Such criteria are usually expressed as qualitative properties and not directly quantifiable, like comprehensiveness and neutrality. Existing work have focused on the automatic assessment of Wikipedia articles, proposing approaches that would quantify the WikiProject criteria.

In [9], the authors map the qualitative criteria proposed into quantitative ones. In particular, findings are that two criteria, “edit times” (the number of times an article has been edited) and the readability, are critical to assess featured articles (FA). Results are that the two criteria work well in distinguishing FA from the other classes, with a precision equal 91%. FA recognition is also the goal of Blumenstock in [10], which proposes 100 criteria divided into four macro-criteria. The author concludes that, using the only “word count” criterion, it is possible to distinguish more than 97% of FA, testing a variety of classification schemes. Similar results are reported in [11]. These works consider a random set of Wikipedia articles: when considering the Medicine Portal, the only evaluation of criteria taking into account the structure of an article is far from be effective, as shown in Section III-B.

Linguistic criteria are discussed in [12], where 8 linguistic metrics are used and analyzed using a decision tree, reaching the 83% of precision to distinguish FA from random non-FA articles. In [13], the authors introduce historical criteria, like “number of editors” (anonymous or registered users that write a revision) and “edit times”, considering that high-quality articles are edited more often and by distinct authors than usual. They observe that such criteria seem to be very useful to distinguish FA articles from the others.

All the above cited work share the commonality to address only FA, while they do not try to assess the articles through the all the classes proposed by WikiProject. Conversely, the authors of [14] consider 28 criteria, grouped as lingual, structural, historical, and reputational. They use seven different neural networks with the aim of distinguishing which, among those criteria, are more useful to assign an article to the various classes. Overall, each criterion is differently weighted according to the considered class. For example, linguistic criteria are more important than others to recognize articles in the lower classes (namely, Stub articles generally have less words and sentences), while richness of content and articulated structure are important to distinguish articles of higher classes. Instead, Warncke-Wang *et al.*, in [15], propose a criteria-based classifier that assigned an article to one of the possible WikiProject classes. In their work, they also make use of the Wikipedia template messages², as new features to assess the quality of the articles. Lastly, they base their analysis on randomly sampled articles.

Differently to [14], [15], in this work we propose a methodology to evaluate the relevance of an article to all the WikiProject classes. Our scenario is more specific than the ones illustrated in all the above cited work. Indeed, we analyze all the articles of the Wikipedia Medicine Portal and, for each of them, we output a vector that (quantitatively) measures the relevance of the article with respect to all the classes. The vector represents

²In Wikipedia, template messages are small notes to inform readers and editors of specific problems within articles or sections.

Intensity	Description	Explanation
1	Equal	Two elements contribute equally to the objective
3	Moderate	One element is slightly more relevant than another
5	Strong	One element is strongly more relevant than another
7	Very strong	One element is very strongly more relevant than another
9	Extreme	One element is extremely more relevant than another

TABLE I. THE AHP FUNDAMENTAL SCALE

what we call the *maturity degree* of the article and it can be considered as a measure of the collaborative effort lavished to produce the article.

III. MULTI-CRITERIA ASSESSMENT OF MEDICAL ARTICLES

In this section, we show how to apply the Analytic Hierarchy Process (AHP) to assess the maturity degree of an article, with particular reference to the medical articles on the Wikipedia Medicine Portal. Before instantiating the methodology, we briefly recall the basic steps of AHP.

A. The Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) [7], [8] is a multi-criteria decision making technique, which has been largely used in several fields of study. Given a decision problem, where several different *alternatives* can be chosen to reach a *goal*, AHP sorts the alternatives from the *most relevant* to the *less relevant*, with respect to a set of *criteria* and *subcriteria*. Practically, the AHP approach is to subdivide a complex problem into a set of sub-problems, criteria and subcriteria, and then to compute the global solution by properly merging the various local solutions for each sub-problem.

The problem is structured as a hierarchy, as shown in Figure 2. Once the hierarchy is built, the method performs pairwise comparisons, from the bottom to the top, in order to compute the relevance, hereafter called *local priority*: i) of each alternatives with respect to each subcriteria, ii) of each subcriterion with respect to the relative criterion, and finally, iii) of each criterion with respect to the goal. Comparisons are expressed in a matricial form, called *pairwise comparison matrix*. A pairwise comparisons matrix M is a square matrix which has positive entries and satisfies the reciprocal property, *i.e.*, $a_{ii} = 1$ and $a_{ij} = \frac{1}{a_{ji}}$. For each level of the hierarchy, the entries of a pairwise comparison matrix represent how much one element at that level is more relevant with respect to another element at the same level. Relevancy is estimated with respect to elements at the upper level, and according to a pre-defined comparison scale typical of AHP (see Table I). The scale indicates how many times an element is *more relevant* than another one, with respect to the element at the upper level. The relevance can be established according either to subjective or objective statements. The normalized eigenvector associated with the largest eigenvalue [16] of each matrix gives a vector of local priorities. To avoid writing inconsistent comparisons and come up with an invalid result, AHP requires a *consistency* check on the matrices. Inconsistency of a reciprocal matrix $n \times n$ can be captured by a *Consistency Index*: $CI = \frac{\lambda_{max} - n}{n - 1}$, where λ_{max} is the maximum eigenvalue of the matrix. CI must be less than 0.1. Once all local priorities are computed, global priorities $P_g^{a_i}$, given the relevances of the alternatives a_i *w.r.t.*

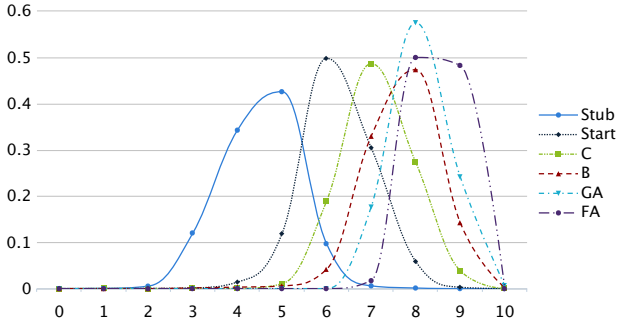


Fig. 1. *word count* subcriterion distribution per Wikipedia Medicine Portal articles and quality classes (x scale is logarithmic)

the goal g , are computed as a weighted sum. For the sake of simplicity, and without loss of generality, we have in mind a hierarchy tree where the leftmost n_1 criteria have a set of subcriteria each, while the rightmost n_2 criteria have no subcriteria below them, and $n_1 + n_2 = n$ is the number of total criteria. Thus, global priorities are computed as:

$$P_g^{a_i} = \sum_{w=1}^{n_1} \sum_{k=1}^{q(w)} p_g^{c_w} \cdot p_{c_w}^{s_{c_k}^{c_w}} \cdot p_{s_{c_k}^{c_w}}^{a_i} + \sum_{j=1}^{n_2} p_g^{c_j} \cdot p_{c_j}^{a_i} \quad (1)$$

where $q(w)$ is the number of subcriteria for criterion c_w , $p_g^{c_w}$ is the local priority of criterion c_w with respect to the goal g , $p_{c_w}^{s_{c_k}^{c_w}}$ is the local priority of subcriterion k with respect to criterion c_w , and $p_{s_{c_k}^{c_w}}^{a_i}$ is the local priority of alternative a_i with respect to subcriterion k of criterion c_w .

B. AHP to assess the maturity of a medical article

The dataset we consider consists of the whole Wikipedia Medicine Portal, made of around 24k medical articles. The dataset is distinctive for more than one reason. For example, it is composed by medical articles that are intrinsically complex and based on a very technical dictionary. This means that standard approaches measuring the comprehension difficulty of contemporary English (like, *e.g.*, the Flesch reading ease, [17]), could produce flat results when applied to the technical dictionary of the Medicine Portal articles, compared to a random subset of the whole Wikipedia articles. Similarly, distinguishing the quality class of a Wikipedia medical article considering only simple sub criteria as “word count”, as proposed in [10], [11], could be limitative: as shown in Figure 1, where the article classes are distributed with respect to the word count subcriterion, there is a strong overlapping between many classes. Further, the dataset is very heterogeneous, being composed both by comprehensive articles (like “Tuberculosis”) and very short ones with an easy structure (an example is “5-Iodowillardiine”). Finally, the Featured and Good Article classes are composed by a very small fraction of the Medicine Portal, mainly because the articles of intermediate quality classes lack of neutrality or readability.

Before instantiating AHP for article assessment, we introduce here the definition of *maturity degree* of a Wikipedia medical article. The maturity degree is a vector $v = [v_i]$, with $1 \leq i \leq 6$ and $\sum v_i = 1$. Each index i , from 1 to 6, represents the WikiProject class *stub*, *start*, *class C*, *class B*, *Good article*, *Featured article*, respectively. The vector quantifies the relevance of a given article to each of those classes,

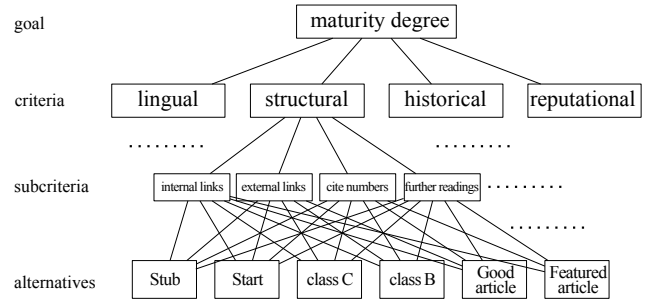


Fig. 2. AHP Hierarchy to assess the maturity degree of a Wikipedia medical article

introduced in Section I. We say that the maturity degree vector is *consistent* when it has exactly one absolute max value. Formally, we require that one of the three conditions holds: either (1) $\forall i, v_i \geq v_{i+1}$, or (2) $\forall i, v_i \leq v_{i+1}$, or (3) $\exists! i$ such that $v_i \geq v_{i-1} \wedge v_i \geq v_{i+1}$ and $\forall j < i, v_{j-1} \leq v_j$ and $\forall j > i, v_j \geq v_{j+1}$. The property of consistency of the vector ensures that the relevance is max either for only one class or for neighboring classes.

Figure 2 shows the AHP instantiation to assess the maturity degree. We consider an AHP hierarchy where the *goal* is to produce a maturity degree vector for a given article. We consider four criteria: *lingual*, *structural*, *historical* and *reputational*. Each criterion has a different number of subcriteria (to save space, only some of them are shown in the figure), mainly considered in the past by Academia, *e.g.*, [9], [14]. Subcriteria belonging to the *lingual* criterion are related to the textual content of an article from a syntactical point of view. Subcriteria are: (1) Flesch reading ease and (2) Flesch-Kincaid grade level (they both measure the comprehension difficulty of English texts); (3) word count and (4) sentence count; (5) multi-syllable words / words ratio; (6) spell error / words ratio. *Structural* criterion is drawn according to the composition of an article and other elements of its structure. As structural subcriteria, we consider: (1) number of categories³; (2) internal and (3) external links; (4) non-textual resources (like sounds, images, and alike); (5) further readings (*i.e.*, the number of additional references); (6) symbols in title; (7) number of section headings; (8) number of citations⁴. The *historical* criterion considers the revision process that yields the current status of the article. The following subcriteria belong to historical criterion: (1) number of edits; (2) number of editors; (3) number of devoted editors (namely, those that edited the article more than once); (4) number of anonymous editors; (5) minor edit ratio (*i.e.*, the fraction of edits marked as *minor* by the editor); (6) article age (namely, the period of time between the first and the last edit of the article); edit frequency (namely, the number of edits over the article age). Finally, the *reputational* criterion takes into account the authorship information of an article: its subcriteria are (1) average active age of editors; (2) average upload amount of editors; (3) average edit times of editors; (4) average talk times of editors. In particular, to collect the reputational subcriteria,

³In Wikipedia, the *categories* are substantially tags attached to articles, representing other related arguments. They are topic categories an article belongs to.

⁴In Wikipedia, *citations* are essentially footnotes referencing sources supporting the assertions of the article.

$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 2 & 5 & 5 & 5 & 7 \\ S & \frac{1}{2} & 1 & 5 & 5 & 5 & 7 \\ C & \frac{1}{5} & \frac{1}{5} & 1 & 1 & 1 & 1 \\ B & \frac{1}{5} & \frac{1}{5} & 1 & 1 & 1 & 1 \\ G & \frac{1}{5} & \frac{1}{5} & 1 & 1 & 1 & 1 \\ F & \frac{1}{7} & \frac{1}{7} & 1 & 1 & 1 & 1 \end{matrix}$ <p>(a) <i>Low matrix</i> (CI=0.0117)</p>	$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 1 \\ S & 3 & 1 & \frac{1}{3} & \frac{1}{3} & 1 & 3 \\ C & 6 & 3 & 1 & 1 & 3 & 6 \\ B & 6 & 3 & 1 & 1 & 3 & 6 \\ G & 3 & 1 & \frac{1}{3} & \frac{1}{3} & 1 & 3 \\ F & 1 & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 1 \end{matrix}$ <p>(b) <i>Intermediate matrix</i> (CI=0.0443)</p>	$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 1 & 1 & 1 & \frac{1}{7} & \frac{1}{7} \\ S & 1 & 1 & 1 & 1 & \frac{1}{5} & \frac{1}{5} \\ C & 1 & 1 & 1 & 1 & \frac{1}{5} & \frac{1}{5} \\ B & 1 & 1 & 1 & 1 & \frac{1}{5} & \frac{1}{5} \\ G & 7 & 5 & 5 & 5 & 1 & \frac{1}{2} \\ F & 7 & 5 & 5 & 5 & 2 & 1 \end{matrix}$ <p>(c) <i>High matrix</i> (CI=0.0117)</p>
--	---	---

Fig. 3. Empirical matrices

we crawled the article metadata, all the discussions (“talks”) occurred among the editors, and many statistics related to the editors of a specific article. For example, the average talk times of editors is related to the number of times that the contributors of an article have taken part to a discussion about any modification. Given the values a Wikipedia medical article exhibits for the considered set of subcriteria, we will assign to the article a maturity degree vector, expressing its relevance to each class.

C. Comparison matrices for maturity assessment

The relative relevance of two classes (*i.e.*, two alternatives) with respect to a subcriterion in the upper level of the hierarchy depends on the value that such subcriterion assumes for a given article. We need to use several different comparison matrices, depending on the values an article exhibits for a given subcriterion. In the following, we describe two ways for defining the comparison matrices for each subcriterion, in order to automatically assess the maturity of each article. In Section IV-B, we apply AHP using the matrices obtained with both the approaches and we compare the result of each AHP instantiation.

1) *Empirical approach*: We divide the possible values that an article can exhibit, for a given subcriterion, in three intervals, namely *low*, *intermediate*, *high*. Then, we define three matrices expressing the relevance of the six WikiProject classes (Featured, Good, etc.), with respect to each subcriterion, as shown in Figure 3. Finally, for each subcriterion, we link one of the three intervals to one of the three matrices. For example, let the reader consider the matrix that compares all the alternatives *w.r.t.* the subcriterion *word count* (belonging to the lingual criterion). If the article under investigation has a *low* number of words, we pick up the *low* comparison matrix, that assigns more relevance to the less mature classes (say Stub and Start). In this case, class Stub has been considered 7 times more relevant than class Featured, for that article. Similarly, when the number of words is several thousands, then, the relevance could be symmetrically considered, namely the Featured article class could be 7 times more relevant than the Stub class. In this case, we adopt the *high* comparison matrix. The values for the scale of relevance have been defined according to Table I. As another example, the *Flesch reading ease* subcriterion says that the more a text is complex, the more the article has a lower value for that subcriterion. If the article has a *low* value for *Flesch reading ease*, we use the matrix *high*, defined to assign more relevance to Featured and Good classes, rather than to Stub and Start classes. Finally, we choose the *intermediate* matrix whenever an article assumes, for the given subcriterion, a value that falls in the intermediate values interval.

criterion	range		
	<i>low</i>	<i>intermediate</i>	<i>high</i>
lingual			
<i>flesch reading ease</i>	(−∞; 50]	(50; 80]	(80; +∞)
<i>flesch kincaid grade level</i>	(−∞; 8]	(8; 16]	(16; +∞)
<i>multi syllable words ratio</i>	(−∞; 0.4]	(0.4; 0.6]	(0.6; +∞)
<i>spell error / words count</i>	(−∞; 0.1]	(0.1; 0.2]	(0.2; +∞)
<i>words count</i>	(−∞; 200]	(200; 1000]	(1000; +∞)
<i>sentences count</i>	(−∞; 20]	(20; 100]	(100; +∞)
structural			
<i>categories count</i>	(−∞; 2]	(2; 4]	(4; +∞)
<i>internal links</i>	(−∞; 40]	(40; 200]	(200; +∞)
<i>external links</i>	(−∞; 0]	(0; 2]	(2; +∞)
<i>non textual resources</i>	(−∞; 0]	(0; 5]	(5; +∞)
<i>further readings</i>	(−∞; 0]	(0; 1]	(1; +∞)
<i>non char symbols in title</i>	(−∞; 0]	(0; 1]	(1; +∞)
<i>section headings count</i>	(−∞; 4]	(4; 10]	(10; +∞)
<i>cite numbers</i>	(−∞; 5]	(5; 40]	(40; +∞)
historical			
<i>edit times</i>	(−∞; 20]	(20; 1000]	(1000; +∞)
<i>update ed. count</i>	(−∞; 10]	(10; 200]	(200; +∞)
<i>repeat ed. count ratio</i>	(−∞; 0.1]	(0.1; 0.33]	(0.33; +∞)
<i>anonym ed. count ratio</i>	(−∞; 0.1]	(0.1; 0.2]	(0.2; +∞)
<i>minor edit ratio</i>	(−∞; 0.1]	(0.1; 0.3]	(0.3; +∞)
<i>article age</i>	(−∞; 1 year]	(1 year; 4 years]	(4 years; +∞)
<i>edit frequency</i>	(−∞; 0.05]	(0.05; 0.25]	(0.25; +∞)
reputational (related to editors)			
<i>avg active age of ed.</i>	(−∞; 1 year]	(1 year; 4 years]	(4 years; +∞)
<i>avg upload amount of ed.</i>	(−∞; 1MB]	(1MB; 3MB]	(3MB; +∞)
<i>avg edit times of ed.</i>	(−∞; 1000]	(1000; 2000]	(2000; +∞)
<i>avg talk times of ed.</i>	(−∞; 100]	(100; 200]	(200; +∞)

TABLE II. EMPIRICAL RANGES

Clearly, a crucial point of the whole procedure is the definition of the intervals for each subcriterion, namely, the definition of those values for which we should switch from low to intermediate and from intermediate to high. We empirically set the intervals for each subcriterion, as shown in Table II. The choice has been intuitive but subjective and it was made without considering the values actually exhibited by the articles in the dataset of the Wikipedia Medicine Portal.

2) *Statistical approach*: Starting from the articles dataset, we build a sample set formed by an equal number of articles belonging to each of the six WikiProject classes (Featured, Good, etc.). For each subcriterion, we sort the values exhibited by our sample set and split such values in intervals with the same number of elements: the values related to the borderline elements are used to define the extremes of the intervals. We decide to use six intervals. In principle, each interval should correspond to each WikiProject class. What happens in practice is that articles of different Wikipedia classes fall into different intervals, for each subcriterion, as clarified in Section IV. Since we a priori know the class each article in our intervals belongs to, we leverage the distribution of the classes in the different intervals to define their relative relevance with respect to each subcriterion. This leads us to build more than 150 comparison matrices, reported online⁵ for the sake of brevity.

Till now, we have defined the comparison matrices comparing the alternatives in our problem (*i.e.*, the six Wikipedia classes) with respect to each subcriterion of the hierarchy shown in Figure 2. Going to the upper levels of the same hierarchy, we need to define comparison matrices both comparing subcriteria with respect to each criterion, and criteria

⁵<http://mobicare.iit.cnr.it/wikiassessment/>

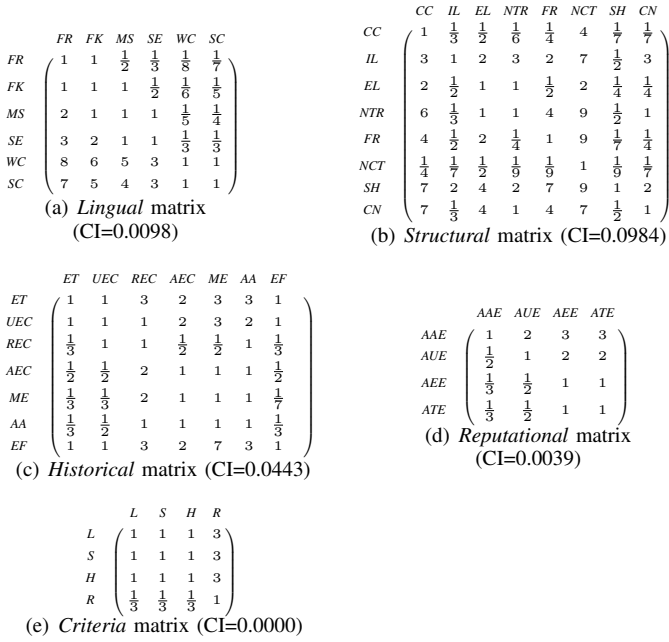


Fig. 4. Comparison matrices for criteria and subcriteria

with respect to the final goal (the top element in the hierarchy).

To define those matrices, we followed the guidelines proposed by WikiProject and also the main observations suggested by Academia, see, *e.g.*, [14], [9], [12], [10] on the importance of each subcriterion. All the obtained matrices are reported in Figure 4. A noteworthy observation is about the comparison matrix that compares criteria with respect to the goal: we suppose that only the reputational criterion is less relevant when compared to the others, since the WikiProject guidelines does not mention any reputational factor about the article editors.

IV. IMPLEMENTATION AND RESULTS

In this section, we describe how we have implemented the AHP algorithm above described to realize our assessment of the Wikipedia Medicine Portal. Then, we report and discuss the obtained results.

A. Assessment implementation

First, we crawled the whole set of articles, leading to the dataset described in Section III-B. Then, we designed a database to manage all the subcriteria described in the same section. We populated the database evaluating all the subcriteria for each single article. In particular, this phase required a lexical analysis, a structural evaluation, and a review of the history of each page. Moreover, we analyzed the history of each editor that had contributed to at least one article in our dataset (this serves to evaluate the reputational subcriteria). The final size of our database amounts to 2GB.

Finally, for each article in the dataset, we have run AHP. As described in Section III-C, we propose both an empirical and a statistical approach. To apply AHP to a single article, we have queried the database to obtain the values of that article for all the subcriteria and we have picked up all the matrices related

WikiProject Class	Consistent maturity degrees	
	Empirical matrices	Statistical matrices
Stub	54%	95%
Start	96%	98%
Class C	97%	94%
Class B	94%	88%
Good Article	86%	94%
Featured Article	75%	86%

TABLE III. PERCENTAGE OF ARTICLES THAT OBTAINED A CONSISTENT MATURITY DEGREE

to those values, according to one of the two approaches. The matrices at the upper levels of the hierarchy are pre-defined and shown in Figure 4. We have evaluated the eigenvector for each matrix and we have used the AHP global formula in Eq. 1 to obtain the final vector of priorities. The final vector of priorities represents the maturity degree of the article, namely, a vector with six components, each representing the relevance of the article to the corresponding WikiProject class.

For each vector, we have verified its consistency, as defined in Section III-B, to check if our assessment produces conflicting results. A conflicting result happens when, for example, an article results to be mature to be a Start as well as a Featured article (indeed, they are not neighboring classes). Finally, we have compared the obtained maturity degrees with the class associated to that article by the WikiProject Medicine. The comparison is useful to check how our quantitative results agree with the quality class qualitatively evaluated by WikiProject.

B. Results

Table III reports the percentage of consistent vectors. The notion of consistency has been introduced in Section III-B. The consistency property ensures that the maturity degree assigns to an article, for a given class, a relevance that does not conflict with all the relevances of the other classes. The table shows the consistency results following the approaches presented in Section III-C, the empirical and the statistical one. Both the approaches exhibit a good level of consistence, thanks to the fact that we have applied AHP with matrices reflecting the progressive growth of an article from the Stub class to the highest one. However, the statistical approach is able to produce more consistent maturity degrees, since it makes use of finer ranges and many more matrices than the empirical approach. For this reason, hereafter we will focus on the results of the statistical approach.

Figure 5 shows a summary of the results of our assessment. For each WikiProject class, we randomly sampled 50 articles belonging to that class and draw their resulting maturity degree as a line following the relevance of each class. Overall, it can be noticed that all the plots have a noticeable shape with only one max, testifying the satisfaction of the consistency property, at least as a general trend. For example, all the articles of the *Stub* class have their max on the *Stub* or on the *Start* class and none of them has two max separated by a lower relevance. A similar observation holds for articles of the other classes. In general, our maturity assessment is in line with the WikiProject quality assessment, *i.e.*, the maturity degrees resulting from applying AHP reflect, overall, the class

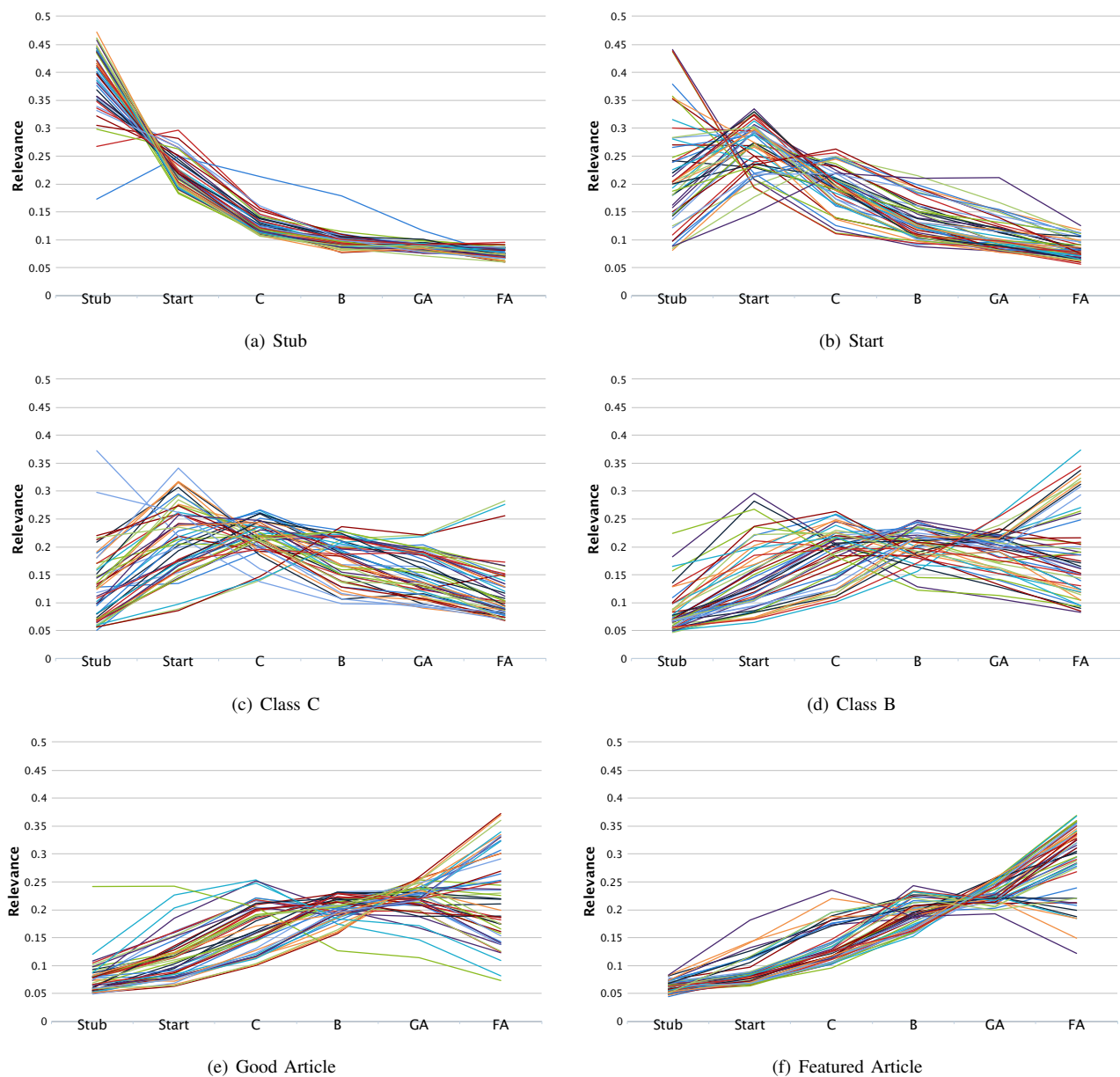


Fig. 5. Maturity degree with respect to the WikiProject assessment

assigned to an article by the project. This is more evident for the utmost classes (*Stub* and *FA*), while the articles tagged by the project as “intermediate classes articles” expose a flatter trend (see sub-figures 4(c) and 4(d) in Figure 5). However, in all the plots, some articles have received a maturity degree significantly different from the WikiProject class they belong to. For example, in Figure 5(f) that shows the results for Featured Articles, we can notice a couple of articles that have their maximum relevance on the corresponding *C* quality class. From the analysis of our dataset, we noticed that those articles, which have a maturity degree max on higher classes than the one associated by WikiProject, exhibit quantitative criteria typical of high quality class articles (*i.e.*, long texts, articulated structure, and many references). However, the manual assessment by WikiProject considers also other qualitative guidelines, as the neutrality and the comprehensiveness that are hard to compute in a quantitative way. On the contrary,

some articles have a maturity degree max on lower classes than the one associated by WikiProject (*e.g.*, “Thyrotoxic periodic paralysis” has a maturity degree max on *Class C* but it is actually a *FA*, with outstanding style and full coverage of the treated topics). They have been assigned by the project to the top classes, even if their structure is relatively simple and expose several low values on the quantitative criteria, when compared to other Good or Featured articles.

From the above observations, we can say that our maturity degree is a different metric than the quality level attached to each article by the WikiProject quality assessment. Thus, we can conclude that those quantitative indicators (like linguistic, structural, historical, and reputational features) are not eventually decisive for the quality assessment for Wikipedia Medicine Portal. However, the same features play an important role when using automatic techniques for article evaluation

(like the approach shown in this paper), since they are easy to extract from the Web and easy to compute. This paves the way for further investigation, going towards either finding affordable methods to render the pretty qualitative guidelines into quantitative criteria, or to achieve a new assessment that makes use of only quantitative measures (like in our approach).

V. CONCLUSIONS

In this paper, we defined the maturity degree of Wikipedia medical articles. We then exploited the AHP methodology to assess such degree over the whole dataset of the Wikipedia Medicine Portal, based on several criteria inherited both from literature and from the WikiProject community. To the best of our knowledge, our attempt to give relevance degrees to an article, for each of the WikiProject classes, is original. In some cases, we found a discrepancy between the quality class attached to them by WikiProject and their maturity degree computed with AHP. This leads us to conclude that the manual assessment took into account other uncountable guidelines, that, in some cases, had a strong impact on the final evaluation of the article. We think that our results, other than being interesting on their own, also call for further investigations, mainly towards the full automation of the current manual article assessment. This could lead to less ambiguity in the evaluation results (mainly due to intrinsic human bias), resulting in a strengthened medicine portal.

REFERENCES

- [1] M. R. Laurent and T. J. Vickers, "Seeking health information online: does Wikipedia matter?" *Journal of the American Medical Informatics Association*, vol. 16, no. 4, 2009.
- [2] J. Heilman *et al.*, "Wikipedia: A key tool for global public health promotion," *Journal of Medical Internet Research* 13(1):e14, 2011.
- [3] L. Vega, T. DeHart, and E. Montague, "Trust between patients and health websites: a review of the literature and derived outcomes from empirical studies," *Health and Technology*, vol. 1, no. 2-4, 2011.
- [4] Rachel Feltman, "The #1 doctor in the world is Dr. Wikipedia," *MSN News online*, Jan 2014, <http://goo.gl/zhTNoa>, Last checked: February 10, 2014.
- [5] M. Aitken, T. Altmann, and D. Rosen, "Engaging patients through social media," IMS Institute for healthcare informatics, Tech. Rep., Jan 2014, <http://goo.gl/BoFJA8>, Last checked: February 10, 2014.

- [6] M. S. Rajagopalan *et al.*, "Patient-oriented cancer information on the Internet: a comparison of Wikipedia and a professionally maintained database," *Journal of Oncology Practice*, vol. 7, no. 5, 2011.
- [7] T. L. Saaty, "Decision-making with AHP: Why is the principal eigenvector necessary," *Journal Operational Research*, vol. 145(1), 2003.
- [8] —, "Decision making with the Analytic Hierarchy Process," *International Journal of Services Sciences*, vol. 1, no. 1, pp. 83–98, 2008.
- [9] B. Stvilia, M. B. Twidale, L. Gasser, and L. C. Smith, "Information quality discussions in Wikipedia," in *Intl. Conference on Knowledge Management*, S. Hawamdeh, Ed., 2005.
- [10] J. E. Blumenstock, "Automatically assessing the quality of Wikipedia," in *UC Berkeley: School of Information Report 2008-021*, 2008.
- [11] H. Dalip *et al.*, "Automatic quality assessment of content created collaboratively by web communities: A case study of Wikipedia," in *9th ACM/IEEE-CS Joint Conference on Digital Libraries*. ACM, 2009.
- [12] Y. Xu and T. Luo, "Measuring article quality in Wikipedia: Lexical clue model," in *Web Society (SWS), 2011 3rd Symposium on*, 2011.
- [13] D. M. Wilkinson and B. A. Huberman, "Cooperation and Quality in Wikipedia," in *Intl. Symposium on Wikis*. ACM, 2007.
- [14] K. Wu, Q. Zhu, Y. Zhao, and H. Zheng, "Mining the factors affecting the quality of Wikipedia articles," in *Information Science and Management Engineering (ISME), 2010 International Conference of*, vol. 1, 2010.
- [15] M. Warncke-Wang, D. Cosley, and J. Riedl, "Tell Me More: An actionable quality model for Wikipedia," in *9th Intl. Symposium on Open Collaboration*. ACM, 2013.
- [16] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, no. 3, 1977.
- [17] R. Flesch, "A new readability yardstick," *Journal of Applied Psychology*, vol. 32(3), Jun 1948.

VI. APPENDIX

In this appendix, we show all the matrices used for the statistical approach presented in Section III-C2. We report the full list of comparison matrices that we used for our experiments. Each figure groups six matrices. Each group of matrices is related to a specific criterion, as an example Figure 6 groups all the matrices related to the criterion "Flesch reading ease". The matrices compare elements at the lower level of the hierarchy (lower with respect to the criterion). For each group, each of the six matrices is instantiated in our implementation of the AHP algorithm when the value for the specific criterion the matrix is related to is in the interval specified in the caption.

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 2 & 2 & 2 & 1 & 3 \\ S & \frac{1}{2} & 1 & 1 & 1 & 1 & 1 \\ C & \frac{1}{2} & 1 & 1 & 1 & 1 & 2 \\ B & \frac{1}{2} & 1 & 1 & 1 & 1 & 1 \\ G & 1 & 1 & 1 & 1 & 1 & 2 \\ F & \frac{1}{3} & 1 & \frac{1}{2} & 1 & \frac{1}{2} & 1 \end{matrix}$$

(a) range [0.194444, 0.411849]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ S & 2 & 1 & 2 & 1 & 1 & 1 \\ C & 2 & \frac{1}{2} & 1 & \frac{1}{2} & 1 & 1 \\ B & 2 & 1 & 2 & 1 & \frac{1}{2} & \frac{1}{2} \\ G & 2 & 1 & 1 & 2 & 1 & 1 \\ F & 2 & 1 & 1 & 2 & 1 & 1 \end{matrix}$$

(b) range [0.411849, 0.431373]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 1 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ S & 1 & 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} \\ C & 1 & 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} \\ B & 2 & 1 & 1 & 1 & \frac{1}{2} & 1 \\ G & 3 & 2 & 2 & 2 & 1 & 1 \\ F & 2 & 2 & 2 & 1 & 1 & 1 \end{matrix}$$

(c) range [0.431373, 0.450649]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ S & 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ C & 2 & 1 & 1 & 1 & 1 & 1 \\ B & 3 & 2 & 1 & 1 & 1 & 1 \\ G & 3 & 2 & 1 & 1 & 1 & 1 \\ F & 3 & 2 & 1 & 1 & 1 & 1 \end{matrix}$$

(d) range [0.450649, 0.469497]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 1 & \frac{1}{2} & 1 & 2 & \frac{1}{2} \\ S & 1 & 1 & \frac{1}{2} & 1 & 4 & 1 \\ C & 2 & 2 & 1 & 2 & 5 & 1 \\ B & 1 & 1 & \frac{1}{2} & 1 & 3 & \frac{1}{2} \\ G & \frac{1}{2} & \frac{1}{4} & \frac{1}{5} & \frac{1}{3} & 1 & \frac{1}{5} \\ F & 2 & 1 & 1 & 2 & 5 & 1 \end{matrix}$$

(e) range [0.469497, 0.494934]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 1 & 2 & 2 & 7 & 9 \\ S & 1 & 1 & 2 & 1 & 4 & 9 \\ C & \frac{1}{2} & \frac{1}{2} & 1 & 1 & 2 & 9 \\ B & \frac{1}{2} & 1 & 1 & 1 & 3 & 9 \\ G & \frac{1}{7} & \frac{1}{4} & \frac{1}{2} & \frac{1}{3} & 1 & 9 \\ F & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 \end{matrix}$$

(f) range [0.494934, 1]

Fig. 8. Multi syllable words ratio matrices

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 1 & 1 & 1 & 1 & 1 \\ S & 1 & 1 & 1 & 1 & 1 & 1 \\ C & 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ B & 1 & 1 & 2 & 1 & 1 & 1 \\ G & 1 & 1 & 2 & 1 & 1 & 2 \\ F & 1 & 1 & 2 & 1 & \frac{1}{2} & 1 \end{matrix}$$

(a) range [0, 0.00680967]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & \frac{1}{8} & \frac{1}{9} & \frac{1}{4} & \frac{1}{9} & \frac{1}{9} \\ S & 8 & 1 & 1 & 2 & \frac{1}{2} & \frac{1}{3} \\ C & 9 & 1 & 1 & 2 & \frac{1}{2} & \frac{1}{2} \\ B & 4 & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} & \frac{1}{5} \\ G & 9 & 2 & 2 & 4 & 1 & \frac{1}{2} \\ F & 9 & 3 & 2 & 5 & 2 & 1 \end{matrix}$$

(b) range [0.00680967, 0.0094162]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ S & 2 & 1 & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} \\ C & 2 & 2 & 1 & 1 & 1 & 2 \\ B & 2 & 1 & 1 & 1 & 1 & 1 \\ G & 3 & 2 & 1 & 1 & 1 & 1 \\ F & 2 & 2 & \frac{1}{2} & 1 & 1 & 1 \end{matrix}$$

(c) range [0.0094162, 0.0125136]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{2} & \frac{1}{3} \\ S & 2 & 1 & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{2} \\ C & 3 & 2 & 1 & 1 & 1 & 1 \\ B & 4 & 2 & 1 & 1 & 2 & 1 \\ G & 2 & 1 & 1 & \frac{1}{2} & 1 & 1 \\ F & 3 & 2 & 1 & 1 & 1 & 1 \end{matrix}$$

(d) range [0.0125136, 0.0165975]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & 1 & 1 \\ S & 2 & 1 & 1 & 1 & 2 & 4 \\ C & 2 & 1 & 1 & 1 & 2 & 3 \\ B & 3 & 1 & 1 & 1 & 2 & 4 \\ G & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 2 \\ F & 1 & \frac{1}{4} & \frac{1}{3} & \frac{1}{4} & \frac{1}{2} & 1 \end{matrix}$$

(e) range [0.0165975, 0.026264]

$$\begin{matrix} & f & S & C & B & G & F \\ f & 1 & 7 & 7 & 9 & 9 & 9 \\ S & \frac{1}{7} & 1 & 2 & 3 & 9 & 9 \\ C & \frac{1}{7} & \frac{1}{2} & 1 & 2 & 3 & 6 \\ B & \frac{1}{9} & \frac{1}{3} & \frac{1}{2} & 1 & 2 & 4 \\ G & \frac{1}{9} & \frac{1}{9} & \frac{1}{3} & \frac{1}{2} & 1 & 2 \\ F & \frac{1}{9} & \frac{1}{9} & \frac{1}{6} & \frac{1}{4} & \frac{1}{2} & 1 \end{matrix}$$

(f) range [0.026264, 1]

Fig. 9. Spell error words count ratio matrices

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 2 & 1 & 2 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{2} & 1 & \frac{1}{3} & \frac{1}{4} \\ 1 & 2 & 1 & 2 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{2} & 1 & \frac{1}{3} & \frac{1}{5} \\ 2 & 3 & 2 & 3 & 1 & 1 \\ 3 & 4 & 3 & 5 & 1 & 1 \end{pmatrix} \end{matrix}$$

(a) range [0, 0.186047]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & \frac{1}{3} & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 3 & 2 & 1 & 1 & 1 & 2 \\ 2 & 2 & 1 & 1 & 1 & 1 \\ 3 & 2 & 2 & 1 & 1 & 2 \\ 2 & 2 & \frac{1}{2} & 1 & \frac{1}{2} & 1 \end{pmatrix} \end{matrix}$$

(b) range [0.186047, 0.245711]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{4} & \frac{1}{7} & \frac{1}{3} & \frac{1}{2} \\ 3 & 1 & 1 & \frac{1}{2} & 1 & 1 \\ 4 & 1 & 1 & \frac{1}{2} & 1 & 2 \\ 7 & 2 & 2 & 1 & 3 & 3 \\ 3 & 1 & 1 & \frac{1}{3} & 1 & 1 \\ 2 & 1 & \frac{1}{2} & \frac{1}{3} & 1 & 1 \end{pmatrix} \end{matrix}$$

(c) range [0.245711, 0.282609]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} \\ 2 & 1 & 2 & \frac{1}{2} & 3 & 1 \\ 1 & \frac{1}{2} & 1 & \frac{1}{3} & 1 & \frac{1}{2} \\ 1 & 2 & 3 & 1 & 3 & 1 \\ 1 & \frac{1}{3} & 1 & \frac{1}{3} & 1 & \frac{1}{3} \\ 2 & 1 & 2 & 1 & 3 & 1 \end{pmatrix} \end{matrix}$$

(d) range [0.282609, 0.325112]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 2 \\ \frac{1}{2} & 1 & 1 & 1 & 1 & 2 \\ \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} & 1 \\ \frac{1}{2} & 1 & 1 & 2 & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{2} & 1 \end{pmatrix} \end{matrix}$$

(e) range [0.325112, 0.413953]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & 2 & 9 & 4 & 8 \\ 1 & 1 & 2 & 9 & 2 & 6 \\ \frac{1}{2} & \frac{1}{2} & 1 & 9 & 2 & 3 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & 6 & 1 & 2 \\ \frac{1}{8} & \frac{1}{6} & \frac{1}{3} & 3 & \frac{1}{2} & 1 \end{pmatrix} \end{matrix}$$

(f) range [0.413953, 1]

Fig. 24. Minor edit ratio matrices

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & 2 & 6 & 2 & 7 \\ 1 & 1 & 1 & 5 & 2 & 5 \\ \frac{1}{2} & 1 & 1 & 3 & 1 & 3 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{3} & 1 & \frac{1}{4} & 1 \\ \frac{1}{2} & \frac{1}{2} & 1 & 4 & 1 & 3 \\ \frac{1}{7} & \frac{1}{5} & \frac{1}{3} & 1 & \frac{1}{3} & 1 \end{pmatrix} \end{matrix}$$

(a) range [0.0050503551496702]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & 1 & 3 & 4 & 2 \\ 1 & 1 & 1 & 3 & 4 & 2 \\ 1 & 1 & 1 & 2 & 2 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 & 2 & 2 & 1 \end{pmatrix} \end{matrix}$$

(b) range [4.7817300862506]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 2 & 1 & 2 & 3 & 3 \\ \frac{1}{2} & 1 & 2 & 2 & 2 & 3 \\ 1 & \frac{1}{2} & 1 & 1 & 2 & 2 \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & 1 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & 1 \end{pmatrix} \end{matrix}$$

(c) range [6.9065826357179]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{2} & 2 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 \\ 2 & 1 & 1 & 2 & 2 & 2 \\ 2 & 1 & \frac{1}{2} & 1 & 2 & 2 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 1 \\ 1 & 1 & \frac{1}{2} & \frac{1}{2} & 1 & 1 \end{pmatrix} \end{matrix}$$

(d) range [8.0628548959919]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ 9 & 1 & \frac{1}{3} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ 9 & 3 & 1 & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \\ 9 & 7 & 3 & 1 & 2 & 2 \\ 9 & 7 & 2 & \frac{1}{2} & 1 & 2 \\ 9 & 7 & 2 & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} \end{matrix}$$

(e) range [9.6137313229325]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ 1 & 1 & \frac{1}{5} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ 9 & 5 & 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{9} \\ 9 & 9 & 2 & 1 & \frac{1}{2} & \frac{1}{3} \\ 9 & 9 & 5 & 2 & 1 & \frac{1}{2} \\ 9 & 9 & 9 & 3 & 2 & 1 \end{pmatrix} \end{matrix}$$

(f) range [11.414115233384]

Fig. 25. Article age matrices

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 9 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & 1 & 4 & 4 & 4 & 4 \\ \frac{1}{9} & \frac{1}{4} & 1 & 1 & 1 & 1 \\ \frac{1}{9} & \frac{1}{4} & 1 & 1 & 1 & 1 \\ \frac{1}{9} & \frac{1}{4} & 1 & 1 & 1 & 1 \\ \frac{1}{9} & \frac{1}{4} & 1 & 1 & 1 & 1 \end{pmatrix} \end{matrix}$$

(a) range [0, 0.0137012]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{3} & 1 & 2 & 5 & 9 \\ 3 & 1 & 2 & 6 & 9 & 9 \\ 1 & \frac{1}{2} & 1 & 3 & 6 & 9 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} & 1 & 2 & 9 \\ \frac{1}{5} & \frac{1}{9} & \frac{1}{6} & \frac{1}{2} & 1 & 7 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{7} & 1 \end{pmatrix} \end{matrix}$$

(b) range [0.0137012, 0.0355535]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{7} & \frac{1}{9} & \frac{1}{7} & \frac{1}{5} & 9 \\ 7 & 1 & \frac{1}{2} & 1 & 2 & 9 \\ 9 & 2 & 1 & 1 & 4 & 9 \\ 7 & 1 & 1 & 1 & 2 & 9 \\ 5 & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} & 1 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 \end{pmatrix} \end{matrix}$$

(c) range [0.0355535, 0.0805251]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ 9 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 \\ 9 & 2 & 1 & 1 & \frac{1}{2} & 2 \\ 9 & 2 & 1 & 1 & \frac{1}{2} & 3 \\ 9 & 2 & 2 & 2 & 1 & 3 \\ 9 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 1 \end{pmatrix} \end{matrix}$$

(d) range [0.0805251, 0.170719]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{7} & \frac{1}{9} & \frac{1}{7} & \frac{1}{9} \\ 2 & 1 & \frac{1}{3} & \frac{1}{9} & \frac{1}{4} & \frac{1}{9} \\ 7 & 3 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{5} \\ 9 & 9 & 2 & 1 & 3 & \frac{1}{2} \\ 7 & 4 & 2 & \frac{1}{3} & 1 & \frac{1}{3} \\ 9 & 9 & 5 & 2 & 3 & 1 \end{pmatrix} \end{matrix}$$

(e) range [0.170719, 0.48096]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & \frac{1}{8} & \frac{1}{8} & \frac{1}{9} & \frac{1}{9} \\ 1 & 1 & \frac{1}{8} & \frac{1}{8} & \frac{1}{9} & \frac{1}{9} \\ 8 & 8 & 1 & \frac{1}{2} & \frac{1}{6} & \frac{1}{9} \\ 8 & 8 & 2 & 1 & \frac{1}{3} & \frac{1}{9} \\ 9 & 9 & 6 & 3 & 1 & \frac{1}{2} \\ 9 & 9 & 9 & 9 & 2 & 1 \end{pmatrix} \end{matrix}$$

(f) range [0.48096, 9.76467]

Fig. 26. Edit frequency matrices

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & \frac{1}{2} & 1 & 1 & 2 \\ 1 & 1 & \frac{1}{2} & \frac{1}{2} & 1 & 2 \\ 2 & 2 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 1 \end{pmatrix} \end{matrix}$$

(a) range [4.0620243531202E-5]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{5} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{8} \\ 5 & 1 & 2 & 1 & \frac{1}{2} & \frac{1}{2} \\ 3 & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 6 & 1 & 2 & 1 & 1 & 1 \\ 6 & 2 & 2 & 1 & 1 & 1 \\ 8 & 2 & 2 & 1 & 1 & 1 \end{pmatrix} \end{matrix}$$

(b) range [5.072203196347]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & 1 & \frac{1}{2} \\ 2 & 1 & 1 & \frac{1}{2} & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & \frac{1}{2} & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{matrix}$$

(c) range [5.3747463216641]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \\ 1 & 1 & 1 & 1 & 1 & \frac{1}{2} \\ 2 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & \frac{1}{2} \\ 3 & 2 & 1 & 1 & 2 & 1 \end{pmatrix} \end{matrix}$$

(d) range [5.6557902080162]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 3 & 2 & 1 \\ 1 & 1 & 1 & 3 & 2 & 1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 1 & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 \end{pmatrix} \end{matrix}$$

(e) range [6.0220700152207]

$$\begin{matrix} & \begin{matrix} f & S & C & B & G & F \end{matrix} \\ \begin{matrix} f \\ S \\ C \\ B \\ G \\ F \end{matrix} & \begin{pmatrix} 1 & 3 & 4 & 6 & 3 & 9 \\ \frac{1}{3} & 1 & 2 & 3 & 1 & 9 \\ \frac{1}{4} & \frac{1}{2} & 1 & 2 & 1 & 7 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 1 & \frac{1}{3} & 4 \\ \frac{1}{3} & 1 & 1 & 3 & 1 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{7} & \frac{1}{4} & \frac{1}{9} & 1 \end{pmatrix} \end{matrix}$$

(f) range [6.5134449518011]

Fig. 27. Average active age of editors matrices

$$\begin{array}{c}
f \\
S \\
C \\
B \\
G \\
F
\end{array}
\begin{pmatrix}
f & S & C & B & G & F \\
1 & 1 & 3 & 1 & 1 & 2 \\
1 & 1 & 4 & 2 & 2 & 3 \\
\frac{1}{3} & \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
1 & \frac{1}{2} & 2 & 1 & 2 & 1 \\
1 & \frac{1}{2} & 2 & \frac{1}{2} & 1 & 1 \\
\frac{1}{2} & \frac{1}{3} & 2 & 1 & 1 & 1
\end{pmatrix}$$

(a) range [0, 1476.23]

$$\begin{array}{c}
f \\
S \\
C \\
B \\
G \\
F
\end{array}
\begin{pmatrix}
f & S & C & B & G & F \\
1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \\
2 & 1 & 1 & \frac{1}{2} & 1 & \frac{1}{2} \\
2 & 1 & 1 & 1 & 1 & \frac{1}{2} \\
2 & 2 & 1 & 1 & 2 & 1 \\
2 & 1 & 1 & \frac{1}{2} & 1 & \frac{1}{2} \\
3 & 2 & 2 & 1 & 2 & 1
\end{pmatrix}$$

(b) range [1476.23, 1997.37]

$$\begin{array}{c}
f \\
S \\
C \\
B \\
G \\
F
\end{array}
\begin{pmatrix}
f & S & C & B & G & F \\
1 & 3 & 1 & 1 & 1 & 1 \\
\frac{1}{3} & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & \frac{1}{2} \\
1 & 2 & 1 & 1 & 1 & 1 \\
1 & 3 & 1 & 1 & 1 & 1 \\
1 & 3 & 1 & 1 & 1 & 2 \\
1 & 2 & 1 & 1 & \frac{1}{2} & 1
\end{pmatrix}$$

(c) range [1997.37, 2644]

$$\begin{array}{c}
f \\
S \\
C \\
B \\
G \\
F
\end{array}
\begin{pmatrix}
f & S & C & B & G & F \\
1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{3} & 1 & \frac{1}{3} \\
2 & 1 & \frac{1}{2} & \frac{1}{2} & 2 & 1 \\
4 & 2 & 1 & 2 & 3 & 1 \\
3 & 2 & \frac{1}{2} & 1 & 3 & 1 \\
1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 1 & \frac{1}{2} \\
3 & 1 & 1 & 1 & 2 & 1
\end{pmatrix}$$

(d) range [2644, 3793.46]

$$\begin{array}{c}
f \\
S \\
C \\
B \\
G \\
F
\end{array}
\begin{pmatrix}
f & S & C & B & G & F \\
1 & 1 & 1 & 1 & \frac{1}{2} & 1 \\
1 & 1 & 1 & 2 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 \\
1 & \frac{1}{2} & 1 & 1 & \frac{1}{2} & \frac{1}{2} \\
2 & 1 & 1 & 2 & 1 & 1 \\
1 & 1 & 1 & 2 & 1 & 1
\end{pmatrix}$$

(e) range [3793.46, 6968.28]

$$\begin{array}{c}
f \\
S \\
C \\
B \\
G \\
F
\end{array}
\begin{pmatrix}
f & S & C & B & G & F \\
1 & 2 & 2 & 3 & 1 & 4 \\
\frac{1}{2} & 1 & 1 & 2 & 1 & 2 \\
\frac{1}{2} & 1 & 1 & 2 & 1 & 3 \\
\frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{3} & 1 \\
1 & 1 & 1 & 3 & 1 & 3 \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{3} & 1 & \frac{1}{3} & 1
\end{pmatrix}$$

(f) range [6968.28, 10276]

Fig. 30. Average talk times of editors matrices