Characterizing Wikipedia Pages Using Edit Network Motif Profiles

Guangyu Wu, Martin Harrigan, Pádraig Cunningham
Clique Research Cluster
School of Computer Science & Informatics
University College Dublin, Ireland
{guangyu.wu, martin.harrigan, padraig.cunningham}@ucd.ie

ABSTRACT

Good Wikipedia articles are authoritative sources due to the collaboration of a number of knowledgeable contributors. This is the *many eyes* idea. The edit network associated with a Wikipedia article can tell us something about its quality or authoritativeness. In this paper we explore the hypothesis that the characteristics of this edit network are predictive of the quality of the corresponding article's content. We characterize the edit network using a profile of network motifs and we show that this network motif profile is predictive of the Wikipedia quality classes assigned to articles by Wikipedia editors. We further show that the network motif profile can identify outlier articles particularly in the 'Featured Article' class, the highest Wikipedia quality class.

Categories and Subject Descriptors: H.5.0 [Information Interfaces and Presentation]: General; H.2.8 [Database Management]: Data Mining

General Terms: Algorithms, Experimentation

Keywords: Wikipedia, Authoritativeness, Network motifs

1. INTRODUCTION

While the impact and widespread uptake of Wikipedia is undeniable, the reliability of much of the content available on Wikipedia has been the subject of intense debate. Perhaps the epicentre of this debate has been the *Nature* article favorably comparing the quality of Wikipedia entries to those in the Encyclopdia Britannica [7] and the robust response to this article from the Britannica owners[1]. The truth is that sometimes Wikipedia is effective at harnessing the *wisdom of crowds* to produce excellent authoritative articles and sometimes Wikipedia articles are quite poor – often because they are not the result of much collaboration.

In this paper we set out to examine to what extent the quality of a Wikipedia article can be determined from looking at the network of edits around an article. We explore the hypothesis that a characterization of the edit network in terms of network motif profiles will be predictive of the quality of the article. For this reason we consider network motifs on their own rather than in combination with other features such as content-based features that have been shown to be predictive of quality in Wikipedia.

The network motif profile comprises a count of the occurrences of the 17 network motifs shown in Figure 2. Two example edit networks are shown in Figure 1. Since these networks are centered on a specific article they can be considered *ego* networks for that article. Both ego networks are from the article covering the 2011 Japanese earthquake and tsunami [2]. The network on the left is from two days after the disaster and the one on the right is after ten days. The article was initially rated C class on the Wikipedia quality scale (see section 4) but was re-rated B class after ten days. A characteristic of the C class article is that (at least at that stage) the contributors had not collaborated on other Wikipedia articles.

While the majority of Wikipedia articles carry these quality class markers, there is a vast array of Wikipedia content that is not rated and much of it is of dubious or poor quality. This is a problem in Wikipedia articles on entertainment, sport and popular culture but it is even a problem in science articles where Wikipedia normally works well. Two examples relevant for this paper are the articles on network motifs and the k-nearest neighbor algorithm. At the time of writing, these articles are not necessarily inaccurate but they are not comprehensive or authoritative. It is clear from an examination of the edit history of these articles that they have not received much attention from Wikipedia contributors.

In the next section we provide a brief review of some relevant research on social network analysis and authoritativeness. In section 3 we outline the procedures we employ to represent Wikipedia articles in terms of edit networks. Section 4 provides some background on the Wikipedia quality scale and describes the datasets we use in our study. The usefulness of this network motif-based representation for classification is evaluated in section 5 and an assessment of the discriminating power of the different network motifs is presented in section 6. We visualize the data using 2-D

[1]Fatally Flawed: Refuting the Recent Study on Encyclopedic Accuracy by the Journal *Nature*, `http://corporate.britannica.com/britannica_nature_response.pdf`

[2]The Tōhoku earthquake and tsunami: `http://bit.ly/grGP84`

SMUC'11, October 28, 2011, Glasgow, Scotland, UK.

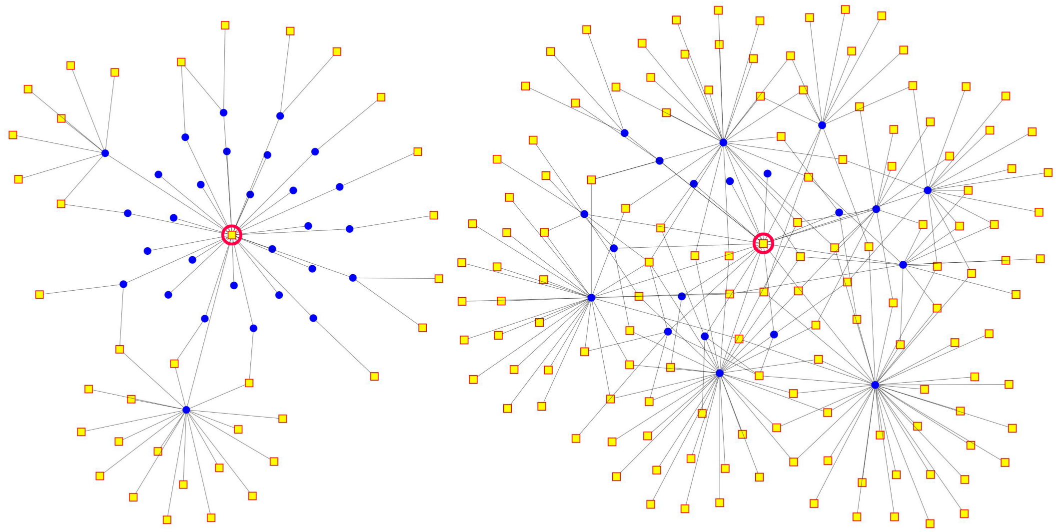

Figure 1: The edit networks for two versions of the 2011 Tōhoku earthquake and tsunami Wikipedia article. The network on the left is from two days after the earthquake and the one on the right is after ten days. The blue circles are the editors and the yellow squares are the articles.

spatializations in section 7. The paper concludes in section 8 with a summary and an outline of future work.

2. SOCIAL NETWORK ANALYSIS AND AUTHORITATIVENESS

The defining characteristic of Web 2.0 is the sharing of user generated content (UGC) in virtual communities. This is enormously disruptive as resources such as TripAdvisor, Wikipedia and MenuPages replace traditional businesses that provided similar services. At the same time, this raises questions regarding the quality of UGC [3]. Is an answer on *Yahoo! Answers* correct? Are details presented in blogs or on Wikipedia credible? Can we trust reviews on TripAdvisor or on Amazon? These questions have received a good deal of attention in recent research. We focus on the work related to Wikipedia here.

In contrast to traditional encyclopedias, where authority derives from expert contributors, Wikipedia depends on collaboration and consensus to produce quality articles. There has been some controversial research that suggests that the quality of Wikipedia articles approaches that of established encyclopedias [7]. The famous quote from Surowiecki's *The Wisdom of Crowds* is that "under the right circumstances, groups are remarkably intelligent" [15]. The challenge when using Wikipedia is to determine whether the circumstances that produced the article in question were *right* or not.

The first requirement for a Wikipedia article to be authoritative is the *many eyes* idea [11] – a significant group of contributors must have cooperated to produce the article. It is also important that this collaboration has been constructive and it is better if the editors have a reasonable reputation as contributors. Adler and De Alfaro [1] have pursued a content driven strategy to assess editor reputation. They have used text survival and edit distance to quantify editor reputation. In later work [2] they show that edit longevity is a good measure of editor contribution.

Korfiatis et al. [9] pursue a network-based strategy to evaluate authoritative sources in Wikipedia. They construct a two-mode network of articles and contributors. The article nodes are linked by hyperlinks and contributors are linked if they have worked on the same article. Contributors are also linked to articles on which they worked. The study proposes article and contributor degree centrality as indicators of authoritativeness. This is similar in spirit to the strategy in our work as degree centrality is captured by a subset of the network motifs we consider.

Brandes et al. [4] have also analysed the collaboration structure in Wikipedia. Their work has focused on the edit interactions on individual articles. Edges between individual contributors represent *delete*, *undelete* and *restore* interactions. The main contribution of this work is to present the notion of bipolarity that captures the level of conflict between the contributors to an article. Thus the work is more directed at the problem of Wikipedia vandalism than the issue of authoritativeness that is the subject of this paper.

Recently, Laniado et al. [10] presented an algorithm that assigns scores to all contributors of a Wikipedia article according to their contribution, and selects the top contributors to build a collaboration network of authors where edges represent the co-authorship between authors. Thus the inexperienced authors are filtered out and the co-authorship networks becomes more informative. With the exception of eigenvector centrality (where edge weights were considered) the features they extracted were taken from unweighted versions of the networks.

Dalip et al. [6] presented a comprehensive of assessment of quality indicators in collaborative content curation with a focus on Wikipedia. In their analysis they considered 69 indicators including text features, review features and basic network features. They used a machine learning approach to discover the most effective indicators and combination of indicators. They found that the easy-to-extract text-based features were most informative – more informative that more complex features based on link analysis.

There also exists non-network based studies, for example, Lipka et al. [12] used machine learning techniques to identify featured articles using character trigram and part-of-speech trigram vectors. These features that are known to be characteristic of writing style out-performed alternatives in both a single domain and a domain transfer situation with F-measure scores of 0.88 across domains and good performance on articles of varying length.

This work by Dalip et al. [6] and Lipka et al. [12] is complementary to ours in that our network-based features can be combined with their content-based features to further improve classification accuracy.

3. NETWORK MOTIF PROFILES

The idea of characterizing networks in terms of network motif profiles is well established and has had a considerable impact in bioinformatics [14]. Our objective is to characterize Wikipedia articles in terms of edit network motif profiles and then examine whether or not articles at different quality levels have characteristic network motif profiles.

3.1 Wikipedia Network Motifs

Our Wikipedia network motifs comprise editor and article nodes and editor-article edges (see Figure 2). The editor-article edges represent edit activities on Wikipedia articles. The network is bipartite since there are no between-editor edges or hyperlink edges between articles. Hyperlink edges were excluded from consideration because earlier analysis has found that hyperlink density can dominate the network motif profiles and, from a quality perspective, this is not an interesting distinction between articles[17].

Given that this is a bipartite network, the edges are undirected with an editor at one end and an article at the other, we generated a set of possible network motifs. The complete set of network motifs of up to five nodes is shown in Figure 2.

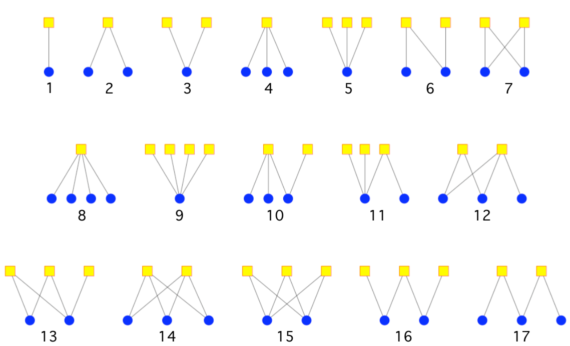

Figure 2: The network motifs used to characterize the Wikipedia articles. All network motifs up to five nodes are considered. Yellow squares are articles and blue circles are editors.

We used `nauty` [13] to enumerate all network motifs up to five nodes without considering node labels. There are 31 unlabeled network motifs with between one and five nodes. When we allow nodes to be either 'editor' or 'article' these 31 unlabeled network motifs produce 419 two-labeled network motifs. When single node network motifs and network motifs with editor-editor or article-article edges are removed the set reduces to the 17 network motifs in Figure 2.

3.2 Data Gathering

It is possible to produce edit networks such as shown in Figure 1 because Wikipedia provides a complete edit history of all Wikipedia articles. If the editor is logged in to Wikipedia at the time of the edit then the edit history references the editor's user name. Editors not logged in to Wikipedia are identified by their IP address – since these will not result in interesting network structure these nodes are dropped from the network. For the purpose of this study we selected almost 2,000 articles from the History and United States categories as summarised in Table 2[3].

The analysis is based on the network motif profiles of the ego networks surrounding these articles using the latest 100 revisions from the edit history. The ego network for an article is developed by considering other articles connected by hyperlinks to the core article. Hyperlinked articles are dropped from the network if they have not been edited by editors of the core article. For both the core article and the hyperlinked articles the edit window considered covers 100 revisions.

For most articles this edit window will cover months or years of activity. However edit activity is quite 'bursty' and some articles can go through 100 revisions in a few days, for instance, the edit histories on the two versions of the Japan earthquake and tsunami article had no overlap even though the second version came just one week after the first one. Nevertheless this snapshot of the edit behavior seems to offer a useful characterization of the article.

The most computationally intensive part of this analysis is the generation of the network motif profiles for each ego network. For this task we use `GraphGrep` [8], a graph query tool that requires a query graph and a target graph and returns the number of occurrences of the query graph in the target graph. Each ego network is considered as a target graph for `GraphGrep` while the 17 network motifs are used as the query graphs. When using `GraphGrep` for network motif counting care must be taken to handle graph automorphisms. For instance, `GraphGrep` returns six when both the target and query graphs are simple triangles. To correct for this, each count is divided by the number of automorphisms of the query graphs.

Because the time taken to extract the network motif profiles increases more than linearly with the size of the network the analysis has been restricted to articles with < 300 nodes in their edit networks. This entailed dropping about 5% of articles from the analysis. As might be expected, there were very few articles with large edit networks in the Start class – less than 1% of these articles exceeded the 300 node threshold.

4. DATASETS

Before describing the datasets used in this evaluation it is worth going into some detail on the Wikipedia quality scale. The main classes in the Wikipedia quality scale are shown in Table 1[4]. In addition to these categories there are about 70,000 "List" articles in the English language Wikipedia and somewhere between a half million and one million articles that have not been assessed. Even though Stub articles form the largest category we do not consider them in this study

[3] `http://bit.ly/m4bLZL` and `http://bit.ly/kdNJKG`
[4] The article counts presented in this table were taken from `http://bit.ly/1avQfU`.

Table 2: Datasets analyzed

	Good		Medium	
Dataset	Classes	Count	Classes	Count
History-F-S	F	122	Start	297
History-FB-S	F&B	572	Start	297
US-FB-S	F&B	713	Start	389
Mixed-FB-S	F&B	1285	Start	686

as they are really just placeholders for content that may be added in the future. So we have taken the Start class as our basic category and Featured Articles and B class articles as representative of good quality content. We have included the B class here as so many articles are included in this category and these articles are "mostly complete" and are useful sources of information.

In this study we have gathered data from the History and United States categories on Wikipedia. As stated already, for our classification evaluation we have taken Start class articles are representative of typical *work in progress* articles while Featured Articles and B class articles are considered authoritative. The details of the four datasets considered in the study are shown in Table 2. The first dataset presents the simplest classification task; this entails distinguishing Featured Articles in the History category from Start articles. In the next section we show that classification accuracies of better than 85% are possible on this task. The second and third datasets entail distinguishing Featured Articles and B class articles from Start class articles in the History and United States collections. The fourth dataset comprises a mix of United States and History articles and is the hardest classification task.

5. CLASSIFICATION EXPERIMENTS

There are two aspects to the machine learning analysis covered in this paper. In the next section we present some results on feature subset selection that highlight the discriminating network motifs in the profile but first we present some results on classification using the network motif profiles.

For the classification analysis (Table 3) we consider three methods: random forest (100 trees); logistic regression and k-nearest-neighbor (k-NN). We report performance from 10-fold cross validation tests in terms of overall accuracy and ROC area – ROC area is relevant because overall accuracy may be misleading when datasets are imbalanced. Random forest is included because it is an ensemble method that can be expected to give very good performance. Logistic regression is included because it is a simple method that should also perform well and offers some insight into how features contribute to the classification. k-NN is considered because the classes may be diverse and a local learner may be expected to work well in some circumstances.

The purpose of this evaluation is not to identify the best classifier for this task but to get an assessment of what classification accuracy is achievable using network motif profiles. The best accuracy figure in the evaluation is the 85.7% figure achieved with logistic regression on the 'easy' Featured versus Start class articles in the History collection. When the B class articles are included alongside the Featured Articles the accuracy falls to 77.4% with the ROC area falling to 0.85 from 0.92. The accuracy is about the same on the United States dataset, the main difference being that the random forest and k-NN classifiers also do well on that dataset. When the two datasets are mixed the accuracy falls by only one or two percent indicating that the edit network characteristics of good articles are reasonably consistent across domains.

6. IDENTIFYING DISCRIMINATING NETWORK MOTIFS

In this section we analyze the contribution of the different network motifs to classification accuracy. We use feature ranking using information gain, wrapper-based feature subset selection and principle component analysis [5] for this assessment. The complete set of network motifs is shown in Figure 2. However the network motif 1 is not considered because enumerating it is simply an edge count in the ego-network. We focus on the Mixed-FB-S dataset as it is the largest and in addition we are interested in identifying features that are discriminating across domains.

The first finding is that all features show some information gain, i.e. all network motifs have some potential to discriminate between good and basic Wikipedia articles in the mixed dataset (see Figure 3). The five best motifs are 9, 5, 10, 3 and 17 and the weakest network motif is 14. 4 and 8 are stars with articles at their centre while 3, 5 and 9 are stars with editors at their centre. This is interesting because it shows that *many eyes* is not really the defining characteristic of quality, instead experience is important – the editors should have worked on many other articles. This point is reinforced when we look at the two densest network motifs, 14 and 15. The network motif with three articles and two editors is better than the one with three editors and two articles. Network motif 15 describes a situation where two contributors have co-edited three articles. This kind of cooperation over a number of different articles is characteristic of good quality articles. Indeed an examination of the B class and C class networks in Figure 1 shows that this is a differentiating factor.

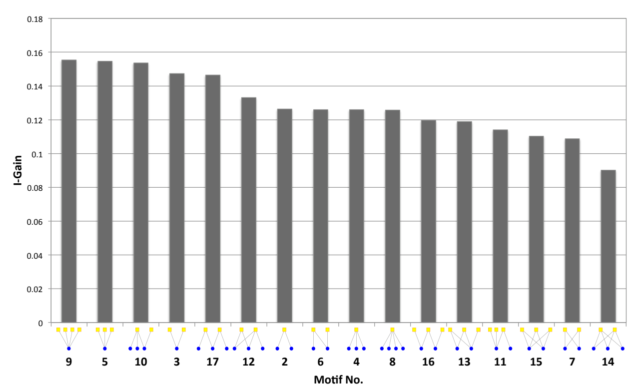

Figure 3: The information gain for each of the network motifs on the Mixed-FB-S dataset. Network motifs 3, 5 and 9 are stars (see Figure 2) indicating editors who have worked on multiple articles.

Following on from this information gain analysis we used wrapper-based feature subset selection to identify a com-

Table 1: Wikipedia quality classes. The final column indicates the percentage of all articles above Stub class made up of that class.

Class	Summary	Count	Percent
Featured Article	Professional, outstanding, and thorough; a definitive source for encyclopedic information.	3,480	0.4%
A	"A fairly complete treatment of the subject."	785	0.1%
Good Article	"Useful to nearly all readers, with no obvious problems..."	11,517	1.3%
B	"The article is mostly complete and without major issues..."	70,854	8%
C	"The article is substantial, but is still missing important content or contains a lot of irrelevant material."	83,567	10%
Start	"An article that is developing, but which is quite incomplete and may require further reliable sources."	687,914	80%
Stub	"A very basic description of the topic."	1,743,128	-

Table 3: Classification Results

	Random Forest		Logistic		k-NN	
Dataset	Accuracy	ROC Area	Accuracy	ROC Area	Accuracy	ROC Area
History-F-S	82.3%	0.88	85.7%	0.92	77.3%	0.79
History-FB-S	74.5%	0.80	77.4%	0.85	68.7%	0.73
US-FB-S	77.9%	0.84	77.8%	0.82	76.5%	0.80
Mixed-FB-S	76.0%	0.80	76.1%	0.81	74.1%	0.76

pact discriminating subset of features. Different wrapper-based feature subset selection strategies [5] (e.g. genetic or greedy search) return different feature subsets that yield classification performance that is roughly equivalent. This indicates that there is considerable redundancy is the representation. Indeed this is evident from an examination of Figure 2. Clearly, the counts of network motifs 5 and 9 will be correlated.

We explore this correlation further through a principle component analysis on the data. This shows that the first three principle components capture respectively 67%, 19% and 9% of the variation in the data. The fact that 95% of the variation in the data is captured in the first three principle components suggests that, for a given analysis task, some network motifs could be dropped from the analysis without losing discriminating power.

7. VISUALIZATION

We visualize the datasets using 2-D spatializations. Before producing the spatializations we need to normalize the network motif profiles. For each network motif profile, we compute a *network ratio profile* [14]: a 17-element vector where each entry is the *normalized ratio* of the corresponding entry in the network motif profile. The ratio profile rp of an ego network is computed using

$$rp_i = \frac{nmp_i - \overline{nmp_i}}{nmp_i + \overline{nmp_i} + \epsilon}$$

where nmp_i is the ith entry of the network motif profile, $\overline{nmp_i}$ is the average of the ith entry of all of the network motif profiles, and ϵ is a small integer that ensures that the ratio is not misleadingly large when the network motif appears very few times in all of the ego networks. To adjust for scaling the normalized ratio profile nrp of an ego network is computed using

$$nrp_i = \frac{rp_i}{\sqrt{\sum rp_j^2}}.$$

A normalized ratio measures the abundance of a network motif in each individual ego network relative to all the ego networks; it is similar to a z-score. It does not require the construction of a random network ensemble as found in related approaches [14]. There are correlations between the elements of a network ratio profile. To adjust for these, we compute a PCA. We use the first two eigenvectors as the axes for our spatializations.

While the spatializations in Figure 4 show reasonably good clustering of the different classes in two dimensions there is also some significant spread of the classes. For the History dataset, there is good clustering of the B class to the top right and of Start class articles to the bottom left. However, the Featured Articles are more spread out with clusters to the left and right marked L and R. To explore this further we examined ten Featured Articles from the extreme left (L) and ten from the extreme right (R) of the figure (see Table 4). The Wikipedia quality scale has a further minor dimension, which reflects the importance of an article and it transpires that the Featured Articles on the left are inclined to be low or mid importance compared to high importance articles on the right. This niche characteristic is emphasized by the fact that these articles are inclined not to have been featured on the Wikipedia main page. We conclude from this that, at least in edit network terms, some low importance Featured Articles *look* like more ordinary articles.

The spatializations can be used to identify anomalous or outlier articles. We selected two Start class articles that are completely surrounded by B class articles or Featured Articles – these are marked as X and Y in Figure 4. It seems that the anomalous position of the article at X[5] is due van-

[5] `http://bit.ly/mhOydo`.

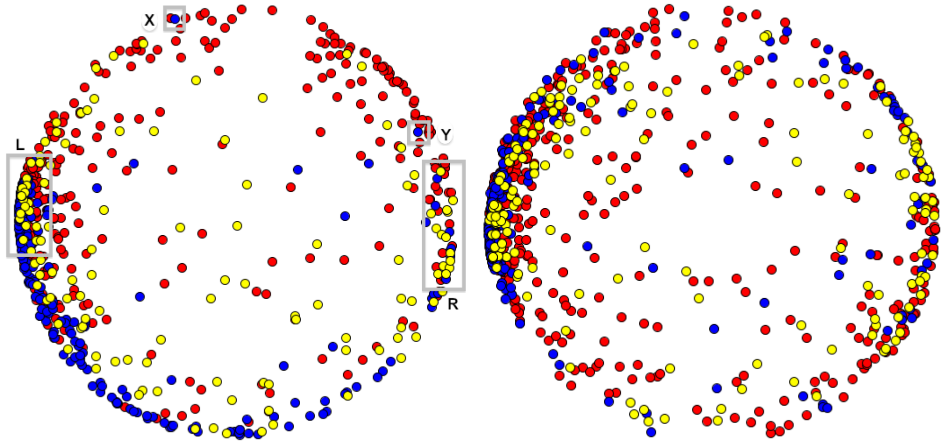

Figure 4: Spatializations of the History (left) and United States (right) articles produced using PCA. The Featured Articles are yellow; the B class articles are red; and the Start class articles are blue.

dalism. It is an article on antisemitism and much of its edit history can be attributed to vandalism. The situation with the article at Y[6] is more difficult to explain. The article is made up mostly of lists of hyperlinks and it may be that this hyperlink structure gives it the superficial characteristics of a higher quality article. Some of the edit activity on Wikipedia is carried out by bots that carry out mundane editing tasks. These bots are quite active on Start class articles so Start class articles with a lot of hyperlinks can quickly build up a dense network – mostly from bot activity.

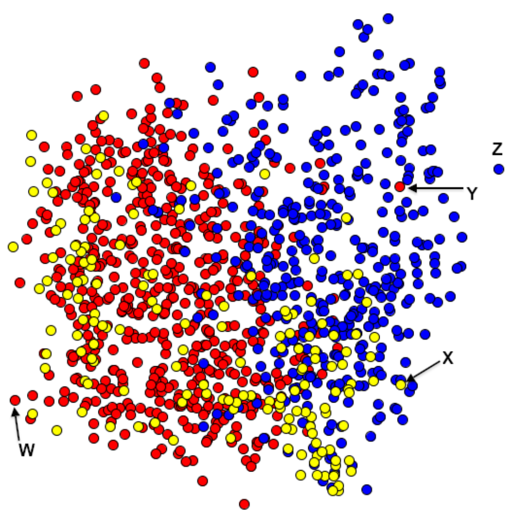

Figure 5: A spatialization of the United States dataset using a locality preserving projection. The Featured Articles are yellow; the B class articles are red; and the Start class articles are blue.

Because the spatialization of the United States dataset is disappointing in terms of separating the B class and Start class articles we present an alternative view of that data using a locality preserving projection in Figure 5. This spatialization was computed using Isomap [16]. This spatialization

Table 4: History Featured Article breakdown

	Importance				Main Page?	
	Low	Mid	High	Top	No	Yes
Left	4	6	0	0	5	5
Right	1	2	6	1	3	7

produces a much better clustering of the B class and Start class articles. Again the Featured Articles separate into two clusters that roughly correspond to important articles on the left and niche interest articles on the right. The rightmost Featured Article (marked X in Figure 5) is a low-importance article that is largely the work of one person[7]. The same is true of the rightmost B class article at Y[8]. It is an article with only 9 edits, two of which were by bots. Similarly the rightmost Start class article at Z has only 7 edits and is little more than a Stub article[9]. By contrast the *leftmost* B class article on the "Compromise of 1850" at W has 2,780 edits[10]. This seems like enough work to bring it up to Featured Article status but in fact much of the edit activity is due to a long series of edit wars.

8. CONCLUSIONS AND FUTURE WORK

We have presented some work on characterizing the quality of Wikipedia articles using network motif profiles. We have demonstrated that this network motif-based characterization can be used to classify good from ordinary quality articles with reasonable accuracy. For instance, a classification accuracy of 75% or better should be useful in a *triage* stage in assessing the more than half a million articles that have not been assigned any quality class.

The network motif-based analysis has offered a number of interesting insights. It has demonstrated that the *many eyes* idea is not of itself a sign of quality; it is important that the editors are experienced as evidenced by collaborations on other articles. The most *virtuous* network motifs are those

[6] `http://en.wikipedia.org/wiki/Archbishop_of_St_Andrews`.

[7] `http://bit.ly/ao5OQ5`

[8] `http://en.wikipedia.org/wiki/Louis_Ayres`

[9] `http://bit.ly/jvIb1C`

[10] `http://en.wikipedia.org/wiki/Compromise_of_1850`

that show contributors who have worked on other articles as well. The spatializations using PCA presented in section 7 are quite informative with regard to the History dataset but less so with regard to the United States dataset. In the History dataset the Start class and B class articles are pretty compact while there is considerable variety in the Featured Article class. It seems that articles on niche topics can reach Featured Article status without a huge amount of collaboration. Because the PCA-based spatialization of the United States dataset is disappointing we used an alternative locality preserving projection that provides reasonable clustering of the classes and reveals some interesting outliers.

In future work we will experiment with removing Wikipedia bots from the analysis as they can produce a false impression of collaborative edit activity. We will also work on improving the efficiency of the network motif counting procedure in order to make the method applicable to much larger networks. However the most promising direction to take this research is to combine these network-based features with the content-based features reported by Dalip et al. [6] and Lipka et al. [12] in order to produce an integrated system that will offer increased accuracy by combining multiple views on the data.

9. ACKNOWLEDGEMENTS

This work is supported by Science Foundation Ireland (SFI) Grant No. 08/SRC/I140 (Clique: Graph and Network Analysis Cluster).

10. REFERENCES

[1] B. Adler and L. De Alfaro. A content-driven reputation system for the Wikipedia. In *Proceedings of the 16th International Conference on World Wide Web*, page 270. ACM, 2007.

[2] B. Adler, L. de Alfaro, I. Pye, and V. Raman. Measuring author contributions to the Wikipedia. In *Proceedings of the 4th International Symposium on Wikis*, pages 1–10. ACM, 2008.

[3] R. Baeza-Yates. User Generated Content: How Good Is It? In *3rd Workshop on Information Credibility on the Web (WICOW 2009)*, pages 1–2, 2009.

[4] U. Brandes, P. Kenis, J. Lerner, and D. van Raaij. Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th International Conference on World Wide Web*, pages 731–740. ACM, 2009.

[5] P. Cunningham. Dimension Reduction. In M. Cord and P. Cunningham, editors, *Machine Learning Techniques for Multimedia*, Cognitive Technologies, pages 91–112. Springer Berlin Heidelberg, 2008.

[6] D. Dalip, M. Gonçalves, M. Cristo, and P. Calado. Automatic quality assessment of content created collaboratively by web communities: a case study of Wikipedia. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 295–304, 2009.

[7] J. Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, 2005.

[8] R. Giugno and D. Shasha. GraphGrep: A fast and universal method for querying graphs. In *International Conference on Pattern Recognition*, volume 16, pages 112–115, 2002.

[9] N. Korfiatis, M. Poulos, and G. Bokos. Evaluating authoritative sources using social networks: an insight from Wikipedia. *Online Information Review*, 30(3):252–262, 2006.

[10] D. Laniado and R. Tasso. Co-authorship 2.0: Patterns of collaboration in Wikipedia. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia*, pages 201–210. ACM, 2011.

[11] A. Lih. Wikipedia as participatory journalism: reliable sources? metrics for evaluating collaborative media as a news resource. In *In Proceedings of the 5th International Symposium on Online Journalism*, pages 16–17, 2004.

[12] N. Lipka and B. Stein. Identifying featured articles in Wikipedia: writing style matters. In *Proceedings of the 19th International Conference on World Wide Web*, pages 1147–1148. ACM, 2010.

[13] B. McKay. Practical graph isomorphism. *Congressus Numerantium*, 30(30):47–87, 1981.

[14] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538, 2004.

[15] J. Surowiecki, M. Silverman, et al. The wisdom of crowds. *American Journal of Physics*, 75:190, 2007.

[16] J. B. Tenenbaum, V. d. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[17] G. Wu, M. Harrigan, and P. Cunningham. A Characterization of Wikipedia Content Based on Motifs in the Edit Graph. Technical Report UCD-CSI-2011-02, University College Dublin, February 2011.