Online Social Networks and Media 27 (2022) 100178

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem

Understanding the characteristics of COVID-19 misinformation communities through graphlet analysis

James R. Ashford [a,*], Liam D. Turner [a], Roger M. Whitaker [a], Alun Preece [b], Diane Felmlee [c]

[a] School of Computer Science and Informatics, Cardiff University, UK
[b] Crime and Security Research Institute, Cardiff University, UK
[c] Department of Sociology & Criminology, Pennsylvania State University, USA

ARTICLE INFO

Keywords:
Reddit
COVID-19
Disinformation
Complex networks

ABSTRACT

Online social networks serve as a convenient way to connect, share, and promote content with others. As a result, these networks can be used with malicious intent, causing disruption and harm to public debate through the sharing of misinformation. However, automatically identifying such content through its use of natural language is a significant challenge compared to our solution which uses less computational resources, language-agnostic and without the need for complex semantic analysis. Consequently alternative and complementary approaches are highly valuable. In this paper, we assess content that has the potential for misinformation and focus on patterns of user association with online social media communities (subreddits) in the popular Reddit social media platform, and generate networks of behaviour capturing user interaction with different subreddits. We examine these networks using both global and local metrics, in particular noting the presence of induced substructures (graphlets) assessing 7,876,064 posts from 96,634 users. From subreddits identified as having potential for misinformation, we note that the associated networks have strongly defined local features relating to node degree — these are evident both from analysis of dominant graphlets and degree-related global metrics. We find that these local features support high accuracy classification of subreddits that are categorised as having the potential for misinformation. Consequently we observe that induced local substructures of high degree are fundamental metrics for subreddit classification, and support automatic detection capabilities for online misinformation independent from any particular language.

1. Introduction

Misinformation is a major cause for concern with potentially dangerous ramifications for social processes, including the stability of democracy [1,2]. As of writing, the COVID-19 pandemic has been the subject of various "fake news" stories and conspiracies resulting in an "infodemic" as described by the WHO (World Health Organisation)[1] [3–5]. This issue has become a serious threat to public health and has triggered multiple public responses including the destruction of 5G cellular towers in the UK[2] [4] and the proposition to reject potential vaccinations.[3] As part of this, the informality of online social media is well-suited to propagation of misinformation, which has been an unforeseen consequence of the technology's role in liberating global participation [6].

Accordingly, in this paper we focus on identifying the latent characteristics of bipartite network structures built from social media behaviour, and use Reddit as our data source. Our hypothesis is that *a network-based approach using frequency of bipartite graphlets can be used to distinguish between user activity surrounding general topics and user behaviour aligned to potential misinformation*. We firstly compare and contrast graphlet frequencies, alongside global metrics, between sets of subreddits potentially associated with misinformation and sets of sample subreddits not associated. From this, we build machine learning models to determine the predictive power of graphlet and global features in distinguishing between the activity of different sets of subreddits. This gives a basis to assess the role of local features, including substructures, in capturing online behaviours aligned to potential misinformation.

* Corresponding author.
E-mail addresses: AshfordJR@cardiff.ac.uk (J.R. Ashford), TurnerL9@cardiff.ac.uk (L.D. Turner), WhitakerRM@cardiff.ac.uk (R.M. Whitaker), PreeceAD@cardiff.ac.uk (A. Preece), Dhf12@psu.edu (D. Felmlee).

[1] https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference
[2] https://www.bbc.co.uk/news/technology-52281315
[3] https://www.telegraph.co.uk/global-health/science-and-disease/one-third-uk-may-not-get-coronavirus-vaccine-one-developed-new/

https://doi.org/10.1016/j.osnem.2021.100178
Received 14 November 2020; Received in revised form 8 October 2021; Accepted 8 October 2021
Available online 9 November 2021

This network-based, language agnostic approach presents important opportunities to support automation in the detection of misinformation in online communities. Our work extends methods that were successful in classifying controversy in Wikipedia articles [7,8] and further contributes to characterising potentially disruptive groups [9,10]. As far as we can establish, little research has addressed communities of misinformation on COVID-19, particularly with reference to Reddit, making the current work both timely and relevant.

2. Background

The relative ease with which misinformation can be produced and become disruptive has motivated recent investigation into this phenomenon. From a psychological viewpoint, it appears that there are individual differences in how misinformation becomes potentially endorsed by individuals [11,12]. This acceptance gives a basis for misinformation to become potentially promoted by like-minded others, leading to a compound effect where an informal group endorses particular content with a reinforcement across its participants. For example, one subreddit in particular, r/Wuhan_flu, a subreddit which actively promoted the use of free speech gained a lot of attention with its anti-censorship agenda. Reddit took action and placed the subreddit in "quarantine" suggesting that it "*may contain misinformation or hoax content*"[4] therefore requiring users to "opt-in" to the community to view its content. The effects of misinformation have taken place in various different contexts, with politics being particularly susceptible, as seen in the 2016 US presidential election [13,14]. In previous events, disinformation concerning the funding of the UK's National Health Service (NHS) was circulated and brought to the attention of various political leaders during the 2019 UK General Election, which originated from Russian actors on Reddit.[5] However, once misinformation is embedded, echo chambers are known to take hold and to support engagement of misinformation, using weak ties [15,16] and lack of effective moderation [17] alongside "soft facts". These occur as a result of users sharing potentially misleading content without knowing the entire facts of an event [18–20].

The impact of misinformation surrounding COVID-19 has been observed on multiple platforms including the micro-blogging site Twitter [21–26]. Evidence of social media analysis suggests that individuals fail to discern between truth and falsehood, prompting the argument that health information shared on social media should be regulated [25,27–32]. Beyond Twitter, Reddit has also been susceptible to promoting misinformation. Reddit is an interesting platform to consider, since it allows self-defined communities to establish themselves, giving a unique basis for the analysis of online interactions. As of July 2021, Reddit is the 19th most popular site on the Internet globally[6] with as many as 330 million registered users[7] using the platform to read and discover topic-based user-generated content. Users are encouraged to join "subreddits" which serve as individual communities dedicated to a topic or theme where users can share and comment on posts submitted to the community. This provides the freedom to connect and reinforce the views of others. In particular, the Reddit platform has played a role in the spreading of hoaxes originating from Wikipedia [33] as well as sharing misinformation across multiple platforms [34]. With respect to misinformation around COVID-19 on Reddit, research has addressed the difference in narrative and language within Reddit communities using natural language processing [35,36] as well as the location [37] to assess the geographical influence.

To summarise, social media platforms provides innovative mechanisms allowing users to promote and share news and stores of current events. As a consequence, they produce conditions in which misinformation can develop at scale such that large audiences are potentially subject to misleading information. This is especially important considering the ease in which information can propagate through user interactions such as sharing and cross-posting across different communities.

3. Methods

We examine the global and local features of network-based representations of user participation in subreddits. This is achieved through bipartite networks that link users with subreddits to which they contribute, which we term as *subreddit association networks*. In doing so, subreddit association networks capture two areas of interest: (1) a user's diversity in posting to different subreddits and (2) the overlap between users in posting to the same subreddit(s). An example of a hypothetical subreddit association network is presented in Fig. 1.

For a corpus of subreddits, we define a *subreddit association network* (also known as SAN) as a bipartite graph $G = (V_1 \cup V_2, E)$ where V_1 represents a set of Reddit users, and V_2 represents the set of subreddits to which the users in V_1 have collectively posted. V_1 and V_2 represent a bipartite set of nodes (i.e., $V_1 \cap V_2 = \emptyset$), and there exists an edge $(i, j) \in E$ if and only if user $v_i \in V_1$ has posted in subreddit $j \in V_2$. Our approach is similar to work performed by Cheng et al. [26] however, we do not limit our network analysis exclusively to centrality and degree-based metrics and extend this to include graphlet analysis.

3.1. Local network features

Our approach to classifying SANs is based on counting the frequency of graphlets that are induced within its structure. Defining *graphlets* to include non-trivial induced substructures with up to six nodes provides a reasonable trade-off between the combinatorial complexity of counting (see [38] concerning the graphlet isomorphism problem) and the presence of useful features for analysis. In the network science literature, the possible induced subgraphs of a fixed size are typically referred to as graphlets [39–41]. We extend the same terminology here to denote all connected bipartite graphs with 3 to 6 nodes, as presented in Fig. 2, resulting in 43 possible alternatives.

The frequency of graphlets present in a given subreddit association network G is denoted by vector V_G where:

$$V_G = (v_1, v_2, \ldots, v_{43}) \tag{1}$$

and where v_i represents the frequency of the ith possible graphlet from Fig. 2. To enable comparison of networks of different size, we normalise V_G according to:

$$V_G = \frac{1}{\sum_{j=1}^{43} v_j}(v_1, v_2, \ldots, v_{43}). \tag{2}$$

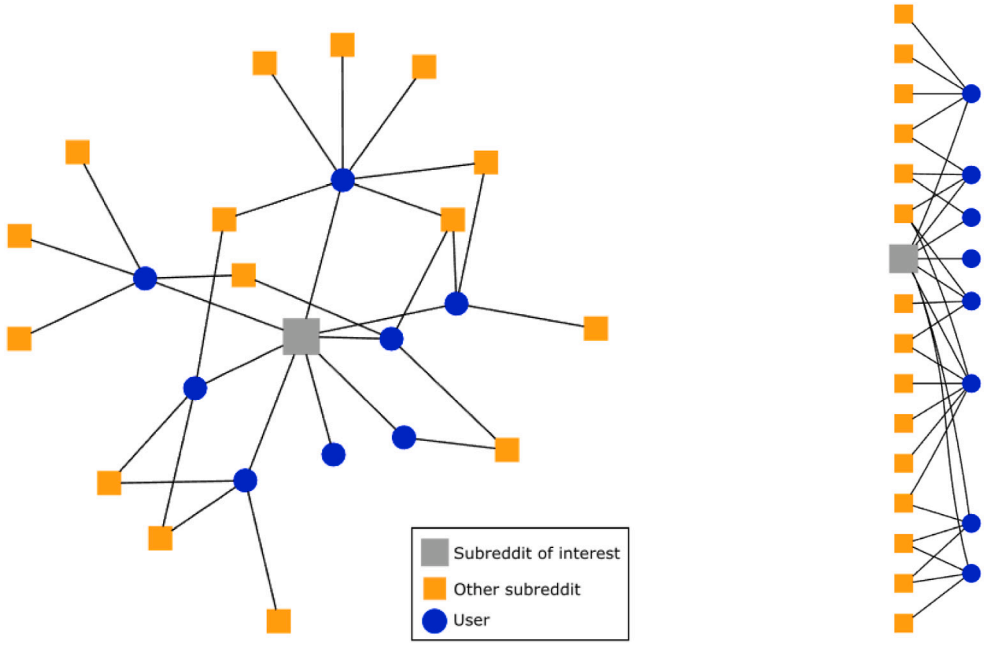
Vector V_G gives a basis to consider the relative under or over-representation of induced graphlets, in comparison to other subreddit association networks. This is similar to network-motif analysis approach for complex networks [42,43], and gives a basis to compare networks based on their latent structural characteristics. The relatively high dimensional space associated with $V(G)$ means that dimensionality reduction is a useful tool to provide further insights into the relationships between different association networks. Therefore we analyse V_G as derived from different subreddits, to establish the extent of similarity between different classes of association network.

[4] https://www.reddit.com/r/Wuhan_Flu

[5] https://www.reddit.com/r/redditsecurity/comments/e74nml/suspected_campaign_from_russia_on_reddit/

[6] https://www.alexa.com/siteinfo/reddit.com

[7] https://www.wired.com/story/reddit-redesign/



(a) (b)

Fig. 1. Example of a randomly-generated bipartite subreddit association network where the subreddit of interest is marked in grey and its surrounding users as blue circles. Other subreddits are represented as orange squares. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Global network features

Alongside the induced subgraphs represented through association networks, we also consider network-based metrics which provide an understanding of how users and subreddits behave collectively. These metrics include: subreddit and user degree, closeness centrality, clustering coefficient (i.e., local density between neighbouring connections), Latapy clustering (to determine heavily clustered interactions between user and subreddits) and Robins-Alexander clustering [44] to determine the clustered interactions through an aggregation of cycles and paths. These metrics provide further ways in which user association networks can be classified and assessed.

3.3. Data collection

Using the Reddit API, we have sampled data from a total of 257 subreddits, resulting in a corpus of data consistent with the scale of other misinformation studies on COVID-19 e.g. [35]. For each subreddit, we sorted posts by date (most recent first) and extracted a list of users who created the posts. Overall including all subreddits, our data spans between September 2017 and May 2020. To preserve anonymity, all usernames were hashed and no personal information was stored. This allowed us to build a subreddit association network. A full list of subreddits used in the paper can be found in Appendix A.

Subreddits were manually classified, aligned to their *potential for misinformation* (PFM) concerning COVID-19. There is no definitive way to achieve this, and in our case, subreddits were sourced from a team effort, and then cross checked to ensure they met the following criteria. Either:

1. The subreddit generally had very few moderators (users who are responsible for maintaining a subreddit community) and applied little or no moderation given the size and age of the subreddit. This is relevant because it allows more freedom for misinformation to go unchecked.
2. The subreddit description used terms such as "anti-censorship" or "freedom of speech" (FOS) in the subreddit description with little moderator involvement. This leaves greater opportunity for misinformation to be established.
3. The subreddit had been placed in "quarantine" by the Reddit administrators for containing potentially misleading or harmful content for the community. This is relevant due to the potential detection of misinformation.

Additionally, we include popular COVID-19 subreddits (e.g. r/Coronavirus, r/CoronavirusUK, etc.) as they are considered subreddits with the potential to contain misinformation. These are highly topic-relevant subreddits that might be attractive to agents who are keen to express misinformation. In total, application of these criteria resulted in 27 subreddits being selected as having potential for misinformation. These are referred to the *PFM subreddits*. Appendix A provides a list of the PFM subreddits used in this study. We further note that alternative criteria for selection of the PFM subreddits could equally be applied.

To provide a basis for comparison of PFM subreddits, we introduce three other sets of subreddits so that we can benchmark against alternative forms of user interaction with this social media platform. We do so with the objective of comparing against alternative subreddits to assess predictive utility. Furthermore, our aim is to discover how our approach can represent subreddits of different taxonomies (e.g. Q&A compared with discussion). Our benchmark subreddits are defined as follows.

- **PFM**: a total of 27 subreddits relating to COVID-19 which may contain misinformation.
- **Ask**: a sample of 30 Ask Q&A-based subreddits that involve interactions within in a highly moderated environment. This allows us to compare with PFM as posts made to Ask subreddits undergo strict moderation due to restricted posting rules meaning that they are unlikely to contain misinformation, contrary to PFM subreddits.
- **New**: a sample of 100 random subreddits created in 2020, covering the time period relevant to the creation of most PFM subreddits. This enables subreddit age to be controlled for in subsequent analysis.
- **Random**: a sample of 100 random subreddits without any constraints for subreddit age. This serves as a random baseline to include the diversity of Reddit content in general.

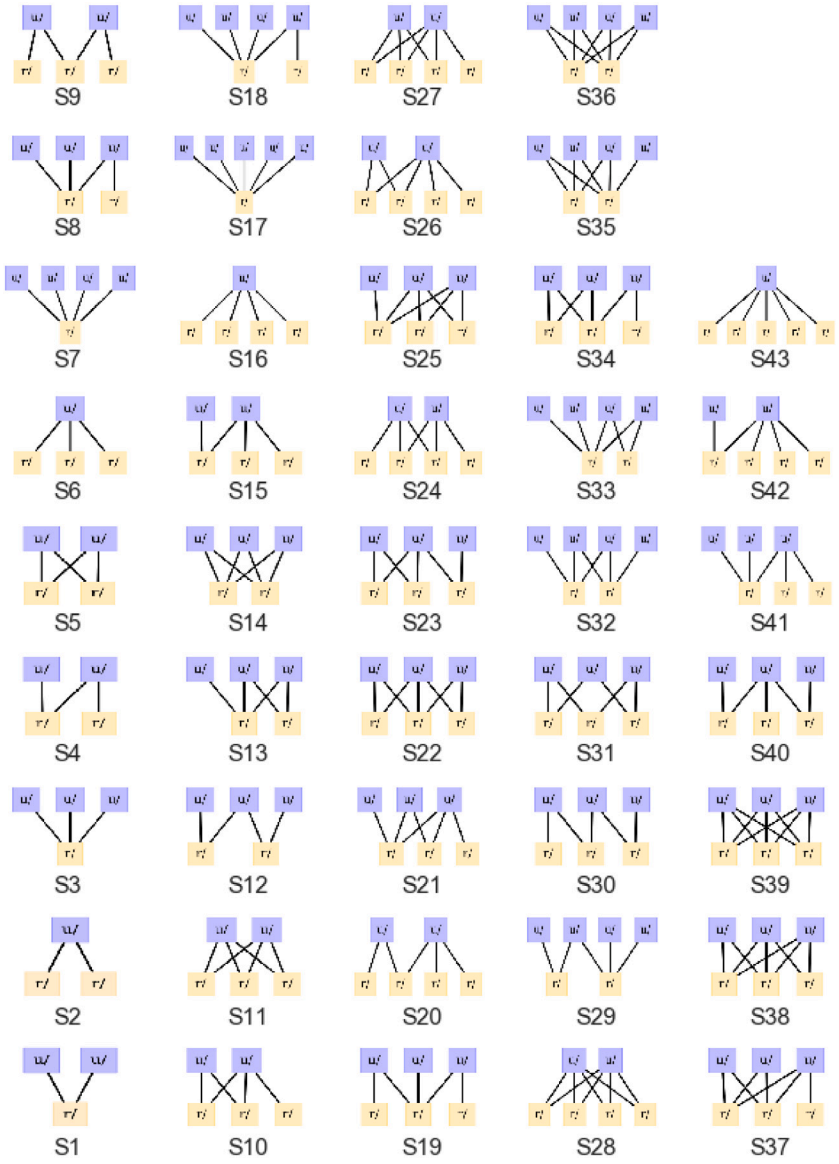


Fig. 2. Collection of all 43 induced bipartite graphlets featuring graphlet sizes from 3 to 6. User nodes are labelled as blue and subreddits as yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For each of the users active in any of the subreddits in any of the four datasets, we determined the list of all other subreddits that they also submitted to over the time period and aggregated these to generate the bipartite user association network across the subreddits in the above samples, as described in Section 3. Across all of the subreddits selected, a total of $7,876,064$ posts were processed across 96,634 users.

Using non-network metrics, we attempt to use metadata (such as age and subscribers) to demonstrate that these features are not necessarily the strongest indicators for clustering as observed in Fig. 3. By comparison, network features provide better spatial relevance. Additionally, the availability of such data is limited and may not always be consistent meaning that the exclusive use of metadata does not serve as a reliable proxy for classification. Furthermore, this justifies the need for considering network-based metrics.

The results in Fig. 4 reveal how PFM subreddits show lower maximum and average in subreddit degree (see Figs. 4(a) and 4(b)) meaning fewer users engage with these subreddits as compared with that of the Ask subreddit communities. Furthermore, we observe that PFM subreddits also have a higher but varied average user degree (Fig. 4(c)) and produce less clustering (Fig. 4(d)), contrary to Ask subreddits.

In Fig. 5 we present the subreddit and user degrees as a normalised ratio of the maximum degree featured in each network. For Ask subreddits, the maximum degree for user and subreddits nodes are fairly balanced with a partial swing towards having a slightly larger subreddit degree maximum (Fig. 5(a)). Furthermore, the PFM subreddits are heavily skewed towards a higher user degree and lower max subreddit degree (Fig. 5(b)).

4. Experimentation and results

Using the data collected in Section 3.3, we firstly examine the profile of graphlets induced by the PFM, Ask, new and random sets of subreddits (Section 4.1). We then apply principal component analysis (PCA) to examine the extent to which these different classes of subreddit can be distinguished under dimensionality reduction. This reveals the contribution made by particular graphlets in support of the resulting new dimensions, indicating the dominance of particular graphlets. For comparison purposes, in Section 4.2 we carry out the analogous analysis but with global network metrics. Finally, in Section 4.3, we explore the extent to which it is possible to predict the classification

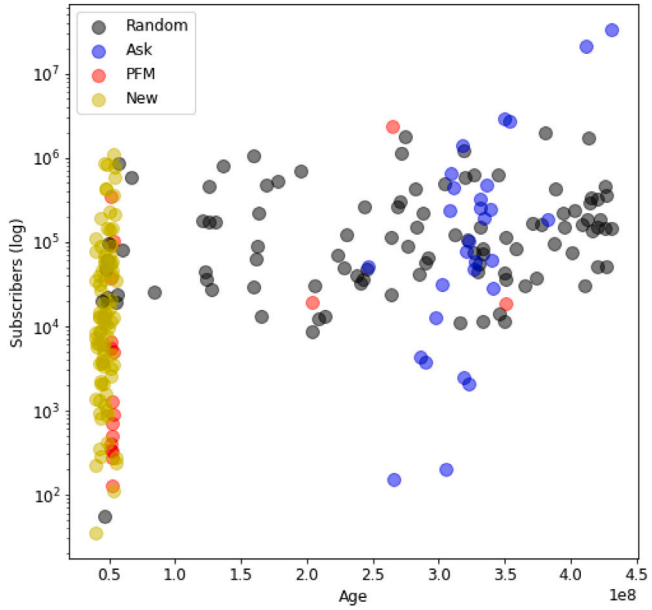


Fig. 3. Using age (x-axis) and subscriber counts (y-axis) shows little clustering potential compared with network-based features. New subreddits are marked in yellow, Ask in blue, PFM in red and Random in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of subbreddits based on the dominant features of the PCA dimensions identified in Sections 4.1 and 4.2 .

4.1. Association network profiling through graphlets

We enumerate over all induced bipartite graphlets (see Fig. 2) to produce normalised vectors of graphlet frequency (see Eq. (2)) for all networks. The results in Fig. 6 show some general similarities between different classes of subreddit. In particular, the S17 and S43 graphlets are dominant across all four sets of subreddit. We also observe such a high variation across all graphlets for new subreddits. Smaller graphlets such as S1 and S2 make more of an appearance in new subreddits. The graphlet profiles also reflect the sparse nature of the subreddit association networks, with high degree graphlets (both user and subreddit) being absent. However despite similarities, at a more granular level significant differences are evident between the graphlet profiles for the different classes. This is apparent when principal component analysis is applied, which reduces the 43-dimensional feature vector to two principal components as shown in Fig. 7.

The results in Fig. 7 show clear differentiation between the Ask and PFM subreddits. Although both PFM and new subreddits were created around a similar time, we observe that their positioning in the scatter plot remains distinct by comparison which reaffirms that factors in addition to age distinguish these subreddits. The Ask subreddits exhibit clustering while the PFM subreddits generally exhibit higher values (greater than zero) against the second principal component. More generally, these results indicate that graphlets can distinguish alternative classes of subreddits that appear similar in face value. Furthermore, we note that principal component analysis positions the official coronavirus subreddits (r/CoronavirusUK and r/Coronavirus) away from clusters of PFM subreddits. Subreddits with poor moderation such as r/CoronavirusUncensored, r/VirusOutbreak and r/CoronavirusFOS, are clearly distinguished from Ask subreddits.

Taking into account data point sizes, we also observe from Fig. 7 that subreddit subscriber count (Fig. 7(a)) and age (Fig. 7(b)) do not necessarily provide a strong indication of reliability or maturity compared with some subreddits which are much older and well established. The scatter plots reveal how younger subreddits share similar structures to that of well-established subreddits (such as the Ask communities). Furthermore, it is possible for older subreddits to align with the suspicious PFM subreddits cluster.

Fig. 9(a) presents the eigenvalues for each graphlet with respect to their influence within each principal component. The first principal component is primarily characterised by the strong presence of graphlets S17 and S43 which describe a one-to-many (and visa versa) relationship. In other words, high degree from users to subreddits and high degree from subreddits to users are influential. Graphlets S18, S20, S29 and S42 are also highlighted, which suggests a partial overlap and mutual ties could contribute to distinguishing these alternative sets of subreddits.

4.2. Global features of association networks

Using the bipartite metrics presented in Section 3.2 we follow a similar approach to Section 4.1. Here we characterise the user association networks using global metrics and apply PCA to create a reduced dimension space. The results provided in Fig. 8 demonstrate similar clustering behaviour to that of the local features however differentiation between the spacing and clustering of particular subreddit groups is less pronounced. For example, the selected principal components provide little improvement in distinguishing between Ask and PFM subreddits. Additionally, greater spread is seen in the resulting dimensions, which is driven mainly by the New subreddits. Extracting the PCA coefficients (see Fig. 9(b)) demonstrates that the first principal component is heavily influenced by maximum degree ratios for both subreddit and user nodes. The second component is mainly positively influenced by Latapy clustering and negatively influenced by maximum user degree ratio.

4.3. Predicting class of subreddit

We use the data processed from our analysis earlier as part of a prediction task to classify PFM, Ask and New subreddits respectively. Consistent with other approaches (e.g., [9]) we apply binary logistic regression (BLR), support vector machine (SVM) and a random forest classifier (RFC) applied with 10-fold cross-validation.

4.3.1. Local features

We train our prediction models using normalised feature vectors of graphlet counts (as seen in Fig. 6). Due to the imbalance between each of the subreddit groups, we perform random under sampling over $N = 100$ trials and report the distribution, mean and standard deviation of various prediction metrics [45]. Our prediction metrics are reported in the form of violin plots (See Figs. 10–12) to capture the distribution of accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). We compare each subreddit group (PFM, Ask and New) to a random baseline for comparison purposes. We also perform a pairwise comparison between each set of subreddits in isolation, such that prediction is performed with PFM vs Ask, PFM vs New and Ask vs New. We do this to assess how robust prediction is between separate groups, in comparison to others. Finally, in addition to this, we perform a pairwise comparison between each set such that the random baseline is removed completely. We do so to understand how well classification of a particular set performs in isolation.

The classification results provided in Figs. 10–12 demonstrate that a RFC consistently outperforms BLR and SVM by comparison. This is reflected in the accuracy metrics as the distribution of values for RFC are much higher than those of BLR and SVM with an average accuracy of $P = 0.74$ for PFM, $P = 0.77$ for Ask and $P = 0.96$ for new using local features. Using global features, a RFC yields average accuracy values of $P = 0.97$ for PFM, 0.89 for Ask and 0.95 for new subreddits.

The results demonstrate a clear performance gain between classifying Ask vs New and PFM vs Ask subreddits as both the Positive

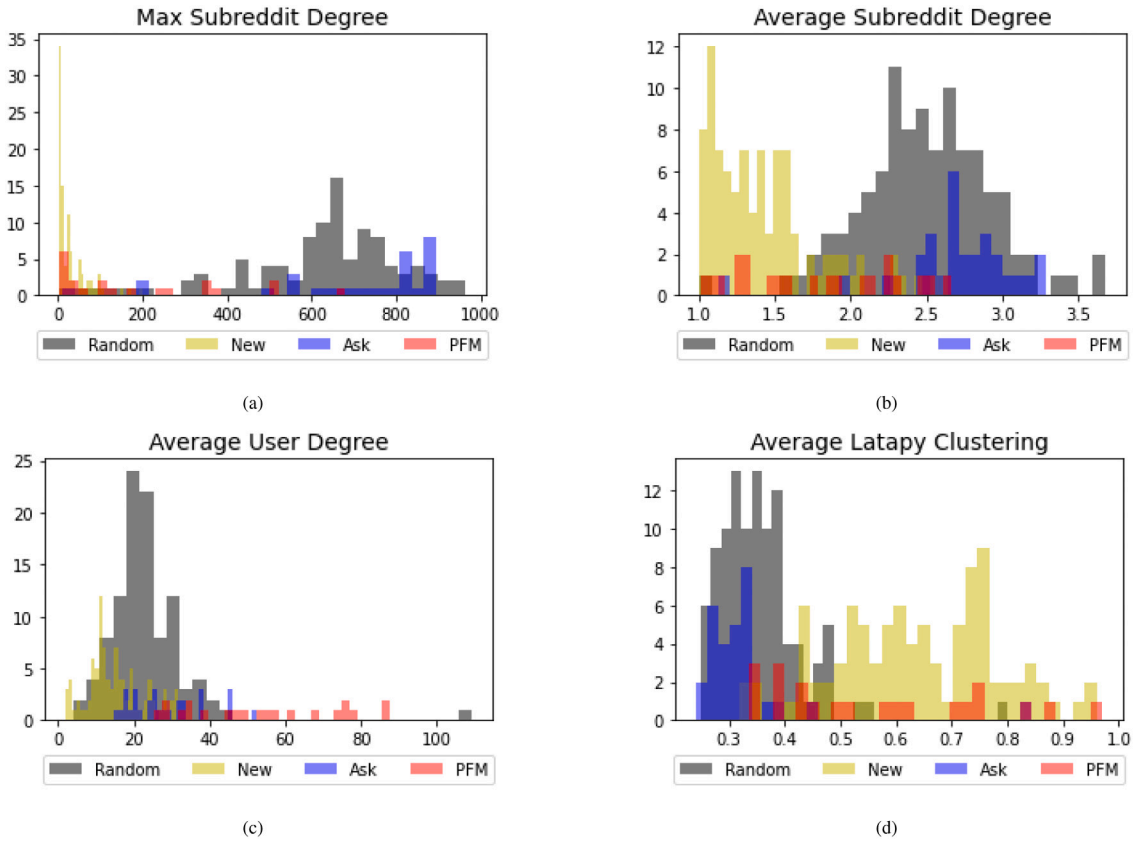


Fig. 4. Subreddit degree distributions, user degree distributions and average latapy clustering distributions for the Ask, PFM, New and Random association networks.

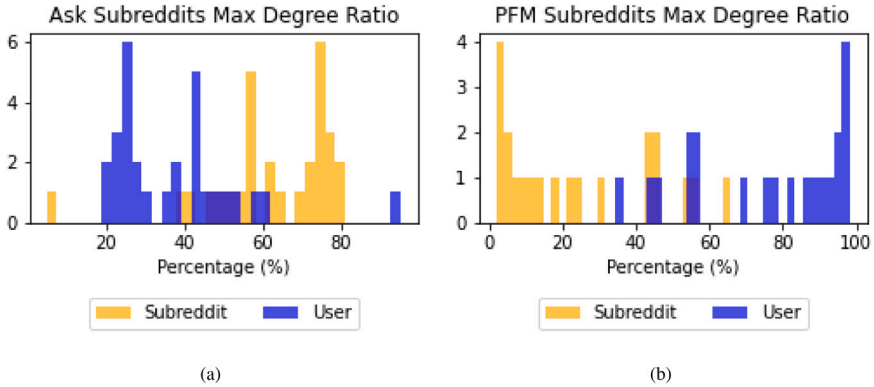


Fig. 5. The maximum degree ratios for the Ask and PFM association networks.

Predictive Rate, PPR (See Figs. 13(d) and 13(e)) and Negative Predictive Rate, NPR (See Figs. 15(d) and 15(e)) are relatively stable with a mean PPV of $P = 0.96$ for Ask vs New and $P = 0.89$ for PFM vs Ask and a mean NPV of $P = 0.9$ for Ask-new and $P = 0.87$ for PFM vs Ask using local features trained on a RFC. This demonstrates the effectiveness of local prediction with the ability to both separate and classify subreddits into groups with relative ease.

4.3.2. Global features

Using the same classifiers as in Section 4.3.1, we train our models using the global network features (see Section 3.2) with a view to understanding their predictive performance. As in Section 3.2, we perform classification in the scenario where each set is compared to a random baseline followed by classification when pairs of subreddit sets are involved.

By comparison to the pairwise prediction in Section 4.3.1, the violin plots indicate much greater dispersion of results in some cases: see Figs. 13–15. Here there is much less consistency in results — for example, Fig. 14(e) reveals only one set of NPV's using a RFC. This reaffirms the idea that RFC performs consistently well across all sets. The results indicate that global features have much less stability as a basis for prediction in comparison to local features for categorisation involving two alternative sets of subreddits. This is especially true for predicting PFM with New subreddits as shown in Fig. 14.

4.3.3. Comparing local and global performance

The classification results from Section 4.3 provide useful insights towards the effectiveness of using machine learning to predict characteristics of subreddits — such as potential behaviours correlating with potential misinformation activity. We compare the classification results by taking the average accuracy of each classifier and task

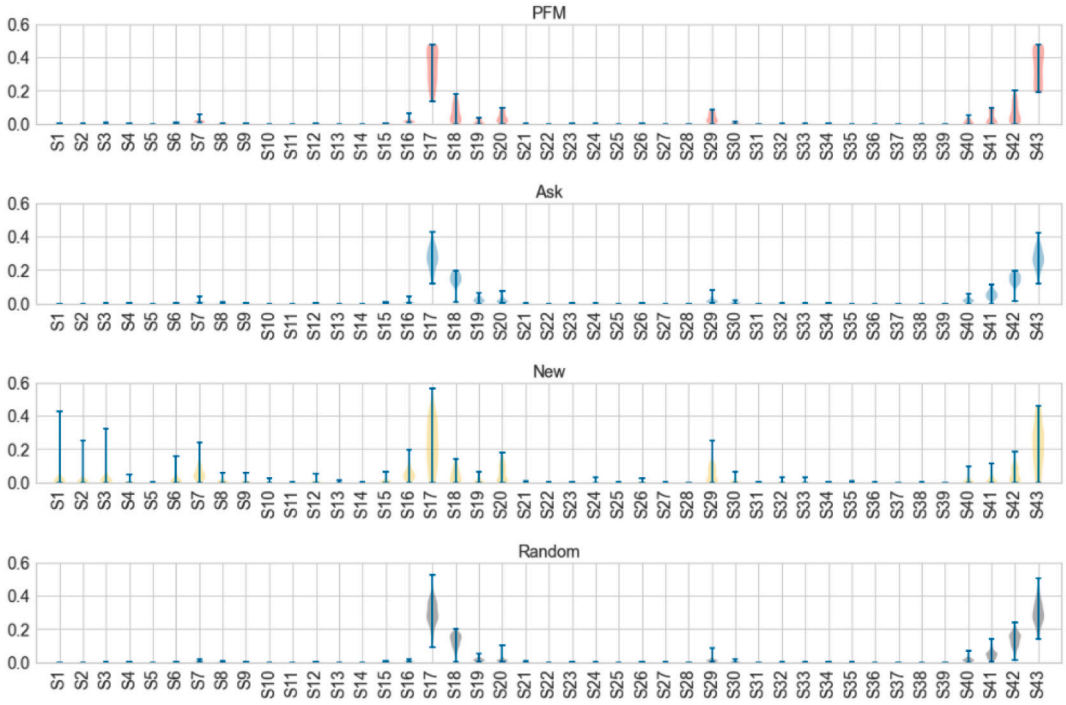


Fig. 6. Normalised frequency as a violin plot of all 43 induced bipartite graphlets with PFM subreddits (first), Ask subreddits (second), New subreddits (third) and Random subreddits (fourth).

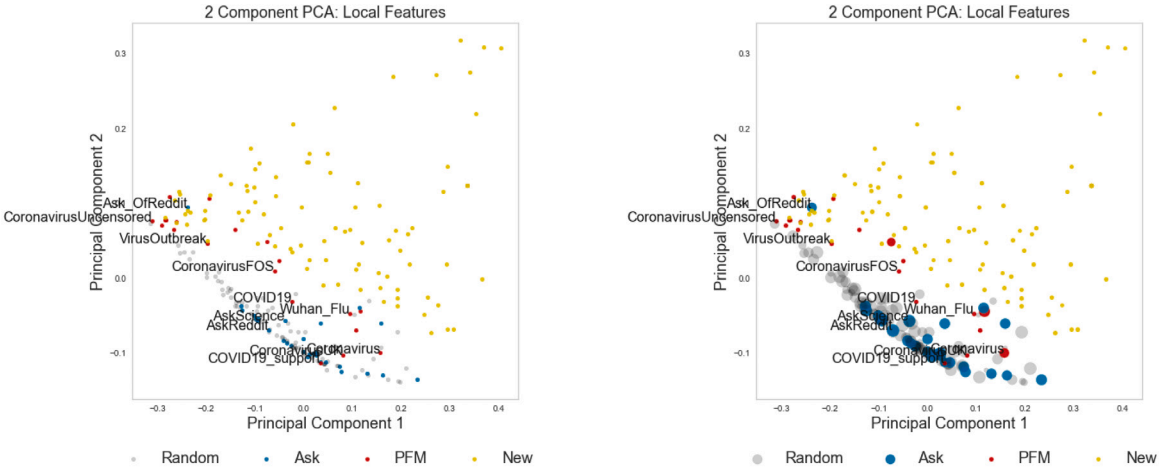


Fig. 7. Scatter plot of two-dimensional PCA of graphlets counts producing distinct clusters with a few significant subreddits labelled. Ask subreddits are marked in blue, new subreddits in yellow, PFM subreddits in red and random subreddits in black. Nodes are sized according to subscriber count 7(a) (largest as most subscribed) and age 7(b) (largest as oldest). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cross-comparing prediction using local features over global features to help interpret comparing values. We chose accuracy specifically as it provides a reasonable metric for analysing prediction performance at a high-level.

The results presented provide evidence that local bipartite network features and induced graphlets play a significant role in understanding users' posting activity and similarity to others through subreddit association networks. We observe that some classification tasks involving PFM subreddits perform better using local features whereas Ask subreddits perform better with global features. This may suggest that PFM subreddits are dependent on more-detailed graphlet formations whereas Ask subreddits rely on simpler metrics, such as degree, using global features.

5. Discussion

The classification results from Section 4.3 provide useful insights on the effectiveness of alternative approaches to accurately predicting the categorisation of different classes of subreddit. A number of key issues are evident and important to highlight. These relate to graphlet frequencies, predictability, the influence of local features, the potential for other classes of subreddits to be recognised and the potential for applicability beyond Reddit.

Firstly, it is important to note that the profile of graphlet frequencies differs between the PFM subreddits and other subreddits having a similar age. This supports the role of graphlets as a useful tool to detect distinguishing structural differences in the underlying subreddit

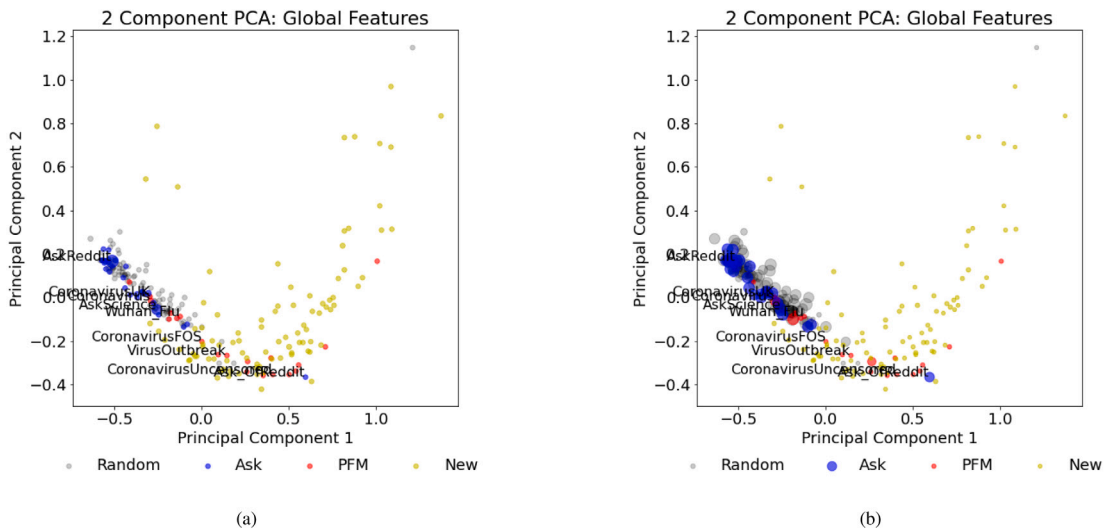


Fig. 8. Scatter plot of two-dimensional PCA using only graph-based metrics as used in Section 3.2. Ask-subreddits are marked in blue, new subreddits in yellow, PFM subreddits in red and random subreddits in black. Nodes are sized according to subscriber count 8(a) (largest as most subscribed) and age 8(b) (largest as oldest). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

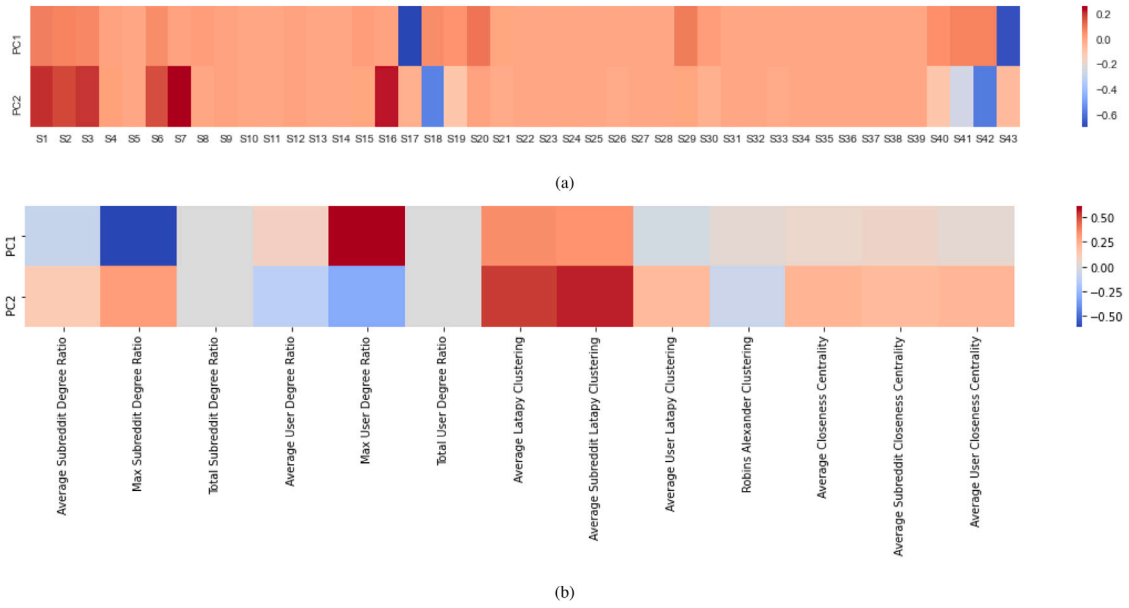


Fig. 9. Eigenvector values from two-dimensional principal component analysis when using alternative variables, based on graphlet count 9(a) and global network features 9(b). These values indicate the positive or negative contribution made by each variable when projecting subreddits into two dimensional space, as presented in Figs. 7 and 8.

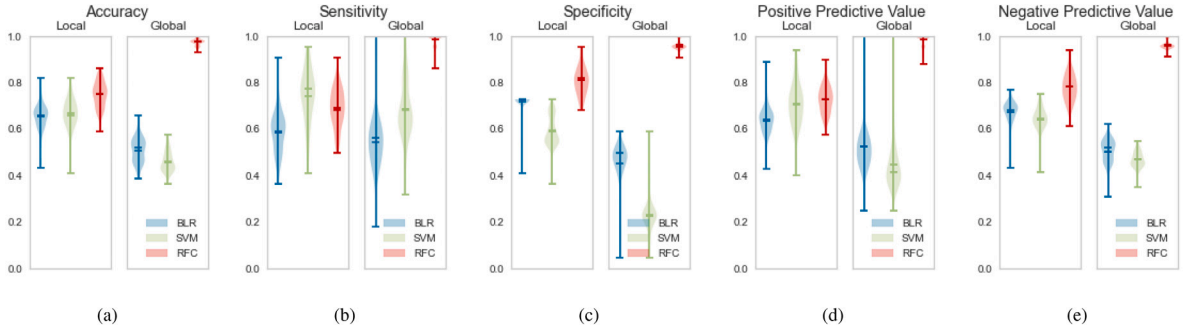


Fig. 10. Classification performance for PFM subreddits comparing local and global features reveals a consistent performance for RFC.

association networks. This also reaffirms the overarching utility in the representation — that a network-based language-agnostic approach has potential for classification purposes. As shown in Figs. 7 and 8, PFM and Ask subreddits remain distinct among new subreddits despite being

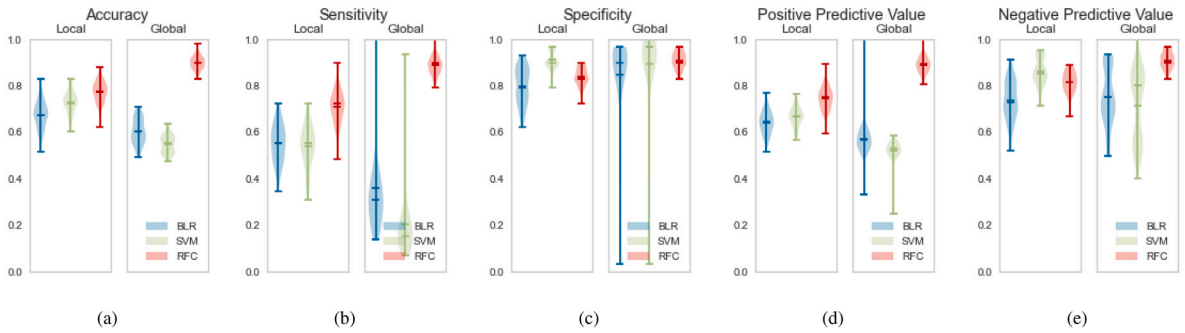


Fig. 11. Classification performance for Ask subreddits comparing local and global features are a little more varied by comparison to PFM subreddits in Fig. 10.

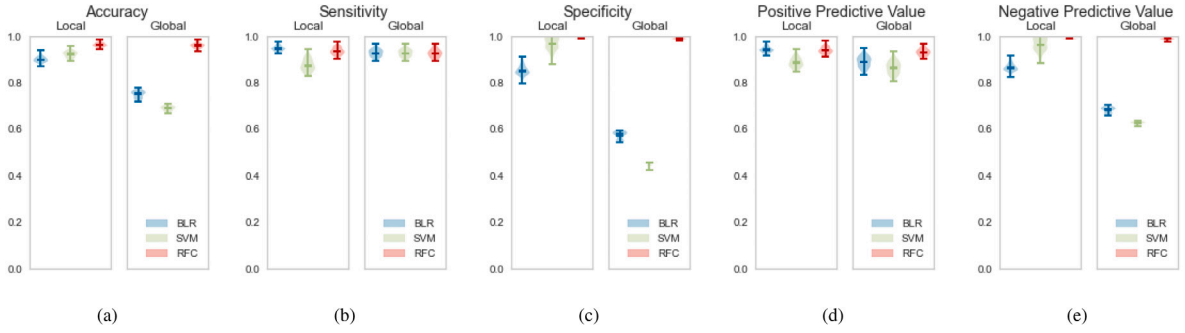


Fig. 12. Classification performance for New subreddits comparing local and global features.

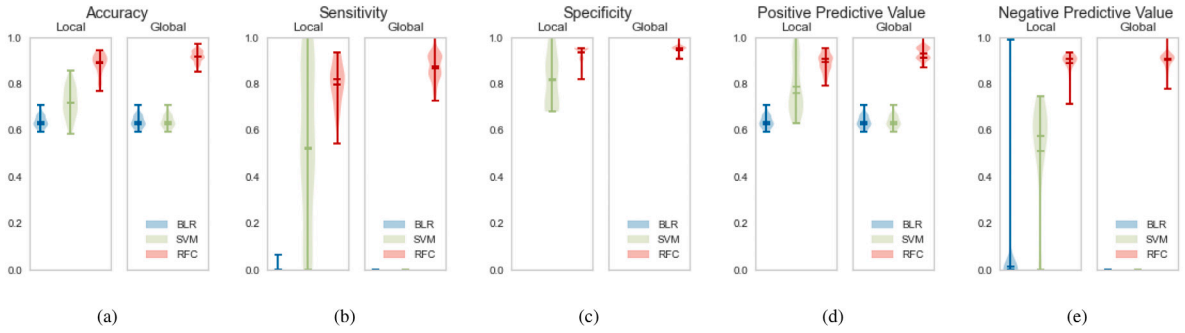


Fig. 13. Classification performance comparing PFM with Ask subreddits comparing local and global features.

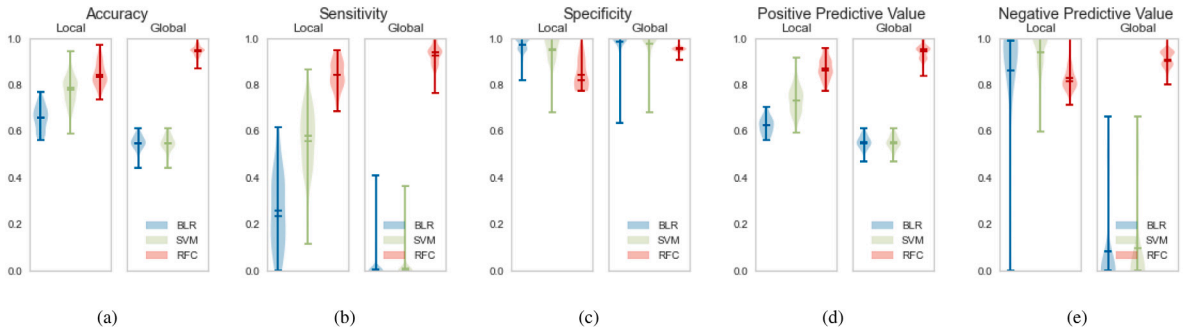


Fig. 14. Classification performance comparing PFM with New subreddits comparing local and global features.

of a similar age. When considering age and subscriber counts (see Fig. 3), the New and PFM classes of subreddit have substantial overlap. However, extracting features from the user association networks enables these classes to be distinguished. Note that this separation is arguably more pronounced when graphlet features are the basis for dimensionality reduction, as compared to standard graph-based metrics (i.e., Fig. 7 as compared to Fig. 8).

Secondly, the overall predictability of classification of PFM (and other) classes of subreddit are generally high while using alternative features and methods, based on both graphlets and global characteristics. This again supports the utility from the underlying representation of user association networks. It also further supports the creation of future monitoring agents for social media, with or without a human in the loop. Our results reveal that a Random Forest Classifier overall

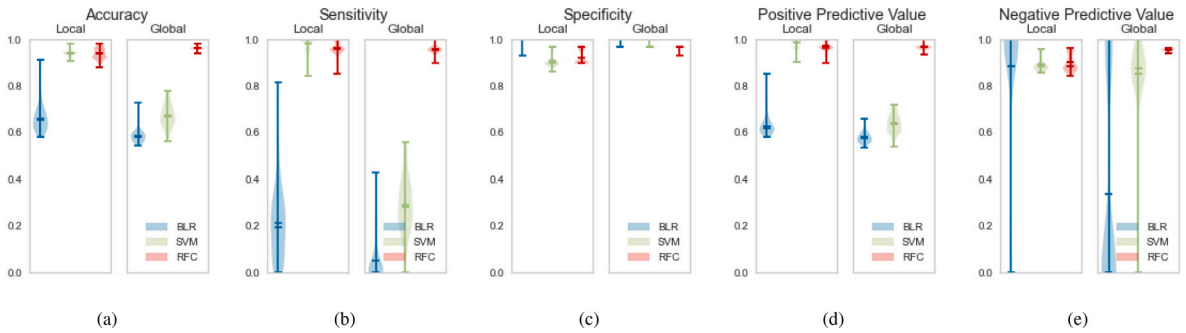


Fig. 15. Classification performance comparing Ask with New subreddits comparing local and global features.

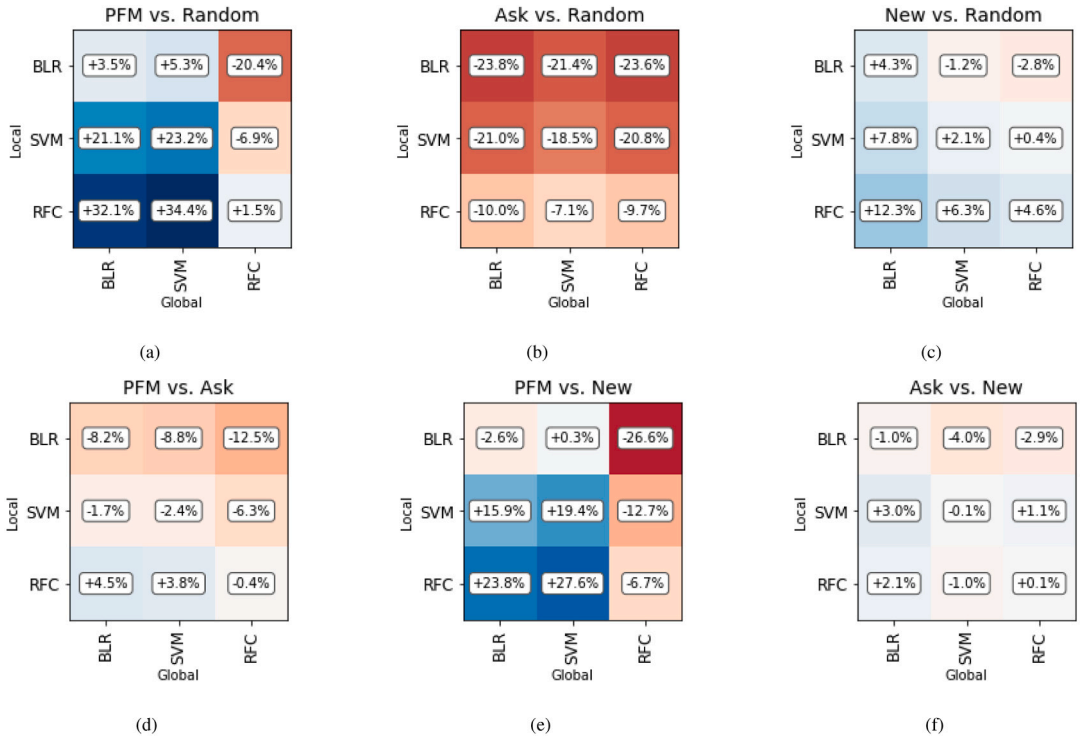


Fig. 16. Pairwise comparison assessing the prediction gain (as a percentage) of local features over global features broken down into six prediction tasks. We derive the percentage differences Δ_{ij} by obtaining the difference between prediction values p_i, p_j scaled by the original value of the prediction task of interest p_i such that $\Delta_{ij} = \frac{(p_j - p_i)}{p_i} * 100$.

provides the best and most consistent classification performance out of three alternative methods used. For example, classification of PFM and Ask subreddits can be achieved with relative ease as mentioned in Section 4.3.1 using this method.

Thirdly, both influential graphlet features and the influential global features for prediction of PFM subreddits appear to be related. The most influential global features involve node degree, while tree-like graphlets are the salient local features for prediction (see Figs. 9(a) and 9(b)). This means that local approaches (such as graphlet analysis) are well-suited to PFM classification and using alternative features (local and global) has helped to reaffirm this observation. The importance of local features opens up prospects for efficient real-time detection based on decentralisation and observation of graphlets in observable subnetworks. From extracting the eigenvector coefficients using PCA, we observe that aligned with global measures of degree, tree-like graphlets structures such as S17 and S43 are particularly important. Furthermore, our prediction results in Section 4.3.3 indicate that local features provide generally strong predictive utility for PFM (see Fig. 16(a)) and new (see Fig. 16(c)) subreddits.

Fourthly, from focussing on PFM subreddits and comparing these to other classes, we hypothesise that different classes of subreddit may leave particular underlying signatures in their corresponding user association networks, which reflect the different forms of user behaviour in which communities participate. For example, PMF and Ask fundamentally differ in the way in which users interact as one is used to distribute new articles whereas the other is primarily representative of stricter Q&A-like discussions. Consequently we speculate that there may be a wider underlying taxonomy of significant graphlets for different classes of interaction in subreddits.

Finally, we note that user association networks are a generalised approach to representing user behaviour in respect of social media content. This requires only an associative link between a user and another entity representing some form of content. Here we have restricted our attention to Reddit (i.e., interactions with subreddits), but we believe that the general approach should yield insights into other forms of social media through modelling in a similar way. For example, these could include creating association networks based on user subscriptions, users tweeting certain hashtags on Twitter or users commenting across different articles. There are wide ranging ways in which associations can be made and examined.

6. Conclusions and future work

In this paper, we have used a general network representation of user association with social media content as a basis for prediction of important sub-classes of content that align with the potential for misinformation (PFM). This has been applied to the Reddit social media platform, using a number of alternative groups of subreddits for benchmarking purposes. The utility in this representation stems from being able to potentially categorise subreddits as having the potential for misinformation without undertaking any semantic analysis of content. This increases opportunities to detect classes of subreddits with agility, for example removing the need for translation in assessing foreign language social media. The analysis carried out in this paper has identified that PFM subreddits are distinguished by the characteristics of their underlying user association networks — in other words the user-interaction with particular types of content has patterns associated with it that leave distinct signatures. These relate to the presence of high degree nodes which induce local tree-like structures that are seen in the form of particular graphlets that are strongly represented as compared to other induced substructures. The predictive capabilities of the graphlet census, alongside the global metrics, has been assessed, while employing PCA decomposition to identify the key features.

The methods included in this paper place an emphasis on the utility of network analysis where induced graphlets provide a fundamental topological representation. Furthermore, the use of the graphlet-based census serves as an ideal potential embedding technique for networks similar to that of tested techniques such as `gl2vec` [46]. By using PCA decomposition of graphlet-count feature vectors, we discovered latent differences across networks which open up insights that are not apparent when considering just the graphlet census in isolation. Through the use of dimensionality reduction we also observed how PFM communities and "anti-censorship" self-align and produce a distinct cluster in high-dimensional space aligned to the representation of induced subgraphs. In addition, we observed that moderation within the Ask subreddits appears to contribute to their distant positioning away from the PFM subreddits in the high dimensional space defined by induced graphlets as opposed to global features.

Numerous global features, such as degree-based metrics and clustering, are inherited as a consequence of simple local substructures, which has motivated the use of graphlets for analysis of complex networks across the wider literature. From our study there is evidence that graphlets have been effective because they are easily able to characterise the salient underlying features of the network related to node degree. More generally, graphlets also lend themselves to application in partially obfuscated network scenarios, giving potential for their flexible deployment in wide ranging scenarios.

We also established that classifying PFM subreddits can be achieved without the need for metadata, such as age and subscriber counts. Interestingly, it appears that subreddits with a low subscriber count or young age (i.e., immaturity) are not necessarily strong indicators of the potential for misinformation — subreddits with similar age or subscriber counts may have different network properties. The underlying behaviour of the users aligned to different classes of subreddit is the significant differentiating factor as this impacts the on the structure of interactions.

Finally, we believe that the exemplar presented in this paper provides a useful proof of concept that could be extended to address other misinformation scenarios where the intent is to undermine public perceptions and rational behaviour. We see the approaches considered in this paper being relevant to future applications where subreddit classification can be performed at scale for situations such as automated moderation for growing communities, without recourse to semantic analysis.

CRediT authorship contribution statement

James R. Ashford: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualisation. **Liam D. Turner:** Conceptualization, Methodology, Validation, Data curation, Writing – original draft, Writing – review & editing, Supervision. **Roger M. Whitaker:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision. **Alun Preece:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision. **Diane Felmlee:** Validation, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorised to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Appendix A. Subreddits used

- **PFM**: COVID19, ncovshills, cvnews, epidemic, Real_Coronavirus, CoronaVirusFFA, Wuhan_Flu, ID_News, VirusOutbreak, CoronavirusUncensored, corovirusdata, novel_coronavirus, Coronavirus UK, CoronavirusGLOBAL, Covid2019, COVID19_support, 2020 WuhanVirus, China_Flu, nCoronaVirus, 2019COVID, Coronavirus, CoronavirusFOS
- **Ask**: AskWomen, AskCulinary, AskDrugs, AskFeminists, AskLiteraryStudies, TrueAskReddit, AskMusic, AskModerators, AskStatistics, AskEngineers, AskDad, AskComputerScience, AskHistorians, AskLosAngeles, AskSeddit, AskAcademia, AskPhotography, AskScience, AskReddit, AskSciTech, AskUk, AskMen, AskTransgender, AskGSM, AskElectronics, AskArt, Ask_OfReddit, AskPhilosophy, AskSocialScience

Table B.1
Global network statistics of the Reddit data.

	Mean	SD	Min	Max
Average Subreddit Degree	2.42	0.508	1.012	3.6
Max Subreddit Degree	575.68	252.175	3.0	964.0
Av. User Degree	28.89	15.943	6.843	87.34
Max User Degree	434.03	222.159	84.0	923.0
Av. Subreddit Degree Ratio	0.09	0.0379	0.014	0.212
Max Subreddit Degree Ratio	0.54	0.206	0.018	0.86
Av. User Degree Ratio	0.9	0.0379	0.787	0.98
Max User Degree Ratio	0.45	0.206	0.139	0.981
Av. Latapy Clustering	0.38	0.131	0.239	0.970
Av. Subreddit Latapy Clustering	0.40	0.13	0.257	0.987
Av. User Latapy Clustering	0.08	0.039	0.037	0.256
Robins Alexander Clustering	0.05	0.021	0.0	0.147
Av. Closeness Centrality	0.54	0.039	0.513	0.955
Av. Subreddit Closeness Centrality	0.52	0.04	0.504	0.953
Av. User Closeness Centrality	0.69	0.037	0.636	1.073

- **New**: OnlyFun4U, BBC4BBWS, breadboysyt, bfatIRL, SpecialHumor, sims2help, Adultcontentcreators, NativePlantGardening, YourWellnessNerd, TradeAnalyzerFF, JuliaBayonetta, Sheismich aelaNSFW, Jord627_, AllSaintsStreet, moreplatesmoredates, HUEstation, KiryuCoco, WallStreetbetsELITE, Cartooncat, USAHotGirls, VALORANT, delta8, ImaginaryAnthro, TopPops, skamtebord, InfluencergossipDK, IndianStreetBets, onlyfansbros, CatfishMe Please, TheWildAtHeart, onlyfanschicks, RedditMasterClasses, Dodocodes, oldhagfashion, SRGroup, MeatoSubincision, SatoshiBets, Promote_Your_YouTube, IPTVresell, onlyfansgirls101, WKHS, naughtychicks, Wallstreetbetsnew, Mya_For_The_Queen_, HeroWarsFB, Spudmode, quarantineactivities, ACVillager, assettopirate, TifaxAerith, LegendofthePhoenix, TheYouShow, PokeMeow, MLFBprospringfootball, BigBoobsAndAssess, Equity ResearchIndia, OnlyfansXXX, DankExchange, AMDLaptops, USTravelBan, xxxycelebs, VictoriasecretGW, AmateurGoneWildPlus, CruelSummer, TgirlHUB, yeagerbomb, onlyfans_get_noticed, Naveljunkies, OnlyFans_Amateurs, ExtremelyHairyWomen, Desihub, mummytummies, Cross_Trading_Roblox, OnlyFansAsstastic, RedditPregunta, Epicentr, Teenpussyx, TLAUNCHER, CaliforniaJobsForAll, Helltaker, buksebule, confidentlyincorrect, AlabamaJobs, Life360, MedicineCommunity, CPTSDFightMode, PPPLoans,

Table C.2
Graphlet prediction results for PFM subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.677347	0.792449	0.864286
SD Accuracy	0.066987	0.030552	0.041260
Mean F1 Score	0.460567	0.745211	0.853024
SD F1 Score	0.176783	0.031276	0.043656
Mean Precision	0.831587	0.836975	0.833129
SD Precision	0.185315	0.065113	0.052828
Mean Recall	0.330909	0.673636	0.875455
SD Recall	0.147160	0.024377	0.047508
Mean Sensitivity	0.959630	0.889259	0.855185
SD Sensitivity	0.029674	0.051452	0.050538
Mean Specificity	0.330909	0.673636	0.875455
SD Specificity	0.147160	0.024377	0.047508
Mean Positive Predictive Value (PPV)	NaN	0.836975	0.833129
SD Positive Predictive Value (PPV)	NaN	0.065113	0.052828
Mean Negative Predictive Value (NPV)	0.641535	0.769555	0.894412
SD Negative Predictive Value (NPV)	0.050188	0.016824	0.038926

Table C.3
Graphlet prediction results for Ask subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.573103	0.594483	0.677069
SD Accuracy	0.060866	0.053835	0.055326
Mean F1 Score	0.577603	0.662721	0.687415
SD F1 Score	0.059650	0.049826	0.057733
Mean Precision	0.572328	0.568504	0.665524
SD Precision	0.058992	0.041630	0.052569
Mean Recall	0.584483	0.802759	0.713103
SD Recall	0.067109	0.102534	0.074578
Mean Sensitivity	0.561724	0.386207	0.641034
SD Sensitivity	0.075279	0.112162	0.063788
Mean Specificity	0.584483	0.802759	0.713103
SD Specificity	0.067109	0.102534	0.074578
Mean Positive Predictive Value (PPV)	0.572328	0.568504	0.665524
SD Positive Predictive Value (PPV)	0.058992	0.041630	0.052569
Mean Negative Predictive Value (NPV)	0.574485	0.678635	0.693155
SD Negative Predictive Value (NPV)	0.064786	0.118185	0.063627

Table C.4
Graphlet prediction results for New subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.893050	0.923150	0.961100
SD Accuracy	0.010557	0.012097	0.007334
Mean F1 Score	0.885511	0.926093	0.962094
SD F1 Score	0.011644	0.012178	0.006946
Mean Precision	0.952498	0.891871	0.939142
SD Precision	0.013654	0.019278	0.012450
Mean Recall	0.827500	0.964000	0.986300
SD Recall	0.015516	0.027532	0.005226
Mean Sensitivity	0.958600	0.882300	0.935900
SD Sensitivity	0.012330	0.024934	0.013935
Mean Specificity	0.827500	0.964000	0.986300
SD Specificity	0.015516	0.027532	0.005226
Mean Positive Predictive Value (PPV)	0.952498	0.891871	0.939142
SD Positive Predictive Value (PPV)	0.013654	0.019278	0.012450
Mean Negative Predictive Value (NPV)	0.847619	0.962011	0.985603
SD Negative Predictive Value (NPV)	0.011891	0.026861	0.005404

Table C.5
Graphlet prediction results for PFM vs Ask subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	7.755102e−01	8.302041e−01	0.882245
SD Accuracy	3.330669e−16	1.629589e−02	0.020379
Mean F1 Score	6.841905e−01	7.832142e−01	0.868471
SD F1 Score	5.167490e−03	1.599712e−02	0.023026
Mean Precision	9.292308e−01	9.212255e−01	0.870916
SD Precision	2.086871e−02	4.315329e−02	0.023729
Mean Recall	5.418182e−01	6.818182e−01	0.866364
SD Recall	1.233151e−02	2.220446e−16	0.027887
Mean Sensitivity	9.659259e−01	9.511111e−01	0.895185
SD Sensitivity	1.004790e−02	2.957402e−02	0.020323
Mean Specificity	5.418182e−01	6.818182e−01	0.866364
SD Specificity	1.233151e−02	2.220446e−16	0.027887
Mean Positive Predictive Value (PPV)	9.292308e−01	9.212255e−01	0.870916
SD Positive Predictive Value (PPV)	2.086871e−02	4.315329e−02	0.023729
Mean Negative Predictive Value (NPV)	7.212865e−01	7.856704e−01	0.891793
SD Negative Predictive Value (NPV)	3.173020e−03	5.375983e−03	0.021321

Table C.6
Graphlet prediction results for PFM vs New subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.666939	0.793469	0.847551
SD Accuracy	0.049813	0.049427	0.026524
Mean F1 Score	0.540839	0.790091	0.822059
SD F1 Score	0.086942	0.051439	0.029424
Mean Precision	0.703240	0.731279	0.865912
SD Precision	0.080590	0.063025	0.043399
Mean Recall	0.444091	0.870455	0.783636
SD Recall	0.093122	0.098202	0.032853
Mean Sensitivity	0.848519	0.730741	0.899630
SD Sensitivity	0.044474	0.091453	0.036400
Mean Specificity	0.444091	0.870455	0.783636
SD Specificity	0.093122	0.098202	0.032853
Mean Positive Predictive Value (PPV)	0.703240	0.731279	0.865912
SD Positive Predictive Value (PPV)	0.080590	0.063025	0.043399
Mean Negative Predictive Value (NPV)	0.653767	0.883523	0.836411
SD Negative Predictive Value (NPV)	0.041225	0.076711	0.023180

Table C.7
Graphlet prediction results for Ask vs New subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.910345	0.947414	0.938621
SD Accuracy	0.033343	0.008577	0.024422
Mean F1 Score	0.916716	0.944772	0.937064
SD F1 Score	0.028057	0.009047	0.024812
Mean Precision	0.867796	0.994276	0.961653
SD Precision	0.054369	0.014777	0.029025
Mean Recall	0.974138	0.900345	0.914138
SD Recall	0.015708	0.018178	0.028641
Mean Sensitivity	0.846552	0.994483	0.963103
SD Sensitivity	0.073372	0.014400	0.028542
Mean Specificity	0.974138	0.900345	0.914138
SD Specificity	0.015708	0.018178	0.028641
Mean Positive Predictive Value (PPV)	0.867796	0.994276	0.961653
SD Positive Predictive Value (PPV)	0.054369	0.014777	0.029025
Mean Negative Predictive Value (NPV)	0.971285	0.909261	0.918486
SD Negative Predictive Value (NPV)	0.017669	0.015529	0.026614

Table C.8
Global feature prediction results for PFM subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.654227	0.643126	0.851116
SD Accuracy	0.047716	0.054630	0.041896
Mean F1 Score	0.660762	0.647285	0.854024
SD F1 Score	0.036657	0.042211	0.041231
Mean Precision	0.651952	0.646075	0.838903
SD Precision	0.054786	0.066341	0.050351
Mean Recall	0.671818	0.653636	0.871818
SD Recall	0.029162	0.050526	0.049950
Mean Sensitivity	0.636700	0.632787	0.830395
SD Sensitivity	0.083369	0.107988	0.060411
Mean Specificity	0.671818	0.653636	0.871818
SD Specificity	0.029162	0.050526	0.049950
Mean Positive Predictive Value (PPV)	0.651952	0.646075	0.838903
SD Positive Predictive Value (PPV)	0.054786	0.066341	0.050351
Mean Negative Predictive Value (NPV)	0.658219	0.644607	0.868011
SD Negative Predictive Value (NPV)	0.042705	0.049602	0.044486

Table C.9
Global feature prediction results for Ask subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.752320	0.729994	0.750444
SD Accuracy	0.039178	0.044819	0.050748
Mean F1 Score	0.781529	0.770427	0.764137
SD F1 Score	0.030536	0.030722	0.041858
Mean Precision	0.700392	0.673178	0.727952
SD Precision	0.040198	0.046165	0.057747
Mean Recall	0.885172	0.903103	0.806207
SD Recall	0.024892	0.019320	0.035428
Mean Sensitivity	0.620023	0.557724	0.694931
SD Sensitivity	0.066762	0.089088	0.083976
Mean Specificity	0.885172	0.903103	0.806207
SD Specificity	0.024892	0.019320	0.035428
Mean Positive Predictive Value (PPV)	0.700392	0.673178	0.727952
SD Positive Predictive Value (PPV)	0.040198	0.046165	0.057747
Mean Negative Predictive Value (NPV)	0.843439	0.850940	0.781313
SD Negative Predictive Value (NPV)	0.034998	0.031749	0.044294

Table C.10
Global feature prediction results for New subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.856791	0.904920	0.919840
SD Accuracy	0.019584	0.012780	0.010216
Mean F1 Score	0.855728	0.908995	0.924333
SD F1 Score	0.019649	0.011571	0.009393
Mean Precision	0.866938	0.876939	0.879953
SD Precision	0.020933	0.018413	0.013673
Mean Recall	0.844894	0.943723	0.973617
SD Recall	0.020307	0.010125	0.011700
Mean Sensitivity	0.868817	0.865699	0.865484
SD Sensitivity	0.021398	0.022733	0.017438
Mean Specificity	0.844894	0.943723	0.973617
SD Specificity	0.020307	0.010125	0.011700
Mean Positive Predictive Value (PPV)	0.866938	0.876939	0.879953
SD Positive Predictive Value (PPV)	0.020933	0.018413	0.013673
Mean Negative Predictive Value (NPV)	0.847188	0.938402	0.970279
SD Negative Predictive Value (NPV)	0.019566	0.010567	0.012685

Table C.11
Global feature prediction results for PFM vs Ask subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.844624	0.850121	0.886028
SD Accuracy	0.039900	0.039865	0.038112
Mean F1 Score	0.884763	0.878343	0.912165
SD F1 Score	0.028106	0.032561	0.029950
Mean Precision	0.846590	0.915432	0.898462
SD Precision	0.045386	0.048242	0.040680
Mean Recall	0.929545	0.846364	0.927273
SD Recall	0.042336	0.043016	0.030829
Mean Sensitivity	0.679384	0.849499	0.810676
SD Sensitivity	0.155103	0.123040	0.082326
Mean Specificity	0.929545	0.846364	0.927273
SD Specificity	0.042336	0.043016	0.030829
Mean Positive Predictive Value (PPV)	0.846590	0.915432	0.898462
SD Positive Predictive Value (PPV)	0.045386	0.048242	0.040680
Mean Negative Predictive Value (NPV)	NaN	0.751909	0.863737
SD Negative Predictive Value (NPV)	NaN	0.088869	0.049020

Table C.12
Global feature prediction results for PFM vs New subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	0.684572	0.664609	0.908104
SD Accuracy	0.063011	0.067679	0.036129
Mean F1 Score	0.765947	0.761987	0.918894
SD F1 Score	0.034403	0.035630	0.031361
Mean Precision	0.656142	0.633654	0.900343
SD Precision	0.049897	0.052448	0.042637
Mean Recall	0.925909	0.961364	0.939545
SD Recall	0.048810	0.033633	0.035241
Mean Sensitivity	0.376018	0.290885	0.868509
SD Sensitivity	0.187424	0.182232	0.065086
Mean Specificity	0.925909	0.961364	0.939545
SD Specificity	0.048810	0.033633	0.035241
Mean Positive Predictive Value (PPV)	0.656142	0.633654	0.900343
SD Positive Predictive Value (PPV)	0.049897	0.052448	0.042637
Mean Negative Predictive Value (NPV)	NaN	NaN	0.921435
SD Negative Predictive Value (NPV)	NaN	NaN	0.043658

Table C.13
Global feature prediction results for Ask vs New subreddits.

Metric	BLR	SVM	RFC
Mean Accuracy	9.195431e−01	0.948298	0.937001
SD Accuracy	2.425640e−02	0.020217	0.022099
Mean F1 Score	9.318403e−01	0.954787	0.943523
SD F1 Score	1.850276e−02	0.017036	0.019267
Mean Precision	9.010704e−01	0.949901	0.958464
SD Precision	3.441313e−02	0.030859	0.027908
Mean Recall	9.655172e−01	0.960345	0.929655
SD Recall	1.110223e−16	0.014113	0.023851
Mean Sensitivity	8.588389e−01	0.932327	0.946216
SD Sensitivity	5.798061e−02	0.043500	0.037929
Mean Specificity	9.655172e−01	0.960345	0.929655
SD Specificity	1.110223e−16	0.014113	0.023851
Mean Positive Predictive Value (PPV)	9.010704e−01	0.949901	0.958464
SD Positive Predictive Value (PPV)	3.441313e−02	0.030859	0.027908
Mean Negative Predictive Value (NPV)	9.499837e−01	0.947613	0.912277
SD Negative Predictive Value (NPV)	5.837108e−03	0.018189	0.028947

BlackOnlyFun, AdoptmyACNLvillagers, Bugsnax, CODWarzone, exfds, DirtySocialMedia

- **Random**: bodybuilding, drunk, summonerswar, googlehome, rails, oscp, bravefrontier, Doom, Buddhism, germany, WeWantPlates, lepin, harrystyles, 3amjokes, hearthstone, Bedbugs, BravoRealHousewives, PUBGMobile, massachusetts, csgomarketforum, Sat, NoPoo, Vinesauce, talesfromcallcenters, shameless, datascience, NelkFilmz, Citrix, btc, Swimming, Buffalo, VitaPiracy, statistics, uofm, Shitty_Car_Mods, Cloud9, NFL_Draft, 8bitdo, Warhammer, outwardgame, minnesota, BoneAppleTea, greentext, zoombackgrounds, MordekaiserMains, girlsfrontline, TNOmod, JRPG, SuperMegaBaseball, M1Finance, botw, deadbydaylight, APUSH, funimation, canadacordcutters, queensuniversity, PHP, drums, mtgfinance, Miata, silenthill, TurkeyJerky, bangtan, TheMidnightGospel, manhwa, RATS, Choices, uruguay, Switzerland, rpghorrorstories, italy, findapath, 23andme, virginvschad, XboxSeriesX, zelda, InternetStars, JohnMayer, PiratedGTA, Nikon, Lexus, Soundbars, TheMonkeysPaw, stimuluscheck, lastimages, TrueCrime, SuperModelIndia, Machinists, Galaxy_S20, fixit, whichbike, IllegallySmolCats, SkyGame, Suomi, ElectricSkateboarding, starwarsrebels, nuzlocke, thedavidpakmanshow, TryingForABaby

Appendix B. General statistics

See Table B.1.

Appendix C. Prediction results

C.1. Local features

See Tables C.2–C.7.

C.2. Global features

See Tables C.8–C.13.

References

[1] Z. Barua, S. Barua, S. Aktar, N. Kabir, M. Li, Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation, Prog. Disaster Sci. (2020) 100119, http://dx.doi.org/10.1016/j.pdisas.2020.100119, URL: http://www.sciencedirect.com/science/article/pii/S2590061720300569.

[2] Y. Rodny-Gumede, Fake it till you make it: The role, impact and consequences of fake news, in: B. Mutsvairo, B. Karam (Eds.), Perspectives on Political Communication in Africa, Springer International Publishing, Cham, 2018, pp. 203–219, URL: https://doi.org/10.1007/978-3-319-62057-2_13.

[3] C. O'Connor, M. Murphy, Going viral: doctors must tackle fake news in the covid-19 pandemic, Bmj 24 (369) (2020) m1587.

[4] W. Ahmed, J. Vidal-Alaball, J. Downing, F.L. Seguí, Covid-19 and the 5G conspiracy theory: social network analysis of Twitter data, J. Med. Internet Res. 22 (5) (2020) e19458.

[5] D. Orso, N. Federici, R. Copetti, L. Vetrugno, T. Bove, Infodemic and the spread of fake news in the COVID-19-era, Eur. J. Emerg. Med. (2020).

[6] G.L. Ciampaglia, Fighting fake news: a role for computational social science in the fight against digital misinformation, J. Comput. Soc. Sci. 1 (1) (2018) 147–153.

[7] J. Ashford, L. Turner, R. Whitaker, A. Preece, D. Felmlee, D. Towsley, Understanding the signature of controversial Wikipedia articles through motifs in editor revision networks, in: Companion Proceedings of the 2019 World Wide Web Conference, 2019, pp. 1180–1187.

[8] G. Wu, M. Harrigan, P. Cunningham, Characterizing Wikipedia pages using edit network motif profiles, in: Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents, 2011, pp. 45–52.

[9] G. Wu, M. Harrigan, P. Cunningham, Classifying Wikipedia articles using network motif counts and ratios, in: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration, 2012, pp. 1–10.

[10] E. Guest, (Anti-)echo chamber participation: Examining contributor activity beyond the chamber, in: Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18, ACM Press, Copenhagen, Denmark, 2018, pp. 301–304, http://dx.doi.org/10.1145/3217804.3217933, URL: http://dl.acm.org/citation.cfm?doid=3217804.3217933.

[11] D.A. Scheufele, N.M. Krause, Science audiences, misinformation, and fake news, Proc. Natl. Acad. Sci. 116 (16) (2019) 7662–7669.

[12] P. Mena, D. Barbe, S. Chan-Olmsted, Misinformation on instagram: The impact of trusted endorsements on message credibility, Soc. Media+ Soc. 6 (2) (2020) 2056305120935102.

[13] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2) (2017) 211–236.

[14] J. Allen, B. Howland, M. Mobius, D. Rothschild, D.J. Watts, Evaluating the fake news problem at the scale of the information ecosystem, Sci. Adv. 6 (14) (2020) eaay3539, Publisher: American Association for the Advancement of Science.

[15] P. Törnberg, Echo chambers and viral misinformation: Modeling fake news as complex contagion, PLoS One 13 (9) (2018) e0203958, Publisher: Public Library of Science San Francisco, CA USA.

[16] H. Webb, M. Jirotka, B.C. Stahl, W. Housley, A. Edwards, M. Williams, R. Procter, O. Rana, P. Burnap, Digital wildfires: hyper-connectivity, havoc and a global ethos to govern social media, ACM SIGCAS Comput. Soc. 45 (3) (2016) 193–201, Publisher: ACM New York, NY, USA.

[17] R. Nithyanand, B. Schaffner, P. Gill, Online political discourse in the trump era, 2017, arXiv:1711.05303 [Cs], URL: http://arxiv.org/abs/1711.05303.

[18] M. Innes, Techniques of disinformation: Constructing and communicating "soft facts" after terrorism, Br. J. Sociol. 71 (2) (2020) 284–299, http://dx.doi.org/10.1111/1468-4446.12735, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-4446.12735.

[19] D. Dobreva, D. Grinnell, M. Innes, Prophets and loss: How "soft facts" on social media influenced the brexit campaign and social reactions to the murder of Jo Cox MP, Policy Internet 12 (2) (2020) 144–164, http://dx.doi.org/10.1002/poi3.203, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/poi3.203, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/poi3.203.

[20] M. Innes, D. Dobreva, H. Innes, Disinformation and digital influencing after terrorism: spoofing, truthing and social proofing, Contemp. Soc. Sci. (2019) 1–15.

[21] L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga, Y. Wang, A first look at COVID-19 information and misinformation sharing on Twitter, 2020, arXiv:2003.13907 [Cs], URL: http://arxiv.org/abs/2003.13907.

[22] R. Kouzy, J. Abi Jaoude, A. Kraitem, M.B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E. Akl, K. Baddour, Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter, Cureus (2020) http://dx.doi.org/10.7759/cureus.7255.

[23] G.K. Shahi, A. Dirkson, T.A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, Online Soc. Netw. Media 22 (2021) 100104.

[24] Y. Sunmoo, M. Odlum, P. Broadwell, N. Davis, C. Hwayoug, D. Nanyi, M. Patrao, D. Schauer, M.E. Bales, C. Alcantara, Application of social network analysis of COVID-19 related tweets mentioning cannabis and opioids to gain insights for drug abuse research, Stud. Health Technol. Inform. 272 (2020) 5.

[25] L.E. Young, E. Sidnam-Mauch, M. Twyman, L. Wang, J.J. Xu, M. Sargent, T.W. Valente, E. Ferrara, J. Fulk, P. Monge, Disrupting the COVID-19 misinfodemic with network interventions: Network solutions for network problems, Am J Public Health 111 (3) (2021) 514–519.

[26] M. Cheng, C. Yin, S. Nazarian, P. Bogdan, Deciphering the laws of social network-transcendent COVID-19 misinformation dynamics and implications for combating misinformation phenomena, Sci. Rep. 11 (1) (2021) 1–14.

[27] G. Pennycook, J. McPhetres, Y. Zhang, J.G. Lu, D.G. Rand, Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention, pp. 11.

[28] J.Y. Cuan-Baltazar, M.J. Muñoz Perez, C. Robledo-Vega, M.F. Pérez-Zepeda, E. Soto-Vega, Misinformation of COVID-19 on the internet: Infodemiology study, JMIR Publ. Health Surveill. 6 (2) (2020) e18444, http://dx.doi.org/10.2196/18444, URL: https://publichealth.jmir.org/2020/2/e18444/, Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada.

[29] J.J. Van Bavel, K. Baicker, P.S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M.J. Crockett, A.J. Crum, K.M. Douglas, J.N. Druckman, et al., Using social and behavioural science to support COVID-19 pandemic response, Nat. Hum. Behav. (2020) 1–12.

[30] K. Hunt, B. Wang, J. Zhuang, Misinformation debunking and cross-platform information sharing through Twitter during hurricanes harvey and irma: a case study on shelters and id checks, Nat. Hazards 103 (2020) 861–883.

[31] K.-C. Yang, F. Pierri, P.-M. Hui, D. Axelrod, C. Torres-Lugo, J. Bryden, F. Menczer, The COVID-19 infodemic: Twitter versus facebook, Big Data Soc. 8 (1) (2021) 20539517211013861.

[32] T. Wilson, K. Starbird, Cross-platform disinformation campaigns: lessons learned and next steps, Harvard Kennedy School Misinform. Rev. 1 (1) (2020).

[33] S. Kumar, R. West, J. Leskovec, Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes, in: Proceedings of the 25th International Conference on World Wide Web - WWW '16, ACM Press, Montral, Québec, Canada, 2016, pp. 591–602, http://dx.doi.org/10.1145/2872427.2883085, URL: http://dl.acm.org/citation.cfm?doid=2872427.2883085.

[34] S. Zannettou, T. Caulfield, E. De Cristofaro, M. Sirivianos, G. Stringhini, J. Blackburn, Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web, in: Companion Proceedings of the 2019 World Wide Web Conference, ACM, San Francisco USA, 2019, pp. 218–226, http://dx.doi.org/10.1145/3308560.3316495, URL: https://dl.acm.org/doi/10.1145/3308560.3316495.

[35] J.S. Zhang, B.C. Keegan, Q. Lv, C. Tan, A tale of two communities: Characterizing reddit response to COVID-19 through /r/China_Flu and /r/coronavirus, 2020, arXiv:2006.04816 [Cs] URL: http://arxiv.org/abs/2006.04816.

[36] X. Song, J. Petrak, Y. Jiang, I. Singh, D. Maynard, K. Bontcheva, Classification aware neural topic model and its application on a new COVID-19 disinformation corpus, 2020, arXiv:2006.03354 [Cs, Stat], URL: http://arxiv.org/abs/2006.03354.

[37] N. Gozzi, M. Tizzani, M. Starnini, F. Ciulla, D. Paolotti, A. Panisson, N. Perra, Collective response to the media coverage of COVID-19 pandemic on reddit and wikipedia, 2020, Physics, URL: http://arxiv.org/abs/2006.06446.

[38] S.A. Cook, The complexity of theorem-proving procedures, in: Proceedings of the Third Annual ACM Symposium on Theory of Computing, 1971, pp. 151–158.

[39] M. Bressan, F. Chierichetti, R. Kumar, S. Leucci, A. Panconesi, Counting graphlets: Space vs time, in: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, in: WSDM '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 557–566, URL: https://doi.org/10.1145/3018661.3018732.

[40] T. Hočevar, J. Demšar, A combinatorial approach to graphlet counting, Bioinformatics 30 (4) (2014) 559–565, arXiv:https://academic.oup.com/bioinformatics/article-pdf/30/4/559/17345305/btt717.pdf, URL: https://doi.org/10.1093/bioinformatics/btt717.

[41] J. Janssen, M. Hurshman, N. Kalyaniwalla, Model selection for social networks using graphlets, Internet Math. 8 (4) (2012) 338–363.

[42] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.

[43] S.S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of escherichia coli, Nature Genet. 31 (1) (2002) 64–68.

[44] G. Robins, M. Alexander, Small worlds among interlocking directors: Network structure and distance in bipartite graphs, Comput. Math. Organiz. Theory 10 (1) (2004) 69–94.

[45] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284, http://dx.doi.org/10.1109/TKDE.2008.239.

[46] K. Tu, J. Li, D. Towsley, D. Braines, L.D. Turner, Gl2vec: Learning feature representation using graphlets for directed networks, in: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2019, pp. 216–221, http://dx.doi.org/10.1145/3341161.3342908.